CHAPTER ONE

Introduction to diffuse scattering and data collection

Xiaokun Pei[a,1], Neti Bhatt[b,1], Haoyue Wang[c], Nozomi Ando[a,b,c,*], and Steve P. Meisburger[d,*]

[a]Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY, United States
[b]Department of Physics, Cornell University, Ithaca, NY, United States
[c]Graduate Field of Biophysics, Cornell University, Ithaca, NY, United States
[d]Cornell High Energy Synchrotron Source, Cornell University, Ithaca, NY, United States
*Corresponding authors. e-mail address: nozomi.ando@cornell.edu; spm82@cornell.edu

Contents

1. Introduction 2
2. Theory 4
3. Samples for diffuse scattering 10
   3.1 Crystallization and transport 16
   3.2 Sample mounting 17
4. Experimental design 19
   4.1 Beamline parameters 19
   4.2 Data collection strategy 23
   4.3 Background measurement 26
5. Data quality assessment 28
   5.1 Examination of diffraction images 28
   5.2 Quantifying and controlling radiation damage 29
   5.3 Diffuse map quality assessment 33
6. Conclusions 36
Acknowledgments 38
References 38

Abstract

A long-standing goal in X-ray crystallography has been to extract information about the collective motions of proteins from diffuse scattering: the weak, textured signal that is found in the background of diffraction images. In the past few years, the field of macromolecular diffuse scattering has seen dramatic progress, and many of the past challenges in measurement and interpretation are now considered tractable. However, the concept of diffuse scattering is still new to many researchers, and a general set of procedures needed to collect a high-quality dataset has never been described in detail. Here, we provide the first guidelines for performing diffuse scattering experiments, which can be performed at any macromolecular crystallography beamline that supports

[1] Equal contribution.

Methods in Enzymology, Volume 688
ISSN 0076-6879, https://doi.org/10.1016/bs.mie.2023.07.007

room-temperature studies with a direct detector. We begin with a brief introduction to the theory of diffuse scattering and then walk the reader through the decision-making processes involved in preparing for and conducting a successful diffuse scattering experiment. Finally, we define quality metrics and describe ways to assess data quality both at the beamline and at home. Data obtained in this way can be processed independently by crystallographic software and diffuse scattering software to produce both a crystal structure, which represents the average atomic coordinates, and a three-dimensional diffuse scattering map that can then be interpreted in terms of models for protein motions.

1. Introduction

An important goal of structural biology is to relate protein structure and function, but this goal cannot be achieved without an understanding of how proteins move. In particular, the collective structural fluctuations of a protein are of fundamental importance for activity, yet they are very challenging to measure directly. Thus far, correlated protein motions have primarily been studied by nuclear magnetic resonance (NMR) spectroscopy, such as by correlating timescales of exchange rates (Lisi & Loria, 2016; Palmer, 2015). Structural techniques including X-ray crystallography and cryo-electron microscopy have been used to infer such motions by fitting ensemble models to electron density (Fraser et al., 2009; Fraser et al., 2011), analyzing difference distance matrices between multiple structures (Schneider, 2000), or analyzing variability among particle images (Zhong, Bepler, Berger, & Davis, 2021). However, only a subset of motions can be inferred in each case, either because of incomplete sampling of snapshots or because correlated motions whose amplitudes are smaller than the imaging resolution cannot be unambiguously determined. On the other hand, in crystallography, any deviation from a perfect crystal leads to a continuous signal that appears in the background of diffraction patterns, known as diffuse scattering (Fig. 1A). Crystal disorders, whether static (displacive, substitutive, etc.) or dynamic (atomic vibrations and correlated motion), produce distinct diffuse scattering patterns (Welberry & Weber, 2016). Thus, by combining diffuse scattering analysis with crystallographic structure determination, it is possible to simultaneously obtain a high-resolution structure and understand how atomic motions within the structure are correlated (Xu, Meisburger, & Ando, 2021).

However, it was only very recently that methods to process and analyze macromolecular diffuse scattering became available (Chapter 2: Meisburger & Ando, in press). In this chapter, we provide the first guidelines for

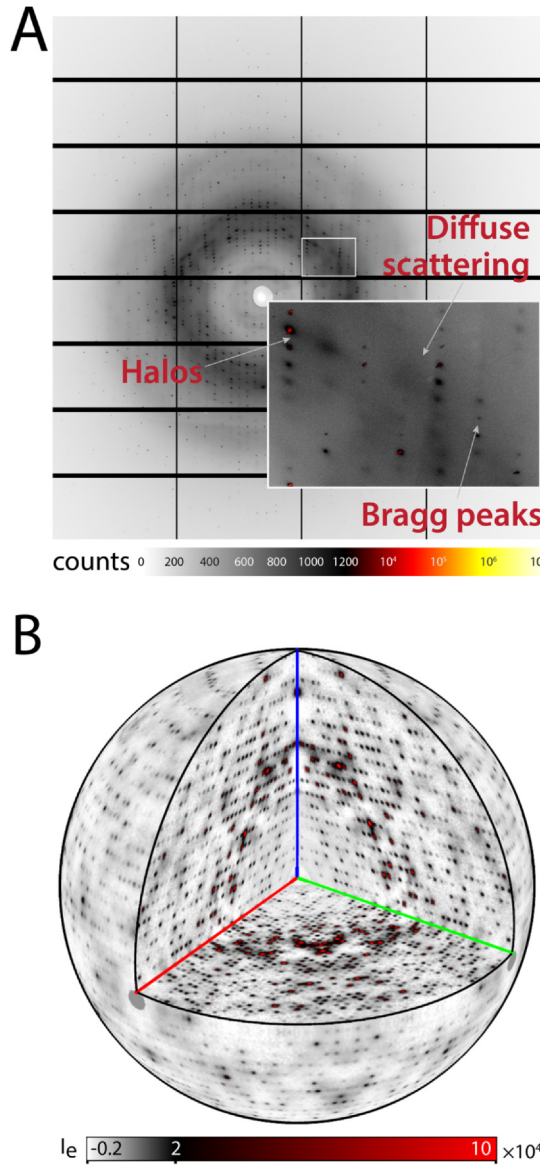

Fig. 1 Appearance of diffuse scattering in diffraction data from tetragonal lysozyme at room temperature. (A) To make the diffuse features clearly visible, a long still exposure was taken of a large crystal. The diffraction image shows Bragg peaks and diffuse scattering features including halos from lattice dynamics and the cloudy diffuse arising from correlated motions. (B) A three-dimensional diffuse scattering map reconstructed from a low-dose, high-redundancy diffraction dataset (Meisburger et al. 2023; reproduced with permission) showing both cloudy and halo features. Bragg peaks have been masked out, and the intensity is shown in absolute electron units (I_e) per asymmetric unit (ASU). A map of this quality is suitable for interpreting lattice dynamics and protein motion.

collecting diffuse scattering data of sufficient quality to generate a diffuse map for quantitative analysis of correlated motion (Fig. 1B). We begin by briefly introducing the theory of diffuse scattering. Following the theory section, we provide suggestions on how to conduct a successful diffuse scattering experiment at room temperature to best capture protein's conformational landscape at a physiologically relevant condition (Fraser et al., 2009, 2011). Although diffuse scattering can be measured at any macromolecular crystallography beamline in principle, special considerations are required in choosing and handling crystals, setting up the measurement, and minimizing excess background scattering and radiation damage. These considerations arise from the fact that diffuse scattering signals are extremely weak (orders of magnitude weaker than diffraction peaks on a per-pixel basis) and they are easily contaminated by the intense diffraction and background scattering. We note that the exact data collection setting will depend on the sample and the beamline. Therefore, we also discuss practical guidelines and potential alternatives for collecting data at various beamlines on different types of samples. Finally, we end by discussing preliminary steps in data processing, which can be followed by more rigorous methods (Chapter 2: Meisburger & Ando, in press) and interpreted by various models (Chapters 3–6: Case, in press; Peck et al., submitted; Wych & Wall, in press a; Wych & Wall, in press b).

2. Theory

A solid grasp of X-ray diffraction theory is important in diffuse scattering research. It is needed to understand the scattering features in detector images to enable better measurements, and for building appropriate models to interpret the data. The theory of macromolecular diffuse scattering has been reviewed recently (Meisburger, Thomas, Watkins, & Ando, 2017). Here, we provide a general overview and emphasize key results relevant to measurement and interpretation.

A typical macromolecular crystallography (MX) experiment consists of placing a protein crystal in an X-ray beam and measuring scattered X-rays using an area detector. The scattering results from the interaction between the X-rays and the electrons in the sample. Protein crystals are made up of repeating structural units, or unit cells. X-ray diffraction from these repeating unit cells causes constructive interference in certain directions, creating intense spots on the detector known as Bragg peaks. The positions

of the Bragg peaks on the detector provide information on the size and shape of the unit cells, while the intensities of the Bragg peaks contain the information about the electron density within the unit cell. In conventional MX, these Bragg peak intensities are analyzed to reconstruct the arrangement of atoms in the unit cell: the protein's structure.

Bragg peaks are not the only features observed in X-ray diffraction images from protein crystals. The images also contain diffuse scattering, a continuous signal between and underneath the Bragg peaks. If all unit cells were identical (and the experimental setup perfect), Bragg peaks would be the only features observed. Diffuse scattering occurs whenever disorder is present, and protein crystals contain significant disorder. For instance, protein crystals have disordered solvent occupying space between protein molecules, defects in crystal packing, and intrinsic heterogeneity in terms of protein conformation, even at cryogenic temperatures. More interestingly, diffuse scattering has a distinctive pattern depending on how the disorder is correlated over distance: long-ranged disorders such as lattice distortions produce very different diffuse scattering from short-ranged disorders such as the thermal vibrations of atoms.

The scattering observed in diffraction images is not limited to diffuse scattering from the protein and solvent content of the crystal; it also includes experimental background. The largest source of background at MX beamlines is typically scattering from air and other materials in the beam path, such as sample mounting materials and excess solvent surrounding the crystal. Atoms can also fluoresce or scatter X-rays inelastically (Compton scattering). In accord with previous work (Meisburger et al., 2017; Meisburger, Case, & Ando, 2020), the diffuse scattering of interest is defined as the elastic component of the non-Bragg signal from all atoms in the crystal (i.e., the protein, ligands/ions, plus ordered/disordered solvent). Here we consider only the theory of elastic scattering for an infinitely large crystal, assuming that other contributions have been measured or otherwise corrected (see Section 4.3 and Chapter 2: Meisburger & Ando, in press).

The elastic scattering process that leads to Bragg diffraction and diffuse scattering is coherent. The scattered wave in any particular direction $\hat{\mathbf{s}}_1$ has an amplitude that is proportional to the structure factor, F, of the illuminated atoms. For X-rays, the structure factor is simply the Fourier transform of the electron density evaluated at a point in reciprocal space $\mathbf{q} = (2\pi/\lambda)(\hat{\mathbf{s}}_1 - \hat{\mathbf{s}}_0)$, where $\hat{\mathbf{s}}_0$ is the incident beam direction and λ is the X-ray wavelength. What is measured is not the X-ray scattering amplitude but rather its intensity, or the absolute square of the amplitude.

The intensity recorded by a detector, after correcting for background, geometric distortions, and other effects (see Chapter 2: Meisburger & Ando, in press) can be written as follows:

$$I_{\text{total}}(\mathbf{q}) = \langle |F_{\text{crystal}}(\mathbf{q})|^2 \rangle_{\text{time}}, \tag{1}$$

where F_{crystal} is the structure factor of the illuminated crystal volume, and the angle brackets denote the time average during the exposure. From the mathematics of Fourier transforms, the structure factor of a crystal will include a lattice of peaks, the reciprocal lattice, which is related to the lattice in real space formed by the repetitive arrangement of unit cells. These peaks correspond to the Bragg diffraction, and for convenience, they can be considered separately as follows:

$$I_{\text{total}} = I_{\text{Bragg}} + I_{\text{diffuse}}. \tag{2}$$

It can be shown using properties of Fourier transforms that I_{Bragg} depends only on the average structure factor of the unit cells (and thus the average electron density). Let F_{crystal} be written as a sum over structure factors of individual unit cells F_n at lattice positions $\mathbf{R}_n$ as follows:

$$F_{\text{crystal}}(\mathbf{q}) = \sum_n F_n(\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{R}_n}. \tag{3}$$

Then, the Bragg intensity is

$$I_{\text{Bragg}}(\mathbf{q}) = \left\langle |\overline{F_{\text{cell}}(\mathbf{q})}|^2 \right\rangle_{\text{time}} \left| \sum_n e^{i\mathbf{q}\cdot\mathbf{R}_n} \right|^2, \tag{4}$$

where $\overline{F_{\text{cell}}(\mathbf{q})}$ is the average unit cell structure factor. The first factor is a continuous function, the transform of the average unit cell, and this is multiplied by the second factor, which is a lattice of sharp peaks (delta functions in infinite crystal limit). Thus, the Bragg intensity is said to "sample" the unit cell transform at the reciprocal lattice points with integer Miller indices h, k, and l.

A similar expression for the diffuse intensity can be obtained by algebraic manipulation of the preceding equations:

$$I_{\text{diffuse}}(\mathbf{q}) = \left\langle \left| \sum_n e^{i\mathbf{q}\cdot\mathbf{R}_n} (F_n(\mathbf{q}) - F_{\text{cell}}(\mathbf{q})) \right|^2 \right\rangle_{\text{time}}, \tag{5}$$

As this equation illustrates, the diffuse intensity depends on the difference between the instantaneous unit cell structure factors and the average, or equivalently, the instantaneous and average electron density. When these are identical (as in the case of a perfect crystal), no diffuse scattering is observed. Any source of disorder produces diffuse scattering.

In biological applications of diffuse scattering, disorder resulting from the motion of atoms in the protein is of particular interest. Such motions affect both the Bragg and diffuse intensities, and must be considered in models used for structure refinement against Bragg data. Typically, each atom in the model is assigned a Gaussian probability distribution describing its position in the unit cell. Neglecting disordered solvent for simplicity, the average unit cell structure factor is modeled as follows:

$$\overline{F_{\text{cell}}(\mathbf{q})} = \sum_j f_j(\mathbf{q})\, T_j(\mathbf{q})\, e^{i\mathbf{q}\cdot\bar{\mathbf{r}}_j}, \tag{6}$$

where the sum runs over all atoms in the unit cell, $\bar{\mathbf{r}}_j$ is the mean atomic position, f_j the atomic scattering factor, and T_j is the Debye–Waller factor that depends on disorder. The Debye–Waller factor is the Fourier transform of the Gaussian distribution of atomic displacements from the mean position, $\mathbf{u} \equiv \mathbf{r} - \bar{\mathbf{r}}$, as follows:

$$T_j(\mathbf{q}) = e^{-(1/2)\mathbf{q}\cdot\mathbf{U}_j\mathbf{q}}, \tag{7}$$

where $\mathbf{U}$ is the symmetric matrix,

$$\mathbf{U} = \langle \mathbf{u}\mathbf{u}^{\mathrm{T}} \rangle = \begin{pmatrix} \langle u_x u_x \rangle & \langle u_x u_y \rangle & \langle u_x u_z \rangle \\ \langle u_y u_x \rangle & \langle u_y u_y \rangle & \langle u_y u_z \rangle \\ \langle u_z u_x \rangle & \langle u_z u_y \rangle & \langle u_z u_z \rangle \end{pmatrix}. \tag{8}$$

The six unique elements of $\mathbf{U}$ are known as Atomic Displacement Parameters (ADPs). Often, ADPs are reduced to an isotropic B-factor:

$$B = (8\pi^2/3)\,\mathrm{trace}(\mathbf{U}). \tag{9}$$

Although ADPs or B-factors contain information about the displacement of an atom, they don't provide insight into the type of motion that is causing the displacement: many different types of motion potentially generate similar patterns of B-factor variation across the protein (Meisburger et al., 2017; Xu et al., 2021).

The diffuse scattering intensity generated by atoms with Gaussian-distributed displacements is as follows:

$$I_{\text{diffuse}}(\mathbf{q}) = \sum_{n}\sum_{n'} e^{i\mathbf{q}\cdot(\mathbf{R}_n-\mathbf{R}_{n'})}\sum_{j}\sum_{j'} f_j(\mathbf{q})f_{j'}(\mathbf{q})\, e^{i\mathbf{q}\cdot(\bar{\mathbf{r}}_j-\bar{\mathbf{r}}_{j'})} T_j(\mathbf{q})\, T_{j'}(\mathbf{q})\left[e^{(1/2)\mathbf{q}\cdot\mathbf{V}_{njn'j'}\mathbf{q}}-1\right], \tag{10}$$

where the sums run over pairs of atoms j and j' in unit cells n and n' respectively. Although the equation for the diffuse intensity is longer to write than its Bragg counterpart, most of the terms are identical. In fact, the only additional quantity is the matrix $\mathbf{V}$. The $\mathbf{V}$ matrix elements are similar to ADPs, except that they describe the displacement *covariances* between pairs of atoms in the crystal, as follows:

$$\mathbf{V}_{njn'j'} = \left\langle \mathbf{u}_{nj}\mathbf{u}_{n'j'}^{\mathrm{T}} \right\rangle + \left\langle \mathbf{u}_{n'j'}\mathbf{u}_{nj}^{\mathrm{T}} \right\rangle. \tag{11}$$

If a pair of atoms move in an uncorrelated fashion with respect to each other, the $\mathbf{V}$ matrix elements for that pair will be zero. Consequently, in Eq. (10) the factor in square brackets becomes zero, and that pair of atoms will not contribute to diffuse scattering (an atom's motion is always correlated with itself, and this produces diffuse scattering, however it can be predicted from the ADPs and thus carries no new information). If, on the other hand, the motions are correlated, then diffuse scattering is produced. Thus, the unique information present in the diffuse intensity is the manner in which atomic motions are correlated, and the problem of analyzing diffuse scattering data amounts to determining the $\mathbf{V}$ matrix elements from measured data. Unfortunately, Eq. (10) cannot be inverted because the number of parameters exceeds the amount of data collected (nine parameters for every pair of atoms). For example, the number of unique atom pairs in a single molecule of lysozyme is approximately 2×10^6. If we then consider all atom pairs between unit cells, this number becomes orders of magnitude larger than the number of unique observations even for a finely sampled diffuse map, such as that reported for triclinic lysozyme ($\sim 50 \times 10^6$) (Meisburger et al., 2020). In general it is necessary to reduce the complexity of Eq. (10) by applying a model.

From Eq. (10), one can imagine how different kinds of disorder are likely to appear in diffraction images. If all the disorder is caused by

protein motion such that displacements are not correlated between different unit cells, all the terms with $n \neq n'$ drop out, and the diffuse intensity becomes:

$$I_{\text{diffuse}}(\mathbf{q}) = N \sum_{j} \sum_{j'} f_j(\mathbf{q}) f_{j'}(\mathbf{q}) e^{i\mathbf{q}\cdot(\bar{\mathbf{r}}_j - \bar{\mathbf{r}}_{j'})} T_j(\mathbf{q}) T_{j'}(\mathbf{q}) \left[e^{(1/2)\,\mathbf{q}\cdot\mathbf{V}_{jj'}\mathbf{q}} - 1 \right], \tag{12}$$

where N is the number of unit cells. Unlike the equation for Bragg intensity (Eq. (4)), Eq. (12) has no dependence on the lattice vectors $\mathbf{R}$. Thus, the diffuse pattern will not have any particular relationship with the reciprocal lattice. The texture of the diffuse pattern is determined by the phase factors $e^{i\mathbf{q}\cdot(\bar{\mathbf{r}}_j - \bar{\mathbf{r}}_{j'})}$, which are functions of the inter-atomic vector. Because atom pairs in this case are within a unit cell, there is a maximum distance cutoff that, in turn, sets how rapidly the phase factors can oscillate as a function of $\mathbf{q}$. The result is a smoothly varying, or "cloudy" diffuse pattern on the scale of the reciprocal lattice.

On the other hand, if atomic displacements are correlated between unit cells, the lattice sums in Eq. (10) mean that diffuse scattering will be related to the reciprocal lattice. Certain models of lattice disorder, for instance, predict broad peaks (or "halos") that overlap with the Bragg peaks. In addition, the diffuse scattering features can be very sharp because the argument of the phase factor is no longer limited to short inter-atomic distances.

Diffuse scattering patterns from protein crystals contain both cloudy and halo features. From a modeling point of view, it is therefore necessary to account for both correlated motion within proteins, as well as lattice disorder, which produce the halo features. From an experimental point of view, the strategy is to measure the diffuse scattering at all values of $\mathbf{q}$ in reciprocal space (to obtain a complete, three-dimensional diffuse map) with sufficiently fine sampling to capture the halo features in addition to the cloudy pattern. Finally, we note that a high-quality structure determined from the Bragg data is a necessary starting point for diffuse scattering analysis. This is because most of the terms in Eq. (10) can be set using the Bragg data, and only the unknown covariance matrix elements need to be determined from the diffuse scattering.

3. Samples for diffuse scattering

In this section, we review best practices for room-temperature sample preparation and provide practical guidance for diffuse scattering based on our own experiences. We first describe considerations that go into sample requirements.

Low vs. high resolution diffraction. The conservation of total scattering (Eq. (2)) implies that crystals that diffract to low resolution will produce greater diffuse scattering. For example, crystals of *Thermosynechococcus vestitus* photosystem II that diffract to ~4.5 Å display strong diffuse scattering past the diffraction limit (Ayyer et al., 2016). However, crystals that diffract to low resolution often have problematic sources of disorder (see below). In contrast, crystals that have sufficient quality for conventional crystallography are generally useful for diffuse scattering studies because, as described above, most parameters needed to describe the diffuse scattering intensity (Eq. (10)) can be determined from the Bragg data. We therefore recommend at least starting with crystals that exhibit low mosaic disorder (see next section), and yield reproducibly high-resolution (~2.5 Å or less) structures at cryogenic temperature or room temperature.

Sources of disorder. The source of disorder that produces diffuse scattering should be considered. If diffuse scattering is observed in images from conventional structure determination, the crystal may be a promising candidate for further study. However, not all scattering is useful for understanding protein internal motions. "Messy" diffraction patterns that appear to have intense diffuse scattering often arise from high mosaicity and problematic forms of lattice disorder (Fig. 2D; detailed discussions in Section 4.1 and Xu et al., 2021), rather than from protein internal motion. These undesirable forms of disorder are commonly found in crystals that diffract to low resolution. Although total scattering analysis can be a valuable tool to characterize pathological crystals and in some cases to solve their structures (Lovelace & Borgstahl, 2020), pathological causes of diffuse scattering will only complicate the analysis of protein internal motion, and thus, such crystals are best excluded at the sample preparation stage. Likewise, other issues that are considered problematic in conventional crystallographic data processing, such as multiple lattices, twinning, and pseudo translation, should be avoided because Bragg diffraction and diffuse scattering will be difficult to separate from each other and the diffuse scattering from different crystalline domains will be averaged together.

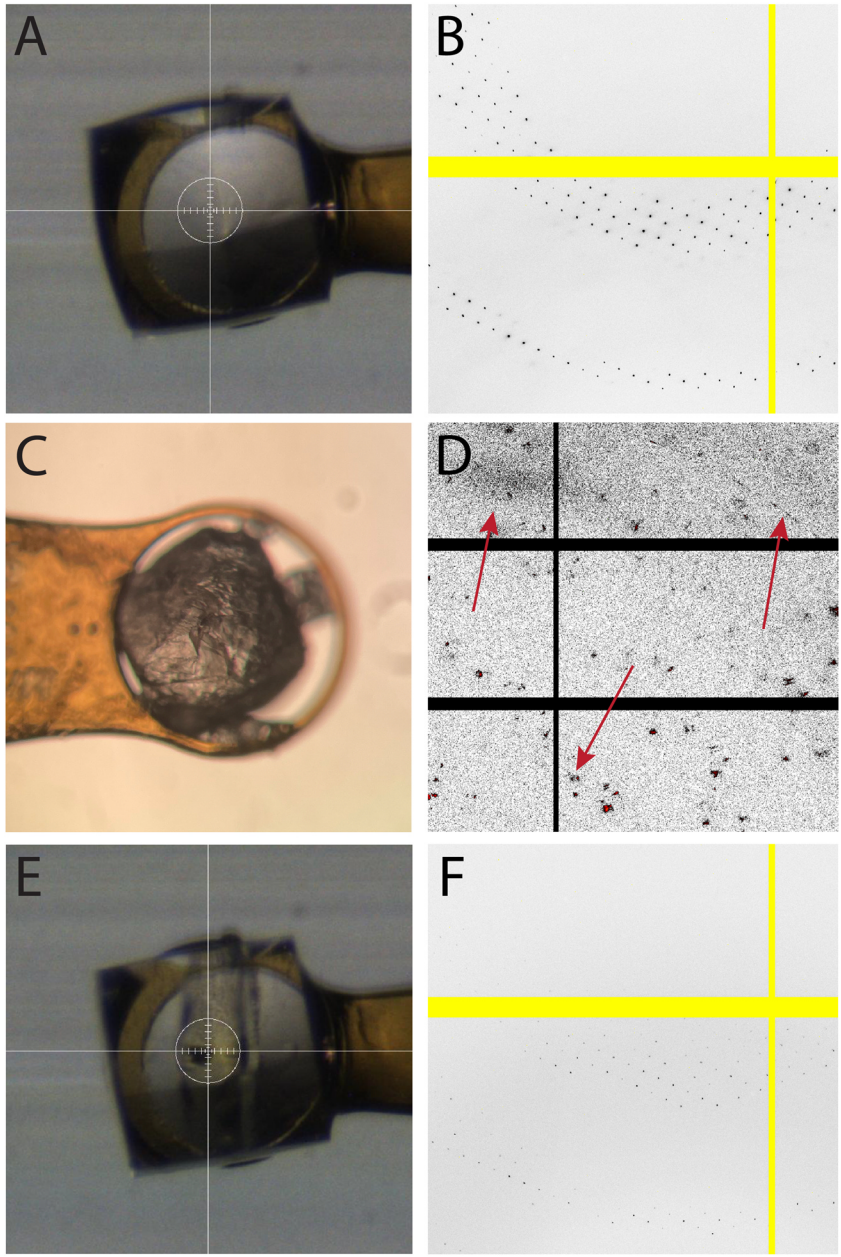

Fig. 2 Visual indicators of sample quality for diffuse scattering. (A) A large, high-quality crystal suitable for single-crystal diffuse scattering (tetragonal lysozyme). The circle in the center has a diameter of 100 µm. (B) Example X-ray diffraction from a crystal of similar quality to (A) showing sharp, well-separated Bragg peaks with no

(Continued)

Cryogenic vs. room temperature. The development of cryo-cooling methods was critical to the widespread adoption of macromolecular crystallography because the effects of radiation damage are greatly reduced at cryogenic temperatures and frozen crystals can be shipped much more easily to remote facilities. However, cryo-cooling is also known to introduce lattice strain and lead to higher mosaicity or non-isomorphism (Halle, 2004; Juers & Matthews, 2001; Kriminski, Caylor, Nonato, Finkelstein, & Thorne, 2002; Thompson, Cascio, & Yeates, 2018; Vahedi-Faridi, Lovelace, Bellamy, Snell, & Borgstahl, 2003). Thus, to avoid undesirable forms of disorder (see above), we advocate non-cryogenic collection of diffuse scattering data, at least initially.

Achieving high signal-to-noise and high redundancy. In conventional crystallography, high redundancy may be achieved by utilizing crystals in high-symmetry space groups. For diffuse scattering analysis, a lower symmetry form is preferable if there are multiple options. This is because orientational averaging from multiple copies of the asymmetric unit will tend to diminish the diffuse signal from internal protein motions (Chapman et al., 2017). Additionally, the final form of processed diffuse scattering data is a three-dimensional diffuse map that samples *all* of reciprocal space within the resolution limits of the crystal of interest. Thus, it is advantageous to have many large (e.g., >100 μm in at least one dimension, see crystal sizes in Table 1), high-quality crystals for measuring such a large amount of data points (i.e., reciprocal space voxels). The increased scattering per unit dose afforded by large crystals makes a significant difference in signal to noise. Moreover, because the diffuse signal competes with background from other sources that may be difficult to eliminate or subtract, the increased relative signal from large crystals can improve the overall data quality. Although a single large crystal may be sufficient for a complete diffuse map, using many crystals has the benefit of extra redundancy for

Fig. 2—Cont'd contaminating diffraction or scattering. (C) Looping the crystal slowly in an environment with low relative humidity quickly dries out the surrounding mother liquor and dehydrates the crystal, resulting in precipitation of buffer components and visible cracking/degradation of the crystal. (D) Data collection from crystals similar to (C) results in problematic features indicated by red arrows, including split Bragg peaks, powder diffraction rings, and possibly contaminating scattering from polymeric components. (E) Crystal in (A) after radiation damage from a long φ-scan. (F) Diffraction image of the pre-exposed crystal taken with the same orientation, beam intensity, and exposure time as (B). Radiation damage causes significant decay in the intensity of Bragg peaks and likely alters diffuse scattering.

Table 1 Examples of crystals and their diffuse scattering dataset measurement conditions.

	Lysozyme (triclinic)	Lysozyme (orthorhombic)	Lysozyme (tetragonal)	Sample A	Sample B
Reference	Meisburger et al. (2020)	Meisburger et al. (2023)	Meisburger et al. (2023)	**Unpublished**	**Unpublished**
Sample					
Crystallization method	Microbatch under oil	Sitting drop vapor diffusion	Hanging drop vapor diffusion	Hanging drop vapor diffusion	Hanging drop vapor diffusion (5 µL drop size)
Crystal size (approx.)	Various: 100–500 µm	$1000 \times 400 \times 200\ \mu m^3$	$500 \times 500 \times 250\ \mu m^3$	$500 \times 100 \times 50\ \mu m^3$	$400 \times 200 \times 100\ \mu m^3$
Mount	Loop in PET capillary	Loop in PET capillary	Loop in PET capillary	Loop in PET capillary	Loop in PET capillary
Unit cell volume	$26\ nm^3$	$127\ nm^3$	$243\ nm^3$	$278\ nm^3$	$2000\ nm^3$
Longest cell dimension	34.5 Å	73.9 Å	79.6 Å	85 Å	130 Å
Amino acids per unit cell	129	516	1032	1224	1400
Solvent content	33.8%	44.4%	41.9%	42%	60.8%
Mosaicity (apparent, sigma)	0.024°	0.03°	0.02°	0.043°	0.055°

(*continued*)

Table 1 Examples of crystals and their diffuse scattering dataset measurement conditions. (*cont'd*)

	Lysozyme (triclinic)	Lysozyme (orthorhombic)	Lysozyme (tetragonal)	Sample A	Sample B
Reference	Meisburger et al. (2020)	Meisburger et al. (2023)	Meisburger et al. (2023)	Unpublished	Unpublished
Beamline					
Location	CHESS F1	CHESS F1	CHESS F1	CHESS ID7B2	CHESS ID7B2
X-ray energy	12.693 keV	12.693 keV	12.693 keV	12.867 keV	12.867 keV
Bandwidth ($\Delta E/E$)	5×10^{-4}	5×10^{-4}	5×10^{-4}	10^{-4}	10^{-4}
Flux	2.5×10^{10}	2.5×10^{10}	2.5×10^{10}	1.7×10^{11}	1.7×10^{11}
Beam size	100 μm (round)	100 μm (round)	100 μm (round)	70×24 μm^2	70×24 μm^2
Beam profile	Top-hat	Top-hat	Top-hat	Gaussian	Gaussian
Detector model	Pilatus 3 6M	Pilatus 3 6M	Pilatus 3 6M	Eiger2 X 16M	Eiger2 X 16M
Sensor (thickness)	Si (1 mm)	Si (1 mm)	Si (1 mm)	Si (450 μm)	Si (450 μm)
Detector distance	230 mm	275 mm	250 mm	180 mm	180 mm

Data collection

Rotation per image	0.1°	0.1°	0.1°	0.1°	0.05°
Exposure per image	100 ms	100 ms	100 ms	12 ms	10 ms
Rotation rate	1°/s	1°/s	1°/s	8.3°/s	10°/s
Multi-motor scan	Sweep and translate	Sweep and translate	Sweep and translate	Vector: 500 μm/21 s	Vector: 400 μm/12 s
Rotation per sweep	50°	50°	50°	180°	60°
Number of sweeps (crystals)	11 (4)	8 (2)	8 (1)	1	1
Images per dataset	5500	4000	4000	1800	1200 (incomplete)

judging repeatability of the measurement (Su et al., 2021) and for reducing artifacts such as detector gaps and blind regions of reciprocal space. Merged datasets should be from isomorphous crystals of consistent quality. Because many crystallization trials may be needed for optimization, it is helpful if sample availability is not a limiting factor.

3.1 Crystallization and transport

It is important to establish a reproducible crystallization protocol when producing crystals for room-temperature experiments. Having good control over crystallization greatly facilitates optimization for producing large crystals, and it allows for samples to be prepared on demand prior to scheduled beamtime. It can be helpful to first map out the crystallization phase diagram and to understand the factors affecting crystal nucleation and growth rates, such as drop volumes and crystallization temperature. In cases where nucleation is difficult to control, seeding strategies can be used (Stura, 1999; Thaller et al., 1981).

The crystallization format should be considered carefully when producing crystals for data collection, especially if crystals will be transported to the synchrotron. It is important to note that crystallization format can influence growth in subtle ways. For instance, when vapor diffusion is used, the drop volume influences the rate at which the solution approaches supersaturation, the rate that protein is depleted from the drop once crystals nucleate and grow, and other factors. Thus, when switching between formats it may be necessary to re-optimize the growth conditions. For this reason, it is often wise to choose an appropriate format early in the project and stick with it. We have had good success with standard plastic 24-well hanging or sitting well trays, dialysis buttons, and 24- or 96-well sitting drop trays for micro-batch under paraffin oil. Strategies for optimizing growth conditions to produce large crystals have been reviewed in the neutron protein crystallography literature (Budayova-Spano, Koruza, & Fisher, 2020; Ng et al., 2015).

Shipping protein crystals at ambient temperature while maintaining diffraction quality is a challenge due to their sensitivity to humidity and temperature change. An ideal scenario would be to set up crystallization trays at the synchrotron. Another common solution is to transport lab-grown crystals in their crystallization trays. Ground transport (car or train) is preferred, because pressure changes during a flight make air travel problematic (Fischer, 2021). Vibrations should be minimized as much as possible. Securing trays in thick Styrofoam boxes can reduce such

vibrations as well as prevent any abrupt temperature changes. If crystals are grown at 4 degrees, an ice slurry may be used to maintain this temperature during transport. Even with these precautions, it can be difficult to control temperature well enough to prevent condensation within hanging drop vapor-diffusion setups, so other crystallization formats such as sitting well and microbatch under oil may be better suited.

Another approach is to transport protein crystals pre-mounted for data collection. The simplest method is to place the crystals in glass capillaries sealed with epoxy, however the high background scattering from glass is not ideal for diffuse scattering measurements. We commonly use plastic capillaries with Kapton loops for sample mounting (MiTeGen MicroRT system). Other groups have reported successfully shipping pre-mounted crystals in this system, with the extra precaution of sealing the capillary to the magnetic base with vacuum grease. The Structural Molecular Biology program at Stanford Synchrotron Radiation Lightsource recommends adding 0.6–1% agarose to the humidity-stabilizing droplet in MicroRT tubing in order to keep the droplet in place during long distance shipping. In-situ crystallization plates have also been developed, which can be directly mounted on a goniometer. Diffuse scattering from crystals in such devices has not been reported, and it is likely that the increased background scattering from excess mother liquor around the crystal as well as X-ray window materials will need careful consideration. Regardless of whether the crystals are pre-mounted or not, control of temperature during shipping is of the utmost importance. A well-insulated thermal shipper has been developed by Crystal Positioning Systems that can maintain the box temperature at 20 °C (± 5 °C) for 7 days.

3.2 Sample mounting

There are many factors to consider when choosing which crystals to loop for a diffuse scattering study. Ideally, crystals should be free of any visible imperfections, such as twinning or cracks. These features—when visible—are not trivial at the ångström scale and require many unnecessary correction steps in data processing. Tiny imperfections can be difficult to discern by visual inspection and might only be discovered when manipulating crystals under the microscope. Ultimately, crystal quality cannot be judged by eye, and the most reliable method is to check the quality of the Bragg diffraction (i.e., conventional crystallographic data statistics) during data collection. Once high-quality, well-diffracting crystals are found, 10–20 nominally identical crystals are commonly needed for a successful

study, especially given the need to establish optimal data collection parameters and the possibility of accidentally damaging samples during handling.

Looping crystals successfully takes practice. For experienced protein crystallographers, it may be second nature, but those new to the field should not underestimate the importance of learning excellent technique. It is important to avoid pressing, plunging, or pushing crystals as this may induce cracks or strain. A general method is to move the loop beneath the crystal then gently lift and glide out of the drop. The mechanical robustness of crystals can be very sample dependent, and different looping techniques may be required for different crystals.

After looping, the crystal must be maintained in an environment having a relative humidity as close as possible to its growth condition. With the MiTeGen MicroRT system, the mounted loop should be transferred immediately into a plastic capillary sleeve that has been pre-loaded with ~10 μL of well solution in the tip. The volume of this solution can be tuned, and the closer the liquid surface is to the crystal the better it preserves the humidity. When using capillary sleeves, it is important to check with the beamline scientist the maximum allowed length for the beamline geometry.

Some crystals are very sensitive to water loss and may degrade in the few seconds while being transferred to a capillary. An ideal solution is to loop crystals in a humid environment (Thorne, 2023) such as a cold room with a humidifier or a commercial sample-loading humid chamber. Alternatively, crystals can be transferred from mother liquor to oil to prevent dehydration, or looped after covering crystallization droplets with oil. Several types of oil have been successfully used for crystal mounting, and include polycarbon or modified polycarbon products (Bautis, 1975; Pflugrath, 2015) with a range of viscosity and vapor permeability. If the oil technique is used, several types should be screened to observe the effect on diffraction quality. Viscous oils such as type NVH immersion oil are useful for removing excess water (Warkentin & Thorne, 2009) and make an excellent vapor barrier, however more strain is introduced when looping crystals through it. For diffuse scattering experiments, it is also important to minimize the amount of oil on the surface of the crystal as it contributes background scattering. A common strategy is to quickly dip the looped crystals in oil and then wipe excess oil as much as possible by touching the loop on a glass surface under microscope. Finally, oil may not be a sufficient vapor barrier at room temperature. Oil may be combined with MicroRT tubing for a double layer of protection.

Finally, it is generally a good idea to visually inspect the crystal after looping and after data collection to assess whether additional imperfections

appear (Fig. 2). These steps help to build a qualitative understanding of the relationship between crystal appearance, crystal sensitivity to certain environments, and diffraction quality.

4. Experimental design

Measuring diffuse scattering signals accurately requires careful attention to the experimental setup and data collection strategy. Until recently, accommodating non-ideal properties of X-ray detectors was a primary concern (Meisburger & Ando, 2017). In addition to having slow frame rates, the charge-coupled device (CCD) detectors used in earlier studies suffered from read-out noise and other artifacts. Nevertheless, 3D diffuse scattering maps of protein crystals were successfully measured using CCDs in pioneering studies (Meisburger et al., 2017; Wall, Clarage, & Phillips, 1997; Wall, Ealick, & Gruner, 1997). The introduction of hybrid photon counting (HPC) detectors in the last decade has made it feasible to collect diffuse scattering under similar conditions to modern Bragg data collection with continuous crystal rotation (Benschoten et al., 2016). With detectors no longer limiting, there has been renewed attention to other aspects of the experimental setup that affect data quality and methods to more accurately subtract the background (Meisburger et al., 2020; Xu et al., 2021). In addition, algorithms for scaling and merging have been developed that take full advantage of the single photon-counting sensitivity of HPCs (Meisburger et al., 2020), and diffuse scattering can now be collected using the modern approach of fine φ-slicing with a low dose per frame and high redundancy (Meisburger, Case, & Ando, 2023). In this section, we discuss in detail how various parameters of an MX beamline can be chosen or optimized for diffuse scattering, considerations for data collection strategy, and methods to evaluate data quality while at the synchrotron. For concrete examples, we refer to three previously published datasets from model protein crystals (triclinic, orthorhombic, and tetragonal lysozyme) and two of our unpublished datasets from more complex proteins (summarized in Table 1).

4.1 Beamline parameters

Current MX beamlines have not been designed with diffuse scattering in mind. However, many of the same considerations that improve Bragg data quality are also needed for diffuse scattering, such as a well-collimated and

monochromatic beam, low-noise detector, and low background scattering. Thus, state-of the-art MX beamlines can be successfully used for diffuse scattering if they can accommodate room temperature data collection. When data of exceptional quality is needed, it can pay to choose a beamline with specific hardware and beam properties that will maximize data quality, and to collaborate with beamline staff to fully optimize the setup. Even if an "off the shelf" MX facility is used, certain beamline parameters must be chosen by the user during data collection, such as X-ray energy, beam size, and detector distance, and thus it is important to understand the likely effect of those choices on data quality.

As mentioned above, a high-quality X-ray detector is perhaps the most important beamline component for diffuse scattering measurements. HPCs such as EIGER 16M and PILATUS 6M are now widely available at MX facilities and provide the essential properties needed: single photon sensitivity, high dynamic range, fast frame rates, and single-pixel point spread function (Förster, Brandstetter, & Schulze-Briese, 2019). For facilities with very high instantaneous flux, such as X-ray free electron lasers (XFELs) and fourth-generation synchrotrons, charge-integrating hybrid pixel array detectors such as JUNGFRAU are now available with sufficiently low noise to be useful for diffuse scattering (Leonarski et al., 2018). Small pixels are useful in diffuse scattering because they allow for finely sampled maps, however they are of little advantage if the features in the diffraction pattern are intrinsically more than one pixel wide, for instance if the diffraction pattern is smeared by a mosaic crystal or a divergent beam.

In a diffuse scattering experiment, one of the primary challenges is to cleanly separate Bragg peaks from diffuse background. When the Bragg peaks are few and far between, this is relatively straightforward, and that is why early experiments focused on protein crystals with small unit cells. When crystals with larger unit cells are investigated, the demands on both crystal quality and measurement increase. The size and number of Bragg peaks appearing in a diffraction image is fundamentally limited by the crystal's mosaicity, a measure of angular broadening of the diffraction pattern due to slight misalignment of microscopic crystalline domains. Divergence of the X-ray beam produces a similar angular broadening, and often data processing software will report an "apparent" mosaicity, which is the combination of true mosaicity and other factors (Fig. 3) (Dauter, 1999; Snell, Bellamy, & Borgstahl, 2003). At room temperature, mosaicities can be extremely small (less than one hundredth of a degree) (Bellamy, Snell, Lovelace, Pokross, & Borgstahl, 2000; Snell et al., 1995). The divergence at

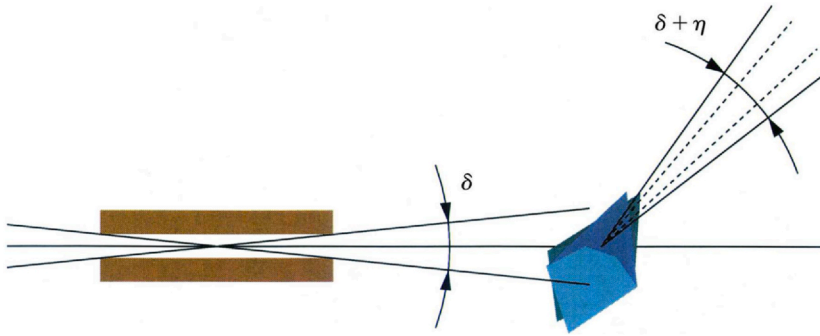

Fig. 3 Illustration showing how beam divergence (δ) and crystal mosaicity (η) contribute to the smearing of diffraction features. Reproduced with permission (Dauter, 1999).

MX beamlines with focusing optics may exceed the intrinsic mosaicity at room temperature, as was likely the case for the lysozyme datasets (Table 1). Finite energy bandwidth of the X-ray beam also smears the diffraction pattern, and for this reason high bandwidth beamlines (e.g., for Laue crystallography) are not suitable for diffuse scattering. For the best performance, crystal monochromators such as Si(111) are preferable to multilayer optics.

As diffuse scattering is carried out at non-cryogenic temperatures, the first obvious challenge is radiation damage, as the dose sensitivity for protein crystals at room temperature is approximately 1–2 orders of magnitude higher than at cryotemperatures (Southworth-Davies, Medina, Carmichael, & Garman, 2007; Thorne, 2023). To maximize the usable signal for a limited dose budget, it is important not to waste photons. Having control over beam size can help ensure that the entire crystal is irradiated evenly during the exposure. Ideally, the beam should not be larger than the crystal because the portion of the beam that does not interact with the crystal contributes extra background scattering (e.g., from the capillary and air in the beam path) without adding to the signal. Multi-motor scans can also be used to distribute the dose evenly when the crystal is larger than the beam. For instance, the tetragonal lysozyme dataset was collected from sweeps of 50 degrees spaced out over eight locations on a large crystal (Table 1, third column). The ability to perform vector or helical scans, where the goniometer translates and rotates at the same time, can be useful for distributing dose, particularly if the crystal is rod-shaped. These strategies are discussed further in Section 5.2.

The resolution of scattering that can be measured is determined by the X-ray energy, detector size, and detector distance. The considerations are similar to cryo-MX and have been reviewed elsewhere (Dauter, 1999). In conventional MX, it is common to choose the detector distance based on the resolution of diffraction so that the diffraction pattern fills the available area of the detector, maximizing the Bragg peak separation. For diffuse scattering, however, it may be preferable to place the detector close to the sample regardless of diffraction resolution. First, if the detector is too far away, diffuse scattering extending beyond the Bragg resolution would be missed. Second, for HPC detectors with negligible point spread function and small pixels, such as the EIGER, the Bragg peaks may be adequately separated even when the detector is close. High energy (short wavelength) can also be employed to collect high resolution data, with the advantage that the Ewald sphere is flatter, resulting in a smaller blind region of reciprocal space. However, the most common MX detectors use silicon sensors, and the quantum efficiency drops significantly at high energy. For example, the quantum efficiency of PILATUS with 450 μm thick silicon sensor drops to around 50% at 17 keV (Donath et al., 2013). In HPCs such as PILATUS and EIGER, the pixels are much thicker than they are wide, and thus a high energy photon striking the panel from an angle will be potentially absorbed by many adjacent pixels, broadening the pattern. HPCs with a high stopping power sensor, such as the EIGER CdTe, are now becoming available at MX beamlines and should enable better performance for diffuse scattering at atomic resolution.

Finally, the beamline should have low background scattering. Typically, air in the beam path between the final aperture and the beamstop is a major contributor to background at MX beamlines. This air path is often made somewhat large to accommodate hardware such as cameras and robotics. If possible, it should be reduced to a minimum. Of particular importance is the gap between the final aperture and the crystal, because air scattering in this region effectively makes a projection image of the sample mount on the detector, which is a non-trivial pattern that depends on the precise position of the sample during the scan and is difficult to subtract out. Air between the sample and beamstop contributes an isotropic scattering pattern that is most intense at low angles and is therefore less of a concern. However, it is often a simple matter to reduce the scattering by moving the beamstop closer to the crystal, at the expense of low-resolution Bragg data.

4.2 Data collection strategy

Once a crystal has been mounted at the beamline, the experimenter must choose the data collection parameters: the exposure time per image, starting angle and rotation speed of the goniometer (or φ increment per frame), total number of images to collect (or total angle of the φ scan), and perhaps other beamline settings such as beam size, κ angle, and attenuation. Assuming the goal of the experiment is to collect a detailed, three-dimensional diffuse map, the considerations are similar to those of conventional MX at room temperature. The dataset should be complete (covering as much of reciprocal space as possible without large gaps) with adequate signal-to-noise while carefully controlling the radiation damage to the crystal. With modern HPCs, parameter choices that improve Bragg data quality usually also improve diffuse scattering data quality. However, the relative importance of a particular parameter for Bragg data and diffuse maps may not be equal. For instance, the fine φ-slicing technique, where the φ angle increment per frame is chosen to be less than the mosaicity, will maximize the signal to noise ratio of the Bragg peaks by minimizing the amount of background scattering that contributes to the integrated peak intensity (Pflugrath, 1999). Fine φ-slicing is often advocated for Bragg data collection but is not essential. In contrast, for diffuse scattering the φ increment per frame determines the spatial resolution of the map (here, we use the term "spatial resolution" to refer to fineness of sampling in reciprocal space). Thus, the φ increment should be chosen small enough to achieve the desired map parameters (considering also the mosaicity, which fundamentally limits the spatial resolution). See Table 1 for examples and Chapter 2 for further discussion.

In MX, the term "strategy" usually refers to an optimal set of scan parameters to achieve high completeness given the crystal symmetry and orientation on the goniometer, and traditional strategies attempt to minimize the total number of diffraction images (Bourenkov & Popov, 2006; Dauter, 1999). At high flux beamlines with HPCs, minimizing the total number of frames is less of a concern, and thus modern strategies attempt to maximize data quality after merging. Such strategies typically involve high redundancy and low dose per frame (i.e., collecting 360 degrees of fine-sliced data) (Winter et al., 2019). When data are collected near room temperature, strategies to optimally distribute the dose gain extra importance, particularly if the crystals are small. For instance, the strategy used in serial oscillation crystallography is to collect small wedges (typically 1–10 degrees)

from many crystals with random orientations and to merge the data to form a complete dataset (Hasegawa et al., 2017). For larger crystals at room temperature, the best strategy might be to use a large, uniform (top-hat) beam to uniformly irradiate the sample. Alternatively, if the diffraction is not uniform across the crystal, a small beam can be used with a multi-motor scan (such as a vector or helical scan) and frames with poor quality diffraction can be filtered out during processing (Flot et al., 2009; Polsinelli et al., 2017).

For measuring diffuse scattering from large crystals at room temperature, we advocate a strategy of high redundancy and low dose per frame with fine φ-slicing (Fig. 4). High redundancy (of a factor of ~10 or more) is especially important for diffuse scattering. It enables averaging over systematic errors such as detector pixel inconsistencies, allows outlier measurements to be identified with high confidence, and fills in the gap regions created by the arrangement of detector panels. In addition, high redundancy can be used in sophisticated scaling algorithms to further correct artifacts such as extra background scattering from solvent on the crystal surface or mounting materials (Meisburger & Ando, in, press; Meisburger et al., 2020). High redundancy can be achieved by collecting a large angular range (e.g., 360 degrees) from a single crystal with high symmetry (see tetragonal lysozyme in Table 1). Or, if the crystal is low symmetry, it can help to merge datasets from multiple crystals in different orientations, as this tends to cover reciprocal space more uniformly (see triclinic lysozyme in Table 1). Crystal shape is also an important consideration. Aligning the incident beam parallel to the longer crystal dimensions helps to improve signal to noise both for Bragg diffraction and diffuse scattering. For example, the diffuse scattering of plate-shaped isocyanide hydratase (ICH) crystals was measured after careful alignment so that the beam passed through the plane of the crystal as it rotated (Su et al., 2021), thus illuminating a larger number of unit cells in each exposure.

Finally, as room-temperature data collection is a dose-budgeting experiment, the dose limit needs to be determined before finalizing the data collection strategy. Often determining an optimal strategy is an iterative process of experimentation and analysis that may require more than one beamtime. The dose limit is determined by the highest flux and total exposure time that a crystal can tolerate while incurring minimal radiation damage. Once the dose tolerance of the sample is characterized (see Section 5.2), a data collection strategy can be designed to provide a complete dataset within the allowed dose. In general, one can either attenuate the beam and expose the crystal for longer and therefore increase

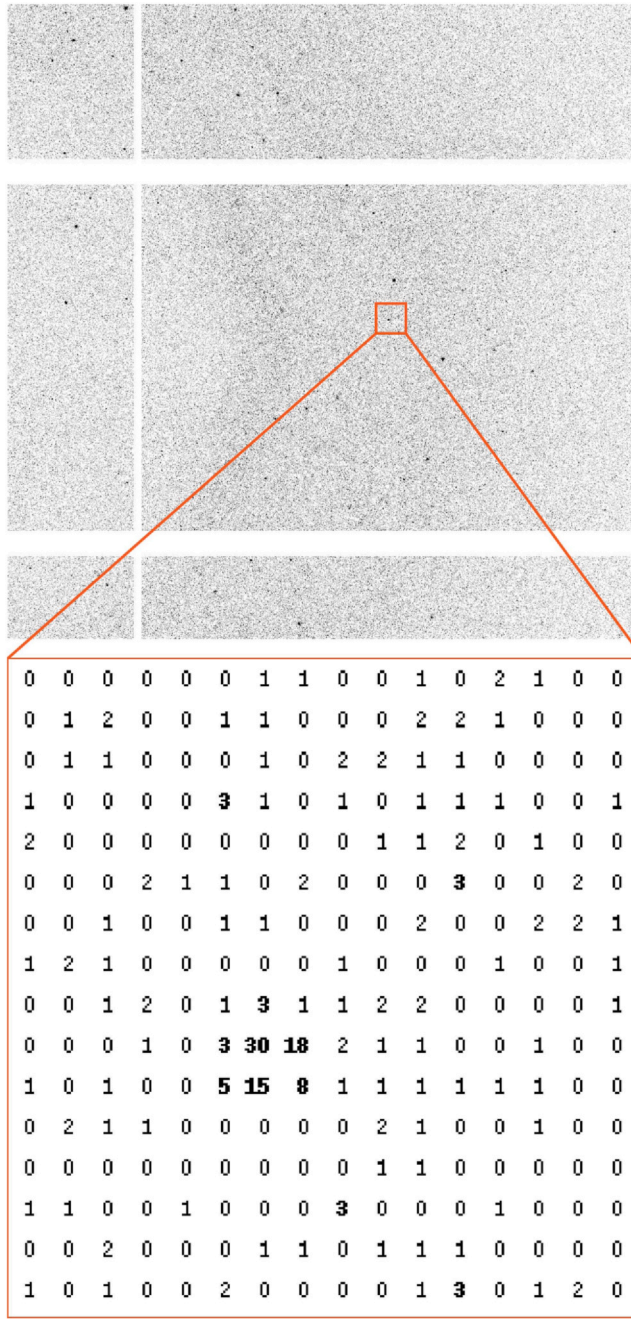

Fig. 4 Section of a typical diffraction image from a fine φ-sliced, low-dose diffuse scattering dataset (crystal B in Table 1). Bragg peaks and diffuse scattering are difficult

(Continued)

reciprocal space coverage in one single dataset, or decrease attenuation and expose the crystal for a shorter time and obtain higher I/sigma on each frame (see sample B in Table 1; beam not attenuated), possibly requiring multiple crystals for a complete dataset (see triclinic lysozyme dataset in Table 1). As discussed in Section 4.1, use of a larger beam size, a multi-motor scan (see sample A in Table 1), and a top-hat beam profile (rather than Gaussian), all help by distributing photons over a larger volume and therefore increase data collection efficiency.

4.3 Background measurement

The signal of interest in a diffuse scattering experiment is the elastic scattering from the atoms in the crystal (i.e., the proteins and the solvent occupying space between them). The ideal measurement conditions would consist of the protein crystal suspended in vacuum, so that no background scattering from the crystal mount or air is present in the diffraction images. Unfortunately, these conditions are difficult to achieve because of the need to maintain hydration, although we note that there is precedence for in-vacuum diffraction measurements using crystals wrapped in graphene (Warren et al., 2015; Wierman, Alden, Kim, McEuen, & Gruner, 2013). Fortunately, when large enough crystals are used, the background scattering from all sources at a typical macromolecular beamline is not overwhelming. In our experience, it can be made comparable to or less than the diffuse scattering signal from a large (>100 μm) crystal. Still, to obtain an accurate diffuse map, the background must be subtracted (Fig. 5).

The background has contributions from air in the beam path, sample mounting materials such as loops and capillaries, and excess water or oil on the crystal's surface. Air scatters isotropically (the intensity depends only on scattering angle, after accounting for polarization of the X-ray beam). However, as mentioned above, scattering from the air between the final aperture and the crystal makes a projection image of the sample mount on the detector. Consequently, the air scattering recorded by the detector may depend on the orientation of the sample during the scan (Meisburger et al., 2020). To measure the scan-varying background from air as well as the

Fig. 4—Cont'd to discern in the noisy image (top). Accumulated photon counts in each pixel for the zoomed-in region (bottom) show a compact Bragg peak that occupies several pixels and quickly decays to background. Most pixels in this image record zero or one photons. The diffuse signal is revealed only after merging to produce a three-dimensional map (not shown).

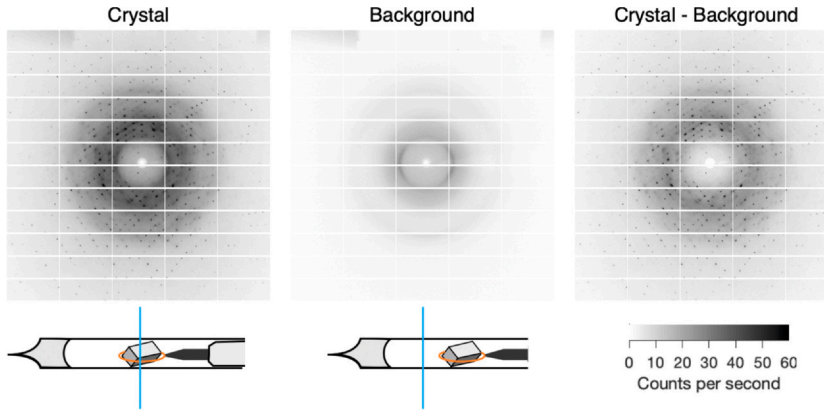

Fig. 5 Illustration of crystal and background data collection for triclinic lysozyme. Diffraction images are shown for crystal (left panel) and background measurements (middle panel) with equivalent exposures of 1 s and 1 degree of rotation. The paths of the X-ray beam (blue lines) relative to the crystal, loop, and capillary are indicated in the diagrams below each panel. The diffuse scattering intrinsic to protein crystals is obtained after background is subtracted (right panel). Shadows of the sample loop are visible in both left and central panels as a bright triangular area right of the beam center. Reproduced under CC BY 4.0 license (Meisburger et al., 2020).

capillary, we have found a reliable method is to translate the goniometer along the spindle axis by several hundred μm so the beam passes beyond the tip of the loop (but still through the capillary) (Fig. 5, middle panel). The measurement is done after each successful data collection, and it is performed over the same angular range. Because the background signal does not change rapidly with angle, the φ increment per image can be coarse (1 degree is adequate). After the background is measured, it is then subtracted from each frame during data processing (see Chapter 2).

After subtracting the air and capillary background scattering, contributions from extraneous water or oil around the crystal and the loop potentially remain. We have not attempted to measure the loop contribution (doing so would require removing the crystal and collecting an identical dataset from the same spot, a challenging task). Instead, we rely on scaling algorithms to estimate and subtract an excess isotropic signal. To use this method, two conditions must be met. First, the residual background scattering must be isotropic; for this reason, Kapton loops are preferred over nylon, whose scattering is strongly anisotropic. Second, the X-ray beam must pass through bare crystal during much of the scan, otherwise the scaling algorithm will not have sufficient information to determine the

excess signal. We therefore strive to remove as much liquid from the crystal surface as possible during mounting, which has the added benefit of a more stable mount (if too much liquid is present, the crystal can slip during data collection).

Finally, even after all sources of non-crystal background are subtracted, there remains the Compton scattering from atoms in the crystal itself. It is arguably not background at all, since it originates from the sample. However, Compton scattering is incoherent and does not contribute structural information, and it is typically not included in disorder models that assume elastic scattering (introduced in Section 1). Fortunately, it is straightforward to predict Compton scattering knowing only the atomic inventory of the unit cell, as estimated for example from the protein sequence and solvent fraction. Before it can be subtracted, however, the data must be placed on an absolute intensity scale, the procedures for which we described previously (Meisburger et al., 2020). Having data on an absolute scale is particularly useful for fitting models of lattice disorder, because the overall scale factor can be used to determine the magnitude of the structural fluctuation (i.e., the ADP) in ångström units.

5. Data quality assessment

When collecting data at the synchrotron, how do you know if a diffraction dataset is sufficient to reconstruct a high-quality diffuse map? This question is best answered by performing the reconstruction. Unfortunately, at present, the reconstruction process cannot be done in real-time at the beamline. However, we have found that visual examination of the diffraction images and a careful assessment of Bragg data quality can give a good indication of whether a dataset is worthy of further analysis. In this section, we first identify general features we expect to see in diffraction images from high quality crystals, as well as common issues that can be identified visually. Second, we describe statistics reported during Bragg data processing of relevance to diffuse data quality, specifically to radiation damage. Finally, we introduce statistical measures used to evaluate diffuse map quality and benchmark model-data comparisons.

5.1 Examination of diffraction images

Although it is always good practice to examine diffraction images when collecting MX data, it has become less of a necessity in recent years with

the prevalence of automated Bragg data processing pipelines (auto-processing). For diffuse scattering, auto-processing results are extremely useful: a complete, high quality Bragg dataset is a precondition for biophysical applications of diffuse scattering. However, they do not substitute for visual examination of diffraction images. Consider that a very small fraction of the detector pixels contributes to the integrated Bragg intensities, and integration programs have become effective at rejecting artifacts that don't fit the dominant lattice model (e.g., from twinning or multiple lattices). For diffuse scattering, we do not have the luxury of rejecting most detector pixels. Moreover, if multiple lattices or twins are present, the diffuse signals will overlap, and it will be difficult if not impossible to separate them.

Problematic diffraction can often be identified by eye. Diffraction from perfect crystals should have sharp and round Bragg peaks. If the Bragg peaks are very elongated, large, or smeared, it usually means that the crystal has high mosaicity. If closely adjacent Bragg peaks are observed, this usually indicates the existence of multiple lattices (Fig. 2D), which might mean that the crystal grew as a twin or the crystal cracked during looping or data collection. On the other hand, intense diffraction peaks that are far from each other usually suggests contamination of salt or ice crystals. If multiple lattices are present, this will often be flagged by Bragg data processing programs as a low fraction of indexed reflections. However, if the crystal is cracked so that the multiple-lattice peaks are very close together, the integration program may lump them into a single mosaic peak, and the issue would be missed without looking at the images.

Finally, we note that when using a strategy of high redundancy, fine φ-slicing, and low dose, the individual images will be extremely noisy (Fig. 4), and this complicates the examination process described above. One useful approach is to bin fine-sliced images together (for instance, every 1 degree) to increase signal-to-noise. Binning can be performed on-the-fly by commonly used image viewers, such as ADXV (http://www.scripps.edu/tainer/arvai/adxv.html) and ALBULA (https://www.dectris.com/detectors/albula-software/). Another approach is to collect longer exposures from test samples to assess diffraction quality before collecting full datasets from fresh crystals.

5.2 Quantifying and controlling radiation damage

Radiation damage processes begin as soon as the sample is exposed to X-rays, and the effects of this damage can be seen in the diffraction data. Radiation damage is known to cause reduction or radiolysis of a protein's

chemical structure, which includes reducing of metal centers (Yano et al., 2005), breaking of disulfide bonds, decarboxylation of aspartate and glutamate, and dihydroxylation of tyrosine (Burmeister, 2000). In addition, when damage is pervasive, the crystalline order can be disrupted, causing a loss of Bragg peak intensity, increase in the Wilson B-factor, increase in mosaicity, and altered unit cell dimensions (Blake & Philips, 1962; Ravelli & McSweeney, 2000). Local and global damage may occur at very different rates, especially at cryogenic temperatures where the glassy solvent state slows the loss of crystalline order. Near room temperature, global damage progresses much more rapidly than at cryogenic temperatures (approximately 1–2 orders of magnitude faster) (Southworth-Davies et al., 2007). Diffuse scattering performed at room temperature must therefore contend with the possible effects of increased disorder caused by X-ray exposure. As Bragg peak intensities fade, a concomitant increase in the diffuse scattering is expected (the total scattering from a crystal is conserved). It is not yet understood how radiation damage would specifically impact diffuse scattering. However, it is unlikely to be a simple increase in background scattering that can be corrected in data processing. The dissociation of chemical bonds modifies the mechanical constraints of internal protein motion, and the disturbance of protein packing might introduce additional diffuse scattering from static disorder; both might influence the interpretation of observed diffuse scattering. Thus, for diffuse scattering it is especially important to monitor radiation damage effects and to choose a maximum dose conservatively.

Metrics have been proposed to quantify radiation damage in Bragg diffraction data. Both local and global damage potentially complicate diffuse scattering experiments at room temperature, where they are expected to occur at similar rates (Gotthard et al., 2019). Global damage results in the fading of high-resolution Bragg reflections, and this effect can be quantified using the dose at which intensity drops by half (Henderson, 1990) or the change in average B-factor between damaged and undamaged sample (Kmetko, Husseini, Naides, Kalinin, & Thorne, 2006). Evidence for site-specific damage can often be observed in the electron density maps, especially the appearance of difference electron density ($F_o - F_c$) in the vicinity of chemical species known to be radiation sensitive. Recently, sensitive metrics have been developed to automatically detect and quantify such features (Mora et al., 2020; Shelley & Garman, 2022).

For a given sample and measurement conditions, the amount of radiation damage assessed using metrics described above have been found

to be roughly proportional to the applied dose (Kmetko et al., 2006). At cryogenic conditions, the rate of global damage progression for a given dose is nearly universal (Leiros, Timmins, Ravelli, & McSweeney, 2006; Owen, Rudiño-Piñera, & Garman, 2006; Sliz, Harrison, & Rosenbaum, 2003). In contrast, at room temperature the dose tolerance can be sample-dependent (Garman, 2010) and can also depend on the dose rate (Rajendran, Dworkowski, Wang, & Schulze-Briese, 2011; Southworth-Davies et al., 2007; Warkentin et al., 2012). For instance, it has been shown that very high flux data collection can outrun certain damage processes at room temperature, leading to an apparently higher dose tolerance under those conditions (Owen et al., 2012; Owen et al., 2014). The benefits of outrunning damage at a synchrotron are relatively modest (a factor of 1–4). At an XFEL, however, when diffraction data are collected in femtoseconds before the atoms have a chance to move, orders of magnitude higher dose may be used. For diffuse scattering experiments, the benefits of high flux data collection may be worth considering when only small crystals are available, since these experiments will be particularly signal-starved. However, we note that a boost in total signal does not solve the problem of high background scattering when small crystals are used. For this reason, it has been most common to collect diffuse scattering from large crystals using modest dose rates.

When starting data collection from a new sample and/or a new beamline, we recommend a systematic test of dose tolerance. A straightforward experiment is to collect narrow wedges (~10 degrees) repeatedly from a single spot on the crystal until the high-resolution peaks begin to disappear. After processing and scaling the data, the B-factor corrections from scaling can be plotted as a function of exposure time. A maximum tolerable B-factor increase can then be chosen, which sets a maximum exposure time for a dataset. For instance, when collecting diffuse scattering data from lysozyme polymorphs (Table 1), a maximum B-factor increase was set at $\sim 2\ \text{Å}^2$, leading to a maximum exposure time of 50 s. The dose limit estimated in this way represents a worst-case scenario, because the same spot on the crystal was irradiated continuously. The next step is to devise a strategy for obtaining a complete dataset by distributing the dose over the full crystal volume or from multiple crystals if necessary.

Because mitigation of radiation damage is so important for room-temperature data collection, here we review in greater detail strategies for dose distribution first introduced in Section 4.2. When thinking about dose, is important to consider that the X-ray intensity is not always evenly

distributed through the cross section of the beam, and many beamlines employ focusing optics so that the profile is approximately Gaussian in one or two dimensions. The benefits of a top-hat (uniform) (Sutter et al., 2016) profile for more evenly distributing dose has gained attention in recent years. Furthermore, when a crystal is rotated during data collection, the dose at the rotation center is higher than at locations further away (Fig. 6A). Such non-uniform irradiation is not ideal for data quality because the signal becomes a mixture of signals from damaged and undamaged structures. Several strategies have been developed to alleviate these issues when the crystal is larger than the beam. For instance, the center of the crystal may be offset from the goniometer rotation center, such that the center is not over-irradiated (Fig. 6D) (Zeldin, Brockhauser, Bremridge, Holton, & Garman, 2013). Alternatively, the crystal can be translated while rotating (known as a 'helical scan') to distribute the dose over the entire crystal (Fig. 6C) (Flot et al., 2009). Similarly, mini-κ goniometers (Brockhauser, Ravelli, & McCarthy, 2013; Waltersperger et al., 2015) can be used to improve dose distribution compared with standard φ-scans. The program *RADDOSE-3D* (Bury, Brooks-Bartlett, Walsh, & Garman, 2018; Zeldin, Gerstel, & Garman, 2013) can be used to calculate dose distributions for different beam profiles, crystal contents, and data collection methods.

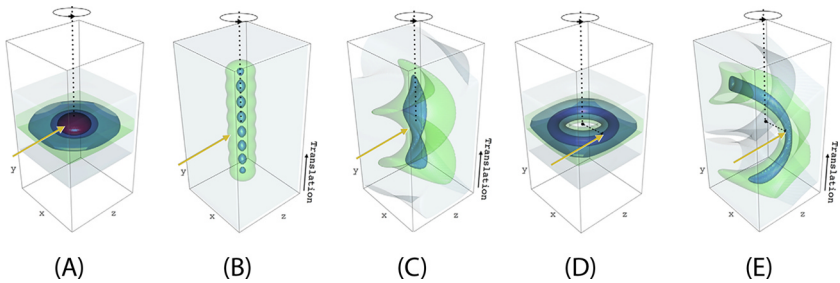

Fig. 6 Simulated dose distribution for various scan strategies. The accumulated X-ray dose on a cuboid crystal (boxes outlined in thin black lines) was simulated using RADDOSE-3D for 5 different scan types: (A) a typical single-point φ scan; (B) multi-point φ scan; (C) helical scan; (D) off-centering scan; and (E) a helical scan with offset centering. The model crystal was $100 \times 200 \times 100\ \mu m^3$ and the X-ray beam had a flux of 5×10^{11} photons/s, energy of 12.4 keV, and a Gaussian intensity profile of $20 \times 20\ \mu m^2$ (full width at half maximum). Contours correspond to dose of 0.0001 MGy (gray), 5 MGy (green), 10 MGy (light blue), 20 MGy (dark blue) and 30 MGy (red). Detailed parameters used for each simulation can be found at the original publication. Reproduced with permission (Bury et al., 2018).

After collecting a complete dataset, the global radiation damage should be assessed using the B-factor correction applied in scaling, as described above. If the dose distribution strategy was successful, the B-factor increase should be below the maximum tolerable value. If not, frames from overly damaged sample may be excluded from the dataset. However, excluding frames may result in an incomplete dataset, and in this case more diffraction data should be collected. A potential strategy to avoid incomplete data is to attenuate the beam (or collect shorter exposures) so that very little damage occurs during a 360-degree rotation, and to repeat the same scan multiple times. The later datasets can then be excluded once the B-factor increase threshold is reached without impacting completeness. If possible, the electron density should be examined for signs of local radiation damage especially at sites of known sensitivity (particularly disulfide bonds and metals). It may be necessary to revise the B-factor cutoff if too much damage is seen. Finally, since radiation damage issues are sometimes subtle and may go unnoticed during data collection, as a rule of thumb we recommend collecting more datasets than necessary to obtain a complete map.

5.3 Diffuse map quality assessment

The process of reconstructing a three-dimensional reciprocal space map from diffraction images follows similar steps as Bragg data reduction, including integration, scaling, and merging. These steps are discussed in greater detail in Chapter 2. Here, we review methods that have been employed to assess the quality of maps.

In general, two factors are important in data quality: precision and accuracy. The precision of a diffuse scattering dataset is related to how noisy it is. Because radiation damage limits the exposure times, the dominant source of noise in diffuse scattering is often the unavoidable Poisson statistics from photon-counting; the uncertainty in a measurement with N photons is $\sqrt{N}$. Precision improves when more data are averaged together according to the same rule of $\sqrt{M}$, where M is the number of measurements. Thus, to improve the precision by a factor of two, four equivalent datasets must be averaged together. Precision can also improve if background scattering is reduced, because the error in a measurement with background is $\sqrt{N_{\text{crystal}} + N_{\text{background}}}$. If the background can be reduced so it is significantly less than the scattering from the crystal, its effect on precision becomes negligible. On the other hand, even a small amount of background can ruin accuracy if it is not subtracted correctly. Although accuracy is harder to determine than precision, often reproducibility is a

good proxy. When data can be reproduced under very different conditions (a different crystal, or even a different beamline), it is likely that the data are accurate. Thus, for both Bragg diffraction and diffuse scattering, statistics related to precision and reproducibility are used to assess data quality.

Because photon-counting statistics often dominate the noise in diffuse maps, data reduction algorithms propagate the Poisson error estimate through data reduction steps, giving an "error bar" for each data point (see Chapter 2 for details). The Poisson error represents a lower bound because other types of systematic error may also contribute. For instance, if a detector were poorly calibrated, the spread in equivalent observations from different detector panels might be larger than expected given Poisson statistics. To estimate the uncertainty including certain systematic errors, a common strategy is to split the dataset into equivalent halves that are merged separately and to compute the correlation between them. If the split is random, the statistic is called $CC_{1/2}$ (Karplus & Diederichs, 2012). For diffraction data, the split can also be performed according to symmetry operations and may be called CC_{Laue} or $CC_{Friedel}$ depending on how operations are grouped, or the data can be split between different crystals in the case of CC_{rep} (Su et al., 2021). These correlation coefficients measure the internal consistency of the dataset. Because signal-to-noise depends strongly on resolution, CCs are often calculated in bins of constant resolution. R-factors have also been used to assess internal consistency (Wall, Ealick, et al., 1997). In general, the switch to R-factors makes sense when correlation coefficients approach ~1 because R-factors are more sensitive to discrepancies. In the context of Bragg diffraction, where the signal-to-noise ratio is very high, R-factors are the preferred statistic (except when judging resolution cutoffs). In contrast, most diffuse scattering datasets reported to date have relatively high noise, and so correlation coefficients are an appropriate statistic to use.

Although statistics used to assess Bragg data and diffuse data are superficially similar (such as $CC_{1/2}$), there is an important distinction. For Bragg data, the $CC_{1/2}$ is a universal measure of data consistency because only Bragg peak intensities (i.e., integrated *hkl* intensities) are compared. When $CC_{1/2}$ is applied to diffuse scattering, it is important to also specify exactly how the data were treated. For instance, if a very coarse map is generated by binning data down to one point per integer Miller index, the noise level will be very low, and $CC_{1/2}$ may be very high. In contrast, if an extremely fine map is generated from the same dataset, each point may be very noisy and $CC_{1/2}$ will be low. $CC_{1/2}$ will also depend on the kinds of

diffuse scattering present. For example, if halo features are very intense, the $CC_{1/2}$ may be very high even if the cloudy diffuse pattern is noisy. Thus, for diffuse scattering, $CC_{1/2}$ is best used to compare datasets from nominally identical crystals that are processed identically (i.e., with the same mask and integration grid). There can be no general recommendation such as *the diffuse map must have* $CC_{1/2} > 0.95$ without also specifying the data processing details. Given these caveats, $CC_{1/2}$ is still a useful statistic. For instance, it has been used to judge whether a scaling algorithm made a meaningful improvement in data quality (see for example Fig. 7).

Finally, we note that $CC_{1/2}$ also plays an important role in quantifying model-data agreement. From $CC_{1/2}$, it is possible to estimate the best possible correlation between a model and the data if the model were perfect. This statistic, known as CC* (Karplus & Diederichs, 2012), was originally developed for Bragg data, but it applies equally well to diffuse data (Fig. 8). It can be derived by considering how the noise in each half-dataset contributes to the correlation coefficient between them, compared with the noise level of the merged dataset vs. the model (which is noise-free). CC* is a simple transformation of $CC_{1/2}$, as follows:

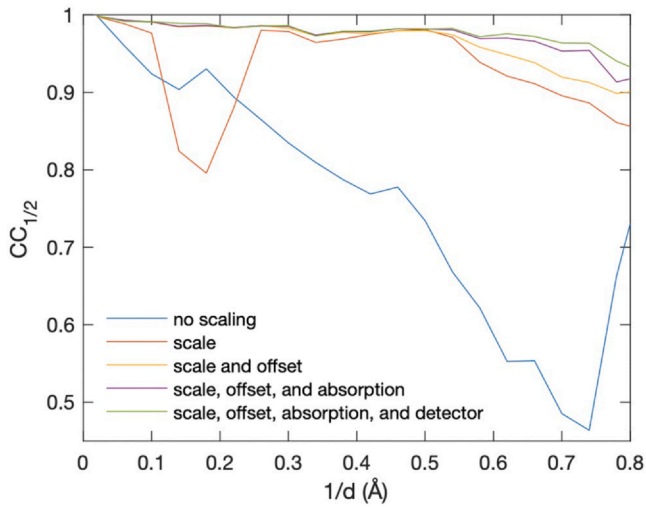

Fig. 7 Using $CC_{1/2}$ to measure data quality improvements during data processing. A scaling model with four correction types (scale, offset, absorption, and detector) was refined in mdx-lib (see Chapter 2) to merge a map of diffuse scattering from triclinic lysozyme with one sample per integer Miller index. The $CC_{1/2}$ statistic computed for each resolution shell (*d*) shows that the data become more self-consistent as each correction is added. Reproduced under CC BY 4.0 license (Meisburger et al., 2020).

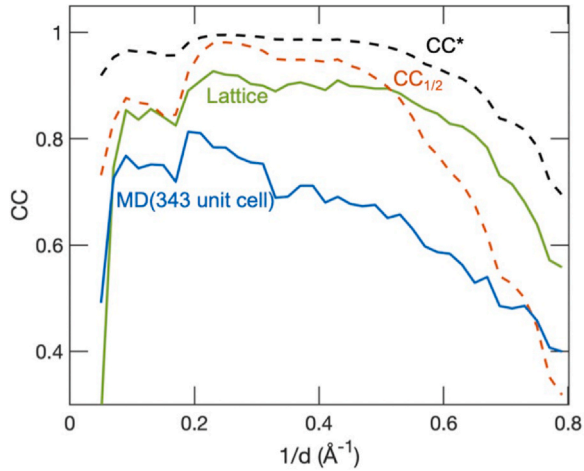

Fig. 8 Correlation coefficients (CC) for model-data comparison. CCs are computed between the experimental diffuse map from triclinic lysozyme and two simulations: a lattice dynamics model (solid green curve) and molecular dynamics (MD) simulation with 343 unit cells (solid blue curve). The CC* statistic (black dashed line) estimates the maximum possible model-data correlation and is calculated from $CC_{1/2}$ (dashed red line) using Eq. (13). Note that the lattice model CC exceeds $CC_{1/2}$ at high resolution (*d*) but always falls below CC*. *Adapted under CC BY 4.0 license (Meisburger et al., 2020).*

$$CC^* = \sqrt{\frac{2CC_{1/2}}{1 + CC_{1/2}}} \tag{13}$$

As the macromolecular diffuse scattering field matures, we expect further work will be done to address shortcomings of these indicators and to standardize the reporting of statistics in publications.

6. Conclusions

Recent years have seen major advances in the field of macromolecular diffuse scattering. Building on decades of seminal studies (Caspar, Clarage, Salunke, & Clarage, 1988; Phillips, Fillers, & Cohen, 1980; Polikanov & Moore, 2015; Wall, Clarage, et al., 1997; Wall, Ealick, et al., 1997), the widespread adoption of direct detectors at crystallography beamlines catalyzed the development of new experimental and computational methods. In 2020, it was shown with high confidence that diffuse scattering from protein crystals can be explained in terms of lattice

dynamics and protein dynamics (Meisburger et al., 2020), achieving a goal that was first envisioned by Don Caspar and co-workers over 30 years ago (Caspar et al., 1988).

With the collective knowledge the field has gained over the past few years, it is now possible to provide general guidelines for collecting diffuse scattering data. In general, a crystal that is suitable for conventional crystallography meets the minimum requirement for a diffuse scattering study. However, as is generally true for all scattering experiments, the requirement for sample quality is more stringent for diffuse scattering than for conventional crystallography. This is because the signal of interest is spread throughout reciprocal space rather than being concentrated in diffraction spots. Additionally, to study protein motions without the complicating factors introduced by cryo-cooling crystals, it is recommended to start diffuse scattering experiments at non-cryogenic temperatures. Ultimately, the goals of diffuse scattering data collection are (1) to achieve the cleanest signal throughout reciprocal space without introducing unnecessary artifacts, (2) to collect data in a way that will resolve the most intense diffuse scattering features (e.g., halos), and (3) to collect data in a way that allows for sophisticated scaling and merging procedures (and generate data quality statistics in the process). Best practices, such as careful sample handling and beamline selection, low-dose data collection for mitigating radiation damage, precise background collection, stringent quality check on diffraction images, should be always taken in order to meet these goals.

There are still many unanswered questions about macromolecular diffuse scattering. For example, a systematic study of the effect of radiation damage has not yet been done. Additionally, it is not yet known whether diffuse scattering from a crystal that is considered "poor quality" can be interpreted in a robust manner. Such crystals may be highly mosaic (producing smeary diffraction) or feature highly disordered regions (leading to diffuse scattering that extends past the resolution limit of the diffraction pattern). Finally, measuring diffuse scattering from a serial crystallography experiment (such as that performed at an XFEL) has its own challenges, which need to be analyzed in detail. In addition to answering these questions, there is still room for improvement on the computational front, both in data processing and modeling. Although the field of macromolecular diffuse scattering has not yet matured, we hope that this chapter provides readers with the knowledge needed to design experiments that will produce datasets that can stand the test of time, such that they can be reanalyzed by newer software as they become available.

Acknowledgments

The authors are grateful to Dr. Lindsey Backman and Prof. Catherine Drennan (MIT) for sharing the image shown in Fig. 2D. Data shown in Figs. 1A, 2B, 2F and 4 were collected at beamline ID7B2 at the Center for High-Energy X-ray Sciences (CHEXS), which is supported by the at National Science Foundation (BIO, ENG and MPS Directorates) under award DMR-1829070, and the Macromolecular Diffraction at CHESS (MacCHESS) facility, which is supported by award P30GM124166 from the National Institute of General Medical Sciences, National Institutes of Health, and by New York State's Empire State Development Corporation (NYSTAR). This work was partially supported by the Cornell Center for Material Research with funding from the NSF MRSEC program (DMR-1719875) and primarily supported by National Institutes of Health grant GM124847 (to N.A.).

References

Ayyer, K., Yefanov, O. M., Oberthür, D., Roy-Chowdhury, S., Galli, L., Mariani, V., & Chapman, H. N. (2016). Macromolecular diffractive imaging using imperfect crystals. *Nature, 530*, 202–206.

Bautis, A. N. (1975). *Immersion oil formulations for use in microscopy and similar fields. U.S. Patent No. 3,929,667*. Washington, DC: U.S. Patent and Trademark Office.

Bellamy, H. D., Snell, E. H., Lovelace, J., Pokross, M., & Borgstahl, G. E. O. (2000). The high-mosaicity illusion: Revealing the true physical characteristics of macromolecular crystals. *Acta Crystallographica. Section D, Biological Crystallography, 56*, 986–995.

Benschoten, A. H. V., Liu, L., Gonzalez, A., Brewster, A. S., Sauter, N. K., Fraser, J. S., & Wall, M. E. (2016). Measuring and modeling diffuse scattering in protein X-ray crystallography. *Proceedings of the National Academy of Sciences of the United States of America, 113*, 4069–4074.

Blake, C. C. F., & Philips, D. C. (1962). Effects of X-irradiation on single crystals of myoglobin. In S. Eklund (Ed.). *Proceedings of the Symposium on the Biological Effects of Ionizing Radiation at the Molecular Level* (pp. 183–191). Vienna: International Atomic Energy Agency.

Bourenkov, G. P., & Popov, A. N. (2006). A quantitative approach to data-collection strategies. *Acta Crystallographica. Section D, Biological Crystallography, 62*, 58–64.

Brockhauser, S., Ravelli, R. B. G., & McCarthy, A. A. (2013). The use of a mini-κ goniometer head in macromolecular crystallography diffraction experiments. *Acta Crystallographica. Section D, Biological Crystallography, 69*, 1241–1251.

Budayova-Spano, M., Koruza, K., & Fisher, Z. (2020). *Large crystal growth for neutron protein crystallography. Methods in Enzymology, 634*, Elsevier, 21–46.

Burmeister, W. P. (2000). Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallographica. Section D, Biological Crystallography, 56*, 328–341.

Bury, C. S., Brooks-Bartlett, J. C., Walsh, S. P., & Garman, E. F. (2018). Estimate your dose: RADDOSE-3D. *Protein Science, 27*, 217–228.

Case, D. A. (in press). MD simulations of macromolecular diffuse crystals: Implications for the analysis of Bragg and diffuse scattering. In *Methods in enzymology* (Vol. 688).

Caspar, D. L. D., Clarage, J., Salunke, D. M., & Clarage, M. (1988). Liquid-like movements in crystalline insulin. *Nature, 332*, 659–662.

Chapman, H. N., Yefanov, O. M., Ayyer, K., White, T. A., Barty, A., Morgan, A., & Pande, K. (2017). Continuous diffraction of molecules and disordered molecular crystals. *Journal of Applied Crystallography, 50*, 1084–1103.

Dauter, Z. (1999). Data-collection strategies. *Acta Crystallographica. Section D, Biological Crystallography, 55*, 1703–1717.

Donath, T., Brandstetter, S., Cibik, L., Commichau, S., Hofer, P., Krumrey, M., & Wernecke, J. (2013). Characterization of the PILATUS photon-counting pixel detector

for X-ray energies from 1.75 keV to 60 keV. *Journal of Physics: Conference Series, 425*, 062001.
Fischer, M. (2021). Macromolecular room temperature crystallography. *Quarterly Reviews of Biophysics, 54*, e1.
Flot, D., Mairs, T., Giraud, T., Guijarro, M., Lesourd, M., Rey, V., & Mitchell, E. (2009). The ID23-2 structural biology microfocus beamline at the ESRF. *Journal of Synchrotron Radiation, 17*, 107–118.
Förster, A., Brandstetter, S., & Schulze-Briese, C. (2019). Transforming X-ray detection with hybrid photon counting detectors. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 377*, 20180241.
Fraser, J. S., Clarkson, M. W., Degnan, S. C., Erion, R., Kern, D., & Alber, T. (2009). Hidden alternative structures of proline isomerase essential for catalysis. *Nature, 462*, 669–673.
Fraser, J. S., Van Den Bedem, H., Samelson, A. J., Lang, P. T., Holton, J. M., Echols, N., & Alber, T. (2011). Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proceedings of the National Academy of Sciences, 108*, 16247–16252.
Garman, E. F. (2010). Radiation damage in macromolecular crystallography: What is it and why should we care? *Acta Crystallographica. Section D, Biological Crystallography, 66*, 339–351.
Gotthard, G., Aumonier, S., De Sanctis, D., Leonard, G., Von Stetten, D., & Royant, A. (2019). Specific radiation damage is a lesser concern at room temperature. *IUCrJ, 6*, 665–680.
Halle, B. (2004). Biomolecular cryocrystallography: Structural changes during flash-cooling. *Proceedings of the National Academy of Sciences, 101*, 4793–4798.
Hasegawa, K., Yamashita, K., Murai, T., Nuemket, N., Hirata, K., Ueno, G., & Yamamoto, M. (2017). Development of a dose-limiting data collection strategy for serial synchrotron rotation crystallography. *Journal of Synchrotron Radiation, 24*, 29–41.
Henderson, R. (1990). Cryo-protection of protein crystals against radiation damage in electron and X-ray diffraction. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 241*, 6–8.
Juers, D. H., & Matthews, B. W. (2001). Reversible lattice repacking illustrates the temperature dependence of macromolecular interactions. *Journal of Molecular Biology, 311*, 851–862.
Karplus, P. A., & Diederichs, K. (2012). Linking crystallographic model and data quality. *Science, 336*, 1030–1033.
Kmetko, J., Husseini, N. S., Naides, M., Kalinin, Y., & Thorne, R. E. (2006). Quantifying X-ray radiation damage in protein crystals at cryogenic temperatures. *Acta Crystallographica. Section D, Biological Crystallography, 62*, 1030–1038.
Kriminski, S., Caylor, C. L., Nonato, M. C., Finkelstein, K. D., & Thorne, R. E. (2002). Flash-cooling and annealing of protein crystals. *Acta Crystallographica. Section D, Biological Crystallography, 58*, 459–471.
Leiros, H.-K. S., Timmins, J., Ravelli, R. B. G., & McSweeney, S. M. (2006). Is radiation damage dependent on the dose rate used during macromolecular crystallography data collection? *Acta Crystallographica. Section D, Biological Crystallography, 62*, 125–132.
Leonarski, F., Redford, S., Mozzanica, A., Lopez-Cuenca, C., Panepucci, E., Nass, K., & Wang, M. (2018). Fast and accurate data collection for macromolecular crystallography using the JUNGFRAU detector. *Nature Methods, 15*, 799–804.
Lisi, G. P., & Loria, J. P. (2016). Solution NMR spectroscopy for the study of enzyme allostery. *Chemical Reviews, 116*, 6323–6369.
Lovelace, J. J., & Borgstahl, G. E. O. (2020). Characterizing pathological imperfections in macromolecular crystals: Lattice disorders and modulations. *Crystallography Reviews, 26*, 3–50.
Meisburger, S. P., & Ando, N. (2017). Correlated motions from crystallography beyond diffraction. *Accounts of Chemical Research, 50*, 580–583.

Meisburger, S.P., & Ando, N. (in press). Processing macromolecular diffuse scattering data. In *Methods in enzymology* (Vol. 688).
Meisburger, S. P., Case, D. A., & Ando, N. (2020). Diffuse X-ray scattering from correlated motions in a protein crystal. *Nature Communications, 11*, 1271.
Meisburger, S. P., Case, D. A., & Ando, N. (2023). Robust total X-ray scattering workflow to study correlated motion of proteins in crystals. *Nature Communications, 14*, 1228.
Meisburger, S. P., Thomas, W. C., Watkins, M. B., & Ando, N. (2017). X-ray scattering studies of protein structural dynamics. *Chemical Reviews, 117*, 7615–7672.
Mora, E., De La Coquelle, N., Bury, C. S., Rosenthal, M., Holton, J. M., Carmichael, I., & Weik, M. (2020). Radiation damage and dose limits in serial synchrotron crystallography at cryo- and room temperatures. *Proceedings of the National Academy of Sciences, 117*, 4142–4151.
Ng, J. D., Baird, J. K., Coates, L., Garcia-Ruiz, J. M., Hodge, T. A., & Huang, S. (2015). Large-volume protein crystal growth for neutron macromolecular crystallography. *Acta Crystallographica. Section F, Structural Biology Communications, 71*, 358–370.
Owen, R. L., Axford, D., Nettleship, J. E., Owens, R. J., Robinson, J. I., Morgan, A. W., & Evans, G. S. (2012). Outrunning free radicals in room-temperature macromolecular crystallography. *Acta Crystallographica. Section D, Biological Crystallography,* 810–818.
Owen, R. L., Paterson, N., Axford, D., Aishima, J., Schulze-Briese, C., Ren, J., & Evans, G. (2014). Exploiting fast detectors to enter a new dimension in room-temperature crystallography. *Acta Crystallographica. Section D, Biological Crystallography, 70*, 1248–1256.
Owen, R. L., Rudiño-Piñera, E., & Garman, E. F. (2006). Experimental determination of the radiation dose limit for cryocooled protein crystals. *Proceedings of the National Academy of Sciences, 103*, 4912–4917.
Palmer, A. G. (2015). Enzyme dynamics from NMR spectroscopy. *Accounts of Chemical Research, 48*, 457–465.
Peck, A., Lane, T. J., & Poitevin, F. (submitted). Modeling diffuse scattering with simple, physically interpretable models. In *Methods in enzymology* (Vol. 688).
Pflugrath, J. W. (1999). The finer things in X-ray diffraction data collection. *Acta Crystallographica. Section D, Biological Crystallography, 55*, 1718–1725.
Pflugrath, J. W. (2015). Practical macromolecular cryocrystallography. *Acta Crystallographica. Section F, Structural Biology Communications, 71*, 622–642.
Phillips, G. N., Fillers, J. P., & Cohen, C. (1980). Motions of tropomyosin. Crystal as metaphor. *Biophysical Journal, 32*, 485–502.
Polikanov, Y. S., & Moore, P. B. (2015). Acoustic vibrations contribute to the diffuse scatter produced by ribosome crystals. *Acta Crystallographica. Section D, Biological Crystallography, 71*, 2021–2031.
Polsinelli, I., Savko, M., Rouanet-Mehouas, C., Ciccone, L., Nencetti, S., Orlandini, E., & Shepard, W. (2017). Comparison of helical scan and standard rotation methods in single-crystal X-ray data collection strategies. *Journal of Synchrotron Radiation, 24*, 42–52.
Rajendran, C., Dworkowski, F. S. N., Wang, M., & Schulze-Briese, C. (2011). Radiation damage in room-temperature data acquisition with the PILATUS 6M pixel detector. *Journal of Synchrotron Radiation, 18*, 318–328.
Ravelli, R. B., & McSweeney, S. M. (2000). The 'fingerprint' that X-rays can leave on structures. *Structure, 8*, 315–328.
Schneider, T. R. (2000). Objective comparison of protein structures: Error-scaled difference distance matrices. *Acta Crystallographica. Section D, Biological Crystallography, 56*, 714–721.
Shelley, K. L., & Garman, E. F. (2022). Quantifying and comparing radiation damage in the Protein Data Bank. *Nature Communications, 13*, 1314.
Sliz, P., Harrison, S. C., & Rosenbaum, G. (2003). How does radiation damage in protein crystals depend on X-ray dose? *Structure, 11*, 13–19.

Snell, E. H., Bellamy, H. D., & Borgstahl, G. E. O. (2003). *Macromolecular crystal quality. 368*, 268–288.
Snell, E. H., Weisgerber, S., Helliwell, J. R., Weckert, E., Hölzer, K., & Schroer, K. (1995). Improvements in lysozyme protein crystal perfection through microgravity growth. *Acta Crystallographica. Section D, Biological Crystallography, 51*, 1099–1102.
Southworth-Davies, R. J., Medina, M. A., Carmichael, I., & Garman, E. F. (2007). Observation of decreased radiation damage at higher dose rates in room temperature protein crystallography. *Structure, 15*, 1531–1541.
Stura, E. (1999). Chapter 14. Seeding. In T. M. Bergfors (Ed.). *Protein crystallization: Techniques, strategies, and tips: A laboratory manual* (pp. 141–153). La Jolla, Calif: International University Line.
Su, Z., Dasgupta, M., Poitevin, F., Mathews, I. I., Van Den Bedem, H., Wall, M. E., & Wilson, M. A. (2021). Reproducibility of protein X-ray diffuse scattering and potential utility for modeling atomic displacement parameters. *Structural Dynamics, 8*, 044701.
Sutter, J. P., Alcock, S. G., Kashyap, Y., Nistea, I., Wang, H., & Sawhney, K. (2016). Creating flat-top X-ray beams by applying surface profiles of alternating curvature to deformable piezo bimorph mirrors. *Journal of Synchrotron Radiation, 23*, 1333–1347.
Thaller, C., Weaver, L. H., Eichele, G., Wilson, E., Karlsson, R., & Jansonius, J. N. (1981). Repeated seeding technique for growing large single crystals of proteins. *Journal of Molecular Biology, 147*, 465–469.
Thompson, M. C., Cascio, D., & Yeates, T. O. (2018). Microfocus diffraction from different regions of a protein crystal: Structural variations and unit-cell polymorphism. *Acta Crystallographica. Section D, Structural Biology, 74*, 411–421.
Thorne, R. E. (2023). Determining biomolecular structures near room temperature using X-ray crystallography: Concepts, methods and future optimization. *Acta Crystallographica. Section D, Structural Biology, 79*, 78–94.
Vahedi-Faridi, A., Lovelace, J., Bellamy, H. D., Snell, E. H., & Borgstahl, G. E. O. (2003). Physical and structural studies on the cryocooling of insulin crystals. *Acta Crystallographica. Section D, Biological Crystallography, 59*, 2169–2182.
Wall, M. E., Clarage, J. B., & Phillips, G. N. (1997). Motions of calmodulin characterized using both Bragg and diffuse X-ray scattering. *Structure, 5*(12), 1599–1612.
Wall, M. E., Ealick, S. E., & Gruner, S. M. (1997). Three-dimensional diffuse X-ray scattering from crystals of Staphylococcal nuclease. *Proceedings of the National Academy of Sciences, 94*, 6180–6184.
Waltersperger, S., Olieric, V., Pradervand, C., Glettig, W., Salathe, M., Fuchs, M. R., & Wang, M. (2015). PRIGo: A new multi-axis goniometer for macromolecular crystallography. *Journal of Synchrotron Radiation, 22*, 895–900.
Warkentin, M., Badeau, R., Hopkins, J. B., Mulichak, A. M., Keefe, L. J., & Thorne, R. E. (2012). Global radiation damage at 300 and 260 K with dose rates approaching 1 MGy s−1. *Acta Crystallographica. Section D, Biological Crystallography, 68*, 124–133.
Warkentin, M., & Thorne, R. E. (2009). Slow cooling of protein crystals. *Journal of Applied Crystallography, 42*, 944–952.
Warren, A. J., Crawshaw, A. D., Trincao, J., Aller, P., Alcock, S., Nistea, I., & Evans, G. (2015). *In vacuo* X-ray data collection from graphene-wrapped protein crystals. *Acta Crystallographica. Section D, Biological Crystallography, 71*, 2079–2088.
Welberry, T. R., & Weber, T. (2016). One hundred years of diffuse scattering. *Crystallography Reviews, 22*, 2–78.
Wierman, J. L., Alden, J. S., Kim, C. U., McEuen, P. L., & Gruner, S. M. (2013). Graphene as a protein crystal mounting material to reduce background scatter. *Journal of Applied Crystallography, 46*, 1501–1507.
Winter, G., Gildea, R. J., Paterson, N. G., Beale, J., Gerstel, M., Axford, D., & Hall, D. R. (2019). How best to use photons. *Acta Crystallographica. Section D, Structural Biology, 75*, 242–261.

Wych, D. C., & Wall, M. E. (in press a). Molecular-dynamics simulations of macromolecular diffraction, part I: Preparation of protein crystal simulations. In *Methods in enzymology* (Vol. 688).

Wych, D. C., & Wall, M. E. (in press b). Molecular-dynamics simulations of macromolecular diffraction, part II: Analysis of protein crystal simulations. In *Methods in enzymology* (Vol. 688).

Xu, D., Meisburger, S. P., & Ando, N. (2021). Correlated motions in structural biology. *Biochemistry, 60*, 2331–2340.

Yano, J., Kern, J., Irrgang, K.-D., Latimer, M. J., Bergmann, U., Glatzel, P., & Yachandra, V. K. (2005). X-ray damage to the Mn4Ca complex in single crystals of photosystem II: A case study for metalloprotein crystallography. *Proceedings of the National Academy of Sciences, 102*, 12047–12052.

Zeldin, O. B., Brockhauser, S., Bremridge, J., Holton, J. M., & Garman, E. F. (2013). Predicting the X-ray lifetime of protein crystals. *Proceedings of the National Academy of Sciences, 110*, 20551–20556.

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

Zhong, E. D., Bepler, T., Berger, B., & Davis, J. H. (2021). CryoDRGN: Reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods, 18*, 176–185.