



UNIVERSITY OF OXFORD

DPhil THESIS

# **Quantifying Radiation Damage in X-ray Diffraction experiments in Structural Biology**

*Jonathan Charles Brooks-Bartlett*

supervised by

Prof. Elspeth GARMAN

February 18, 2016

# Contents

<b>1</b>	<b>Introduction</b>	<b>6</b>
1.1	Macromolecular X-ray crystallography (MX)	6
1.1.1	Producing a protein crystal	8
1.1.2	Crystals	8
1.1.3	The diffraction experiment	9
1.1.4	Understanding diffraction from a crystal	10
1.1.5	From diffraction patterns to electron density - The theory	13
1.1.6	From diffraction patterns to electron density - In practice	14
1.2	Small Angle X-ray Scattering (SAXS)	15
1.3	Limitations and alternatives to MX and SAXS	17
1.4	Radiation damage in MX	18
1.4.1	Types of X-ray interactions with atoms: Primary damage	18
1.4.2	Secondary damage	19
1.4.3	Quantifying energy absorbance: Dose	20
1.4.4	Manifestations of damage: Global damage	22
1.4.5	Manifestations of damage: Specific damage	25
1.4.6	Experimental methods for dealing with radiation damage	26
1.4.7	Modelling intensity decay	27
1.5	Radiation damage in SAXS	31
1.6	This thesis in context	31
<b>A</b>	<b>Appendix Title</b>	<b>41</b>



# List of Figures

1.1	Typical protein structure solution pipeline in X-ray crystallography [40]. . . . .	7
1.2	(a) Crystals are formed by repeating unit cells related by translations in all 3 dimensions. The intersection points of the lines are the lattice points. (b) A single unit cell with axes $a, b$ and $c$ , with angles between them $\alpha, \beta$ and $\gamma$ . [31] . . . . .	8
1.3	Two dimensional Bravais lattices. (a) Rectangular lattice system with no centering. (b) Centred rectangular lattice system. . . . .	9
1.4	Diffraction image from a crystal of bovine pancreatic insulin crystallised in space group $I2_13$ recorded in January 2014 on beamline P14 at PETRA III synchrotron, Hamburg. Individual spots known as reflections or Bragg peaks are clearly visible. . . . .	10
1.5	(a) Scattering from a system of two electrons labelled 1 and 2. Incident beam direction is denoted $s_0$ and the scattered beam direction is denoted $s$ . The origin is defined at the position of electron 1. Electron 2 is at a position $r$ with respect to electron 1. The path length difference of the beam scattered by electron 1 and electron 2 is $p + q$ . (b) The angle $\theta$ is the angle between the incident beam and the reflecting plane. $S = s - s_0$ and is perpendicular to the reflecting plane. The lengths of the vectors $s$ and $s_0$ are arbitrary but are chosen to be $1/\lambda$ for convenience, where $\lambda$ is the wavelength of the incident X-ray beam. This gives $ S  = 2 \sin(\theta)/\lambda$ [30]. . . . .	11
1.6	(a) An atom with a centrosymmetric electron density distribution (i.e. $\rho(r) = \rho(-r)$ ) about the centre of the nucleus. (b) atomic scattering factor, $f$ , of carbon as a function of $ S  = 2 \sin(\theta)/\lambda$ . $f$ is expressed in number of electrons. Notice an angle of $\theta = 0$ , $f = 6$ corresponds to the atomic number of carbon. For this figure, $0^\circ \leq \theta \leq 90^\circ$ and $\lambda = 1.0$ . . . . .	12
1.7	(a) X-ray beam is incident on a SAXS sample and the scattered radiation is collected on a detector. The symbol $s$ in the figure is equivalent to the momentum transfer denoted $q$ in equation 1.20. (b) Radially averaged intensity curves from a solution of bovine serum albumin (BSA). The radially averaged intensity of the solution containing only the buffer (red curve) is subtracted from the radially averaged intensity of the solution containing BSA and buffer (black curve) to obtain the resulting intensity curve for BSA (blue curve) [12]. . . . .	16
1.8	(a) Kratky plots. The red curve shows unfolded RNA domain P4-P6 from the Tetrahymena ribozyme where the salt concentration was low. Blue curve implies the domain was folded and monomeric [91]. (b) Distance distribution functions for several geometrical volumes: sphere (red), dumbbell (blue), cylinder (green) disk (yellow) [12]. . . . .	17

---

1.9	Relative contribution to the overall X-ray cross section as a function of incident beam energy for chicken egg-white lysozyme [85]. The photoelectric effect dominates the interaction cross section at typical wavelengths used in MX (about 12 keV). However, at higher energies the elastic contribution increases but the Compton effect increases faster and begins to dominate. . . . .	20
1.10	Dose distributions in crystals represented as a polyhedron. (a) An oblate ellipsoid ( $x = z = 10 \mu m$ , $y = 6 \mu m$ ) crystal that has been irradiated with a Gaussian profile beam with a full width half maximum (FWHM) of $5 \mu m \times 5 \mu m$ , flux of $10^{10}$ photons per second and an incident photon energy of 12.7 keV for a rotation of $360^\circ$ . The collimation was set to $20 \mu m \times 20 \mu m$ and hence the entire volume of the crystal was irradiated. The 3 visible contours represent dose values of 20 MGy, 15 MGy and 5 MGy from inner to outer contour levels. (b) An icosahedral crystal with maximum dimensions of 10 $\mu m$ in each of the $x, y$ and $z$ dimensions, which has been irradiated with a Gaussian profile beam with FWHM 4 $\mu m$ in both the $z$ and $y$ directions, flux $2 \times 10^{10}$ photons per second and an incident photon energy of 12.7 keV. The rotation axis was offset 5 $\mu m$ from the centre of the crystal in the $x$ -direction and the crystal was rotated $360^\circ$ for a total exposure time of 180 seconds. Contouring levels correspond to 2.59 MGy, 12.5 MGy and 18.75 MGy from inner to outer contour levels. It is evident that different crystal, beam and experiment design parameters lead to very different distributions of absorbed dose within a protein crystal. . . . .	21
1.11	Intensity decay of reflections on diffraction images of cubic insulin collected at 100 K with increasing radiation exposure. The intensity of reflections corresponding to high resolution information (spots closer to the edges of the images) decay faster than low resolution reflections. . . . .	23
1.12	Various global radiation damage metrics as a function of dose for four holoferritin crystals. Work performed by [84], reproduced from [39]. . . . .	24
1.13	Transition dynamics of the compartments in the Blake and Phillips model. Note here that the states are irreversible and it is possible to transition directly from the undamaged fraction to the amorphous fraction. . . . .	28

## List of Tables

# Chapter 1

## Introduction

1. **TODO:** Need to sort the title page
2. **TODO:** Need to sort out font for thesis.
3. **TODO:** Need to sort page style for thesis.
4. **TODO:** The style for the references need to be sorted.

Determination of the three-dimensional (3D) structure of biological macromolecules is essential in developing our understanding of vital processes that occur within the cells of living organisms. This is because the 3D structure of a macromolecule is one of the major factors that determine its function [9, s 3.6], not just its composition [47]. However, macromolecules are too small to be observed under a light microscope (the typical diameter for a soluble protein monomer is about  $3 - 6 \text{ nm}$  [89], whereas light microscopes can only resolve objects that are at least a few hundred nanometres in size [102, p 54]) so alternative methods must be used to probe the molecular structure. Several techniques to determine 3D molecular structure have been discovered and developed since the beginning of the 20th century. The research presented in this thesis is concerned with method development for the structural techniques of Macromolecular X-ray crystallography (MX) and small angle X-ray scattering (SAXS) applied to protein molecules. In particular, quantitative methods for assessing and correcting radiation damage for these techniques are explored.

### 1.1 Macromolecular X-ray crystallography (MX)

Macromolecular X-ray crystallography (MX) is by far the most common technique for solving the atomic structure of 3D macromolecules. As of 15th January 2016 the Protein Data Bank (PDB) contained 115031 structures and just over 89% of those structures were solved by MX. Despite the dominance of MX, there are many stages in the structure solution pipeline that present their own challenges. The typical protein structure solution pipeline in X-ray crystallography is outlined in Figure 1.1 [40].

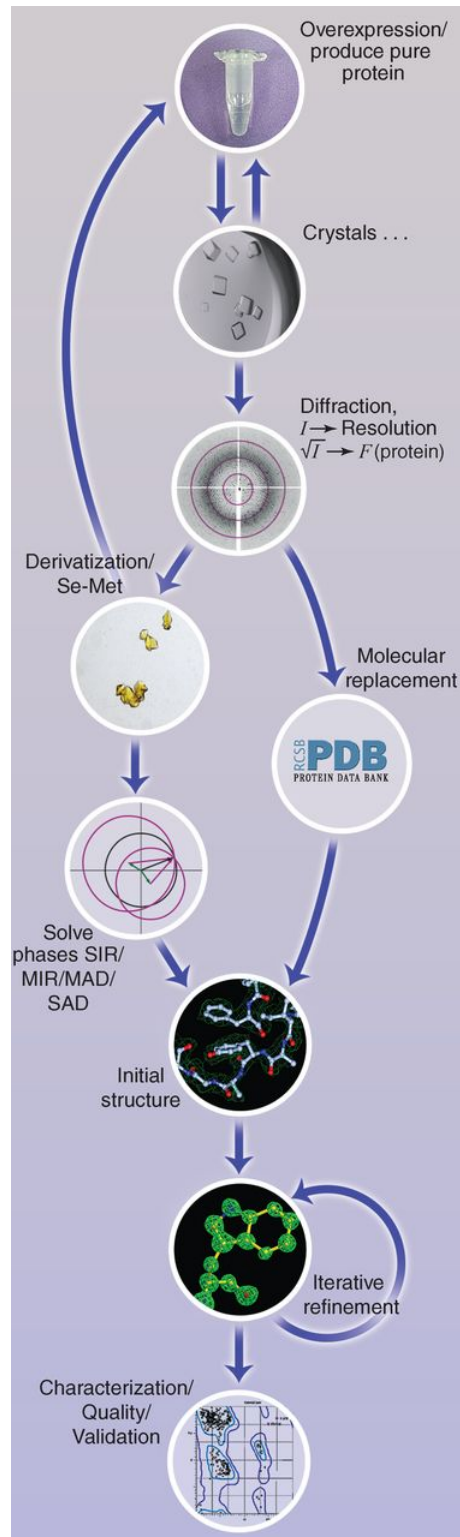


Figure 1.1: Typical protein structure solution pipeline in X-ray crystallography [40].



### 1.1.1 Producing a protein crystal

When a particular protein target has been identified, the first step is to produce it in an appropriate expression system. *E.coli* is the most common system since it is cheap, fast and produces higher quantities of protein compared to other systems [92]. However other expression systems such as yeast, insect and mammalian cells are used when bacterial systems are not suitable e.g. bacterial cells lack the capacity to perform certain types of post-translational modification such as glycosylation. The target protein is generally not the only product formed during the expression phase so it has to be isolated. The process of isolating the target protein is known as purification, whose methods include filtration and chromatography [45]. The next step is crystallisation, which is generally regarded as the major bottleneck in crystallography [40]. At this stage of the pipeline, a solution containing the purified target protein is mixed with a precipitant solution to achieve suitable chemical and environmental conditions for crystallogenesis. However, in general the exact conditions for crystallogenesis for a given protein is unknown, hence many laboratories use robots to screen several conditions in the hope to find the correct one(s) [68].

### 1.1.2 Crystals

A conventional crystal is essentially a 3D repeating array of identical subunits known as unit cells (Figure 1.2a). A unit cell is defined by six parameters -  $a, b, c, \alpha, \beta, \gamma$  (Figure 1.2b).  $a, b, c$  represent the lengths of the edges of the unit cells and  $\alpha, \beta, \gamma$  are the angles between them [31].

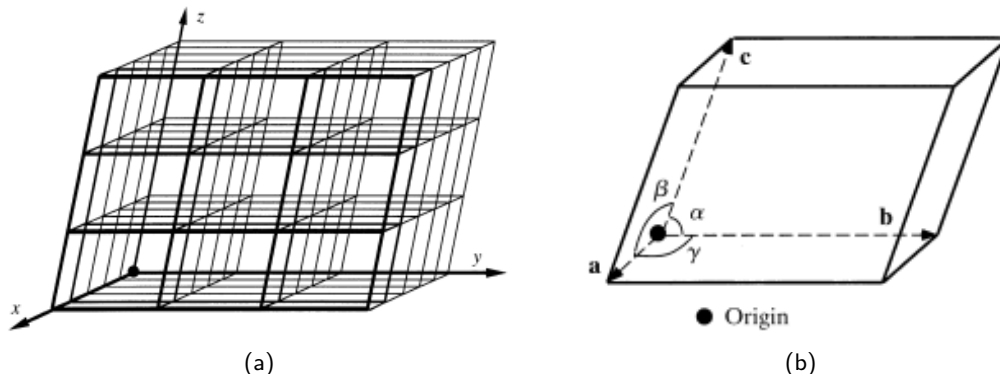


Figure 1.2: (a) Crystals are formed by repeating unit cells related by translations in all 3 dimensions. The intersection points of the lines are the lattice points. (b) A single unit cell with axes  $a, b$  and  $c$ , with angles between them  $\alpha, \beta$  and  $\gamma$ . [31]

If  $a, b, c$  are viewed as directional vectors (vectors will be generally represented by lowercase bold letters  $\mathbf{a}, \mathbf{b}, \mathbf{c}$ ) instead of physical unit cell lengths, and only integer multiples of them are considered, then the resulting set of points is referred to as a (Bravais) lattice. Visually the lattice is a set of points, which can be viewed as the points where the edges of the unit cells intersect each other in Figure 1.2a. Mathematically this is equivalent to:

$$\{n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c} \mid n_1, n_2, n_3 \in \mathbb{N}\}, \quad (1.1)$$

where  $\mathbb{N}$  denotes the set of natural numbers  $\{0, 1, 2, 3, \dots\}$  (Note: sometimes the set of natural numbers is defined as excluding 0 but here it is included). The terms 'crystal' and 'lattice' are sometimes used interchangeably because they are very closely related. However, the distinction between the physical crystal and the hypothetical lattice is made clear by noticing that the set defined in 1.1 implies that

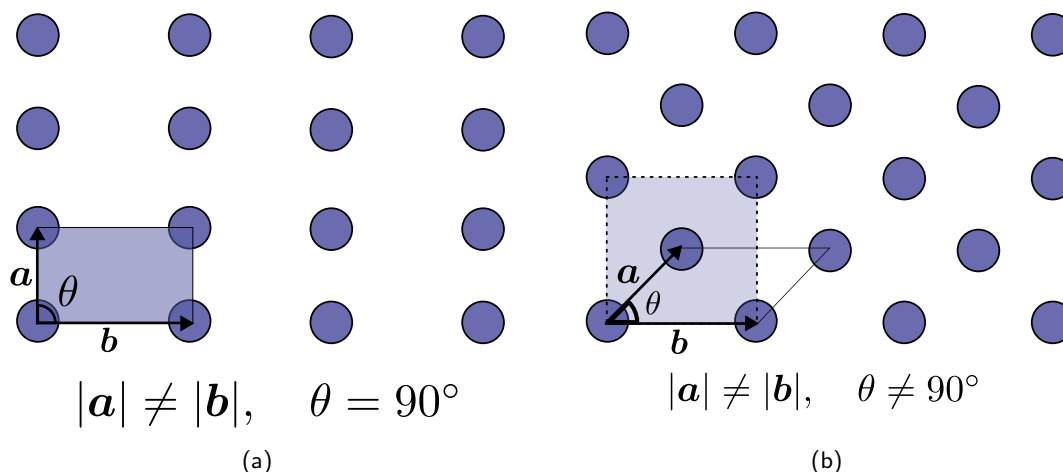


Figure 1.3: Two dimensional Bravais lattices. (a) Rectangular lattice system with no centering. (b) Centred rectangular lattice system.

the array of lattice points is infinite whereas the physical crystal is not. Despite this distinction, the properties of lattices are applicable to crystals.

There are different types of lattices defined by their lattice (axial) system and their types (centerings) (examples given in Figure 1.3). In three dimensions there are 7 lattice systems and 7 lattice types and these combine to make the set of 14 Bravais Lattices [18]. (Note: a stricter way to define the 14 possible Bravais lattices is to realise that two lattices are equivalent if their symmetry groups are isomorphic. Bravais' original criterion only classifies 11 types of lattice [90]).

Thus far only translational symmetries have been considered for unit cells in a crystal. These transformations move every point in space from their original position. However, Bravais lattices also possess symmetry operations that leave at least one point in space fixed. These are called the point group symmetries and consist of reflections and rotations. Protein molecules are chiral and hence symmetry operations that reverse the chirality of the molecule are not allowed for protein crystals (these correspond to the reflections - symmetry operations for which the determinant of the transformation matrix is equal to -1). Further to this, the crystallographic restriction theorem only allows for 2, 3, 4, and 6 fold rotations [26, p 63]. This leaves only 11 enantiomorphic point groups for protein crystals [31].

The combination of fractional translations along a unit cell axis with rotations result in screw axis symmetry operations. When these operations are combined with the point group, the resulting set of operations is referred to as the space group. There are 65 enantiomorphic space groups for protein crystals [31].

### 1.1.3 The diffraction experiment

The final experimental stage of the structure solution pipeline is the data collection experiment. Once a suitable crystal composed of the target protein has been grown, it is collected using a loop that will also keep the crystal hydrated in the solution in which it was grown: the mother liquor. It is then mounted and irradiated with a beam of intense X-rays, typically whilst being rotated, although other collection protocols are becoming increasingly common (e.g. helical scans). The diffracted X-ray photons over a small angular rotation range are collected on a detector and produce a diffraction pattern (a diffraction image) that is unique to that protein (Figure 1.4). The spots that are observed on the images are known as reflections. These observations are in fact the intensities of the reflections. The goal of

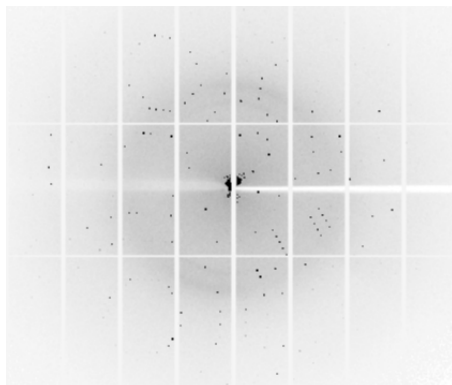


Figure 1.4: Diffraction image from a crystal of bovine pancreatic insulin crystallised in space group  $I2_13$  recorded in January 2014 on beamline P14 at PETRA III synchrotron, Hamburg. Individual spots known as reflections or Bragg peaks are clearly visible.

the data collection experiment is to accurately observe the intensities of as many reflections (unique and multiplicitous) as possible. The space that contains these reflections is known as reciprocal space. Therefore to sample as much of reciprocal space as possible multiple diffraction patterns are collected as the crystal rotates (often the number of images range in the hundreds). Understanding how the diffraction of the X-ray photons arise from their interaction with the atoms in the protein crystal allows crystallographers to interpret the pattern and solve the atomic structure of the protein.

### 1.1.4 Understanding diffraction from a crystal

#### Scattering from single electron

The diffraction of the X-ray beam with the crystal results from the interaction of the electric component of the X-ray beam with the electrons in the crystal. When the beam hits the electron, the electron begins to oscillate. When the scattering is elastic (the desired scattering type for data collection), the electron then emits an X-ray with the same wavelength as the incident X-ray photon. The amplitude of the electric component,  $E_{el}$  of the scattered X-ray photon at a distance  $r$  from a free (unbound) electron is given by

$$E_{el} = E_0 \frac{1}{r} \frac{e^2}{mc^2} \sin(\varphi), \quad (1.2)$$

where  $E_0$  is the amplitude of the electric vector of the incident X-ray photon,  $e$  is the electron charge,  $m$  is the electron mass,  $c$  is the speed of light and  $\varphi$  is the component of  $E_0$  perpendicular to the scattering electron [31].

#### Scattering from two electron system

Suppose a beam of X-rays is incident on a system of two electrons as depicted in Figure 1.5a. The direction of the incident X-ray beam,  $s_0$  is altered after the diffraction event and the direction of the scattered beam is denoted  $s$ . The lengths of these vectors can be arbitrarily chosen. The convenient choice is  $|s| = |s_0| = 1/\lambda$  where  $\lambda$  is the wavelength of the incident beam. The path length difference of the beam scattered by electron 1 and electron 2 is  $p + q = \lambda[r \cdot (s_0 - s)]$ . So the second beam lags behind the first and the resulting phase difference is

$$\frac{-2\pi}{\lambda} \times \lambda[r \cdot (s_0 - s)] = 2\pi r \cdot S, \quad (1.3)$$

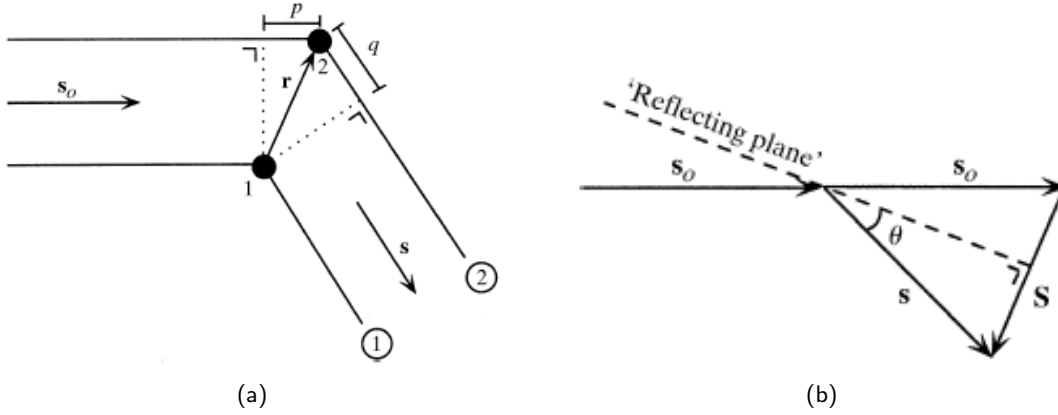


Figure 1.5: (a) Scattering from a system of two electrons labelled 1 and 2. Incident beam direction is denoted  $\mathbf{s}_0$  and the scattered beam direction is denoted  $\mathbf{s}$ . The origin is defined at the position of electron 1. Electron 2 is at a position  $\mathbf{r}$  with respect to electron 1. The path length difference of the beam scattered by electron 1 and electron 2 is  $p + q$ . (b) The angle  $\theta$  is the angle between the incident beam and the reflecting plane.  $\mathbf{S} = \mathbf{s} - \mathbf{s}_0$  and is perpendicular to the reflecting plane. The lengths of the vectors  $\mathbf{s}$  and  $\mathbf{s}_0$  are arbitrary but are chosen to be  $1/\lambda$  for convenience, where  $\lambda$  is the wavelength of the incident X-ray beam. This gives  $|\mathbf{S}| = 2 \sin(\theta)/\lambda$  [30].

where  $\mathbf{S} = \mathbf{s} - \mathbf{s}_0$  [30]. Figure 1.5b is a graphical description showing

$$|\mathbf{S}| = 2 \sin(\theta)/\lambda. \quad (1.4)$$

where  $\theta$  is the angle of reflection of the incident beam with the reflecting plane. It is important to note here that the phase of a wave with respect to another one is dependent on the relative position of the electrons. This demonstrates the importance of the phases for generating accurate structural information [105, 106].

The amplitudes of the scattered waves from both electrons in the system is the same, they only differ in phase. If the amplitude of the scattered wave is equal to 1, then the total scattered wave from the system is  $1 + 1 \times \exp[2\pi i \mathbf{r} \cdot \mathbf{S}]$  where  $i$  is the imaginary number,  $i = \sqrt{-1}$ .

### Scattering from an atom

The electron cloud of an atom scatters the incident X-ray beam. The strength of the scattered beam is dependent on the number of electrons and their position in the electron cloud. The electrons in an atom are not free as was assumed previously, but continuing to treat them as free electrons gives sufficient accuracy provided the wavelength of the incident beam is not too close to an absorption edge for the atom [31] (which generally is not the case for most data collection experiments). Setting the origin of the system at the centre of the atom, the total scattering from the atom is calculated as

$$f = \int_{\mathbf{r}} \rho(\mathbf{r}) \exp[2\pi i \mathbf{r} \cdot \mathbf{S}] d\mathbf{r}, \quad (1.5)$$

where  $\rho(\mathbf{r})$  is the electron density at position  $\mathbf{r}$  and the integral is over all space.  $f$  is known as the atomic scattering factor [30].

The electron density of an atom is assumed perfectly centrosymmetric i.e.  $\rho(\mathbf{r}) = \rho(-\mathbf{r})$  (Figure 1.6a). So equation 1.5 can be simplified to

$$f = 2 \int_{\mathbf{r} \text{ half space}} \rho(\mathbf{r}) \cos[2\pi \mathbf{r} \cdot \mathbf{S}] d\mathbf{r}. \quad (1.6)$$

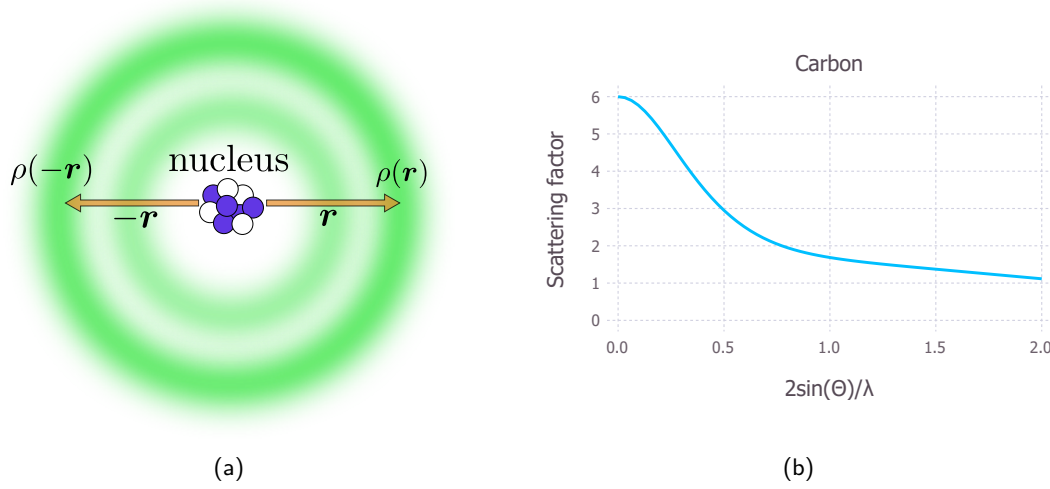


Figure 1.6: (a) An atom with a centrosymmetric electron density distribution (i.e.  $\rho(\mathbf{r}) = \rho(-\mathbf{r})$ ) about the centre of the nucleus. (b) atomic scattering factor,  $f$ , of carbon as a function of  $|\mathbf{S}| = 2 \sin(\theta)/\lambda$ .  $f$  is expressed in number of electrons. Notice an angle of  $\theta = 0$ ,  $f = 6$  corresponds to the atomic number of carbon. For this figure,  $0^\circ \leq \theta \leq 90^\circ$  and  $\lambda = 1.0$ .

The integral is only over half of the entire space and the quantity does not contain any imaginary terms, hence the atomic scattering factor is a real quantity.

The atomic scattering factor of an atomic is only dependent on the length of the vector  $\mathbf{S}$  (equation 1.4). The dependence of the atomic scattering factor,  $f$ , on  $|\mathbf{S}|$  is shown for carbon in Figure 1.6b.

### Scattering from a unit cell

Imagine an atom placed in a unit cell where the origin of system is placed at the origin of the unit cell. Now the atom has positional vector  $\mathbf{r}$  to define its location. The scattering from this atom is now

$$f = f \exp[2\pi i \mathbf{r} \cdot \mathbf{S}]. \quad (1.7)$$

Now suppose there are  $N$  atoms in the unit cell. The total scattering from the unit cell is now

$$\mathbf{F}(\mathbf{S}) = \sum_{j=1}^N f_j \exp[2\pi i \mathbf{r}_j \cdot \mathbf{S}]. \quad (1.8)$$

where  $\mathbf{r}_j$  and  $f_j$  are the position and atomic scattering factor of the  $j$ -th atom respectively.  $\mathbf{F}(\mathbf{S})$  is referred to as the structure factor because it depends on the structure of the atoms within the unit cell [31].

### Scattering from a crystal

Suppose an origin is chosen to be at an arbitrary unit cell but the scattering from a different unit cell from the same crystal is to be calculated. Unit cells are related by translations along the unit cell basis vectors  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\mathbf{c}$  as described in section 1.1.2. So the unit cell of interest is located at position  $t\mathbf{a} + u\mathbf{b} + v\mathbf{c}$ , where  $t, u, v \in \mathbb{Z}$  i.e.  $t, u, v$  are integers. So the scattering from this unit cell is

$$\mathbf{F}(\mathbf{S}) \times \exp[2\pi i t \mathbf{a} \cdot \mathbf{S}] \times \exp[2\pi i u \mathbf{b} \cdot \mathbf{S}] \times \exp[2\pi i v \mathbf{c} \cdot \mathbf{S}]. \quad (1.9)$$

---

If this is extended to include all unit cells then the total wave scattered by the crystal is

$$\mathbf{K}(\mathbf{S}) \times \sum_{t=0}^{n_1} \exp[2\pi i t \mathbf{a} \cdot \mathbf{S}] \times \sum_{u=0}^{n_2} \exp[2\pi i u \mathbf{b} \cdot \mathbf{S}] \times \sum_{v=0}^{n_3} \exp[2\pi i v \mathbf{c} \cdot \mathbf{S}], \quad (1.10)$$

where  $n_1, n_2, n_3$  are the number of unit cells along the direction of the basis vectors  $\mathbf{a}, \mathbf{b}$  and  $\mathbf{c}$  respectively. Since  $n_1, n_2$  and  $n_3$  are typically very large, the summations  $\sum_{t=0}^{n_1} \exp[2\pi i t \mathbf{a} \cdot \mathbf{S}]$ ,  $\sum_{u=0}^{n_2} \exp[2\pi i u \mathbf{b} \cdot \mathbf{S}]$  and  $\sum_{v=0}^{n_3} \exp[2\pi i v \mathbf{c} \cdot \mathbf{S}]$  are usually equal to zero unless  $\mathbf{a} \cdot \mathbf{S} = h$ ,  $\mathbf{b} \cdot \mathbf{S} = k$  and  $\mathbf{c} \cdot \mathbf{S} = l$  where  $h, k, l \in \mathbb{Z}$ . So a crystal scatters X-rays if

$$\mathbf{a} \cdot \mathbf{S} = h, \quad (1.11)$$

$$\mathbf{b} \cdot \mathbf{S} = k, \quad (1.12)$$

$$\mathbf{c} \cdot \mathbf{S} = l. \quad (1.13)$$

These are known as the Laue conditions [30]. The  $h, k, l \in \mathbb{Z}$  are referred to as Miller indices and they define individual reflections (Figure 1.4). The result is that the amplitude of the total scattered wave from the crystal is proportional to the structure factor  $\mathbf{F}(\mathbf{S})$  and the number of unit cells in the crystal [30].

### 1.1.5 From diffraction patterns to electron density - The theory

The goal of structure determination is to obtain the atomic structure of the target molecule. So given that it is understood how X-ray diffraction arises from the interaction with a crystal, the inverse problem must be solved i.e. determining the atomic structure from the diffraction pattern.

Recall equation 1.8 calculates the structure factor as a summation over atoms in the unit cell. It is possible to write this as an integral over the electron density in the cell instead, giving

$$\mathbf{F}(\mathbf{S}) = \int_{\text{cell}} \rho(\mathbf{r}) \exp[2\pi i \mathbf{r} \cdot \mathbf{S}] d\mathbf{v}. \quad (1.14)$$

Introducing fractional coordinates  $x, y, z$  (i.e.  $0 \leq x < 1$  and similarly for  $y$  and  $z$ ) and given the unit cell has volume  $V$ , the volume element,  $d\mathbf{v}$ , can be rewritten as

$$d\mathbf{v} = V \times dx dy dz. \quad (1.15)$$

The position  $\mathbf{r}$  can also be rewritten as  $\mathbf{r} = \mathbf{a}x + \mathbf{b}y + \mathbf{c}z$ , so this means

$$\mathbf{r} \cdot \mathbf{S} = \mathbf{a} \cdot \mathbf{S} x + \mathbf{b} \cdot \mathbf{S} y + \mathbf{c} \cdot \mathbf{S} z \quad (1.16)$$

$$= hx + ky + lz. \quad (1.17)$$

So  $\mathbf{F}(\mathbf{S})$  can be written as a function of the Miller indices  $\mathbf{F}(h, k, l)$  giving

$$\mathbf{F}(h, k, l) = V \int_{x=0}^1 \int_{y=0}^1 \int_{z=0}^1 \rho(x, y, z) \exp[2\pi i(hx + ky + lz)] dx dy dz. \quad (1.18)$$

Equation 1.18 shows explicitly that the structure factor  $\mathbf{F}(h, k, l)$  is the Fourier transform of the electron density  $\rho(x, y, z)$ . Since the Fourier transform is the inverse of itself, it means that the electron density in the unit cell can be obtained by taking the Fourier transform of the structure factor

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |\mathbf{F}(h, k, l)| \exp -2\pi i(hx + ky + lz) + i\alpha(h, k, l) \quad (1.19)$$

where  $|\mathbf{F}(h, k, l)|$  is the structure factor amplitude and  $\alpha(h, k, l)$  is the phase. Equation 1.19 is known as the electron density equation and accurately calculating this is the ultimate goal of structure

---

determination in MX. This is because it gives the density of electrons at every point in space in the unit cell. However, this equation can only be calculated once the amplitudes and phases are known. Notice that the summation is over reflections ( $hkl$ s) which explains why it is desirable to accurately collect as many reflections as possible in the data collection experiment. The more (accurate)  $hkl$  terms that are used in the electron density equation, the better the Fourier series representation of the electron density in the unit cell (i.e. the better the electron density map). The amplitudes can be derived from the experimentally observed intensities in the data collection experiment but the phases are lost. This is known as the phase problem [106].

### 1.1.6 From diffraction patterns to electron density - In practice

In practice, much of the theory is abstracted from the crystallographer and many software programs perform the necessary calculations 'behind the scenes'.

A set of diffraction images from a data collection experiment is known as a dataset. Extracting the amplitudes from the dataset is the ultimate aim of the first series of programs. The first step is to calculate the intensities of the observations from the images programs. Programs such as MOSFLM [67], XDS [60] and more recently DIALS, are used to find the spots on the images, assign Miller indices to each reflection (indexing) and finally integrate the intensity of the observations on the images. Commonly it is the case that a single observation is observed on multiple images because the rotation range of the crystal that results in a diffraction image (the oscillation angle) does not sample the entirety of a reflection. The fraction of the reflection that is recorded on a given image, the partiality of a reflection, has to be calculated and the experimental parameters refined in a process called post-refinement [97,98].

Although the space group of the crystal is determined during the indexing stage, it is still only a guess. A more informed guess can be made about the space group after the integration stage with a program called POINTLESS [33].

The intensities of reflections are affected not only by the number of electrons in the unit cell but also by the variation of other systematic factors such as the rotation rate, incident beam intensity, the path length of the X-ray beam through the crystal, secondary absorption etc. [32]. These factors must be taken into account so that the intensities of all reflections are on the same scale. To achieve this the difference between intensities of reflections that should be identical according to the symmetry of the crystal are used to estimate scaling parameters in a process called scaling. This is carried out by programs such as AIMLESS [34] and XSCALE [60].

Now that the intensity estimates have been put on an internally consistent scale it is possible to derive the amplitudes. In theory the amplitudes are equal to the square root of the intensity. However the conversion is not as straight forward as this because subtraction of background noise in the diffraction images during the integration step leads to weakly measured intensity observations having negative values. This is a problem because it has been established that the amplitude should be a real quantity. One way to deal with negative intensity reflections is to set the intensity values to zero. However French and Wilson developed a treatment that uses Bayesian analysis [37] and Wilson statistics [112] to calculate a better estimate of the amplitude given that a negative intensity has been observed. Programs such as CTRUNCATE [33] perform this analysis and calculate further data assessment statistics to check for crystal pathologies such as twinning.

Although phases are not observed directly in the diffraction experiment, they can be derived in multiple ways. If the sequence of the target protein is similar to another structure that has already been solved (about 30% [114]) or the structure is thought to be similar to another structure, then initial phases can be taken from the homologous structure. First, the homologue structure is transformed via a series of rotations and translations so the amplitudes best match the ones that have been obtained

---

experimentally. Programs such as PHASER perform this sort of analysis [69].

If no structural homologue exists then the phases must be derived experimentally. Isomorphous replacement is a method in which experimental data is collected on the native crystal and then again with a crystal in which a heavier element has been incorporated into the crystal without changing the structure of the atoms in the unit cells [87]. The phases can then be derived by analysing the difference in the intensity observations [106].

Another method known as anomalous dispersion can also be used in which a heavy element is incorporated into the structure (sometimes this is not necessary if the structure already contains enough elements like sulphur) and multiple datasets are collected at different X-ray photon energies. Some of the datasets are collected at close to the absorption edge of the heavy element so appreciable differences between the native dataset and the new dataset can be measured [10]. If only 1 dataset is collected in this way, this is known as single wavelength anomalous diffraction/dispersion (SAD) and differences in intensities are analysed between Friedel pairs of reflections. If more datasets are collected it is known as multiple wavelength anomalous diffraction/dispersion (MAD) [50, 106].

With both amplitudes and phases obtained, it is possible to perform the Fourier transform defined in equation 1.19 to get an electron density map. The atomic structure can then be built into the resulting map. The initial map is unlikely to provide a satisfactory structure in general and so a process of refinement is carried out to improve the agreement of the amplitude calculated from the model that has been built, with those that were derived from the experimental intensities. Programs such as REFMAC [76] and PHENIX.REFINE [3] perform this analysis until a satisfactory structure has been obtained.

## 1.2 Small Angle X-ray Scattering (SAXS)

Small angle X-ray scattering (SAXS) is a technique used to determine the macromolecular structure, which like X-ray crystallography, requires a sample containing the molecule to be irradiated with a beam of X-rays. However, unlike X-ray crystallography, SAXS does not require the growth of crystals. Instead, the protein solution is irradiated and the diffracted X-rays are collected on a detector. In solution the movement of molecules is not restricted and they can adopt random orientations with respect to one another [12]. Therefore the data observed on the detector are not individual Bragg peaks. Although SAXS does not provide atomic level detail of a structure, it does give information about the interatomic distances between atoms, an overall envelope of the structure and the molecular weight of the molecule [91].

The X-ray scattering intensity is radially isotropic and hence it is determined by taking a radial average of the intensities recorded on the detector image [36]. The radially averaged intensity is usually written as a function of the momentum transfer,  $q$ , which itself is defined in terms of the scattering angle:

$$q = \frac{4\pi \sin(\theta)}{\lambda} \quad (1.20)$$

The final scattering pattern is the result of subtracting the radially averaged intensity of the solution containing the protein with the radially averaged intensity from the solution containing the buffer without protein (Figure 1.7).

Several graphs are produced during the analysis of SAXS data which can give a lot of information about the state of the protein. For example, Kratky plots give information about whether or not a protein is in a folded state (Figure 1.8a). The Fourier transform of the SAXS pattern gives the pair-distance distribution function, a function that provides information about the interatomic distances in the protein:

$$p(r) = \frac{r^2}{2\pi^2} \int_0^\infty q^2 I(q) \frac{\sin(qr)}{qr} dq. \quad (1.21)$$



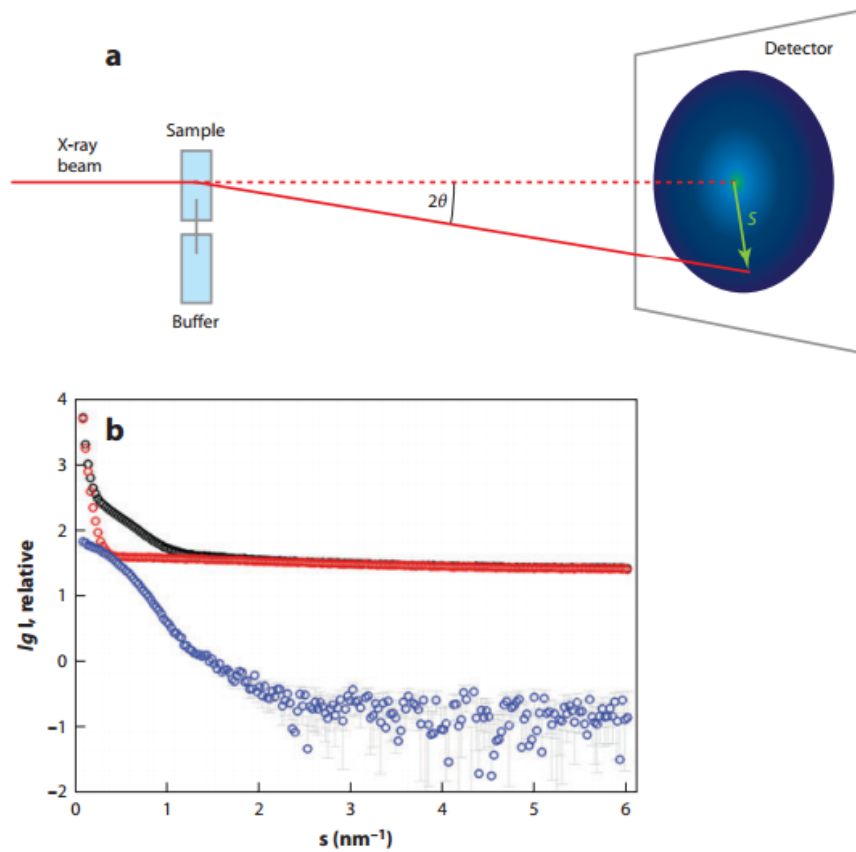


Figure 1.7: (a) X-ray beam is incident on a SAXS sample and the scattered radiation is collected on a detector. The symbol  $s$  in the figure is equivalent to the momentum transfer denoted  $q$  in equation 1.20. (b) Radially averaged intensity curves from a solution of bovine serum albumin (BSA). The radially averaged intensity of the solution containing only the buffer (red curve) is subtracted from the radially averaged intensity of the solution containing BSA and buffer (black curve) to obtain the resulting intensity curve for BSA (blue curve) [12].

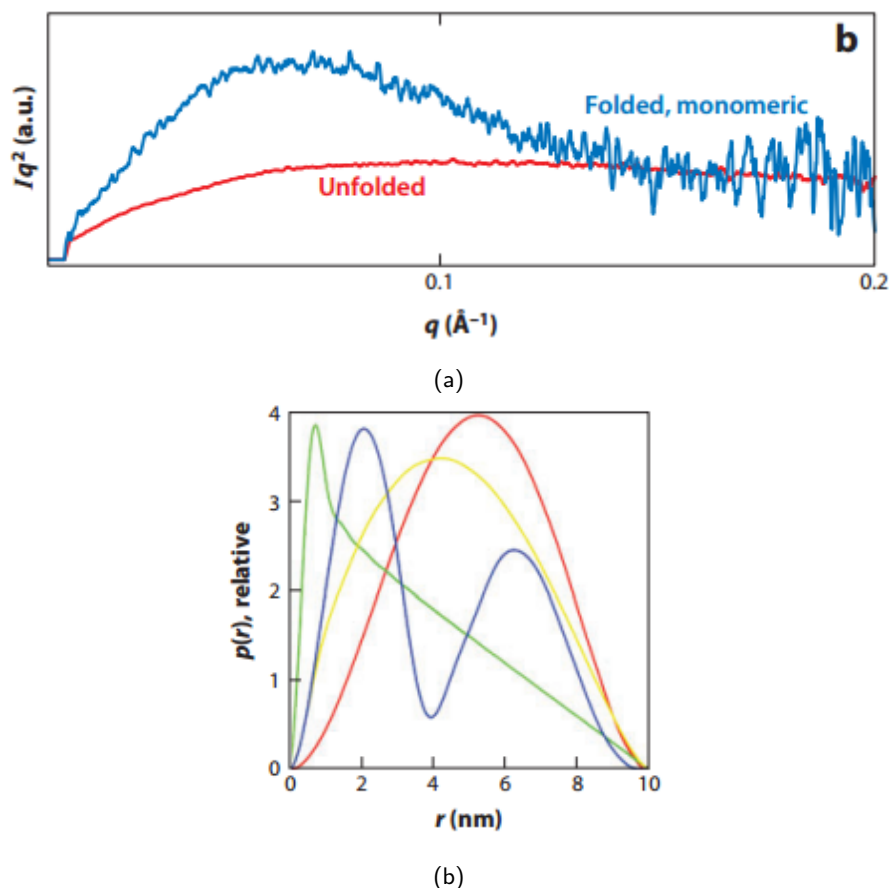


Figure 1.8: (a) Kratky plots. The red curve shows unfolded RNA domain P4-P6 from the Tetrahymena ribozyme where the salt concentration was low. Blue curve implies the domain was folded and monomeric [91]. (b) Distance distribution functions for several geometrical volumes: sphere (red), dumbbell (blue), cylinder (green) disk (yellow) [12].

where  $r$  is the distance between electrons in the molecule. Examples of distance distribution functions for various geometrical shapes is shown in Figure 1.8b.

### 1.3 Limitations and alternatives to MX and SAXS

Despite the dominance of MX as a method for structural analysis and the recent emergence of SAXS as a reliable method to provide complimentary structural and functional information, both methods have drawbacks. It has already been mentioned that producing protein crystals is a bottleneck of the pipeline in MX (section 1.1.1). This does not account for the necessity to grow suitably sized, well diffracting crystals. Often it is the case that there is a limit to the size and quality of the crystal that can be grown. Hence the hardware and software for data collection is constantly being developed to handle these situations. Theoretical work on the minimum crystal size for which sufficient data can be collected on a single crystal in a typical synchrotron radiation experiment [53] has inspired work on various synchrotron beamlines to achieve this limit. Another drawback of MX is the fact that it is essentially a method that only provides a temporal and spatial average of the molecule so no dynamical information is readily obtained from the experiment. However a relatively new technique making use of the Hadamard transform is being developed that promises to give dynamic structural information

---

about the molecule [113].

With regards to SAXS, the major disadvantage is that it does not provide atomic resolution information. Therefore relatively subtle structural dynamics of molecules will not be uncovered from this technique.

Several alternatives to MX and SAXS exist. These include X-ray free electron lasers (XFELs) and single particle cryo-electron microscopy (cryo-EM). In XFEL experiments very intense beams of X-rays irradiate crystals over a very short time period (about  $10^{12}$  photons in a single pulse in about 10 femtoseconds ( $fs$ ) [24]). XFELs can obtain high resolution diffraction information from nanocrystals. However the technique is still relatively immature. The data processing is not trivial (the crystals not being rotated during exposure and unknown and dynamic X-ray beam wavelengths per pulse are just a couple of the factors complicating data processing), and access to XFELs is very limited. Cryo-EM, like SAXS, does not require the production of crystals but can additionally provide atomic resolution detail due to recent hardware and software advances [7]. In cryo-EM a set of projection images of single molecules is taken with an electron microscope (after some sample preparation). The images are then computationally combined to obtain a density distribution of the molecule [72]. Again, cryo-EM comes with its drawbacks. Obtaining high resolution structures from small ( $< 200 - 300\text{ kDa}$ ), unstable or flexible molecules is not trivial and it is not yet accessible enough [7].

## 1.4 Radiation damage in MX

A common theme amongst all of the structural techniques mentioned thus far is that the sample is probed with ionising radiation. This results in the sample suffering radiation damage (even samples studied at XFEL sources also show signs of radiation damage with a long enough pulse [77]). Being the most mature of these structural techniques however, much more literature exists about radiation damage for synchrotron based MX. Radiation damage is a problem because limits the amount of useful data that can be collected from a single crystal during an MX experiment [39]. It is therefore the major cause of unsuccessful data collection at synchrotron sources given a sufficiently diffracting crystal [115].

### 1.4.1 Types of X-ray interactions with atoms: Primary damage

In a fairly typical MX experimental set-up only a small fraction of the incident X-ray photons will interact in any way with the atoms in the crystal. Garman reported that for a  $100\text{ }\mu m$  thick protein crystal only 2% of the incident photons of a  $12.4\text{ keV}$  ( $1\text{ }\text{\AA}$ ) incident beam will interact with it [39]. This section will explain the possible interactions that occur and how they can give rise to the data or result in radiation damage

#### Elastic scattering

Elastic scattering (also referred to as Rayleigh, Thompson or coherent scattering) is a type of interaction in which no energy from the incident X-ray photon is deposited in the sample [78]. The resulting scattered waves interfere to give rise to the observed diffraction pattern. It is the type of interaction that the experimenter would like to maximise. However, of the 2% of the incident X-ray photons that interact with the sample, only about 8% of that results in elastic scattering [94].

---

## Compton Scattering

Another 8% of the interaction is due to Compton scattering. This type of scattering occurs when the X-ray photon scatters incoherently from the crystal, thereby transferring some of its energy to an electron. The resulting scattered X-ray photon leaves with less energy (higher wavelength). It is possible for the recoil electron (the name given to the electron in which the energy is deposited) to also be ejected from the atom [78].

## Photoelectric effect

By far the most common interaction is the photoelectric effect which is responsible for the other 84% of the X-ray-electron interactions. In this type of interaction, the incident X-ray photon is completely absorbed and an inner shell electron is ejected [39]. The ejected electron is called a photoelectron. The vacancy left by the ejection of the photoelectron is filled by another electron. This transition can lead to two different outcomes: either the production of a characteristic X-ray, known as fluorescence, or the ejection of an electron from an outer shell [78]. The ejected electron is known as the Auger electron. The relative probability of either fluorescence or Auger emission occurring is dependent on the atomic species. With lighter elements that constitute the vast majority of protein molecules the more likely process is Auger emission.

An ionisation event caused by the photoelectric effect or Compton scattering is referred to as primary damage, although the use of this term varies in the literature [39]. The relative probability of each of the three interactions (cross section) described above varies with incident X-ray beam energy (Figure 1.9). The atomic number also affects the amount of energy absorption in the sample with heavier elements contributing more to the overall absorption per atom. To maximise the amount of elastic scattering it would initially seem advisable to increase the energy of the incident X-ray beam from the range of typical values used in X-ray crystallography, usually around 12 *keV* [85]. However, the diffracted intensity per incident photon is lower for incident beams with higher energy. In practice, the optimal energy for an MX experiment is highly dependent on the sample composition and the aims of the experiment (i.e. the design may be different if the experiment aims to collect suitable data for phasing) and no general consensus has been established [85].

### 1.4.2 Secondary damage

In contrast to primary damage (ionisation due to the incident photon), secondary damage is the damage caused by the energetic electrons produced as a result of the former. In fact it is secondary damage that is responsible for the majority of the damage that is observed in MX experiments. A single photoelectron has enough energy to cause  $\approx 500$  ionisation events [80]. The ionisation event is referred to as direct damage if the event occurs within the protein molecule. If the event occurs within the surrounding mother liquor then it is referred to as indirect damage. The radical species produced as a result of these events can diffuse through the crystal disrupting the structure of the proteins by breaking chemical bonds, causing redox processes and producing more radical species [70]. Some of the radical species produced are hydroxyl radicals, electrons and hydrogen ions [39]. At temperatures below  $\approx 180\text{ K}$  (the glass transition) it is thought that the only mobile radical species are electrons [58] due to the solvent possessing a quasi-infinite viscosity [109]. Electrons are able to overcome energy barriers by quantum mechanical tunnelling mechanisms [41] and cause the majority of the damage observed in MX experiments [39].

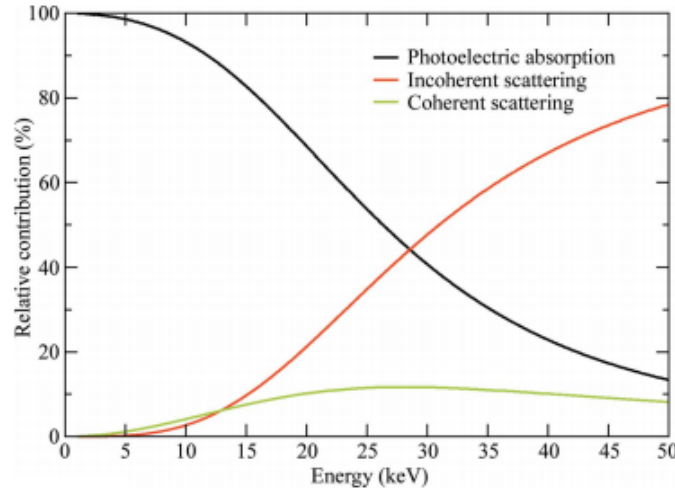


Figure 1.9: Relative contribution to the overall X-ray cross section as a function of incident beam energy for chicken egg-white lysozyme [85]. The photoelectric effect dominates the interaction cross section at typical wavelengths used in MX (about 12 *keV*). However, at higher energies the elastic contribution increases but the Compton effect increases faster and begins to dominate.

### 1.4.3 Quantifying energy absorbance: Dose

Being able to quantify the amount of energy absorbed by the crystal is vital to track the progression of radiation damage in an MX experiment [11, 52]. Dose is the metric by which the energy absorbed in the crystal is quantified. The dose is defined as the energy absorbed per unit mass and the SI unit used in MX is the gray (*Gy*), where 1 *Gy* = 1 *J kg*<sup>-1</sup>. In typical cryo-MX experiments it is common for a crystal to receive a dose that ranges in the order of millions of gray (*MGy*) [39] due to the brilliance of third generation synchrotron sources [73]. However, due to the complex processes (e.g. the stochastic nature of the absorption events, electron cascades, etc.), direct measurement of the dose is infeasible and so it must be calculated instead. The formula used to calculate the dose for a small volume of sample is

$$Dose = \frac{E_{incident}}{M_{vol}} (1 - e^{-\mu_{abs}z}), \quad (1.22)$$

Where  $E_{incident}$  is the energy incident on the sample,  $M_{vol}$  is the mass of the irradiated volume and  $1 - e^{-\mu_{abs}z}$  is the fraction of the incident beam absorbed through the volume of thickness  $z$ ,  $\mu_{abs}$  is the absorption coefficient of the sample (a value dependent on the atomic cross sections that determines the likelihood of absorbance of the X-ray beam) and  $e = 2.718\dots$  is Euler's number [114].

RADDOS is a software program that was developed to improve the ease of calculating the absorbed dose within a protein crystal [75]. It increases the accuracy by implementing a 2D model of the experiment as opposed to the 1D model in equation 1.22. Information about the crystal (size, unit cell parameters, number of molecules per unit cell, number of residues, solvent content), the beam (energy, flux, and profile), exposure time per image and the total number of images is required for the calculation. RADDOS uses these parameters to determine the maximum beam intensity and simulates an experiment by which the crystal is exposed homogeneously to the beam to provide a worst case scenario dose estimate: the maximum dose. Several improvements to the model were incorporated in successive versions of RADDOS. Version two [86] calculated the fluorescent emission probability and the probability that these fluorescent photons could escape the crystal. Version three [85] took into account energy loss in the crystal due to the Compton Effect.

Despite these improvements RADDOS has several drawbacks. One of the major problems with the dose calculation is that some parameters of the data collection experiment are known with more

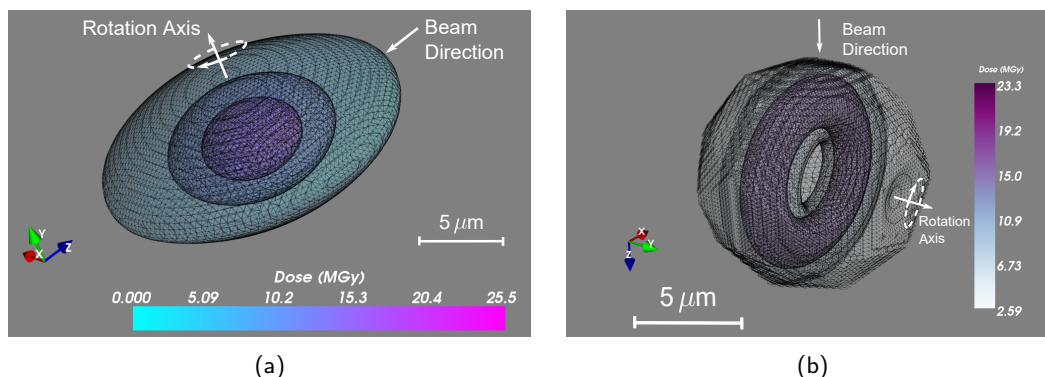


Figure 1.10: Dose distributions in crystals represented as a polyhedron. (a) An oblate ellipsoid ( $x = z = 10 \mu\text{m}$ ,  $y = 6 \mu\text{m}$ ) crystal that has been irradiated with a Gaussian profile beam with a full width half maximum (FWHM) of  $5 \mu\text{m} \times 5 \mu\text{m}$ , flux of  $10^{10}$  photons per second and an incident photon energy of  $12.7\text{keV}$  for a rotation of  $360^\circ$ . The collimation was set to  $20 \mu\text{m} \times 20 \mu\text{m}$  and hence the entire volume of the crystal was irradiated. The 3 visible contours represent dose values of  $20 \text{ MGy}$ ,  $15 \text{ MGy}$  and  $5 \text{ MGy}$  from inner to outer contour levels. (b) An icosahedral crystal with maximum dimensions of  $10 \mu\text{m}$  in each of the  $x$ ,  $y$  and  $z$  dimensions, which has been irradiated with a Gaussian profile beam with FWHM  $4 \mu\text{m}$  in both the  $z$  and  $y$  directions, flux  $2 \times 10^{10}$  photons per second and an incident photon energy of  $12.7 \text{ keV}$ . The rotation axis was offset  $5 \mu\text{m}$  from the centre of the crystal in the  $x$ -direction and the crystal was rotated  $360^\circ$  for a total exposure time of 180 seconds. Contouring levels correspond to  $2.59 \text{ MGy}$ ,  $12.5 \text{ MGy}$  and  $18.75 \text{ MGy}$  from inner to outer contour levels. It is evident that different crystal, beam and experiment design parameters lead to very different distributions of absorbed dose within a protein crystal.

certainty than others. Parameters such as the crystal composition, beam flux and exposure time are known to a relatively high degree of accuracy. However, the extent to which parameters such as the crystal volume [52] and beam profile [64] are known can be significantly less. This can lead to large errors in the calculated dose values and inferences about the radiation susceptibility of protein crystals [64]. The uncertainty can be accounted for by assigning a “factor of 2” decision threshold [52]. This means that a particular radiation damage effect described by quantitative values (e.g. intensity loss or dose) can be considered the same if the values are within a factor of 2 of each other. Methods are being developed to better parametrise the data collection experiment for the crystal shape [19, 61, 103] and the beam [17] but these methods are not routine on the majority of beamlines.

Another problem encountered with the calculation of dose in RADDOSE is the two-dimensional nature of the model. It means that it lacks the three-dimensional information about the damage state of the crystal. This information is important for optimising the use of crystal volume to collect better quality data [115, 116]. Furthermore it assumes that the crystal is always immersed within the beam and does not take into account crystal rotation.

The geometric issues are addressed in the RADDOSE-3D [117] software program, a successor to RADDOSE. RADDOSE-3D takes into account the three-dimensional geometry of the MX experiment to provide a temporally and spatially resolved dose distribution within a protein crystal. Initially RADDOSE-3D was written to model cuboid or spherical crystals, but now it has been extended to be able to capture any polyhedron (Figure 1.10).

The raw output of RADDOSE-3D is a 3D scalar field which assigns a dose to every voxel (3D volume element) in the crystal. In its raw form, these data are not very useful for the experimenter to assess the damage state of the crystal, especially when the dose distributions are highly inhomogeneous (Figure 1.10). So several metrics were proposed to provide useful summaries of the data. These include the maximum dose (the same dose value as output by the original versions of RADDOSE),

---

average dose for the whole crystal and the dose inefficiency (i.e. the maximum dose divided by the absorbed energy), amongst others [116]. The metric that was found to be most promising in faithfully representing the damage state of the crystal is the diffraction weighted dose (DWD) [115]. The DWD is a weighted average, where the weight at each voxel position in the crystal is given by the X-ray fluence through that voxel. It considers the effective dose absorbed by the crystal and its impact on the diffraction pattern for any given image of the dataset. Mathematically the DWD is defined for each image as

$$DWD^i = \frac{\int_{t_{i-1}}^{t_i} \int_{crystal} D(\mathbf{x}, t) \times F(\mathbf{x}, t) \, d\mathbf{x} dt}{\int_{t_{i-1}}^{t_i} \int_{crystal} F(\mathbf{x}, t) \, d\mathbf{x} dt} \quad (1.23)$$

where  $i$  is the image number,  $t$  is the time,  $\mathbf{x}$  is the position in the crystal,  $D$  is the total cumulative absorbed dose ( $Gy$ ) at that position, and  $F$  is the fluence (photons per unit area).

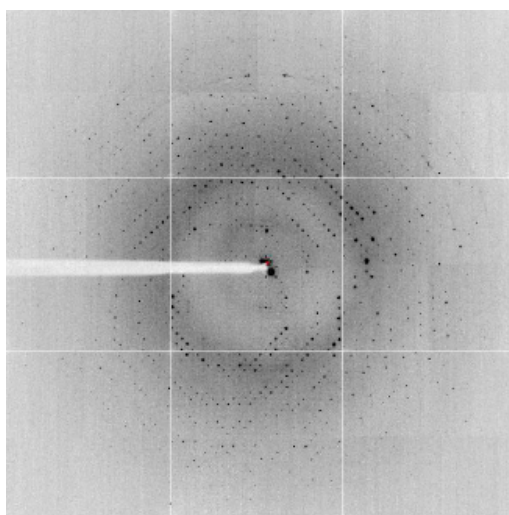
The DWD metric only takes into account the fluence incident on a unit volume during the collection of data for a given image, but does not take into account the loss of diffraction efficiency of the volume due to global radiation damage. As the dose increases, a particular volume of the crystal will only contribute to the background [11].

#### 1.4.4 Manifestations of damage: Global damage

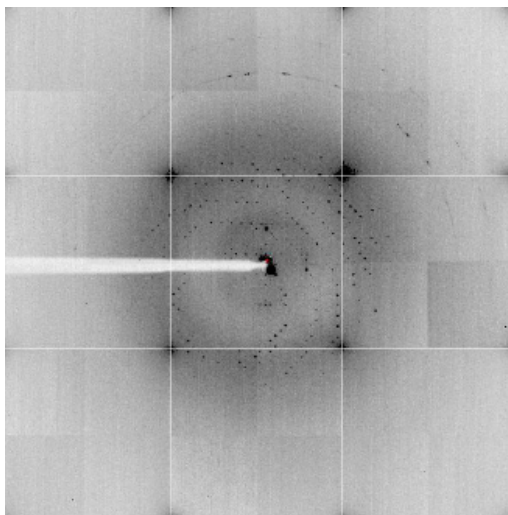
The damage processes described in sections 1.4.1 and ?? are observed in a variety of ways in MX which are classified accordingly. Global radiation damage is observed in reciprocal space and results in changes in mosaicity, increase in the scaling and atomic  $B$  factors, unit cell expansion, increases in data quality  $R$  values ( $R_{merge}$ ,  $R_{meas}$ ,  $R_{pim}$ ) and decreasing  $CC_{1/2}$  and  $CC^*$  [39]. Perhaps the most iconic symptom of global radiation damage is the loss of intensity of reflections in the diffraction pattern, with the reflections corresponding to higher resolution information fading the quickest (Figure 1.11). Given that the diffraction images constitute the experimental data, it is this loss of reflection intensity that is generally attributed to the failure of structure solution as a result of radiation damage. This is particularly problematic in data collection for experimental phasing where multiple datasets are collected for comparison of reflection amplitudes. For experimental determination of phases in isomorphous replacement or anomalous scattering experiments, it is necessary to distinguish intensity changes of around 4% [106]. However it was calculated that for a 0.5% change in all unit cell dimensions or a  $0.5^\circ$  rotation about a single axis of a  $100\text{\AA}^3$  unit cell, there would be a change in the intensity of a  $3\text{\AA}$  general reflection by 15% and 16% respectively [27].

To observe appreciable changes in reciprocal space, radiation damage must be affecting the long range crystalline order of the crystal [70]. At room temperature, many more radical species are mobile and damage occurs much more quickly than at cryo-temperatures ( $100\text{ K}$ ) [48, 110]. However, global radiation damage still occurs at cryo-temperatures and it has been proposed that this is due to hydrogen gas build up. This is because diffusion rates are smaller at lower temperatures. This build up of hydrogen exerts a disruptive force on the crystalline structure leading to the reduction in structural integrity [70].

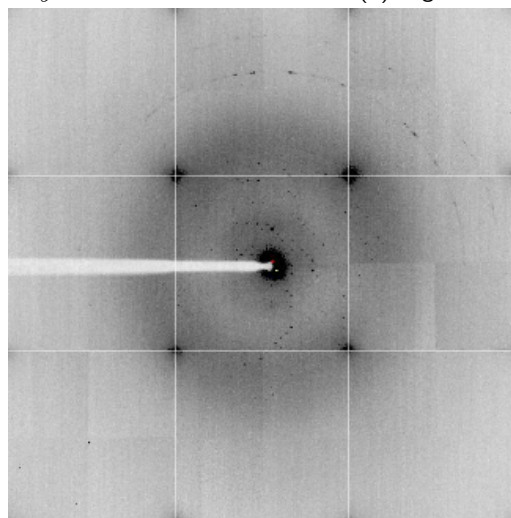
The various symptoms or variants of them have been proposed as metrics to assess the extent of radiation damage as a function of the dose (Figure 1.12). An ideal metric would change reproducibly and monotonically with the increase in the dose. This rules out using the mosaicity as a suitable metric as it violates both of these criteria [39]. Unit cell volume expansion was once thought to be a suitable metric [96], however it was determined later that it was not reliable because the expansion was not consistent amongst crystals of the same protein and the same size [74]. Three promising metrics that seem to adhere to the criteria (within a factor of 2) are: the decay  $R$  factor [28], relative  $B$  factor [62] and the relative intensity [84].



(a) Low dose  $\sim kGy$  dataset.



(b) High dose (30  $MGy$ ) dataset.



(c) Extremely high dose (60  $MGy$ ) dataset.

Figure 1.11: Intensity decay of reflections on diffraction images of cubic insulin collected at 100  $K$  with increasing radiation exposure. The intensity of reflections corresponding to high resolution information (spots closer to the edges of the images) decay faster than low resolution reflections.



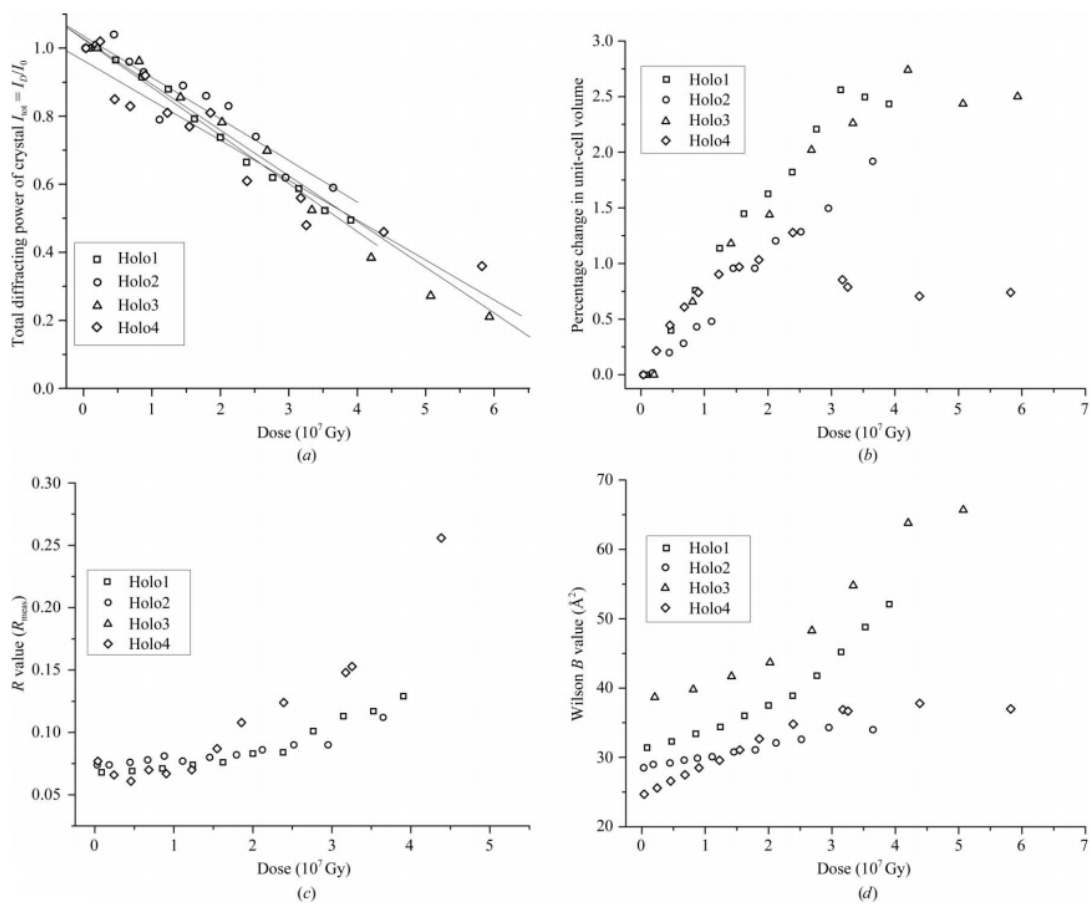


Figure 1.12: Various global radiation damage metrics as a function of dose for four holoferritin crystals. Work performed by [84], reproduced from [39].

---

## Decay R factor

The decay R factor,  $R_d$  is decay metric that assesses the pairwise difference between symmetry related reflections on different images [28]. This results in a relatively smooth function against image number (used as a proxy for dose) that can be interpreted as follows: if  $R_d$  stays constant for each image then no significant radiation damage has occurred, whereas if the  $R_d$  increases then radiation damage is in progressing throughout the dataset. A desirable feature of this metric is that it can be used with a single dataset. Therefore radiation damage analysis can be performed on highly sensitive crystals that may only withstand enough dose for a single dataset. The problem is that it requires high multiplicity which is not always achievable. This metric is less commonly used in radiation damage studies than the other two.

## Relative B factor

The relative B factor,  $B_{rel}$ , is a metric that calculates the difference between the isotropic B factor for the initial dataset and a later dataset collected on the same crystal [62]. Various studies have shown that  $B_{rel}$  increases linearly with dose [14, 15, 62, 66]. Another closely related metric which is derived from  $B_{rel}$  is the coefficient of sensitivity to absorbed dose,  $s_{AD} = \Delta B_{rel} / \Delta D 8\pi^2$  where  $\Delta B_{rel}$  and  $\Delta D$  is the change in relative B factor and dose respectively. The advantage of  $s_{AD}$  is that it is robust (to within a factor of 2) for protein crystals of differing molecular weights and solvent contents [62].

## Relative intensity

The relative intensity is defined as  $I_n/I_1$ , where  $I_n$  is the summed mean intensity of a complete data set  $n$  (or equivalent sections of data) after a dose  $D$ , and  $I_1$  is the mean intensity of the first data set [39]. This metric is desirable because it directly analyses the decay of the experimental data: the intensities. Furthermore it exhibits a linear dependence on the dose for a suitably low dose [84, 115]. An experimental dose limit of 30 *MGy* was established based on the observation that crystals of proteins with different compositions decay at the same rate at cryo-temperatures [84]. At this dose the relative intensity of the crystals had decayed to 70% of their initial value, at which point it is regarded that the structural information that can be determined from the sample has been significantly compromised [13, 84].

### 1.4.5 Manifestations of damage: Specific damage

In contrast to global radiation damage, specific radiation damage is classified as damage that is observed in real space in the electron density maps and occurs up to  $\approx 60$  times faster [52]. It is characterised by what appear to be specific chemical changes to the molecule and occurs in a reproducible order [44, 95, 111]:

1. Metallo-centres are reduced at doses as low as 45 *kGy* [83].
2. Elongation and breakage of disulphide bonds [21, 95].
3. Decarboxylation of glutamate and aspartate residues [21, 95, 111].
4. The  $S^\delta - C^\epsilon$  bond of methionine residues is cleaved [21].
5. Covalent metal bonds are cleaved [93].

---

Dehydroxylation of tyrosine residues have also been observed [21], but the physical basis and the evidence of this being a general symptom of specific radiation damage is contentious (Charlie Bury and Elspeth Garman, personal communication).

Due to the fact that specific damage changes structural information, it can lead to incorrect biological interpretation from structures. Metrics to assess the progression of specific damage also exist but are not always described as a function of dose.

### Peaks in Fourier difference maps

Fourier difference maps,  $F_n - F_1$ , are electron density maps where the amplitudes of the first dataset  $F_1$  are subtracted from the amplitudes of a later dataset  $F_n$ . Negative values correspond to areas where electron density has been lost as the experiment progressed. Sometimes this can be due to noise in the data, but significantly high values (difference peaks) that are coincident with the structural model from the first dataset suggest specific structural damage has occurred. In most studies, this sort of analysis is done via manual inspection of the electron density map [21,95,111]. However work is being done towards automating this analysis and developing new metrics to better characterise the level of specific damage [22].

### Atomic B-factors

Atomic B-factors (also referred to as atomic displacement parameters) describe the level of dynamic disorder of atoms, with each atom having its own B-factor. Generally this value increases as radiation damage progresses and the level of disorder of the structure increases. In practice the B-factor is also substantially influenced by other factors such as static disorder (occupancy), errors in model building and the packing density of the atom [44]. This means that atoms with higher B-factors may not necessarily be more damaged than atoms with a smaller B-factor.

### B damage

B damage,  $B_{damage}$ , is a metric that aims to deconvolute the dependence of the atomic B factor on its packing density. It is defined as the ratio of an atom's B-factor and the average B-factor of atoms that have a similar packing density environment [44]. This is a normalised metric, therefore atoms that show damage will have  $B_{damage}$  values significantly greater than 1.

Analysis of PDB structures with  $B_{damage}$  showed a positive correlation between radiation damage and solvent accessibility [44], in agreement with statements made in [104]. However other studies have claimed otherwise [25,54] demonstrating the lack of consensus of metrics and conclusions in radiation damage studies.

## 1.4.6 Experimental methods for dealing with radiation damage

Due to the challenge presented by radiation damage in MX to achieve successful structure determination, various parameters of the experiment have been varied in an attempt to mitigate its effects.

---

## Temperature

It has been known for decades that diffraction experiments carried out at cryo-temperatures (100 K) improves the lifetime of biological samples [20, 48] by around  $\sim 25 - 110$  times over room temperature (RT) experiments [101]. Cryo-cooling requires careful protection of the sample to prevent freezing, because ice also exhibits a crystalline structure that will too diffract coherently in the MX experiment. Cryo-protection techniques have been developed since the late 1980s [42, 55, 107] and cryo-crystallography is now more routine than RT experiments [40]. The main problem with cryo-cooling a crystal is that it can decrease the crystalline order of the crystal [79] although this is not always the case [38].

There is evidence to suggest that using temperatures below 100 K may further improve the lifetime of crystals and the data quality [70, 108]. However, whether the benefits of cooling to temperatures of 40 K or below are significant enough to warrant the costs required to achieve those temperatures routinely is still not obvious [110].

## Dose rate

Damage processes occur over a finite time period and this has been exploited by increasing the rate at which the dose is deposited in the crystal (dose rate) in an attempt to outrun radiation damage. At cryo-temperatures dose rate effects have been shown to improve crystal lifetimes by up to 10% [84] although some report that there is no effect at all [99]. At RT the evidence that there is a dose rate dependence seems quite convincing [81, 82, 101]. With the increasingly popularity of *in situ* data collection [5, 6], particularly for protein crystals that are not amenable to cryo-cooling, high dose rate experiments are likely to become even more routine.

Serial femtosecond crystallography (SFX) at XFELs use incredibly fast X-ray pulses (as little as 5 fs [16]) to probe protein crystals to obtain “diffraction before destruction” [23]. Although experimental data have now shown indications that radiation damage is still visible at XFELs if the pulse length is too long [77]. SFX at XFELs are still in their infancy but progress is continuously being made and is a promising field for the development of structural biology [20, 40].

## Scavengers

Scavengers are small molecules introduced to the crystal to mitigate the effects of secondary damage by “soaking” up radicals. At present, evidence of the efficacy of scavengers is contradictory. For example the studies presented in [8, 63] are conflicting in their evidence as to whether scavengers have any effect on global radiation damage metrics. It is worth noting that the studies used different global radiation damage metrics ( $I_n/I_1$  improved with use of scavengers, whereas  $B_{rel}$  did not) to come to their conclusions. The effect of scavengers to protect against specific radiation damage is more promising [100]. A comprehensive summary of all scavenger literature to date is provided in [4].

### 1.4.7 Modelling intensity decay

Structure factor amplitudes are derived from the integrated intensity measurements from the experimental data. Therefore accurate intensity values are required to obtain reliable structural information. The problem is that the intensity measurements are affected by several systematic factors [32]. In particular, global radiation damage of the sample can be a significant factor contributing to the overall intensity variation. For this reason attempts to correct for this effect have been proposed and

implemented during data processing.

The first systematic study of radiation damage was carried out on crystals of sperm whale myoglobin in 1962 by Blake and Phillips [11]. In this study they determined that radiation damage progression is proportional to the absorbed dose and predicted that specific damage was occurring without knowledge of the sequence of the molecule. Repeated intensity measurements of a subset of reflections along the  $h0l$  zone were taken with increasing X-ray exposure.

5. **TODO:** Check that the reflection zone is  $h0l$ , not  $010$  as written in [101]. Elspeth: do you have the Blake and Phillips paper to hand?

It was noted that there was a general decrease in the intensity values with reflections corresponding to higher resolution decaying quickest. If the variation was due solely to thermal disorder then the intensity decay was expected to obey the form

$$I = I_0 \exp \left( -\frac{B \sin^2(\theta)}{\lambda^2} \right), \quad (1.24)$$

where  $I$  is the experimentally measured intensity,  $I_0$  is the initial intensity,  $B$  is a measure of the disorder,  $\theta$  is the diffraction angle and  $\lambda$  is the wavelength of the incident beam. However this form did not explain the observation that the intensity flattened out at high  $\theta$  angles. So a compartmental model of the crystal was proposed. The three states that were used to describe the evolution of the crystal damage were

1.  $A_1$  - an undamaged fraction of the crystal which contributes to diffraction at all angles.
2.  $A_2$  - a highly disordered fraction that only contributes to diffraction at low angles.
3.  $A_3 = 1 - (A_1 + A_2)$  - an amorphous fraction that no longer diffracts coherently.

These compartments were suggested to contribute to the diffracted intensity,  $I$ , at time  $t$  as

$$\frac{I(t)}{I(0)} = A_1(t) + A_2(t) \exp \left( -\frac{B \sin^2(\theta)}{\lambda^2} \right). \quad (1.25)$$

To explain the crystal compartment dynamics (i.e. how the crystal transitions between damaged, highly disordered and amorphous states) it was assumed that radiation damage effects are irreversible. The transitions are described according to Figure 1.13 where  $k_1$ ,  $k_2$  and  $k_3$  are rate constants to be determined.

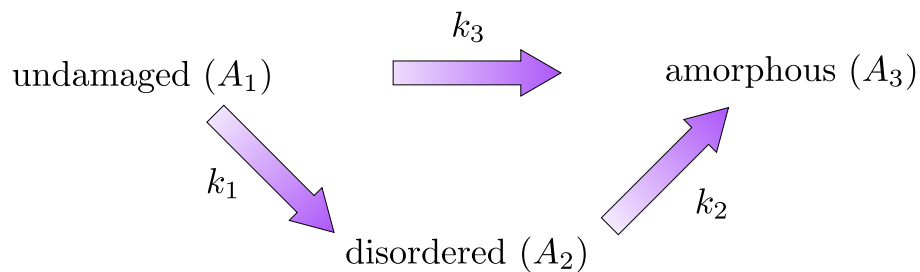


Figure 1.13: Transition dynamics of the compartments in the Blake and Phillips model. Note here that the states are irreversible and it is possible to transition directly from the undamaged fraction to the amorphous fraction.

---

The intensity measurements corrected with equation 1.25 generally agreed well with the measurements from the relatively undamaged crystal.

The Blake and Phillips model was also used to correct for radiation damage for a study on crystals of lamprey haemoglobin [51]. In this study, 150 reflections were observed to determine the change in intensity as a function of the dose. The data they recorded agreed well with the model of Blake and Phillips.

In 1976, Fletterick *et al.* proposed a modification to the Blake and Phillips model for their study on glycogen phosphorylase a crystals [35]. The modification disallowed a direct transition from the undamaged state to the amorphous state (i.e.  $k_3 = 0$  in Figure 1.13). This gives the sequential transition model

$$A_1 \xrightarrow{k_1} A_2 \xrightarrow{k_2} A_3. \quad (1.26)$$

In 1976, Hendrickson studied all possible transitions from the Blake and Phillips model (Figure 1.13) assuming irreversible transition states [49]. This included  $k_2 = k_3$  [51],  $k_1 = k_2$  and  $k_3 = 0$ ,  $k_3 = 0$  [35],  $k_1 + k_3 = k_2$  and the arbitrary time-dependence model. These models were fitted to the data from the Blake and Phillips myoglobin study [11]. It was concluded that all models explain the data well for moderate dose values ( $\approx 26\%$  damage), but none explain radiation damage at high doses very well ( $\approx 72\%$  damage). The general model resulted in refined  $k_3$  values that were very close to zero, and the model with  $k_3 = 0$  consistently resulted in the best fit to the data at higher dose values. This suggests that the radiation damage process is sequential.

In 1988 Sygusch and Allaire proposed an extension to the Fletterick *et al.* model. They suggested that a fourth fraction,  $A'_1$ , that has only suffered small perturbations of surface residues and disulphide bonds, still has the capability of contributing to diffraction that occurs at all angles [104]. This is because these surface perturbations are caused before any significant changes in molecular conformation. The resulting sequential model is expressed as

$$A_1 \xrightarrow{k_1} A'_1 \xrightarrow{k_1} A_2 \xrightarrow{k_2} A_3. \quad (1.27)$$

This model was fitted with data collected on crystals of rabbit skeletal muscle aldolase. The authors show that the proposed model provides a very good fit for radiation damage even at high doses.

In 2009 James Holton proposed an resolution dependent model of intensity decay for a reflection [52]. This model considers the absorbed dose within the crystal and describes the intensity decay of a reflection,  $I$ , as

$$I = I_0 \exp \left( -\ln(2) \frac{D}{Hd} \right) \quad (1.28)$$

where  $I_0$  is the intensity at zero dose,  $D$  is the absorbed dose in  $MGy$ ,  $d$  is the resolution of the reflection in  $\text{\AA}$  corresponding to the distance between successive Bragg planes, and  $H$  is the Howells *et al.* criterion which proposed a resolution dependent dose limit of  $10 \text{ } MGy$  per  $\text{\AA}$  [56]. This model was shown to produce results in agreement with those from [62, 84] when tested with their data.

In 2012 Leal *et al.* modified a traditional scaling model to account for radiation damage [66]. The scaling model defined as

$$J(D, \mathbf{h}) = J(h_d) \times K(D) \times \exp(-B(D)h_d^2/2) \quad (1.29)$$

where  $J(D, \mathbf{h})$  is the expected intensity of a given reflection with Miller indices  $\mathbf{h} = hkl$  after the crystal has absorbed a dose  $D$ , and  $J(h_d)$  is the expected reflection intensity at reciprocal distance  $h_d = |\mathbf{h}| = 1/d$  from the origin in the absence of any radiation damage.  $K(D)$  and  $B(D)$  are the scale and B-factors respectively and both are dependent on the absorbed dose. The authors assumed a linear model of B-factor increase given by

$$B(D) = B_0 + D\beta, \quad (1.30)$$

---

where  $B_0$  and  $\beta$  are empirical constants to be determined. The authors empirically deduced the functional form of the scale factor,  $K$  as

$$K(D) = C \exp(-\gamma^2 D^2), \quad (1.31)$$

where  $C$  and  $\gamma$  are empirical constants to be determined. The  $D^2$  in equation 1.31 means that the model proposed by Leal *et al.* is essentially Gaussian decay behaviour. This model was shown to agree well with the overall intensity decay for data collected on 15 different crystals at RT to a relative intensity down to 60% of the original value.

Despite the differences between these models, they each rely on an accurate calculation of the dose absorbed within a crystal. Of particular note, the models proposed by Sygusch and Allaire and Leal *et al.* allow for a delayed intensity decay. For the Leal *et al.* model, this behaviour follows from the Gaussian form of the equation. The Sygusch and Allaire model was shown to exhibit this behaviour by fitting the model to experimental data [82].

Current scaling methods employ a scaling equation in the same form as equation 1.29 to place multiple datasets on the same scale [59]. However the functional forms of the scale and B-factors are not necessarily assumed to take the forms given by equations 1.30 and 1.31. The scaling of intensities on images within a dataset also takes a form very similar to equation 1.29. Take, for example the scaling model used in AIMLESS:

$$g = g_1(\varphi) \times g_2(\mathbf{s}_2) \times \exp\left(\frac{-2B(\varphi) \sin^2(\theta)}{\lambda^2}\right), \quad (1.32)$$

where  $g$  is the factor that puts the intensity on an internally consistent scale with the other images in the dataset,  $g_1$  and  $g_2$  are scaling factors,  $\varphi$  is the crystal rotation angle which is used as a proxy for the primary beam direction and the dose and  $\mathbf{s}_2$  is the secondary beam direction [34]. The crystal rotation angle,  $\varphi$ , is used as a proxy for dose because the dose is not routinely calculated in MX experiments. It is noted by Evans and Murshudov that the B-factor correction in this model is largely an average correction.

Many studies, including the first by Blake and Phillips, found that the decay of individual reflection intensities is not necessarily monotonic [11, 49, 51]. As well as diffraction being anisotropic [2], some reflection reflections also increase in intensity with increasing dose [1].

Therefore the average radiation damage correction will not account for the specific changes that are known to occur. Diederichs *et al.* used a linear function to correct individual reflection intensities up to a dose of 10 *MGy* and showed that this reflection specific correction could improve results of SAD phasing [29]. Diederichs later used quadratic and exponential functions to correct individual reflection intensities [28]. It was concluded that the quadratic model suffers from the problem that it requires two parameters to be fitted for each reflection and it is not suitable when the unit cell expands because the Fourier transform of the molecule is no longer consistent between observations of the same reflection. The linear model suffers from the problem that the fitted parameter doesn't have a physical interpretation and the correction can result in negative extrapolated intensities. The exponential model has the advantage that it only results in positive values but again it lacks physical interpretation.

Despite some of the drawbacks, these specific correction models have shown some promise, but as of yet these types of corrections are not standard. This is because they are not necessarily straight forward to implement and they don't always seem to improve the data quality (Phil Evans, personal communication).

---

## 1.5 Radiation damage in SAXS

The radiation damage literature for SAXS experiments is far less extensive than that of MX and is less well understood. What is known is that the absorption events that occur in MX (primary damage) will occur in SAXS since the physics of the interactions will not change. SAXS experiments are generally carried out at RT which means that in addition to electron radicals, hydroxyl (OH) and hydroperoxyl (HO<sub>2</sub>) radicals created by the interaction of X-rays with water will also be mobile in the buffer solution [43, 57]. These radicals attach to the protein backbone and/or sidechains which in turn induce protein aggregation through covalent or non-covalent bonding [65]. The radiation damage is observed in the experimental data as a lack of overlap between 1D intensity curves for successive frames.

There are several methods to mitigate the radiation damage caused in SAXS experiments. These include:

- Flowing the sample through a capillary to limit the X-ray exposure for a given volume of sample. An alternative approach to this is to translate the capillary in the beam [57].
- In a similar manner to MX the beam can be attenuated or defocussed [57].
- Radioprotectants such as DTT, TCEP, glycerol, ethylene glycol, ascorbate or sucrose can be added. These are thought to be able to capture hydroxyl radicals [46].
- Carry out the SAXS experiment at cryo-temperatures which reduces the damage per unit dose by up to five orders of magnitude compared to RT experiments [71]. However performing cryo-SAXS is not routine because the experimental set-up is far from trivial [57]. It should be noted that dose estimates in SAXS experiments are calculated using equation 1.22, which is a one dimensional representation of the experiment.

Programs for primary data analysis do exist to assess quality of the data [88]. In particular, DATCMP has recently been improved to provide correlation maps (CorMaps) to assess the similarity of the 1D scattering curves produced in a SAXS experiment [36]. However, there is still room to improve the assessment of radiation damage progression online during the experiment.

## 1.6 This thesis in context

The work presented in this thesis extends previous work on radiation damage correction models. A new data reduction algorithm to process MX data has been developed that explicitly tracks the changes in the amplitude values throughout the diffraction experiment. The several benefits of this approach include:

- Explicit tracking of the amplitude uncertainties, even for negative and weak positive reflections.
- Estimates of amplitude values at every point in the diffraction experiment.
- The possibility to explicitly correct the amplitude values for radiation damage, as opposed to the intensities.

Furthermore, RADDOS-3D has been extended to calculate doses for SAXS experiments. This has been used to assess the efficacy of several radioprotectants for SAXS experiments. As a result of this work, a new metric has been developed for defining the threshold at which 1D SAXS curves are significantly damaged.



- 
- *Chapter 2* describes the experiment that was performed to collect data at 100  $K$  from several crystals of different proteins with a specially designed top-hat profile X-ray beam. The decay models presented in section 1.4.7 are then tested to assess their validity with these data. Finally the chosen model is used to improve the DWD metric proposed by Zeldin *et al.*
  - *Chapter 3* describes the regression model used to make corrections to intensities of individual reflections using RADDOSSE-3D and DWD. It also describes work assessing the correlations of structure factor amplitudes derived from a small polypeptide with structural modifications.
  - *Chapter 4* reports the main results of the thesis. In this chapter, the new data reduction algorithm is presented. The results are also presented and compared to those using the current data reduction pipeline (AIMLESS & CTRUNCATE). Considerations of the algorithm and extensions are also discussed.
  - *Chapter 5* describes the work on assessing errors associated with calculating doses in RADDOSSE-3D. This includes the handling of experimentally measured beam profiles with a significant background count and errors associated with uncertainties of crystal shape and other input parameters. Also the development of a graphical user interface for RADDOSSE-3D is also discussed and how it can be used to compare the results of several strategies to improve experimental design.
  - *Chapter 6* describes the extension of RADDOSSE-3D to calculate dose values for SAXS experiments. This extension is used to assess the dose suffered by samples in an experiment to determine the efficacy of radioprotectants in SAXS experiments. A new metric to define the threshold at which 1D SAXS curves are significantly damaged is also presented.
6. **TODO:** Find out from Elspeth if I am allowed Chapters 5 and 6 seem significant enough to be separate chapters. (I think they are)
- *Chapter 7* Summarises the work and conclusions drawn from the previous chapters and suggests future directions as a result of the work presented.

7. **TODO:** Add appendix entries

**Appendix entries:**

- mathematical group
- isomorphic groups

# Bibliography

- [1] S. C. Abrahams. International Union of Crystallography Commission on Crystallographic Apparatus single-crystal radiation damage survey. *Acta Crystallographica Section A*, 29(2):111–116, Mar 1973. pages
- [2] SC Abrahams and P Marsh. Anisotropy in the variation of serially-measured integrated intensities. *Acta Crystallographica Section A: Foundations of Crystallography*, 43(2):265–269, 1987. pages
- [3] Paul D Adams, Pavel V Afonine, Gábor Bunkóczi, Vincent B Chen, Ian W Davis, Nathaniel Echols, Jeffrey J Headd, L-W Hung, Gary J Kapral, Ralf W Grosse-Kunstleve, et al. Phenix: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 66(2):213–221, 2010. pages
- [4] Elizabeth G Allan, Melissa C Kander, Ian Carmichael, and Elspeth F Garman. To scavenge or not to scavenge, that is still the question. *Journal of synchrotron radiation*, 20(1):23–36, 2012. pages
- [5] Danny Axford, James Foadi, N-J Hu, HG Choudhury, S Iwata, K Beis, G Evans, and Y Alguel. Structure determination of an integral membrane protein at room temperature from crystals in situ. *Acta Crystallographica Section D: Biological Crystallography*, 71(6):0–0, 2015. pages
- [6] Danny Axford, Robin L Owen, Jun Aishima, James Foadi, Ann W Morgan, James I Robinson, Joanne E Nettleship, Raymond J Owens, Isabel Moraes, Elizabeth E Fry, et al. In situ macromolecular crystallography using microbeams. *Acta Crystallographica Section D: Biological Crystallography*, 68(5):592–600, 2012. pages
- [7] Xiao-chen Bai, Greg McMullan, and Sjors HW Scheres. How cryo-em is revolutionizing structural biology. *Trends in biochemical sciences*, 40(1):49–57, 2015. pages
- [8] Adam I Barker, Robert J Southworth-Davies, Karthik S Paithankar, Ian Carmichael, and Elspeth F Garman. Room-temperature scavengers for macromolecular crystallography: increased lifetimes and modified dose dependence of the intensity decay. *Journal of synchrotron radiation*, 16(2):205–216, 2009. pages
- [9] J.M. Berg, J.L. Tymoczko, and L. Stryer. *Biochemistry, Fifth Edition*. Biochemistry. W. H. Freeman, 2002. pages
- [10] JM Bijvoet. Structure of optically active compounds in the solid state. *Nature*, 173, 1954. pages
- [11] Colin C.F. Blake and David C. Phillips. Effects of x-irradiation on single crystals of myoglobin. *Proceedings of the Symposium on the Biological effects of Ionizing Radiation at the Molecular Level*, pages 183–191, 1962. pages
- [12] Clement E Blanchet and Dmitri I Svergun. Small-angle x-ray scattering on biological macromolecules and nanocomposites in solution. *Annual review of physical chemistry*, 64:37–54, 2013. pages

- 
- [13] T.L. Blundell and L.N. Johnson. *Protein Crystallography, by T[om] L Blundell and L[ouise] N Johnson*. Molecular biology. Academic Press, 1976. pages
- [14] Dominika Borek, Stephan L Ginell, Marcin Cymborowski, Wladek Minor, and Zbyszek Otwinowski. The many faces of radiation-induced changes. *Journal of synchrotron radiation*, 14(1):24–33, 2007. pages
- [15] Gleb P Bourenkov and Alexander N Popov. Optimization of data collection taking radiation damage into account. *Acta Crystallographica Section D: Biological Crystallography*, 66(4):409–419, 2010. pages
- [16] Sébastien Boutet, Lukas Lomb, Garth J Williams, Thomas RM Barends, Andrew Aquila, R Bruce Doak, Uwe Weierstall, Daniel P DePonte, Jan Steinbrener, Robert L Shoeman, et al. High-resolution protein structure determination by serial femtosecond crystallography. *Science*, 337(6092):362–364, 2012. pages
- [17] Matthew W Bowler, Didier Nurizzo, Ray Barrett, Antonia Beteva, Marjolaine Bodin, Hugo Caserotto, Solange Delagenière, Fabian Dobias, David Flot, Thierry Giraud, et al. Massif-1: a beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules. *Journal of synchrotron radiation*, 22(6), 2015. pages
- [18] Auguste Bravais. Memoire sur les systemes formes par des points distribues regulierement sur un plan ou dans l'espace. *J. Ecole Polytech*, 19:1–128, 1850. pages
- [19] Sandor Brockhauser, Marco Di Michiel, John E McGeehan, Andrew A McCarthy, and Raimond BG Ravelli. X-ray tomographic reconstruction of macromolecular samples. *Journal of Applied Crystallography*, 41(6):1057–1066, 2008. pages
- [20] Jonathan C Brooks-Bartlett and Elspeth F Garman. The nobel science: One hundred years of crystallography. *Interdisciplinary Science Reviews*, 40(3):244–264, 2015. pages
- [21] Wilhelm Pascal Burmeister. Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallographica Section D: Biological Crystallography*, 56(3):328–341, 2000. pages
- [22] Charles Bury, Elspeth F Garman, Helen Mary Ginn, Raimond BG Ravelli, Ian Carmichael, Geoff Kneale, and John E McGeehan. Radiation damage to nucleoprotein complexes in macromolecular crystallography. *Journal of synchrotron radiation*, 22(2):0–0, 2015. pages
- [23] Henry N Chapman, Carl Caleman, and Nicusor Timneanu. Diffraction before destruction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1647):20130313, 2014. pages
- [24] Henry N Chapman, Petra Fromme, Anton Barty, Thomas A White, Richard A Kirian, Andrew Aquila, Mark S Hunter, Joachim Schulz, Daniel P DePonte, Uwe Weierstall, et al. Femtosecond x-ray protein nanocrystallography. *Nature*, 470(7332):73–77, 2011. pages
- [25] Nicolas Coquelle, Emanuela Fioravanti, Martin Weik, Frédéric Vellieux, and Dominique Madern. Activity, stability and structural studies of lactate dehydrogenases adapted to extreme thermal environments. *Journal of molecular biology*, 374(2):547–562, 2007. pages
- [26] H.S.M. Coxeter. *Regular Polytopes*. Dover books on advanced mathematics. Dover Publications, 1973. pages
- [27] FHC Crick and Beatrice S Magdoff. The theory of the method of isomorphous replacement for protein crystals. i. *Acta Crystallographica*, 9(11):901–908, 1956. pages
- [28] Kay Diederichs. Some aspects of quantitative analysis and correction of radiation damage. *Acta Crystallographica Section D: Biological Crystallography*, 62(1):96–101, 2006. pages

- 
- [29] Kay Diederichs, Sean McSweeney, and Raimond BG Ravelli. Zero-dose extrapolation as part of macromolecular synchrotron data reduction. *Acta Crystallographica Section D: Biological Crystallography*, 59(5):903–909, 2003. pages
- [30] Jan Drenth. *Principles of Protein X-Ray Crystallography*. Springer Advanced Texts in Chemistry. Springer-Verlag GmbH, 1999. pages
- [31] Jan Drenth. Introduction to basic crystallography. *International Tables for Crystallography, volume F*, pages 45–63, 2012. pages
- [32] Philip Evans. Scaling and assessment of data quality. *Acta Crystallographica Section D: Biological Crystallography*, 62(1):72–82, 2005. pages
- [33] Philip R Evans. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D: Biological Crystallography*, 67(4):282–292, 2011. pages
- [34] Philip R Evans and Garib N Murshudov. How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography*, 69(7):0–0, 2013. pages
- [35] RJ Fletterick, J Sygusch, N Murray, NB Madsen, and LN Johnson. Low-resolution structure of the glycogen phosphorylase a monomer and comparison with phosphorylase b. *Journal of molecular biology*, 103(1):1–13, 1976. pages
- [36] Daniel Franke, Cy M Jeffries, and Dmitri I Svergun. Correlation map, a goodness-of-fit test for one-dimensional x-ray scattering spectra. *Nature methods*, 2015. pages
- [37] SIMON French and KEITH Wilson. On the treatment of negative intensity observations. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 34(4):517–525, 1978. pages
- [38] Elspeth Garman. Cool data: quantity and quality. *Acta Crystallographica Section D: Biological Crystallography*, 55(10):1641–1653, 1999. pages
- [39] Elspeth F Garman. Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallographica Section D: Biological Crystallography*, 66(4):339–351, 2010. pages
- [40] Elspeth F Garman. Developments in x-ray crystallographic structure determination of biological macromolecules. *Science*, 343(6175):1102–1108, 2014. pages
- [41] Elspeth F Garman and Colin Nave. Radiation damage in protein crystals examined under various conditions by different methods. *Journal of synchrotron radiation*, 16(2):129–132, 2009. pages
- [42] Elspeth F Garman and Thomas R Schneider. Macromolecular cryocrystallography. *Journal of Applied Crystallography*, 30(3):211–237, 1997. pages
- [43] Warren M Garrison. Reaction mechanisms in the radiolysis of peptides, polypeptides, and proteins. *Chemical Reviews*, 87(2):381–398, 1987. pages
- [44] Markus Gerstel, Charlotte M Deane, and Elspeth F Garman. Identifying and quantifying radiation damage at the atomic level. *Synchrotron Radiation*, 22(2), 2015. pages
- [45] Susanne Gräslund, Pär Nordlund, Johan Weigelt, James Bray, Opher Gileadi, Stefan Knapp, Udo Oppermann, Cheryl Arrowsmith, Raymond Hui, Jinrong Ming, et al. Protein production and purification. *Nature methods*, 5(2):135–146, 2008. pages
- [46] Alexander Grishaev. Sample preparation, data collection, and preliminary data analysis in biomolecular solution x-ray scattering. *Current Protocols in Protein Science*, pages 17–14, 2012. pages

- 
- [47] Hedi Hegyi and Mark Gerstein. The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. *Journal of molecular biology*, 288(1):147–164, 1999. pages
- [48] Richard Henderson. Cryo-protection of protein crystals against radiation damage in electron and x-ray diffraction. *Proceedings: Biological Sciences*, pages 6–8, 1990. pages
- [49] Wayne A. Hendrickson. Radiation damage in protein crystallography. *Journal of Molecular Biology*, 106(3):889–893, September 1976. pages
- [50] Wayne A Hendrickson. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, 254(5028):51–58, 1991. pages
- [51] Wayne A Hendrickson, Warner E Love, and Jerome Karle. Crystal structure analysis of sea lamprey hemoglobin at 2 Å resolution. *Journal of molecular biology*, 74(3):331–361, 1973. pages
- [52] James M Holton. A beginner's guide to radiation damage. *Journal of Synchrotron Radiation*, 16(2):133–142, 2009. pages
- [53] James M Holton and Kenneth A Frankel. The minimum crystal size needed for a complete diffraction data set. *Acta Crystallographica Section D: Biological Crystallography*, 66(4):393–408, 2010. pages
- [54] Christina Homer, Laura Cooper, and Ana Gonzalez. Energy dependence of site-specific radiation damage in protein crystals. *Journal of synchrotron radiation*, 18(3):338–345, 2011. pages
- [55] H Hope. Cryocrystallography of biological macromolecules: a generally applicable method. *Acta Crystallographica Section B: Structural Science*, 44(1):22–26, 1988. pages
- [56] Malcolm R Howells, Tobias Beetz, Henry N Chapman, C Cui, JM Holton, CJ Jacobsen, J Kirz, Enju Lima, Stefano Marchesini, Huijie Miao, et al. An assessment of the resolution limitation due to radiation-damage in x-ray diffraction microscopy. *Journal of Electron Spectroscopy and Related Phenomena*, 170(1):4–12, 2009. pages
- [57] Cy M Jeffries, Melissa A Graewert, Dmitri I Svergun, and Clément E Blanchet. Limiting radiation damage for high-brilliance biological solution scattering: practical experience at the embl p12 beamline petraiii. *Journal of synchrotron radiation*, 22(2):273–279, 2015. pages
- [58] George DD Jones, Jeremy S Lea, Martyn CR Symons, and Fatai A Taiwo. Structure and mobility of electron gain and loss centres in proteins. *Nature*, 330(6150):772–773, 1987. pages
- [59] Wolfgang Kabsch. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D: Biological Crystallography*, 66(2):133–144, 2010. pages
- [60] Wolfgang Kabsch. XDS. *Acta Crystallographica Section D: Biological Crystallography*, 66(2):125–132, 2010. pages
- [61] I Khan, R Gillilan, I Kriksunov, R Williams, WR Zipfel, and U Englich. Confocal microscopy on the beamline: novel three-dimensional imaging and sample positioning. *Journal of applied crystallography*, 45(5):936–943, 2012. pages
- [62] Jan Kmetko, Naji S Hussein, Matthew Naides, Yevgeniy Kalinin, and Robert E Thorne. Quantifying x-ray radiation damage in protein crystals at cryogenic temperatures. *Acta Crystallographica Section D: Biological Crystallography*, 62(9):1030–1038, 2006. pages
- [63] Jan Kmetko, Matthew Warkentin, Ulrich Englich, and Robert E Thorne. Can radiation damage to protein crystals be reduced using small-molecule compounds? *Acta Crystallographica Section D: Biological Crystallography*, 67(10):881–893, 2011. pages

- 
- [64] Tobias Krojer and Frank von Delft. Assessment of radiation damage behaviour in a large collection of empirically optimized datasets highlights the importance of unmeasured complicating effects. *Journal of synchrotron radiation*, 18(3):387–397, 2011. pages
- [65] Shigeo Kuwamoto, Shuji Akiyama, and Tetsuro Fujisawa. Radiation damage to a protein solution, detected by synchrotron x-ray small-angle scattering: dose-related considerations and suppression by cryoprotectants. *Journal of synchrotron radiation*, 11(6):462–468, 2004. pages
- [66] Ricardo Miguel Ferraz Leal, Gleb Bourenkov, Silvia Russi, and Alexander N Popov. A survey of global radiation damage to 15 different protein crystal types at room temperature: a new decay model. *Journal of Synchrotron Radiation*, 20(1):14–22, 2012. pages
- [67] Andrew GW Leslie and Harold R Powell. Processing diffraction data with MOSFLM. In *Evolving methods for macromolecular crystallography*, pages 41–51. Springer, 2007. pages
- [68] Joseph R Luft, Jennifer R Wolfley, Meriem I Said, Raymond M Nagel, Angela M Lauricella, Jennifer L Smith, Max H Thayer, Christina K Veatch, Edward H Snell, Michael G Malkowski, et al. Efficient optimization of crystallization conditions by manipulation of drop volume ratio and temperature. *Protein science*, 16(4):715–722, 2007. pages
- [69] Airlie J McCoy, Ralf W Grosse-Kunstleve, Paul D Adams, Martyn D Winn, Laurent C Storoni, and Randy J Read. Phaser crystallographic software. *Journal of applied crystallography*, 40(4):658–674, 2007. pages
- [70] Alke Meents, Sascha Gutmann, Armin Wagner, and Clemens Schulze-Bries. Origin and temperature dependence of radiation damage in biological samples at cryogenic temperatures. *Proceedings of the National Academy of Sciences*, 107(3):1094–1099, 2010. pages
- [71] Steve P Meisburger, Matthew Warkentin, Huimin Chen, Jesse B Hopkins, Richard E Gillilan, Lois Pollack, and Robert E Thorne. Breaking the radiation damage limit with cryo-saxs. *Biophysical journal*, 104(1):227–236, 2013. pages
- [72] Jacqueline LS Milne, Mario J Borgnia, Alberto Bartesaghi, Erin EH Tran, Lesley A Earl, David M Schauder, Jeffrey Lengyel, Jason Pierson, Ardan Patwardhan, and Sriram Subramaniam. Cryo-electron microscopy—a primer for the non-microscopist. *FEBS Journal*, 280(1):28–45, 2013. pages
- [73] Edward Mitchell, Peter Kuhn, and Elspeth Garman. Demystifying the synchrotron trip: a first time user’s guide. *Structure*, 7(5):R111–R122, 1999. pages
- [74] James Murray and Elspeth Garman. Investigation of possible free-radical scavengers and metrics for radiation damage in protein cryocrystallography. *Journal of Synchrotron Radiation*, 9(6):347–354, 2002. pages
- [75] James W Murray, Elspeth F Garman, and Raimond BG Ravelli. X-ray absorption by macromolecular crystals: the effects of wavelength and crystal composition on absorbed dose. *Journal of Applied Crystallography*, 37(4):513–522, 2004. pages
- [76] Garib N Murshudov, Pavol Skubák, Andrey A Lebedev, Navraj S Pannu, Roberto A Steiner, Robert A Nicholls, Martyn D Winn, Fei Long, and Alexei A Vagin. Refmac5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, 67(4):355–367, 2011. pages
- [77] Karol Nass, Lutz Foucar, Thomas RM Barends, Elisabeth Hartmann, Sabine Botha, Robert L Shoeman, R Bruce Doak, Roberto Alonso-Mori, Andrew Aquila, Sas Bajt, et al. Indications of radiation damage in ferredoxin microcrystals using high-intensity x-fel beams. *Synchrotron Radiation*, 22(2), 2015. pages

- 
- [78] C Nave. Radiation damage in protein crystallography. *Radiation Physics and Chemistry*, 45(3):483–490, 1995. pages
- [79] Colin Nave and Elspeth F Garman. Towards an understanding of radiation damage in cryocooled macromolecular crystals. *Journal of Synchrotron Radiation*, 12(3):257–260, 2005. pages
- [80] Peter O'Neill, David L Stevens, and E Garman. Physical and chemical considerations of damage induced in protein crystals by synchrotron radiation: a radiation chemical perspective. *Journal of synchrotron radiation*, 9(6):329–332, 2002. pages
- [81] Robin L Owen, Danny Axford, Joanne E Nettleship, Raymond J Owens, James I Robinson, Ann W Morgan, Andrew S Dore, Guillaume Lebon, Christopher G Tate, Elizabeth E Fry, et al. Outrunning free radicals in room-temperature macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 68(7):810–818, 2012. pages
- [82] Robin L Owen, Neil Paterson, Danny Axford, Jun Aishima, Clemens Schulze-Bries, Jingshan Ren, Elizabeth E Fry, David I Stuart, and Gwyndaf Evans. Exploiting fast detectors to enter a new dimension in room-temperature crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 70(5):0–0, 2014. pages
- [83] Robin L Owen, Briony A Yorke, James A Gowdy, and Arwen R Pearson. Revealing low-dose radiation damage using single-crystal spectroscopy. *Journal of synchrotron radiation*, 18(3):367–373, 2011. pages
- [84] Robin Leslie Owen, Enrique Rudiño-Piñera, and Elspeth F Garman. Experimental determination of the radiation dose limit for cryocooled protein crystals. *Proceedings of the National Academy of Sciences of the United States of America*, 103(13):4912–4917, 2006. pages
- [85] Karthik S Paithankar and Elspeth F Garman. Know your dose: Raddose. *Acta Crystallographica Section D: Biological Crystallography*, 66(4):381–388, 2010. pages
- [86] Karthik S Paithankar, Robin Leslie Owen, and Elspeth F Garman. Absorbed dose calculations for macromolecular crystals: improvements to raddose. *Journal of Synchrotron Radiation*, 16(2):152–162, 2009. pages
- [87] MF Perutz. Isomorphous replacement and phase determination in non-centrosymmetric space groups. *Acta Crystallographica*, 9(11):867–873, 1956. pages
- [88] Maxim V Petoukhov, Daniel Franke, Alexander V Shkumatov, Giancarlo Tria, Alexey G Kikhney, Michal Gajda, Christian Gorba, Haydyn DT Mertens, Petr V Konarev, and Dmitri I Svergun. New developments in the atsas program package for small-angle scattering data analysis. *Journal of applied crystallography*, 45(2):342–350, 2012. pages
- [89] R. Philips. How big is the "average" protein? [online]. <http://book.bionumbers.org/how-big-is-the-average-protein/>, 2015. Accessed: 30 Dec. 2015. pages
- [90] M Pitteri and G Zanzotto. On the definition and classification of bravais lattices. *Acta Crystallographica Section A: Foundations of Crystallography*, 52(6):830–838, 1996. pages
- [91] Lois Pollack. Saxs studies of ion-nucleic acid interactions. *Annual review of biophysics*, 40:225–242, 2011. pages
- [92] Meena Rai and Harish Padh. Expression systems for production of heterologous proteins. *CURRENT SCIENCE-BANGALORE*, 80(9):1121–1128, 2001. pages
- [93] Udupi A Ramagopal, Zbigniew Dauter, Radhakannan Thirumuruhan, Elena Fedorov, and Steven C Almo. Radiation-induced site-specific damage of mercury derivatives: phasing and implications. *Acta Crystallographica Section D: Biological Crystallography*, 61(9):1289–1298, 2005. pages

- 
- [94] Raimond BG Ravelli and Elspeth F Garman. Radiation damage in macromolecular cryocrystallography. *Current opinion in structural biology*, 16(5):624–629, 2006. pages
- [95] Raimond BG Ravelli and Sean M McSweeney. The “fingerprint” that x-rays can leave on structures. *Structure*, 8(3):315–328, 2000. pages
- [96] Raimond BG Ravelli, Pascal Theveneau, Sean McSweeney, and Martin Caffrey. Unit-cell volume change as a metric of radiation damage in crystals of macromolecules. *Journal of synchrotron radiation*, 9(6):355–360, 2002. pages
- [97] MG Rossmann, Andrew GW Leslie, SS Abdel-Meguid, and T Tsukihara. Processing and post-refinement of oscillation camera data. *Journal of Applied Crystallography*, 12(6):570–581, 1979. pages
- [98] MICHAEL G Rossmann. Processing oscillation diffraction data for very large unit cells with an automatic convolution technique and profile fitting. *Journal of Applied Crystallography*, 12(2):225–238, 1979. pages
- [99] Piotr Sliz, Stephen C Harrison, and Gerd Rosenbaum. How does radiation damage in protein crystals depend on x-ray dose? *Structure*, 11(1):13–19, 2003. pages
- [100] Robert J Southworth-Davies and Elspeth F Garman. Radioprotectant screening for cryocrystallography. *Journal of synchrotron radiation*, 14(1):73–83, 2007. pages
- [101] Robert J Southworth-Davies, Melissa A Medina, Ian Carmichael, and Elspeth F Garman. Observation of decreased radiation damage at higher dose rates in room temperature protein crystallography. *Structure*, 15(12):1531–1541, 2007. pages
- [102] C. Starr, C. Evers, and L. Starr. *Biology: Concepts and Applications without Physiology*. Cengage Learning, 2010. pages
- [103] Olof Svensson, Stéphanie Malbet-Monaco, Alexander Popov, Didier Nurizzo, and Matthew W Bowler. Fully automatic characterization and data collection from crystals of biological macromolecules. *Acta Crystallographica Section D: Biological Crystallography*, 71(8):1757–1767, 2015. pages
- [104] J Sygusch and M Allaire. Sequential radiation damage in protein crystallography. *Acta Crystallographica Section A: Foundations of Crystallography*, 44(4):443–448, 1988. pages
- [105] Garry Taylor. The phase problem. *Acta Crystallographica Section D: Biological Crystallography*, 59(11):1881–1890, 2003. pages
- [106] Garry L Taylor. Introduction to phasing. *Acta Crystallographica Section D: Biological Crystallography*, 66(4):325–338, 2010. pages
- [107] T-Y Teng. Mounting of crystals for macromolecular crystallography in a free-standing thin film. *Journal of Applied Crystallography*, 23(5):387–391, 1990. pages
- [108] T-Y Teng and Keith Moffat. Radiation damage of protein crystals at cryogenic temperatures between 40 k and 150 k. *Journal of synchrotron radiation*, 9(4):198–201, 2002. pages
- [109] M Weik, RBG Ravelli, I Silman, JL Sussman, P Gros, and J Kroon. Specific protein dynamics near the solvent glass transition assayed by radiation-induced structural changes. *Protein Science*, 10(10):1953–1961, 2001. pages
- [110] Martin Weik and J-P Colletier. Temperature-dependent macromolecular x-ray crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 66(4):437–446, 2010. pages



- 
- [111] Martin Weik, Raimond BG Ravelli, Gitay Kryger, Sean McSweeney, Maria L Raves, Michal Harel, Piet Gros, Israel Silman, Jan Kroon, and Joel L Sussman. Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proceedings of the National Academy of Sciences*, 97(2):623–628, 2000. pages
- [112] AJC Wilson. The probability distribution of x-ray intensities. *Acta Crystallographica*, 2(5):318–321, 1949. pages
- [113] Briony A Yorke, Godfrey S Beddard, Robin L Owen, and Arwen R Pearson. Time-resolved crystallography using the hadamard transform. *Nature methods*, 11(11):1131–1134, 2014. pages
- [114] Oliver B. Zeldin. *Methods Development for Structural Biology*. Dphil thesis, University of Oxford, UK, 2013. pages
- [115] Oliver B Zeldin, Sandor Brockhauser, John Bremridge, James M Holton, and Elspeth F Garman. Predicting the x-ray lifetime of protein crystals. *Proceedings of the National Academy of Sciences*, 110(51):20551–20556, 2013. pages
- [116] Oliver B Zeldin, Markus Gerstel, and Elspeth F Garman. Optimizing the spatial distribution of dose in x-ray macromolecular crystallography. *Journal of Synchrotron Radiation*, 20(1):49–57, 2012. pages
- [117] Oliver B. Zeldin, Markus Gerstel, and Elspeth F. Garman. *RADDOSE-3D*: time- and space-resolved modelling of dose in macromolecular crystallography. *Journal of Applied Crystallography*, 46(4):1225–1230, 2013. pages

## **Appendix A**

# **Appendix Title**

Just some random stuff.

## Appendix B

### ToDoS

1. Need to sort the title page
2. Need to sort out font for thesis.
3. Need to sort page style for thesis.
4. The style for the references need to be sorted.
5. Check that the reflection zone is  $h0l$ , not 010 as written in [101].
6. Find out from Elspeth if I am allowed Chapters 5 and 6 seem significant enough to be separate chapters. (I think they are)
7. Add appendix entries