EXPERIMENTAL CHARACTERISTICS OF DIFFUSE X-RAY SCATTERING IN LYSOZYME CRYSTALS

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

Veronica Drabek Pillar

August 2019

EXPERIMENTAL CHARACTERISTICS OF DIFFUSE X-RAY SCATTERING IN LYSOZYME CRYSTALS

Veronica Drabek Pillar, Ph.D.

Cornell University 2019

X-ray crystallography, a popular and powerful technique for determining the atomic structures of proteins and other macromolecules, is traditionally limited to reporting static structures, whereas proteins *in vivo* are constantly shifting between conformations. Such conformational dynamics give rise to structured diffuse scattering, which is observable in protein crystallography experiments. Because the measurement of diffuse scattering involves numerous experimental challenges beyond those inherent to traditional crystallography, some basic questions about the behavior of diffuse intensities remain unanswered: how reproducible is the diffuse intensity between ostensibly identical crystals? What happens to diffuse intensity when the protein binds to an inhibitor molecule, when the crystal is cryocooled, or when it suffers radiation damage? Here, we present the results of diffuse scattering experiments on 23 different lysozyme crystals under a variety of experimental conditions. It is shown that diffuse intensity is highly consistent between ostensibly identical room temperature crystals. Binding to an inhibitor reduces the diffuse intensity non-uniformly, suggesting a change in the molecular motions that produce diffuse scattering. Cryocooling drastically changes the diffuse intensity pattern, indicating that cryocooled crystals are poor representatives of the correlated displacements in room temperature crystals. Radiation dose effects in a single image's diffuse scattering are also discussed.

BIOGRAPHICAL SKETCH

Veronica was born in Washington, D.C. in 1988. After earning her A.B. at Princeton University, she began her Ph.D. program at Cornell in August 2010, where she remains after some twists and turns.

To Betty Pillar (1921–2019), who would have been proud.

ACKNOWLEDGEMENTS

Many, many people have been instrumental in supporting me to finish this work. Thanks first to Sol Gruner, my advisor, for the project idea, for the connections to other scientists in the field, for your patience, and for believing in me when I didn't believe in myself. Thanks to my committee members Michelle Wang and Jim Sethna for your academic and emotional support over the years.

Thanks to Jeney Wierman and Dave Schuller for taking some of this data while I slept during a multi-day CHESS run and for teaching me how to do crystallography experiments. Thanks to Marian Szebenyi for editing data collection software and guiding me through various hurdles in data processing. Thanks to Irina Kriksunov and Qingqiu Huang for their help at MacCHESS, and to Mike Cook, Bill Miller, and Scott Smith for settingup beamlines and fixing things I broke.

Thanks to Mark Tate for reminding me that my work is interesting and for helpful feedback over the years. Thanks to Kate Shanks for helping me through some breakdowns, and to Joel Weiss, Hugh Phillip, Prafull Purohit, Marty Novak, Divya Gadkari, Darol Chamberlain, and the rest of the Gruner Group over the years for their support.

Thanks to Rob Thorne, Jesse Hopkins, and Hakan Atakisi for their help figuring out how to improve our protein crystal growth. Thanks to Tiit Lukk and Chae Un Kim for teaching me the basics of protein model building.

Thanks to Michael Wall for taking a full day in Santa Fe to teach me *Lunus*, and for continued support over the years. Thanks to Eddie Snell for taking me on my first diffuse data collection run, suggesting the inhibitor, and maintaining interest in the project. Thanks to Nozomi Ando and Steve Meisburger for doing more data collection with me and many helpful conversations and inspira-

tion about the frontiers of diffuse scattering research. Thanks to James Holton, Elspeth Garman, Arwen Pearson, Mitch Miller, David Case, Rahel Woldeyes, Artem Lyubimov, Daniel Keedy, Andrew Van Benschoten, Bob Fischetti, and Bob Sweet for helpful and memorable interactions in the crystallography world.

Thanks to Kacey Acquilano for being my rock in the Physics department and to Deb Hatfield for her continued support. Thanks to Jeevak Parpia, Lawrence Gibbons, and Maxim Perelstein for their much-needed support as Directors of Graduate Studies.

Thanks to Andrew Alexander for partnership in climbing, getting work done, and emotional support. Thanks to Robin Bjorkquist for struggling through the end of grad school with me and reminders that our worlds are bigger than this. Thanks to Caitlin McDonold for accountability partnership, loving friendship, and needed critical feedback. Thanks to Kathryn McGill for being by my side through the entire long road of graduate school and for all the love, tough and otherwise. Thanks to Scott Messick for last-minute talk feedback and emotional support, compassionate perspective shifting, and fun times away from work. Thanks to Jenny Wurster for the ad hoc office, relentless and necessary check-ins, and continued faith in me.

Thanks to Bob De Alba, Nathan Mirman, Jen Grab, Dave MacNeill, Brian Leahy, and Kathryn McGill again for being the best first-year cohort I could hope for. Thanks to Ethan Susca for deep friendship, a year of morning routines, and assistance with MATLAB. Thanks to Mike Saelim, Mike Bono, Erin Wissink, Jared Maxson, Matt Bierbaum, Sam Posen, Zach Lamberty, Yariv Yanay, Jim Shanks, Jenni Drewes, Eva Smith, Katherine Quinn, Michelle Kelley, Amanda Ozckowski, Andrea Bruns, Rachel Steinhardt, Meghan Witherow, Emily Cheng, Susie Kahn, C Cowman-Eggert, Emily Laing, Allison Truhlar, Myriam Figueroa,

Kate Cardona, Kim Knight, Sarah D'Antonio, Eve Vavagiakis, Joanna Fisher, Meredith Silberstein, and Kass Urban-Mead for their support and friendship here in Ithaca. Thanks to Adrian Ruskin, Annie Martin, Chloe Shea, Megan Zagrobelny, Christina Powell, Anna Lamine, Sajid Mehmood, Amy Anton, Will Berkeley, Chris Yarnell, Jacob Lewellen, Katie Hsih, Ben Scandella, Yuchen Zhang, Andrew Weintraub, Lucas Pillar, and Katie Walsh for their support and friendship from afar. Thanks to the many other friends and colleagues not listed by name who have been important and positive parts of my life through this whole journey.

Thanks to the Long Trail and the White Mountains for refreshing my soul so I could come back and finish this degree.

Finally, thanks to my parents, Paul Pillar and Cynthia Drabek, for their unwavering love and support through my entire life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

For over half a century, x-ray crystallography has been used to determine the structure, or spatial arrangement of atoms, of specific protein molecules. Though structure determination by other methods such as nuclear magnetic resonance (NMR) spectroscopy or cryogenic electron microscopy (cryo-EM) has become increasingly popular in recent years, as of this writing most new protein structures deposited in the Protein Data Bank [Berman et al., 2000] are still solved using x-ray diffraction.

Traditional x-ray crystallography involves collecting series of diffraction images from a protein crystal and isolating the Bragg diffraction peaks, which show up as small bright spots, from these images. The locations and intensities of these diffraction peaks, combined with information about the experimental setup and often some *a priori* knowledge about the protein of interest, can be analyzed to produce an atomic-resolution picture of a single protein molecule.

Often, x-ray images of protein crystals will also contain weak, broad smudges between and around the sharp Bragg peaks. These smudges, termed diffuse intensity, are typically considered part of the background and thrown out during data processing. However, since much of this diffuse intensity originates from the protein molecules themselves, it contains information about the protein molecules that can be useful for learning about their static and dynamic structure.

The remainder of this chapter is divided into three parts. Section 1.1 provides a mathematical explanation of protein-derived diffuse scatter and dis-

cusses its characteristics and what we can learn from it. Section 1.2 summarizes previous research in this field. Finally, section 1.3 explains the basic open questions about protein-derived diffuse scattering to be addressed in this dissertation.

1.1 Theory

The relationship between diffuse x-ray scattering and correlated atomic motions has long been theoretically understood and applied in small molecule crystallography [Guinier, 1963; Welberry, 2004] while remaining less clear for macromolecules such as proteins. Here, the basics of the underlying theory, without regard to the size of the diffracting molecule, are briefly reviewed.

1.1.1 X-ray diffraction in ideal protein crystals

Consider an object with electron density $\rho(\vec{x})$ in three-dimensional space that remains static over time. Diffraction theory states that when this object is illuminated with a beam of perfectly collimated, monochromatic photons, the pattern of the scattered photons can be described as

$$\vec{F}(\vec{q}) = \mathrm{FT}(\rho) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \rho(\vec{x}) e^{2\pi i \vec{q}\cdot\vec{x}} \, dx \, dy \, dz \tag{1.1}$$

where $\vec{q}$ is the scattering vector, i.e., the difference between the scattered and incident wave vectors, and FT denotes the Fourier transform.[1] The scattered wave $\vec{F}$ at position $\vec{q}$ is called the structure factor of the original object because it depends on the object's structure, represented by the electron density $\rho(\vec{x})$.[2] Structure factors are commonly used in protein crystallography to indirectly describe

the structure of either a single protein molecule or a set of protein molecules that, together, make up one unit cell of a crystal.

Now suppose that a large number of identical objects, each with the same electron density $\rho(\vec{x})$ and, therefore, the same structure factors $\vec{F}(\vec{q})$, are arranged in a perfect three-dimensional lattice structure with lattice translation vectors $\vec{a}$, $\vec{b}$, and $\vec{c}$, i.e., as a crystal. The scattering from this crystal is now the sum of the scattering from each unit cell (assume for now that each unit cell contains exactly one aforementioned object), which turns out to be

$$\vec{K}(\vec{q}) = \vec{F}(\vec{q}) \times \sum_{j=0}^{N_a} e^{2\pi i j \vec{q}\cdot\vec{a}} \times \sum_{k=0}^{N_b} e^{2\pi i k \vec{q}\cdot\vec{b}} \times \sum_{m=0}^{N_c} e^{2\pi i m \vec{q}\cdot\vec{c}} \tag{1.2}$$

where N_a, N_b, and N_c are the numbers of unit cells in the $\vec{a}$, $\vec{b}$, and $\vec{c}$ directions, respectively. Each of the summation terms adds up the scattering from all unit cells along a single dimension.

If N_a is sufficiently large, as in most protein crystals (where $N \approx 2000 - 200{,}000$) the corresponding summation term will be effectively zero unless $\vec{q} \cdot \vec{a}$ is an integer. The same is true for the N_b and N_c summation terms, meaning that $K(\vec{q})$ is nonzero only for specific, discrete values of $\vec{q}$, termed Miller indices.

What does this look like? In a crystallography experiment, the detector records scattered photon intensities, which are proportional to $|\vec{K}(\vec{q})|^2$, along a thin two-dimensional slice through the three-dimensional space defined by $\vec{q}$ (called reciprocal space). In a perfect crystal, x-rays will only be diffracted to certain localized spots, so the detector will record a set of small bright spots of varying intensity, with zero intensity in the spaces between those spots. The pat-

[1] For a more complete introduction to x-ray scattering by a crystal, refer to [Drenth, 1999] or similar.

[2] X-rays incident upon atoms scatter almost exclusively off of electrons, not nuclei, because scattering intensity is inversely proportional to m^4 where m is the scatterer's mass.

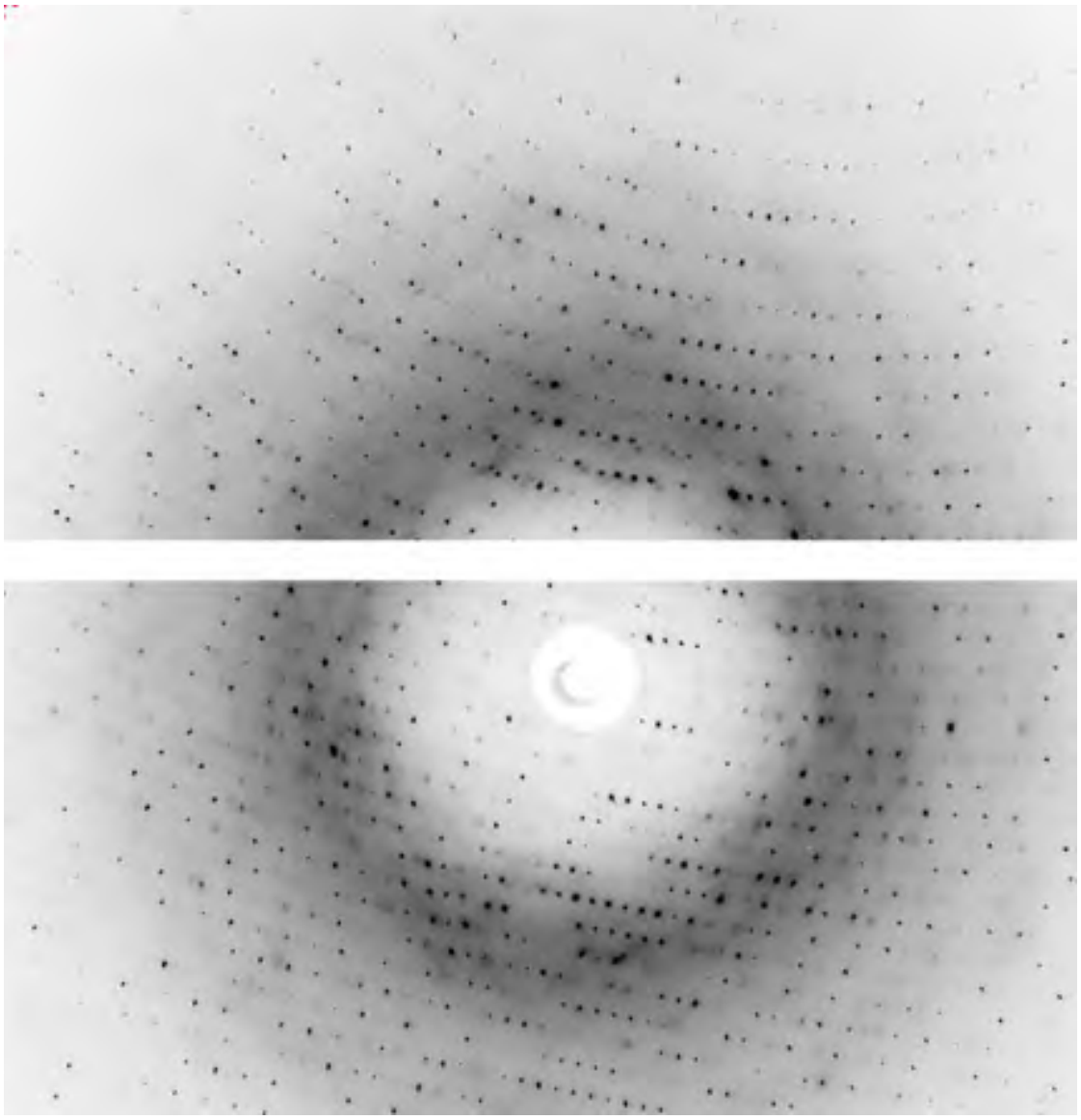


Figure 1.1: Diffraction pattern from a typical protein crystal. Note the bright (black) spots and visible diffuse features between them, indicating correlated displacements within a regular crystalline structure.

tern of the spots depends on the specific crystal geometry and the relative orientation of crystal, x-ray beam, and detector. Typically, a long series of images will be recorded, each with the crystal at a different orientation and therefore each with a different pattern of spots.

1.1.2 How does disorder change the diffraction pattern?

The idealized picture above is very neat, but real protein crystals do not consist of perfectly identical molecules arranged in a perfectly ordered pattern. Rather, the configurations and positions of molecules can vary in numerous ways due to several sources of disorder; a typical diffraction pattern from such a dynamic protein crystal is shown in Figure 1.1. The following mathematical treatment of this general phenomenon is taken from [Clarage and Phillips, Jr., 1997].

If ρ is now the electron density of an entire macromolecular crystals, disorder in that crystal means that ρ is changing in time, so FT(ρ) is also changing in time. A typical x-ray diffraction pattern is recorded much more slowly than the motion of a molecule, so the observed intensity is $\langle|\mathrm{FT}(\rho)|^2\rangle$. The Bragg peaks—that is, the intensities recorded at the specific spots corresponding to integer Miller indices—only capture the Fourier transform of the mean electron density, or $|\mathrm{FT}\langle\rho\rangle|^2$. The diffuse intensity is all scattered intensity that isn't in Bragg peaks, or $\langle|\mathrm{FT}(\rho)|^2\rangle - |\mathrm{FT}\langle\rho\rangle|^2$. This can be rearranged to $\langle(\mathrm{FT}(\rho) - \langle\mathrm{FT}(\rho)\rangle)^2\rangle$, which is simply the variance in the Fourier transform of the electron density. In the perfect crystal example above, the variance was zero, but in a real crystal, variance will be nonzero and diffracted intensity will be found outside of the Bragg peaks.

Many types of disorder contribute to this variance. Clarage and Phillips (1997) modeled several types of disorder in a two-dimensional crystal and calculated the resulting diffuse scattering; specific examples are shown in Figure 1.2. Theirs are models of static disorder rather than dynamic disorder, but since a typical crystallography experiment averages over both space (illuminating about a billion molecules) and time (exposure times are seconds or tenths of seconds while domain motions occur on nanosecond timescales), the static

or dynamic quality of a particular disorder mode is irrelevant to the resulting diffraction pattern. Note that different types of disorder produce obviously different diffuse intensity patterns. Of course, a real crystal will contain multiple modes of disorder, resulting in a much more complex diffuse intensity pattern that cannot be easily directly interpreted.

The entire preceding discussion assumes that the x-ray beam only scatters off the protein molecules and nothing else. In practice, the beam will also scatter off of other things such as a sample holder and water present in the crystal, all of which scattering contributes to the total intensity recorded on the detector. These background contributions are discussed further in section 2.1.1.

1.1.3 What can diffuse intensity tell you that Bragg peaks can't?

As seen above, diffuse intensity carries information about correlated displacements in a crystalline sample. Bragg peak intensities will also change upon displacement from a perfect lattice, but different displacement patterns often result in the same Bragg intensity pattern and therefore cannot be distinguished by Bragg diffraction alone.

For example,[3] consider the simplified example of a molecule with two rigid arms, each of which can move up and down. Bragg diffraction from a crystal of these molecules can, for each arm, determine what fraction of the time the arm is up, down, or in some intermediate configuration. Suppose that each arm is continuously moving up and down, spending equal time in each extreme configuration. It is likely to be biologically relevant whether the arms

[3]This example was inspired by a talk given by Andrew Van Benschoten at the Diffraction Methods in Structural Biology Gordon Research Conference in Lewiston, ME in July 2014.

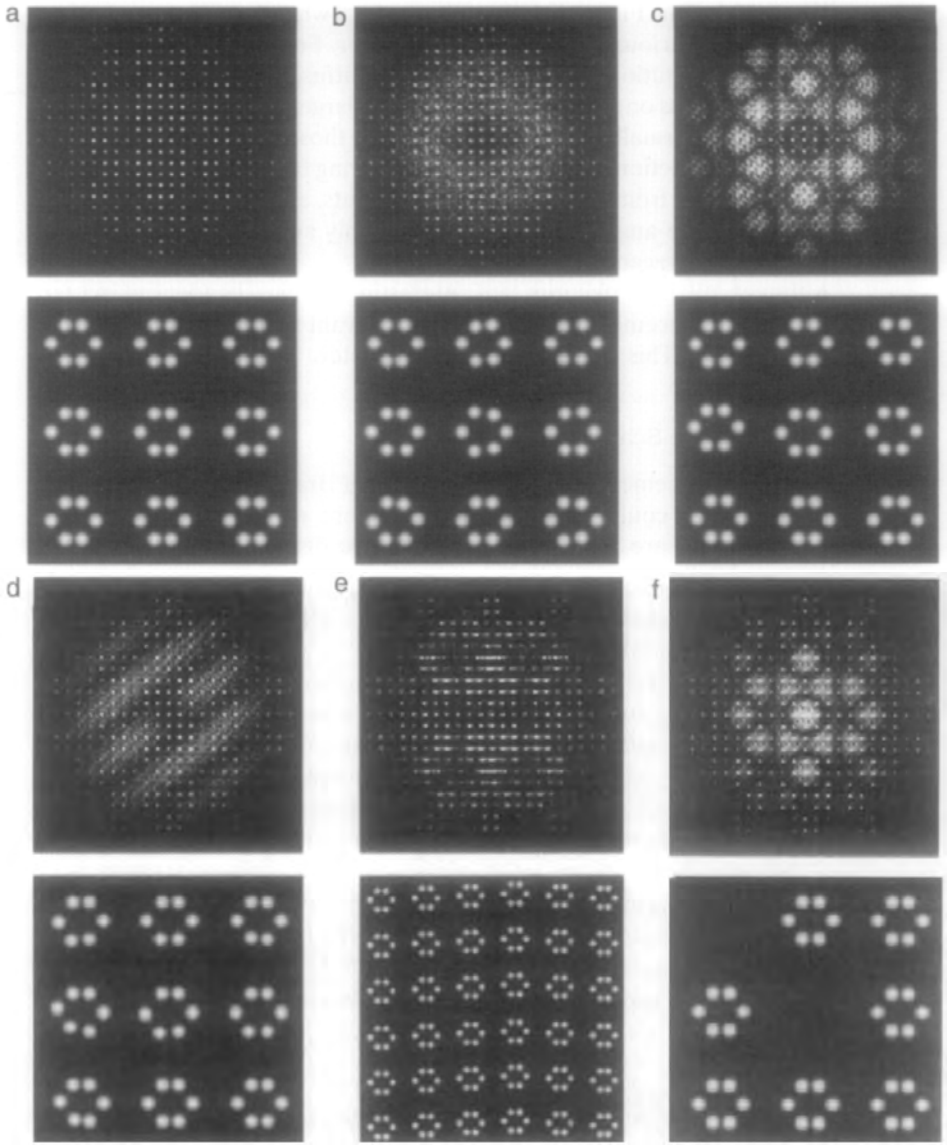


Figure 1.2: Simulated two-dimensional crystals with different patterns of disorder and the resulting calculated scattering intensities. (a) A perfect crystal comprising quasihexagonal "molecules", each made up of six "atoms." No diffuse scattering results. (b) Random perturbation of each atom, resulting in unstructured diffuse intensity. (c) Random perturbation of each molecule as a rigid body, resulting in structured diffuse intensity displaying the Fourier transform of a single molecule. (d) Random perturbation of the lower two atoms of each molecule as a rigid unit, mimicking domain motion in proteins. Structured diffuse features larger than the spacing between Bragg peaks result. (e) Vertical displacement of entire molecules as rigid bodies according to a transverse wave, resulting in smaller diffuse features. (f) Substitution disorder, resulting in diffuse intensity similar to but distinct from (c). Reprinted from *Methods in Enzymology*, Vol. 277, James B. Clarage and George N. Phillips, Jr., Analysis of diffuse scattering and relation to molecular motion, pages 407-432, copyright (1997), with permission from Elsevier.

move up and down such that they are in sync, perfectly out of sync, or uncorrelated. All three options have the same mean electron density, and thus will produce identical Bragg scattering patterns, but they have different variances in their electron density and thus will produce different diffuse scattering patterns. Diffuse intensity is therefore often necessary to distinguish between conformational and dynamic models of a molecule using crystallographic data; in fact, it has been shown that Bragg-based molecular model refinement consistently underestimates conformational heterogeneity [Kuzmanic et al., 2014].

In principle, one could directly calculate molecular motions of interest, in the form of electron density variance, from experimentally measured diffuse scattering patterns and knowledge of the average molecular structure. However, this is an extremely hard problem in practice, for reasons discussed below. The next section summaries progress to date on measuring and interpreting diffuse scattering patterns from proteins.

1.2 Previous work

The study of diffuse x-ray scattering in protein crystals began four decades ago and continues through significant improvements in relevant experimental technology. In recent years, a renewed interest has arisen in diffuse scattering as a tool to solve various types of structural biology problems, but with this new work comes an abundance of fundamental questions about how the diffuse signal behaves.

1.2.1 Early work

In 1980, Phillips et al. reported the first use, to this author's knowledge, of diffuse x-ray scattering to characterize protein motions [Phillips, Jr. et al., 1980]. Working with single images of tropomyosin, a coiled double α-helix protein, they applied a simple analytic model to diffuse streaks in one image to characterize the flexibility of the helix fibers. Subsequent studies used diffuse intensity to characterize displacement waves along tropomyosin filaments [Boylan and Phillips, Jr., 1986] and intermolecular rigid-body motions in orthorhombic lysozyme crystals [Doucet and Benoit, 1987].

In 1988, Caspar et al. published a rigorous analysis of the diffuse intensities from a diffraction image of insulin, explicitly separating the Bragg, diffuse, and background components of the signal [Caspar et al., 1988]; this was to become a standard component of future analysis methods. They used a liquid-like motions model, in which atomic displacements are correlated with a correlation strength that decreases with distance, to extract amplitudes and correlation lengths describing the molecular dynamics. Over the next decade, several more studies emerged showing that significant diffuse intensity arises in many protein crystals [Kolatkar et al., 1994; Faure et al., 1994]. Experimental highlights include noting that protein-derived diffuse intensity was comparable in cumulative magnitude to Bragg intensities [Glover et al., 1991] and that crystal form affected the intramolecular motions implied by the diffuse scattering pattern [Clarage et al., 1992]. Wall et al. capped off this progress by publishing the first analyses of three-dimensional diffuse data sets in *Staphylococcal nuclease* [Wall et al., 1997b] and calmodulin [Wall et al., 1997a], using a charge-coupled device (CCD) detector to collect the images.

At this point, the field was burgeoning; Clarage and Phillips wrote in a 1997 review that "structural biology is on the verge of a major transformation" and expressed the hope that diffuse scattering would rise to the level of rigor exemplified by Bragg crystallography. Yet it was also at this point that diffuse scattering progress stalled for approximately the next 15 years. One severe limitation of the aforementioned studies was data quality. Most diffuse scattering data had been collected on image plates, the reading of which takes too long for multi-image data set collection. CCDs are faster, but have their own issues, discussed in section 1.2.2. There was also limited understanding of what to do with the data–what models would fit the data, and what types of motions did it reveal? Though most of the diffuse scattering studies thus far included interpretations of the diffuse intensity according to one or more models, the field lacked consensus over which model fit the data best. Tetragonal hen egg white lysozyme alone spawned explicit disagreement over whether the diffuse intensity originated from rigid-body displacements or liquid-like motions [Pérez et al., 1996]. Without more and better data, these questions and others remained impossible to resolve.

1.2.2 Detector technology advancements

Charge-coupled device (CCD) detectors were the state of the art in the late 1990s, and while they enabled three-dimensional data collection via fast image readout, they limited diffuse intensity measurements in two significant ways. First, CCDs have a dark current—that is, some small current flowing to the sensor even with zero photons incident on the detector—which introduces a low level of background noise into all measurements. When measuring Bragg

diffraction peaks, the resulting detector noise is rarely a concern because the Bragg peaks are about four orders of magnitude stronger than the detector noise. Diffuse intensities are much weaker, however; often within an order of magnitude of the detector noise. Dark current therefore has to be carefully calibrated out to avoid increasing uncertainty in diffuse measurements; this was typically done successfully.

Second, when a standard phosphor-coupled CCD is hit with an x-ray, the light that arises in the detector material tends to spread out laterally before traveling to the actual pixels of the detector. Therefore, x-rays incident on a spot corresponding to a certain pixel are not guaranteed to be read as belonging to that pixel. This spreading of signal, and therefore detector output, in response to photons at a single point is described by the detector's point spread function (PSF). In the worst cases, PSFs are large enough that bright Bragg peaks bleed into one another; even when they don't, they still bleed into the spaces between each other, which is exactly where diffuse intensity measurements are taken. To make matters worse, a sufficiently bright Bragg peak will saturate one or more pixels with charge, leaving the rest of the charge to spill over entirely into neighboring pixels in a phenomenon known as blooming. Such saturation is difficult to avoid in diffuse scattering experiments because Bragg peaks are frequently 10^3-10^4 times brighter (per pixel) than the structured diffuse signal, so collecting significant photon counts in the diffuse regions typically requires overexposing Bragg peaks.

The development of pixel array detectors (PADs) in the early 2000s addressed both of these problems and offered a few additional advantages. As described by Broennimann et al. in [Broennimann et al., 2006], the PAD archi-

tecture is fundamentally different from that of a CCD, leading to significantly different properties. First, PADs have essentially zero readout noise, eliminating dark current from the recorded intensities. Second, PADs have PSFs less than the width of a single pixel, meaning that the PSF does not noticeably affect the recorded image. They also typically have large dynamic ranges (the Pilatus 6M can record photon counts from one to 10^6), so diffuse and Bragg intensities can be collected in the same image without overexposing Bragg peaks. This means that a well-diffracting crystal produces Bragg peaks confined to a small region on the detector, facilitating easier removal of Bragg scattering. That said, if the crystal itself is poorly ordered (high mosaicity) or the x-ray beam's range of wavelengths is non-negligible (high dispersion), the Bragg peaks will be spread across a greater region of reciprocal space and, therefore, a greater area on the detector, regardless of detector quality.

Finally, PADs are fast, with almost negligible readout speed compared to typical exposure times. This, combined with the high dynamic range enabling simultaneous collection of diffuse and Bragg intensities, reduces the total time necessary for data collection. For room temperature crystals, faster data collection reduces propagation of radiation damage through the crystal and improves data quality; this is explained further in section 1.3.3.

1.2.3 Recent developments

In the last several years, a community has reconvened around diffuse scattering research, now poised to take advantage of state-of-the-art pixel array detectors at existing synchotron beamlines. A 2013 meeting brought researchers together

to discuss challenges to be met by future diffuse scattering research [Wall et al., 2014]; these challenges included understanding temperature dependence of diffuse features, developing data integration methods as robust as those available for Bragg peak integration, interpreting diffuse scattering in terms of protein motions, and simply increasing the number of diffuse scattering data sets available. Much of the work done then and since has focused on comparing experimental diffuse intensities to various theoretical models. These studies have produced minimal agreement regarding how to model the conformational heterogeneity that contributes to diffuse scattering, but they have made significant progress in understanding the various possible sources of protein-derived diffuse scattering and in developing new methods of extracting diffuse intensities from raw data.

Attempts at theoretical modeling have mostly involved liquid-like motions models (LLM), the translation-libration-screw model of rigid-body displacements (TLS), and elastic normal mode models (ENM; also seen as NM or EN). In a mostly theoretical work, Riccardi et al. [Riccardi et al., 2010] compared all three types of models and found LLM and ENM most useful as contributors to diffuse scatter; they refer to Wall's 1997 published *S. nuclease* data set for experimental comparison. Van Benschoten et al. [Van Benschoten et al., 2016] collected two new room-temperature diffuse data sets, from trypsin and cyclophilin A (CypA), and again attempted to fit all three types of models to the data; they also found TLS to fit poorly compared to the other two. Peck et al. [Peck et al., 2018] used the same CypA data as well as data sets from two crystals at 100 K temperature to fit various theoretical models, finding that a LLM model including long-range correlations (that is, correlations between, rather than within, unit cells) best fit the diffuse data. However, when Wall et

al. [Wall et al., 2018] refined a new ENM model against the old *S. nuclease* data, refinement improved the fit to the data and significantly reduced the model's primary correlation lengths, indicating that perhaps internal motions dominate diffuse scattering after all.

Most recently, De Klijn et al. [De Klijn et al., 2019] used a supercell modelling approach involving more frequent sampling of reciprocal space than had been done previously, and they found that rigid-body motions contributed more to diffuse scattering than liquid-like motions, in contrast with previous work. They also distinguish clearly between four qualitatively different types of diffuse scattering, which by now have been observed repeatedly in several different proteins: weak spherical scattering caused by uncorrelated displacements, cloudy features caused by correlated displacements within a unit cell (and likely of greatest biological relevance), halos around Bragg peaks caused by intermolecular correlations, and long, thin streaks caused by long-range correlations across the crystal. Notably, though they used the same CypA data as two previous studies (in addition to a new lysozyme data set), their diffuse intensities "look completely different" than Peck et al.'s; they suggest that this is due to using different strategies for removing Bragg peaks from the images. This indicates the need for improved understanding of the sources of diffuse scattering and how one might distinguish between them in a single data set.

Progress has also been made in using molecular dynamics (MD) simulations to predict diffuse intensities [Wall et al., 2014]. As simulation capabilities improve, so has the fit to experimental data; Wall's recent MD results also point toward internal motions dominating diffuse intensity [Wall, 2018].

Interest has also arisen in using diffuse scattering to improve conventional

structural models, without a specific focus on dynamics. In a study of photosystem II, Ayyer et al. [Ayyer et al., 2016] used diffuse intensities extending to higher resolutions than the Bragg peaks to extend the resolution of their structural model. This type of use for diffuse scattering could potentially contribute to closing the "R-factor gap", which refers to the relatively high incongruency between macromolecular models and their corresponding diffraction data compared to that found in small-molecule crystallography [Wall et al., 2018].

Much of this recent work has been reviewed in detail by Meisburger et al. [Meisburger et al., 2017] and in overview by Wall et al. [Wall et al., 2018]. Both stress the need for many more diffuse data sets, as many diffuse scattering studies to date reuse the same small number of data sets. Wall et al. expect that regardless of what type of disorder truly underlies observable diffuse scattering, there are cases in which diffuse scattering will be important in improving models and better understanding biochemical mechanisms. This indicates the need for an improved understanding of the fundamental behavior of diffuse scattering and the effects of data collection and processing choices on diffuse maps, even in the absence of attempts at model fitting. Meisburger et al. list several questions yet to be answered regarding diffuse scattering fundamentals, many of which are to be addressed in this work (see section 1.3). In sum, amidst all the previous work comparing models to diffuse data, little if any work has been done comparing diffuse data to other diffuse data, leaving a lack of raw data and lack of understanding of the data's reliability and dependence on certain experimental parameters. This dissertation seeks to address both lacks.

1.3 Questions addressed in this dissertation

1.3.1 Reproducibility

Before using the information gathered from a single protein crystal to draw conclusions about that protein in general, one must have reason to believe that that crystal appropriately represents all crystals of that protein. Furthermore, one must understand to what extent even a representative protein crystal, or a set of protein crystals, exhibits the properties of interest of the protein in its setting of interest (usually, in a living cell). With respect to diffuse scattering, the second consideration (the relationship between a crystalline protein and the same protein *in vivo*) will be merely touched upon in this dissertation (see section 1.3.3), but the first consideration will be considered thoroughly.

In traditional x-ray crystallography, reproducibility of signal between crystals is rarely measured directly; any individual crystal yielding sufficiently high-quality data is considered representative. Though unit cell parameters are known to vary noticeably between otherwise identical cryocooled crystals [Farley et al., 2014], a rare study of the reproducibility of protein structures determined via crystallography showed excellent agreement by almost all metrics, except in certain flexible regions of the protein [Liebschner et al., 2013]. Given that flexible regions are likely responsible for much of the structured diffuse scattering that originates from correlated motions within molecules, it is unclear what implications these observations have for the reproducibility of diffuse signal.

The reproducibility of the diffuse signal has yet to be quantitatively studied,

to the best of the author's knowledge. Doucet et al. noted promisingly in 1987 that fresh and old (i.e., previously irradiated) crystals display qualitatively the same diffuse streak pattern, which implies significant consistency across crystals [Doucet and Benoit, 1987]. However, they focused on thinner features corresponding to intermolecular displacements (similar to those in Figure 1.2e), and it does not necessarily follow that features corresponding to intramolecular correlations would be similarly consistent. The first aim of this dissertation is, therefore, to evaluate the reproducibility of diffuse intensity across ostensibly identical protein crystals.

1.3.2 Response to inhibitor binding

As discussed in section 1.1, the motion of individual domains or residues in protein molecules results in observable diffuse intensity. Consider the active site of an enzyme that can bind to other molecules. Typically, the residues around the active site can be found in at least two different configurations: a more open configuration that allows a ligand to find its way into the active site, and a more closed configuration in which the ligand, if present, could be bound to the enzyme. Without a ligand present, the enzyme's active site could potentially be found in either of these configurations or any number of intermediate or alternate positions. When bound to a ligand, however, the enzyme's active site will necessarily be found in the closed position. It is hypothesized that the reduction in available configurations upon ligand binding will reduce the magnitude of diffuse intensity, but this has yet to be tested; evaluating this hypothesis is the second aim of this dissertation.

1.3.3 Effect of cryocooling

The very x-ray radiation used to probe a protein crystal's structure also breaks down that structure. Photons can damage bonds they impact directly and create free radicals that diffuse through the crystal, causing further damage. With enough irradiation, a protein crystal's structure will break down such that it is no longer usable for diffraction experiments.

The practice of flash-cooling crystals in liquid nitrogen has become common practice among crystallographers in the last twenty years. Cryocooling's primary advantage is that it severely limits free radical diffusion, thereby slowing down radiation damage and extending the effective lifetime of crystals in an x-ray beam by a factor of approximately 70 on average [Garman, 2010].

Though cryocooling has been immensely valuable in enabling the collection of more and better crystallographic data from thousands of crystals, concerns remain about the effect of cryocooling on crystal structures. The biologically relevant conformations of any given protein are those undertaken by the protein *in vivo*, while it is performing its function in an organism. Crystallizing a protein at room temperature already constrains the possible conformations, and cryocooling is expected to constrain them further by removing most available thermal energy. The structure of ribonuclease-A, at least, has been shown to depend on temperature [Tilton et al., 1992].

For diffuse scattering, cryocooling is likely to have an even stronger effect. If all the molecules in a protein crystal are free to individually shift among an ensemble of conformations (a common model for thinking about conformational heterogeneity), then the motions between conformations will be re-

flected in measured diffuse intensity. Upon cryocooling, the molecules will lose the energy necessary to switch between conformations and the energy landscape of configurations will change; in other words, the motions will be "frozen out". It is theoretically possible for the room-temperature diffuse scattering pattern to be preserved if the crystal cools faster than the molecules are able to re-equilibrate into a new distribution of configurations, in which case each molecule will be trapped in whatever configuration it happened to be in when the crystal was plunged into liquid nitrogen. The original ensemble of configurations will then be sampled only over space by different molecules, not over time by any individual molecule, but this will not change the x-ray diffraction pattern. However, it is predicted that crystalline protein molecules equilibrate faster than they cool, in which case the molecules will settle into a different conformational ensemble, with a different diffuse scattering pattern, before coming to rest at their cryocooled temperature. A theoretical model suggests that this is in fact the case, with the added complication that different motions are quenched at different temperatures, such that the resulting diffraction pattern may reflect a composite configuration that would not exist in equilibrium at any temperature [Halle, 2004]. Some experimental evidence for this prediction exists: the model system dihydrofolate reductase exhibits conformations when cryocooled that are not seen at room temperature [Keedy et al., 2014], and cryocooling produces surprising effects on ligand binding patterns in cytochrome *c* peroxidase that suggests the relative energies of its conformations depend on temperature [Fischer et al., 2015]. A meta-analysis of data from 30 different proteins revealed consistent changes in conformational ensembles upon cryocooling [Fraser et al., 2011].

If cryocooling changes protein conformational dynamics, and diffuse scat-

tering reveals said conformational dynamics, then cryocooling is expected to change diffuse scattering, but this has yet to be tested. Hence, investigating the impact of cryocooling on the diffuse scattering from a protein crystal is the final aim of this work.

CHAPTER 2

EXPERIMENTAL METHODS

To address the aforementioned questions, several sets of crystallographic x-ray data were collected from protein crystals under various conditions. All data collection was performed at the Cornell High Energy Synchrotron Source (CHESS) on beamlines run by MacCHESS (Macromolecular x-ray science at the Cornell High Energy Synchrotron Source). This chapter details the experimental procedures used, the reasoning behind certain decisions, and some recommendations for future data collection. Because the data presented here was collected over the course of two separate visits to CHESS, during which the experimental procedures were continually revised, the methods of each visit will be described separately with each visit referred to by its month and year.

2.1 Challenges particular to diffuse data collection

2.1.1 Background minimization

Protein-derived diffuse scattering unfortunately qualitatively resembles background scattering off of other items in the beam path, most notably the materials used to mount the crystal on the beamline. Therefore, while background scatter may be a minor concern in traditional crystallography experiments (in which a local background may fairly easily be subtracted from any Bragg peak), for diffuse scattering experiments it is a significant concern. To reasonably accurately extract protein-derived diffuse scattering from diffraction images, other background scattering must be minimized during the experiment and/or sub-

tracted after the fact with some knowledge of what the background scattering looks like. Either way, this requires attention during data collection.

We chose crystal mounting materials and beamstop holders, both of which must lie in the path of both the direct beam and the scattered x-rays, with an eye toward minimizing their influence on the structured diffuse scattering. Such influence can be absorption of protein-derived diffuse scattering by materials downstream of the crystal and/or extra scattering directly off of the mounting materials themselves. Scattering off the mounting materials can have isotropic and anisotropic components; since we chose to focus on the anisotropic (structured) component of the protein-derived diffuse scatter, isotropic scattering off of mounting materials was less of a concern, but it could still overwhelm the desired signal if its magnitude were too great. Details of the choices of mounting materials and beamstop holder are discussed in the CHESS visit-specific sections below.

2.1.2 Room temperature crystallography

As discussed in Section 1.3.3, the modern standard practice of cryocooling protein crystals is likely to significantly alter the conformational dynamics that contribute to diffuse intensity. Most of our data was therefore taken at room temperature. While room temperature crystallography likely better approximates the *in vivo* conformational landscape of the crystallized enzyme, it leaves the crystal open to two significant sources of damage: dehydration and accelerated radiation damage.

To avoid crystal dehydration, all room temperature crystals were mounted

inside some kind of capillary which was then sealed at both ends. Details are provided in the sections below. In general, this involves more material in the beam path than does mounting cryocooled crystals (which can be left open to the cold nitrogen stream without dehydrating), adding to the background scatter. Sometimes the seal would fail mid-experiment, resulting in a suddenly messy diffraction pattern and visible cracked crystal; care and perseverance are required to minimize these occurrences.

From the moment a room temperature protein crystal is exposed to x-rays, radiation damage and its propagation are unavoidable. Damage worsens with x-ray dose and with time since initial x-ray exposure. To mitigate radiation damage in room temperature crystals, the crystal position was shifted during data collection, usually perpendicularly to the beam path, so that different areas of the crystal were exposed at different times. This limited the total dose received by any one spot on the crystal. At any one exposure location, care was taken to collect data in one continuous sweep (within the limitations of the detector and software) to minimize elapsed time between initial and final x-ray exposure. Again, details of the data collection strategy are provided in the sections below.

2.1.3 Low intensity relative to Bragg peaks

As discussed in section 1.2.2, diffuse intensity is three or four orders of magnitude weaker than Bragg reflections, which can lead to a host of problems for diffuse scattering depending on the detector properties. While PAD architecture mostly avoids the side effects of this disparity in intensity, such as blooming, no

detector can remove the disparity itself. In a diffuse scattering experiment, it is necessary to accurately measure Bragg intensities in addition to diffuse intensities to reconstruct the average static molecular structure. This can be done with a single set of images, all with the same exposure time, if the detector has a sufficiently high dynamic range.

Some of the data used in this work was taken with a Pilatus 6M detector [Broennimann et al., 2006], whose dynamic range of 10^6 can capture diffuse and Bragg scattering simultaneously with good statistics for each. The February 2014 data was taken with a beta version of the Eiger detector [Dinapoli et al., 2011], which was at the time configured to a dynamic range of only 4×10^3. Therefore, two sets of images with different exposure times had to be collected from each of the February 2014 crystals; the specific procedure is described below.

2.2 Protein crystals

All protein crystals used in these experiments were tetragonal crystals of hen egg white lysozyme (HEWL). HEWL was chosen as our model system because it is easy to obtain and crystallize, its structure and function are well-understood, and it has been shown to cause diffuse scattering [Faure et al., 1994; Clarage et al., 1992; Pérez et al., 1996].

A chicken egg white is about 88% water and 10% protein, with lysozyme comprising about 3% of the protein [Yamamoto et al., 1997]. Hen egg white lysozyme primarily functions to cleave the polysaccharides of bacterial cell walls, protecting the potential chicken embryo from infection.

HEWL crystallizes in several space groups; the tetragonal form was chosen here because it is the easiest to grow. The tetragonal form's high degree of crystallographic symmetry also means that we can record several symmetry-related observations of diffuse intensity from a single location on one crystal. Since protein-derived diffuse scatter obeys crystallographic symmetry (see section 3.3), we can use these multiple observations to improve data precision. Because crystal contacts between unit cells may affect protein motions, comparing the diffuse intensity from HEWL crystals in various space groups would be an interesting future experiment, but it is beyond the scope of this work.

To investigate the effect of inhibitor binding on diffuse intensity, some of the lysozyme crystals were co-crystallized with the small molecule 3-N-acetylglucosamine (3NAG), also known as N,N',N"-triacetylchitotriose, chosen at the suggestion of Dr. Edward Snell. 3NAG binds strongly to HEWL at specific sites [Ford et al., 1974]. The intent here was to grow crystals in which every HEWL molecule was bound to a 3NAG molecule and determine whether this binding reduced diffuse scattering.

2.3 February 2014 crystals

The oldest data presented in this dissertation was taken in February 2014 from eighteen lysozyme crystals grown under various conditions in an effort to determine the effects of inhibitor binding and cryocooling on the diffuse intensity.

CHESS run:	February 2014	July 2015
Beam diameter (μm)	~ 100	~ 100
Beam divergence (°)	0.03	0.02
Beam flux (photons/sec)	$\sim 6 \times 10^{10}$	$\sim 2 \times 10^{10}$
Wavelength (Å)	0.9179	0.9774
Detector	Eiger 1M	Pilatus 6M
Crystal-to-detector distance (mm)	72	250
Beamline	CHESS F1	CHESS F1

Table 2.1: Universal (i.e., crystal-independent) parameters for each CHESS run. Beam diameter and flux are approximate.

2.3.1 Crystal growth and mounting

The crystals from which data sets were collected at this CHESS visit can be sorted into four categories:

- Room temperature, no ligand, no cryoprotectant
- Room temperature, 3NAG-bound, no cryoprotectant
- Room temperature, no ligand, ethylene glycol added
- Cryocooled, no ligand, ethylene glycol added.

The expectation was that each of these four categories could have noticeably different diffuse intensity patterns, while differences in crystal growth conditions within each category were unlikely to measurably affect the protein dynamics or scattering pattern.

The cryocooled crystals and some of the cryoprotectant-free room temperature crystals were grown using the standard hanging drop method in 24-well trays. Well solutions consisted of 0.1 M sodium acetate (pH 4.6) + 0.6–1.5 M NaCl + 25% ethylene glycol (EtGly) if using, mixed in the drops in 1:1 or 1:2 ratio with 60–75 mg/mL lysozyme (Sigma-Aldrich, powder) + 3–4 mg/mL 3NAG

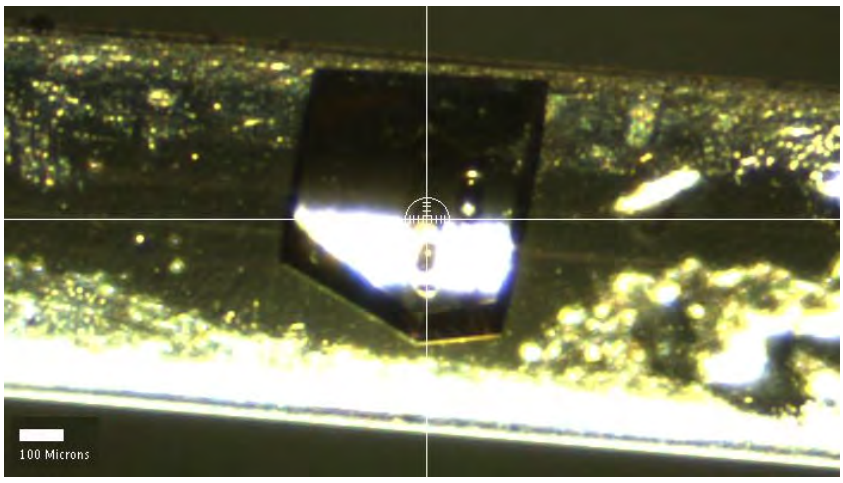
Crystal	Dimensions (μm)	Temperature (K)	EtGly/3NAG	CHESS run
lyso14	500 x 300 x 400	room	-	Feb 2014
lyso15		room	-	Feb 2014
lyso30	600 x 400 x 200	room	-	Feb 2014
lyso34	550 x 450 x 200	room	-	Feb 2014
lyso38		room	3NAG	Feb 2014
lyso47	750 x 350 x 400	room	3NAG	Feb 2014
lyso49	500 x 200 x 400	room	3NAG	Feb 2014
lyso50	300 x 550 x 250	room	3NAG	Feb 2014
lyso16	550 x 300 x 650	room	EtGly	Feb 2014
lyso17	600 x 500 x 150	room	EtGly	Feb 2014
lyso27	500 x 400 x 450	room	EtGly	Feb 2014
lyso28	500 x 400 x 200	room	EtGly	Feb 2014
lyso35	600 x 750 x 400	room	EtGly	Feb 2014
lyso5		100	EtGly	Feb 2014
lyso8X		100	EtGly	Feb 2014
lyso19		100	EtGly	Feb 2014
lyso21		100	EtGly	Feb 2014
lyso22		100	EtGly	Feb 2014
lyso6	450 x 350 x 250	300	-	July 2015
lyso7	350 x 200 x 250	300	-	July 2015
lyso8	500 x 400 x 100	300	-	July 2015
lyso9	800 x 500 x 200	300	3NAG	July 2015
lyso10	900 x 500 x 250	300	3NAG	July 2015

Table 2.2: List of crystals and their experimental parameters. Crystal dimensions are listed with the dimension held perpendicular to the beam (parallel to the rotation axis) first. Dimensions are approximated from photographs; where dimensions are missing, photographs were not taken. In this table and thereafter, February 2014 crystal names will be printed in red (**lyso14**) and July 2015 crystal names will be printed in black (**lyso6**).

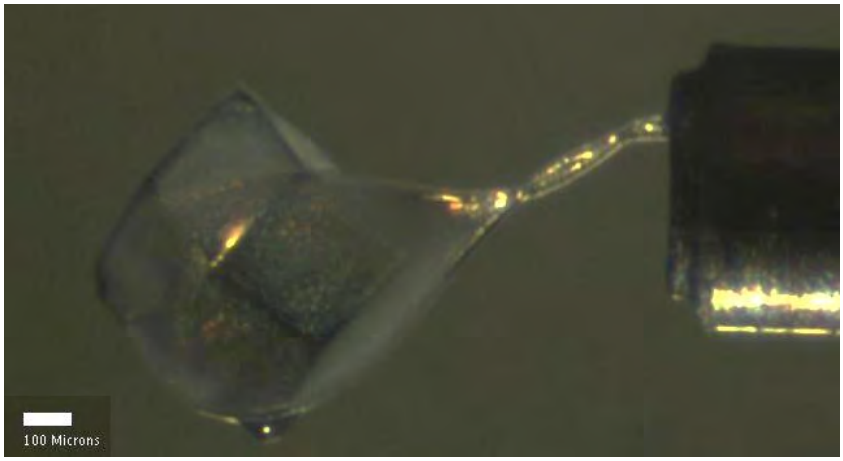
(Sigma-Aldrich, powder) if using. Crystals to be cryocooled were then mounted one at a time on the beamline via scooping out of the drop with a nylon loop (Hampton Research) and plunging into a tub of liquid nitrogen, which kept the crystals from drying out. Crystals to be x-rayed at room temperature were transferred with a loop into glass capillaries containing at least two disjoint plugs of well solution from that crystal's well. The crystal was dropped into one plug of the well solution, and excess liquid around the crystal was removed with a paper wick. The remaining plug(s) of well solution served to keep the crystal hydrated. The ends of the capillary were sealed either with grease, delivered via syringe needle, or with five-minute epoxy. Figures 2.1a and 2.1b show crystals inside a capillary after sealing and mounted on a nylon loop after cryocooling.

Rigid capillaries were chosen for crystal mounting because they preserve crystal hydration long-term without cryocooling and they allow loop-free crystal mounting (avoiding anisotropic scattering from the loop itself). Regular glass capillaries were chosen over more specialized capillary materials (quartz and boron-doped glass) because they are easier to handle and, when illuminated with Cu Kα radiation (1.54 Å) in an in-house test, they displayed lower scattering intensity than the other capillary types. This indicates, though does not confirm, that the specialized capillaries would have scattered more x-rays on the synchrotron beamline (0.918 Å) , which would reduce clarity of the protein-derived diffuse signal.

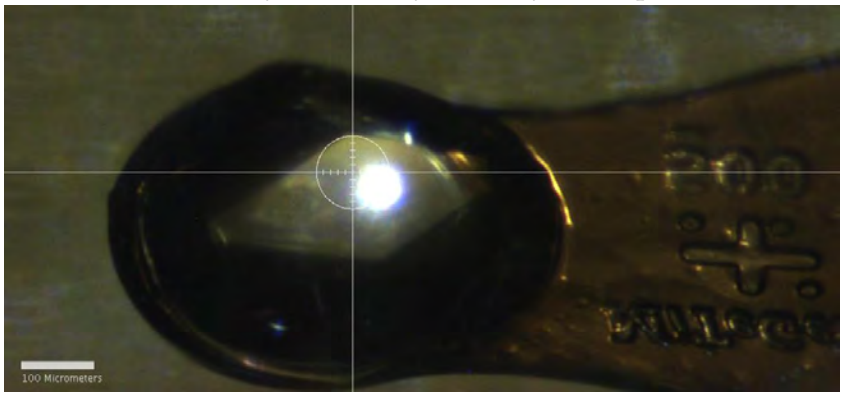
The rest of the room temperature crystals were grown directly in glass capillaries. Well solutions consisted of 0.1 M sodium acetate (pH 4.6) + 0.9–1.3 M NaCl + 25% EtGly if using; for each capillary, 1–2 mL of well solution was placed in the bottom of a 15 mL centrifuge tube and partially soaked up with a



(a) Room temperature crystal in glass capillary



(b) Cryocooled crystal in nylon loop



(c) Room temperature crystal in polyimide loop and sheath

Figure 2.1: The three crystal mounts used. (a) Glass capillary for February 2014 room temperature crystals. (b) Nylon loop for cryocooled crystals. The pin, ultimately attached to the goniometer, is visible on the right side. (c) Polyimide MicroMount for July 2015 room temperature crystals. The crystal and loop are inside a plastic sheath, the edges of which are outside the image.

Kimwipe. A solution of 60–75 mg/mL lysozyme + 3.6 mg/mL 3NAG if using was mixed with well solution in a 1:1 or 2:1 ratio and added to a glass capillary with both ends open; the capillary was then placed in the centrifuge tube and let to rest for a few days. Once crystals grew, a plug of well solution was added to the capillary, liquid was wicked away from the crystal of interest inside the capillary, and both ends of the capillary were sealed as described above. This growth method took more time, both for setup and for crystal growth, than the hanging drop method, but it required less manipulation of the crystals and therefore theoretically reduced the risk of crystal damage during mounting.

Based on photographs taken through the beamline camera, the crystals ranged in size from approximately 0.04–0.18 mm^3. Most had an asymmetric chunky shape, with shortest dimensions in the 200–400 μm range and longest dimensions in the 500–750 μm range. The crystals tended to settle in the loops or capillaries with their longest dimension roughly aligned with the spindle axis, perpendicular to the beam at all rotation angles.

2.3.2 Beamline setup

These data were taken at beamline F1 at MacCHESS. For the most part, the experimental setup was a standard macromolecular crystallography setup, with the sample mounted on a goniometer whose rotation axis was perpendicular to the x-ray beam. Cryocooled crystals on loops, attached in turn to metal pins, were mounted on standard pin bases that attach magnetically to the goniometer. Glass capillaries containing crystals were attached to pin bases using modeling clay. A beamstop was placed just downstream of the sample, with the detector

72 mm beyond the sample. For the cryocooled crystals, a cryostream (100 K) was pointed at the sample.

Two features distinguished this setup from a standard crystallography setup. First, for the room temperature data, the cryostream was moved entirely out of the way, so the sample's temperature was entirely uncontrolled. Second, an atypical beamstop mount was used in an attempt to avoid anisotropic scattering off of the edges of the beamstop holder. The standard beamstop holder used at MacCHESS is a vertical strip of kapton tape a few millimeters wide with the beamstop itself—a small, dense piece of metal—stuck on in the middle. The tape is attached to metal rods several inches above and below the beam path, which in turn are mounted on a base that can slide in both horizontal dimensions. The advantage of using a thin strip of kapton to hold the beamstop is that most scattered x-rays from sample to detector will pass to one side or the other without risking absorption by the tape, but the disadvantage is that the resulting non-uniform absorption by the tape and scattering off the tape edges will cast a shadow in the diffraction images. In traditional crystallography experiments, this disadvantage is irrelevant, because the shadow is much weaker than the Bragg peak intensities, but it poses a problem for diffuse scattering experiments because diffuse intensities are of a similar magnitude to the beamstop holder shadow.

In an attempt to mitigate such contamination of the diffuse intensity pattern, a different beamstop holder consisting of a full sheet of mylar, mounted on the same type of metal rods and base, was used. The mylar sheet was wide enough that all scattered x-rays eventually incident on the detector would pass through the mylar, which reduced all recorded intensities but avoided any extra scat-

tering off the edge of the sheet. Unfortunately, nobody checked carefully how the beamstop itself was attached to the mylar, and it was later discovered that a piece of standard adhesive tape had been involved. The tape was not visually obvious to experimenters, but it became clear during image processing that the tape both absorbed and scattered x-rays in a certain region, ultimately negating the purpose of the mylar sheet beamstop. The tape effects were accounted for in image processing (see section 3.2.6).

2.3.3 Data collection

The detector used in this CHESS run was an early version of an Eiger 1M [Dinapoli et al., 2011], which was not yet configured to use its full dynamic range (though it had all the other useful features of a pixel array detector as described in Section 1.2.2). The available dynamic range was only 4×10^3, which is insufficient to capture both Bragg and diffuse intensities to reasonable precision. Therefore, two alternating exposure schemes were used for all data collection: a shorter exposure (0.05 seconds) over 1 degree of oscillation to collect Bragg data, followed by a longer still exposure (5 seconds, or 10 seconds for some cryocooled crystals) to collect diffuse data. Diffuse data could be collected in stills without loss of relevant information because the protein-derived signal of interest does not vary significantly over 1 degree of oscillation. The alternation of short oscillations and long stills was done automatically by a custom software modification written by Dr. Marian Szebenyi.

For room-temperature crystals, the total x-ray exposure required for one crystal would have caused too much radiation damage if applied to one spot

on the crystal, so the crystal was translated relative to the beam every 10 degrees or so. The beam was 100 μm in diameter while the crystals were typically much larger than that, so there was usually ample fresh or mostly fresh space on the crystal to illuminate at any time.

For cryocooled crystals, radiation damage was not a significant concern, but the nylon loop used to mount the crystals had the potential to introduce strong anisotropic scattering amidst the protein-derived diffuse scattering. To avoid this, each cryocooled crystal was oriented before data collection such that it could rotate through 90 degrees (sufficient for tetragonal symmetry) with the beam always passing through the loop and never hitting the nylon.

2.4 July 2015 data

The rest of the data presented in this dissertation was taken in July 2015 from five lysozyme crystals, all grown and measured at room temperature, in an effort to further analyze reproducibility and determine the effect of inhibitor binding on diffuse intensity.

2.4.1 Crystal growth and mounting

All crystals were grown by the hanging drop method in 24-well trays. Well solutions consisted of 0.1 M sodium acetate (pH 4.5) + 0.9–1.4 M NaCl, mixed in the drops in 1:1 ratio with 30 mg/mL lysozyme + 4.3 mg/mL 3NAG if using. Crystals were mounted on MiTeGen MicroMounts and quickly covered with a polyester MiTeGen MicroRT capillary, which was closed around a blob of well

solution at the top end and sealed with grease at the bottom end (near the pin holding the MicroMount). The capillary served to keep the crystal from drying out. Crystals had similar size and settling behavior to those in the February 2014 data set. Figure 2.1c shows a crystal mounted on a MicroMount.

2.4.2 Beamline setup

The beamline setup was the same as that in February 2014, also at beamline F1 at MacCHESS, except that instead of moving the cryostream out of the way, the cryostream was warmed to 300 K and pointed at the crystal in order to keep the crystal temperatures more consistent. The crystal-to-detector distance was also increased to 250 mm because the detector used had a larger surface area.

2.4.3 Data collection

The detector used in this CHESS run was a fully operational Pilatus 6M, with a dynamic range of 10^6. Since this range was sufficient to capture both Bragg and diffuse intensities simultaneously, a single exposure time of 0.1 second per 0.1° was used for all data collection. All data was taken in 180° continuous sweeps, resulting in series of 1800 images. Between sweeps, the crystal was translated perpendicular to the beam so as to image a new spot. Three to six sweeps were collected per crystal.

CHAPTER 3

DATA PROCESSING

This chapter describes the procedures used to convert raw diffraction data into three-dimensional maps of the diffuse intensity. Steps are detailed in the following sections, with associated validation and data quality checks as appropriate.

3.1 Traditional Bragg data

Each full data set from a unique protein crystal was first processed traditionally, i.e. with software designed for structure determination from Bragg peaks. Software packages used were XDS [Kabsch, 2010], CCP4 (Collaborative Computational Project No. 4 Software for Macromolecular X-Ray Crystallography) software suite version 7.0 [Winn et al., 2011], and PHENIX version 1.15.2 [Adams et al., 2010].

Traditional processing was used to:

- confirm the crystal space group;
- evaluate data quality;
- confirm inhibitor binding when the inhibitor molecule 3NAG was present;
- correct for variation in absorption by the crystal, the detector itself, and anything else in the beam path;
- and place all images on a common scale.

Traditional processing was also used to create a real-space model of the static protein structure for each crystal.

3.1.1 Integration and scaling

Raw images as recorded on the beamline were provided to XDS for integration and parameter refinement with mostly default settings.

For the February 2014 data, the short-exposure oscillation images were the ones used. Because the crystal was translated every 8-10 images, the raw data comprised several "runs", each of which was too short for XDS to index and integrate correctly. Therefore, the raw images were renamed as if part of a single "run", or single continuous rotation of the spindle during data collection.

For the July 2015 data, the original 0.1° oscillation images were used, and runs were not redefined. Each crystal therefore had 3-6 overlapping runs' worth of data, processed separately by XDS.

For each run, XDS produced an output file in .HKL format consisting of refined crystal parameters and data collection parameters followed by a list of Bragg peaks, each with their intensity, reciprocal space location in hkl coordinates, and real-space location on the detector.

The .HKL files were then input to the CCP4 program AIMLESS [Evans and Murshudov, 2013], which accepts multiple .HKL files as input and merges all observations of each Bragg peak into a single intensity value for that Bragg peak. Since AIMLESS uses knowledge of run boundaries to assign scale factors to Bragg peak observations, the February 2014 run boundaries were redefined

upon input to reflect where data collection was actually stopped and started when the crystal was translated.

AIMLESS also computes several quality metrics for the input data. These metrics, presented and defined below, were used to determine which images to discard for poor quality (possibly due to excessive radiation damage, a secondary salt or protein crystal passing through the beam, or other microscale sources of disorder) and what diffraction resolution cutoff to apply to the data. Once these cuts were applied, AIMLESS was rerun to compute quality metrics on the final data set.

3.1.2 Data quality

Standard parameters and quality metrics for each data set are presented and defined in Table 3.1. Overall, the February 2014 data (all those with the Eiger 1M detector) are decent and the July 2015 data (all those with the Pilatus 6M detector) are excellent. The July 2015 data exhibit extremely low mosaicity, high multiplicity, high $\langle I/\sigma \rangle$, low R_{merge} and R_{meas}, and extremely high $\text{CC}_{1/2}$, all of which indicate a large well-ordered crystal with ample and self-consistent data recorded.

The February 2014 data sets all have higher reported mosaicity (0.14°–0.29°), which indicates broader Bragg peaks; broader Bragg peaks in turn require more aggressive masking when extracting diffuse intensities (section 3.2.3). It is worth noting, however, that extremely low mosaicities such as those observed in the July 2015 data (0.03°–0.05°) are difficult to measure with an oscillation angle of 1.0° because fewer reflections appear on multiple consecutive images.

Therefore, it is possible that the difference in reported mosaicity between the February 2014 and July 2015 data is due partially to the difference in data collection strategy, rather than due entirely to differences in crystal quality.

That said, R_{meas}, the measure of internal inconsistency in the data set that is fairly insensitive to multiplicity, is higher (worse) for all the February 2014 crystals, but still well within an acceptable range. The correlation between halves of the data set, $CC_{1/2}$, which is sometimes considered a better quality metric [Evans and Murshudov, 2013], remains high (good) across all data sets except for a dip to 0.976 in **lyso30**. In the highest resolution shells, $CC_{1/2}$ drops as low as 0.260 (**lyso14**), indicating that data quality is significantly reduced but that usable information is likely still present. Resolution cutoffs were chosen to be relatively high, keeping as much data as possible, for the sake of similarity with the diffuse intensity analysis, in which data was kept out to relatively high resolution since no traditional cutoff metric exists.

While the July 2015 data sets and nine of the February 2014 data sets had over 90% completeness, the other nine February 2014 data sets are relatively incomplete (as low as 69.4% for **lyso27**) due to an insufficient rotation range during data collection. Therefore, not all of reciprocal space in the reported resolution range is available for either Bragg or diffuse analysis. The February data sets also suffer from relatively low multiplicity; for the diffuse intensities, this results in greater uncertainty in the final corrected and averaged diffuse intensity values than in the July 2015 data.

Unit cell constants differ slightly but significantly between each of the four conditions. Room temperature crystals with neither 3NAG nor EtGly had average unit cell dimensions of a, b = 79.1 Å and c = 37.9 Å; room temperature

crystals with 3NAG had average unit cell dimensions of 78.6 Å and 38.4 Å; room temperature crystals with EtGly had average unit cell dimensions of 79.6 Å and 37.7 Å; and cryocooled crystals with EtGly had average unit cell dimensions of 78.8 Å and 36.8 Å. All of these averages have uncertainty bounds of ± 0.1 Å and are consistent between February 2014 and July 2015 data sets. This indicates consistent small structural shifts in response to crystal growth conditions, which could cause different diffuse scattering patterns.

	lyso14	lyso15	lyso30	lyso34	lyso38
Data collection					
Temperature (K)	room	room	room	room	room
EtGly/3NAG	-	-	-	-	3NAG
Wavelength (Å)	0.9179	0.9179	0.9179	0.9179	0.9179
Detector	Eiger 1M	Eiger 1M	Eiger 1M	Eiger 1M	Eiger1M
Bragg statistics					
Cell a, b, c (Å)	79.0, 79.0, 37.7	79.2, 79.2, 37.9	79.1, 79.1, 37.9	79.2, 79.2, 37.9	78.5, 78.5, 38.6
Resolution range (Å)	21.59–1.57 (1.60–1.57)	20.88–1.50 (1.53–1.50)	20.88–1.57 (1.60–1.57)	20.90–1.65 (1.68–1.65)	21.76–1.53 (1.56–1.53)
Oscillation range (°)	1.0	1.0	1.0	1.0	1.0
Mosaicity (°)	0.19	0.14	0.20	0.18	0.15
Multiplicity	6.2 (3.8)	4.5 (1.8)	3.9 (2.5)	4.4 (3.4)	6.2 (2.9)
Completeness (%)	94.2 (97.7)	90.1 (83.8)	95.7 (88.5)	97.8 (90.1)	99.7 (98.7)
$\langle I/\sigma(I)\rangle$	7.4 (1.2)	6.8 (0.7)	7.7 (1.6)	6.6 (1.1)	8.3 (1.1)
R_{merge}	0.148 (1.263)	0.125 (1.125)	0.164 (1.386)	0.148 (1.301)	0.132 (1.091)
R_{meas}	0.161 (1.471)	0.138 (1.476)	0.187 (1.766)	0.167 (1.521)	0.143 (1.329)
$CC_{1/2}$	0.989 (0.260)	0.989 (0.384)	0.976 (0.625)	0.988 (0.338)	0.994 (0.404)
Structure statistics					
Resolution range (Å)	19.79–1.50	19.79–1.50	19.77–1.57	17.56–1.65	21.76–1.53
Average B factor (Å^2)	21.7	21.5	19.5	22.0	19.1
$R_{\text{work}}/R_{\text{free}}$	0.174/0.210	0.174/0.209	0.187/0.222	0.166/0.205	0.166/0.209

	lyso47	lyso49	lyso50	lyso16	lyso17
Data collection					
Temperature (K)	room	room	room	room	room
EtGly/3NAG	3NAG	3NAG	3NAG	EtGly	EtGly
Wavelength (Å)	0.9179	0.9179	0.9179	0.9179	0.9179
Detector	Eiger 1M	Eiger 1M	Eiger 1M	Eiger 1M	Eiger 1M
Bragg statistics					
Cell a, b, c (Å)	78.5, 78.5, 38.3	78.8, 78.8, 38.4	78.7, 78.7, 38.3	79.8, 79.8, 37.8	79.6, 79.6, 37.7
Resolution range (Å)	21.60–1.50 (1.53–1.50)	21.67–1.58 (1.61–1.58)	20.87–1.57 (1.60–1.57)	21.77–1.35 (1.37–1.35)	21.69–1.54 (1.57–1.54)
Oscillation range (°)	1.0	1.0	1.0	1.0	1.0
Mosaicity (°)	0.15	0.19	0.19	0.22	0.18
Multiplicity	8.0 (3.1)	7.1 (4.7)	7.3 (4.7)	3.6 (1.1)	4.9 (3.0)
Completeness (%)	99.6 (97.0)	99.7 (99.9)	99.9 (100.0)	79.0 (23.6)	80.4 (68.8)
$\langle I/\sigma(I)\rangle$	10.1 (0.9)	9.5 (1.2)	9.1 (1.2)	7.2 (0.0)	8.2 (0.9)
R_{merge}	0.127 (1.469)	0.123 (1.392)	0.143 (1.511)	0.123 (1.252)	0.125 (1.323)
R_{meas}	0.135 (1.755)	0.132 (1.575)	0.153 (1.705)	0.136 (1.770)	0.139 (1.597)
$CC_{1/2}$	0.994 (0.310)	0.995 (0.345)	0.994 (0.323)	0.990 (0.739)	0.992 (0.348)
Structure statistics					
Resolution range (Å)	21.60–1.50	21.67–1.58	18.55–1.57	19.36–1.35	21.69–1.54
Average B factor (Å^2)	21.1	22.7	21.2	24.8	22.8
$R_{\text{work}}/R_{\text{free}}$	0.165/0.196	0.165/0.185	0.164/0.185	0.200/0.231	0.160/0.180

	lyso27	lyso28	lyso35	lyso5	lyso8X
Data collection					
Temperature (K)	room	room	room	100	100
EtGly/3NAG	EtGly	EtGly	EtGly	EtGly	EtGly
Wavelength (Å)	0.9179	0.9179	0.9179	0.9179	0.9179
Detector	Eiger 1M	Eiger 1M	Eiger 1M	Eiger 1M	Eiger 1M
Bragg statistics					
Cell a, b, c (Å)	79.6, 79.6, 37.7	79.6, 79.6, 37.7	79.6, 79.6, 37.7	78.8, 78.8, 36.8	78.8, 78.8, 36.8
Resolution range (Å)	21.70–1.40 (1.42–1.40)	21.70–1.50 (1.53–1.50)	20.94–1.54 (1.57–1.54)	21.86–1.46 (1.49–1.46)	21.39–1.37 (1.39–1.37)
Oscillation range (°)	1.0	1.0	1.0	1.0	1.0
Mosaicity (°)	0.20	0.18	0.20	0.19	0.29
Multiplicity	3.8 (1.6)	4.7 (3.1)	6.5 (3.4)	4.5 (1.5)	4.2 (1.8)
Completeness (%)	69.4 (29.9)	81.7 (47.6)	99.8 (99.5)	85.8 (77.7)	78.6 (23.9)
$\langle I/\sigma(I)\rangle$	9.8 (0.3)	7.9 (0.7)	10.0 (1.0)	7.4 (0.8)	11.0 (0.9)
R_{merge}	0.138 (3.099)	0.116 (1.534)	0.111 (1.238)	0.114 (0.855)	0.072 (0.865)
R_{meas}	0.154 (4.195)	0.128 (1.796)	0.120 (1.471)	0.126 (1.203)	0.080 (1.133)
$CC_{1/2}$	0.989 (0.432)	0.992 (0.347)	0.995 (0.357)	0.995 (0.354)	0.998 (0.343)
Structure statistics					
Resolution range (Å)	17.81–1.40	21.70–1.50	19.31–1.54	21.37–1.46	19.12–1.37
Average B factor (Å^2)	23.5	22.6	21.9	16.2	18.3
$R_{\text{work}}/R_{\text{free}}$	0.187/0.238	0.170/0.204	0.169/0.195	0.187/0.237	0.181/0.199

	lyso19	lyso21	lyso22	lyso6	lyso7
Data collection					
Temperature (K)	100	100	100	300	300
EtGly/3NAG	EtGly	EtGly	EtGly	-	-
Wavelength (Å)	0.9179	0.9179	0.9179	0.9774	0.9774
Detector	Eiger 1M	Eiger 1M	Eiger 1M	Pilatus 6M	Pilatus 6M
Bragg statistics					
Cell a, b, c (Å)	79.0, 79.0, 36.8	78.8, 78.8, 36.8	78.8, 78.8, 36.8	79.1, 79.1, 37.9	79.1, 79.1, 37.9
Resolution range (Å)	21.41–1.30 (1.32–1.30)	21.38–1.32 (1.34–1.32)	21.86–1.31 (1.33–1.31)	21.95–1.20 (1.22–1.20)	21.94–1.20 (1.22–1.20)
Oscillation range (°)	1.0	1.0	1.0	0.1	0.1
Mosaicity (°)	0.18	0.20	0.18	0.05	0.04
Multiplicity	5.0 (1.2)	4.7 (1.1)	5.0 (1.1)	40.2 (9.9)	30.2 (9.4)
Completeness (%)	76.6 (6.1)	84.4 (13.5)	79.7 (11.5)	96.2 (70.8)	92.1 (54.2)
$\langle I/\sigma(I)\rangle$	13.2 (1.3)	10.7 (1.1)	10.8 (0.8)	49.9 (1.5)	37.7 (1.0)
R_{merge}	0.075 (0.342)	0.082 (0.657)	0.082 (0.813)	0.036 (1.557)	0.043 (1.865)
R_{meas}	0.082 (0.484)	0.090 (0.929)	0.089 (1.150)	0.037 (1.641)	0.043 (1.969)
$CC_{1/2}$	0.996 (0.660)	0.996 (0.604)	0.997 (0.595)	1.000 (0.558)	1.000 (0.461)
Structure statistics					
Resolution range (Å)	19.74–1.30	19.11–1.32	21.86–1.31	21.95–1.20	21.95–1.20
Average B factor (Å^2)	16.0	17.2	16.8	24.1	23.8
$R_{\text{work}}/R_{\text{free}}$	0.184/0.205	0.188/0.218	0.177/0.210	0.165/0.181	0.167/0.176

	lyso8	lyso9	lyso10
Data collection			
Temperature (K)	300	300	300
EtGly/3NAG	-	3NAG	3NAG
Wavelength (Å)	0.9774	0.9774	0.9774
Detector	Pilatus 6M	Pilatus 6M	Pilatus 6M
Bragg statistics			
Cell a, b, c (Å)	79.1, 79.1, 38.0	78.7, 78.7, 38.3	78.7, 78.7, 38.3
Resolution range (Å)	21.93–1.20 (1.22–1.20)	21.82–1.20 (1.22–1.20)	21.83–1.20 (1.22–1.20)
Oscillation range (°)	0.1	0.1	0.1
Mosaicity (°)	0.03	0.03	0.03
Multiplicity	39.8 (11.5)	59.6 (14.9)	60.3 (16.7)
Completeness (%)	93.8 (60.1)	95.0 (67.9)	94.1 (63.1)
$\langle I/\sigma(I)\rangle$	36.9 (1.7)	55.0 (4.1)	60.7 (3.9)
R_{merge}	0.050 (1.425)	0.046 (0.648)	0.042 (0.689)
R_{meas}	0.050 (1.491)	0.047 (0.671)	0.042 (0.711)
$CC_{1/2}$	1.000 (0.666)	1.000 (0.917)	1.000 (0.910)
Structure statistics			
Resolution range (Å)	21.93–1.20	21.82–1.20	21.83–1.20
Average B factor (Å^2)	23.5	20.8	22.2
$R_{\text{work}}/R_{\text{free}}$	0.172/0.185	0.162/0.184	0.162/0.174

Table 3.1: Major parameters and quality metrics for the Bragg data. All crystals were in space group $P4_32_12$ with unit cell angles α, β, γ all equal to 90°. Cell a, b, and c are the unit cell constants (side lengths of one unit cell). Resolution range gives the cutoffs applied in scaling with AIMLESS (under Bragg statistics) or in model-building with PHENIX (under Structure statistics). Numbers in parentheses refer to the higest resolution shell only. Mosaicity is the angular spread of orientations of unit cells. Multiplicity is the average number of individual observations of a given reflection and its symmetry mates. Completeness is the percentage of unique (not symmetry-equivalent) reflections recorded. $\langle I/\sigma(I)\rangle$ is the average single-reflection signal-to-noise ratio. R_{merge} and R_{meas} are both measures of internal inconsistency in the scaled and merged data, defined in [Evans and Murshudov, 2013]. $CC_{1/2}$ is the linear (Pearson) correlation coefficient between arbitrary halves of the data set. B factor is the same as the temperature factor discussed in section 1.1.2. R_{work} and R_{free} are both crystallographic R-factors, which report the relative difference between the structure factors calculated from the model and measured experimentally. R_{work} uses the experimental structure factors used to generate the model, and R_{free} uses a small set of experimental structure factors not used in model generation.

3.1.3 Mean structure determination

The PHENIX software suite [Adams et al., 2010; Williams et al., 2018; Afonine et al., 2012] was used to build and refine real-space structure models from the processed Bragg diffraction data. For each crystal, a structure was built as follows:

First, the scaled and merged reflection data from AIMLESS were used to perform molecular replacement against an existing model, using Phaser-MR [McCoy et al., 2007] with the default settings in the simple one-component interface. For crystals unbound to 3NAG, the existing model used was Protein Data Bank (PDB) entry 193L [Vaney et al., 1996], a published model of a tetragonal HEWL structure to 1.33 Å resolution. For crystals bound to 3NAG, the existing model used was PDB entry 5NJQ [Beyerlein et al., 2017], a published model of a tetragonal HEWL structure bound to 3NAG (referred to as N,N′,N″-triacetylchitotriose) to 1.7 Å resolution.

The model created by molecular replacement was then refined using phenix.refine, including hydrogens and water molecules in the model but otherwise using default settings. This produced a real-space molecular model of mean atom positions along with a wealth of information about how well the model conformed to expected protein geometry. Refinement was repeated, with manual model adjustments if necessary, until the geometry was reasonable; often this required only a single five-cycle run of phenix.refine.

The Structure Statistics portion of Table 3.1 gives statistics related to the final moelecular model for each crystal. Better agreement between the model and the data is indicated by lower R_{work} and R_{free} values and by a smaller gap between

the two values; values for all crystals here are comparable to, if not better than, typical published values for hen egg white lysozyme.

The average B factor is a rough measure of overall disorder in the structure. In the February 2014 data, no significant differences in B factor were observed across the room temperature conditions, but the B factor decreased upon cryocooling (17 ± 1 Å at 100 K versus 23 ± 1 Å with EtGly at room temperature); this is expected given the significant decrease in available thermal energy. In the July 2015 data, average B factor did decrease slightly upon binding to 3NAG (21.5 ± 1.0 Å bound versus 23.8 ± 0.3 Å unbound).

3.2 Diffuse data: two-dimensional images

As previously discussed, a raw diffraction image contains plenty of signals that are not the protein-derived diffuse intensity of interest; the diffuse intensity itself is also subject to modulation from multiple sources. The processing steps described in this section convert the raw images collected as described in the previous chapter into images containing, within certain uncertainty limits, only anisotropic protein-derived diffuse intensity, suitable for integration into a three-dimensional map of the diffuse scattering.

Most of the diffuse data processing was performed using the software package *Lunus* by Michael Wall [Wall, 2009]. Some methods were used without alternation; many were edited and added to create the protocol described in the rest of this chapter. Note that *Lunus* has been updated since it was downloaded for the use described herein [Wall, 2018].

3.2.1 Image formats and masking

All images were initially recorded in Crystallographic Binary File (CBF) format, the current standard for x-ray crystallography [Bernstein H. J. and Hammersley, 2005]. At the time of initial processing, *Lunus* did not support the CBF format but did support the older SMV image format (with plain-text headers and uncompressed binary data), so all images were initially converted to SMV format.

The July 2015 images, all of which were 0.1° oscillations, averaged only 1–4 photons per pixel (depending on image region) between Bragg peaks, for a total of less than 1000 photons per Bragg peak in the diffuse region. This yielded signal-to-noise ratios insufficient to perform meaningful image processing on the diffuse intensities, so groups of ten images were summed into new images that each covered 1° of oscillation. The summed images were used for all image processing discussed below.

Because the image format used stores values as positive integers, but subsequent image processing involved multiplying by non-integer factors and subtraction potentially resulting in negative values, all pixel values were scaled up by a factor chosen to take best advantage of the available value range and then placed on a pedestal (i.e., added to a constant value). This way, potentially negative pixel values could be stored as positive integers with minimal loss of precision. In the future, it would be better to begin by converting images to a floating-point format in order to store negative and decimal values directly.

Finally, a mask was applied to each image to mark detector chip boundaries, malfunctioning pixels, and the beamstop shadow to be ignored in all further image processing. Figure 3.1 shows two images with and without their masks.

Note that some of the masked regions, particularly the thinner chip boundaries, appear normal in the original image; these regions reveal themselves as artifacts upon, for instance, subtraction of the average pixel value at each radial distance from the beam center.

3.2.2 Polarization correction

The actual diffracted x-ray intensity pattern is a convolution of the incoming x-ray beam profile with the Fourier transform of the crystal's electron density. Because the x-ray beam is partially linearly polarized, the beam profile introduces a direction-dependent component to the diffracted intensity. To correct for this, the beam polarization was found by fitting an average azimuthal intensity trace to the following equation from [Wall, 2009]:

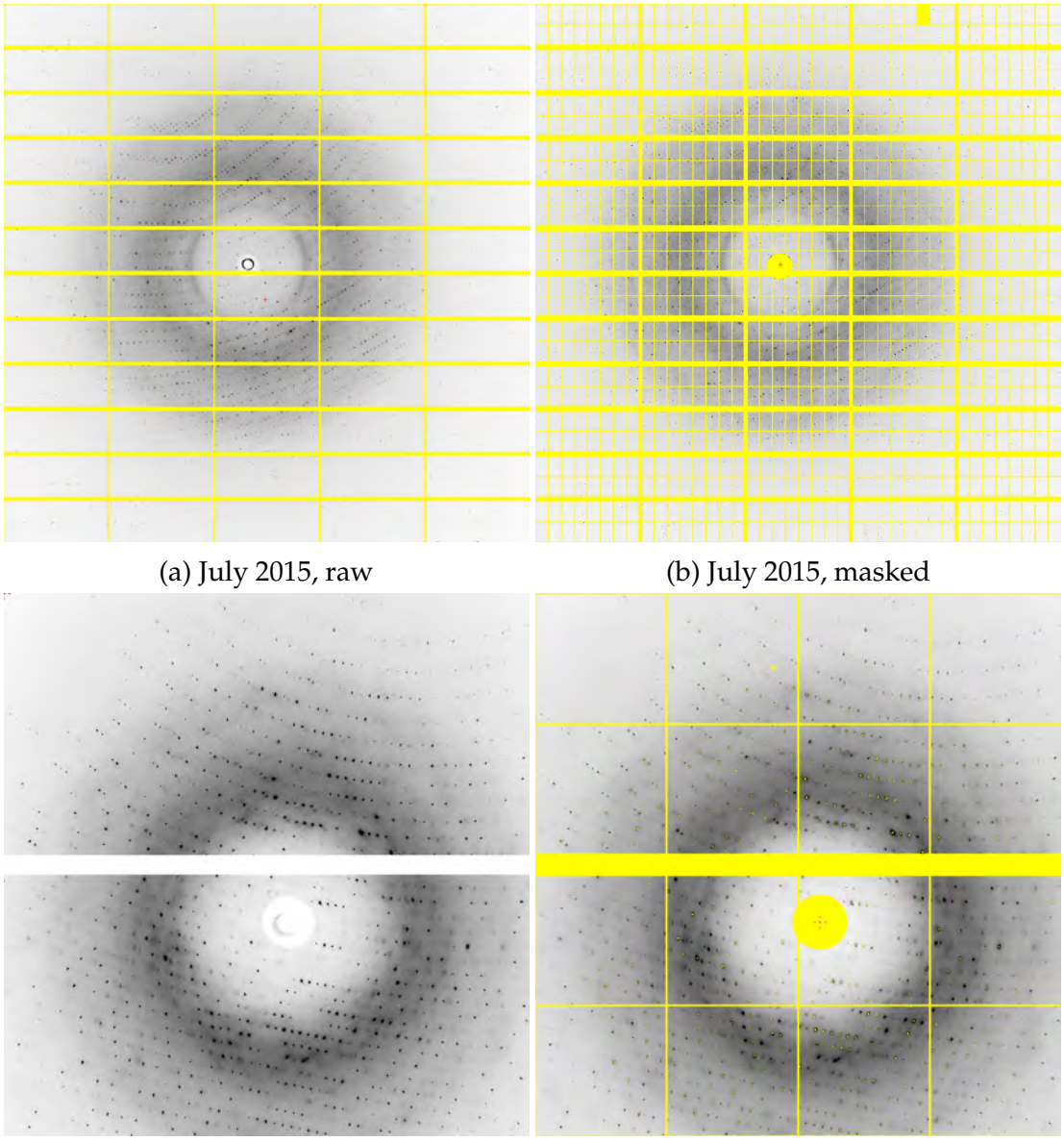
$$I(\phi, \theta) = \frac{a}{2}\left[1 + \cos^2 2\theta - \epsilon \cos 2(\phi - \phi_0) \sin^2 2\theta\right] \tag{3.1}$$

where I is the diffracted intensity, ϕ and θ are the azimuthal and scattering angles as defined in Figure 3.2, ϵ is the beam polarization, and ϕ_0 is the azimuthal angle of a vector parallel to the synchrotron ring.

The average azimuthal trace was constructed as follows: Every image in the first sweep through crystal **lyso6** from the July 2015 data was mode-filtered[1] to remove Bragg peaks and added into a single composite image. Separately, two images of the empty kapton loop and plastic capillary, taken through the thickest and thinnest dimensions of the kapton, were smoothed[2], added together,

[1] A mode filter replaces each pixel value with the mode of values in a box of a certain size around the pixel. Values are first assigned to bins so that a significant fraction of pixels in the box belong to the mode bin.

[2] Each pixel was replaced with the average of the surrounding pixel values in a box of a certain size.



(a) July 2015, raw

(b) July 2015, masked

(c) February 2014, raw

(d) February 2014, masked

Figure 3.1: A representative image from each major CHESS run, with and without masking and polarization correction. Yellow pixels are those masked, i.e., marked to be ignored. (a) First 1° from the **lyso6** crystal from July 2015 (Pilatus 6M detector). This image consists of the first 10 0.1° oscillations added together but otherwise unchanged. (c) First raw still image from the **lyso14** crystal from February 2014 (Eiger 1M detector). (b, d) Same images as in (a, c) respectively, but with untrusted regions masked and a polarization correction applied.

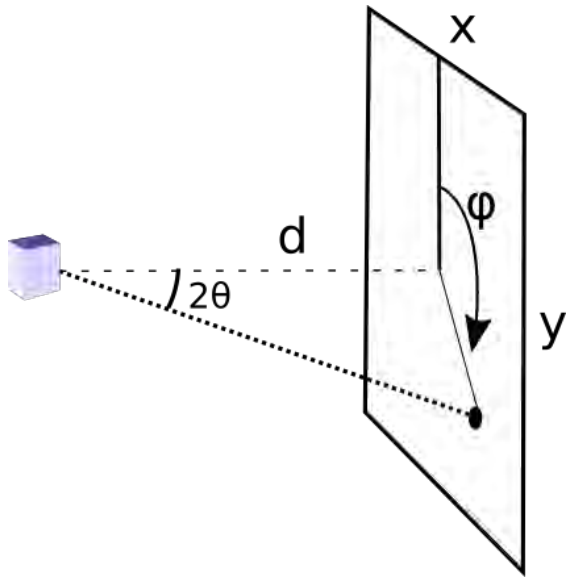


Figure 3.2: Diffraction geometry. The incident beam travels from the left along the z-axis. The detector is assumed perpendicular to the incoming beam, so positions on the detector are identified by x and y coordinates. The scattering angle 2θ is the angle between the scattered beam and the z-axis, and ϕ is the azimuthal angle clockwise from the vertical as shown. d is the distance between the crystal and the nearest point on the detector.

and scaled such that the equivalent exposure time of the kapton composite image matched that of the composite mode-filtered crystal image. The kapton image was subtracted from the crystal image, and a 100-pixel wide circle toward the edge of the resulting image was averaged into a one-dimensional azimuthal trace of intensity vs. ϕ.

Equation 3.1, using the appropriate value for θ, was then fit to this trace using gnuplot to find ϵ and ϕ_0. The best-fit values were $\epsilon = 0.93$ and $\phi_0 = 0.0$, which agree with the approximate values assumed for this beamline during traditional data processing, as communicated by a beamline scientist. Since polarization is a property of the beamline and all July 2015 and February 2014 data were taken on the same beamline, the same values were used to correct all images from both CHESS runs. Actual correction was performed using the *polarim* method

from *Lunus*.

It is worth noting that in traditional crystallography data processing software, it is standard to apply this conjunction to Bragg peak intensities in conjunction with the Lorentz correction, which adjusts for the fact that peaks further from the beam center pass more quickly through the recorded region of reciprocal space as the crystal rotates, and therefore appear weaker on the detector face. Since diffuse features of interest are much larger in reciprocal space than the Bragg reflections, the distance between recorded images in reciprocal space is smaller than the width of a diffuse feature, so there is no equivalent effect and the Lorentz correction is not needed.

3.2.3 Masking of Bragg peaks

The next step to extracting diffuse intensity was to remove the Bragg peaks. Two removal methods were applied in sequence: one to remove especially sharp, bright peaks regardless of location on the detector, and one to remove any amount of intensity at the specific locations corresponding to possible Bragg peaks given the crystal's unit cell constants and orientation. The first method misses weak Bragg peaks but catches spots in unexpected locations (e.g., in cases where the edge of a different crystal is caught in the beam) and extremely broad and bright Bragg peaks, while the second method ensures masking of all Bragg diffraction from the main crystal within a certain distance from the center of each peak.

The first removal method consists of the following steps:

CHESS run	*modeim*		*debraggim*	
	Box size	Bin size	Threshold (σ^2)	m (pixels)
February 2014	22	50	4	5
July 2015	23	100	4	6

Table 3.2: Parameters used for the first Bragg peak-masking method. *modeim* parameters: Box size is the side length of the square surrounding each pixel used to find the local mode. Bin size is the size of the range of intensity values counted as identical for purposes of finding the mode. *debraggim* parameters: The listed threshold value is in units of σ^2, where σ^2 is the overall image variance. A pixel is marked in step 3 if its value in the difference image exceeds the threshold. A pixel is masked in step 4 if it is within m pixels of of a pixel marked in step 3.

1. Mode-filter the image to (aggressively) remove Bragg peaks, using the *Lunus* method *modeim*.
2. Subtract the mode-filtered image from the original image, creating a difference image.
3. Locate all pixels in the difference image with values above a certain threshold.
4. Mask all pixels within a certain distance of those located in step 3.

Steps 2-4 were written as a new *Lunus* method called *debraggim*. The relevant parameters for mode-filtering and masking were chosen by hand for each CHESS run using a few representative images; values are reported in Table 3.2.

For the second removal method, unit cell parameters and crystal orientations were read from the output of XDS integration. Since minor orientation shifts, likely due to slight crystal slippage, did occur during data collection, a different orientation matrix was collected approximately every degree of rotation by adjusting the XDS input parameter DELPHI (which affects how many images are integrated simultaneously as a single batch) and reading orientations from

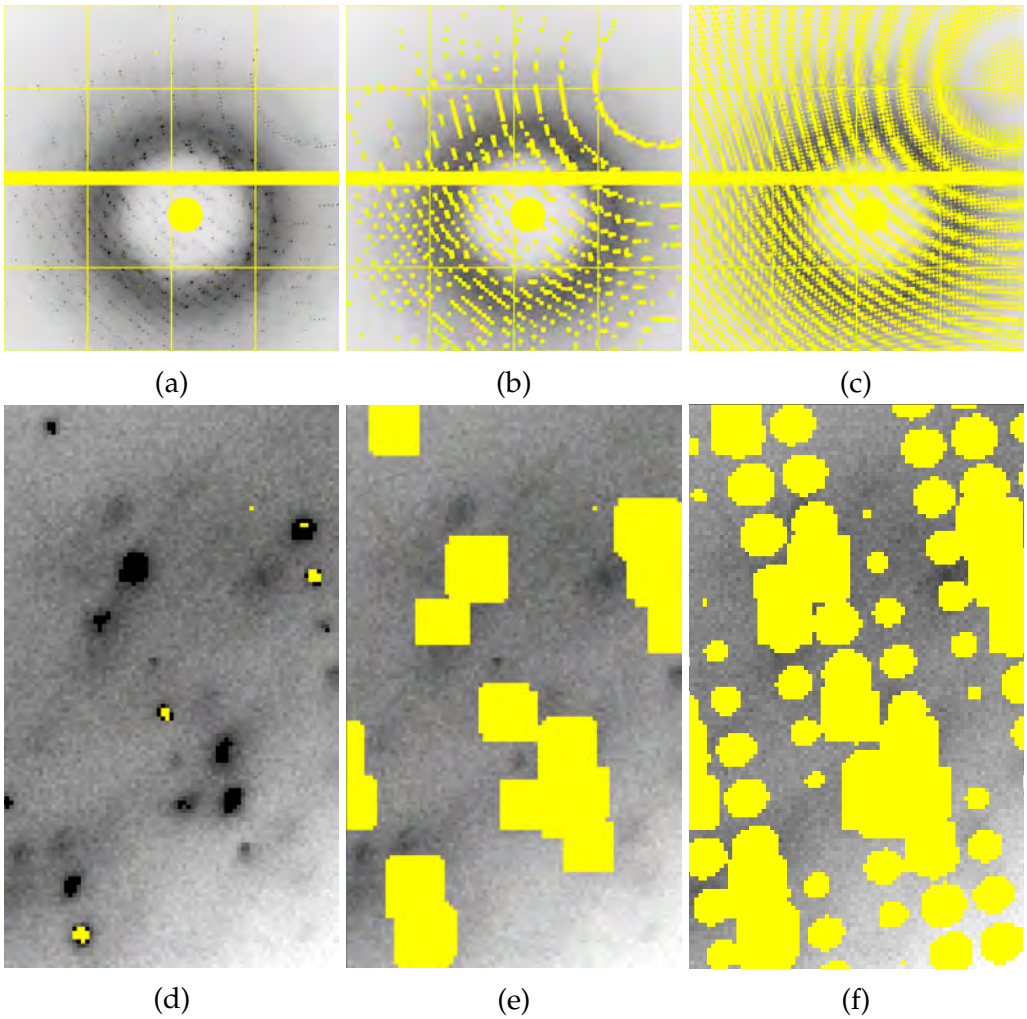


Figure 3.3: Masking of Bragg peaks in an image from crystal **lyso47**, a 3NAG-bound room temperature crystal from the February 2014 data. (a-c) The entire image. (d-f) One interesting portion of the same image. (a,d) Prior to any masking. (b,e) After applying the first, orientation-blind masking method. (c,f) After applying both masking methods.

the INTEGRATE log file, which prints a new crystal orientation for each batch. Given this information, every pixel on every image was then mapped to a spot in reciprocal space. With reciprocal space distances measured in units of the reciprocal space lattice parameters, such that the distance between adjacent points along any of the three lattice dimensions was 1 unit, pixels within 0.45 units of an integer reciprocal lattice point (i.e., a point at which a Bragg peak could pos-

sibly be centered) were masked.

The value 0.45 was chosen by hand to conservatively ensure masking of the overwhelming majority of Bragg diffraction while still leaving plenty of unmasked pixels to actually measure diffuse intensities. This radius is fairly large because Bragg spots often take up a relatively wide spread of pixels, as seen in Figure 3.3d, due to a combination of brightness, crystal mosaicity, and wavelength dispersion in the x-ray beam. Completeness of Bragg masking was validated by comparing the three-dimensional diffuse maps created from the masked images with the Bragg intensities extracted by XDS; results are discussed in section 3.3.1.

Figure 3.3 shows a representative image before and after each step of Bragg masking. Prior to Bragg masking, 7.6% of the pixels in the image were already masked as previously described. The first step masked an additional 10.5% of the image, and the second step masked an additional 27.9% of the image, such that 54% of the detector face remains available for measuring diffuse intensities after masking. The bottom row of Figure 3.3 shows the relatively conservative masking in the patches completely blanked out due to broad Bragg peaks.

Among the parameters that contribute to Bragg spot size is the size of the x-ray beam itself. Even in a perfectly diffracting crystal, a Bragg spot cannot be narrower than the intersection of the beam and the crystal; this becomes relevant for beams and crystals wider than one pixel (72 μm or 172 μm in the detectors used here). A smaller beam will yield a smaller maximum Bragg peak width. However, once the beam is smaller than a pixel and smaller than the crystal (and therefore not unnecessarily illuminating empty space or mounting materials around the crystal), reducing the beam size further will weaken the

desirable diffraction intensity and, eventually, broaden Bragg peaks again if the number of illuminated unit cells in each dimensions becomes small. All experiments described here were performed with an x-ray beam approximately 100 μm in diameter, which is approximately optimal: similar to the pixel size, and never larger than the crystal.

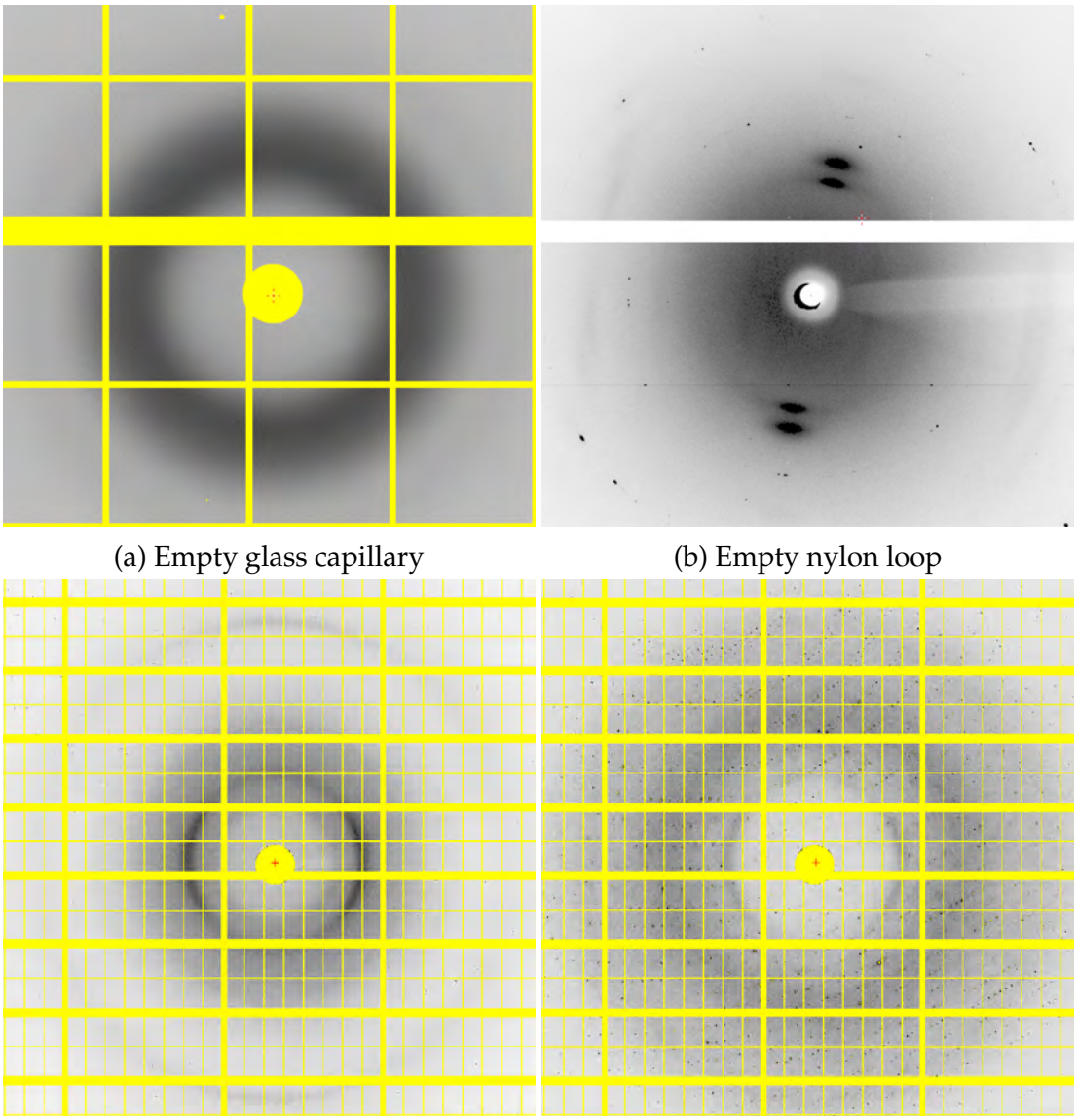
Previous studies of diffuse scattering have used various methods to mask Bragg peaks [Peck et al., 2018; Ayyer et al., 2016], including the straightforward method of simply mode-filtering the image [Wall et al., 1997b]. Mode-filtering was not selected here because it obscures any features smaller than the mode filter box size. For much of the data presented here, the mode filter box side length had to exceed the distance between Bragg peaks to reliably remove Bragg scattering, so diffuse features of a similar size would be lost. While most biologically interesting diffuse features are expected to be much larger than the distance between Bragg peaks, it was still interesting to retain the possibility of observing smaller features. Figure 3.3d-f shows one instance of relatively small streaky features, still visible after masking. Larger darker and lighter patches are also present and vaguely visible in these images, though they are much more easily seen after subtraction of the radially symmetric portion of the signal.

3.2.4 Background subtraction: mounting materials

The remaining diffuse intensity in an image comprises the desired protein-derived intensity and scattering from other materials in the beamline. Here, we focus on anisotropic intensity alone, so the anisotropic intensity from non-protein materials must be subtracted.

The mounting materials used in the February 2014 CHESS run (see section 2.3.1) did not contribute significant anisotropic scattering to the diffraction patterns. Figure 3.4a shows the radially symmetric glass scattering pattern from one of the capillaries used for the room temperature crystals. Figure 3.4b shows the characteristic scattering pattern produced by hitting a nylon loop used for the cryocooled crystals. This pattern completely disappears when the loop is rotated such that the beam passes through the open center of the loop without actually hitting it. All crystals used were large enough to rotate through at least 90 degrees without the loop intersecting the beam, so none of the diffraction images used were contaminated with nylon loop scattering. The lighter shadow on the right side of the image, most likely from the pin holding the crystal, does appear regardless of loop orientation; since it shifted position slightly as crystals rotated, and from crystal to crystal, the entire wedge-shaped region over which it appeared was simply masked.

The kapton and plastic used in the July 2015 CHESS run did produce noticeable anisotropic scattering, shown in Figure 3.4c. Because the anisotropic features in this fiber scattering pattern are qualitatively similar to, but brighter than, protein-derived diffuse features, they had to be subtracted to avoid noticeably contaminating the data. The following "fiberfit" algorithm was used to perform this subtraction: A series of images of the scattering off an empty kapton loop and plastic sheath, taken every 5° over a rotation range of 180°, were added together into a single image representing the average background scatter over the whole intensity range. This composite image was then masked, polarization-corrected, and smoothed to minimize noise with a mean-filter that replaced every pixel with the average of the pixels in a surrounding 15x15 pixel box. An average radial trace was computed from all input images (after Bragg



(a) Empty glass capillary

(b) Empty nylon loop

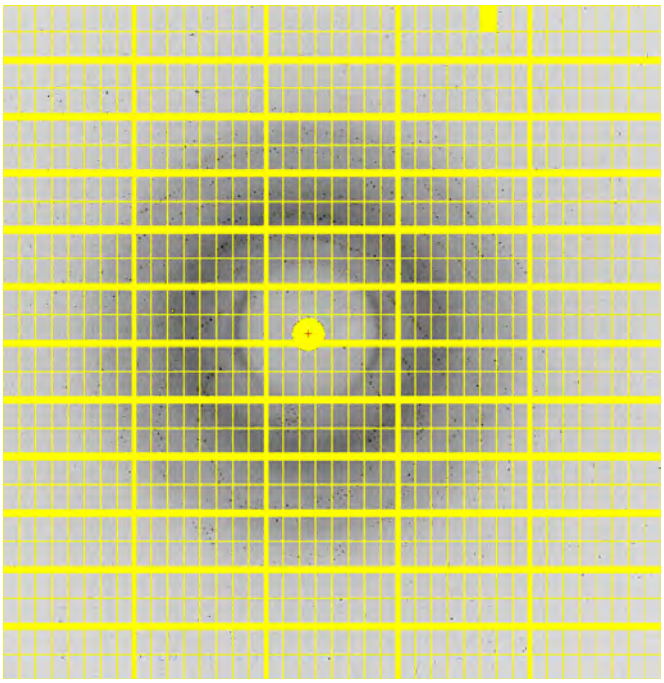
(c) Empty kapton loop and plastic sheath

(d) Crystal on kapton loop in plastic sheath

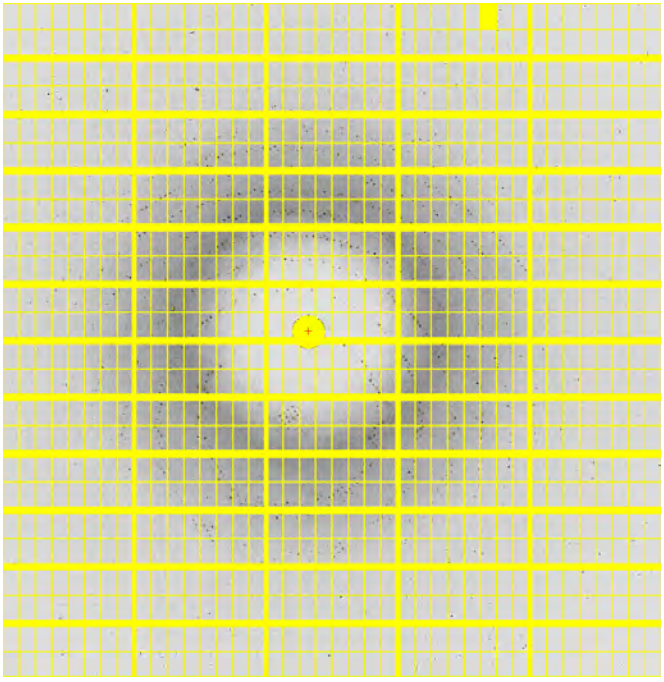
Figure 3.4: Diffraction images of each type of mounting materials used, displayed at arbitrary intensity scales to highlight relevant features. (a) An empty glass capillary, from the February 2014 CHESS run. (b) An empty nylon loop hit directly with the x-ray beam, from the February 2014 CHESS run. The two pairs of large dark spots are the characteristic signature of loop scattering; the smaller disperse spots are likely from tiny salt crystals on the loop. (c) An empty kapton loop in a sheath, from the July 2015 CHESS run. The two sets of dark asymmetric bands are the characteristic signature of fiber scattering from this setup. (d) A crystal mounted on a kapton loop in a plastic sheath. Note the signature dark bands toward the center of the image.

peak removal and all previous steps) and from the composite background image. All radial traces show a noticeable peak at a radius of approximately 270 pixels (which corresponds to the same diffraction resolution in all images, as the detector distance and x-ray wavelength were held constant). The peak region from each radial trace was isolated, and MATLAB's regress function was used to fit the selected region of each crystal's radial trace to the sum of a quadratic function and the background image trace multiplied by some coefficient. A quadratic function, rather than a linear function, was found by trial and error to do a better job fitting the non-background portion of the crystal's radial trace. For each crystal image, the composite background image was then multiplied by the regression coefficient and subtracted. A characteristic image before and after this subtraction is shown in Figure 3.5.

Future diffuse scattering studies would benefit from more advanced methods of background subtraction. Chapman et al. [Chapman et al., 2017] have published a method of background subtraction that relies on assumptions about the statistical distributions of the different sources of diffuse intensity in an image and was shown to improve the quality of one diffuse map. They assumed that each molecule behaved as a rigid body undergoing translational disorder, which is incompatible with discovering the types of motions of fundamental interest to the current work. However, if their method is extended to other crystals, detectors, and types of disorder, it will likely be valuable to the progress of this field. A cruder, untested, but promising experimental approach to background subtraction is to record actual scattering from the mounting materials used as closely as possible; this approach is detailed in Appendix A. Unlike Chapman et al.'s method, the experimental approach would not remove isotropic scattering from mother liquor around the crystal. A standard-



(a)



(b)

Figure 3.5: An image from crystal **lyso7** (July 2015 data), (a) before and (b) after subtracting the fiber scattering due to its mounting materials. Intensity scale is arbitrary but identical for both images. For clarity, these images are shown without the Bragg peak mask; background subtraction was otherwise performed after Bragg peak masking.

ized method for mounting crystals and subtracting the resulting background scatter to extract diffuse intensities would be very helpful, but does not yet exist.

3.2.5 Scaling and crystal absorption corrections

Because the crystal size is nonuniform in all dimensions, the overall intensity of all protein-derived scattering changes as the crystal rotates and as the crystal is translated perpendicular to the beam between sets of images. Furthermore, the crystal itself absorbs a fraction of the scattered x-rays, and this absorption effect is similarly nonuniform, not only between images but also within a given image. XDS, the software used for traditional Bragg data processing, uses symmetry-related integrated peak intensities to compute multiplicative factors to correct for these effects when merging observed Bragg peak intensities. Under the assumption that all protein-derived x-ray scattering, whether into Bragg peaks or diffuse features, has intensity proportional to the illuminated crystal volume and subject to the same absorption patterns, the per-image scale factors and absorption correction factors reported by XDS were applied to the July 2015 diffuse images by the process described in this section.

When scaling together multiple runs (i.e., series of images), XSCALE, the scaling portion of XDS, produces an overall scale factor and B-factor for each run of images from a given crystal. Within each run, XDS and XSCALE each optionally compute three sets of multiplicative correction factors, designed to correct for absorption effects, detector modulation, and radiation damage, respectively. The absorption correction factors depend on image number and de-

tector position (from a set of 13 positions). Since many of the crystals used here had widths along the beam path that varied by 200-400 μm as the crystal rotated, absorption by the crystal also varied significantly. Computed absorption corrections for one typical crystal averaged to approximately 5%. When applied to the diffuse intensities themselves, this correction has a small influence, but it has a potentially large influence on resultant anisotropic diffuse intensity when applied to the isotropic background (e.g., scattering from water) underneath, which can be about 10 times as strong as the diffuse features in a given image (estimated by eye from a characteristic image).

It is worth noting that at this point in image processing, protein-derived diffuse scattering and solvent scattering have not been distinguished. Assuming that the solvent is approximately evenly distributed within and around the protein crystal, absorption patterns affecting the solvent scattering will match those affecting the protein scattering, so it is appropriate to apply the same absorption corrections to the solvent pattern as to the protein-derived intensities. The volume of solvent illuminated by the beam, and therefore the total amount of solvent-derived intensity per image, is not necessarily proportional to the volume of crystal, so the per-image scale factors may change the overall amount of solvent scattering in each image. However, since this scattering is isotropic, it will be subtracted later in its entirety; incorrect solvent scaling with an overall scale factor per image is therefore acceptable.

The detector modulation correction factors, which depend on the X and Y position on the detector and correct for uneven response to x-rays by the detector itself or absorption by anything fixed relative to the detector, were turned off. The detector itself is sufficiently well-calibrated to not require further cor-

rection, and the modulation correction factors are reported by XDS at too coarse an interval to capture the absorption effects of tape in the beam path (see section 3.2.6 for how the tape was actually corrected for).

The "decay" correction factors, designed to account for radiation damage, depend on image number (or, equivalently, the rotation angle phi) and resolution. Since it is neither obvious nor yet demonstrated that structured diffuse intensity responds to radiation damage in the same way that Bragg peaks do, applying a decay correction to diffuse data might obscure effects of interest, so this correction was turned off when running XDS and XSCALE to obtain scales and correction factors for diffuse data processing. However, even with this correction off, XDS still computes a decay correction factor based on image number alone, to correct for changing illuminated crystal volume as the crystal rotates; these correction factors were used in diffuse image processing along with the absorption factors and the overall per-run scale factor and B-factor.

The absorption and decay correction factors were only reported for bins of a certain size along each dependent variable dimension (i.e., image number or detector region), so to estimate an appropriate correction value for every pixel in every image, they were interpolated (and extrapolated) to fill in the entire range of the variable using methods from Python's scipy interpolation sub-package (see Appendix B for details). Both types of corrections are expected to vary smoothly, since crystal volume and shape varies smoothly.

The overall value adjustment for a given pixel is thus given by

$$\frac{Ke^{B\left(\frac{2sin(\theta)}{\lambda}\right)^2}}{D \cdot A},$$

where K and B are read directly from XSCALE output; D and A are the interpo-

lated decay and absorption correction factors; θ is the scattering angle; and λ is the wavelength.

By eye, application of these correction factors appears to smooth out the overall intensity variation across a single image (see Figure 3.6) and across a series of images. Their effect on the symmetry of the three-dimensional map from **lyso6** (see below for how these were constructed) was tested by computing a Pearson correlation coefficient between the unsymmetrized and symmetrized map with and without correction factors, and by computing the mean difference between the unsymmetrized and symmetrized map, scaled by the mean intensity in the map, with and without correction factors. By both metrics, the symmetry increased only negligibly. This raises some suspicions about the accuracy of these scales; however, since their output looks reasonable by eye and the assumptions underlying their appropriateness appear to be sound, they were maintained.

For the February 2014 data, application of scales and absorption corrections from XDS by the above method was also attempted, but they consistently decreased three-dimensional map symmetry even through various iterations of the method details. One key relevant difference between the February 2014 data and the July 2015 data is that the room-temperature February 2014 crystals were translated sideways every several frames, so they lack long series of images under which scale factors are expected to vary smoothly. While it is possible to obtain scale factors for each small continuous series of images from XDS, these were insufficiently accurate for the diffuse intensities, producing "corrected" images of sharply varying intensity. Therefore, this procedure was not used for any February 2014 crystals. Instead, to correct for overall diffuse intensity

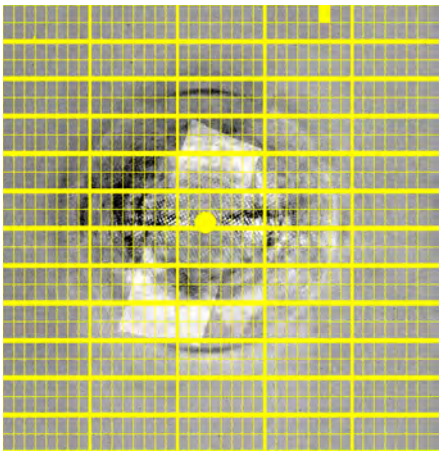
changes between images, a scale factor was computed at the final image processing step, described in section 3.2.7.

One could in principle further validate the above scaling procedure for the July 2015 data by reprocessing the data without XSCALE-derived factors and instead scaling the images the way the February 2014 images were scaled. This was not done here, but a previous poor version of the XSCALE scaling method was initially applied to all the data and retroactively found not to signfiicantly affect the main conclusions of this study. Therefore, it is reasonable to conclude that the details of the scaling method play a minor role in the final diffuse intensity maps. Better quality data (greater completeness and redundancy with less background scattering and fewer artifacts) would perhaps depend relatively more heavily on the details of scale factors and absorption corrections.

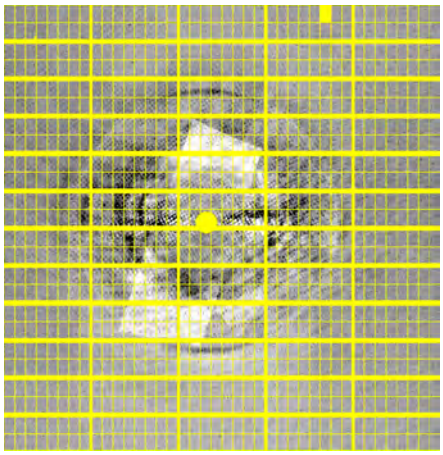
3.2.6 Absorption corrections: tape

As discussed in section 2.3.2, it was discovered during data processing that at least one piece of standard adhesive tape was attached to the beamstop holder during both February 2014 and July 2015 data collection. The primary effect of the tape was to simply absorb some x-rays, reducing scattering intensity in a certain region, so to correct for this, intensities in the affected region of the detector were increased by a small factor chosen by hand. The tape shadow is not visible in individual images, but the shadow and its removal are visible upon averaging many images, as shown in Figure 3.6.

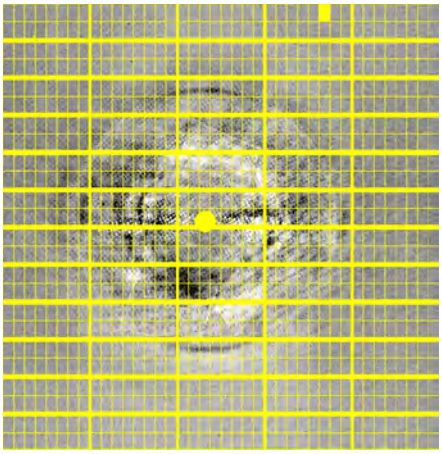
In the February 2014 data, one edge of one piece of tape also contributed significant additional scattering to a small region of every image. Since the ad-



(a) No absorption correction after background subtraction



(b) XSCALE absorption corrections



(c) XSCALE absorption correction and tape correction

Figure 3.6: Visualization of the effects of absorption correction and tape correction. Each image is constructed by averaging the first 99 images from **lyso6**, then subtracting the isotropic component to highlight features. In this averaging, actual signal that varies with crystal rotation is smoothed out, and artifacts remain. (a) Images after fiber background subtraction (and previous manipulations). (b) Images after XSCALE-based absorption corrections. The broad dark/light gradient in the background is flatter than before. Much of the absorption effects will have already been averaged out in (a), so individual images may show a larger proportional change, but this is difficult to see by eye due to the low signal per image. (c) Images after XSCALE-based absorption corrections and tape correction. Note the near absence of the bright white tape shadow. The dark features at top and bottom are remnants of fiber background scattering; other features close to the horizontal midline of the image (corresponding to the rotation axis) and the Bragg-like pattern of spots (sweeping arcs most visible on the right) are artifacts of averaging images after Bragg peak masking but before isotropic subtraction.

ditional scattering was complex and contained in a small region, that region of every image was simply masked and not included in further processing.

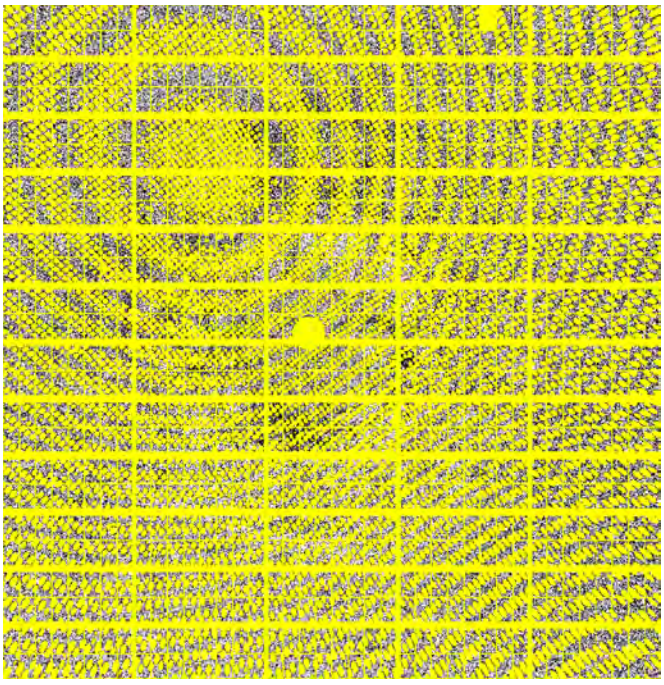
3.2.7 Isotropic signal removal and solid-angle normalization

To extract anisotropic intensities, from each image a one-dimensional radial average trace was computed, smoothed via local averaging to reduce noise, and subtracted from the image. This removed solvent scattering as well as the isotropic component of protein-derived diffuse scattering.

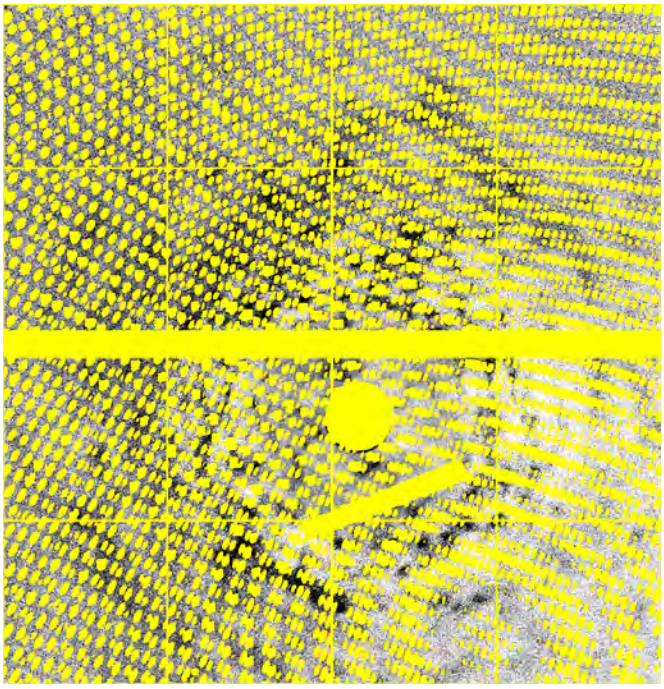
To facilitate assembly of three-dimensional maps, a solid-angle normalization was then applied to each image using the *normim* method in Lunus, detailed in [Wall, 2009]. This normalization corrects for the fact that different pixels on the detector subtend different solid angles around the scattering sample. Therefore, a sample that scatters x-rays in a truly spherically uniform pattern will not produce a uniform, flat image on a flat detector. The solid-angle normalization adjusts pixel values so that they are, in fact, proportional to the intensity expected at the corresponding location in reciprocal space (in other words, proportional to $|F(q)|^2$, the square of the crystal structure factor).

3.3 Diffuse data: sparser and denser three-dimensional lattices

To compile the images from a given crystal into a three-dimensional lattice, every pixel in every image was first mapped to a voxel in reciprocal space using the pixel's coordinates on the detector, the crystal orientation and unit cell structure as reported by XDS, the crystal-detector distance, and the x-ray wave-



(a)



(b)

Figure 3.7: Example images from (a) **lyso6** and (b) **lyso47** after all described processing steps, just prior to 3D integration.

length. The intensity from that pixel was then added to the corresponding voxel, and a counter for that voxel was incremented. Once all pixels in all images had been processed, two lattices were written to file: one unsymmetrized and one symmetrized. For the unsymmetrized lattice, voxel values were computed by dividing the cumulative sum of intensities by the counter value, i.e. by averaging all individual pixel intensities corresponding to that voxel. Voxels with too few corresponding pixels were left blank.

The symmetrized lattice was calculated by summing the intensities and pixel counts of all symmetry-equivalent voxels, and dividing as described above. Diffuse intensity due to correlated displacements within and between protein molecules is theoretically expected to exhibit the same symmetry as scattering from the average molecular structure [Welberry, 2004]. Such consistency in symmetry has been observed experimentally in other crystals by correlating diffuse maps with their symmetry-averaged versions [Peck et al., 2018], and the author has qualitatively observed such symmetry in the diffuse scattering from well-oriented lysozyme crystals as shown in Figure 3.8. Therefore, it is expected that a perfect three-dimensional map of the diffuse scattering from a tetragonal crystal will exhibit a high degree of symmetry. It is also expected that averaging over symmetry-equivalent voxels will improve the accuracy of measurements of protein-derived diffuse intensity by reducing relative noise while reducing the intensity of any artifacts that do not obey crystal symmetry.

Three-dimensional maps from these crystals clearly exhibit symmetric features before symmetry-averaging, but the symmetry is far from perfect, particularly with regards to the relative intensity of features. An example slice[3]

[3]Slices will be referred to by their Miller indices (H,K,L), which give locations along the (a*,b*,c*) axes of recicprocal space. (H,K,L) = (0,0,0) is the center of reciprocal space, representing no deflection of the incident beam, and integer Miller indices correspond to the possible

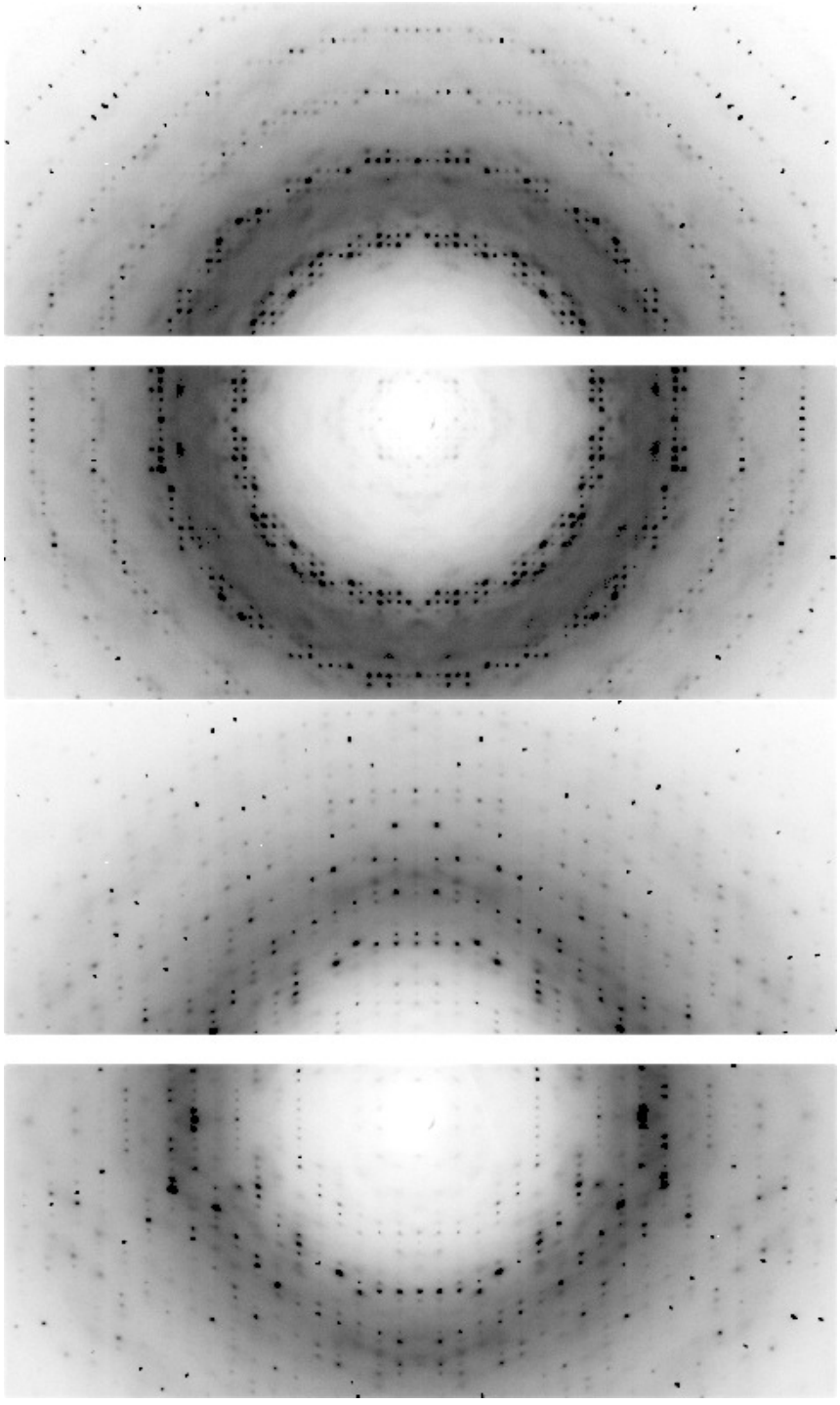


Figure 3.8: Raw diffraction images oriented down two symmetry axes of a tetragonal HEWL crystal (not one of the ones otherwise used in this work). The visible diffuse scattering obeys the symmetry evident in the Bragg peaks. The crystal was mounted on a precise two-axis goniometer stage, which allowed manual adjustment of the crystal orientation to image a symmetry plane.

through an unsymmetrized map is shown in Figure 3.9. Symmetry averaging not only improves accuracy but also fills in the portions of reciprocal space not sampled by the original images. The final maps therefore ideally contain intensities that are spherically anisotropic but tetragonally symmetric according to the specific symmetry rules of the crystals' space group.

For each crystal, lattices at two densities were constructed: a denser lattice with voxel side lengths of 1/4 Miller indices, yielding 64 voxels per Bragg peak, and a sparser lattice with voxel side lengths of whole Miller indices, yielding one voxel per Bragg peak. The denser lattice allows observation of smaller features, while the sparser lattice is less noisy. Figure 3.10 shows an example of this effect in **lyso19**, one of the cryocooled crystals from February 2014: horizontal streaking is evident in the denser lattice when viewed down the K axis, but it cannot be seen in the sparser lattice. While these streaks may indicate insufficient masking of Bragg peaks, it is interesting that they only appear in L planes, perpendicular to the largest reciprocal space dimension, and not uniformly–compare to the view down the L axis in Figure 3.11.

3.3.1 Validation

Before comparing the diffuse maps from different crystals, a few checks were performed on the quality of individual maps. All checks involved computing resolution-dependent Pearson correlation coefficients, described further in section 4.1.1. Checks were performed on three arbitrarily chosen representative crystals, all in the unbound state: **lyso8** (July 2015), **lyso19** (February 2014, cryocooled), and **lyso34** (February 2014, room temperature, no EtGly).

locations of Bragg peaks.

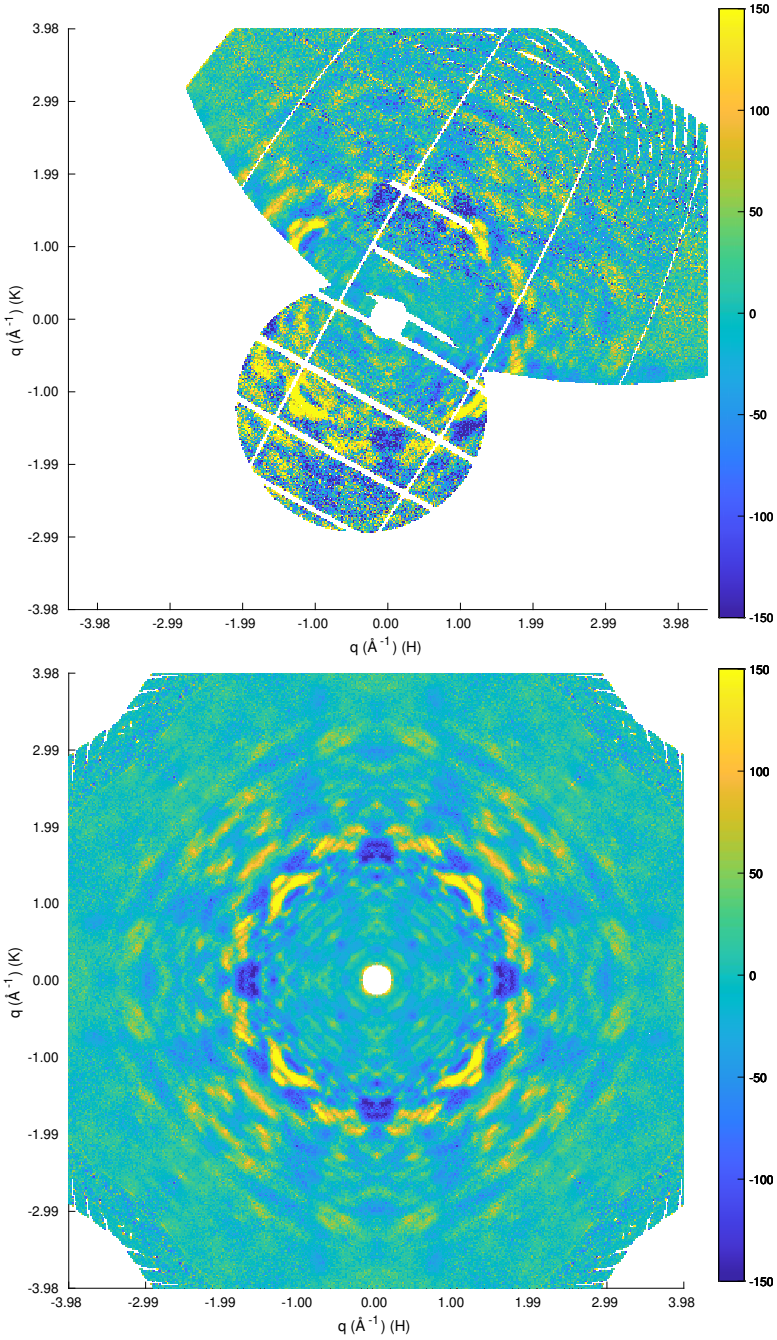


Figure 3.9: Slices at L = 0.5 through the denser (one voxel every 1/4 Miller index) map of **lyso8** (July 2015 data, unbound). Top: Before symmetry-averaging. Symmetry across vertical and horizontal mirror planes through the center of the map is visually evident, but imperfect. Bottom: After symmetry-averaging. Intensities are in arbitrary units.

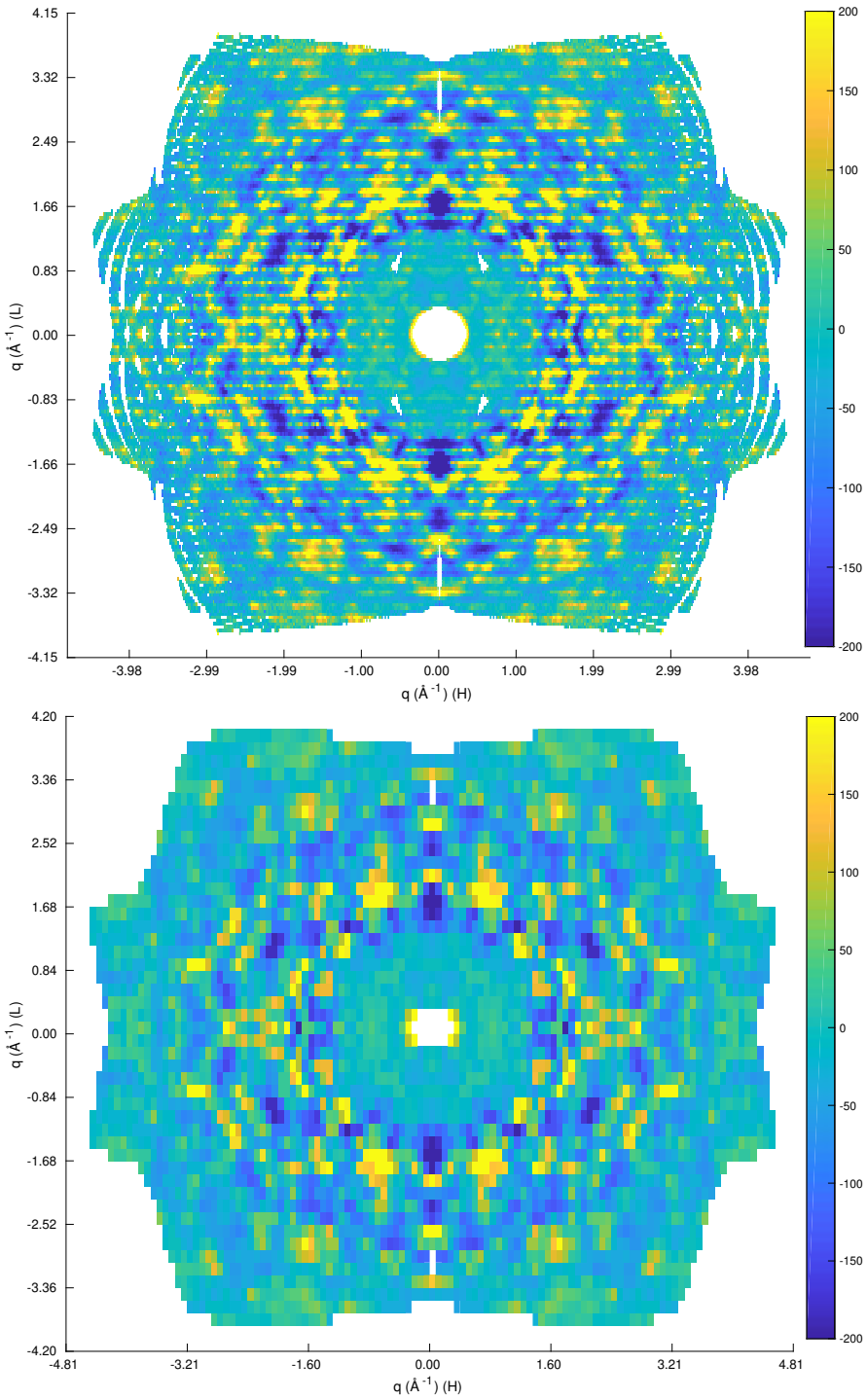


Figure 3.10: Slices at K = 0.5 (top, denser lattice) and K = 0 (bottom, sparser lattice) through the diffuse lattices from **lyso19** (cryocooled February 2014 crystal). Intensity is in arbitrary units. Horizontal streaks visible in the denser lattice are not at all evident in the sparser lattice with only one voxel per integer Miller index. Intensities are denoted by color; scale is arbitrary but consistent between both images.

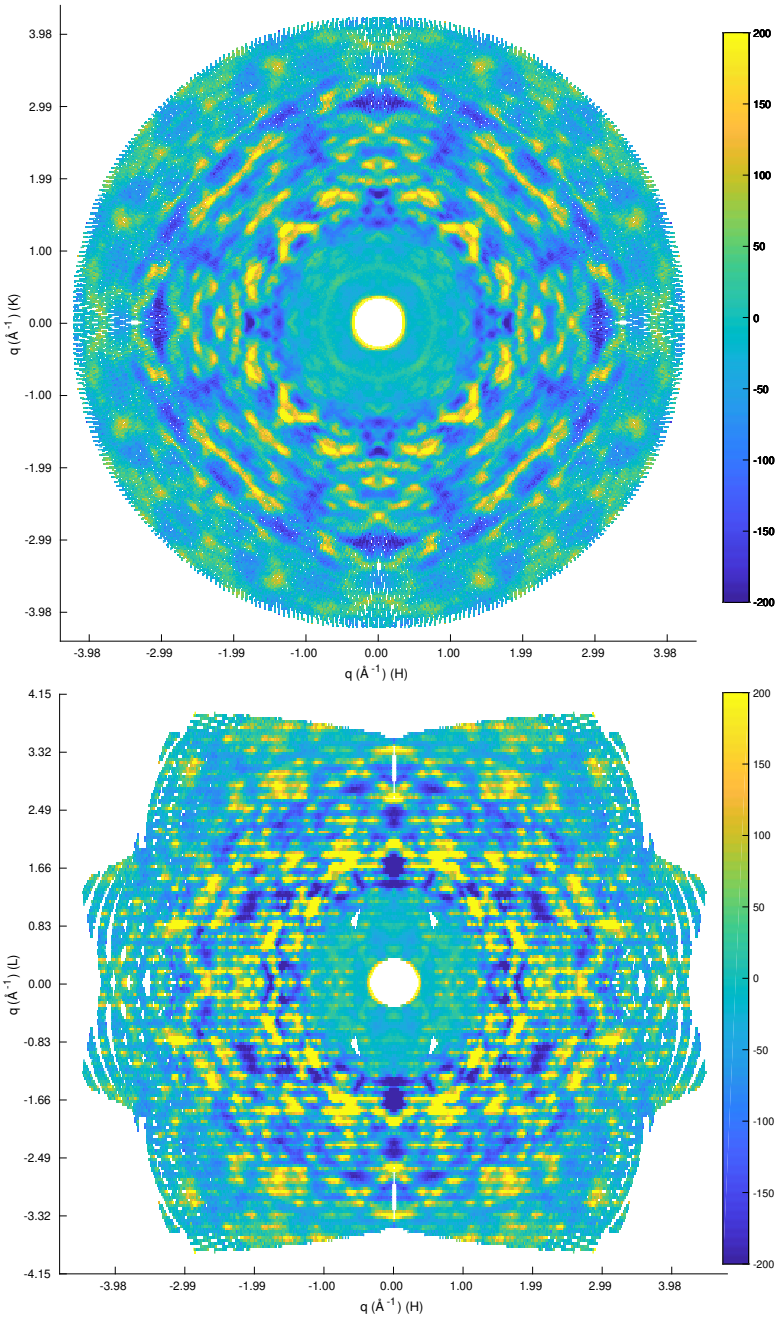


Figure 3.11: Slices through the denser **lyso19** lattice at L = 0.5 (top) and K = 0.5 (bottom).

First, to check that the Bragg peaks were appropriately masked and not somehow dominating the resulting diffuse intensities, a three-dimensional lattice of Bragg intensities was constructed from the integrated and merged peak intensities reported by XSCALE for each of three crystals. By necessity, these maps had only one voxel per integer Miller index. A resolution-dependent Pearson correlation coefficient was then computed between the Bragg map and the lower-density diffuse map for each of the three crystals. Results are presented in Figure 3.12. Some significant correlation persists between the pattern of Bragg peaks and that of diffuse intensities; this could be due to spreading of Bragg peaks beyond their masks, but it could also be that true diffuse intensity—that is, that which excludes the squared Fourier transform of the average molecular electron density—is not completely unrelated to Bragg intensity. In any case, these correlation values, around 0.5, are sufficiently low to confirm that the diffuse maps are at least not dominated by Bragg intensities.

Next, to examine the symmetry of diffuse maps prior to averaging, the unsymmetrized maps from each of these three crystals were split along the L = 0 plane, and a correlation coefficient was computed between the two halves. Results for both the denser and sparser lattices are displayed in Figure 3.13. With the exception of the denser **lyso8** lattice, the correlations all start out fairly high at low resolution and drop off at varying rates with increasing resolution. The cryocooled crystal, **lyso19**, maintains a fairly high correlation between its symmetry-related halves even in the denser lattice, with CC > 0.7 in all but one resolution shell. The denser **lyso8** lattice exhibits a surprisingly low level of correlation, indicating the power of averaging over both pixels and symmetry-related voxels in improving data quality.

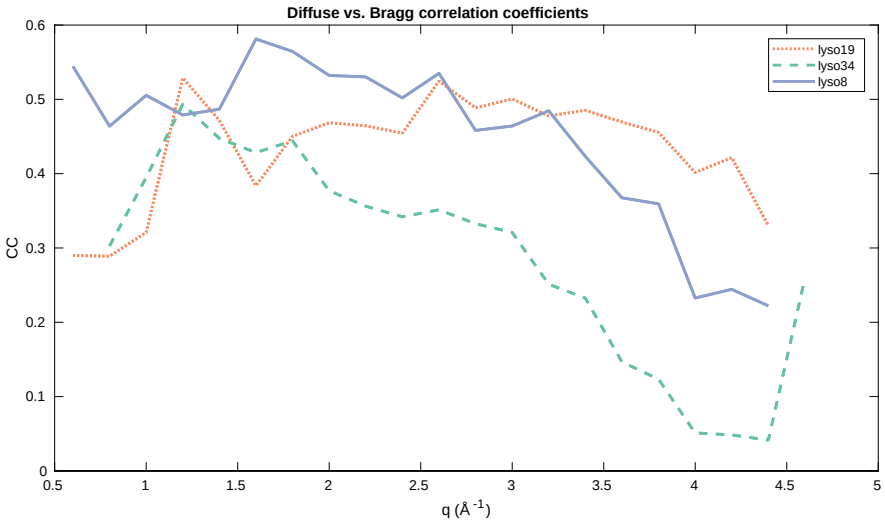


Figure 3.12: Correlation coefficients between three-dimensional diffuse maps and the corresponding Bragg-inclusive maps for each of three unbound crystals (**lyso19**: cryocooled, Feb 2014; **lyso34**: room temp, no EtGly, Feb 2014; **lyso8**: July 2015). All are strictly below 0.6 at all resolutions, indicating that Bragg patterns do not dominate the diffuse intensity.

Finally, for each of the same crystals, the input image set was split in half by alternating images, and a symmetry-averaged lattice was constructed from each half of the dataset. Since there is no reason for systematic variation between a pair of lattices constructed in this way, correlation coefficients below one may here be taken as a measure of the noise inherent in the data. They also provide an estimate of the maximum correlation one might expect to be able to observe between diffuse lattices from different crystals. Correlations between lattices constructed from alternating images are presented in Figure 3.14. The correlations between sparser lattices from interleaved sets of images are extremely

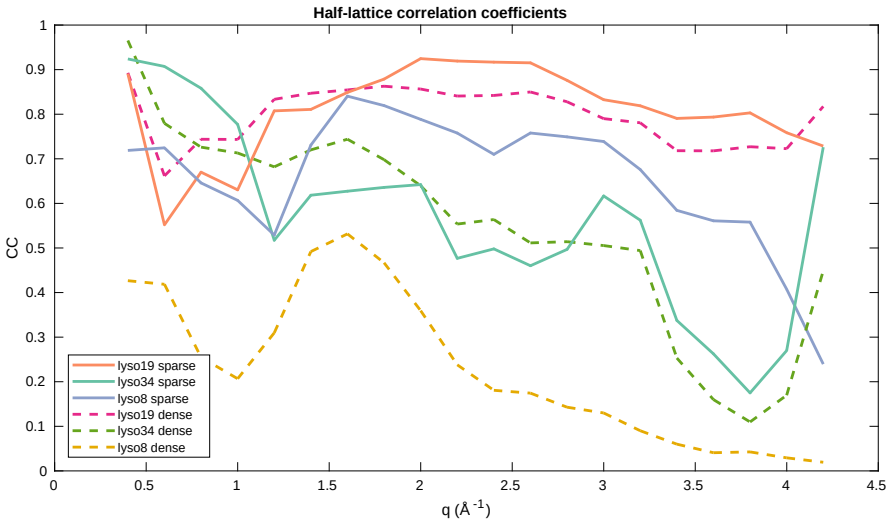


Figure 3.13: Correlation coefficients between the $L \geq 0$ and $L \leq 0$ halves of unsymmetrized three-dimensional maps for each of three unbound crystals (**lyso19**: cryocooled, Feb 2014; **lyso34**: room temp, no EtGly, Feb 2014; **lyso8**: July 2015). All correlations are resonably high in at least some resolution range except for the denser **lyso8** map, which likely has poor cumulative signal strength prior to symmetry-averaging.

high for all crystals. In the denser lattices, correlations in **lyso19** remain higher than for the other two lattices at all resolutions. One contributing factor could be the distinctive streaky features in **lyso19**'s diffuse map (Figure 3.10), which are not present in either **lyso34** or **lyso8** and may be a source of increased signal consistency at higher map density.

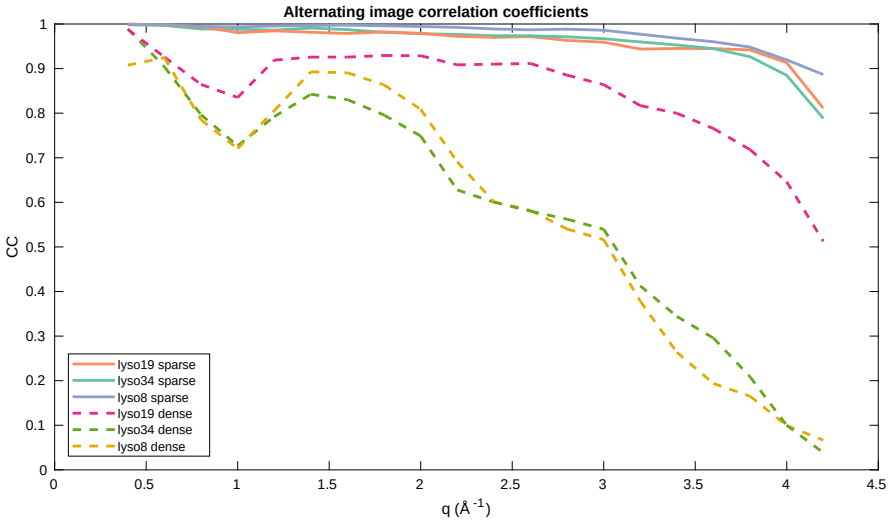


Figure 3.14: Correlation coefficients between symmetrized three-dimensional maps made from alternating halves of the input image series for each of three unbound crystals (**lyso19**: cryocooled, Feb 2014; **lyso34**: room temp, no EtGly, Feb 2014; **lyso8**: July 2015). All correlations are fairly high in the denser maps and nearly 1.0 in the sparser maps, as desired.

CHAPTER 4

COMPARISON OF DATA SETS

A small variety of calculations were performed on the finalized three-dimensional diffuse maps in order to evaluate the reproducibility of signal between ostensibly identical crystals and to deduce the effects of inhibitor binding and cryocooling on diffuse x-ray intensity. Section 4.1 explains the metrics used, and the subsequent three sections present the results of their use to address the aforementioned questions.

4.1 Metrics

4.1.1 Correlation coefficients

Whole-map correlation coefficients appear to be as much the norm as anything for comparing experimental diffuse data to diffuse maps computed from theoretical models, so correlation coefficients are an obvious choice for comparing multiple diffuse data sets. Here, the lunus function *ccrlt*[1] was used to compute Pearson correlation coefficients between three-dimensional maps. The Pearson correlation coefficient between sets A and B is defined as

$$\mathrm{CC} = \frac{\langle (A - \mu_A)(B - \mu_B) \rangle}{\sigma_A \sigma_B}$$

where μ_A and μ_B are the means of A and B, respectively, and σ_A and σ_B are the standard deviations of A and B, respectively. The function *ccrlt* reports a separate correlation coefficient for each resolution shell of width ≈ 0.02 Å^{-1};

[1] Lightly edited from the published version to handle multiple voxels per unit hkl and to report mean values in resolution shells.

these resolution-dependent coefficients were preferred over whole-map correlations because a) the qualitative characteristics of the diffuse intensity pattern are highly resolution-dependent and b) the amount of information averaged into individual voxels also varies greatly with resolution, and summing and averaging improves correlations.

Upper bounds on correlation coefficients

Protein crystal diffraction data is inherently noisy. Two data sets with the same underlying pattern but different Poisson noise will have a Pearson correlation coefficient somewhat below 1; the actual value depends on parameters of the data. To estimate the limits this places on our results, simulations were run to determine the correlation coefficient between two Poisson-noisy distributions drawn from the same underlying pattern. The underlying pattern was a 1000-point line of baseline value M with four Gaussian features (positive or negative) giving an average deviation from M of μM. In terms of real crystal data, M is the mean intensity (isotropic intensity) in a given resolution shell and μ is the ratio of diffuse feature intensity to isotropic intensity. Two forms of averaging were included in the simulation: α_a neighboring elements in a single line were averaged together, creating samples of $1000/\alpha_a$ points apiece, and then α_b independent lines of data were averaged together. In terms of real crystal data, α_a represents the summing of neighboring pixels in a two-dimensional image (or multiple images, in the case of multiple runs covering the same rotation angle) into a single voxel, and α_b represents the averaging of different voxels together during symmetry-averaging. These two forms of averaging do not produce distinguishable effects in the simulations, so they were combined into a single

M	μ	α	CC	Correspondence
200	0.025	40	0.761	**lyso34**, low intensity
500	0.025	40	0.888	...
1000	0.025	40	0.941	...
1500	0.025	40	0.961	**lyso34**, high intensity
10	0.04	2500	0.962	**lyso6**, low intensity
20	0.04	2500	0.979	...
30	0.04	2500	0.989	**lyso6**, high intensity

Table 4.1: Correlation coefficient limits due to Poisson noise as a function of relevant parameters, based on simulations. M is the baseline (nearly mean) value of the simulated data, μ is the ratio of feature size to baseline, and α is the number of points with independent Poisson noise averaged together before computing CC, the Pearson correlation coefficient.

parameter $\alpha = \alpha_a \alpha_b$.

In each simulation, Poisson noise was added independently to copies of the baseline data, the copies were averaged as described above, and then the correlation coefficient between two independently created and averaged results was calculated. Increasing any of the three parameters M, μ, and α increased the resulting correlation coefficient when the other parameters were held constant.

Results are shown in Table 4.1. The parameters in the first four rows and last three rows correspond roughly to the characteristics of the symmetrized (but not further averaged) maps generated from **lyso34** data (February 2014) and **lyso6** data (July 2015), respectively, at different resolutions. Parameters are far more similar between crystals in a single CHESS run than they are between the two primary CHESS runs discussed in this document because of the consistency in beam flux and data collection strategy within each CHESS run; these crystals can reasonably be taken as representatives of their respective CHESS runs for this purpose. Note that μ is similar between the two crystals; differences are as likely to stem from differences in amount of mother liquor around the

crystal (which contributes isotropic diffuse intensity that was not distinguished from protein-derived isotropic diffuse intensity) as from meaningful differences in amount of protein-derived anisotropic diffuse intensity. Values of M, corresponding to the mean diffuse intensity at some resolution in a single image, are much higher for **lyso34** because images were taken only once per degree of rotation with high exposure times; values of α are correspondingly higher for **lyso6** because there were ten images per degree rather than 1, multiple images taken at the same rotation angle, and overall a wider rotation range covered by the raw data. Overall, the increased averaging in the July 2015 is powerful, resulting in notably higher correlation coefficients in the simulations.

The correlation coefficients in the table should be taken as approximate upper bounds on the correlation coefficients measurable in the real data. In other words, two different crystals whose symmetrized diffuse maps correlate with coefficients similar to the corresponding ones in this table can be considered identical to within available precision.

4.1.2 Diffuse:Bragg intensity ratios

Two of the hypotheses tested in this dissertation concerned a change in relative amount of diffuse scattering under different conditions (see Section 1.3). Therefore, one important metric by which to evaluate three-dimensional maps is the amount of diffuse intensity relative to the amount of Bragg diffraction. Both types of scattering should be affected identically by changes in beam intensity, illuminated crystal volume, and absorption effects, so the ratio between them ideally reveals something about the crystal's level of structured disorder.

As with the correlation coefficients, Bragg:diffuse intensity ratios were calculated in resolution shells. The diffuse intensity was calculated by averaging the absolute value of all voxels in the denser three-dimensional diffuse map in a given resolution shell. The Bragg intensity was calculated by summing the intensities of Bragg peaks within the same resolution shell as reported by XSCALE, then dividing the result by the volume of that resolution shell to obtain an average. The intensities reported by XSCALE have been polarization-corrected, background-subtracted, and adjusted by similar correction factors to those applied to the diffuse intensities (see section 3.2.5); that is, they have been through a comparable set of processing steps to the diffuse intensities. The volume of the resolution shell was calculated approximately in units of pixel-equivalents by calculating the shell volume in inverse cubic Angstroms and using the experimental parameters to convert between inverse Angstroms and pixels on the detector face; one pixel-to-inverse Angstroms conversion factor was used for all July 2015 data and another was used for all February 2014 data.

In this way, Bragg intensities and diffuse intensities for a single crystal were placed on approximately, though not precisely, the same scale. The ratios of diffuse intensity to Bragg intensity can therefore be taken as roughly the ratio of total amount of anisotropic diffuse scattering to total amount of Bragg scattering within a resolution shell. Factors that cause these ratios to be inexact (different correction schemes, imprecision in volume scaling factors, overcounting of random variance as anisotropic diffuse scatter) should be similar across crystals from the same CHESS run, so these ratios are most useful as a comparison metric between crystals, rather than as way to characterize individual crystals.

4.2 Reproducibility of diffuse scatter among ostensibly identical crystals

The most fundamental issue addressed in this work is whether measured diffuse intensity is reproducible from crystal to crystal. For an initial birds-eye view of the diffuse scattering behavior in each crystal, Figure 4.1 shows average Bragg and diffuse intensities as a function of scattering vector magnitude for every individual crystal in the February 2014 and July 2015 data sets. Average intensities were calculated as described in section 4.1.2, without dividing one by the other. The approximate shape of each type of intensity curve looks similar across all crystals, as expected. (At low q, the curves for February data look qualitatively different from the curves for July data because the beamstop shadow continued out to higher resolution, affecting the q range shown in these graphs.) The July curves show much more apparent consistency than do the February curves, likely because more total signal was collected and averaged together in the July data sets. Note that though the relative magnitudes of Bragg and diffuse intensities appears different in the July data than in the February data, this is an artifact of estimating scale factors and signifies nothing of interest.

Next, diffuse maps are visually compared by displaying slices through the same planes in the three-dimensional lattice. Before comparing maps from different crystals, maps were constructed through the procedure described in the previous chapter for single sweeps through each of the July 2015 crystals, which each had multiple sweeps cover the same rotation range at different points along the crystal's length. Figure 4.2 shows slices through the L = 0.5 and K

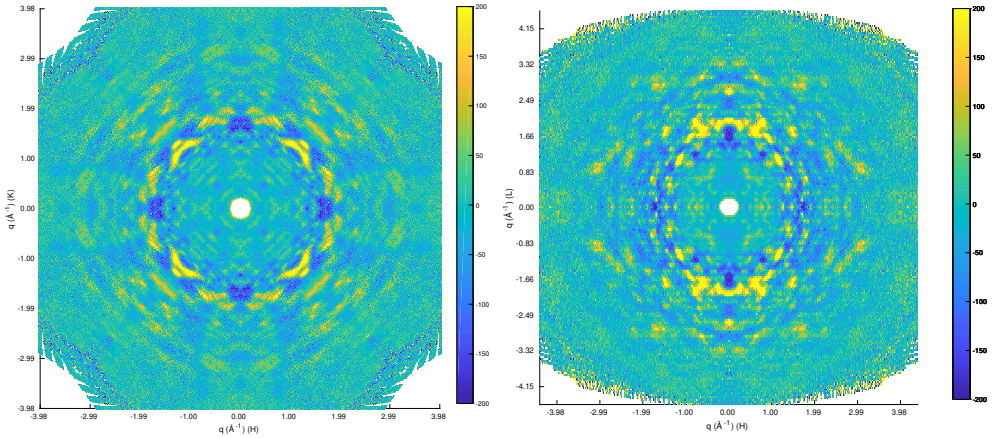
Figure 4.1: Mean Bragg and diffuse intensities in individual crystals. Vertical axes are arbitrary intensity units, scaled for approximately direct comparison between the two intensities in a given crystal. July data are **lyso6**, **lyso7**, **lyso8**, **lyso9**, and **lyso10** (bottom left); February data are all other crystals.

= 0.5 planes of the denser single-sweep maps from **lyso8**, an unbound crystal. These slices display excellent consistency, as expected.

To visually determine consistency between crystals, Figure 4.3 shows slices in the L = 0.5 plane through four room temperature, unbound crystals, two of which are from July 2015 and two from February 2014. By eye, the overall diffuse feature pattern is extremely similar between crystals out to approximately q = 3 Å.

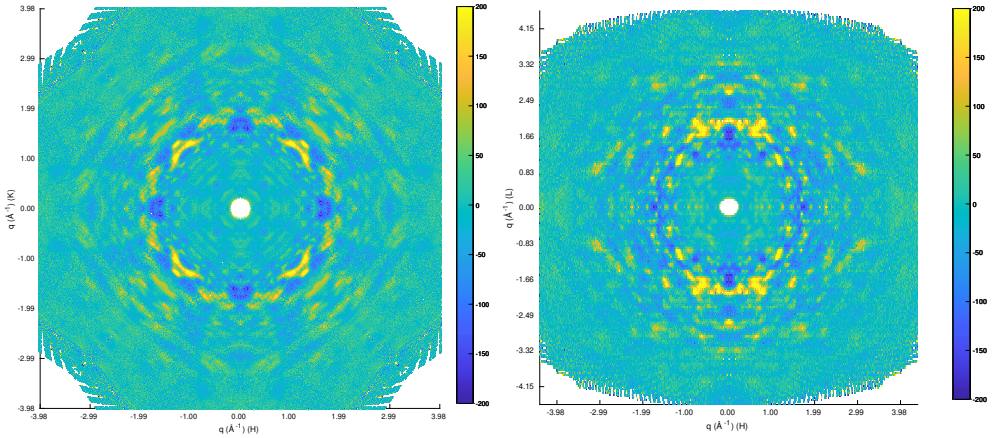
Figure 4.4 shows resolution-dependent correlation coefficients between pairs of unbound, room-temperature crystals from the same CHESS run and under ostensibly identical conditions. Among the denser lattices, correlations are comparable to, or better than, the self-consistency of room temperature lattices as reported in Figure 3.14. Among the sparser lattices, in the July data, correlations remain nearly as high as between alternating frames from the same crystal. Correlations between February crystals are lower but still reasonably high; they remain mostly above 0.8 out to a resolution of 2.5 Å.

These plots indicate that averaging neighboring voxels does not significantly improve cross-crystal correlations between February crystals, contrary to expectations. Reasons for this are not obvious; it is possible that unlike in the July data, inconsistency between same-condition denser diffuse maps from February data is dominated not by noise but by actual changes in the scattering pattern. One possibly important difference between the July and February data is that the temperature of the "room temperature" July crystals was held constant at 300 K, while the temperature of the room temperature February crystals was let to vary with the room itself. Room temperature near the sample holder was not measured at any time; based purely on memory of feel, it likely varied in



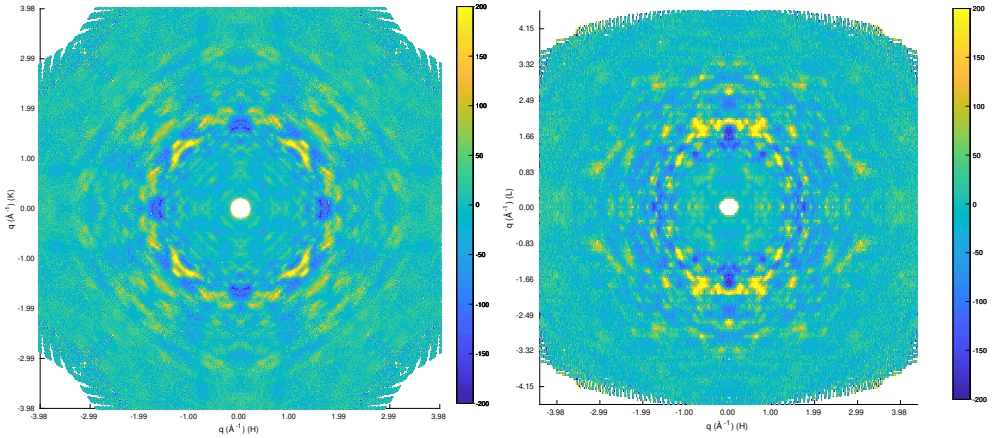
(a) Sweep 1, L = 0.5

(b) Sweep 1, K = 0.5



(c) Sweep 2, L = 0.5

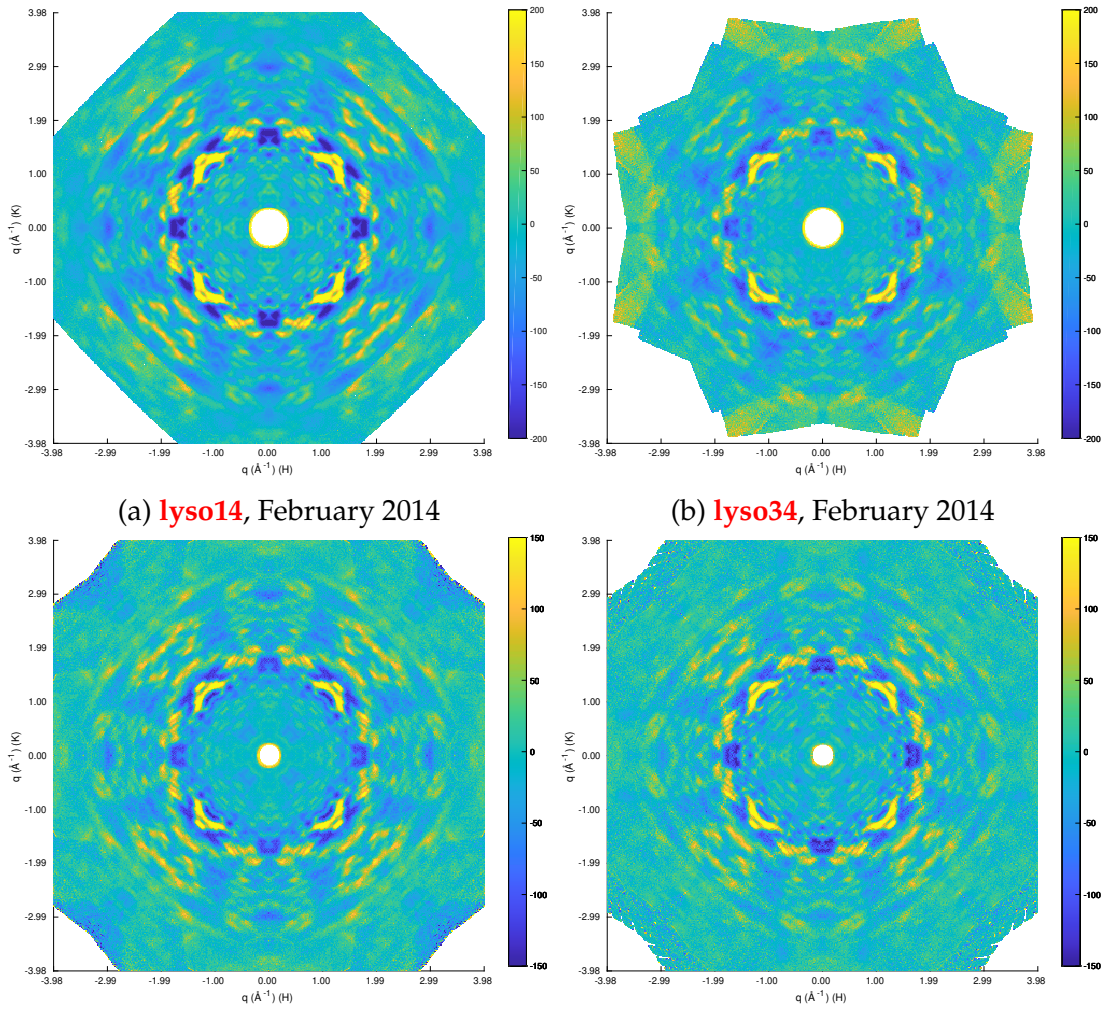
(d) Sweep 2, K = 0.5



(e) Sweep 3, L = 0.5

(f) Sweep 3, K = 0.5

Figure 4.2: Slices at L = 0.5 and K = 0.5 through denser single-sweep maps of **lyso8**, an unbound crystal from July 2015. Color-coded for intensity on an arbitrary scale, where yellow is positive and blue is negative. Maps are visually extremely similar across sweeps.



(a) **lyso14**, February 2014

(b) **lyso34**, February 2014

(c) **lyso6**, July 2015

(d) **lyso8**, July 2015

Figure 4.3: Slices at L = 0.5 through four room temperature, unbound crystals without EtGly. Color-coded for intensity on an arbitrary scale, where yellow is positive and blue is negative. Maps look extremely similar out to q = 3 Å, indicating good reproducibility between crystals.

a range of 3–5 K. Temperature is known to impact protein conformations, so temperature variation may have altered the diffuse scattering among February crystals.

Figure 4.5 shows resolution-dependent correlation coefficients between unbound room-temperature crystals from different CHESS runs, alongside the same-run correlations. While the July data correlates better with itself than with

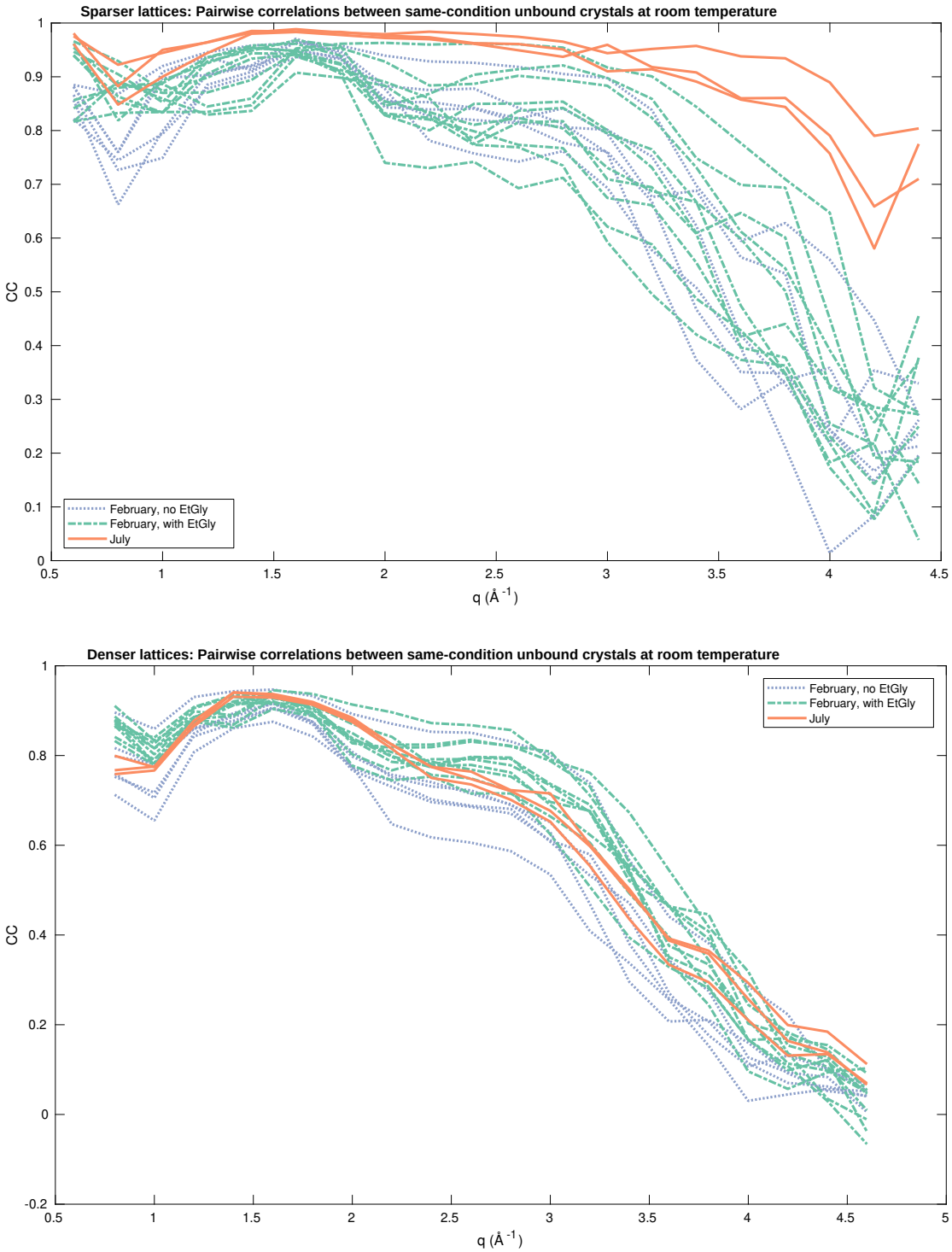


Figure 4.4: Correlation coefficients between individual pairs of unbound, room-temperature crystals under ostensibly identical conditions. Correlations are calculated within (but not between) the July crystals, the February crystals with EtGly, and the February crystals without EtGly. Top: correlations between sparser lattices. Bottom: correlations between denser lattices.

the February data, the February data correlates just as well with July data as with other February data. This supports the idea that there is more variation among the February room temperature crystals than among the July crystals, but it also suggests that there is no systematic difference in the February diffuse maps as compared to the July ones.

4.3 Effect of inhibitor binding

Correlation coefficients within the February 2014 room temperature crystals without ethylene glycol, bound and unbound to 3NAG, are shown in Figure 4.6. Correlations between the bound crystals are slightly but consistently higher than those involving either one or two unbound crystals, implying that binding to an inhibitor may restrict the scope of intramolecular motions contributing to diffuse scattering. However, as some pairs of crystals in different bindings conditions are, at some resolutions, better correlated than some pairs of crystals in the same binding condition, the effect may not be very strong. Figure 4.7 shows the same correlations computed among July 2015 crystals, where there is a sharp gap in correlations between same-binding-state and opposite-binding-state crystal pairs. Still, the drop in correlations for opposite-binding-state pairs is small.

Figure 4.8 provides one look at the distinction between the diffuse scattering pattern in an unbound and a bound crystal. The top image is the L = 0 slice of the sparser diffuse map of **lyso8**, an unbound crystal from the July 2015 data set. The sparser diffuse map of **lyso9**, a 3NAG-bound crystal from the July 2015 data set, was scaled to match the overall intensity of the **lyso8** map and then

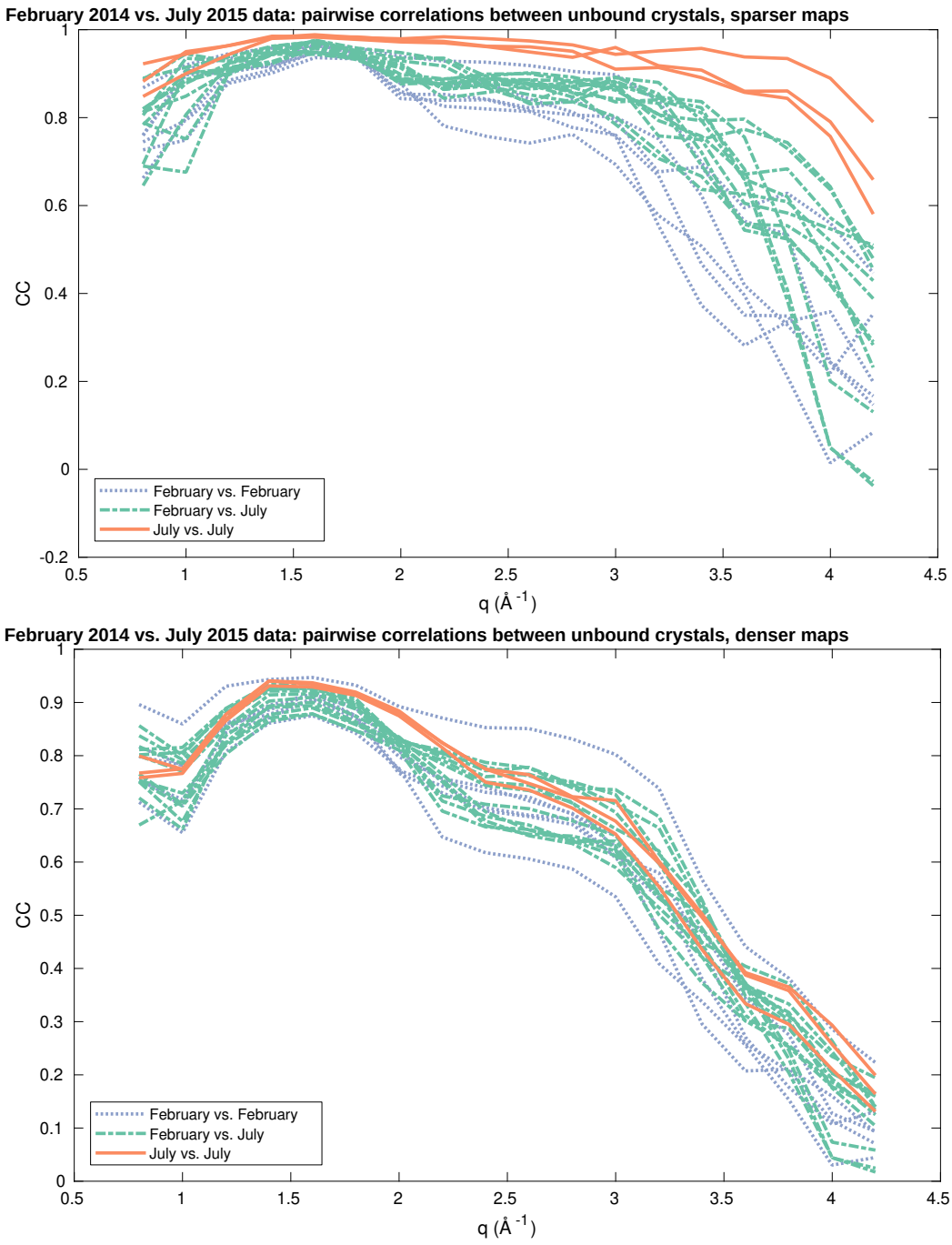


Figure 4.5: Correlation coefficients between individual pairs of unbound, room-temperature crystals without EtGly, sorted by CHESS run(s). Top: correlations between sparser lattices. Bottom: correlations between denser lattices.

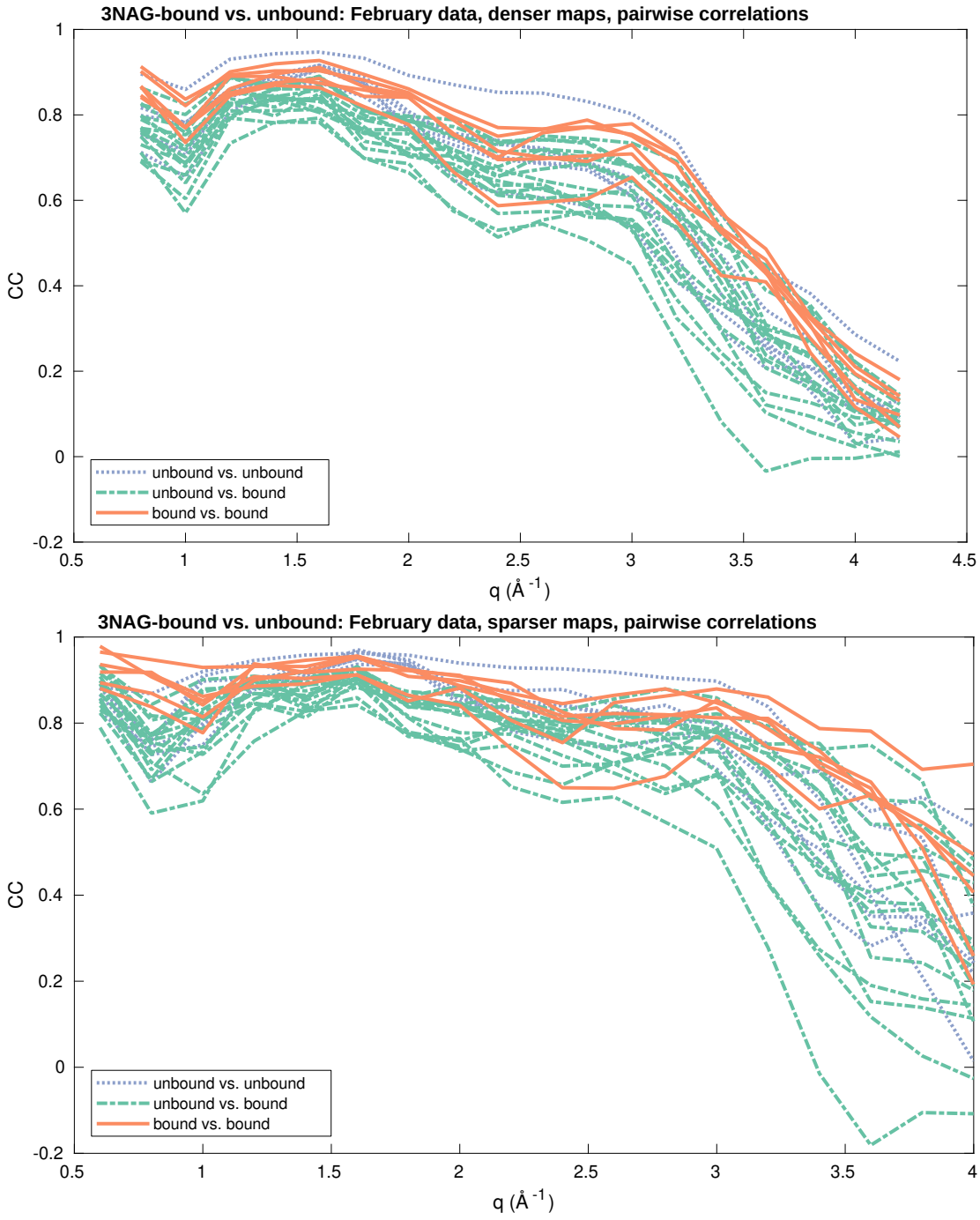


Figure 4.6: Correlation coefficients between single-crystal 3D maps as a function of scattering vector, within the February 2014 data set. All crystals are room-temperature without EtGly; correlations are color-coded by 3NAG binding states. Top: correlations between denser maps. Bottom: correlations between sparser maps.

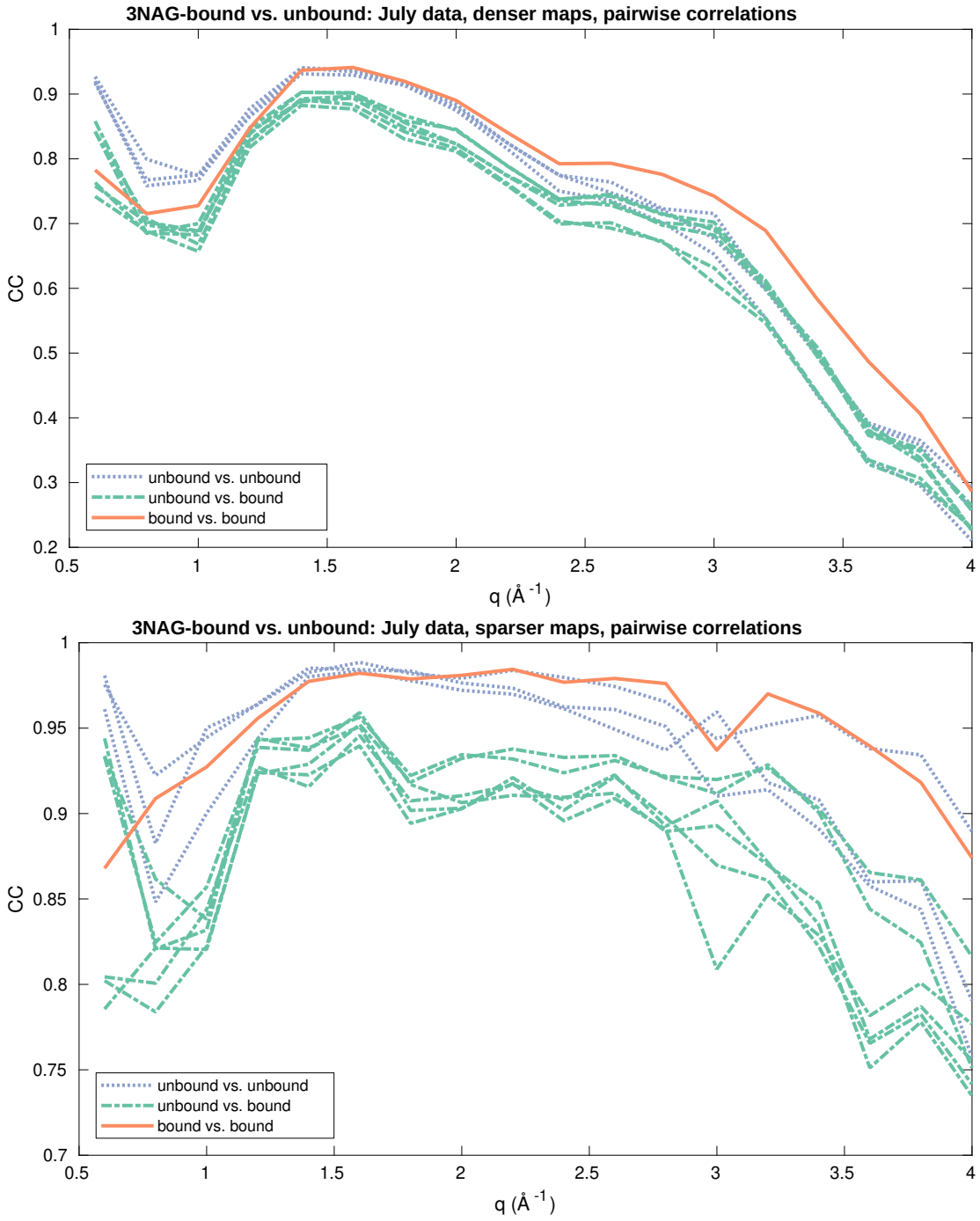


Figure 4.7: Correlation coefficients between single-crystal 3D maps as a function of scattering vector, within the July 2015 data set. All crystals are room-temperature without EtGly; correlations are color-coded by 3NAG binding states. Top: correlations between denser maps. Bottom: correlations between sparser maps.

subtracted from it. The bottom image in Figure 4.8 shows the L = 0 slice of the result. If the maps exhibited the same diffuse pattern, one would expect the difference map to be flat, or, if scaling were poor, to be a weaker (and possibly inverted) version of the pattern in the top image. However, what actually appears is a pattern sharing only some of its features with the original map. In contrast, a difference map created by subtracting **lyso9** from **lyso10**, the other 3NAG-bound July 2015 crystal, looks much flatter (Figure 4.9). This suggests that binding lysozyme to the inhibitor 3NAG does alter diffuse scattering in a small yet consistent way.

For closer comparison with the L = 0.5 and K = 0.5 maps in section 4.2, Figures 4.10 and 4.11 show difference maps constructed as described above, but from the denser lattices. Differences are weaker in these slices (which are essentially one quarter the thickness of the denser map slices), but the additional difference between the bound and unbound crystals remains.

Since the February 2014 data also contained crystals with and without inhibitor, similar example maps and difference maps are displayed in Figures 4.12 and 4.13. As in the July 2015 data, the difference map between the unbound and bound crystals shows more structured diffuse intensity than the difference map between two bound crystals. However, both difference maps also show some characteristics of uneven absorption effects, particularly the broad lighter and darker regions in the **lyso47** minus **lyso34** map and the long, thin, faint lines in the L = 0.5 slice of the **lyso47** minus **lyso38** map. These features are more likely artifacts of the data collection methods rather than meaningful diffuse scattering signatures.

The hypothesis regarding inhibitor binding was not merely that binding af-

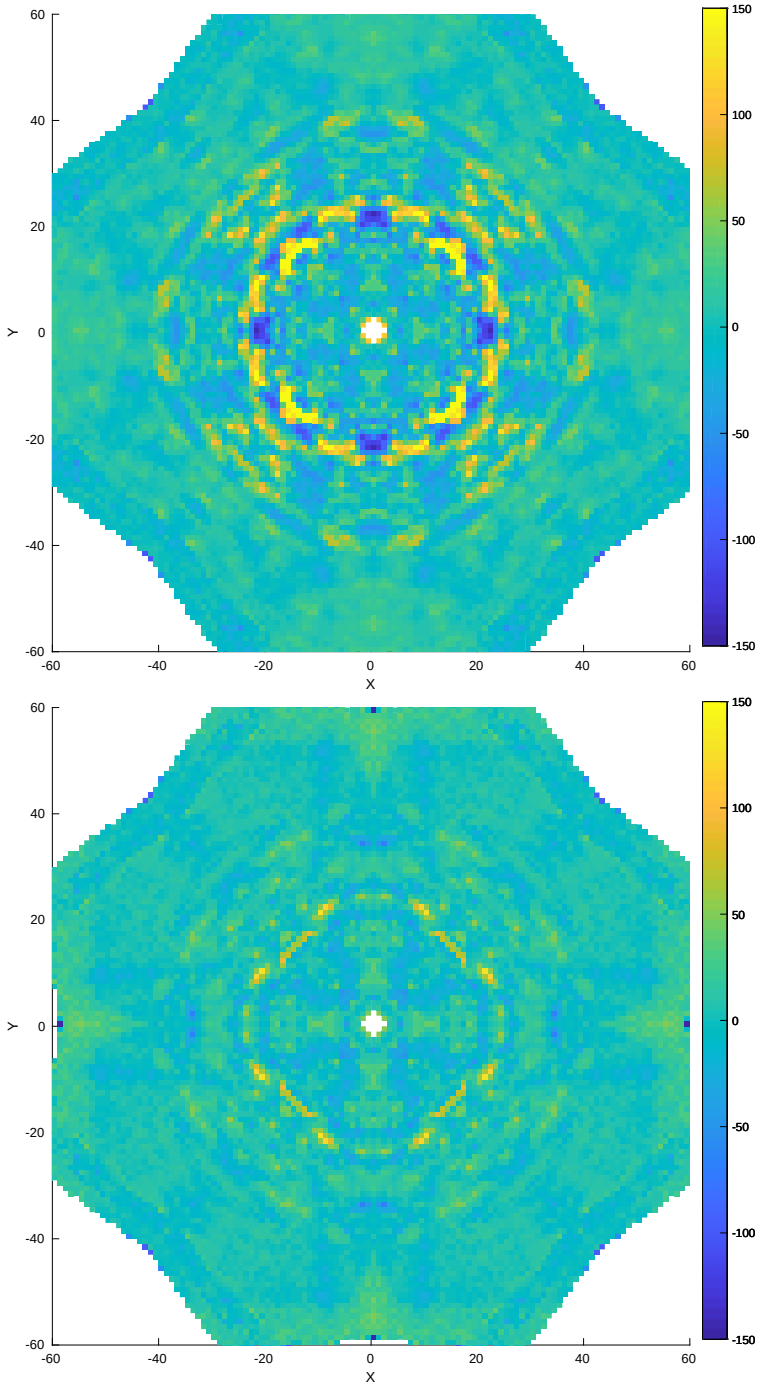


Figure 4.8: Top: L = 0 slice of the sparser diffuse map of **lyso8** (July 2015 unbound crystal). Bottom: L = 0 slice of the difference between the sparser diffuse maps of **lyso8** and **lyso9** (July 2015 3NAG-bound). An overall scale factor was applied to put the two maps on approximately the same scale before subtracting **lyso9** from **lyso8**. Display is on an arbitrary but consistent scale, where yellow denotes positive intensity and blue negative.

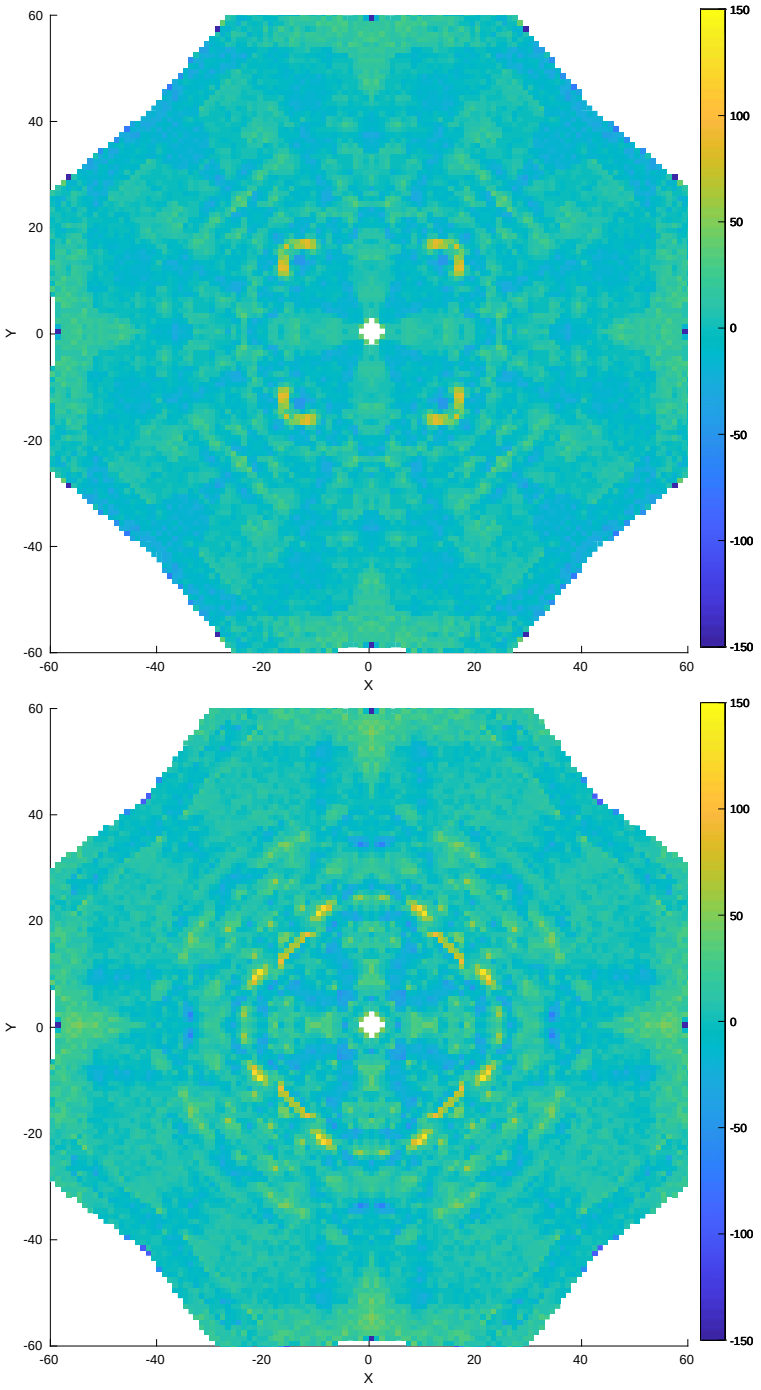
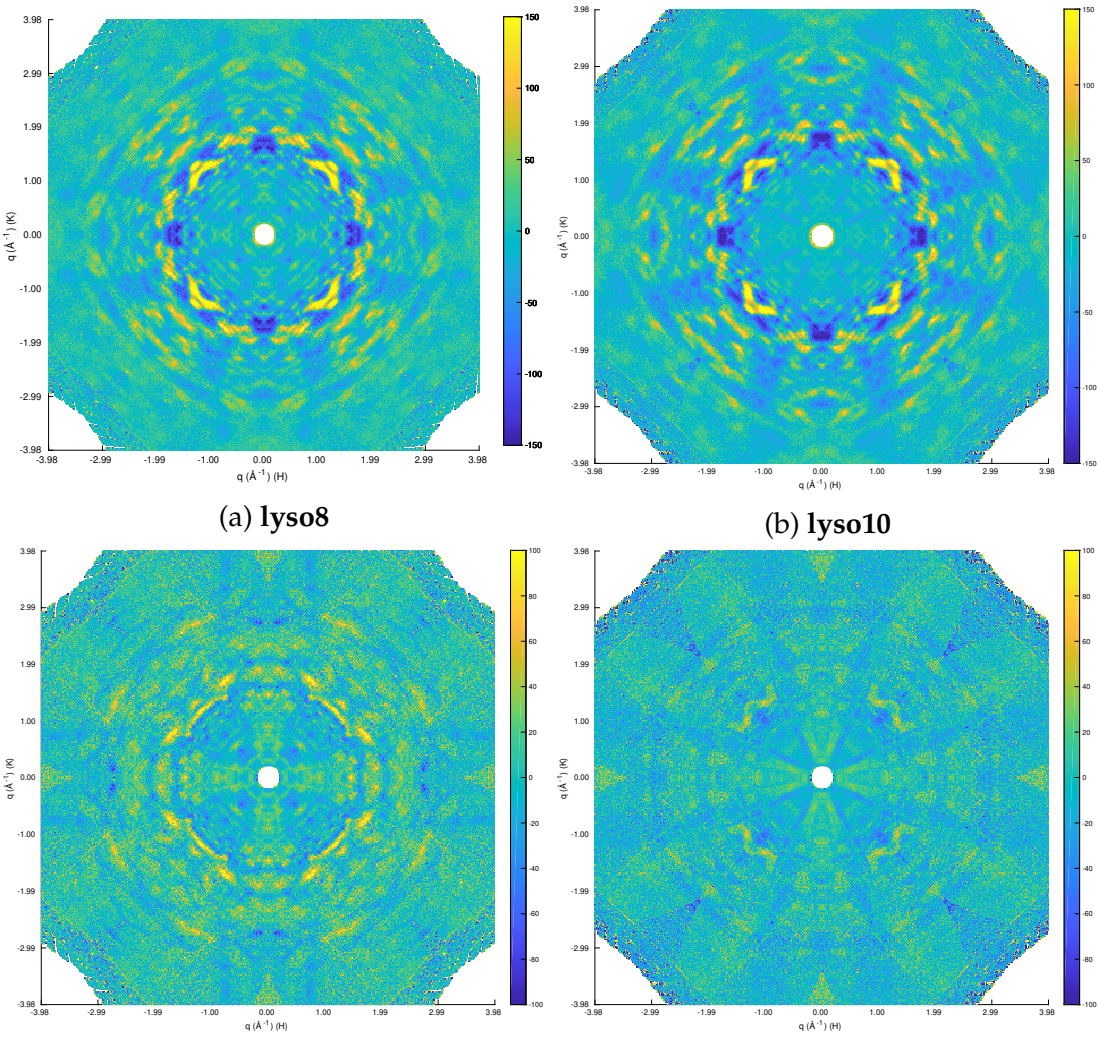


Figure 4.9: L = 0 slices of difference maps between sparser lattices of July 2015 crystals. Top: **lyso10** minus **lyso9** (both 3NAG-bound). Bottom: **lyso8** (unbound) minus **lyso9** (3NAG-bound). Display is on an arbitrary but consistent scale, where yellow denotes positive intensity and blue negative. The bottom map shows much more difference between a bound and unbound crystal than the top map shows between two bound crystals.



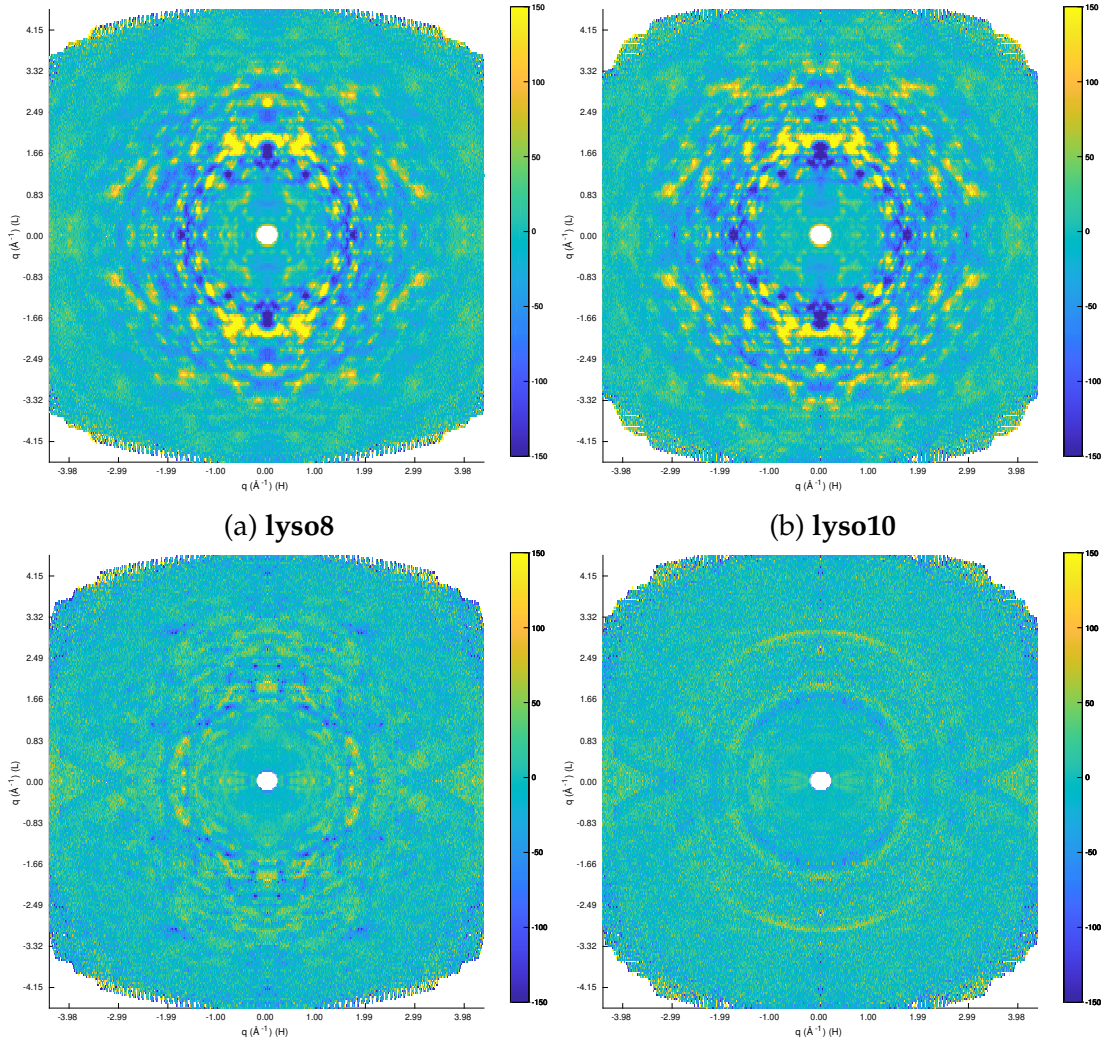
(a) **lyso8**

(b) **lyso10**

(c) **lyso8** minus **lyso9**

(d) **lyso10** minus **lyso9**

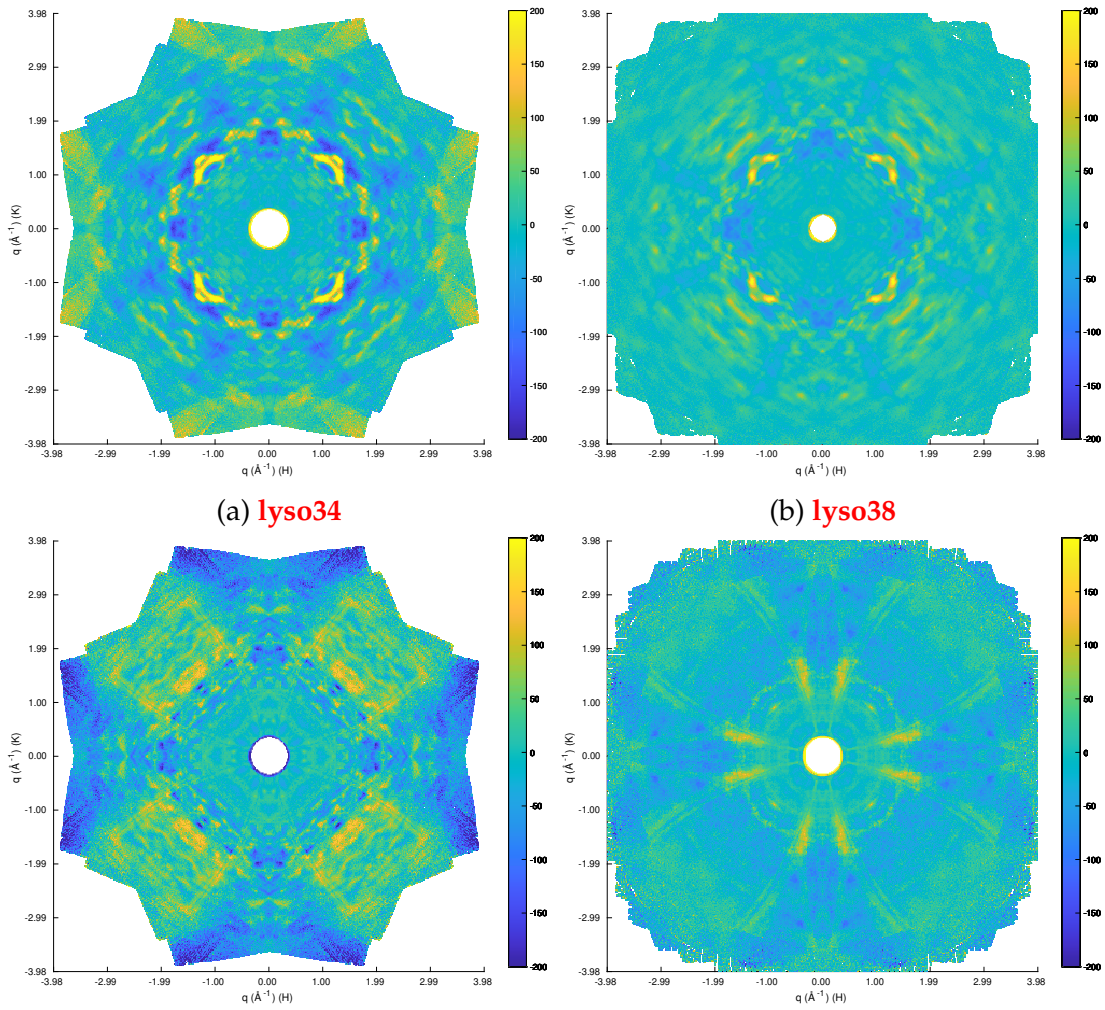
Figure 4.10: L = 0.5 slices of maps between denser lattices of July 2015 crystals. (a) **lyso8** (unbound). (b) **lyso10** (3NAG-bound). (c) **lyso8** (unbound) minus **lyso9** (3NAG-bound). (d) **lyso10** minus **lyso9** (both 3NAG-bound). Display is on an arbitrary but consistent scale, where yellow denotes positive intensity and blue negative. The two bound maps show less difference than the bound and unbound pair.



(a) **lyso8** (b) **lyso10**

(c) **lyso8** minus **lyso9** (d) **lyso10** minus **lyso9**

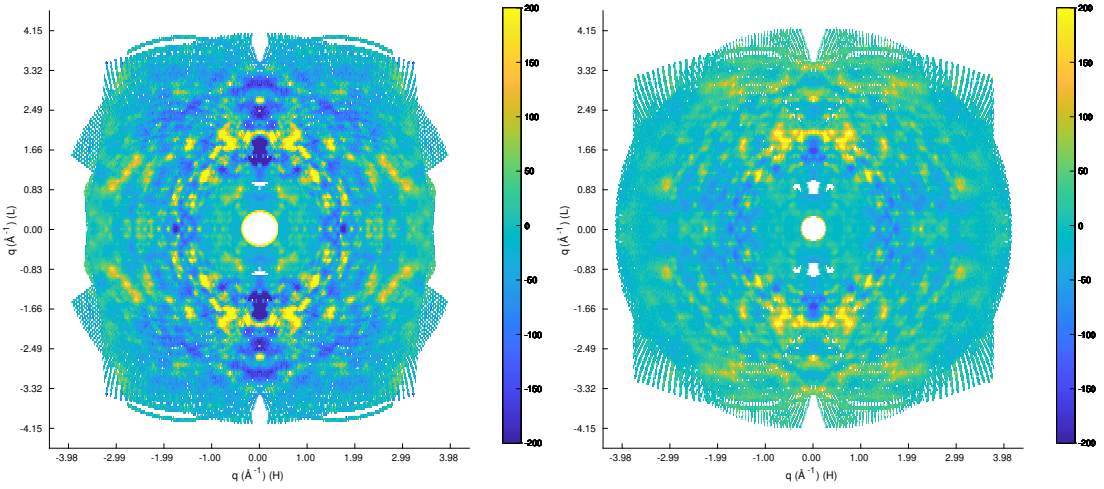
Figure 4.11: K = 0.5 slices of maps between denser lattices of July 2015 crystals. (a) **lyso8** (unbound). (b) **lyso10** (3NAG-bound). (c) **lyso8** (unbound) minus **lyso9** (3NAG-bound). (d) **lyso10** minus **lyso9** (both 3NAG-bound). Display is on an arbitrary but consistent scale, where yellow denotes positive intensity and blue negative. The two bound maps show less difference than the bound and unbound pair.



(a) **lyso34** (b) **lyso38**

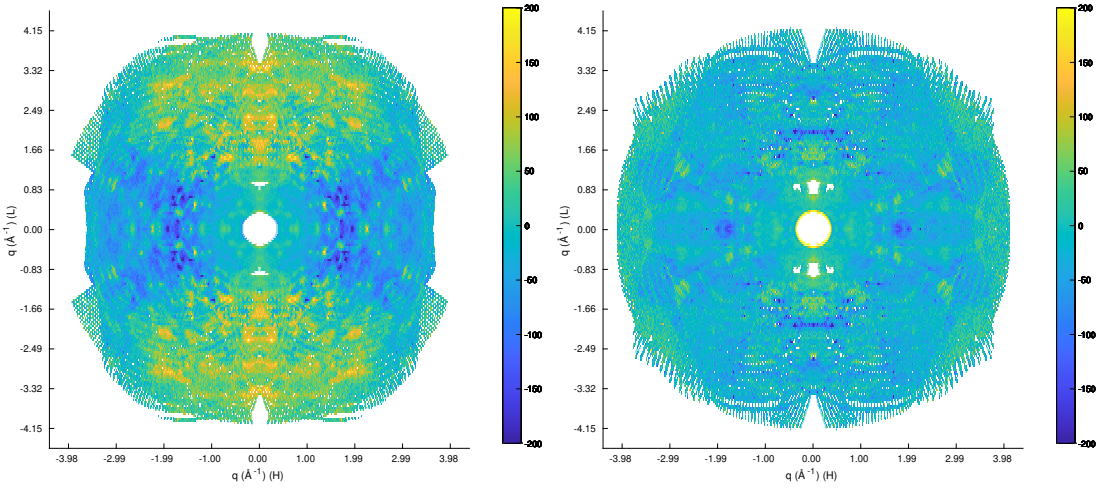
(c) **lyso47** minus **lyso34** (d) **lyso47** minus **lyso38**

Figure 4.12: L = 0.5 slices of maps between denser lattices of February 2014 crystals. (a) **lyso34** (unbound). (b) **lyso38** (3NAG-bound). (c) **lyso47** (3NAG-bound) minus **lyso34** (unbound). (d) **lyso47** minus **lyso38** (both 3NAG-bound). Display is on an arbitrary but consistent scale, where yellow denotes positive intensity and blue negative. Much less signal remains in the difference map (d) between two bound crystals than in the difference map (c) between a bound and an unbound crystal, indicating consistency in a change in diffuse scattering upon 3NAG binding. The long thin lines and broad bright regions in the difference maps are likely marks of the absorption effects which were not corrected in the February 2014 data.



(a) lyso34

(b) lyso38



(c) lyso47 minus lyso34

(d) lyso47 minus lyso38

Figure 4.13: K = 0.5 slices of maps between denser lattices of February 2014 crystals. (a) lyso34 (unbound). (b) lyso38 (3NAG-bound). (c) lyso47 (3NAG-bound) minus lyso34 (unbound). (d) lyso47 minus lyso38 (both 3NAG-bound). Display is on an arbitrary but consistent scale, where yellow denotes positive intensity and blue negative. The broad bright regions in the difference maps are likely marks of the absorption effects which were not corrected in the February 2014 data.

fected the diffuse scattering, but that it reduced it. To test this, diffuse:Bragg intensity ratios were calculated as described previously for all individual crystals. Results are displayed in Figures 4.14 and 4.15. On average, diffuse:Bragg intensity ratios are lower in 3NAG-bound crystals than in unbound crystals, in both the February and July data sets. This pattern holds at most resolutions for most individual crystals, though exceptions are evident in Figure 4.14.

It would be interesting to determine which molecular motions correspond to the changes in diffuse signal between the bound and unbound maps, but such detailed interpretation is beyond the scope of this work. When the relationship between specific diffuse intensity patterns and specific molecular motions is better understood, the comparison between inhibitor-bound and unbound diffuse scattering patterns will again be an important check on our understanding.

4.4 Effect of cryocooling

4.4.1 Ethylene glycol effects

The cryocooled crystals were co-crystallized with ethylene glycol, unlike the room temperature crystals discussed thus far. To distinguish between the effects of the ethylene glycol and the effects of the cryocooling, room temperature crystals with and without ethylene glycol were compared. Correlation coefficients between pairs of these crystals are presented in Figure 4.16. Among the denser maps, it appears that crystals containing EtGly exhibit slightly higher correlation coefficients with any other crystals, but the effect is very small. Among the sparser maps, there appears to be no change in pairwise correlation due to

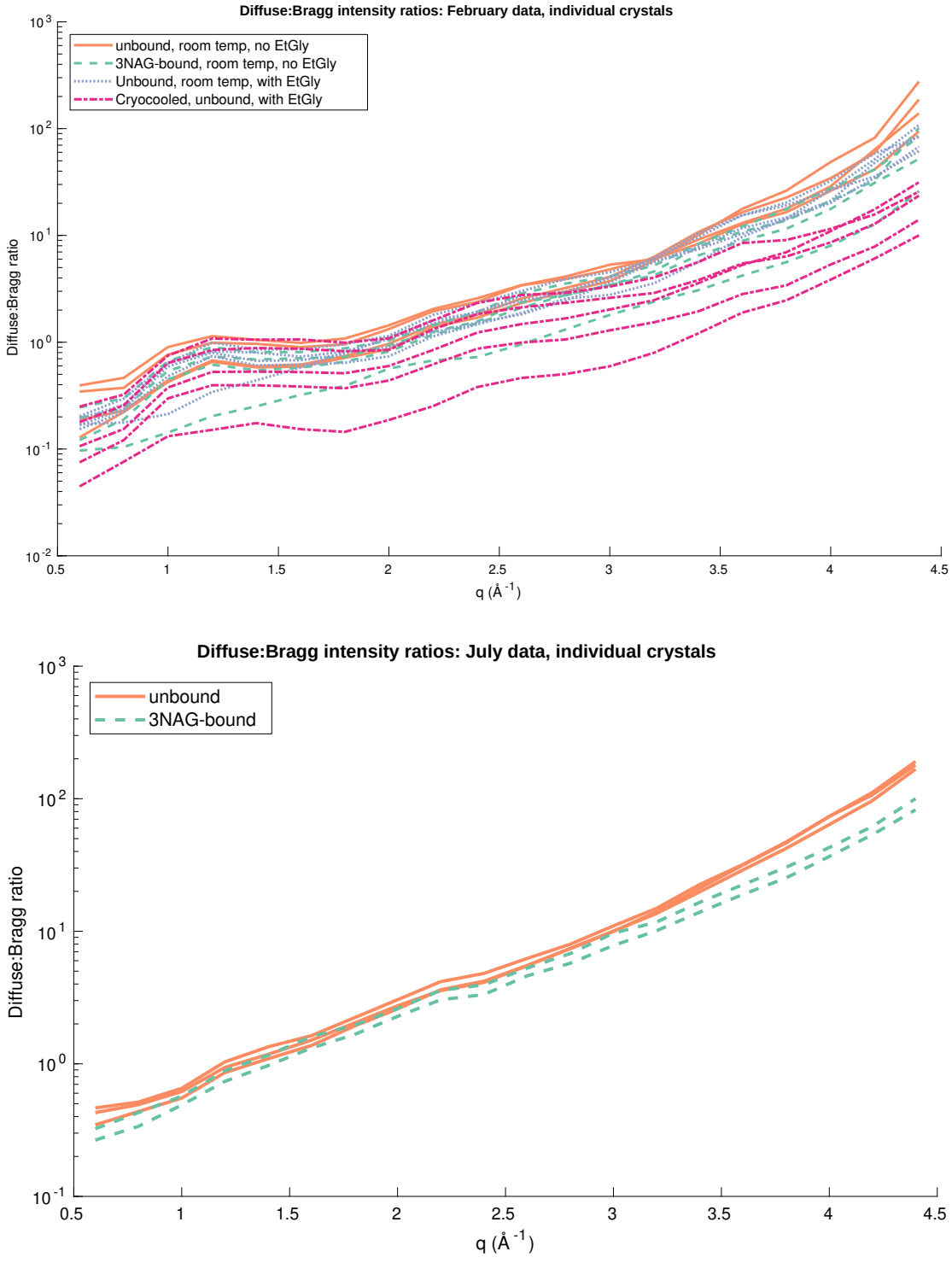


Figure 4.14: Diffuse:Bragg intensity ratios in individual crystals, color-coded by experimental condition.

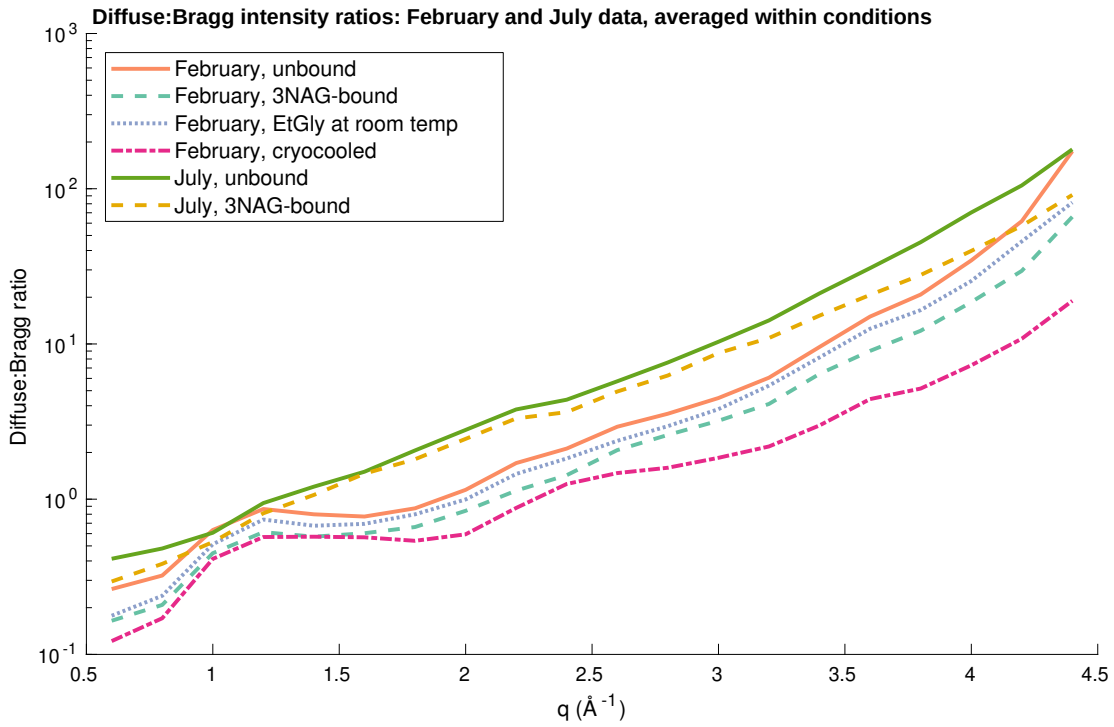


Figure 4.15: Diffuse:Bragg intensity ratios, averaged within conditions within CHESS runs.

EtGly presence. Figure 4.15 shows that diffuse:Bragg ratios drop slightly when EtGly is added to a room temperature crystal, suggesting that EtGly may possibly dampen protein motions even without cryocooling. If so, the effect is likely more uniform than the effect of inhibitor binding, given the consistent correlation coefficients across EtGly states.

4.4.2 Comparisons between room temperature and cryocooled crystals

Figure 4.17 shows pairwise correlation coefficients between EtGly-containing crystals from February 2014, sorted by temperature. The cryocooled crystal **lyso5** was left out of this data, as it demonstrated uniquely poor correlation with

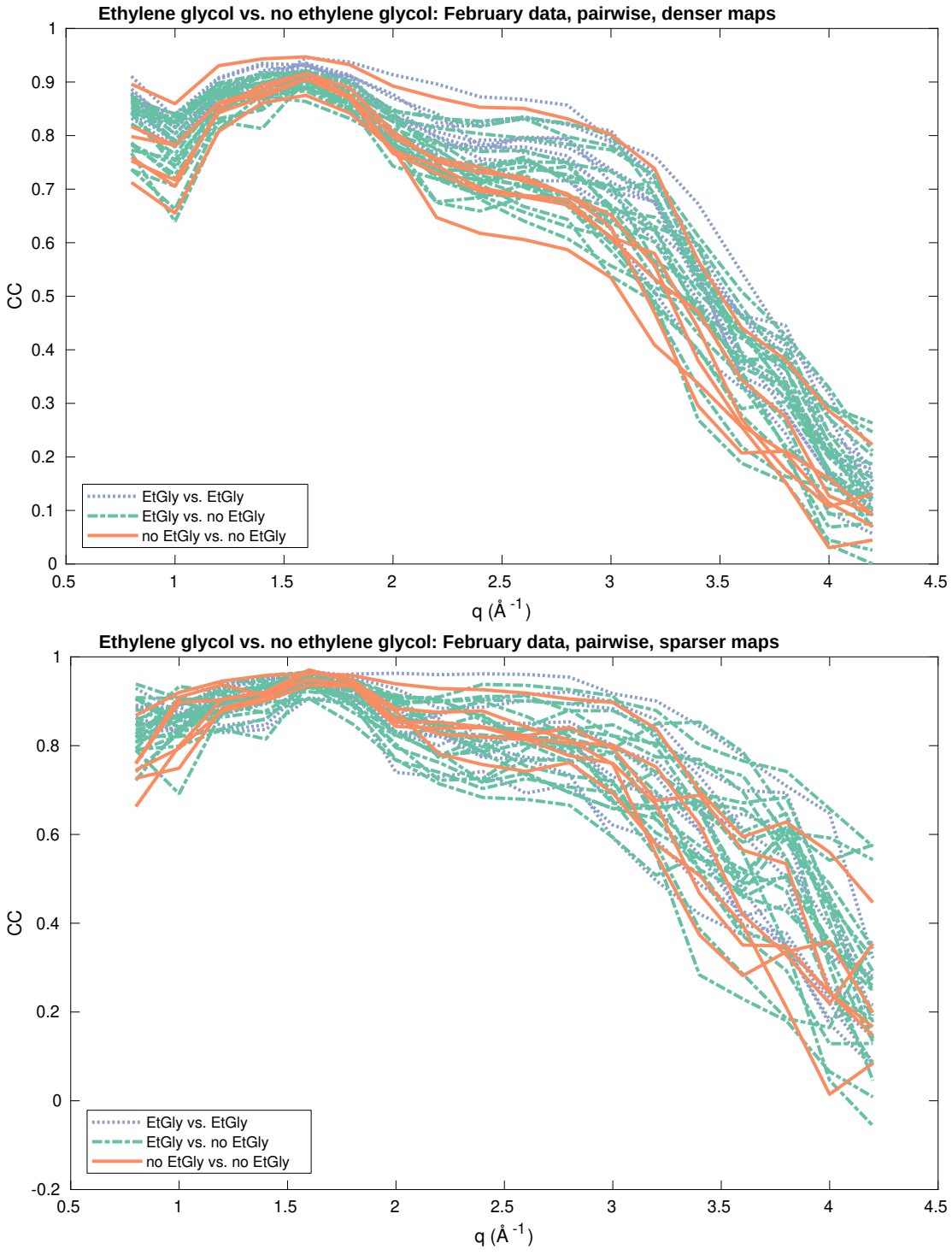


Figure 4.16: Correlation coefficients between pairs of single-crystal 3D maps as a function of scattering vector, within the February 2014 data set. Correlations are color-coded by EtGly presence. Top: correlation coefficients between denser maps. Bottom: correlation coefficients between sparser maps.

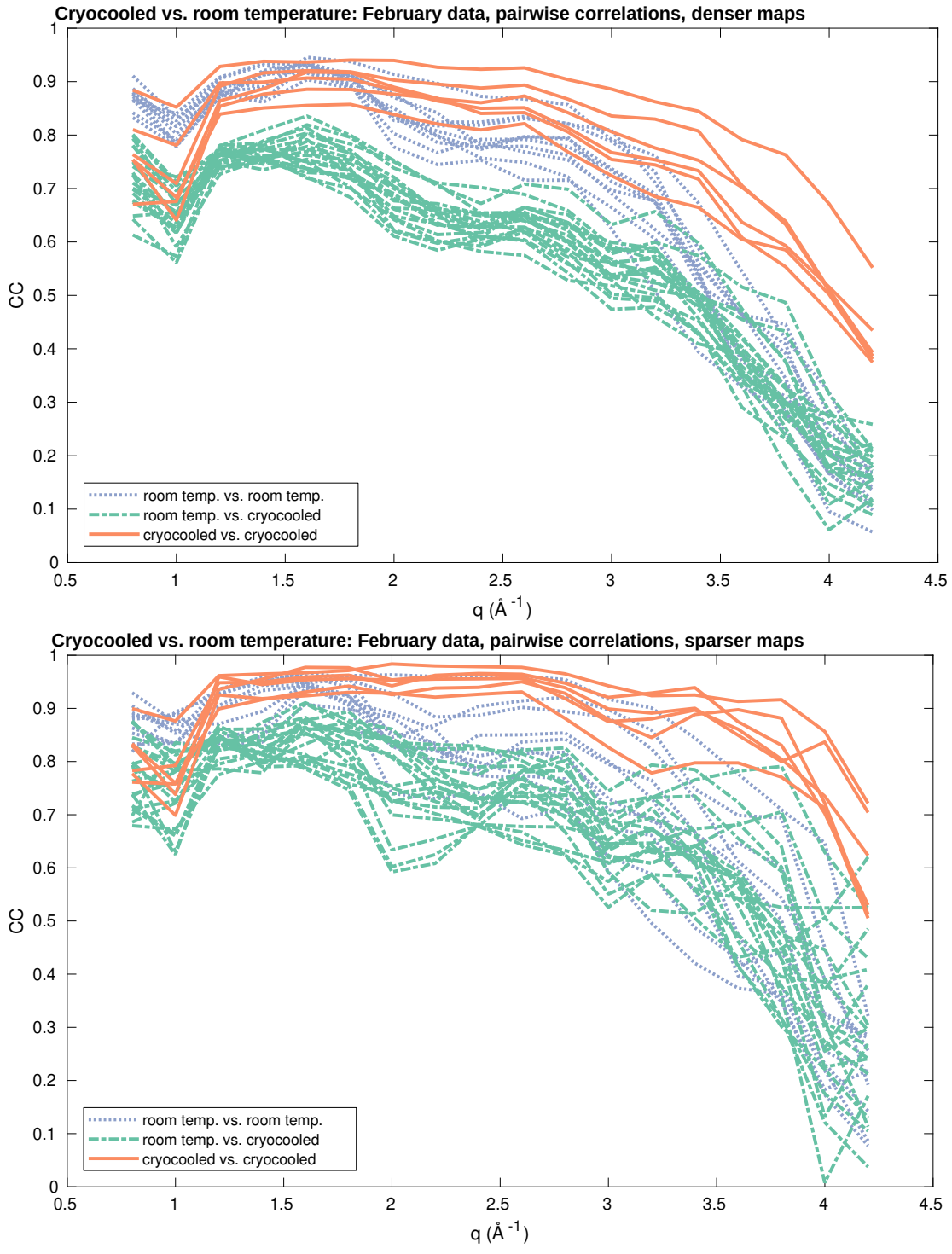


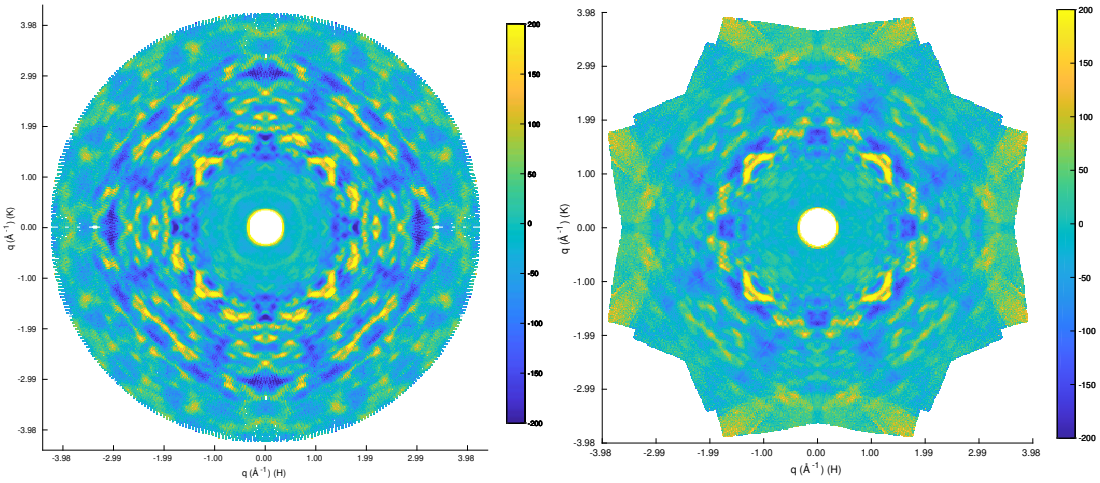
Figure 4.17: Correlation coefficients between pairs of single-crystal 3D maps as a function of scattering vector, among February 2014 crystals containing ethylene glycol. Correlations are color-coded by temperature. Top: correlation coefficients between denser maps. Bottom: correlation coefficients between sparser maps.

everything, as well as surprisingly weak diffuse intensity (Figure 4.1). Since common crystallographic practice is to collect data from many crystals and analyze just one data set, discarding an outlier is consistent with analyses that can produce useful recommendations for future crystallographic studies, as this work aims to do.

In both the sparser and denser maps, cryocooled crystals consistently correlate better with each other than with room temperature crystals for $q > 0.1$ Å, without exception. For the most part, they also correlate better with each other than the room temperature crystals do with each other, indicating both a shift in diffuse scattering and a possible increased consistency of diffuse scattering upon cryocooling.

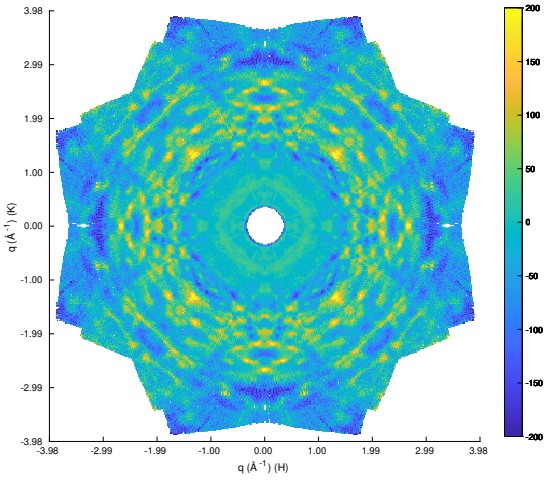
Figures 4.14 and 4.15 show that while the ratio of diffuse to Bragg intensity is, on average, lower for cryocooled crystals than for any of the room temperature conditions, it varies significantly among individual cryocooled crystals except at high resolutions.

Figures 4.18 and 4.19 each show a side-by-side comparison between slices through the same reciprocal space plane in a cryocooled crystal's diffuse map and a room temperature crystal's diffuse map. While many of the features are the preserved between the two maps, there are clear differences in the pattern. Figures 4.18c and 4.19c show the differences between the two maps. Because the patterns were so different, they were difficult to scale together; this is about as flat as the difference map can be. Note that all three images in each figure are displayed on the same (arbitrary) intensity scale; the differences are relatively large.



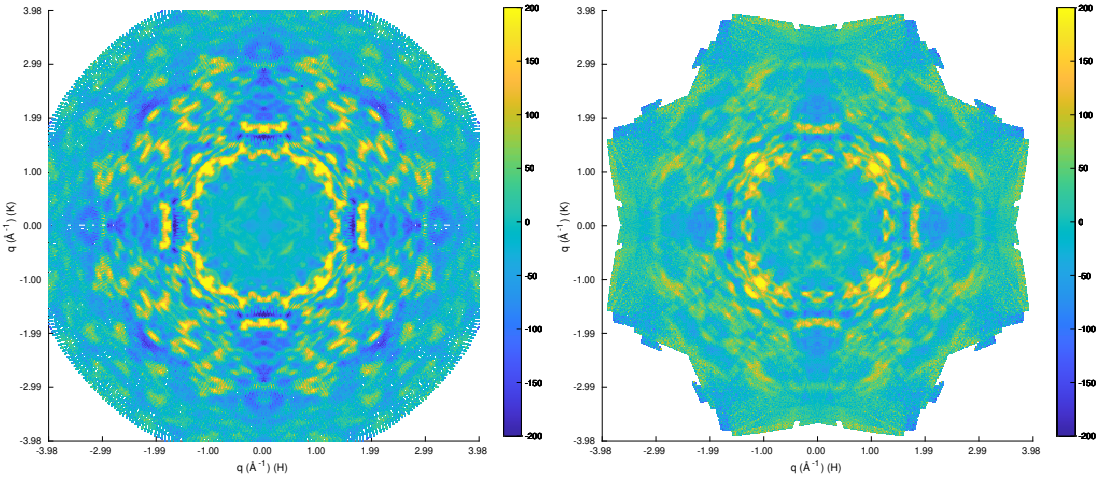
(a) **lyso19**, cryocooled

(b) **lyso34**, room temperature



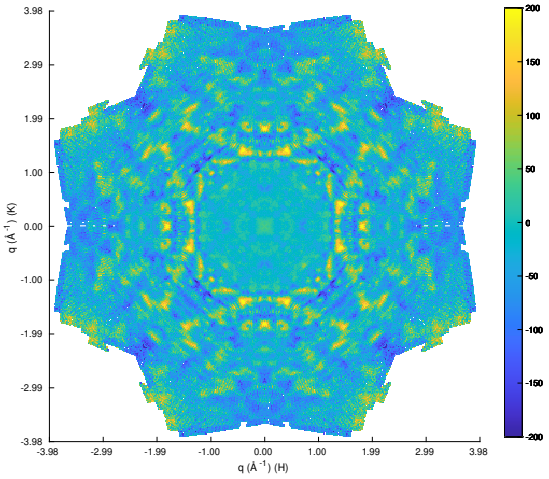
(c) **lyso19** minus **lyso34**

Figure 4.18: Slices in the L = 0.5 plane through denser lattices from February 2014. (a) **lyso19**, a cryocooled crystal. (b) **lyso34**, a room temperature crystal. (c) **lyso34** scaled and subtracted from **lyso19**. Remaining strong features indicate poor agreement between the cryocooled and room temperature maps. All images are displayed on the same arbitrary color intensity scale, where yellow is positive and blue is negative.



(a) **lyso19**, cryocooled

(b) **lyso34**, room temperature



(c) **lyso19** minus **lyso34**

Figure 4.19: Slices in the L = 3.5 plane through denser lattices from February 2014. (a) **lyso19**, a cryocooled crystal. (b) **lyso34**, a room temperature crystal. (c) **lyso34** scaled and subtracted from **lyso19**. Remaining strong features indicate poor agreement between the cryocooled and room temperature maps. All images are displayed on the same arbitrary color intensity scale, where yellow is positive and blue is negative.

These data show that the diffuse scattering pattern is not consistent between the room temperature and cryocooled crystals measured here. It is possible that the differences are at least partially due to crystals at different temperatures suffering different amounts of radiation damage, rather than to the cryocooling itself. The data presented here are insufficient to make this distinction clearly; further comparison between crystals at the same temperature but different radiation doses would help.

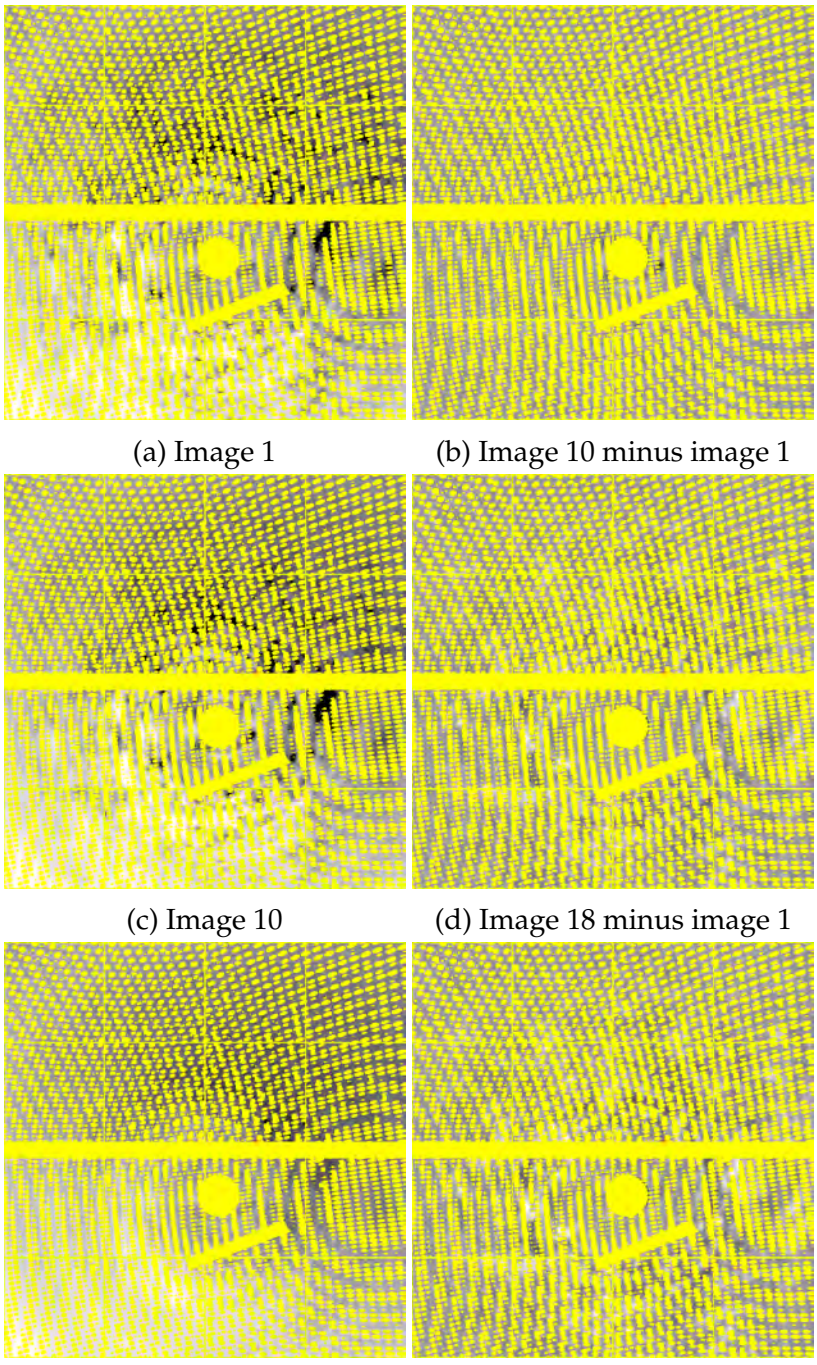
4.5 Radiation damage effects

To evaluate the effects of radiation damage on diffuse intensity, several crystals from the February 2014 CHESS visit were taken through a "death run", in which a series of relatively long exposures were recorded while the crystal sat stationary (i.e., no oscillation in phi during or between images). In most cases, these death runs were recorded after standard data collection, so the initial image is already not of a fresh crystal.

For each crystal, 25-30 still frames of 5 seconds each were recorded, for an x-ray dose of approximately 0.011 MGy per frame. X-ray dose was calculated by obtaining the x-ray beam flux using the Ion Chamber Flux Calculator available on the CHESS website[2] and then providing the flux and other experimental parameters to RADDOSE [Zeldin et al., 2013] to obtain the average dose in the exposed region of the crystal. Time elapsed between frames was negligible.

The death run for **lyso30**, an unbound room temperature crystal without Et-

[2]http://www.chess.cornell.edu/users/calculators/ion-chamber-flux-calculator as of July 11, 2019



(a) Image 1 (b) Image 10 minus image 1

(c) Image 10 (d) Image 18 minus image 1

(e) Image 25 (f) Image 25 minus image 1

Figure 4.20: Images from **lyso30**'s death run. (a, c, e) Individual images with the isotropic component subtracted. At a dose of 0.011 MGy/image, the observable effects of radiation damage plateau by image 25. Structured diffuse intensity has mostly vanished by image 25. (b, d, f) Differences between individual images, with the isotropic component of the difference subtracted. Figures (b, c) together show that the first significant change is increased structure at very low resolution. (f) is nearly the inverse of (a) except for the overall dark-light gradient, which indicates uncorrected absorption effects.

Gly from the February 2014 CHESS run, is presented here as a characteristic example. Individual images were processed as described in section 3.2, with an additional smoothing step to facilitate display and subtraction. Prior to subtracting the radial average from each image, but after smoothing, the first image in the death run was subtracted from every other image in the death run to obtain a series of difference images. This subtraction gets rid of any constant background scatter and mitigates the influence of uncorrected absorption effects on the data.

Figure 4.20 shows a selection of original images and difference images through the death run, all with their isotropic components subtracted to highlight structured diffuse features. By image 25, the visible changes due to x-ray exposure have plateaued and the anisotropic diffuse intensity has almost entirely vanished. Figure 4.20b shows that the first anisotropic change is at very low resolution, and Figure 4.20c in comparison with Figure 4.20a shows that this change is specifically an increase in the anisotropic diffuse intensity. Low-resolution features generally correspond to large length scale motions or disorder; this observation therefore implies that radiation exposure causes larger structural shifts in the molecules before it destroys smaller-scale motions, but this is by no means confirmed.

Figure 4.21 shows how the average isotropic and anisotropic diffuse intensities change with radiation dose as a function of resolution. Each trace in each graph is a radial average of magnitudes in one of the subtracted images created as described above. At most resolutions, the anisotropic diffuse intensity change increases with increasing dose, and Figure 4.20 confirms that this is actually a decrease in visible structure. Around $q = 0.5$ Å, the anisotropic intensity

difference (compared to the first image) is largest after 0.1 MGy; the difference quickly drops again. In constrast, the isotropic intensity change is nearly monotonic with dose everywhere. Beyond approximately $q = 2.2$ Å, the isotropic intensity seems to change only between 0.05 MGy and 0.10 MGy. At lower resolutions, the isotropic change is steadily increasing, leveling off at the highest dosages. Looking at the individual images confirms that these isotropic changes are all increases in intensity. In the raw images, it is apparent by eye that the Bragg peaks fade out with dosage, with higher-resolution peaks disappearing entirely, so an increase in diffuse intensity is expected because the scattered photons must go somewhere.

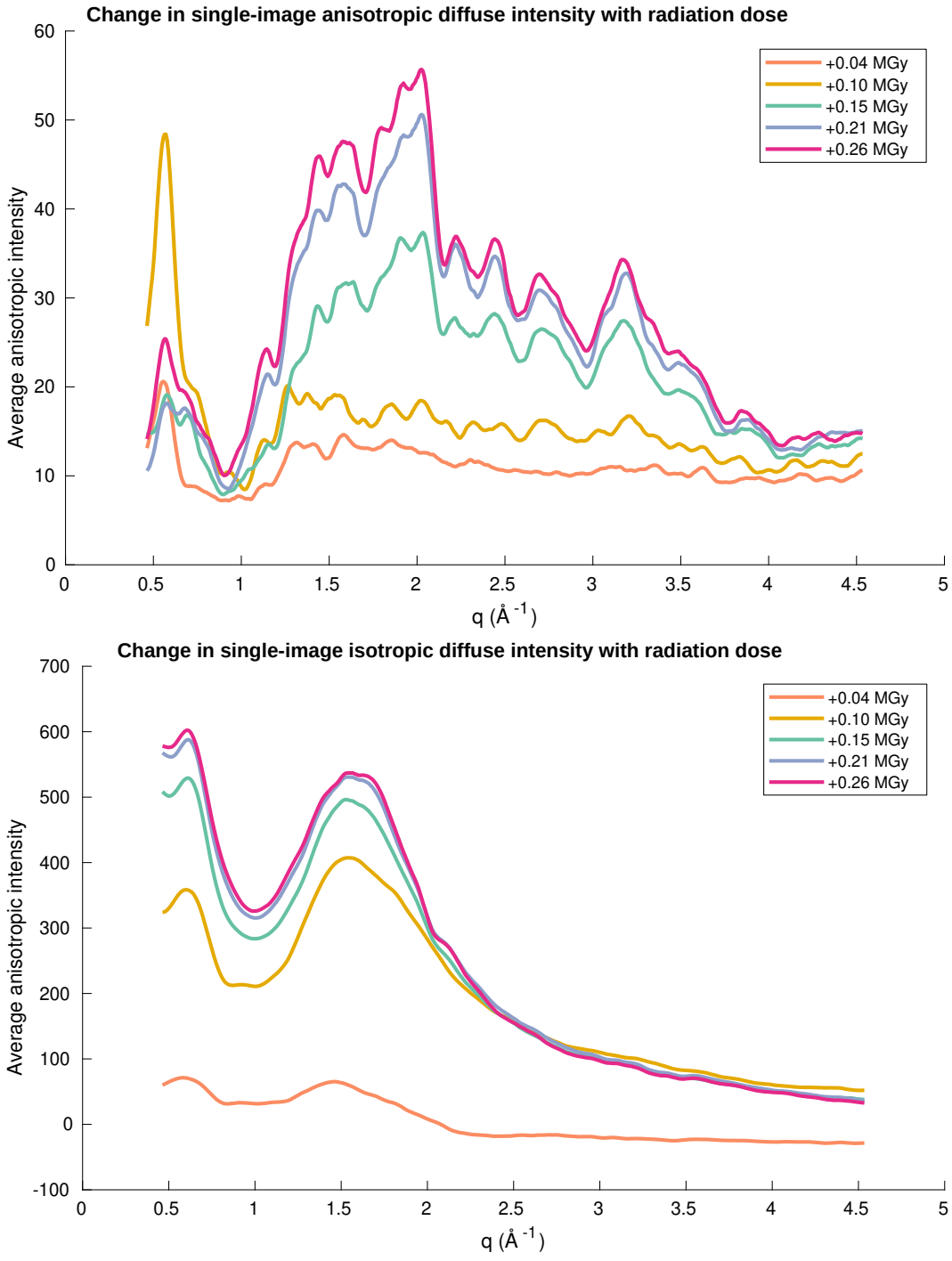


Figure 4.21: Change in average (top) anisotropic and (bottom) isotropic diffuse intensity with radiation dose, as a function of resolution. Each trace is the radial average magnitude of each type of intensity in a subtracted image (e.g., image 10 minus image 1). The legend gives the approximate additional dose received by the crystal between images. Intensity units are arbitrary but consistent between graphs.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Reproducibility

Diffuse intensity appears to be nearly as consistent across crystals under the same experimental conditions as within the images recorded of a single crystal, particuarly when sampling the diffuse intensity four times as often as the Bragg peaks in each dimension. When sampled only once per Bragg peak, the February 2014 data exhibited clearly lower reproducibility, which may be due to the lack of a consistent temperature during data collection.

The structure of diffuse intensity maps does not appear to depend on which CHESS run (February or July) they belong to, which is good—this means that the method of data collection does not have undue effect on the resulting signal. The data quality did vary with CHESS run, however; higher correlations and greater consistency were observed in the July 2015 data, in which more x-ray observations were averaged together into every measurement. This suggests that averaging over more regions of reciprocal space, as was done in the July 2015 data, is helpful for removing or reducing artifacts from, e.g., uncorrected absorption effects or anisotropic background scattering.

It so happened that most of the metrics in chapter 4 were also applied to a previous iteration of all of the described diffuse maps, in which some incorrect scales and absorption corrections were applies to the maps. When the data were updated and the metrics recalculated, the patterns of correlation coefficients and diffuse:Bragg ratios across crystal conditions did not actually change

significantly. It appears, therefore, that at least for this data, results are not highly sensitive to the image processing methods, which is heartening given that diffuse image processing methods remain relatively crude and unstandardized compared to the methods available for Bragg data. That said, more robust and consistent data, with less variation of its own, is likely to be more sensitive to processing method, so development of standardized and validated methods to extract diffuse intensities from diffraction images is still an important immediate direction for further work.

5.2 Effects of inhibitor binding and cryocooling

Binding to an inhibitor caused a small yet consistent change in the resulting diffuse scattering. Furthermore, it reduced the overall amount of diffuse intensity relative to the total Bragg intensity in each resolution shell, supporting the hypothesis that binding to an inhibitor would reduce diffuse scattering. It makes sense that inhibitor binding would change the diffuse scattering pattern rather than simply reducing its intensity uniformly, since an inhibitor molecule binds to only one site of a protein with multifaceted conformational heterogeneity. However, the data presented here cannot confirm what protein motions are suppressed by inhibitor binding; it merely suggests that our general working model may be correct.

Cryocooling altered the diffuse intensity patterns much more than did binding to an inhibitor, as shown by direct comparison of three-dimensional maps and by the stark drop in correlation coefficients between cryocooled and room temperature crystals. While nothing can be said about which temperature re-

sults in a more meaningful diffuse map based on the data presented here, it is widely accepted that conditions closer to those *in vivo* faciliate strutural dynamics closer to that *in vivo*. That assumption combined with the analysis presented here suggests that cryocooling would be a huge impediment to discerning biologically relevant motions from diffuse scattering.

5.3 Recommendations and comments

For future diffuse scattering work, it is recommended first and foremost to focus on acquiring high quality data on the beamline. This includes collecting as many diffracted photons as possible over a large range of reciprocal space. Even for high-symmetry crystals, collecting raw data at as many angles as possible allows for better removal, or at least smoothing, of absorption artifacts. Fine-slicing images is not useful for diffuse scattering, as it necessitates such a low exposure time per image that diffuse intensity will be hard for even a computer to discern from noise in a single image. Reduction of background scattering is also important, and can be accomplished both by minimizing scattering materials in the beam path (partly through selection of mounting materials) and by directly measuring any anisotropic background scattering as closely as possible, even to the point of collecting as much background data as crystal data.

With sufficient data quality, a single crystal's diffuse scattering can likely be taken as representative of the diffuse scattering from any other crystal of that protein under the same conditions. Temperature is a relevant condition here; it is likely that different temperatures all above freezing will lead to different diffuse intensiy maps, and certainly cryocooling will severely change the diffuse

intensity. More work on the specific temperature dependence of "room temperature" diffuse scattering would be helpful here.

Finally, as has been noted elsewhere: the overall magnitude of diffuse intensity in typical protein crystal diffraction is much larger than one might think by looking at a single image. In these data, the ratio of total anisotropic diffuse intensity to total Bragg intensity increases almost logarithmically with resolution, and the diffuse intensity exceeded the Bragg for most of the resolution range–and this is excluding protein-derived isotropic diffuse intensity. There is a wealth of information to be gathered here with robust data collection protocols.

APPENDIX A

SUGGESTED EXPERIMENTAL METHOD FOR BACKGROUND SUBTRACTION

The following is a suggested yet untested method to record an accurate fiber scattering pattern for each individual crystal for which mounting materials scatter x-rays onto the detector: Before mounting a crystal, mark a spot on the pin base and align the loop to the pin base in a reproducible way (for instance, the center of the top of the loop—they are concave—aligned with the marked spot). Then, mount the crystal on the loop, cover it with the plastic sheath, and mark the spot on the sheath that lines up with the marked spot on the base. Place the mounted crystal on the goniometer and collect all desired data. When finished, remove the sheath without disturbing the orientation of the pin base with respect to the goniometer. Clean the loop thoroughly using whatever combination of solvents and paper wicks is preferred, again without rotating the pin base or translating the spindle supporting the sample. Return the loop to its original orientation if it shifted during cleaning, and replace the plastic sheath in its original orientation. Finally, collect another set of images in the same set of orientations as the crystal data, just without any crystal in place. This set of background images should display the same changing intensity pattern with rotation as the fiber scatter in the original data set, except for relatively small absorption effects. Overall scale factors may still need to be applied to the images, perhaps with the original fiberfit algorithm, to correct for beam intensity changes and different exposure times if used.

APPENDIX B

INTERPOLATION AND EXTRAPOLATION METHODS FOR ABSORPTION CORRECTION AND SCALE FACTORS

Decay and absorption correction factors, as reported by XSCALE, were interpolated and extrapolated to fill all necessary parameter space using methods from Python's scipy Interpolation sub-package.

The decay correction factor array was expanded to each image number with the InterpolatedUnivariateSpline method, which extrapolates beyond the input points (i.e., to the smallest and largest image numbers in this case) without introducing local extrema in the extrapolation region.

XSCALE gives absorption correction factors in terms of only 13 regions on the detector, each denoted by the indices of a single pixel. To obtain approximate correction values for all pixels on the detector, these values were interpolated over the entire detector space. Since the outermost listed detector positions are at a resolutions as low as 2.1 Å in the 2014 data (1.7 Å in the 2015 data) while the crystals all diffract well past 2 Å, the interpolation method must also perform extrapolation; Python's radial basis function-based method (scipy.interpolate.rbf) was chosen for this capability. The extrapolation result and, to a lesser extent, the interpolation result depend significantly on the functional form chosen, so four methods were tested on several sets of absorption correction factors. Results from two of the sets, one from each of the two detectors used, are shown in Figure B.1. In both sets, the interpolated values look similar using all four methods while the extrapolated values vary widely. The linear and cubic methods were rejected for producing extreme values at the edge of the detector in both data sets. In the **lyso6** data set (taken in 2015 on the Pilatus 6M), the mul-

tiquadric and thin plate methods produced values with opposite trend (i.e., increasing vs. decreasing). Given the original crystal shape and scattering angles involved, it seems unlikely that the absorbance would have local extrema at intermediate resolutions, so the thin plate extrapolation is probably better. In the **lyso14** data set (taken in 2014 on the Eiger 1M), both the multiquadric and thin plate methods produce values that increase with increasing resolution, but the thin plate values are less extreme, so again the thin plate approximation seems more likely. Therefore, thin plate basis functions were chosen for all subsequent absorption correction interpolation.

Once each set of absorption correction factors, corresponding to one bin of image numbers, was thus expanded to cover the entire detector face, the correction factors for a specific input image were found by linear interpolation between the two nearest expanded sets of factors. For images at the beginning and end of the series, the first or last set of absorption correction factors was applied without further adjustment.

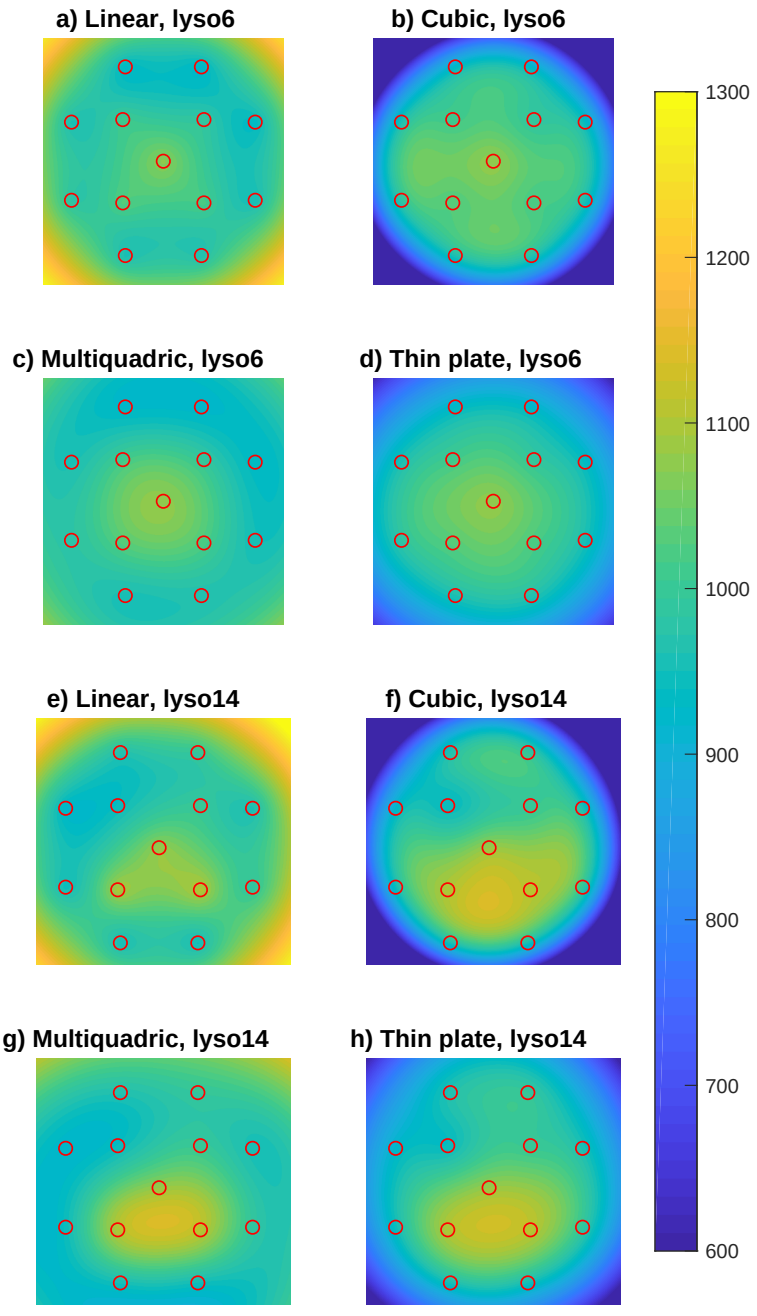


Figure B.1: Interpolated and extrapolated absorption correction factors for: (a-d) one image-number bin in crystal **lyso6** (July 2015 data, Pilatus 6M); (e-h) another image-number bin in **lyso14** (February 2014 data, Eiger 1M). Images display the entire detector face; red circles mark the locations of absorption correction values given by XSCALE (held constant in interpolation). Image titles give the basis function type used by scipy.interpolate.rbf. Values represent a reciprocal multiplicative correction factor multiplied by 1000.

APPENDIX C

CODES AND DATA FILES

Codes used for data processing and the raw and processed data files are all stored in a folder called "Veronica Pillar thesis codes and data", under the care of Professor Sol Gruner. Specific files and where to find them are listed below.

The document "Instructions for overall data processing.odt", in the same folder, gives some step-by-step instructions for processing the diffraction images; some of the information in it is out of date, but much is correct.

Table C.1 lists all the code files by path and brief description. The python scripts depend on libraries available in PHENIX, available at http://www.phenix-online.org. The C codes, in the lunus subdirectory, function in the context of the full *Lunus* software suite, available at mewall.github.io/lunus/.

File path	Description
Codes/lunus/avgrim_aniso.c	Calculates average anisotropic intensity vs. radius in an image
Codes/lunus/ccrlt.c	Calculates Pearson correlation coefficients and average intensities in resolution shells for a 3D lattice
Codes/lunus/debeamim.c	Masks the beamstop shadow in an image
Codes/lunus/debraggim.c	Removes bright spots in an image
Codes/lunus/detectormask.c	Masks chip boundaries in an image
Codes/lunus/lavgrim_aniso.c	Library function
Codes/lunus/lavgrim.c	Library function
Codes/lunus/lccrlt.c	Library function
Codes/lunus/ldebeamim.c	Library function
Codes/lunus/ldebraggim.c	Library function
Codes/lunus/lgetpix.c	Library function
Codes/lunus/lmathim.c	Library function

Codes/lunus/lmergeim.c	Library function
Codes/lunus/lnormim.c	Library function
Codes/lunus/lp6mmask.c	Library function
Codes/lunus/lpolarim.c	Library function
Codes/lunus/lrmpixim.c	Library function
Codes/lunus/lsetped.c	Library function
Codes/lunus/lsmoothim.c	Library function
Codes/lunus/lsmoothrf.c	Library function
Codes/lunus/lunus.h	Header file with parameters used in this work
Codes/lunus/mathim.c	Adds or subtracts two images with optional scale factors
Codes/lunus/mergeim.c	Averages a series of images together
Codes/lunus/multim.c	Multiples two images together with optiona scale factor
Codes/lunus/rmpixim.c	Removes Bragg peaks from an image according to a list of pixel indices
Codes/lunus/setped.c	Sets a pedestal value for an image
Codes/lunus/smoothim.c	Smoothes an image with local averaging
Codes/lunus/smoothrf.c	Smoothes a 1D rfile with local averaging
Codes/MATLAB/displat.m	Displays slices of 3D maps
Codes/MATLAB/fiberfit.m	Finds scale factor to fit fiber scattering image to July 2015 data image for background subtraction
Codes/MATLAB/ get_crystal_type.m	Dependency of readccrlt.m
Codes/MATLAB/readccrlt.m	Reads text file output of *Lunus*'s ccrlt
Codes/MATLAB/readimg.m	Reads a diffraction image (.img)
Codes/MATLAB/readlat.m	Reads a 3D map
Codes/MATLAB/readrf.m	Reads an rf file written by *Lunus*
Codes/MATLAB/ readsymlat16.m	Reads a symmetrized 3D map written in compressed form
Codes/MATLAB/snippets.m	Miscellaneous useful code snippets
Codes/MATLAB/testcorrs.m	Simulates Poisson noise as described in section 4.1.1
Codes/MATLAB/ work_with_ccrs.m	Reads and plots resolution-dependent graphs of correlation coefficients and intensity magnitudes
Codes/MATLAB/writeimg.c	Writes a diffraction image (.img)

Codes/MATLAB/writelat.m	Writes a 3D map
Codes/MATLAB/ writesymlat16.m	Writes a symmetrized 3D map in compressed form
Codes/python/ apply_xscale_corrs.py	Applies correction factors computed from XSCALE to a series of images
Codes/python/cbf2img.py	Converts a CBF (.cbf) to SMV (.img) format; best used with cbf2img.sh
Codes/python/ genlat_xds_parallel_hires.py	Generates a 3D lattice from input images
Codes/python/ genlat_xds_parallel_punch.py	Generates lists of pixels to punch for Bragg peak removal
Codes/python/ get_orientations.py	Gets crystal orientation information from an XDS file
Codes/python/latfromhkl.py	Generates a 3D lattice of Bragg intensities from an XDS output file
Codes/python/ procimgcmodule.c	Library functions to support the python scripts
Codes/python/setup.py	Installs procimgmodule.c into a local copy of PHENIX
Codes/shell_scripts/cbf2img.sh	Converts a CBF (.cbf) to SMV (.img) format
Codes/shell_scripts/myxds.sh	Runs XDS multiple times for multiple purposes

Table C.1: Codes used for the data processing described in this dissertation. Path names are relative to the "Veronica Pillar thesis codes and data" folder.

Data is stored in the subfolder Data within the main folder. The subfolder Data contains a subfolder for each CHESS run, each of which contains subfolders for each crystal. Inside each crystal's folder are the original images after format conversion and basic masking, images fully processed for anisotropic diffuse intensity extraction, the denser and sparser three-dimensional maps, an .mtz file containing the scaled and merged reflection data, and a .pdb file containing the real-space model. For the July 2015 crystals, original raw images prior to format conversion are also included alongside the post-conversion images. Table C.2 lists these files for **lyso14**; all other crystal folders contain the

same pattern of files.

File path	Description
raw_data/lyso14_*.img	Original images after format conversion
processed_data/ lyso14_*pdxtsn.img	Fully processed images of anisotropic diffuse intensity
slyso14.pdxtsn.v*.lat	3D diffuse maps
lyso14_all_weak.mtz	Scaled and merged reflection data
lyso14_v2_refine_2.pdb	Real-space structure

Table C.2: Data archived for **lyso14**. Path names are relative to the "Veronica Pillar thesis codes and data/Data/February2014/lyso14" folder. All other crystals have an equivalent set of data stored in their respective folders.

BIBLIOGRAPHY

[Adams et al., 2010] Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 66(2):213–221.

[Afonine et al., 2012] Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D: Biological Crystallography*, 68(4):352–367.

[Ayyer et al., 2016] Ayyer, K., Yefanov, O. M., Oberthür, D., Roy-Chowdhury, S., Galli, L., Mariani, V., Basu, S., Coe, J., Conrad, C. E., Fromme, R., Schaffer, A., Dörner, K., James, D., Kupitz, C., Metz, M., Nelson, G., Xavier, P. L., Beyerlein, K. R., Schmidt, M., Sarrou, I., Spence, J. C. H., Weierstall, U., White, T. a., Yang, J.-H., Zhao, Y., Liang, M., Aquila, A., Hunter, M. S., Robinson, J. S., Koglin, J. E., Boutet, S., Fromme, P., Barty, A., and Chapman, H. N. (2016). Macromolecular diffractive imaging using imperfect crystals. *Nature*, 530(7589):202–206.

[Berman et al., 2000] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242.

[Bernstein H. J. and Hammersley, 2005] Bernstein H. J. and Hammersley, A. P. (2005). Specification of the Crystallographic Binary File (CBF/imgCIF). In Hall S. R., , and McMahon, B., editors, *International Tables for Crystallography Volume G: Definition and exchange of crystallographic data*, pages 37–43. Springer Netherlands, Dordrecht.

[Beyerlein et al., 2017] Beyerlein, K. R., Dierksmeyer, D., Mariani, V., Kuhn, M., Sarrou, I., Ottaviano, A., Awel, S., Knoska, J., Fuglerud, S., Jönsson, O., Stern, S., Wiedorn, M. O., Yefanov, O. M., Adriano, L., Bean, R., Burkhardt, A., Fischer, P., Heymann, M., Horke, D. A., Jungnickel, K. E. J., Kovaleva, E., Lorbeer, O., Metz, M., Meyer, J., Morgan, A., Pande, K., Panneerselvam, S., Seuring, C., Tolstikova, A., Lieske, J., Aplin, S., Roessle, M., White, T. A., Chapman, H. N., Meents, A., and Oberthür, D. (2017). Mix-and-diffuse serial synchrotron crystallography. *IUCrJ*, 4(6):769–777.

[Boylan and Phillips, Jr., 1986] Boylan, D. and Phillips, Jr., G. N. (1986). Motions of Tropomyosin. *Biophysical Journal*, 49(1):76–78.

[Broennimann et al., 2006] Broennimann, C., Eikenberry, E. F., Henrich, B., Horisberger, R., Huelsen, G., Pohl, E., Schmitt, B., Schulze-Briese, C., Suzuki, M., Tomizaki, T., Toyokawa, H., and Wagner, A. (2006). The PILATUS 1M detector. *Journal of Synchrotron Radiation*, 13(2):120–130.

[Caspar et al., 1988] Caspar, D. L. D., Clarage, J. B., Salunke, D. M., and Clarage, M. S. (1988). Liquid-like movements in crystalline insulin. *Nature*, 332:659–662.

[Chapman et al., 2017] Chapman, H. N., Yefanov, O. M., Ayyer, K., White, T. A., Barty, A., Morgan, A., Mariani, V., Oberthür, D., and Pande, K. (2017). Continuous diffraction of molecules and disordered molecular crystals. *Journal of Applied Crystallography*, 50:1084–1103.

[Clarage et al., 1992] Clarage, J. B., Clarage, M. S., Phillips, W. C., Sweet, R. M., and Caspar, D. L. D. (1992). Correlations of atomic movements in lysozyme crystals. *Proteins*, 12(2):145–157.

[Clarage and Phillips, Jr., 1997] Clarage, J. B. and Phillips, Jr., G. N. (1997). Analysis of Diffuse Scattering and Relation to Molecular Motion. *Methods in Enzymology*, 277:407–432.

[De Klijn et al., 2019] De Klijn, T., Schreurs, A. M. M., and Kroon-Batenburg, L. M. J. (2019). Rigid-body motion is the main source of diffuse scattering in protein crystallography. *IUCrJ*, 6:277–289.

[Dinapoli et al., 2011] Dinapoli, R., Bergamaschi, A., Henrich, B., Horisberger, R., Johnson, I., Mozzanica, A., Schmid, E., Schmitt, B., Schreiber, A., Shi, X., and Theidel, G. (2011). EIGER: Next generation single photon counting detector for X-ray applications. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 650:79–83.

[Doucet and Benoit, 1987] Doucet, J. and Benoit, J.-P. (1987). Molecular dynamics studied by analysis of the X-ray diffuse scattering from lysozyme crystals. *Nature*, 325:643–646.

[Drenth, 1999] Drenth, J. (1999). *Principles of Protein X-Ray Crystallography*. Springer-Verlag, New York, 2nd edition.

[Evans and Murshudov, 2013] Evans, P. R. and Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography*, 69(7):1204–1214.

[Farley et al., 2014] Farley, C., Burks, G., Siegert, T., and Juers, D. H. (2014). Improved reproducibility of unit-cell parameters in macromolecular cryocrystallography by limiting dehydration during crystal mounting. *Acta Crystallographica Section D: Biological Crystallography*, 70(8):2111–2124.

[Faure et al., 1994] Faure, P., Micu, A., Pérahia, D., Doucet, J., Smith, J. C., and Benoit, J.-P. (1994). Correlated intramolecular motions and diffuse x-ray scattering in lysozyme. *Nature Structural Biology*, 1(2):124–128.

[Fischer et al., 2015] Fischer, M., Shoichet, B. K., and Fraser, J. S. (2015). One Crystal, Two Temperatures: Cryocooling Penalties Alter Ligand Binding to Transient Protein Sites. *ChemBioChem*, 16(11):1560–1564.

[Ford et al., 1974] Ford, L. O., Johnson, L. N., Machin, P. A., Phillips, D. C., and Tjian, R. (1974). Crystal structure of a lysozyme-tetrasaccharide lactone complex. *Journal of Molecular Biology*, 88(2):349–371.

[Fraser et al., 2011] Fraser, J. S., Van Den Bedem, H., Samelson, A. J., Lang, P. T., Holton, J. M., Echols, N., and Alber, T. (2011). Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proceedings of the National Academy of Sciences of the United States of America*, 108(39):16247–16252.

[Garman, 2010] Garman, E. F. (2010). Radiation damage in macromolecular crystallography: What is it and why do we care? *Acta Crystallographica Section D: Biological Crystallography*, 66:339–351.

[Glover et al., 1991] Glover, I. D., Harris, G. W., Helliwell, J. R., and Moss, D. S. (1991). The variety of x-ray diffuse scattering from macromolecular crystals and its respective components. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials*, 47(6):960–968.

[Guinier, 1963] Guinier, A. (1963). *X-Ray Diffraction in Crystals, Imperfect Crystals, and Amorphous Bodies*. W. H. Freeman, San Francisco.

[Halle, 2004] Halle, B. (2004). Biomolecular cryocrystallography: Structural changes during flash-cooling. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14):4793–4798.

[Kabsch, 2010] Kabsch, W. (2010). XDS. *Acta Crystallographica Section D: Biological Crystallography*, 66(2):125–132.

[Keedy et al., 2014] Keedy, D. A., Van Den Bedem, H., Sivak, D. A., Petsko, G. A., Ringe, D., Wilson, M. A., and Fraser, J. S. (2014). Crystal cryocooling distorts conformational heterogeneity in a model Michaelis complex of DHFR. *Structure*, 22(6):899–910.

[Kolatkar et al., 1994] Kolatkar, A. R., Clarage, J. B., and Phillips, Jr., G. N. (1994). Analysis of diffuse scattering from yeast initiator tRNA crystals. *Acta Crystallographica Section D: Biological Crystallography*, 50(2):210–218.

[Kuzmanic et al., 2014] Kuzmanic, A., Pannu, N. S., and Zagrovic, B. (2014). X-ray refinement significantly underestimates the level of microscopic heterogeneity in biomolecular crystals. *Nature Communications*, 5:3220.

[Liebschner et al., 2013] Liebschner, D., Dauter, M., Brzuszkiewicz, A., and Dauter, Z. (2013). On the reproducibility of protein crystal structures: five atomic resolution structures of trypsin. *Acta Crystallographica Section D: Biological Crystallography*, 69(8):1447–1462.

[McCoy et al., 2007] McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, 40(4):658–674.

[Meisburger et al., 2017] Meisburger, S. P., Thomas, W. C., Watkins, M. B., and Ando, N. (2017). X-ray scattering studies of protein structural dynamics. *Chemical Reviews*, 117(12):7615–7672.

[Peck et al., 2018] Peck, A., Poitevin, F., and Lane, T. J. (2018). Intermolecular correlations are necessary to explain diffuse scattering from protein crystals. *IUCrJ*, 5(2):211–222.

[Pérez et al., 1996] Pérez, J., Faure, P., and Benoit, J.-P. (1996). Molecular rigid-body displacements in a tetragonal lysozyme crystal confirmed by x-ray diffuse scattering. *Acta Crystallographica Section D: Biological Crystallography*, 52(4):722–729.

[Phillips, Jr. et al., 1980] Phillips, Jr., G. N., Fillers, J. P., and Cohen, C. (1980). Motions of tropomyosin. Crystal as metaphor. *Biophysical Journal*, 32(1):485–502.

[Riccardi et al., 2010] Riccardi, D., Cui, Q., and Phillips, Jr., G. N. (2010). Evaluating elastic network models of crystalline biological molecules with temperature factors, correlated motions, and diffuse x-ray scattering. *Biophysical Journal*, 99(8):2616–2625.

[Tilton et al., 1992] Tilton, R. F., Dewan, J. C., and Petsko, G. A. (1992). Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at nine different temperatures from 98 to 320 K. *Biochemistry*, 31(9):2469–2481.

[Van Benschoten et al., 2016] Van Benschoten, A. H., Liu, L., González, A., Brewster, A. S., Sauter, N. K., Fraser, J. S., and Wall, M. E. (2016). Measuring and modeling diffuse scattering in protein x-ray crystallography. *Proceedings of the National Academy of Sciences of the United States of America*, 113(15):4069–4074.

[Vaney et al., 1996] Vaney, M., Maignan, S., Ries-Kautt, M., and Ducriux, A. (1996). High-resolution structure (1.33 A) of a HEW lysozyme tetragonal crystal grown in the APCF apparatus. Data and structural comparison with a crystal grown under microgravity from SpaceHab-01 mission. *Acta Crystallographica Section D: Biological Crystallography*, 52(3):505–517.

[Wall, 2009] Wall, M. E. (2009). Methods and Software for Diffuse X-Ray Scattering from Protein Crystals. In Foote, R. S. and Lee, J. W., editors, *Micro and Nano Technologies in Bioanalysis*, volume 544 of *Methods in Molecular Biology*, pages 269–279. Humana Press, Totowa, NJ.

[Wall, 2018] Wall, M. E. (2018). Internal protein motions in molecular-dynamics simulations of Bragg and diffuse X-ray scattering. *IUCrJ*, 5:172–181.

[Wall et al., 1997a] Wall, M. E., Clarage, J. B., and Phillips, Jr., G. N. (1997). Motions of calmodulin characterized using both Bragg and diffuse x-ray scattering. *Structure*, 5(12):1599–1612.

[Wall et al., 1997b] Wall, M. E., Ealick, S. E., and Gruner, S. M. (1997). Three-dimensional diffuse x-ray scattering from crystals of Staphylococcal nuclease. *Proceedings of the National Academy of Sciences of the United States of America*, 94(12):6180–6184.

[Wall et al., 2014] Wall, M. E., Van Benschoten, A. H., Sauter, N., Adams, P. D., Fraser, J. S., and Terwilliger, T. C. (2014). Conformational dynamics of a crystalline protein from microsecond-scale molecular dynamics simulations and

diffuse x-ray scattering. *Proceedings of the National Academy of Sciences of the United States of America*, 111(50):17887–17892.

[Wall et al., 2018] Wall, M. E., Wolff, A. M., and Fraser, J. S. (2018). Bringing diffuse X-ray scattering into focus. *Current Opinion in Structural Biology*, 50:109–116.

[Welberry, 2004] Welberry, T. R. (2004). Diffuse X-Ray Scattering and Models of Disorder. In *IUCr Monographs on Crystallography*, volume 16, pages 69–70. Oxford University Press, Oxford.

[Williams et al., 2018] Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (2018). MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science*, 27(1):293–315.

[Winn et al., 2011] Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A. J., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*, 67(4):235–242.

[Yamamoto et al., 1997] Yamamoto, T., Juneja, L. R., Hatta, H., and Kim, M., editors (1997). *Hen Eggs: Their Basic and Applied Science*. CRC Press LLC, Boca Raton.

[Zeldin et al., 2013] Zeldin, O. B., Gerstel, M., and Garman, E. F. (2013). RADDOSE-3D : time- and space-resolved modelling of dose in macromolecular crystallography . *Journal of Applied Crystallography*, 46(4):1225–1230.