



---

---

## CHAPTER 2

---

# Dose Decay Modelling

## 2.1 Introduction

The diffraction weighted dose (DWD) is a dose metric that exhibits a consistent and reproducible relationship with the relative decay in total diffraction intensity from a protein crystal down to a relative intensity of 40% (?) (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). This is because DWD spatially resolves the exposure of the crystal to the X-ray beam in addition to accounting for the absorbed dose. Mathematically this spatial resolution of the exposure is incorporated via the fluence weighting in the DWD calculation (equation ??). However, the DWD currently does not take into account the loss of diffraction power due to the current damage state of the crystal. The *relative diffraction efficiency* (RDE),  $\eta$ , is introduced as a function of the absorbed dose that describes the diffracting power of any given volume of the crystal. The RDE is defined as the ratio of the proportion of incident photons that are currently being diffracted to the proportion that initially diffracted. This is graphically depicted in Figure 2.1.

The RDE is incorporated into the DWD equation as an additional weighting term. Explicitly this is

$$DW D^i = \frac{\int_{t_{i-1}}^{t_i} \iiint_{\text{crystal}} D(\mathbf{x}, t) \times F(\mathbf{x}, t) \times \eta(D(\mathbf{x}, t)) \text{ d}\mathbf{x} \text{ d}t}{\int_{t_{i-1}}^{t_i} \iiint_{\text{crystal}} F(\mathbf{x}, t) \times \eta(D(\mathbf{x}, t)) \text{ d}\mathbf{x} \text{ d}t}. \quad (2.1.1)$$

The work presented in this chapter describes the experiment and analysis carried out to determine a suitable functional form for the RDE. The RDE is then incorporated into the DWD using equation 2.1.1 and compared with the simple DWD (equation ??) and applied to the data used in the previous DWD study (?).

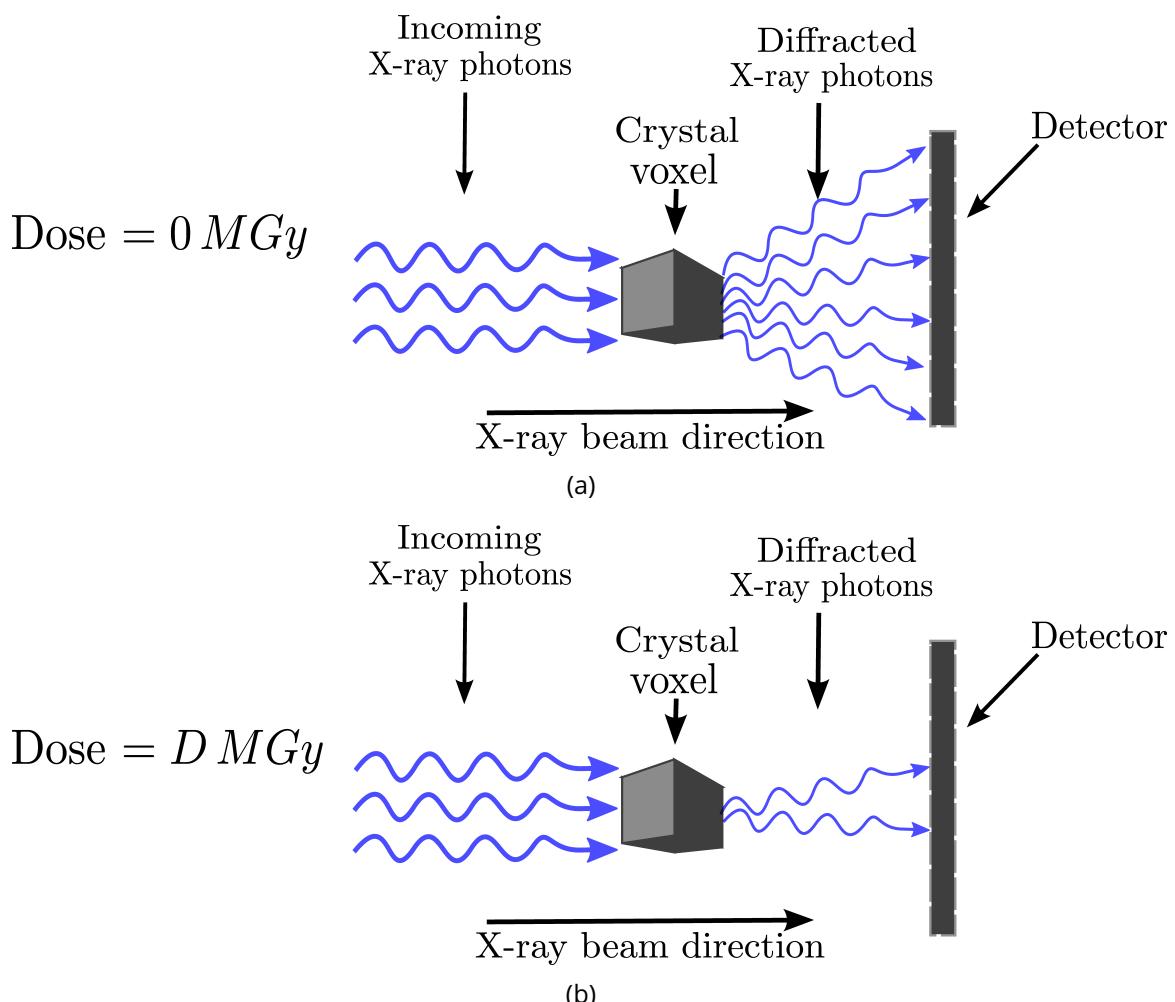


Figure 2.1: A visual example of the relative diffraction efficiency (RDE). (a) Initially at dose =  $0 MGy$  the crystal voxel elastically scatters 6 photons for a given number of incident photons. (b) Later in the experiment when the crystal voxel has absorbed a dose =  $D MGy$ , it elastically scatters only 2 photons for the same number of incident photons. The ratio of the current number of diffracted photons with the initial number is  $2/6 = 1/3$ , therefore in this example the RDE is  $1/3$ . The components in the figure are not drawn to scale.

## 2.2 Experimental methods

### 2.2.1 Considerations

To explore the behaviour of  $\eta$  as a function of absorbed dose it is necessary to understand the role of  $\eta$  in the calculation of the DWD (equation 2.1.1). The crystal is represented as a collection of voxels\* in RADDOS-3D.  $\eta$  is calculated for each voxel within the crystal and the absorbed dose within each voxel is assumed to be homogeneous. Therefore to experimentally determine the behaviour of  $\eta$  within a voxel as a function of absorbed dose, the crystals in the diffraction experiment must be irradiated uniformly to produce a homogeneous dose distribution. To achieve this it is necessary to use a flat (top-hat) beam profile with the entire crystal volume completely immersed within the X-ray beam throughout the rotation. At the time of the experiment (January 2014) RADDOS-3D was only able to model cuboid or spherical crystal shapes. Therefore the crystals used in the experiment were grown to be as close to cuboid in shape as possible.

### 2.2.2 Crystallization

Crystals of bovine pancreatic insulin purchased from Sigma-Aldrich (Lot # SLBJ0654V) were grown by the sitting-drop vapour diffusion method. The well solution consisted of 0.243  $M$  Na<sub>2</sub>HPO<sub>4</sub>, 0.007  $M$  Na<sub>3</sub>PO<sub>4</sub> at pH 10, and 0.01  $M$  Na<sub>3</sub>EDTA. 2  $\mu l$  of the well solution was added to an equal volume of the protein solution which consisted of 20 mg/ml insulin protein, 0.0195  $M$  Na<sub>2</sub>HPO<sub>4</sub>, 0.0005  $M$  Na<sub>3</sub>PO<sub>4</sub> at pH 10, and 0.01  $M$  Na<sub>3</sub>EDTA. The crystals were stored at room temperature ( $\approx 293\text{ K}$ ) and grew in a morphologically cuboid shape within 48 hours (Figure 2.2a). Cuboid shaped crystals less than 140  $\mu\text{m}$  in each dimension were selected and soaked for 30 - 60 seconds in a cryoprotectant solution with an identical composition to that of the well solution except with 30%  $v/v$  glycerol substituted for water, before being flash cooled into liquid nitrogen (77 K).

Human Haspin and myelocytomatosis (MYC) induced nuclear antigen (MINA) protein crystals that were cuboid in shape were kindly provided by the Structural Genomics Consortium (SGC) (Figures 2.2b and 2.2c). These crystals were selected for their cubic shape out of the

\* A voxel is the smallest distinguishable volume element in a three-dimensional representation of a computationally modelled object.

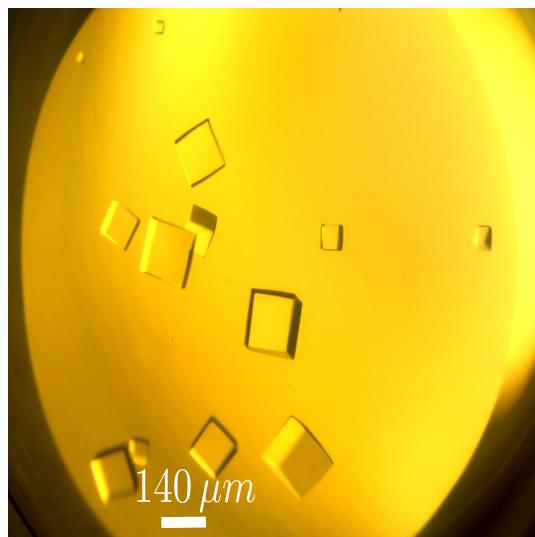
many human proteins crystallised at the SGC. They were cryoprotected in their native well solution with 25% *v/v* glycerol substituted for water.

### 2.2.3 Data collection and dose calculation

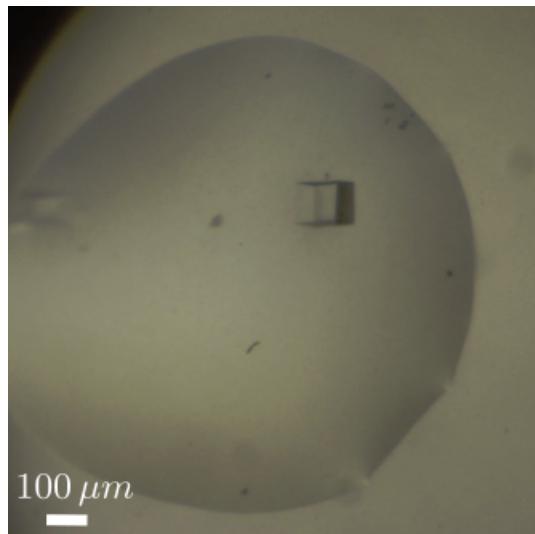
All data were collected at 100 K on the PETRA III Hamburg beamline P14 using an X-ray energy of 12.7 keV ( $\lambda = 0.9764 \text{ \AA}$ ), in collaboration with beamline scientist Dr. Gleb Bourenkov and EMBL beamline director Dr. Thomas Schneider. The experimentally measured beam profile was determined by placing a scintillator combined with an Allied Vision GC1350C CCD camera directly in the beam path. This produced a quantitative map of the beam profile in a portable graymap (pgm) file. The flat profile of the beam (coefficient of variation<sup>†</sup> of the beam is 2.09% vertically and 2.24% horizontally) was achieved by removing the focusing mirrors, and the slits were adjusted to achieve an aperture of  $140 \mu\text{m} \times 140 \mu\text{m}$  (Figure 2.3). The beam current was measured using a  $500 \mu\text{m}$  thick silicon PIN diode placed in the sample position from which the photon flux could be calculated (?). Before the data were collected from each crystal, an indexing set (100 frames of  $0.1^\circ$  rotation and  $0.1 \text{ s}$  exposure time per frame) was acquired. These frames were then indexed to provide the information necessary to reorient the crystal. One of the crystal faces was then aligned perpendicular to the beam direction such that the plane containing the beam direction vector was perpendicular to two of the edges of the aligned face (Figure 2.4). Alignment was performed using an Arinax MD3 mini kappa goniometer with the spindle axis mounted in a vertical and downward configuration. The crystal was centred on the beam position to make sure the entire crystal volume was completely immersed in the beam during the experiment. The dimensions of the crystals were measured on the screen prior to data collection. Table 2.1 contains the details of the data collection strategies for each crystal type.

Dose values were calculated using RADDSE-3D. The photon flux was determined to be  $1.9 \times 10^{11} \text{ ph/s}$ , the composition of the crystal was obtained from Dr. Oliver B. Zeldin's thesis (?) and from the constituents of the crystallisation solution as in Section 2.2.2. Although the crystal composition in Zeldin's thesis is incorrect (he specifies that there is a zinc atom for every insulin monomer in the unit cell whereas the true composition has only two zinc atoms per insulin hexamer), using the same composition allows direct comparison with the

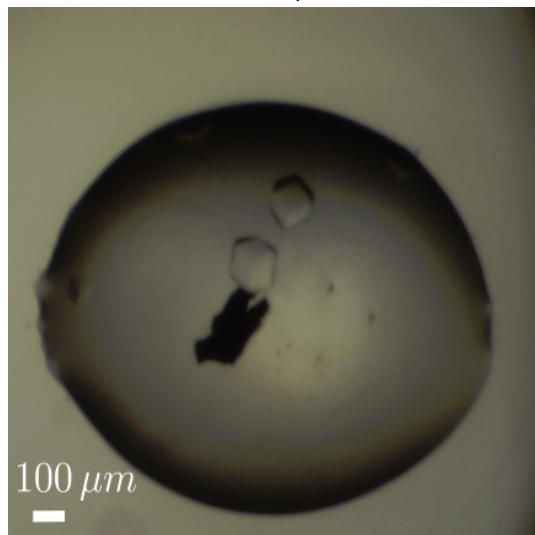
<sup>†</sup>The coefficient of variation is defined as the ratio of the standard deviation to the mean of a set of values which can also be expressed as a percentage by multiplying by 100%.



(a) Insulin



(b) Haspin



(c) MINA

Figure 2.2: Crystals used in the experiment.

results obtained in (?). Functionality to handle the experimentally measured beam profile was added to RADDOSE-3D to further improve the simulation of the absorbed dose.

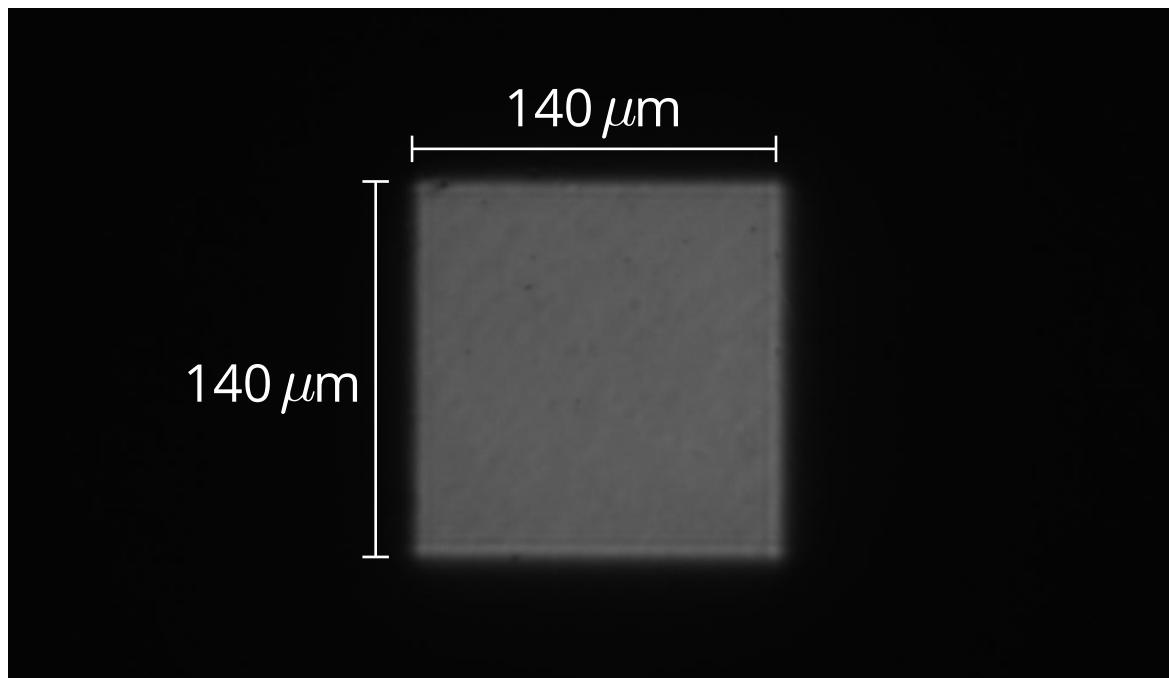
Table 2.1: Data collection strategy for each protein crystal type

Protein Crystal Type	Number of Crystals	Total Number of Frames per crystal	Rotation per image (°)	Total rotation (°)	Exposure Time per image (seconds)	Approx resolution
Insulin	8	14400	0.1	1440	0.5	1.38 Å
Haspin	5	7200	1	7200	1	2.9 Å
MINA	3	7200/3600	0.1	720/360	1	3.0 Å

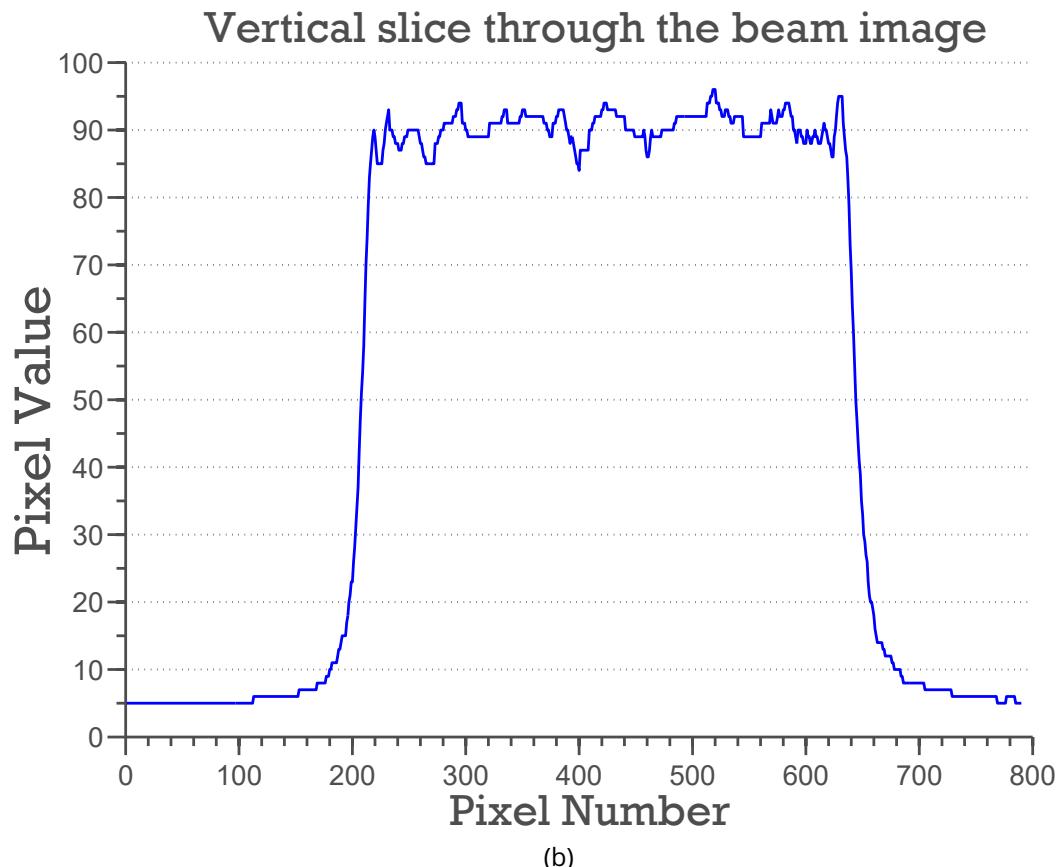
## 2.2.4 Data processing

The data collected for the SGC MINA and haspin crystals were non-trivial to analyse and hence the data processing procedures described here are relevant only for the insulin crystals. Data were processed using the Collaborative Computational Project No. 4 (CCP4) suite (?) with a standardised script used to call each program from within the suite to ensure identical treatment of all crystals. MOSFLM (?) was run manually with the space group set to  $I2_13$ . All images collected from a crystal were integrated together, fixing the unit cell angles and the detector distance but allowing the unit cell dimensions and mosaicity to be refined during integration. This was done because it is well documented that the unit cell expands and the mosaicity increases as radiation damage progresses (?).

The data were then scaled with AIMLESS (?) in batches of 900 frames (equivalent to 90° rotations) separated 50 frames apart (equivalent to 5°). This resulted in several overlapping datasets being extracted which allowed the progression of radiation damage to be tracked in much more detail than would be the case if datasets did not overlap. It is important to note however that these datasets will have sampled different regions of reciprocal space and so will therefore have different numbers of total observations, however, the differences are not expected to be significant. 14400 images were collected from each insulin crystal, so processing the data in this way produced 271 datasets per crystal. Despite the fact that the insulin crystals diffracted to 1.38 Å, the data were scaled to a resolution limit of 1.8 Å. This was to ensure that the processing for the highly damaged datasets were less likely to fail and to allow more direct comparison of the intensity values between datasets. Only five of the eight insulin datasets produced data that were processed straightforwardly in MOSFLM.



(a)



(b)

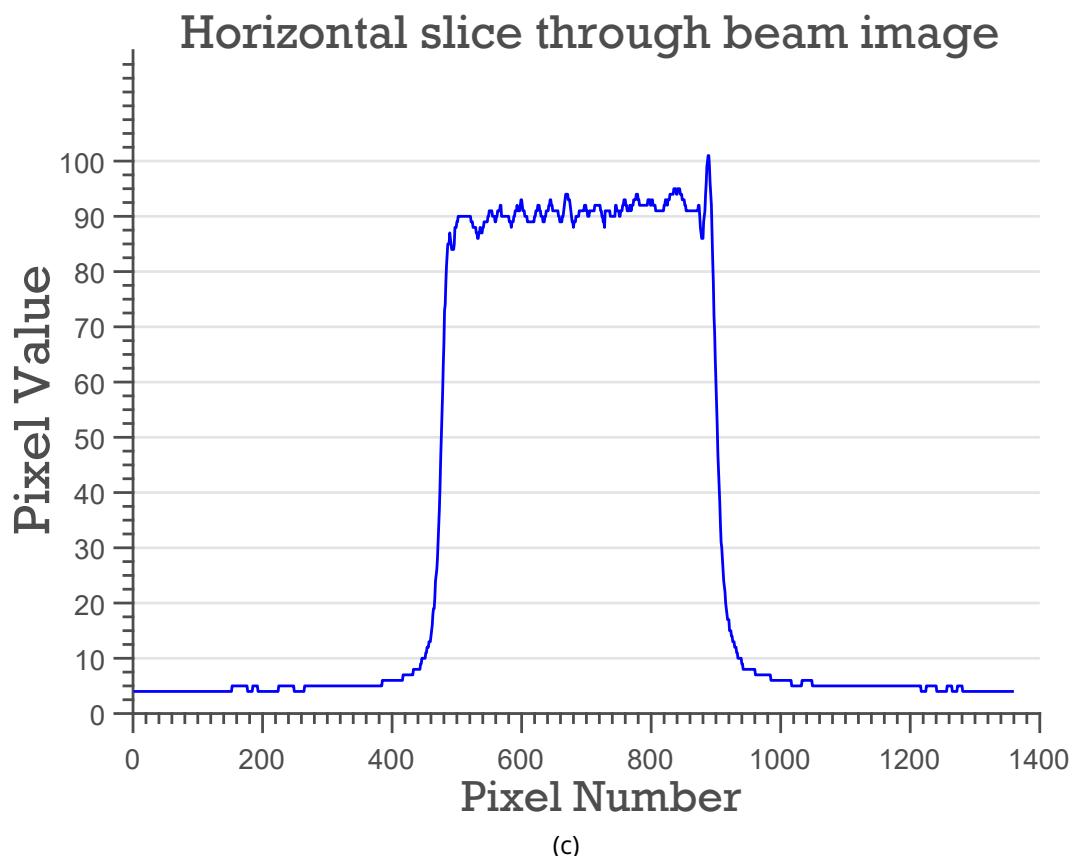


Figure 2.3: (a) Image (790 x 1360 pixels) of the experimentally determined beam profile. (b) Vertical slice through the beam image showing the flat profile of the beam. (c) Horizontal slice through the beam image showing the flat profile of the beam.

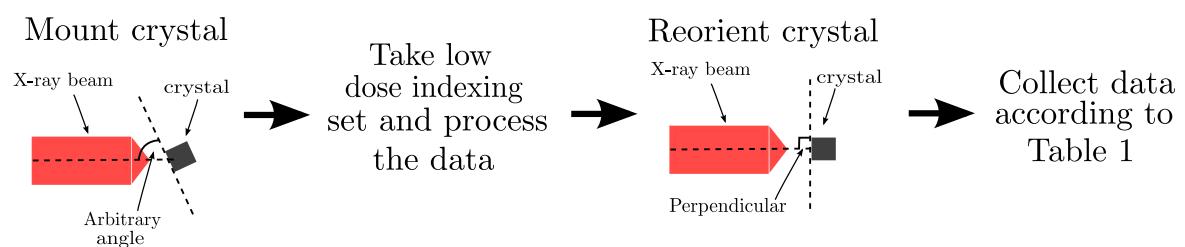


Figure 2.4: Flow diagram of the crystal reorientation process prior to data collection.

Data processing statistics for those five insulin crystals are shown in Table 2.2.

Table 2.2: Overall data processing statistics for the first data set collected from each of the processed insulin crystals (the completeness and multiplicity values are similar for the other datasets because the angular range is the same and the highest observable resolution data, out to 1.4 Å, has been discarded for this part of the analysis).

Values in parentheses are for the outer shell (1.83–1.79 Å). Unit cell and mosaicity are average values for all 14400 images

Crystal	0259	128	172	137	180
Space group	$I2_13$	$I2_13$	$I2_13$	$I2_13$	$I2_13$
Unit-cell parameters					
$a = b = c$ (Å)	78.28	78.28	78.35	78.36	78.40
$\alpha = \beta = \gamma$ (°)	90	90	90	90	90
Total No. of reflections	71955 (4528)	70860 (4208)	70757 (4423)	71968 (4554)	71580 (4463)
No. of unique reflections	7446 (474)	7436 (445)	7478 (471)	7478 (473)	7468 (471)
Completeness (%)	99.8 (100)	99.9 (100)	100 (100)	99.9 (100)	99.9 (100)
Multiplicity	9.7 (9.6)	9.5 (9.5)	9.5 (9.4)	9.6 (9.6)	9.6 (9.5)
$I/\sigma(I)$	33.3 (12.3)	37.9 (20.4)	32.3 (11.5)	43.6 (22.7)	33.9 (13.2)
$R_{\text{merge}}$	0.040 (0.128)	0.041 (0.097)	0.041 (0.158)	0.036 (0.086)	0.041 (0.128)
$CC_{1/2}$	0.999 (0.991)	0.998 (0.995)	0.994 (0.986)	0.988 (0.997)	0.999 (0.989)
Mosaicity (°)	0.42	0.30	0.29	0.34	0.28

## 2.2.5 Calculating the relative intensity

The relative intensity of a uniformly irradiated crystal is effectively the same as the RDE. Recall that 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. Implicit in this definition is that the total intensity is integrated for the same reflections for each dataset. Note: the relative intensity in this thesis has been calculated incorrectly and is described below.

Figure 2.5 shows a table from the log file of an AIMLESS job. The overall mean intensity given in the table (highlighted red in Figure 2.5) is calculated from intensities of the reflections within that particular dataset. The relative intensity can be calculated by dividing this value by the corresponding overall mean intensity from the first dataset, referred to in Figure 2.6 as *Aimless average intensity*.

Another way to calculate the relative intensity is to calculate the total summed intensity of a dataset, then divide that by the total summed intensity of the first dataset, referred to

in Figure 2.6 as *total intensity*. This was achieved by multiplying the number of measured reflections in each resolution bin ( $N_{meas}$  - highlighted orange in Figure 2.5) by the average intensity in that resolution bin ( $A_{vl}$  - highlighted blue in Figure 2.5), then summing them. The latter method of calculating the relative intensity was preferred over the previous method because it is more representative of the overall intensity loss, as opposed to the mean intensity loss. Therefore the relative intensity was always calculated using the total intensity method. Figure 2.6 shows the results of calculating the relative intensity using both methods and shows that the Aimless average intensity method decays more slowly than the total intensity method as the relative intensity drops below 0.6.

It is important to note that both of the methods described above for calculating the relative intensity are incorrect for two main reasons:

1. each dataset was scaled separately in Aimless. However Aimless assigns a different scale factor for each run and therefore the scaled values are not necessarily comparable between datasets.
2. the summed intensity values are derived from all reflections observed in a particular dataset. However these datasets sample different sections of reciprocal space and hence the same reflections are not necessarily observed in all datasets.

To calculate the relative intensity correctly, all datasets should be scaled together and only the reflections that are observed in all datasets should be used for the total summed intensity calculation. This method of calculating the relative intensity has not been used for the data in this thesis because this was discovered during the viva and it was determined that it is likely that the conclusions of the work will not change using the new method.

N	1/d^2	Dmid	Rmrg	Rfull	Rcum	Rmeas	Rpim	Nmeas	AVI	RMSdev	sd	I/RMS	Mn(I/sd)	FrcBias
1	0.0064	12.53	0.027	-	0.027	0.029	0.010	600	12681	811	498	15.6	60.5	-0.002
2	0.0191	7.23	0.030	-	0.028	0.032	0.010	1172	4961	247	226	20.1	53.4	-0.008
3	0.0319	5.60	0.029	-	0.029	0.031	0.011	1323	7403	367	316	20.2	54.1	-0.002
4	0.0446	4.74	0.030	-	0.029	0.032	0.010	1591	10684	540	434	19.8	61.3	0.008
5	0.0573	4.18	0.031	-	0.030	0.033	0.010	2017	12067	613	485	19.7	66.3	0.012
6	0.0701	3.78	0.033	-	0.031	0.035	0.011	2182	10410	599	429	17.4	61.8	0.003
7	0.0828	3.47	0.034	-	0.031	0.036	0.011	2478	7527	441	331	17.1	58.2	0.008
8	0.0956	3.23	0.033	-	0.031	0.035	0.011	2439	5525	293	262	18.8	53.8	0.002
9	0.1083	3.04	0.037	-	0.032	0.039	0.012	2739	3962	231	207	17.2	48.1	-0.005
10	0.1210	2.87	0.035	-	0.032	0.036	0.012	2901	3047	163	174	18.7	44.3	0.000
11	0.1338	2.73	0.035	-	0.032	0.037	0.012	3099	2753	158	165	17.5	42.2	-0.015
12	0.1465	2.61	0.036	-	0.032	0.038	0.012	3190	2201	125	145	17.5	38.5	-0.020
13	0.1593	2.51	0.039	-	0.033	0.041	0.013	3270	2005	123	139	16.3	36.7	-0.021
14	0.1720	2.41	0.041	-	0.033	0.044	0.014	3410	1717	113	130	15.2	34.2	-0.023
15	0.1847	2.33	0.048	-	0.033	0.051	0.016	3685	1482	114	124	13.0	30.9	-0.031
16	0.1975	2.25	0.051	-	0.034	0.054	0.017	3558	1370	110	121	12.5	29.2	-0.041
17	0.2102	2.18	0.054	-	0.035	0.057	0.018	3967	1249	104	118	12.0	27.3	-0.031
18	0.2230	2.12	0.064	-	0.035	0.067	0.021	3667	1115	113	114	9.9	25.4	-0.047
19	0.2357	2.06	0.075	-	0.036	0.079	0.025	3906	859	102	105	8.4	21.2	-0.049
20	0.2485	2.01	0.086	-	0.037	0.092	0.030	3722	780	108	101	7.2	19.9	-0.042
21	0.2612	1.96	0.092	-	0.037	0.097	0.031	4115	638	94	95	6.8	17.6	-0.055
22	0.2739	1.91	0.099	-	0.038	0.105	0.034	4070	571	91	93	6.3	16.6	-0.067
23	0.2867	1.87	0.113	-	0.039	0.119	0.038	4315	463	85	88	5.5	14.3	-0.088
24	0.2994	1.83	0.128	-	0.040	0.135	0.043	4524	375	78	85	4.8	12.3	-0.100
<b>Overall:</b>														
N	1/d^2	Dmid	Rmrg	Rfull	Rcum	Rmeas	Rpim	Nmeas	AVI	RMSdev	sd	I/RMS	Mn(I/sd)	FrcBias

Figure 2.5: A table given in the log file of an AIMLESS job from an insulin crystal. The table gives data for 24 resolution bins including the number of measured reflections in each resolution bin, Nmeas (highlighted orange), and the Average Intensity in each resolution bin, AvI (highlighted blue). The value highlighted in red is the overall average intensity for a dataset calculated using reflections that are measurable in that data set.

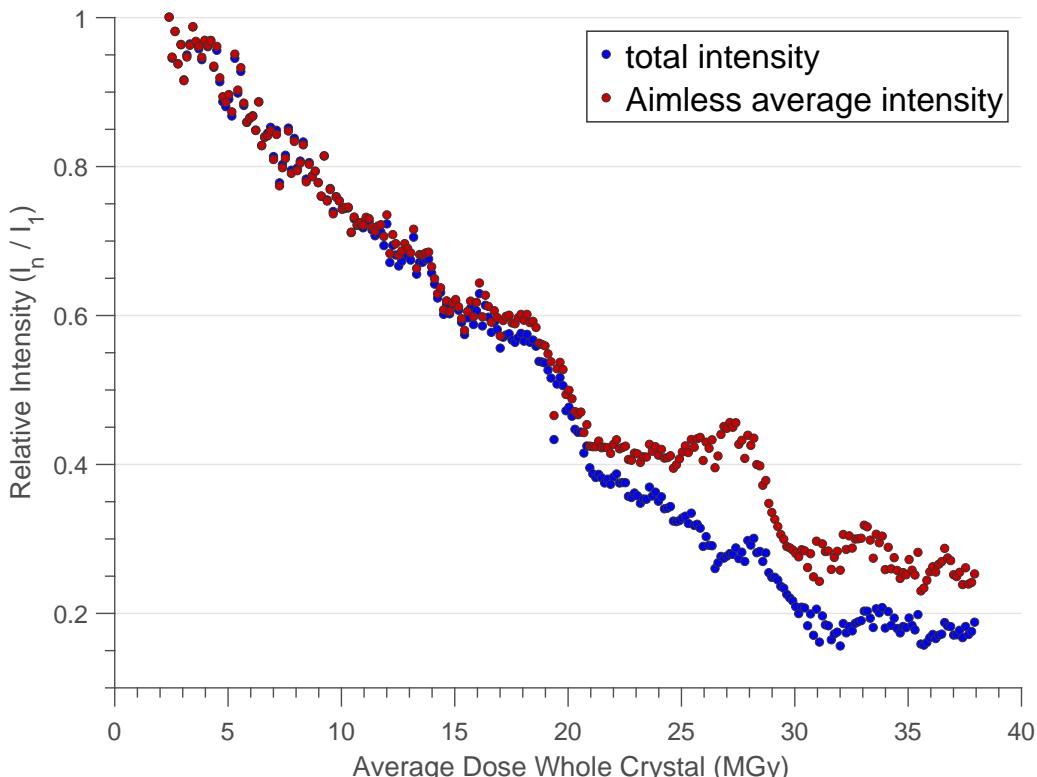


Figure 2.6: Relative intensity plotted against the average absorbed dose for one insulin crystal using the two methods described in section 2.2.5.

## 2.3 Dose decay models

With a calculated dose and a relative intensity value for each dataset, a relationship between radiation damage progression and absorbed dose can be determined via dose decay models (DDMs). A DDM is a function that describes the change of reflection intensity as a function of the absorbed dose. Several DDMs have been proposed over the last few decades as enumerated in section ???. In particular, three different DDMs have been successful at describing radiation damage progression. These models were analysed and compared, to determine which one best described the data collected in this experiment.

As mentioned in chapter ??, the first of the three models was proposed by Sygusch & Al-laire (1988) (?), which was developed from the original model proposed by Blake & Phillips (1962) (?) and subsequently altered by Hendrickson, Fletterick and others. Sygusch & Al-laire's model assumes that the intensity of a reflection from a protein crystal is a linear combination of scattering contributions from 4 states; an undamaged fraction,  $A_1$ , that contributes to diffraction at all angles, a fraction that has undergone 'surface modification' but still conformationally resembles the undamaged protein,  $A'_1$ , and hence also contributes to diffraction at all angles, a disordered fraction,  $A_2$ , that mainly contributes to diffraction at low angles, and finally an amorphous fraction,  $A_3$ , that is no longer capable of coherent scattering. The model also assumes that radiation damage is a sequential and irreversible process, so the crystal transitions between states as follows:



Mathematically the intensity of an individual reflection,  $I$ , as a function of absorbed dose,  $D$  is given by:

$$I(D)/I(0) = A_1(D) + A'_1(D) + A_2(D) \exp\left[-\frac{B}{2}h^2\right], \quad (2.3.2)$$

assuming isotropic atomic vibrations, where  $B$  is the thermal parameter related to the mean square displacement of atomic vibration (?), and  $h = 1/d$ , where  $d$  is the distance between successive Bragg planes.

The crystal fractions are assumed to evolve according to the following system of coupled

first order linear ordinary differential equations (ODEs)

$$\frac{dA_1}{dt} = -k_0 I_0, \quad (2.3.3a)$$

$$\frac{dA'_1}{dt} = k_0 I_0 - k_1 A'_1, \quad (2.3.3b)$$

$$\frac{dA_2}{dt} = k_1 A'_1 - k_2 A_2, \quad (2.3.3c)$$

$$\frac{dA_3}{dt} = k_2 A_2, \quad (2.3.3d)$$

subject to the following constraints:

$$A_1(D) + A'_1(D) + A_2(D) + A_3(D) = A_0, \quad (2.3.4a)$$

$$A'_1(0) = A_2(0) = A_3(0) = 0, \quad (2.3.4b)$$

$$A_1(0) = A_0, \quad (2.3.4c)$$

where  $I_0$  represents the intensity of the incident irradiation and  $A_0$  is the quantity of protein in the irradiated sample.

The solution of the set of ODEs (2.3.3) subject to the constraints (2.3.4) is presented here:

$$A_1 = A_0 - k_0 I_0 D, \quad (2.3.5a)$$

$$A'_1 = \frac{k_0 I_0}{k_1} \left( 1 - e^{-k_1 D} \right), \quad (2.3.5b)$$

$$A_2 = \frac{k_0 I_0}{k_2} \left( 1 - e^{-k_2 D} \right) + \frac{k_0 I_0}{k_2 - k_1} \left( e^{-k_2 D} - e^{-k_1 D} \right), \quad (2.3.5c)$$

$$A_3 = k_0 I_0 D + \left( \frac{k_0 I_0}{k_2 - k_1} - \frac{k_0 I_0}{k_2} \right) \left( 1 - e^{-k_2 D} \right) - \frac{k_2 k_0 I_0}{k_1 (k_2 - k_1)} \left( 1 - e^{-k_1 D} \right). \quad (2.3.5d)$$

Equations (2.3.5a), (2.3.5b) and (2.3.5c) are explicitly given in (?), whereas equation (2.3.5d) is my own work for this thesis. Since the transition  $A_1 \rightarrow A'_1$  is zero-order, eventually all of the undamaged crystal fraction,  $A_1$ , will be converted into  $A'_1$ . Above this dose,  $D = D_L = A_0/k'_0$ , where  $k'_0 = k_0 I_0$  (set  $A_1 = 0$  and rearrange equation 2.3.5a), the solutions given by equation 2.3.5 will no longer be valid since  $A_1$  will become negative. To obtain a solution for  $D > D_L$  I observed that  $k_0 \equiv 0$  and hence solved the set of ODEs (2.3.3) subject to the

constraints (2.3.4) with the additional constraint on  $k_0$  yields:

$$A_1 = 0, \quad (2.3.6a)$$

$$A'_1 = A'_{10} e^{-k_1(D-D_L)}, \quad (2.3.6b)$$

$$A_2 = \frac{k_1 A'_{10}}{k_2 - k_1} e^{-k_1(D-D_L)} + \left( A_{20} - \frac{k_1 A'_{10}}{k_2 - k_1} \right) e^{-k_2(D-D_L)}, \quad (2.3.6c)$$

$$A_3 = A_{30} + \left( A_{20} - \frac{k_1 A'_{10}}{k_2 - k_1} \right) \left( 1 - e^{-k_2(D-D_L)} \right) + \frac{k_2 A'_{10}}{k_2 - k_1} \left( 1 - e^{-k_1(D-D_L)} \right), \quad (2.3.6d)$$

where  $A'_{10}$ ,  $A_{20}$  and  $A_{30}$  are the values of the crystal fractions  $A'_1$ ,  $A_2$  and  $A_3$  at dose  $D = D_L$  respectively. Although a solution for the model is undefined for  $k_1 = k_2$ , Table 2.5 shows that the best fit values for the two parameters are different enough to neglect this case.

The strength of this model lies in the fact it explicitly predicts the proportions of damaged states of the crystal as a function of the absorbed dose via the system of ODEs, and uses that solution to determine the intensity of a reflection. The drawbacks of this model are that it requires 3 parameters to be determined and also that its analytical form is not easy to intuitively interpret.

The second model was proposed by Dr. James Holton and assumes that the average intensity of a reflection  $I(D)$  decays exponentially (?). Mathematically this is expressed as:

$$I(D) = I_{ND} \exp \left[ -\ln(2) \frac{Dh}{H} \right], \quad (2.3.7)$$

where  $I_{ND}$  is the average reflection intensity expected in the absence of radiation damage,  $\ln(2)$  is the natural logarithm of 2 ( $\approx 0.693$ ),  $D$  is the dose,  $h = 1/d$ , where  $d$  is the distance between successive Bragg planes and  $H$  is Howell's criterion ( $10 \text{ MGy } \text{\AA}^{-1}$ ). The advantages of this model are that it is very simple and only has a maximum of two parameters to be determined,  $I_{ND}$  and  $H$ . To simplify the parameter value extraction, the value of  $H$  can be assumed to be  $10 \text{ MGy } \text{\AA}^{-1}$  and  $I_{ND}$  can be approximated to the intensity values from the first data set. Therefore this model is relatively easy to apply and does not require any further simplifying assumptions.

However the simplicity of this model is also the cause of its main disadvantage: it does not predict a lag phase for intensity decay. At cryotemperatures this prediction may be correct but recent results suggest otherwise for room temperature MX (?). Therefore this model already has apparent limitations in the applicability of its predictive power.

The third DDM was proposed by Leal *et al.* (?) and uses a similar radiation damage model to those used in many scaling programs (??). It is:

$$J(D) = J(h) \times scale(D) \times \exp \left[ -\frac{B(D)h}{2} \right], \quad (2.3.8)$$

with

$$scale(D) = K \exp \left[ -\gamma^2 D^2 \right] \quad (2.3.9)$$

$$B(D) = B_0 + \beta D \quad (2.3.10)$$

where  $J(D)$  is the expected intensity after the crystal has absorbed a dose  $D$ ,  $J(h)$  is the expected reflection intensity at reciprocal distance  $h$  from the origin in the absence of any radiation damage and  $B_0$ ,  $\beta$ ,  $K$  and  $\gamma$  are parameters to be determined by fitting the model to the data, and are completely empirical.

This model essentially describes a Gaussian decay of the intensity with dose. This gives it the potential to predict lag phases, which will be dependent on the parameter values obtained from the data. It has already been shown to successfully predict relative intensity decay at room temperature (?) at dose rates below those used in (?) ( $0.05 - 300 \text{ kGy s}^{-1}$ ). The main disadvantage of this model is that the scale function  $K \exp [-\gamma^2 D^2]$  is completely empirical and has no obvious physical interpretation.

### 2.3.1 Validity test

Since the parameters for the models are fitted, judging the models based on how well they fit the data would not necessarily validate any particular model. A more robust test would be to check whether the data transforms in the way that the model predicts they should. In particular, models given by equations 2.3.7 and 2.3.8 can be transformed into linear forms

$$\ln(I) = \ln(I_{ND}) - \frac{h \ln(2)}{H} \times D, \quad (2.3.11)$$

$$\ln \left( \frac{J(D)}{J(h)} \right) = \ln(scale(D)) - \frac{B(D)}{2} \times h^2. \quad (2.3.12)$$

$$(2.3.13)$$

These equations are of the linear form  $y = mx + c$ . Note that the Sygusch & Allaire model does not transform easily into a linear form. If the data follow the relationships described by the models, then transforming the data according to equations 2.3.11 and 2.3.12 means that plots of  $\ln(I)$  against  $D$  (equation 2.3.11) and  $\ln\left(\frac{J(D)}{J(h)}\right)$  against  $h^2$  (equation 2.3.12) should resemble straight lines. A Pearson Correlation Coefficient (PCC) value<sup>‡</sup> can then be used to determine the strength of the linear relationship. PCC values were found for all data in each resolution shell where the relative intensity values ( $I_n/I_1$ ) were above the experimental dose limit, 0.7 (beyond this point the data are likely to be biologically compromised (?)). Table 2.3 shows the results of this test. The Leal *et al.* model seems to explain the data better. However the correlation coefficients determined for Holton's model are generally strong enough not to reject the possibility that this model may still sufficiently describe the data.

Table 2.3: Pearson Correlation Coefficient (PCC) values for linearly transformed intensity data. The values are negative, showing the negative correlation between intensity and dose i.e. as the dose increases, the intensity decreases.

	Holton Model	Leal <i>et al.</i> Model
Mean PCC	-0.9240	-0.9915
Max PCC	-0.2599 <sup>§</sup>	-0.9896
Min PCC	-0.9819	-0.9935

### 2.3.2 Obtaining model parameter values

The transformations given by equations (2.3.11) and (2.3.12) produce suitably linear plots (Table 2.3), which allow the parameters for the Holton and Leal *et al.* models to be determined by calculating the gradient and intercepts for each plot respectively. Data collected before the experimental dose limit was reached were used to obtain fitted parameters for each of the models.  $I_{ND}$  was obtained by exponentiating (with Euler's number  $e = 2.718\dots$ ) the intercept of the plot of  $\ln(I)$  against dose, where  $I$  is the average reflection intensity in a given resolution bin. Note that  $I_{ND}$  will be different for each resolution bin (Table 2.4).  $H$  was obtained by calculating the gradient from the straight line plot for each resolution bin and rearranging according to equation 2.3.11. However  $H$  should be a constant value across all resolution bins, so the 'best' estimate of  $H$  was taken from the resolution bin that gave the best straight line fit.

<sup>‡</sup>The Pearson Correlation Coefficient is a measure of linear correlation between two variables and gives values between -1 and 1 inclusive. A value of 1 represents a positive correlation, -1 represents a negative correlation and 0 represents no correlation.

<sup>§</sup>This value is an outlier from the lowest resolution shell. The PCC value for the next shell is -0.7578.

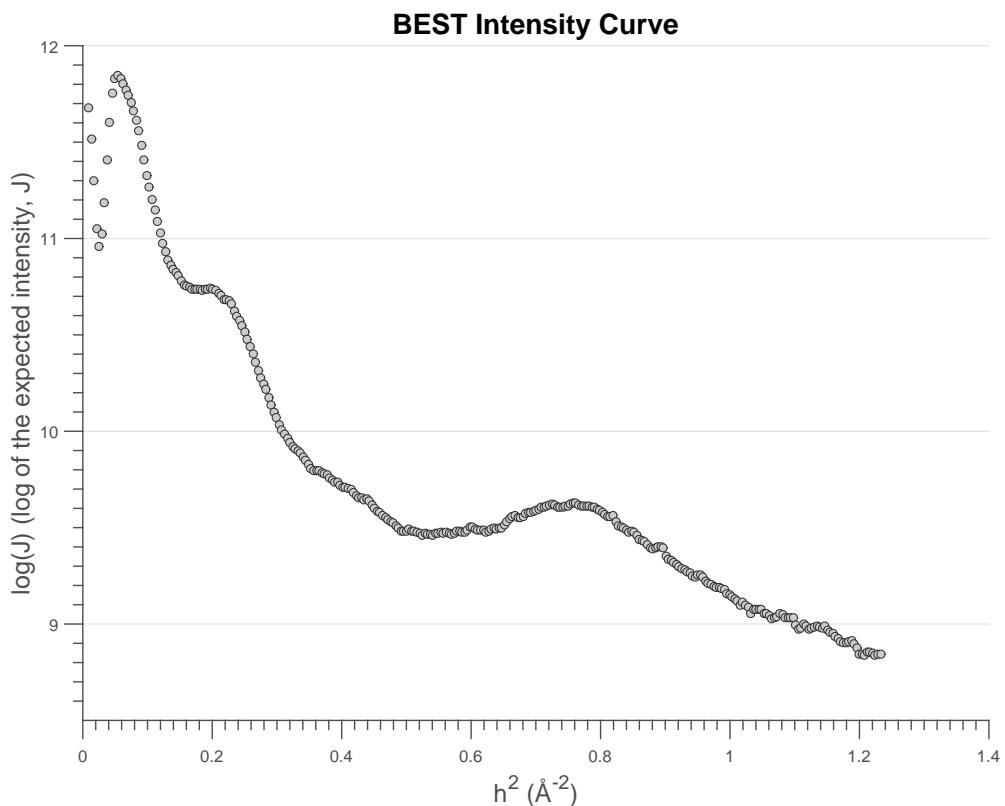


Figure 2.7: Logarithm of the expected intensity values against resolution for 72 different protein crystals which were scaled together (?). The proteins had different folds, molecular masses, space groups, data collection resolutions and were collected at different temperatures (both RT and cryo). This curve is reproduced from data kindly provided by Dr. Gleb Bourenkov.

The  $B(D)$  and  $scale(D)$  values for the Leal *et al.* model were found for each dose by calculating the gradient and intercept from a plot of  $\ln\left(\frac{J(D)}{J(h)}\right)$  against  $h^2$  for each dose. The  $J(h)$  values are given by the BEST intensity curve (?), derived from data collected from 72 different protein crystals at the DESY protein crystallography beamlines (Figure 2.7), and the  $J(D)$  values are the measured mean intensities in resolution bins. The values of  $B_0$ ,  $\beta$ ,  $K$  and  $\gamma$  are found by fitting the functions given by equations (2.3.9) and (2.3.10) to the data values of  $scale(D)$  and  $B(D)$  using the Cauchy M-estimator (Figure 2.8). The Cauchy M-estimator<sup>¶</sup> was used instead of the commonly used least-squares fitting procedure to reduce the influence of errors on the fit.

The parameter values for the Sygusch & Alliare model ( $k'_0$ ,  $k_1$  and  $k_2$ ) were found using an iterative numerical minimisation procedure for which an objective function (i.e. a function that returns a value to be minimised) was created. The objective function was constructed

<sup>¶</sup>M-estimation is a fitting technique designed to be insensitive to outliers. The idea is to try to minimise a function of the residual values that increases less than the square, because the square of the residual of an outlier is large. Thus using a function that increases less than the square decreases the influence of outliers in the fitting.

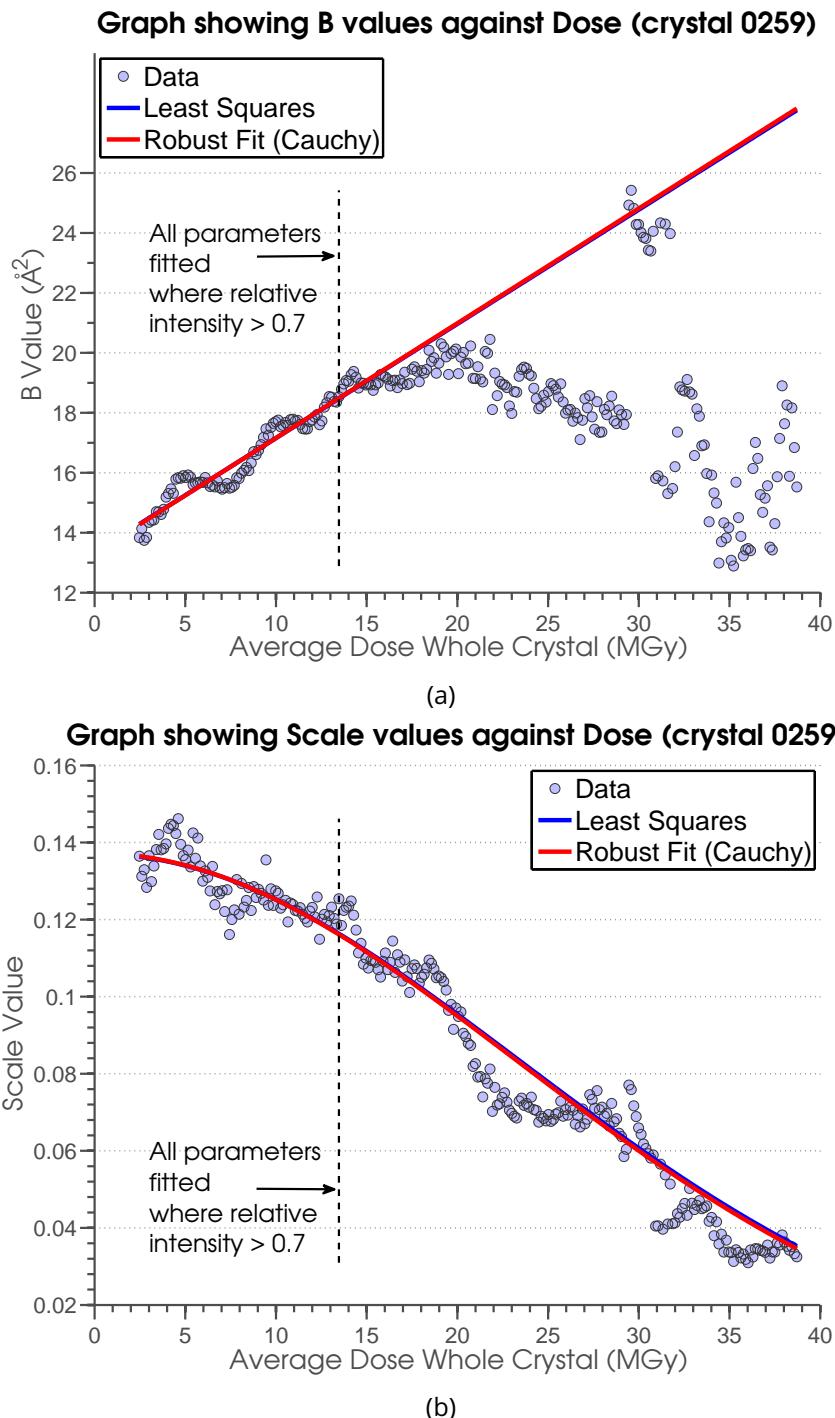


Figure 2.8: (a)  $B$  values plotted against average dose (whole crystal). Initially the  $B$  values increase linearly as reported in the literature (???). However the linearity begins to break down as radiation damage progresses. The group of data points in the top right are possibly an artefact of unstable data processing for highly damaged datasets. This was a major factor in the decision to fit the parameters in Tables 2.5 and 2.4 using only data where the relative intensity was above a given threshold value. The threshold relative intensity value of 0.7 was chosen since other data quality indicators suggest that data beyond this point become significantly biologically compromised (?). (b) Scale values plotted against the average dose (whole crystal). The blue and red curves are fits to the data using equation 2.3.9 and are almost coincident. The same function was also used successfully to fit data from crystals irradiated at room temperature (?), suggesting that this function is suitable for parameterising both cryo and room temperature data.

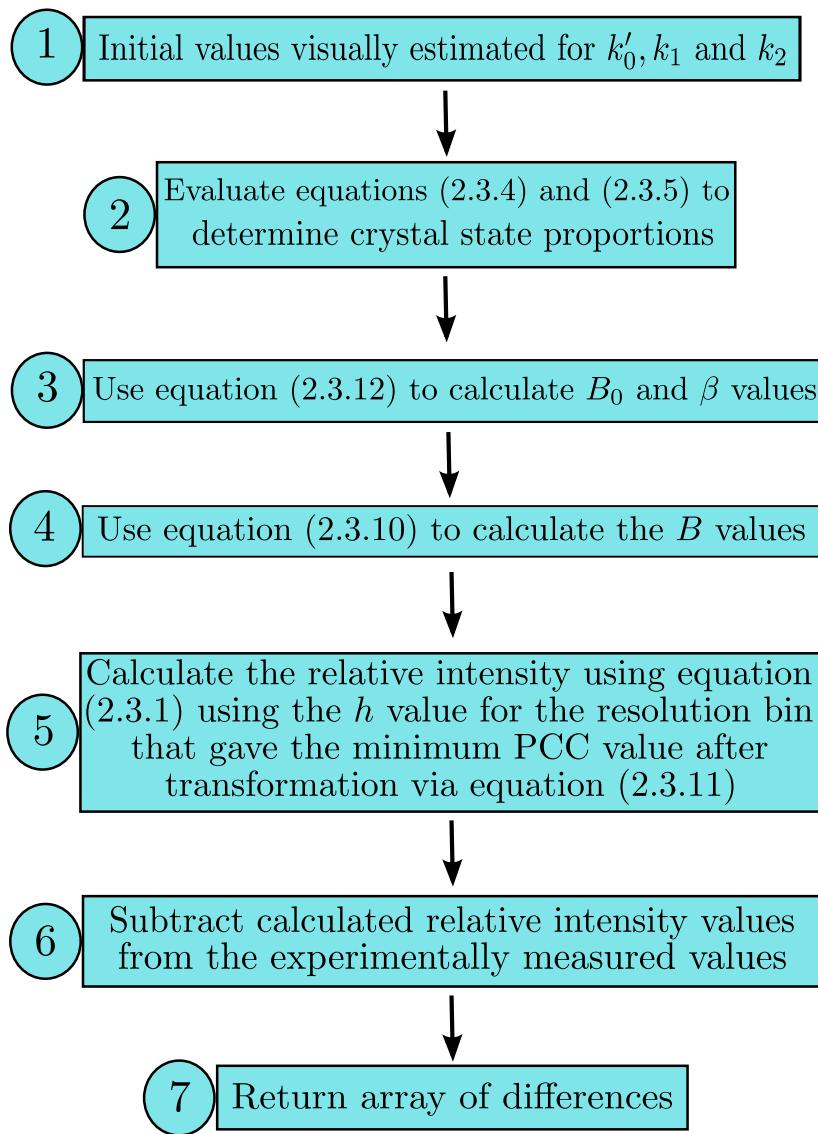


Figure 2.9: Flow chart outlining the structure of the objective function used to find best fit parameters for  $k'_0, k_1$  and  $k_2$ .

according to the flowchart in figure 2.9. This function was minimised using the Matlab `lsqnonlin` procedure which in turn uses the trust-region-reflective algorithm (?). The results are shown in Figure 2.10. The fit to the relative intensity data is very good and the crystal proportions change in a similar manner to that found in Owen *et al.* (2014). It is worth noting that numerical minimisation procedures can return values that correspond to local minima in the parameter space, and therefore the initial choice of values for  $k'_0, k_1$  and  $k_2$  will affect the resulting values returned by the procedure. Initial values in this case were determined by inspection to make sure that the calculated relative intensity values were a close match to the measured intensities as possible. A full parameter analysis has not been performed and therefore the values for  $k'_0, k_1$  and  $k_2$  presented in Table 2.5 may not be the combination that provide the best overall fit to the data. Thus more analysis is required before the model

can be fully evaluated, but these results show the model to be very promising.

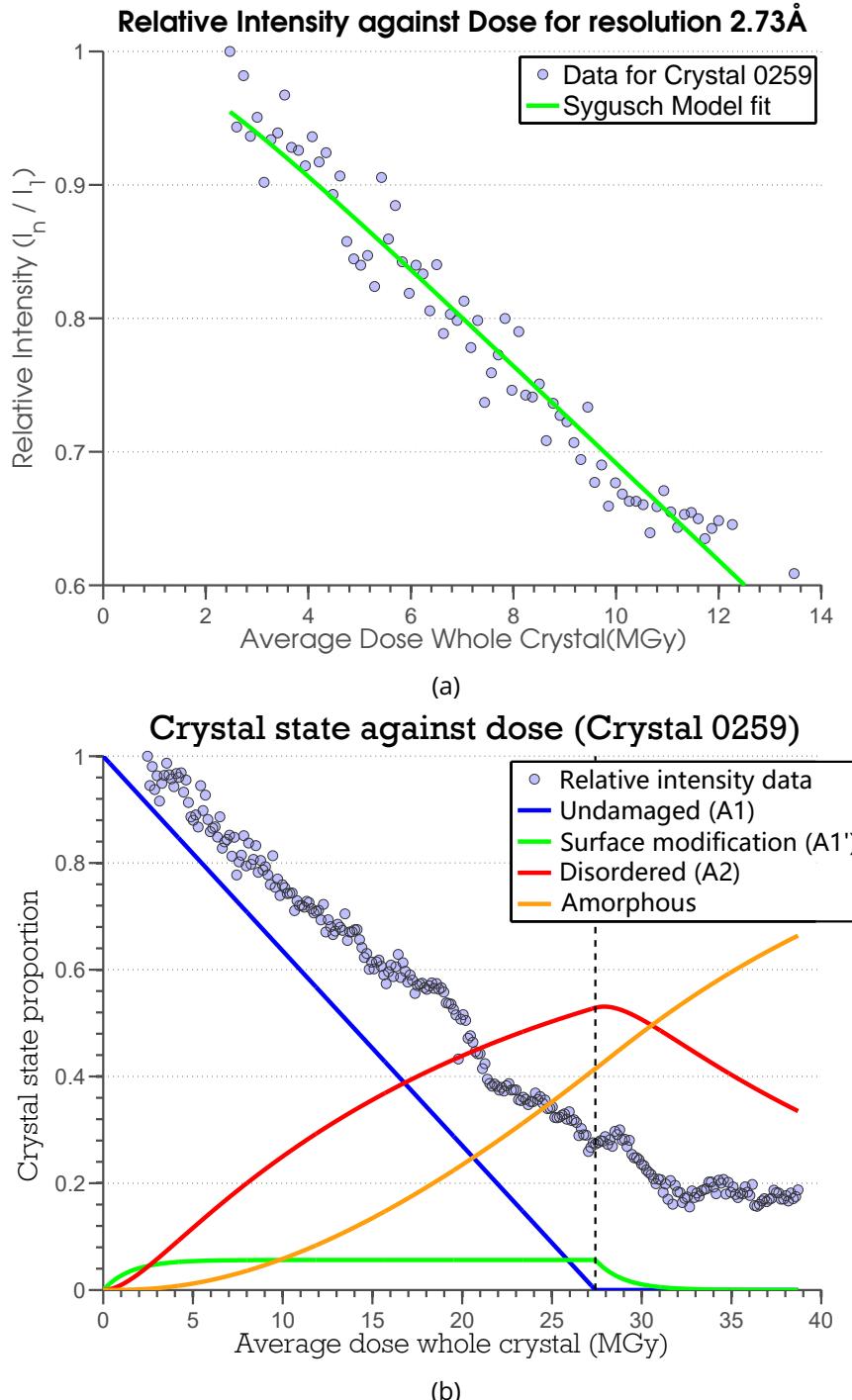


Figure 2.10: (a) Relative intensity against dose (whole crystal) for the 2.73 Å resolution bin. The Sygusch & Allaire model was fitted to these data to obtain parameter values for  $k'_0$ ,  $k_1$  and  $k_2$  using a non-linear least squares procedure. This resolution bin was chosen for this crystal because the data in the resolution bin gave the best PCC value after transformation via equation (2.3.11). (b) Crystal state proportions as a function of the average absorbed dose (whole crystal) according to the Sygusch & Allaire model. The dose at which all of the  $A_1$  proportion has been converted,  $D_L = 27.42 \text{ MGy}$ , is marked as a black vertical dashed line. The overall relative intensity data,  $I_n / I_1$ , for the crystal are also overlaid on the graph.

## 2.4 DDM comparison results

### 2.4.1 Uniform irradiation

To validate the subsequent analysis performed to find a suitable DDM, it was essential that the crystals were uniformly irradiated. A homogeneous dose distribution resulting from uniform irradiation would mean that the various dose metrics: average dose, maximum dose and DWD, should all give the same value. Simulations performed in RADDOSE-3D show that all crystals were completely immersed in the X-ray beam and the predictions displayed showed a very homogeneous dose distribution (Figure 2.11).

### 2.4.2 Calculating the RDE

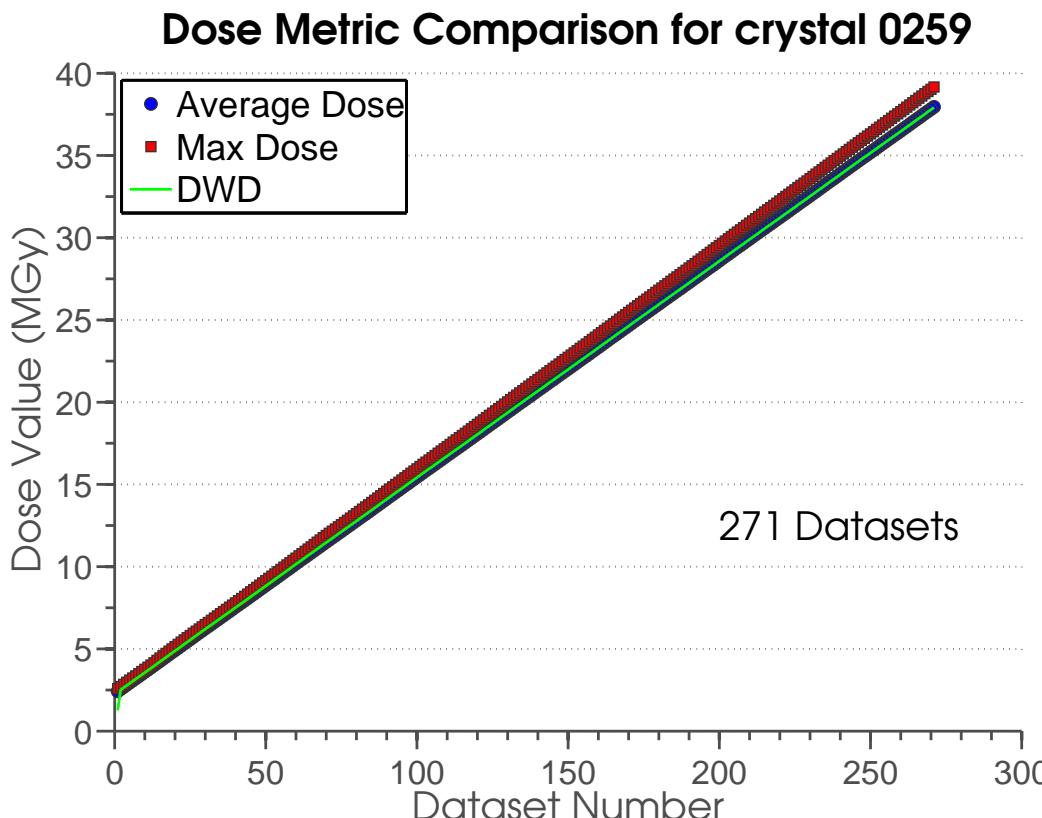
To compare the theoretical RDEs calculated from the DDMs with the experimentally measured relative intensities, a spherically symmetric volume integral is performed on each of equations (2.3.2), (2.3.7) and (2.3.8) up to the resolution limit of the data (1.8 Å) at each of the recorded dose values. These values are then divided by the same integral at dose  $D = 0$  MGy to give an estimate of the relative intensities. The resulting equations are given below and plotted in Figure 2.12 along with the experimentally measured relative intensity values  $I_n/I_1$  for the insulin crystals using the parameter values given in tables 2.5 and 2.4.

$$\text{RDE Holton} = \frac{\int_{h_{min}}^{h_{max}} h^2 I_{ND}(h) \exp\left[\frac{-\ln(2) \times D \times h}{H}\right] dh}{\int_{h_{min}}^{h_{max}} h^2 I_{ND}(h) dh} \quad (2.4.1)$$

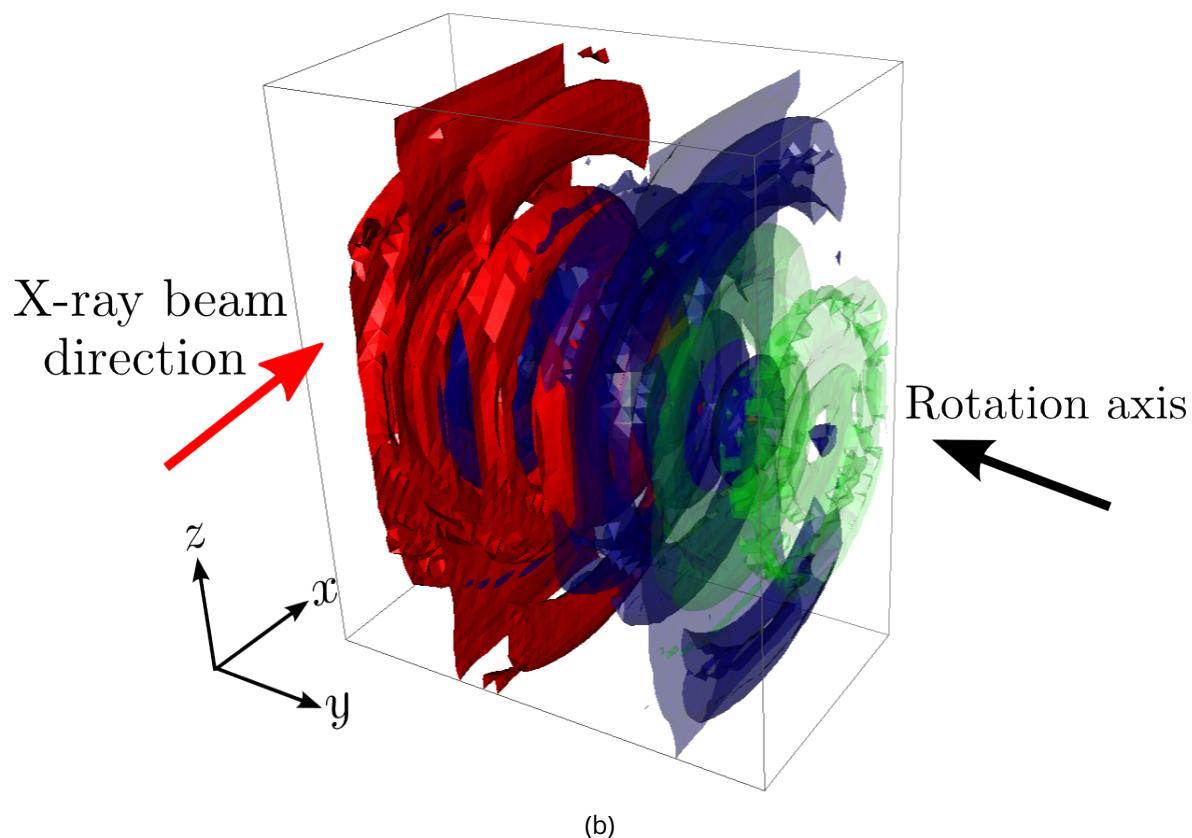
$$\text{RDE Leal} = \frac{\exp[-\gamma^2 D^2] \int_{h_{min}}^{h_{max}} h^2 J(h) \exp\left[-\frac{1}{2}(B_0 + \beta D)h^2\right] dh}{\int_{h_{min}}^{h_{max}} h^2 J(h) \exp\left[-\frac{1}{2}B_0 h^2\right] dh} \quad (2.4.2)$$

$$\text{RDE Sygusch} = \frac{\int_{h_{min}}^{h_{max}} h^2 I(D = 0, h) \left[A_1(D) + A'_1(D) + A_2(D) \exp\left(\frac{-Bh^2}{2}\right)\right] dh}{\int_{h_{min}}^{h_{max}} h^2 I(D = 0, h) \left[A_1(0) + A'_1(0) + A_2(0) \exp\left(\frac{-Bh^2}{2}\right)\right] dh} \quad (2.4.3)$$

The root mean squared deviation (RMSD) was used to assess the fit of the RDE models to the data, with a lower RMSD suggesting a superior fit. Although the parameter values were determined using only the data down to 70% of the initial intensity, the RMSD values are found using the entire range of data for each crystal. Figure 2.12 shows that the RDE using



(a)

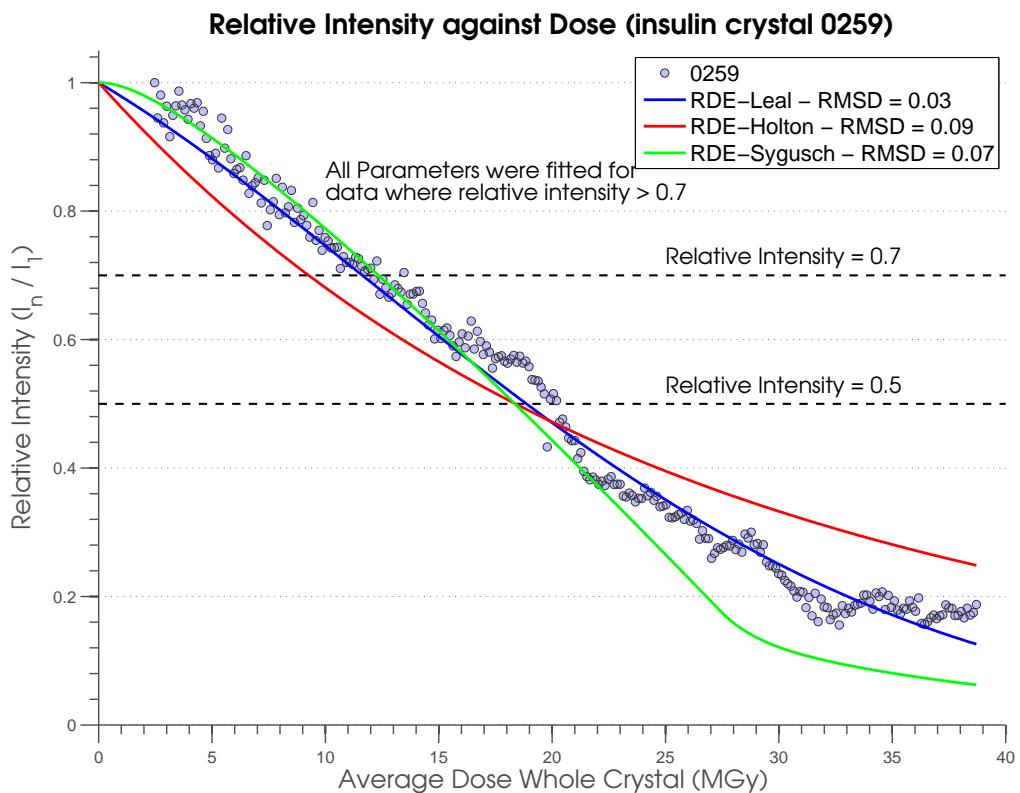


(b)

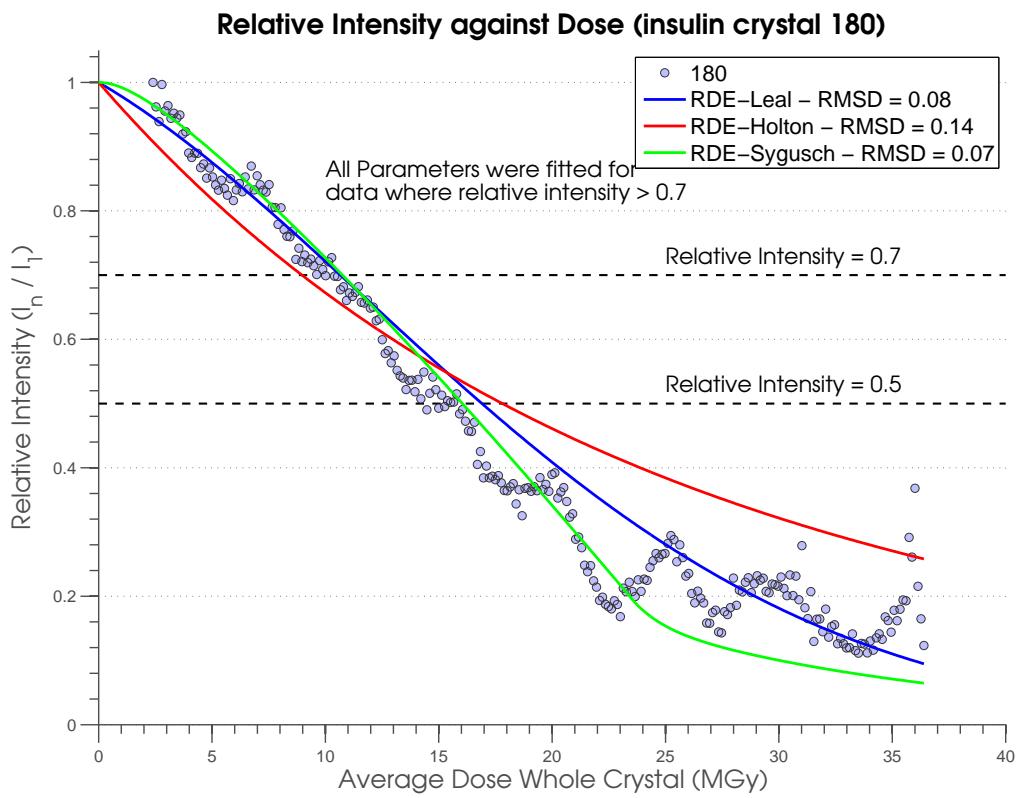
Figure 2.11: (a) The three dose metric values are predicted to be very similar throughout the experiment. Here the DWD values are calculated using equation ???. (b) The calculated dose map from an MX simulation in RADDOS-3D of an insulin crystal ( $89 \mu\text{m} \times 74 \mu\text{m} \times 40 \mu\text{m}$ ) exposed to the beam shown in Figure 2.3a over a  $1440^\circ$  rotation. The dose isosurfaces are: green =  $38 \text{ MGy}$ , dark blue =  $38.5 \text{ MGy}$  and red =  $39 \text{ MGy}$ . The dose distribution is very homogeneous.

the Leal *et al.* intensity decay model best describes the crystal intensity decay for three out of the five crystals (crystal IDs: 0259, 137 and 128) according to the RMSD. For the other two crystals (crystal IDs: 180 and 172), the RDE Sygusch model gives the lowest RMSD, suggesting that this model is also adequate at describing the relative intensity decay. An important feature of the RDE Sygusch model is that it displays at least two phases during the relative intensity decay. Both phases appear to decay linearly, with the first linear phase decaying faster than the second phase. The RDE Holton model consistently gives the highest RMSD value for all five crystals and the decay curve does not correlate well visually with the decay shown in the data. It should be noted that the RMSD does not account for the different degrees of freedom of the models. The Holton model has fewer degrees of freedom than the other two models and so it would be expected to perform worse. A more representative metric to assess the quality of the fit would be something like the Pearson's chi-squared test, however this has not been performed in this analysis.

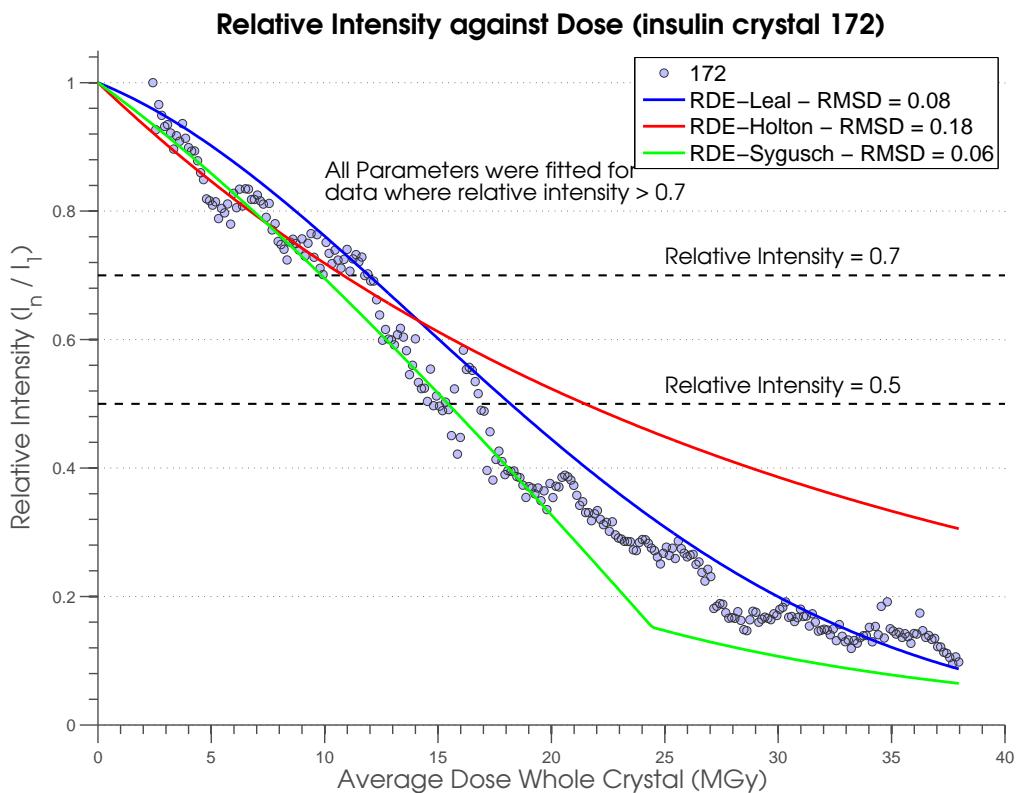
The RDE Leal model was chosen as the one to investigate with the DWD due to its simplicity and superior predictive ability. Figure 2.13 shows the relative intensity values predicted by the RDE Leal model with parameters obtained as described in section 2.3.2 and averaged for each of the 5 insulin crystals processed. However, this time the data were processed to a resolution limit of 1.4 Å (as opposed to the 1.8 Å limit used previously) to ensure that the difference in resolution would not affect the predictive ability of the model.



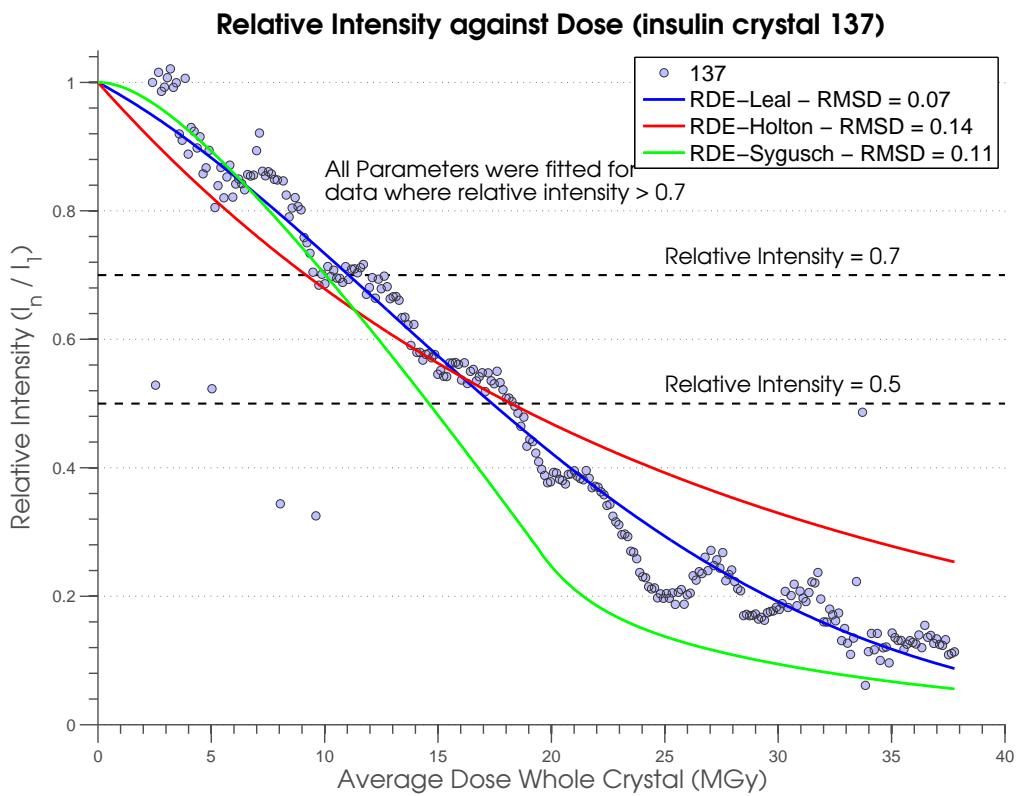
(a)



(b)



(c)



(d)

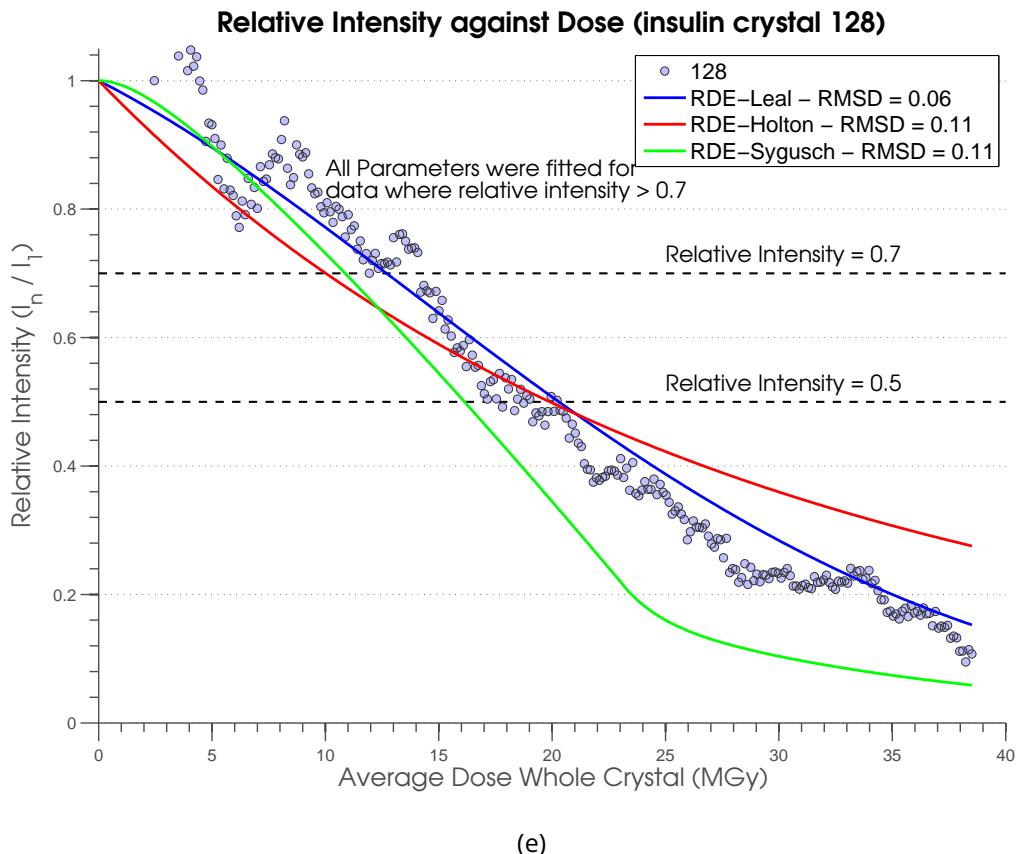


Figure 2.12: (a)-(e) Relative intensity plotted against the average dose over the whole crystal for each of the insulin crystals. Grey circles represent the experimental data processed to  $1.8\text{\AA}$ . Blue, red and green solid lines represent the RDE-Leal, RDE-Holton and RDE-Sygusch models respectively. The RMSDs for each model are given in the figure legend on each plot.

Table 2.4: Best fit zero dose average intensity values ( $I_{ND} = I(D = 0, h)$  (arb. units)) for each resolution bin.  
Note that  $1/h_{mid} = d_{mid}$  where  $d_{mid}$  is the mid-point resolution in each resolution bin

resolution (Å) (1/h <sub>mid</sub> )	Crystal ID				
	0259	180	172	137	128
12.53	12984	8271	20087	29840	37536
7.23	5161	3342	7857	12277	14732
5.60	7764	4946	11258	17550	21769
4.74	11244	7043	15363	25479	31497
4.18	12772	8000	18242	28230	35589
3.78	11325	7050	16401	25625	31156
3.47	8234	5197	11827	19044	22606
3.23	6035	3787	9008	13885	16543
3.04	4293	2673	6447	9662	11656
2.87	3328	2104	5075	7592	9092
2.73	3021	1904	4426	6903	8122
2.61	2417	1522	3569	5509	6473
2.51	2232	1369	3130	4907	5909
2.41	1913	1181	2697	4261	5101
2.33	1659	1010	2332	3668	4406
2.25	1525	959	2170	3479	3990
2.18	1376	853	1982	3041	3619
2.12	1240	746	1697	2656	3237
2.06	958	607	1341	2159	2451
2.01	871	540	1217	1898	2172
1.96	719	429	922	1483	1783
1.91	645	398	835	1367	1594
1.87	524	324	674	1091	1257
1.83	422	263	534	891	1041

Table 2.5: Parameter values for the dose decay models. Note:  $K$  does not explicitly appear in equation (2.4.2) and hence is not required for the analysis. The large  $k_1$  parameter value for crystal 172 seems highly unphysical when compared with the other values. This large value implies that any site specific damaged crystal proportion in the crystal immediately becomes disordered, which suggests that the site specific state is present in negligible quantities for this crystal. Accounting for the intrinsic variability of crystals and the fact that the surface modification fraction of crystal 0259 was never higher than 5% of the total crystal fraction, it is not too surprising that at least one of the crystals may show a negligible surface modification fraction.

units:  $k'_0, k_1, k_2 \equiv MGy^{-1}, H \equiv MGy\text{\AA}^{-1}, B_0 \equiv \text{\AA}^2, \beta \equiv \text{\AA}^2 MGy^{-1}, \gamma \equiv MGy^{-1}$

Crystal	$k'_0$	$k_1$	$k_2$	$H$	$B_0$	$\beta$	$\gamma$
0259	$3.6 \times 10^{-2}$	0.648	$5.0 \times 10^{-2}$	5.466	13.329	0.383	0.030
180	$4.3 \times 10^{-2}$	0.772	$5.0 \times 10^{-2}$	5.274	13.408	0.377	0.036
172	$4.1 \times 10^{-2}$	2310	$5.0 \times 10^{-2}$	6.345	14.818	0.266	0.037
137	$5.1 \times 10^{-2}$	0.473	$5.0 \times 10^{-2}$	5.391	14.136	0.353	0.036
128	$4.3 \times 10^{-2}$	0.624	$5.0 \times 10^{-2}$	5.842	14.538	0.338	0.029

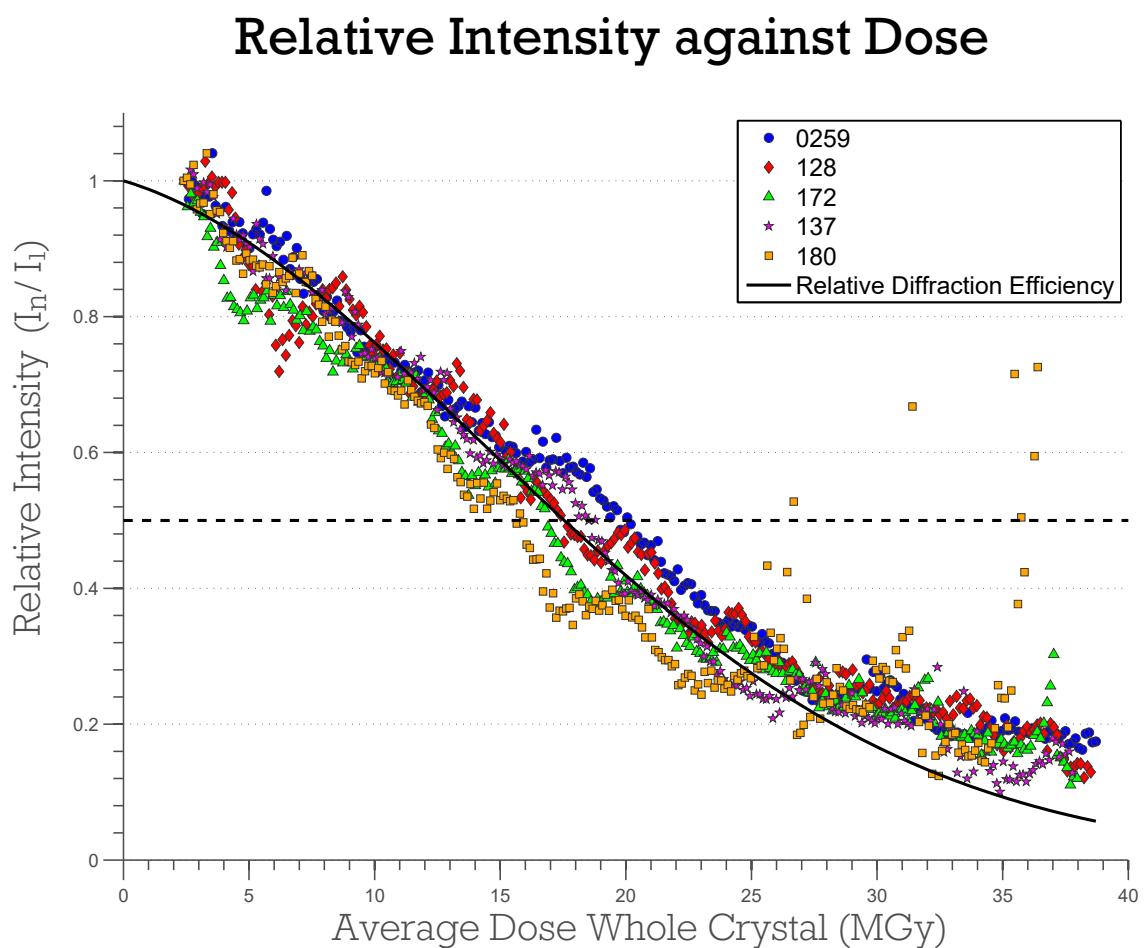


Figure 2.13: Relative intensity plotted against the average dose over the whole crystal for all of the insulin crystals. Discrete data points represent the experimental data for each crystal processed to  $1.4\text{\AA}$ . The black solid line is the RDE Leal model with parameters obtained as described in section 2.3.2 and with averaging of the parameter values for each crystal.

## 2.5 Further investigation of the RDE Leal model

One of the factors that will affect the parameter values that are obtained from the analysis is the resolution of the data collected (?). It is well known that the intensity of higher resolution reflections decay more quickly than for lower resolution reflections. Therefore if the data were collected at a lower resolution, the relative intensity throughout the experiment may be expected to be systematically higher than data collected from the same crystal at a higher resolution. To determine the effect that the resolution dependence would have on the parameter values, the data were scaled to three different resolutions: 1.8 Å, 3 Å, and 4 Å. The parameter values found for each resolution are given in Table 2.6 and the resulting RDE models plotted with the experimental data are shown in Figure 2.14a. It is clear that the calculated relative intensity predictions become worse as the resolution of the data decreases.

The other factor that will affect the results of the RDE model are the resolution limits used to perform the RDE Leal integrals (equation 2.4.2). The resolution of the *BEST* data extend from 12345.68 Å to 0.66 Å (Figure 2.7), thus the integrals can be performed between these limits provided the parameter values have been determined. What is unclear is whether performing the integral to the high resolution *BEST* limit, 0.66 Å, would allow the model to converge for the high resolution relative intensity data despite obtaining the parameter values using lower resolution data. Figure 2.14b shows the RDE models derived from processing the data to the three different resolutions (1.8 Å, 3 Å, and 4 Å) but with the integration performed over the resolution limits of the *BEST* data. The magenta, red and green curves correspond to the RDE models using the parameters determined from the data processed to 1.8 Å, 3 Å, and 4 Å respectively. The green and red curves clearly do not converge towards the expected values of intensity decay (blue circles). The green curve suggests that the crystal is much more resilient than the measured 1.8 Å data. The red curve displays an exponential decay shape that is not seen in relative intensity decay data of crystals at cryotemperatures. On the other hand the magenta curve looks much more reasonable for the decay of cryotemperature data.

These results suggest that to obtain the best predictions of the RDE from the data, it is necessary to collect the data to the highest resolution possible. Furthermore, if the data are

collected to a high enough resolution, then calculating the RDE using the *BEST* limits should give a more representative model of the true RDE of the crystal. This is because the integral is performed over a large resolution range regardless of the diffraction resolution, which will affect the apparent radiation sensitivity.

Table 2.6: Parameter values for Leal *et al.* model determined by the method described in section 2.3.2 with data scaled to different resolution limits.

Parameter Values			
Scaled resolution limit ( $\text{\AA}$ )	$B_0 (\text{\AA}^2)$	$\beta (\text{\AA}^2 \text{MGy}^{-1})$	$\gamma (\text{MGy}^{-1})$
1.8	13.329	0.383	0.030
3.0	1.190	0.621	0.010
4.0	46.277	0.536	0.006

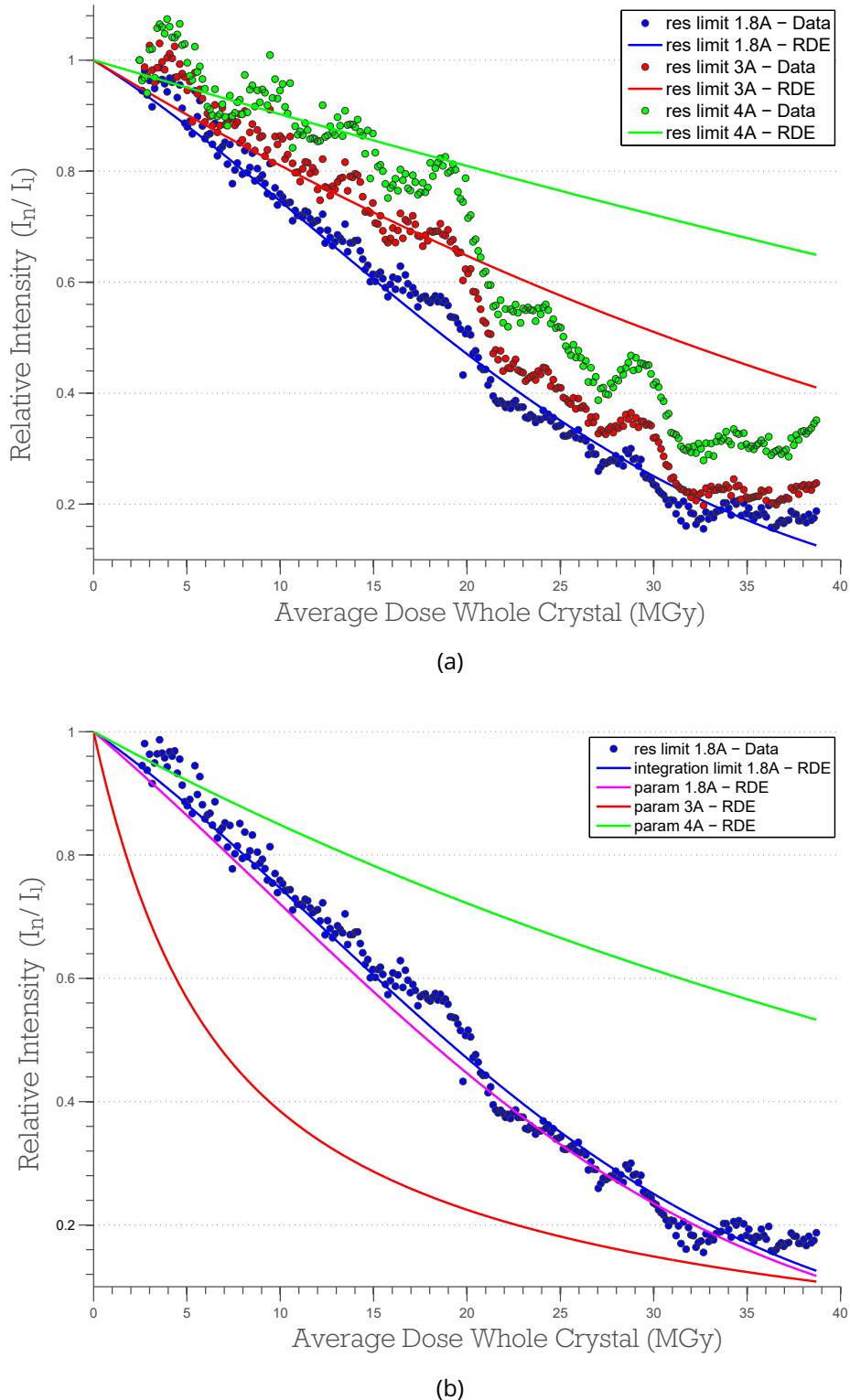


Figure 2.14: (a) Relative intensity data (circles) plotted with the calculated relative intensity using equation 2.4.2, with the corresponding parameter values for each resolution limit from Table 2.6. (b) Relative intensities for the data processed to 1.8 Å (circles) plotted with the calculated relative intensity using equation 2.4.2 with the corresponding parameter values for each resolution limit from Table 2.6, but the high resolution integral limit used was the limit of the *BEST* intensity data, 0.66 Å. The blue solid line corresponds to the calculated relative intensity value with the high resolution integral limit set at 1.8 Å.

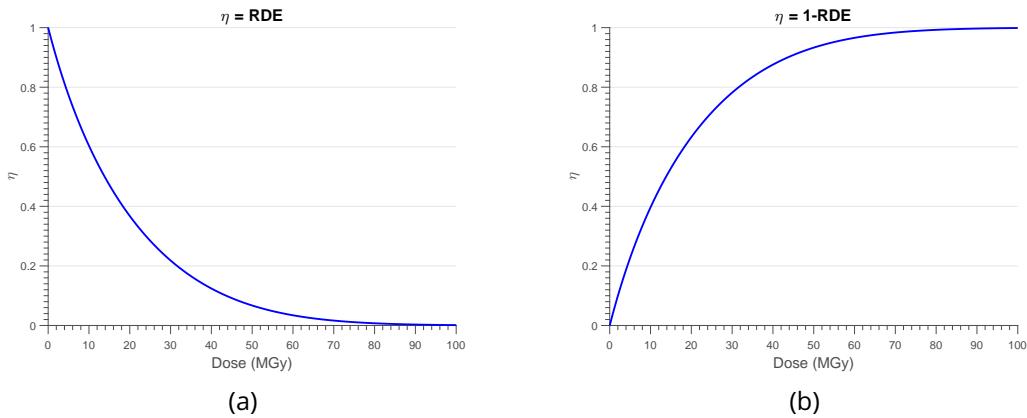


Figure 2.15: Two different forms of  $\eta$  used in equation 2.1.1. (a)  $\eta = \text{RDE}$ . (b)  $\eta = 1 - \text{RDE}$ .

## 2.6 Incorporating the RDE into DWD calculations

The RDE is a monotonically decreasing function of the dose which is bounded between the values of 0 and 1 (equation 2.4.2). If  $\eta = \text{RDE}$  then the decreasing nature of the RDE means that as the dose increases, the DWD defined in equation 2.1.1 will decrease at some point. This may be a desirable characteristic if only diffraction is to be considered. However if the DWD is used to characterise the damage in a crystal, then it would be expected to always increase with increasing radiation exposure. For this reason two forms of  $\eta$  were investigated along with the simple DWD where  $\eta$  had no dose dependence:

$$\eta = 1 \quad \text{simple } \eta \text{ form,} \quad (2.6.1)$$

$$\eta = \text{RDE} \quad \text{decreasing } \eta \text{ form,} \quad (2.6.2)$$

$$\eta = 1 - \text{RDE} \quad \text{increasing } \eta \text{ form} \quad (2.6.3)$$

Both dose dependent forms are bounded between 0 and 1. The difference between them is that equation 2.6.2 decreases from 1 to 0 whereas equation 2.6.3 increases from 0 to 1 as shown in Figures 2.15a and 2.15b respectively.

The remainder of this section presents the results of the analysis carried out with the data from Zeldin *et al.* (2013) using the forms of  $\eta$  given above, comparing their performance with the simple DWD (equation ??).

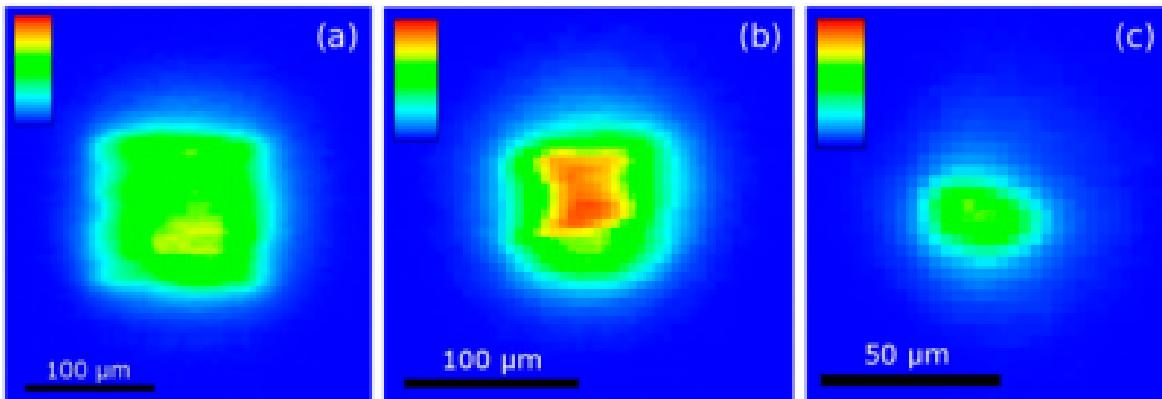


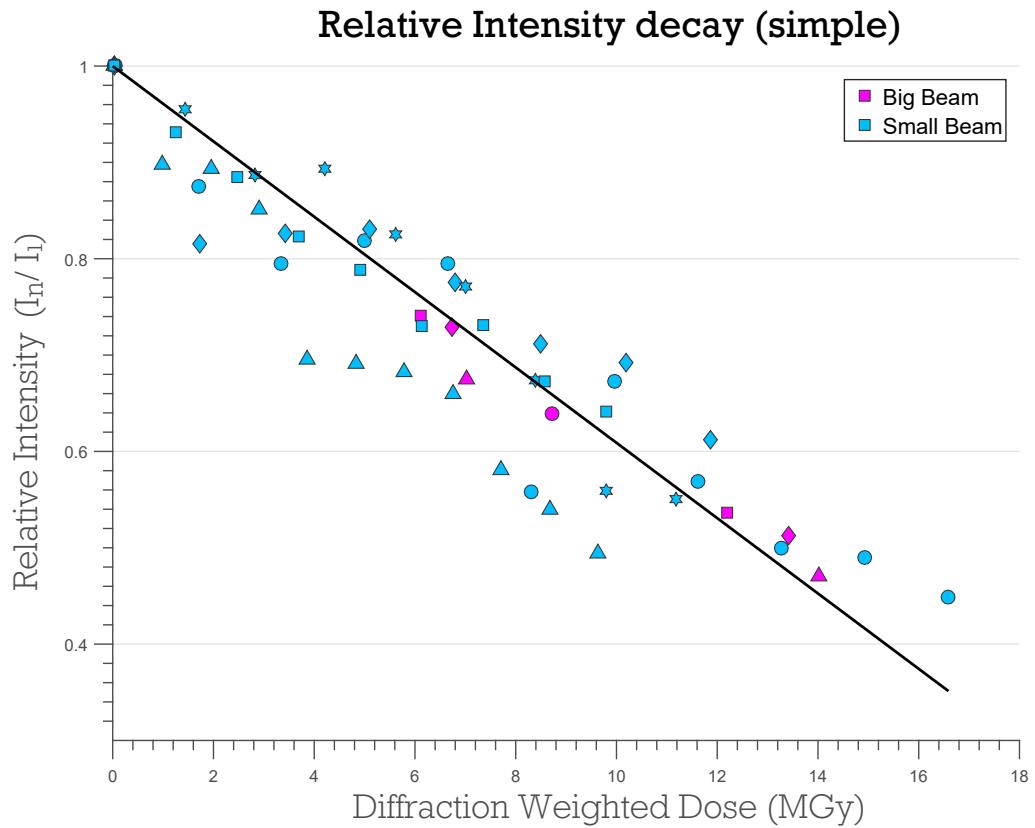
Figure 2.16: False color images of the beam profiles used for the experiments in Zeldin *et al.* (2013). The pixel size is  $5 \times 5 \mu\text{m}^2$  in all cases: (a) big beam, (b) medium beam and (c) small beam.

### 2.6.1 Predicting intensity loss

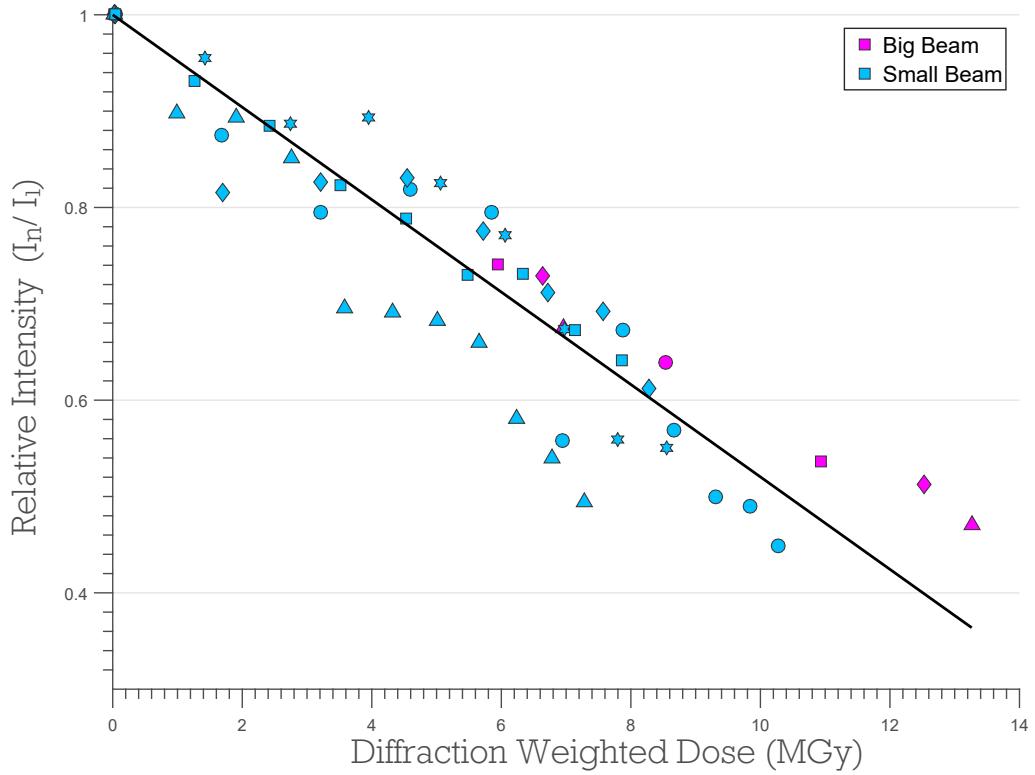
In the study carried out by Zeldin *et al.* (?) cubic crystals of bovine pancreatic insulin were irradiated under different dose contrast conditions. The three beams used in the study: big, medium and small, are shown in Figure 2.16. It was shown that the DWD is a significantly better metric for assessing the extent of radiation damage compared to the average dose for the whole crystal (AD-WC) or the maximum dose, because the spread of relative intensity values around the line of best fit of the data was greatly reduced (plots A-C in Figure 2.21 reproduced from Zeldin *et al.* (2013.)). The data for the big and small beams were available so the analysis performed by Zeldin *et al.* (2013) was repeated for each of the  $\eta$  forms. The results are shown in Figure 2.17. The main difference that can be seen in the plots is the DWD range. The increasing  $\eta$  function results in the largest range of DWD values. Conversely, the decreasing  $\eta$  function reduces the range of DWD values when compared to the simple DWD. Another difference is that introducing a dose dependent  $\eta$  function shifts the big beam data relative to the small beam data. The decreasing  $\eta$  function shifts the big beam data towards higher dose values relative to the small beam data, whereas the increasing function shifts the big beam data towards lower dose values. This suggests that there is a signature from the different beams. However this is an undesirable feature because DWD should already account for the different beam conditions via the flux weighting.

A line of best fit was calculated using a least squares fitting procedure and plotted (solid black line Figure 2.17). A measure of the overall deviation of the data from the line was obtained by calculating the squared Euclidean norm<sup>||</sup> for each DWD form (Table 2.7). The

<sup>||</sup>The squared Euclidean norm is defined as  $\sum_i (f(x_i) - y_i)^2$ .



(a)  
**Relative Intensity decay (decreasing)**



(b)

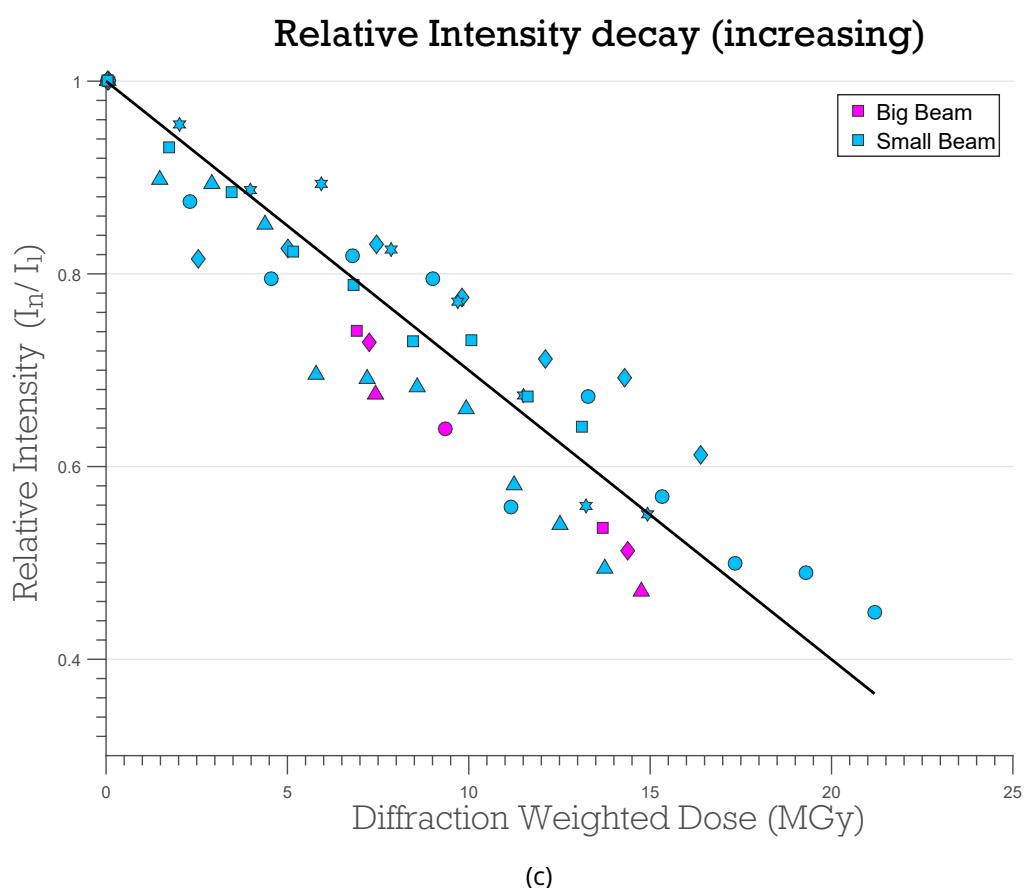


Figure 2.17: Relative intensity decay against the DWD using the different forms of  $\eta$  given by equations 2.6.1, 2.6.2 and 2.6.3. (a) Simple  $\eta$ . (b) Decreasing  $\eta$ . (c) Increasing  $\eta$ .

DWD form that gave the lowest value for the squared Euclidean norm was the simple DWD form. This suggests that the data are less spread overall without adding a dose dependent form for  $\eta$ .

Table 2.7: Squared Euclidean norm values of the line of best fit with the data in Figure 2.17.

$\eta$ form	Squared Euclidean norm
$\eta = 1$ (simple)	0.203
$\eta = RDE$ (decreasing)	0.211
$\eta = 1 - RDE$ (increasing)	0.206

$D_{1/2}$  is a metric of the radiation sensitivity of a protein crystal that is defined as the dose at which the relative intensity falls to 50%.  $D_{1/2}$  values were calculated from the line of best fit for each of the DWD forms. Furthermore  $s_{AD}$  values (as defined in section ??) were also calculated and both sets of values are given in Table 2.8. The spread of the  $D_{1/2}$  values

Table 2.8:  $D_{1/2}$  and  $s_{AD}$  values calculated using DWD with different  $\eta$  forms. The DWD with the simple  $\eta$  form results in the most similar  $D_{1/2}$  values, whereas using the increasing or decreasing  $\eta$  forms gives significantly different  $D_{1/2}$  values. The differences in the ranges of  $s_{AD}$  are relatively large and are not significantly altered by including the various  $\eta$  forms.

Beam size	$D_{1/2}$ (MGy)			$s_{AD}$ ( $\text{\AA}^2/\text{MGy}$ )		
	Simple	Decreasing	Increasing	Simple	Decreasing	Increasing
Big beam	12.94	12.18	13.97	0.0125	0.0133	0.0116
Small beam	13.32	9.91	17.84	0.0068	0.0092	0.0050

for the simple DWD, 0.38 MGy, is much smaller than the spread for the decreasing  $\eta$  form, 2.27 MGy, and the increasing  $\eta$  form, 3.87 MGy. This confirms the result that the spread of the data is increased by incorporating the dose dependent  $\eta$  forms. The ranges of  $s_{AD}$  values are not greatly improved by adding the dose dependent forms of  $\eta$  to the DWD equation. However, given the fact that DWD does not significantly reduce the data spread for the  $B_{rel}$  metric (?), a reduction in the spread of  $s_{AD}$  values was not expected.

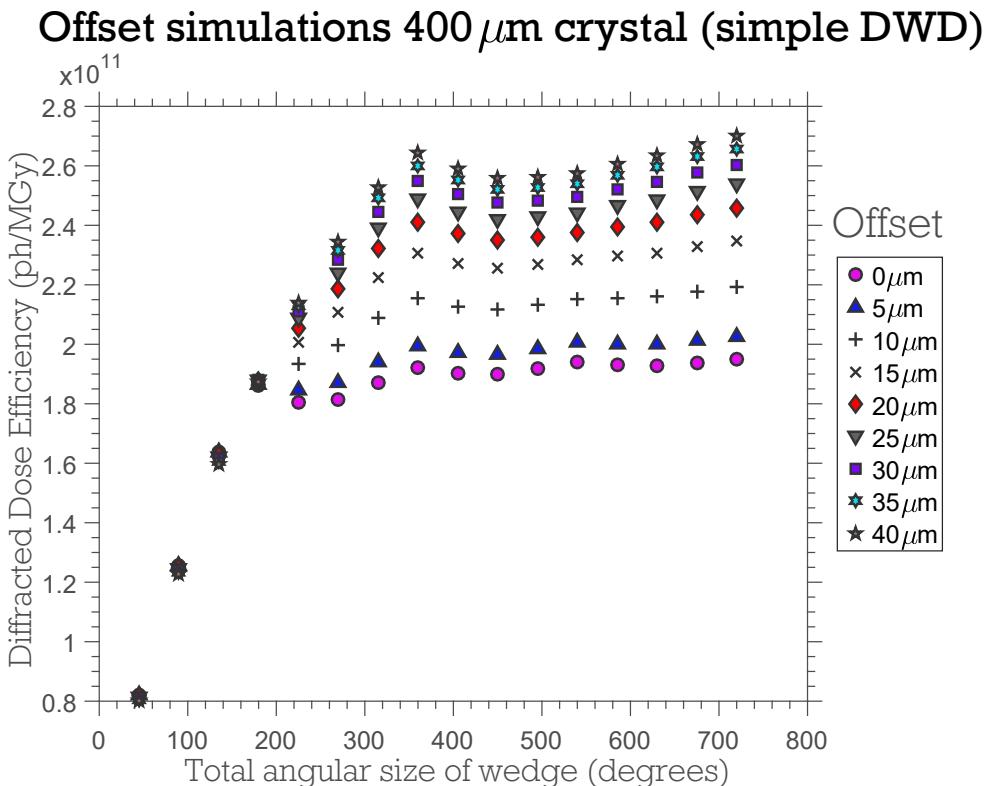
## 2.6.2 Offset simulations

To quantify the efficiency of a given data collection strategy, Zeldin *et al.* introduced a metric called the diffracted dose efficiency (DDE) defined as the ratio of elastically scattered photons to DWD. The DDE states the number of elastically scattered photons that are diffracted per unit dose and it is this quantity that should be maximised for a given experiment. To explore this metric, the authors simulated experiments where the rotation axis was offset from the beam axis by various distances to determine how spreading the dose affected the

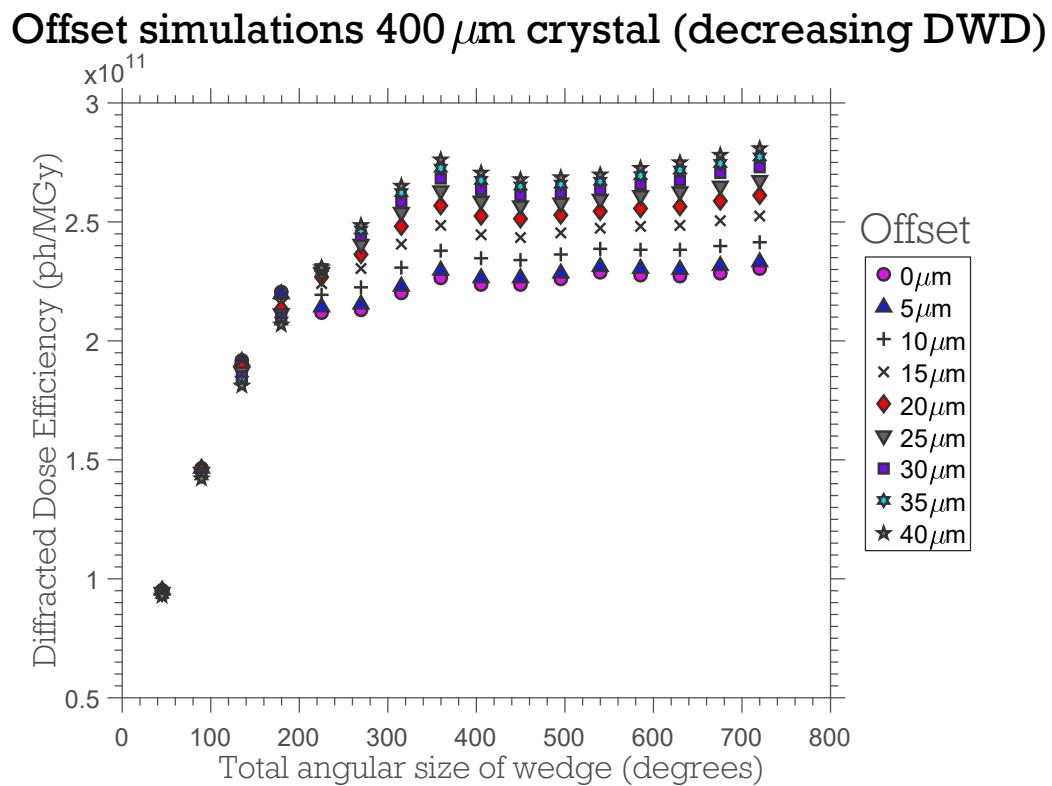
DDE. These simulations were repeated for this work with each  $\eta$  form incorporated into the DWD. In all simulations a cuboid crystal of various sizes and “average” crystal absorption (absorption coefficient =  $0.237 \text{ mm}^{-1}$  (?)) was exposed to a Gaussian profile beam with a  $20 \times 20 \mu\text{m}^2$  FWHM, a flux of  $5 \times 10^{11} \text{ ph/s}$  and incoming photon energy of  $12.4 \text{ keV}$  ( $1 \text{ \AA}$ ). The collimation was rectangular and set to  $40 \mu\text{m} \times 40 \mu\text{m}$  and the total exposure time was 60 seconds. The results of these simulations with each of the forms of DWD are presented in Figure 2.18 for two cubic crystals with edge lengths of  $400 \mu\text{m}$  and  $60 \mu\text{m}$ .

For the  $400 \mu\text{m}$  crystal, Figures 2.18a, 2.18b and 2.18c show that the DDE increases with a greater rotation range up to a  $360^\circ$  rotation. However the decreasing  $\eta$  form of the DWD suggests that the DDE values are generally higher than for the simple form, whereas the opposite is true for the increasing  $\eta$  form. This means that the relative improvements in the diffraction efficiency are different for the different  $\eta$  forms. Even more important is that the inferences that would be made from the simulations are different depending on which form of  $\eta$  is used. The simple  $\eta$  form suggests that there is an almost undetectable difference in the DDE values for the different offsets until the rotation angle is larger than  $180^\circ$ . On the other hand, using the decreasing  $\eta$  form, suggests that offsetting the rotation axis is detrimental for a rotation range below  $180^\circ$ , but for angles larger than  $180^\circ$  the offsetting becomes beneficial. The situation is very different for the increasing  $\eta$  form which suggests that there is negligible difference in offsetting strategy for angles up to  $90^\circ$ , but for rotations ranges larger than  $90^\circ$  it is beneficial to offset the rotation axis.

For the  $60 \mu\text{m}$  crystal (Figures 2.18d, 2.18e and 2.18f) the differences are less pronounced. Of course the DDE values differ, again the decreasing  $\eta$  form suggests an increase in DDE values whereas a decrease in DDE is predicted with the increasing  $\eta$  form. All  $\eta$  forms suggest that no improvement in DDE is gained by offsetting with rotation angles up to  $180^\circ$ , but from  $270^\circ$  upwards it is best to offset the rotation axis by  $15 \mu\text{m}$  from the beam axis. The main differences result from offsetting the crystal by either  $5 \mu\text{m}$  or  $25 \mu\text{m}$  with a rotation range larger than  $270^\circ$ . The simple  $\eta$  form suggests that offsetting by either distance gives similar DDE values but the better distance to offset fluctuates with the overall rotation angle. In contrast, the decreasing  $\eta$  form suggests that offsetting by  $5 \mu\text{m}$  produces superior DDE values than offsetting by  $25 \mu\text{m}$ . However the increasing  $\eta$  form suggests the opposite i.e. for a rotation range larger than  $360^\circ$  it is always better to offset by  $25 \mu\text{m}$ .

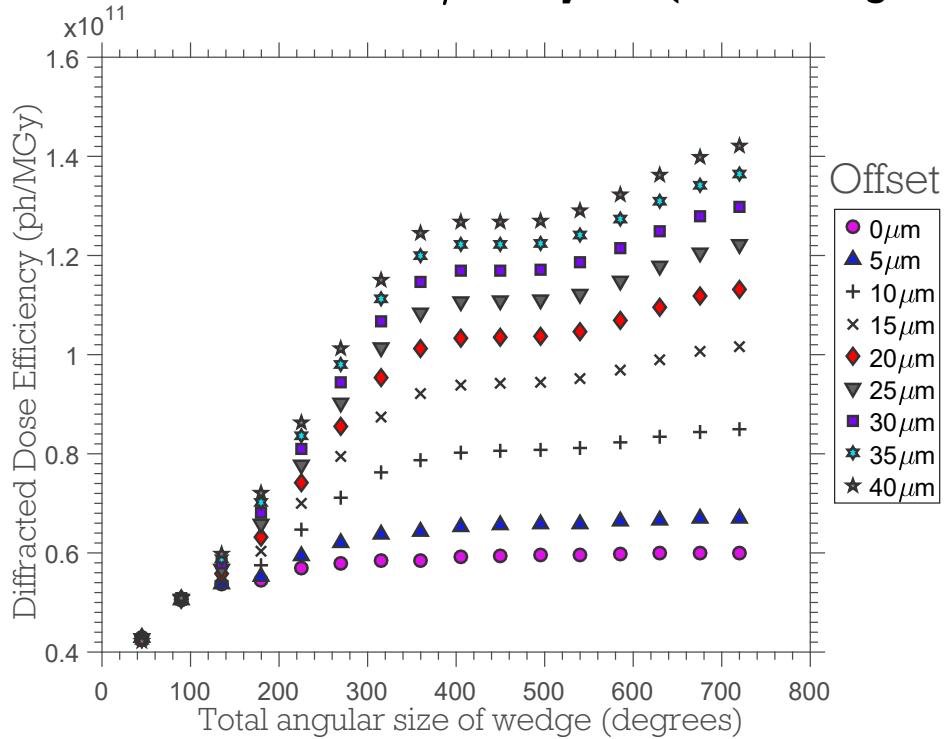


(a)



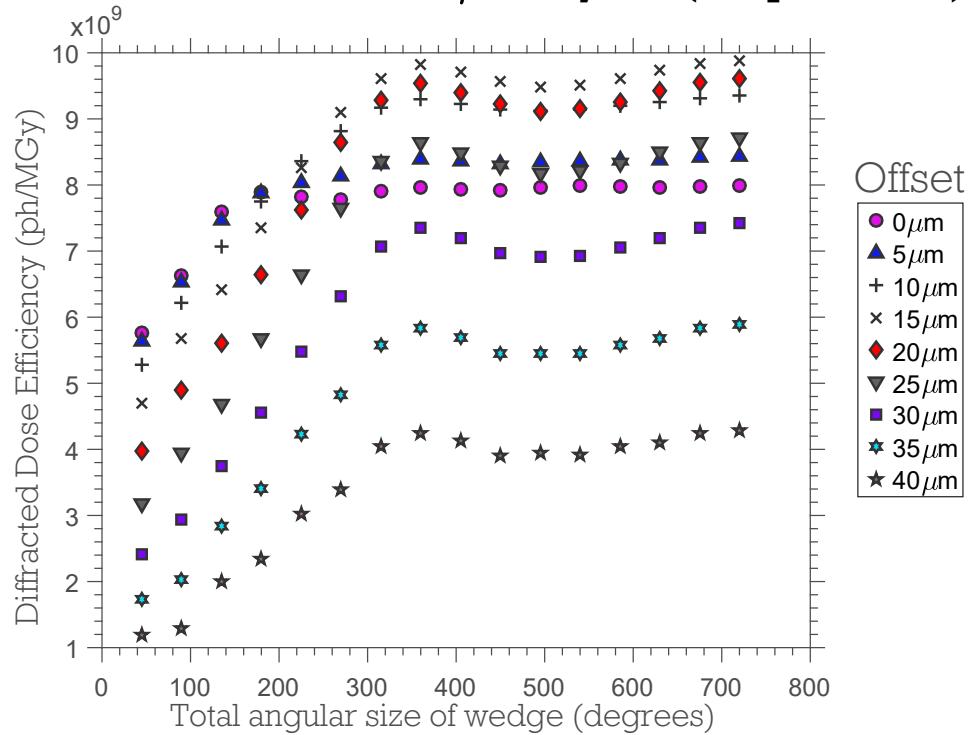
(b)

**Offset simulations 400  $\mu\text{m}$  crystal (increasing DWD)**



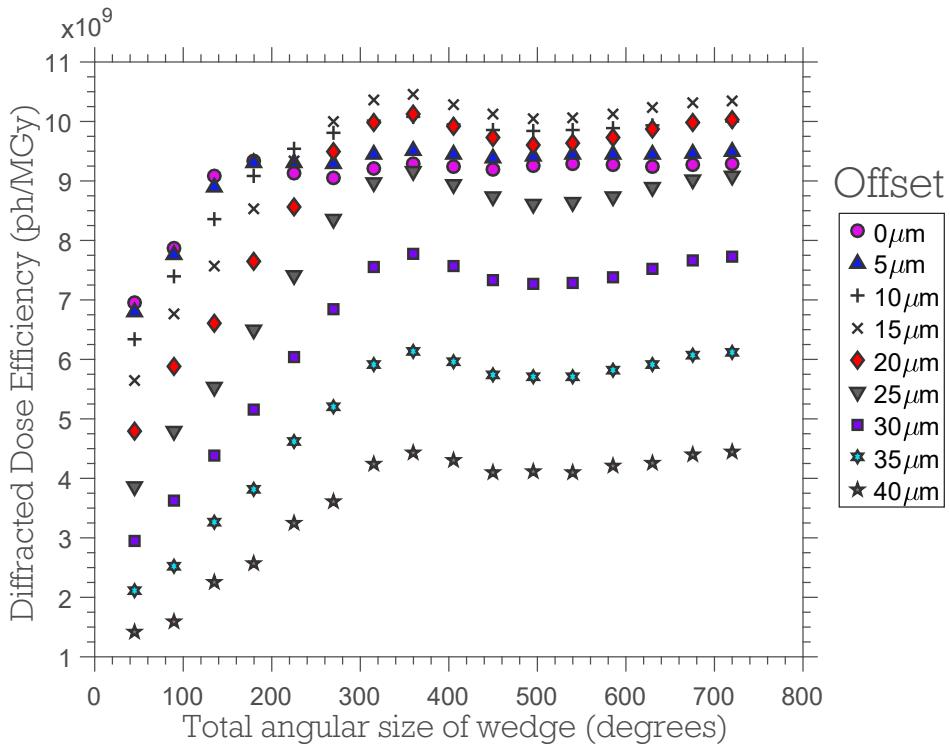
(c)

**Offset simulations 60  $\mu\text{m}$  crystal (simple DWD)**



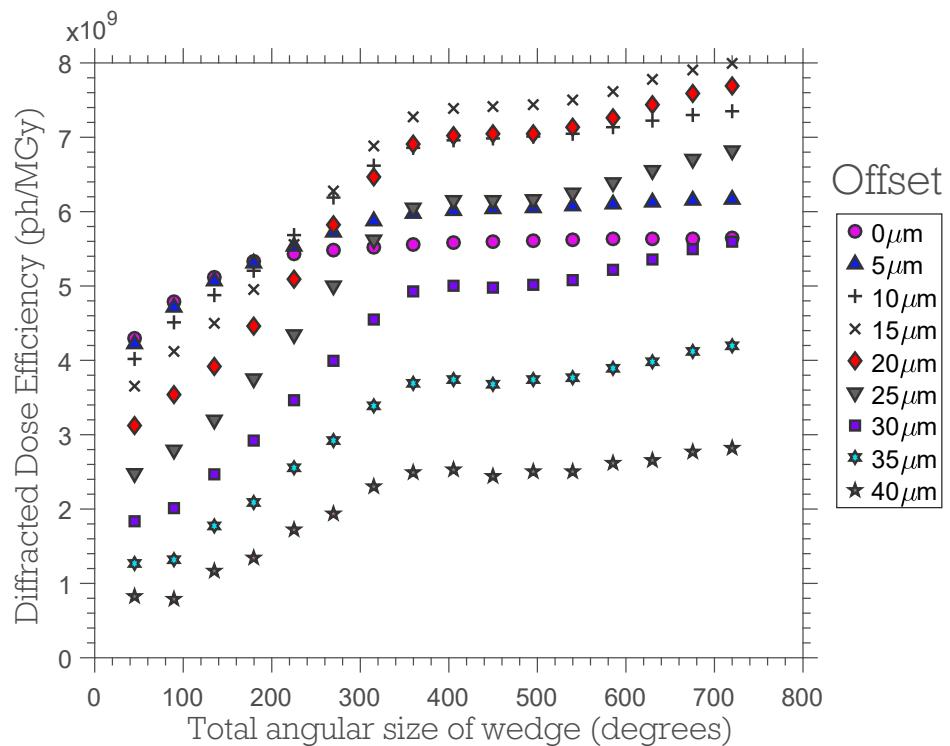
(d)

### Offset simulations 60 $\mu\text{m}$ crystal (decreasing DWD)



(e)

### Offset simulations 60 $\mu\text{m}$ crystal (increasing DWD)



(f)

Figure 2.18: Results from the offset simulations showing the diffracted dose efficiency (DDE) plotted against the total rotation range for two different sized cubic crystals: 400  $\mu\text{m}$  and 60  $\mu\text{m}$  edge lengths. The DDE values are calculated from the DWD using the different forms of  $\eta$  given by equations 2.6.1, 2.6.2 and 2.6.3

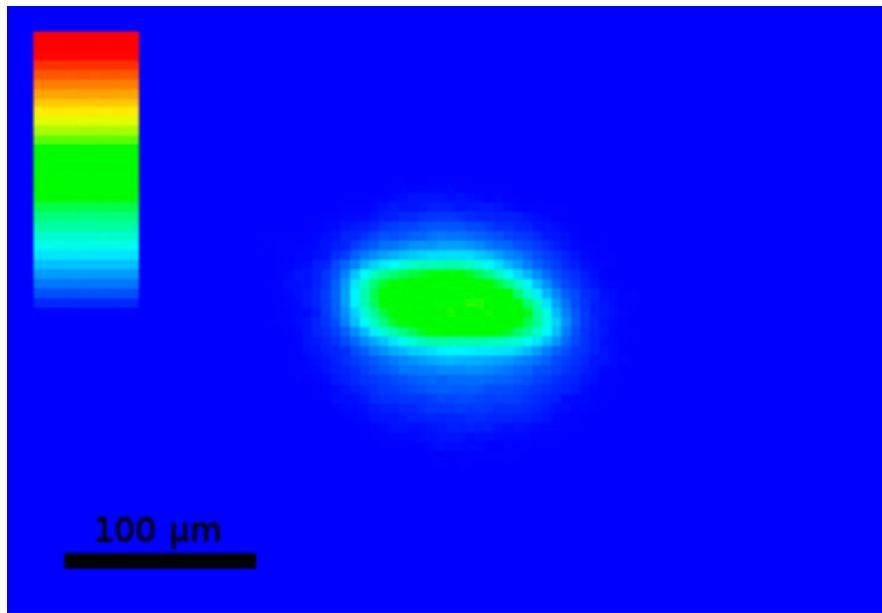


Figure 2.19: False colour images of the beam profile used in the offset experiment by Zeldin *et al.* (2013). The colour bar represents 0-255 intensity units.

### 2.6.3 Offset experiment

An experiment to validate the offset simulations was performed by Zeldin *et al.* and reported in (?). In the experiment, a cuboid crystal of bovine pancreatic insulin ( $460 \times 550 \times 260 \mu\text{m}^3$ ) was irradiated in two regions with a beam of approximate size  $40 \times 70 \mu\text{m}^2$  (Figure 2.19). The first position corresponded to the rotation axis aligned with the beam axis (standard strategy). The second position corresponded to the rotation axis offset from the beam axis by 1.25 times the beam FWHM ( $50 \mu\text{m}$  - offset strategy). At each of the two positions three datasets were collected: first a  $180^\circ$  low dose *probe* dataset, then a high-dose burn dataset, and then finally another low dose *probe* dataset to evaluate the damage state of the crystal after it had been subjected to a high dose X-ray exposure. To achieve the same DWD value (for the simple  $\eta = 1$  form) for each of the high dose datasets, the standard strategy exposed the crystal for 126 seconds whereas for the offset strategy the exposure was 162 seconds. The final dose state of the crystal as calculated in RADDOSE-3D is shown in Figure 2.20. The results from the data processing of this experiment with the different forms of DWD can be seen in Table 2.9. The resulting DWD values for this experiment tell very different stories about the states of the crystal. For the first probe datasets (P1-S and P1-O), the DWD values obtained using the decreasing  $\eta$  form agrees perfectly (to 2 d.p.) with the resulting DWD values using the simple  $\eta$  form. The DWD values for the increasing  $\eta$  form are practically

\*\*Equivalent at 100% transmission ( $1.4 \times 10^{12} \text{ ph/s}$ )

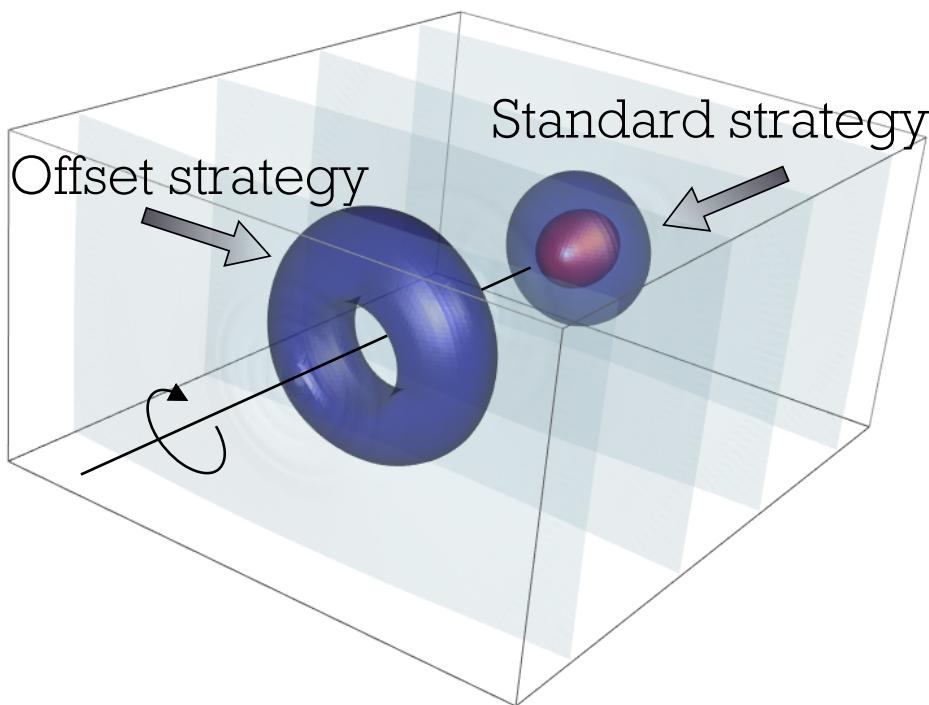


Figure 2.20: Dose isosurface map for the crystal used in the offset experiment. Isosurfaces are at 0.1 MGy (light blue), 5 MGy (dark blue) and 10 MGy (red). Note that the dose at any point in the crystal for the offset experiment is lower than the maximum dose for the standard strategy.

Table 2.9: The rows are presented in chronological order of the experiment. Syntax of the first column is: P1: 1st probe dataset, P2: 2nd probe dataset, HD: high dose dataset, -S: standard strategy, -O: offset strategy.

Wedge	Total time** (s)	Elastic yield (ph)	DWD simple $\eta$ (MGy)	DWD decreasing $\eta$ (MGy)	DWD increasing $\eta$ (MGy)
P1-S	5.4	$5.5 \times 10^{10}$	0.08	0.08	0.16
P1-O	5.4	$5.3 \times 10^{10}$	0.08	0.08	0.14
HD-O	162.0	$1.6 \times 10^{12}$	1.86	1.77	2.85
HD-S	126.0	$1.3 \times 10^{12}$	2.00	1.80	3.73
P2-S	5.4	$5.5 \times 10^{10}$	3.81	1.77	7.03
P2-O	5.4	$5.3 \times 10^{10}$	3.56	3.34	4.83

double that value. For the high dose datasets (HD-S and HD-O) that were designed so that the simple  $\eta$  form DWD was similar for both strategies, the decreasing  $\eta$  form DWD values for the two experiments are the most similar. On the other hand, the increasing  $\eta$  form DWD values differ by almost 1 MGy. This result suggests either that the increasing  $\eta$  form may not actually be representative of the real DWD (i.e. this is the incorrect form), or that the estimated DWD values with the other  $\eta$  forms are misleading. Finally there are several differences in the results for the second probe datasets (P2-S and P2-O). The simple and increasing  $\eta$  form DWD values suggest that the standard strategy results in a higher DWD value than the offset experiment. This is the expected result. The main difference however, is the large range in DWD values for the increasing  $\eta$  DWD (2.2 MGy) compared to the range for the simple  $\eta$  DWD (0.25). Given that the relative intensity ( $I_n/I_1$ ) for the standard and

offset strategies for the second probe datasets are 0.79 and 0.85 respectively, the higher DWD difference could account for the relative intensity difference. The DWD values using the decreasing  $\eta$  form for the second probe datasets could be considered counterintuitive at first sight. The result states that the DWD for the standard strategy is lower than the DWD for the high dose dataset which was taken before the second probe dataset. If radiation damage is considered progressive then this result rules out using the decreasing form of  $\eta$ . However it may be that the diffraction quality is better for the probe dataset than the high dose dataset, in which case this form of  $\eta$  does not represent the level of damage, but instead, it is a dose value that describes the quality of the diffraction. This result suggests that the relationship between the relative intensity and the DWD using the decreasing  $\eta$  form is not a one-to-one function. There are many possible relative intensity values for a given DWD value because the DWD can increase and then decrease again. The DWD values are similar for the simple and decreasing  $\eta$  forms for the offset strategy in the second probe experiment. This is due to the fact that the dose for the offset experiment is smaller and hence the reduction of  $\eta$  for the decreasing form was not significant enough to lower the DWD, as occurred for the standard experiment.

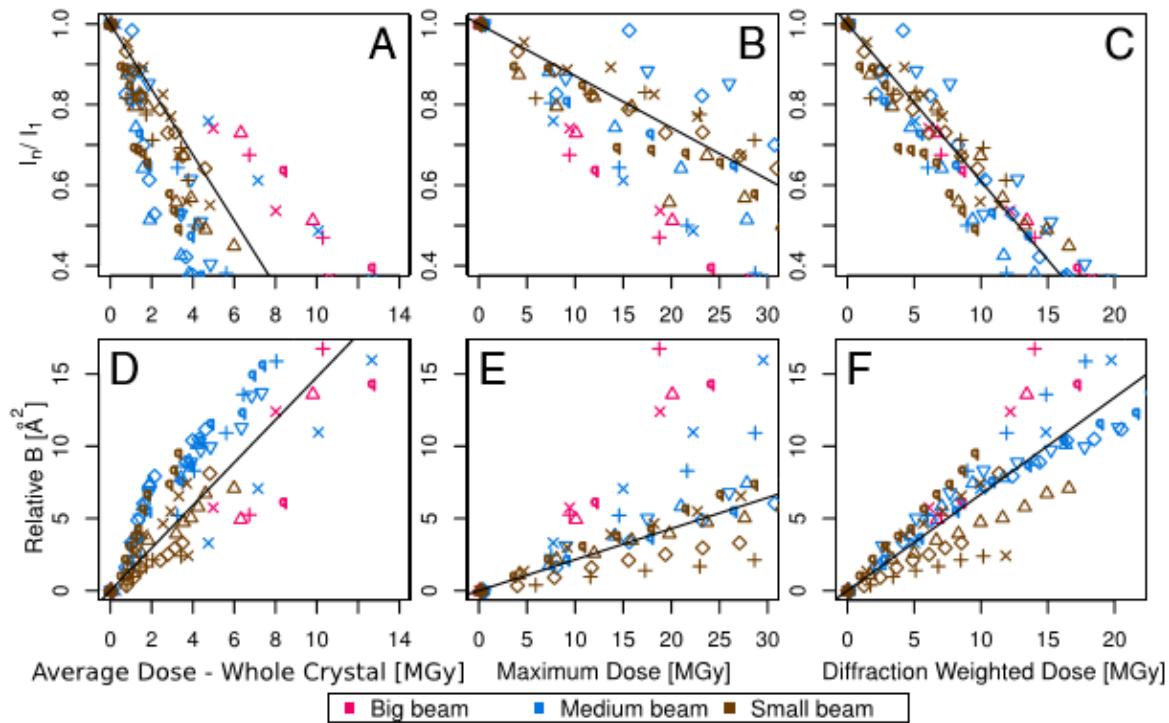


Figure 2.21: (A-C)  $I_n/I_1$  against dose. The reduced scatter along the line of best fit in (C) is evidence that DWD is invariant to the dose distribution in the crystal. (D-F)  $B_{rel}$  against dose. There is no obvious reduction in scatter around the line of best fit for these metrics. This suggests that DWD offers no significant improvement over other dose metrics when using  $B_{rel}$  as a measure of radiation damage progression. The symbols represent individual crystals. Reproduced from (?).

## 2.7 Discussion

The DWD is a dose metric that was introduced by Zeldin *et al.* which represents “the average dose in the diffracting crystal volume from which the photons making up a given image have scattered” (?). In essence, the DWD is a weighted average of the dose where the weights are given by the fluence at each point in the crystal (equation ??). Zeldin *et al.* demonstrated that DWD was superior to other metrics (average dose - whole crystal and the maximum dose) in predicting the relative intensity loss of protein crystals whilst accounting for different beam profiles. The improvement in relative intensity decay prediction makes DWD a promising metric for comparing radiation damage studies and also for zero dose extrapolation. However, data points are still scattered around the line of best fit in Figure 2.21 (C). The scatter is thought to be caused by two factors: the first is the intrinsic variation in crystal quality, and the second is the fact that the DWD does not account for the effects of inhomogeneous dose distribution, such as an uneven loss of diffraction efficiency throughout the crystal (?).

The incorporation of an additional weighting term in the definition of the DWD, which ac-

counts for the loss of diffracting efficiency, is hoped to reduce the remaining scatter (equation 2.1.1). The RDE was introduced as a function of the absorbed dose that would represent the loss of diffraction efficiency. Note that equations ?? and 2.1.1 are equivalent when  $\eta = 1$  (the simple  $\eta$  form). To obtain experimental estimates of the RDE of a crystal it was necessary to perform an experiment where the crystals are irradiated uniformly. This would eliminate any dependence on  $F$  and  $\eta$  in equation 2.1.1, and thus the result of any intensity changes being described solely by the absorbed dose. At the time of the experiment, RADDOS-3D could only model cuboid and spherical crystals, so only cuboid shaped crystals were used in the experiment. Diffraction data were processed for five insulin crystals and the results data were used to study the intensity decay as a function of the dose.

Three dose decay models that describe the relationship of intensity decay of a reflection with absorbed dose were investigated for their suitability as models for the RDE: the Sygusch & Allaire model, the Holton model and the Leal *et al.* model. The Sygusch & Allaire model required the solution of a system of coupled first order linear ordinary differential equations for two domains. The interesting feature of this model is that it predicts two different dynamic regions (Figure 2.10b). The Holton and Leal *et al.* models are relatively simple in comparison and they can be converted into linear forms to readily obtain parameter values. To study these functions as RDEs and compare them to the relative intensity decay, these functions had to be integrated over a sphere in reciprocal space and then be normalised with respect to the zero dose integral. The RDE forms of these models were then assessed for their ability to explain the observed data for the five insulin crystals using the RMSD. The Leal *et al.* model narrowly outperformed the Sygusch & Allaire model overall. However, the behaviour of the relative intensity data above around 27 MGy, particularly for crystals 0259 (Figure 2.12a) and 172 (Figure 2.12c), exhibits a slower decay than the decay below that dose. This suggests that the Sygusch & Allaire model prediction of two behavioural domains may be valid and perhaps the model is a more accurate description of the true dynamics.

The resulting RDE using the Leal *et al.* model (equation 2.4.2) was investigated to determine how resolution changes would affect the calculated parameter values and its ability to explain the observed data. It was found that collecting and processing data to the highest resolution possible, as well as performing the spherical integration to the maximum resolution limits of the *BEST* data is the best approach to faithfully representing the observed data.

The RDE model was incorporated into the DWD using equation 2.1.1 and tested with the data from the Zeldin *et al.* study. On its own, the RDE is a monotonically decreasing function of the dose. This is sensible when describing intensity decay but not necessarily when being used as a dose metric to represent damage in the crystal. Hence by defining  $\eta = 1 - RDE$ , an increasing function could be used to describe the progression of damage as the dose increases. The main difference in the resulting DWD values using the new forms of  $\eta$  is that the dose values are increased when using the increasing function of  $\eta$ , whereas the dose values are decreased for the decreasing  $\eta$  function. One of the important results from the analysis is that the observed scatter in the data is increased when using the dose dependent forms of  $\eta$ , a result that contradicts the hypothesis that adding the terms would decrease the scatter. Another interesting result from including the decreasing  $\eta$  form to the DWD is that it not only decreases the range of the resulting DWD values, but the DWD values decrease once an upper threshold dose limit is reached.

To determine which  $\eta$  form is correct, the offset simulations should be carried out as actual experiments. This is because the different  $\eta$  forms suggest slightly different results regarding the expected benefits obtained by offsetting the rotation axis from the beam axis with crystals of different sizes. In particular, the differences in the benefits of offsetting the rotation axis by  $0\mu m$  and  $40\mu m$  for the  $400\mu m$  are more pronounced than for the  $60\mu m$  crystal. What would need to be established first is how the DDE correlates with the observed intensities i.e. if the difference in DDE between two offsets is  $x$  ph/MGy, then what is the expected difference in the observed (relative) intensities? Furthermore to make the results statistically significant, at least three repeat experiments would need to be carried out to give increased confidence in the conclusions that could be drawn.

Given that the simple DWD (equation ??) already gives a measure of the absorbed dose in the crystal accounting for the crystal composition, it is reasonable to ask what would be added by introducing another dose dependent term into the DWD equation. It may help to think about two crystals of the same protein that crystallise in the same space group and are irradiated under the same conditions by the same source. The quality of the crystals will differ due to the intrinsic crystal variation, resulting in different data quality and possibly different rates of intensity decay. This could be due to different mosaicity or unit cell volumes of the two crystals, which may also alter differentially during the experiment. The absorbed dose alone does not account for these factors, and hence a function that takes them into

account, in theory, should account for the additional variation between crystals. In the analysis, the scatter of the data was increased by incorporating a term that was supposed to reduce it. This result could be due to incorrect parameter values for the RDE model. The method used to determine the parameter values (transforming the data to a linear form before performing a linear fit) may not give the optimum results. The values may be improved by fitting the function to the original data without any transformation. It is unknown how sensitive the dose values are to small deviations from the true parameter values, hence any perturbations of the parameter values could lead to misleading results.

Another question is whether the  $\eta$  function should be an increasing or decreasing function. From the results presented in this chapter, the answer depends on whether DWD should describe the extent of damage or the quality of diffraction. Radiation damage is generally a progressive process which means that reducing DWD values do not make sense when thinking about the metric in terms of describing damage. Hence to describe damage,  $\eta$  should be an increasing function. The results presented here provide some unconvincing evidence for using the increasing function of  $\eta$  over the simple form. Hence an experiment, such as the offsetting one described above, would need to be carried out to determine which form to use.

When framing the DWD as a metric to describe the quality of diffraction, it helps to think about an experiment where a crystal is irradiated by a Gaussian beam. In this case, the Gaussian beam has tails that are being diffracted from relatively undamaged regions of the crystal, even at late stages of the experiment. So when “highly damaged” data are processed, the structure factors that are obtained are more similar to the zero-dose case than they were in the middle of the dataset when the crystal in the bright part of the beam was contributing significantly (James Holton, personal communication). In this case it makes sense for the DWD to decrease because it tells us that the quality of the diffraction has improved, despite the overall damage state of the crystal being worse.

This presents a case for using two metrics concerned with different aspects of radiation damage. One metric that assesses the *damage* caused by the X-rays, which is generally the interpretation of the dose that has been used until now. A second metric could be used to assess the *diffraction quality*. A first step towards this would be to use the RDE as defined in this chapter.