Diffuse Features in X-ray Diffraction from Protein Crystals

Michael E. Wall

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of Physics

January, 1996

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Sol Gruner, for his foresight, understanding, and endless patience in guiding me in this research.

Special thanks to Shyam for introducing me to biological physics, and to Uncle Bob for taking an interest in seeing me "saved."

Formal thanks to Rick Walter for technical support on all levels; Steve Ealick for access to the many resources at Cornell; Sandor Barna, Eric Eikenberry, Shyamsunder Erramilli, Damian Hajduk, Yuqi Han, Fredrik Österberg, John Shepherd, Mark Tate, Dan Thiel and Paul Urayama for assistance in data-taking at CHESS; Mark Tate for his part in creating beautiful x-ray detectors; Teresa Gamble, Apo Gittis, Ed Lattman, and Patrick Loll for excellent crystals of *Staph.* nuclease; David Shortle for a hands-on explanation of the bacterial expression of *Staph.* nuclease; Wayne Volkmuth for help in navigating me through the UNIX labyrinth; Eric Eikenberry for useful criticism and suggestions; and Joel Berendzen for the timely suggestion of using a mode filter to remove Bragg peaks from diffraction images.

Informal thanks to Benk for incredible quantities of love, Mom and Dad for continued faith and support, Steve for being a high-adventure role model, Irwin and Sarah for early encouragement, Sarah for diner enthusiasm, Damian for lunch-time consistency and excellent dinner parties, John for great beer tastings, Tim and Ieva for documentary appreciation, JC for inexplicably engaging conversations about odd topics, Richard and Hilary for a timely trip to Colorado, and everyone else who made the past six years incredibly worthwhile.

Abstract

The use of diffuse x-ray scattering to characterize crystalline protein dynamics is investigated. A method is developed for obtaining three-dimensional maps of diffuse intensity from images of x-ray diffraction from protein crystals. Maps of diffuse scattering from *Staphylococcal* nuclease crystals have been obtained experimentally, and a measure of internal consistency based upon symmetry constraints is defined. The maps have the symmetry one would expect from the $P4_1$ space group of the unit cell. The reproducibility of diffuse scattering measurements is quantified; the difference between diffuse maps obtained from nominally identical crystals is as low as 20% in a resolution shell spanning 4.1 **Å** – 3.6 **Å**. A polar-coordinate *shell image* graphing scheme, used to observe diffuse features in three-dimensional reciprocal space, is described. Differences are observed in diffuse scattering from crystalline *Staph.* nuclease with and without Ca^{++} and the substrate-analog pdTp (thymidine-3′,5′-diphosphate) bound. These differences are interpreted as the observation of a change in the dynamics of *Staph.* nuclease upon binding of Ca^{++} and pdTp.

Some models of diffuse scattering are discussed, and a new model of diffuse scattering based upon independent rigid-body motions of crystalline protein molecules is described. Simulated three-dimensional diffuse maps of *Staph.* nuclease are generated, and these maps are compared to the maps obtained experimentally. By these comparisons, it is shown that the nature of disorder which gives rise to diffuse scattering in *Staph.* nuclease is most like both the liquid-like correlated motions described by Caspar *et. al.* [Nature **332** (1988) 659], and the more traditional thermal excitation of crystalline normal modes discussed in Glover *et. al.* [Acta Cryst **B47** (1991) 960].

Contents

List of Figures

List of Tables

SYMBOLS AND ABBREVIATIONS

$\mathbf{a}, \mathbf{b}, \mathbf{c}$,	primitive lattice vectors
$\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$,	primitive reciprocal lattice vectors
$f(\mathbf{q})$,	structure factor
$\mathbf{k}$,	incident x-ray momentum
$\mathbf{k}'$,	scattered x-ray momentum
$\mathbf{q}$,	$\mathbf{k}' - \mathbf{k}$
$\mathbf{s}$,	$\frac{\mathbf{q}}{2\pi}$ = scattering vector
$I(\mathbf{q})$,	scattering intensity
I_B,	Bragg intensity
I_D,	diffuse intensity
Q,	charge
$\mathbf{R}$,	lattice vector
V,	variance matrix
$\rho(\mathbf{x})$,	electron-density distribution

ADU,	analog-to-digital unit = 1 count
CHESS,	Cornell High-Energy Synchrotron Source
NMR,	nuclear magnetic resonance
RMS,	root-mean-square

Chapter 1

Introduction

Knowledge of enzymatic structure encourages us to think of catalysis mechanically as well as chemically. In modern papers offering a discussion of the nature of the binding of a substrate to a protein, one will find a description of how the substrate is meant to "sit" in the vicinity of the *active site*, the place where the key catalytic groups are found. Accompanying this description will be an analysis of the proximities of atoms on the substrate and protein which are thought to create favorable conditions for binding, *e.g.*, by presenting opposite charges or offering the correct geometry for hydrogen bonds. In mechanical models, however, objects are allowed to have moving parts, and a complete description of a mechanical system necessitates a characterization of internal dynamics as well as kinetics.

Given the importance of understanding protein function in understanding disease and improving rational drug design, there is still much debate about whether or not the role of dynamics will be central or peripheral in describing how proteins work. A characterization of internal dynamics may be necessary to understand the function of proteins at even a primitive level: for example, meaningful dynamical parameters may be necessary to quantify the relation between structure and enzymatic activity. On the other hand, static structural models may prove to be adequate to usefully characterize the problem.

In order to experimentally determine the role of internal dynamics in protein function, it will be necessary to identify dynamical parameters of proteins which can be systematically varied and measured in a controlled fashion. Such a parameter may be something as specific as the amplitude of the displacement of a particular domain of the protein, or as general as a measure of protein elasticity. The connection with protein function must then be made experimentally by monitoring a measure of activity while varying the observed dynamical parameters.

This work experimentally demonstrates that three-dimensional maps of diffuse x-ray scattering may be used to identify the nature of disorder in protein crystals. It is shown that, in some cases, the disorder can be related to the internal dynamics of the protein, in which case careful measurements of diffuse scattering can provide measurements of dynamical parameters. An extension of the methods described here to a range of experimental systems may suggest new ways of quantitifying the relation between protein dynamics and protein function, thereby assessing the significance of dynamical parameters in predicting protein activity.

1.1 The Dynamic Enzyme

There is little disagreement that proteins are intrinsically dynamic objects – a point which has been asserted in the literature time and time again [1, 2]. There is ample evidence for the existence of spontaneous fluctuations in proteins. Lakowicz and Weber [3] in 1973 found evidence for nanosecond structural fluctuations using fluorescence quenching techniques. Experiments of Englander et *et. al.*[4] and Woodward and Hilton [5] indicate that nearly all sterically “inaccessible” protons in native proteins can be substituted in the presence of solvent, requiring transient openings of 4 **Å** channels in the structure [1]. The non-exponential recombination kinetics of cryogenic myoglobin and carbon monoxide observed by Austin et *et. al.*[6] imply the existence of a distribution of functionally distinguishable conformations in an ensemble under biological conditions.

In a 1976 note to PNAS[7], Alan Cooper reflected on some remarkable aspects of thermodynamic fluctuations in protein molecules. The note was written to resolve a "paradox" observed in experimental data on proteins. By this time, there was ample evidence for a rather fluid structure for globular proteins, as given by fluorescence[3, 8, 9], phosphorescence[10] and NMR experiments[11, 12, 13]. In apparent contrast to this were crystallographic studies, which revealed solid-like packing densities in proteins, and measurements of the heat capacities of proteins, which at $0.30-0.35\ cal/g/C^{\circ}$ were larger than those of organic liquids but comparable to those of crystalline amino acids ($0.316\ cal/g/C^{\circ}$) at $25^{\circ}C$.

Against this background, there were some incorrect thermodynamic arguments being floated at the time, claiming that large fluctuations in single proteins would necessarily cause large fluctuations in thermodynamic properties of protein ensembles. It is in the course of resolving these issues that Cooper points out that the RMS fluctuations in both the internal energy and volume of a protein are comparable to the changes in these parameters upon protein denaturation. There was a misunderstanding that this implied that macroscopic measurements would be correspondingly imprecise.

Cooper also points out that the third moment of the energy distribution of protein ensembles is high enough to make the distribution asymmetric, implying that measured mean thermodynamic parameters, such as energy or volume, do not necessarily represent the most likely value for the parameter in the ensemble. He points out that the reason why measurements of thermodynamic parameters are so precise is that there is no coherence in the fluctuations of individual protein molecules in an ensemble. RMS fluctuations scale like $N^{-\frac{1}{2}}$, where N is the number of proteins in the ensemble, and N is large enough to account for experimental precisions observed. Even though the issue being resolved was quite trivial, he nevertheless brought up an extremely important point in the process – even in a macroscopic sense, a complete understanding of protein function requires a characterization of the distribution of states in the ensemble. Measurements of mean thermodynamic parameters do not suffice.

1.2 Diffuse X-ray Scattering

Recent experiments in x-ray diffraction from protein crystals suggest that the technique yields information about the distribution of protein conformational states[14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24]. For example, traditionally, measurements of Bragg reflections in x-ray diffraction from protein crystals have been used to obtain atomic structural models of protein molecules. In the work of Frauenfelder *et. al.*[14] on metmyoglobin, measurements of Bragg reflections were used to measure independent atomic B-factors, which which are related to rms fluctuations of atomic positions about equilibrium.

Early experiments by Sternberg *et. al.*[15] on lysozyme and later work by Kuriyan and Weis[16] on a number of proteins showed a good correlation between independent atomic B-factors obtained by Bragg reflection measurements and the predictions of a model where proteins independently execute rigid-body rotations and translations in the crystal. Kuriyan reported remarkable consistency between B-factors simulated using only ten independent free parameters of a rigid-body motion model and those obtained experimentally by modelling atomic motions independently with orders of magnitude more free parameters.

In addition, diffracted x-ray intensity between Bragg peaks, termed *diffuse scattering*, has been recognized as a source of information about protein dynamics. Images of diffuse scattering have been simulated using various models of molecular motion, including crystal acoustical modes [21], coherent unit-cell subdomain motions[18], protein subdomain motions[17], "liquid-like" motions correlated over a characteristic length scale[19, 22, 23], and molecular dynamics and normal modes analysis of isolated proteins[24]. In all cases, simulated images have been obtained which bear common visual features with measured diffraction images.

Although previous diffuse scattering studies have used single x-ray exposures as the source of diffraction data, there is a great advantage to having a full three-dimensional data set constructed from exposures taken at multiple crystal orientations. Firstly, the full three-dimensional data set can be checked for internal consistency using symmetry constraints implied by the unit-cell space group. Secondly, one ignores most of the available data by limiting the measurement to a slice of reciprocal space probed by a single exposure. A full three-dimensional measurement provides good discrimination between the various models in simulating the data; although the various studies of diffuse scattering have convincingly reproduced features observed in single exposures, there is as yet no convincing argument in the literature for the validity of one proposed model over another.

We initially measured three-dimensional maps of diffuse x-ray scattering from *Staphylococcal* nuclease, an extracellular protein which cleaves both DNA and RNA by hydrolyzing P-O bonds[25], with both Ca^{++} and pdTp (thymidine-$3'$,$5'$-diphosphate) bound, in order to quantify the internal consistency in the measured diffuse scattering and to determine the reproducibility of measurements of diffuse scattering, both of which are difficult to do with only single x-ray exposures from a crystal. Since the features we observed in these maps were seen to be reproducible, we obtained a map of diffuse scattering from *Staph.* nuclease without Ca^{++} and pdTp bound, and observed differences between this map and the original ones. These differences are interpreted as evidence for changing dynamics in crystalline *Staph.* nuclease upon binding of Ca^{++} and pdTp.

We chose *Staph.* nuclease for several reasons:

- Crystals of *Staph.* nuclease yield measurable diffuse x-ray scattering.
- A high-resolution crystallographic structure of the protein had been previously determined both with and without Ca^{++} and pdTp.[26, 27, 28, 29]
- The protein crystallizes in space group $P4_1$, providing many symmetry-related measurements to check for the internal consistency of diffuse scattering maps.
- The stability of the protein to denaturation has been mapped out in pressure[30], providing a good background for further experiments.

In order to interpret our measured diffuse maps in terms of protein dynamics, we investigated the predictions of a range of models of diffuse scattering. Included in this work is a derivation of a new model of diffuse scattering based upon independent rigid-body rotations and translations of protein molecules in the crystal. No comparison was made with a normal-modes model of protein fluctuations.[1] In order to compare the predictions of models with our data, simulated three-dimensional maps were generated with the aid of the package XPLOR by Axel Brunger. Our simulations show that our data best agrees with both the liquid-like motions model of Caspar *et. al.*[19], and the crystalline acoustical modes model used by Glover *et. al.*[21]. The equivalence of these models is discussed in section 3.9.

---

[1] See, *e.g.*, Levitt *et. al.*[31] and Go *et. al.*[32] for a discussion of normal-modes models of protein fluctuations.

1.3 Historical Survey

Probing protein dynamics with x-rays elastically scattered from protein crystals has a relatively recent history, spanning only about the last fifteen years. What follows is a survey of some studies of protein dynamics using this experimental technique. The papers below discuss both B-factor analysis using Bragg reflections, and analysis of diffuse x-ray scattering exposures to reveal characteristics of the structural fluctuations of proteins in a crystalline state.

1.3.1 Frauenfelder, Petsko and Tsernoglou

Frauenfelder *et. al.*[14] in 1979 reported the observation of protein structural dynamics in x-ray diffraction from myoglobin. The RMS deviations of all nonhydrogen atoms (1261 in all) were added as free parameters in a structural model refined against measurements of Bragg intensities. The result was a map of RMS deviations for the entire protein, in addition to a structural model, which was then analyzed to give evidence for the existence of conformational substates of myoglobin in the crystal.

Indirect evidence for a distribution of conformational substates was given through a temperature study whose results were used in conjunction with Mössbauer spectroscopy results cited from Debruner *et. al.*[33]. After subtraction of a component due to lattice disorder which was estimated by Mössbauer measurements, an interesting temperature profile for the remaining mean square deviation, $\langle x^2 \rangle$, averaged over all atoms in the crystal was observed. Below a critical temperature T_c, $\langle x^2 \rangle$ has a more gentle slope than above T_c, which is interpreted in terms of the "freezing in" of the distribution of conformational substates in the crystal, or the arrest of the large-scale motions of the molecule. This is in accord with the putative glass transition in myoglobin at cryogenic temperatures identified by Austin *et. al.*[6].

1.3.2 Sternberg, Grace and Phillips

About the same time as the Frauenfelder *et. al.*[14] myoglobin analysis was published, Sternberg *et. al.*[15] published an analysis of the atomic temperature factors of lysozyme. Based upon an as yet unpublished refinement of individual isotropic atomic temperature factors in a 2 **Å** structural model of tetragonal hen egg-white lysozyme, they obtained a least-squares best estimate of the parameters for models of rigid-body rotational and translational disorder, and for a couple of proposed breathing modes. The rigid-body model used was based upon the so-called TLS model of Schomaker and Trueblood[34], in which three 3x3 matrices are proposed to parametrize the disorder. In the general model, the rotations are parametrized by six independent elements of the symmetric librational matrix L, translations are parametrized by six independent elements of the symmetric translational matrix T, and the coupling between translations and rotations are parametrized by eight independent elements of the matrix S, which has an indefinite trace. These matrices, combined with the position vector of the atom in question, yield the anisotropic B-factor, from which the rms displacement of the atom along any direction can be calculated. In the work of Sternberg *et. al.*[15], the *isotropic* B-factors are obtained during refinement, which only depend upon ten of the twenty free parameters of the TLS model. It is these ten free parameters which are refined against the 129 B-factors of the lysozyme model.

The model for breathing modes was a hinge-bending mode, where two lobes of the molecule independently execute libration about an assumed hinge axis. In the implemented model, there is only one free parameter once the hinge axis is given, and that parameter is essentially the angular amplitude h of the librations of the two lobes, which are assumed to be equal. The isotropic B-factors are determined solely by h and the distance d from the atom in question to the hinge axis. It can be shown that the hinge-bending model produces the same B-factors as a model which uses only T and L with a center of libration on the hinge axis, as is pointed out in the paper.

One should note that without translational parameters, the B-factors are essentially anisotropic for atoms fixes to a rigid rotator, since the motion is confined to a plane perpendicular to the vector to the center of libration. Isotropic B-factors describe a spherically symmetric distribution of atomic displacements, so that fitting atomic displacements due to rigid-body rotational disorder using isotropic B-factors may be a risky procedure. Nevertheless, Sternberg *et. al.*[15] report that the measured isotropic B-factors are consistent with both the TLS and breathing/TL models, with a slightly better fit using the TLS model, even accounting for the use of additional free parameters. In the end, they suggest that diffuse scattering measurements may provide further evidence for the presence of rigid-body motions.

1.3.3 Phillips, Fillers and Cohen

In 1980, Phillips *et. al.*[17] reported observing diffuse features in x-ray diffraction from tropomyosin crystals, which they then interpreted in terms of transverse motions of the filaments. By simulating diffraction data from a structural model undergoing 8 **Å** fluctuations in the long arm and 5 **Å** fluctuations in the short arm of the protein, they were able to qualitatively reproduce the streaks observed in the diffraction pattern. These studies were later followed up by a more detailed analysis of the diffuse scattering, which characterized the motions to a greater extent[35]. These studies of tropomyosin provide one of the first mentions of evidence for direct observation of large-scale motions in a protein through identification of diffuse features.

1.3.4 Doucet and Benoit

In 1987, an interpretation of x-ray diffuse scattering from protein crystals came from Doucet and Benoit[18]. They observed strong diffuse streaks concentrated along two families of Bragg planes in $P2_12_12_1$ orthorhombic hen egg-white lysozyme. Using these features, they argue for the existence of intermolecular rigid-body displacements which are correlated along the a- and c-axes of the crystal.

To model the data, the scattering intensity was considered in three parts: the Bragg intensity I_B, giving rise to sharp diffraction peaks from the long-range periodic order of the crystal; the thermal diffuse scattering, I_{TDS}, which comes from "collective modes" in the crystal and gives rise to the smoother intensity gradient in the neighborhood of the peaks; and the diffuse scattering I_D, which is due to the motion of independent correlated "superunits" of electron density, and gives rise to the larger scale features such as the streaks identified in the data. The authors report the best agreement between simulation and experiment when they assume that there are four pairs of lysozyme molecules contributing to the diffuse streaks: two pairs stacked along the a-axis executing displacements along the a-axis, and two pairs stacked along the b-axis executing displacements along the c-axis. The mean square amplitudes of the displacements along each of these axes is "extracted" from the data, and is given as $\langle \mathbf{u}_a^2 \rangle = (5 \pm 3)10^{-4}\text{nm}^2$ for **a**-axis displacements and $\langle \mathbf{u}_c^2 \rangle = (7 \pm 3)10^{-4}\text{nm}^2$ for **c**-axis displacements.

In a recent review article[36], Benoit and Doucet present a broad summary of diffuse scattering in protein crystallography, in which they consider many different models of disorder, and cover many previous experiments. In the review, there is an emphasis on simulating diffuse scattering through normal modes analysis and molecular dynamics simulations, neither of which is touched on in this work. In addition, there is a description of some of the experimental issues involved in collecting diffuse scattering data, providing useful suggestions for measurement procedures.

1.3.5 Caspar, Clarage, Salunke and Clarage

Caspar *et. al.*[19] in 1988 reported using a different modelling scheme to interpret diffuse scattering from a R3 hexagonal 2Zn pig insulin crystal. They describe the diffuse features in terms of two kinds of motion: one in which lysozyme molecules move rigidly, independent of all other molecules but nearest neighbors; and one in which the entire crystal softens to atomic displacements which are correlated within a 6 **Å** sphere about any atom. The former kind of motion is meant to produce "haloes" about Bragg peaks, and is analogous to the I_{TDS} described in the work of Doucet and Benoit. The latter kind of motion is proposed as the cause for the large-scale, smoothly varying, asymmetric background.

Caspar *et. al.*[19], like Doucet and Benoit[18], demonstrate a qualitative agreement between the results of simulations and the x-ray exposure.

1.3.6 Glover, Harris, Helliwell and Moss

In 1991, Glover *et. al.*[21] published the results of a survey of diffuse scattering from a variety of protein crystals. The theme of the paper is that macromolecules, being large and complex in structure, will have complex motions which are likely to vary depending upon the molecule under study. Furthermore, different macromolecules will give different signatures in the diffuse scattering due to the different nature of the motions.

Many examples of diffuse scattering are given, including exposures from ribonuclease (RNAse), transferrin, γ-II crystallin and t-RNA(MET). They find a wide variation in features from crystal to crystal, with a common feature of strong diffuse scattering in the "solvent" ring. They also observed the effects of chemically altering the *gamma*-II crystallin crystal by initiating crosslinking through the addition of 2.5% gluteraldehyde, and by preparing heavy-atom derivatives. Both modifications suppress Bragg peaks at high resolution and enhance diffuse features.

Two models are discussed as a means of interpreting the data. The first is an "Einstein crystal" model, where each atom vibrates independently with the same mean-squared displacement. Since each atom has an approximately spherically-symmetric electron density distribution, the resulting diffuse intensity has spherical symmetry in such a model (see chapter 3 for a discussion of this model). This model is dismissed by the authors on the grounds that simulated densitometered radial profiles calculated using a wide range of parameters fail to reproduce the features of the profiles generated from diffraction images. One may also dismiss such a model by observing that there is significant asymmetry in the diffuse scattering, particularly in the neighborhood of the solvent ring.

The second model discussed involves an analysis of the scattering in the neighborhood of Bragg peaks. Bragg peak analysis reveals a q^{-2} dependence in the intensity distribution, which is characteristic of scattering from thermally excited acoustic modes in a crystal. The form of the scattered intensity that they suggest agrees with that expected in scattering from an electron density distribution on a lattice with thermally excited acoustic modes.

The paper succeeds in demonstrating the wide variety of diffuse scattering which is visible in x-ray scattering from macromolecular crystals. It also presents convincing evidence for the presence of first-order thermally excited acoustic modes in RNAse(MET), which shows q^{-2} tails in low-resolution Bragg peaks.

Chapter 2

Experimental Methods

Three-dimensional maps of diffuse x-ray scattering were obtained experimentally from three crystals described in this work. Procedures in obtaining these maps are categorized into *data collection*, *image processing*, and *map generation*. The discussion of these procedures below defines what is meant by a *diffuse map*, and provides a useful background for discussing the models of chapter 3 and the analysis in chapter 4.

2.1 Data Collection

2.1.1 Materials

Diffuse data was collected using two crystals of *Staph.* nuclease grown over 23% MPD (2-methyl-2,4-pentanediol) 10.5 μM potassium phosphate with 1.1 M equivalent pdTp (thymidine-3$'$,5$'$-diphosphate) and 2 M equivalent Ca^{++}, and one crystal of *Staph.* nuclease grown over 23% MPD 10.5 μM potassium phosphate. All crystals had a tetragonal unit cell with space group $P4_1$. The first co-crystal measured 2.0 $mm \times 0.4\ mm \times 0.4\ mm$, with unit cell parameters $a = b = 48.5$ **Å**, $c = 63.4$ **Å**, $\alpha = \beta = \gamma = 90°$, as determined by the program DENZO written by Zbyszek Otwinowski and Wladek Minor. The second co-crystal measured 2.2 $mm \times 0.4\ mm \times 0.4\ mm$, with unit cell parameters $a = b = 48.3$ **Å**, $c = 63.4$ **Å**, $\alpha = \beta = \gamma = 90°$. The bare *Staph.*

nuclease crystal was selected from a batch of average size approximately 0.4 $mm \times 0.2\ mm \times 0.2\ mm$, and had unit cell parameters $a = b = 48.2$ **Å**$, c = 63.9$ **Å**$, \alpha = \beta = \gamma = 90°$.

Co-crystals of *Staph.* nuclease were provided by Teresa Gamble of Genentek, who mounted them in capillaries with mother liquor and shipped them on cold packs to Princeton. The mother liquor was wicked away at the beamline and the capillary was fixed to a goniometer head.

The crystal of bare *Staph.* nuclease were was grown by Jennifer Ekstrom in Professor Steve Ealick's lab at Cornell University. It was transferred from a hanging drop to a capillary shortly before experimentation at CHESS.

All crystals were mounted with the $c-$axis nearly parallel to the capillary.

2.1.2 Beamline issues

We used a CCD area x-ray detector designed at Princeton and installed on the F1 and A1 beamlines at the Cornell High Energy Synchrotron Source (CHESS) for data collection. The detector was similar in design to the CCD detector described by Tate *et. al.*[37]. It has a dynamic range of roughly 10^4 x-rays/pixel, and an active area of about 80×80 mm^2, subdivided into an array of 2048×2048 pixels. Pixels values were binned both horizontally and vertically during readout, producing a 1024×1024 image; this was done to save disk space, decrease image-transfer times, and improve statistics. Pixel values mapped to x-ray counts by a ratio of roughly one-to-one.

An "anti-blooming" procedure[1], where overflowing electrons from saturated wells in the CCD are channelled off the device during integration, was implemented to improve the handling of strong Bragg reflections, which can otherwise distort intensity

---

[1] Described in a personal communication by James Janesick of Pixel Vision, Huntington, CA 92649

measurements in the neighborhood of the peak. The beam was tuned to 0.91 **Å**, had a polarization of 0.8-0.93 perpendicular to the beam in the plane of the synchrotron ring, and was collimated to a 100 μm diameter.

The data collection protocol for each crystal involved interleaving the collection of two data sets, one being a set of two degree *oscillation exposures* to be used in calculating the orientation of the crystal and in refining a structural model, and the other being a set of *stills* spaced one degree apart in spindle rotation, from which diffuse intensity was obtained. Since we are not interested in studying variations in diffuse intensity on a smaller scale than the separation between Bragg peaks, the set of stills adequately samples reciprocal space for our purposes.

All exposures were five seconds long. The data set for the first co-crystal spanned 94 degrees of spindle rotation, and the data set for the second co-crystal spanned 96 degrees of spindle rotation. The data set for the bare *Staph.* nuclease crystal spanned 90 degrees of spindle rotation. Figure 2.1 shows a typical still exposure from these data sets. Data was collected only at room temperature.

Great care was taken to minimize parasitic scattering in x-ray exposures (see figure 2.2 for a reference schematic). The beam stop was adjusted such that the tails of the main beam were blocked as well as possible. A lead sheath with a 1 mm hole at the end was slipped over the collimator, which was a known source of parasitic scattering. A large lead shield was placed before the collimator in order to eliminate static, hard x-ray patterns which contaminated early exposures. In addition, to minimize systematic asymmetries in the background, a rectangular mylar sheet was used in place of the usual strip of kapton tape to hold the beam stop.

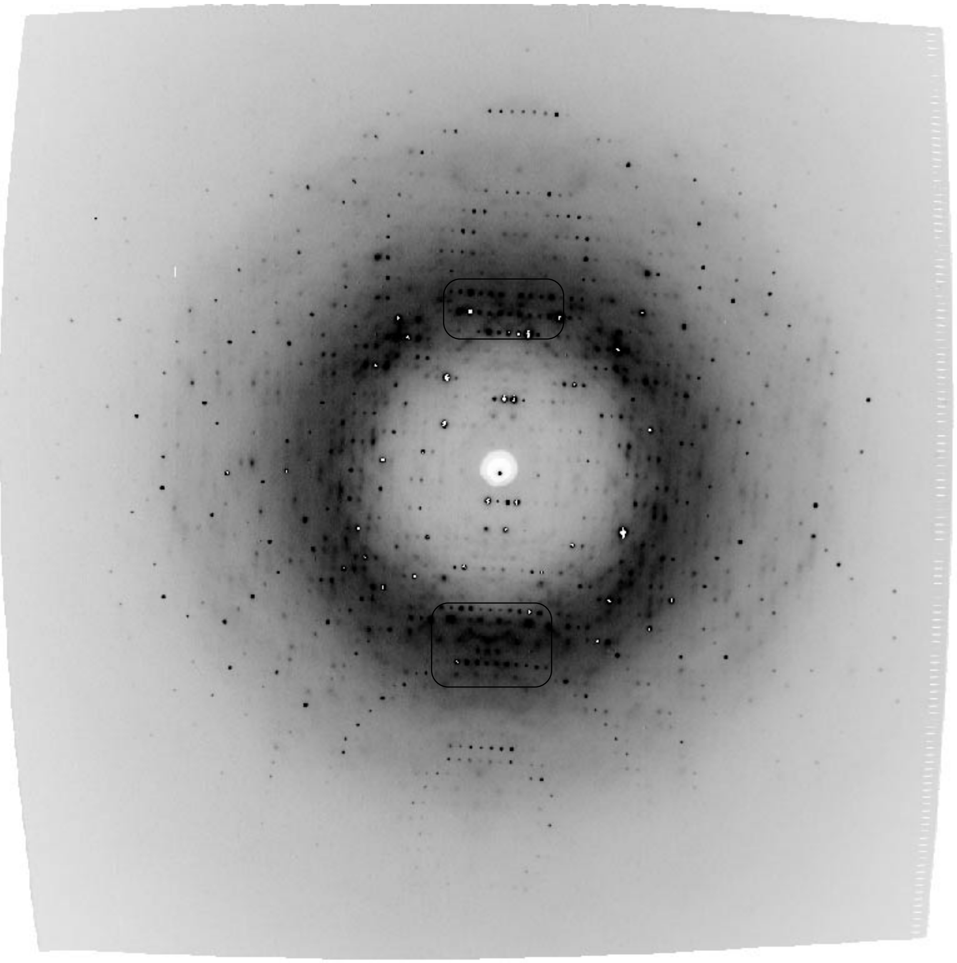

Figure 2.1: *Raw image of a still exposure of* Staph. *nuclease with* Ca^{++} *and pdTp.* Note the structure in the diffuse scattering. Two strong features are indicated which suggest that the symmetry in the diffuse scattering may be consistent with the unit cell space group $P4_1$. We wish to investigate the internal consistency and reproducibility of such features in the absence of the Bragg peaks.

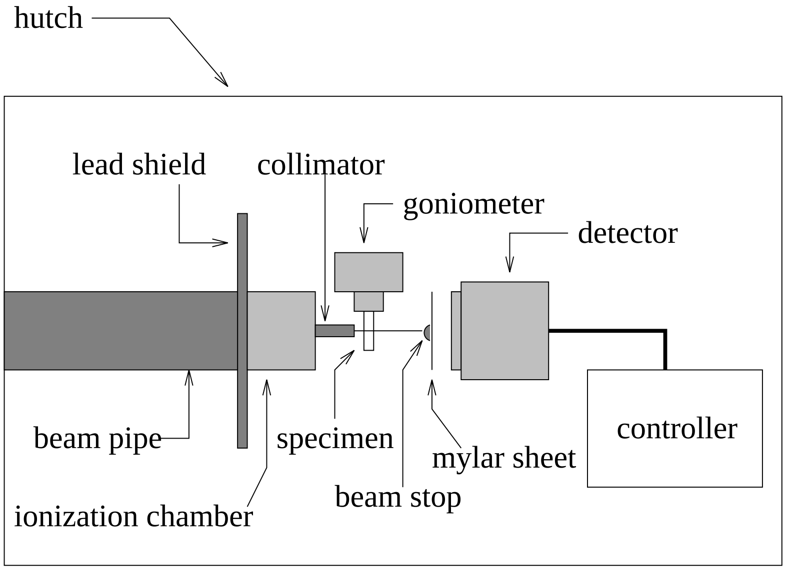

Figure 2.2: *Experimental schematic.*

2.2 Image Processing

2.2.1 Bragg peak indexing

The intensities of Bragg reflections were measured and assigned Miller indices from oscillation exposures using the programs DENZO and SCALEPACK by Zbyszek Otwinowski and Wladek Minor. Estimates of the sample-to-detector distance and beam position were entered as inital parameters to DENZO for the first image in each data set. After a suitable fit was found to the unit cell parameters, crystal orientation, detector face rotation, and other parameters, these parameters were used as a starting point for indexing the rest of the data set in batch mode. In this way, a set of refined experimental parameters were produced for each exposure in the data set.

SCALEPACK was used to assimilate the parameters from all data sets into a single best-fit prediction for the unit cell parameters, the mosaicity, and the orientation of

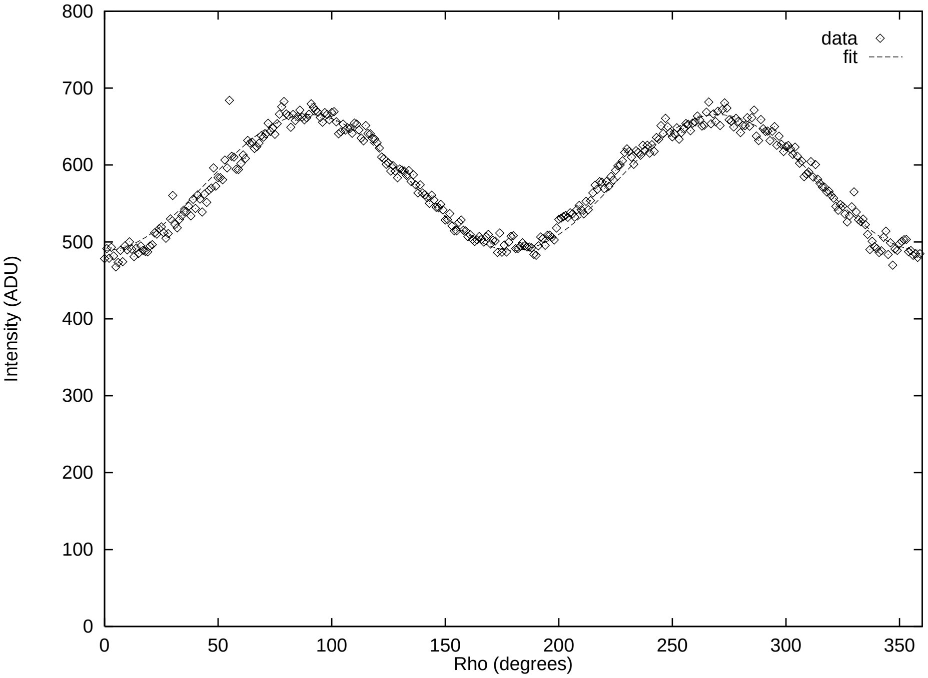

Figure 2.3: *Determination of beam polarization for crystal 2.* The azimuthal intensity profile was obtained at a scattering angle of $\theta = 16°$ from a mode-filtered diffraction image (see section 2.2.4). The least-squares best fit to the expression $\frac{a}{2}(1 + \cos^2 2\theta - \epsilon \cos 2(\rho - \rho_0) \sin^2 2\theta)$ is shown, with a fixed parameter $\theta = 16°$, and variable parameters $\rho_0 = 0.0$, $a = 672$, and polarization $\epsilon = 0.93$.

the crystal at the beginning of the data set. These parameters were fed into DENZO for a second run, and for each data set the process was repeated until the fit did not improve after several iterations. In this way, a list of measured Bragg reflections was obtained, along with the unit cell parameters, crystal orientation, and mosaicity.

2.2.2 Polarization correction

Images were corrected for synchrotron beam polarization using the formula[2]

$$I' = \frac{2I_{pixel}}{(1 + \cos^2 2\theta - \epsilon \cos 2\rho \sin^2 2\theta)}, \tag{2.1}$$

---

[2] See P. 304 Giacovazzo[38].

where ϵ parametrizes the polarization of the beam, θ is the scattering angle to the pixel, and ρ is the angle between the plane of the synchrotron ring and the vector in an image from the beam fiducial mark to the pixel. The polarization ϵ was determined experimentally by analyzing the azimuthal variation of the background intensity at a fixed scattering angle θ. An azimuthal intensity profile was obtained from a mode-filtered diffraction image, and was modelled by the expression $\frac{a}{2}(1+\cos^2 2\theta-\epsilon\cos 2(\rho-\rho_0)\sin^2 2\theta)$, where the additional parameter ρ_0 has been added to account for detector-face rotation. Figure 2.3 shows the results of a fit obtained for crystal 2 at $\theta = 16^\circ$, where the best-fit polarization was $\epsilon = 0.93$.

2.2.3 Solid angle normalization

Pixel values were also normalized to the solid angle subtended by the pixel about the specimen, in order to produce a map that, at every position in reciprocal space, was simply proportional to $I(\mathbf{q}) = |F(\mathbf{q})|^2$, the square of the Fourier transform of the electron density of the crystal. This yields a correction of the form

$$I'' = I' \cos^3 \psi, \tag{2.2}$$

where $\psi = 2\theta$ is the angle between the scattered and incident beam (see figure 2.4). The correction comes about by an analysis of how to relate pixel values to the electron density of the crystal, as is described in the following pages.

Our digitized images of x-ray diffraction consist of about one million 16-bit integers which map to the number of x-rays incident on each of 1024 x 1024 pixels at the face of an area x-ray detector. Each integer corresponds to a measurement $I(x, y)$ of the

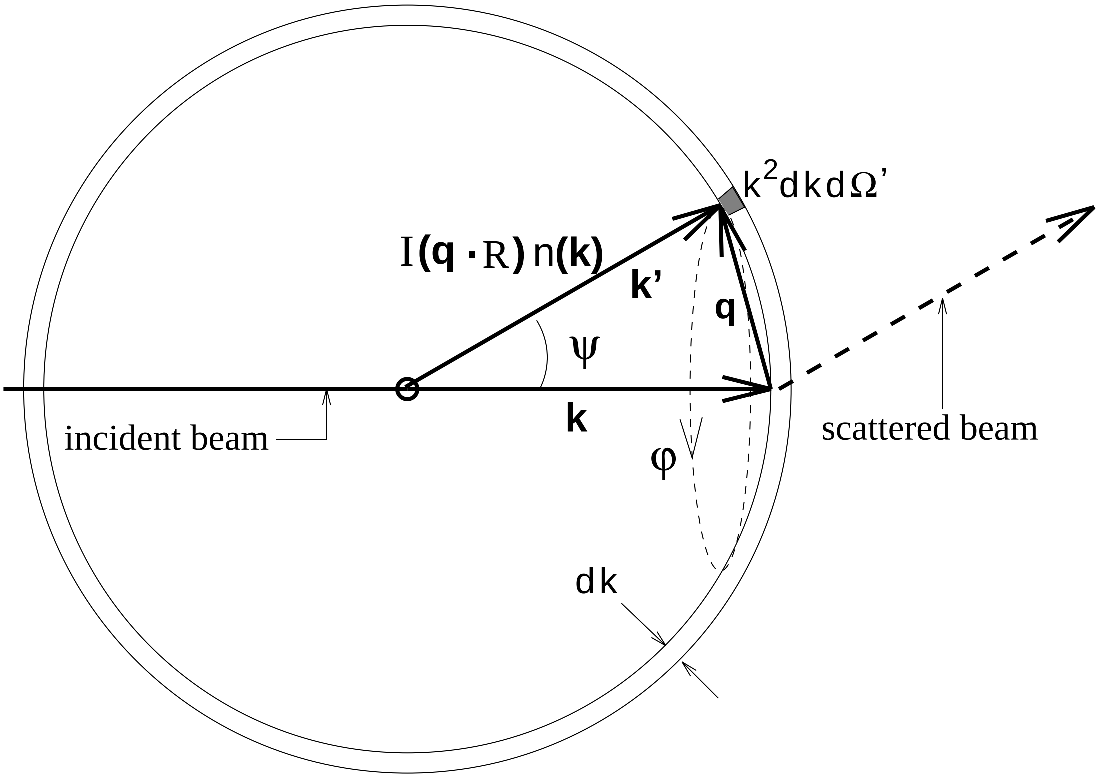

Figure 2.4: *Ewald sphere representation of scattering geometry.* X-rays of momentum $\mathbf{k}$ are scattered into a direction (ψ, ϕ) with probability $I(\mathbf{q} \cdot \mathsf{R})\, n(\mathbf{k})\, k^2\, dk\, d\Omega'$, where $I(\mathbf{q} \cdot \mathsf{R})$ is the scattering intensity for a crystal at orientation R, and $n(\mathbf{k})$ is the momentum distribution of the incident x-rays, which is very sharply peaked in these experiments, since we use monochromatic x-rays. The shaded area is the differential volume $k^2\, dk\, d\Omega'$ pointed to by the scattered x-rays. The illustrated differential spherical shell contains a family of Ewald spheres which define the experimentally accessible region of reciprocal space for momentum k. Given a specified incident beam bandwidth δk and solid angle $\delta\Omega'$ which accounts for crystal mosaicity and beam divergence, the scattered intensity is proportional $\langle I(\mathbf{q} \cdot \mathsf{R}) \rangle_{\delta k\, \delta\Omega'}$.

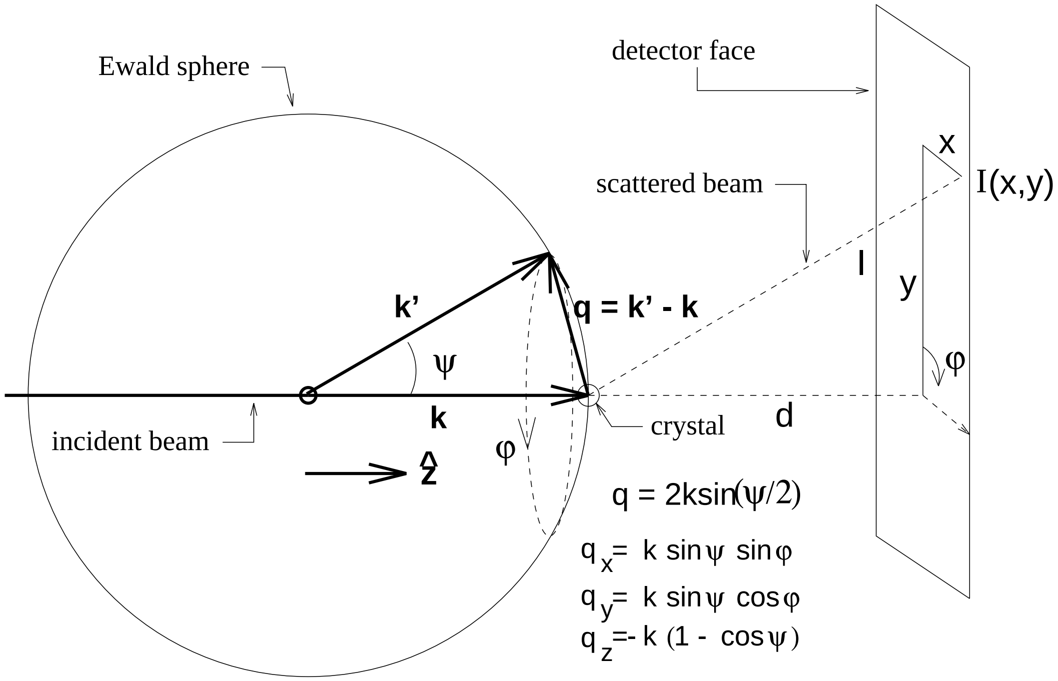

Figure 2.5: *Experimental scattering geometry.*

x-rays scattered into the solid angle $\delta\Omega$ spanned by pixel (x, y):

$$I(x, y) = \int_{\delta\Omega} d\Omega \, \langle I(\mathbf{q} \cdot \mathsf{R}) \rangle_{\delta k, \delta\Omega'}, \tag{2.3}$$

where (x, y) is the position in cartesian coordinates of a pixel in the detector plane referenced to the beam fiducial (see figure 2.5). The distribution $I(\mathbf{q})$ is averaged over a spatial frequency bandwidth δk and solid angle $\delta\Omega'$ in a neighborhood of $\mathbf{q} \cdot \mathsf{R}$, where R is a matrix representing the orientation of the crystal with respect to the coordinate system in which $I(\mathbf{q})$ is calculated. The solid angle $\delta\Omega'$ parametrizes the smearing due to beam divergence and crystal mosaicity (see figure 2.4). If $I(\mathbf{q})$ varies sufficiently slowly over the reciprocal space region spanned by δk and $\delta\Omega'$, we may replace $\langle I(\mathbf{q}) \rangle$ with $I(\mathbf{q})$:

$$I(x, y) = \int_{\delta\Omega} d\Omega \, I(\mathbf{q} \cdot \mathsf{R}), \tag{2.4}$$

which we take to be the case for the diffuse scattering we are studying.

Figure 2.5 shows the scattering geometry for our diffraction experiments. The scattering vector **q** can be expressed in terms of the experimental parameters as follows:

$$\begin{aligned} q_x &= k \sin\psi \sin\phi \\ q_y &= k \sin\psi \cos\phi \\ q_z &= -k(1-\cos\psi), \end{aligned} \tag{2.5}$$

where $\psi = 2\theta$ is the angle between the incident and scattered beam (twice the scattering angle θ), $\hat{\mathbf{z}}$ points along the incident beam, k is the spatial frequency of the x-rays, and ϕ is the azimuthal angle about the **z**-axis. We can express this transformation in terms of the pixel coordinates (x, y) and the sample-detector distance d as follows:

$$\begin{aligned} q_x &= \frac{kx}{\sqrt{x^2+y^2+d^2}} \\ q_y &= \frac{ky}{\sqrt{x^2+y^2+d^2}} \\ q_z &= -k\left(1-\frac{d}{\sqrt{x^2+y^2+d^2}}\right). \end{aligned} \tag{2.6}$$

Assuming that $I(\mathbf{q})$ varies slowly over the solid angle $\delta\Omega$, equation 2.4 becomes

$$I(x, y) = \delta\Omega\, I(\mathbf{q}\cdot\mathsf{R}). \tag{2.7}$$

We refer to figure 2.5 to determine $\delta\Omega$, noting that a pixel area A at (x, y) subtends a solid angle

$$\delta\Omega = \frac{A}{l^2}\cos\psi. \tag{2.8}$$

The distance l is given in terms of the angle ψ and the sample-detector distance d:

$$l = \frac{d}{\cos\psi}, \tag{2.9}$$

so that

$$\delta\Omega = \frac{A}{d^2}\cos^3\psi \tag{2.10}$$

and

$$I(x,y) = \frac{A}{d^2} I(\mathbf{q})\cos^3\psi, \tag{2.11}$$

which is what we were seeking to show. Using the fact that

$$\begin{aligned} \cos\psi = &\quad 1 - 2\sin^2\frac{\psi}{2} && (2.12)\\ = &\quad 1 - \frac{1}{2}\frac{q^2}{k^2}, && (2.13) \end{aligned}$$

we can again rewrite $I(x,y)$ in terms of $\mathbf{q}$, so that

$$I(x,y) = \frac{A}{d^2} I(\mathbf{q}\cdot\mathsf{R})\left(1 - \frac{1}{2}\frac{q^2}{k^2}\right)^3. \tag{2.14}$$

Together with the transformation in 2.6, equation 2.14 gives us a simple expression for converting our measured pixel values $I(x,y)$ in diffraction images to physically interpretable measurements of $I(\mathbf{q})$.

2.2.4 Mode filtering

Bragg peaks were suppressed in the stills in order to accurately measure the background diffuse intensity. This was achieved using a mode filtering image processing technique, where pixels in a new image are given the value of the mode (most common

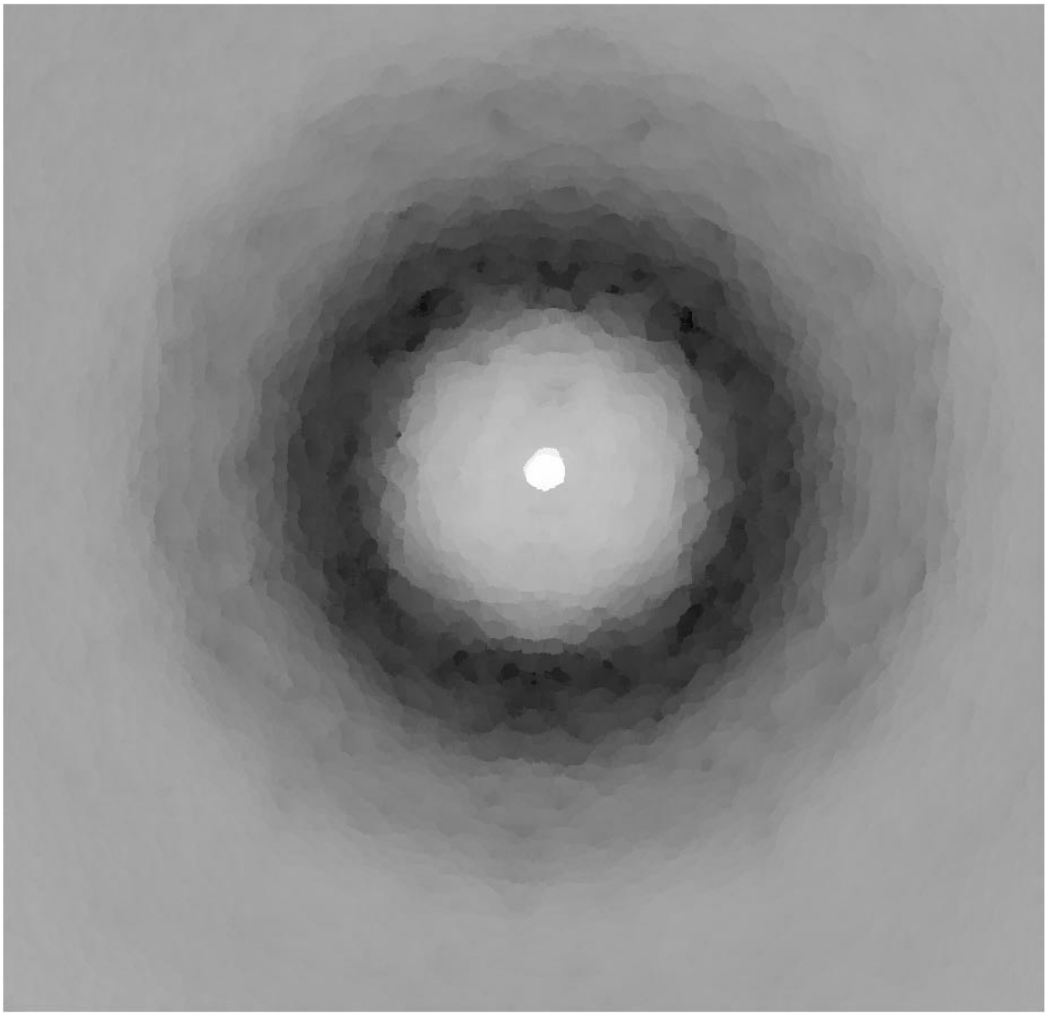

Figure 2.6: *Image of an exposure of* Staph. *nuclease after polarization correction, solid angle normalization and mode filtering to remove Bragg peaks.*

value) of the distribution of pixel values in a 15×15 patch about the pixel in the original image. The filter is often used to "de-star" astronomical sky images[39], where one is interested in measuring background intensities in an image "contaminated" by stars, which bears close resemblance to the problem of measuring diffuse intensity. Figure 2.6 shows the exposure in figure 2.1 after polarization correction, solid-angle normalization and mode filtering.

An important question to ask is how close the mode-filtered image is to the true background scattering pattern. If the differences are at the level of the noise in the diffraction image, the technique is adequate for our needs. If the differences are at the level of the background intensity, the technique is no good.

To understand the mode filtering process, an image of a simulated background patten with noise was generated (see figure 2.7). The length scale of the pattern was chosen to match the length scale of diffuse features in diffraction images. The amplitude of the pattern was 100 counts, and the added noise was uniformly distributed over the range $(-30, 30)$ counts, chosen to match the noise in diffraction images.

A large number of simulated Gaussian peaks were added to the simulated background pattern in order simulate Bragg reflections, thus forming the simulated diffraction image in figure 2.8. The peaks are on a square lattice, spaced thirteen pixels apart, similar to the spacing in *Staph.* nuclease diffraction images. From left to right, the peaks increase in integrated intensity from 0 to $100,000$ counts, while from top to bottom the full-width at half-maximum (FWHM) of the Gaussian increases from 0.1 to 2.6 pixels. Information in DENZO output files indicates that real peak intensities vary from 0 to $100,000$ and have a FWHM of one pixel.

The simulated diffraction image was then mode filtered (see figure 2.9), and the simulated background image *without* noise was subtracted, yielding an image of the residual. This image was displayed with a lower-cutoff threshold of 30 counts (see

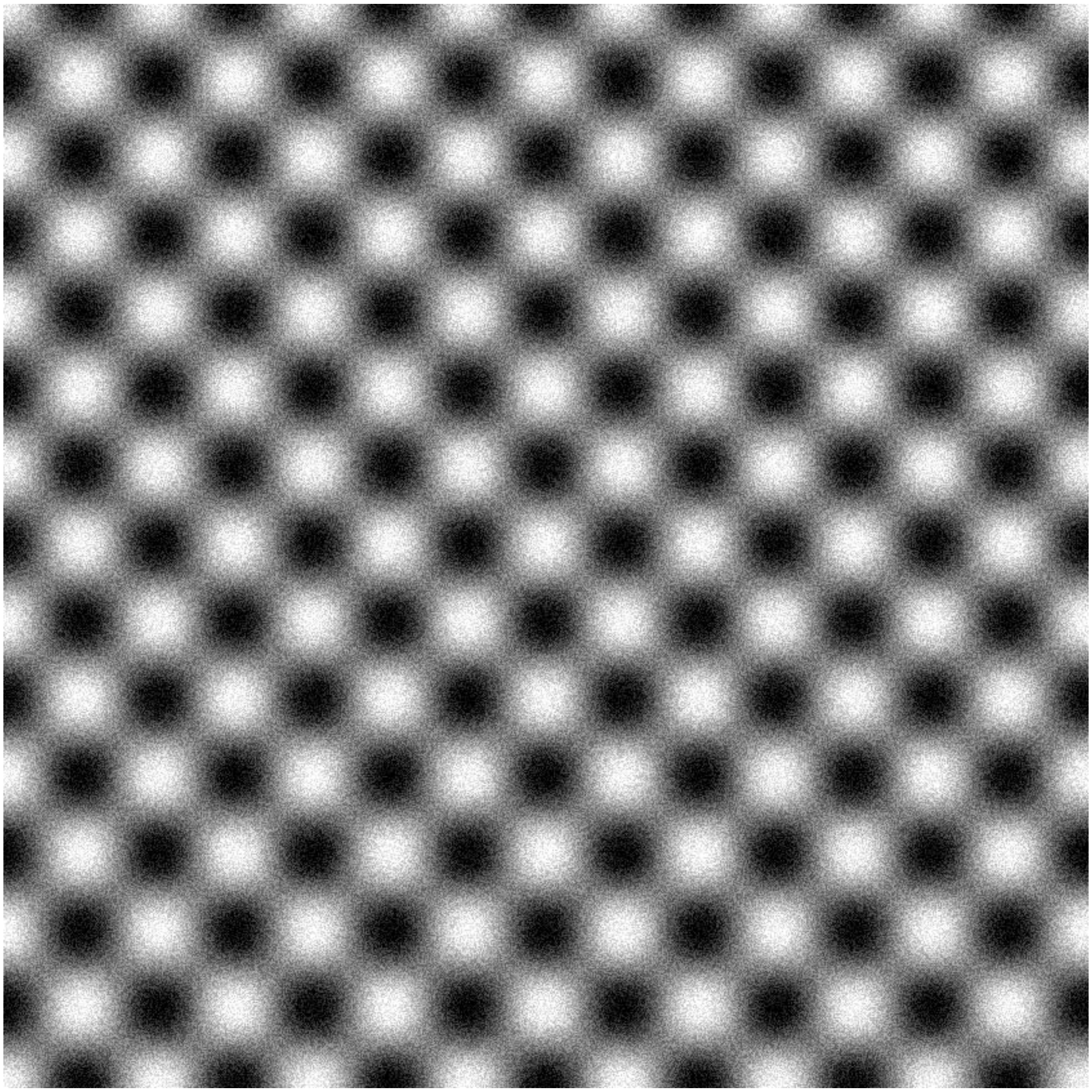

Figure 2.7: *Simulated background pattern with noise.* The background is sinusoidal in two dimensions with an amplitude of 100 counts and a wavelength of 120 pixels. The noise is uniformly distributed in the range $(-30, 30)$. These parameters were chosen to match the observed diffuse scattering in diffraction images. The image is displayed using a linear grey scale, with black corresponding to -100, and white corresponding to 100.

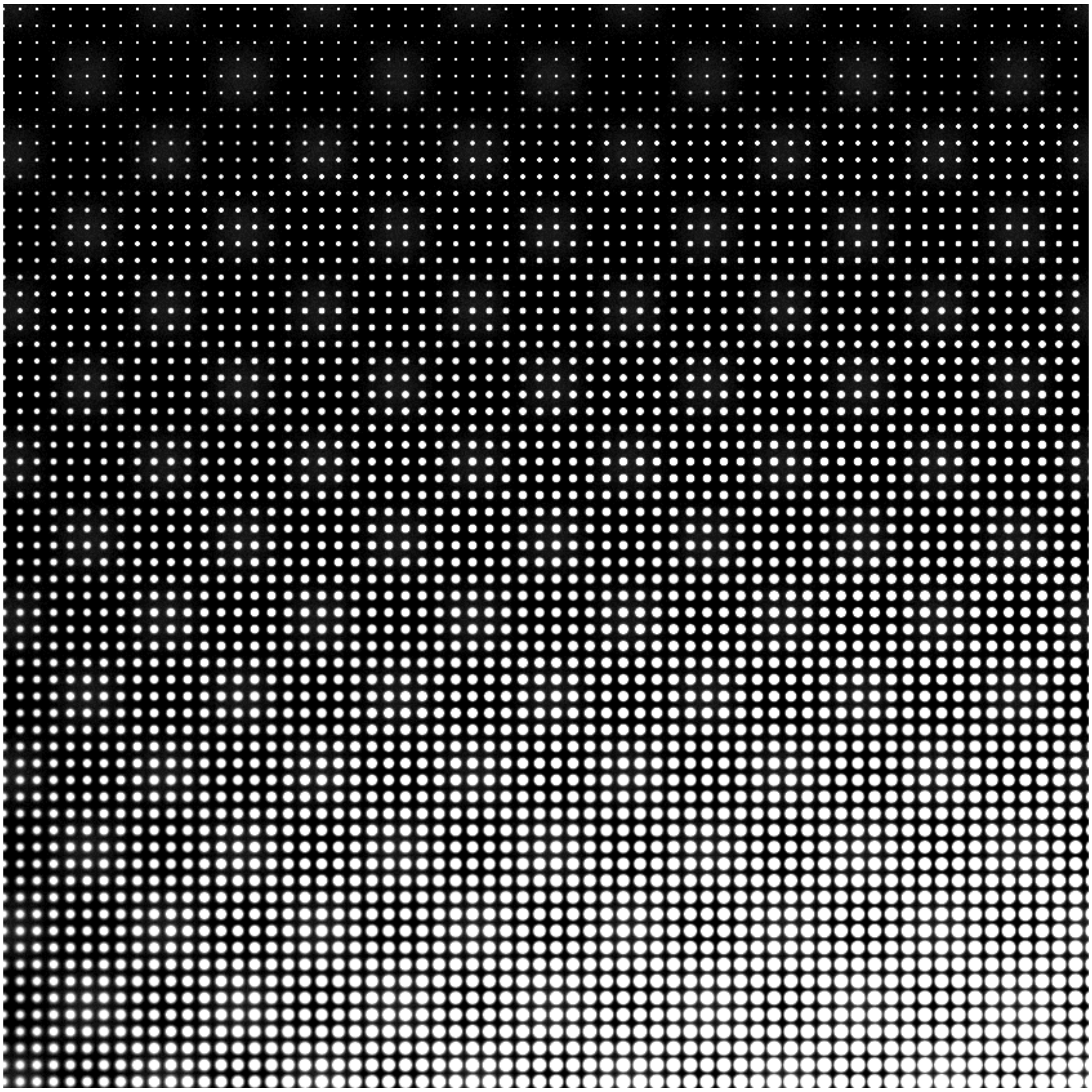

Figure 2.8: *Simulated diffraction image with background pattern and noise.* This is the background pattern of figure 2.7 with a superimposed array of Bragg peaks. The Bragg peaks are ordered from left to right increasing in integrated intensity from 0 to 10^5. The peaks are Gaussian in profile, and their widths increase from 0.1 to 2.6 pixels from top to bottom. Experimentally obtained Bragg reflection measurements yield integrated intensities in the range $(0, 10^5)$ with a Gaussian width of one pixel. The background intensity should be recovered when this image is mode filtered. The image is displayed using a linear grey scale, with black corresponding to 0 and below, and white corresponding to 1000 and above.

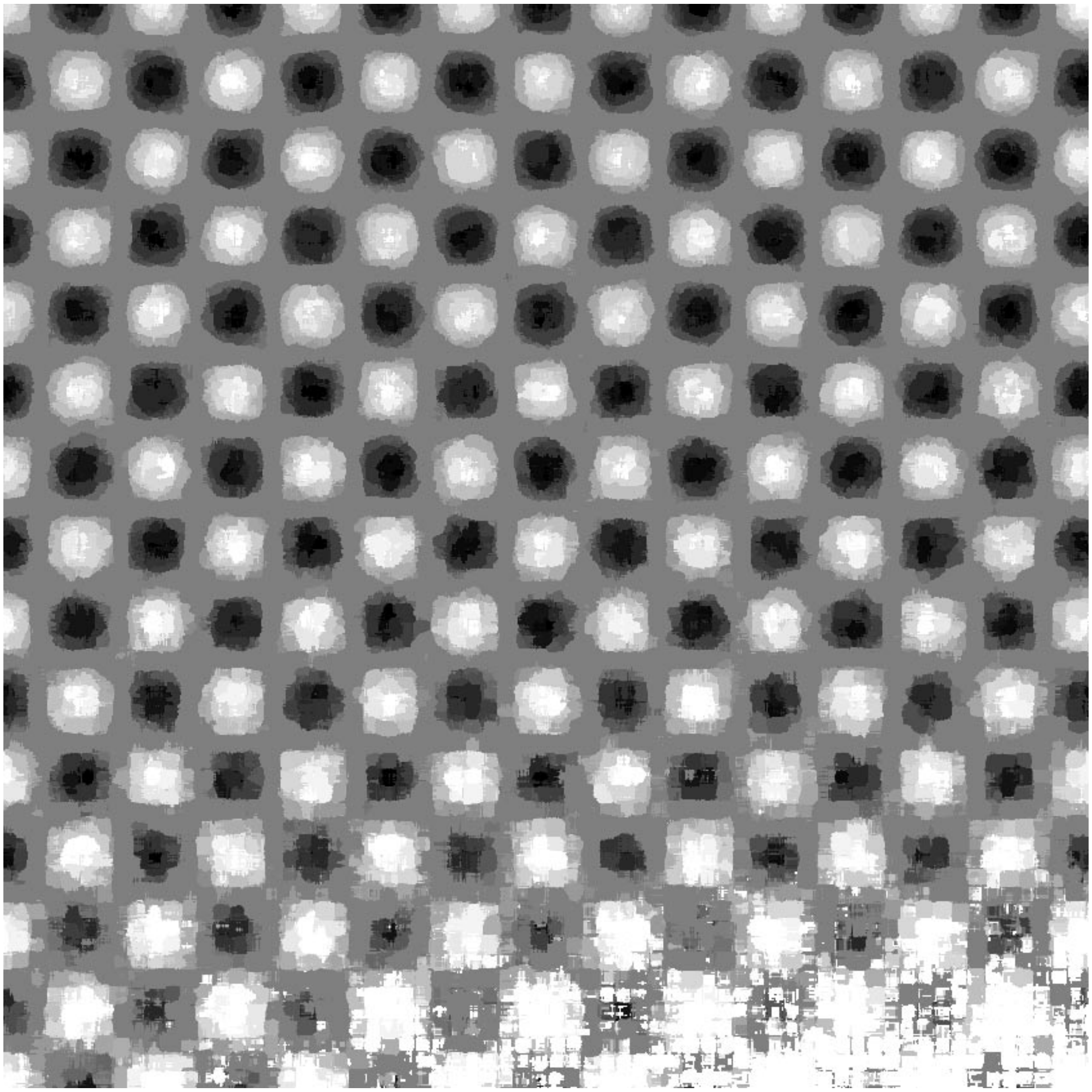

Figure 2.9: *Mode-filtered simulated diffraction image.* This is the image of figure 2.8 after mode filtering. The pixel mask used was 15×15, and pixel value bins spanned ten counts. The technique breaks down in the lower right hand portion of the image.

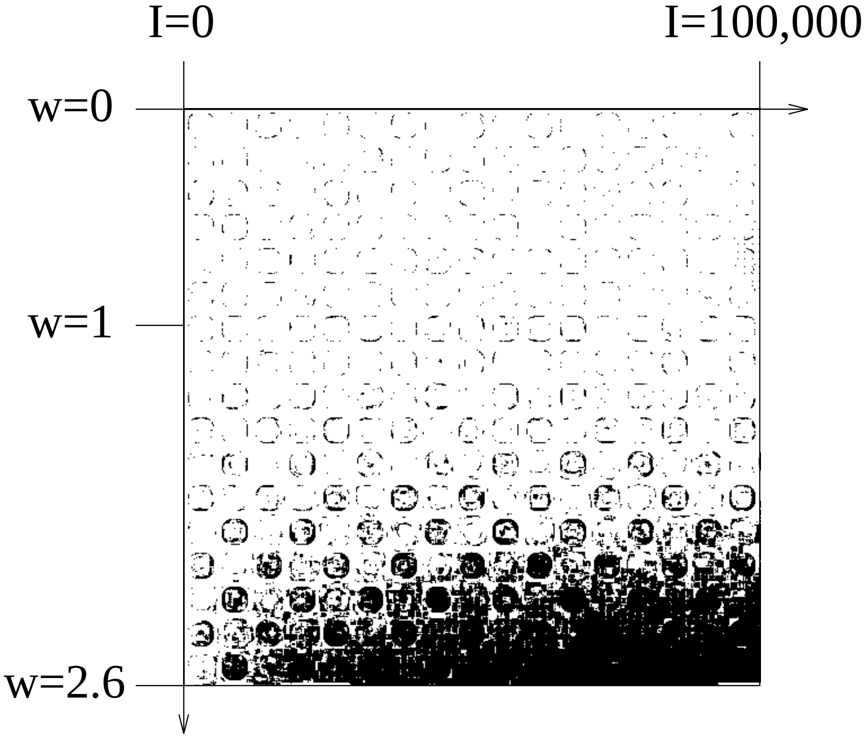

Figure 2.10: *Pixels in the difference image which are above the noise threshold.* All pixels with pixel value greater than 30 are displayed as black, and all other pixels are displayed as white. The axes label the peaks where they would occur in figure 2.8, w being the peak width in pixels and I being the integrated peak intensity. Our experiments had peaks in the neighborhood of $w = 1$, which is marked on the $w-$axis. This indicates that the mode filtered image is the same as the background image to within noise levels.

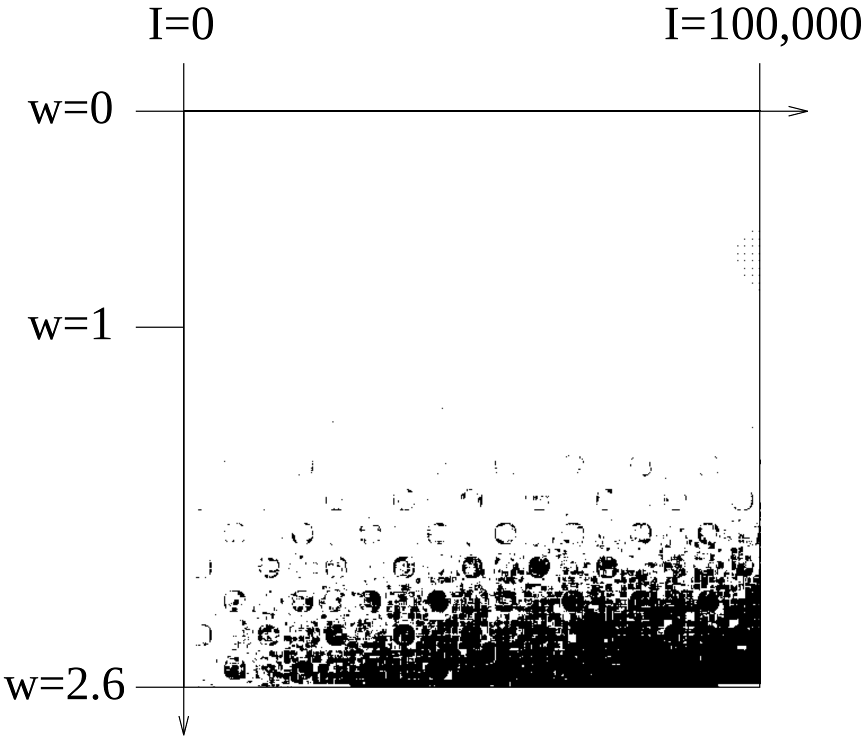

Figure 2.11: *Pixels in the difference image which are above half the background amplitude.* All pixels with pixel value greater than 50 are displayed as black, and all other pixels are displayed as white. The axes label the peaks where they would occur in figure 2.8, w being the peak width in pixels and I being the integrated peak intensity. Our experiments had peaks in the neighborhood of $w = 1$, which is marked on the $w-$axis. The difference between the mode-filtered image and the original background image is less than half the amplitude of the background for $w = 1$, no matter how large the Bragg peak.

figure 2.10) to compare it with the noise level, and and was displayed with a threshold of 50 counts (see figure 2.11) to compare it with the level of half of the background signal amplitude. The results were positive: the differences are at the level of the noise, indicating that the technique is useful for our purposes.

Unfortunately, saturated peaks can leave residual intensity in mode-filtered images. In the case of the CCD detector used in the experiments described here, a non-exponential tail in the point-spread function of the detector would be the culprit. As is described in Tate *et. al.*[37], a similar detector showed that at a distance of 450 μm from a peak, the pixel value was still 0.1% of the maximum. Assuming that saturated peaks have 10^5 ADU maximum values, this amounts to a 100 ADU effect, which is on the order of the measured intensity of diffuse features.

Fortunately, 450 μm corresponds to less than 6 pixels on the detector, which has a pixel size of 80 μm, while the 15×15 mode-filter mask extends between 7.5 pixels (on a side) and 10.6 pixels (on a diagonal) from the center of the square. Even if the mask is centered on a saturated peak, therefore, there is a good chance that the mode will not contain significant contributions from the tails of the peak. If there is any observable effect at all, moreover, it would be expected only in the immediate neighborhood of the saturated pixel, and would have a high likelihood of being eliminated in the rejection of measurements too close to a Bragg peak (see section 2.3). Contamination of diffuse maps due to saturated Bragg peaks, therefore, is expected to be an insignificant contribution to systematic errors in the measurements performed here, although it may present problems in measuring diffuse features which are weaker than those observed in *Staph.* nuclease.

2.2.5 Image scaling

In order to correct for beam intensity variations over time and for absorption as the crystal is rotated, all images in a diffuse data set are scaled to an arbitrary reference image using a single multiplicative constant per image. For each pixel in an image, the distance in pixels to the beam fiducial mark is calculated and rounded to the nearest integer. All pixels with the same distance are binned together, and their values are averaged. An average radial intensity profile for the image is generated by plotting the average pixel value *vs.* distance from the fiducial mark (see figure 2.12 for some sample profiles).

One image is arbitrarily chosen as a reference image, and its profile is chosen as a reference profile. For each image, a linear least-squares fit is used to find a multiplicative constant which scales the intensity profile to the reference profile. Figure 2.12 shows some sample average radial intensity profiles, and a plot of the scale factor *vs.* image number in the data set. [3]

2.3 Generating the Diffuse Map

Based upon the scattering geometry, each pixel in each image is mapped to a scattering vector (h', k', l') which is oriented relative to the crystal lattice using both the detector face rotation angles and the crystal orientation matrix elements calculated using DENZO. The Bragg peak nearest to (h', k', l') is identified, associating the scattering vector (h', k', l') with miller indices (h, k, l). If a $\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}$ cube centered on the Bragg peak (h, k, l) contains the point (h', k', l'), the pixel is rejected as measuring

---

[3] This procedure may be used as an alternative method for calculating image scale factors for a standard crystallographic data set. Due to increased statistics, this method should provide a more precise measurement of the scale factors than is obtained by sole analysis of the Bragg peaks, such as is routinely done in crystallographic analysis.

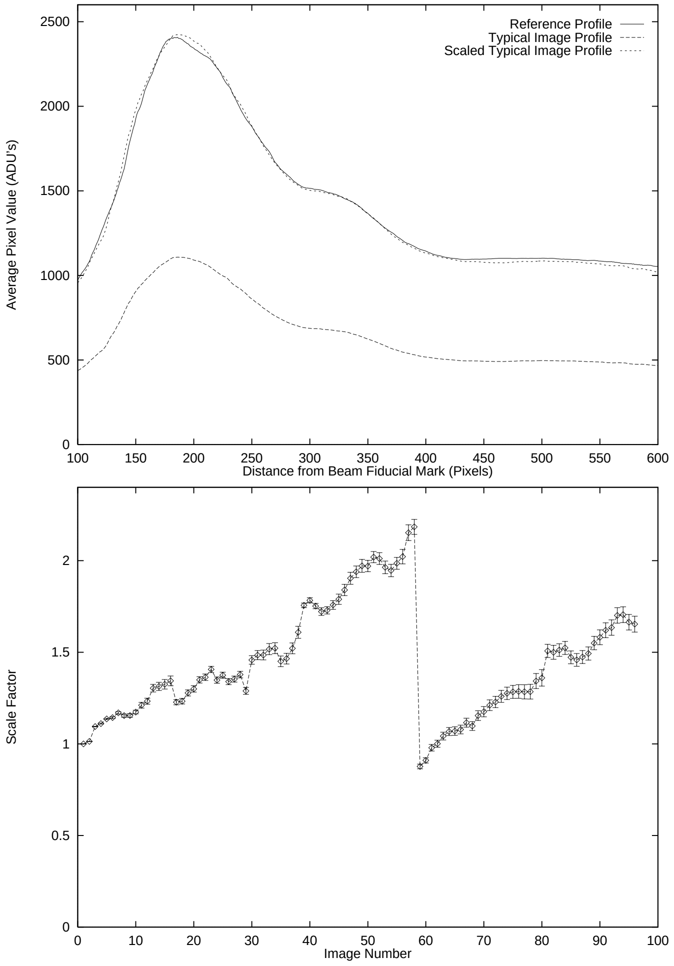

Figure 2.12: *Image scaling. Top*: An image's scale factor is found by scaling its average radial intensity profile to that of a reference image (see text for a definition of the profile). The reference profile for the first *Staph.* nuclease diffuse data set is shown, along with a typical image profile before and after scaling. *Bottom*: Scale factors calculated for all exposures in the first *Staph.* nuclease diffuse data set. All pixels in an image are multiplied by the scale factor to correct for variations in beam intensity and crystal absorption. The large dip between images fifty-eight and fifty-nine is due to a beam intensity increase at a new "fill," when electrons and positrons are re-injected into the synchrotron ring at CHESS. The source of the other systematic variations is unknown, but is attributed to fluctuations in the beam intensity at the F1 station.

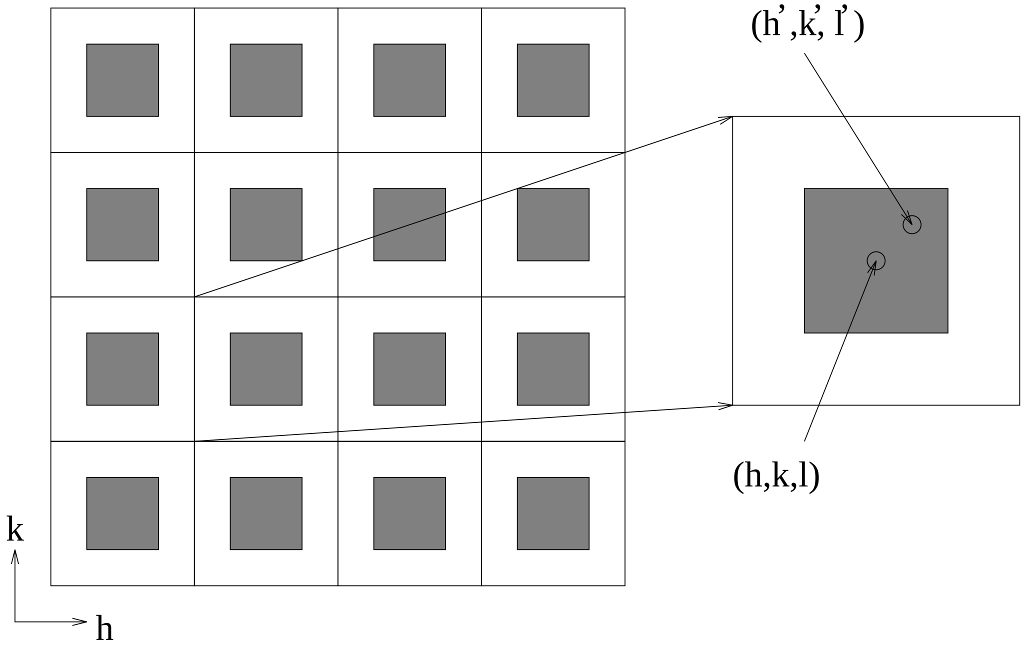

Figure 2.13: *Illustration of a subdivision of reciprocal space into ignored regions around Bragg peaks (shaded) and measured regions (white).* The inset illustrates a situation where (h', k', l') falls within the $\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}$ box around (h, k, l), indicating a rejection. All pixel values which measure diffuse intensity in the white bordered region about (h, k, l) are averaged together for a single measurement to generate a three-dimensional diffuse map.

intensity too close to a Bragg peak (see figure 2.13). All pixels from all images which are not rejected are multiplied by a scale factor (see above) and are averaged together as a measurement of the diffuse intensity in the neighborhood of (h, k, l) in reciprocal space. The collection of values of diffuse intensity for all coordinates (h, k, l) which span the data set is stored as a three-dimensional lattice in a computer, and represents a three-dimensional map of diffuse scattering from *Staph.* nuclease. See section 5.1 for a description of the maps which were experimentally obtained from crystals of *Staph.* nuclease.

These three-dimensional maps are the objects which are compared and characterized. Changes in the maps, representing changes in diffuse scattering, indicate possible changes in the dynamics of the crystalline protein. The symmetry in the maps can be used to characterize the self-consistency of the diffuse data. Simulation of features in these maps will be used to provide insight into the nature of the underlying disorder in the crystal which causes them.

Chapter 3

Models of Diffuse Scattering

In this section, various disordered states of the crystal are defined, and scattering theory is used to predict what diffuse features arise from these states. Features of the models are related to properties of diffuse maps. After describing the models, we will have some understanding of what distinguishing features can be used to identify the nature of disorder in the crystal based upon measurements of diffuse scattering.

3.1 Diffraction from an Imperfect Crystal

Figure 3.1 illustrates a two-dimensional representation of a disordered crystal. X-ray diffraction theory describes what form the scattering amplitude from an electron density distribution $\rho(\mathbf{x}, t)$ takes (see figure 3.2)[1]

$$F(\mathbf{q}, t) = \int d^3\mathbf{x} e^{i\mathbf{q}\cdot\mathbf{x}} \rho(\mathbf{x}, t). \tag{3.1}$$

---

[1] See, *e.g.*, chapter 2 of Guinier [40] or chapter 3 of Giacovazzo [38], for an introduction to the theory of x-ray diffraction.

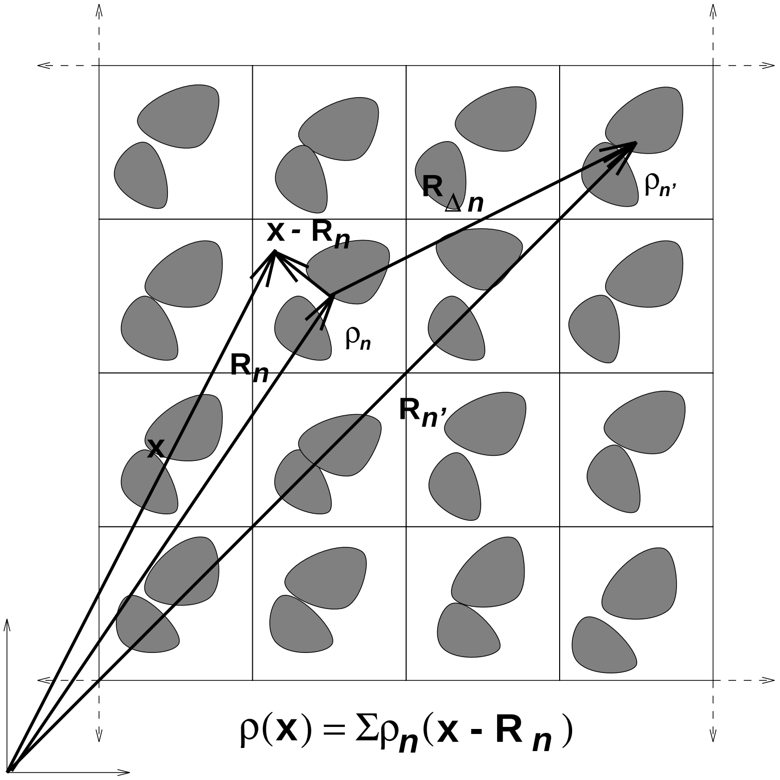

Figure 3.1: *Representation of a disordered crystal.* The crystal is composed of unit cells n of electron density $\rho_n(\mathbf{x})$ at displacements $\mathbf{R}_n$ from the origin, so that the crystal electron density is given by $\rho(\mathbf{x}) = \sum \rho_n(\mathbf{x} - \mathbf{R}_n)$. Unit cells n and n' are separated by a displacement $\mathbf{R}_{\Delta n} = \mathbf{R}_{n'} - \mathbf{R}_n$. Note that the unit cells are not all exactly the same.

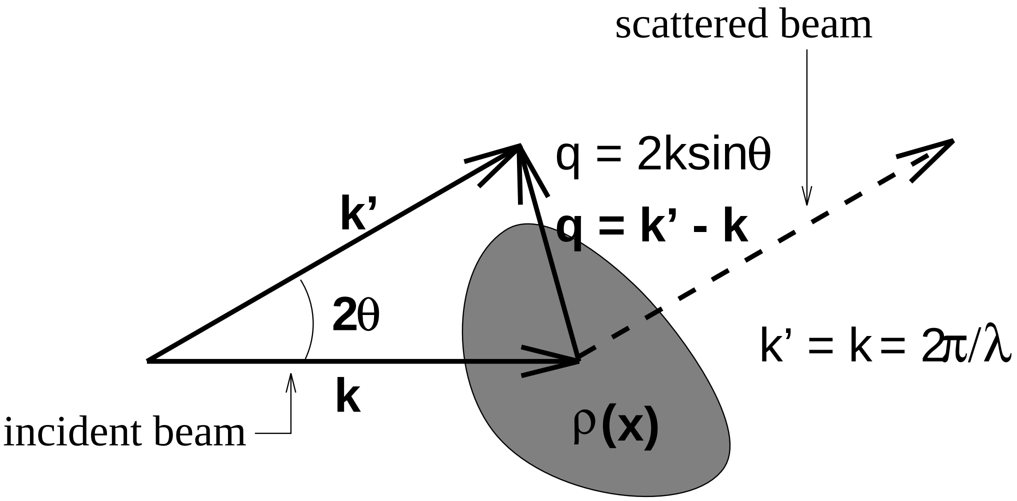

Figure 3.2: *Definition of parameters of elastic x-ray scattering.* Incident x-rays of momentum $\mathbf{k}$ scatter from an electron density distribution $\rho(\mathbf{x})$ through an angle of 2θ to an outgoing momentum $\mathbf{k}'$, yielding a scattering vector $\mathbf{q} = \mathbf{k}' - \mathbf{k}$.

For a crystal composed of unit cells n of electron density $\rho_n(\mathbf{x})$ located at positions $\mathbf{R}_n$ with respect to the origin, the total distribution $\rho(\mathbf{x})$ can be represented by

$$\rho(\mathbf{x}, t) = \sum_n \rho_n(\mathbf{x} - \mathbf{R}_n, t) \tag{3.2}$$

so that

$$\begin{aligned} F(\mathbf{q}, t) &= \sum_n \int d^3\mathbf{x} e^{i\mathbf{q}\cdot\mathbf{x}} \rho_n(\mathbf{x} - \mathbf{R}_n, t) && (3.3) \\ &= \sum_n \left(\int d^3\mathbf{x} e^{i\mathbf{q}\cdot\mathbf{x}} \rho_n(\mathbf{x}, t) \right) e^{i\mathbf{q}\cdot\mathbf{R}_n} && (3.4) \\ &= \sum_n f_n(\mathbf{q}, t) e^{i\mathbf{q}\cdot\mathbf{R}_n} && (3.5) \end{aligned}$$

where the unit cell structure factor $f_n(\mathbf{q}, t)$ has been defined as

$$f_n(\mathbf{q}, t) \equiv \int d^3\mathbf{x} e^{i\mathbf{q}\cdot\mathbf{x}} \rho_n(\mathbf{x}, t). \tag{3.6}$$

The scattering intensity, given by

$$I(\mathbf{q},t) = |F(\mathbf{q},t)|^2 \tag{3.7}$$

is then

$$I(\mathbf{q},t) = \sum_{n,n'} f_n(\mathbf{q},t) f^*_{n'}(\mathbf{q},t) e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})}, \tag{3.8}$$

so that the scattering intensity at any time can be calculated exactly in the case where all of the instantaneous structure factors $f_n(\mathbf{q})$ are known. For a perfect, finite-sized crystal, in which $f_n(\mathbf{q},t) = f_{n'}(\mathbf{q},t)$ for all n and n' in the crystal, this reduces to

$$I(\mathbf{q}) = |f(\mathbf{q})|^2 \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})}, \tag{3.9}$$

where $f(\mathbf{q})$ is the structure factor of the common unit cell.

If the crystal is sufficiently large, the scattering intensity $I(\mathbf{q},t)$ becomes effectively time-independent *even* in the disordered crystal. As long as the experimental resolution of δq for scattering vectors $\mathbf{q}$ is such that

$$\delta q \gg \frac{2\pi}{L}, \tag{3.10}$$

where L is the size of the crystal, it is shown in appendix A.1 that the scattered intensity $I(\mathbf{q},t)$ can be expressed as

$$I(\mathbf{q},t) = I(\mathbf{q}) = \sum_{n,n'} \langle f_n(\mathbf{q}) f^*_{n'}(\mathbf{q}) \rangle_{f_n,f_{n'}} e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})}, \tag{3.11}$$

where the average is over all possible values of $f_n(\mathbf{q})f^*_{n'}(\mathbf{q})$ for a given pair (n, n'):

$$\langle f_n(\mathbf{q})f^*_{n'}(\mathbf{q})\rangle = \int \delta f_n \int \delta f_{n'} P_{nn'}(f_n, f_{n'}) f_n(\mathbf{q}) f^*_{n'}(\mathbf{q}), \tag{3.12}$$

$P_{nn'}(f_n, f_{n'})$ being the joint probability of finding structure factors f_n and $f_{n'}$ in unit cells n and n'. An important result here is that the exact distribution of electron density in each unit cell in the entire crystal is not needed to calculate the scattered intensity. It suffices to know the distribution of states of unit cell pairs (n, n') which are separated by a common displacement vector $\mathbf{R}_{\Delta n}$.

How do the numbers work out for a real crystal, where the effective size is that of a mosaic block? Using the pixel size of the detector in the experiments described in this thesis to determine the angular resolution $\delta\theta$ of scattered x-rays at typical sample-to-detector distances, the estimate is $\delta\theta = 10^{-3}$. The estimated error in measurements of q due to error in θ is

$$\delta q = 2k\delta\theta, \tag{3.13}$$

which is gives $\delta q \approx 10^{-2}$ Å^{-1} for a beam wavelength of 1 Å. By the relation in equation 3.10, this leads to

$$L \gg 2\pi \times 10^2 \text{ Å}, \tag{3.14}$$

or, assuming a roughly 50 Å unit cell, which is the case for *Staph.* nuclease,

$$N \gg 10, \tag{3.15}$$

where N is the number of unit cells on a side in a *mosaic block*, which describes the effective size of the crystal due to defects which disturb the long-range order in the crystal. This relation is satisfied in protein crystals, where mosaic blocks are unusually large due to plasticity in the unit cell structure, which acts to relieve the stresses which would ordinarily create defects.

The scattered intensity can be calculated at any time using any snapshot of the crystalline electron density. At the same time, crystal dynamics cannot be resolved in real-time, since the scattered intensity is time-independent. This necessitates the study of, *e.g.*, temperature variations in order to truly distinguish between x-ray signatures of dynamic and static disorder.

3.2 Classes of Diffuse Scattering

Models of diffuse scattering can broadly be sorted into two classes: those in which the distribution of individual unit-cell electron densities are independent, and those in which there are correlations among unit-cell electron-density distributions. For models in the former class, the problem can be simplified by isolating the term due to diffuse scattering from the term due Bragg scattering. Returning to equation 3.11,

$$I = \sum_{n,n'} \langle f_n f^*_{n'} \rangle_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})}, \tag{3.16}$$

where the functional dependence of f_n and I on $\mathbf{q}$ is now assumed in the notation, and the average is now over all unit cell pairs (n, n'). Independent unit-cell electron-density distributions imply independent unit-cell structure factors f_n, so that, from the fact that

$$\langle f_n \rangle_n = \langle f_{n'} \rangle_{n'}, \tag{3.17}$$

it follows that

$$\langle f_n f^*_{n'} \rangle_{n,n'} = \begin{cases} \langle |f_n|^2 \rangle_n & n = n' \\ |\langle f_n \rangle_n|^2 & n \neq n' \end{cases} \tag{3.18}$$

or

$$\langle f_n f_{n'}^* \rangle_{n,n'} = (\langle |f_n|^2 \rangle_n - |\langle f_n \rangle_n|^2)\delta_{nn'} + |\langle f_n \rangle_n|^2 \tag{3.19}$$

where $\delta_{nn'}$ is the Kronecker delta distribution

$$\delta_{nn'} = \begin{cases} 1 & n = n' \\ 0 & n \neq n' \end{cases} .$$

Substituting this expression for $\langle f_n f_{n'}^* \rangle_{n,n'}$ into equation 3.16,

$$\begin{aligned} I = & \sum_{n,n'} \left[(\langle |f_n|^2 \rangle_n - |\langle f_n \rangle_n|^2)\delta_{nn'} + |\langle f_n \rangle_n|^2 \right] e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})} & (3.20) \\ = & N(\langle |f_n|^2 \rangle_n - |\langle f_n \rangle_n|^2) + |\langle f_n \rangle_n|^2 \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})} & (3.21) \\ = & I_D + I_B & (3.22) \end{aligned}$$

where N is the number of unit cells in the crystal. The diffuse component I_D has been defined as

$$I_D = N(\langle |f_n|^2 \rangle_n - |\langle f_n \rangle_n|^2), \tag{3.23}$$

which is equivalent to Guinier's expression for diffuse scattering in his discussion of the scattering intensity from a crystal with no correlations in fluctuations from unit cell to unit cell.[2] This is the expression used in many of the recent interpretations of data in diffuse scattering experiments discussed in the introduction[17, 18, 19, 21, 20, 23, 22, 24].

The Bragg component I_B is defined as

$$I_B = |\langle f_n \rangle_n|^2 \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})}, \tag{3.24}$$

---

[2] See pp.163-166, chapter 6 of Guinier[40].

which gives rise to the usual sharp peaks characteristic of x-ray diffraction from three-dimensional crystals.

In the class of models where the unit-cell electron densities are independent, we can predict that, since all unit cell structure factors f_n refer to distributions smaller than one unit cell in size, the length scale of variations in diffuse features in reciprocal space must be greater than or equal to the spacing between Bragg peaks. In order to look for signatures of fluctuations within a unit cell, therefore, features in the diffraction pattern which vary on a length scale greater than the distance between Bragg peaks should be studied. This observation provided partial motivation for the chosen method of mapping diffuse features in three dimensions, which ignores features on length scales smaller than Bragg peak spacings.

Of the more general class of models, where correlations among unit-cell electron densities are allowed, two will be discussed. One is the liquid-like motions model of Caspar *et. al.*[19], which associates a correlation length with atomic displacements, while the other is a model based on the excitation of normal modes in the crystal, as described in Glover *et. al.*[21]. It is also shown that the liquid-like motions model is a special case of the model of normal modes in the crystal, so that they describe the same type of disorder using different mathematical expressions.

A feature of the models of Caspar *et. al.*[19] and the normal modes model is that they both use the unit-cell structure factor $f(\mathbf{q})$ in their expressions for diffuse scattering. The fact that the experimental diffuse maps show a high correlation with $|f(\mathbf{q})|^2$, as is shown in section 5.3, favors these models, as well as any other general models which use $|f(\mathbf{q})|^2$, over the class of models where unit cells have independent electron-density distributions, which generally contain structure factors from unit-cell subdomains. Correlations with a handful of subdomains were sought, but none did nearly as well at mimicking the data as the $|f(\mathbf{q})|^2$ map.

What follows is first a discussion of models of diffuse scattering with independent unit-cell fluctuations. Included among these is a new model describing rigid-body motions of arbitrary structural units. Then, the two more general diffuse scattering models will be discussed: first the liquid-like motions model of Caspar *et. al.*[19], and then the crystalline normal modes model described by Glover *et. al.*[21].

3.3 Independent Molecular Domains

As was discussed in the introduction, proteins are likely to exhibit significant internal fluctuations under biological conditions. There is ample evidence that these fluctuations occur within the protein crystalline state as well. One way to model these internal motions is to identify atomic clusters in the unit cell and assume that the internal motions can be subdivided into the independent motions of these domains.

If $\rho_{nj}(\mathbf{x})$ represents the electron density of a domain j in unit cell n, then the unit cell electron density $\rho_n(\mathbf{x})$ is given by (see figure 3.3)

$$\rho_n(\mathbf{x}) = \sum_j \rho_{nj}(\mathbf{x}), \tag{3.25}$$

and the corresponding unit cell structure factor $f_n(\mathbf{q})$ is given by

$$f_n(\mathbf{q}) = \sum_j f_{nj}(\mathbf{q}), \tag{3.26}$$

where $f_{nj}(\mathbf{q})$ is the structure factor corresponding to the electron density of domain j in unit cell n. As is derived in appendix A.2, it follows that the diffuse intensity I_D

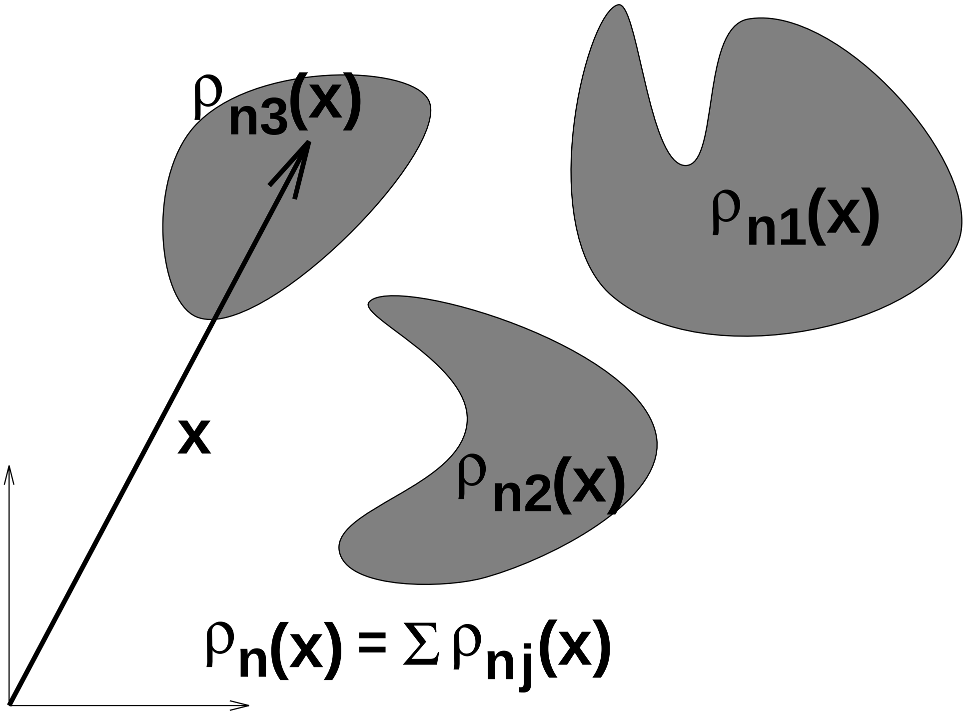

Figure 3.3: *Illustration of a distribution* ρ_n *composed of multiple, independent, non-overlapping domains* ρ_{nj}.

is given by

$$\begin{aligned} I_D = \quad & N \sum_j (\langle |f_{nj}|^2 \rangle_n - |\langle f_{nj} \rangle_n|^2) && (3.27) \\ = \quad & \sum_j I_{Dj}. && (3.28) \end{aligned}$$

The diffuse intensity I_D due to multiple independent domains is equal to the sum of the diffuse intensities I_{Dj} due to each domain j considered independently. Each domain contributes incoherently to the diffuse intensity.

What if there is some correlation among the domains? In this case, the diffuse intensity I_D becomes

$$I_D = N \sum_j (\langle |f_{nj}|^2 \rangle_n - |\langle f_{nj} \rangle_n|^2) + N \sum_{j \neq j'} \langle f_{nj} \rangle_n \langle f^*_{nj'} \rangle_n \Gamma_{jj'}, \tag{3.29}$$

where the correlation coefficient $\Gamma_{jj'}$ is defined by

$$\Gamma_{jj'} \equiv \frac{\langle f_{nj} f^*_{nj'} \rangle}{\langle f_{nj} \rangle \langle f_{nj'} \rangle} - 1. \tag{3.30}$$

This would be the case, for instance, if there were a significant effect on the fluctuations of neighboring proteins due to excluded volume effects. In general, correlations among domains are difficult cases to parametrize analytically, but may be tractable by molecular dynamics simulations. If the correlation coefficients $\Gamma_{jj'}$ are small enough, the effect of correlations can be ignored.

Regardless of the significance of the correlations among domains, the Bragg intensity is given by

$$I_B = \quad |\langle f_n \rangle_n|^2 \sum_{nn'} e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})} \tag{3.31}$$

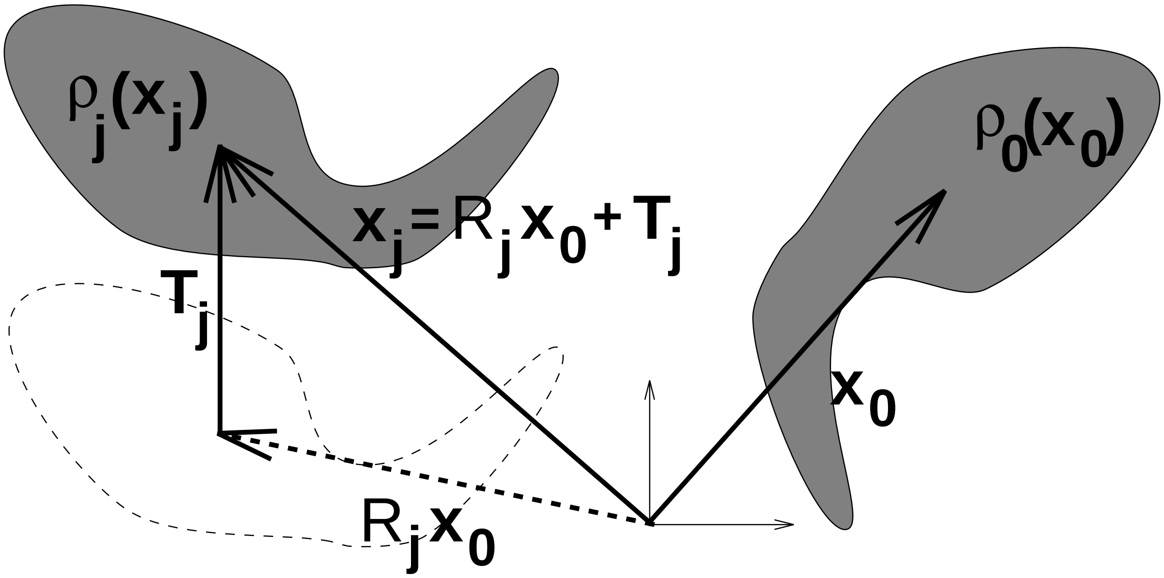

Figure 3.4: *Generation of a distribution* $\rho_j(\mathbf{x}_j)$ *isomorphous to a distribution* $\rho_0(\mathbf{x}_0)$ *using a rotation* R_j *and a translation* T_j. The dashed line illustrates the distribution ρ_0 after rotation but before translation.

$$= \quad |\sum_j \langle f_{nj}\rangle_n|^2 \sum_{nn'} e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})}, \tag{3.32}$$

so that each independent domain contributes coherently to the Bragg scattering.

In analogy with the arguments presented in section 3.1, we can comment on the length scale of variations in diffuse scattering images. Here, we note that the smallest variations will come from the largest independent domain in the unit cell, so that if one can identify the smallest visible diffuse features in a map, the size of the largest coherent, mobile structural unit can be identified.

3.4 Independent Isomorphous Molecules

A special case of independently fluctuating domains is the case where the unit cell contains a number of isomorphous molecules which are related by simple rotations and translations, which is the case for most protein crystals, including *Staph.* nuclease. The effect on Bragg reflections of introducing these copies of the reference structure into the unit cell is described, for example, in chapter 3 of Giacovazzo[38], but it is difficult to find a description of the effect on diffuse scattering. Here, some expressions will be derived for both Bragg scattering and diffuse scattering when there are independently fluctuating, isomorphous molecules.

In each unit cell n, an isomorphous copy $\rho_j(\mathbf{x}_j)$ of an electron density distribution $\rho_0(\mathbf{x}_0)$ can be generated by a rotation R_j and a translation $\mathbf{T}_j$ (see figure 3.4):

$$\rho_j(\mathbf{x}_j) = \rho_0(\mathbf{x}_0) \tag{3.33}$$

where

$$\mathbf{x}_j = \mathsf{R}_j \cdot \mathbf{x}_0 + \mathbf{T}_j. \tag{3.34}$$

The structure factor of each molecule j is then

$$\begin{aligned} f_j(\mathbf{q}) = & \int d^3\mathbf{x}_j e^{i\mathbf{q}\cdot\mathbf{x}_j} \rho_j(\mathbf{x}_j) && (3.35) \\ = & \int d^3\mathbf{x}_0 e^{i\mathbf{q}\cdot(\mathsf{R}_j\cdot\mathbf{x}_0+\mathbf{T}_j)} \rho_0(\mathbf{x}_0) && (3.36) \\ = & \; e^{i\mathbf{q}\cdot\mathbf{T}_j} f_0(\mathbf{q}\cdot\mathsf{R}_j) && (3.37) \end{aligned}$$

where $f_0(\mathbf{q}\cdot\mathsf{R}_j)$ is the structure factor of the reference molecule evaluated at the scattering vector $\mathbf{q}\cdot\mathsf{R}_j$. Using this equation, expressions for the Bragg intensity I_B and the diffuse intensity I_D can be derived as in appendix A.3. The Bragg intensity I_B is given by

$$\begin{aligned} I_B = & \; |\langle f_n\rangle_n|^2 \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})} && (3.38) \\ = & \; |\sum_j e^{i\mathbf{q}\cdot\mathbf{T}_j} \langle f_{n0}(\mathbf{q}\cdot\mathsf{R}_j)\rangle_n|^2 \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})}, && (3.39) \end{aligned}$$

where f_{n0} is the structure factor of the reference molecule in unit cell n, while the diffuse intensity I_D is given by

$$\begin{aligned} I_D = {} & N(\langle |f_n|^2 \rangle_n - |\langle f_n \rangle_n|^2) && (3.40) \\ = {} & \sum_j I_{D0}(\mathbf{q} \cdot \mathsf{R}_j), && (3.41) \end{aligned}$$

where I_{D0} is the diffuse intensity evaluated from the reference structure.

These results can be used to comment on the symmetry of diffuse maps. Consider the case of *Staph.* nuclease, which has four molecules per unit cell in space group P4_1. Each molecule j is related to the asymmetric unit by a rotation R_j and a translation $\mathbf{T}_j$ along the screw axis $\mathbf{c}$. If each molecule moves independently, the diffuse scattering is given by

$$\begin{aligned} I_D = {} & \sum_j I_{D0}(\mathbf{q} \cdot \mathsf{R}_j) && (3.42) \\ = {} & I_{D0}(\mathbf{q} \cdot \mathsf{R}_0) + I_{D0}(\mathbf{q} \cdot \mathsf{R}_1) + I_{D0}(\mathbf{q} \cdot \mathsf{R}_2) + I_{D0}(\mathbf{q} \cdot \mathsf{R}_3), && (3.43) \end{aligned}$$

where $\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2$, and R_3 correspond to rotations of $0°, 90°, 180°$ and $270°$ about the screw axis. This expression is invariant under $90°$ rotations R about the c^*-axis, since

$$\begin{aligned} I_D(\mathbf{q} \cdot \mathsf{R}) = {} & I_{D0}(\mathbf{q} \cdot \mathsf{R} \cdot \mathsf{R}_0) + I_{D0}(\mathbf{q} \cdot \mathsf{R} \cdot \mathsf{R}_1) + I_{D0}(\mathbf{q} \cdot \mathsf{R} \cdot \mathsf{R}_2) + I_{D0}(\mathbf{q} \cdot \mathsf{R} \cdot \mathsf{R}_3) \\ = {} & I_{D0}(\mathbf{q} \cdot \mathsf{R}_1) + I_{D0}(\mathbf{q} \cdot \mathsf{R}_2) + I_{D0}(\mathbf{q} \cdot \mathsf{R}_3) + I_{D0}(\mathbf{q} \cdot \mathsf{R}_0) \\ = {} & I_D(\mathbf{q}), && (3.44) \end{aligned}$$

where the relation $\mathsf{R} \cdot \mathsf{R}_j = \mathsf{R}_{j+1}$ has been used, with $\mathsf{R}_4 \equiv \mathsf{R}_0$. This implies that the diffuse map from *Staph.* nuclease is four-fold symmetric about the c^*-axis in

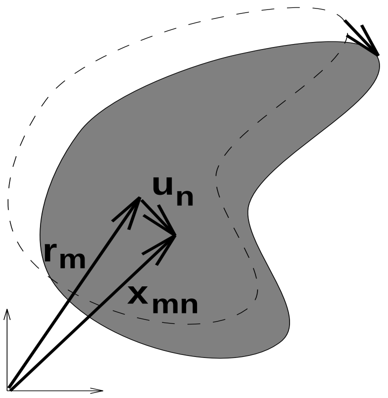

Figure 3.5: *Illustration of a translation by a vector* $\mathbf{u}_n$ *of a rigid body.* Vectors $\mathbf{r}_m$ to atoms m on the reference distribution are mapped to vectors x_{mn} on the shifted distribution.

reciprocal space. In addition, the inherent Friedel symmetry

$$I_D(\mathbf{q}) = I_D(-\mathbf{q}) \tag{3.45}$$

combined with the four-fold c^*-axis implies that the a^*b^*-plane is a mirror plane. These are symmetries which one would expect to see in diffuse maps of *Staph.* nuclease.

3.5 Rigid-Body Translations

So far, I have discussed expressions of diffuse scattering in terms of the distributions $\langle|f_n|^2\rangle_n$ and $|\langle f_n\rangle_n|^2$ without introducing any parameters of disorder. It is possible to use these expressions in concert with molecular dynamics simulations in order to simulate diffuse scattering, as is done in Fauré *et. al.*[24]. Using some assumptions about the nature of the dynamics in the unit cell, however, it is possible to derive some expressions of diffuse scattering which contain explicit parameters of disorder.

One of the simplest such expressions is for the case where structural domains in the unit cell execute independent rigid-body translations, which, along with rigid-body rotations, has been used to interpret B-factors from real protein crystals (see the discussion of Kuriyan and Weis[16] and Sternberg *et. al.*[15] in the introduction). The mathematics is similar to that found in derivations of anisotropic B-factors for crystallographic structural models, but here the tools are used to derive expressions for diffuse scattering. By assuming that the entire unit-cell electron-density distribution undergoes rigid translations, expressions for diffuse scattering from a single structural unit will be derived. Using the equations derived in sections 3.3 and 3.4, it is possible to extend the results to multiple structural units moving independently, as would be the case in a real protein crystal, where multiple molecules can move in the unit cell.

Consider a crystal whose unit cell electron density varies like a rigid body displaced by a translation vector $\mathbf{u}_n$ ($\langle \mathbf{u}_n \rangle_n = 0$) which shifts the origin of the density distribution $\rho(\mathbf{x})$ (see figure 3.5):

$$\rho_n(\mathbf{x}) = \rho(\mathbf{x} - \mathbf{u}_n). \tag{3.46}$$

Such a translation of the unit-cell electron-density distribution will multiply the unit cell structure factor by a phase factor:

$$f_n(\mathbf{q}) = e^{i\mathbf{q}\cdot\mathbf{u}_n} f(\mathbf{q}), \tag{3.47}$$

where $f(\mathbf{q})$ is the unperturbed structure factor of each unit cell.

If the displacements $\mathbf{u}_n$ are sufficiently small, the Bragg intensity I_B and the diffuse intensity I_D can be derived as in appendix A.4. The Bragg intensity I_B is

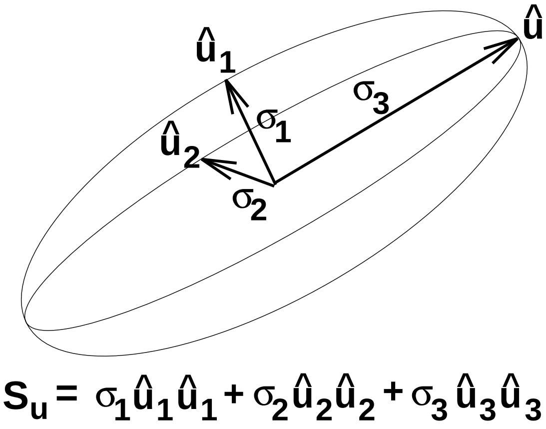

Figure 3.6: *Illustration of a translational disorder ellipsoid.* Axes of the ellipsoid are $\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \hat{\mathbf{u}}_3$, with lengths of $\sigma_1, \sigma_2, \sigma_3$. The projection of any displacement $\mathbf{u}_n$ along an axis $\hat{\mathbf{u}}_i$ is Gaussian distributed with width σ_i.

given by

$$I_B = |\langle f_n \rangle_n|^2 \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})} \tag{3.48}$$

$$= e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}} |f|^2 \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})}, \tag{3.49}$$

where the variance matrix V is defined by the relation

$$\mathsf{V} \equiv \langle \mathbf{u}_n \mathbf{u}_n \rangle_n. \tag{3.50}$$

The diffuse intensity I_D is given by

$$I_D = N(\langle |f_n|^2 \rangle_n - |\langle f_n \rangle_n|^2) \tag{3.51}$$

$$= N(1 - e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}})|f|^2. \tag{3.52}$$

Some original notation can be used for convenience: the symmetric matrix S can be defined such that

$$\mathsf{V} = \mathsf{S} \cdot \mathsf{S} \tag{3.53}$$

so that the Bragg intensity I_B becomes

$$I_B = e^{-|\mathbf{q}\cdot\mathsf{S}|^2} |f|^2 \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})}, \tag{3.54}$$

while the diffuse intensity I_D becomes

$$I_D = N(1 - e^{-|\mathbf{q}\cdot\mathsf{S}|^2})|f|^2. \tag{3.55}$$

S parametrizes an ellipsoid which represents the width of the translational motions along an arbitrary direction (see figure 3.6). Represented diagonally in terms of the ellipsoid axes,

$$\mathsf{S} = \sigma_{u1}\hat{\mathbf{u}}_1\hat{\mathbf{u}}_1 + \sigma_{u2}\hat{\mathbf{u}}_2\hat{\mathbf{u}}_2 + \sigma_{u3}\hat{\mathbf{u}}_3\hat{\mathbf{u}}_3, \tag{3.56}$$

where σ_{ui} are the Gaussian widths of the displacement distributions along axes $\hat{\mathbf{u}}_i$. S is the three-dimensional tensor analog of the standard deviation, so that the projection of any displacement $\mathbf{u}_n$ along an axis $\hat{\mathbf{u}}_i$ is Gaussian distributed with width σ_i.

In the case of isotropically distributed displacements, the sigma matrix S is proportional to the identity matrix $\mathbf{I}$

$$\mathsf{S} = \sigma_u \mathbf{I} \tag{3.57}$$

so that

$$|\mathbf{q} \cdot \mathsf{S}|^2 = q^2\sigma_u^2, \tag{3.58}$$

leaving

$$I_D = N(1 - e^{-q^2\sigma_u^2})|f|^2 \tag{3.59}$$

and

$$I_B = e^{-q^2\sigma_u^2}|f|^2 \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})}. \tag{3.60}$$

This can be compared with the usual crystallographic isotropic B-factor notation,

$$I_B = e^{-2B\frac{\sin^2\theta}{\lambda^2}}|f|^2 \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})} \tag{3.61}$$

by which

$$\mathsf{B} = 8\pi^2\sigma_u^2. \tag{3.62}$$

In the case where there are multiple structural domains in the unit cell which are translating independently, the methods outlined in sections 3.3 and 3.4 can be used to augment the expression for diffuse scattering. Each independent unit j will contribute a component I_{Dj} of the form

$$I_{Dj} = N(1 - e^{-\mathbf{q}\cdot\mathsf{V}_j\cdot\mathbf{q}})|f_j|^2, \tag{3.63}$$

where V_j is the variance matrix for displacements of structural domain j with structure factor $f_j(\mathbf{q})$. If single proteins are translating independently, then j will index the copies of the protein in the unit cell, and the diffuse map will be a superposition of the maps due to each protein's translational disorder. In this case, the diffuse scattering would have a high correlation with a properly symmetrized map of $|f_0(\mathbf{q})|^2$, the square of the structure factor of an individual protein. If, on the other hand, all of the proteins in a unit cell move as a coherent unit, the diffuse scattering would have a high correlation with $|f(\mathbf{q})|^2$, the squared structure factor of the entire unit cell, making discrimination between these two cases possible on the basis of comparison of the diffuse scattering with simulated maps of $|f_0(\mathbf{q})|^2$ and $|f(\mathbf{q})|^2$. Note that the

Bragg scattering would be identical in the two cases, so long as the variance matrices were equivalent, making them indistinguishable on the basis of measurements of Bragg reflections alone.

3.6 Independent Atomic Fluctuations

In standard crystallographic refinement of structural models of proteins, individual atomic B-factors enter as free parameters in the model. The B-factors are related to the rms displacements of independently fluctuating atoms. Here is a derivation of the diffuse scattering from a crystal where all atoms execute independent fluctuations about their equilibrium positions.

Consider the case where each atom m in the unit cell undergoes independent position fluctuations described by a variance matrix V_m, as defined in Equation 3.50. Considering each atom to be an independent rigid body, the results of section 3.3 can be drawn upon, using equation 3.32:

$$I_B = |\sum_m \langle f_m e^{i\mathbf{q}\cdot\mathbf{r}_{mn}}\rangle_n|^2 \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})} \tag{3.64}$$

and equation 3.27:

$$I_D = N\sum_m (\langle |f_m e^{i\mathbf{q}\cdot\mathbf{r}_{mn}}{}_n|^2\rangle - |\langle f_m e^{i\mathbf{q}\cdot\mathbf{r}_{mn}}\rangle_n|^2), \tag{3.65}$$

where f_m is the structure factor of atom m, $e^{i\mathbf{q}\cdot\mathbf{r}_{mn}}$ is the phase factor due to the displacement r_m of atom m from the origin in unit-cell n, and N is the number of unit cells in the crystal.

The phase factors can be handled in a manner analogous to that in section 3.5, so that

$$I_B = \sum_{m,m'} f_m f_{m'}^* e^{-\frac{1}{2}\mathbf{q}\cdot(\mathsf{V}_m+\mathsf{V}_{m'})\cdot\mathbf{q}} e^{i\mathbf{q}\cdot(\mathbf{r}_m-\mathbf{r}_{m'})} \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})} \tag{3.66}$$

and

$$I_D = N\sum_m |f_m|^2(1-e^{-\frac{1}{2}\mathbf{q}\cdot(\mathsf{V}_m+\mathsf{V}_{m'})\cdot\mathbf{q}}). \tag{3.67}$$

In the case where all of the variance matrices V_m are identical, so that

$$\mathsf{V}_m = \mathsf{V}_{m'} = \mathsf{V}, \tag{3.68}$$

the expressions for the Bragg scattering I_B and diffuse scattering I_D simplify, so that

$$I_B = e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}} \sum_{m,m'} f_m f_{m'}^* \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})} \tag{3.69}$$

and

$$I_D = N(1-e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}})\sum_m |f_m|^2. \tag{3.70}$$

The Bragg scattering I_B is identical in form to that in equation 3.49 for rigid-body translations. The diffuse scattering I_D differs, however: in the case where atoms execute independent fluctuations, the diffuse scattering has much less structure than the case where larger domains move coherently. As is seen in equation 3.70, in this case the structure in diffuse scattering comes from the atomic structure factors and the factor $(1-e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}})$. Both of these vary on very long length scales in reciprocal space compared with the higher-frequency variations which appear when larger units are moving coherently. The broadness of diffuse features in this model makes it a poor candidate for describing the diffuse scattering observed from *Staph.* nuclease, which shows a great deal of structure on length scales on the order of the size of the reciprocal unit cell.

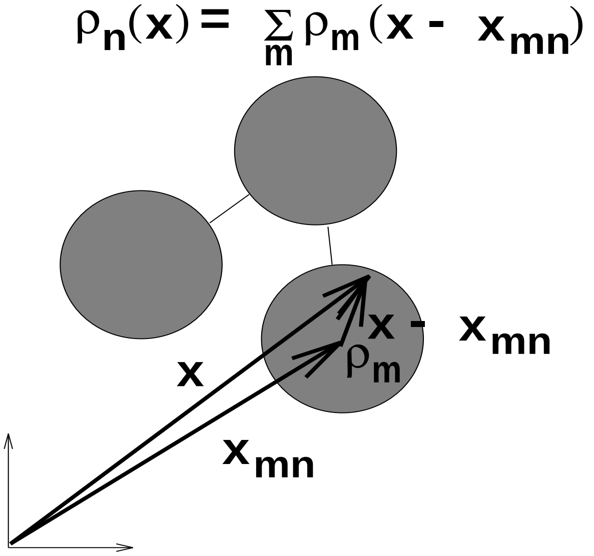

Figure 3.7: *Atomic electron densities contributing to the unit cell electron density.* Individual atomic electron densities contribute $\rho_m(\mathbf{x}-\mathbf{x}_{mn})$ to the n^{th} unit cell electron density ρ_n. $\mathbf{x}_{mn}$ is the atomic position of atom m in unit cell n with respect to the origin of unit cell n.

3.7 Rigid-Body Rotations

To fully understand what signatures of rigid-body motion are visible in diffuse scattering from protein crystals, it is necessary to consider the case of rigid-body rotations. Using a combination of translational and rotational disorder models, one can begin to look for features in diffuse scattering which would be expected if protein molecules in a crystal behave in the way suggested by experiments of Kuriyan and Weis[16] and Sternberg[15], where individual molecules execute independent rigid-body motions. Rigid-body rotation models, as described in Sternberg[15] for the case of lysozyme, can be useful in understanding large-scale motions of proteins by describing them in terms of independently moving structural domains. In order to further understand the effects of rigid-body motions on diffuse scattering, a model of rigid-body rotations, where proteins rotate about a center of libration, was derived.

Consider the case where there are rigid-body translations *and* rotations in the unit cell, with a single molecule per unit cell.[3] The unit-cell electron density $\rho_n(\mathbf{x})$ is the sum of the electron densities $\rho_m(\mathbf{x})$ of all atoms m in the unit cell. The center of atomic distribution ρ_m is located at $\mathbf{x}_{mn}$, which is the position with respect to the unit cell origin of atom m in unit cell n (see figure 3.7), so that

$$\rho_n(\mathbf{x}) = \sum_m \rho_m(\mathbf{x} - \mathbf{x}_{mn}) \tag{3.71}$$

and the unit cell structure factor becomes

$$\begin{aligned} f_n(\mathbf{q}) = & \sum_m \int d^3\mathbf{x} e^{i\mathbf{q}\cdot\mathbf{x}} \rho_m(\mathbf{x} - \mathbf{x}_{mn}) && (3.72) \\ = & \sum_m \left(\int d^3\mathbf{x} e^{i\mathbf{q}\cdot\mathbf{x}} \rho_m(\mathbf{x}) \right) e^{i\mathbf{q}\cdot\mathbf{x}_{mn}} && (3.73) \\ = & \sum_m f_m(\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{x}_{mn}} && (3.74) \end{aligned}$$

where the atomic structure factor $f_m(\mathbf{q})$ has been defined as

$$f_m(\mathbf{q}) \equiv \int d^3\mathbf{x} e^{i\mathbf{q}\cdot\mathbf{x}} \rho_m(\mathbf{x}). \tag{3.75}$$

The squared structure factor $|f_n(\mathbf{q})|^2$ is given by

$$|f_n(\mathbf{q})|^2 = \sum_{m,m'} f_m(\mathbf{q}) f_{m'}(\mathbf{q}) e^{i\mathbf{q}\cdot(\mathbf{x}_{mn} - \mathbf{x}_{m'n'})}. \tag{3.76}$$

In the case where the protein executes a rotation about a center of libration, (see figure 3.8)

$$\mathbf{x}_{mn} = \mathbf{r}_0 + \mathsf{R}_n \cdot \mathbf{r}'_m + \mathbf{u}_n, \tag{3.77}$$

---

[3] See, *e.g.*, p.150, chapter 3 of Giacovazzo[38] for a description of scattering by a molecule.

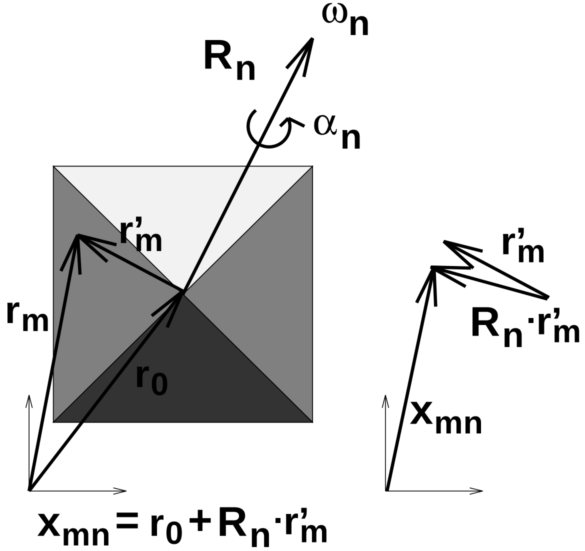

Figure 3.8: An illustration of a rotation R_n by an angle α_n about a direction $\hat{\omega}_n$ for an object with a rotationally invariant moment of inertia tensor in unit cell n. The vector $\mathbf{r}_0$ from the unit cell origin to the center of libration and the vector $\mathbf{r}_m$ from the origin to the reference position of atom m resolve the vector $\mathbf{r}'_m$, the average displacement of the atom m from the center of libration. R_n rotates the vector $\mathbf{r}'_m$ about the center of libration to $\mathsf{R}_n \cdot \mathbf{r}'_m$, translating the atomic position $\mathbf{r}_m$ to the position $\mathbf{x}_{mn} = \mathbf{r}_0 + \mathsf{R}_n \cdot \mathbf{r}'_m$.

where $\mathbf{r}_0$ is the vector to the center of libration,

$$\mathbf{r}'_m = \mathbf{r}_m - \mathbf{r}_0 \tag{3.78}$$

is the vector from the center of libration $\mathbf{r}_0$ to atom m at average position $\mathbf{r}_m$ in the unit cell, R_n is the rotation matrix for unit cell n, and $\mathbf{u}_n$ is a unit cell translation vector of the form introduced in section 3.5. Thus,

$$\begin{aligned} f_n = & \sum_m f_m e^{i\mathbf{q}\cdot\mathbf{x}_{mn}} && (3.79) \\ = & \sum_m f_m e^{i\mathbf{q}\cdot\mathbf{r}_0} e^{i\mathbf{q}\cdot(\mathsf{R}_n\cdot\mathbf{r}'_m + \mathbf{u}_n)}. && (3.80) \end{aligned}$$

As is derived in appendix A.5, the Bragg intensity I_B in this case is given by

$$\begin{aligned} I_B = & \sum_{m,m'} f_m f_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m - \mathbf{r}_{m'})} e^{-\frac{1}{2}(|\mathbf{q}\times\mathbf{r}'_m|^2 + |\mathbf{q}\times\mathbf{r}'_{m'}|^2)\sigma_\alpha^2} e^{-\mathbf{q}\cdot\mathsf{V}_\mathbf{u}\cdot\mathbf{q}}, && (3.81) \\ & \times \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})} && (3.82) \end{aligned}$$

where σ_α is the half-width of the Gaussian distribution of small angular displacements α, and $\mathsf{V}_\mathbf{u}$ is the variance matrix of the translations. Using the relation

$$|\mathbf{A}\times\mathbf{B}|^2 = |\mathbf{A}|^2|\mathbf{B}|^2 - |\mathbf{A}\cdot\mathbf{B}|^2, \tag{3.83}$$

and defining V_m by

$$\mathsf{V}_m \equiv (\mathbf{r}_m\cdot\mathbf{r}_m - \mathbf{r}_m\mathbf{r}_m)\sigma_\alpha^2, \tag{3.84}$$

this can be rewritten as

$$I_B = \sum_{m,m'} f_m f^*_{m'} e^{-\frac{1}{2}\mathbf{q}\cdot(\mathsf{V}_m + \mathsf{V}_{m'})\cdot\mathbf{q}} e^{i\mathbf{q}\cdot(\mathbf{r}_m - \mathbf{r}_{m'})} \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})}, \tag{3.85}$$

which is identical to equation 3.66 for independent atomic fluctuations.

The diffuse intensity I_D is given by

$$\begin{aligned} I_D = \quad & N \sum_{m,m'} f_m f_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m - \mathbf{r}_{m'})} \\ & \times \left(e^{-\frac{1}{2}|\mathbf{q}\times(\mathbf{r}'_m - \mathbf{r}'_{m'})|^2\sigma_\alpha^2} - e^{-\frac{1}{2}(|\mathbf{q}\times\mathbf{r}'_m|^2 + |\mathbf{q}\times\mathbf{r}'_{m'}|^2)\sigma_\alpha^2} e^{-\mathbf{q}\cdot\mathsf{V}_\mathbf{u}\cdot\mathbf{q}}\right). \end{aligned} \tag{3.86}$$

In order to make this expression more suitable for simulations, it would be good to re-express it in terms of the unit-cell structure factor $f(\mathbf{q})$. Expanding the exponentials in equation 3.86, it follows that

$$\left(e^{-\frac{1}{2}|\mathbf{q}\times(\mathbf{r}'_m - \mathbf{r}'_{m'})|^2\sigma_\alpha^2} - e^{-\frac{1}{2}(|\mathbf{q}\times\mathbf{r}'_m|^2 + |\mathbf{q}\times\mathbf{r}'_{m'}|^2)\sigma_\alpha^2} e^{-\mathbf{q}\cdot\mathsf{V}_\mathbf{u}\cdot\mathbf{q}}\right) \approx (\mathbf{q}\times\mathbf{r}'_m)\cdot(\mathbf{q}\times\mathbf{r}'_{m'})\sigma_\alpha^2 + \mathbf{q}\cdot\mathsf{V}_\mathbf{u}\cdot\mathbf{q} \tag{3.87}$$

giving

$$I_D = N(1 - e^{-\mathbf{q}\cdot\mathsf{V}_\mathbf{u}\cdot\mathbf{q}})|f(\mathbf{q})|^2 + N \sum_{m,m'} f_m f_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m - \mathbf{r}_{m'})} (\mathbf{q}\times\mathbf{r}'_m)\cdot(\mathbf{q}\times\mathbf{r}'_{m'})\sigma_\alpha^2, \tag{3.88}$$

from which the translational component

$$I_{DT} \equiv N(1 - e^{-\mathbf{q}\cdot\mathsf{V}_\mathbf{u}\cdot\mathbf{q}})|f(\mathbf{q})|^2 \tag{3.89}$$

and the rotational component

$$I_{DR} \equiv N \sum_{m,m'} f_m f_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m - \mathbf{r}_{m'})} (\mathbf{q}\times\mathbf{r}'_m)\cdot(\mathbf{q}\times\mathbf{r}'_{m'})\sigma_\alpha^2 \tag{3.90}$$

are identified. The rotational component I_{DR} can be rewritten as

$$I_{DR} = \quad N|\mathbf{q}\times\sum_m f_m \mathbf{r}'_m e^{i\mathbf{q}\cdot\mathbf{r}_m}|^2\sigma_\alpha^2 \tag{3.91}$$

$$= \quad N|\mathbf{q}\times[\sum_m f_m \mathbf{r}_m e^{i\mathbf{q}\cdot\mathbf{r}_m} - \mathbf{r}_0 \sum_m f_m e^{i\mathbf{q}\cdot\mathbf{r}_m}]|^2\sigma_\alpha^2 \tag{3.92}$$

$$= \quad N|\mathbf{q}\times[\boldsymbol{\nabla}_q f(\mathbf{q}) - i\mathbf{r}_0 f(\mathbf{q})]|^2\sigma_\alpha^2 \tag{3.93}$$

Thus, I_D can be expressed as

$$I_D = \quad I_{DT} + I_{DR} \tag{3.94}$$

$$= \quad N(1 - e^{-\mathbf{q}\cdot\mathbf{V_u}\cdot\mathbf{q}})|f(\mathbf{q})|^2 + N|\mathbf{q}\times[\boldsymbol{\nabla}_q f(\mathbf{q}) - i\mathbf{r}_0 f(\mathbf{q})]|^2\sigma_\alpha^2 \tag{3.95}$$

The effect of adding a small amount of rotational disorder is to introduce a term of the form of I_{DR} above to the diffuse scattering. Although the above expressions have been derived for the case where the entire unit cell is moving as a rigid body, the results of sections 3.3 and 3.4 can be used to extend the results to multiple molecules or structural domains in the unit cell. Thus, the case where individual proteins in the unit cell are moving like rigid bodies can be studied, in order to look for diffuse features to complement studies like those performed by Sternberg *et. al.*[15] and Kuriyan and Weis[16]. In addition, the internal dynamics of multiple, isomorphous molecules in the unit cell can be modelled by subdividing the reference molecule into independently moving, rigid domains and applying symmetry operations as discussed in section 3.4.

3.8 Liquid-Like Correlated Motions

In the more general class of models of diffuse scattering, correlations are allowed among unit-cell electron-density distributions. The most noticeable effect of this on the diffuse scattering is the fact that it introduces features which can vary on length scales smaller than the dimensions of the reciprocal unit cell. As will be shown, to first order, these large-scale motions in the crystal produce features whose intensities are closely tied to the intensity of the neighboring Bragg peak, and thus, like the Bragg peaks, show correlations with the map of $|f(\mathbf{q})|^2$, the squared Fourier-transform of the unit cell.

In the case where the correlations in fluctuations are solely determined by the displacement $\mathbf{r}$ between the points in consideration, Caspar *et. al.*[19] have derived an expression for modelling the diffuse intensity, which was used to measure parameters of disorder in the crystal. A more detailed derivation of the expression is given in Clarage *et. al.*[22]. This model allows for correlations between unit-cell electron densities, since correlations between any two atoms in the crystal are solely determined by the vector distance between them. As is described in the introduction, their measurements come about by visually comparing simulated diffraction images with a single image of x-ray diffraction from an insulin crystal. Extending their measurements into three dimensions is possible with the techniques developed in this thesis. A modified derivation of their expression for diffuse scattering is presented here.

One can model the electron density of the crystal as

$$\rho(\mathbf{x}) = \rho_{point}(\mathbf{x}) = \sum_m Q_m \delta(\mathbf{x} - \mathbf{x}_m), \tag{3.96}$$

where $\mathbf{x}_m = \mathbf{r}_m + \mathbf{u}_m(t)$, $\mathbf{r}_m$ being the average position of atom m, $\mathbf{u}_m$ being its time-dependent displacement, and m indexing all atoms in the crystal. Q_m is the charge of atom m as seen by the x-rays. It follows that

$$\begin{aligned} I(\mathbf{q}) = \langle I(\mathbf{q})\rangle = & \left\langle \sum_{m,m'} Q_m Q_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m - \mathbf{r}_{m'})} \times e^{i\mathbf{q}\cdot(\mathbf{u}_m - \mathbf{u}_{m'})} \right\rangle_t & (3.97) \\ = & \sum_{m,m'} Q_m Q_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m - \mathbf{r}_{m'})} \times \langle e^{i\mathbf{q}\cdot(\mathbf{u}_m - \mathbf{u}_{m'})} \rangle_t, & (3.98) \end{aligned}$$

which, for small displacements $\mathbf{u}_m$, becomes

$$I(\mathbf{q}) \approx \sum_{m,m'} Q_m Q_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m - \mathbf{r}_{m'})} \times e^{-\frac{1}{2}\langle[\mathbf{q}\cdot(\mathbf{u}_m - \mathbf{u}_{m'})]^2\rangle_t}. \tag{3.99}$$

The last exponent in equation 3.99 can be expanded to give

$$e^{-\frac{1}{2}\langle[\mathbf{q}\cdot(\mathbf{u}_m - \mathbf{u}_{m'})]^2\rangle_t} = e^{-\frac{1}{2}\mathbf{q}\cdot\mathsf{V}_m\cdot\mathbf{q} - \frac{1}{2}\mathbf{q}\cdot\mathsf{V}_{m'}\cdot\mathbf{q} + \langle\mathbf{q}\cdot\mathbf{u}_m\mathbf{u}_{m'}\cdot\mathbf{q}\rangle_t}, \tag{3.100}$$

where the individual atomic variance matrices V_m are defined as in section 3.6. It is further assumed that each atom has identical RMS displacements:

$$\mathsf{V}_m = \mathsf{V}_{m'} = \mathsf{V}. \tag{3.101}$$

Then, a correlation matrix $\mathsf{C}_{mm'}$ is defined such that

$$e^{-\frac{1}{2}\langle[\mathbf{q}\cdot(\mathbf{u}_m - \mathbf{u}_{m'})]^2\rangle_t} = e^{-\mathbf{q}\cdot(\mathsf{V} - \mathsf{C}_{mm'})\cdot\mathbf{q}}, \tag{3.102}$$

so that

$$\mathsf{C}_{mm'} = \langle\mathbf{u}_m\mathbf{u}_{m'}\rangle_t \tag{3.103}$$

It is assumed that the correlation matrices $\mathsf{C}_{mm'}$ solely depend upon the average displacement $\mathbf{r}_m - \mathbf{r}_{m'}$ between atoms:

$$\mathsf{C}_{mm'} = \mathsf{C}(\mathbf{r}_m - \mathbf{r}_{m'}) \tag{3.104}$$

so that

$$I(\mathbf{q}) = \sum_{m,m'} Q_m Q_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m - \mathbf{r}_{m'})} \times e^{-\mathbf{q}\cdot[\mathsf{V} - \mathsf{C}(\mathbf{r}_m - \mathbf{r}_{m'})]\cdot\mathbf{q}}. \tag{3.105}$$

Making use of the Dirac delta function, this can be rewritten as:

$$\begin{aligned} I(\mathbf{q}) = & \int d^3\mathbf{r} e^{i\mathbf{q}\cdot\mathbf{r}} e^{-\mathbf{q}\cdot[\mathsf{V}-\mathsf{C}(\mathbf{r})]\cdot\mathbf{q}} \sum_{m,m'} Q_m Q_{m'} \delta(\mathbf{r}-\mathbf{r}_m+\mathbf{r}_{m'}) \\ = & \int d^3\mathbf{r} e^{i\mathbf{q}\cdot\mathbf{r}} e^{-\mathbf{q}\cdot[\mathsf{V}-\mathsf{C}(\mathbf{r})]\cdot\mathbf{q}} P_0(\mathbf{r}) \end{aligned}$$

where

$$P_0(\mathbf{x}) \equiv \sum_{m,m'} Q_m Q_{m'} \delta(\mathbf{x}+\mathbf{r}_m-\mathbf{r}_{m'}) \tag{3.106}$$

is the Patterson function for the unperturbed crystal. For $\mathbf{q}\cdot\mathsf{C}\cdot\mathbf{q} \ll 1$,

$$\begin{aligned} I(\mathbf{q}) \approx & \quad e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}} \int d^3\mathbf{r} e^{i\mathbf{q}\cdot\mathbf{r}} (1+\mathbf{q}\cdot\mathsf{C}(\mathbf{r})\cdot\mathbf{q}) P_0(\mathbf{r}) \\ = & \quad e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}} \int d^3\,\mathbf{r} e^{i\mathbf{q}\cdot\mathbf{r}}\, P_0(\mathbf{r}) \\ + & \quad e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}} \mathbf{q}\cdot\left[\int d^3\,\mathbf{r} e^{i\mathbf{q}\cdot\mathbf{r}}\,\mathsf{C}(\mathbf{r})\,P_0(\mathbf{r})\right]\cdot\mathbf{q} \qquad (3.107) \\ = & \quad e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}} I_0(\mathbf{q}) \\ + & \quad e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}} \mathbf{q}\cdot[I_0(\mathbf{q}) * \mathsf{c}(\mathbf{q})]\cdot\mathbf{q}, \qquad (3.108) \end{aligned}$$

where the scattering intensity $I_0(\mathbf{q})$ from the unperturbed crystal is given by

$$I_0(\mathbf{q}) = \int d^3\mathbf{r} e^{i\mathbf{q}\cdot\mathbf{r}} P_0(\mathbf{r}), \tag{3.109}$$

and the correlation matrix $\mathsf{c}(\mathbf{q})$ is defined as

$$\mathsf{c}(\mathbf{q}) \equiv \int d^3\mathbf{r} e^{i\mathbf{q}\cdot\mathbf{r}} \mathsf{C}(\mathbf{r}). \tag{3.110}$$

Finally, separating the Bragg component I_B and the diffuse component I_D,

$$I_B = e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}} I_0(\mathbf{q}) \tag{3.111}$$

and

$$I_D = e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}}\mathbf{q}\cdot[I_0(\mathbf{q}) * \mathsf{c}(\mathbf{q})]\cdot\mathbf{q}. \tag{3.112}$$

The above expression for I_D is valid in the case where both

1. The displacements of all of the atoms in the crystal can be described by a single variance matrix V, and
2. The correlation $\mathsf{C}_{mm'}$ between any two atoms m and m' only depends upon the vector displacement between them.

If the atomic displacements are isotropic, so that

$$\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q} = q^2\sigma^2, \tag{3.113}$$

and the correlation $\mathsf{C}_{mm'}$ between any two atoms is only dependent upon the distance between them, so that

$$\mathbf{q}\cdot\mathsf{c}(\mathbf{q})\cdot\mathbf{q} = q^2 c(\mathbf{q}), \tag{3.114}$$

the diffuse scattering I_D can be rewritten as

$$I_D = e^{-q^2\sigma^2} q^2\sigma^2 I_0(\mathbf{q}) * \Gamma(\mathbf{q}), \tag{3.115}$$

where $\Gamma(\mathbf{q})$ has been defined as

$$\Gamma(\mathbf{q}) \equiv \frac{c(\mathbf{q})}{\sigma^2}. \tag{3.116}$$

As a check, it is shown in appendix A.6 that this expression correctly predicts the diffuse scattering in the case of independent atomic fluctuations discussed in section 3.6.

Clarage *et. al.*[22] used an equation analogous to equation 3.115 to model disorder in lysozyme crystals. They assumed that the correlations in the crystal were such that the diffuse scattering took on the form:

$$I_D(\mathbf{q}) = e^{-q^2[\sigma_1^2+\sigma_2^2]} q^2 \times I_0(\mathbf{q}) * [\sigma_1^2 \Gamma_1(\mathbf{q}) + \sigma_2^2 \Gamma_2(\mathbf{q})], \tag{3.117}$$

where $\Gamma_1(\mathbf{q})$ and $\Gamma_2(\mathbf{q})$ are distributions of the form

$$\Gamma(\mathbf{g}) = \frac{8\pi\gamma^3}{(1+g^2\gamma^2)^2}, \tag{3.118}$$

which is the Fourier representation of a simple exponential decay function with widths γ_1 and γ_2 respectively. By simulating an image of x-ray diffraction from tetragonal lysozyme and visually comparing it to an experimentally obtained diffraction image, they obtained the values $\sigma_1 = 0.11$ **Å**, $\sigma_2 = 0.5$ **Å**, $\gamma_1 = 50$ **Å** and $\gamma_2 = 6$ **Å**. The correlation function Γ_1 was associated with small, long-range correlated motions, while Γ_2 was associated with larger, short-range correlated motions, as are seen in a liquid. Although the simulations are only of single two-dimensional diffraction images, reciprocal space is sampled at a finer resolution in these simulations than is done in the techniques outlined in this thesis, enabling comparison of Bragg peak profiles to estimate the correlation length of the long-range motions.

Using the three-dimensional maps obtained from *Staph.* nuclease, a comparison can be made between the predictions of short-range correlated motions and experimental data. Maps of $I_0(\mathbf{q})$ in equation 3.115 will have the form

$$I_0(\mathbf{q}) = |f(\mathbf{q})|^2 \sum_{hkl} \delta(\mathbf{q} - 2\pi[\mathbf{a}^* h + \mathbf{b}^* k + \mathbf{c}^* l]), \tag{3.119}$$

where $|f(\mathbf{q})|^2$ is the squared Fourier transform of the unit cell, (h, k, l) are Miller index coordinates for Bragg reflections, and $\mathbf{a}^*, \mathbf{b}^*$ and $\mathbf{c}^*$ are basis vectors for the reciprocal unit cell. Experimental diffuse maps in this model, therefore, will show a high correlation with smoothed maps of predicted squared structure factors $|f(\mathbf{q})|^2$ of the unit cell, where discrete peaks at positions $2\pi[\mathbf{a}^*h+\mathbf{b}^*k+\mathbf{c}^*l]$ are broadened by an amount defined by the width of the function $\Gamma(\mathbf{q})$. Section 4.5 describes how smoothed experimental diffuse maps of *Staph.* nuclease were compared with smoothed simulated maps of $|f(\mathbf{q})|^2$ to demonstrate a correlation between them, providing evidence for the type of disorder described by this model.

3.9 Crystalline Normal Modes

The standard way to deal with thermally-induced disorder in a crystal is to expand the perturbations about equilibrium in terms of excited crystalline normal modes. In fact, it is possible to represent the liquid-like correlated motions model as an equivalent model in terms of crystalline normal modes. Rendering the model in this alternate way allows interpretation of the correlated motion identified in Caspar *et. al.*[19] in terms of periodic lattice vibrations.

The normal-modes treatment not only provides an alternate way of modelling liquid-like motions: the most general expressions allow for motions of such great complexity that one could not hope to separate their contributions to the diffuse scattering in order to fingerprint them. What follows is a derivation of a general expression for the diffuse intensity from a crystal undergoing normal-mode excitations.[4] Afterwards is a demonstration of the equivalence of the liquid-like correlated motions model and a limiting case of the normal-modes model, which is explicitly shown by introducing a spatial-frequency dependence to the elastic modulus of the crystal. The

---

[4] James[41], Warren[42] and Ashcroft and Mermin[43] are three of many texts which describe the normal-modes treatment of perturbations. The extension to molecular crystals, which proved difficult to locate in the literature, used suggestions found on pp. 442-443 of Ashcroft and Mermin[43].

diffuse maps predicted by this model are identical to those predicted by the liquid-like motions model, but the resulting interpretation is an original one.

Consider a macromolecular crystal undergoing harmonic structural perturbations. The average intensity from such a crystal has the following form:

$$\langle I(\mathbf{q},t)\rangle_t = \sum_{m,m'} Q_m Q_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m - \mathbf{r}_{m'})} \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})} \langle e^{i\mathbf{q}\cdot(\mathbf{u}_{mn} - \mathbf{u}_{m'n'})}\rangle_t, \tag{3.120}$$

where Q_m and $\mathbf{r}_m$ are the charge and average position of atom m in the unit cell, $\mathbf{R}_n$ is the position of unit cell n with respect to the origin, and $\mathbf{u}_{mn}$ is the displacement of atom m in unit cell n. The atomic displacements $\mathbf{u}_{mn}$ can be expanded in terms of crystalline normal modes, so that

$$\mathbf{u}_{mn} = \sum_{\mathbf{g}s} a_{s,\mathbf{g}} \mathbf{e}_{ms,\mathbf{g}} e^{i(\mathbf{g}\cdot\mathbf{R}_n - \omega_s(\mathbf{g})t)}, \tag{3.121}$$

where $\mathbf{e}_{ms,\mathbf{g}}$ is the complex polarization vector for atom m in branch s of the mode with wave vector $\mathbf{g}$, $a_{s,\mathbf{g}}$ is the amplitude of branch s of mode $\mathbf{g}$, and $\omega_{s,\mathbf{g}}$ is its frequency. As is shown in appendix A.7, in an approximation which is first order in $\mathbf{u}_{mn}$, two additive contributions I_B and I_D, such that

$$I = I_B + I_D, \tag{3.122}$$

can be identified in the expression for the scattering intensity. The Bragg intensity I_B is given by

$$I_B = e^{-\frac{1}{2}\mathbf{q}\cdot(\mathsf{V}_m + \mathsf{V}_{m'})\cdot\mathbf{q}} I_0(\mathbf{q}), \tag{3.123}$$

where the atomic variance matrices V_m are given by

$$\mathsf{V}_m = \sum_{\mathbf{g},s} |a_{s,\mathbf{g}}|^2 \mathbf{e}_{ms,\mathbf{g}} \mathbf{e}_{ms,-\mathbf{g}}, \tag{3.124}$$

and $I_0(\mathbf{q})$ has been defined as

$$I_0(\mathbf{q}) = \sum_{m,m'} Q_m Q_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m - \mathbf{r}_{m'})} \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})}. \tag{3.125}$$

The diffuse intensity I_D is given by

$$\begin{aligned} I_D = \quad & \sum_{\mathbf{g},s} \sum_{n,n'} e^{i(\mathbf{q}+\mathbf{g})\cdot(\mathbf{R}_n - \mathbf{R}_{n'})} \sum_{m,m'} Q_m Q_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m - \mathbf{r}_{m'})} \\ \times \quad & e^{-\frac{1}{2}\mathbf{q}\cdot(\mathsf{V}_m + \mathsf{V}_{m'})\cdot\mathbf{q}} \\ \times \quad & \mathbf{q}\cdot\left[|a_{s,\mathbf{g}}|^2 \mathbf{e}_{ms,\mathbf{g}} \mathbf{e}_{m's,-\mathbf{g}}\right]\cdot\mathbf{q} \end{aligned} \tag{3.126}$$

The above expressions for I_B and I_D represent the first-order scattering from a macromolecular crystal undergoing normal-mode excitations. As is shown in appendix A.7, the connection with the liquid-like motions model is established by both assumimg that all of the atomic variances V_m are the same, so that

$$\mathsf{V}_m = \mathsf{V}_{m'} = \mathsf{V}, \tag{3.127}$$

and assuming that the polarization vectors $\mathbf{e}_{ms,\mathbf{g}}$ all have the form

$$\mathbf{e}_{ms,\mathbf{g}} = \mathbf{e}_{s,\mathbf{g}} e^{i\mathbf{g}\cdot\mathbf{r}_m}, \tag{3.128}$$

which is true for an elastically homogeneous crystal, where waves continuously modulate atomic displacements within the unit cell. It can be shown that the diffuse intensity I_D can be expressed as

$$I_D = e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}}\mathbf{q}\cdot[I_0(\mathbf{q}) * \mathsf{c}_{NM}(\mathbf{q})]\cdot\mathbf{q}, \tag{3.129}$$

which is the same as equation 3.112. The correlations $\mathsf{c}_{NM}(\mathbf{g})$ are given by

$$\mathsf{c}_{NM}(\mathbf{g}) \equiv n(\mathbf{g})\sum_s |a_s(\mathbf{g})|^2 \mathbf{e}_s(\mathbf{g})\mathbf{e}_s(-\mathbf{g}), \tag{3.130}$$

where $n(\mathbf{g})$ is the density of states. Thus, the liquid-like correlated motions model is seen to be equivalent to a limiting case of the crystalline normal modes model, where the displacements of all atoms are equivalently distributed, and where the elastic properties of the crystal are homogeneous.

Finally, a connection can be established between the parameters of disorder in the liquid-like motions model and the elastic modulus of the crystal. Let s label the polarization of mode $\mathbf{g}$, where $s = 1$ stands for the longitudinally polarized mode, and $s = 2$ and $s = 3$ stand for the two transversely polarized modes. For each of these modes, the total energy contribution $E_{s,\mathbf{g}}$ of mode $\mathbf{g}$ is given by

$$E_{s,\mathbf{g}} = \frac{1}{2}n(\mathbf{g})K_s g^2|a_s(\mathbf{g})|^2, \tag{3.131}$$

where K_s is the elastic modulus for modes of polarization s, and V is the volume of the crystal. By the equipartition theorem,

$$E_{s,\mathbf{g}} = \frac{1}{2}k_B T, \tag{3.132}$$

where T is the temperature and k_B is Boltzmann's constant. This leads to

$$|a_s(\mathbf{g})|^2 = \frac{k_B T}{n(\mathbf{g}) K_s g^2} \tag{3.133}$$

as an expression for the amplitude $a_s(\mathbf{g})$ of branch s of mode $\mathbf{g}$. This leads to the usual g^{-2} spreading of Bragg peaks due to first-order excitation of normal modes in the crystal, as used by Glover *et. al.*[21] to describe the peak profiles in x-ray diffraction from protein crystals.

If there is a significant difference in the amplitudes of the transversely and longitudinally polarized modes in the crystal, it will be reflected in the distribution of diffuse intensity about the Bragg peaks. From equation 3.130, one can predict that longitudinally polarized modes will broaden peaks along a line connecting the peak and the origin, while transversely polarized modes broaden peaks in directions perpendicular to this line. The observation of long radial streaks in diffraction images, therefore, would correspond to the excitation of relatively soft longitudinal modes in the crystal, while the observation of axial streaks would indicate that transverse modes were softer.

If the direction of observed streaks around Bragg peaks appears to be independent of the peak position, it indicates that there are absolute directions in the crystal which correspond to high-amplitude modes. For example, if the contacts between unit cells were particularly soft along the $c-$axis, one might expect to see streaks along the direction of $\mathbf{c}$ in reciprocal space. Since the techniques developed in this thesis measure diffuse features on length scales larger than that of the reciprocal unit cell, well-defined streaks would most likely not be observed in the three-dimensional diffuse maps. Broad asymmetries, however, could possibly be observed, although none were obviously visible in the maps obtained from *Staph.* nuclease.

In the case of the liquid-like motions model, the correlation function $\Gamma(\mathbf{g})$ is given the form

$$\Gamma(\mathbf{g}) = \frac{8\pi\gamma^3}{(1+g^2\gamma^2)^2}, \tag{3.134}$$

by Caspar *et. al.*[19] and Clarage *et. al.*[22], which is normalized to $8\pi^3$ in order to satisfy the real-space relation $\mathsf{C}(0) = \sigma^2$. To demonstrate an explicit connection between equation 3.130 and this expression, the assumption is made that the elastic modulus K_s for each polarization is the same value of K, making amplitudes of vibration $a_s(\mathbf{g})$ due to all polarizations s the same value $a(\mathbf{g})$, so that, from equation 3.116,

$$\Gamma_{NM}(\mathbf{g}) = n(\mathbf{g})\frac{|a(\mathbf{g})|^2}{\sigma^2} \tag{3.135}$$

where σ^2 is defined by

$$\sigma^2 q^2 = \mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}. \tag{3.136}$$

Substituting the expression for $|a_s(\mathbf{g})|^2$ from equation 3.133:

$$\Gamma_{NM}(\mathbf{g}) = \frac{k_B T}{K g^2 \sigma^2}. \tag{3.137}$$

To make this look like the distribution in equation 3.134, it is necessary to introduce a dependence of the elastic modulus K on the mode $\mathbf{g}$, such that

$$K(\mathbf{g}) = K_{min}\frac{(1+\gamma^2 g^2)^2}{4\gamma^2 g^2}, \tag{3.138}$$

so that

$$\Gamma_{NM}(\mathbf{g}) = \frac{4\gamma^2 k_B T}{\sigma^2 K_{min}(1+\gamma^2 g^2)^2}, \tag{3.139}$$

where K_{min} is the minimum value for the elastic modulus, and γ^{-1} is the width of the distribution in g. Choosing K_{min} such that

$$K_{min} = \frac{1}{2\pi}\frac{k_B T}{\sigma^2\gamma} \tag{3.140}$$

yields

$$\Gamma_{NM}(\mathbf{g}) = \frac{8\pi\gamma^3}{(1+\gamma^2 g^2)^2}, \tag{3.141}$$

establishing the form of the distribution, and showing the connection with the explicit form of correlations assumed by Caspar *et. al.*[19] and Clarage *et. al.*[22] in modelling diffuse scattering in terms of liquid-like motions. For the short-range correlated motions in lysozyme discussed at the end of section 3.8, $\sigma^2 = 0.25\ \mathbf{\AA}^2$, and $\gamma = 6\ \mathbf{\AA}$, yielding

$$K_{min} = \quad 3\times 10^{-3}\ ev\,\mathbf{\AA}^{-3} \tag{3.142}$$

$$= \quad 4\times 10^{8}\ N\,m^{-2} \tag{3.143}$$

at 300 K. For comparison, the bulk modulus of water is roughly $2\times 10^9\ N\,m^{-2}$ at room temperature.

Chapter 4

Analysis Methods

This chapter describes the analysis methods which were developed and which yielded the results described in chapter 5. A combination of three-dimensional visualization tools and number-crunching code was used to understand the diffuse maps which were measured as described in chapter 2. Interactive manipulation of the map rendered in three dimensions allowed inspection of salient features which were recognized in all three data sets, and allowed an understanding and identification of the symmetries present in the maps. A graphical *shell image* representation of the diffuse maps allowed a more complete three-dimensional visualization of diffuse features.

Measures of the internal symmetry and reproducibility of diffuse maps were defined and applied to the data sets collected. These same measures were used to look for changes in diffuse scattering between the bare crystal and the co-crystal, and to characterize the quality of fit of simulated maps to experimentally obtained maps. The results of these fits were used to identify the types of disorder which are most consistent with our observed diffuse maps.

4.1 Visualization Using EXPLORER

Primary visualization of diffuse maps was performed using the EXPLORER data visualization package[1] (see figure 5.11). The program renders a 3D representation of the map, and allows interactive manipulations to explore the symmetry of the visible features. Using this program, it was possible to get a qualitative feel for the data and to become familiar with the distribution of the features in the maps. Comparison of two rendered views of the two different measured diffuse maps provided visible evidence for the reproducibility of the data (see figure 5.12).

4.2 Visualization Using Shell Images

Although EXPLORER provided a good way to qualitatively study the data, a more quantitative way of displaying the data was sought. What was desired was a way to observe features in a diffuse map, and to easily answer the question: "Where in reciprocal space is this diffuse feature located?" The method should also display the maps in such a way that symmetries are readily apparent, and such that features identified using EXPLORER could be found on the resulting graphs. The following describes the method which was developed.

In order to compactly display an entire diffuse map, the lattice was divided into spherical shells about the origin. Each shell had a width of one reciprocal-unit-cell diagonal. For each shell at scattering vector $|\mathbf{q}|$, a *shell image* was generated which represented voxel value measurements on the spherical shell. Pixel coordinates in the image linearly correspond to polar angles (θ, ϕ) on the shell: θ ranges from 0 at the top of the image to π at the bottom of the image, while ϕ ranges from $-\pi$ to π from

---

[1] Available through the IRIX 5.2 distribution for Silicon Graphics workstations

left to right. The pixel value in an image is calculated by averaging the values of all voxels whose polar angles lie in the angular region spanned by the pixel, and whose radius falls between the inner and outer radius of the shell.

For *Staph.* nuclease, a section of the shell restricted to $(0, \pi)$ in θ and $(-\pi, -\frac{\pi}{2})$ in ϕ was mapped to the cartesian shell image, to increase clarity at the cost of eliminating redundant data points. Images of the diffuse map obtained from the first crystal are displayed in figure 4.1. The symmetry of the map is still visible in the images generated in this way. In addition, the coordinates corresponding to an arbitrary feature are easily calculated once θ and ϕ are measured from a shell image at scattering vector $|\mathbf{q}|$:

$$\begin{aligned} q_x &= |\mathbf{q}| \cos\phi \sin\theta \\ q_y &= |\mathbf{q}| \sin\phi \sin\theta \\ q_z &= |\mathbf{q}| \cos\theta. \end{aligned} \tag{4.1}$$

4.3 Self-Consistency of Diffuse Maps

There are a large number of symmetry-related points in maps of diffuse scattering. Most models of diffuse scattering predict that diffuse maps bear the same symmetry as that predicted for Bragg reflections. In the case of *Staph.* nuclease, the $P4_1$ symmetry of the unit cell predicts that each point in the lattice has seven symmetry-related points. If a point is labelled by (x, y, z), and the $z-$axis is the four-fold axis, symmetry-equivalent points are $(-y, x, z)$, $(-x, -y, z)$, $(y, -x, z)$, $(-x, -y, -z)$, $(y, -x, -z)$, $(x, y, -z)$, and $(-y, x, -z)$, the first three being related by four-fold symmetry about the $z-$axis, and the latter four being the Friedel mates of the others.

It is possible that diffuse maps would not bear the same symmetry as that pre-

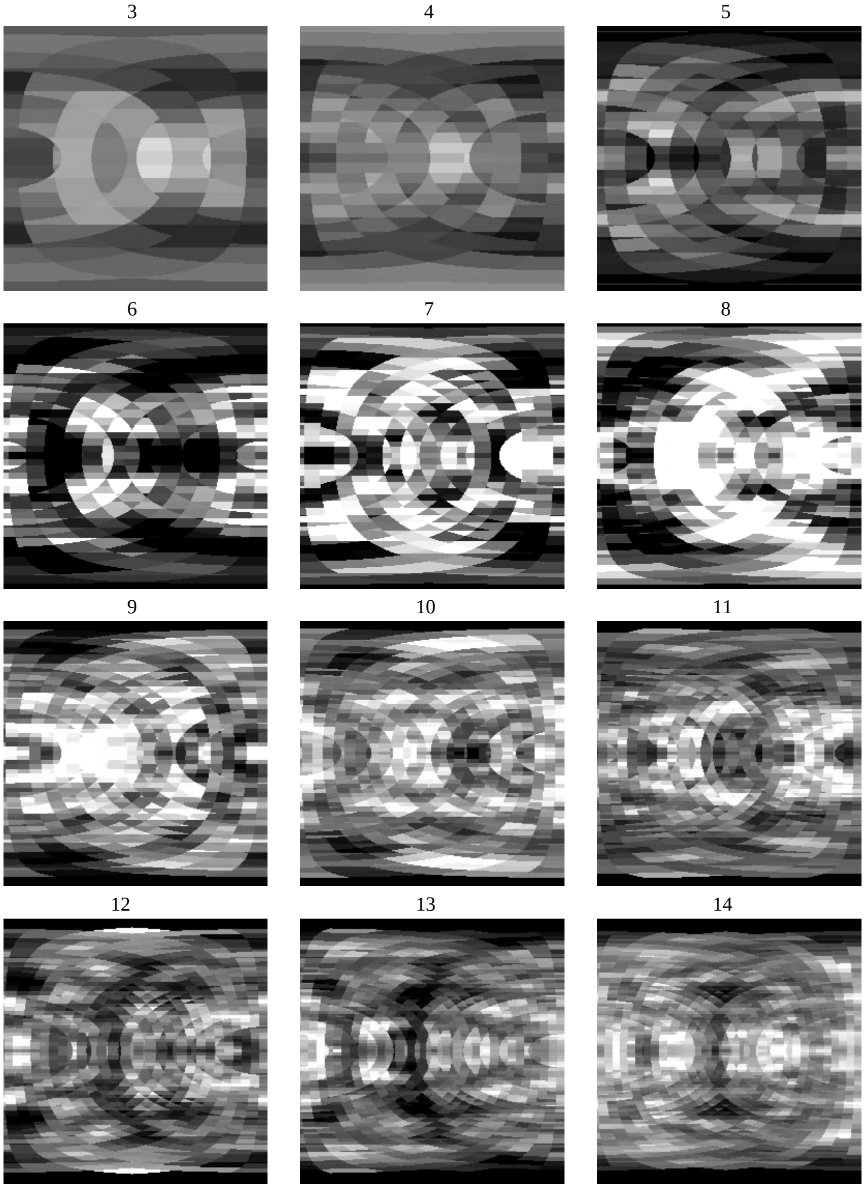

Figure 4.1: *Shell images of the diffuse map from* Staph. *nuclease crystal 1.* As described in the text, an image of shell $|\mathbf{q}|$ represents a map of diffuse intensity indexed by polar-coordinate angles (θ, ϕ). The numbers n above each image are related to the shell's scattering vector length s by the relation $s = 0.033n$ $\textbf{Å}^{-1}$. The images are displayed on a linear grey scale, with a pixel value of -100 corresponding to black, and 100 corresponding to white.

dicted by the space group of the unit cell. For this to be the case, however, there would have to be a symmetry in the average structure of the unit cell without a corresponding symmetry in the dynamics. Although such a situation is unlikely, it is possible, and experimental observation of diffuse features which do not bear the same symmetry as that of the Bragg peaks would give evidence for this unusual case.

Due to the expected symmetry of the maps, a measure of the degree to which a map is symmetric is used to characterize the self-consistency of the data. First, a symmetric lattice is generated by symmetry-averaging the original map according to the above predictions of a P4_1 space group. This was done by replacing each voxel value at (x, y, z) with the average value of voxels (x, y, z), $(-y, x, z)$, $(-x, -y, z)$, $(y, -x, z)$, $(-x, -y, -z)$, $(y, -x, -z)$, $(x, y, -z)$, and $(-y, x, -z)$.

For both the original and the symmetry-averaged maps, the average voxel value was then evaluated within thin spherical shells about the origin, producing profiles of the spherically-averaged background scattering at given resolution values in reciprocal space (see figure 5.8)[2]. Finally, for each voxel, the spherically-averaged component of the scattering was interpolated from the profile and was subtracted from the voxel value.

A difference lattice was constructed by subtracting the two resulting lattices voxel-by-voxel. The RMS voxel value of the difference lattice was calculated in spherical resolution shells, creating a profile of the amplitude of the difference between the two lattices[3]. The thickness of the shells was equal to the length of the diagonal of the reciprocal unit cell in the Bragg lattice. Similar profiles were calculated from the original and symmetrized maps, and all were displayed simultaneously for comparison (see figure 4.2).

---

[2] These profiles may hereafter be referred to as *average intensity profiles.*

[3] These profiles may hereafter be referred to as *RMS intensity profiles.*

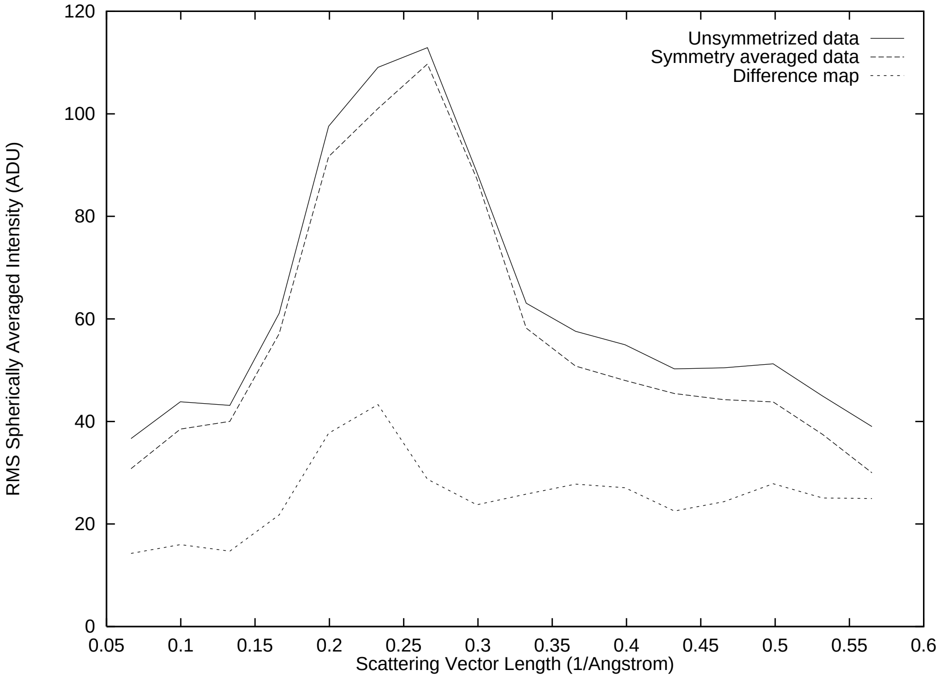

Figure 4.2: *A measure of the self-consistency of diffuse maps.* The difference between the original diffuse map and a symmetry-averaged map is calculated and its RMS amplitude is plotted in resolution shells. Profiles of the RMS amplitude of the original and symmetrized maps are shown for comparison.

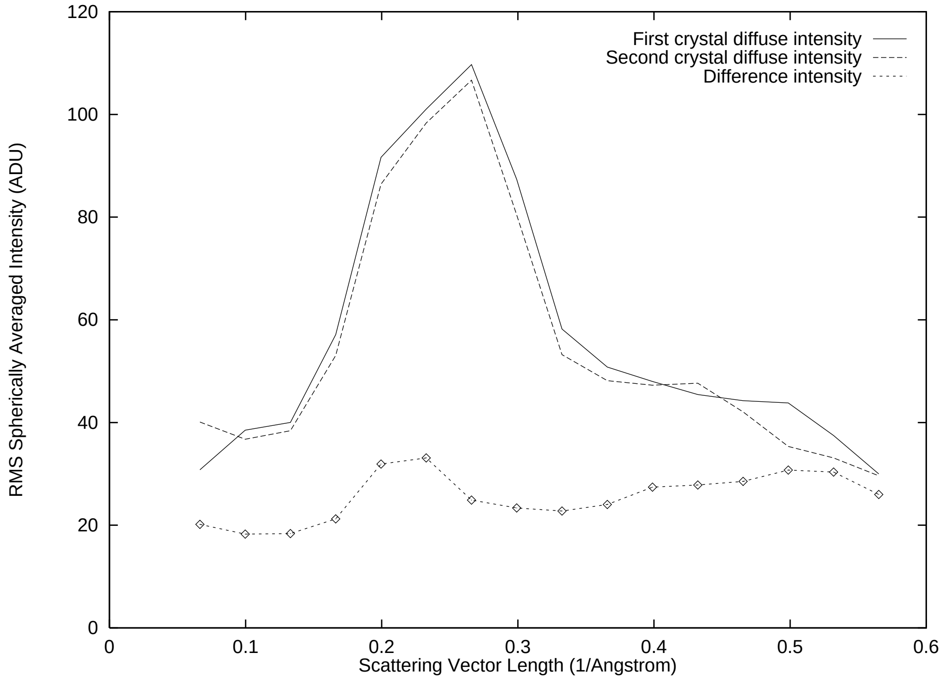

Figure 4.3: *Graphical representation of the reproducibility of diffuse data.* Shown are the RMS amplitudes of diffuse maps from the first and second crystals, after the maps have been symmetry averaged, and the spherically averaged component has been subtracted. Also shown is the RMS amplitude of the difference map.

4.4 Reproducibility of Diffuse Maps

The reproducibility of diffuse maps was tested by comparing two data sets obtained from different crystals grown under identical conditions. Each map was symmetry-averaged in order to reduce systematic errors due to, *e.g.*, the rotation between experiments of the detector face with respect to the incident beam. Then, the spherically-averaged component of each lattice was subtracted (see section 4.3).

A difference lattice was constructed by subtracting the two lattices voxel-by-voxel in an identical manner to that described in section 4.3. The RMS voxel value of the difference lattice was calculated in spherical resolution shells, creating a profile of the amplitude of the difference between the two lattices. Using a program written for the task, a scale factor between the two lattices was adjusted until a minimum difference profile was detected. By applying this method to the two data sets, a difference profile was generated which shows that the maps differ by as little as 20% in the neighborhood of 0.26 $\mathbf{\AA}^{-1}$, where many of the interesting diffuse features are seen (see figure 4.3).

Comparison of figure 4.2 with figure 4.3 shows that symmetrized maps from different crystals are slightly more consistent with each other than they are with their unsymmetrized counterparts. This implies that there is some small systematic asymmetry present in the maps as measured. The most likely source for this asymmetry is a detector-face rotation: if the detector normal is rotated by an angle $\delta\psi$ with respect to the incident beam, then the solid-angle correction of section 2.2.3 can swing from $\cos^3(\psi - \delta\psi)$ to $\cos^3(\psi + \delta\psi)$, yielding a difference

$$\begin{aligned}\cos^3(\psi+\delta\psi) - \cos^3(\psi-\delta\psi) \approx&\quad 6\delta\psi\cos^2\psi\sin\psi \\ =&\quad 6\delta\psi(1-\frac{q^2}{2k^2})^2(1-\frac{q^2}{4k^2})^{\frac{1}{2}}\frac{q}{k} \\ \approx&\quad 6\delta\psi\frac{q}{k} \qquad (4.2)\end{aligned}$$

Thus, for a resonable detector-face rotation $\delta\psi$ of 0.035 radians (2 degrees), at $\frac{q}{k} = 0.25$ (about 4 $\mathbf{\AA}$ in our case) the difference is 0.05. If the background intensity is 2000 at $\frac{q}{k} = .25$, the difference in ADU is $2000\times 0.05 = 100$, implying an azimuthal variation of amplitude 50 ADU, which is roughly consistent with the observed difference.

This same method was used to detect whether or not there was a difference between the measured diffuse maps from crystalline *Staph.* nuclease both with and without Ca^{++} and pdTp bound. First, each map is symmetrized and its spherically-averaged component is subtracted. Then, one map is scaled to the other to minimize whatever differences there are. This sets a lower limit on the differences observed between the diffuse maps. As is detailed in chapter 5, significant differences were observed between the maps of diffuse scattering from these crystals.

4.5 Simulation of Diffuse Maps

In order to identify what kinds of disorder in the crystal were consistent with the observed diffuse maps, simulated diffuse maps were generated for comparison to the data. The first key ingredient in simulating diffuse maps is an acceptable atomic structural model of the protein, from which electron densities can be calculated. In the case of *Staph.* nuclease, a structural model was available from the Brookhaven Protein Data Bank (PDB).[4]

Using the PDB structure as a starting point, a refinement of the structure was performed using XPLOR by Axel Brünger[44]. XPLOR takes a PDB structure as an input, and can be directed to generate an electron-density map, calculate its Fourier transform, generate a list of predicted Bragg reflections, compare the predictions with measured Bragg reflections, and make adjustments to the structural model iteratively to minimize the difference between measured and predicted reflections. The measured reflections were those obtained from oscillation exposures during data collection (see section 2.1).

Once a good structural model has been obtained, three-dimensional simulations of maps of diffuse scattering can be generated. Using the XPLOR-refined structural model, maps of the Fourier-transformed electron density of arbitrary regions of the protein are generated and output, depending upon the requirements of the model. Once these maps are converted to the same format as the diffuse maps obtained experimentally, they can be transformed and manipulated in an identical fashion to

---

[4] The ID for the structure was 1STN, deposited February 17, 1993 by T. R. Hynes and R.O. Fox.

that used for diffuse maps. Mathematical operations indicated by a chosen model of diffuse scattering can be performed to turn the maps of Fourier transforms into simulated maps of diffuse intensity.

For example, consider the model where individual proteins execute independent, isotropic rigid-body translational motions in a monomolecular unit cell. As derived in section 3.5, the diffuse scattering is described by equation 3.59:

$$I_D(\mathbf{q}) = N(1 - e^{-q^2\sigma_u^2})|f(\mathbf{q})|^2, \tag{4.3}$$

where N is the number of unit cells in the crystal, σ_u is the Gaussian width of the distribution of molecular displacements, and $f(\mathbf{q})$ is the Fourier transform of the electron density of the molecule. Once the structural model for the molecule has been refined, the square of the Fourier transform $|f(\mathbf{q})|^2$ can be calculated using XPLOR, and the results placed on a lattice. Having $|f(\mathbf{q})|^2$, $I_D(\mathbf{q})/N$ can be calculated by choosing a value for σ_u and multiplying each voxel value in the map of $|f(\mathbf{q})|^2$ by $(1 - e^{-q^2\sigma_u^2})$.

For *Staph.* nuclease, the procedure involves an additional step to account for the four-molecule unit cell with space group $P4_1$. The screw $c-$axis becomes a four-fold c^*-axis in the reciprocal space map. Therefore, a single molecule in the unit cell is first chosen as a reference to generate a simulated map of diffuse scattering from a crystal where each protein molecule in the unit cell executes independent translations. Once a diffuse map has been calculated for the single molecule, it is symmetry-averaged to generate the proper diffuse map for the crystal. The Friedel symmetry of the simulated diffuse map is guaranteed already, but the symmetry-equivalent points due to the four-fold symmetry about the c^*-axis must be averaged by hand. This is done by replacing each voxel value at each point (x, y, z) with the average voxel values of the points $\{(x,y,z),(-y,x,z),(-x,-y,z),(y,-x,z)\}$. The Friedel symmetry guarantees that points $\{(-x,-y,-z),(y,-x,-z),(x,y,-z),(-y,x,-z)\}$ will be equivalent at the end of the four-fold averaging, but in practice these are also redundantly included in the symmetry-averaging procedure.

One does not yet expect these simulated maps to bear a great resemblance to the experimentally obtained diffuse scattering maps. This is because of the huge spherically-symmetric component in each of the maps. While in experimental maps, there are spherically-symmetric contributions coming from disordered solvent, air scattering and capillary scattering, in simulated maps the only contribution comes from the ordered protein in contrast to a flat solvent background.

In order to compare the two maps, therefore, they are transformed such that they both represent scattering from ordered structures in the crystal. One accomplishes this by subtracting the spherically-averaged component of the scattering from both lattices. This removes the spherically-averaged component of the *protein* scattering from both lattices, but it also removes any intensity due to disordered solvent, air and capillary scattering which may be present in the experimentally obtained map, leaving comparable maps.

One possible problem in this approach would come from significantly different absorption of x-rays scattered into different angles, which would introduce asymmetries in the spherically-symmetric component of the scattering. The specific pattern of absorption would depend upon the shape and orientation of the crystal, but generally the differential absorption would cause asymmetries on long length scales in reciprocal space. We can estimate the effect on *Staph.* nuclease by considering the crystal to be a 400 μm thick cylinder oriented perpendicular to the beam, which will approximate the case for the largest crystal studied in this thesis, giving the maximum effect. The largest asymmetry will occur for x-rays scattered on the upstream

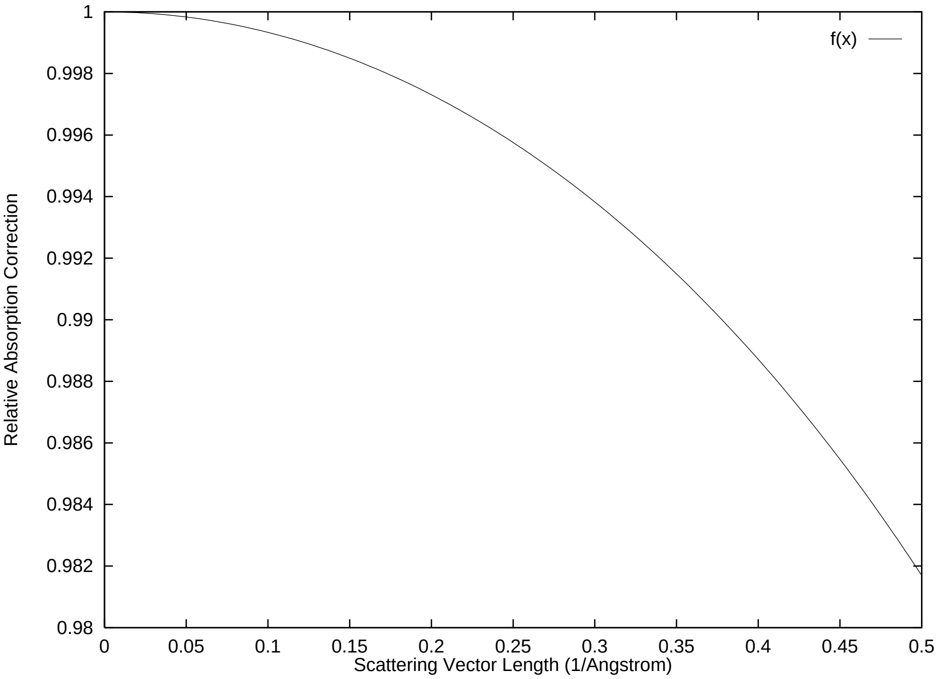

Figure 4.4: *Effect of differential absorption on diffuse maps.* This plot shows the worst possible ratio of absorption corrections for two pixels in an image *versus* scattering vector length, as calculated for the 400 μm cylindrical crystal with an underestimated x-ray absorption length of 2.5 mm described in the text. Absorption corrections are ignored on the basis of this plot.

surface of the cylinder: the ratio R of absorption by x-rays scattered into an angle ψ in the equatorial plane versus forward-scattered x-rays is given by

$$R = e^{-\frac{t}{\tau}\left(\frac{1}{\cos\psi}-1\right)}, \tag{4.4}$$

where t is the thickness of the crystal and τ is the absorption length of the x-rays in the crystal. In terms of the scattering vector length s, this becomes

$$R = e^{-\frac{t}{\tau}\left(\frac{1}{\cos(2\sin^{-1}(0.5s\lambda))}-1\right)}, \tag{4.5}$$

which is plotted in figure 4.4 for an underestimated absorption length of 2.5 mm. This absorption length is lower than what it would be if the crystal were composed solely of oxygen, which at a wavelength of 0.92 **Å** has the highest cross-section of oxygen, nitrogen and carbon.

At 0.26 $\textbf{Å}^{-1}$, there is less than a 0.6% asymmetric perturbation caused by differential absorption. Even at 1 **Å** resolution, the effect is below the 2% level. Overestimating the spherically-symmetric component of the scattering at 2000 ADU, there would be only a 10 ADU asymmetry in the data caused by differential absorption at 0.26 $\textbf{Å}^{-1}$, which is much less than the greater-than 100 ADU RMS intensity of the diffuse scattering. By comparison, the single-pixel photon noise at 2000 ADU is roughly $\sqrt{2000} = 45$ ADU. Based on these calculations, absorption effects were ignored.

An additional benefit of subtracting the spherically-averaged component from calculated and measured maps is that it allows them to be Fourier transformed without edge effects, since the subtracted maps fall to zero at the edges. This is particularly useful in smoothing the maps by convolution with a three-dimensional Gaussian, which is used to compare maps at different levels of detail. The relationship

$$A(\mathbf{q}) * B(\mathbf{q}) = \mathrm{FT}^{-1}\left(\mathrm{FT}\left(A(\mathbf{q})\right) \times \mathrm{FT}\left(B(\mathbf{q})\right)\right) \tag{4.6}$$

is used to speed up the convolution, which would be prohibitively slow otherwise. This became extremely important in modelling the diffuse scattering, since no agreement was observed between a simulated and experimental map at "high detail," with no smoothing. In order to observe any agreement between a simulation and an experimental map, it was necessary to smooth both the simulated and experimental maps at "medium detail," using a Gaussian whose half-width was equal to one reciprocal unit-cell diagonal. Even better agreement was seen by smoothing at "low detail," with a Gaussian of half-width equal to two reciprocal unit-cell diagonals.

Once a simulated map has been generated, then, the procedure for comparing it to the data is outlined as follows:

- Calculate the spherically-averaged component of both the simulated and experimental maps by averaging, and subtract that component from the original maps.
- Choose a level of detail at which to compare the maps. Smooth both maps by convolution with the same Gaussian map, choosing the width of the Gaussian as a measure of the detail level.
- In an identical manner to the way that reproducibility measurements were described in section 4.4, find the best scale factor relating the two maps. Subtract the scaled maps and plot the RMS difference intensity vs. resolution, comparing it to the RMS amplitudes of both the simulated and experimental diffuse maps.

An observation that the difference map is smaller in amplitude than the scaled simulated and experimental maps is interpreted as evidence that there is some similarity between the two. By observing the effect of changing the scale factor on the RMS amplitude of the difference map, it can be determined whether or not the best-fit scale factor brings the difference to a local minimum, providing further evidence for similarity between the maps. By comparing the relative amplitudes of the difference maps generated by using simulated maps based upon various different models, a choice can be made as to which model best agrees with what is observed experimentally.

In practice, one can predict whether or not two maps will agree with one another by two other means. One is to simply visualize the maps using the means described in the beginning of this chapter. If one does not observe broadly similar features in the simulated and experimental maps, it is a good sign that the fit will be poor.

Another check to see that things are OK is whether or not the scale factors agree sufficiently well when one constrains the fit over different spherical shells in reciprocal space. The written routines allow one to specify an inner and outer radius in reciprocal space to confine the region over which a best scale factor is to be found. If one observes significantly different scale factors between resolution shells, or if one observes negative numbers in the scale factor, these may be signs that the lattices are not sufficiently similar to be able to quantify their agreement.

In this thesis, comparison of simulated diffuse maps with experimental data are used to place constraints on the possible nature of disorder in *Staph.* nuclease crystals. The method used was one of simply comparing the various $|F(\mathbf{q})|^2$ maps output by XPLOR with experimentally obtained diffuse maps. For models where individual subdomains of the unit cell move independently, diffuse maps resemble a superposition of $|F(\mathbf{q})|^2$ calculated for each independently moving unit. For the liquid-like motion model of Clarage *et. al.*[22], and for the acoustic modes model discussed by Glover *et. al.*[21], the diffuse maps most resemble $|F(\mathbf{q})|^2$ calculated from the electron density of the entire unit cell. [5]

There are a huge number of individual subdomains of *Staph.* nuclease which can be considered for such a comparison. If one allows multiple domains to move, the combinations quickly become too many to compare reasonably by brute-force trial-and-error. One needs independent information to motivate testing whether a particular model successfully reproduces the experimentally observed diffuse map.

The structural components of crystalline *Staph.* nuclease chosen for this study were

- *The entire unit cell with four proteins.* This was chosen to test agreement with liquid-like motion and acoustical modes models.
- *The entire protein.* This was chosen to test agreement with rigid-body motion models.
- *The residue 53-70 alpha-helical domain.* This was chosen by observing a backbone rendering of the structural model.
- *The residue 92-113 beta-barrel domain.* This was chosen by observing a backbone rendering of the structural model.
- *The residue 6-15 high-B-factor tail region.* This was chosen by reported high B-factors in the literature, and by observing a backbone rendering of the structural model.
- *The residue 42-53 high-B-factor flexible loop region.* This was chosen by reported high B-factors in the literature, and by observing a backbone rendering of the structural model.

---

[5] See chapter 3 for a discussion of models of diffuse scattering.

A comparison of simulated maps of $|F(\mathbf{q})|^2$ with the data yielded a quantifiable agreement using only one of the maps: that generated from the entire unit cell. None of the maps agreed with the data at the highest level of detail. The unit-cell simulated map alone, however, showed a measurable agreement when smoothed both to the lowest level of detail and to an intermediate level of detail. These results are presented in greater detail in chapter 5, and were interpreted as indicating that our measured maps best agree with the predictions of the liquid-like motions model of Clarage *et. al.*[22], and the crystalline normal modes model described in Glover *et. al.*[21].

Chapter 5

Results

Experimental maps of diffuse scattering from *Staph.* nuclease crystals were obtained and analyzed as described in chapters 2 and 4. The degree to which these maps showed the internal symmetry predicted by the $P4_1$ space group of the unit cell was determined: at 0.26 $\mathbf{Å}^{-1}$, the most symmetric map differed from its symmetry-averaged counterpart by roughly 25%, while the least symmetric differed by roughly 40%. The asymmetries are attributed to systematic measurement errors which are primarily due to a rotation of the detector plane with respect to the incident beam.

The reproducibility of measurements of diffuse scattering was determined by measuring the difference between two symmetry-averaged diffuse maps obtained from nominally identical crystals. The difference between the maps was as low as 20% at 0.26 $\mathbf{Å}^{-1}$, which sets a limit on the sensitivity to changes in diffuse scattering from *Staph.* nuclease which can be observed using the methods outlined in this thesis. This difference is smaller than the measured deviation from the $P4_1$ symmetry for maps obtained from single crystals, which was at best 25% at 0.26 $\mathbf{Å}^{-1}$.

In order to better understand the source of errors, differences were also measured between diffuse maps which were smoothed to remove high-frequency components of the features. Below 0.3 $\mathbf{Å}^{-1}$, diffuse maps grew more reproducible at lower detail levels. Above 0.3 $\mathbf{Å}^{-1}$, however, maps became *less* reproducible at lower levels of detail, perhaps due to systematic errors in measuring diffuse intensity, which tend to be exaggerated at high resolution.

Changes in diffuse scattering were observed by measuring differences in diffuse maps obtained from crystals of *Staph.* nuclease both with and without Ca^{++} and the substrate analog pdTp (thymidine-3′,5′-diphosphate) bound. In order to better understand the source of the changes, differences were also measured between diffuse maps which had been smoothed as in the reproducibility studies. Significant changes were observed at all detail levels, with greater differences observed at low detail than at high detail. The primary difference observed is an enhancement of diffuse scattering along the c^*-axis upon the binding of Ca^{++} and pdTp. These changes are interpreted as reflecting a change in the dynamics of crystalline *Staph.* nuclease upon the binding of Ca^{++} and pdTp.

Simulated maps of $|f(\mathbf{q})|^2$ were generated from the structural domains listed at the end of section 4.5. Only the map generated from the entire four-protein unit cell showed a significant correlation with an experimentally obtained map of diffuse scattering from *Staph.* nuclease. Of the models considered in chapter 3, the liquid-like motions model of Caspar *et. al.*[19] and the crystalline modes model of Glover *et. al.*[21] are therefore the candidates for describing the nature of disorder giving rise to diffuse scattering from *Staph.* nuclease. Models based on substitution disorder, and rigid-body motions of the entire unit cell, however, are not ruled out as being consistent with the experimental maps.

Table 5.1: Description of all crystals studied.

Crystal #	Ca^{++}/pdTp?	Size (mm^3)	Unit Cell ($\mathbf{Å}^3$)	Total Rotation Range
1	YES	2.0×0.4×0.4	48.5×48.5×63.4	94°
2	YES	2.2×0.4×0.4	48.3×48.3×63.4	96°
3	NO	0.4×0.2×0.2	48.2×48.2×63.9	90°

5.1 *Staphylococcal* Nuclease Diffuse Maps

Diffuse maps were obtained from three crystals of *Staph.* nuclease. The first, crystal 1, contained Ca^{++} and pdTp. Crystal 2, used to determine the reproducibility of diffuse scattering measurements, also contained Ca^{++} and pdTp. Crystal 3 contained neither Ca^{++} nor pdTp, and was used to measure differences in diffuse scattering upon changing the crystalline protein. Table 5.1 describes some features of interest in the data sets.

In order to generate maps of diffuse scattering from the three *Staph.* nuclease crystals, diffraction images within each data set were first scaled as described in section 2.2.5. For crystal 1 (see figure 5.1) and crystal 2 (see figure 5.2), which were exposed using the F1 station at CHESS, images needed to be scaled by more than a factor of two in some cases, while for crystal 3 (see figure 5.3), exposed at the A1 station at CHESS, the scale factor is remarkably consistent throughout the experiment. Rumors of a less stable beam at the F1 station are consistent with this observation.

Radiation damage prevented the collection of a full 180° data set for each of the crystals. As is shown in table 5.1, the measured range of reciprocal space for each crystal was about 90°, which is the minimum required to generate a complete data set using the symmetry predicted by the $P4_1$ space group of the unit cell. Due to the limited rotation range in each experiment, different regions of reciprocal space were

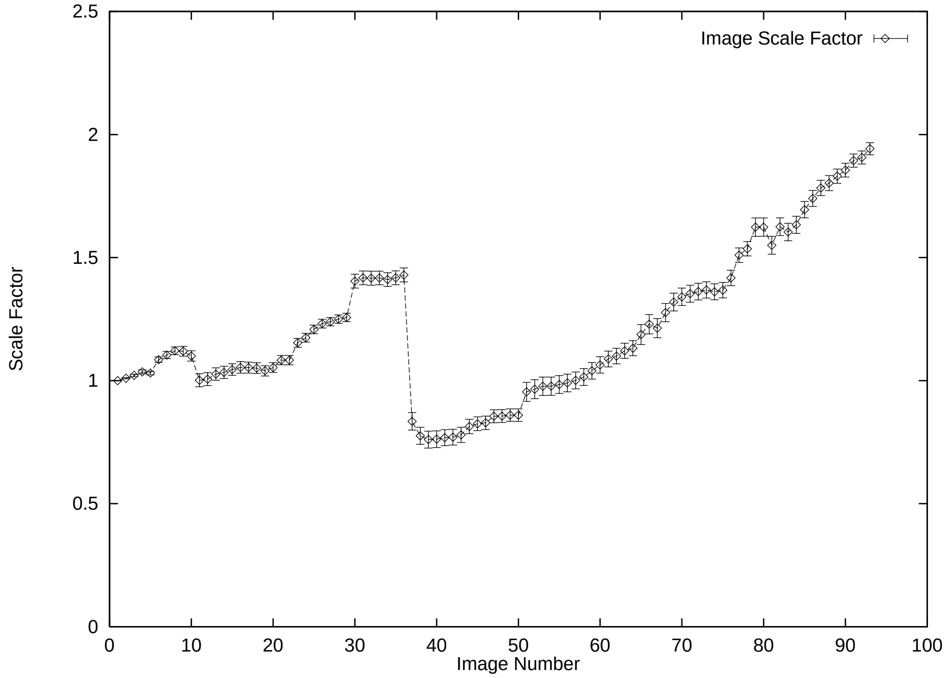

Figure 5.1: *Crystal 1 image scale factors.* The scale factor indicates by how much to multiply each pixel in the corresponding image in order to best agree with image number one. The sudden dip between images thirty-six and thirty-seven occurs between "fills" at CHESS, when the x-ray flux at F1-station is interrupted to refill the ring with positrons and electrons for high-energy physics experiments at CLEO. Error bars indicate the degree to which the scale factor is to be trusted, so that much of the fine structure in the plot is real, and is most likely due to fluctuations in the beam intensity.

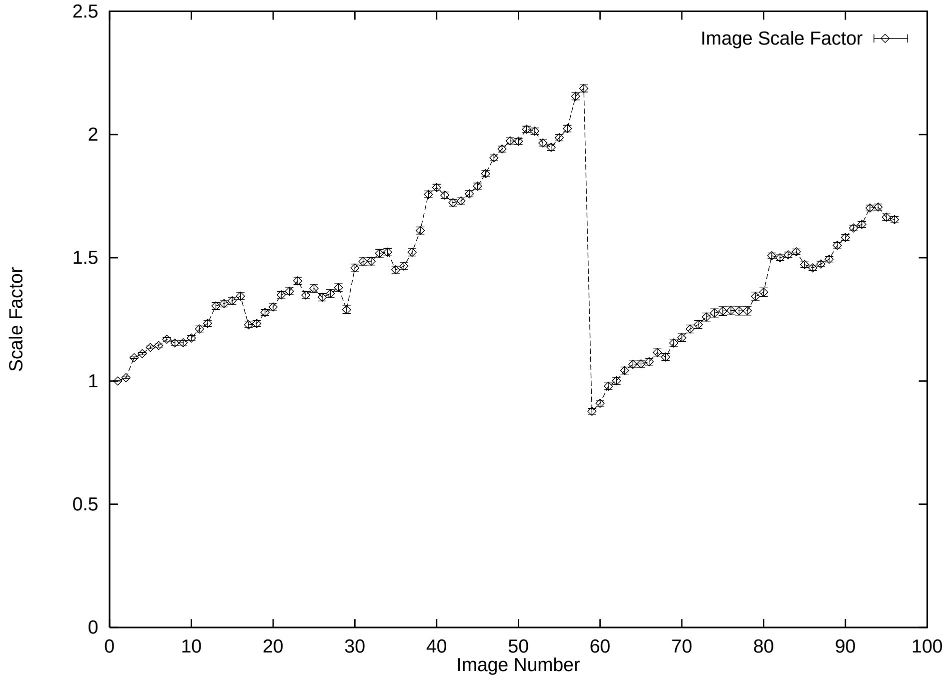

Figure 5.2: *Crystal 2 image scale factors.* See figure 5.1 for a description of the meaning of the scale factor. The sudden dip in the scale factor between images fifty-eight and fifty-nine is the signature of a new fill.

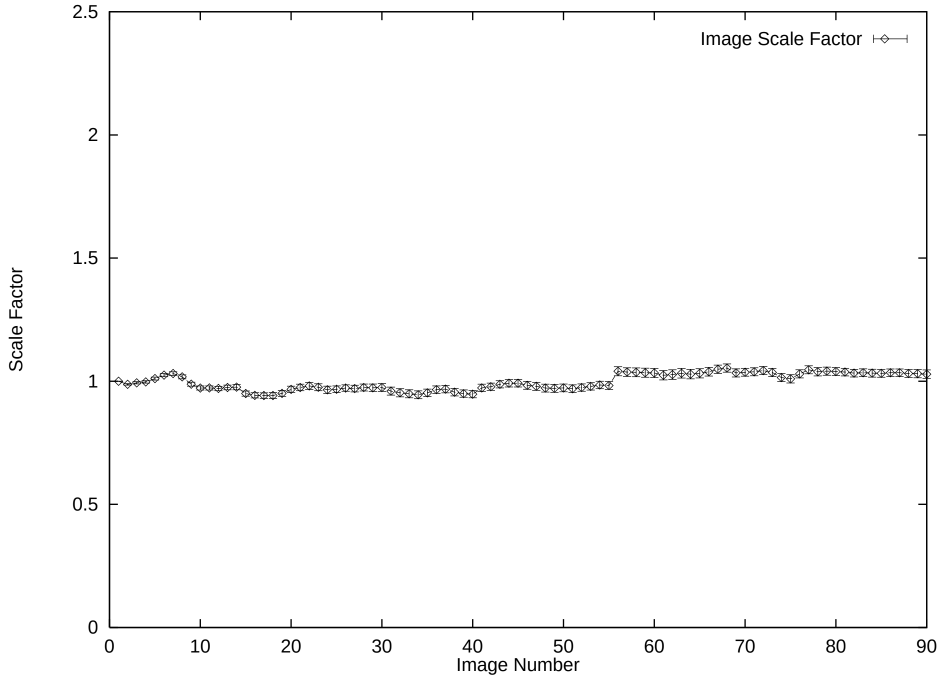

Figure 5.3: *Crystal 3 image scale factors.* See figure 5.1 for a description of the meaning of the scale factor. The superior stability of the beam at A1-station is indicated by the remarkably consistent scale factor for this data set.

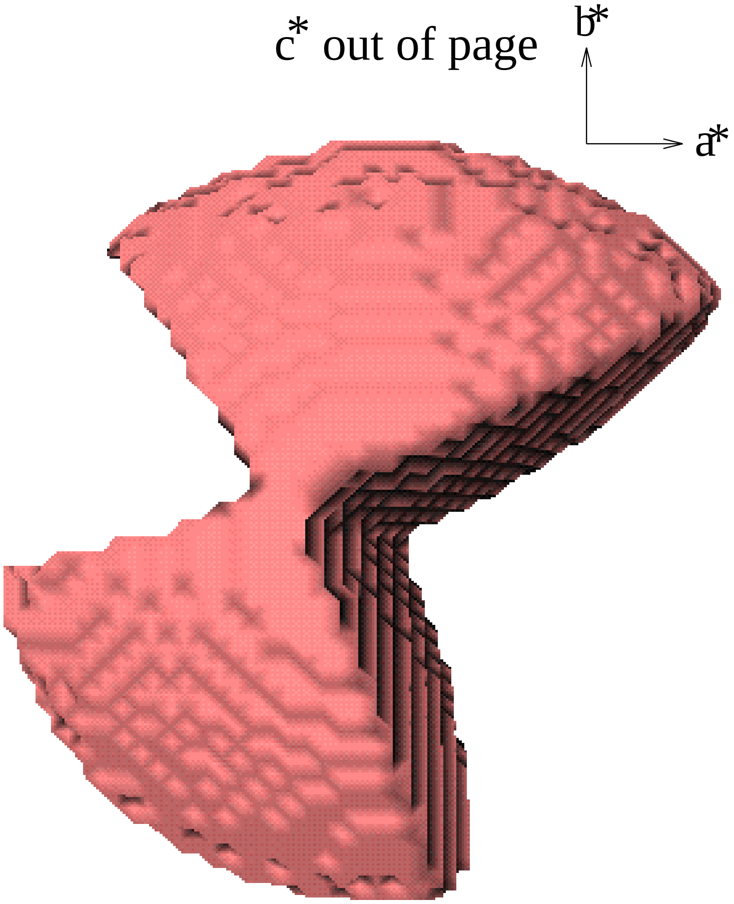

Figure 5.4: *Reciprocal-space volume sampled by experiments on crystal 1.* The rendered surface bounds the sampled volume of reciprocal space for crystal 1. The four-fold c^*-axis points out of the page, while a^* points to the right in the page, and $b*$ points up in the page. Symmetery averaging about the four-fold will complete the data set. The outer radius corresponds to a cutoff of 0.6 $\mathbf{\AA}^{-1}$ which is enforced when the map is created.

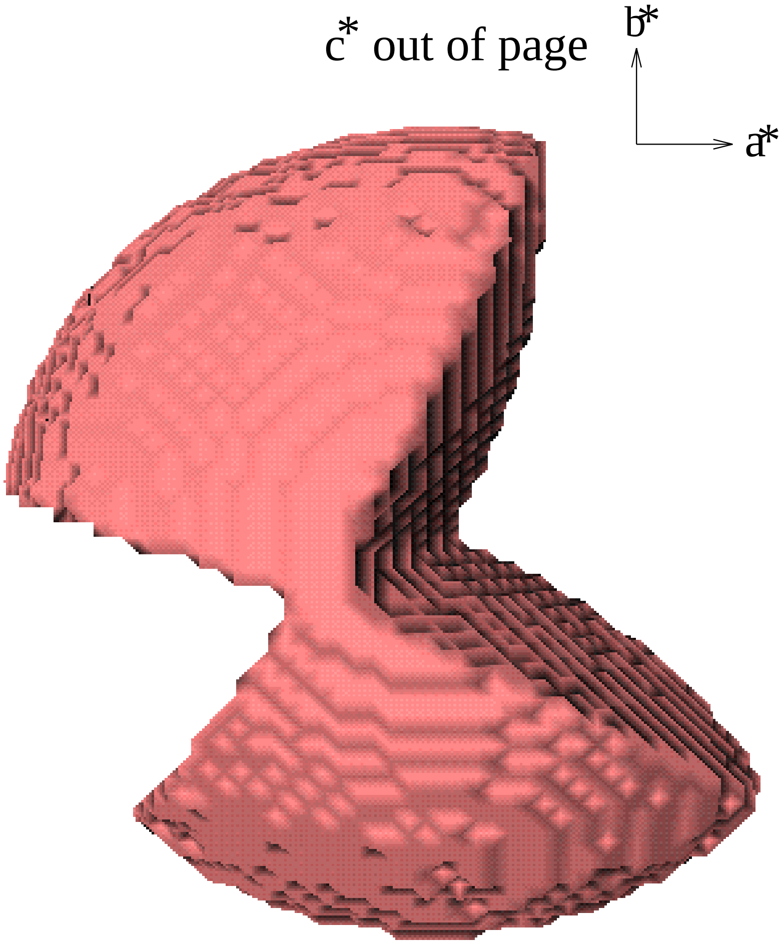

Figure 5.5: *Reciprocal-space volume sampled by experiments on crystal 2.* The rendered surface bounds the sampled volume of reciprocal space for crystal 2 (see figure 5.4 for an explanation). This data set is rotated by about 45 degrees about the four-fold with respect to the data set from crystal 1.

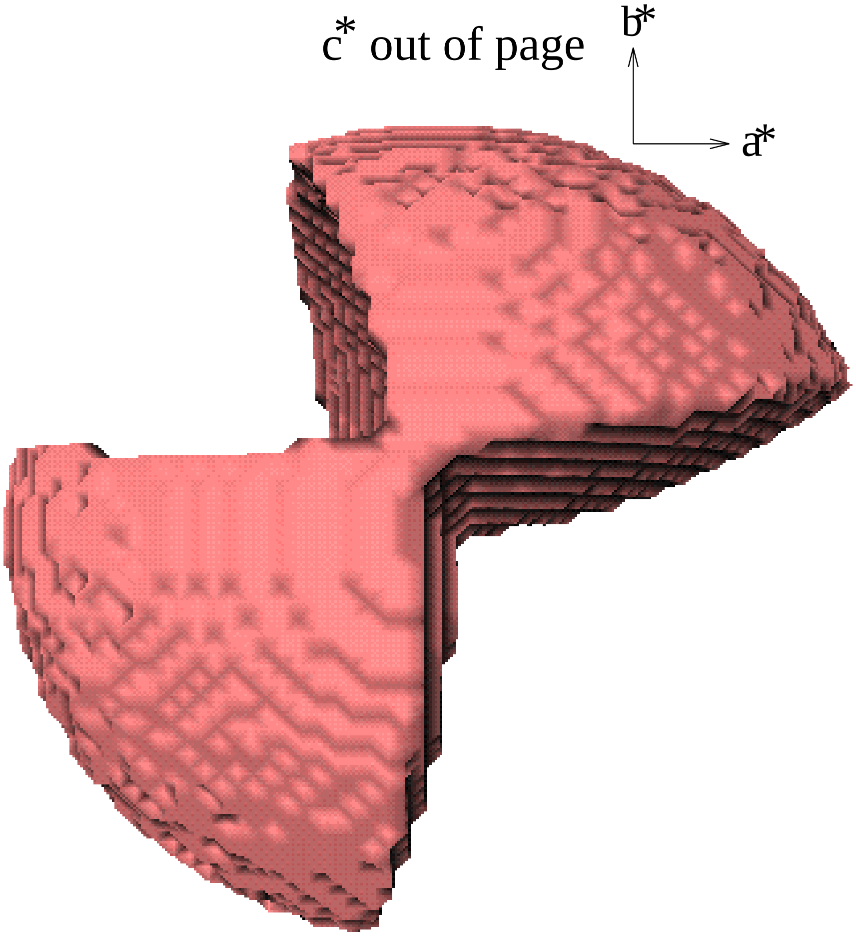

Figure 5.6: *Reciprocal-space volume sampled by experiments on crystal 3.* The rendered surface bounds the sampled volume of reciprocal space for crystal 3 (see figure 5.4 for an explanation).

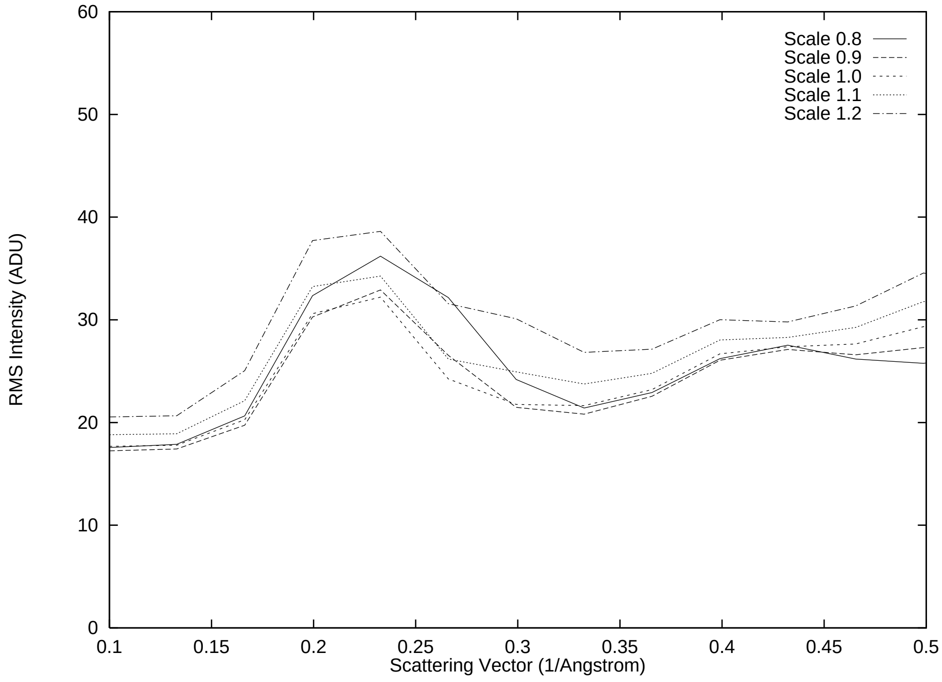

Figure 5.7: *Verification of scaling between diffuse maps from crystal 1 and crystal 2.* In order for a scale factor between two lattices to be meaningful, the difference profile must be minimized with respect to changes in the scale factor. After multiplying the crystal 2 map by a trial scale factor, difference profiles are generated between the crystal 1 map and several crystal 2 maps which are scaled by slightly more or less than the trial scale factor. The *scale 1.0* plot corresponds to the difference profile calculated using the trial scale factor, while the *scale 1.1* plot, for example, is calculated using a scale factor 1.1 times as large. If the *scale 1.0* plot shows the smallest difference over a given range in scattering vector, the fit is verified over that range. In this figure, the scale factor is verified below 0.3 $\mathbf{Å}^{-1}$.

probed by each of the experiments. Figures 5.4, 5.5 and 5.6 show the regions where all of the experimentally measured diffuse intensity lies in diffuse maps obtained from each of the crystals. Symmetry averaging completes all of the data sets, since there is a four-fold axis pointing out of the page.

As described in section 4.4, the diffuse map obtained from crystal 1 was used as a reference for scaling the maps from crystal 2 and crystal 3. Perturbations on the scale factor were made to ensure that the difference was locally minimized. Difference

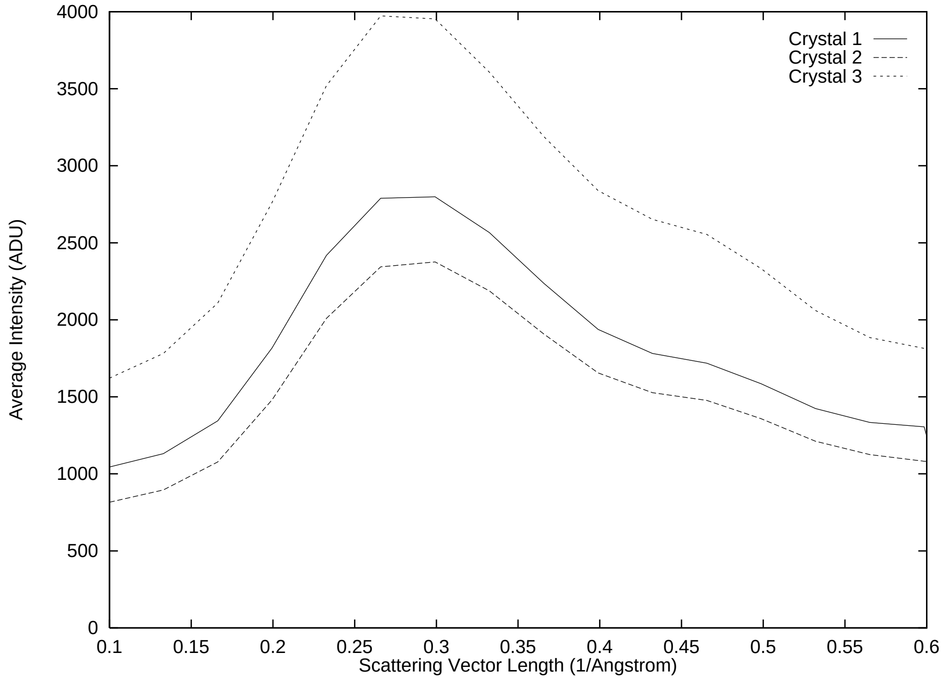

Figure 5.8: *Spherically-averaged scattering from* Staph. *nuclease crystals.* The average intensity profile for each crystal has a similar shape, but its scale is determined by things such as the beam intensity and crystal size. See the text for a definition of the average intensity profile.

maps were calculated using scale factors in the neighborhood of the best-fit value, and RMS intensity profiles were calculated from the difference maps. The profiles were simultaneously plotted for inspection to ensure that the lowest profile corresponded to the best-fit scale factor. If the best-fit scale factor did not produce the lowest profile, the scale factor was adjusted in order to make it produce the lowest profile. If no scale factor was found which could produce a local minimum, the scale factor was interpreted as meaningless, and the maps were considered too different to have their difference quantified in this way. Figure 5.7 shows the results of a successful scaling between the diffuse maps obtained from crystal 1 and crystal 2.

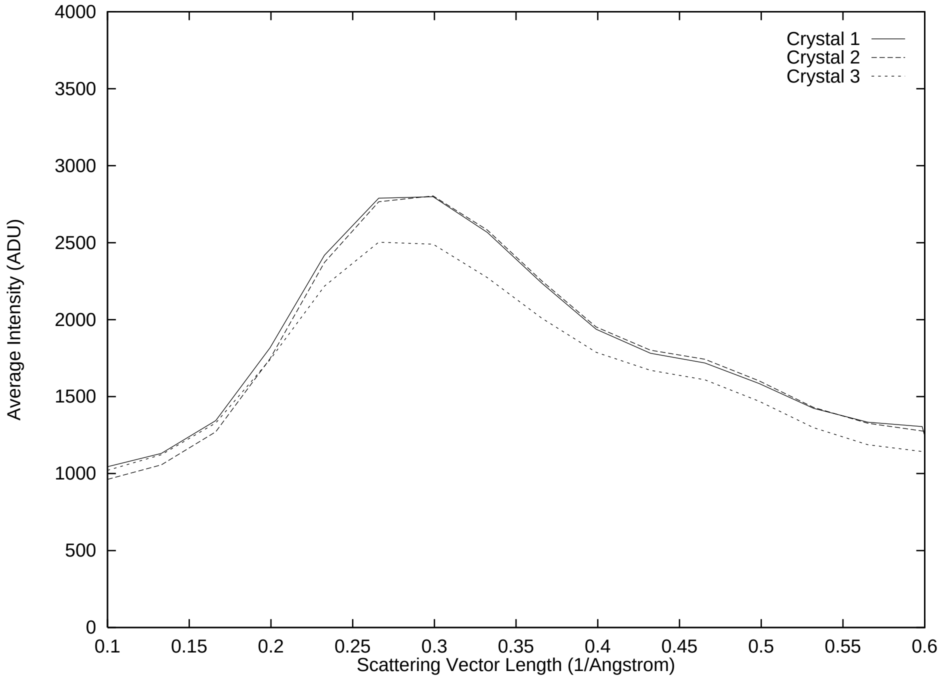

Figure 5.9: *Spherically-averaged scattering evaluated from scaled diffuse maps.* The same profiles as those in figure 5.8, but evaluated after the lattices have been scaled to minimize the voxel-by-voxel differences.

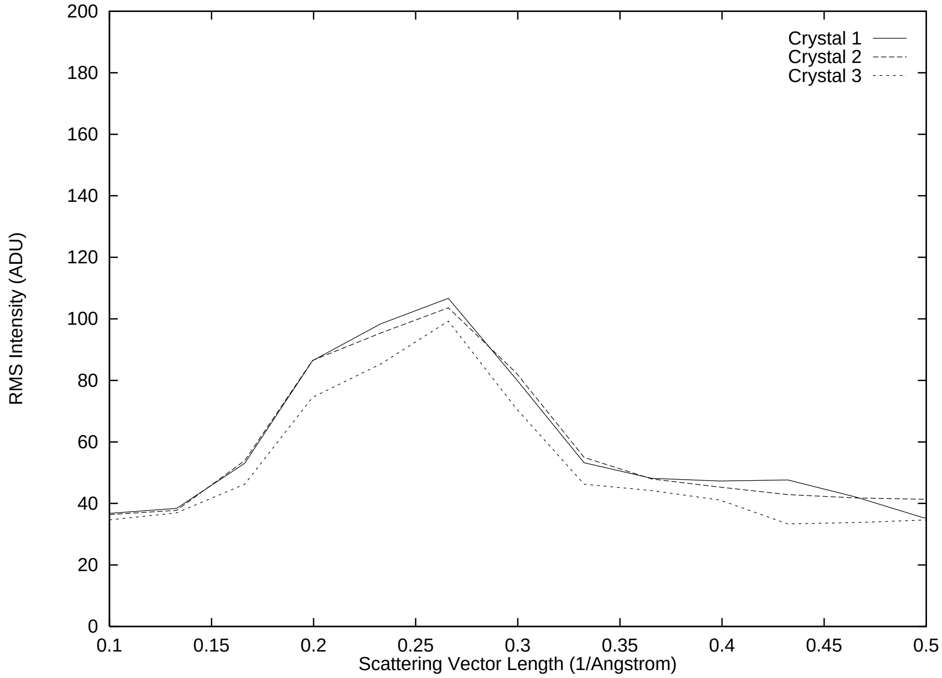

Figure 5.10: *RMS intensity profiles from diffuse maps.* These plots are generated by first evaluating the spherically-averaged intensity profile of each scaled map, as displayed in figure 5.9. Each voxel value in the map then has subtracted from it the spherical average, determined by interpolation of the average intensity profile at the voxel's scattering vector. The RMS intensity of the resulting map is then calculated in spherical shells, and displayed in a way analogous to figure 5.9. All of the profiles peak at 0.26 $\text{\AA}^{-1}$, and are of similar amplitude after scaling.

The average intensity profiles[1] from the scaled maps revealed the similarity of the crystal 1 and crystal 2 maps, and a difference between these maps and the crystal 3 map. Figure 5.8 shows the profiles from all of the experimental diffuse maps before scaling. After scaling, the profiles are as shown in figure 5.9. The profiles from the scaled crystal 1 and crystal 2 maps overlap almost perfectly, while the crystal 3 profile lies consistently lower than the other two, with the greatest difference above about 0.23 $\text{\AA}^{-1}$.

---

[1] See section 4.3.

RMS intensity profiles[2] were obtained from scaled maps with the spherically-averaged intensity subtracted, and showed good agreement for all crystals as well, with the greatest differences noted at high resolution. While it is not surprising that the profiles from crystal 1 and crystal 2 agree so well, it is somewhat surprising that they both agree with the profile from crystal 3, given the marked differences observed in the average intensity profiles. As is shown below, a comparison of the full three-dimensional maps also shows significant differences between crystal 3, which is grown without a substrate analog bound, and the other crystals, which contain the substrate analog pdTp.

The first indication of the quality of the three-dimensional data came from viewing the diffuse maps using EXPLORER. Clear features were observed which were reproduced when maps from nominally identical crystals were viewed. Figure 5.11 and figure 5.12 show views of the maps studied in this way, and provide a comparative view to confirm the reproducibility of diffuse features. Both views were displayed by selecting a voxel value of 2850 at which to draw an interpolated isosurface of diffuse intensity.

Figure 5.13 shows a view of the diffuse map from the third crystal, which differs in that there is no Ca^{++} and pdTp bound to the protein. Similar features are noticeable in this map, but the similarity is not as striking as that between the crystals grown with Ca^{++} and pdTp. Unfortunately, there is no data in the region defined by the overlap between the crystal 1 and crystal 2 data sets, making comparison before symmetry-averaging difficult.

These views also suggest that the diffuse scattering obeys the symmetry predicted by the $P4_1$ space group of the unit cell. In all maps, there is a visible mirror symmetry

---

[2] See section 4.3.

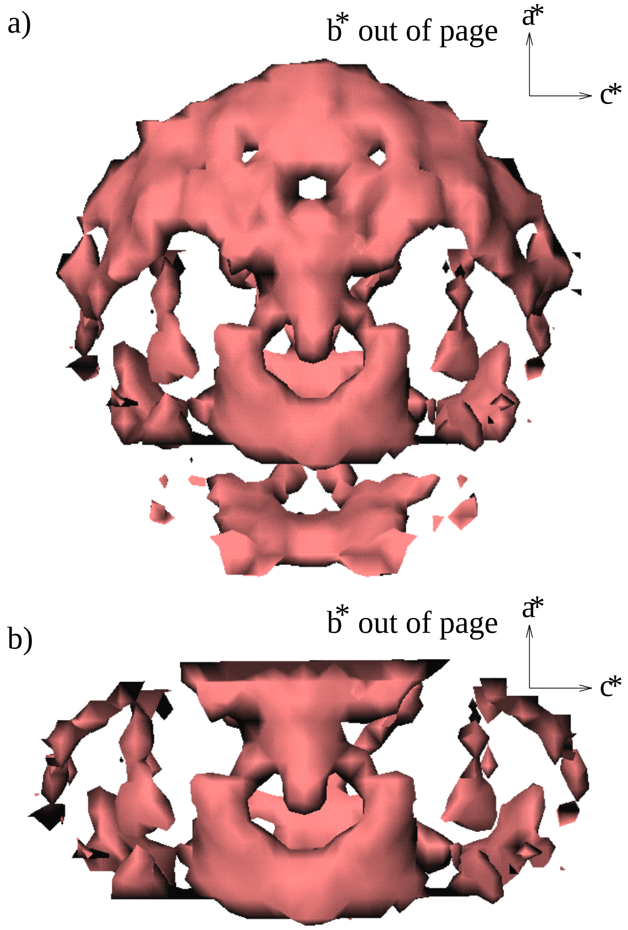

Figure 5.11: a) *An interpolated isosurface in the 3D diffuse map of* Staph. *nuclease crystal 1.* Features are easily distinguished above noise. The visible mirror symmetry across the a^*b^*-plane is consistent with the $P4_1$ space group of the unit cell. Reciprocal axes a^*, b^* and c^* are drawn to orient the map. Two opposing wedges of the map are unfilled, since the data only spans 96 degrees of spindle rotation. b) *The same view restricted to the overlapping volume of reciprocal space sampled for both crystals.* Compare with figure 5.12 to observe the reproducibility of the measurement.

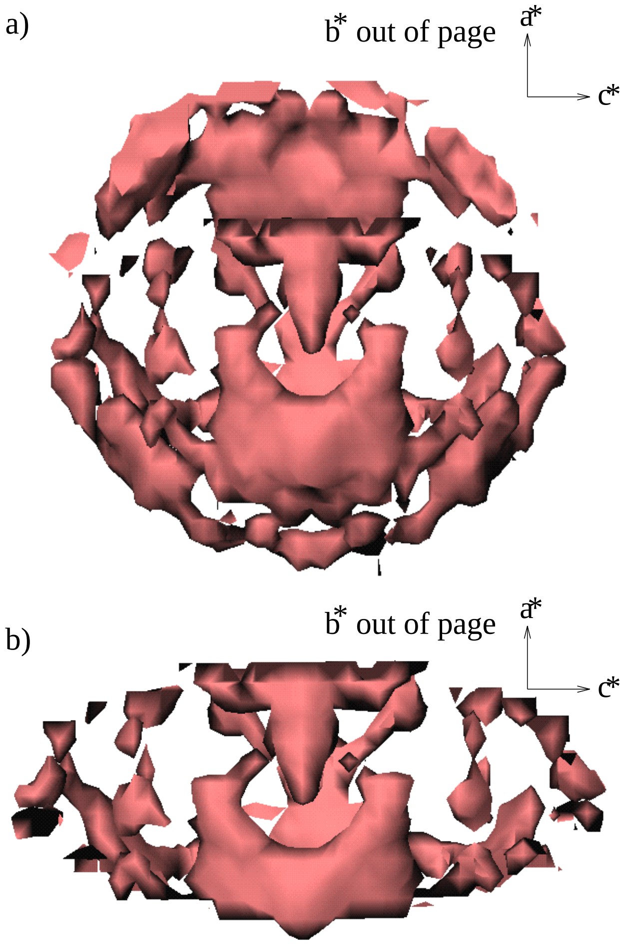

Figure 5.12: a) *An interpolated isosurface in the 3D diffuse map of* Staph. *nuclease crystal 2.* The data spans a 94 degree region in reciprocal space rotated roughly 45 degrees about the c^*-axis with respect to that of the first crystal. The combination of views suggests that c^* is a four-fold rotational symmetry axis, which is consistent with the P4_1 space group of the unit cell. Reciprocal axes a^*, b^* and c^* are drawn to orient the map. b) *The same view restricted to the overlapping volume of reciprocal space sampled for both crystals.* Comparison with figure 5.11 shows the reproducibility of the measurement.

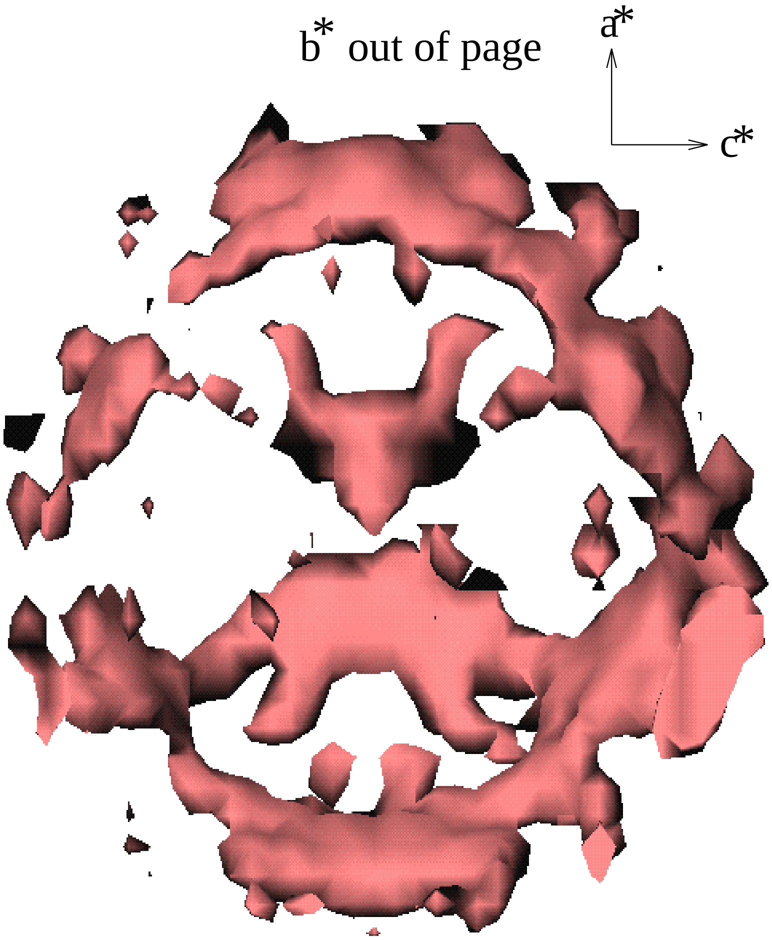

Figure 5.13: *An interpolated isosurface in the 3D diffuse map of* Staph. *nuclease crystal 3.* The scaled diffuse map from crystal 3 is here displayed in the same orientation as that from crystals 2 and 1, using the same method as in figure 5.11. There are recognizable features here which are similar to the features in maps of diffuse scattering from the other crystals, but the similarity is not as striking as that between crystals 1 and 2.

in features across the a^*b^*-plane. In addition, the combination of views from crystal 1 and crystal 2 suggest a four-fold c^*-axis. Both of these symmetries are predicted by the $P4_1$ space group.

In order to both characterize the degree of internal symmetry and reproducibility of diffuse maps, symmetry-averaged maps were generated from the original data, as described in section 4.3. These symmetry-averaged maps were also used in modelling diffuse scattering by simulations, since all of the models considered generate maps which bear the symmetry predicted by $P4_1$. Once again, by viewing the maps using EXPLORER, similarities were observed among all three crystals. Figures 5.14,5.15, and 5.16 show views which illustrate this observation. Figure 5.17 displays stereo pairs of perspective views of each of the maps, in order to better show the three-dimensionality of the features. Even among symmetrized maps, the scattering from crystal 3 appears to be significantly different from that from the other two crystals. Section 5.2 describes this difference quantitatively.

As described in section 4.3, the degree of internal symmetry in diffuse maps was characterized by measuring the difference between a map and its symmetry-averaged counterpart. Profiles of the RMS difference intensity for this measurement are displayed figure 5.18, along with the RMS intensity profile for crystal 1, as shown in figure 5.10. By comparing the difference profiles to the RMS intensity profile, one can measure the degree of internal symmetry in the diffuse maps. For example, at 0.26 $\mathbf{\AA}^{-1}$, the difference profile for crystal 2 reads roughly 25 ADU, while the crystal 1 RMS profile reads roughly 110 ADU, so that the difference is 25% of the “signal”. In other words, the crystal 2 diffuse map differs from its symmetry-averaged counterpart by only 25% at 0.26 $\mathbf{\AA}^{-1}$, which gives a measure of the internal symmetry of the map.

A good way of visually comparing maps is by using the shell images described in section 4.2. Images of the same shell from two different crystals can be placed side-

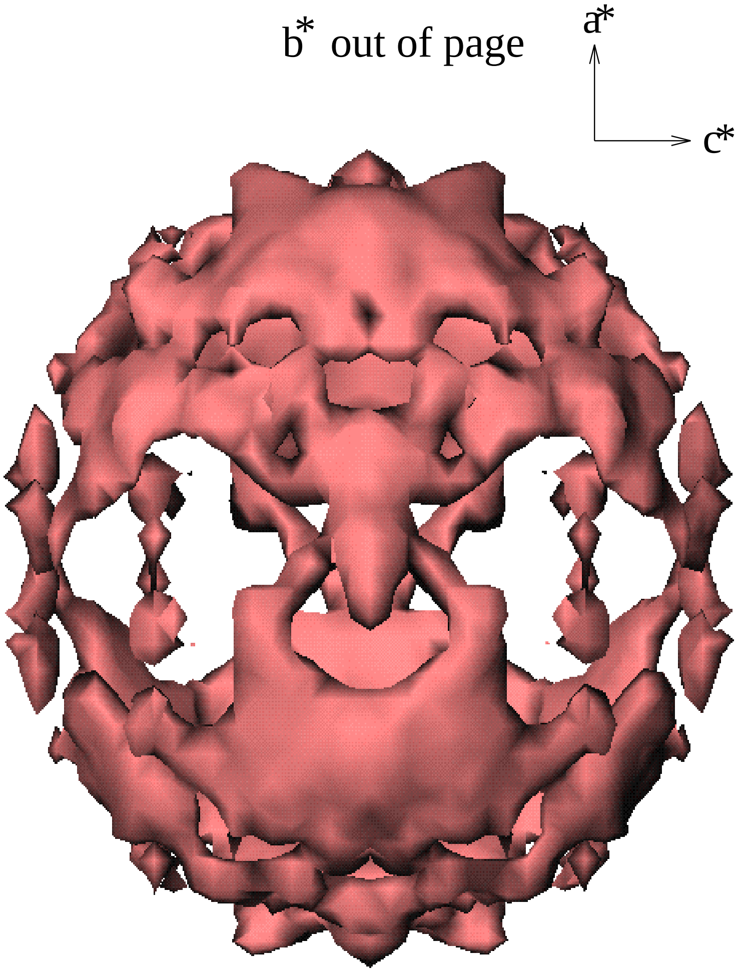

Figure 5.14: *A view of the symmetry-averaged diffuse map from crystal 1.* The symmetry-averaged diffuse map is displayed in a manner identical to that described in figure 5.11, including the voxel-value threshold of 2850. Symmetry averaging is performed as described in section 4.3.

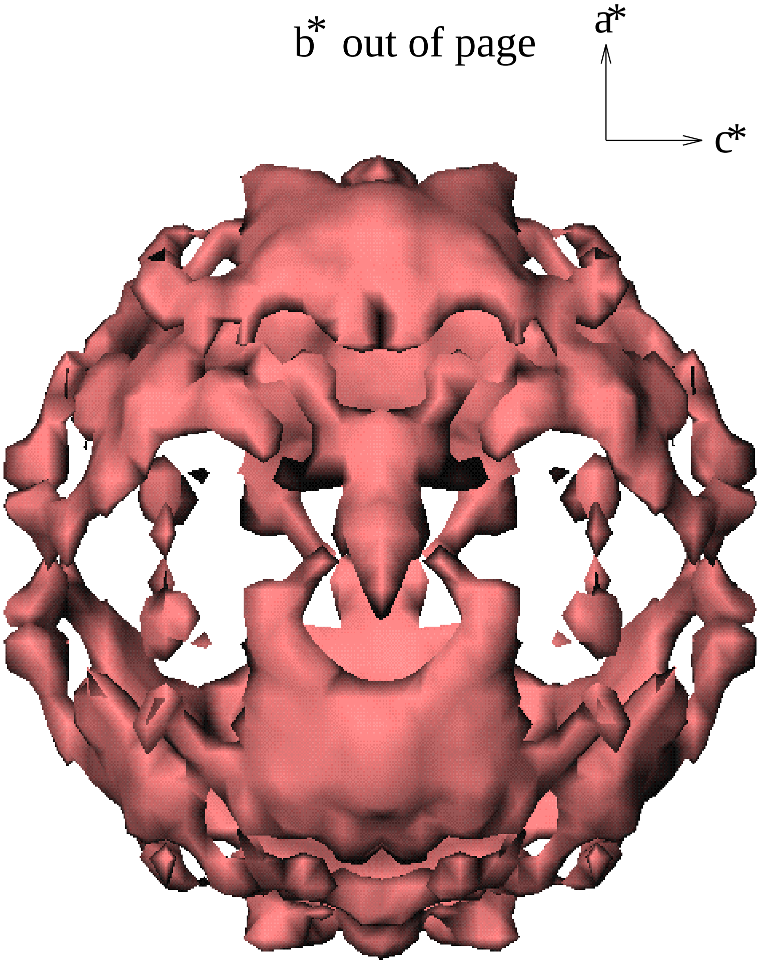

Figure 5.15: *A view of the symmetry-averaged diffuse map from crystal 2.* The symmetry-averaged diffuse map is displayed in a manner identical to that described in figure 5.11, including the voxel-value threshold of 2850. Note the great similarity between the features observable in this view and those observable in figure 5.14, which indicates good reproducibility of diffuse maps.

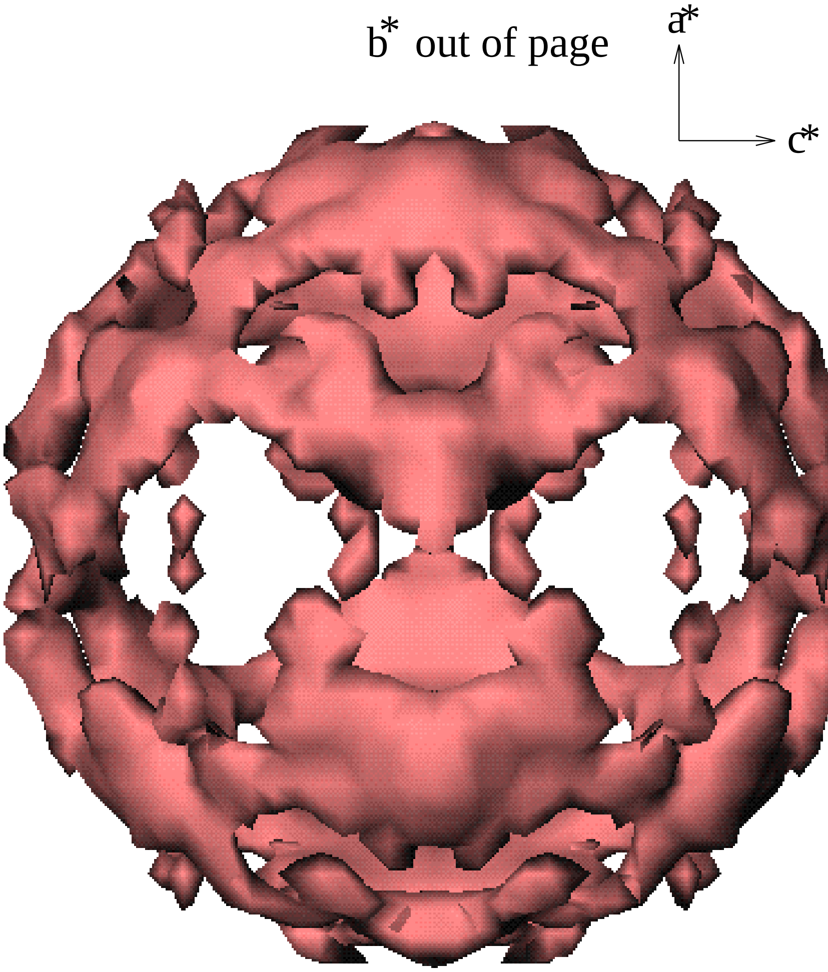

Figure 5.16: *A view of the symmetry-averaged diffuse map from crystal 3.* The symmetry-averaged diffuse map is displayed in a manner identical to that described in figure 5.11, except a voxel-value threshold of 2550 is used instead of 2850. Similarities are visible between this view and those in figures 5.14 and 5.15, but there are clear observable differences.

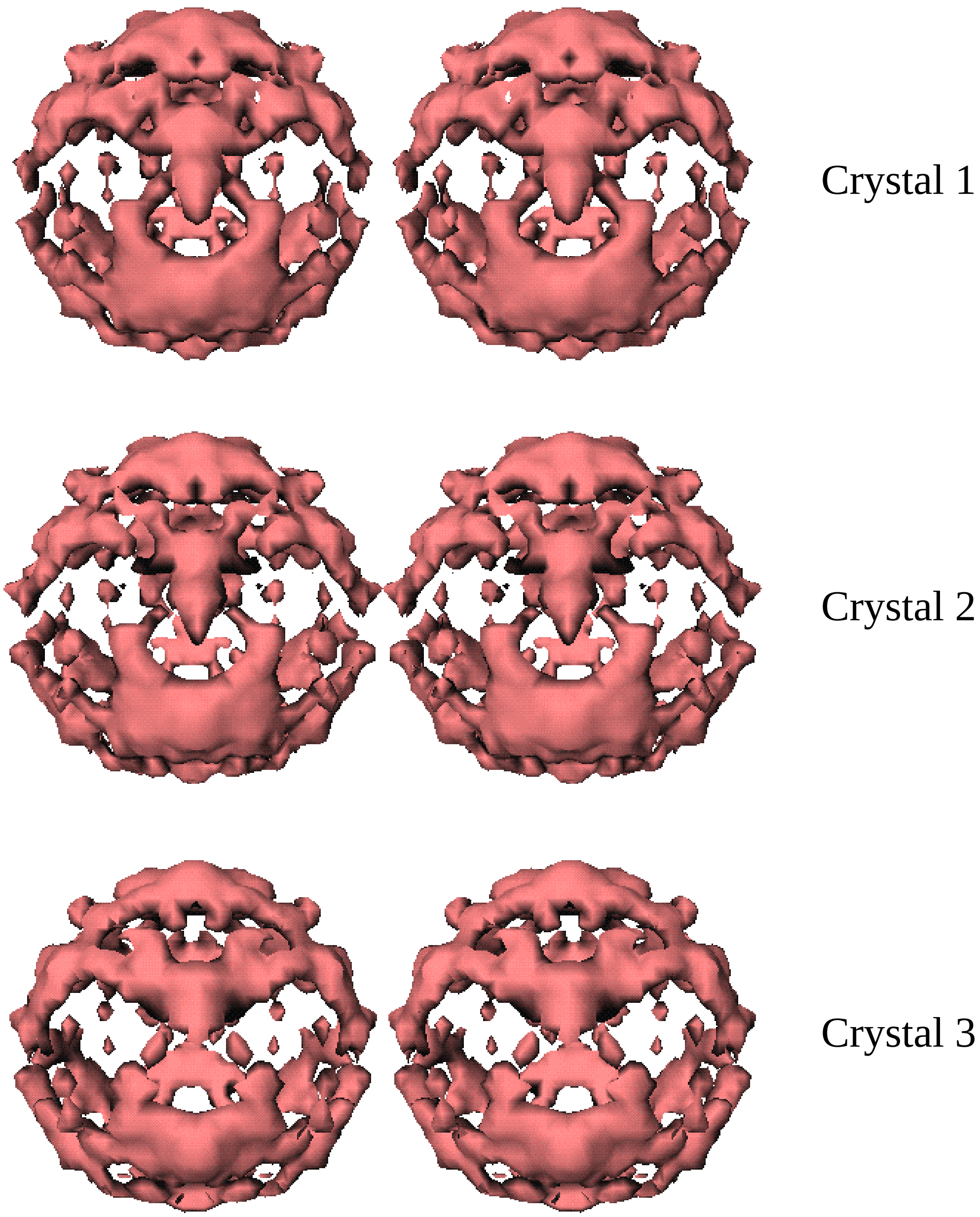

Figure 5.17: *Stereo views of the symmetry-averaged diffuse maps.* The same maps of figure 5.14, figure 5.15 and figure 5.16, displayed in stereo pairs, to better show the three-dimensionality of the features in the maps.

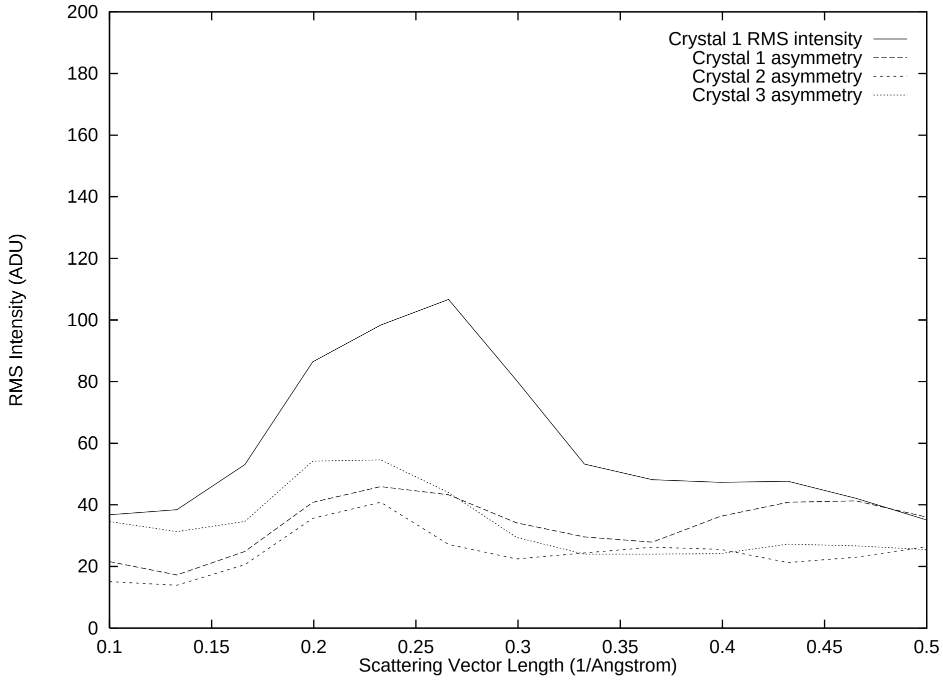

Figure 5.18: *Internal symmetry of* Staph. *nuclease diffuse maps.* RMS difference profiles between symmetrized and unsymmetrized diffuse maps constitute measures of the internal symmetry in *Staph.* nuclease diffuse maps. The profiles are shown for all three crystals, side-by-side with the RMS non-spherically-symmetric diffuse intensity profile of crystal 1 for comparison.

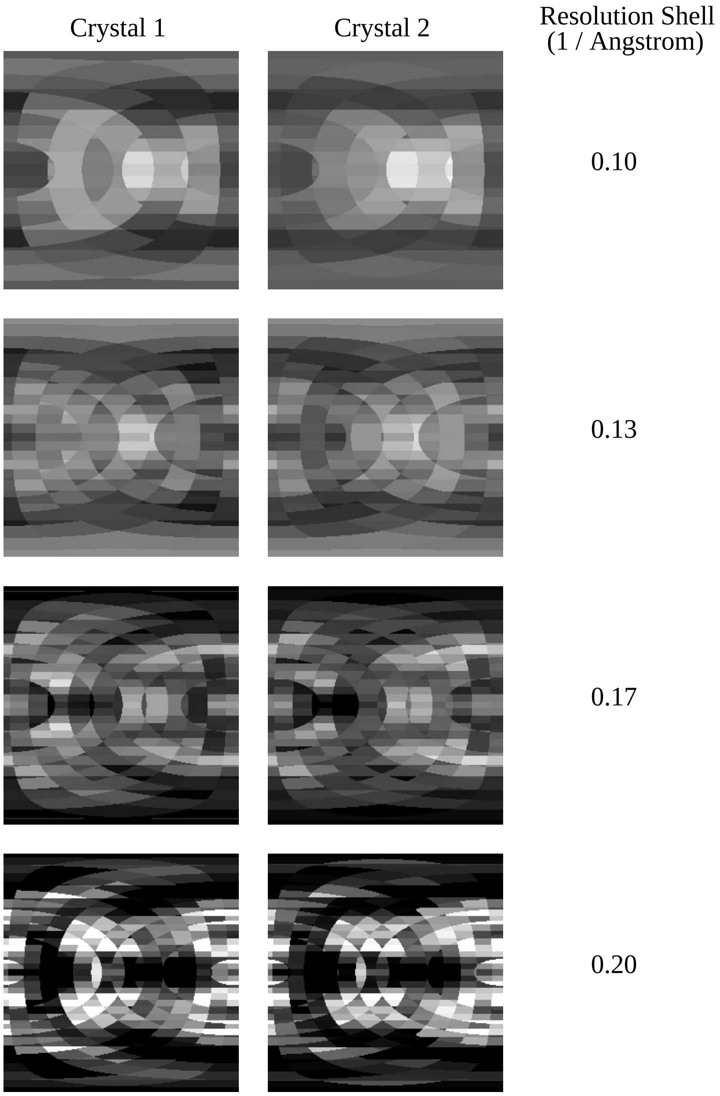

Figure 5.19: *Low-resolution shell images obtained from the symmetrized diffuse maps of crystal 1 and crystal 2.* Diffuse maps from crystal 1 and crystal 2 are compared side-by-side in shells between 0.1 $\mathbf{\AA}^{-1}$ and 0.2 $\mathbf{\AA}^{-1}$. In order to observe the variations in diffuse scattering within the shells, the maps have had the spherically-averaged intensity subtracted. The polar angle θ varies from 0 at the top to π at the bottom of each image, while, in order to eliminate the four-fold redundancy in the view, the azimuthal angle ϕ only varies from $-\pi$ at the left to $-\frac{1}{2}\pi$ at the right of each image. The images are displayed on a linear grey scale, with -100 corresponding to black, and 100 corresponding to white.

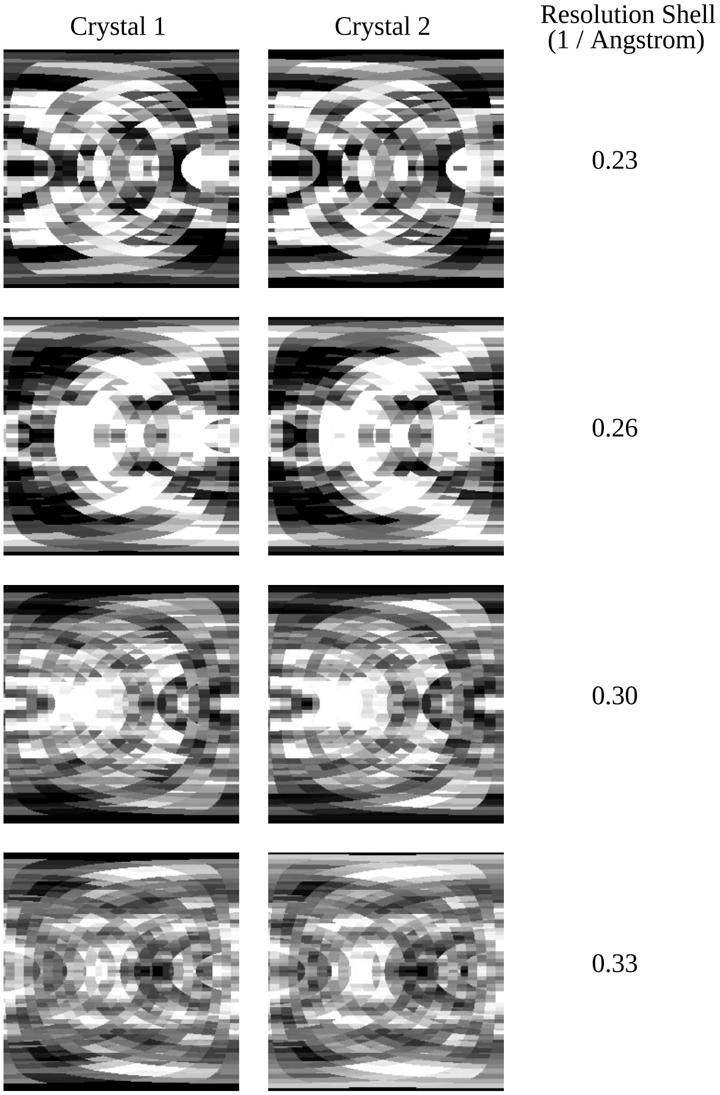

Figure 5.20: *Medium-resolution shell images obtained from the symmetrized diffuse maps of crystal 1 and crystal 2.* Diffuse maps from crystal 1 and crystal 2 are compared side-by-side in shells between 0.23 $\mathbf{Å}^{-1}$ and 0.33 $\mathbf{Å}^{-1}$. In order to observe the variations in diffuse scattering within the shells, the maps have had the spherically-averaged intensity subtracted. The polar angle θ varies from 0 at the top to π at the bottom of each image, while, in order to eliminate the four-fold redundancy in the view, the azimuthal angle ϕ only varies from $-\pi$ at the left to $-\frac{1}{2}\pi$ at the right of each image. The images are displayed on a linear grey scale, with -100 corresponding to black, and 100 corresponding to white.

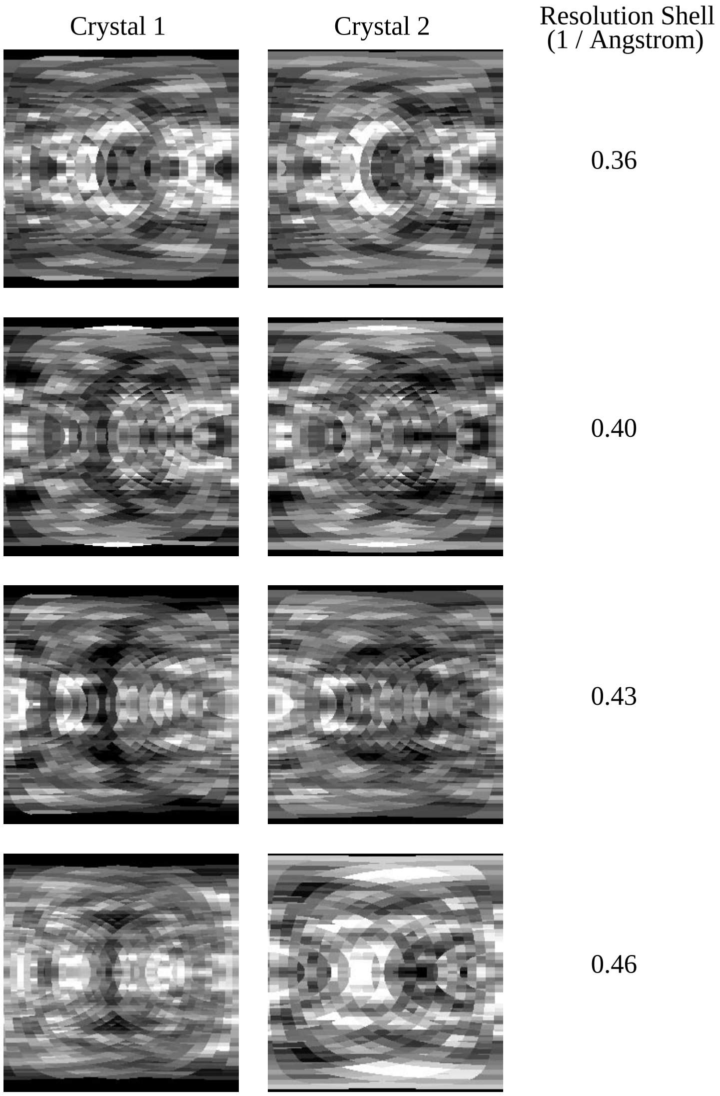

Figure 5.21: *High-resolution shell images obtained from the symmetrized diffuse maps of crystal 1 and crystal 2.* Diffuse maps from crystal 1 and crystal 2 are compared side-by-side in shells between 0.36 $\mathbf{\AA}^{-1}$ and 0.46 $\mathbf{\AA}^{-1}$. In order to observe the variations in diffuse scattering within the shells, the maps have had the spherically-averaged intensity subtracted. The polar angle θ varies from 0 at the top to π at the bottom of each image, while, in order to eliminate the four-fold redundancy in the view, the azimuthal angle ϕ only varies from $-\pi$ at the left to $-\frac{1}{2}\pi$ at the right of each image. The images are displayed on a linear grey scale, with -100 corresponding to black, and 100 corresponding to white.

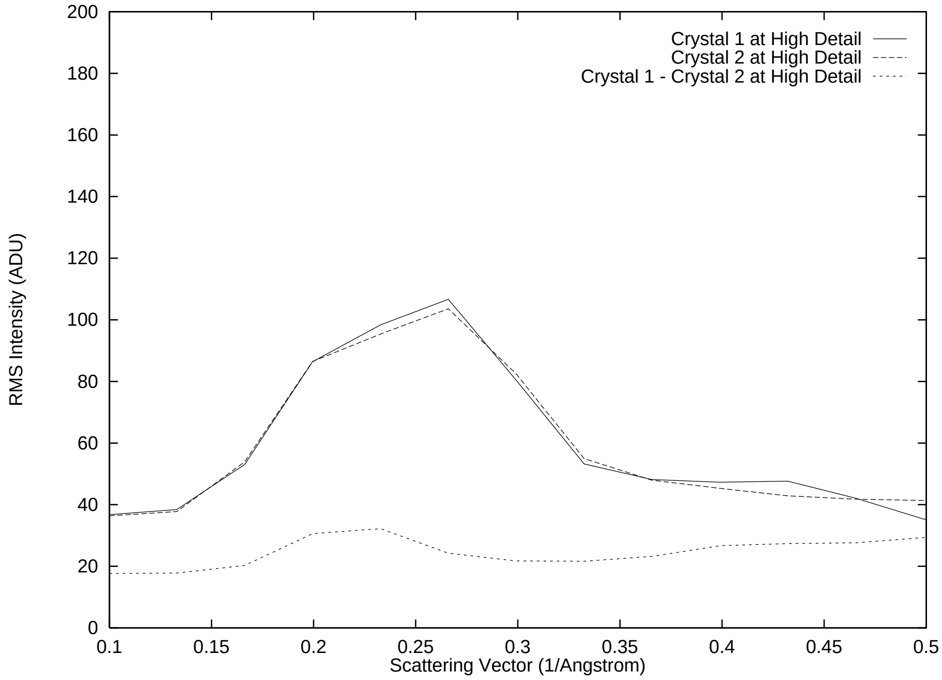

Figure 5.22: *Reproducibility of* Staph. *nuclease diffuse maps at high detail.* The difference profile between crystals 1 and 2 is much smaller than the profile of RMS non-spherically-symmetric diffuse intensity from either of the two crystals, indicating good reproducibility. One can estimate that the reproducibility is as good as 20% at 0.26 $\mathbf{\AA}^{-1}$ by comparing the RMS profiles to the difference profile.

by-side to see where they differ, and a three-dimensional comparison can be made by comparing images from many shells which span the resolution range of interest. For example, figures 5.19, 5.20 and 5.21 show a three-dimensional visual comparison of the diffuse maps from crystal 1 and crystal 2 over the range 0.1 - 0.5 $\mathbf{\AA}^{-1}$. These figures give a more complete visualization of the reproducibility of diffuse features than the views rendered using EXPLORER.

In order to quantify the observations seen by visualization techniques, the reproducibility of diffuse maps was characterized in a manner similar to that used to characterize the internal symmetry, as described in section 4.4. Initially, maps were symmetry averaged, and the spherically-averaged component was subtracted. By the method described above, the resulting diffuse maps were scaled to minimize the difference profile. The resulting profile was compared to the RMS intensity profile in order to measure the reproducibility.

In order to scale the maps from crystal 1 and crystal 2, the crystal 2 map had to be multiplied by a factor of 1.2. The confirmation of the fit for the scale factor is illustrated in figure 5.7. Figure 5.22 shows the resulting difference profile compared to the RMS profiles for crystal 1 and crystal 2, from which the reproducibility can be quantified. At 0.26 $\text{\AA}^{-1}$, the difference profile reads roughly 20 ADU, while the RMS intensity profile reads roughly 100 ADU, yielding a difference of only 20%. This is a better number than that obtained from the internal symmetry characterization, indicating that there is some systematic asymmetry in the measured diffuse maps. As is described in section 4.4, this is attributed to a measurement error, correctable by symmetry-averaging, rather than to an inherent asymmetry in maps of diffuse scattering from *Staph.* nuclease.

Reproducibility measurements were also performed on maps which were smoothed by convolution with a three-dimensional Gaussian map. To avoid edge effects due to Fourier transforms in the smoothing algorithm, maps were first symmetry-averaged, and then had their spherically-averaged intensity subtracted. Medium-detail diffuse maps were generated by convoluting the resulting maps with a Gaussian of half-width equal to one reciprocal unit-cell diagonal, and low-detail maps were generated by convolution with a Gaussian of half-width equal to two reciprocal unit-cell diagonals. Due to the discrete representation of the Gaussian on a three-dimensional lattice, the normalization was incorrect, which caused the scale of the maps to change after smoothing. Since maps were only compared at the same detail level, however, this effect was ignored for the studies performed here.

In order to scale the medium-detail maps from crystal 1 and crystal 2, the crystal 2

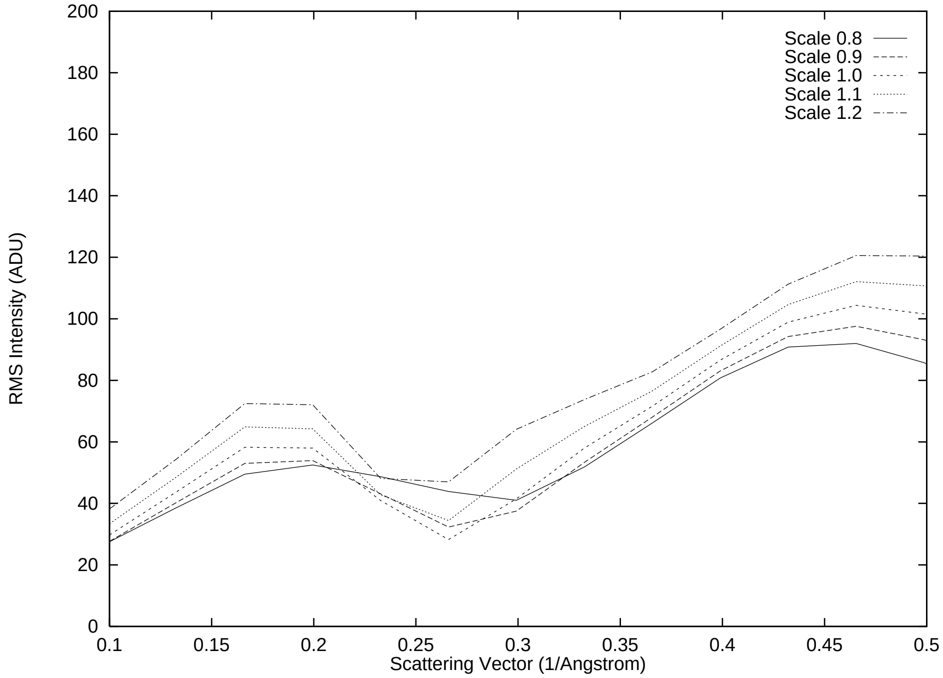

Figure 5.23: *Verification of scaling between medium-detail diffuse maps from crystal 1 and crystal 2.* As in figure 5.7, this verifies the fit for the scale factor between the medium-detail maps from the second and first crystals. The best fit is in the range 0.23-0.26 $\mathbf{Å}^{-1}$.

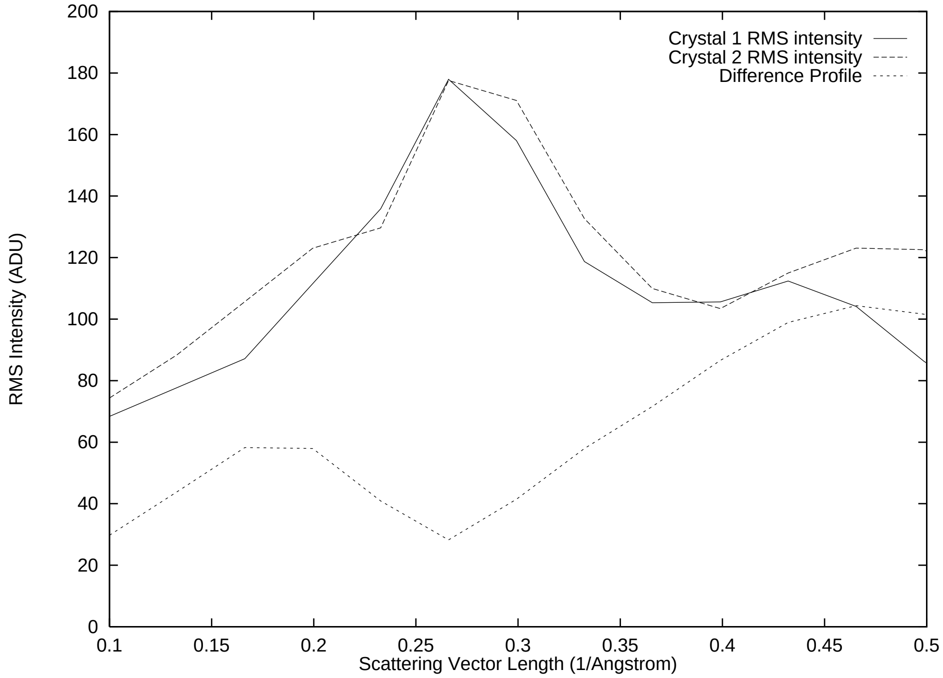

Figure 5.24: *Reproducibility of* Staph. *nuclease diffuse maps at medium detail.* In the neighborhood of 0.26 $\mathbf{\AA}^{-1}$, the difference between the maps is even smaller compared with the RMS profiles than at high detail. At higher resolution, however the differences become greater than at high detail.

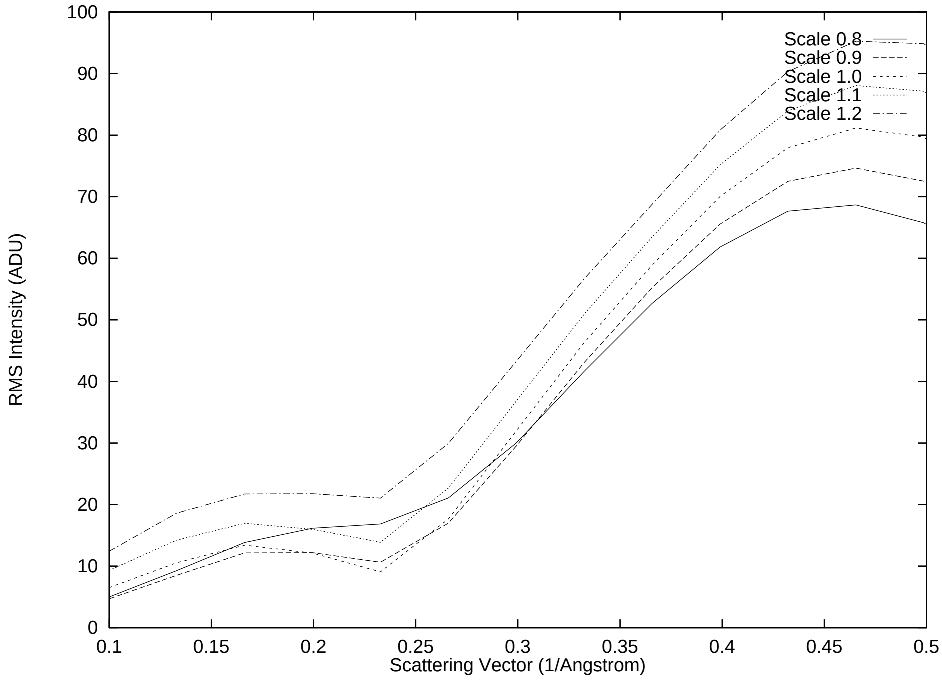

Figure 5.25: *Verification of scaling between low-detail diffuse maps from crystal 1 and crystal 2.* As in figure 5.7, this verifies the fit for the scale factor between the low-detail maps from the second and first crystals. The best fit is in the range 0.20-0.26 $\mathbf{\AA}^{-1}$.

map had to be multiplied by a factor of 1.2. The confirmation of the fit for the scale factor is illustrated in figure 5.23. Figure 5.24 shows the resulting difference profile compared to the RMS profiles for crystal 1 and crystal 2, from which the reproducibility can be quantified. At 0.26 $\mathbf{\AA}^{-1}$, the difference profile reads roughly 25 ADU, while the RMS intensity profile reads roughly 180 ADU, yielding a difference of only 15%, which is better than the difference at high detail. At high resolution, however, the differences are larger at medium detail than at high detail.

In order to scale the low-detail maps from crystal 1 and crystal 2, the crystal 2 map also had to be multiplied by a factor of 1.2. The confirmation of the fit for the scale factor is illustrated in figure 5.25. Figure 5.26 shows the resulting difference profile compared to the RMS profiles for crystal 1 and crystal 2, from which the

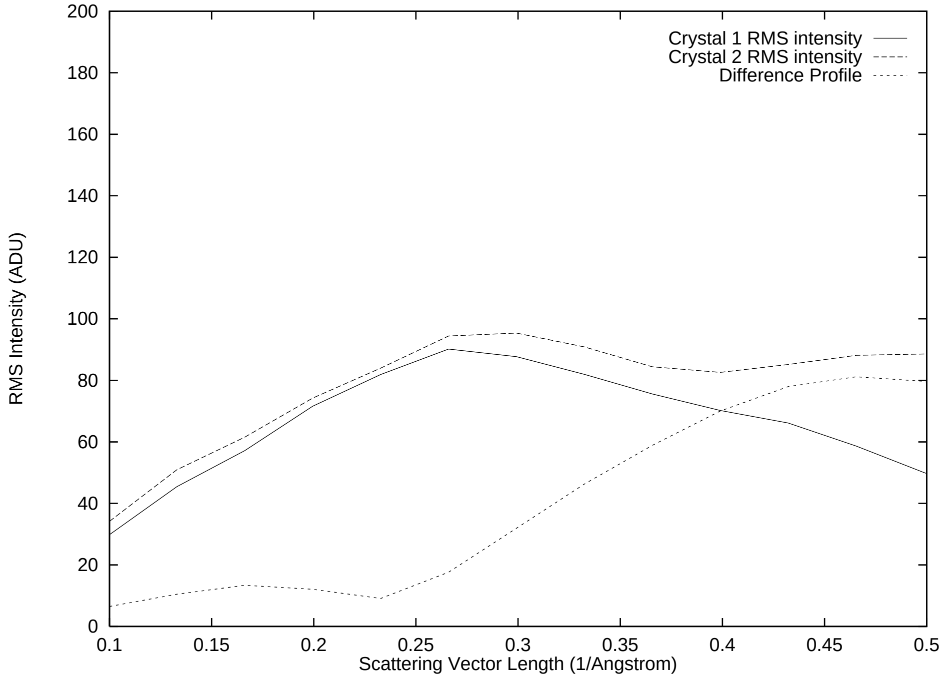

Figure 5.26: *Reproducibility of* Staph. *nuclease diffuse maps at low detail.* Differences between the maps become almost as small as 10% below 0.3 $\mathbf{\mathring{A}}^{-1}$, but at high resolution the difference profile sharply rises.

reproducibility can be quantified. At 0.23 $\text{\AA}^{-1}$, the difference profile reads roughly 10 ADU, while the RMS intensity profile reads roughly 90 ADU, yielding a difference of near 10%, which is the best agreement found among all maps. At high resolution, however, the differences are even larger at low detail than at medium detail.

In short, the differences between diffuse maps obtained from nominally identical crystals were measured as sufficiently small to suggest that significant changes in diffuse maps may be observable in controlled difference experiments. The quality of the maps suggested that modelling the three-dimensional diffuse scattering by generating simulated maps may provide meaningful insight into the nature of disorder in *Staph.* nuclease crystals. In addition, by smoothing the maps to medium- and low-detail levels, which proved useful in modelling the diffuse scattering, higher reproducibility was observed at low resolution, and lower reproducibility was observed at high resolution. Fortunately, in the resolution range 0.2-0.3 $\text{\AA}^{-1}$, where most of the interesting features were observed, the reproducibility was not harmed due to smoothing.

5.2 Difference Measurement.

Significant differences were observed between crystalline *Staph.* nuclease with and without Ca^{++} and pdTp bound. By comparing diffuse maps obtained from crystal 1 and crystal 3, the differences were quantified using the techniques developed to measure the reproducibility of diffuse maps (See section 4.4). These differences were significantly larger than those observed in the reproducibility measurement performed using crystal 1 and crystal 2, giving evidence for changes in the dynamics of the protein upon binding of Ca^{++} and pdTp.

Diffuse maps from crystal 1, crystal 2 and crystal 3 were initially compared using shell images, generated as described in section 4.2. Images from shells up to

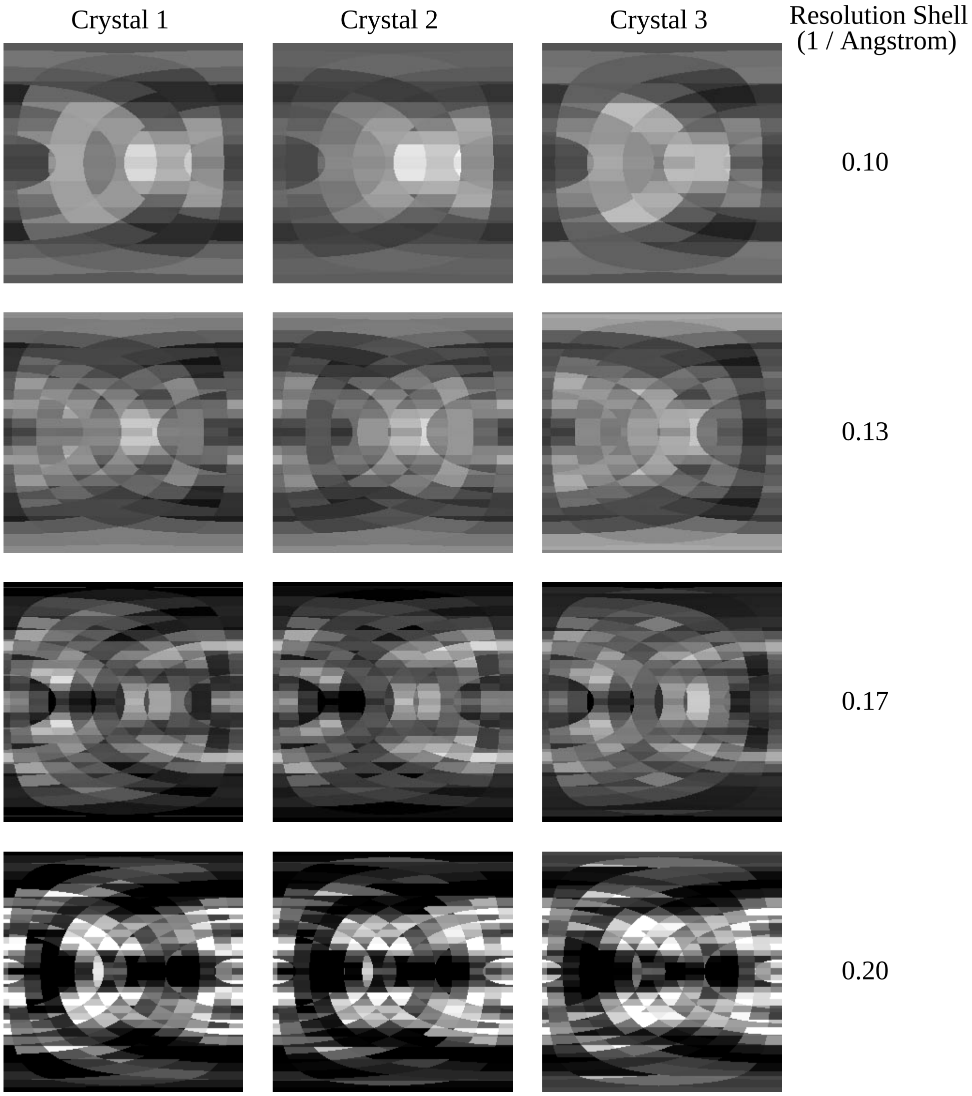

Figure 5.27: *Low-resolution shell images obtained from the symmetrized diffuse maps of crystal 1, crystal 2 and crystal 3.* Diffuse maps from crystal 1, crystal 2 and crystal 3 are compared side-by-side in shells between 0.1 $\text{\AA}^{-1}$ and 0.2 $\text{\AA}^{-1}$. In order to observe the variations in diffuse scattering within the shells, the maps have had the spherically-averaged intensity subtracted. The polar angle θ varies from 0 at the top to π at the bottom of each image, while, in order to eliminate the four-fold redundancy in the view, the azimuthal angle ϕ only varies from $-\pi$ at the left to $-\frac{1}{2}\pi$ at the right of each image. The images are displayed on a linear grey scale, with -100 corresponding to black, and 100 corresponding to white.

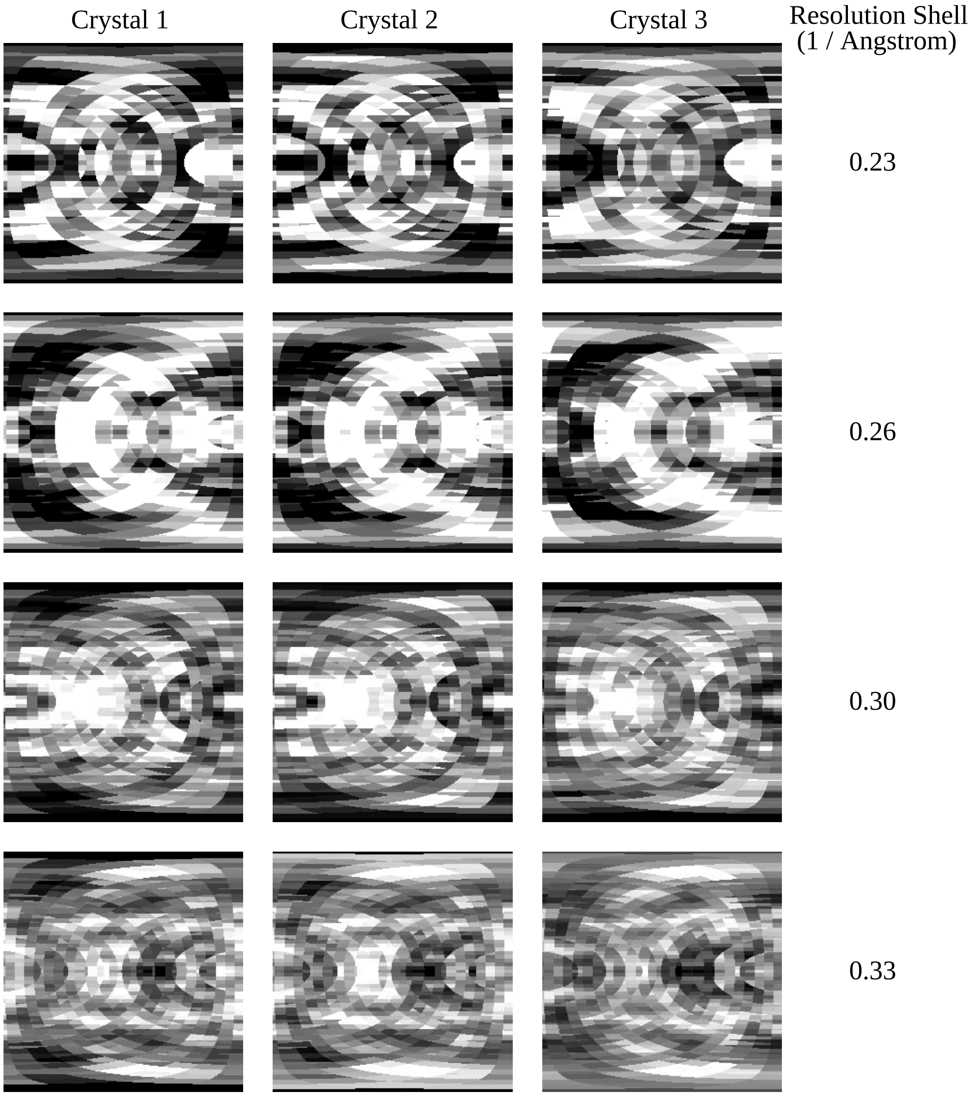

Figure 5.28: *Medium-resolution shell images obtained from the symmetrized diffuse maps of crystal 1, crystal 2 and crystal 3.* Diffuse maps from crystal 1, crystal 2 and crystal 3 are compared side-by-side in shells between 0.23 $\textbf{\AA}^{-1}$ and 0.33 $\textbf{\AA}^{-1}$. In order to observe the variations in diffuse scattering within the shells, the maps have had the spherically-averaged intensity subtracted. The polar angle θ varies from 0 at the top to π at the bottom of each image, while, in order to eliminate the four-fold redundancy in the view, the azimuthal angle ϕ only varies from $-\pi$ at the left to $-\frac{1}{2}\pi$ at the right of each image. The images are displayed on a linear grey scale, with -100 corresponding to black, and 100 corresponding to white.

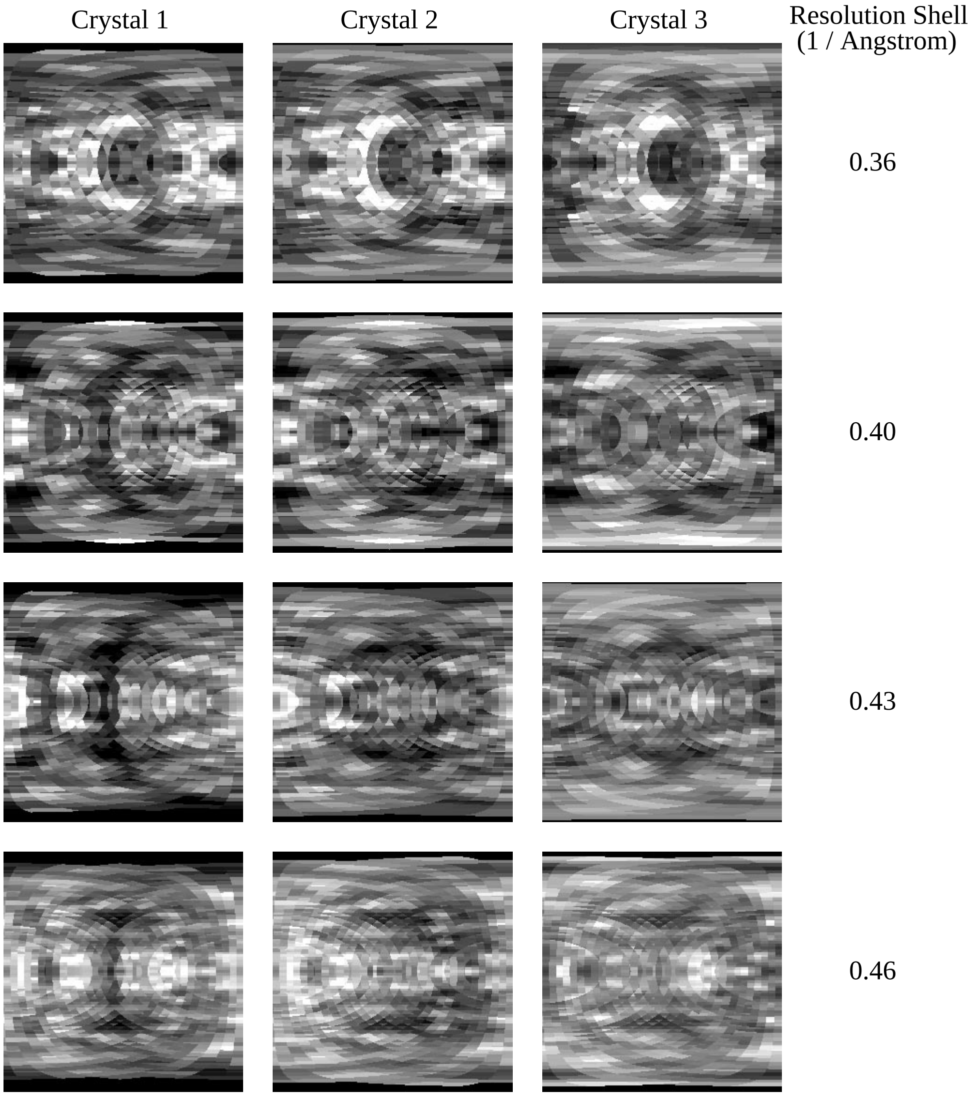

Figure 5.29: *High-resolution shell images obtained from the symmetrized diffuse maps of crystal 1, crystal 2 and crystal 3.* Diffuse maps from crystal 1, crystal 2 and crystal 3 are compared side-by-side in shells between 0.36 $\mathbf{Å}^{-1}$ and 0.46 $\mathbf{Å}^{-1}$. In order to observe the variations in diffuse scattering within the shells, the maps have had the spherically-averaged intensity subtracted. The polar angle θ varies from 0 at the top to π at the bottom of each image, while, in order to eliminate the four-fold redundancy in the view, the azimuthal angle ϕ only varies from $-\pi$ at the left to $-\frac{1}{2}\pi$ at the right of each image. The images are displayed on a linear grey scale, with -100 corresponding to black, and 100 corresponding to white.

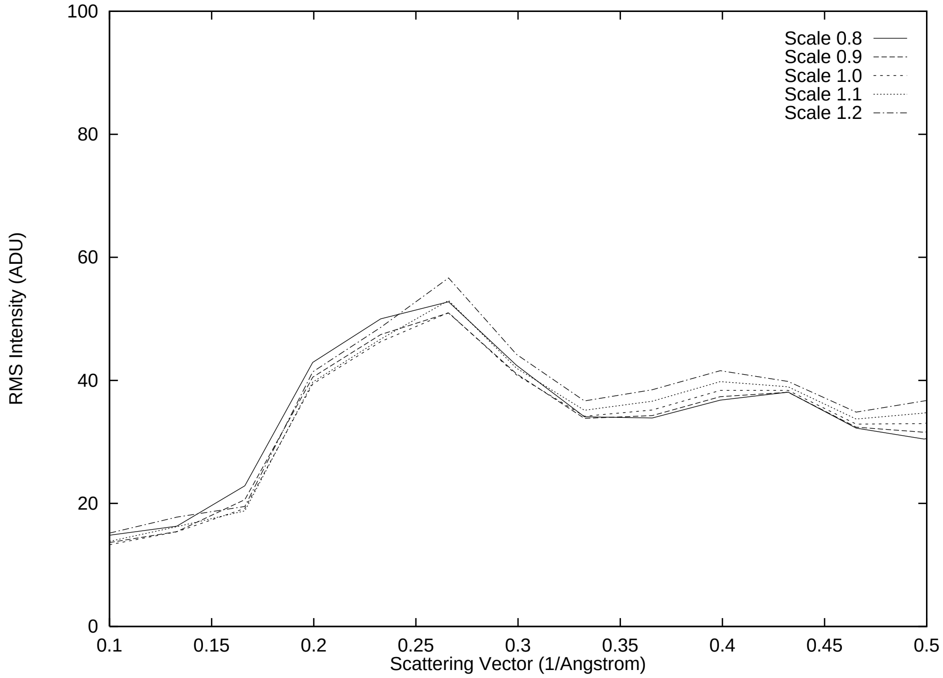

Figure 5.30: *Verification of scaling between diffuse maps from crystal 1 and crystal 3.* As in figure 5.7, this verifies the fit for the scale factor between the diffuse maps obtained from the third and first crystals. The fit is best below 0.33 $\mathbf{\AA}^{-1}$.

0.33 $\mathbf{\AA}^{-1}$, shown in figure 5.27 and figure 5.28, are visually very similar, with greater differences observable at higher resolution. Above 0.33 $\mathbf{\AA}^{-1}$, significant differences are observable between crystal 3 and the other crystals: most notably, there appears to be an enhancement of intensity at the top and bottom of the images, indicating enhanced diffuse scattering along the $\mathbf{c}^*$ direction. Differences are also visible in the more detailed patterns in the images.

To quantify the degree of differences, a procedure was followed which is analogous to that described for the reproducibility measurement in section 5.1. First, the maps from crystal 1 and crystal 3 were symmetry-averaged and had their spherically-averaged component subtracted. They were then scaled together by multiplying the crystal 3 map by 0.6, using figure 5.30 to verify the fit for the scale factor.

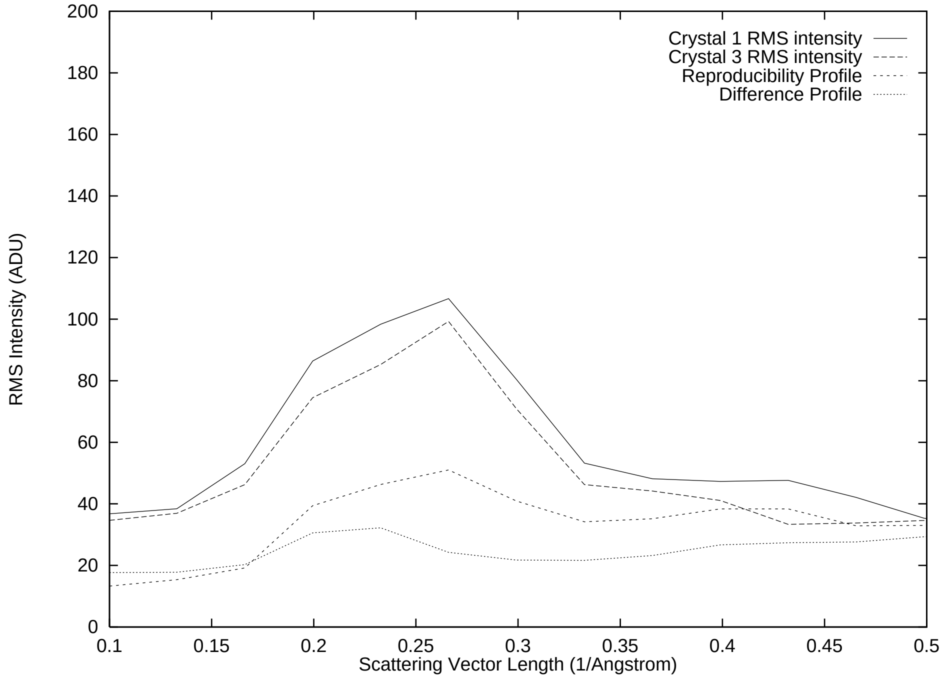

Figure 5.31: *Difference between maps of* Staph. *nuclease with and without* Ca^{++} *and pdTp.* Above 0.2 $\mathbf{\AA}^{-1}$, the difference profile between crystal 1 and crystal 3 diffuse maps is larger than the reproducibility profile. RMS intensity profiles from crystal 1 and crystal 3 are shown for scale. This plot indicates that the diffuse maps from crystal 1 and crystal 3 are significantly different.

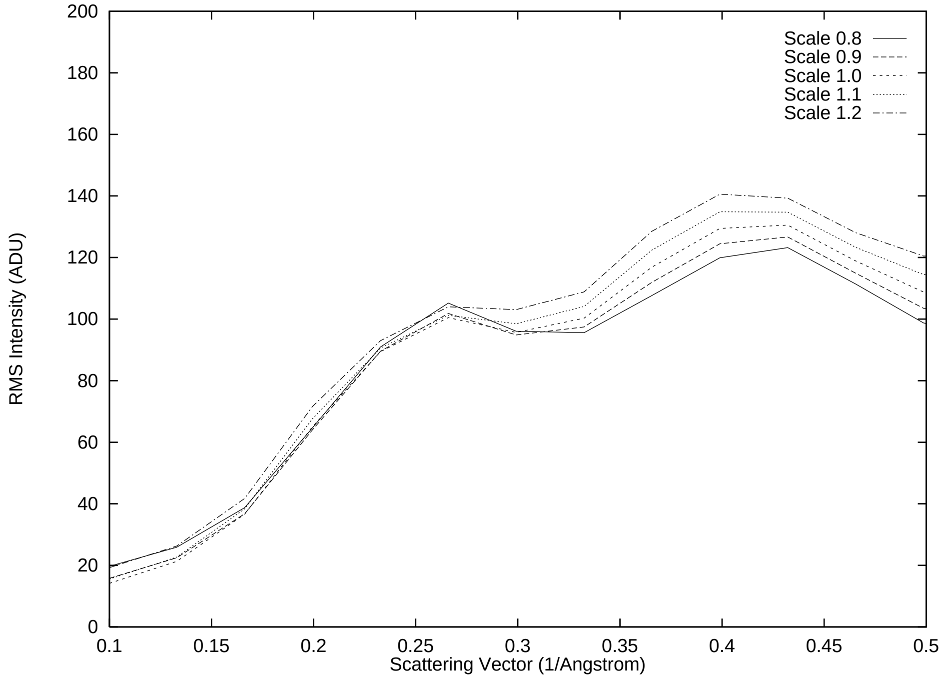

Figure 5.32: *Verification of scaling between medium-detail diffuse maps from crystal 1 and crystal 3.* As in figure 5.7, this verifies the fit for the scale factor between the medium-detail diffuse maps obtained from the third and first crystals. The fit is best below 0.30 $\text{\AA}^{-1}$.

The difference profile between the scaled maps is shown in figure 5.31, compared with the reproducibility profile and the RMS intensity profiles from crystal 1 and crystal 3. At 0.17 $\text{\AA}^{-1}$ and below, the difference profile is comparable to the profile obtained from the reproducibility measurement, indicating no significant difference. Above 0.17 $\text{\AA}^{-1}$, however, the difference significantly exceeds the profile from the reproducibility: at 0.26 $\text{\AA}^{-1}$, for instance, the difference profile reads roughly 50 ADU, while the reproducibility profile reads only 20 ADU.

A comparison of smoothed maps from crystal 1 and crystal 3 was also done. The confirmation of the fit for the scale factor between maps scaled to medium detail is illustrated in figure 5.32. As is seen in this figure, the scale factor does not minimize the difference above 0.3 $\text{\AA}^{-1}$, preventing comparisons at high resolution. Figure 5.33

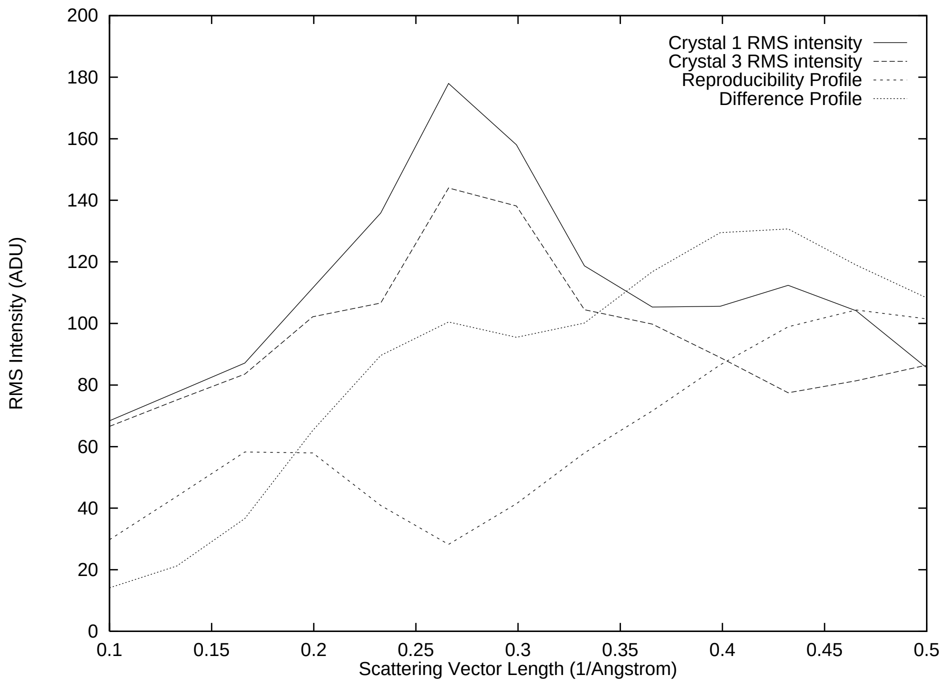

Figure 5.33: *Difference between medium-detail maps of* Staph. *nuclease with and without Ca^{++} and pdTp.* Above 0.2 $\text{\AA}^{-1}$, the difference profile between crystal 1 and crystal 3 medium-detail diffuse maps is larger than the reproducibility profile. Medium-detail RMS intensity profiles from crystal 1 and crystal 3 are shown for scale. This plot indicates that the medium-detail diffuse maps from crystal 1 and crystal 3 are significantly different.

shows the difference profile compared to the medium-detail RMS profiles for crystal 1 and crystal 3, along with the medium-detail reproducibility profile. At 0.26 Å^{-1}, the difference profile reads roughly 100 ADU, while the reproducibility profile reads roughly 25 ADU, showing that there is a significant difference between the medium-detail diffuse maps.

At low detail, no scale factor was found which minimized the difference between the maps, so that the degree of difference could not be measured. In general, the lower the level of detail obtained in smoothing diffuse maps, the greater the difference was between them. This trend was the same as observed in characterizing the reproducibility of diffuse maps, where low-detail maps differed more than high-detail maps at high resolution.

The observed difference between diffuse maps obtained from *Staph.* nuclease with and without Ca^{++} and pdTp bound can, in principle, indicate many different things, only one of which is a change in the dynamics of the crystalline protein. For example, changes in the average electron density of the unit cell could cause changes in diffuse scattering, either due to the the addition of Ca^{++} and pdTp to the crystal, or due to conformational changes in the protein.

As is described below, simulations indicate that the entire unit-cell electron density contributes to the diffuse scattering. Since Ca^{++} and pdTp represent only a small fraction of the total electron density of the unit cell, these units alone cannot account for the size of the change observed. In addition, although structural comparisons were not done in this work, previous high-resolution structural studies on *Staph.* nuclease show only minor changes upon binding Ca^{++} and pdTp, confined to a handful of residues and an ill-defined flexible loop in the protein[29]. The size of these domains similarly argues against their contributing to the changes.

Finally, it is important to note that a change in the dynamics of the crystalline protein is perfectly consistent with the observed change in diffuse scattering. Considering the results of the simulations described below, the nature of the change may be described by changes in the elasticity of the bulk protein crystal upon binding of Ca^{++} and pdTp. Such an effect could change the distribution of diffuse intensity while maintaining similarities in the patterns observable in diffuse maps. Ignoring the results of the simulations, the changes could be due to changes in the internal fluctuations of the protein upon binding of Ca^{++} and pdTp. Both kinds of changes could be potentially significant in understanding the function of the protein as an enzyme: the bulk elasticity of the crystal depends in part on the flexibility of the protein itself, which can significantly affect, *e.g.*, binding energies, while the nature of the fluctuations in the protein will directly affect reaction rates if different protein conformations are functionally distinguishable.

5.3 Modelling Results

As was described at the end of section 4.5, simulated maps of $|f(\mathbf{q})|^2$ were generated from many structural subdomains of the unit cell, and were compared with experimental maps of diffuse scattering. The subdomains are listed at the end of section 4.5, along with the motivation for selecting each one. Without smoothing, no simulated map could be successfully scaled against an experimentally obtained diffuse map.

After smoothing by convolution with a Gaussian of half-width equal to one reciprocal unit-cell diagonal, only the map generated from the entire unit cell could be scaled against the crystal 1 experimental diffuse map. The difference between the simulated and experimental maps was further reduced by smoothing using a Gaussian of half-width equal to two reciprocal unit-cell diagonals. The models most consistent with these results are the liquid-like motions model of Caspar *et. al.*[19] and the crys-

Figure 5.34: *Structural model of* Staph. *nuclease.* This view was generated from the PDB structure of *Staph.* nuclease, using the program Ribbons by Mike Carson[45]. Hydrogen-bonded regions of alpha-helices and beta-sheets are displayed as flat ribbons, while the remainder of the backbone is displayed as a thin, winding cylinder. Notable secondary structure includes the three alpha-helices displayed on top of the model, and a beta-barrel seen in the lower left.

talline normal modes model discussed in Glover *et. al.*[21]. A type of substitution order, where a fraction of the unit cells are randomly replaced with solvent, also cannot be ruled out.

As described in section 4.5, the structural model of Hynes and Fox[29] was downloaded from the Brookhaven Protein Data Bank and used as a starting point for simulating diffuse maps of *Staph.* nuclease. The program XPLOR by Axel Brünger[44] was used to refine the model against measured the Bragg reflections, which were obtained during diffuse scattering data collection (see section 2.1). The R-factor as measured by DENZO from internal symmetry considerations was 6% to 2.2 **Å**. The data set was 76% complete as determined by XPLOR after enforcing a 2σ cut on the Bragg peaks.

A forty-step rigid-body refinement was used to generate a structural model of

Staph. nuclease with XPLOR: the final R-factor was 33%. Given the quality of the diffuse data, this was considered adequate for generating simulated maps. A PDB-format structure was output and used as the input structure for simulations. Figure 5.34 shows a rendered view of the model.

By specifying a range of residues in XPLOR, a domain of the protein was selected from the list at the end of section 4.5. Maps of $|f(\mathbf{q})|^2$ were generated by zeroing all individual atomic B-factors and requesting an output of the calculated Bragg reflections based on the PDB model of *Staph.* nuclease. The resulting maps were converted to the same format as the experimentally obtained diffuse maps.

Following the arguments in section 3.4, the $|f(\mathbf{q})|^2$ maps were symmetry-averaged according to the predictions of space group $P4_1$. The only exception to this was the map generated from the full unit cell, which automatically had the correct symmetry. For this map alone, XPLOR was directed to generate $P4_1$ isomorphous copies of the asymmetric unit before calculating the $|f(\mathbf{q})|^2$ map.

As described in section 4.5, in order to compare the simulated maps with the experimental diffuse maps, the spherically-averaged component of the maps was subtracted. The maps were then scaled in an identical manner to that described in the reproducibility measurement of section 5.1. Without smoothing, no proper scale factor was found between any simulated and experimental maps.

After smoothing using a Gaussian of half-width equal to one reciprocal unit-cell diagonal, only the simulated map from the entire unit cell could be scaled with the experimental map. Attempts to scale simulated maps from other structural units failed to show a locally minimized difference profile. A verification of the fit for the unit-cell simulated map is shown in figure 5.35, in which it can be seen that it is only good over a limited range of roughly 0.26-0.3 $\mathbf{Å}^{-1}$. A comparison of the medium-detail shell images generated from the crystal 1 diffuse map and the simulated diffuse

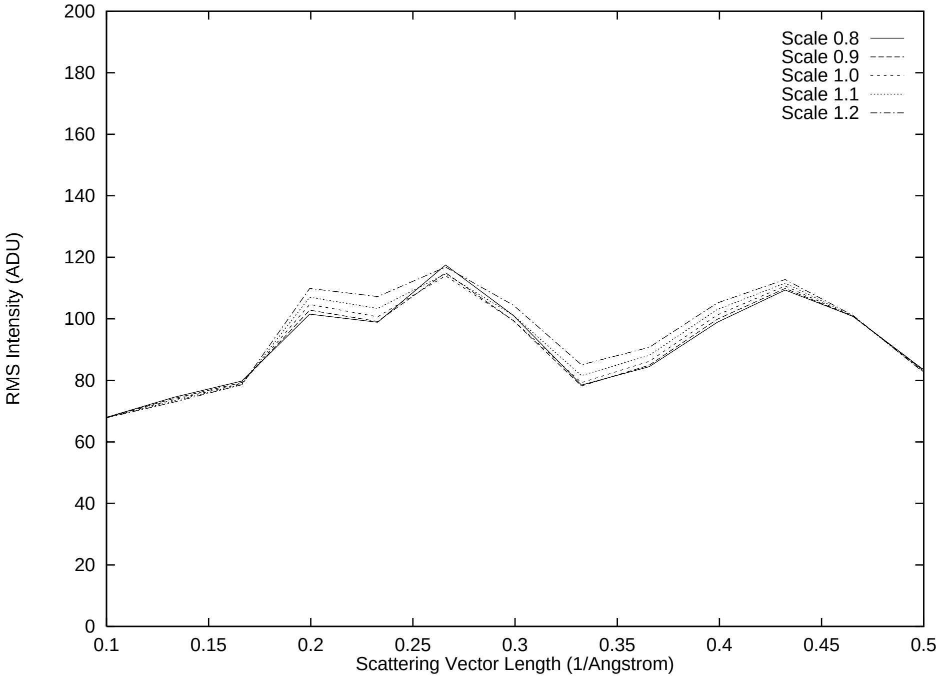

Figure 5.35: *Verification of scaling between medium-detail diffuse maps from crystal 1 and the simulated unit cell.* As in figure 5.7, this verifies the fit for the scale factor between the medium-detail diffuse maps. The fit is best in the range 0.26-0.33 $\mathbf{\AA}^{-1}$.

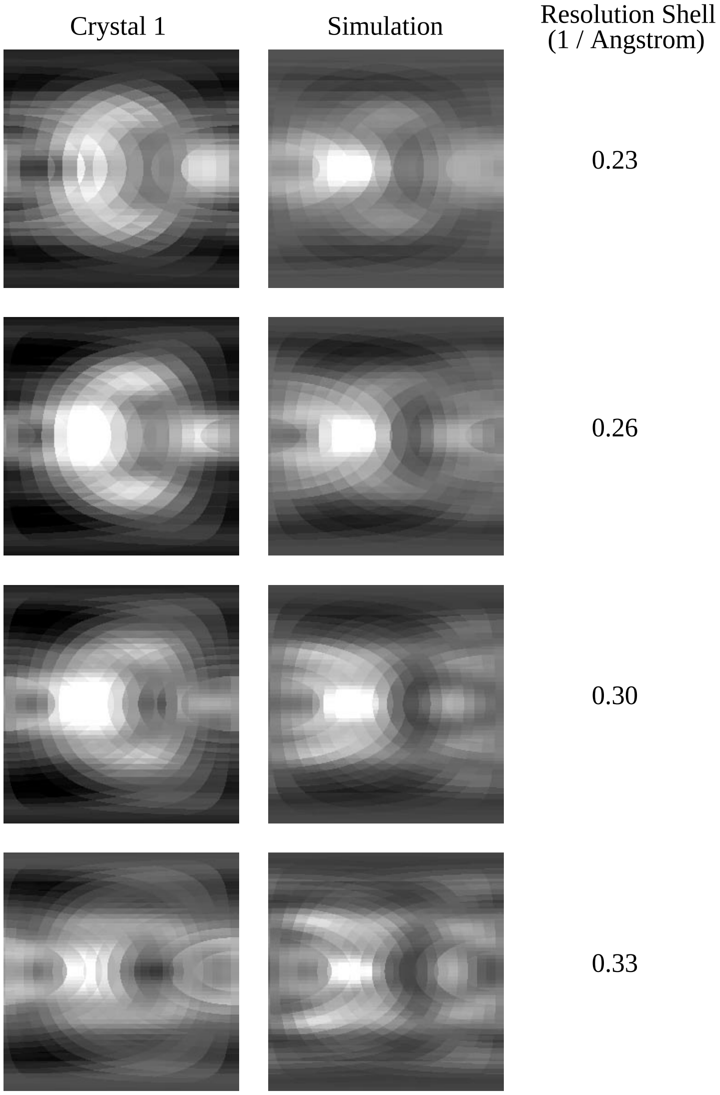

Figure 5.36: *Medium-resolution shell images obtained from the symmetrized diffuse map of crystal 1 and the unit-cell simulated map.* Diffuse maps from crystal 1 and the unit-cell simulated map are compared side-by-side in shells between 0.23 $\mathbf{Å}^{-1}$ and 0.33 $\mathbf{Å}^{-1}$. In order to observe the variations in diffuse scattering within the shells, the maps have had the spherically-averaged intensity subtracted. The polar angle θ varies from 0 at the top to π at the bottom of each image, while, in order to eliminate the four-fold redundancy in the view, the azimuthal angle ϕ only varies from $-\pi$ at the left to $-\frac{1}{2}\pi$ at the right of each image. The images are displayed on a linear grey scale, with -300 corresponding to black, and 300 corresponding to white.

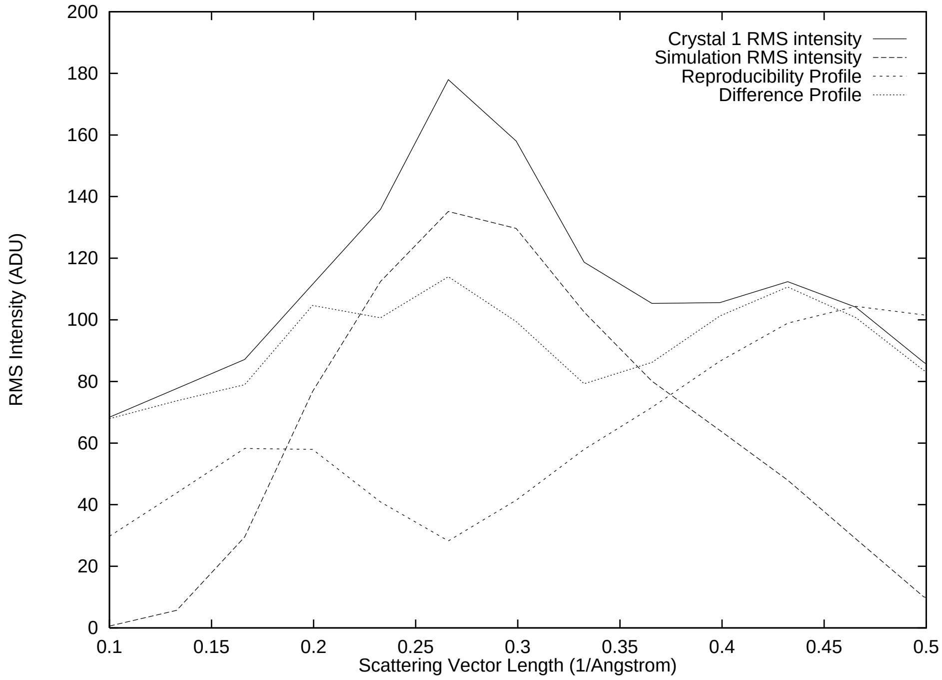

Figure 5.37: *Difference between medium-detail experimental and simulated maps of* Staph. *nuclease diffuse scattering.* Between 0.23 $\mathbf{\AA}^{-1}$ and 0.33 $\mathbf{\AA}^{-1}$, the difference profile is smaller than either of the RMS intensity profiles, indicating a degree of success in the simulation.

map is shown in figure 5.36. The similarities are not as striking as those observed in the reproducibility measurement, but common features are definitely visible in these images.

The methods developed for the reproducibility measurement were used to quantify the difference between the simulated and experimental maps. The difference profile calculated using the medium-detail scaled maps is shown in figure 5.37, and is plotted side-by-side with the reproducibility profile of figure 5.24 and the RMS intensity profiles of both the crystal 1 diffuse map and the simulated diffuse map. The fact that the difference profile is lower than either of the RMS intensity profiles is evidence for the correlation between the simulated and experimental maps.

An improved fit of the unit-cell simulated map with the crystal 1 diffuse map

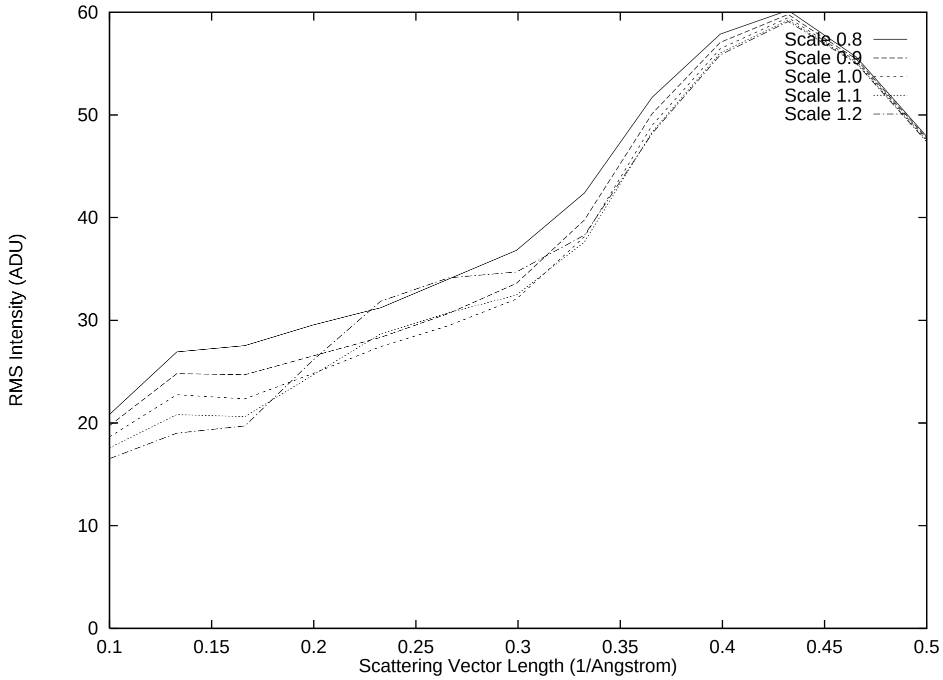

Figure 5.38: *Verification of scaling between low-detail diffuse maps from crystal 1 and the simulated unit cell.* As in figure 5.7, this verifies the fit for the scale factor between the low-detail diffuse maps. The fit is best in the range 0.20-0.33 $\mathbf{\AA}^{-1}$.

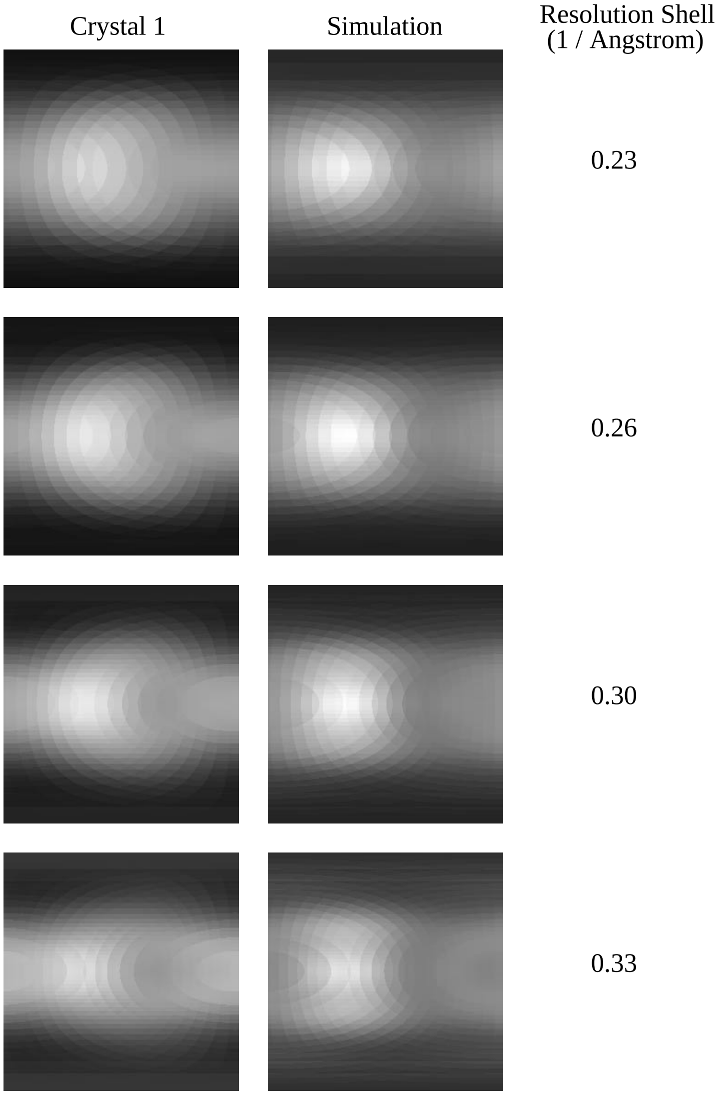

Figure 5.39: *Medium-resolution shell images obtained from the symmetrized diffuse map of crystal 1 and the unit-cell simulated map.* Diffuse maps from crystal 1 and the unit-cell simulated map are compared side-by-side in shells between 0.23 $\mathbf{\AA}^{-1}$ and 0.33 $\mathbf{\AA}^{-1}$. In order to observe the variations in diffuse scattering within the shells, the maps have had the spherically-averaged intensity subtracted. The polar angle θ varies from 0 at the top to π at the bottom of each image, while, in order to eliminate the four-fold redundancy in the view, the azimuthal angle ϕ only varies from $-\pi$ at the left to $-\frac{1}{2}\pi$ at the right of each image. The images are displayed on a linear grey scale, with -200 corresponding to black, and 200 corresponding to white.

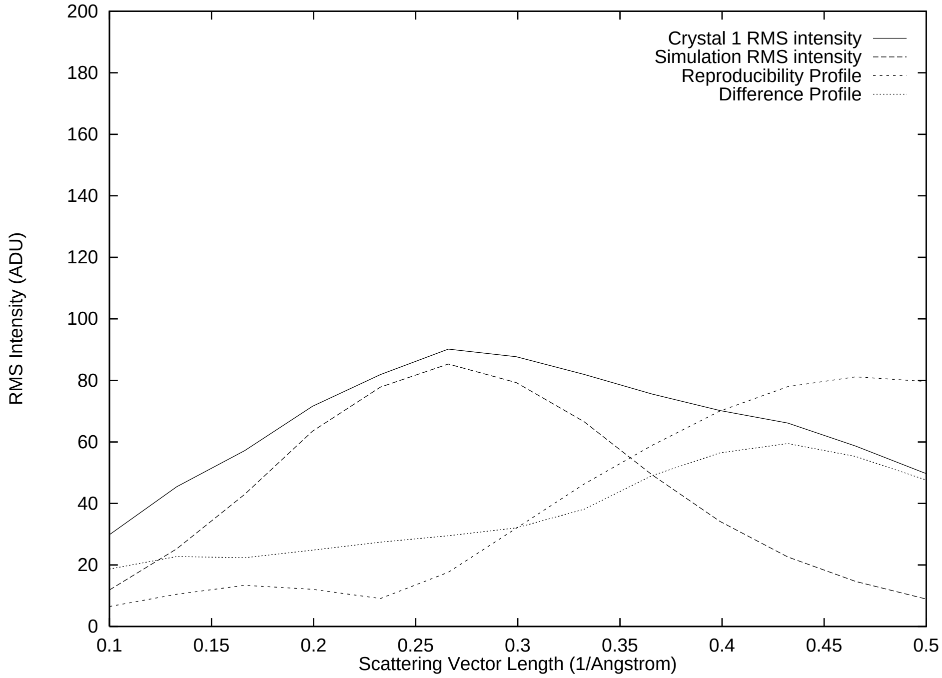

Figure 5.40: *Difference between low-detail experimental and simulated maps of* Staph. *nuclease diffuse scattering.* Between 0.13 Å^{-1} and 0.33 Å^{-1}, the difference profile is much smaller than either of the RMS intensity profiles, indicating good agreement between the simulation and experiment.

was seen when the maps were smoothed to low detail using a Gaussian of half-width equal to two reciprocal unit-cell diagonals. A verification of the low-detail fit is shown in figure 5.38, where it is seen that the fit holds over the region 0.23-0.33 Å^{-1}. A comparison of the low-detail shell images generated from the crystal 1 diffuse map and the simulated diffuse map is shown in figure 5.39. At this detail level, the two maps look somewhat similar.

Once again, the differences were quantified using the methods developed in the reproducibility measurement. The difference profile calculated using the low-detail scaled maps is shown in figure 5.40, and is plotted side-by-side with the reproducibility profile of figure 5.26 and the RMS intensity profiles of both the crystal 1 diffuse map and the simulated diffuse map. By comparing the height of the difference profile with the RMS intensity profiles, the fit at low detail is seen to compare favorably with the 33% R-factor of the rigid-body refinement, which sets a lower limit on the observed difference between the simulation and the data.

In summary, these results indicate that the three-dimensional diffuse scattering from *Staph.* nuclease crystals most resembles a map of $|f(\mathbf{q})|^2$, the squared structure factor, generated from the electron density of the entire four-protein unit cell. This fact alone can be used to place constraints on which models of disorder are consistent with the observed diffuse scattering from *Staph.* nuclease. For example, independent rigid-body motions of single proteins the unit cell, such as is described in the B-factor analysis of Kuriyan and Weis [16], give rise to diffuse maps which are derived from the structure factor of single proteins (see section 3.4). Since no correlation was observed between experimental maps and simulated maps generated from single-protein structure factors, such a model is inappropriate for describing the disorder which gives rise to diffuse scattering in *Staph.* nuclease crystals.

By similar arguments, sole, independent motions of each structural subdomain in the list at the end of section 4.5 are ruled out as the cause of the observed diffuse features. Arbitrary combinations of motions of these domains were not tested, and therefore are not ruled out. The strong correlation observed between the unit-cell simulated map and the crystal 1 diffuse map, however, suggests that subdomain motions do not contribute significantly to the diffuse scattering. Given the fact that many different models of diffuse scattering have been successfully used to simulate diffraction images in the past, however, more studies may need to be done to understand how robust the three-dimensional modelling is before such suggestions are given too much weight.

Models which are derived from the unit-cell structure factor are the ones which are indicated by the results of simulations of *Staph.* nuclease diffuse scattering. Four kinds of disorder are consistent with this result – two from each class described in section 3.2. In the class of independent unit cells, models based on both substitution disorder and independent, rigid-body translations of entire unit cells are indicated. In the class of correlated unit-cell motions, the liquid-like motions model of Caspar *et. al.*[19] and the crystalline normal modes model discussed in Glover *et. al.*[21] are indicated.

The substitution disorder is derived from equation 3.23 by assuming that the unit cell has two states: one where the unit cell is occupied by the four protein molecules, and one where it is unoccupied. Equation 3.23 can be rewritten as

$$I_D = N\langle |f_n - \langle f_n \rangle_n|^2 \rangle_n, \tag{5.1}$$

so that

$$I_D = N\langle |f_n - \nu f|^2 \rangle_n, \tag{5.2}$$

where f is the structure factor of the occupied unit cell, and ν is the fraction of unit cells which are occupied. The unit-cell structure factor f_n can assume a value of either f or 0. After some algebra, the expression for diffuse intensity becomes

$$I_D = N(1-\nu)\nu|f|^2, \tag{5.3}$$

which is proportional to the squared unit-cell structure factor $|f|^2$. For high occupancies, as one may expect for protein crystals, the diffuse scattering is proportional to the fraction $(1-\nu)$ of unoccupied unit cells, which makes it very small compared with the Bragg scattering. One way to definitively distinguish this form of disorder from others would be to look for temperature-induced changes in diffuse scattering, since one would not expect any changes if substitution disorder were giving rise to the observed diffuse features. Although no temperature study was performed in this work, experimental maps of *Staph.* nuclease show significant diffuse scattering: substitution disorder is therefore an unlikely, but possible, candidate as the cause of the observed diffuse scattering.

As is described in section 3.5, unit-cell rigid-body translations give rise to diffuse scattering of the form

$$I_D = N(1 - e^{-|\mathbf{q}\cdot\mathsf{S}|^2})|f|^2, \tag{5.4}$$

where f is the structure factor of the unit cell. No attempt was made to fit for a matrix S which parametrizes the motion of the unit cell, but the correlation of the experimental map with $|f|^2$ suggests this model as a possibility. In order for the unit cell to behave like a rigid unit, however, the contacts between proteins within the unit cell must be stronger than the contacts across unit cells – if they were of equal magnitude, the motions would more resemble those of individual molecules. If the four-molecule clustering of *Staph.* nuclease were energetically favorable, one would expect that under certain conditions clusters would form without crystallizing. There has been no suggestion that *Staph.* nuclease would tend to cluster in this fashion, but the results of these studies may be interpreted to provide evidence for such a state.

The remaining models which are consistent with the observed diffuse scattering in *Staph.* nuclease crystals are the liquid-like motions model of Caspar *et. al.*[19] and the crystalline normal modes model described in Glover *et. al.*[21], which are shown to be equivalent in section 3.9. To first order, these models predict diffuse scattering of the form

$$I_D = e^{-q^2\sigma^2} q^2\sigma^2 I_0(\mathbf{q}) * \Gamma(\mathbf{q}). \tag{5.5}$$

Since $I_0(\mathbf{q})$ is proportional to the squared structure factor $|f(\mathbf{q})|^2$ of the unit cell, this form of diffuse scattering resembles a map of $|f(\mathbf{q})|^2$ which has been convoluted with a smoothing function $\Gamma(\mathbf{q})$. The function $\Gamma(\mathbf{q})$ is related to the amplitudes of normal modes in the crystal as described in section 3.9.

In principle, it would be possible to characterize the correlation length, as was done in Caspar *et. al.*[19], or the elasticity of the crystal by establishing the best-fit functional dependence of $\Gamma(\mathbf{q})$. No explicit fit was performed in this work, but some insight can be gained into the length scales involved by observing the dependence of the success of simulations upon the detail level of smoothing. As mentioned above, without smoothing, no agreement was seen between experimental and simulated maps. At medium detail, where the first agreement was seen, the smoothing was performed by convolution with a Gaussian of half-width equal to one reciprocal unit-cell diagonal, which is 0.033 $\text{\AA}^{-1}$ long. In terms of $\Gamma(\mathbf{q})$, this would correspond to a roughly $1/(2\pi\times0.033) = 5$ Å correlation length, which is close to the 6 Å correlation length in insulin reported by Caspar *et. al.*[19].

Chapter 6

Conclusions

A significant result of the work presented in this thesis is a complete method for generating three-dimensional diffuse-scattering maps from images of x-ray diffraction from protein crystals. The method has been used here to study *Staph.* nuclease crystals, but in principle any crystal can be studied by the same techniques. The format of the measured diffuse maps is such that they can be represented in an identical fashion to a collection of measured Bragg reflections, making it easy to take advantage of existing code, such as XPLOR, for simulations.

Two methods are presented for characterizing the quality of measured diffuse maps: one which measures the degree of internal symmetry in a single map, and another which measures the degree of difference between diffuse maps from nominally identical crystals. These methods provide a benchmark with which to gauge the significance of changes observed in controlled difference experiments. They also provide a scale against which to compare differences between simulated and experimental diffuse maps. Being completely general, the methods can be used to characterize maps of diffuse scattering from any protein crystal.

A limitation to the method presented is that it only records diffuse features which are larger than the size of the reciprocal unit cell. Features on smaller length scales are either averaged over or removed. Thus, although the method adequately samples diffuse features which are due to motions uncorrelated across unit cells, features which are due to motions correlated across unit cells are not properly sampled, making quantitative analysis of these motions difficult. Future refinement of the method may increase the resolution of the maps, allowing the characterization of smaller diffuse features.

Certain steps in generating diffuse maps may need modification in future experiments. For certain crystal morphologies, for example, absorption corrections may be required for an accurate measurement. In addition, detectors with a large point-spread function may require more sophisticated methods than mode filtering in order to eliminate Bragg peaks. Perhaps other astronomical "de-starring" techniques would provide further insight into this problem.

Although *Staph.* nuclease lasted well in the intense x-ray beams at CHESS, other crystals may not behave as kindly. Therefore, it may become necessary to freeze crystals for diffuse scattering experiments. Although freezing techniques are now widely used, they may not be controlled enough for diffuse scattering experiments, since the common way of bringing a crystal to cryogenic temperatures is to simply dunk it in, *e.g.*, liquid nitrogen. Not only must the reproducibility of diffuse maps obtained from frozen crystals be characterized, but comparisons of diffuse maps from cryogenic and room-temperature crystals may be necessary to resolve the experimental issues involved.

Although a state-of-the-art CCD-based x-ray detector was used on a world-class synchrotron beamline for this experiment, diffuse scattering experiments have been performed on standard rotating anode sources and more common detectors, such as film and image plates. It would be interesting and important to see whether or not the quality of the three-dimensional diffuse maps is the same using these resources. If it is, diffuse scattering experiments would more easily become widespread.

The diffuse scattering in *Staph.* nuclease appears to change upon binding of Ca^{++} and pdTp. Arguments have been presented to show that the change is indicative of a change of dynamics in the crystal, but what are the resulting implications? Although *Staph.* nuclease has become a model system for studying protein folding, the effect of the internal dynamics of the protein on its activity is not yet understood. This crucial connection must be established in order for any characterization of protein dynamics to contribute significantly to an understanding how proteins work.

As described in the introduction, in order to make an experimental dent in this problem, it will be necessary to identify dynamical parameters of proteins which can be systematically varied and measured in a controlled fashion. Such a parameter may be something as specific as the amplitude of the displacement of a particular domain of the protein, or as general as a measure of protein elasticity. The connection with protein function must then be made experimentally by monitoring a measure of activity while varying the observed dynamical parameters.

The parameters of disorder in the models which best agree with the observed diffuse scattering from *Staph.* nuclease tend to describe material properties of the crystal, rather than specific motions of known functionally relevant groups in the protein. Understanding the functional relevance of the results through these models would come in two steps: first, a connection would need to be established between the parameters of crystalline disorder, such as correlation lengths or elasticities, and some dynamical property of the protein. If one assumes that the bulk properties of the crystal reflect the bulk properties of the protein, the measured parameters of disorder will apply to individual proteins, providing the link to dynamical properties. Elasticity, for instance, is related to the energy cost for the protein to adopt non-equilibrium configurations, which can directly affect reaction rates.

The second step is to relate the dynamical property of the protein to its functionality. This connection would have to be established through separate experiments, since diffuse scattering is unlikely to provide any direct information about protein activity. The ideal situation would involve crystallizing a protein which retains its functionality in the crystal, and characterizing both the activity and the protein dynamics simultaneously while changing the temperature or pressure of the specimen, varying the pH or the salt concentration, or even by changing the chemical potential of the water by immersing the crystal in a bath of dissolved organic polymers.

It is possible that some protein besides *Staph.* nuclease is a better candidate for establishing a complete connection between protein dynamics and protein function. Other proteins may show significantly different diffuse scattering, possibly connected with internal motions of the molecule. The methods developed here can be used to survey protein crystals which show significant diffuse scattering. It would be interesting, for example, to obtain three-dimensional diffuse maps from crystals of tRNA, which has been shown by Kolatkar *et. al.*[23] to exhibit signatures of intramolecular motion in images of x-ray diffraction, or from one of the many other crystals which have shown interesting diffuse features in x-ray diffraction experiments.

Further work will reveal how useful three-dimensional maps of diffuse scattering are in understanding how proteins work. The results here show how the maps can reveal the nature of disorder in a protein crystal, and suggest that protein dynamics can be characterized by diffuse scattering measurements. A survey of three-dimensional diffuse maps from protein crystals promises both to reveal much about the nature of protein dynamics, and to suggest how to quantitatively connect protein dynamics with protein function.

Appendix A

Supplemental Calculations

The following sections contain calculations supplemental to the description of models in chapter 3. Where they are missing, symbol definitions can be found in the cross-referenced section where the appropriate model is discussed.

A.1 Diffraction from an Imperfect Crystal

As is discussed in section 3.1, if the crystal is sufficiently large, the scattering intensity $I(\mathbf{q}, t)$ becomes effectively time-independent *even* in the disordered crystal[1]. To demonstrate this, the substitution $n' = n + \Delta n$ is first made in equation 3.8, where Δn points to a unit cell separated from n by $\mathbf{R}_{\Delta n}$ (see figure 3.1). It follows that

$$\begin{aligned} I(\mathbf{q}, t) = & \sum_{n,\Delta n} f_n(\mathbf{q}, t) f^*_{n+\Delta n}(\mathbf{q}, t) e^{-i\mathbf{q}\cdot\mathbf{R}_{\Delta n}} && \text{(A.1)} \\ = & \sum_{\Delta n} e^{-i\mathbf{q}\cdot\mathbf{R}_{\Delta n}} \sum_{n} f_n(\mathbf{q}, t) f^*_{n+\Delta n}(\mathbf{q}, t) && \text{(A.2)} \\ = & \sum_{\Delta n} e^{-i\mathbf{q}\cdot\mathbf{R}_{\Delta n}} N(\Delta n) \langle f_n(\mathbf{q}, t) f^*_{n+\Delta n}(\mathbf{q}, t)\rangle_n && \text{(A.3)} \end{aligned}$$

where $N(\Delta n)$ is the number of unit cells n in the crystal which are separated by a displacement vector $\mathbf{R}_{\Delta n}$.

---

[1] The following argument draws on a discussion on p. 155 of Guinier[40].

If $N(\Delta n)$ is sufficiently large, and the joint electron densities of all pairs (n, n') separated by a given Δn evolve identically, the set of products $\{f_n(\mathbf{q}, t) f^*_{n+\Delta n}(\mathbf{q}, t)\}$ will effectively span the space of all such products dynamically available. In this case the average $\langle f_n(\mathbf{q}, t) f^*_{n+\Delta n}(\mathbf{q}, t)\rangle_n$, and thus the scattering intensity, becomes time-independent. This condition is never met for the largest displacements $\mathbf{R}_{\Delta n}$ in the crystal, since there will always be too few unit cell pairs (n, n') which share the same displacement, yielding the resolution limitation discussed in the text.

A.2 Independent Molecular Domains

In order to derive an expression for the diffuse intensity I_D in section 3.3, it is necessary to obtain expressions for $\langle |f_n|^2\rangle_n$ and $|\langle f_n\rangle_n|^2$:

$$\langle f_n\rangle_n = \sum_j \langle f_{nj}\rangle_n \tag{A.4}$$

so that

$$|\langle f_n\rangle_n|^2 = \sum_{j,j'} \langle f_{nj}\rangle_n \langle f^*_{nj'}\rangle_n . \tag{A.5}$$

While for $\langle |f_n|^2\rangle_n$

$$\begin{aligned}
\langle |f_n|^2\rangle_n = \quad & \langle | \sum_j f_{nj}|^2\rangle_n & \text{(A.6)}\\
= \quad & \sum_j \langle |f_{nj}|^2\rangle_n + \sum_{j\neq j'} \langle f_{nj} f^*_{nj'}\rangle_n & \text{(A.7)}\\
= \quad & \sum_j \langle |f_{nj}|^2\rangle_n + \sum_{j\neq j'} \langle f_{nj}\rangle_n \langle f^*_{nj'}\rangle_n & \text{(A.8)}\\
= \quad & \sum_j (\langle |f_{nj}|^2\rangle_n - |\langle f_{nj}\rangle_n|^2) + |\langle f_n\rangle_n|^2, & \text{(A.9)}
\end{aligned}$$

where the substitution

$$\langle f_{nj} f^*_{nj'} \rangle_n = \langle f_{nj} \rangle_n \langle f^*_{nj'} \rangle_n \tag{A.10}$$

has been made based on the assumption of independent domains. The diffuse intensity I_D is given by

$$\begin{aligned} I_D = & \quad N(\langle |f_n|^2 \rangle_n - |\langle f_n \rangle_n|^2) & \text{(A.11)} \\ = & \quad N \sum_j (\langle |f_{nj}|^2 \rangle_n - |\langle f_{nj} \rangle_n|^2) & \text{(A.12)} \\ = & \quad \sum_j I_{Dj}. & \text{(A.13)} \end{aligned}$$

A.3 Independent Isomorphous Molecules

Using the equation

$$f_j(\mathbf{q}) = e^{i\mathbf{q}\cdot\mathbf{T}_j} f_0(\mathbf{q} \cdot \mathsf{R}_j) \tag{A.14}$$

derived in section 3.4, transformations of the average and average squared structure factor under rotations and translations can be calculated:

$$\begin{aligned} \langle f_{nj} \rangle_n = & \quad \langle e^{i\mathbf{q}\cdot\mathbf{T}_j} f_{n0}(\mathbf{q} \cdot \mathsf{R}_j) \rangle_n & \text{(A.15)} \\ = & \quad e^{i\mathbf{q}\cdot\mathbf{T}_j} \langle f_{n0}(\mathbf{q} \cdot \mathsf{R}_j) \rangle_n, & \text{(A.16)} \end{aligned}$$

where f_{nj} is the j^{th} copy of the reference structure factor f_{n0} in unit cell n, and the average is over all unit cells. This means that

$$\begin{aligned} |\langle f_{nj} \rangle_n|^2 = & \quad |e^{i\mathbf{q}\cdot\mathbf{T}_j} \langle f_{n0}(\mathbf{q} \cdot \mathsf{R}_j) \rangle_n|^2 & \text{(A.17)} \\ = & \quad |\langle f_{n0}(\mathbf{q} \cdot \mathsf{R}_j) \rangle_n|^2 & \text{(A.18)} \end{aligned}$$

and

$$\langle|f_{nj}|^2\rangle_n = \langle|e^{i\mathbf{q}\cdot\mathbf{T}_j}f_{n0}(\mathbf{q}\cdot\mathsf{R}_j)|^2\rangle_n \tag{A.19}$$

$$= \langle|f_{n0}(\mathbf{q}\cdot\mathsf{R}_j)|^2\rangle_n \tag{A.20}$$

From this it can be shown how the Bragg intensity I_{B0} and the diffuse intensity I_{D0} transform under a unit-cell rotation R_j and translation $\mathbf{T}_j$ into I_{Bj} and I_{Dj}. Using the results of section 3.3,

$$I_{Bj} = |\langle f_{nj}\rangle_n|^2 e^{i\mathbf{q}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})} \tag{A.21}$$

$$= |\langle f_{n0}(\mathbf{q}\cdot\mathsf{R}_j)\rangle_n|^2 e^{i\mathbf{q}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})}, \tag{A.22}$$

while

$$I_{Dj} = N(\langle|f_{nj}|^2\rangle_n - |\langle f_{nj}\rangle_n|^2) \tag{A.23}$$

$$= N(\langle|f_{n0}(\mathbf{q}\cdot\mathsf{R}_j)|^2\rangle_n - |\langle f_{n0}(\mathbf{q}\cdot\mathsf{R}_j)\rangle_n|^2) \tag{A.24}$$

$$= I_{D0}(\mathbf{q}\cdot\mathsf{R}_j) \tag{A.25}$$

For multiple isomorphous molecules j in the unit cell, the Bragg intensity I_B is then

$$I_B = |\langle f_n\rangle_n|^2 \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})} \tag{A.26}$$

$$= |\sum_j \langle f_{nj}\rangle_n|^2 \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})} \tag{A.27}$$

$$= |\sum_j e^{i\mathbf{q}\cdot\mathbf{T}_j}\langle f_{n0}(\mathbf{q}\cdot\mathsf{R}_j)\rangle_n|^2 \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})} \tag{A.28}$$

while the diffuse intensity I_D is given by

$$\begin{aligned} I_D = & \quad N(\langle|f_n|^2\rangle_n - |\langle f_n\rangle_n|^2) & \text{(A.29)} \\ = & \quad N\sum_j(\langle|f_{nj}|^2\rangle_n - |\langle f_{nj}\rangle_n|^2) & \text{(A.30)} \\ = & \quad N\sum_j(\langle|f_{n0}(\mathbf{q}\cdot\mathsf{R}_j)|^2\rangle_n - |\langle f_{n0}(\mathbf{q}\cdot\mathsf{R}_j)\rangle_n|^2) & \text{(A.31)} \\ = & \quad \sum_j I_{D0}(\mathbf{q}\cdot\mathsf{R}_j). & \text{(A.32)} \end{aligned}$$

A.4 Rigid-Body Translations

In order to obtain expressions for the diffuse intensity I_D and the Bragg intensity I_B for the case of rigid-body translations discussed in section 3.5, expressions for $|\langle f_n\rangle_n|^2$ and $\langle|f_n|^2\rangle_n$ must first be found. By equation 3.47,

$$\langle f_n\rangle_n = \langle e^{i\mathbf{q}\cdot\mathbf{u}_n}\rangle_n f. \tag{A.33}$$

For small displacements $\mathbf{u}_n$, the exponential can be expanded, so that

$$\begin{aligned} \langle e^{i\mathbf{q}\cdot\mathbf{u}_n}\rangle_n = & \quad \left\langle 1 + i\mathbf{q}\cdot\mathbf{u}_n - \frac{1}{2}\mathbf{q}\cdot\mathbf{u}_n\mathbf{u}_n\cdot\mathbf{q}\right\rangle_n & \text{(A.34)} \\ = & \quad 1 + i\mathbf{q}\cdot\langle\mathbf{u}_n\rangle_n - \frac{1}{2}\mathbf{q}\cdot\langle\mathbf{u}_n\mathbf{u}_n\rangle_n\cdot\mathbf{q}. & \text{(A.35)} \end{aligned}$$

The second term vanishes, since $\langle\mathbf{u}_n\rangle = 0$. The average of the tensor $\mathbf{u}_n\mathbf{u}_n$ can be written as

$$\begin{aligned} \langle\mathbf{u}_n\mathbf{u}_n\rangle_n = & \quad \langle u_{nx}u_{nx}\rangle_n\hat{\mathbf{x}}\hat{\mathbf{x}} + \langle u_{nx}u_{ny}\rangle_n\hat{\mathbf{x}}\hat{\mathbf{y}} + \langle u_{nx}u_{nz}\rangle_n\hat{\mathbf{x}}\hat{\mathbf{z}} \\ + & \quad \langle u_{ny}u_{nx}\rangle_n\hat{\mathbf{y}}\hat{\mathbf{x}} + \langle u_{ny}u_{ny}\rangle_n\hat{\mathbf{y}}\hat{\mathbf{y}} + \langle u_{ny}u_{nz}\rangle_n\hat{\mathbf{y}}\hat{\mathbf{z}} \\ + & \quad \langle u_{nz}u_{nx}\rangle_n\hat{\mathbf{z}}\hat{\mathbf{x}} + \langle u_{nz}u_{ny}\rangle_n\hat{\mathbf{z}}\hat{\mathbf{y}} + \langle u_{nz}u_{nz}\rangle_n\hat{\mathbf{z}}\hat{\mathbf{z}} & \text{(A.36)} \end{aligned}$$

Defining the symmetric variance matrix V as

$$\mathsf{V} \equiv \langle \mathbf{u}_n \mathbf{u}_n \rangle_n, \tag{A.37}$$

it follows that

$$\begin{aligned} \langle e^{i\mathbf{q}\cdot\mathbf{u}_n} \rangle_n &= 1 - \frac{1}{2}\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q} && \text{(A.38)} \\ &= e^{-\frac{1}{2}\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}} && \text{(A.39)} \end{aligned}$$

which is in the form of the familiar Debye-Waller factor[2]. The matrix V can be compared with the usual crystallographic anisotropic B-factor:

$$\langle e^{i\mathbf{q}\cdot\mathbf{u}_n} \rangle_n = e^{-\hat{\mathbf{q}}\cdot\mathsf{B}\cdot\hat{\mathbf{q}}\frac{\sin^2\theta}{\lambda^2}} \tag{A.40}$$

so that, using the fact that $q = 2k\sin\theta = \frac{4\pi}{\lambda}\sin\theta$ (see figure 3.2),

$$\mathsf{B} = 8\pi^2\mathsf{V}. \tag{A.41}$$

It follows from equation A.33 that

$$\langle f_n \rangle_n = e^{-\frac{1}{2}\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}} f \tag{A.42}$$

so that

$$|\langle f_n \rangle_n|^2 = e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}} |f|^2. \tag{A.43}$$

---

[2] See, *e.g.*, pp.148-154, chapter 2 of Giacovazzo[38]; pp.35-38, chapter 3 of Warren[42]; pp.186-193, chapter 7 of Guinier[40]; and pp. 20-25, chapter 2 of James[41]

The phase factor $e^{i\mathbf{q}\cdot\mathbf{u}_n}$ in equation 3.47 vanishes when calculating $|f_n(\mathbf{q})|^2$:

$$|f_n(\mathbf{q})|^2 = |e^{i\mathbf{q}\cdot\mathbf{u}_n} f(\mathbf{q})|^2 = |f(\mathbf{q})|^2, \tag{A.44}$$

so that

$$\langle |f_n|^2 \rangle_n = |f|^2. \tag{A.45}$$

Substituting the expressions for $|\langle f_n \rangle_n|^2$ and $\langle |f_n|^2 \rangle_n$ in equations A.43 and A.45 into the expression for the diffuse intensity I_D in equation 3.23:

$$I_D = N(\langle |f_n|^2 \rangle_n - |\langle f_n \rangle_n|^2) \tag{A.46}$$

$$= N(|f|^2 - e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}}|f|^2) \tag{A.47}$$

$$= N(1 - e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}})|f|^2. \tag{A.48}$$

The Bragg intensity I_B is given by substitution in equation 3.24:

$$I_B = |\langle f_n \rangle_n|^2 \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})} \tag{A.49}$$

$$= e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}}|f|^2 \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})}. \tag{A.50}$$

A.5 Rigid-Body Rotations

In order to calculate I_B and I_D for the case of rigid-body rotations discussed in section 3.7, $\langle f_n \rangle_n$ and $\langle |f_n|^2 \rangle$ must be calculated. If the rotations are decoupled from the translations, the following expressions hold:

$$\langle f_n \rangle_n = \sum_m f_m e^{i\mathbf{q}\cdot\mathbf{r}_0} \langle e^{i\mathbf{q}\cdot\mathsf{R}_n\cdot\mathbf{r}'_m} \rangle_n \langle e^{i\mathbf{q}\cdot\mathbf{u}_n} \rangle_n \tag{A.51}$$

and

$$\langle |f_n|^2 \rangle_n = \sum_{m,m'} f_m f_{m'} \langle e^{i\mathbf{q}\cdot \mathsf{R}_n \cdot (\mathbf{r}'_m - \mathbf{r}'_{m'})} \rangle_n, \tag{A.52}$$

so that the phase factor $e^{i\mathbf{q}\cdot\mathbf{r}_0}$ has dropped out. Let us first look at the term

$$A_m = \langle e^{i\mathbf{q}\cdot \mathsf{R}_n \cdot \mathbf{r}'_m} \rangle_n \tag{A.53}$$

from equation A.51. An angle α_n can be defined such that

$$\mathbf{q} \cdot \mathsf{R}_n \cdot \mathbf{r}'_m = q r'_m \cos(\theta'_m + \alpha_n), \tag{A.54}$$

where θ'_m is defined by the relation

$$\mathbf{q} \cdot \mathbf{r}'_m = q r'_m \cos(\theta'_m). \tag{A.55}$$

It is now assumed that the angle α is Gaussian distributed with a width σ_α, so that

$$A_m = N_\alpha \int d\alpha e^{iqr'_m \cos(\theta'_m + \alpha)} e^{-\frac{\alpha^2}{2\sigma_\alpha^2}}, \tag{A.56}$$

where N_α is a normalization constant. If the distribution in α is sufficiently narrow, the cosine term in the phase factor can be expanded:

$$\cos(\theta'_m + \alpha) \approx \cos(\theta'_m) - \alpha \sin(\theta'_m) \tag{A.57}$$

so that

$$A_m = e^{i\mathbf{q}\cdot\mathbf{r}'_m} N_\alpha \int d\alpha e^{-iqr'_m \sin(\theta'_m)\alpha - \frac{\alpha^2}{2\sigma_\alpha^2}}. \tag{A.58}$$

Completing the square in the exponential and integrating the resulting Gaussian to unity, one obtains

$$A_m = e^{i\mathbf{q}\cdot\mathbf{r}'_m} e^{-\frac{1}{2}q^2(r'_m)^2\sigma_\alpha^2\sin^2(\theta'_m)}, \tag{A.59}$$

or

$$A_m = e^{i\mathbf{q}\cdot\mathbf{r}'_m} e^{-\frac{1}{2}|\mathbf{q}\times\mathbf{r}'_m|^2\sigma_\alpha^2} \tag{A.60}$$

A similar derivation follows for

$$A_{mm'} = \langle e^{i\mathbf{q}\cdot\mathsf{R}_n\cdot(\mathbf{r}'_m-\mathbf{r}'_{m'})}\rangle, \tag{A.61}$$

where one obtains

$$A_{mm'} = e^{i\mathbf{q}\cdot(\mathbf{r}'_m-\mathbf{r}'_{m'})} e^{-\frac{1}{2}|\mathbf{q}\times(\mathbf{r}'_m-\mathbf{r}'_{m'})|^2\sigma_\alpha^2}. \tag{A.62}$$

From equations A.51 and A.52, it then follows that

$$|\langle f_n\rangle_n|^2 = \sum_{m,m'} f_m f_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m-\mathbf{r}_{m'})} e^{-\frac{1}{2}(|\mathbf{q}\times\mathbf{r}'_m|^2+|\mathbf{q}\times\mathbf{r}'_{m'}|^2)\sigma_\alpha^2} e^{-\mathbf{q}\cdot\mathsf{V}_\mathbf{u}\cdot\mathbf{q}} \tag{A.63}$$

and

$$\langle|f_n|^2\rangle_n = \sum_{m,m'} f_m f_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m-\mathbf{r}_{m'})} e^{-\frac{1}{2}|\mathbf{q}\times(\mathbf{r}'_m-\mathbf{r}'_{m'})|^2\sigma_\alpha^2} \tag{A.64}$$

so that

$$\begin{aligned} I_B = \quad & \sum_{m,m'} f_m f_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m-\mathbf{r}_{m'})} e^{-\frac{1}{2}(|\mathbf{q}\times\mathbf{r}'_m|^2+|\mathbf{q}\times\mathbf{r}'_{m'}|^2)\sigma_\alpha^2} e^{-\mathbf{q}\cdot\mathsf{V}_\mathbf{u}\cdot\mathbf{q}}, && \text{(A.65)} \\ & \times\sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})} && \text{(A.66)} \end{aligned}$$

and

$$I_D = \quad N(\langle|f_n|^2\rangle_n - |\langle f_n\rangle_n|^2) \tag{A.67}$$

$$
\begin{aligned}
= \quad & N \sum_{m,m'} f_m f_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m - \mathbf{r}_{m'})} \\
& \times \left(e^{-\frac{1}{2}|\mathbf{q}\times(\mathbf{r}'_m - \mathbf{r}'_{m'})|^2 \sigma_\alpha^2} - e^{-\frac{1}{2}(|\mathbf{q}\times\mathbf{r}'_m|^2 + |\mathbf{q}\times\mathbf{r}'_{m'}|^2)\sigma_\alpha^2} e^{-\mathbf{q}\cdot\mathsf{V}_\mathbf{u}\cdot\mathbf{q}}\right) \qquad \text{(A.68)}
\end{aligned}
$$

A.6 Liquid-like Correlated Motions

As a sanity check, in the case where all of the atoms execute independent, isotropic displacements, an expression for I_D analogous to that of equation 3.70 should be obtained from the expression for diffuse scattering due liquid-like correlated motions in section 3.8.. The correlation coefficient $\Gamma(\mathbf{r})$ in this case will be 1 for $r = 0$, and 0 for $r \neq 0$, so that

$$\Gamma(\mathbf{r}) = \delta(\mathbf{r}). \qquad \text{(A.69)}$$

Working from equation 3.107,

$$
\begin{aligned}
I_D = \quad & e^{-q^2\sigma^2} \int d^3\,\mathbf{r} e^{i\mathbf{q}\cdot\mathbf{r}}\, q^2\sigma^2\delta(\mathbf{r})\, P_0(\mathbf{r}) && \text{(A.70)} \\
= \quad & e^{-q^2\sigma^2} q^2\sigma^2 P_0(0) && \text{(A.71)} \\
\approx \quad & (1 - e^{-q^2\sigma^2}) P_0(0) && \text{(A.72)}
\end{aligned}
$$

which is the isotropic equivalent to equation 3.70 after using equation 3.106 to show that

$$
\begin{aligned}
P_0(0) = \quad & \sum_{m,m'} Q_m Q_{m'} \delta(\mathbf{r}_m - \mathbf{r}_{m'}) && \text{(A.73)} \\
= \quad & \sum_m Q_m^2, && \text{(A.74)}
\end{aligned}
$$

since

$$\sum_m |f_m|^2 = \sum_m Q_m^2 \qquad \text{(A.75)}$$

for a point-like charge distribution.

A.7 Crystalline Normal Modes

Given the expression

$$\langle I(\mathbf{q},t)\rangle_t = \sum_{m,m'} Q_m Q_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m-\mathbf{r}_{m'})} \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})} \langle e^{i\mathbf{q}\cdot(\mathbf{u}_{mn}-\mathbf{u}_{m'n'})}\rangle_t, \tag{A.76}$$

a first-order normal modes expansion can be performed to give the results discussed in section 3.9. For small perturbations $\mathbf{u}_{mn}$, the approximation

$$\langle e^{i\mathbf{q}\cdot(\mathbf{u}_{mn}-\mathbf{u}_{m'n'})}\rangle_t = e^{-\frac{1}{2}\langle[\mathbf{q}\cdot(\mathbf{u}_{mn}-\mathbf{u}_{m'n'})]^2\rangle}{}_t \tag{A.77}$$

can be made, so that the expression

$$\langle\,[\mathbf{q}\cdot(\mathbf{u}_{mn}-\mathbf{u}_{m'n'})]^2\,\rangle_t = \langle(\mathbf{q}\cdot\mathbf{u}_{mn})^2\rangle_t + \langle(\mathbf{q}\cdot\mathbf{u}_{m'n'})^2\rangle_t - 2\langle\mathbf{q}\cdot\mathbf{u}_{mn}\mathbf{u}_{m'n'}\cdot\mathbf{q}\rangle_t \tag{A.78}$$

must be calculated. The atomic displacements $\mathbf{u}_{mn}$ are now expanded in terms of crystalline normal modes:

$$\mathbf{u}_{mn} = \sum_{\mathbf{g}s} a_{s,\mathbf{g}}\mathbf{e}_{ms,\mathbf{g}} e^{i(\mathbf{g}\cdot\mathbf{R}_n-\omega_{s,\mathbf{g}}t)}, \tag{A.79}$$

where $\mathbf{e}_{ms,\mathbf{g}}$ is the complex polarization vector for atom m in branch s of the mode with wave vector $\mathbf{g}$, $a_{s,\mathbf{g}}$ is the amplitude of branch s of mode $\mathbf{g}$, and $\omega_{s,\mathbf{g}}$ is its frequency. This yields

$$\langle(\mathbf{q}\cdot\mathbf{u}_{mn})^2\rangle_t = \sum_{\mathbf{g},\mathbf{g}'}\sum_{s,s'} a_{s,\mathbf{g}}a_{s',\mathbf{g}'}$$

$$\times \quad \mathbf{q}\cdot\mathbf{e}_{ms,\mathbf{g}}\mathbf{q}\cdot\mathbf{e}_{ms',\mathbf{g}'}\langle e^{i[(\mathbf{g}+\mathbf{g}')\cdot\mathbf{R}_n-(\omega_{s,\mathbf{g}}+\omega_{s',\mathbf{g}'})t]}\rangle_t \tag{A.80}$$

$$= \quad \sum_{\mathbf{g},\mathbf{g}'}\sum_{s,s'} a_{s,\mathbf{g}}a_{s',\mathbf{g}'}$$

$$\times \quad \mathbf{q}\cdot\mathbf{e}_{ms,\mathbf{g}}\mathbf{q}\cdot\mathbf{e}_{ms',\mathbf{g}'}e^{i(\mathbf{g}+\mathbf{g}')\cdot\mathbf{R}_n}\delta_{\mathbf{g},-\mathbf{g}'}\delta_{s,s'} \tag{A.81}$$

$$= \quad \sum_{\mathbf{g},s} a_{s,\mathbf{g}}a_{s,-\mathbf{g}}\mathbf{q}\cdot\mathbf{e}_{ms,\mathbf{g}}\mathbf{q}\cdot\mathbf{e}_{ms,-\mathbf{g}}, \tag{A.82}$$

where in order to introduce $\delta_{\mathbf{g},-\mathbf{g}'}\delta_{s,s'}$ it is assumed that each mode $(\mathbf{g}, s)$ has a unique frequency $\omega_{s,\mathbf{g}}$. Remembering that

$$\langle(\mathbf{q}\cdot\mathbf{u}_{mn})^2\rangle_t = \mathbf{q}\cdot\mathsf{V}_m\cdot\mathbf{q}, \tag{A.83}$$

where V_m is the variance matrix for displacements of atom m, it follows that

$$\mathsf{V}_m = \sum_{\mathbf{g},s}|a_{s,\mathbf{g}}|^2\mathbf{e}_{ms,\mathbf{g}}\mathbf{e}_{ms,-\mathbf{g}}. \tag{A.84}$$

In a similar fashion,

$$\langle(\mathbf{q}\cdot\mathbf{u}_{m'n'})^2\rangle_t = \mathbf{q}\cdot\left[\sum_{\mathbf{g},s}|a_{s,\mathbf{g}}|^2\mathbf{e}_{m's,\mathbf{g}}\mathbf{e}_{m's,-\mathbf{g}}\right]\cdot\mathbf{q}. \tag{A.85}$$

The cross term becomes

$$\langle(\mathbf{q}\cdot\mathbf{u}_{mn})(\mathbf{q}\cdot\mathbf{u}_{m'n'})\rangle_t = \quad \sum_{\mathbf{g},\mathbf{g}'}\sum_{s,s'} a_{s,\mathbf{g}}a_{s',\mathbf{g}'}\mathbf{q}\cdot\mathbf{e}_{ms,\mathbf{g}}\mathbf{q}\cdot\mathbf{e}_{m's',\mathbf{g}'}$$

$$\times \quad \langle e^{i[\mathbf{g}\cdot\mathbf{R}_n+\mathbf{g}'\cdot\mathbf{R}_{n'}-(\omega_{s,\mathbf{g}}+\omega_{s',\mathbf{g}'})t]}\rangle_t \tag{A.86}$$

$$= \quad \sum_{\mathbf{g},\mathbf{g}'}\sum_{s,s'} a_{s,\mathbf{g}}a_{s',\mathbf{g}'}\mathbf{q}\cdot\mathbf{e}_{ms,\mathbf{g}}\mathbf{q}\cdot\mathbf{e}_{m's',\mathbf{g}'}$$

$$\times \quad e^{i(\mathbf{g}\cdot\mathbf{R}_n+\mathbf{g}'\cdot\mathbf{R}_{n'})}\delta_{\mathbf{g},-\mathbf{g}'}\delta_{s,s'} \tag{A.87}$$

$$= \quad \mathbf{q}\cdot\sum_{\mathbf{g},s}|a_{s,\mathbf{g}}|^2\mathbf{e}_{ms,\mathbf{g}}\mathbf{e}_{m's,-\mathbf{g}}e^{i\mathbf{g}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})}\cdot\mathbf{q} \tag{A.88}$$

The expression for the scattered intensity becomes

$$\begin{aligned}\langle I(\mathbf{q},t)\rangle_t = & \sum_{m,m'} Q_m Q_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m-\mathbf{r}_{m'})} \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})} \\ \times & \; e^{-\frac{1}{2}\mathbf{q}\cdot(\mathsf{V}_m+\mathsf{V}_{m'})\cdot\mathbf{q}} \\ \times & \; e^{\mathbf{q}\cdot\left[\sum_{\mathbf{g},s}|a_{s,\mathbf{g}}|^2\mathbf{e}_{ms,\mathbf{g}}\mathbf{e}_{m's,-\mathbf{g}}e^{i\mathbf{g}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})}\right]\cdot\mathbf{q}}. \end{aligned} \tag{A.89}$$

Expanding the exponential yields

$$e^{\mathbf{q}\cdot\left[\sum_{\mathbf{g},s}|a_{s,\mathbf{g}}|^2\mathbf{e}_{ms,\mathbf{g}}\mathbf{e}_{m's,-\mathbf{g}}e^{i\mathbf{g}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})}\right]\cdot\mathbf{q}} \approx \\ 1+\mathbf{q}\cdot\left[\sum_{\mathbf{g},s}|a_{s,\mathbf{g}}|^2\mathbf{e}_{ms,\mathbf{g}}\mathbf{e}_{m's,-\mathbf{g}}e^{i\mathbf{g}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})}\right]\cdot\mathbf{q}, \tag{A.90}$$

so that two additive contributions I_B and I_D can be identified, such that

$$I = I_B + I_D, \tag{A.91}$$

where

$$\begin{aligned} I_B = & \sum_{m,m'} Q_m Q_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m-\mathbf{r}_{m'})} \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})} \\ \times & \; e^{-\frac{1}{2}\mathbf{q}\cdot(\mathsf{V}_m+\mathsf{V}_{m'})\cdot\mathbf{q}} && \text{(A.92)} \\ = & \; e^{-\frac{1}{2}\mathbf{q}\cdot(\mathsf{V}_m+\mathsf{V}_{m'})\cdot\mathbf{q}} I_0(\mathbf{q}), && \text{(A.93)} \end{aligned}$$

where $I_0(\mathbf{q})$ has been defined as

$$I_0(\mathbf{q}) = \sum_{m,m'} Q_m Q_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m-\mathbf{r}_{m'})} \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})}, \tag{A.94}$$

and

$$\begin{aligned} I_D = & \sum_{m,m'} Q_m Q_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m - \mathbf{r}_{m'})} \sum_{n,n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})} \\ \times & \; e^{-\frac{1}{2}\mathbf{q}\cdot(\mathsf{V}_m + \mathsf{V}_{m'})\cdot\mathbf{q}} \\ \times & \; \mathbf{q}\cdot\left[\sum_{\mathbf{g},s} |a_{s,\mathbf{g}}|^2 \mathbf{e}_{ms,\mathbf{g}} \mathbf{e}_{m's,-\mathbf{g}} e^{i\mathbf{g}\cdot(\mathbf{R}_n - \mathbf{R}_{n'})}\right]\cdot\mathbf{q} \qquad \text{(A.95)} \\ = & \sum_{\mathbf{g},s}\sum_{n,n'} e^{i(\mathbf{q}+\mathbf{g})\cdot(\mathbf{R}_n - \mathbf{R}_{n'})} \sum_{m,m'} Q_m Q_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m - \mathbf{r}_{m'})} \\ \times & \; e^{-\frac{1}{2}\mathbf{q}\cdot(\mathsf{V}_m + \mathsf{V}_{m'})\cdot\mathbf{q}} \\ \times & \; \mathbf{q}\cdot\left[|a_{s,\mathbf{g}}|^2 \mathbf{e}_{ms,\mathbf{g}} \mathbf{e}_{m's,-\mathbf{g}}\right]\cdot\mathbf{q} \qquad \text{(A.96)} \end{aligned}$$

In order to demonstrate the connection with the liquid-like motions model, it is necessary to make some simplifying assumptions. The first is that all of the atomic variances V_m are the same, so that

$$\mathsf{V}_m = \mathsf{V}_{m'} = \mathsf{V}. \qquad \text{(A.97)}$$

The diffuse term I_D in the normal modes model then becomes

$$\begin{aligned} I_D = & \; e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}} \sum_{\mathbf{g},s}\sum_{m,m'} Q_m Q_{m'} e^{i\mathbf{q}\cdot(\mathbf{r}_m - \mathbf{r}_{m'})} \sum_{n,n'} e^{i(\mathbf{q}+\mathbf{g})\cdot(\mathbf{R}_n - \mathbf{R}_{n'})} \\ \times & \; \mathbf{q}\cdot\left[|a_{s,\mathbf{g}}|^2 \mathbf{e}_{ms,\mathbf{g}} \mathbf{e}_{m's,-\mathbf{g}}\right]\cdot\mathbf{q}. \qquad \text{(A.98)} \end{aligned}$$

The next is that the polarization vectors $\mathbf{e}_{ms,\mathbf{g}}$ all have the form

$$\mathbf{e}_{ms,\mathbf{g}} = \mathbf{e}_{s,\mathbf{g}} e^{i\mathbf{g}\cdot\mathbf{r}_m}, \qquad \text{(A.99)}$$

which would be true for an elastically homogeneous crystal[3], where waves continuously

---

[3] The opposite extreme would be where the entire unit cell is uniformly displaced as a rigid unit,

modulate atomic displacements within the unit cell. This would imply

$$\begin{aligned} I_D = \quad & e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}} \sum_{\mathbf{g},s} \mathbf{q} \cdot \left[|a_{s,\mathbf{g}}|^2 \mathbf{e}_{s,\mathbf{g}} \mathbf{e}_{s,-\mathbf{g}} \right] \cdot \mathbf{q} \\ \times \quad & \sum_{m,m'} Q_m Q_{m'} e^{i(\mathbf{q}+\mathbf{g})\cdot(\mathbf{r}_m - \mathbf{r}_{m'})} \sum_{n,n'} e^{i(\mathbf{q}+\mathbf{g})\cdot(\mathbf{R}_n - \mathbf{R}_{n'})}. \end{aligned} \tag{A.100}$$

In this, one can recognize

$$I_0(\mathbf{q}+\mathbf{g}) = \sum_{m,m'} Q_m Q_{m'} e^{i(\mathbf{q}+\mathbf{g})\cdot(\mathbf{r}_m - \mathbf{r}_{m'})} \sum_{n,n'} e^{i(\mathbf{q}+\mathbf{g})\cdot(\mathbf{R}_n - \mathbf{R}_{n'})}, \tag{A.101}$$

so that

$$I_D = e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}} \sum_{\mathbf{g}} I_0(\mathbf{q}+\mathbf{g}) \sum_s \mathbf{q} \cdot \left[|a_{s,\mathbf{g}}|^2 \mathbf{e}_{s,\mathbf{g}} \mathbf{e}_{s,-\mathbf{g}} \right] \cdot \mathbf{q}. \tag{A.102}$$

Converting the sum over modes $\mathbf{g}$ into an integral,

$$\begin{aligned} I_D = \quad & e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}} \int d^3\mathbf{g}\, n(\mathbf{g}) I_0(\mathbf{q}+\mathbf{g}) \sum_s \mathbf{q} \cdot \left[|a_s(\mathbf{g})|^2 \mathbf{e}_s(\mathbf{g}) \mathbf{e}_s(-\mathbf{g}) \right] \cdot \mathbf{q} \\ = \quad & e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}} \int d^3\mathbf{g}\, I_0(\mathbf{q}+\mathbf{g}) \mathbf{q} \cdot \mathsf{c}(\mathbf{g}) \cdot \mathbf{q} \\ = \quad & e^{-\mathbf{q}\cdot\mathsf{V}\cdot\mathbf{q}} \mathbf{q} \cdot [I_0(\mathbf{q}) * \mathsf{c}(\mathbf{q})] \cdot \mathbf{q}, \end{aligned} \tag{A.103}$$

which is the same as equation 3.112, where $\mathsf{c}(\mathbf{g})$ has been defined as

$$\mathsf{c}(\mathbf{g}) \equiv n(\mathbf{g}) \sum_s |a_s(\mathbf{g})|^2 \mathbf{e}_s(\mathbf{g}) \mathbf{e}_s(-\mathbf{g}), \tag{A.104}$$

and the density of states $n(\mathbf{g})$ has been introduced. Thus, the liquid-like correlated motions model is seen to be equivalent to a limiting case of the crystalline normal modes model, where the displacements of all atoms are equivalently distributed, and where the elastic properties of the crystal are homogeneous.

---

leading to $\mathbf{e}_{ms}(\mathbf{g}) = \mathbf{e}_s(\mathbf{g})$.

Bibliography

[1] H. Frauenfelder and G. Petsko, Biophys J **32**, 465 (1980).

[2] G. Careri, P. Fasella, and E. Gratton, Ann Rev Biophys Bioeng **8**, 69 (1979).

[3] J. R. Lakowicz and G. Weber, Biochemistry **12**, 4171 (1973).

[4] S. W. Englander, N. W. Downer, and H. Teitelbaum, Ann Rev Biochem **41**, 903 (1972).

[5] C. K. Woodward and B. D. Hilton, Ann Rev Biophys Bioeng **8**, 99 (1979).

[6] R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder, and I. C. Gunsalus, Biochemistry **14**, 5355 (1975).

[7] A. Cooper, Proc Natl Acad Sci USA **73**, 2740 (1976).

[8] M. R. Eftnik and C. A. Ghiron, Proc Natl Acad Sci USA **72**, 3290 (1975).

[9] A. Grinvald and I. Z. Steinberg, Biochemistry **13**, 5170 (1974).

[10] M. L. Saviotti and W. C. Galley, Proc Natl Acad Sci USA **71**, 4154 (1974).

[11] A. Allerhand et al., J Am Chem Soc **93**, 544 (1971).

[12] O. Oster, G. W. Neireiter, A. O. Clouse, and F. R. N. Gurd, J Biol Chem **250**, 7990 (1975).

[13] I. D. Campbell, C. M. Dobson, and R. J. P. Williams, Proc R Soc London Ser B **189**, 503 (1975).

[14] H. Frauenfelder, G. A. Petsko, and D. Tsernoglou, Nature **280**, 558 (1979).

[15] M. J. E. Sternberg, D. E. P. Grace, and D. C. Phillips, J. Mol. Biol. **130**, 231 (1979).

[16] J. Kuriyan and W. I. Weis, Proc. Natl. Acad. Sci. USA **88**, 2773 (1991).

[17] G. N. Phillips Jr., J. P. Fillers, and C. Cohen, Biophys. J. **32**, 485 (1980).

[18] J. Doucet and J. P. Benoit, Nature **325**, 643 (1987).

[19] D. L. D. Caspar, J. Clarage, D. M. Salunke, and M. Clarage, Nature **332**, 659 (1988).

[20] D. L. D. Caspar and J. Badger, Curr. Opin. Struct. Biol. **1**, 877 (1991).

[21] I. D. Glover, G. W. Harris, J. R. Helliwell, and D. S. Moss, Acta Cryst. **B47**, 960 (1991).

[22] J. B. Clarage, M. S. Clarage, W. C. Phillips, R. M. Sweet, and D. L. D. Caspar, Proteins **12**, 145 (1992).

[23] A. R. Kolatkar, J. B. Clarage, and G. N. Phillips Jr., The Rigaku Journal **9**, 4 (1992).

[24] P. Fauré et al., Struct. Biol. **1**, 124 (1994).

[25] P. W. Tucker, E. E. Hazen Jr., and F. A. Cotton, Mol. and Cell. Biochem. **22**, 67 (1978).

[26] F. A. Cotton, E. E. Hazen Jr., and M. J. Legg, Proc. Natl. Acad. Sci. USA **76**, 2551 (1979).

[27] P. A. Loll and E. E. Lattman, Proteins **5**, 183 (1989).

[28] A. Arnone et al., J. Biol. Chem. **245**, 2302 (1971).

[29] T. R. Hynes and R. O. Fox, Proteins **10**, 92 (1991).

[30] C. A. Royer et al., Biochemistry **32**, 5222 (1993).

[31] M. Levitt, C. Sander, and P. Stern, J Mol Biol **181**, 423 (1985).

[32] N. Go, T. Noguti, and T. Nishikawa, Proc Natl Acad Sci USA **80**, 3696 (1983).

[33] A. Dwivedi, T. Pederson, and P. G. Debrunner, Proc Mössbauer conf, Kyoto, 1978.

[34] V. Schomaker and K. N. Trueblood, Acta Cryst **B24**, 63 (1968).

[35] S. Chacko and G. N. Phillips Jr., Biophys. J. **61**, 1256 (1992).

[36] J. P. Benoit and J. Doucet, Quarterly Reviews of Biophysics **28**, 131 (1995).

[37] M. W. Tate et al., J. Appl. Cryst. **28**, 196 (1995).

[38] C. Giacovazzo et al., *Fundamentals of Crystallography*, Oxford University Press, Oxford, England, 1992.

[39] P. B. Stetson, *User's Manual for DAOPHOT II: The Next Generation*, Dominion Astrophysical Observatory, Herzberg Institute of Astrophysics, Victoria, British Columbia, 1995.

[40] A. Guinier, *X-ray Diffraction*, W. H. Freeman and Company, San Francisco, 1963.

[41] R. W. James, *The Optical Principles of the Diffraction of X-Rays*, Ox Bow Press, Woodbridge, Connecticut, 1962.

[42] B. E. Warren, *X-ray Diffraction*, Dover Publications, INC., New York, 1963.

[43] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Holt, Rinehart and Winston, Inc., Orlando, 1976.

[44] A. Brünger, *X-PLOR: A System for X-ray Crystallography*, Yale University Press, New Haven and London, 3.1 edition, 1987.

[45] University of Alabama at Birmingham, Birmingham, AL, *Ribbons 2.0 Manual*, 1994.