Methods Development for Structural Biology



Robert Oliver Zeldin

St Cross College

University of Oxford

A thesis submitted in partial fulfilment of the degree of

Doctor of Philosophy

Trinity 2013

Abstract

Two research questions are investigated here: the first, major, section addresses the problem of uneven distributions of dose (absorbed energy per unit mass) in crystals used for macromolecular crystallography (MX), and the second presents the development of a high-throughput metalloprotein characterisation technique, HT microPIXE.

In MX, the advent of X-ray microbeam data collection has led to uneven distributions of dose within the crystal volume becoming increasingly common. In these cases, the rotation method creates a highly damaged central region of crystal that stays within the beam throughout exposure, and less damaged outer regions, which are introduced during rotation. This thesis presents a new software program, RADDOSE-3D, which performs a full 3D simulation of the profile of absorbed energy (the dose state) within a crystal during X-ray exposure. In order to utilise this time resolved, 3D picture of the dose state of the crystal, a new metric – Diffraction Weighted Dose – is proposed. This metric is then experimentally validated, and is found to summarise the dose state into a single dose value, which reflects the damage state of the crystal. Simulations are performed using RADDOSE-3D and Diffraction Weighted Dose to compare possible dose spreading strategies, and generalised recommendations for MX experimentalists are offered.

Uniquely identifying the species and stoichiometry of bound metals in protein samples is a significant challenge for biophysical characterisation. Low throughput microbeam Proton Induced X-ray Emission (microPIXE) provides an unambiguous analysis of these properties, but has a limited throughput of ~10 samples per day. As a consequence, its applicability has been restricted to niche cases. This thesis presents significant progress, including proof of principle experiments, on developing sample preparation methods, data acquisition systems, and data analysis protocols to increase this throughput by an order of magnitude, opening up major new applications for the technique.

To my father, for teaching me not to be afraid of equations that require the page to be turned into landscape, when I was eleven.

Acknowledgements

I thank my supervisor, Professor Elspeth Garman, for her patient guidance during the last fours years, and for teaching me not just how to do good science, but how to be a good scientist.

I am indebted to Markus Gerstel for his constant stream of advice and explanations of the more technical aspects of computing. There is no doubt that without him, the computational quality of this work (and my programming ninja skills) would be much lower. I also owe him my thanks for letting me use this LaTeX thesis template.
I owe Edward Lowe my thanks for helping me with data processing, and for always taking the time explain things properly to me. The microPIXE part of this thesis would not have happened without Geoff Grime's input, guidance, and perpetual good humour at the Surrey Ion Beam Centre.

I also owe a great deal to the many people who have humoured my incessant stream of scientific questions, at moments appropriate and otherwise, during the last four years. In particular, the development of DWD benefited greatly from some wonderful conversations with James Holton and Arwen Pearson. Finally, I would like the thank all the members of the Garman group for their help throughout.

I am grateful to the Life Sciences Interface program of the EPSRC funded Doctoral Training Centre for funding the studentship that has made this work possible, to the John Fell Fund at the University of Oxford for helping to fund the microPIXE work, and to the Surrey Ion Beam Centre for generously providing the beamtime necessary for these studies.

Contents

1 Introduction to Radiation Damage 1
- 1.1 Atomic resolution structures in biology . . . 3
  - 1.1.1 Macromolecular crystals . . . 4
  - 1.1.2 From crystals to electron density maps . . . 5
  - 1.1.3 Limitations of crystallography . . . 8
- 1.2 Radiation damage as a bottleneck of successful structure determination . . . 9
- 1.3 Physical origin of radiation damage . . . 11
  - 1.3.1 Primary X-ray–crystal interactions . . . 11
  - 1.3.2 Electron cascades, secondary damage and temperature dependence . . . 13
  - 1.3.3 Dose, and modelling it using RADDOSE v1-3 . . . 16
- 1.4 Manifestation of damage: Global and Specific damage . . . 20
- 1.5 Global damage . . . 20
  - 1.5.1 Room temperature . . . 21
  - 1.5.2 Cryo-temperatures . . . 25
- 1.6 Specific damage . . . 29
- 1.7 Other variables . . . 32
  - 1.7.1 Dose rate . . . 32
  - 1.7.2 Scavengers . . . 34
  - 1.7.3 Temperatures <100K . . . 34
- 1.8 Challenges in the field: why is it so hard to obtain reproducible results? . . . 35
  - 1.8.1 Choice of metrics . . . 35
  - 1.8.2 Intrinsic variation in crystal quality . . . 37
  - 1.8.3 Dose modelling . . . 37
- 1.9 This thesis in context . . . 37

2 Historical Development of RADDOSE-3D 41
- 2.1 Background, and limitations of 1-D models . . . 43
- 2.2 A note on coordinates . . . 44
- 2.3 Early development in Matlab . . . 45
  - 2.3.1 Representing the crystal . . . 46
  - 2.3.2 Representing the beam . . . 47
  - 2.3.3 Assigning exposure to voxels . . . 47
  - 2.3.4 Assigning dose to voxels . . . 50
  - 2.3.5 Moving from 2- to 3D . . . 57
- 2.4 Move to Java and development of object oriented structure . . . 60

3 RADDOSE-3D: the Code 63
3.1 Design principles and overview of the code . . . 65
3.2 Technical description of the code . . . 66
3.2.1 The Crystal class . . . 69
3.2.2 The Beam interface . . . 71
3.2.3 The Wedge class . . . 73
3.2.4 Input . . . 74
3.2.5 Output . . . 75
3.3 Aggregate Metrics . . . 79
3.3.1 Simple dose metrics . . . 79
3.3.2 Diffraction Weighted Dose [DWD] . . . 82
3.3.3 Strategy evaluation metrics . . . 85
3.4 A Typical MX experiment . . . 87
3.5 Two further examples . . . 91
3.5.1 Minimal input file . . . 91
3.5.2 Complex input file . . . 94
3.6 Online implementation . . . 98
3.7 Conclusions . . . 98

4 Predicting Damage in the Uneven-dose Regime 99
4.1 Experimental design . . . 101
4.2 Methods . . . 103
4.2.1 Crystallisation . . . 103
4.2.2 Beam profiling . . . 104
4.2.3 Crystal modelling and alignment . . . 106
4.2.4 Data collection strategies . . . 110
4.2.5 Dose calculation . . . 110
4.2.6 Data processing . . . 112
4.3 Results . . . 114
4.3.1 Data collection statistics . . . 114
4.3.2 General consideration on dose metrics . . . 119
4.3.3 Comparison of dose metrics for I_n/I_1 . . . 123
4.3.4 Analysis of B_{rel} . . . 136
4.3.5 Real space analysis . . . 146
4.4 Proof of concept . . . 156
4.4.1 About $PepT_{St}$ . . . 157
4.4.2 Methods . . . 158
4.4.3 Current state of the art: EDNA strategy. . . . 161
4.4.4 Evaluating the crystal, and planning an optimised strategy. . . . 161
4.4.5 RADDOSE-3D Strategy 1 . . . 163
4.4.6 RADDOSE Strategy 2: offset . . . 164
4.4.7 Discussion . . . 171

5 Optimising the Spatial Distribution of Dose 175
5.1 Diffraction Efficiency of Standard, Helical, and Translational strategies. . . . 179
5.2 The Offset strategy . . . 183
5.2.1 Simulations . . . 185
5.2.2 Discussion and results . . . 185

6 Development of High Throughput micron beam Proton Induced X-ray Emmission (HT microPIXE) 191
6.1 The role of metals in structural biology . . . 193
6.1.1 Alternative techniques . . . 194
6.2 X-ray fluorescence as a metalloprotein assay . . . 195
6.3 Experimental set-up . . . 200
6.4 Other samples . . . 202
6.4.1 Whole cells . . . 202
6.4.2 Homogenates . . . 204
6.4.3 Fluids: Cerebral Spinal Fluid (CSF), serum and others . . . 205
6.5 The need for HT microPIXE . . . 206
6.6 Feasibility study of HT microPIXE . . . 207
6.7 Requirements of HT microPIXE . . . 208
6.8 Sample Preparation . . . 209
6.9 ArrayJet printing proof of principle . . . 212
6.10 Development of new sample holders . . . 216
6.10.1 Sample holder design parameters . . . 216
6.10.2 Sample jig . . . 217
6.10.3 Sample stick . . . 219
6.11 Optimisation of printing protocols . . . 221
6.11.1 Verification that there is no cross-talk between drops . . . 221
6.11.2 Sample additives for better print quality . . . 223
6.11.3 Drop multiplicity . . . 229
6.12 HT data acquisition methods . . . 232
6.12.1 Spatial localisation of the spots in the sample chamber. . . . 234
6.12.2 Data analysis: choice of metrics . . . 236
6.13 Final protocol and method for HT micro-PIXE . . . 237
6.14 Measurements in collaboration . . . 239

7 Conclusions 243
7.1 Radiation Damage . . . 245
7.1.1 Comparing dose metrics . . . 245
7.1.2 Using RADDOSE-3D and DWD in routine MX . . . 245
7.1.3 Working with challenging crystals . . . 247
7.1.4 Photoelectrons, and what RADDOSE-3D does not do . . . 248
7.2 MicroPIXE . . . 248

Appendix 1: RADDOSE-3D command reference 248
8.1 General syntax considerations . . . 251
8.2 `Crystal` block . . . 252
8.2.1 **`TYPE`** . . . 252
8.2.2 **`DIMENSION`** . . . 253
8.2.3 **`DDM/DIFFRACTIONDECAYMODEL`** . . . 254
8.2.4 **`PIXELSPERMICRON`** . . . 254
8.2.5 **`ANGLEP`** . . . 254
8.2.6 **`ANGLEL`** . . . 255
8.2.7 **`ABSCOEFCALC`** . . . 255
8.2.8 **`UNITCELL`** . . . 256

8.2.9 NUMMONOMERS . . . 257
8.2.10 NUMRESIDUES . . . 257
8.2.11 NUMRNA . . . 258
8.2.12 NUMDNA . . . 259
8.2.13 PROTEINHEAVYATOMS . . . 259
8.2.14 SOLVENTHEAVYCONC . . . 260
8.2.15 SOLVENTFRACTION . . . 260
8.3 Beam block . . . 261
8.3.1 TYPE . . . 261
8.3.2 FLUX . . . 262
8.3.3 FWHM . . . 262
8.3.4 ENERGY . . . 262
8.3.5 COLLIMATION . . . 262
8.4 Wedge block . . . 263
8.4.1 EXPOSURETIME . . . 263
8.4.2 ANGULARRESOLUTION . . . 263
8.4.3 STARTOFFSET . . . 263
8.4.4 TRANSLATEPERDEGREE . . . 264
8.4.5 ROTAXBEAMOFFSET . . . 264

Appendix 2: publications derived from the work presented in this thesis 264
Zeldin, Gerstel, and Garman 2013a . . . 267
Zeldin, Gerstel, and Garman 2013b . . . 276
Garman and Zeldin 2013 . . . 283

Bibliography 295

List of Figures

1.1 Schematic of decay paths after photoelectric absorption, from Paithankar *et al.* (2009). An atom *(a)* absorbs an incident photon, leading to the ejection of an inner shell electron (*b*). The excited atom can then decay emitting either an Auger electron, or a fluorescent photon (*c*). The relative probability of the two escape routes depends on the atomic species, with Auger electron emission being much more likely for the light elements that make up the bulk of biological samples. . . . 14
1.2 Energy dependence of the attenuation cross sections for the three processes relevant to MX for a crystal of chicken egg-white lysozyme from Paithankar and Garman (2010). Photoelectric absorption dominates, and the Compton effect only becomes significant at higher energies. 14
1.3 Examples of raw images collected with the same exposure parameters from a crystal of cubic insulin at 100K after various levels of radiation damage. Note how the intensity drops in the high resolution bins, and how the remaining spots increase in size due to mosaicity increases. 22
1.4 Three state Blake and Phillips model. Figure from Southworth-Davies *et al.* (2007). The model adds an alternate route to the amorphous state (A_3) *via* a disordered state (A_2) which contributes to diffraction only at low angles. 23
1.5 Comparison of radiation damage metrics for four holoferritin crystals the work of Owen *et al.* (2006), figure reproduced from Garman (2010). 30

2.1 Schematic of right-handed RADDOSE-3D coordinate system used throughout this thesis. As viewed from above for horizontal goniometer rotation axis. 45
2.2 Example of the projected Gaussian beam profile used in early Matlab development of RADDOSE-3D. 47
2.3 Evolution of exposure and dose assignment. The early (all 2D and 3D codes until June 2011) method simply found the point in the beam with the smallest z value and defined this as $surf(x_i, y_i)$ for all points within the crystal which are in the beam. This leads to errors in the depth required for the exposure estimation of size Δz_1 for a point p. These are small for low surface normal - beam angles, but this is not generally the case, and so this error cannot be ignored. 51
2.4 Example of the errors in an exposure slice map caused by badly assigned front surface points. The colour map shows the incident fluence at each voxel in photons. The exposed region (red and yellow) should be evenly exposed, but due to the shortcomings of the early surface point assignment protocol, jagged regions of voxels are mis-assigned highly different exposures, since the 'depth' is calculated from the rear, rather than the front face. Colour bar units are in photons/voxel. . . 51

2.5 Flow schematic of the first iteration of the improvements to RADDOSE. The problem is reduced from 3D to 2D by projecting the 2D Gaussian beam profile into 1D and assuming that the beam is fully immersed in the y-axis of the crystal. This uses the iteration-by-iteration assignment of dose which, as shown in figure 2.6, was inadequate. Subsequent versions skipped the "find total beam path in crystal, calculate dose for iteration", and performed a dose calculation for each voxel, as detailed in section 2.3.3. 52
2.6 Errors introduced by assigning the dose based on an estimate for a whole angular iteration at a time, and not accurately estimating the depth of the crystal. This was overcome by assigning dose on a voxel-by-voxel basis. 53
2.7 Diagram for dose calculation for a normally incident beam on a voxel. 55
2.8 Diagram for dose calculation for 45° incident beam the same voxel as figure 2.7, except rotated by 45° along the axis into the page. The half-height $d/\sqrt{2}$ is from Pythagoras, based on voxel edges of length d. 56
2.9 Dose maps for the early 2D simulations, showing the large effect on dose distribution of rotational strategies. The translations were made along the x axis, which is experimentally atypical. Nonetheless, the large effects were a major motivator for moving to a 3D model earlier. The blue box represents the crystal. 59

3.1 Flow chart illustrating the structure of the RADDOSE-3D code. 67
3.2 Structure of a RADDOSE-3D input file. The crystal is first described by a `Crystal` block (colour salmon), which remains active throughout. Next, a beam (the green block), followed by one or more wedges (blue), is described. As illustrated in figure 3.1, each wedge defines how to simulate an exposure of the previously specified beam with the crystal. Here the first two wedges will act with the first beam, and the last two with the second beam. The full input syntax is described in Appendix 1. 68
3.3 Voxel resolution testing for large (up to 15×10^6 voxels for the 250^3 μm ^{3} crystal) and smaller crystals, with 20×20 and 40×40 μm ^{2} uncollimated FWHM beams (left and right columns respectively). The top row is for a 90° rotation, the middle row is for 360°, and the bottom row is also for 360°, but with a full translation. Note that voxel resolution is quoted in voxels per micron. In all cases, the simulations converge towards their high voxel resolution values at around $1/10^{th}$ of the beam FWHM, shown by the vertical dashed line (0.25 vox/ μm for the larger beam, and 0.5 vox/ μm for the smaller beam). Horizontal lines have been added at ± 5% of the converged values for visual clarity. The metrics are described in section 3.3 72
3.4 Dose map for the input configuration described by code box 3.3. Iso-surfaces are shown at 0.05 and 5 MGy, and a slice map is shown in the plane perpendicular to the rotation axis and along the beam axis. The crystal has been rotated 90° about the y-axis, with the beam starting down the z-axis, and ending along the x-axis, as labelled in the figure. The figure was generated from the output-dosestate.csv file using Paraview (http://www.paraview.org/) 89
3.5 Dose isosurface for an evenly exposed spherical crystal completely bathed in a uniform Top-Hat beam, as modelled using the minimum input file described in code box 3.5. The dose field is, as expected for a case when the beam is larger than the crystal, even throughout the crystal volume. The figure was generated from the output-DoseState.csv file using Paraview (http://www.paraview.org/). 93

3.6 Dose isosurface map for the complicated example described by the input file in figure 3.2. The solid red isosurface represents a 60 MGy (very high dose) surface, and the dotted red surface represents the Experimental Dose limit of 30 MGy (Owen et al., 2006). This crystal was oriented at 45 to the goniometer axis (y axis in figure) in the plane normal to the beam axis. All wedges were for 360 rotations. The rightmost two wedges, with a first beam, show a (i) standard wedge, and a (ii) helical wedge. Here helical scanning was defined along the (1,1,0) vector, to allow for the orientation of the crystal in the loop. Improvements in dose contrast for this method are described in detail in Chapter 5. The two leftmost wedges, both simulated with a different beam to the first two simulations, contrast (iii) a standard strategy and (iv) an offset strategy. The offset strategy offsets the beam and rotation axes, so that they do not intersect. This leads to a lowering of the maximum dose for a given average dose. The figure was generated from the output-dosestate.csv file using Paraview (http://www.paraview.org/) 97

4.1 Example of a morphologically suitable crystal of cubic insulin in its crystallisation drop, image taken by John Bremridge. 104

4.2 False colour images of the beam profiles used for the experiments presented here. Pixel size is 5×5 μm ^{2} in all cases, and the colour scale is from 0-255 (blue-red), and is the same on all the images. Images were subsequently scaled to the full flux, as measured by the in-line calibrated diode I_1 (see figure 4.3). Beam features are due to a multiple factors such as collimation settings and mirror slope errors. The beam was focused by adjusting the upstream mirrors and narrowing the collimation slits. 105

4.3 Arrangement of diodes at ID14-4. I_1, previously calibrated in-house, was checked against I_{Garman}, and found to be in good agreement. The 300 μm collimator after I_1 is shown, but not labelled in the figure for clarity. 107

4.4 Online microscope image of well-aligned cuboid crystal 191. Thanks to the crystal morphology and mini-kappa goniometry, it is possible to accurately parametrise the shape of the crystal in RADDOSE-3D. Note the small triangular wedge of solvent on the right. 107

4.5 Alignment of images to establish the relationship between the central pixel of the PGM images, and the beam axis used to centre the crystal. 109

4.6 Plots of I_n/I_1 against DWD, with different forms of the DDM within the DWD equation. All plots were fitted with a straight line through (1,0), for datasets with $I_n/I_1 > 0.4$ (see main text), and crystal P3S3 was discarded (see section 4.3.3). Varying the form of the diffraction efficiency term in the DWD equation has a large effect for (*c*) in the high damage region. It does, however, not have a significant effect in the biologically important $I_n/I_1 > 0.4$ region, as shown by the similar values of $D_{1/2}$ derived from the linear fits shown in table 4.6. 122

4.7 Overall intensity decay data for the big beam data collection. Linear fits have been performed with crystal P3S3 excluded (see section 4.3.3 for explanation), and for points with $I_n/I_1 > 0.4$. This threshold was set in order to match the range of intensities found in the small beam data series, and in order to remain within the linear region of decay, as identified by Teng and Moffat (2000), and validated in figure 4.11b. 128

4.8 Overall intensity decay data for the medium beam data collection, with the linear fits once again only for those datasets with $I_n/I_1 > 0.4$. Of particular interest is crystal 2004, which had smaller dimensions than the other crystals present, meaning that the crystal was illuminated relatively less by the weak tails of the beam. This is manifested in an artificicially increased lifetime for ADWC and TAD, the metrics which consider averages for volume. The fact that DWD is unaffected by this change in crystal size is a powerful validation of the metric. 130
4.9 Overall intensity decay data for the small beam data collection. Total range of I_n/I_1 is smaller than for the other two cases, due to the lower flux of the beam. 132
4.10 Plots of dose against I_n/I_1, with all three beam conditions shown. This clearly illustrates how the DWD is invariant to dose distribution. Note that the misaligned crystal P3S3 was excluded. In all cases, the x axis is in MGy 133
4.11 Residual analysis for the I_n/I_1 vs DWD data with $I_n/I_1 > 0.4$. These are the data shown in figure 4.10b. The `lm(I0s - 1 ~ DWDs -1)` explicitly shows the fit command used in R: a linear fit through $(0, 1)$ in I_n/I_1 *vs* DWD. 134
4.12 $D_{1/2}$ for each individual crystal, shown with a box and whiskers plot. The bold line with a value above it is the median, the box encloses the first and third quartile, and the whiskers extend to the furthest datum within 1.5 interquartile ranges from the upper and lower interquartile respectively. The much decreased interquartile range for the DWD is apparent, and the DWD median is 12.9 MGy, coincidentally exactly the same as the $D_{1/2}$ derived from the whole dataset. Note that the scatter plots have been rounded so that points stack up rather than overlap to help with visual clarity, whereas the box and whiskers plot shows the raw data. 135
4.13 B_{rel} increase data for the big beam data collection. Linear fits have been performed with crystal P3S3 excluded (see section 4.3.3 for explanation), and for points with $I_n/I_1 > 0.4$, as was used for the analysis presented in section 4.3.3. Each crystal appears to have a linear increase in B_{rel} for the first 4-5 datasets, but there is significant scatter between crystals. The red dotted line shows the postulated general S_{AD} of 1 Å^2/MGy (Kmetko *et al.*, 2006). 139
4.14 Increase of B_{rel} for the medium beam data collection. Linear fits have been performed for points with $I_n/I_1 > 0.4$, as was used for the analysis in section 4.3.3. This set of crystals appear to cluster well for DWD and, again, 2004 is a strong outlier in the other metrics due to its small size affecting the dose distribution. The red dotted line shows the postulated general S_{AD} of 1 Å^2/MGy (Kmetko *et al.*, 2006). 141
4.15 B_{rel} increase data for the small beam data collection. Linear fits have been performed for points with $I_n/I_1 > 0.4$, as was used for the analysis in section 4.3.3. Despite the relatively linear behaviour of the individual crystals, there is not a strong correlation between the samples. The red dotted line shows the postulated general S_{AD} of 1 Å^2/MGy (Kmetko *et al.*, 2006). 143
4.16 Plots of dose against B_{rel} for all three beam conditions. Because of the higher level of intrinsic scatter when using B_{rel}, the trends are not as stark as for I_n/I_1, however there is still a reduced spread for the DWD, particularly when compared to the Max Dose. In all cases, the y axis is in Å^2, and the x axis in MGy 144

4.17 S_{AD} for each individual crystal, shown with a box and whiskers plot. The bold line with a value above it is the median, the box encloses the first and third quartile, and the whiskers extend to the furthest datum within 1.5 interquartile ranges from the upper and lower interquartile respectively. The three different beams appear to group more than for I_n/I_1, showing that B_{rel} is potentially more sensitive to dose distributions. Note that the scatter plots have been rounded so that points stack up rather than overlap to help with visual clarity, whereas the box and whiskers plot shows the raw data. 145
4.18 Real space correlation coefficients for the small beam, plotted against each of the four dose metrics. There appears to be a fairly linear decrease for different crystals, but different starting points. 149
4.19 Relative real space correlation coefficients for the big beam ($RSCC_n^{rel} = RSCC_n - RSCC_1$), plotted against each of the four dose metrics. The outlier P3S10 is not an outlier when plotted against I_n/I_1 or B_{rel}, so its behaviour is assigned to intrinsic crystal variability. 151
4.20 Relative real space correlation coefficients for the medium beam ($RSCC_n^{rel} = RSCC_n - RSCC_1$), plotted against each of the four dose metrics. 153
4.21 Relative real space correlation coefficients for the small beam ($RSCC_n^{rel} = RSCC_n - RSCC_1$), plotted against each of the four dose metrics. 155
4.22 Cartoon structure of $PepT_{St}$, showing one of the two monomers that make up the asymmetric unit, along with it's coordinated cadmium ion (bottom left). Figure taken from the PDB 4APS, resolution 3.3Å. 158
4.23 The crystal used for comparative evaluation of RADDOSE-3D based strategies. EDNA indicates the region where the benchmark EDNA dataset was collected, $D_{1/2}$ the region used to characterise the crystal, RD3D-1 where the first RADDOSE-3D optimised, 180° dataset was collected, and Offset the region where the 360° offset data collection was collected. Full details of each data collection can be found in the main text. The blue box indicates the approximate FWHM of the uncollimated beam. The rotation axis is across the width of the page. 162
4.24 False colour image of the beam used to collect data from the crystals of $PepT_{St}$. The FWHM of the beam ($\sim 40 \times 60\ \mu m^2$) was set to match the crystal dimension. Dark current was 20 intensity units, and the corrected dynamic range is 0-106. . . 162
4.25 $PepT_{st}$ data statistics and dose estimation, using the EDNA suggested strategy. . 166
4.26 Progression of the dose state of the crystal during the four exposures. The translation, based on the approximate beam size was not large enough, as only the beam collimation was taken into account, not the experimental profile, leading to the RADDOSE-3D and offset strategies starting at a non-negligible DWD. 167
4.27 Initial estimation of $D_{1/2}$ calculated online at the ESRF, which was then used to design optimised strategies. 168
4.28 $PepT_{st}$ data statistics and dose estimation for the first RADDOSE-3D strategy. . . . 169
4.29 $PepT_{st}$ data statistics and dose estimation for the offset strategy. 170

5.1 Comparison of Diffraction Efficiency (elastically scattered photons/MGy) and Dose Inefficiency (Max Dose/Absorbed energy). Dose inefficiency, as used in Zeldin *et al.* (2013a), shows the same broad trends as the more sophisticated, experimentally validated, Diffraction Efficiency, but Diffraction Efficiency suggests that there is a more significant advantage to Translational strategies, especially with smaller beams. This comes at the cost of a very large number of wedges, and is thus only considered practical with the more cuboid shaped crystals, as discussed in the text. 182
5.2 Comparison of dose distribution with and without offset. In *(c)* and *(d)*, note the larger light blue wire-frame 8 MGy isosurface in the Offset strategy, and the absence of any regions in the Offset map above the 20 MGy isosurface, shown as a red isosurface red in *(c)* 184
5.3 Offset data for crystals up to 100 μm. In these cases, a bigger offset can lead to a drop in Diffraction Efficiency, due to the crystal being unevenly exposed with relatively high damage around the edges. 188
5.4 Offset data for larger crystals, showing the general trends more clearly. In both cases, significant improvements are only achieved for rotations above $\sim 360^{\circ}$, and there are diminishing gains for offsetting the crystal beyond 1 FWHM (here 20 μm).189

6.1 The principle of PIXE: 1. A proton passes through the atom without interacting. 2. An inelastically scattered proton ejects an electron then, as another electron falls into the (lower energy) hole, a photon is emitted with an energy that is unique to the element from which the proton scatters. These photons are used to identify and quantify elements present. Adapted from Garman and Grime (2005) 195
6.2 Example of an RBS spectrum and fit for the sample of mouse cerbral spinal fluid in fig.6.9. The position of the high energy (right hand side) edge of elemental peaks indicates the species, and the depth of the peak indicates the sample thickness. Note the low energy end of the (red) fit which is trying to fit a single thickness to a sample of varying depth. 198
6.3 Example of a fitted PIXE spectrum for a sample of Niemann-Pick infected astroglia cells. Note the logarithmic count scale and simultaneous measurement of all elements heavier than neon. 198
6.4 The Y-G plot empirically relating liquid protein concentration to protein size/number of S-containing residues. Blue diamonds: measurement possible. Pink squares: measurement not possible. The green line is an empirical model currently used as a bound between possible and impossible samples. Adapted from Garman and Grime (2005) 198
6.5 Schematic of the sample chamber. The light microscope is used to approximately position the sample and the Faraday cup absorbs the protons that pass through the sample. Adapted from Garman and Grime (2005) 201
6.6 Example of a problematic Ca hot spot in a map of a population of NPC astroglial cells. The obvious high intensity spot of Ca (several orders of magnitude higher than the rest) was removed before quantification. Analysis indicated that it was probably a particle of calcium carbonate. Colour bar is in arbitrary X-ray count units.204
6.7 Slight differential precipitation in a sulphur map of a drop of mouse brain homogenate. Relatively homogenous, but evidently more concentrated at the edges. Colour bar as for figure 6.6. 205

6.8 Iron map of the same sample as in fig.6.7 showing strong differential precipitation. This sample was more inhomogeneous so average levels could be taken across the whole drop. Colour bar as for figure 6.6. 205
6.9 Cl map of Cerebro-Spinal Fluid showing the differential precipitation problems associated with mixed solutions leading to both concentration and depth variation. See fig.6.2 for RBS. Colour bar as for figure 6.6. 205
6.10 Elemental map showing absence of cross talk between two closely spaced (~200 μm separation) 0.2μl drops of C_{550} (top) and pseudoazurin (bottom). Blue: S, Red: Fe and Green: Cu. Colour bar as for figure 6.6. 208
6.11 Proof of principle print run, November 2010. 24 protein samples were printed, each six times (12 in the top 6 rows, 12 in the bottom 6 rows), making up two sets of 6 rows. The missing half columns in the top half of positions 4, 5, 10, and 11 were an obvious problem and this was subsequently addressed by changing the protocol (see text). 213
6.12 Image of print buffer array showing no missing rows. This indicated that the polypropylene film itself was not the cause of the missing rows seen in figure 6.11. The dark region at the top of the image is glue on the edge of the aluminium LT sample holder. 213
6.13 Optical image of the test array from figure 6.11 taken after data collection. The rows that are not visible in the elemental map, figure 6.14, are not present here either, indicating either complete evaporation or that the sample was never correctly dispensed (i.e. only buffer was printed). The vertical and horizontal scanning pattern of the beam path is visible where it has etched away some of the polypropylene film. 214
6.14 Elemental map of the array shown in figure 6.11. Sulphur: green, bromine: red and phosphorous: blue. Sulphur is an indicator of protein being present and bromine and phosphorous were key buffer components for many of the samples. The rectangular regions are scans showing that there was no elementally detectable sample in these regions (they show background, stronger in sulphur because there is more bremsstrahlung there). The results for the samples present are consistent with those expected given the sample buffers and protein concentrations. Columns 9-12 and 21-24 were not scanned due to beam time constraints. 215
6.15 Overlay of figure 6.13 and figure 6.14 (linearly transformed due to different geometry of PIXE detector and optical microscope). This clearly shows the additional missing samples to the left of sample 8 and to the right of sample 19. Columns 9-12 and 21-24 were not scanned due to beam time constraints. 215
6.16 Design of new aluminium sample holder. With five 8 mm $\times$ 8 mm windows, the slide design allows flexibility for multiple sample dispensing protocols. The notch guarantees the correct orientation when placed into the beamline sample stick. . . 217
6.17 Jig for assembly of sample holders. The film is kept clean by being retained on a hop (A). The first tensioning roller (B) holds one end of the film taut, the sample holder (C) is secured in the middle and the second tensioning roller (D) adds light tension to the film while the glue dries. The experimenter is then free to gently smooth wrinkles out of the film a few times during drying, ensuring a consistent result. 218
6.18 3D representations of the new sample stick design showing the mounting point where this stick forms a seal with the sample chamber flange (A), the sample holder (B) and the in-built calibration standard samples (C). Adapted from figures kindly provided by Vladimir Palitsin, University of Surrey. 220

6.19 Engineering drawing of the new sample stick assembly designed to securely hold new sample holders and eliminate thermal drift. Figure kindly provided by Vladimir Palitsin, University of Surrey. 220
6.20 Microscope image of standard salt array showing the successful printing of multiple samples onto the polypropylene film. This demonstrates that the film can be fixed flat enough onto the mount for good spatial resolution and that the surface of the film is suitable for drop adhesion. The green line shows where the data for figure 6.21 (adjacent row verification) were collected and the red line from where those for figure 6.22 were collected (adjacent column verification). 222
6.21 Elemental data from the analysis of 30 PIXE spectra collected across two drops (CaCl and CsI) of our standard salt array. The clear absence of any calcium or chlorine in the CsI drop and *vice versa* demonstrates the absence of any cross talk. Errors in peak regions (above LoDs) are within the thickness of the line – note the logarithmic y scale – and errors below the LoD are meaningless, since the signal is of the same order as the background. The dotted lines indicate the LoDs for each element. 222
6.22 Elemental data from the analysis of 30 PIXE spectra collected across two drops ($Ba(C_2H_3O_2)$ and CaCl) of our standard salt array. Although the barium acetate drop has spread a little, probably due to the effect of the acetate on the viscosity, this again shows that there is no cross talk at all above the minimum detectable limit between the drops. Errors in peak regions (above LoDs) are within the thickness of the line and errors below the LoD are meaningless since the signal is of the same order as the background. Note the logarithmic y scale. The dotted lines indicate the Minimum Detectable Limit (same as LoD) for each element. 224
6.23 Microscope image (a), sulphur map (b), and potassium map (c) for the $1\times$ glycrerol and protein concentration gradient experiment. 230
6.24 Microscope image (a), sulphur map (b), and potassium map (c) for the $3\times$ glycrerol and protein concentration gradient experiment. 231
6.25 flow-chart of MAP PIXE protocol. Different physical locations are colour coded: the user's institution in red, Oxford in blue and the Ion Beam Centre in Guildford in pale blue. 240

8.1 Schematic of **ANGLEP**. 255
8.2 Schematic of **ANGLEL**. 255

Chapter 1

Introduction to Radiation Damage

1.1 Atomic resolution structures in biology

Atomic resolution structures of biologically significant macromolecules underpin vast swathes of the advances made in the life sciences over the last 50 years. Due to the challenges inherent to collecting data at resolutions thousands of times higher than the Rayleigh criterion for visible light[1], advanced techniques must be used to acquire this information. As of June 2013[2], the Protein Data Bank (PDB) contained 91,550 macromolecular structures, of which 88.1% have been solved by Macromolecular Crystallography (MX), 10.9% by NMR, 0.7% by electron microscopy, and the remaining by hybrid and other techniques. These figures clearly show that MX is the dominant method for obtaining atomic resolution data from macromolecules, yet it is not without its major challenges, the first, and largest, of which is getting macromolecules to actually crystallise. Crystallisation is achieved by slowly increasing the concentration of a protein solution until the solution is supersaturated under just the right combination of chemical and environmental conditions.

There is no generalised protocol for predicting under what conditions a protein will crystallise, or indeed if it will crystallise at all, and in many cases, the conditions for obtaining crystals of important biological macromolecules remain illusive. This thesis is concerned with optimising the way we handle these hard won crystals, in particular with respect to the radiation damage caused by incident X-rays during the subsequent diffraction experiment.

[1] For visible light of wavelength 400 nm, this is approximately 200 nm. A typical resolution achieved by protein crystallography is 2 Å. It is not possible to simply reduce the wavelength of the imaging source, since there is no method to effectively focus high energy electromagnetic radiation.

[2] http://www.rcsb.org/pdb/statistics/holdings.do

1.1.1 Macromolecular crystals

A crystal is made up of a large number of identical unit cells, which are related to each other by translational symmetry: for a perfect crystal, looking one (or an integer number) of unit cells vectors away from any origin, the crystal will appear the same as at the origin. This unit cell is defined by the length of the three unit vectors (**a, b, c**) and the angles between these unit vectors (α, β, γ) that bound it. Beyond this minimum definition, there exist several more levels of symmetry:

1. *Bravais Lattices.* All the possible arrays of points created by applying the translations defined by (**a, b, c**) and (α, β, γ) can be described using only 14 basic Bravais lattices (Bravais, 1850). These lattices define the position of all spots in reciprocal space, but not their intensities.

2. *Point Group.* A further layer of symmetry, which can be applied about the lattice points defined by the Bravais lattice, is the point group symmetry. In a general sense, these are rotation and reflection operations about an invariant centre. In three dimensions and under the constraint that translational symmetry must be observed, there are 32 unique sets of 2,3,4, and 6 fold rotations and mirror planes. Because biological molecules are chiral, mirror symmetry is not observed in MX. This leaves the 11 enantiomorphic point groups made up of multiple rotations about one or more axes: 1, 2, 222, 4, 422, 3, 32, 6, 622, 23, 432.

3. *Space Group*. It is further possible to extend these point groups with screw axes, where a rotation is combined with a partial translation within the unit cell. Together with the point groups and Bravais lattices, this leads to the 65 allowed enantiomorphic space groups in which macromolecules can theoretically crystallise.

The typical interaction energies between protein molecules in a unit cell are around 63×10^{-21} J, comparable to those in small molecule crystals (Drenth, 2012). However, macromolecular unit cells are comparatively large (~100 Å), and contain around 50% by volume of disordered solvent. The interaction energy is roughly equivalent to ten hydrogen bonds, four salt bridges, or a 400 Å^2 buried hydrophobic surface (Drenth, 2012). As a consequence of these relatively weak interactions relative to the molecule size, protein crystals are much more sensitive than small molecule crystals to distorting forces, and small monomer rotations and side-chain disorder are likely.

1.1.2 From crystals to electron density maps

Because of the repeating unit cells that define a crystal, the interference pattern from X-rays that have been elastically scattered in a crystal only has non-negligible values under the Laue conditions:

$$\mathbf{a}.\mathbf{S} = h,\ \mathbf{b}.\mathbf{S} = k,\ \mathbf{c}.\mathbf{S} = l$$

where **a, b, c** are the unit cell basis vectors, *h, k, l* are integers, and **S** is the scattering vector, which points in a direction intersecting the incident and scattered wave and has magnitude $2sin(\theta)/\lambda$. The location of these peaks on the detector informs on the Bravais lattice, and their intensities are

those of the complex structure factors of the whole unit cell for the associated scattering vectors. Mathematically, this interference pattern is the complex Fourier transform of the electron density within the crystal. This three dimensional complex scalar field is commonly known as reciprocal space. MX allows us to project the magnitude - but not the complex phases - of thin wedges of reciprocal space onto a two dimensional detector. In modern experimental set-ups, this is achieved by rotating the crystal, usually about an axis perpendicular to the beam, while illuminating it with a source of X-rays that has a coherence length of at least several unit cells at the sample position. Data are then collected on a detector placed downstream of the crystal. This has traditionally been done in small wedges of order 1°, but modern detectors with extremely fast readout times and negligible readout noise have allowed data to be collected continuously, without the need to open and close the X-ray shutter, enabling much finer sampling of reciprocal space. Once a sufficient total volume of reciprocal space has been sampled, the size of which is dictated by the symmetry of the crystal itself (Dauter, 2005), the 2D detector images are then processed to reconstruct a picture of reciprocal space.

First, a program such as MOSFLM (Leslie and Powell, 2007) or XDS (Kabsch, 2010) is used to find coordinates for spots in the images, find the unit cell, assign Miller indices ((*h, k, l*) indeces) to each peak, and integrate the total intensity of each reciprocal space point. Since some reciprocal lattice points will only partially fall within the wedge of Fourier space sampled by a given image, it is important to be able to determine what fraction of the point it recorded on the image: its partiality. This is performed during a post-refinement phase (Rossmann, 1979), where the phys-

ical experimental parameters are refined in order to allow accurate prediction of partiality based on crystal orientation and rotation values. Each image is collected with the crystal in a different orientation and after a different level of exposure to X-ray radiation, leading to different relative levels of intensity at the detector. As a consequence, the intensities of the diffraction patterns must be scaled together. This is achieved by using the prior knowledge that certain symmetry related reflections should have the same intensity, and trying to match these using a physical model of the experiment. This process is performed by programs such as XSCALE in the XDS suite (Kabsch, 2010), or SCALA (Evans, 2006).

After correcting this list of reciprocal lattice points for negative values of intensity and testing it for undesirable crystal phenomena such as twinning[3], these intensities are converted to structure factor amplitudes. Phases are assigned to these structure factors using a variety of methods, dependent on the particular details of the sample: if the protein has a high sequence identity, typically at least 30%, or is thought to be a member of a structurally homologous family, molecular replacement is often used, implemented in programs such as Phaser (McCoy *et al.*, 2007). This technique takes a single or ensemble of structural homologues, and performs a rotation and translational search to match the observed structure factor amplitudes to those calculated from the homologue. It has the advantage of not requiring any additional experimental data beyond a native data set. If this is not possible, then phases must be experimentally determined. The earliest method of experimental phasing involved isomorphically inserting an ordered, highly scattering

[3] Where two or more orientations of the unit cells exit in a symmetry-related conformation such that the diffraction patterns for the twins are superimposed. The superimposed diffraction patterns must then be deconvoluted using a program such as `detwin` from the CCP4 suite. (Winn *et al.*, 2011).

heavy atom into the crystal, and solving the heavy atom substructure directly, then using these phases to calculate the weaker remaining ones (Crick and Magdoff, 1956). Alternatively, the weak anomalous signal which is the cause of differences in intensity between Friedel pairs of reflections[4] can be used to obtain phases. Techniques such as Multiwavelength Anomalous Dispersion (Kahn *et al.*, 1985; Hendrickson and Ogata, 1997), or Single Wavelength Anomalous Dispersion (Bijvoet, 1954; Wang, 1985) can be used here. This set of structure factors can be Fourier transformed to yield an electron density map, into which the protein sequence, with its known atomic composition, is built. The model is then refined using chemical and physical parameters such as bond lengths and torsion angles in either real or reciprocal space, so that the calculated structure factor amplitudes from the model are in the best possible agreement with the observed ones. Programs for Refinement include REFMAC (Murshudov *et al.*, 2011) or PHENIX.refine (Adams *et al.*, 2010).

1.1.3 Limitations of crystallography

Although MX has seen huge methodological advances over recent decades, in particular in the automation of effective data collection and processing, it still has several fundamental limitations. As a corollary to the ubiquitous challenge of growing suitably diffracting crystals, our known database of crystal structures is highly biased towards proteins that are able to form crystals, in conformations that will crystallise! This presents a particular challenge for protein complexes and membrane proteins, which are often highly unstable in the chemical environments most likely to yield crystals, and yet represent a vitally important family of proteins for biomedical applications.

[4] $(h,\ k,\ l)$ and $(-h,\ -k,\ -l)$.

Even with the best crystals, the fact that contributions are made to a typical diffraction image by around a thousand billion molecules[5], means that MX involves intrinsic spatial and temporal averaging. The main consequence of this is that less-ordered loops and side-chains are often not distinguishable in the final electron density maps. Efforts to overcome this through the use of ensemble refinement, where a superposition of final structures is used, have yielded some success (see for example Burnley *et al.* (2012)), especially for high resolution structures, but crystallography is still a long way from being able to offer dynamic information in the way that techniques such as NMR can.

1.2 Radiation damage as a bottleneck of successful structure determination

Considering the significant challenges to be overcome in order to obtain a well diffracting crystal, it is important that methods are optimised to minimise the number of cases when a structure cannot be solved from a dataset. Unpublished figures from beamline 8.3.1 at the Advanced Light Source in Berkeley, CA, USA, kindly provided by James Holton, reveal that for the ~ 2500 datasets collected each year, there are only ~ 100 submissions to the PDB. This shows that there is still plenty of room for improvement in our structure solving pipeline. The three dominant causes of failure, as identified by Holton, are:

[5] Assuming 1 monomer per unit cell, a typical unit cell size of 10 nm, and a typical crystal size of 100μm, this gives $(100 \times 10^{-6})^3 \div (10 \times 10^{-9})^3 = 10^{12}$ protein molecules per crystal.

1. *Overlaps.* When the oscillation range chosen for the data collection is too large, multiple reciprocal space points will be projected on top of each other in the detector image, making it impossible to correctly assign intensities to individual *h,k,l*s. This is a human error in strategy selection, and the problem has been solved for some time: see, for example, pages 94-95 of Arndt and Wonacott (1977). The worst case condition[6], derived with the longest unit cell dimension a, along the beam axis, is $\delta\phi \leq 180d/\pi a - \zeta$, where $\delta\phi$ is the maximum oscillation range for an image in degrees, d is the maximum resolution of the diffraction in Å, and ζ the angular size of the spot in degrees in reciprocal space.

2. *Radiation Damage.* A fundamental limitation of MX is that X-rays will damage proteins. This can compromise the data to the extent that a biologically reliable structure can not be determined. Excessive damage caused by the absorption of X-rays in the sample can lead to an inability to obtain interpretable electron density maps. The causes and symptoms of this undesirable phenomena are discussed in detail in the remaining sections of this chapter.

3. *Signal to noise.* Too low a signal to noise ratio can also yield uninterpretable electron density maps. Assuming that radiation damage is not the limiting factor, a better signal to noise ratio can be achieved by using a higher X-ray fluence at the crystal, or by using more crystals. In either case, if the limitations caused by radiation damage are well handled, signal to noise is simply a problem of selecting an appropriate strategy.

[6] As presented in Dauter and Wilson (2012)

1.3 Physical origin of radiation damage

For a typical 100 μm thick crystal, only about 2% of a 12.4 keV incident beam will interact at all with the crystal (Garman, 2010). Out of this 2%, only 8% — 0.16% of the incident beam — will be elastically scattered, contributing to diffraction. The remaining 92% of the interacting photons will contribute to radiation damage. This section explores the various physical processes which contribute to these figures.

1.3.1 Primary X-ray–crystal interactions

At the energies relevant to MX, three physical processes occur with non-negligible probabilities: Rayleigh scattering, Compton scattering, and photoelectric absorption. **Rayleigh scattering** is the elastic scattering of incident photons by the electron cloud. This type of interaction gives rise to the interference effects which are observed as the diffraction pattern on the detector. **Compton scattering** is when the photon is *in*elastically scattered from the electron cloud: the incident photon transfers some of its energy to the electron from which it is scattered (the recoil electron), and is then scattered at an energy lower than that with which it was incident. The final type of interaction, **photoelectric absorption** (see figure 1.1), is the dominant cause of damage in MX. Here, the energy from the incident photon is fully absorbed by the atom, and an electron is ejected from an inner shell of the atom, leaving the atom in an excited state. An outer shell electron will then drop down into the hole left by the ejected photoelectron *via* one of two decay paths: either with the associated emission of an Auger electron, whereby another outer shell electron is also

ejected from the atom, or by fluorescent decay, which is when a photon is emitted with energy equal to the difference between the original and final energy shells of the electron that drops down to fill the hole left by photoelectric emission. Auger electrons are the most likely decay path for the light atoms prevalent in proteins; however for heavy atoms, fluorescent decay can also be an important route and must be taken into account. For example, the probability of fluorescent decay from a *K*-shell excited iron atom is 30%. This probablilty increases with the atomic number of the atom.

The probabilities of each of these phenomena depend on the atomic species, and is energy dependent, as shown in figure 1.2. Since atomic composition is often dictated by which crystallisation conditions give crystals, this is a difficult parameter to explore. Table 1.1 shows the relative atomic contributions to the absorption coefficient for an insulin crystal, illustrating how most of the absorption occurs in the oxygen (from the water in the solvent channels), and in any heavy atoms present in the crystal. Here there are two zinc atoms associated with each insulin molecule, and these contribute significantly to the absorption. The heavy zinc atoms in this example are ordered and are part of the diffracting structure. They are necessary for the protein, but in other cases, it is worthwhile trying to avoid heavy atoms in crystallisation conditions wherever possible; see Holton (2009) for a list of heavy atom concentrations required to double the probability that an X-ray will be absorbed in a typical protein crystal, calculated with RADDOSE v2 (Paithankar *et al.*, 2009).

Because cross-sections for all three main processes are energy dependent, it is desirable to con-

Element	Contribution to absorption coefficient
Oxygen	42.0%
Zinc	33.7%
Sulphur	9.9%
Carbon	5.7%
Phosphorous	4.3%
Nitrogen	3.2%
Sodium	1.2%

Table 1.1: Relative contributions to total absorbed energy by element for cubic insulin crystallised in 425 mM sodium phosphate, calculated for 12.6 keV X-rays using RADDOSE v3 (Paithankar and Garman, 2010). (Calculated solvent content: 65.1 %).

sider if there is an optimum energy at which to collect data in MX. Although the elastic cross-section increases gradually with energy, the inelastic, damage-causing, Compton cross section also increases. Phasing considerations and the elemental composition of a crystal will affect the theoretical optimal energy for data collection, and there is no clear consensus within the field for a 'best-practice' with respect to collecting data at very high or low energies (Paithankar and Garman, 2010). Furthermore, despite the promise of potential advantages with high energy data collection (Cowan and Nave, 2008), most MX beamlines are optimised for the ~10 keV energy range (principally for phasing reasons), and so significant practical hurdles remain.

1.3.2 Electron cascades, secondary damage and temperature dependence

The damage caused by these initial absorption events of the incident beam in the crystal are referred to as primary damage. However, after these have taken place, we are left with highly energetic electrons still in the crystal: photoelectrons, Auger electrons, and recoil electrons. These particles will progress through the crystal, losing energy as they interact with atoms, knocking off more electrons, and forming energetic species. This secondary damage is the dominant driver of

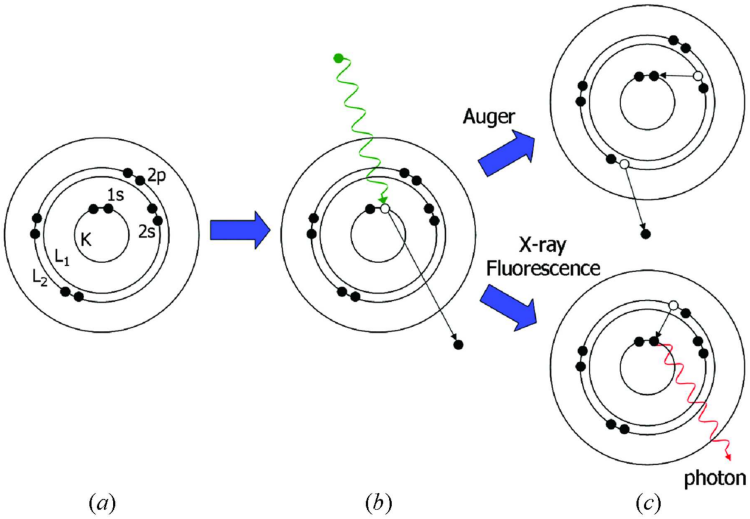


Figure 1.1: Schematic of decay paths after photoelectric absorption, from Paithankar *et al.* (2009). An atom *(a)* absorbs an incident photon, leading to the ejection of an inner shell electron (*b*). The excited atom can then decay emitting either an Auger electron, or a fluorescent photon (*c*). The relative probability of the two escape routes depends on the atomic species, with Auger electron emission being much more likely for the light elements that make up the bulk of biological samples.

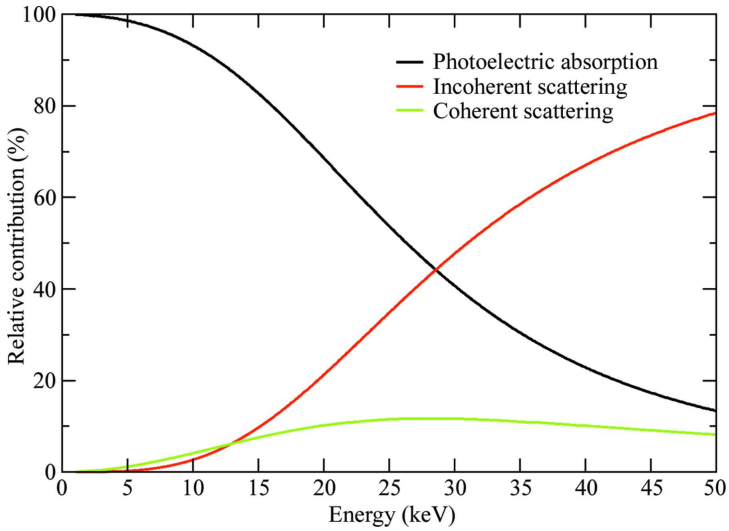


Figure 1.2: Energy dependence of the attenuation cross sections for the three processes relevant to MX for a crystal of chicken egg-white lysozyme from Paithankar and Garman (2010). Photoelectric absorption dominates, and the Compton effect only becomes significant at higher energies.

damage in MX experiments. The specific mechanisms that cause damage are highly temperature dependent, with a change in behaviour associated with the glass transition at around 180 K (Weik *et al.*, 2001):

1. *180 K up to Room Temperature (RT)* In this regime, there is a great deal of complex radiation chemistry that can occur, and we observe an extremely wide range of sensitivities to damage (Ferraz Leal *et al.*, 2013; Warkentin *et al.*, 2013). This variability is principally attributed to the diffusion of reactive species created in the buffer, mother liquor, and protein (Allan *et al.*, 2013).

2. *Below 180 K & Cryo-Crystallography* Below ~180 K, the diffusion of many larger reactive species appears to be frozen out, as suggested based on experimental observations of OH radicals using UV-vis microspectrophotometry by Owen *et al.* (2012). Because of this, cryo-cooled crystals are typically ~70 times more radiation resistant than RT crystals (Nave and Garman, 2005). Due to the simplified radiation chemistry, cryo-cooled protein crystals are often considered as an amorphous glass, with a more consistent decay behaviour as they are exposed to an X-ray beam compared with crystals at room temperature.

Since most of the damage occurs away from the initial interaction site, and diffraction originates at the interaction site, it is interesting to consider if this can be used to our advantage to reduce observed damage. In particular, can the total amount of energy deposited in the crystal be reduced by using smaller crystals, where a significant proportion of the photoelectrons may escape from the crystal, therefore not contributing to damage? Computational studies (Nave and Hill, 2005;

Cowan and Nave, 2008) examining photoelectron escape from small samples at varying energies have predicted large improvements in dose lifetime for small samples by increasing the energy[7]. Even at a typical MX energy of 12.5 keV, figure 3 of Cowan and Nave (2008) shows a predicted ~2 fold improvement in the dose to diffraction ratio for a 2.5μm crystal *vs* a 20μm crystal, suggesting a potential advantage of micro-beam MX collection from small crystals. Experimental validation of the spreading of damage due to photoelectron path lengths came in 2011 when Sanishvili *et al.* (2011) used a 1μm, 18.5 keV X-ray beam to damage a region of crystal. The damage state of the surrounding crystal volume was then probed stepwise both along and perpendicular to the polarization axis of the beam, leading them to observe decreasing damage, originating from the irradiated region, with a range of ~4 μm. The anisotropy of the damage predicted by Monte Carlo simulations – more damage along the horizontal polarization vector – was also observed.

1.3.3 Dose, and modelling it using RADDOSE v1-3

The energy initially absorbed by a crystal volume through the primary interactions lies at the start of all analysis of radiation damage. Dose, the measure of absorbed energy per unit mass[8], has been used as the appropriate independent variable against which to plot various measures of damage ever since the first RT study of radiation damage in MX by Blake and Phillips (1962). Dose can be

[7] e.g. A 7-fold improvement in going from 12 keV to 25keV for a 2.5 μm crystal with a 25 Å unit cell
[8] Dose = Joules/kg, SI unit: gray, Gy

calculated for a small, evenly exposed volume of crystal using the equation:

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

where $E_{incident}$ is the incident energy on the volume, M_{vol} the mass of the irradiated volume and $1 - e^{-\mu_{abs}z}$ is the fraction of the incident beam absorbed through the volume of depth z. To calculate this, the following parameters are required: $E_{incident}$, M_{vol}, μ_{abs}, and z.

1. $E_{incident}$ is the product of the beam flux, exposure time, and photon energy. The latter two quantities are known to a high degree of accuracy, and total beam flux can be measured using an ion chamber or a calibrated photo-diode (Owen *et al.*, 2009). However, the profile of many beams is not well described (see, for example, Krojer and von Delft (2011)), and this can lead to inaccuracies if the volume being considered is not under the 'evenly-exposed' regime.

2. M_{vol} is the product of the irradiated crystal volume and its density (ρ_{cryst}). The density can be computed from the unit cell composition, which is itself a function of the macromolecule, the space group, and the buffer contents. The crystal volume requires knowledge of the crystal shape, which can be estimated using an on-line microscope, but can be difficult to quantify accurately if the refractive index of the crystal is similar to that of the crystallisation medium surrounding it (e.g. lipidic cubic phase crystallisation). This can be overcome by using online tomography (Brockhauser *et al.*, 2008), or potentially with online

confocal microscopy (Khan *et al.*, 2012a), but these techniques are not yet routine in MX.

3. μ_{abs}, the absorption coefficient. If the atomic composition of the sample is known, this can be calculated by summing the interaction cross-sections for each of the constituent atoms in the sample: $\mu_{abs} = (1/V_{cell}) \sum \sigma$, where $\sigma = \sigma_{photoelectric} + \sigma_{Compton}$, and V_{cell} is the volume of the unit cell.

4. z, the thickness of the crystal as seen by the X-ray beam. This is found using identical methods to those used to find the crystal volume.

In order to simplify the routine calculation of dose for biological applications, RADDOSE was developed in 2004 (Murray *et al.*, 2004). RADDOSE calculates the absorption coefficient from the unit cell size, the number of monomers, the monomer primary sequence length, the number of heavy atoms in the monomer (sulphur from cysteines and methionines, or selenium from seleno-methionine substitutions and/or known bound metal ions), the millimolar concentration of heavy atoms in the solvent, and, optionally, the solvent content. If no solvent content is provided, the program calculates this based on volume considerations and fills any empty space with water. The first version of RADDOSE only took into account photoelectric absorption, which is by far the dominant effect at typical incident X-ray energies. The second version of RADDOSE (Paithankar *et al.*, 2009) calculated both the fluorescent emission probability, and the probability that these fluorescent photons might escape the crystal rather than be reabsorbed, making the simplifying assumption that all the fluorescent photons were produced in the centre of the crystal. The energy loss in the crystal due to the Compton effect (inelastic X-ray scattering) was included in version

3 of RADDOSE with the appropriate angular distribution and self-absorption treatment (Paithankar and Garman, 2010). This version was made available to groups working on high energy macromolecular crystallography on request. RADDOSE v2 has become a widely used tool for estimating the absorbed dose in MX experiments, due to its convenient use of biologically relevant parameters as inputs, and its detailed consideration of the relevant physics. An alternative approach (Warkentin *et al.*, 2012), yielding results within 5% of those output by RADDOSE, is to manually compile the crystal composition, and use mass-energy absorption coefficients from the NIST tables (http://www.nist.gov/pml/data/xray_gammaray.cfm).

These one-dimensional treatments are correct for a crystal that is completely immersed in a beam with a flat, top-hat profile. In the case of a Gaussian beam, RADDOSE calculates the peak beam intensity, and uses this to provide a maximum dose; a worst case scenario for damage. This approximation has obvious limitations in an era when many beams are smaller than the crystals being irradiated[9], and when the beams can have highly featured profiles (Krojer and von Delft, 2011). One solution to this problem is to use a sacrificial crystal to experimentally characterise radiation damage sensitivity before determining an optimal data collection strategy. This approach is one of the options implemented in the program BEST (Bourenkov and Popov, 2010). Experimental characterisation has the obvious disadvantage of requiring a spare crystal of the desired type to be available, and can add a time burden since a data collection and burn series must be performed before the desired data collection can proceed.

[9] Allowing the best diffracting regions from large crystals to be selected for data collection.

1.4 Manifestation of damage: Global and Specific damage

Thus far the basic microscopic physical processes going on within the crystal have been described: the initial, well-understood, interactions, and how the more complex radiation chemistry arising from the electron cascade that follows leads to temperature dependence of crystal sensitivity. These processes are, however, not directly observed in an MX experiment. We observe a diffraction image on the detector, and then process a set of these to produce a spatially and temporally averaged model of the macromolecule of interest. This section concerns itself with how the damage processes described above manifest themselves in the diffraction data, classifying damage by how it is observed. Global damage is observed in reciprocal space: in the diffraction image, and in the statistics of a data set. Specific damage refers to structural changes observed in the final electron density maps as a consequence of damage.

1.5 Global damage

The main consequences of global damage on a MX dataset are:

1. A loss of intensity throughout the diffraction pattern, in particular at high resolution. This can be seen by comparing the three images in figure 1.3. This is widely used to quantify the effect of radiation damage: for a series of identical, progressive damage datasets, the metric I_n/I_1 relates the average intensity over all resolution shells of the n'th dataset to the 1st dataset.

2. An increase in mosaicity, which is so significant across these three images that we can also directly see the spots become larger and less well defined in the extreme dose image 1.3c.

3. An increase in unit cell volume. This is not visible by eye from figure 1.3.

4. Increase in R factors. As the structure becomes more damaged, it is intuitive that metrics measuring its fidelity to an undamaged model will get worse, so R factors (R_{free}, R_{cryst}, R_{pim}, R_{sim}) increase.

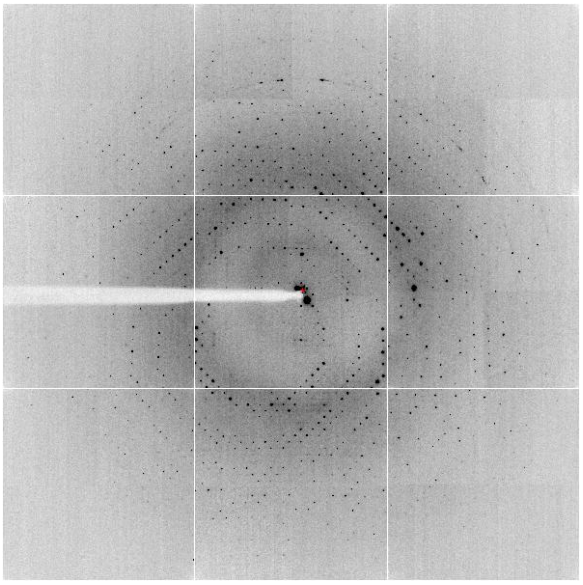
5. An increase in B factor. The Wilson B-factor increases with damage, as the average atomic positions defining the structure become less well defined.

6. Loss of isomorphism. Individual monomers in the unit cell can rotate, or move slightly as damage progresses. This is a particularly nefarious, since, along with the increase in unit cell volume, it can overwhelm the small changes in intensity that must be measured for anomalous phasing techniques.

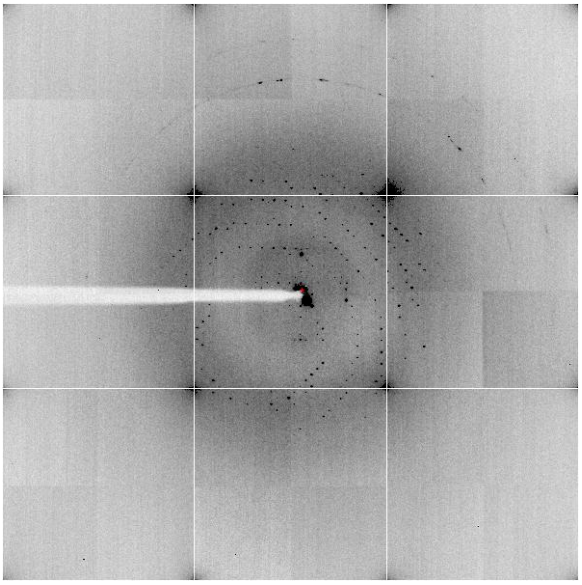
Because the progression of radiation damage at RT and 100 K is so different, they are treated separately below.

1.5.1 Room temperature

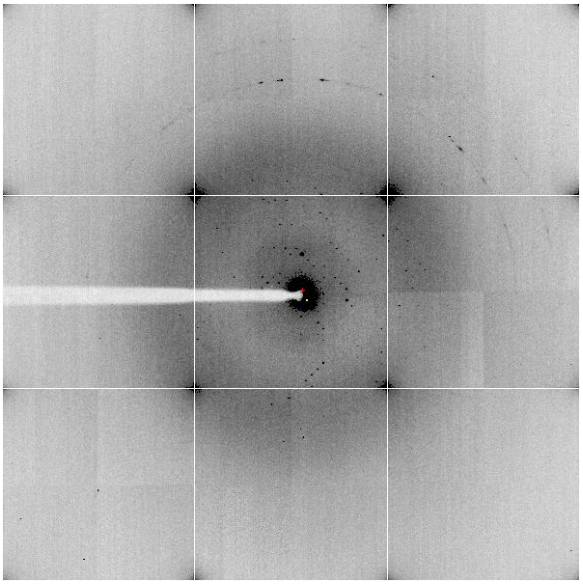
The first, and most easily observed manifestation of radiation damage is the loss of diffraction intensity, particularly at high resolution. Blake and Phillips (1962) observed the loss of intensity at RT for several individual diffraction spots from sperm-whale myoglobin of increasing resolution in



(a) Low dose (~kGy) dataset.



(b) High dose (30 MGy) dataset.



(c) Extreme dose (60 MGy) dataset.

Figure 1.3: Examples of raw images collected with the same exposure parameters from a crystal of cubic insulin at 100K after various levels of radiation damage. Note how the intensity drops in the high resolution bins, and how the remaining spots increase in size due to mosaicity increases.

the 010 zone[10] at seven time points of increasing dose. They observed that damage was proportional to the absorbed dose. Intuitively, one would expect the resolution-dependent loss of decay to follow the form for random thermalisation:

$$I = I_0\, e^{-\frac{Bsin^2\theta}{\lambda^2}}$$

where I is the original of a spot, I_0 the undamaged intensity of a spot, B the level of disorder, and $sin^2\theta/\lambda^2$ is proportional to the resolution. This is, however, not the form that was observed. Experimentally, the intensity drop-off with increasing scattering vector flattens out at high θ. This led them to suggest a 3-state model of radiation damage:

$$\begin{array}{ccc} Undamaged\,(A_1) & \xrightarrow{k_3} & Amorphous\,(1-[A_1+A_2]) = (A_3) \\ {\scriptstyle k_1}\searrow & & \nearrow{\scriptstyle k_2} \\ & Disordered(A_2) & \end{array} \qquad (6)$$

Figure 1.4: Three state Blake and Phillips model. Figure from Southworth-Davies *et al.* (2007). The model adds an alternate route to the amorphous state (A_3) *via* a disordered state (A_2) which contributes to diffraction only at low angles.

where there is a contribution to the total diffraction by an undamaged part, A_1 and a disorded part, A_2, which only diffracts at low angles (low resolution). A third, amorphous, part A_3 does not contribute to diffraction at all. This yields the following form for the resolution-dependent intensity decay:

$$\frac{I}{I_0} = A_1(t) + A_2(t)e^{-\frac{Bsin^2\theta}{\lambda^2}}$$

[10] The centrosymetric projections of the structure down the b axis.

with A_1 and A_2 indicating the undamaged and disordered parts, as shown in figure 1.4. This was subsequently validated for a larger RT data-set by Hendrickson *et al.* (1973), with $k_2 = k_3$ in figure 1.4. Fletterick *et al.* (1976) suggested a $k_3 = 0$ model for glycogen phosphorylase a crystals based on a linear decrease of intensity with time at all resolutions, with a higher rate of decrease at higher resolutions, according to the scheme:

$$A_1 \overset{k_1}{\rightarrow} A_2 \overset{k_2}{\rightarrow} A_3$$

Hendrickson (1976) subsequently performed an analysis of the original Blake and Phillips myoglobin data for all possible reaction paths in figure 1.4, concluding that all models fitted well at low-moderate doses (down to ~0.75 I_n/I_1), that only the Fletterick *et al.* (1976) $k_3 = 0$ model fitted for high doses (to ~0.5 I_n/I_1), and that none of the models are appropriate for very high doses (down to ~0.3 I_n/I_1). The Fletterick *et al.* (1976) model was later extended by Sygusch and Allaire (1988), who added a dose-dependent, still-diffracting state A_1' to the scheme:

$$A_1 \overset{k_0}{\rightarrow} A_1' \overset{k_1}{\rightarrow} A_2 \overset{k_2}{\rightarrow} A_3$$

This model fitted the data up to very high doses, and was thought to provide a good description of RT radiation damage progression. Although these models can provide a good fit to the data, it is worth noting that there is, so far, no microscopic physical model of the processes leading to global damage effects in MX.

Thanks to the possibility of in-plate MX (Axford *et al.*, 2012) there has been a recent resurgence in interest in RT MX. Using modern high throughput data collection and processing techniques to generate large data sets for analysis, Ferraz Leal *et al.* (2012) have developed a sample-dependent two-parameter model for the intensity decay of RT crystals, with a resolution independent exponential decay, $e^{-\gamma^2 D^2}$, and a resolution dependent decay term $e^{-\frac{1}{2}(B_0+\beta D)h^2}$, where the $(B_0 + \beta D)$ term models a linear increase in B factor with dose, and h is the distance to the origin in reciprocal space, proportional to the resolution. Another recent paper (Warkentin and Thorne, 2010) has also validated the Hendrickson (1976) model, and additionally found a linear increase in scaling B factors (see section 1.5.2) with dose for thaumatin crystals at RT. In all cases, the predictive power of these models relies on accurately quantifying the dose.

1.5.2 Cryo-temperatures

With the advent of cryo-crystallography (Haas and Rossmann, 1970; Hope, 1988; Garman and Schneider, 1997), a reduction in radiation-dose sensitivity of almost two orders of magnitude was observed, making this the method of choice for most modern MX experiments. Damage effects at 100K were first systematically observed in the 1990's using a Laue beam on a first generation synchrotron to achieve sufficient fluence (photons/unit area) (Gonzalez *et al.*, 1992; Gonzalez and Nave, 1994). Gonzalez and Nave (1994) damaged a lysozyme crystal held at 100 K, and observed a decrease in the correlation coefficient[11] with increasing dose. At 100 K, significant radiation damage occurs at doses of the order of 10^6-10^7 Gy, many times higher than the 10^5 Gy

[11] $[\sum(F_0 - \langle F_0\rangle)(F_n - \langle F_n\rangle)]/[\sum(F_0 - \langle F_0\rangle)^2 \sum(F_n - \langle F_n\rangle)^2]^{\frac{1}{2}}$

range typical of significant damage being observed during room temperature MX. As described in section 1.3.2, this is primarily due to the localisation of reactive chemical species, preventing them from diffusing and causing higher levels of 'knock-on' damage. As a consequence of this simplified chemistry, protein crystals have a much more homogeneous behaviour at 100 K than at RT. A major benefit of this is that it has been possible to find certain metrics that behave predictably with dose, and thus develop generalised guidelines for the dose tolerance of crystals in cryo-MX. However, many earlier research articles use either fluence ($photons/mm^2$), data set number, or time as a proxy for dose, making it difficult to directly compare experiments. Some of the key publications are summarised in table 1.2.

Much of the variation in results that can be seen in table 1.2 can be attributed to the small number of samples and absence of replicates for crystals under identical conditions. This was due to the significant experimental challenges and time-consuming nature of performing these experiments on 2nd generation synchrotrons with slow readout detectors. The first major dose calibrated systematic study of global damage phenomena by Owen *et al.* (2006), using four holo-ferritin and three apoferritin crystals, established a linear loss of total summed intensity with dose. These proteins were chosen because apoferritin represents the typical absorption properties for a protein crystal, and holoferritin has an exceptionally large absorption cross-section due to the ~1600 iron atoms for every 48 kD of protein (24 monomers enclosing the amorphous Fe core). This leads to holoferritin crystals absorbing twice as many photons per unit fluence as apoferritin. The seven crystals were analysed for I_n/I_1, Wilson B, R_{meas}, unit cell expansion, and mosaicity.

I_n/I_1 was modelled with a linear decay proportional to dose. The average dose required to reduce the intensity of the integrated diffraction pattern to 0.5 of its original value (the $D_{1/2}$) was 43 MGy, with a standard deviation of 3 Mgy. These figures were calculated from 7 crystals. The overall error is stated to be dominated by the dose estimation, despite the use of RADDOSE v1 and a manual correction for fluorescent photon escape. The 7 crystals showed a remarkable agreement, given the vast differences in decay rate with respect to fluence. Wilson B, R_{meas} and unit cell size increased for all crystals, but in all cases was not systematic enough to be used as a damage metric, as illustrated in figure 1.5. The mosaicity did not show a consistent behaviour at all between crystals, increasing in some cases, and decreasing in others.

By analysing the real-space electron density maps from their data sets, Owen *et al.* (2006) established a limit of biological fidelity at $0.7I/I_0$, corresponding to 30 MGy. This is the well-known 'experimental dose limit' which is often quoted as a cryo-MX radiation damage guideline. This is 50% higher than the estimate of 20 MGy made by Henderson (1990), which was derived based on the observation of $5e^-/\text{Å}^2$ to reach half-intensity in 2D electron diffraction.

Around the same time, Kmetko *et al.* (2006) independently performed a similar analysis, this time using Hen Egg White Lysozyme (HEWL) crystals soaked in a variety of heavy atoms to achieve a range of absorption coefficients, as well as catalase, thaumatin and apoferritin crystals to compare crystal species. They proposed the use of the relative B factor, B_{rel}, which is the difference in scaling B factor when scaling together successively damaged datasets. They found a good linear fit, with a similar gradient, for all samples, and the authors suggested the rate of change of B_{rel} be

Reference	Dose Calculation	Intensity Decay	Wilson B factor	Unit Cell Expansion	Mosaicity	R factors	Non-Isomorphism	Comments
Burmeister (2000)	Fluence as a proxy	Single exponential: $I/I_0 \propto e^{D/D_0}$, where D_0 is a constant.	Exponential decay. Assumed to be the cause of intensity loss.	Linear		R_{free} and R_{work} get worse, but no analysis.	Uniform expansion of the crystal	Single crystal
Teng and Moffat (2000)	Calculated to give gray	Linear decrease to 10MGy in all resolution shells, to higher dose at low resolution	Linear increase to 10MGy	Linear increase to 10 MGy		R_{merge} increases		Single Crystal. Also observed linear decrease in number of reflections $I > 20\sigma$
Ravelli and McSweeney (2000)	Calculated to give gray	No analysis	Increases, but no clear trend between crystals	Increases linearly, but different gradients for different crystals	Increases, but no clear trend between crystals	R_{merge} increases linearly	Refined models reveal rotation and translation relative to 1st dataset	4 crystals, 3 different proteins
Sliz *et al.* (2003)	Fluence as a proxy	Resolution dependent decays using the Hendrickson (1976) model (4 parameters)	B factor	Unit Cell Expansion	Increases, no data shown	R_d, the R-factor between subsequent datasets, increases exponentially at high resolution. No change at low resolution. Different for different crystals		3 crystal species

Table 1.2: Summary of selected radiation damage research at 100 K. Although most research agrees on the trends listed in section 1.5, there is no consensus on the specific rates and form of these phenomena.

used as a coefficient of sensitivity ($S_{AD} = \frac{\Delta B_{rel}}{8\pi^2 \Delta D}$), found to be approximately 0.014 Å^{2}MGy^{-1}[12]. These two metrics, I_n/I_1 and B_{rel}, are the currently most widely used measures of radiation damage.

It is worth noting that both of these important studies went to significant lengths to ensure that the distribution of dose within the crystal volume was even. In the case of Owen *et al.* (2006), using a very flat beam profile (top-hat beam) that was larger than the crystal ensured that it was evenly exposed, and in the case of Kmetko *et al.* (2006), using the same small angular ranges of $5°$ so that unevenness of dose distribution due to rotation was kept to a minimum.

1.6 Specific damage

Because, unlike global damage, specific damage is highly dependent on the local environment, a quantitative theory has not yet been developed. Due to the difficulties in creating multiple electron density maps, and in collecting successive full datasets at RT, systematic studies are scarce. Blake and Phillips (1962) did observe changes in the electron density patterns, but were not able to accurately characterise them, since they did not know the myoglobin primary sequence. Burley *et al.* (1984) first observed disulphide bridge breaking at 10 Mrad (10 kGy) in ribonuclease at RT: the most well known sign of specific radiation damage. At 100 K there is broad agreement on what types of damage are most commonly observed, and on the sequential order in which these take place:

1. *Reduction of metallocentres* occurs very early in diffraction experiments. Berglund *et al.* (2002) used a form of dose-multiplexing to collect full datasets in small dose ranges on

[12] Average value across characterised crystals.

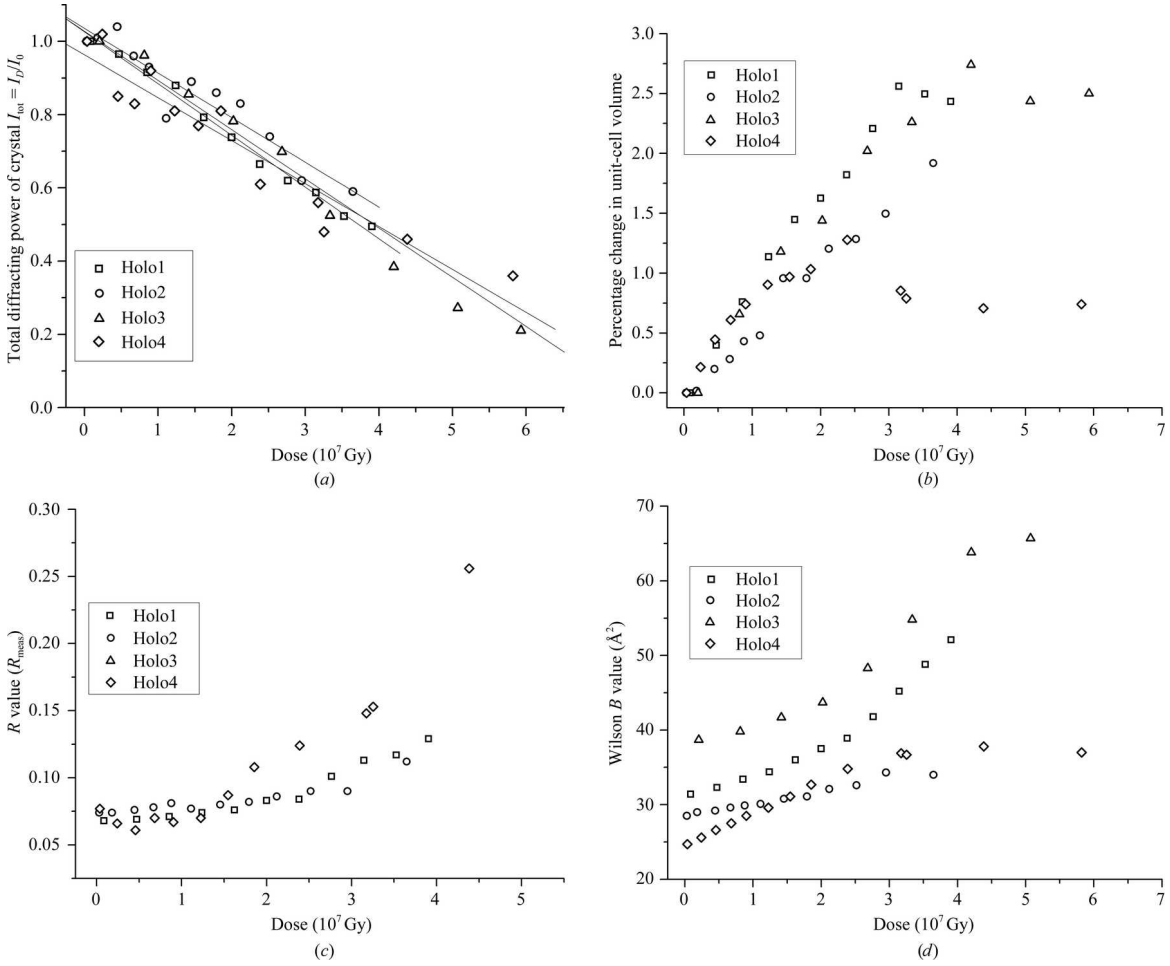


Figure 1.5: Comparison of radiation damage metrics for four holoferritin crystals the work of Owen *et al.* (2006), figure reproduced from Garman (2010).

horseradish peroxidase, revealing the step-by step reduction of the haem group. Unfortunately, no quantitative information on the dose is presented. Corbett *et al.* (2007) also observed this systematically using x-ray spectroscopy, and found a large reduction in metallocentre reduction at 40 K *vs* 100 K. Later, dose-calibrated work (Owen *et al.*, 2011) concluded that these changes would happen at as low a dose as ~45 kGy at 100 K.

2. *Elongation and breaking of disulfide bridges* was observed by several groups (Weik *et al.*, 2000; Ravelli and McSweeney, 2000; Burmeister, 2000), with the breakage occurring at around 5 MGy (Weik *et al.*, 2000; Ravelli and McSweeney, 2000), although dependent on protein and differing for the various disulphides in the structure.

3. *Decarboxylation of glutames and aspartates* was again observed by (Weik *et al.*, 2000; Ravelli and McSweeney, 2000; Burmeister, 2000). These amino acids lost density around 10 MGy (Weik *et al.*, 2000; Ravelli and McSweeney, 2000), and were completely cleaved by 50 MGy (Ravelli and McSweeney, 2000).

4. *Dehydroxylation of tyrosine* was observed by (Ravelli and McSweeney, 2000; Burmeister, 2000), with density being lost around 10 MGy (Ravelli and McSweeney, 2000).

5. *Cleavage of covalent bonds to heavy atoms* is common, as was observed for the cleavage of a mercury derivative at around 3-4 MGy by Ramagopal *et al.* (2005).

A particularly interesting phenomena observed by these studies is the specific order in which identical chemical transformations take place within specific proteins (Weik *et al.*, 2000; Ravelli and

McSweeney, 2000). There is no obvious correlation to solvent accessibility for decarboxylation (Fioravanti *et al.*, 2007), and the precise mechanism of this sequential specific damage remains unresolved (Weik *et al.*, 2002; Leiros *et al.*, 2001; Schiltz *et al.*, 2004). Although specific damage is normally undesirable due to its possible effect on the biological fidelity of the model, it can, under certain circumstances be used to advantage. Ravelli *et al.* (2003) proposed a method for using a damaged and undamaged pair of datasets to phase a structure, which has been employed successfully on several occasions (see Garman (2010) for a compilation of references).

1.7 Other variables

Several other parameters of the MX experiment have been considered and investigated as variables which might lead to a reduction of radiation damage if they were changed. However none other than cryocooling have been adopted as standard practice for radiation damage mitigation.

1.7.1 Dose rate

At RT, there are several studies that show a significant dose-rate effect in two seperate regimes: around $10^1 - 10^3$ Gy/s, improvements of several fold in dose lifetime have been observed, and the suggested mechanism is the self-termination of reactive chemical species interacting with each other (Southworth-Davies *et al.*, 2007; Barker *et al.*, 2009). At higher dose rates in the region of $10^5 - 10^6$ Gy/s, where many damage processes are diffusion limited, it may be possible to 'outrun' radiation damage by collecting data faster than the diffusion limited damage timescales. In all

cases, the potential improvement is much smaller than that offered by cryocooling. None the less, there are advantages to RT data collection, in particular with the advent of $in\ situ$ data collection (Axford *et al.*, 2012), and for certain virus crystals that are not suitable for cryocooling due to difficulties associated with cryocooling them because of their large unit cells and the presence of internal cavities. The large unit cells further makes these crystals very sensitive to mosaicity increases, compounding the challenges. Warkentin *et al.* (2011) observed 'dark progression' of damage at temperatures around 200 K, where damage processes appeared to continue after the X-ray beam was switched off, suggesting that collecting data over a shorter period of time would lead to less observed damage. They were not able to observe dark progression above 220 K due to their experimental set-up which did not allow them to collect probe data sets sufficiently fast. Owen *et al.* (2012) were able to observe a reduced level of radiation damage at RT using a super-fast Pilatus detector running at a 25 Hz frame rate, thus providing a proof-of-principle for the advantages of super-fast data collection at RT.

At 100K, there have been several studies showing either no, or a small <10% effect of dose-rate on global damage (Sliz *et al.*, 2003; Leiros *et al.*, 2006; Owen *et al.*, 2006). It is thus widely considered not to be variable at current synchrotron flux density levels (e.g. $\sim 10^{15}$ ph/s/mm^2 in Sliz *et al.* (2003)).

Taken to the extreme, Serial Femtosecond Crystallography (SFX) using a free electron laser can collect high fluence diffraction images from individual crystals on timescales of the order of 20 fs Chapman *et al.* (2011). This has been shown to lead to damage free diffraction images (Boutet

et al., 2012; Lomb *et al.*, 2011). However, this method is in the very early stages of development, and there are significant challenges to overcome in sample handling, and in data collection and processing. SFX relies on hundreds of thousands of randomly oriented individual images from single, non-rotating, micro-crystals, as opposed to the continuous sequence of images obtained by the rotation method, and is sufficiently different as a structural biology technique to be outside the scope of this thesis.

1.7.2 Scavengers

The efficacity of small molecule compounds to act as 'sinks' for reactive chemical species is currently an open question in the field. There is contradictory evidence regarding the efficacity of these compounds: see, for example, Murray and Garman (2002) and Kmetko *et al.* (2011). A summary of all work to date regarding scavengers can be found in Allan *et al.* (2013), which also points out the potential role of commonly used buffer components in the radiation chemistry, adding a further layer of complexity to this issue, and highlighting the need for more research.

1.7.3 Temperatures <100K

There is some evidence that helium temperatures may further reduce radiation damage by small amounts (Meents *et al.*, 2007; Teng and Moffat, 2002), however these effects are not large enough to warrant the cost of using lower temperature cooling systems for routine applications. Corbett *et al.* (2007), however, observed that 40 K had a significant effect on the sensitivity of metallocentres, and helium cooling may thus be advisable for the study of proteins with metals in

the active site, if their oxidation state is of interest.

1.8 Challenges in the field: why is it so hard to obtain reproducible results?

One of the fundamental goals of radiation damage research is to develop tools that allow us to ensure that experiments will not fail due to excessive loss of diffraction quality as a consequence of irradiation. Ideally, this would involve the predictive modelling of damage, which will allow the experimenter to plan a strategy that will allow the acquisition of the maximum data without losing biological fidelity. Although this has been achieved under carefully optimised circumstances (Owen *et al.*, 2006; Kmetko *et al.*, 2006; Leal *et al.*, 2011), due to experimental and practical limitations, this has not yet been translated into the standard pipelines for MX (Krojer and von Delft, 2011).

1.8.1 Choice of metrics

We have seen that different radiation damage studies use different damage metrics ($I/I_0, B_{rel}$ etc.), which may not be directly comparable. For example, I_n/I_1 is necessarily more biased by the high intensity, low resolution reflections than B_{rel}, which is derived from the change in the rate of drop-off of intensity with resolution from the Wilson plot. Furthermore, B_{rel} can be calculated in several different ways, with the choice of data processing strategy having a potential impact on the results: Kmetko *et al.* (2006) and Southworth-Davies *et al.* (2007) use $MOSFLM$ and $SCALEIT$, Warkentin *et al.* (2012) and Meents *et al.* (2009) use XDS and

$XSCALE$, Warkentin and Thorne (2010) used both $MOSFLM$ and $DENZO$ to integrate, then $SCALEIT$ and $SCALEPACK$[13] to scale, with $SCALEPACK$ giving results 1.33 times larger than $SCALEIT$. Additionally, scaling B factor (B_{scale}), the frame-by-frame scaling B factor calculated within a single data set is sometimes used as an online metric (Bourenkov and Popov, 2006; Hilgart *et al.*, 2011). This scaling B has, however, been shown to not be a reliable metric (Krojer and von Delft, 2011; Allan *et al.*, 2013).

Howells *et al.* (2009), in a paper about X-ray diffraction microscopy, combine data from the literature with additional measurements across a wide resolution range (0.2-100 nm) to analyse the maximum tolerable dose for multiple techniques (MX, X-ray Diffraction Imaging, electron crystallography and single particle reconstruction). This leads them to suggest a criterion of 10 MGy/Å as a rate of loss of resolution with dose, *e.g.* if 2 Å data are required, the maximum tolerable dose would be 20 MGy. Extending this approach, Holton (2009) suggests an exponential, resolution-dependent form of intensity decay with dose, $I = I_0 e^{-ln(2)\frac{D}{Hd}}$ where I is the spot intensity, I_0 is the spot intensity at zero dose, D is the dose, H the 'Howells criteron' of 10 MGy / Å, and d the resolution, such that $I = 0.5I_0$ when D is ten times the achieved resolution. This leads to good agreement with the studies cited above (Owen *et al.*, 2006; Kmetko *et al.*, 2006), and is an interesting functional form of decay for further investigation.

[13] Applied twice: firstly to a single wedge to internally scale, then between wedges to give relative scales used for B_{rel}.

1.8.2 Intrinsic variation in crystal quality

Although it has been shown that crystals behave broadly the same at 100 K, there is still significant variability between crystals prepared under identical conditions (see, for example Nowak *et al.* (2009), which describes a scavenger study, or Pozharski (2012)), and it is therefore important to ensure maximum redundancy in measurements.

1.8.3 Dose modelling

Perhaps the most significant reason that it is difficult to routinely apply the lessons learnt from radiation damage research to the MX workflow is that the stringent requirements of optimising an experiment for the even dose condition – where these rules apply – is often at odds with the practical requirements of obtaining the data required from whatever crystals are available in a timely fashion. Inadequate modelling of beam profiles and crystal shapes means that, away from the even dose condition, much of the knowledge from radiation damage studies is of little routine applicability. This is addressed in detail in Chapters 2-5.

1.9 This thesis in context

The work presented in this thesis represents a significant step forward in allowing the application of radiation damage research to real life crystallography. A software program, RADDOSE-3D has been developed and made accessible to the community, the size of the errors in dose estimation due to the absence of a 3D model of the crystal, beam, and experiment are highlighted, and a new

method for accurately quantifying dose that is independent of experimental specifics (Diffraction Weighted Dose, DWD) is presented.

- *Chapter 2* describes the historical development of the RADDOSE-3D code during the first 2 years of this research. It charts how the final data structures and algorithms were arrived at, and the early development in Matlab. It can be omitted without any loss of information relating to how to use the current versions of RADDOSE-3D.

- *Chapter 3* describes the software program RADDOSE-3D, which was written in Java and is now available to the community. This program is also described in Zeldin *et al.* (2013b) (See Appendix 2 for this and other papers published during this thesis). The concept of dose metrics to summarise a dose field as a single number is introduced here, and the metrics used in later chapters are described in detail.

- *Chapter 4* reports the main experimental results for this section of the thesis. The metrics described in Chapter 3 are investigated experimentally, and a protocol to use RADDOSE-3D for data collection optimization is described. A preliminary investigation in which RADDOSE-3D was applied to a challenging biological system, the membrane protein $PepT_{St}$, is presented. These results are currently in preparation for publication.

- *Chapter 5* presents the results of simulations performed using RADDOSE-3D to quantify the effect on dose distributions of various optimisation strategies that can be employed during data collection. These results are partially adapted from Zeldin *et al.* (2013a), making use

of the findings from Chapter 4. New results on offsetting the rotation and beam axes are also presented.

- *Chapter 6* Introduces and describes the development of high throughput microPIXE (ht-microPIXE), a side project of the work presented here.

- *Chapter 7* offers conclusions drawn from the previous chapters, and suggestions for future work based on this thesis.

Chapter 2

Historical Development of RADDOSE-3D

2.1 Background, and limitations of 1-D models

As described in Chapter 1, the program RADDOSE v1-3 provides a one dimensional model of absorbed dose in a non-rotating macromolecular crystal, and allows predictive estimates of crystal lifetime to be made based on an experimentally determined dose limit. For cases where the crystal is smaller than the beam, and where the beam has a uniform profile, this provides a valuable piece of information for use by strategy optimisation tools.

Since the original development of RADDOSE in 2004, advances in synchrotron technology, in particular data collection using microbeams between 1 and 30 μm in size (e.g. Riekel *et al.* (2005); Axford *et al.* (2012)), have led to the dose distribution in crystals often being highly inhomogeneous. The consequence of not being able to model the distribution of dose in three dimensions has been that, for systematic radiation damage studies, crystals must always be chosen that are smaller than the beam, or that experimental design must be carefully optimised to create an even dose profile (e.g. Kmetko *et al.* (2006); Owen *et al.* (2006)). Thus, existing guidelines for experiments taking into account the effect of dose on data quality are largely derived for even-dose situations, and do not represent a general case.

For the wide application of dose-lifetime guidelines to the crystallographic determination of novel structures, the 1D model in RADDOSE v1-3 is very powerful within the regime for which it was designed. However, it offers limited predictive power for cases where dose contrast is high. For example, work presented in Chapter 3 and in Zeldin *et al.* (2013a) describes a 'typical' MX experiment with a $100 \times 100 \times 100$ (μm)^3 crystal and a 20×20 (μm)^2 FWHM beam where the dose at

the centre of a 90° wedge is an order of magnitude higher than the average dose over the whole crystal volume. In cases like this, it is clear that a one dimensional model is of little use.

The work presented in this chapter chronicles the early development of a full 3D model of the diffraction experiment, eventually released as RADDOSE-3D (Zeldin *et al.*, 2013b). This deals with the beam intensity distribution, and crystal size and rotation explicitly, and allows a full treatment of the simulation for a more faithful estimation of dose. A further minor improvement in the accuracy of the simulation provided by the 3D model is made by taking into account both the absorption and the attenuation coefficient when calculating the dose at each depth point in the crystal. For a given small volume element in the crystal, the dose is a function of the intensity at the leading edge of that volume element, parametrised by the attenuation coefficient, and of the proportion of the beam absorbed by that volume element; parametrised by the absorption coefficient. Previous versions of RADDOSE have treated these as equal, since the difference is very small for the conditions normally used in MX, where 2% attenuation of the beam is typical (Garman, 2010). A full three dimensional dose field can treat these two coefficients separately, and thus corrects this approximation.

2.2 A note on coordinates

Throughout this thesis, the goniometer coordinate system is defined as: the z axis increasing downstream along the beam path, the y axis as the rotation axis of the goniometer used for data collection, perpendicular to z, and normally horizontal. Finally, the x axis completes the right-

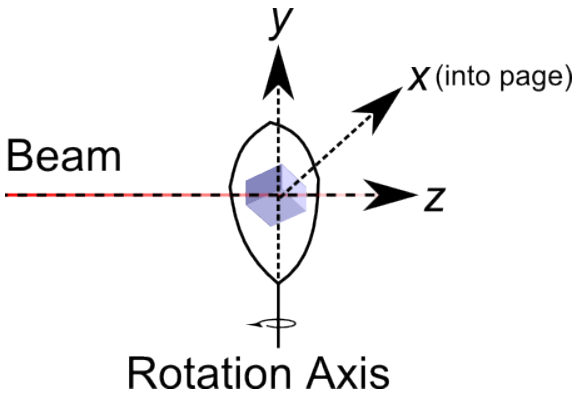


Figure 2.1: Schematic of right-handed RADDOSE-3D coordinate system used throughout this thesis. As viewed from above for horizontal goniometer rotation axis.

handed set, and is normally vertical. This is also shown schematically in figure 2.1. Note that this is different to the axis convention used at synchrotrons such as the Diamond light source, where x is this horizontal (rotation) axis, y the vertical, and z the beam axis.

2.3 Early development in Matlab

Early developmental work on the project was performed in Matlab[1], which offers a highly integrated development environment and a large number of visualisation libraries, making it ideal for fast prototyping and trying out ideas. Furthermore, it was already known to the author, which was an advantage. The original plan for the development was to release the code in two steps:

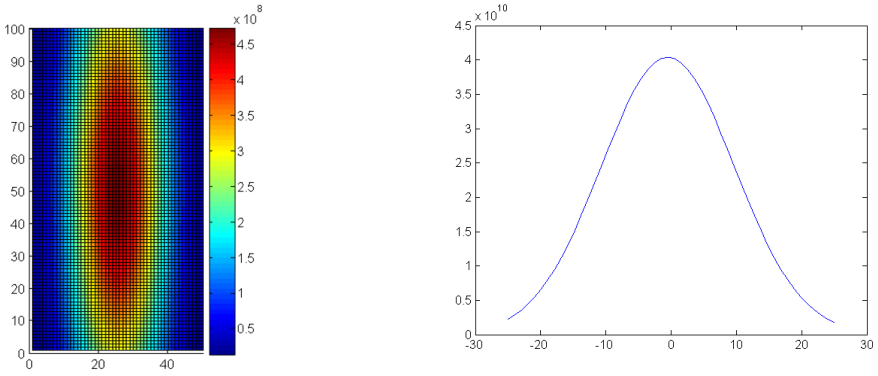
1. A reduced dimensionality model based on $x - y - z$ estimations of crystal size where the crystal was visualized down the goniometer axis, so projecting the y-coordinate of the experiment onto a plane. This had the advantage of being computationally N times faster (N^2 pixels vs N^3 voxels). It would also be simpler to implement as a mock up.

[1] http://www.mathworks.co.uk/

2. A full 3D model of the crystal, allowing for real crystal geometries to be implemented eventually, which could also model translations during exposure and more sophisticated data collections strategies.

2.3.1 Representing the crystal

In 2D, it is assumed that the whole y-width of the beam is incident on the crystal, so that the exposed region – the effective crystal size – can be projected onto a 2D plane. This originally started with a pixel model of the crystal projected along the y axis onto an M by N matrix (in 2D) with dose values at each index. However, this proved very inefficient since, after bulk matrix rotations on the whole array, the position of individual crystal pixels did not generally line up with the discretised points representing the beam intensity, requiring a lot of interpolation to assign intensity values. The program thus moved to a finite-element style model using a 2- or 3-dimensional vector to hold coordinates associated with the centre of each voxel. This then allowed a simple two (or three) dimensional rotation matrix to be applied to each coordinate very efficiently. Various attributes for each voxel, such as dose or total incident fluence (photons/unit area) were then stored in one dimensional vectors of the same length. The vector of voxel coordinates was made up of evenly spaced voxels, since although some regions (e.g. corners) are topologically more important, dose distribution intensity was required for the whole of the bulk, and so coarse graining inner regions would be counter-productive. Evenly spaced elements also had the advantage of being simpler to deal with since they all have the same associated volume element. This general



(a) Gaussian 2D beam profile used for early simulations. Axes are in μm and colour bar units are photons/(s$\times$ μm ^{2})

(b) 1D projection of adjacent Gaussian beam. Units of x-axis are μm and of y axis photons/(sec$\times$ μm).

Figure 2.2: Example of the projected Gaussian beam profile used in early Matlab development of RADDOSE-3D.

data structure for the crystal (a 2- or 3D coordinate vector, and additional vectors for each scalar attribute), was then kept for all future versions.

2.3.2 Representing the beam

For the early 2D model, the 2D profile of the beam was projected onto the x-vertical axis so that it was a 1D function normalized to the total beam flux. An example of this is shown in figure 2.2. Later, 3D Matlab developments used a pixel image of the beam calculated from a 2D Gaussian profile, and returned the intensity for the pixel in which the x, y coordinates of interest lay. This general approach allowed easy inclusion of arbitrary experimental beam profiles at a future date, at the expense of a more precise or computationally efficient direct Gaussian representation.

2.3.3 Assigning exposure to voxels

Given the basic structures for the beam and the crystal described above, the next step is to calculate the intensity of the beam at each point in the crystal during a simulated MX experiment,

and, from this, to assign a dose to each voxel. This is then repeated for many small angular iterations that make up the full rotational range of a wedge. The beam is, by definition, centred on the origin, and the crystal is initially rotated or translated to its starting position (with the obvious limitation that, in 2D, it is not possible to translate along y). After rotating the crystal to the position associated with the iteration, the Matlab code then finds all the voxels that are within the beam area, and finds the instantaneous intensity at each of these. For the ith voxel, this is the product of the unattenuated intensity – the interpolated value of the beam profile at coordinates (x_i, y_i) – and the attenuation factor $1 - e^{-\mu_{att} \times depth}$, where μ_{att} is the attenuation coefficient of the crystal, and $depth = surf(x_i, y_i) - z_i$ is the depth of the ith voxel from the surface of the crystal. All of the terms except for $surf(x_i, y_i)$ can readily be looked up in the beam image or the rotated crystal coordinate vectors. This absorption correction is often small, typically around 2% (Garman, 2010), but becomes highly significant if long wavelength X-rays are used, due to the wavelength-dependence of the atomic absorption cross-sections.

Efforts were made to avoid introducing a separate set of surface planes to define the crystal outline, since this would necessitate performing all rotation and translation operations identically on both the coordinate vector and on the planes, introducing an additional source of complexity in the data structures and code. This is principally because any change to the crystal orientation and position would have to be made simultaneously at several places in the source code each time these methods are altered: a recipe for the introduction of errors.

Finding an effective form for $surf(x_i, y_i)$ then went through several iterations:

1. The earliest versions found the coordinates of the voxel within the subset of voxels in the beam with the most upstream z value, and used this. This was a fair assumption for thick crystals where the front face of the crystal was close to normal to the beam, but led to large errors when this was not the case, as illustrated in figure 2.3.

2. The next step was to try to define a specific surface value of z for each voxel i in the beam at any given iteration. Our first effort was to define (at the same time as the crystal vector was being created) a logical vector labeling all the surface points. During each rotational iteration, this was logically paired with another logical vector of Points In the Beam (PIB), to give the surface PIB. The challenge was then to separate the points into those belonging to the front and to the rear faces of the crystal. Since there was no definition of a face in this representation (and even had there been, it was not clear how to deal with a corner or multiple faces in the beam), this was not trivial. An attempt was made to separate the surface PIB by taking all those upstream of the mean z value of the surface PIB to be the front face, but this fails in the case of high surface normal-beam angles and z-thin crystals, as illustrated in figure 2.4 where computational rounding errors with voxel assignment leads to incorrect assignment in certain crystal regions.

3. After this, a solution was implemented which searches upstream in z of each surface PIB in a 1.98x1.98 voxel size window for any other surface PIB. If this search came back negative, the point was defined as being part of the front surface. These front surface points were then used to generate a 2D searchable surface (using built-in Matlab functions), $surf(x, y)$,

that was searched using the x and y coordinates of each PIB to find the associated z surface coordinate for that point.

2.3.4 Assigning dose to voxels

The initial method of assigning dose, as shown in figure 2.5 for the 2D model, was to calculate the dose at each angular iteration in a similar way to that implemented in RADDOSE v1-3, but to do this for each angular iteration making up a wedge. This involved finding the effective crystal thickness at that angle, calculating the total absorbed energy for that iteration, and then distributing the absorbed energy amongst the illuminated voxels, by applying a weighting which was proportional to the incident fluence on them for that iteration. Essentially, this takes the dose for a given iteration (by using the total distance the beam travels through the crystal) and normalises the exposure profile to this so that the sum of the dose on all the nodes is equal to the total dose, and the distribution is the same as the distribution of exposure. If the total absorbed energy for the iteration is calculated accurately, this would be a valid method. However, as illustrated in figure 2.3, thickness estimation was not reliable, and so significant errors were propagated into dose distributions, as shown in figure 2.6, where mis-estimation of total absorbed energy per iteration led to a 21 % error in the peak dose along this line.

In order to correct this, a method to calculate the dose directly from the voxel fluence was

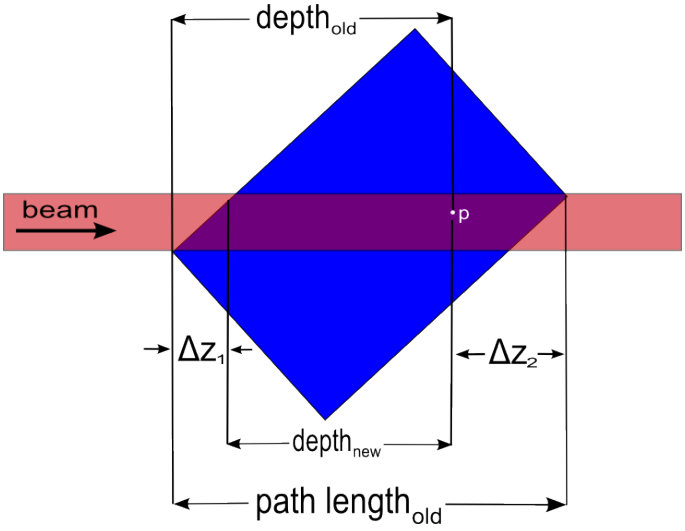


Figure 2.3: Evolution of exposure and dose assignment. The early (all 2D and 3D codes until June 2011) method simply found the point in the beam with the smallest z value and defined this as $surf(x_i, y_i)$ for all points within the crystal which are in the beam. This leads to errors in the depth required for the exposure estimation of size Δz_1 for a point p. These are small for low surface normal - beam angles, but this is not generally the case, and so this error cannot be ignored.

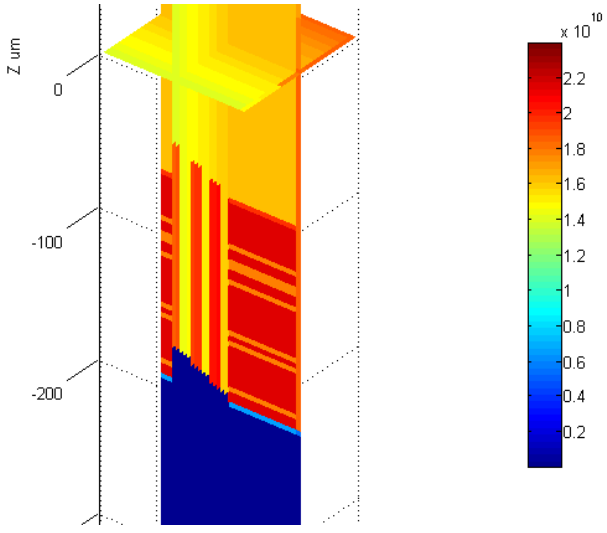


Figure 2.4: Example of the errors in an exposure slice map caused by badly assigned front surface points. The colour map shows the incident fluence at each voxel in photons. The exposed region (red and yellow) should be evenly exposed, but due to the shortcomings of the early surface point assignment protocol, jagged regions of voxels are mis-assigned highly different exposures, since the 'depth' is calculated from the rear, rather than the front face. Colour bar units are in photons/voxel.

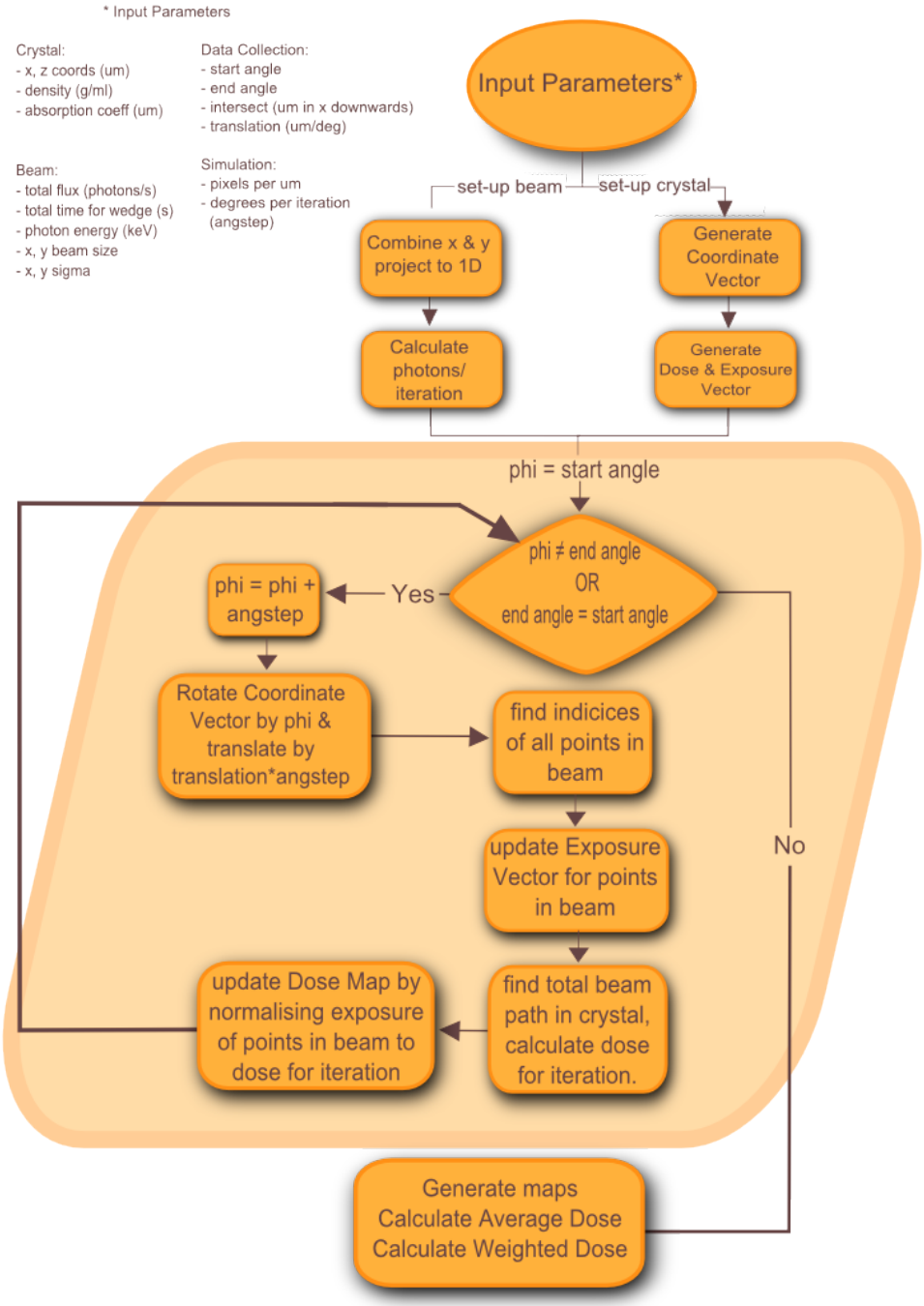
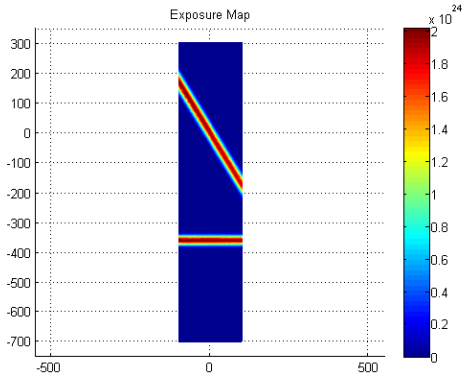
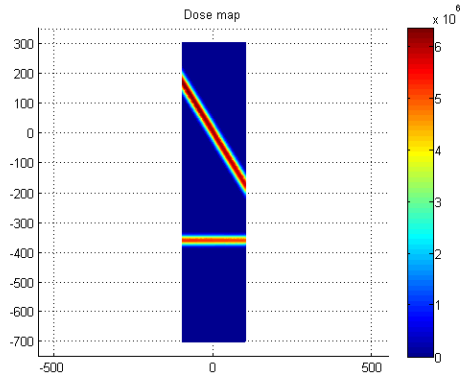


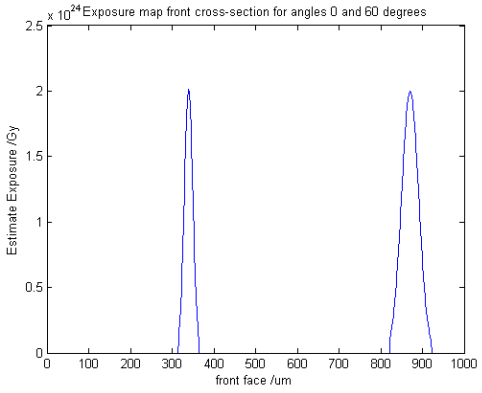
Figure 2.5: Flow schematic of the first iteration of the improvements to RADDOSE. The problem is reduced from 3D to 2D by projecting the 2D Gaussian beam profile into 1D and assuming that the beam is fully immersed in the y-axis of the crystal. This uses the iteration-by-iteration assignment of dose which, as shown in figure 2.6, was inadequate. Subsequent versions skipped the "find total beam path in crystal, calculate dose for iteration", and performed a dose calculation for each voxel, as detailed in section 2.3.3.



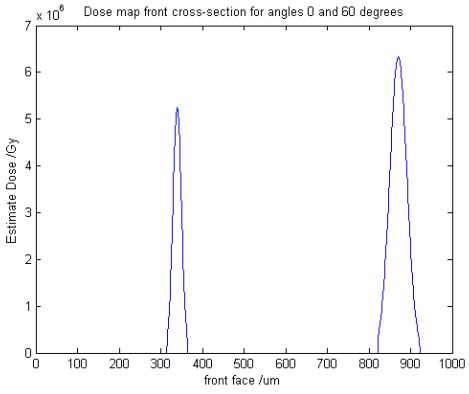
(a) Exposure map for 2 beams, one at $0°$ and one at $60°$. Parameters show in table 2.1.



(b) Dose map for same conditions as figure 2.6a. The difference is not immediately obvious from this representation



(c) Exposure profile for a line scan along the leading edge (min z) of figure 2.6a. As expected the peaks for the $0°$ (left) and $60°$ (right) are the same height.



(d) Dose profile for the leading edge (min z) of figure 2.6b showing the error caused by overestimating the dose for large angles by using path $length_{old}$ as explained in figure 2.3. The $60°$ peak is 1.21 times the size of the $0°$ peak, so this method gives an error of 21%.

Figure 2.6: Errors introduced by assigning the dose based on an estimate for a whole angular iteration at a time, and not accurately estimating the depth of the crystal. This was overcome by assigning dose on a voxel-by-voxel basis.

devised. Since each cubic voxel is small and equally sized, the dose for that voxel will be:

$$\frac{E_{incident}}{M_{voxel}}(1 - e^{-\mu_{abs} * z_{voxel}}) \tag{2.1}$$

Equation 2.1 is for a small unit volume with the front face of the voxel normal to the incident beam. If we wish to apply this to voxels in any general rotation, it is necesarry to show that this is still true for the limiting case of a voxel with its front face at $45°$ to the beam.

Crystal z	200 μm
Crystal x	1000 μm
Crystal y	>200* μm
Beam x FWHM	25 μm
Beam y FWHM	125 μm
Beam x aperture	50 μm
Beam y aperture	200 μm
Total photons	10^{14}
μ_{abs}	0.2 mm^{-1}
Angle wedge 1	$60°$
Angle wedge 2	$0°$

Table 2.1: Simulation parameters for the case illustrated in figure 2.6. * not explicitely modelled, but must be larger than the beam y aperture for the assumptions of the 2D model to hold.

This will be shown by considering the bounding cases, with the front face at $0°$ and $45°$ to the beam. First, for the normal case, with the origin at the centre of a cubic voxel, as shown in figure 2.7:

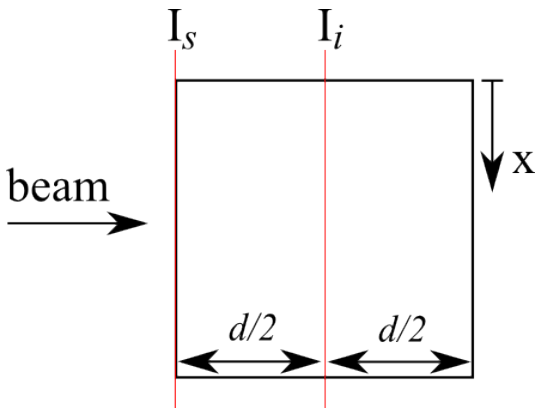


Figure 2.7: Diagram for dose calculation for a normally incident beam on a voxel.

where I_i is the incident energy density (J/m^2) at the centre of the voxel, $I_s = I_i e^{\frac{\mu d}{2}}$ the incident energy density at the surface of the voxel, and μ is the absorption coefficient. The dose for the voxel is thus:

$$Dose_{norm} = \frac{d}{M_{voxel}} \int_0^d I_s \left(1 - e^{-\mu d}\right) dx \tag{2.2}$$

Where M_{voxel} is the mass of the crystal, and the d in front of the integral is for the depth (into the page) of the voxel.

$$= \frac{d}{M_{voxel}} \int_0^d I_i e^{\frac{\mu d}{2}} \left(1 - e^{-\mu d}\right) dx \tag{2.3}$$

$$= \frac{2dI_i}{M_{voxel}} \sinh\left(\frac{\mu d}{2}\right) \int_0^d dx \tag{2.4}$$

$$= \frac{2d^2 I_i}{M_{voxel}} \sinh\left(\frac{\mu d}{2}\right) \tag{2.5}$$

$$= \frac{d^2 I_i}{M_{voxel}} \mu d \left(1 + \frac{(\mu d)^2}{4!} + \frac{(\mu d)^4}{16 \times 5!} + ...\right) \tag{2.6}$$

from a series expansion of $sinh\left(\frac{\mu d}{2}\right)$[2].

Next, the dose can be calculated for the corner case with the beam at $45°$ to the voxel, as shown in figure 2.8:

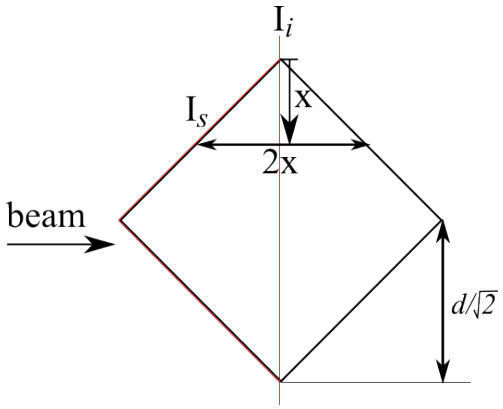


Figure 2.8: Diagram for dose calculation for $45°$ incident beam the same voxel as figure 2.7, except rotated by $45°$ along the axis into the page. The half-height $d/\sqrt{2}$ is from Pythagoras, based on voxel edges of length d.

$$I_s = I_i e^{\mu x} \tag{2.7}$$

$$Dose_{45} = 2\frac{d}{M_{voxel}} \int_0^{\frac{d}{\sqrt{2}}} I_s \left(1 - e^{-2\mu x}\right) dx \tag{2.8}$$

$$= \frac{4 I_i d}{M_{voxel}} \int_0^{\frac{d}{\sqrt{2}}} \sinh(\mu x)\, dx \tag{2.9}$$

$$= \frac{4 I_i d}{\mu M_{voxel}} \left(\cosh\left(\frac{\mu d}{\sqrt{2}}\right) - 1\right) \tag{2.10}$$

$$= \frac{d^2 I_i}{M_{voxel}} \mu d \left(1 + \frac{(\mu d)^2}{4!} + \frac{(\mu d)^4}{2 \times 6!} + ...\right) \tag{2.11}$$

from the series expansion of $\cosh x$[3].

[2] $\sinh x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + ...$
[3] $\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + ...$

So we can see that they only differ at the third order of expansion. Even for high absorption values (μ), and a typical voxel size, (e.g. $\mu = 5 \times 10^{-4} \mu m^{-1}$ and $d = 2\mu m$), the ratio of the third to the first term is of the order $(\mu d)^4 \simeq 10^{-12}$ i.e. negligible.

Having shown that the orientation of a fully enclosed voxel has only a negligible effect on the dose value, the dose was subsequently calculated from the fluence following equation 2.1. Further validation was performed by checking the dose for a no-rotation data collection against RADDOSE v2, and the two were in good agreement to within 3 %. The agreement is not perfect since RADDOSE-3D treats μ_{abs} and μ_{att} differently (leading to a lower dose), RADDOSE v2 deals with fluorescent escape explicitly (leading to a lower dose), and RADDOSE v1-3 uses a 21-point grid to simulate a Gaussian beam, while RADDOSE-3D treats this at a higher level of accuracy, which introduces rounding errors into the dose estimates in RADDOSE v1-3.

2.3.5 Moving from 2- to 3D

Using the 2D MATLAB code, the effect of data collection strategies on dose distribution was modelled for the first time. A comparison of the old model (no rotation) versus taking rotation and varying levels of translation during exposure into account was performed. The following input parameters were used:

- Crystal: 1000 μm x 400 μm x 200 μm (xyz) with y perpendicular to the beam throughout data collection. μ_{abs} of $2 \times 10^{-4} \mu m^{-1}$.

- Collection strategies:

1. Crystal does not rotate (c.f RADDOSE v1-v3).
2. 90° wedge, no translation.
3. 90° wedge, 2 μm translation in x per $^\circ$.
4. 90° wedge, 5 μm translation in x per $^\circ$.

- Beam: 100 μm x 50 μm collimated 2D Gaussian (σ_x = 50 μm, σ_y = 10 μm). 10^{12} total photons/s, 100 s per data set, 10 keV photons.

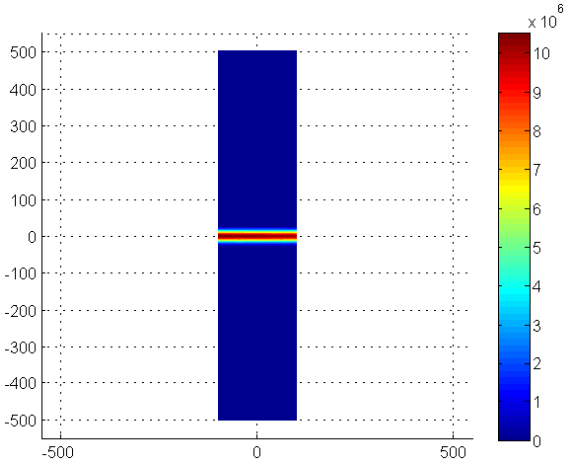
Dose maps for these simulations are shown in figure 2.9, and the resulting average doses in table 2.2. It is clear that it is important to be able to model translation, since the effect of translation on dose is significant.

	No rotation	90° wedge, 0 μm/deg	90° wedge, 2 μm/deg	90° wedge, 5 μm/deg
Average Dose (Exposed Region only)	5.2 MGy	1.9 MGy	1.3 MGy	0.61 MGy
Old Dose (RADDOSE v2)	5.2 MGy	5.2 MGy	5.2 MGy	5.2 MGy

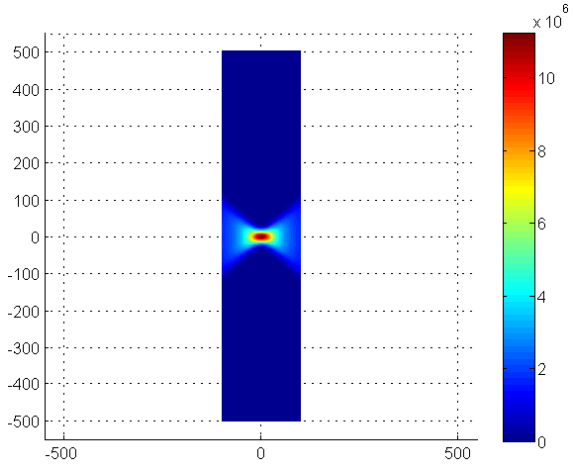
Table 2.2: Simulation results for the 2D simulations, with translations along the x axis. The Old Dose is that reported by RADDOSE v2, without using the `GAUSS` keyword, so average rather than maximum dose (as would be reported with the `GAUSS` keyword in RADDOSE v1-3), and highlights the limitations of a 1D approach.

Shortly after submission of the author's transfer report (end of June 2011), it was decided to skip the originally planned 2D release of RADDOSE, and to move straight to a 3D model. Reasons for this were:

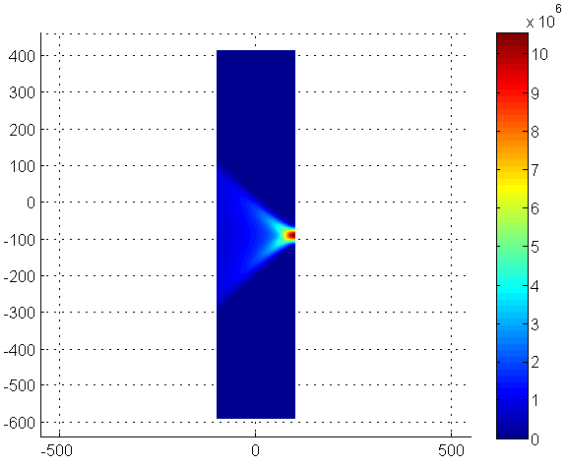
1. Translation is normally along y. This requires 3D (main reason).
2. Most camera geometries are not down the y-axis so reconstruction of real experimental geometries would be difficult.



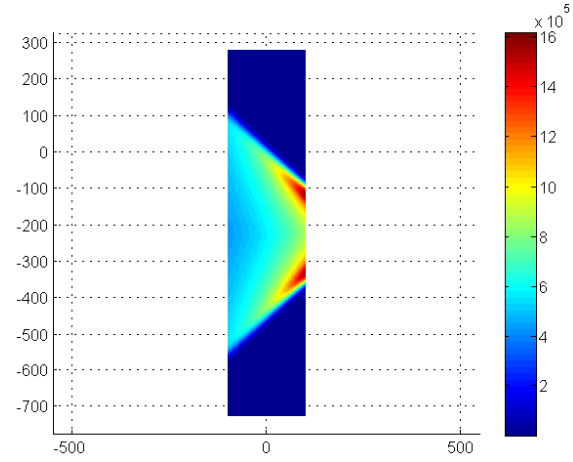
(a) Dose map for collection strategy 1: no rotation. Units are μm on x and y axes and Gy on colour bar.



(b) Dose map for collection strategy 2: $90°$ wedge, no translation. This represents a typical dataset. There is clearly a larger region of low damage and a smaller region of high damage than for collection strategy 1. Units are μm on x and y axes and Gy on colour bar.



(c) Dose map for collection strategy 3: $90°$ wedge, 2 μm translation/deg along the x axis. This shows some spreading of dose and shifts the high damage region to the back of the crystal. Units are μm on x and y axes and Gy on colour bar.



(d) Dose map for collection strategy 4: $90°$ wedge, 5 μm translation/deg along x. This shows a large decrease in the maximum dose found within the crystal compared with all of the above figures, illustrating the importance of these effects. Note that the scale of the colour bar is one order of magnitude lower than the others in this figure. Units are μm on x and y axes and Gy on colour bar.

Figure 2.9: Dose maps for the early 2D simulations, showing the large effect on dose distribution of rotational strategies. The translations were made along the x axis, which is experimentally atypical. Nonetheless, the large effects were a major motivator for moving to a 3D model earlier. The blue box represents the crystal.

3. Depth in y is rarely constant.

4. It was not desirable to develop the 2D code too far, as this would have made the change to 3D a much larger job than necesarry.

5. 3D is more reliable and reflects real geometry better, and is easier to validate.

At the same time as this move, the code was separated into 3 main functions: $makebeam$, $makecrystal$ and RADDOSE$3DROTATE$. $makebeam$ and $makecrystal$ initialised the crystal and beam data structures, and RADDOSE$3DROTATE$ used the first two functions along with data collection parameters as inputs to perform the simulation, and to output results. A flow schematic for the 2D version is shown in figure 2.5. After moving the Matlab code to 3D, updating it to use the final versions of exposure and dose assignment described above, and checking the results were consistent in the appropriate limits (e.g. orthogonal beam-crystal face for dose assignment), the code was re-written for public release using object oriented design principles in Java.

2.4 Move to Java and development of object oriented structure

The main reason why it was necessary to move away from Matlab was that Matlab code cannot run without a local installation, which is (a) proprietary, and (b) makes online integration at X-ray facilities problematic. Additionally, using a different language allows a potential increase in speed and a better encapsulated code-base, producing a more robust piece of software. Java was selected as the language of choice for several reasons:

1. It is completely object-oriented, allowing the code to be highly modular.

2. It is natively cross-platform, so it will run optimally on any operating system with a Java Runtime Environment.

3. It is widely used, and so the codebase can easily be utlised by other researchers.

4. Local expertise on the language was available in the group from Markus Gerstel, another DPhil student who has a computer science background.

An object oriented structure allows each of the key elements of the code to be encapsulated into classes, keeping the data structures and associated functions within a single 'object'. For example, this means that rotating all the voxels in a crystal and the vertices defining planes that enclose the crystal at the same time can be abstracted away into a single $rotate()$ method for a 'crystal' object[4]. As a consequence, the challenges associated with having to rotate both of these structures simultaneously are eliminated, and the current version of RADDOSE-3D calculates depth for cuboid crystals by finding the minimum positive (i.e. upstream) distance to any one of the enclosing planes, using basic linear algebra. A full description of the current version of the code and its object structure is presented in Chapter 3.

[4] This crystal object also holds the coordinates, dose vector, and exposure vector data structures along with methods to rotate the crystal, return a description of the crystal, calculate absorption coefficients, and to find the depth of a voxel in the crystal etc.

Chapter 3

RADDOSE-3D: the Code

Sections of this chapter are adapted from Zeldin et al. *(2013a) and Zeldin* et al. *(2013b). All adapted or reproduced sections were written by the author, with no more input from co-authors than would be appropriate for a thesis.*

3.1 Design principles and overview of the code

The design principle behind RADDOSE-3D was to create an open source package that not only provides a full 3D simulation of MX experiments for the calculation of dose profiles, but also creates a general framework that could be extended to deal with many other macroscopic phenomena involving macromolecular crystals and X-ray beams. In its most general sense, this package takes a voxel representation of a crystal, a computational model of a beam, and then simulates the interaction of the two via a strategy, here described by a wedge which specifies the angular range and translations associated with a single strategy within an MX experiment. The crystal has various properties associated with each voxel: position in three dimensional space, dose, and fluence (photons/unit area). The beam is described by an intensity profile, the photon energy, and the total flux. The logical flow of the code, as shown in figure 3.1, starts by using the parameters defined in the input file or through the online interface available at www.raddo.se to set up the three components (crystal, beam and wedge) described above. Input files describe a single crystal, and one or more beams and wedges of contiguous data collection. An example is shown in figure 3.2. The full grammar for the input files is detailed in Appendix 1. A single run of the program thus always acts on a single crystal, but can simulate several wedges (e.g. in different places on the

crystal), each wedge being associated with the most recently defined beam. For instance, different beams might be required, and involve changes in energy for Multi-wavelength Anomalous Diffraction (MAD) collection, or changes in flux due to a difference in transmission or an electron storage ring refill. The code works its way through the simulation in the order described in the input file. The output files are updated after calculating each wedge, so that a step-by step report of the dose state of the crystal can be provided at the end of the simulation. This step-by-step output allows information to be obtained on the time course of the dose accumulation. The later stages of code development were performed in collaboration with Markus Gerstel (MG), and his contributions are outlined explicitly below. Principally he implemented the use of standard design patterns for the UML structure of the code, which means that the structure should be easily comprehensible to coders versed in this approach, and reduces the need to alter source code at multiple places in order to extend the capabilities of RADDOSE-3D.

3.2 Technical description of the code

RADDOSE-3D is written in Java 1.6 with three main classes[1] forming the core of the code: a `Crystal` class, a `Beam` interface and a `Wedge` class. A single strategy for an MX experiment is thus fully described by a `beam` acting on a `crystal` according to the strategy described in a `wedge`

[1] In object oriented programming, a `class` is a computational framework which describes the general case of an `object`, which is the specific instantiation of that `class`, and is what is then dealt with by other `objects` in the code. In Java, methods (functions for a class) for an `object` are accessed with commands of the type `object.method()`, and, as is best practice in object oriented programming, variables stored in objects are not directly accessible, but are interacted with *via* 'getter' and 'setter' methods which, respectively get or set a variable; for example `thisCrystal.getSizeXUM()` will return the size of the crystal `thisCrystal` object along its x dimension in μm. This only has meaning for an object and not a class, since although the method may be defined in the class, there is no value of the crystal size in x specified until the crystal object is actually created using a `constructor` method written into the class. Classes are conventionally capitalised, and objects are referred to in lower case.

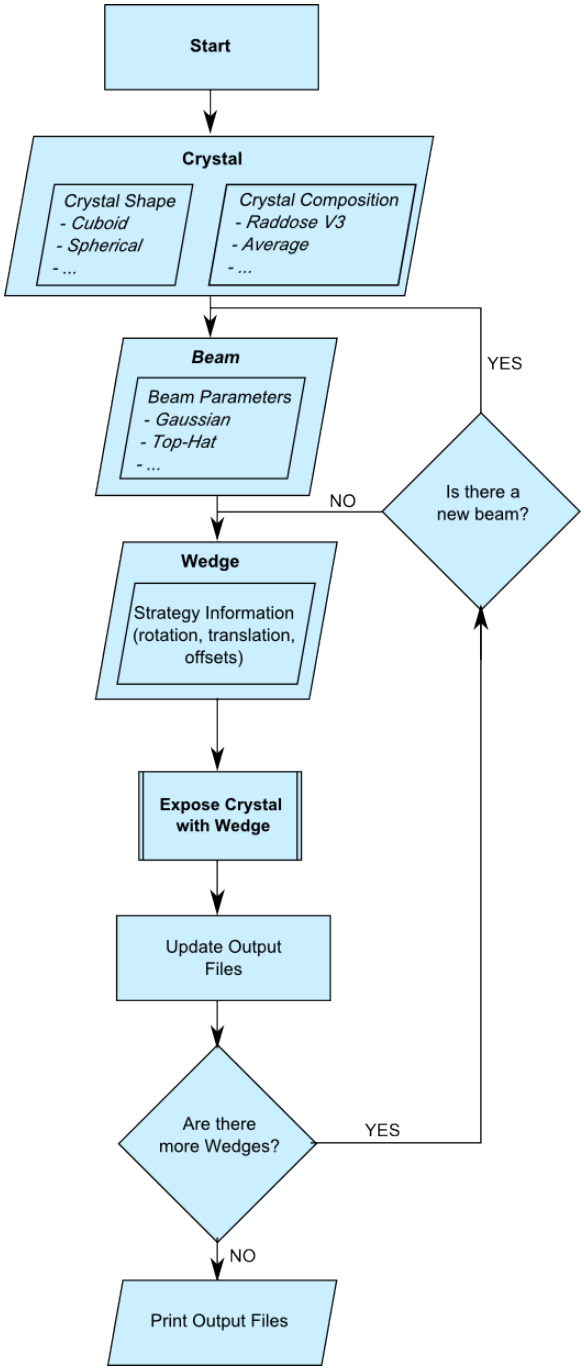


Figure 3.1: Flow chart illustrating the structure of the RADDOSE-3D code.

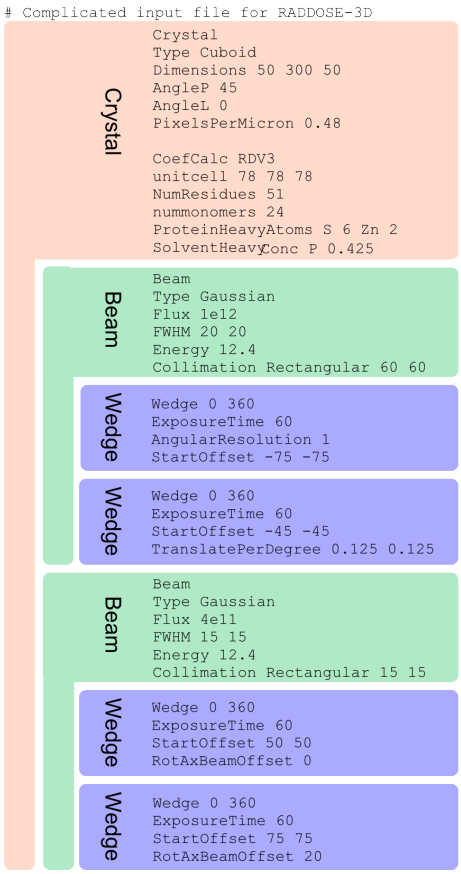


Figure 3.2: Structure of a RADDOSE-3D input file. The crystal is first described by a `Crystal` block (colour salmon), which remains active throughout. Next, a beam (the green block), followed by one or more wedges (blue), is described. As illustrated in figure 3.1, each wedge defines how to simulate an exposure of the previously specified beam with the crystal. Here the first two wedges will act with the first beam, and the last two with the second beam. The full input syntax is described in Appendix 1.

object. The full javadoc code documentation is available at www.raddo.se/doc. The key classes are described below individually.

3.2.1 The Crystal class

Crystals are always represented by an abstract class which contains the general computational description of the crystal. Specific implementations of this class currently exist for cuboid crystals (`CrystalCuboid`), and spherical crystals (`CrystalSpherical`). The actual voxel data structures are abstracted away, and other classes access these through methods acting on individual voxel indexes, such as `getCrystCoord(i, j, k)`, which returns the x, y, z coordinates for voxel (`i, j, k`). For non-cuboid crystals, an occupancy vector (queried *via* the `isCrystalAt(i, j, k)` method which has value 1 if there is crystal present, and 0 if not) allows the crystal to sparsely fill a bounding box; and this is how non-cuboid crystals can be specified. Three 1D vectors record the fluence, dose, and elastic yield for each voxel, calculated as described in Chapter 2[2]. The value at any index (`i, j, k`) can be incremented with addition methods, and queried with getters. This has the advantage of, for example, making it impossible for another part of the program to decrease the dose at a voxel[3], which would be physically impossible. The `Crystal` class has a method to `expose(Beam, Wedge)`, which takes a `beam` and `wedge` object and performs the iterative rotate-calculate dose cycle. Pseudo code for this method is shown in code box 3.1. Each implementation of the `Crystal` class (i.e. currently `Cuboid` or `Spherical`) has its own method to return how far

[2] Except for elastic yield, which was not previously described: this is calculated in the same way as dose, but using the elastic rather than absorption cross-section.

[3] Because there are no public (accessible to other classes) methods for subtracting dose from the values at an index.

a voxel `i, j, k` is from the front surface of the crystal for an angle ϕ (the rotation axis of the goniometer).

Dose-dependent diffraction intensity models, that can be used to deal with the loss of diffraction intensity as different parts of the crystal are damaged, exist as objects (Dose Decay Models: DDMs) that are stored within a `crystal` object. They are thus accessed with `crystal.getDDM().someMethodForTheDDM(...)` type commands.

Absorption, elastic, and attenuation coefficients are dealt with in a similar way, with a `coefCalc` object which can return energy dependent cross sections. The code currently has two possible `CoefCalc` classes: `CoefCalcDummy` which contains average crystal values[4], and `CoefCalcRDV2` which takes crystal composition information from the input, and calls RADDOSE v2 to explicitly calculate the coefficients for that particular crystal.

Code box 3.1 Outline of the basic exposure method.

```
FOR each angle within range and resolution set in Wedge
    FOR each voxel in Crystal
        IF voxel is in the crystal
            let coords be (x,y,z) coordinates of voxel
            Translate coords by angle*(transX, transY)
            Rotate coords by angle
            Calculate fluence at voxel(x,y) using Beam
            find depth of voxel along beam to upstream face of crystal
            Calculate dose, elastic yield, and fluence at voxel(x,y,z)
            Update dose, elastic yield, and fluence for voxel
        END
    END
END
```

[4] Absorption coefficient as $0.237mm^{-1}$, and attenuation coefficient as $0.281mm^{-1}$. These were calculated (Holton and Frankel, 2010), for 1 Å (12.4 keV) incident radiation, using an average composition from a survey of the PDB (Berman *et al.*, 2002). When using these averages, the crystal density is taken as 1.2 g/ml with an assumed 50% solvent content with solvent mass ~1 g/ml, and a protein mass of 1.41 g/ml (Fischer *et al.*, 2004). It is worth noting that the variation in μ_{abs} caused by a fixed energy (a ~50% decrease between 11 and 13 keV) is within the same regime as the changes due to buffer and protein content, which can easily be a factor of 2 (Holton, 2009).

Pixel resolution, defined in pixels per micron, is a particularly important parameter to be specified. An insufficient resolution will lead to significant errors in the dose metrics (described in section 3.3). The recommended voxel size for convergence of these metrics is ~1/10th of the beam FWHM (minimum of x and y), as demonstrated by the simulation results shown in figure 3.3. The specific meaning of each metric is described later in this chapter, but the key point is that at they are stable at resolutions higher than 1/10th of the beam FWHM. Setting this value too low will lead to errors in dose estimation, particularly for Max Dose (the maximum dose at any point in the crystal), and too high a value will increase the code runtime needlessly.

3.2.2 The Beam interface

Beam classes implement the `Beam` interface which defines the methods that they must be able to fulfil, so that they all interact with other parts of the code in a standard way. These are:

- `beamIntensity(x, y, offset)` which returns the intensity, in J/ μm ^{2}/s, for a voxel at position x, y, with a rotation-axis to beam offset of `offset` μm, which is defined as the distance between the beam centre and the rotation axis of the crystal, and can be offset to spread dose more efficiently, as described in Chapter 5.

- `getDescription()` which returns a description string for the beam to be used in the output.

- `getPhotonEnergy()` which returns the photon energy. For example, this is called while finding absorption and attenuation coefficients for a new beam energy if the `CoefCalcRDV2`

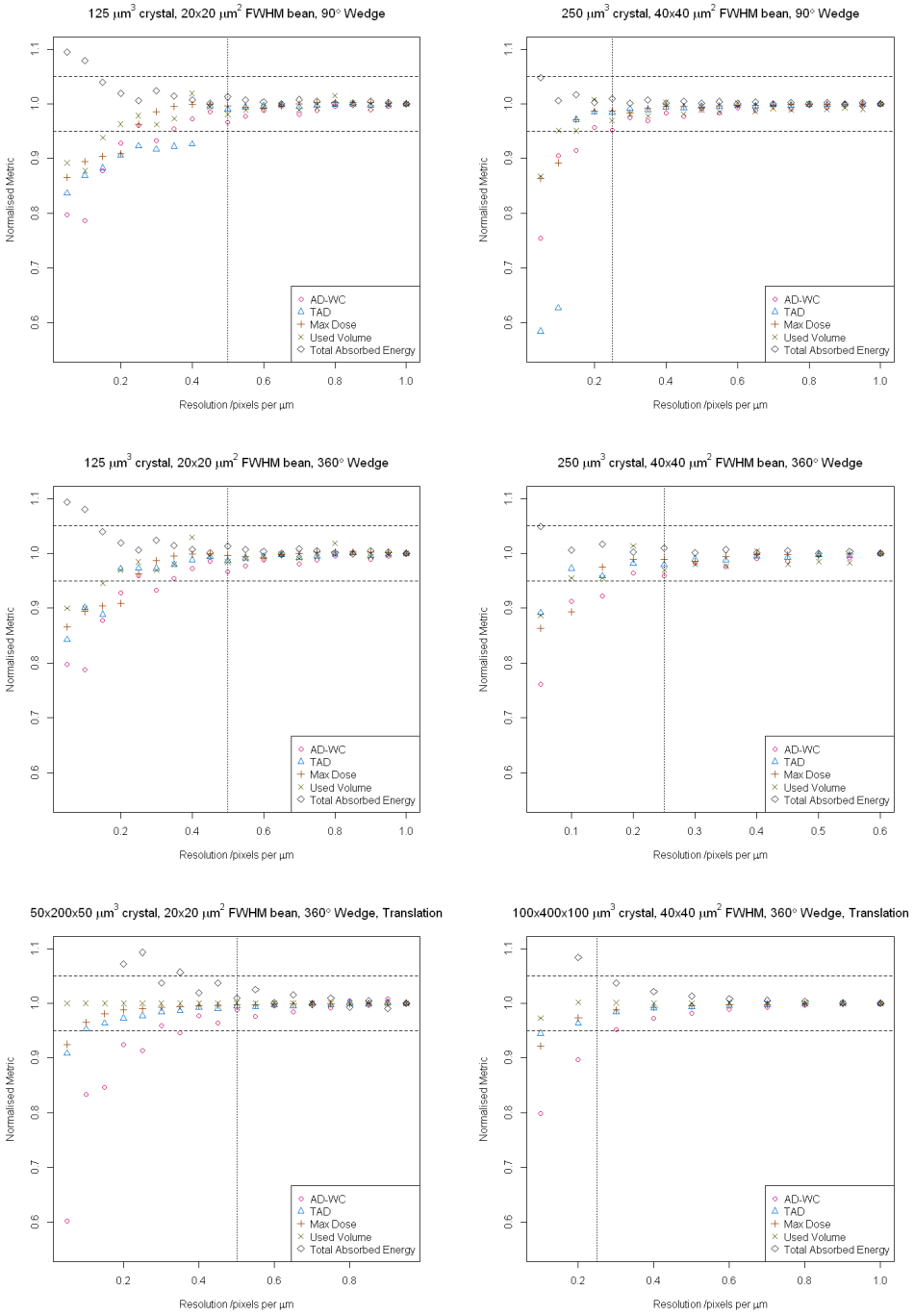


Figure 3.3: Voxel resolution testing for large (up to 15×10^6 voxels for the 250^3 μm ^{3} crystal) and smaller crystals, with 20×20 and 40×40 μm ^{2} uncollimated FWHM beams (left and right columns respectively). The top row is for a 90° rotation, the middle row is for 360°, and the bottom row is also for 360°, but with a full translation. Note that voxel resolution is quoted in voxels per micron. In all cases, the simulations converge towards their high voxel resolution values at around $1/10^{th}$ of the beam FWHM, shown by the vertical dashed line (0.25 vox/ μm for the larger beam, and 0.5 vox/ μm for the smaller beam). Horizontal lines have been added at $\pm 5\%$ of the converged values for visual clarity. The metrics are described in section 3.3

class is being used.

- `getPhotonsPerSec()` this returns the photon flux.

Currently three implementations of `Beam` exist: `BeamTopHat`, which describes a rectangular beam with a constant flux within a collimated region, and zero flux outside, `BeamGaussian`, which describes a 2D Gaussian shaped beam, and `BeamExperimental`, which uses a 2D grid of beam intensity as input. `BeamExperimental` is itself an abstract class, describing the general case of an experimentally determined beam defined through a 2D grid of relative intensity values, and `BeamExperimentalPGM` was written (see Chapter 4) specifically to handle input from the Portable GreyMap (PGM) files generated by the beam profiling facilities at the ESRF, Grenoble, France. These implementations differ principally in their `beamIntensity` methods, with `BeamTopHat` returning a constant value if x, y is within the beam area, `BeamGaussian` explicitly calculating intensities using the highly optimised Apache Commons Mathematics Library[5], and `BeamExperimental` using bilinear interpolation to find intensity values.

3.2.3 The Wedge class

The `Wedge` class describes a fairly simple type of object which is only used to store the information needed to perform a `crystal.expose(...)`, and to output a string describing the details of the strategy for that wedge. Wedge properties are: angular resolution, start angle, end angle, offset between the rotation axis and beam axis, starting position of the crystal relative to its centre, total

[5] http://commons.apache.org/math

time, and translation for this wedge to be performed during data collection. These are all described in detail in Appendix 1. Combined with a `beam` object, this fully describes a data collection to be performed on a crystal.

3.2.4 Input

Input was initially handled through an `Input` class which parsed a text file using a series of logical statements to assign the relevant parameters to the Crystal, Beam, Wedge and Output options. In a sub-project initiated by MG[6] this was upgraded to an LL* parser, generated using ANTLR (www.antlr.org). This has two principle advantages: it is much more robust to errors in input files, and it is easily extensible to take input from other sources such as a graphical interface or a real-time streaming input provided by experimental apparatus.

The basic input structure of a file is made up of a single crystal block of text, describing the crystal, followed by one or more sets of: a beam block, followed by one or more wedge blocks. This can be written as a regular expression: $\{Crystal\,(Beam\,Wedge+)+\}$, where the + indicates that one or more of the previous elements must be present. An example input file has already been shown in figure 3.2. The full syntax user guide is presented in Appendix 1, and should be referred to in order to write input files for use with the code.

The specific `crystal` and `beam` interfaces are created during run-time by `BeamFactory` or `CrystalFactory` classes, following the factory method design pattern (Gamma *et al.*, 1994),

[6] OBZ designed the input structure EBNF metasyntax notation, and wrote the first draft version, MG debugged and coded the full version, MG and OBZ updated it as the project evolved.

which allows a single `factory` object to interact with the parser and to instantiate a `crystal` or `beam` of any existing subtype, without having to explicitly name the individual type. This means that to add an extra `Crystal` or `Beam` class, only the implementation of the abstract `Crystal` class or the `Beam` interface must be written. The parser passes the arguments in the input file 'blind' to the `factory`, which then attempts to create a `crystal` or `beam` object of the type specified by the `Type` keyword in the inputfile. The parser thus only needs to be updated if a new keyword is needed, as was the case, for example, to take beam inputs from a file rather than *via* parameters. This design pattern was implemented by MG[7].

3.2.5 Output

The raw output of the 3D dose simulation is a scalar field describing the dose state of the crystal at every voxel and at every time-step. Since each wedge is broken into time steps corresponding to a defined angular resolution (default: $2°$), a fully time-resolved picture yields a dose field for each angular step computed during the simulation. Since it is not practical (or very illuminating) to manually look at several hundred (e.g. 180 dose fields for a single $360°$ wedge at default resolution) four dimensional arrays, each with hundreds of thousands of entries (typical number of voxels: ~125,000), it is vital to compute aggregate metrics. A full list of these is described in detail section 3.3. The code currently has four selectable output modules, which print to file or STDOUT some combination of aggregate metrics, a histogram of dose state, and the full dose field

[7] Previously, the code explicitly checked if the 'type' of the `crystal` matched a list of `crystal/beam` types, then created the objects explicitly: structurally simpler code, but less robust, less extensible, and requiring synchronised changes of the input module and of the new class in order to implement a new `crystal/beam` class.

of the crystal. These are described in table 3.1.

File	Contents	Originating Class
output-summary.txt and STDOUT	Human readable description of the `crystal`, and of each `wedge` and `beam`. Aggregate metrics and a dose state histogram of the crystal after each wedge.	`OutputSummaryText`, which calls the `getInfo()` methods from `Crystal, Beam`, and `Wedge` classes for descriptions, and uses these classes 'getters' to access the aggregate metrics. The dose state histogram is generated from the `crystal` object after each wedge by looping over the whole Dose vector and populating a histogram of non-zero voxels.
output-summary.csv	Aggregate metrics in an easily parsable comma separated variable file.	`OutputSummaryCSV` which gets the values from the `Crystal, Beam`, and `Wedge` classes using their 'getter' methods.
output-DoseState.csv	5-element CSV file with $x, y, z, dose, fluence, totalelasticyield$ for each voxel, output after the last wedge is processed.	`OutputFinalDoseStateCSV`, which loops over the data structures in the `crystal` object.
output-DoseState.r (Entirely MG's work.)	Prints to file an R script (www.r-project.org) that will display a 3D iso-surface visualisation of the crystal.	`OutputFinalDoseStateR`, which loops over the data structures in the `crystal` object.
output-FluencePerDose-HistCSV.csv	An iteration-by-iteration breakdown of the exposure state of the crystal. Each row contains Wedge number, ϕ angle of iteration, the Diffraction Weighted Dose for that image (see section 3.3.2), and a fine grain fluence weighted dose histogram (default 0-100 MGy in 0.1 MGy steps), of the effective contribution of each dose bin to the diffraction for that iteration. See section 3.3.2.	`OutputFluencePerDoseHistCSV`, which updates the histogram for each voxel of each iteration of each wedge.

Table 3.1: Current output modules for RADDOSE-3D. Their final class architecture, using the Observer design pattern (Gamma *et al.*, 1994), as implemented by MG, is described below.

Output was originally handled by several output classes called sequentially after the crystal was exposed, and the desired information was printed to files or the terminal directly. This was subsequently upgraded by MG to use an Observer Design Pattern (Gamma *et al.*, 1994), which allows individual classes to monitor the evolving properties of the crystal without being directly coded into the `crystal.expose(...)` method. Classes 'subscribe' to the crystal object, and are then notified of changes that occur to the crystal, either voxel-by-voxel for maximum detail by

instantiating the `ExposeObserver` interface, or at a more global level through the `Output` interface when the crystal is instantiated, the beam is set or updated, or a wedge is processed. The specific places where `ExposeObserver` is called are outlined in code box 3.2.

For example, after each wedge is exposed, the subscribed observers that implement `Output` are all notified. This takes the form of calling a method `thisOutput.publishWedge(wedge)`, but what each observer does with this information is dependent on its role. For example, the `OutputSummaryCSV` class will append a line to the summary CSV file, but the `OutputFinalDoseState` does nothing (since it only reports on the final dose state, not intermediary ones between wedges) until it gets the the final `thisObserver.close()` call, at which point it prints the full final dose state of the crystal to file. On the other hand, the `thisObserver.close()` method for `OutputSummaryCSV` simply closes the CSV file and resets its status. The advantage of this structure is that any output format can be obtained with only the creation of a method which implements one or both of the `Output` and the `ExposeObserver` interfaces, without having to change the source code of the `Crystal` class. Essentially it encapsulates output statistics away from the actual dose field calculations.

A second advantage of this is that, if only high level statistics are desired, the user can specify which ones are required in the command line[8], and if none of the desired output modules implement the fine-grain `ExposeObserver` interface, then Java's Just In Time compilation engine (JIT) will not actually perform those calls, leading to gains in speed. In contrast, languages such as C++ would still call the method, but the outputs would be passed to a null object. An example of this

[8] With the `-o` option.

would be if only the summary CSV file and the final dose state were required, since neither of these output modules requires the `ExposeObserver` interface. Calls to the `Output` interfaces are handled by the `Experiment` class that ensures that `OutputObservers` are created at the start of the simulation, processes the input `Crystal-Beam-Wedge` object stream from the parser, and then resets the `OutputObservers` cleanly after each step.

Overall, these behind the scenes changes led by MG have made the source code more extensible, and brought it in line with the computer science design pattern software engineering philosophy, but do not affect the absolute values output from the code, as described below.

Code box 3.2 Outline of the various methods called through the `ExposureObserver` interface. This allows an output class to get voxel-by-voxel information for every voxel in every iteration, or for every voxel after every wedge. This dense probing of the scalar dose (or exposure, or elastic yield) field can then be used for the calculation of any desired statistic.

```
FOR ALL ExposureObservers
   Report start of wedge
END
FOR each angle within range and resolution set in Wedge
    FOR each voxel in Crystal
        IF voxel is in the crystal
            Translate, Rotate, expose and update voxel
            FOR ALL ExposureObservers
                Report changes and current voxel state
            END
        END
    END
    FOR ALL ExposureObservers
        Report that the image is complete
    END
END
FOR each voxel in Crystal
    FOR ALL ExposureObservers
        Report end of wedge state
    END
END
FOR ALL ExposureObservers
    Report that  exposure is complete
END
```

3.3 Aggregate Metrics

The code outputs 10 aggregate metrics, of which five are different ways of calculating dose (MGy). The purpose of an aggregate dose metric is to describe the state of the crystal in a way which informs the user on how damaged they can expect collected data to be. The proposed dose metrics are tested experimentally in Chapter 4.

3.3.1 Simple dose metrics

These are described below, and summarised in table 3.2. They are all based on different ways of implementing the dose equation:

$$Dose = \frac{Energy\ Absorbed}{Unit\ Mass}$$

- **Average Dose (Whole Crystal) [AD-WC]** this is the average of the dose values for the whole crystal volume, i.e. the total absorbed energy over the total mass of the crystal. It is a crude metric that does not reflect the fact that the vast majority of the data may have been collected from one highly damaged region, while some of the volume of the crystal may not have been irradiated at all. Changes in beam shape or collection strategy for the same exposure time, beam energy, and crystal composition should not result in large effects on this value since the total absorbed energy will be approximately constant. It is only meaningful

Metric	Description
Average Dose (Whole Crystal) [AD-WC]	The average dose within the whole crystal volume, irrespective of exposure patterns.
Average Dose (Exposed Region) [AD-ER]	The average dose within all the voxels with non-zero exposure.
Average Dose (95% of Absorbed energy threshold) [TAD]	The average dose within the smallest region to enclose 95% of the total absorbed energy.
Max Dose	The maximum dose value within the whole crystal volume.
Diffraction Weighted Dose [DWD]	A weighting scheme for dose that describes the average dose of the regions from which diffraction has taken place.
Dose Contrast	The ratio of the Max Dose to the TAD.
Used Volume	The fraction of total voxels with non-zero dose.
Elastic Yield	The total number of photons elastically scattered during the exposure.
Dose Inefficiency	The ratio of Max Dose to Absorbed energy.
Diffraction Efficiency	The ratio of total elastic yield to DWD. i.e. photons elastically scattered per MGy of dose.

Table 3.2: Summary of the dose metrics output by RADDOSE-3D, described in full in section 3.3.

if the crystal is smaller than the beam, and is then a good metric if the beam has a top-hat profile. This is reported for the whole crystal after each wedge, and thus increases monotonically with exposure, since the denominator is fixed and the numerator can only increase.

- **Average Dose (Exposed Region) [AD-ER]** this is the average dose calculated with the same absorbed energy as for the AD-WC, but with the mass in the denominator of the dose equation calculated from the volume swept out by the beam, as defined by all voxels with a non-zero dose. This is reported for the whole crystal after each wedge. This can both increase and decrease with each wedge, since both the denominator and the numerator can increase, as new regions are exposed, and more energy is absorbed respectively.

- **Average Dose (95.0 % of total absorbed energy threshold) [TAD]** this is the average dose

for the region defined by the isodose surface which encloses 95% of the absorbed energy. This has the advantage of leaving out the very weakly exposed regions arising from the tails of the Gaussian beam profile, but the absolute value is clearly sensitive to the percentage (here 95%) chosen for the absorbed energy threshold. This can both increase and decrease with exposure time, since the denominator is fixed as 0.95 of the total absorbed energy, and the numerator is a function of the isodose threshold, the crystal size and density, the beam FWHM, flux and energy, and the rotation and translations of the crystal during the data collection, which all contribute to where the 95% absorbed energy isosurface lies.

Comparing these three average dose metrics, it is noted that there is a constant ratio between the numerators of the equations defining them, and all have similar values (1:1[9]:0.95). The main difference between them is in the denominator of the dose equation (the mass term). For AD-WC this is only a function of crystal size and density, for AD-ER, this is also a function of how the beam area is defined, and of any rotation or translation of the crystal during exposure.

- **Max Dose** this is the highest value of the dose reached anywhere in the crystal (i.e. voxel with the highest dose). It is a useful metric since very high maximum doses can have a large negative impact on data quality and can significantly increase crystal non-isomorphism (Ravelli and McSweeney, 2000; Ravelli *et al.*, 2002; Murray and Garman, 2002). This is reported for the whole crystal after each wedge, and can thus only stay the same or increase with new wedges. It is the value reported by RADDOSE v1-3 for a Gaussian beam, calcu-

[9] Assuming the exposed region fully encloses the beam.

lated by finding the expected peak flux density based on total flux, collimation, and Gaussian FWHM of the beam.

3.3.2 Diffraction Weighted Dose [DWD]

Any way of surmising a dose field is effectively a weighting scheme: AD-WC assigns the same weight to all voxels in the whole crystal, and Max Dose assings a weight of 1 to the highest dose voxel, and 0 to all other voxels. In order to fully utilise the time and space resolved description of dose provided by RADDOSE-3D, we propose a new metric, based on a more sophisticated weighing scheme, against which to quantify radiation damage: Diffraction Weighted Dose. This metric combines information from the dose state of the crystal at a given time with the way the crystal is being exposed at that moment. This is necessary because where a crystal is damaged and from where data are collected do not necessarily colocalise. For example, a crystal may have half its volume highly damaged from a previous data set, then have a subsequent data set collected from a fresh region, and yet in each case, average dose metrics would only reflect the final dose state of the crystal, not its history. For this simple case, it would be possible to consider the crystal as two separate halves, but in the general case, wedges may overlap, or be collected in the same region several times, and it is thus necessary to consider how much fluence each part of the crystal receives, and how damaged each of these parts is.

Diffraction Weighted Dose is calculated for each angular iteration by weighting the dose in each voxel that makes up the whole crystal volume by how much it diffracts during the iteration be-

ing considered. The dose is then updated, and the next iteration considered. This allows DWD to bridge the gap between the dose state of the crystal at a given time, and the way in which data are collected. Essentially, DWD represents the effective average dose that is observed in the diffraction pattern. This is the average dose in the crystal volume from which the photons making up a given image have scattered.

In an MX experiment, this can be described as a function of the dose state of the crystal and the spatial distribution of exposure over a short angular range. Mathematically, we define the Diffraction Weighted Dose (DWD) for the angular range $\phi_{i-1} \rightarrow \phi_i$ as:

$$DWD = \frac{\int_{\phi_{i-1}}^{\phi_i} d\phi \iiint_{crystal} d\mathbf{x}\, D(\mathbf{x},\phi) F(\mathbf{x},\phi) \eta(D(\mathbf{x},\phi))}{\int_{\phi_{i-1}}^{\phi_i} d\phi \iiint_{crystal} d\mathbf{x} F(\mathbf{x},\phi) \eta(D(\mathbf{x},\phi))} \tag{3.1}$$

where ϕ is the angle of the rotation axis (the experimental coordinate, which is proportional to time for a constant rotation rate), $D(\mathbf{x},\phi)$ and $F(\mathbf{x},\phi)$ are the total cumulative dose (MGy) and flux density (photons/unit area/radian) respectively at position x, and experimental coordinate ϕ.

Finally, $\eta(D(\mathbf{x},\phi))$ is the relative diffraction intensity, described by a Dose Decay Model (DDM), a function of the dose which returns the current diffraction intensity as a fraction of the initial diffraction intensity. The Dose Decay Model should be parametrised in terms of the intensity integrated over all resolution shells, since each real-space position contributes to all the diffraction spots seen on the detector. This provides a theoretically more satisfying model of the 'effective contribution' of a given voxel than simply weighting by the fluence, since highly damaged regions of

the crystal will be downweighted by η due to their lower diffracting power relative to undamaged voxels.

Thus $F(\mathbf{x}, \phi)\eta(D(\mathbf{x}, \phi))$ is the weighting function, which corresponds to how much the vector position $\mathbf{x}$ contributes to the diffraction pattern at position ϕ. The numerator is then normalised by the flux density and the DDM, normalising the weighting operation, leaving the units of DWD as MGy, unchanged from the units of dose. It is important to note that this is a proposed weighting scheme, and that although it is based on physical considerations of the contribution to diffraction from the voxels for each image, it is not the only possible way of weighing voxels in order to optimally surmise the dose state of the crystal.

This integral is implemented as a summation in the code where angular ranges become angular steps at the defined angular resolution, and the volume integral becomes a summation over all the voxels:

$$DWD^i = \frac{\Sigma_{v=allvoxels}\left(D^{i-1}(v) + \Delta D^i(v)/2\right) \times F^i(v) \times DDM\left(D^{i-1}(v) + \Delta D^i(v)/2\right)}{\Sigma_{v=allvoxels} F^i(v) \times DDM\left(D^{i-1}(v) + \Delta D^i(v)/2\right)} \tag{3.2}$$

where $\left(D^{i-1}(v) + \Delta D^i(v)/2\right)$ is the total cumulative dose at the voxel v before the current iterative step plus half the dose to be assigned for this step. This formulation allows a linear interpolation of the two previous and current steps, making the DWD more robust to the use of larger angular ranges. $F^i(v)$ is the fluence that is incident upon the voxel during the current iterative step. The code output module `OutputFluencePerDoseCSV` creates a .csv file with the DWD for each

angular step, and a histogram containing the total DDM weighted fluence (effectively the relative contribution of each dose bin to the DWD for that step) within the whole crystal for each dose range specified by the FineHist parameters in the source code (0-200 MGy in 0.1 MGy steps). Specifically, for each voxel v, a contribution $F^i(v) \times DDM\left(D^{i-1}(v) + \Delta D^i(v)/2\right)$ is added to the dose bin in which $D^{i-1}(v)+\Delta D^i(v)/2$ falls. Finally, the angular-step wise DWDs are averaged to give an average DWD for each wedge as specified in the input, and this is output in the main summary file.

3.3.3 Strategy evaluation metrics

Several metrics are output which can aid in evaluating how efficiently and evenly a data collection strategy uses the available crystal volume. These are enumerated bellow, and summarised in table 3.2:

- **Dose Contrast (Max/TAD)** this is the ratio of the Maximum Dose to the 95% Threshold Average Dose (TAD). It is a useful quick metric which indicates the range of the dose values across the region from which most of the diffraction has originated. Thus a lower, rather than a high number, implies that most of the data have come from a more homogeneously damaged region. This metric was used in Zeldin *et al.* (2013a) to compare data collection strategies, as described in Chapter 5.

- **Used Volume** this is the fraction of voxels with non-zero dose, indicating how much of the available crystal volume has been irradiated.

- **Elastic Yield** This is the total elastic scattering from the crystal for the previous wedge. It is calculated analogously to the dose calculation described in Chapter 2, but using the elastic rather than absorption cross-section. Elastic yield should be proportional to diffraction, and should thus be optimised. Of course, the proportionality is not constant since a damaged crystal will diffract less well than an undamaged one (but this could be accommodated for with the use of a DDM).

- **Absorbed Energy (this Wedge)** This is the total amount of energy absorbed by the crystal for the previous wedge, and provides a useful proxy for elastic scattering (before this was explicitly added to the code), as well as a useful, intuitive measure of how much the crystal has been exposed during a given wedge.

- **Dose Inefficiency (Max Dose/Absorbed Energy)** Minimising the Maximum Dose is not a sufficient criterion by which to judge the effectiveness of a given data collection strategy, since lowering the dose can trivially be achieved by reducing the exposure. A better metric to consider is the Maximum Dose per unit diffraction. Since, within the narrow incident X-ray energy range normally used in MX (7-15 keV), the cross section for diffraction and the cross section for absorption events (and thus the absorbed energy) maintain an approximately proportional relationship, the absorbed energy can be used as a proxy for the diffraction yield of the crystal. A 'Dose Inefficiency' metric can thus be defined as the Maximum Dose over the total absorbed energy (i.e. over the total diffracted intensity). The SI unit for this metric is 1/kg, which reflects the fact that scaling a given crystal-strategy system up

(larger crystal volume, higher total flux, unchanged flux density) will increase the amount of diffraction for the same maximum dose. This is intuitive since a bigger crystal illuminated with a bigger beam would be expected to diffract more than if both are smaller. A lower Dose Inefficiency implies that for each mJ absorbed, the peak dose in the crystal is lower, and so the strategy is deemed better. Thus halving the Dose Inefficiency means that the maximum dose is halved for a given amount of absorbed energy (and thus to 1st order, also for the same amount of diffraction). This metric was used in Zeldin *et al.* (2013a) to compare data collection strategies, as described in Chapter 5.

- **Diffraction Efficiency (Elastic Yield/DWD)** This is a metric similar to Dose Inefficiency, but more sophisticated. It was developed after the publication of Zeldin *et al.* (2013a), since it required the experimental validation of DWD provided in Chapter 4 before it could be used with confidence. Because DWD is shown to be proportional to loss of average diffraction intensity (see Chapter 4), this gives an exposure-time independent measure of how many photons are elastically scattered per unit damage. *i.e.* if one strategy has a Diffraction Efficiency twice that of another, then for a given dose limit, it will provide twice as much elastic scattering, making it significantly more optimised.

3.4 A Typical MX experiment

To illustrate the workflow of using the RADDOSE-3D software, a simple 90° wedge data collection is simulated here. The input file for this is shown in code box 3.3. As described above, the crystal

block specifies that the crystal is a cuboid of size $100 \times 100 \times 100$ μm^3, and the absence of any explicit crystal composition instructions will cause the program to default to average absorption and attenuation coefficients[10]. The beam block describes a Gaussian profile with a vertical FWHM of 40 μm and a horizontal FWHM of 20 μm, with 10^{12} 12.1 keV photons per second, and an outer beam collimation of 80×40 μm^2 (vertical by horizontal). This collimation refers to the size of the slits used to collimate the beam, but for uncollimated Gaussians, setting this to twice the beam FWHM will speed up the run-time of the code, while only introducing an error that is much smaller than the likely fidelity of the true beam profile to a perfect Gaussian. Finally, the wedge block defines a $90°$ data collection over 30 seconds that will be processed in default angular steps of $2°$. It is only worth changing the $2°$ default angular resolution in the case of very small wedge data collection (e.g. for repeated $5°$ wedges used to probe relative B-factors). Code box 3.4 gives the summary output file for this simulation, and figure 3.4 shows a dose map created from the DoseState.csv file. As can be seen both visually from the dose map, and through the variation in the dose metrics in the summary file, this is a highly inefficient use of the crystal volume, with a Maximum Dose of 12.9 MGy, and an AD-WC of 1.2 MGy. The TAD of 3.7 MGy represents the average dose within the region that has absorbed 95% of the energy, and so offers an average that effectively weights the weakly diffracting tails of the Gaussian beam profile to zero. This large variation serves as a powerful example of the need for 3D dose simulations to optimise the use of the available crystal volume, and of effecitve dose metrics in the field of radiation damage research.

[10] For 12.4 keV.

Code box 3.3 Simple input file describing a 90° wedge of data, for which the final dose field is visualised in figure 3.4.

```
Crystal
Type Cuboid
Dimensions 100 100 100

Beam
Type Gaussian
Flux 1e12
FWHM 40 20
Energy 12.1
Collimation Rectangular 80 40

Wedge 0 90
ExposureTime 30
```

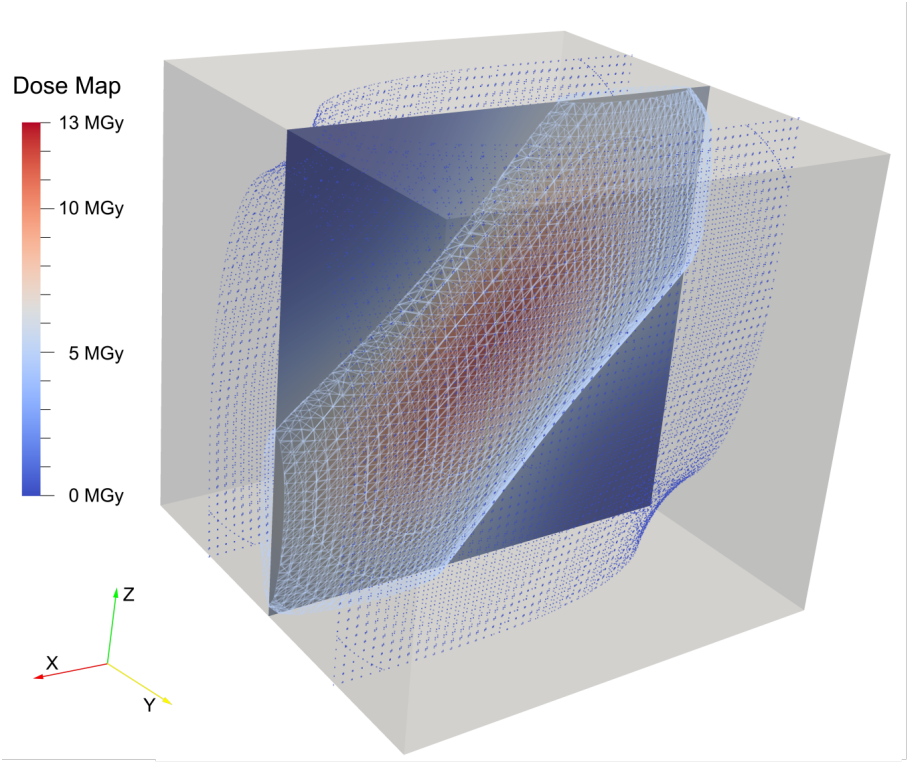


Figure 3.4: Dose map for the input configuration described by code box 3.3. Iso-surfaces are shown at 0.05 and 5 MGy, and a slice map is shown in the plane perpendicular to the rotation axis and along the beam axis. The crystal has been rotated 90° about the y-axis, with the beam starting down the z-axis, and ending along the x-axis, as labelled in the figure. The figure was generated from the output-dosestate.csv file using Paraview (http://www.paraview.org/)

Code box 3.4 Summary output file for the input shown in code box 3.3. Note the large spread between the Maximum Dose (12.9 MGy) and the Average Dose – Whole Crystal (1.2 MGy), and the inefficient use of the crystal volume: only 40% has a non-zero dose, and, as can be seen from the dose histogram, most of this is within the 0.1-3.4 MGy bin, so is weakly exposed.

```
Cuboid Crystal of size [100, 100, 100] um [x, y, z] at a resolution of 2.00
microns per voxel edge.
Simple Dose Decay Model used. Diffraction intensity does not decay with
increasing dose.
Gaussian beam, 80.0x40.0 um with 40.00 by 20.00 FWHM (x by y) and 1.0e+12
photons per second at 12.10 keV.
Wedge 1:
Collecting data for a total of 30.0s from phi = 0.0 to 90.0 deg.
Dummy crystal coefficients used.
Absorption coefficient: 2.37e-04 /um.
Attenuation coefficient: 2.81e-04 /um.
Elastic coefficient: 1.80e-05 /um.
Density: 1.20 g/ml.
Average Diffracted Dose           : 2.94 MGy
Average Dose (Whole Crystal)      : 1.17 MGy
Average Dose (Exposed Region)     : 2.89 MGy
Elastic Yield                     : 5.82e+10 photons
Max Dose                          : 12.865 MGy
Average Dose (95.0 % of total absorbed energy threshold (0.60 MGy)): 3.65 MGy
Dose Contrast (Max/Threshold Av.) : 3.53
Used Volume                       : 40.4%
Absorbed Energy (this Wedge)      : 1.49e-03 J.
Dose Inefficiency (Max Dose/mJ Absorbed) : 8.7 1/g
Final Dose Histogram:
Bin  1,  0.0 to  0.1 MGy:  6.8 %
Bin  2,  0.1 to  3.4 MGy: 58.9 %
Bin  3,  3.4 to  6.7 MGy: 23.0 %
Bin  4,  6.7 to 10.1 MGy:  9.0 %
Bin  5, 10.1 to 13.4 MGy:  2.2 %
Bin  6, 13.4 to 16.7 MGy:  0.0 %
Bin  7, 16.7 to 20.0 MGy:  0.0 %
Bin  8, 20.0 to 23.4 MGy:  0.0 %
Bin  9, 23.4 to 26.7 MGy:  0.0 %
Bin 10, 26.7 to 30.0 MGy:  0.0 %
Bin 11, 30.0 MGy upwards:  0.0 %
```

3.5 Two further examples

3.5.1 Minimal input file

In its simplest form, users define only the crystal dimensions to create a crystal, using the average absorption coefficients specified above. The simplest beam is a uniform intensity Top-Hat shaped beam, with specified flux, energy, and rectangular collimation. The wedge can be defined simply by the beginning and end angles, and the total irradiation time. The input and output files for this simple case are shown in code boxes 3.5 and 3.6 respectively, and an iso-surface map of the resulting dose field with a 0.1 MGy threshold is shown in figure 3.5. Since, in this case, the crystal is smaller than the beam, the average doses for the whole crystal and for the exposed region are the same. Dose contrast is low, and the rest of the metrics are tightly bunched, with the variation being due to absorption of the beam along the beam direction in the crystal leading to slight inhomogeneity in the dose distribution.

Code box 3.5 Minimum input file for RADDOSE-3D. Full details of input grammar can be found in Appendix 1. This describes the shortest type of input that can be used for RADDOSE-3D.

```
Crystal
Type Spherical
Dimensions 100
Beam
Type TopHat
Flux 2e12
Energy 12.4
Collimation Rectangular 120 120
Wedge 0 360
ExposureTime 100
```

Code box 3.6 Summary output file for the input shown in code box 3.5. As expected from a Top-Hat beam, the average dose is highly even.

```
Spherical crystal of diameter 100 um, at a resolution of 2.00 um per voxel edge.
Top hat beam, 120.0 by 120.0 (x by y) um with 2.0E12 photons per second at 12.4 keV.
Wedge 1:
Collecting data for a total of 100.0s from phi = 0.0 deg to phi = 360.0 deg.
Dummy crystal coefficients used.
Absorption Coefficient: 2.37e-04 /um.
Attenuation Coefficient: 2.81e-04 /um.
Density: 1.20 g/ml.

Average Dose (Whole Crystal)      : 5.39 MGy
Average Dose (Exposed Region)     : 5.39 MGy
Average Dose (95.0 % of total absorbed energy threshold (5.12 MGy)): 5.12 MGy
Max Dose                          : 5.448 MGy
Dose Contrast (Max/Threshold Av.) : 1.06
Used Volume                       : 100.0%
Total Absorbed Energy             : 3.38e-03 J.
Final Dose Histogram:
Bin  1,  0.0 to  0.1 MGy:  0.0 %
Bin  2,  0.1 to  3.4 MGy:  0.0 %
Bin  3,  3.4 to  6.7 MGy: 100.0 %
Bin  4,  6.7 to 10.1 MGy:  0.0 %
Bin  5, 10.1 to 13.4 MGy:  0.0 %
Bin  6, 13.4 to 16.7 MGy:  0.0 %
Bin  7, 16.7 to 20.0 MGy:  0.0 %
Bin  8, 20.0 to 23.4 MGy:  0.0 %
Bin  9, 23.4 to 26.7 MGy:  0.0 %
Bin 10, 26.7 to 30.0 MGy:  0.0 %
Bin 11, 30.0 MGy upwards:  0.0 %
```

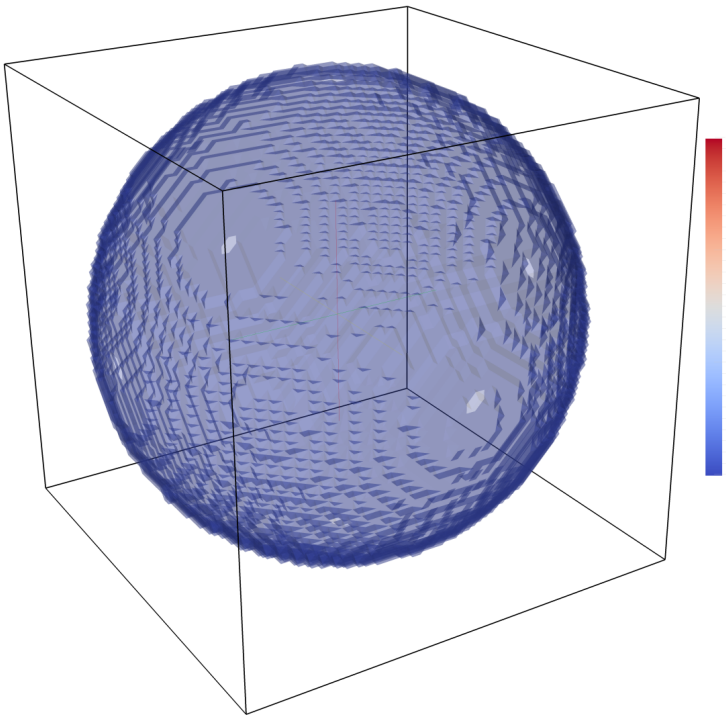


Figure 3.5: Dose isosurface for an evenly exposed spherical crystal completely bathed in a uniform Top-Hat beam, as modelled using the minimum input file described in code box 3.5. The dose field is, as expected for a case when the beam is larger than the crystal, even throughout the crystal volume. The figure was generated from the output-DoseState.csv file using Paraview (http://www.paraview.org/).

3.5.2 Complex input file

A more complex input file, simulated here, has already been shown in figure 3.2. It demonstrates some of the more advanced input options detailed in Appendix 1, and illustrates how the program processes multiple wedges, each spatially separated along the crystal. This simulation is made up of two wedges with a first beam specification, then two wedges with a different second beam. The first pair of wedges compares a standard and helical data collection, and the second pair a standard and offset data collection. The offset strategy describes cases when the rotation axis and the beam axis are deliberately offset in order to spread the high dose region – but not the total exposed region – across a larger crystal volume. Offset strategies can be highly effective in reducing the maximum dose for a given diffraction yield, and this is investigated further in Chapter 5.

An iso-surface rendering of the final dose field for this simulation is shown in figure 3.6, and the main metrics are presented wedge by wedge in table 3.3. As described above, Average Dose (Whole Crystal), Max Dose, and TAD are calculated over the whole crystal volume after each wedge, and absorbed energy only refers to the specified wedge. Due to the cumulative nature of multiple wedge simulations, this is not an appropriate data set to directly contrast different dose distribution strategies. Comparative simulations for the helical and standard strategies are presented in Chapter 5.

As expected, the Average Dose (Whole Crystal) metric increases proportionally to the absorbed energy with each wedge, but the more sophisticated metrics that segment the crystal volume (ex-

posed region, TAD) do not increase in the same way. The TAD and Average Dose (Exposed Region) remain approximately constant, since both the numerator (absorbed energy) and the denominator (mass) of the dose equation increase as wedges are added. The Max Dose keeps a running tally of the highest dose voxel: first for the centre of the less intense beam's standard data collection from wedge 1, then from the centre of the second, more intense beam's standard collection hot-spot from wedge 3. The Dose Contrast, being the ratio of Max Dose to TAD, increases slightly for the first pair, and is then roughly constant for the second pair of wedges. As expected, used volume increases as more wedges are collected from the same crystal. The increase is larger when the helical method has been used, and proportional to beam size for all other wedges. Because Dose Inefficiency tracks the overall maximum dose and the wedge-wise absorbed energy, it is of limited use for cases where more than one wedge is collected from a single crystal, and is not shown, since it could use the maximum dose from a previously exposed part of the crystal, and the total absorbed energy from the most recent wedge. The DWD reveals a much reduced dose with the helical or offset strategies relative to their respective standard strategies. Finally, the Diffraction Efficiency shows (1) that the larger beam used for wedges 1 and 2 is overall more efficient at spreading dose, and that, relative to each standard strategy, both the helical and offset strategies increase the number of elastically scattered photons per MGy (DWD), with the helical one being the most effective. This can be explained by the fact that the helical scan uses up a much larger total crystal volume (since it translated during data collection) than the offset. The offset, however, offers improvement over standard without any increase in crystal volume usage.

Since these different strategies were applied to the same crystal successively, it is not appropriate to compare their effectiveness in reducing dose contrast here. Comparisons of several dose spreading strategies are made, in detail, in Chapter 5.

Wedge	Notes	DWD [MGy]	Elastic Yield [ph]	Diffraction Efficiency [ph/MGy]	AD-WC [MGy]	AD-ER [MGy]	Max Dose [MGy]	TAD [MGy]	Dose Contrast	Used Volume
1	Beam 1, Standard	16.1	8.04E+10	4.98E+09	3.4	12.8	93.5	22.1	4.2	26.5 %
2	Beam 1, Helical	8.4	8.05E+10	9.60E+09	6.8	11.3	93.5	17.7	5.3	60.2 %
3	Beam 2, Standard	17.3	3.50E+10	2.03E+09	8.3	12.2	114.7	17.7	6.5	67.7 %
4	Beam 2, Offset	10.9	2.77E+10	2.55E+09	9.4	12.7	114.7	17.5	6.5	74.4 %

Table 3.3: Summary of the outputs from the input defined in figure 3.2. All metrics except for DWD, Elastic Yield and Diffraction Efficiency are calculated over the whole crystal volume after each wedge. This means that the absorbed energy numerator of the dose equation always increases monotonically, but that the denominator can either stay constant (AD-WC, Max Dose), or increase as more crystal is exposed (AD-ER, TAD), making them unsuitable for comparing wedges. Note how the helical and offset strategies offer lower DWD and higher Diffraction Efficiency relative to their respective standard strategies. This is not reflected in the 'whole crystal' metrics, since they are analysing the cumulative sum of the dose field, *e.g.* for the first two wedges, wedge 1 describes the dose state after first standard wedge, and wedge 2, the cumulative dose state after the first and second wedges. See the main text for a metric-by-metric description of the simulation.

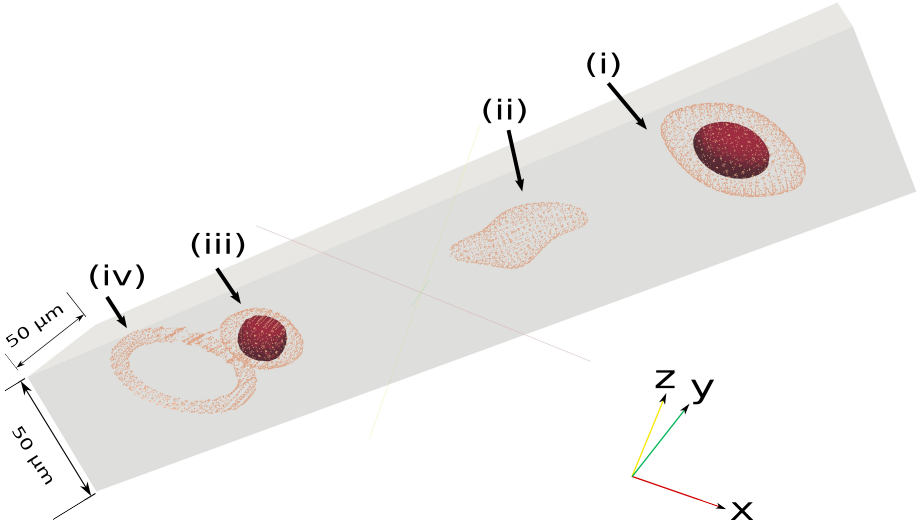


Figure 3.6: Dose isosurface map for the complicated example described by the input file in figure 3.2. The solid red isosurface represents a 60 MGy (very high dose) surface, and the dotted red surface represents the Experimental Dose limit of 30 MGy (Owen et al., 2006). This crystal was oriented at 45° to the goniometer axis (y axis in figure) in the plane normal to the beam axis. All wedges were for 360° rotations. The rightmost two wedges, with a first beam, show a (i) standard wedge, and a (ii) helical wedge. Here helical scanning was defined along the (1,1,0) vector, to allow for the orientation of the crystal in the loop. Improvements in dose contrast for this method are described in detail in Chapter 5. The two leftmost wedges, both simulated with a different beam to the first two simulations, contrast (iii) a standard strategy and (iv) an offset strategy. The offset strategy offsets the beam and rotation axes, so that they do not intersect. This leads to a lowering of the maximum dose for a given average dose. The figure was generated from the output-dosestate.csv file using Paraview (http://www.paraview.org/)

3.6 Online implementation

The website www.raddo.se, SVN repository (http://subversion.apache.org) for version control, online Apache ANT compilation (http://ant.apache.org/), Jenkins continuous integration server (http://jenkins-ci.org/), automated Checkstyle (http://checkstyle.sourceforge.net/), automated Javadoc generation, and automated regression testing using TestNg (http://testng.org/) were implemented by MG. The PhP interface at http://www.raddo.se/php/ was designed and coded by Jean-Paul Ebejer from the Department of Statistics, University of Oxford and also MG.

3.7 Conclusions

A new tool has been created to simulate the macroscopic properties of the dose absorbed during a diffraction experiment. Aggregate metrics to describe the effect of proposed data collection strategies on radiation damage have been suggested, and examples of the use of the code have been presented. The next step was to experimentally investigate the relationship between these metrics and the observed diffraction quality, and this is the subject of Chapter 4.

Chapter 4

Predicting Damage in the Uneven-dose Regime

The previous two chapters have established the tools to simulate 3D dose distributions, and suggested metrics to summarise these into appropriate descriptions of the damage state of a crystal at a given time. This chapter presents an experimental investigation into the effectiveness of these metrics at modelling damage, with the goal of developing a protocol for predictively modelling the lifetime of protein crystals, allowing experimentalists to avoid failed data collections caused by radiation damage.

This was achieved by using broadly the same approach as Owen *et al.* (2006), where multiple datasets are collected from the same crystal volume. I_n/I_1 and B_{rel} metrics were then calculated from the datasets, and plotted against the various dose metrics. The goal was to find a dose metric which was proportional do the damage metric, allowing predictive modeling to be performed.

4.1 Experimental design

In order to test the putative dose metrics presented in the last chapter, sets of crystals must be compared under several dose contrast regimes. The requirement is for a dose metric which faithfully reflects the damage observed in the data under a wide variety of contrast strategies. Because of the intrinsic variability of crystals, five crystals were used for each dose contrast regime, and, initially, two dose contrast regimes were explored. A third, intermediate regime was then used to provide a validation of the initial conclusions. Data were collected from each crystal using an initial low-dose $Probe$ data-set, then 10 sets of $Burn + Probe$, where $Burn$ was a high dose exposure with no data collected.

Accurate calculations of dose fields are fundamental to this type of experiment. This requires experimental characterisation of the beam and of the crystal, as well as the computational implementation of this experimental information. Ideally, a combination of quantitative 2D beam profiling and a 3D crystal tomogram would be used. Although tomography, and other experimental techniques such as online confocal microscopy have been, and are being, developed (Brockhauser *et al.*, 2008; Khan *et al.*, 2012b) for use in MX, such measurements are currently far from routine. The implementation of these data into RADDOSE-3D is quite involved, and is beyond the scope of this thesis. An alternative is to use crystals that can be accurately characterised using only the online microscope at a beamline. This requires the preparation of crystals that are faithfully cuboid, and the ability to orientate them into a position that can be described by the RADDOSE-3D `CrystalCuboid` class; this re-orientation can most easily be achieved by a beamline equipped with a multi-axis goniometer, such as the EMBL/ESRF mini-kappa goniometer.

Based on the requirements for a beamline with a tunable beam size (to achieve multiple dose contrast regimes), mini-kappa goniometry, and 2D beam profiling, all experiments were conducted on the ID14-4 beamline at the ESRF, Grenoble, France in collaboration with Dr. Sandor Brockhauser, the EMBL Scientist in Charge at the beamline.

Three sets of five crystals were then used to test the dose metrics proposed earlier. In all cases, using the same beamline, crystals grown under identical conditions, and an identical data collection and processing protocol ensured that the only extrinsic variable was the dose contrast, as determined by the beam size and profile.

4.2 Methods

4.2.1 Crystallisation

Crystals of bovine pancreatic insulin were grown by John Bremridge, a Part II undergraduate Biochemistry student, with guidance from Elspeth Garman and myself. Bovine pancreatic insulin (Sigma I6634) was prepared by the hanging drop vapour diffusion method, screening phosphate precipitant (Na_2HPO_4) from 0.225 M to 0.45 M, and pH from 10.0 to 10.7, based on a protocol for growing morphologically cubic insulin established by Debreczeni *et al.* (2003). The final selected conditions were 0.425 M Na_2HPO_4 at pH 10.4 for the well solution. Protein drops were made up of 4 μl of 12.5 mg/ml insulin in 0.02M Na_2HPO_4 and 0.01M Na_3EDTA, added to 4 μl of well solution. Crystals grew in 24-36 hours. In order to select morphologically suitable crystals larger than (100 μm)^3 in size from the population of plate-like and cubic crystals, multiple trays were prepared identically, and from these, 30 suitable candidates were identified and harvested. Crystals were mounted in nylon cryo-loops (Hampton Research[1]), and cryoprotected for 2 minutes in a solution of identical concentration to that of the well solution, but with 30% v/v glycerol substituted for water, before being flash cooled in liquid nitrogen. An example of the selection of suitable crystals is shown in figure 4.1.

[1] http://hamptonresearch.com/product_detail.aspx?cid=24&sid=136&pid=381

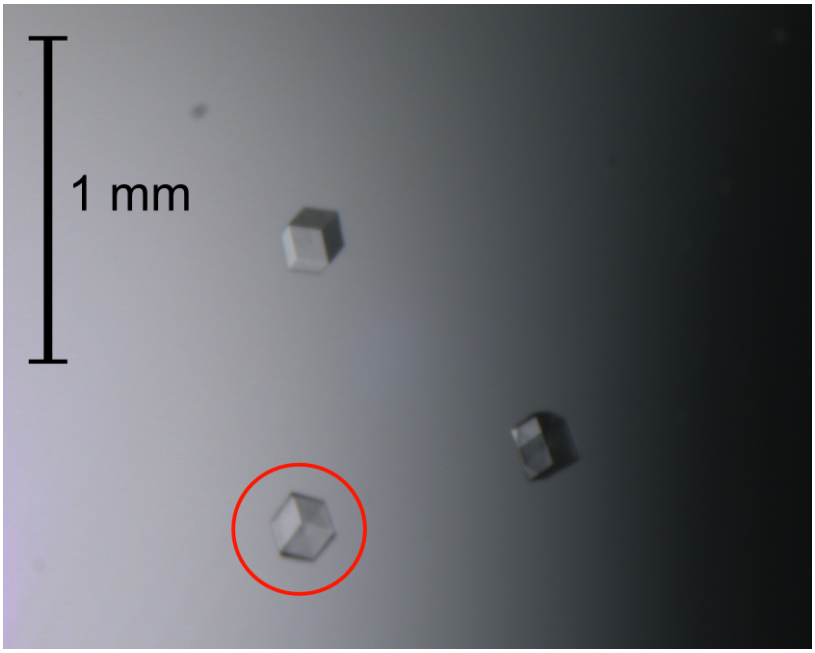
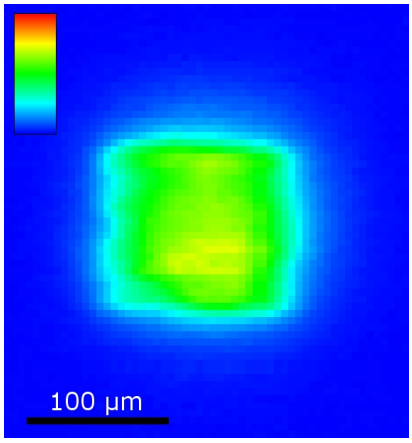


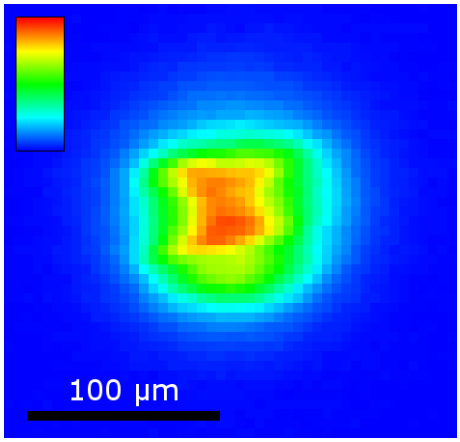
Figure 4.1: Example of a morphologically suitable crystal of cubic insulin in its crystallisation drop, image taken by John Bremridge.

4.2.2 Beam profiling

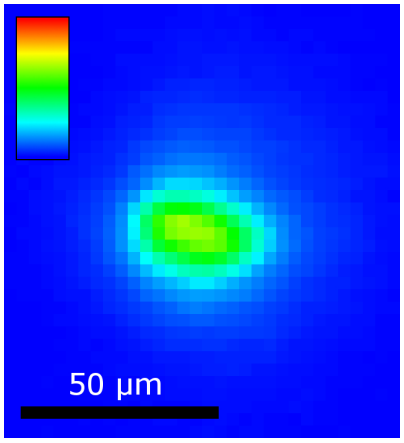
2D beam profiling was carried out on ESRF beamline ID14-4 (McCarthy *et al.*, 2009) using a high-speed online 8-bit CCD detector (SONY XC-HR50) with an effective pixel size of 5 μm, combined with a scintillator directly in the beam path, placed in front of the ADSC Q315r diffraction detector, to provide a quantitative map of the beam profile in a portable graymap file (PGM). Dark current was subtracted from the image, using the average of the background far away from the main beam centre. Representations of the resulting measured profiles for the three beam sizes used in these experiments can be found in figure 4.2.



(a) Big beam, approximately 115 × 115 μm^2 FWHM, slits set to 120 × 120 μm, dark current value: 21. Dynamic range after dark current subtraction: 0-185.



(b) Medium beam, approximately 60 × 60 μm^2 FWHM. Dark current value: 18. Dynamic range after dark current subtraction: 0-237, but only 1 pixel saturated (value after DC subtraction: 237).



(c) Small beam, approximately 25 × 40 μm^2 FWHM (vertical, horizontal). Dark current value: 21. Dynamic range after dark current subtraction: 0-222.

Figure 4.2: False colour images of the beam profiles used for the experiments presented here. Pixel size is 5×5 μm^2 in all cases, and the colour scale is from 0-255 (blue-red), and is the same on all the images. Images were subsequently scaled to the full flux, as measured by the in-line calibrated diode I_1 (see figure 4.3). Beam features are due to a multiple factors such as collimation settings and mirror slope errors. The beam was focused by adjusting the upstream mirrors and narrowing the collimation slits.

Absolute values of the flux were obtained using the diode I_1, which is located downstream of the slits and the shutter (see figure 4.3). I_1 is a silicon diode which measures the flux indirectly *via* the current produced by electron emission from a 25 μm aluminium foil placed in-line at $45°$ to the beam (McCarthy *et al.*, 2009). This signal is then amplified, and the output values are converted to photon fluxes using a calibrated[2] voltage-flux curve. The flux values from the I_1 diode were verified against a previously calibrated 500 μm silicon diode owned by the Garman group (Owen *et al.*, 2009) placed in the sample position, and found to be in good agreement. Measurements of I_1 were subsequently recorded both before and after each crystal was irradiated, and experiments were timed so that they did not take place during a storage ring top-up, ensuring consistent levels of flux throughout.

In order to accurately apply these profiles to the dose calculation, a new implementation of the `Beam` interface class was written for RADDOSE-3D. This class takes an input PGM file, pixel size, and total beam flux, and then uses bilinear interpolation to calculate the flux density at an (x, y) coordinate, as required by the `beamIntensity(x, y)` method of the `Beam` interface, detailed in Chapter 3.

4.2.3 Crystal modelling and alignment

With the optimised crystal morphologies grown as described in section 4.2.1, it was possible to obtain cuboid crystals suitable for accurate modelling in RADDOSE-3D. However, it is still very difficult to reliably ascertain the crystal orientation in terms of the P and L angles described in

[2] Previously calibrated in-house against a 300 μm silicon diode.

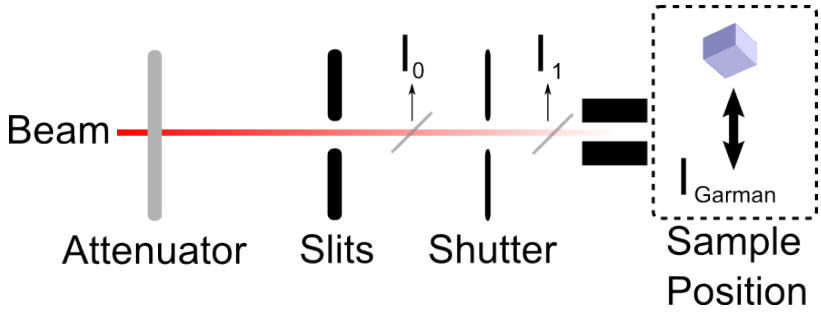
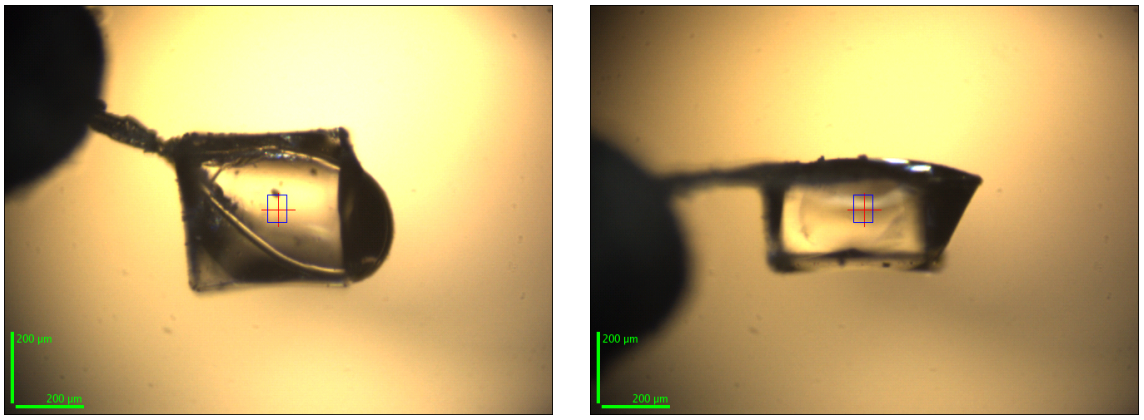


Figure 4.3: Arrangement of diodes at ID14-4. I_1, previously calibrated in-house, was checked against I_{Garman}, and found to be in good agreement. The 300 μm collimator after I_1 is shown, but not labelled in the figure for clarity.



(a) Face on. (b) Side on.

Figure 4.4: Online microscope image of well-aligned cuboid crystal 191. Thanks to the crystal morphology and mini-kappa goniometry, it is possible to accurately parametrise the shape of the crystal in RADDOSE-3D. Note the small triangular wedge of solvent on the right.

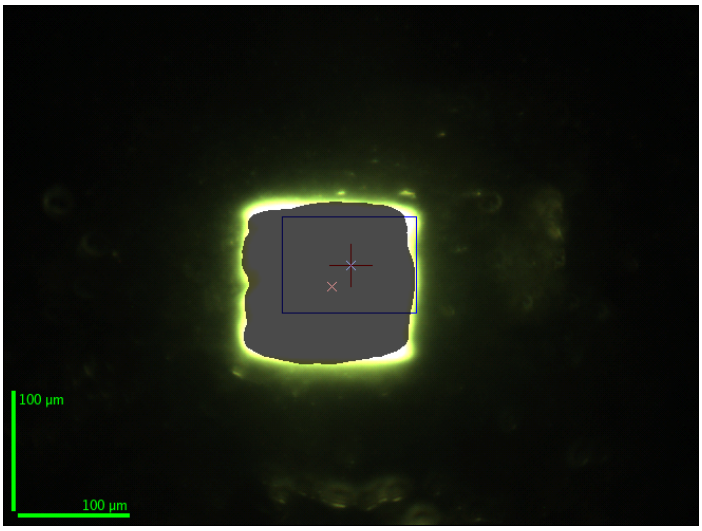
Appendix 1. By using the mini-kappa goniometer available on ID14-4, and the STAC software (Brockhauser *et al.*, 2011) it was possible to reorientate the crystals either so that they were face on at 0° ($P = L = 0^\circ$), or so that the orientation angles could be determined by inspection of several online microscope images. An example of a crystal of well known orientation in the loop is shown in figure 4.4. The crystal sizes, along with the P and L angles for all crystals, are shown in table 4.1.

Because the beam centroid does not necessarily correspond with the central pixel from the PGM image of the beam (*i.e*, the profiling CCD is not accurately centred to the beam axis), it

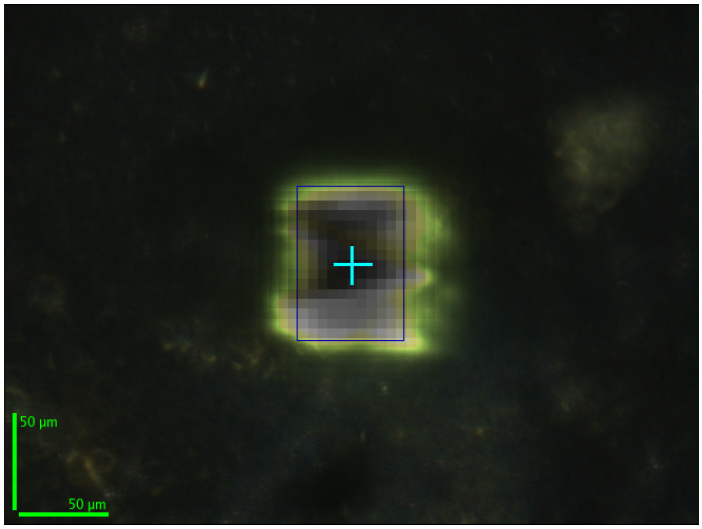
Beam	Crystal ID	Size $x \times y \times z$ (μm ^{3})	P	L
Big	P2S5	$150 \times 150 \times 80$	$5°$	$10°$
	P3S1	$166 \times 120 \times 167$	$4°$	$0°$
	P3S3	$150 \times 150 \times 90$	$5°$	$0°$
	P3S9	$189 \times 209 \times 97$	$0°$	$0°$
	P3S10	$126 \times 126 \times 110$	$0°$	$0°$
Mid	191	$420 \times 460 \times 270$	$0°$	$0°$
	225	$550 \times 570 \times 315$	$30°$	$0°$
	2004	$75 \times 73 \times 97$	$45°$	$0°$
	2012	$162 \times 250 \times 120$	$0°$	$0°$
	2019	$396 \times 417 \times 213$	$60°$	$45°$
	2024	$441 \times 447 \times 210$	$15°$	$30°$
Small	55	$180 \times 180 \times 95$	$0°$	$45°$
	197	$171 \times 190 \times 100$	$0°$	$0°$
	5185	$180 \times 180 \times 100$	$0°$	$0°$
	P1S10	$130 \times 130 \times 130$	$35°$	$45°$
	P3S2	$140 \times 140 \times 90$	$50°$	$35°$

Table 4.1: Crystal sizes and orientations for all three beam conditions. Crystal sizes were estimated using the online microscope, and have errors of $\sim$5 μm in the x and y directions (face on), and $\sim$10 μm in the z direction, since for this axis, the loop makes measurement difficult. Orientation estimations are correct to $\sim 10°$. Although these errors are not small, doses for non-translational data collections are not very sensitive to these effects, since the exposed crystal volume does not change significantly with P and L. Furthermore, these errors could be overcome with the introduction of improved crystal profiling.

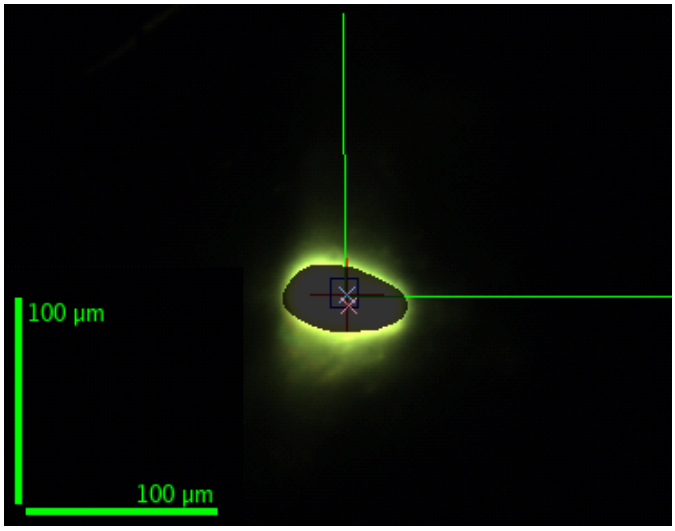
was necessary to perform a coordinate translation on the RADDOSE-3D simulations *via* the $YStartOffset$ and the $RotAxBeamOffset$ keywords. This was achieved by aligning an image of the beam profile taken from a YAG crystal placed in the sample position using the on-axis microscope, which shows the qualitative beam shape, with the quantitative PGM profile. The images were scaled by using the known size of the pixels in the PGM, and the known size of the online scale bar from the YAG image (Gabadinho *et al.*, 2010). The offset between the central pixel and the alignment cross from the online microscope image (if any) was then used to adjust the origin of the beam in the PGM file away from the middle pixel, as interpreted by RADDOSE-3D. This procedure, and the compensating offsets, are shown for the three beams in figure 4.5.



(a) Big beam alignment. Offset in y (horizontal) applied: -20.0 μm Offset in x (vertical) applied: -17.5 μm.



(b) Medium beam alignment. No Offset required.



(c) Small beam alignment. Zero offset in y (horizontal) applied, Offset in x (vertical) applied: -9.6 μm.

Figure 4.5: Alignment of images to establish the relationship between the central pixel of the PGM images, and the beam axis used to centre the crystal.

4.2.4 Data collection strategies

For each beam condition, the same general strategy was employed: complete ($\Delta\Phi = 60°$), low-dose, datasets were used to probe the damage state of the crystal, and high-dose burn sets were used between each set. This took the form of $Probe + 10 \times (Burn + Probe)$. Details for the $Burn$ and $Probe$ protocols are presented in table 4.2.

Beam	Typical total flux [ph/s]	$Probe$	$Burn$
Big	2×10^{12}	60 $1°$ images, 0.1 s/image, 20% beam transmission.	$360°$ exposure over 150s at 100 % transmission.
Medium	10^{12}	60 $1°$ images, 0.1 s/image, 100% beam transmission.	$360°$ exposure over 90s at 100 % transmission.
Small	1.5×10^{11}	60 $1°$ images, 0.1 s/image, 100% beam transmission.	$360°$ exposure over 150s at 100 % transmission.

Table 4.2: Strategy used for the $Probe$ and $Burn$ data collections. Each $Probe$ wedge was taken over the same $60°$. The variations in the total fluences for the $Probe$ sets ensured sufficient data quality for analysis up to high doses. The data for the Medium beam were collected at a later date than those for the Large and Small beam, and so the fluence for the $Burn$ steps was adjusted using the DWD and the $D_{1/2}$ from the other two conditions to create optimally dose spaced data-sets. Dose per dataset is not quoted, since it does not inform on the total diffraction intensity from the crystal, whereas incident fluence does.

4.2.5 Dose calculation

Based on the above parameters, doses were calculated using RADDOSE-3D, with the absorption cross-sections being calculated by RADDOSE v2. An example input file for a single crystal is given in code box 4.1

Code box 4.1 Input file for RADDOSE-3D used to calculate dose for the $Probe - Burn$ strategy

```
# DoseDistr
# Med Beam mx1414 FWHM 40 25
# 197

Crystal
Type Cuboid
DiffractionDecayModel Simple
Dimensions 171 190 100
PixelsPerMicron 0.5
CoefCalc RDV3
NumMonomers 24
NumResidues 51
ProteinHeavyAtoms Zn 2 S 6
SolventHeavyConc P 425
UnitCell 78.02 78.02 78.02
SolventFraction 0.6436

Beam
Type ExperimentalPGM
File medb-15nov-21dc.pgm
PixelSize 5 5
Flux 1.48e11
FWHM 25 40
Energy 13.2
Collimation Rectangular 120 120

# Collect Burn Strategy

Wedge 0 60
ExposureTime 6
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 360
ExposureTime 150
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 60
ExposureTime 6
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 360
ExposureTime 150
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 60
ExposureTime 6
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 360
ExposureTime 150
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 60
ExposureTime 6
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 360
ExposureTime 150
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 60
ExposureTime 6
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 360
ExposureTime 150
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 60
ExposureTime 6
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 360
ExposureTime 150
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 60
ExposureTime 6
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 360
ExposureTime 150
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 60
ExposureTime 6
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 360
ExposureTime 150
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 60
ExposureTime 6
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 360
ExposureTime 150
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 60
ExposureTime 6
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 360
ExposureTime 150
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649

Wedge 0 60
ExposureTime 6
AngularResolution 1
StartOffset 0.0 0.0
TranslatePerDegree 0.0 0
RotAxBeamOffset -9.649
```

4.2.6 Data processing

Data were processed using the CCP4 suite (Winn *et al.*, 2011) with a standardised script[3] being used to call each program for maximum control over data processing parameters, ensuring identical treatment of all crystals. All programs not directly referenced can be found in the reference for CCP4 above. $MOSFLM$ (Leslie and Powell, 2007) was run manually using the $IPMOSFLM$ interface, with the space group set to $I2_13$ on the first $Probe$ dataset of each crystal in order to generate a $.sav$ file which was used to define the starting conditions throughout the processing of data for all datasets from that particular crystal. Postrefinement and integration were then performed for each $Probe$ dataset. After sorting with $SORTMTZ$[4], the data were integrated using $SCALA$ (Evans, 2006), with a resolution limit of 1.8 Å in all cases, despite the fact that the resolution for the low dose first $Probe$ sets was typically ~1.5 Å. This fixed resolution limit was used to allow the worsening statistics caused by radiation damage to be clearly apparent. Previous work in the group over the last 10 years has shown that fixing the resolution is important if progressive damage datasets are to be compared. This ensures that the same set of reflections are being analysed for each progressive damage dataset, and forms part of standard radiation damage data analysis protocols. After correction for negative intensity values, and conversion to structure factors using $CTRUNCATE$, an identical free R set of reflections was defined from a reference MTZ for all datasets. The full logfile for each run was then parsed using a Python script to extract the data required for systematic analysis of intensities. A small number of the high dose

[3] Provided by Edward Lowe, from the LMB Oxford. All other scripts for B_{rel}, RSCC's etc. were written by the author.
[4] Allowing the space group to be easily forced to $I2_13$ throughout

datasets failed to process when certain errors exceeded the thresholds used by $MOSFLM$ or $SCALA$. These were omitted from the subsequent analysis in order to ensure that all full data sets were processed in an identical fashion. Each low-dose dataset was phased by molecular replacement in $Phaser\ MR$ (McCoy *et al.*, 2007), using the high resolution (1.4 Å) insulin structure 2BN3 (Nanao *et al.*, 2005) from the PDB for the first crystal processed, and then using the refined structure from the first crystal processed for these experiments as a model from which to phase subsequent crystals. Each low-dose structure was then TLS and rigid body refined using $Refmac5$ (Murshudov *et al.*, 2011), and manually inspected and adjusted in $COOT$ 0.6 (Emsley *et al.*, 2010). This refinement-build cycle was repeated 2-4 times until the refinement parameters were satisfactory, defined here by R and R_{free} both being below 0.2. All subsequent damaged datasets from each crystal were assigned the low dose phases from the first $Probe$, as recommended in Weik *et al.* (2000). Difference maps were calculated using the program FFT. Relative B factors were calculated separately by merging all the output files from $CTRUNCATE$ using CAD (twice when there were more than 9 datasets successfully processed), then scaling together these internally scaled datasets using $SCALEIT$ to give relative isotropic B factors. Real space correlation coefficients were calculated using $SFCHECK$, correlated against the refined coordinates calculated for the most undamaged dataset from each crystal. R[5] was used to generate plots and to perform statistical analyses.

[5] http://www.r-project.org/

4.3 Results

4.3.1 Data collection statistics

Full statistics for all fifteen crystals can be found in tables 4.3 (big beam), 4.4 (mid beam), and 4.5 (small beam). Note that some of the statistics in high dose datasets are unphysical ($R_{meas} > 1$, negative $I/\sigma(I)$) due to the data being so degraded that the processing software was operating out of its range of convergence. The general trends in data statistics are all observed in each crystal: increasing mosaicity, unit cell dimensions, and $CC_{1/2}$, and decreasing $I/\sigma(I)$. In each case, changes occurred fastest in the high resolution shell (1.8 - 1.9 Å). The fixed resolution limit for all crystals was chosen so that statistics could be compared directly for the analysis of radiation damage. A dynamic high resolution limit, starting around the diffraction limit of the crystals (~1.3 Å), and allowed to change to a lower resolution as damage degrades the diffraction, would have yielded better absolute statistics, but at the cost of compromising the direct comparability of the samples. Changing the resolution limit would lead to different resolution bins being defined for each dataset, making the average intensity change with the number of reflections, and preventing the direct comparison of intensity per resolution bin across multiple datasets. This is evident in the somewhat high R_{meas} values found for several of the crystals. It is interesting to note that the statistics appear to get worse at higher I_n/I_1 values for the smaller beams: for the big beam, mosaicity estimation only fails at around I_n/I_1 = 0.1, but for the mid and small beams, at values up to ~0.5. It is important to note that it is not possible to draw meaningful conclusions from a

comparison of these statistics, since the fluence (and hence total diffraction) for each of the probe sets was not equal, and this will have a large effect on the statistics. This should not, however, affect the I_n/I_1 metric, which takes into account changes in diffraction quality, and so is robust to these effects.

Crystal	Probe	I_n/I_1	Unique Observations	$CC_{1/2}$	Multiplicity	$I/\sigma(I)$	R_{meas}	Unit Cell [Å]	Mosaicity [°]
P2S5	1	1.0	7513	0.998 (0.985)	7.1 (7.3)	22.3 (10.7)	0.066 (0.176)	78.14	0.13
	2	0.64	7572	0.998 (0.952)	7.1 (7.3)	17.9 (6.0)	0.072 (0.322)	78.32	0.12
	3	0.4	7613	0.998 (0.386)	7.1 (7.2)	11.6 (1.2)	0.105 (1.832)	78.46	0.42
	4*	0.088	7515	0.558 (0.017)	7.1 (7.3)	1.2 (0.1)	1.005 (25.785)	78.14	0.05
	5*	0.075	7513	0.644 (0.036)	7.0 (7.1)	0.7 (-0.1)	1.254 (-9.990)	78.14	0.05
P3S1	1	1.0	7469	0.998 (0.971)	7.0 (7.2)	20.2 (7.5)	0.069 (0.250)	78.02	0.49
	2	0.73	7545	0.998 (0.882)	7.0 (7.2)	15.9 (3.4)	0.080 (0.554)	78.31	0.51
	3	0.51	7606	0.998 (0.237)	7.0 (7.2)	11.2 (1.0)	0.107 (2.020)	78.56	0.57
	4	0.34	7616	0.998 (0.071)	6.9 (7.1)	7.6 (0.3)	0.161 (6.710)	78.60	0.65
	5	0.23	7625	0.997 (0.078)	6.8 (7.0)	5.2 (0.2)	0.253 (9.169)	78.65	0.78
	6	0.16	7683	0.993 (0.011)	6.7 (6.8)	4.1 (0.1)	0.390 (28.006)	78.77	0.91
	7	0.12	7689	0.990 (0.098)	6.5 (6.1)	2.5 (-0.4)	0.626 (-3.458)	78.89	0.98
	8	0.034	7791	0.971 (-0.013)	6.5 (6.5)	1.5 (0.1)	1.072 (58.404)	79.21	1.05
P3S3	1	1.0	7440	0.998 (0.994)	7.1 (7.2)	30.8 (17.5)	0.049 (0.105)	78.06	0.27
	2	0.78	7485	0.999 (0.987)	7.1 (7.3)	27.2 (11.7)	0.052 (0.163)	78.22	0.28
	3	0.61	7521	0.999 (0.940)	7.1 (7.3)	21.6 (5.7)	0.061 (0.372)	78.36	0.36
	4	0.5	7525	0.999 (0.706)	7.1 (7.2)	16.8 (2.5)	0.080 (0.889)	78.41	0.50
	5	0.39	7539	0.998 (0.276)	7.0 (7.1)	12.9 (1.1)	0.110 (2.296)	78.46	0.64
	6	0.3	7547	0.997 (0.031)	6.9 (7.1)	10.3 (0.5)	0.163 (5.992)	78.54	0.77
	7	0.26	7540	0.996 (-0.004)	6.9 (7.0)	8.2 (0.2)	0.223 (24.470)	78.54	0.87
	8	0.21	7587	0.994 (-0.036)	6.8 (6.8)	6.8 (0.1)	0.306 (-9.990)	78.67	0.94
P3S9	1	1.0	7492	0.999 (0.971)	7.1 (7.2)	21.1 (7.9)	0.061 (0.251)	78.04	0.26
	2	0.74	7547	0.999 (0.853)	7.1 (7.3)	16.6 (3.3)	0.073 (0.633)	78.24	0.34
	3	0.54	7569	0.999 (0.211)	7.0 (7.2)	11.1 (0.9)	0.114 (2.499)	78.32	0.49
	4	0.37	7586	0.998 (0.029)	7.0 (7.1)	8.0 (0.3)	0.181 (7.011)	78.34	0.62
	5	0.21	7596	0.996 (-0.037)	6.9 (6.9)	4.1 (0.1)	0.311 (20.043)	78.39	0.72
	6	0.14	7611	0.988 (0.023)	6.1 (5.9)	2.4 (0.0)	0.438 (-9.990)	78.46	0.73
	7*	0.092	7576	0.932 (0.015)	6.7 (7.3)	1.4 (0.0)	1.228 (34.847)	78.38	0.05
	8*	0.096	7570	0.972 (-0.001)	6.9 (7.1)	1.6 (0.1)	1.362 (26.384)	78.30	0.20
	9*	0.034	7580	0.827 (0.063)	6.6 (5.5)	0.9 (-0.4)	3.866 (-1.311)	78.38	0.05
	10*	0.016	7582	0.826 (0.032)	7.0 (7.1)	0.9 (0.0)	2.382 (-9.990)	78.38	0.05
P3S10	1	1.0	7516	0.998 (0.875)	7.1 (7.2)	15.3 (3.5)	0.093 (0.560)	78.18	0.53
	2	0.68	7582	0.997 (0.447)	7.0 (7.2)	11.0 (1.2)	0.129 (1.671)	78.45	0.66
	3	0.47	7602	0.996 (0.061)	6.9 (7.0)	6.8 (0.3)	0.230 (7.417)	78.53	0.99
	4	0.32	7599	0.990 (0.052)	6.6 (6.5)	4.4 (-0.1)	0.441 (-9.467)	78.54	1.19
	5	0.23	7505	0.988 (-0.046)	5.7 (3.5)	2.6 (-1.4)	0.596 (-0.657)	78.32	1.34
	6*	0.076	7582	0.903 (0.080)	6.5 (5.0)	0.7 (-1.0)	2.004 (-1.050)	78.48	0.05
	7*	0.015	7577	0.878 (-0.021)	7.0 (7.2)	0.6 (0.0)	2.865 (-9.990)	78.48	0.05
	8*	0.0036	7576	0.706 (-0.029)	6.9 (6.7)	0.3 (-0.1)	11.254 (-3.461)	78.48	0.05

Table 4.3: Statistics for the big beam data collection. These are highly damaged crystals, with each $Burn$ between datasets representing ~10 MGy, and so later datasets can exhibit unphysical statistics. Typical $Probe$ doses were around 30 kGy. The use of the $I_n/I_1 > 0.4$ cutoff for later data analysis ensured that data from crystals that were too damaged to be meaningful were not included. Once the data are sufficiently degraded ($I/\sigma(I) < 1.5$, I_n/I_1 0.1), $MOSFLM$ is often unable to compute a mosaicity for the crystal, so resets the value to its default, indicated by an asterix in the table. All datasets had completeness >99.2% (98.2% in the highest resolution shell). Below the dashed line for each crystal, data are deemed unreliably processed.

Crystal	Probe	I_n/I_1	Unique Observations	$CC_{1/2}$	Multiplicity	$I/\sigma(I)$	R_{meas}	Unit Cell [Å]	Mosaicity [°]
191	1	1.0	7505	0.974 (0.976)	6.9 (7.2)	12.1 (9.7)	0.158 (0.174)	78.26	0.43
	2	0.86	7550	0.980 (0.974)	6.9 (7.2)	11.9 (8.7)	0.157 (0.179)	78.40	0.43
	3	0.88	7572	0.980 (0.975)	6.9 (7.1)	12.4 (8.1)	0.145 (0.181)	78.42	0.41
	4	0.85	7597	0.982 (0.970)	6.9 (7.1)	11.9 (7.2)	0.148 (0.209)	78.51	0.46
	5	0.53	7597	0.984 (0.963)	6.9 (7.1)	11.3 (7.1)	0.147 (0.232)	78.55	0.43
	6	0.61	7598	0.985 (0.949)	6.9 (7.1)	10.8 (5.6)	0.143 (0.283)	78.54	0.46
	7	0.51	7599	0.987 (0.941)	6.9 (7.1)	11.0 (5.9)	0.141 (0.332)	78.58	0.48
	8	0.4	7609	0.986 (0.908)	6.9 (7.1)	9.9 (5.0)	0.155 (0.426)	78.59	0.47
	9	0.33	7615	0.987 (0.870)	6.9 (7.1)	9.1 (4.1)	0.165 (0.548)	78.60	0.51
	10	0.3	7621	0.985 (0.809)	6.9 (7.1)	8.8 (3.7)	0.164 (0.701)	78.62	0.54
	11	0.26	7632	0.984 (0.761)	6.9 (7.1)	8.5 (3.2)	0.168 (0.828)	78.65	0.49
225	1	1.0	7465	0.998 (0.994)	6.8 (7.1)	25.6 (16.7)	0.060 (0.097)	78.03	0.34
	2	0.83	7482	0.996 (0.991)	6.8 (7.1)	21.2 (13.1)	0.075 (0.118)	78.11	0.35
	3	0.98	7508	0.998 (0.989)	6.8 (7.1)	23.4 (13.6)	0.063 (0.129)	78.25	0.38
	4	0.82	7551	0.998 (0.984)	6.8 (7.1)	21.6 (11.0)	0.067 (0.163)	78.36	0.43
	5	0.7	7578	0.997 (0.972)	6.8 (7.1)	19.3 (8.6)	0.072 (0.209)	78.44	0.45
	6	0.61	7587	0.997 (0.959)	6.8 (7.1)	17.9 (7.1)	0.077 (0.273)	78.47	0.49
	7	0.53	7607	0.997 (0.930)	6.8 (7.1)	16.4 (5.9)	0.085 (0.351)	78.54	0.51
	8	0.42	7612	0.996 (0.886)	6.8 (7.1)	14.7 (4.8)	0.092 (0.461)	78.58	0.54
	9	0.38	7627	0.996 (0.836)	6.8 (7.1)	13.1 (3.7)	0.101 (0.605)	78.62	0.55
	10	0.33	7625	0.996 (0.756)	6.8 (7.1)	12.2 (3.1)	0.110 (0.772)	78.61	0.57
	11	0.31	7666	0.996 (0.621)	6.8 (7.1)	11.2 (2.4)	0.120 (1.043)	78.67	0.58
2004	1	1.0	7483	0.998 (0.994)	7.0 (7.2)	29.7 (16.1)	0.053 (0.110)	78.05	0.41
	2	0.76	7532	0.998 (0.988)	7.0 (7.2)	27.1 (12.6)	0.057 (0.151)	78.25	0.34
	3	0.61	7596	0.998 (0.963)	7.0 (7.2)	23.1 (7.1)	0.063 (0.285)	78.45	0.34
	4	0.49	7632	0.998 (0.856)	7.0 (7.2)	18.2 (3.2)	0.074 (0.624)	78.61	0.40
	5	0.37	7700	0.998 (0.491)	7.0 (7.2)	13.8 (1.5)	0.087 (1.354)	78.76	0.51
	6	0.31	7702	0.998 (0.134)	6.9 (7.1)	10.9 (0.7)	0.104 (3.022)	78.80	0.61
	7	0.24	7709	0.998 (0.014)	6.9 (7.1)	8.6 (0.3)	0.129 (5.384)	78.86	0.72
	8	0.19	7712	0.998 (-0.029)	6.8 (7.0)	6.9 (0.2)	0.164 (11.106)	78.90	0.83
	9	0.16	7739	0.998 (0.024)	6.8 (7.0)	5.6 (0.1)	0.211 (20.484)	78.98	0.89
	10*	0.039	7527	0.749 (-0.015)	7.0 (7.2)	1.0 (0.1)	1.033 (49.363)	78.24	0.05
2012	1	1.0	7522	0.996 (0.996)	6.9 (7.2)	30.8 (25.0)	0.065 (0.078)	78.20	0.07
	2	0.8	7545	0.996 (0.994)	7.0 (7.3)	29.7 (19.7)	0.063 (0.095)	78.29	0.14
	3	0.64	7569	0.997 (0.985)	7.0 (7.2)	25.1 (11.7)	0.064 (0.149)	78.35	0.25
	4	0.5	7584	0.997 (0.969)	7.0 (7.2)	20.9 (8.6)	0.069 (0.236)	78.41	0.34
	5	0.38	7601	0.998 (0.932)	7.0 (7.2)	16.4 (5.3)	0.078 (0.374)	78.44	0.42
	6	0.32	7606	0.997 (0.860)	7.0 (7.2)	13.1 (3.2)	0.089 (0.569)	78.46	0.47
	7	0.27	7607	0.996 (0.736)	7.0 (7.2)	10.4 (2.3)	0.108 (0.846)	78.48	0.51
	8	0.25	7611	0.995 (0.614)	7.0 (7.1)	8.3 (1.7)	0.130 (1.185)	78.48	0.51
	9	0.2	7591	0.993 (0.491)	7.0 (7.2)	6.9 (1.4)	0.156 (1.446)	78.40	0.46
	10*	0.15	7577	0.969 (0.496)	7.1 (7.3)	4.9 (1.2)	0.248 (1.483)	78.37	0.03
	11*	0.16	7567	0.984 (0.306)	7.0 (7.2)	5.3 (1.0)	0.212 (2.083)	78.32	0.34
2019	1	1.0	7429	0.993 (0.996)	6.9 (7.2)	28.4 (23.6)	0.071 (0.079)	77.95	0.07
	2	0.88	7483	0.996 (0.995)	7.0 (7.2)	30.3 (19.6)	0.063 (0.097)	78.10	0.13
	3	0.74	7522	0.998 (0.990)	7.0 (7.2)	26.7 (13.7)	0.062 (0.137)	78.23	0.18
	4	0.64	7547	0.998 (0.982)	7.0 (7.2)	23.3 (9.5)	0.067 (0.199)	78.33	0.20
	5*	0.51	7568	0.992 (0.957)	7.0 (7.2)	10.9 (5.3)	0.117 (0.300)	78.42	0.02
	6*	0.43	7553	0.984 (0.915)	6.9 (7.2)	8.2 (3.8)	0.153 (0.434)	78.44	-0.01
	7*	0.38	7556	0.978 (0.847)	6.9 (7.2)	6.8 (2.7)	0.183 (0.604)	78.43	-0.01
	8*	0.32	7529	0.969 (0.748)	6.9 (7.2)	5.9 (2.0)	0.203 (0.842)	78.35	0.00
2024	1	1.0	7419	0.996 (0.996)	6.6 (7.1)	29.8 (22.4)	0.060 (0.078)	78.04	0.23
	2	0.81	7486	0.996 (0.993)	6.6 (7.2)	26.8 (17.0)	0.064 (0.104)	78.18	0.25
	3	0.73	7518	0.996 (0.989)	6.8 (7.2)	23.5 (12.6)	0.068 (0.138)	78.26	0.29
	4	0.65	7547	0.997 (0.982)	6.8 (7.2)	20.8 (10.0)	0.071 (0.184)	78.35	0.31
	5	0.53	7559	0.997 (0.967)	6.9 (7.2)	18.5 (6.7)	0.077 (0.259)	78.41	0.32
	6	0.48	7581	0.996 (0.933)	6.8 (7.2)	15.7 (5.6)	0.087 (0.368)	78.50	0.32
	7	0.38	7578	0.995 (0.890)	6.9 (7.2)	12.4 (4.6)	0.106 (0.479)	78.49	0.20
	8	0.34	7611	0.995 (0.801)	6.9 (7.2)	11.6 (3.4)	0.112 (0.709)	78.62	0.28
	9*	0.29	7617	0.985 (0.687)	6.9 (7.2)	6.8 (2.5)	0.180 (0.953)	78.65	0.09
	10*	0.23	7642	0.979 (0.562)	6.9 (7.2)	5.9 (2.0)	0.199 (1.253)	78.68	0.08
	11*	0.21	7659	0.992 (0.419)	6.9 (7.1)	7.7 (1.8)	0.154 (1.601)	78.68	0.32

Table 4.4: Statistics for the mid beam data collection. An asterix indicates unreliable data, a consequence of the damage. All datasets had >99.3% completeness. Below the dashed line for each crystal, data are deemed unreliably processed.

Crystal	Probe	I_n/I_1	Unique Observations	$CC_{1/2}$	Multiplicity	$I/\sigma(I)$	R_{meas}	Unit Cell [Å]	Mosaicity [°]
55	1	1.0	7498	0.997 (0.986)	7.1 (7.3)	20.0 (10.9)	0.078 (0.171)	78.07	0.03
	2	0.9	7508	0.997 (0.984)	7.1 (7.3)	19.4 (9.6)	0.074 (0.195)	78.13	0.06
	3	0.89	7540	0.997 (0.971)	7.1 (7.3)	18.5 (7.6)	0.075 (0.251)	78.19	0.09
	4	0.85	7560	0.997 (0.948)	7.1 (7.3)	16.8 (5.5)	0.081 (0.350)	78.25	0.12
	5	0.7	7566	0.997 (0.921)	7.1 (7.3)	15.6 (4.5)	0.087 (0.458)	78.33	0.13
	6	0.69	7577	0.997 (0.872)	7.1 (7.3)	14.5 (3.6)	0.094 (0.588)	78.35	0.14
	7	0.68	7576	0.997 (0.824)	7.1 (7.3)	13.3 (2.9)	0.101 (0.732)	78.36	0.16
	8	0.66	7604	0.997 (0.723)	7.1 (7.3)	12.1 (2.3)	0.111 (0.958)	78.45	0.17
	9	0.58	7609	0.996 (0.615)	7.1 (7.3)	11.3 (1.9)	0.121 (1.180)	78.49	0.18
	10	0.54	7571	0.997 (0.629)	7.0 (7.3)	11.4 (1.9)	0.119 (1.150)	78.37	0.19
	11	0.49	7600	0.997 (0.488)	7.0 (7.3)	10.5 (1.4)	0.131 (1.573)	78.46	0.22
197	1	1.0	7558	0.993 (0.984)	7.1 (7.3)	17.0 (10.2)	0.099 (0.164)	78.28	0.06
	2	0.87	7565	0.993 (0.987)	7.1 (7.3)	17.2 (10.6)	0.096 (0.173)	78.32	0.06
	3	0.8	7580	0.995 (0.979)	7.1 (7.3)	16.8 (9.1)	0.093 (0.200)	78.35	0.07
	4	0.82	7641	0.990 (0.959)	7.1 (7.4)	10.7 (5.9)	0.124 (0.290)	78.60	0.02
	5	0.79	7591	0.996 (0.961)	7.1 (7.3)	15.8 (6.8)	0.093 (0.283)	78.38	0.09
	6	0.56	7590	0.996 (0.953)	7.1 (7.4)	15.5 (6.0)	0.096 (0.332)	78.39	0.10
	7	0.67	7590	0.996 (0.952)	7.1 (7.3)	16.1 (6.0)	0.092 (0.330)	78.41	0.11
	8	0.57	7606	0.996 (0.926)	7.1 (7.3)	14.7 (5.0)	0.097 (0.417)	78.43	0.11
	9	0.5	7606	0.997 (0.892)	7.1 (7.3)	14.2 (4.2)	0.100 (0.510)	78.45	0.12
	10*	0.49	7640	0.995 (0.857)	7.1 (7.4)	12.4 (3.6)	0.113 (0.593)	78.60	0.06
	11*	0.45	7640	0.996 (0.831)	7.1 (7.4)	12.2 (3.4)	0.113 (0.655)	78.60	0.07
5185	1	1.0	7428	0.996 (0.988)	7.1 (7.3)	19.3 (12.4)	0.082 (0.151)	77.92	0.03
	2	0.82	7428	0.996 (0.982)	7.1 (7.3)	19.5 (11.6)	0.082 (0.173)	77.92	0.03
	3	0.83	7430	0.996 (0.983)	7.1 (7.3)	18.0 (10.0)	0.084 (0.186)	77.92	0.03
	4	0.83	7429	0.995 (0.976)	7.1 (7.3)	16.2 (8.7)	0.087 (0.209)	77.92	0.02
	5	0.78	7429	0.994 (0.974)	7.1 (7.3)	15.0 (7.7)	0.094 (0.238)	77.92	0.02
	6	0.71	7430	0.992 (0.963)	7.1 (7.3)	13.0 (6.7)	0.102 (0.275)	77.92	0.02
	7	0.69	7498	0.997 (0.960)	7.1 (7.3)	17.2 (6.8)	0.082 (0.306)	78.15	0.11
	8	0.61	7530	0.997 (0.947)	7.1 (7.3)	16.2 (5.7)	0.086 (0.363)	78.18	0.12
P1S10	1	1.0	7445	0.998 (0.986)	7.1 (7.2)	22.4 (11.4)	0.065 (0.164)	77.94	0.14
	2	0.96	7465	0.998 (0.983)	7.1 (7.2)	22.3 (9.8)	0.062 (0.193)	78.03	0.14
	3	0.89	7485	0.998 (0.974)	7.1 (7.2)	20.5 (8.2)	0.064 (0.231)	78.10	0.16
	4	0.89	7485	0.998 (0.965)	7.1 (7.2)	18.7 (6.6)	0.069 (0.283)	78.16	0.21
	5	0.83	7517	0.998 (0.950)	7.1 (7.2)	17.1 (5.6)	0.076 (0.344)	78.20	0.25
	6	0.77	7520	0.998 (0.932)	7.1 (7.2)	15.6 (4.9)	0.082 (0.398)	78.24	0.26
	7	0.67	7536	0.997 (0.905)	7.1 (7.2)	14.5 (4.2)	0.089 (0.475)	78.27	0.28
	8	0.56	7542	0.997 (0.882)	7.1 (7.2)	13.9 (3.6)	0.094 (0.568)	78.30	0.29
	9	0.55	7557	0.997 (0.845)	7.1 (7.2)	13.1 (3.2)	0.098 (0.642)	78.34	0.30
P3S2	1	1.0	7489	0.998 (0.987)	7.0 (7.1)	23.1 (12.2)	0.062 (0.152)	78.03	0.60
	2	0.93	7500	0.998 (0.986)	7.0 (7.1)	22.7 (11.1)	0.063 (0.168)	78.08	0.61
	3	0.89	7510	0.998 (0.980)	7.0 (7.1)	21.7 (9.9)	0.063 (0.193)	78.10	0.61
	4	0.82	7510	0.998 (0.975)	7.0 (7.1)	21.2 (8.8)	0.064 (0.213)	78.15	0.62
	5	0.79	7510	0.998 (0.973)	7.0 (7.1)	20.4 (8.1)	0.066 (0.234)	78.16	0.63
	6	0.73	7542	0.998 (0.965)	7.0 (7.1)	19.8 (7.4)	0.068 (0.259)	78.18	0.63
	7	0.73	7542	0.998 (0.964)	7.0 (7.1)	19.3 (6.7)	0.068 (0.277)	78.20	0.65
	8	0.67	7545	0.998 (0.954)	7.0 (7.1)	18.3 (6.3)	0.072 (0.310)	78.25	0.65
	9	0.64	7561	0.998 (0.948)	7.0 (7.1)	17.8 (5.7)	0.074 (0.341)	78.26	0.66

Table 4.5: Statistics for the small beam data collection. An asterix indicates that there were very few reflection available for post-refinement, leading to problems in obtaining the correct mosaicity. All datasets had >99.6% completeness. Below the dashed line for each crystal, data are deemed unreliably processed.

4.3.2 General consideration on dose metrics

The data were analysed both in terms of I_n/I_1, and in terms of B_{rel}. In each case, the four dose metrics AD-WC, TAD, Max Dose, and DWD, described in Chapter 3, were compared, to identify a metric which showed the same rate of change of damage (*i.e.* I_n/I_1 or B_{rel}) with dose – as calculated by RADDOSE-3D– irrespective of dose distribution. A metric which fulfils this condition, accurately calculated, could be used to reliably predict radiation damage progression.

For both I_n/I_1 and B_{rel}, the data were only fitted for values of I_n/I_1 greater than 0.4. This range was chosen so that it extended beyond the threshold of biological fidelity suggested by Owen *et al.* (2006), but was cut off before the damage was so severe that the diffraction data could no longer be considered of any biological relevance at all. Values of $D_{1/2}$ and B_{rel} were calculated by least-squares fitting a straight line fixed through (0,1) for I_n/I_1 and (0,0) for B_{rel} to the data being considered. Values for the coefficients associated with the fits, as well as a Pearson correlation coefficient is shown in the graph figures.

Of particular importance is the functional form of the Dose Decay Model (DDM) used in the calculation of the DWD[6]. Because of the absence of any physical models for the decay of intensity with dose, there is currently no widely accepted form for the DDM. A preliminary set of tests, using a linear decay model, were performed to see if this simple form had a significant effect on the values

[6] As a reminder:

$$DWD = \frac{\int d\phi \iiint d\mathbf{x} D(\mathbf{x},\phi) F(\mathbf{x},\phi) \eta(D(\mathbf{x},\phi))}{\int d\phi \int\int\int d\mathbf{x} F(\mathbf{x},\phi) \eta(D(\mathbf{x},\phi))}$$

where $D(\mathbf{x},\phi)$ is the cumulative dose at position $\mathbf{x}$, calculated from the start of the simulation up to the experimental coordinate defined by the angular position ϕ, $F(\mathbf{x},\phi)$ is the instantaneous fluence incident on $\mathbf{x}$, and coordinate ϕ, and $\eta(D(\mathbf{x},\phi))$ is the dose decay model of the volume element at $(\mathbf{x},\phi)$, as a function of the dose at that position.

of $D_{1/2}$ found by fitting the DWD (with different DDMs) to the I_n/I_1 data. The data from the big and small beams were plotted using three different linear forms for η: no decay ($\eta = 1$), the Owen *et al.* (2006) 43 MGy limit for $D_{1/2}$[7], and the $D_{1/2}$ found by fitting the early, linear, part of I_n/I_1- DWD graph, with $\eta = 1$, which gave a $D_{1/2}$ of 13 MGy[8].

The DDM calculated using the Owen *et al.* (2006) half dose of 43 MGy, shown in figure 4.6*b* did not have a visible effect on the shape of the plot compared to one calculated with no decay (figure 4.6*a*), and also did not significantly affect the $D_{1/2}$ found by fitting the plot, as shown in table 4.6. The more aggressive DDM, with a half dose of 13 MGy (figure 4.6*c*) leads to a clearly unphysical shape of the graph, with the higher dose data sets having a lower, or nearly constant DWD. This is because as the central region of the crystal becomes more damaged, voxels reach 26 MGy (twice the $D_{1/2}$), and stop contributing at all to the DWD. They are thus discarded from the sum, and the effective dose is left fairly constant. Despite this unphysical behaviour, the effect on the $D_{1/2}$ fit is still small, as shown in table 4.6.

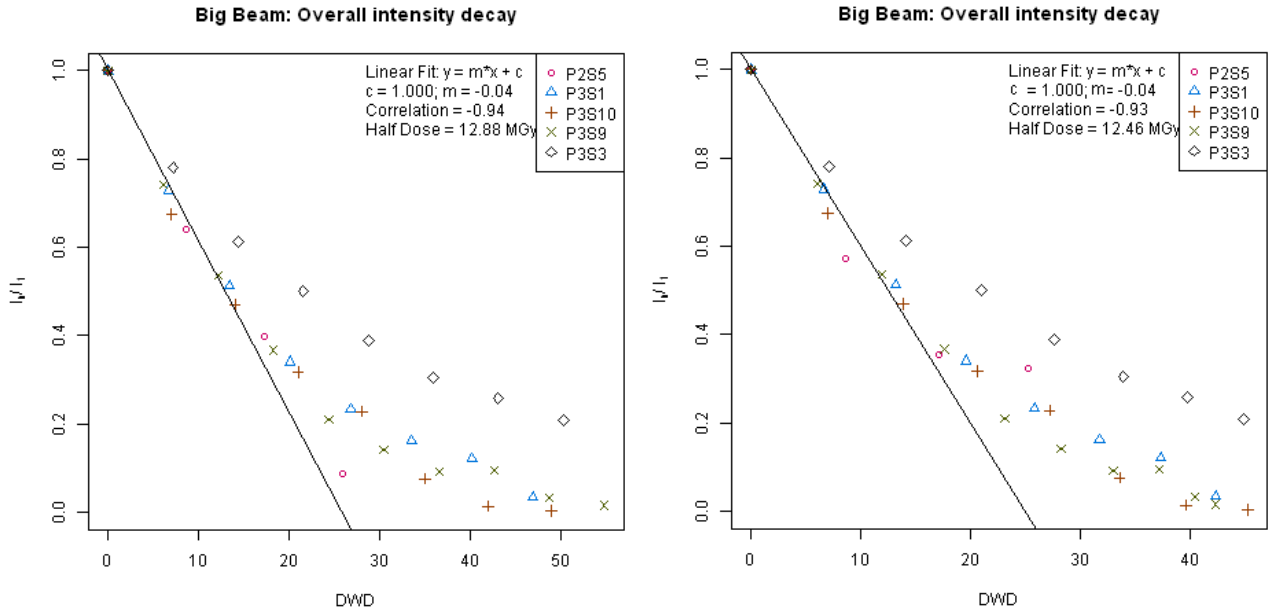
These two models were thus not considered to be valid (one had no significant effect, and the other was unphysical), and also appeared to have a small effect on the $D_{1/2}$. It was thus decided to perform the following analyses with a simplified weighting scheme for DWD, with $\eta = 1$ for all dose values, giving DWD the effective form of:

$$DWD = \frac{\int d\phi \iiint d\mathbf{x} D(\mathbf{x},\phi) F(\mathbf{x},\phi)}{\int d\phi \int\int\int d\mathbf{x} F(\mathbf{x},\phi)}$$

[7] $D_n < 2\times 43 :\ \eta = 1 - \frac{D_n}{2\times 43}\ ;\ D_n > 2\times 43 :\ \eta = 0$
[8] $D_n < 2\times 13 :\ \eta = 1 - \frac{D_n}{2\times 13}\ ;\ D_n > 2\times 13 :\ \eta = 0$

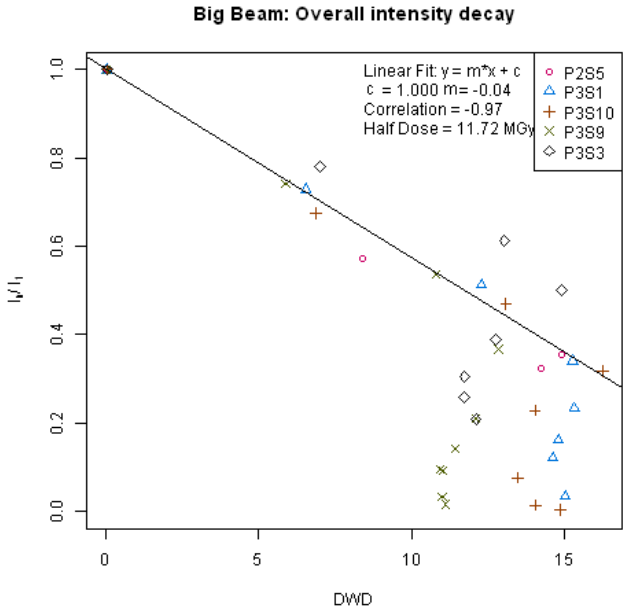
	$D_{1/2}$ [MGy]		
Model	$\eta = 1$	$\eta = 1 - \frac{D_n}{2\times 43}$	$\eta = 1 - \frac{D_n}{2\times 13}$
Big Beam	12.6	12.5	11.7
Small Beam	12.9	12.9	10.4

Table 4.6: Effect on $D_{1/2}$ for $I_n/I_1 > 0.4$ of a linear decay η function in the DWD equation. The changes in $D_{1/2}$ are small in this region, even for a very fast loss of intensity, despite very large changes at higher damage states, as can be seen from figure 4.6.



(a) I_n/I_1 *vs* DWD with $\eta = 1$

(b) I_n/I_1 *vs* DWD with $\eta = 1 - \frac{D_n}{2\times 43}$



(c) I_n/I_1 *vs* DWD with $\eta = 1 - \frac{D_n}{2\times 13}$

Figure 4.6: Plots of I_n/I_1 against DWD, with different forms of the DDM within the DWD equation. All plots were fitted with a straight line through (1,0), for datasets with $I_n/I_1 > 0.4$ (see main text), and crystal P3S3 was discarded (see section 4.3.3). Varying the form of the diffraction efficiency term in the DWD equation has a large effect for (*c*) in the high damage region. It does, however, not have a significant effect in the biologically important $I_n/I_1 > 0.4$ region, as shown by the similar values of $D_{1/2}$ derived from the linear fits shown in table 4.6.

4.3.3 Comparison of dose metrics for I_n/I_1

Plots of I_n/I_1 for each family of crystals, with doses as calculated by RADDOSE-3D, are shown in figures 4.7 (Big Beam), 4.8 (Mid Beam), and 4.9 (Small Beam). The summary of the fitted half doses for $I_n/I_1 > 0.4$ is shown in table 4.7. For the big beam data collection, crystal P3S3 was an outlier for all dose metrics. The notes taken during data collection indicated that there may have been an error in the centring of the crystal[9]. Simulations were performed for offsets up to 25 μm in steps of 5 μm, and, with a 25 μm offset, the intensity decay of the crystal was close to that of the other crystals. This suggests that the higher calculated dose tolerance was a consequence of the experiment having a 25 μm offset, whilst the model used 0 μm offset. Despite the good fit of the data from the crystal after this correction, it was left out of subsequent analysis, since this correction could not be experimentally verified. When plotted, crystal P3S3 is shown with the un-corrected doses.

Looking at table 4.7, it is immediately obvious that, of the four dose metrics, DWD is invariant for the three dose conditions, and that the other three metrics show a large variance as the dose contrast is changed. This is also illustrated very clearly by figure 4.10, which presents all three beam conditions together.

This is the key result of this thesis: it has been shown that, for the biologically relevant $I_n/I_1 >$ 0.4 regime, DWD is an effective, strategy-invariant measure of radiation damage, as reported

[9] An accidental misalignment of the beam and rotation axes would reduce the dose per unit incident photons, as described in Chapter 5.

by I_n/I_1.

Another important aspect to be validated is the use of a linear model to characterise the decay of the crystals for $I_n/I_1 > 0.4$. In order to confirm that this was the correct approach, residual analysis (figure 4.11a) and normal probability plots (figure 4.11b) were generated for I_n/I_1 against DWD (the data are shown in figure 4.10b). The residuals appeared normally distributed, with no trends in the data, and with a RMSD on the residuals of 0.067 MGy. The normal distribution of these residuals is confirmed by the Q-Q plot in figure 4.11b, which shows the quantiles of the residuals against their expected (normal) theoretical values. It is clearly a straight line, validating the linear fit to the I_n/I_1 data, down to $I_n/I_1 = 0.4$.

	$D_{1/2}$				
Beam	ADWC [MGy]	DWD [MGy]	Max Dose [MGy]	TAD [MGy]	$N_{crystals}$
Big	10.8	12.9	18.7	10.1	4
Medium	6.9	12.9	42.1	5.2	6
Small	5.6	12.8	38.0	4.5	5

Table 4.7: Summary of the I_n/I_1 data for the three beam classes. In each case, the $D_{1/2}$ was calculated by performing a linear fit on the average summed intensity of each full dataset which had an I_n/I_1 value above 0.4. ADWC, Max Dose and TAD vary by around a factor of 2 between beam conditions, but the $D_{1/2}$ using DWD varies by less than 1%, showing how it is robust for use in strategy considerations.

In light of the fact that the $D_{1/2}$ using DWD is the same for all three beam conditions, it is worth considering the potential impact of the different strategies on data quality. A simple way of considering this is by examining diffraction efficiency: here defined as the number of elastically scattered photons per MGy (DWD) of absorbed dose. As shown here, one MGy of absorbed dose corresponds to a constant loss of I_n/I_1, for $I_n/I_1 > 0.4$. Thus, a higher diffraction efficiency indicates a more effective strategy. Table 4.8 shows that the large, more even beam will give a

more efficient strategy, consistent with the hypothesis that, for non-small crystals[10], a more even exposure strategy is most efficient.

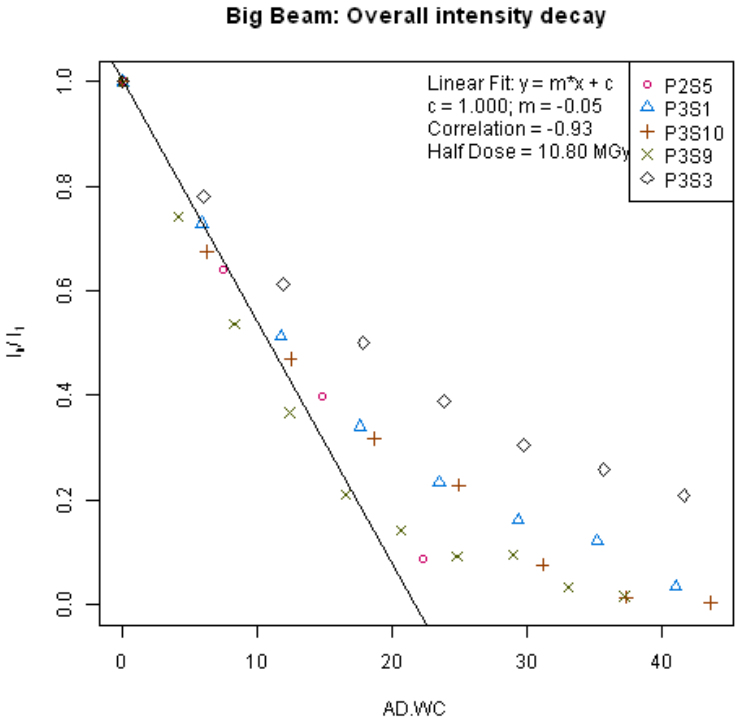
Beam size:	Big	Medium	Small
Diffraction Efficiency [photons/MGy]	3.45×10^{11}	2.27×10^{11}	1.6×10^{11}

Table 4.8: Simulated relative Diffraction Efficiencies of the three beams, calculated for an identical 200 μm crystal using RADDOSE-3D, to allow direct comparison. It is clear that for each MGy of dose absorbed, many more photons are elastically scattered with the more even large beam.

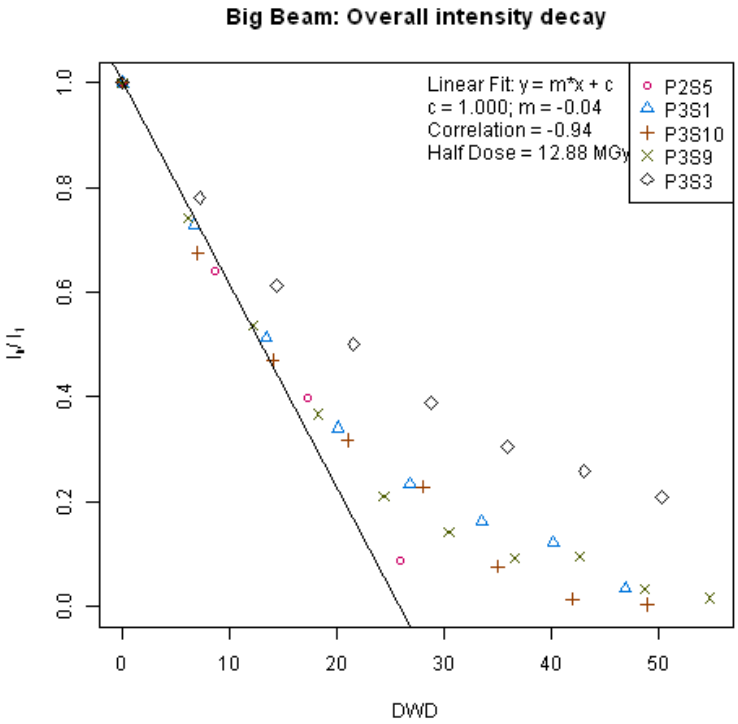
A final piece of analysis is to consider the distribution of individual $D_{1/2}$ for each crystal, as opposed to the global $D_{1/2}$ from fitting the data from all the crystals at once. This is shown in figure 4.12. The first, striking, point is that the median DWD $D_{1/2}$ is 12.9 MGy, the same value as for the fit to the whole dataset. This level of precision is unexpected, and probably a coincidence, however it is a strong reinforcement that the $D_{1/2}$ for DWD derived from different approaches are in agreement. DWD also has a much tighter interquartile range than any of the other metrics, with outliers only present from the higher contrast medium and small beams. The outliers are perhaps a consequence of the smaller, more Gaussian beams being more sensitive to small errors in crystal centring, leading to differences between the modelled and actual dose fields. The very wide spread for the other metrics is at least in part due to the variety of crystal sizes having a strong effect on AD-WC and TAD. TAD and AD-WC look similar, with a smaller interquartile range for TAD; a consequence of the 'tail-chopping' effect of this metric. In both of these cases, a higher half-dose is essentially a mark of a more even dose distribution, so the most uneven dose distributions are for the medium beam with the very large crystals, then for the small beam with smaller – but

[10] So that photoelectron escape can be ignored.

not quite proportionally smaller – crystals, and then for the big beam and for 2004, which was a small crystal in the medium beam. Max Dose also behaves as expected, with the big beam having clearly the smallest values, followed by the medium and then the small beam: a reflection of the increasing peak doses experienced by these crystals, as can be seen from the heat maps of the beams in figure 4.2.

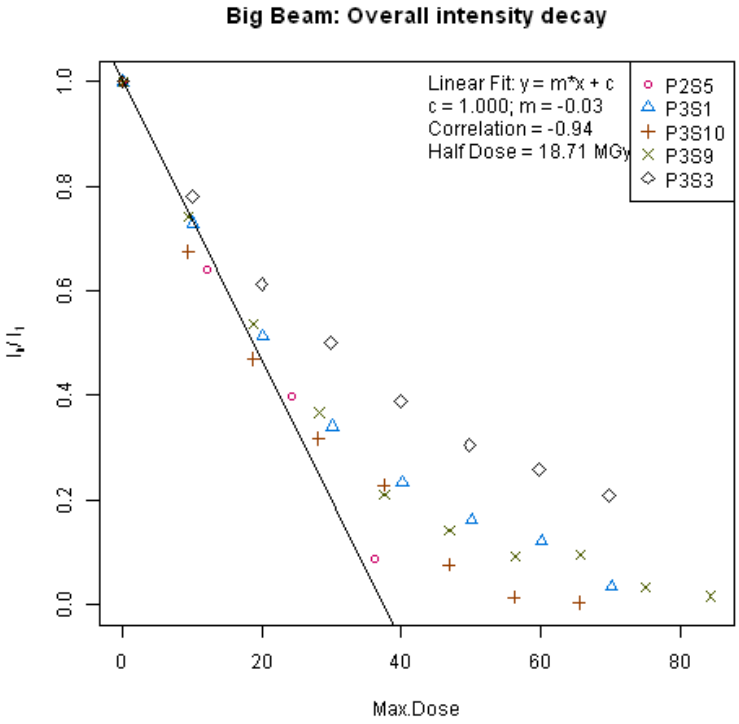


(a) I_n/I_1 against Average Dose (Whole Crystal) for the big beam data collection. The x axis is in MGy.

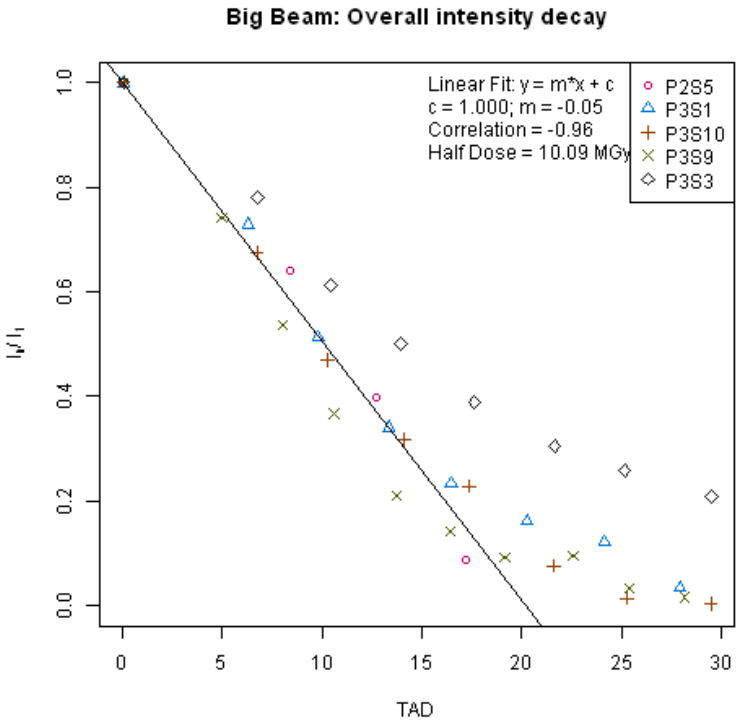


(b) I_n/I_1 against DWD for the big beam data collection. The x axis is in MGy.

Figure 4.7

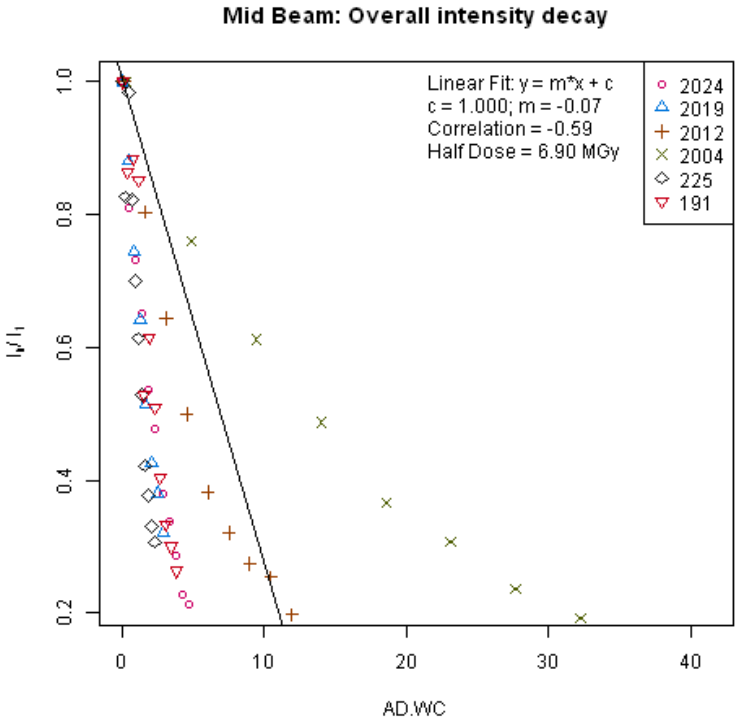


(c) I_n/I_1 against Maximum Dose for the big beam data collection. The x axis is in MGy.

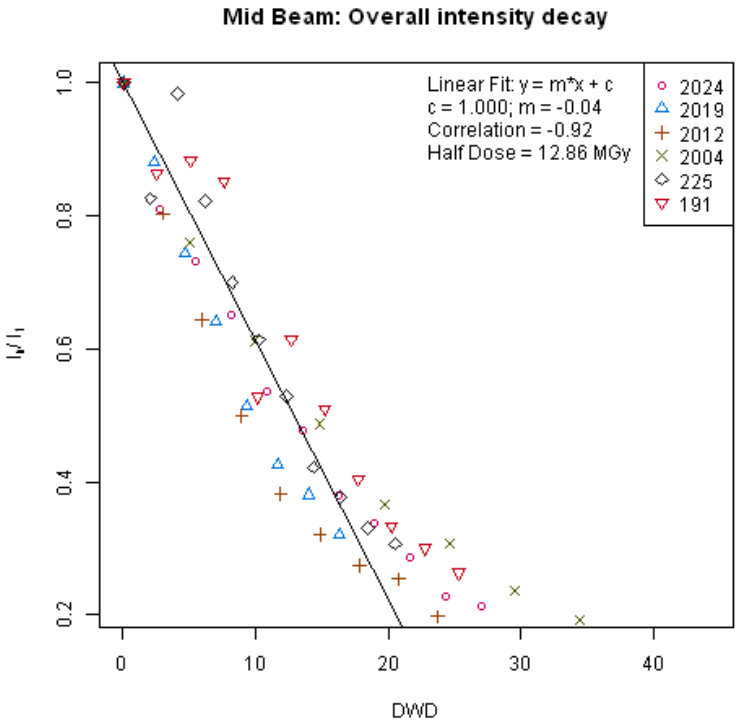


(d) I_n/I_1 against Threshold Average Dose (95%) for the big beam data collection. The x axis is in MGy.

Figure 4.7: Overall intensity decay data for the big beam data collection. Linear fits have been performed with crystal P3S3 excluded (see section 4.3.3 for explanation), and for points with $I_n/I_1 > 0.4$. This threshold was set in order to match the range of intensities found in the small beam data series, and in order to remain within the linear region of decay, as identified by Teng and Moffat (2000), and validated in figure 4.11b.

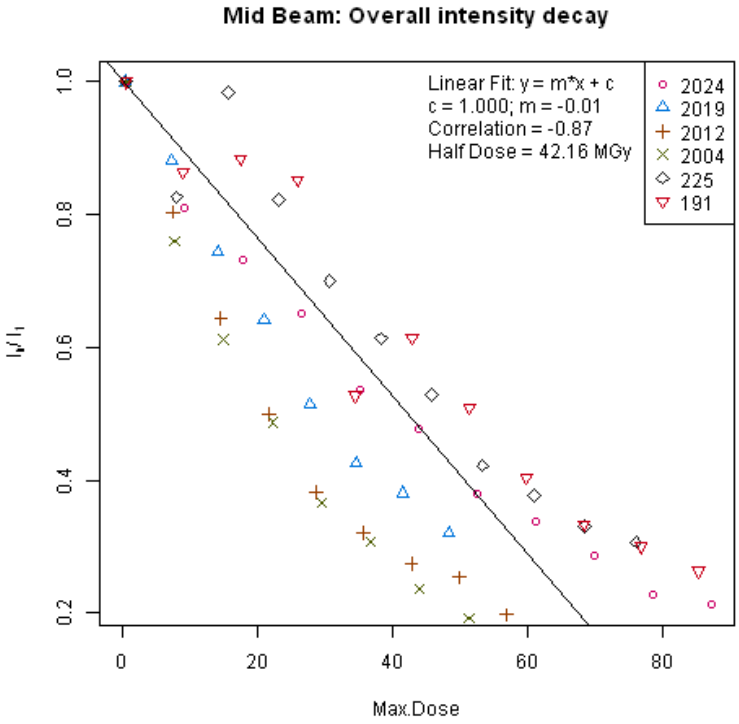


(a) I_n/I_1 against Average Dose (Whole Crystal) for the medium beam data collection. The x axis is in MGy.

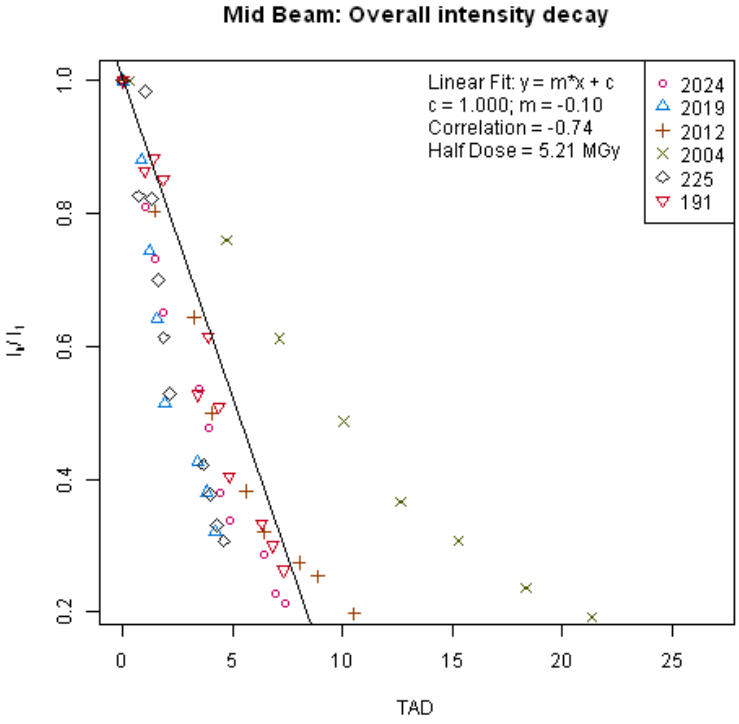


(b) I_n/I_1 against DWD for the medium beam data collection. The x axis is in MGy.

Figure 4.8

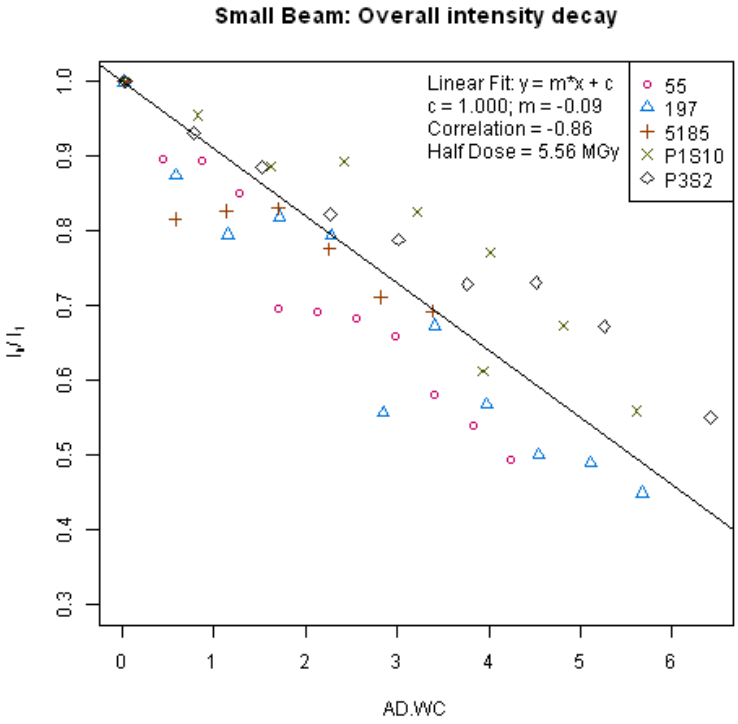


(c) I_n/I_1 against Maximum Dose for the medium beam data collection. The x axis is in MGy.

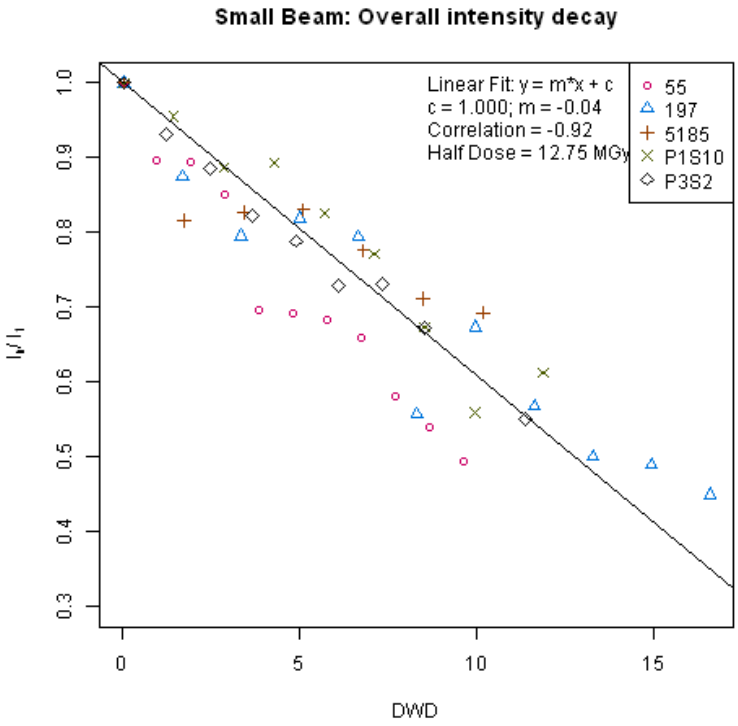


(d) I_n/I_1 against Threshold Average Dose (95%) for the medium beam data collection. The x axis is in MGy.

Figure 4.8: Overall intensity decay data for the medium beam data collection, with the linear fits once again only for those datasets with $I_n/I_1 > 0.4$. Of particular interest is crystal 2004, which had smaller dimensions than the other crystals present, meaning that the crystal was illuminated relatively less by the weak tails of the beam. This is manifested in an artificicially increased lifetime for ADWC and TAD, the metrics which consider averages for volume. The fact that DWD is unaffected by this change in crystal size is a powerful validation of the metric.

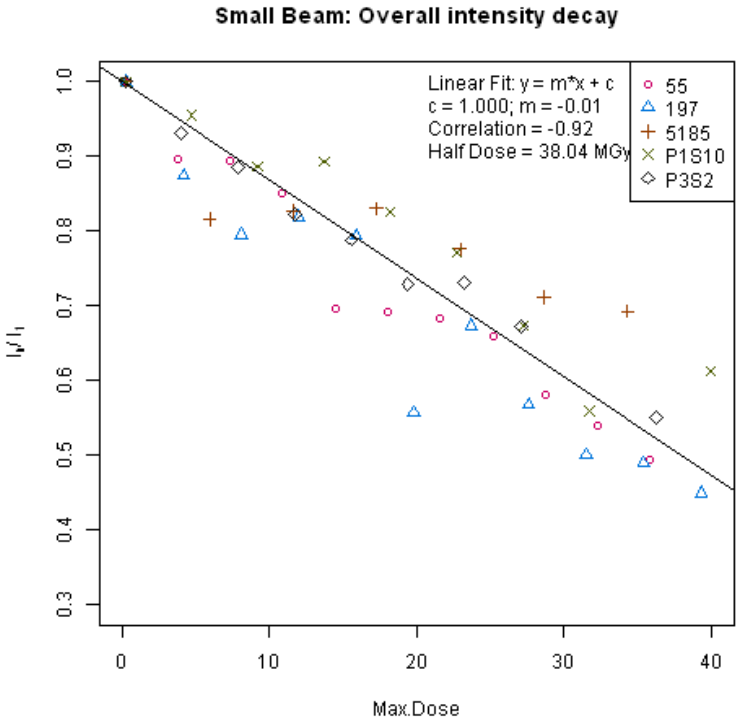


(a) I_n/I_1 against Average Dose (Whole Crystal) for the small beam data collection. The x axis is in MGy.

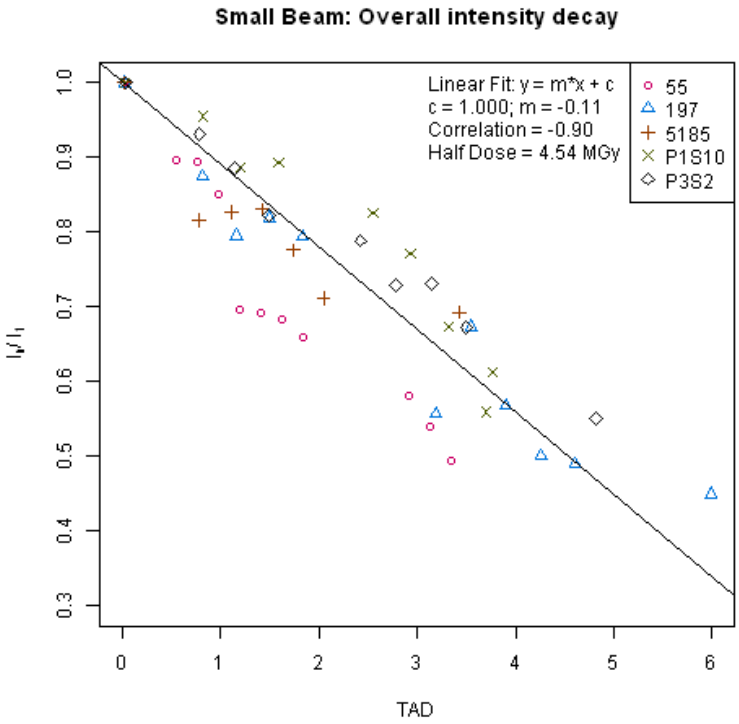


(b) I_n/I_1 against DWD for the small beam data collection. The x axis is in MGy.

Figure 4.9

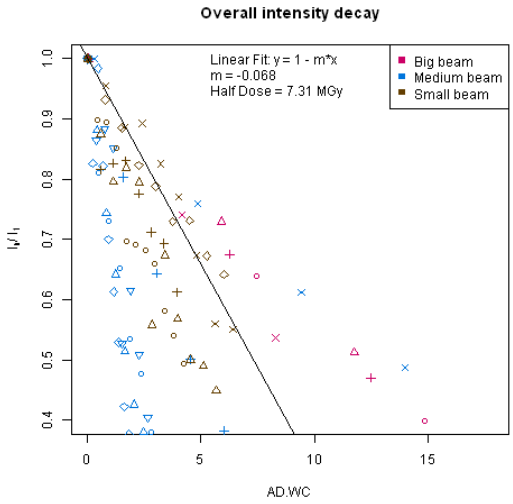


(c) I_n/I_1 against Maximum Dose for the small beam data collection. The x axis is in MGy.

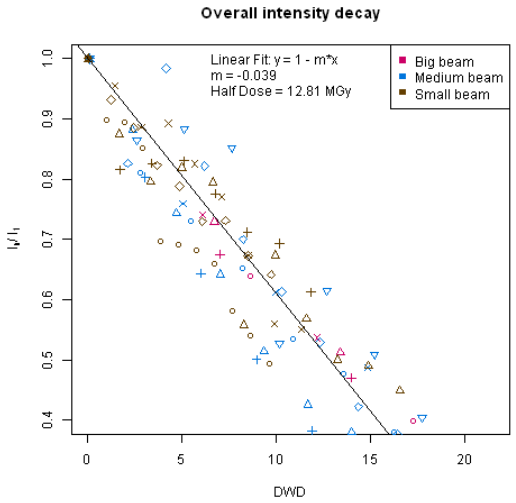


(d) I_n/I_1 against Threshold Average Dose (95%) for the small beam data collection. The x axis is in MGy.

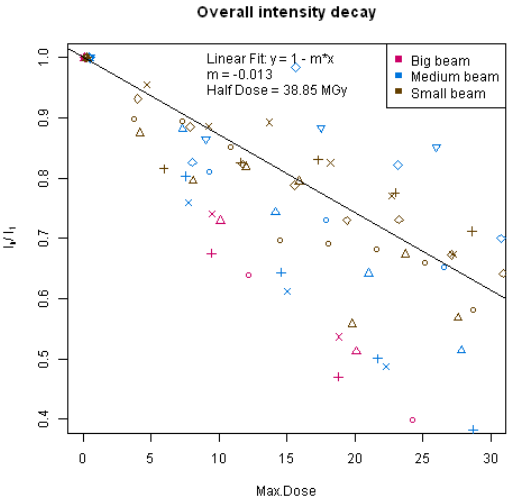
Figure 4.9: Overall intensity decay data for the small beam data collection. Total range of I_n/I_1 is smaller than for the other two cases, due to the lower flux of the beam.



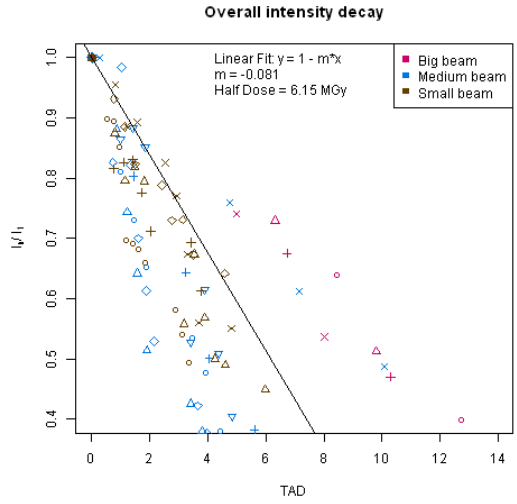
(a) Average Dose (Whole Crystal) *vs* I_n/I_1 for all three dose profiles.



(b) DWD *vs* I_n/I_1 for all three dose profiles.

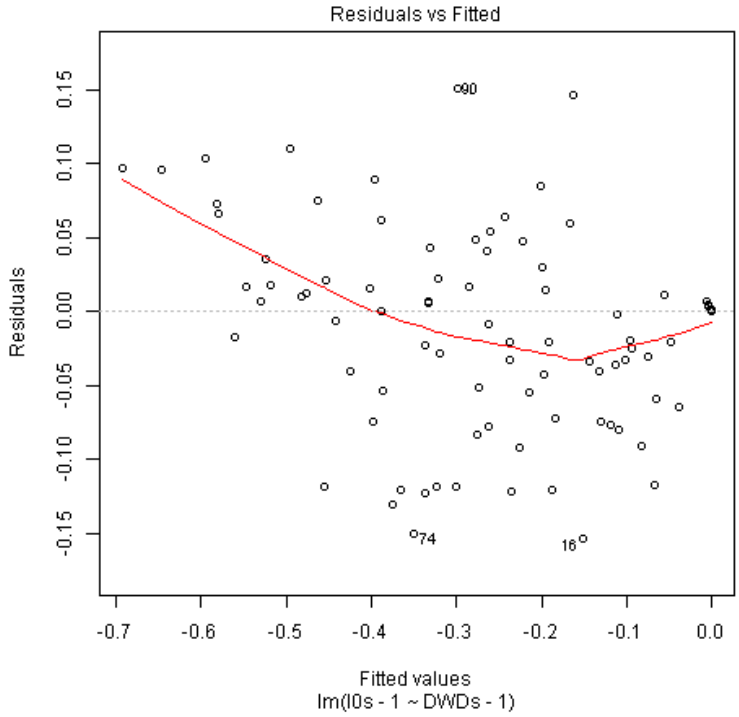


(c) Max Dose *vs* I_n/I_1 for all three dose profiles.

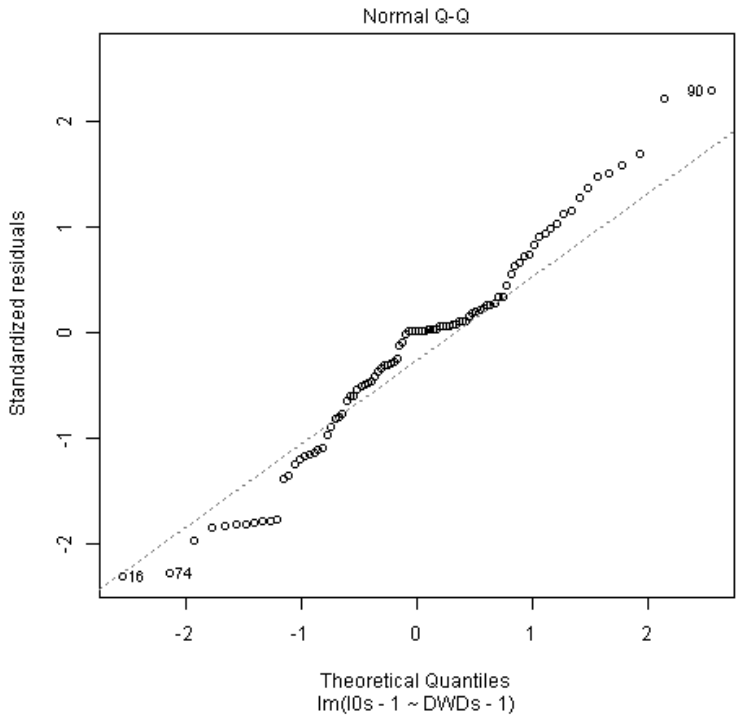


(d) TAD *vs* I_n/I_1 for all three dose profiles.

Figure 4.10: Plots of dose against I_n/I_1, with all three beam conditions shown. This clearly illustrates how the DWD is invariant to dose distribution. Note that the misaligned crystal P3S3 was excluded. In all cases, the x axis is in MGy



(a) I_n/I_1 residuals against for the DWD average half-dose. No trends signaling a non-linear pattern are present.



(b) Q-Q plot showing the quantiles of the residuals against their expected (normal) theoretical values. The quantiles forming a straight line indicates that the residuals were normally distributed about the fit. The slight presence of points below the line at low quantiles and above at high quantiles indicates that there are long tails on the distribution, but the symmetry clearly validates this as linear.

Figure 4.11: Residual analysis for the I_n/I_1 vs DWD data with $I_n/I_1 > 0.4$. These are the data shown in figure 4.10b. The `lm(I0s - 1 ~ DWDs -1)` explicitly shows the fit command used in R: a linear fit through $(0, 1)$ in I_n/I_1 *vs* DWD.

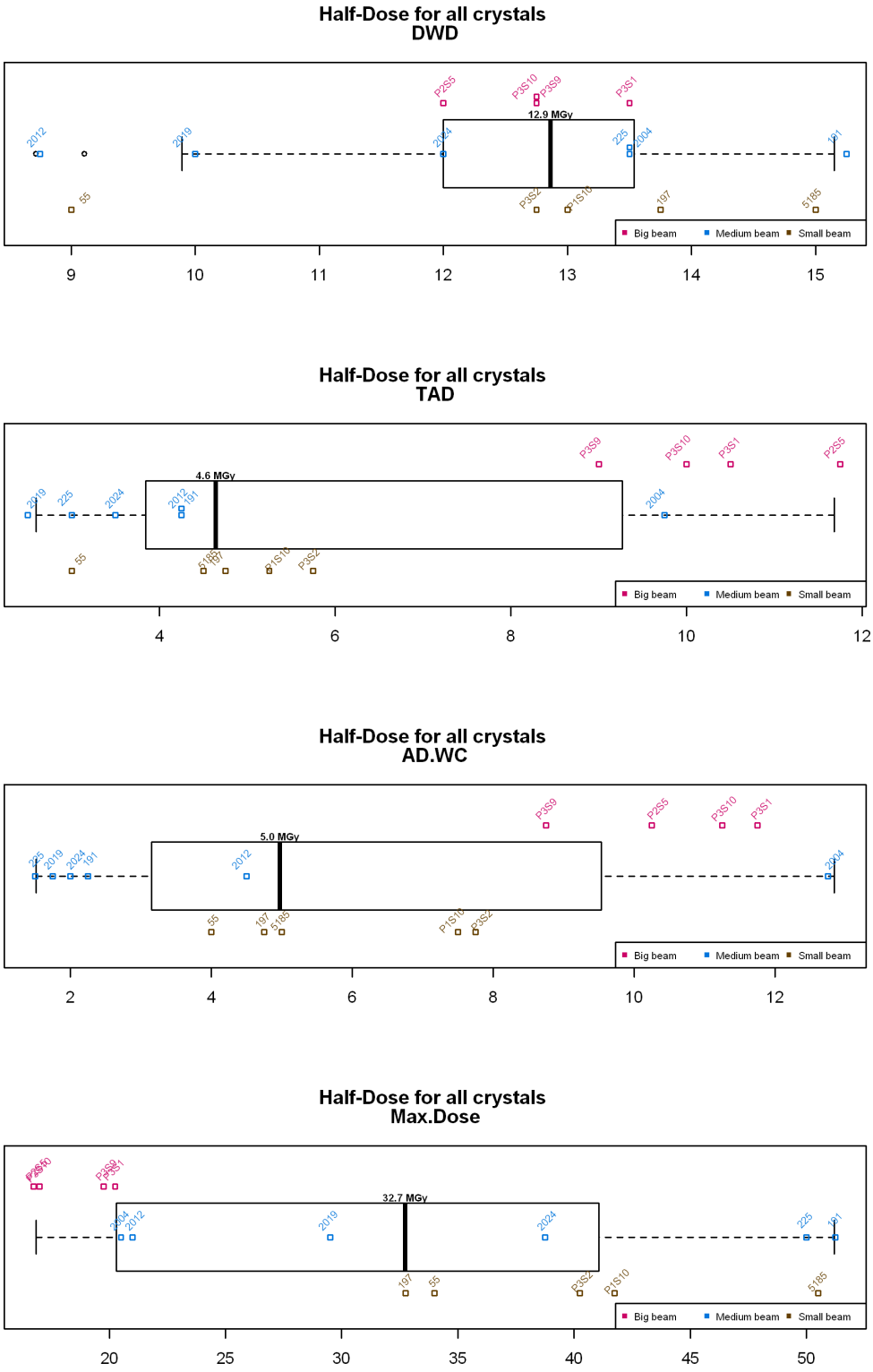


Figure 4.12: $D_{1/2}$ for each individual crystal, shown with a box and whiskers plot. The bold line with a value above it is the median, the box encloses the first and third quartile, and the whiskers extend to the furthest datum within 1.5 interquartile ranges from the upper and lower interquartile respectively. The much decreased interquartile range for the DWD is apparent, and the DWD median is 12.9 MGy, coincidentally exactly the same as the $D_{1/2}$ derived from the whole dataset. Note that the scatter plots have been rounded so that points stack up rather than overlap to help with visual clarity, whereas the box and whiskers plot shows the raw data.

4.3.4 Analysis of B_{rel}

For completeness, a similar analysis to that of I_n/I_1 shown above was performed using B_{rel} as the damage metric. In order to allow direct comparison, the same $I_n/I_1 > 0.4$ subset of the data were used for the fits, and crystal P3S3 (also an outlier in these graphs) was again excluded from the analysis. A summary of coefficients of sensitivity[11] ($S_{AD} = \Delta B_{rel}/\Delta D$) are shown in table 4.9. The lack of any clear trends can be explained by the fact that B_{rel}, being theoretically related to the Wilson B factor by $B_{rel}^{D_n} = B_{Wilson}^{D_n} - B_{Wilson}^{D_0}$, is essentially a measure of intensity loss as a function of resolution, and so is very sensitive to the change in intensity within the weaker, high resolution, reflections. It is thus unsurprising that it is not consistent across high contrast dose distributions, despite being internally consistent for individual crystals, and, to a lesser extent, across multiple crystals measured under identical conditions, as can be seen in figures 4.13, 4.14, and 4.15. Figure 4.16 presents this as a summary, showing all three beams together, and highlighting the beam-dependent variability caused by dose distribution. This is validated by figure 4.17, which shows the individually calculated S_{AD} for each crystal, similar to the way in which figure 4.12 does for I_n/I_1. There is a large spread in results against all dose metrics of ~a factor of five, compared to a factor of 2 for the DWD I_n/I_1, showing that, by the B_{rel} metric, there is little predictive power against dose. Further analysis of the individual S_{AD}s is shown in figure 4.17. In this figure, the three different beams appear to group more than for I_n/I_1, showing that B_{rel} is potentially more

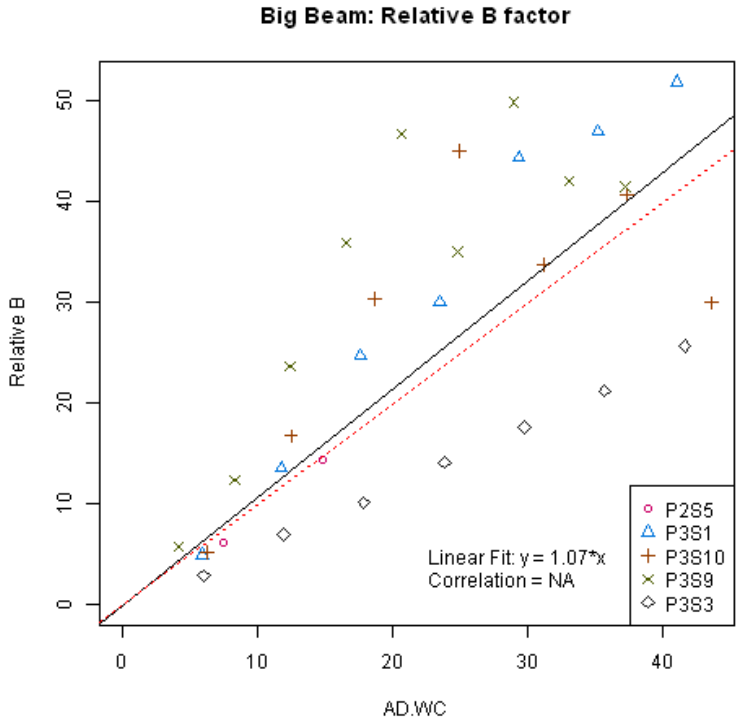
[11] Here quoted in terms of changes in B-factor as opposed to the original formulation, in terms of RMS atomic displacement: $S_{AD} = \Delta B_{rel}/\Delta D\ 8\pi^2$ (Kmetko *et al.*, 2006). Change in B-factor is the more widely adopted form, since in these units, the original Kmetko *et al.* (2006) S_{AD} = 0.014 Å^2/MGy RMS displacement works out to the more memorable ~1 Å^2/MGy change in B-factor.

sensitive to dose distributions: an important point for future work.

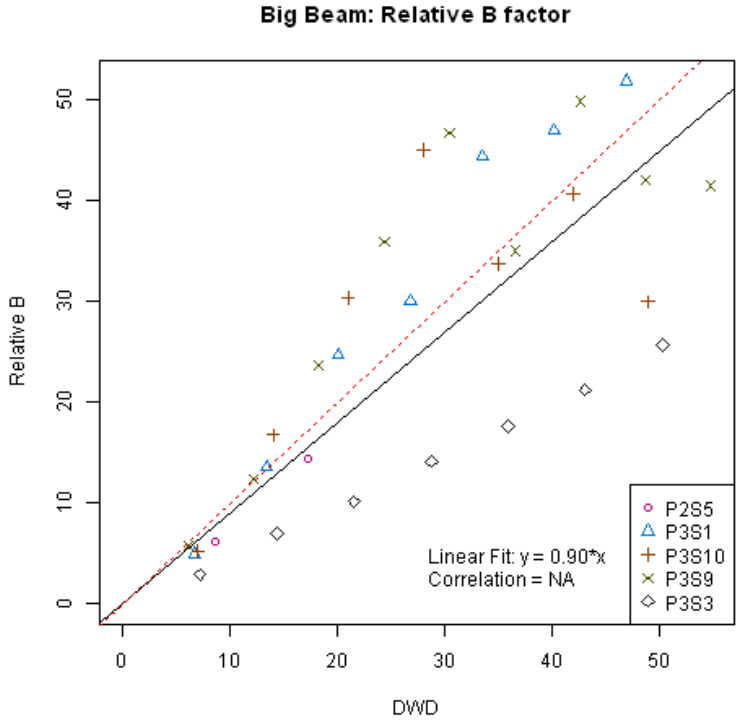
It is possible that the lack if agreement found with B_{rel} against DWD is caused by the simplifying assumption that $\eta = 1$ used here, and the lack of a resolution-dependent treatment. Since B_{rel} is a measure of the rate of change of intensity with resolution, as opposed to average intensity (as measured by I_n/I_1), an extended form of DWD, considering the resolution-dependent effect on the diffraction data of the combination of voxels at different dose states that make up each image may lead to a better agreement with the B_{rel}.

	S_{AD} [Å^2/MGy]				
Metric:	AD-WC	DWD	Max Dose	TAD	$N_{crystals}$
Big Beam	1.07	0.90	0.62	1.11	4
Medium Beam	1.32	0.70	0.21	1.73	6
Small Beam	1.29	0.54	0.18	1.54	5

Table 4.9: Summary of the B_{rel} data for the three beam classes. In each case, S_{AD} was calculated by performing a linear fit on the relative B factor for each full dataset which had an I_n/I_1 value above 0.4. None of the dose metrics is consistent across all beam sizes, which can be explained by the large spread in the gradients of the B_{rel} increases shown in figures 4.13, 4.14, and 4.15.

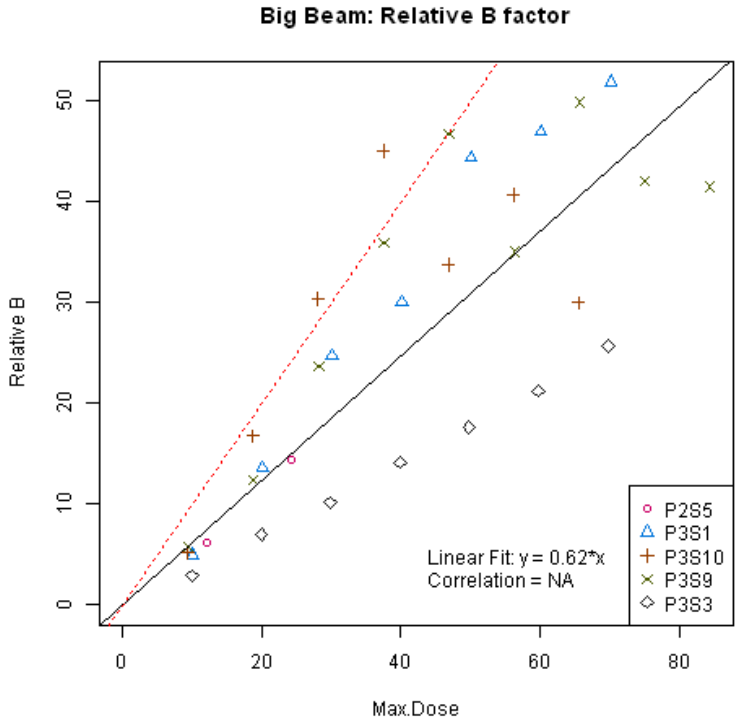


(a) B_{rel} against Average Dose (Whole Crystal) for the big beam data collection. The y axis is in Å^2, and the x axis in MGy.

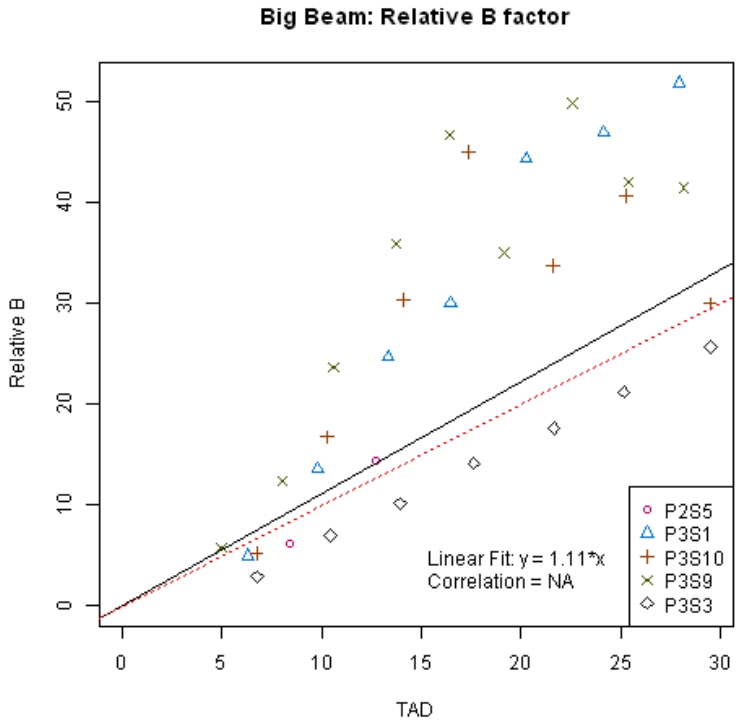


(b) B_{rel} against DWD for the big beam data collection. The y axis is in Å^2, and the x axis in MGy.

Figure 4.13

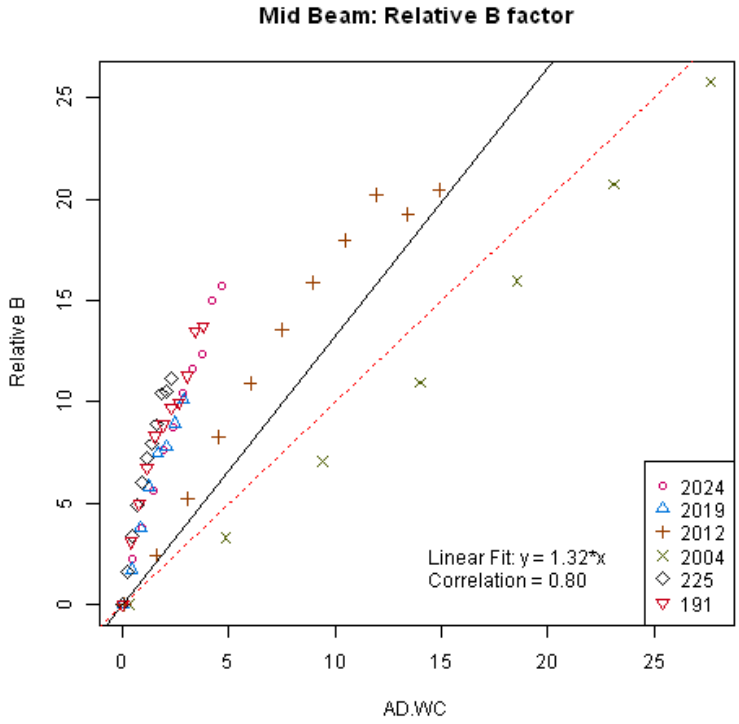


(c) B_{rel} against Maximum Dose for the big beam data collection. The y axis is in Å^2, and the x axis in MGy.

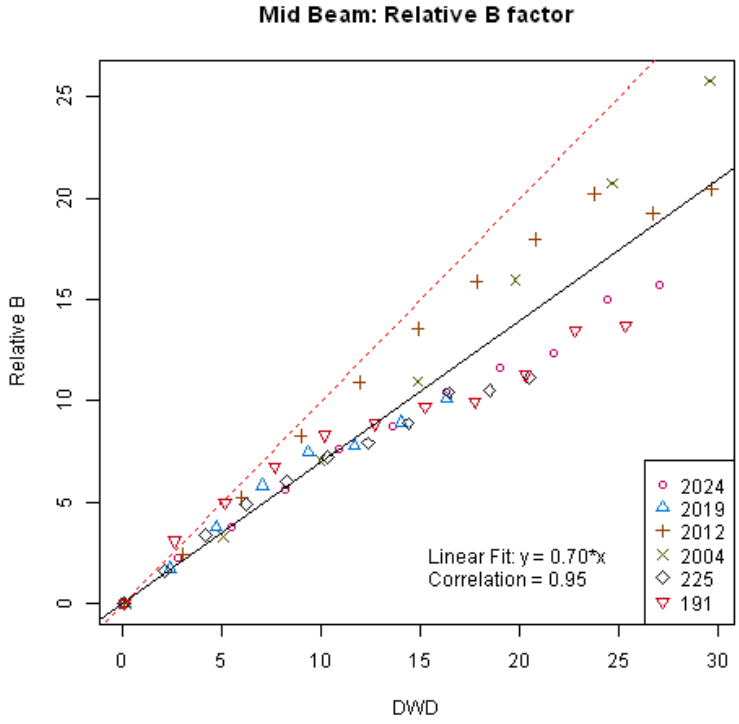


(d) B_{rel} against Threshold Average Dose (95%) for the big beam data collection. The y axis is in Å^2, and the x axis in MGy.

Figure 4.13: B_{rel} increase data for the big beam data collection. Linear fits have been performed with crystal P3S3 excluded (see section 4.3.3 for explanation), and for points with $I_n/I_1 > 0.4$, as was used for the analysis presented in section 4.3.3. Each crystal appears to have a linear increase in B_{rel} for the first 4–5 datasets, but there is significant scatter between crystals. The red dotted line shows the postulated general S_{AD} of 1 Å^2/MGy (Kmetko *et al.*, 2006).

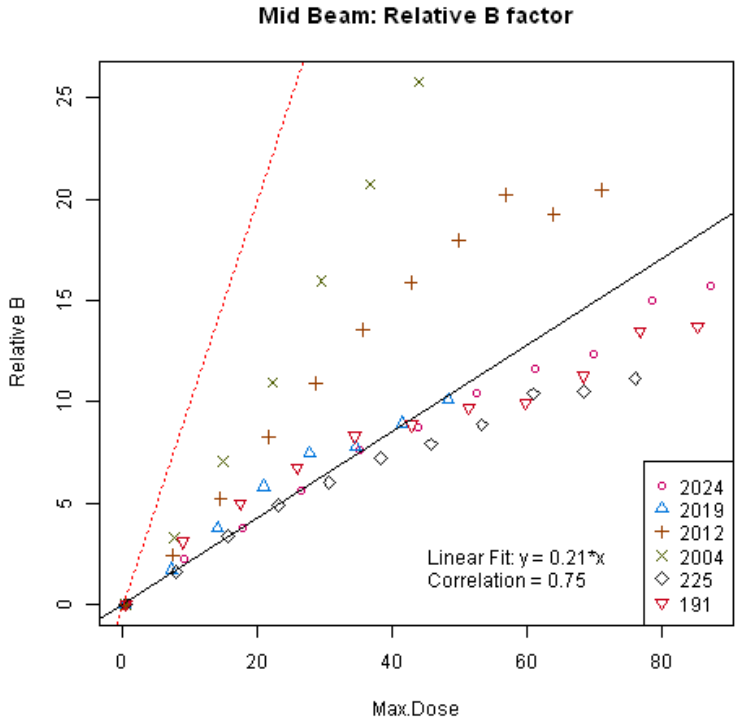


(a) B_{rel} against Average Dose (Whole Crystal) for the medium beam data collection. The y axis is in Å^2, and the x axis in MGy.

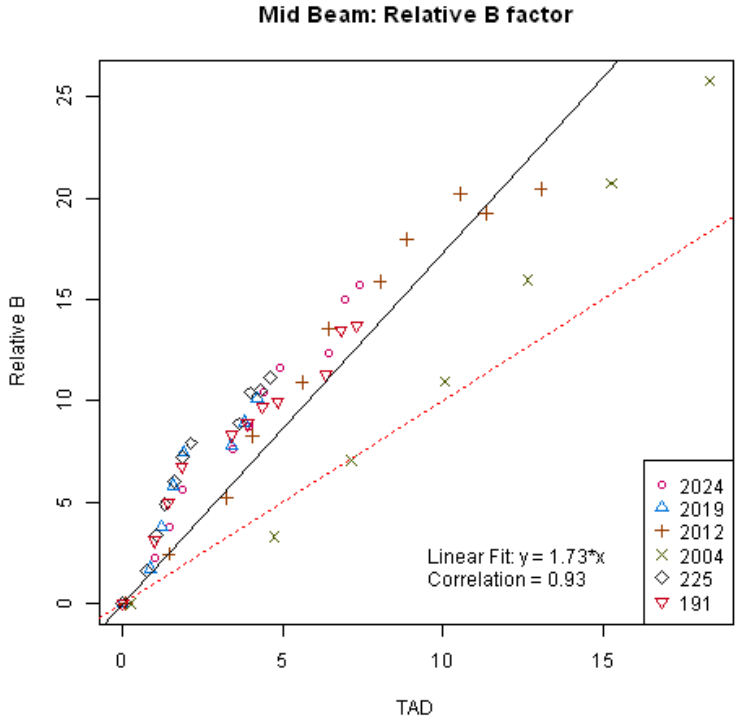


(b) B_{rel} against DWD for the medium beam data collection. The y axis is in Å^2, and the x axis in MGy.

Figure 4.14

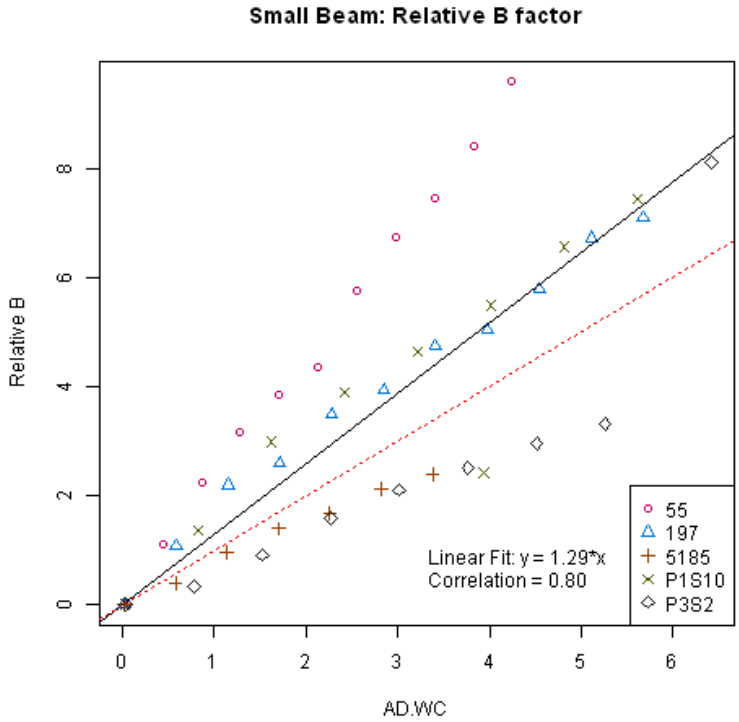


(c) B_{rel} against Maximum Dose for the medium beam data collection. The y axis is in Å^2, and the x axis in MGy.

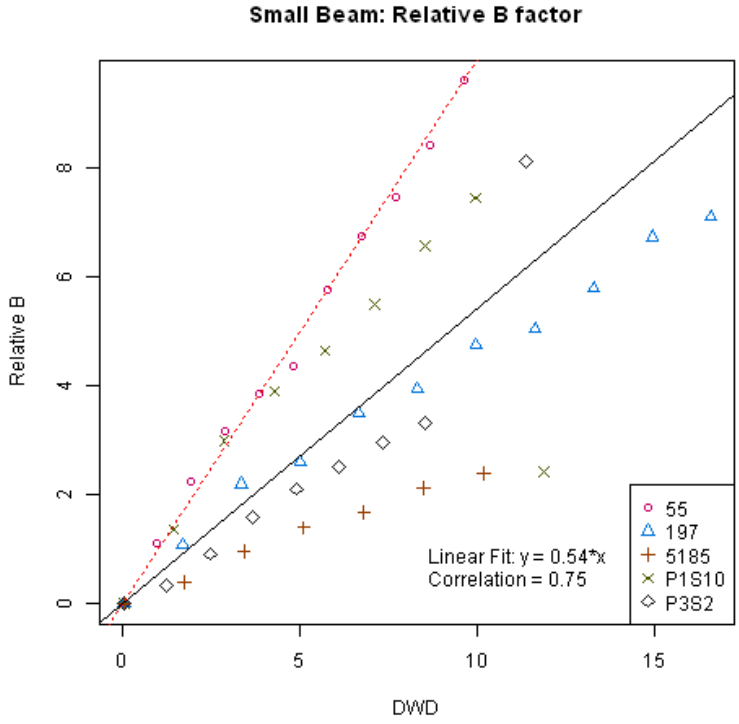


(d) B_{rel} against Threshold Average Dose (95%) for the medium beam data collection. The y axis is in Å^2, and the x axis in MGy.

Figure 4.14: Increase of B_{rel} for the medium beam data collection. Linear fits have been performed for points with $I_n/I_1 > 0.4$, as was used for the analysis in section 4.3.3. This set of crystals appear to cluster well for DWD and, again, 2004 is a strong outlier in the other metrics due to its small size affecting the dose distribution. The red dotted line shows the postulated general S_{AD} of 1 Å^2/MGy (Kmetko *et al.*, 2006).

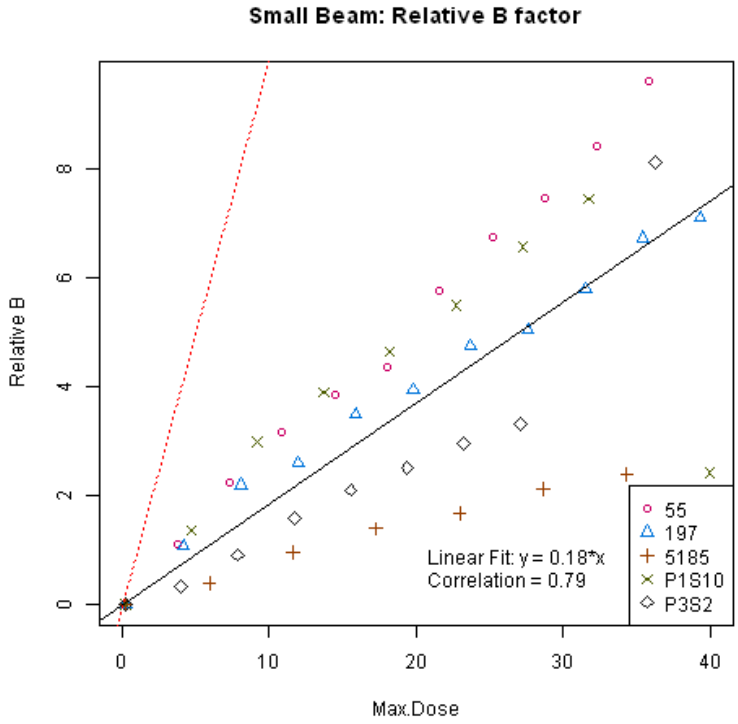


(a) B_{rel} against Average Dose (Whole Crystal) for the small beam data collection. The y axis is in Å^2, and the x axis in MGy.

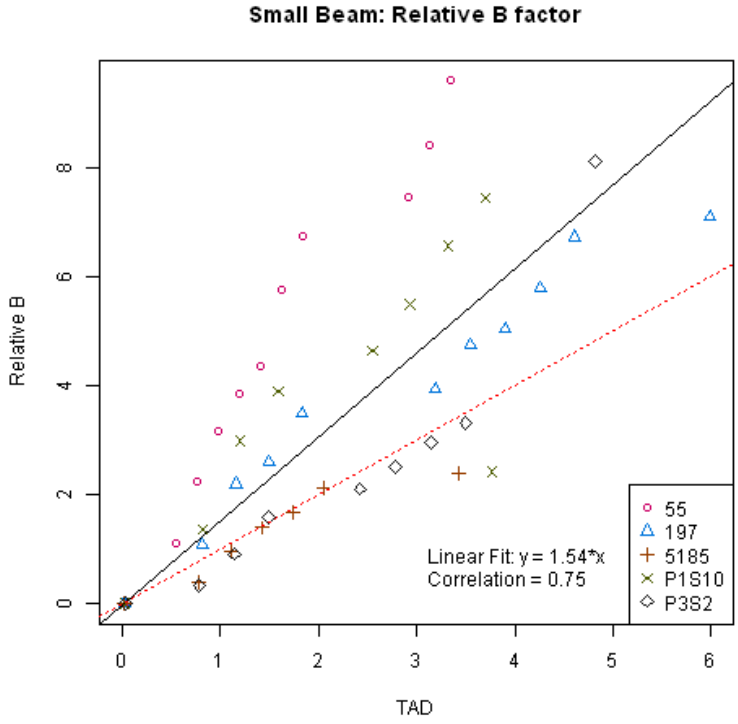


(b) B_{rel} against DWD for the small beam data collection. The y axis is in Å^2, and the x axis in MGy.

Figure 4.15

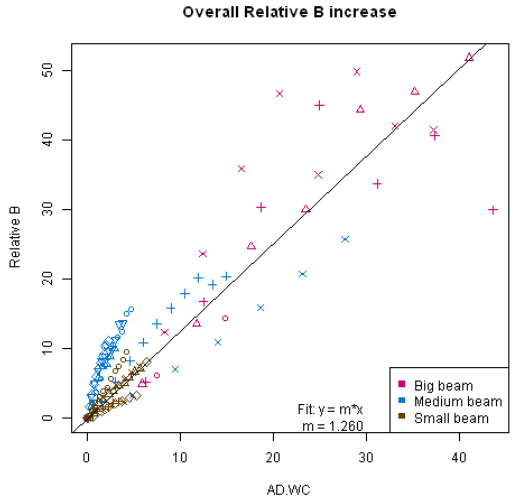


(c) B_{rel} against Maximum Dose for the small beam data collection. The y axis is in Å^2, and the x axis in MGy.

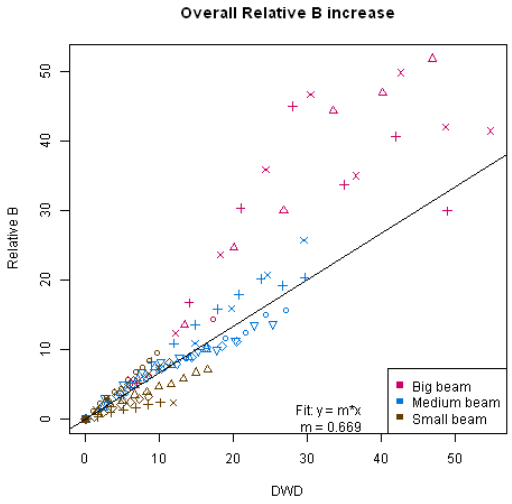


(d) B_{rel} against Threshold Average Dose (95%) for the small beam data collection. The y axis is in Å^2, and the x axis in MGy.

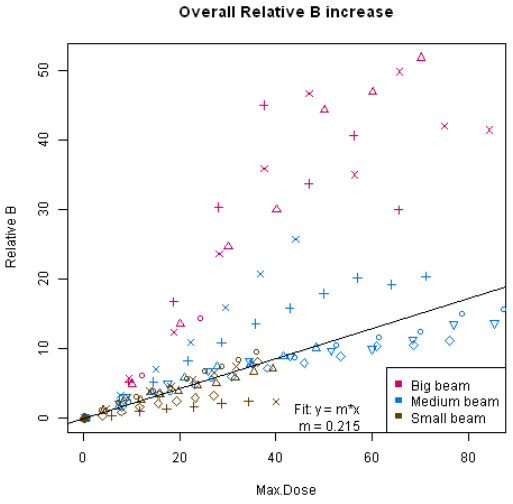
Figure 4.15: B_{rel} increase data for the small beam data collection. Linear fits have been performed for points with $I_n/I_1 > 0.4$, as was used for the analysis in section 4.3.3. Despite the relatively linear behaviour of the individual crystals, there is not a strong correlation between the samples. The red dotted line shows the postulated general S_{AD} of 1 Å^2/MGy (Kmetko *et al.*, 2006).



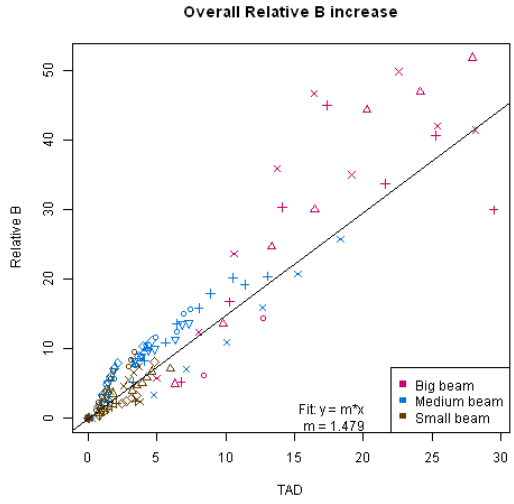
(a) Average Dose, Whole Crystal *vs* B_{rel} for all three dose profiles.



(b) DWD *vs* B_{rel} for all three dose profiles.



(c) Max Dose *vs* B_{rel} for all three dose profiles.



(d) TAD *vs* B_{rel} for all three dose profiles.

Figure 4.16: Plots of dose against B_{rel} for all three beam conditions. Because of the higher level of intrinsic scatter when using B_{rel}, the trends are not as stark as for I_n/I_1, however there is still a reduced spread for the DWD, particularly when compared to the Max Dose. In all cases, the y axis is in Å^2, and the x axis in MGy

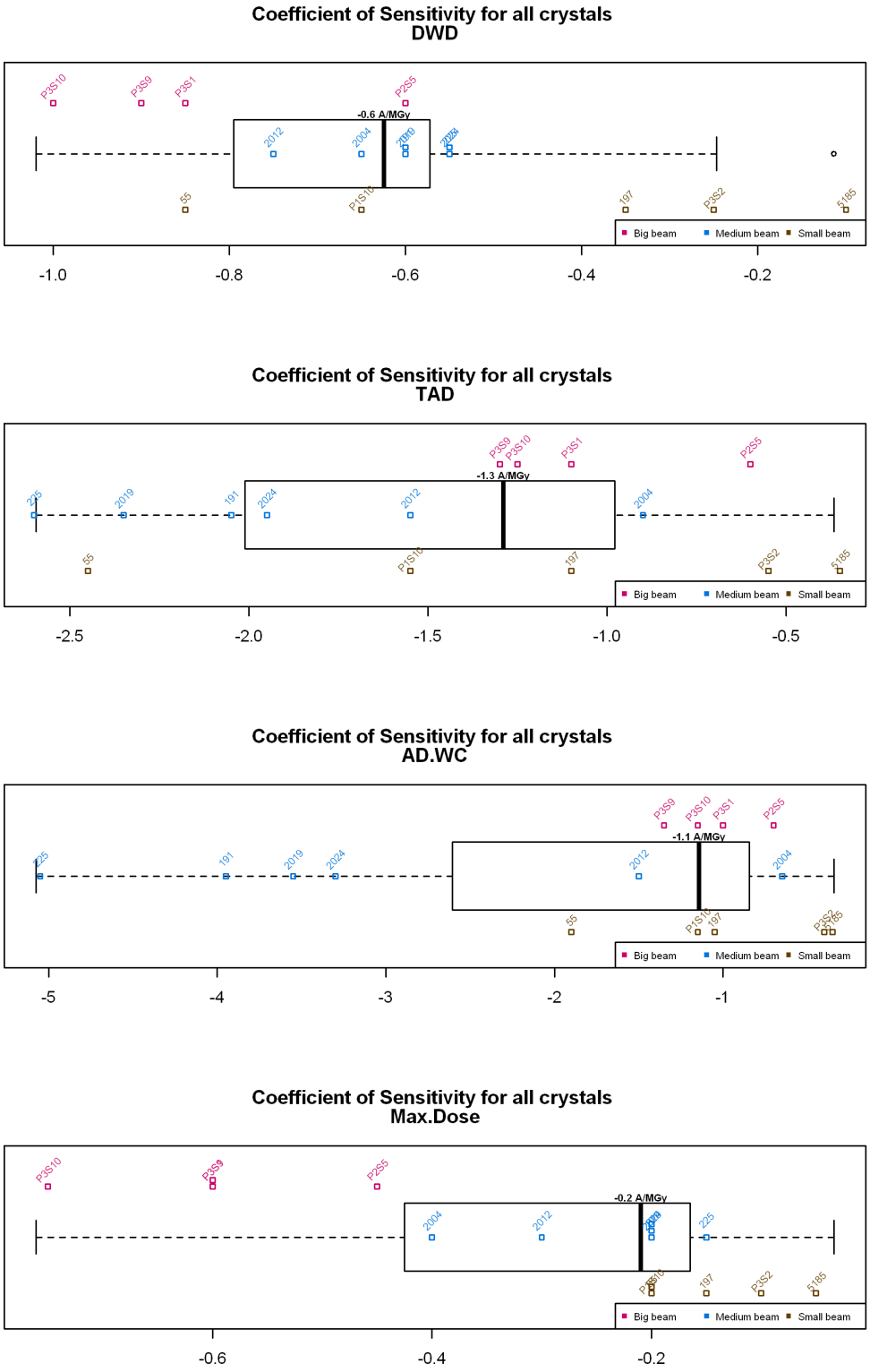


Figure 4.17: S_{AD} for each individual crystal, shown with a box and whiskers plot. The bold line with a value above it is the median, the box encloses the first and third quartile, and the whiskers extend to the furthest datum within 1.5 interquartile ranges from the upper and lower interquartile respectively. The three different beams appear to group more than for I_n/I_1, showing that B_{rel} is potentially more sensitive to dose distributions. Note that the scatter plots have been rounded so that points stack up rather than overlap to help with visual clarity, whereas the box and whiskers plot shows the raw data.

4.3.5 Real space analysis

Although the main focus of this thesis was on global damage metrics, specifically the effect of dose distributions on them, it would be highly desirable to examine if differences in the final electron density can be systematically quantified. For this reason, a preliminary investigation was carried out using whole chain Real Space Correlation Coefficients (RSCC)[12], which were explored as a potential metric. Although the approach used to calculate electron density here is the one suggested by Weik *et al.* (2000), and is the commonly accepted method for comparing subsequent damage data sets in radiation damage studies, the fixed phases may affect the quality of the damaged maps. This is because phase will affect the electron density, so the 'true' phases will also change. The optimal protocols for generating electron density maps for radiation damage studies is an interesting subject for future work.

Because the unit cell can change between data-sets during post-refinement of the damaged data, $SFCHECK$ was unable to compute RSCCs for some high damage datasets. These data were thus excluded, since by the time the damage was sufficient for a significant change in unit cell size, RSCCs would be of limited applicability anyway due to the large non-isomorphism introduced by the unit cell change. Initial plots of the data are shown in figure 4.18 for the small beam, and this demonstrated a promising level of linearity, with each individual crystal appearing to have a similar gradient, but a different intercept. For this reason, relative RSCCs ($RSCC_n^{rel} =$

[12] As implemented in $SFCHECK$ as:

$$D_{corr} = \frac{\langle \rho_{obs} \rangle \langle \rho_{calc} \rangle}{\sqrt{\langle \rho_{obs}^2 \rangle \langle \rho_{calc}^2 \rangle}}$$

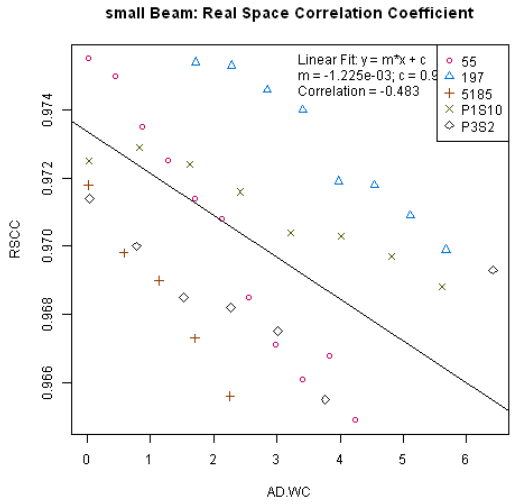
where ρ_{obs} is the mean observed density, defined by the $2F_o - F_c$ map, and ρ_{calc} is the calculated density derived from the PDB file of the model.

$RSCC_n - RSCC_1$) were plotted for all three beam classes, as shown in figures 4.19, 4.20, and 4.21. In each case, there is a general linear decrease with dose, but the spread between crystals and beam types that is observed does not offer any generalised conclusion, as shown in table 4.10. The main issues with these data are outliers, and inter-crystal variability. The big beam datasets, shown in figure 4.19, are all well bunched, except for crystal P3S10, which is a strong outlier. There is no obvious reason why this crystal demonstrates such different behaviour – despite behaving similarly to other crystals when assessed by I_n/I_1 or by B_{rel}– and the analysis was repeated to check the result. For the medium beam (figure 4.20), it is interesting to note the much reduced spread when DWD is applied, since it corrects for overall crystal size effects, as described in the caption to figure 4.8. Results from the small beam (figure 4.21) show a much larger spread between the crystals than the other two, despite a constant decreasing trend.

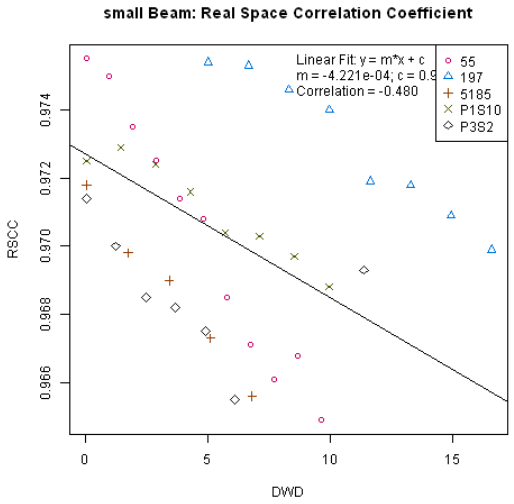
This simple approach to real space analysis of radiation damage was thus not sufficient to allow conclusions to be drawn on the effect of dose distribution on the RSCCs generated following the protocol used here. A significant limitation is that the whole-molecule RSCC used here will demonstrate small changes of the order of 0.001 CC per MGy. A more suitable candidate may be side chain RSCCs. Furthermore, since a real space analysis essentially considers an average rate of structural change, it can be expected to be closely dependent on the crystal's susceptibility to specific damage, which is highly variable, rendering this metric challenging as a predictive tool.

Beam	AD-WC	DWD	Max Dose	TAD	$N_{crystals}$
Big [$10^{-3}\Delta$CC / MGy]	-1.53	-1.14	-0.76	-1.99	4
Medium [$10^{-3}\Delta$CC / MGy]	-2.78	-1.20	-0.41	-3.09	6
Small [$10^{-3}\Delta$CC / MGy]	-1.59	-0.72	-0.24	-1.95	5

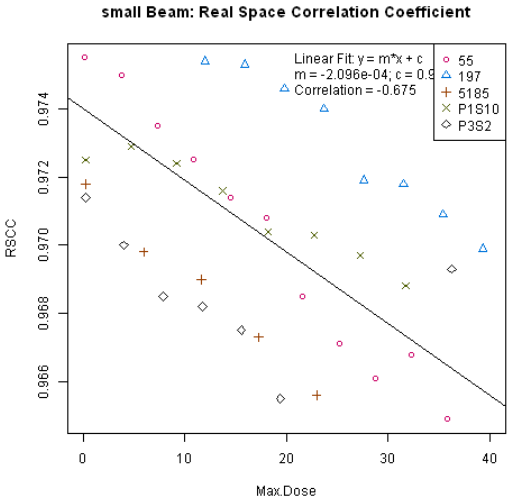
Table 4.10: Summary of the fitted slopes of the RSCC data for the three beam shapes. There are no clear trends across the various dose metrics, potentially due to the large inter-crystal variations seen in the data (figures 4.19, 4.20, and 4.21).



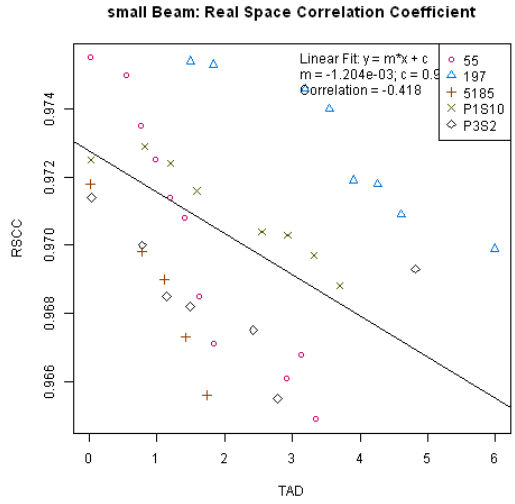
(a) RSCC *vs* Average Dose (Whole crystal) for the small beam. The x axis is in MGy.



(b) RSCC *vs* DWD for the small beam. The x axis is in MGy.

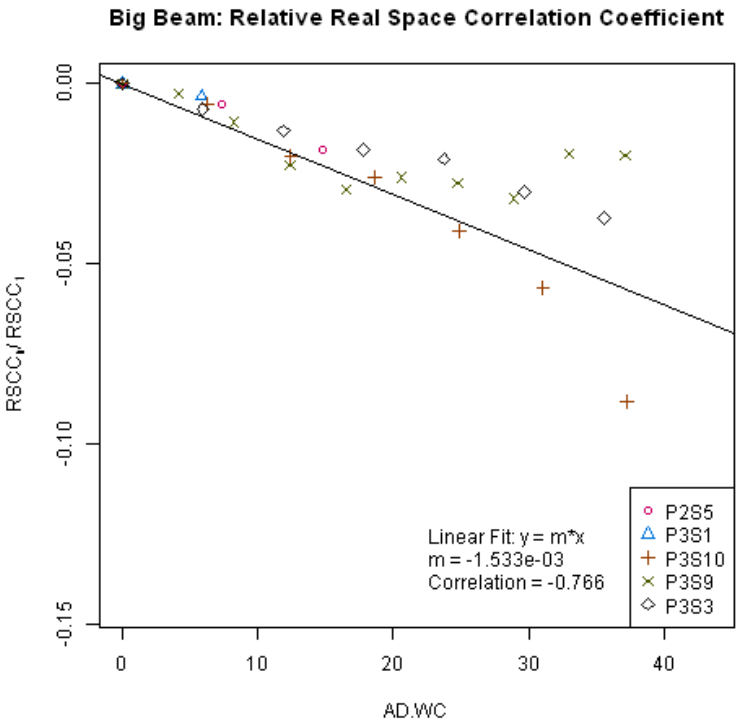


(c) RSCC *vs* Maximum Dose for the small beam. The x axis is in MGy.

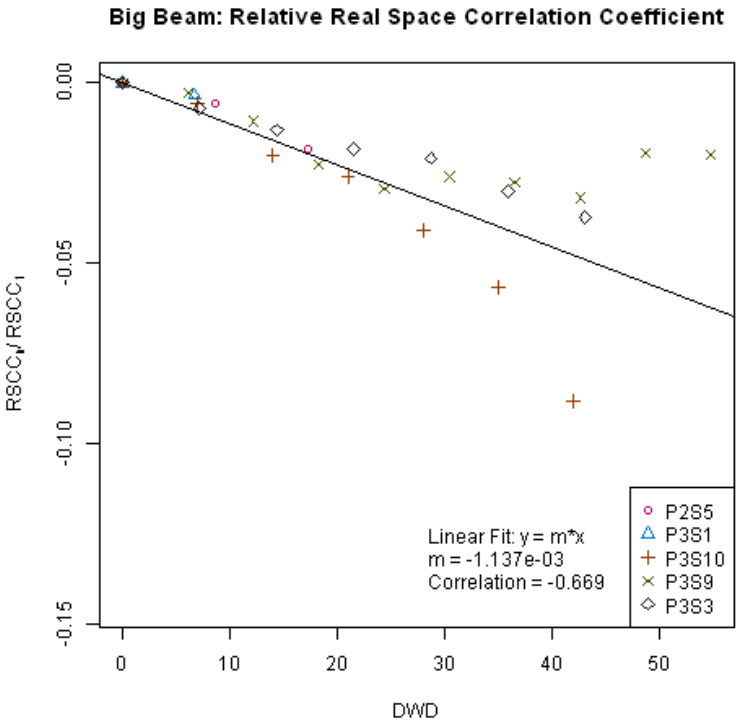


(d) RSCC *vs* TAD for the small beam. The x axis is in MGy.

Figure 4.18: Real space correlation coefficients for the small beam, plotted against each of the four dose metrics. There appears to be a fairly linear decrease for different crystals, but different starting points.

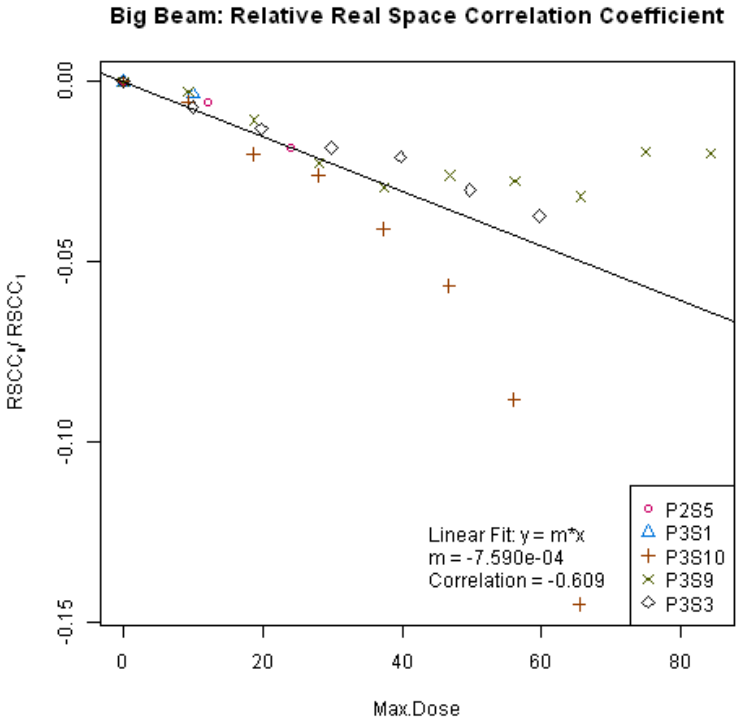


(a) Relative RSCC *vs* Average Dose (Whole crystal) for the big beam. The x axis is in MGy.

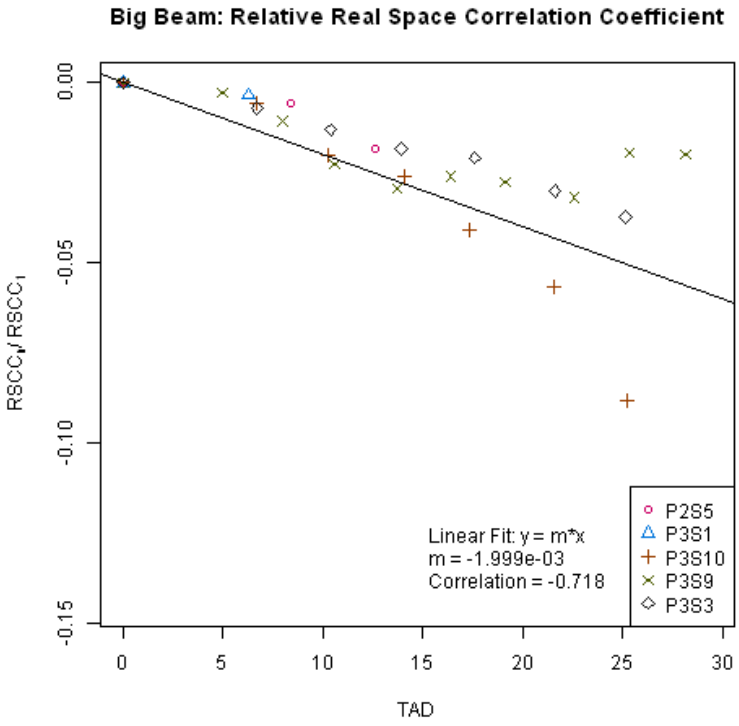


(b) Relative RSCC *vs* DWD for the big beam. The x axis is in MGy.

Figure 4.19

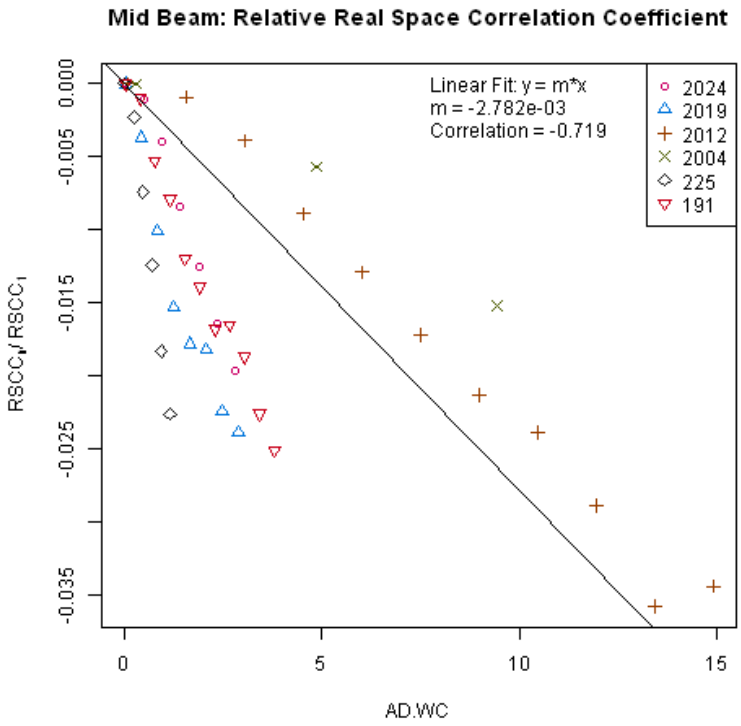


(c) Relative RSCC *vs* Maximum Dose for the big beam. The x axis is in MGy.

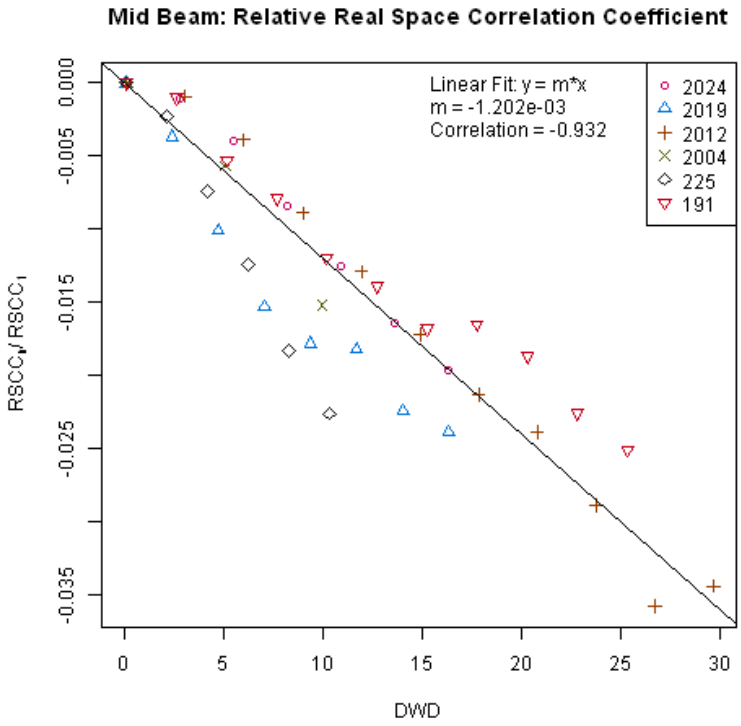


(d) Relative RSCC *vs* Threshold Average Dose (95%) for the big beam. The x axis is in MGy.

Figure 4.19: Relative real space correlation coefficients for the big beam ($RSCC_n^{rel} = RSCC_n - RSCC_1$), plotted against each of the four dose metrics. The outlier P3S10 is not an outlier when plotted against I_n/I_1 or B_{rel}, so its behaviour is assigned to intrinsic crystal variability.

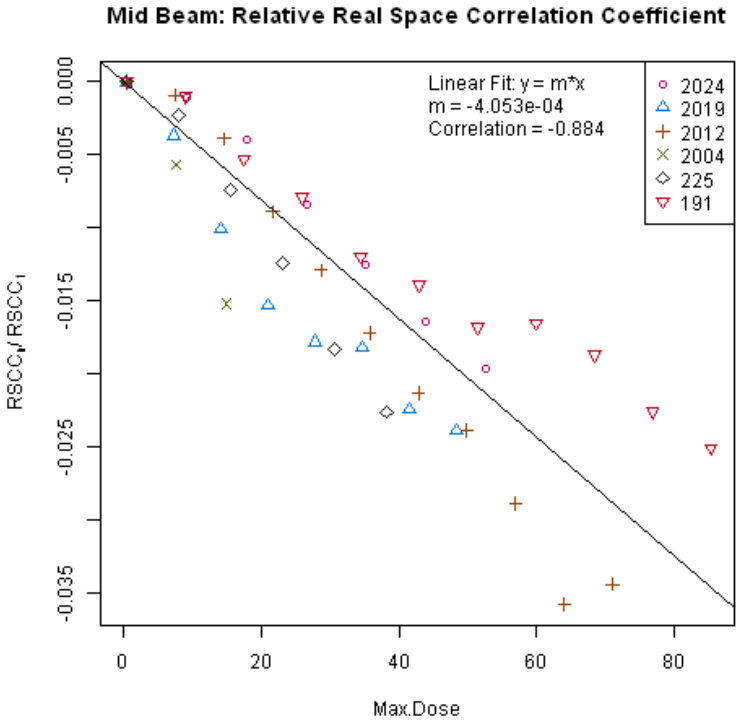


(a) Relative RSCC *vs* Average Dose, Whole crystal for the medium beam. The x axis is in MGy.

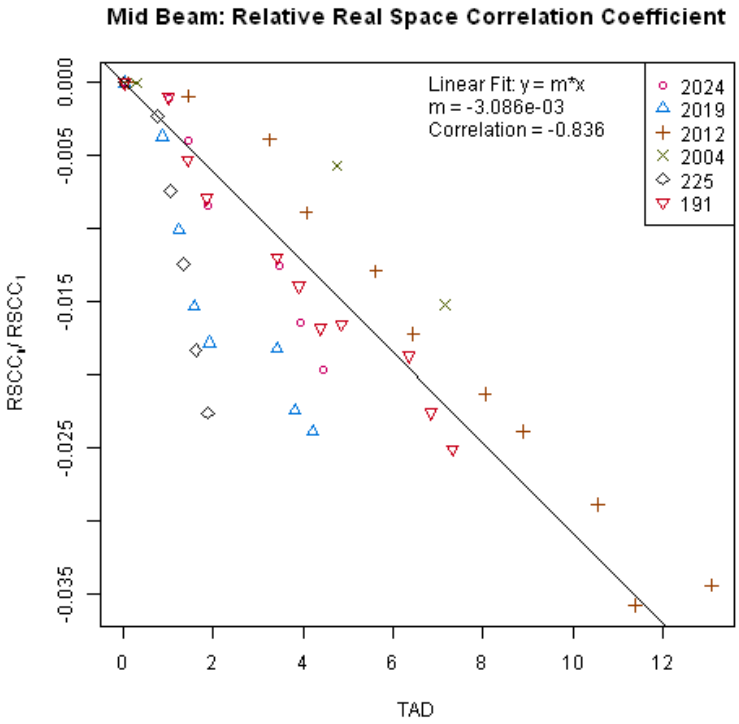


(b) Relative RSCC *vs* DWD for the medium beam. The x axis is in MGy.

Figure 4.20

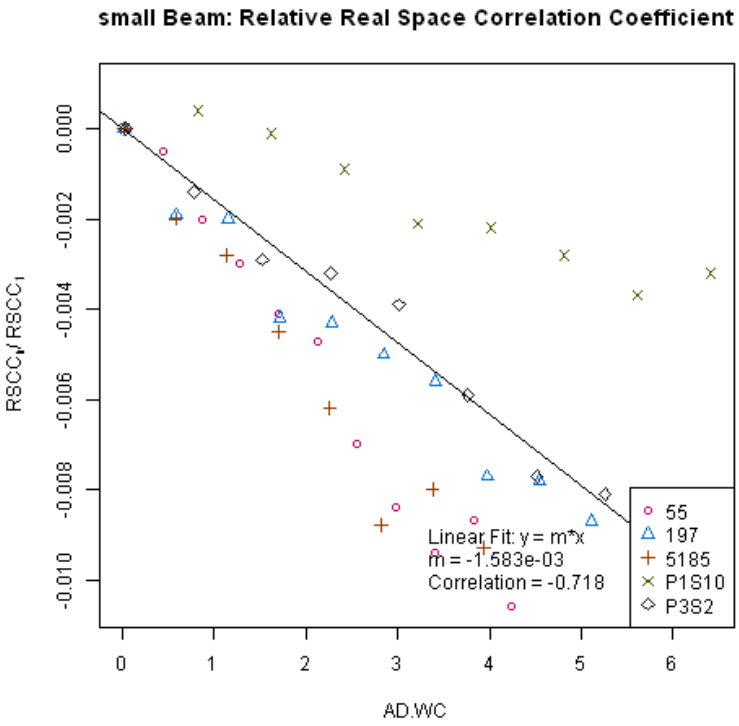


(c) Relative RSCC *vs* Maximum Dose for the medium beam. The x axis is in MGy.

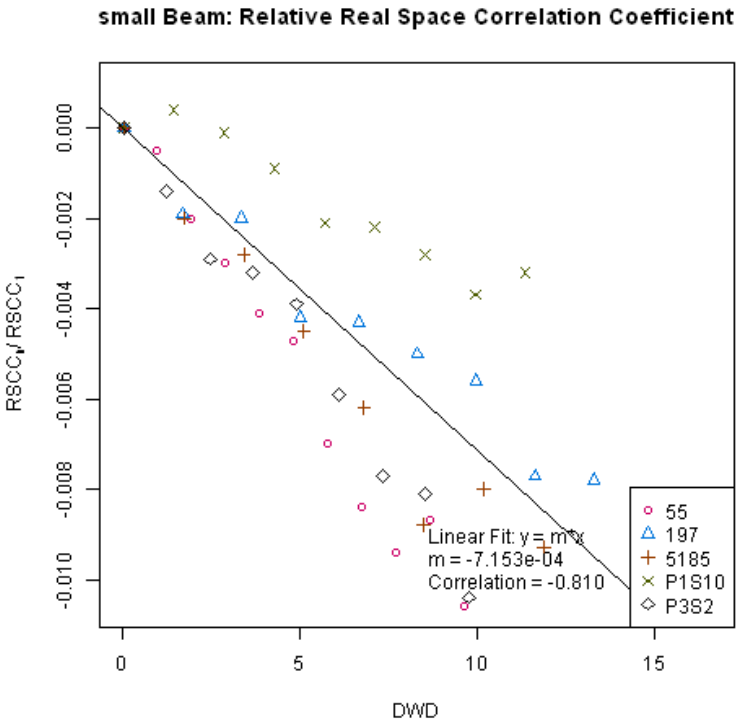


(d) Relative RSCC *vs* Threshold Average Dose (95%) for the medium beam. The x axis is in MGy.

Figure 4.20: Relative real space correlation coefficients for the medium beam ($RSCC_n^{rel} = RSCC_n - RSCC_1$), plotted against each of the four dose metrics.

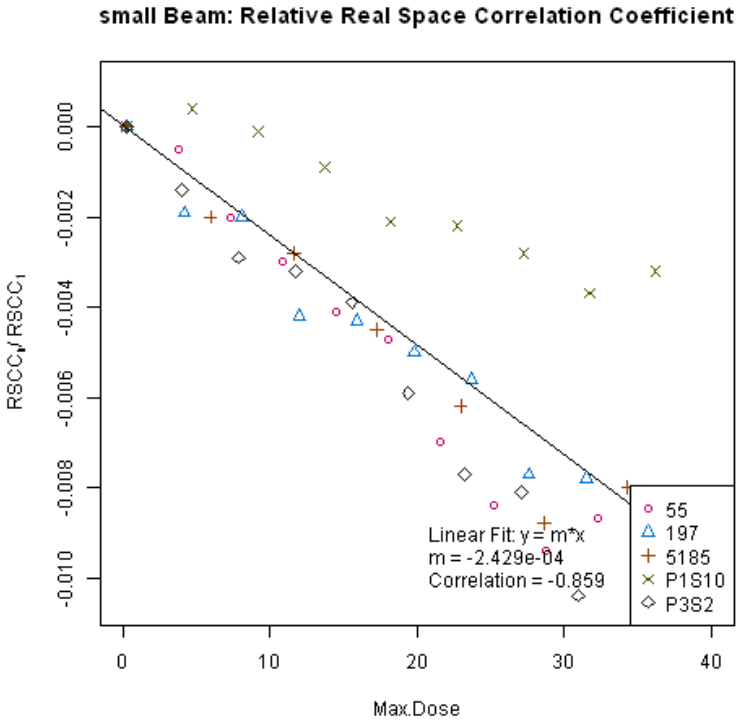


(a) Relative RSCC *vs* Average Dose, Whole crystal for the small beam. The x axis is in MGy.

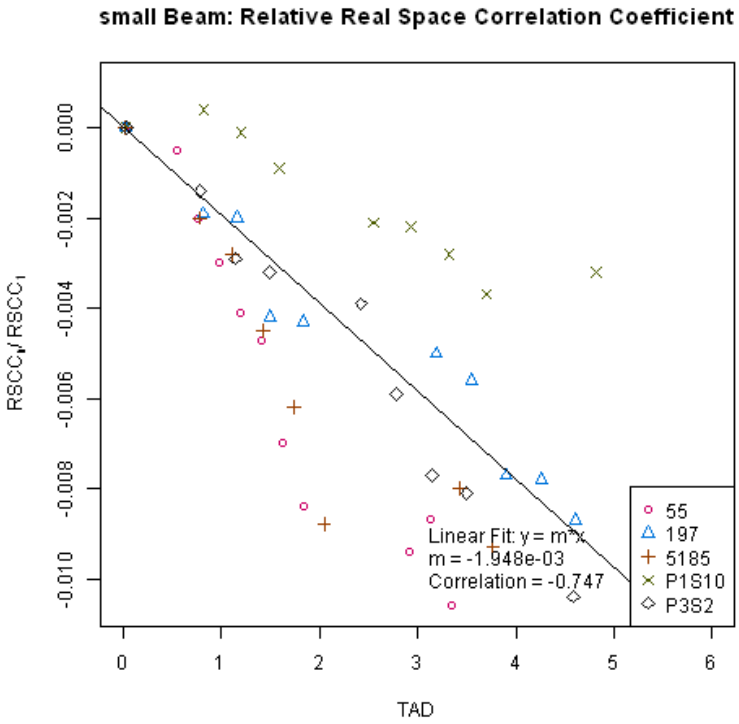


(b) Relative RSCC *vs* DWD for the small beam. The x axis is in MGy.

Figure 4.21



(c) Relative RSCC *vs* Maximum Dose for the small beam. The x axis is in MGy.



(d) Relative RSCC *vs* Threshold Average Dose for the small beam. The x axis is in MGy.

Figure 4.21: Relative real space correlation coefficients for the small beam ($RSCC_n^{rel} = RSCC_n - RSCC_1$), plotted against each of the four dose metrics.

4.4 Proof of concept

Showing that DWD is an effective metric against which to quantify radiation damage in a strategy-independent fashion is an important result, but unless it can be demonstrated that this has the potential to improve the protocols used by structural biologists, it remains of niche interest. For this reason, a proof of principle experiment was performed to apply the results found above to a challenging, biologically relevant test case.

Samples for this experiment needed to be: *(1)* biologically relevant, so representative of the kind of structures that are commonly termed 'hot structures', *(2)* crystallise in a size and shape that is suitable for modelling in RADDOSE-3D without the use of tomography, and *(3)* produce crystals which are homogenous enough that it is possible to collect several datasets from comparable crystals, or large enough to be able to collect several datasets from a single crystal.

Based on these requirements, crystals of $PepT_{St}$ (Solcan *et al.*, 2012) were provided by Nicolae Solcan and Dr. Simon Newstead from the Department of Biochemistry, University of Oxford. These crystals represent a recent challenging crystallographic problem, and are a good example of a biologically important sample that could benefit from improved methodology during the collection of diffraction data.

4.4.1 About $PepT_{St}$

$PepT_{St}$, from the bacterium *Streptococcus thermophilus*, is a member of the proton-dependent oligopeptide transporter (POT) family of transmembrane proteins. There is high sequence conservation between $PepT_{St}$ and the mammalian peptide transporters $PepT1$ and $PepT2$, such that insights gained from structural studies of bacterial POT family proteins have important implications for understanding peptide absorption and retention in humans. Additionally, these proteins are important for the absorption of drug compounds that are sterically similar to peptides, such as β-lactam antibiotics (as described in Solcan *et al.* (2012)). The first crystal structure of a POT family transporter ($PepT_{So}$, from *Shewanella oneidensis*, PDB code 2XUT, resolution 3.6 Å) was recently solved (Newstead *et al.*, 2011), and the additional structure of $PepT_{St}$ (PDB code 4APS, resolution 3.3 Å) reveals a different, inward facing conformation that offers new insight into the role of salt bridges during conformational changes allowing peptide transport. The cartoon structure of $PepT_{St}$ is shown in figure 4.22.

$PepT_{St}$ presents several major crystallographic challenges:

- The diffraction is highly anisotropic. For example, the $I/\sigma(I)$ in the high resolution shell for the EDNA strategy (see below) is 0.06 along a^*, 4.24 along b^*, and 0.09 along c^*, in space group $P2_12_12_1$.

- Diffraction is poor: only one crystal in 50 diffracts to ~3.2 Å (Simon Newstead, private communication). Typical diffraction is in the range of 4-6.5 Å.

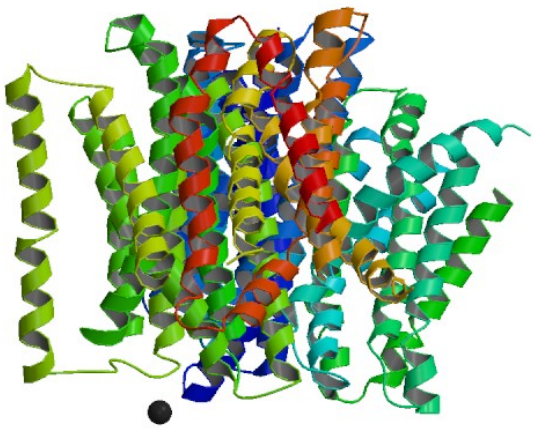


Figure 4.22: Cartoon structure of $PepT_{St}$, showing one of the two monomers that make up the asymmetric unit, along with it's coordinated cadmium ion (bottom left). Figure taken from the PDB 4APS, resolution 3.3Å.

- The protein is on the large side, with 491 residues and 12 transmebrane helices.

It is therefore very important to be able to obtain the maximum possible number of data from the few crystals that are well ordered. This is a good example of a typical challenging case in crystallography, where – once the correct conditions have been obtained – many crystals can be grown, but the best diffracting ones are rare, and so the strategy must be highly optimized to collect the best data from them.

4.4.2 Methods

In order to validate the use of RADDOSE-3D as a tool to guide improved data collection protocols, a 'state of the art' dataset based on current best practices in crystallography was used as a benchmark against which to compare a RADDOSE-3D guided data collection. Experiments were performed at the ESRF beamline ID14-4, with the beamline parameters set as described in section

4.2. The EDNA strategy prediction package available online at the ESRF allows highly optimised data collection, and makes use of the $BEST$ strategy program (Popov and Bourenkov, 2003), and also uses RADDOSE v2 (Paithankar *et al.*, 2009) to estimate the radiation damage tolerance of crystals (Bourenkov and Popov, 2010). The EDNA guidelines are those that a vast majority of crystallographers would use to predict optimal strategies for collection of data on a CCD detector with non-negligible readout noise. The proposed RADDOSE-3D protocol is superficially similar to the characterisation procedure suggested by Leal *et al.* (2011), but with two important differences: firstly, it uses the I_n/I_1 metric rather than B_{rel}, since we have shown above that this is more transferable crystal-to-crystal, and secondly, the use of DWD allows a characterisation performed using one strategy (e.g. 10° wedges) to be applied to complex, three-dimensional strategies such as helical or offset scans. This second point is particularly important, since it presents the possibility of truly optimising total elastic scattering over available crystal volume and diffraction lifetime. Not having to use the same strategy for characterisation as for collection is thus transformative for the application of predictive damage models, and moves the use of this from niche interest to potentially mainstream. The protocol used to characterise the $PepT_{St}$ crystal for these validation experiments was:

1. Characterise the loss of diffraction intensity of the crystal, either using a poorly diffracting crystal of the same type, or a corner region of the crystal that will be used for collection. Characterisation is performed using the Probe-Burn protocol.

2. Use this characterisation to establish a $D_{1/2}$, and then design a strategy that optimises

elastic yield within a tolerable level of damage. For the preliminary experiments here, the average DWD corresponding to I_n/I_1 of 0.85 was used as a cutoff (Blundell and Johnson, 1976). This corresponds roughly to an I_n/I_1 of 0.7 for the last frame of the dataset, the lower limit suggested in both Blundell and Johnson (1976) (0.67) and Owen *et al.* (2006) (0.7)[13].

A large crystal ($56 \times 350 \times 125$ μm ^{3}) of $PepT_{St}$ that diffracted to better than 4 Å (along the best direction, the b^* axis) was identified, allowing multiple datasets to be collected with a well matched beam size (FWHM $\sim 40 \times 60$ μm ^{2}, slits set to 50×60 μm ^{2}), as shown in figure 4.23 and figure 4.24. The crystal was oriented using the mini-kappa goniometer at ESRF ID14-4 so that its long axis was close to the rotation axis (to allow multiple datasets to be collected without overlapping the footprints of the beam on the crystal) while also ensuring that the scattering from the well diffracting b^* axis of the crystal was not shadowed by the backstop on the diffraction pattern. The final orientation, with a P angle of approximately $17°$, offered the best compromise. Four positions on the crystal were used , as shown in figure 4.23, to collect: a best-practice EDNA dataset, a characterisation procedure, and two RADDOSE-3D guided data collections, one being a $180°$ standard collection, and the other a $360°$ offset strategy. Separation between collection sites was estimated allowing for a 4 μm gap[14] beyond the beam delimiting box (set to the approximate beam FWHM using the phosphor screen and the online microscope) as a minimum

[13] This is significantly higher than the $I_n/I_1 > 0.4$ criterion used to fit the data in the previous section. This is because the 0.7 is a limit for 'good data', whereas 0.4 was selected to extend to beyond this, allowing more global trends to be analysed.

[14] For photoelectron propagation.

for each centroid. For each dataset, crystal quality was assessed using the EDNA prediction tool, and data were processed using $MOSFLM$ and $AIMLESS$ from the CCP4 suite (Winn *et al.*, 2011).

4.4.3 Current state of the art: EDNA strategy.

EDNA characterisation of the data at position 1, and $\Phi = 35.6°$ indicated a high resolution limit of 3.89 Å, based on average radiation damage rates and the quality of a test image. 71 images of width $1.25°$, with 2.65 s of exposure at a flux of 5.15×10^{11} ph/s, at 12.658 keV were collected, starting at $\Phi = 169°$. The wedge was subsequently manually processed in $MOSFLM$ without post-refinement (since the resolution was too low for this to be a valid approach). The $CC_{1/2}$ = 0.5 resolution was 3.86Å[15], in close agreement with the EDNA prediction. The statistics for this collection, along with a progressive, image-by-image simulation of DWD are reported in figure 4.25. A table containing the summary of the table 1's for all three compared strategies can be found in section 4.4.7. The dose distribution for this strategy can be seen in figure 4.26a.

4.4.4 Evaluating the crystal, and planning an optimised strategy.

In order to characterise the crystal decay, 30 repeated $10°$ wedges were collected with 1s/ ° at a flux of 5.15×10^{11} ph/s, starting at $\Phi = 169°$, and ending at $\Phi = 179°$. The total intensity loss after this exposure set was still quite low (~0.75 I_n/I_1), so a further 10 sets of Probe + Burn $10°$ wedges were collected, with 1 s/ ° for the probes, and 5 s/ ° for the burns. A

[15]The alternative Mn(I/sd) > 2.00 limit was 4.42 Å

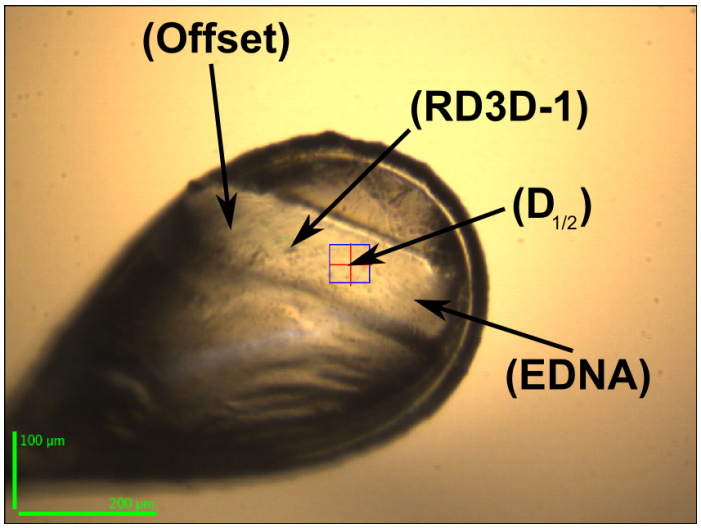


Figure 4.23: The crystal used for comparative evaluation of RADDOSE-3D based strategies. EDNA indicates the region where the benchmark EDNA dataset was collected, $D_{1/2}$ the region used to characterise the crystal, RD3D-1 where the first RADDOSE-3D optimised, 180° dataset was collected, and Offset the region where the 360° offset data collection was collected. Full details of each data collection can be found in the main text. The blue box indicates the approximate FWHM of the uncollimated beam. The rotation axis is across the width of the page.

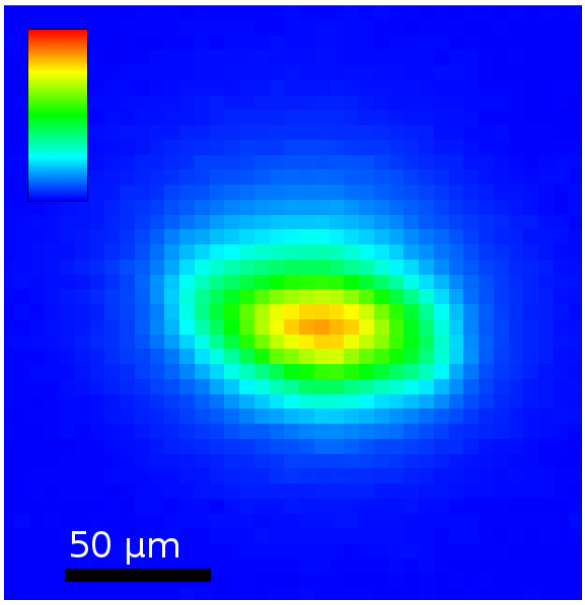


Figure 4.24: False colour image of the beam used to collect data from the crystals of $PepT_{St}$. The FWHM of the beam ($\sim 40 \times 60$ μm ^{2}) was set to match the crystal dimension. Dark current was 20 intensity units, and the corrected dynamic range is 0-106.

few datasets (1, 15, 30, 35, and 41) were then processed manually ($MOSFLM$ + $AIMLESS$) during data collection to characterise $D_{1/2}$, which was found to be 18.8 MGy, as shown in figure 4.27. This corresponds to a DWD of ~5.5 MGy for a I_n/I_1 of 0.85. An important question is then what to do with this value: the results above show that 5.5 MGy will correspond to an I_n/I_1 of 0.85, but should this be the dose target for the last image of the dataset, or for the average? Furthermore, the I_n/I_1 guideline of 0.85 is highly empirical, and there is the potential to improve this with more work. Two strategies were thus devised for the RADDOSE-3D guided data collections: one using a simple $180°$ wedge, allowing higher multiplicity and the spreading of dose over more measurements of each hkl. with the average DWD set to 5.5 MGy, and a second strategy with a lower average DWD, and the dose more efficiently spread by offsetting the rotation axis by 10 μm using the z-offset motor in the goniometer.

4.4.5 RADDOSE-3D Strategy 1

For the first RADDOSE-3D data collection, 180 images of width $1°$, with 2.2 s of exposure per image with a flux of 5.2×10^{11} ph/s, at 12.658 keV were collected, starting at motor position $\Phi = 154.6°$. In what follows, this data collection wedge is referred to as RADDOSE-3D-1. The wedge was subsequently manually processed in $MOSFLM$ without post-refinement (since the resolution was too low for this to be a valid approach). The $CC_{1/2}$ = 0.5 resolution was 4.18 Å, and the Mn(I/sd) > 2.00 limit was 6.04 Å, which is worse than for the EDNA strategy. Subsequent analysis, running the EDNA software to find a ranking resolution (defined as the estimated best

Position:	EDNA strategy	RADDOSE-3D-1	Offset
Ranking Resolution:	5.05 Å	>6.5Å	6.58 Å

Table 4.11: Expected achievable resolutions for the three positions, as estimated by EDNA. The RADDOSE-3D-1 was not possible to characterise reliably on-line due to a sub-optimal choice of test image orientation.

resolution achievable with $I/\sigma(I) > 3$ in the high resolution shell, given the total exposure time estimated by BEST Bourenkov and Popov (2010)) for the RADDOSE-3D-1 position on the crystal showed that, for $\Phi = 155°$, the RADDOSE-3D-1 wedge had a ranking resolution below 6.5 Å. In contrast, the ranking resolution for wedge collected at the position where the EDNA suggested strategy was exposed was 5.05 Å, at orientation $\Phi = 350°$, which is nearly $180°$ apart from $\Phi = 155°$, offering comparative metrics. Images along the same Φ were not collected, since it was erroneously assumed that crystal diffraction would be of similar quality in both positions. A summary of ranking resolutions for all three positions is shown in table 4.11. The statistics for this collection, along with a progressive, image-by-image simulation of DWD are reported in figure 4.28. The initial DWD of 2.5 MGy shown in figure 4.28b highlights how the crystal was not optimally translated after the highly damaging probe datasets. This explains why the DWD of 7.35 MGy in figure 4.28a is higher than intended[16], since the simulation used to design the strategy was performed on a fresh crystal.

4.4.6 RADDOSE Strategy 2: offset

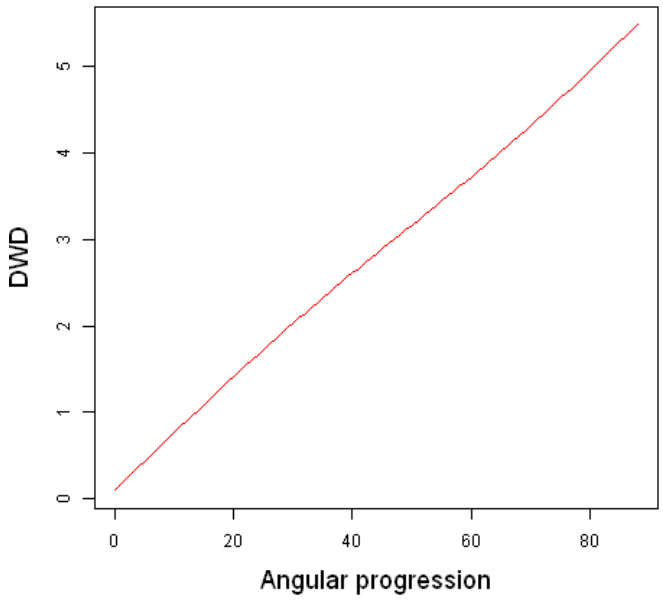
For the final section of available crystal volume, the same total exposure as the first RADDOSE-3D strategy was used, but a more efficient dose distribution was applied: a full $360°$ rotation, with

[16] Simulated DWD is 5.5 MGy for a fresh crystal.

a 10 μm mis-seting of the rotation axis, leading to a dose distribution with more 'spread'. 360 images of width $1°$, with 1.1 s of exposure at a flux of 5.5×10^{11} ph/s, at 12.658 keV were collected, starting at $\Phi = 154.6°$. The wedge was subsequently manually processed in $MOSFLM$ without post-refinement (since the resolution was too low for this to be a valid approach). The $CC_{1/2}$ = 0.5 resolution was 3.87 Å, and the Mn(I/sd) > 2.00 limit was 4.26 Å. Subsequent analysis, running EDNA to find a ranking resolution showed that, at $\Phi = 155°$, the offset strategy had a ranking resolution of 6.58 Å *vs* 5.05 Å for the EDNA location at $\Phi = 350°$, which is nearly $180°$ apart. The statistics for this collection, along with a progressive, image-by-image simulation of DWD are reported in figure 4.29. The comparable quality of the statistics in table 4.29a with those in table 4.25a shows that, despite being collected from a much worse part of the crystal, that was also already somewhat damaged, the offset strategy, with a higher dose was able to yield comparable results to those from the EDNA suggested strategy.

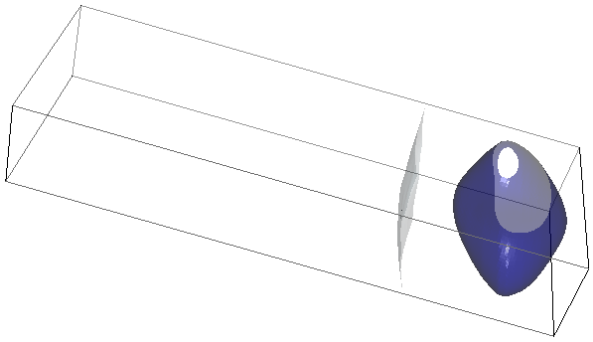
	Overall	InnerShell	OuterShell
Average DWD (MGy)	2.8 MGy	-	-
Low res limit (Å)	48.28	48.28	3.93
High res limit (Å)	3.68	10.40	3.68
Rmerge	0.166	0.018	1.782
Rmeas	0.200	0.022	2.190
Rpim	0.109	0.013	1.253
Total obs	83005	3265	12998
Total unique	25946	1118	4540
Mean((I)/sd(I))	5.7	57.0	0.6
CC(1/2)	0.991	0.999	0.405
Completeness (%)	98.3	89.1	96.3
Multiplicity	3.2	2.9	2.9
Unit Cell (Å)	$(91.09 \times 113.7 \times 227.7\ \mu\text{m}^3)$		
Mosaicity (°)	0.89	-	-

(a) Statistics for the EDNA data collection.

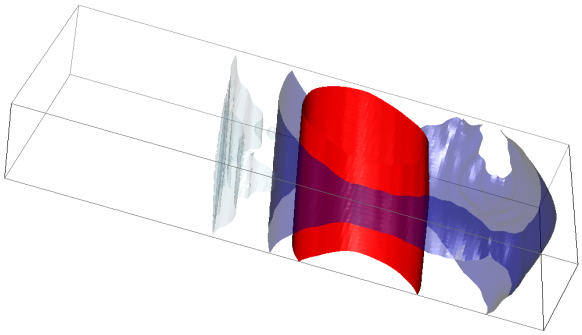


(b) Instantaneous DWD for each angular position during data collection for the EDNA strategy.

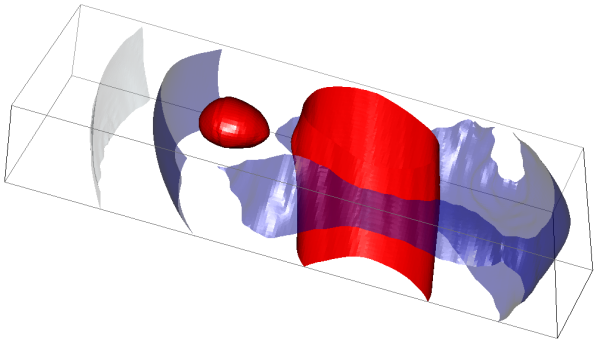
Figure 4.25: $PepT_{st}$ data statistics and dose estimation, using the EDNA suggested strategy.



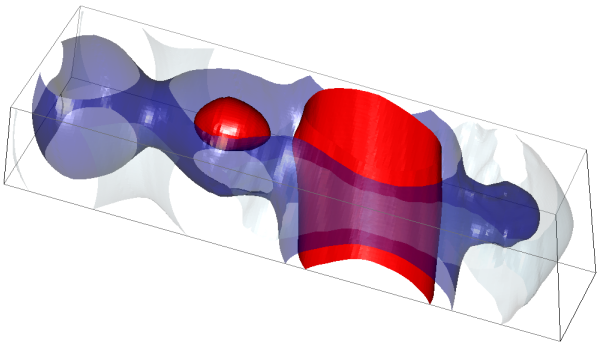
(a) Dose state after the EDNA collection. Light blue: 0.5 MGy, dark blue: 5.5 MGy, red: 18.8 MGy.



(b) Dose state after characterisation.Light blue: 0.5 MGy, dark blue: 5.5 MGy, red: 18.8 MGy.



(c) Dose state after the first RADDOSE-3D strategy. Light blue: 0.5 MGy, dark blue: 5.5 MGy, red: 18.8 MGy.



(d) Dose state after the final offset strategy. Light blue: 5.5 MGy, dark blue: 10 MGy, red: 18.8 MGy. Note the different isosurfaces, and the better spread of the dose for the leftmost offset strategy.

Figure 4.26: Progression of the dose state of the crystal during the four exposures. The translation, based on the approximate beam size was not large enough, as only the beam collimation was taken into account, not the experimental profile, leading to the RADDOSE-3D and offset strategies starting at a non-negligible DWD.

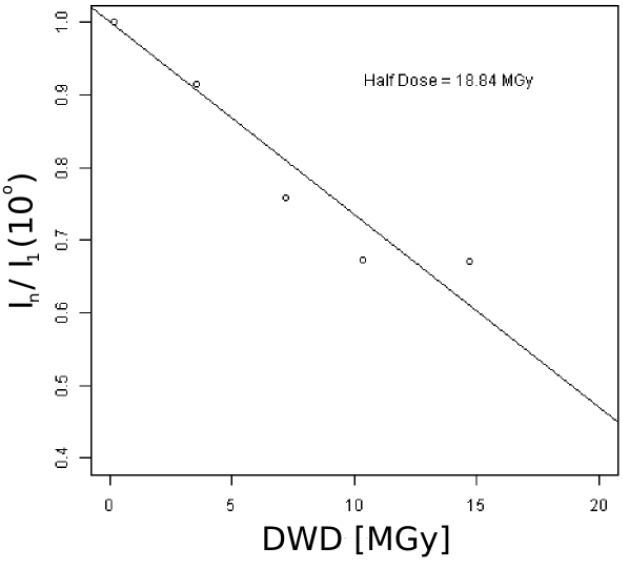
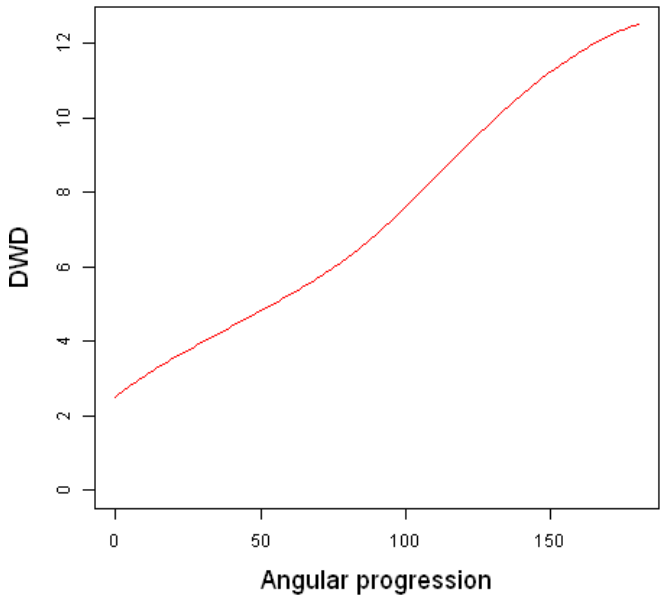


Figure 4.27: Initial estimation of $D_{1/2}$ calculated online at the ESRF, which was then used to design optimised strategies.

	Overall	InnerShell	OuterShell
Average DWD (MGy)	7.35 MGy	-	-
Low res limit (Å)	48.52	48.52	3.93
High res limit (Å)	3.68	10.41	3.68
Rmerge	0.307	0.045	3.099
Rmeas	0.350	0.050	3.611
Total obs	134889	7173	20277
Total unique	26171	1159	4564
Mean((I)/sd(I))	1.0	24.3	-3.7
CC(1/2)	0.956	0.998	0.127
Completeness (%)	98.4	90.1	96.3
Multiplicity	5.2	6.2	4.4
Unit Cell (Å)	$(91.34 \times 113.7 \times 229\ \mu m^3)$		
Mosaicity (°)	0.95	-	-

(a) Statistics for the first RADDOSE-3D based data collection. The dose was higher than planned due to overlap with previous characterisation wedge, as shown in figure 4.26.

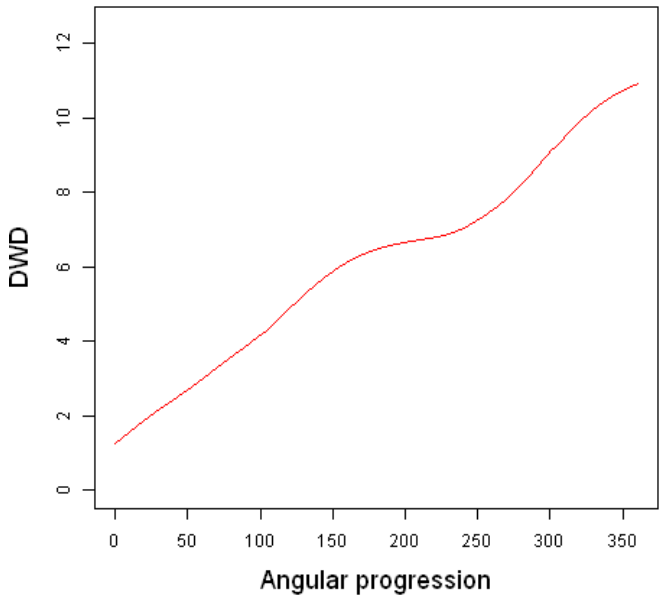


(b) Instantaneous dose for each angular position during data collection for the first RADDOSE-3D strategy. DWD has been calculated starting with the dose state at the end of the EDNA and characterisation wedges, as shown in figure 4.26.

Figure 4.28: $PepT_{st}$ data statistics and dose estimation for the first RADDOSE-3D strategy.

	Overall	InnerShell	OuterShell
Average DWD (MGy)	6.10 MGy	-	-
Low res limit (Å)	48.43	48.43	3.93
High res limit (Å)	3.68	10.41	3.68
Rmerge	0.326	0.060	5.527
Rmeas	0.339	0.062	5.745
Total obs	355762	15490	63733
Total unique	26481	1210	4723
Mean((I)/sd(I))	5.0	25.8	0.8
CC(1/2)	0.997	0.998	0.421
Completeness (%)	99.7	93.9	99.8
Multiplicity	13.4	12.8	13.5
Unit Cell (Å)	$(991.23 \times 113.7 \times 228.6\ \mu\text{m})$		
Mosaicity (°)	0.70	-	-

(a) Statistics for the offset data collection. This DWD was ~ 2 MGy higher than estimates based on simulations with a fresh crystal, but still close to the 5.5 MGy limit that was targeted.



(b) Instantaneous dose for each angluar position during data collection for the offset strategy. DWD has been calculated starting with the dose state at the end of the first RADDOSE-3D wedges. The 'lump' in the distribution is due to the deeper orientation of the crystal passing through the beam, bringing in undamaged volume at a higher rate. Starting DWD is 1.25 MGy.

Figure 4.29: $PepT_{st}$ data statistics and dose estimation for the offset strategy.

4.4.7 Discussion

It is difficult to draw clear conclusions from these preliminary data, since *(a)* the crystals are poorly diffracting enough to make a real-space analysis using difference maps inconclusive, *(b)* there was a larger than anticipated inhomogeneity of crystal quality within the single crystal used for data collection, and *(c)* the individual collections overlapped more than was anticipated. The wedge separation was based on estimations using the beam's approximate FWHM, plus 4 μm, as a putative sufficient separation: clearly, this was insufficient, and predictive simulations would have prevented this.

Additionally, the use of EDNA ranking resolution for non-identical Φ values is sub-optimal for assessing crystal quality, and more work must be done to develop a better protocol for benchmarking different crystal regions before comparison. Despite this, the size of the shifts ($\sim$1.5Å) are sufficient to be confident that the offset location was qualitatively worse than the EDNA location, and thus to have faith in the result that the RADDOSE-3D optimised offset strategy improved the quality of obtainable data.

Point *(c)* is itself a valuable result, demonstrating how, even for users who are highly aware of radiation damage issues, it is easy to mis-estimate the separation of multiple wedges on a single crystal. When used online – automated into data collection systems – RADDOSE-3D would be able to ensure that multiple wedges collected from a single crystal are optimally spaced; an immediate and useful application of the program. The discrepancy between the expected profile based on

the collimation, and the experimental profile, could have many causes[17], and further investigation is warranted.

Despite the challenges presented by the inhomogeneity of the crystal, it is clear that current, state of the art strategies can underestimate the allowable dose to a single location on a crystal due to not taking the whole, 3D picture into account. As a comparison, the EDNA strategy had a total elastic yield[18] of 1.5×10^{11} photons, the RADDOSE-3D-1 strategy 3.1×10^{11} photons, and the Offset strategy 2.5×10^{11} photons. This increased yield is of course only valuable if there is not significant radiation damage in the later images that make up the dataset. In order to check this, the first and second $180°$ of the offset strategy were processed independently, and table 4.12 shows that there is no significant difference in the data collection statistics for these two wedges, despite the second being at a higher dose state; in particular, the resolution limit as assessed by $CC_{1/2} = 0.5$ does not change.

For the RADDOSE-3D-1 wedge, statistics for the first $90°$ were compared to the whole $180°$, both at 3.86 Å (for comparison to the other data collections) and at its own limiting resolution of 4.18 Å. The R_{meas} and $CC_{1/2}$ are consistently better for the whole $180°$ *vs* the first $90°$. The second $90°$ of this strategy was not independently analysed, since it probes a different region of the crystal volume to the first $90°$, and would thus not offer any comparative value.

Overall, although not comprehensive, this proof of principle experiment has highlighted a number

[17] For example scattered X-rays, inaccuracies in the slit motors, bleed between pixels on the detector, issues with the aluminium scintilator, diffraction from the collimator

[18] Which is a measure of the total elastic scattering from the crystal, not the total diffraction.

	Hires limit (Å)	R_{meas}	$I/\sigma(I)$	$CC_{1/2}$	Completeness (%)
EDNA	3.68	0.20 (1.8)	5.7 (0.6)	0.991 (0.405)	98.3 (96.3)
RADDOSE-1	3.68	0.35 (3.6)	1.0 (-3.7)	0.956 (0.127)	98.4 (96.3)
RADDOSE-1 $0-90°$	3.68	0.73 (4.8)	9.2 (0.6)	0.476 (0.010)	89.7 (88.4)
RADDOSE-1	4.18	0.21 (1.2)	4.6 (1.1)	0.988 (0.796)	99.0 (99.3)
RADDOSE-1 $0-90°$	4.18	0.37 (1.6)	3.5 (0.9)	0.937 (0.502)	90.3 (90.1)
OFFSET	3.68	0.34 (5.7)	5.0 (0.8)	0.997 (0.421)	99.7 (99.8)
OFFSET $0-180°$	3.68	0.24 (4.5)	6.3 (0.7)	0.997 (0.328)	99.6 (99.6)
OFFSET $180-360°$	3.68	0.25 (4.1)	5.4 (0.7)	0.996 (0.385)	99.1 (97.6)

	Multiplicity	Res limit (Å): $CC_{1/2} > 0.5$	Res limit (Å): $I/\sigma(I) > 2$	Mosaicity (°)	Unit Cell (a,b,c) (Å)
EDNA	3.2 (2.9)	3.86	4.42	0.89	(91.09, 113.7, 227.7)
RADDOSE-1	5.2 (4.4)	4.18	6.04	0.95	(91.34, 113.7, 229)
RADDOSE-1 $0-90°$	3.5 (3.5)	4.44	4.69	0.81	(91.34, 113.7, 229.1)
RADDOSE-1	5.8 (5.5)	4.18	5.68	0.95	(91.34, 113.7, 229)
RADDOSE-1 $0-90°$	3.7 (3.8)	4.18	6.04	0.81	(91.34, 113.7, 229.1)
RADDOSE-1 $0-90°$	3.7 (3.8)	4.18	6.04	0.81	(91.34, 113.7, 229.1)
OFFSET	13.4 (13.5)	3.85	4.25	0.7	(91.23, 113.7, 228.6)
OFFSET $180-360°$	6.3 (6.1)	3.9	4.32	0.79	(91.23, 113.7, 228.5)
OFFSET $180-360°$	5.8 (5.6)	3.88	4.38	0.92	(91.4, 113.7, 228.5)

Table 4.12: Comparison of statistics for all wedges collected on the $PepT_{St}$ crystal. Note that the early, complete, parts of the RADDOSE-3D-1 and Offset datasets are both worse than the overall wedges, indicating that they were not excessively damaged in the latter half of data collection. The Offset and EDNA stategies are also highly comparable, especially in terms of $CC_{1/2}$, despite the worse diffraction for the Offset region of the crystal, as Characterised by EDNA.

of valuable points:

- Current state of the art strategy prediction programs can underestimate as well as overestimate the dose tolerance of a crystal, meaning that it is sometimes possible, as shown by the offset collection, to improve upon these protocols with dose modeling.

- Real time modelling of data collections using experimentally parametrised beams can allow multiple wedges to be collected at optimal spacings, without wasting any crystal volume.

- More work is required to better understand how to use the predictive power of DWD to generate optimal strategies. In particular, either a better understanding of I_n/I_1 thresholds, or an alternative metric to I_n/I_1 which faithfully reflects data quality.

Chapter 5

Optimising the Spatial Distribution of Dose

Selecting the optimal strategy for data collection is an important problem in macromolecular crystallography (MX), and guidelines exist for defining suitable experimental settings based on crystal parameters (Dauter, 2005; Bourenkov and Popov, 2006). Under the assumption that crystal size is matched to the beam, Bourenkov and Popov (2010) have extended this work to take account of the loss of diffraction intensity during data collection.

In cases where the even-dose approximation – with the crystal size matched to a uniform beam profile – does not hold, an extra parameter for the optimisation of MX experiments presents itself: three-dimensional strategy, whereby how efficiently dose is spread throughout the crystal volume is also taken into account. Research on comparing three strategies: Standard[1], Helical[2], and Translational[3] was the subject of work after the development of the first, feature complete version of the code (Zeldin *et al.*, 2013a). Since this was undertaken before the full experimental investigation of the metrics, and before elastic yield was output from RADDOSE-3D, Zeldin *et al.* (2013a) focused on using a metric named 'Dose Inefficiency', the Max Dose per unit absorbed energy, as described in Chapter 3. The three strategies were simulated for four common crystal shapes (cube, plate, long and short needles) and for a wide range of FWHM beams, and the results tabulated for reference.

[1] Standard: a 90° wedge collected symmetrically across one edge of the crystal with the beam centred for all dimensions of the crystal.

[2] Helical: as for Standard, but translating the crystal along the goniometer axis during data collection. The translation was defined so that at the start and end positions, the beam centre was half of its FWHM from the crystal edge. For example, the centre of a 10 μm $\times$ 10 μm beam on a Plate (200 μm $\times$ 200 μm $\times$ 25 μm) was translated from -95 μm to +95 μm (centre of crystal side being defined as 0 μm) along the goniometer axis.

[3] Translational: as for Standard, except that the total exposure was broken up into several individual wedges at different positions on the crystal along the goniometer axis. The first wedge was collected half a beam profile FWHM from the edge of the crystal, then the crystal was translated by one beam profile FWHM for each subsequent wedge until the beam centre was within half a FWHM of the other edge of the crystal.

Zeldin *et al.* (2013a) concluded that the pro□le of the beam has a pivotal effect on the dose distribution within an irradiated crystal , and that top-hat shaped beams will always result in more uniform dose distributions than Gaussian beams. In terms of minimising the maximum dose per unit diffraction, the most efficient way of collecting data with a Gaussian beam was found to be to closely match the beam size to the crystal in order to get the most even dose profile possible while not excessively exposing the non-crystal regions of the loop (cryobuffer/mother liquor).This effect was also found to be achievable by matching only the vertical (perpendicular to the rotation axis) size of the beam to the crystal dimensions and performing a Helical scan. This has an important implication for beamline design, since it only requires collimation in one dimension to optimise dose distribution. Relative to collimating in two dimensions, this would result in a linear reduction in the amount of beam area lost to the collimator for a given Cube crystal smaller than the vertical beam dimensions. When this beam shape is not available, either a Helical or a Translational strategy was found to provide an improvement in both the TAD-95 and the Dose Inefficiency. Judged by these metrics, a Translational strategy is was found to be more effective than a Helical one, and this was more pronounced the smaller the beam was relative to the crystal.

The original data from this work, considering the six metrics Average Dose (Whole Crystal), Average Dose (Exposed Region)[4], Threshold Average Dose (95%), Maximum Dose, Dose Contrast (Max/TAD), and Dose Inefficiency (Max/Absorbed Energy) are presented in full in Appendix 2. This chapter first presents an extention of the simulations in Zeldin *et al.* (2013a), using the now experimentally validated Diffraction Efficiency (Total elastic yield/DWD), and then introduces and

[4] Defined as 3× the FWHM along each axis.

analyses an additional strategy, as already employed in Chapter 4: 'Offset', where the rotation and beam axes do not intersect, but are offset by some small displacement, so that the dose is not concentrated in the hot-spot at the intersection of these axis.

5.1 Diffraction Efficiency of Standard, Helical, and Translational strategies.

The simulations reported in Zeldin *et al.* (2013a) were repeated using the latest version of RADDOSE-3D (v1.0.950), providing the additional required outputs. Because the DWD based metrics consider intermediate dose fields, not just the final state of the crystal, the wedge-by-wedge results must be combined for the Translational strategy: the average over all the wedges was taken for DWD, and the sum over all the wedges for the Elastic Yield, and for the Absorbed Energy. Finally, the value after all the wedges were complete (the final dose state) was taken for remaining metrics (AD-WC, AD-ER, Max Dose, and TAD). In this way, Diffraction Efficiency was calculated from the sum of the elastic scattering over the average DWD, and Dose Inefficiency was calculated using the total absorbed energy over all the wedges and the Max Dose as before.

Figure 5.1 shows that the Diffraction Efficiency has the same major trend as the Dose Inefficiency: the optimal strategy is to match the beam size to the crystal. Other than this, the major difference is that a Translational strategy with many wedges is significantly (up to an order of magnitude for the smallest beams) more efficient than a helical data collection. This can be explained by the more efficient use of the crystal volume for a Translational strategy, which uses the full angular

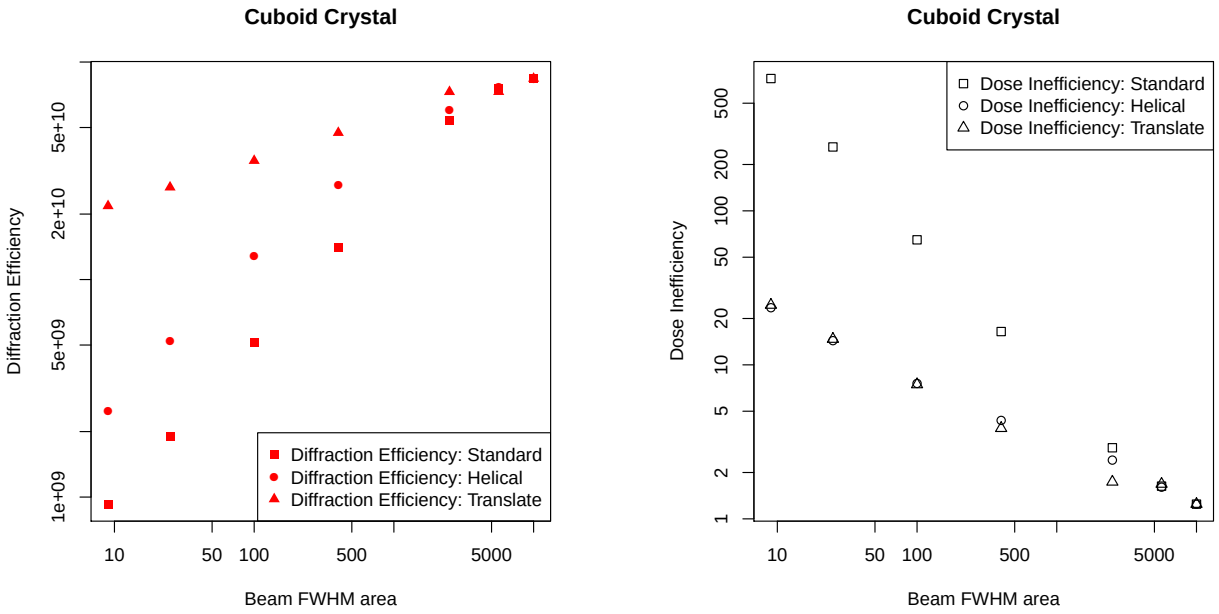
range at each point along the crystal. In contrast, a helical scan, using a beam of identical x and y FWHM, only exposes the crystal along one direction (or a very small angular range) at any given position. For a helical scan with a beam that matches the crystal size in the vertical dimension, the Diffraction Efficiency would be considerably higher. This is, however, counteracted by the fact that these simulations are for a very large number of wedges. Table 5.1 shows that, for the biggest advantage in Diffraction Efficiency, over 100 wedges must be collected! A more practical comparison can be made by placing an arbitrary limit of 10 wedges on collection. For this regime, there is a large improvement for the cuboid crystal, a factor of 2.75, compared to ~1.5 for the other crystal shapes, as shown in table 5.2.

These results using Diffraction Efficiency allow us to make a small update to the recommendations offered in Zeldin *et al.* (2013a): for cuboid crystals, if the experimenter is willing to collect ~10 wedges, it is better to perform multiple wedges than a helical strategy[5]. The tables presented in this thesis also represent an updated look-up chart for the dose advantage offered by Helical or Translational strategies in various contexts.

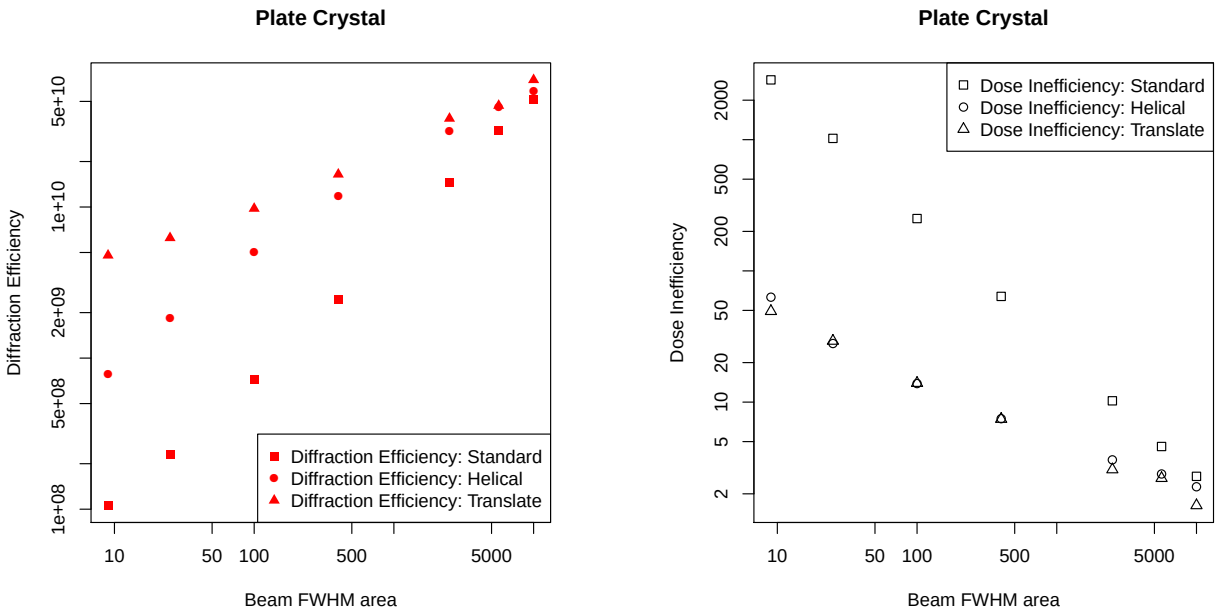
Beam FWHM [μm ^{2}]	Cuboid	Plate/Short Needle	Long Needle
3×3	33	66	266
5×5	20	40	160
10×10	10	20	80
20×20	5	10	40
50×50	2	4	16
75×75	1	2	10
100×100	1	2	8

Table 5.1: Number of wedges employed by the Translational strategy. The larger numbers are unrealistic for modern data collection systems, but there is still an advantage to be weighed against the extra complexity of data collection.

[5] As clearly stated in Zeldin *et al.* (2013a), these recommendations are only from a radiation damage perspective, and do not take into account other potential disadvantages of the Helical strategy, such as crystal inhomogeneity.

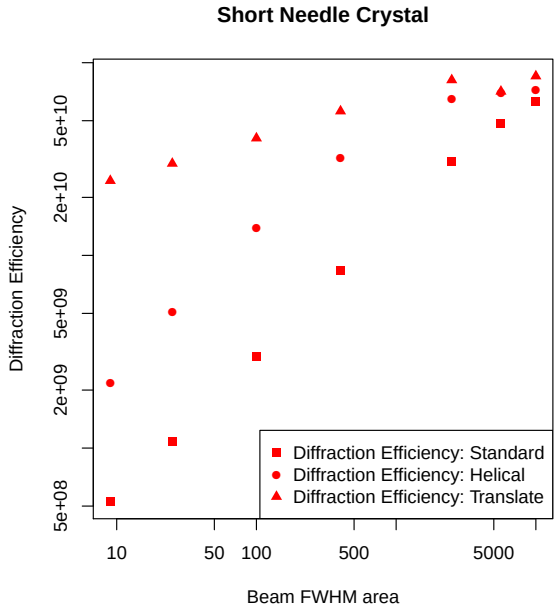


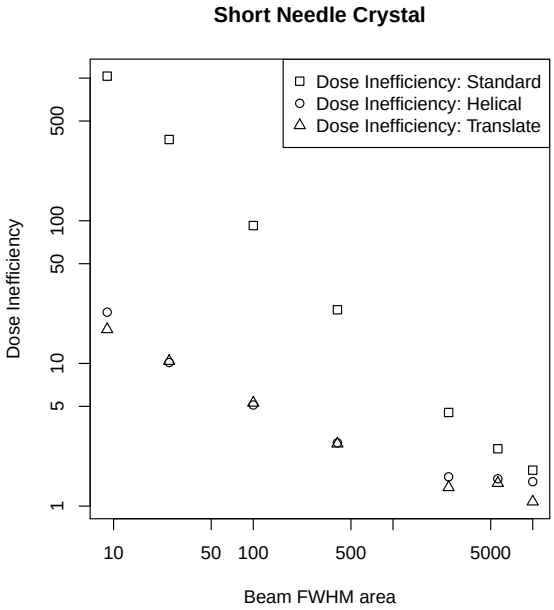
(a) Comparison between Dose Inefficiency and Diffraction Efficiency for the Cuboid crystal shape. The x axis is in μm^2, and the y axis in photons/MGy (DWD) for Diffraction Efficiency and in MGy/mJ for Dose Inefficiency.



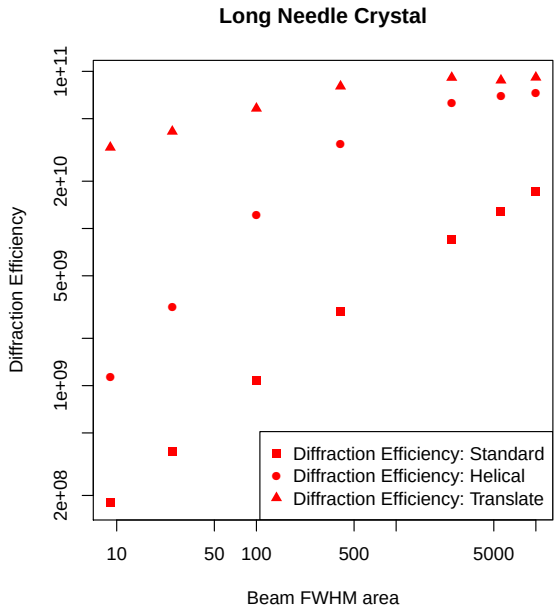
(b) Comparison between Dose Inefficiency and Diffraction Efficiency for the Plate crystal shape. The x axis is in μm^2, and the y axis in photons/MGy (DWD) for Diffraction Efficiency and in MGy/mJ for Dose Inefficiency.

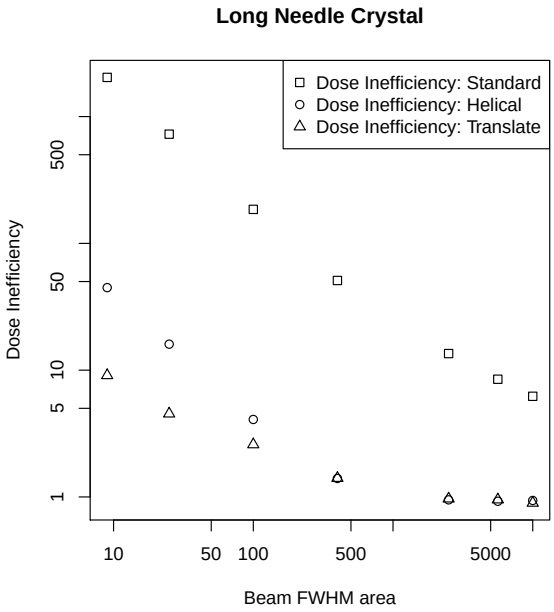
Figure 5.1





(c) Comparison between Dose Inefficiency and Diffraction Efficiency for the Short Needle crystal shape. The x axis is in μm ^{2}, and the y axis in photons/MGy (DWD) for Diffraction Efficiency and in MGy/mJ for Dose Inefficiency.





(d) Comparison between Dose Inefficiency and Diffraction Efficiency for the Long Needle crystal shape. The x axis is in μm ^{2}, and the y axis in photons/MGy (DWD) for Diffraction Efficiency and in MGy/mJ for Dose Inefficiency.

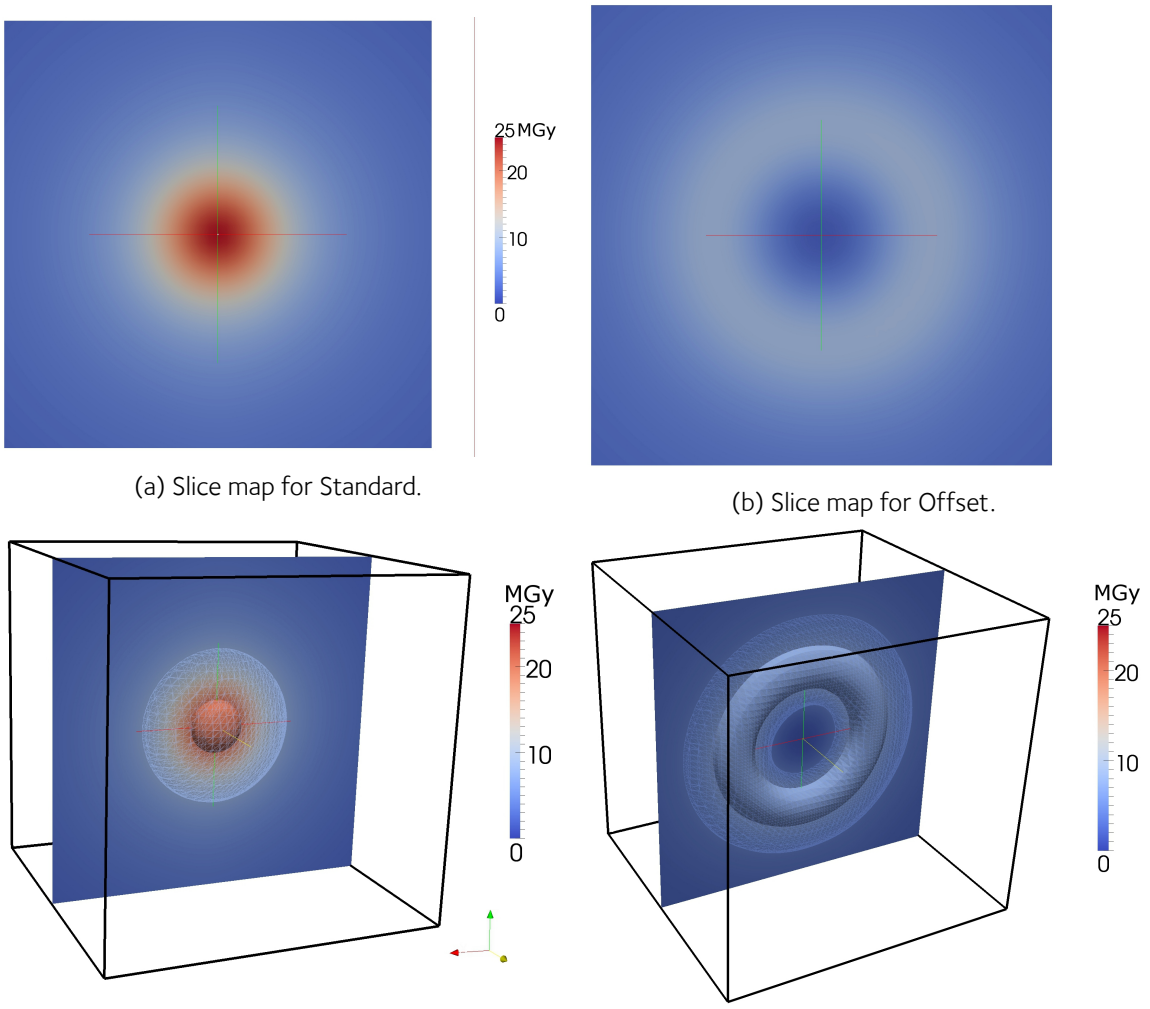
Figure 5.1: Comparison of Diffraction Efficiency (elastically scattered photons/MGy) and Dose Inefficiency (Max Dose/Absorbed energy). Dose inefficiency, as used in Zeldin *et al.* (2013a), shows the same broad trends as the more sophisticated, experimentally validated, Diffraction Efficiency, but Diffraction Efficiency suggests that there is a more significant advantage to Translational strategies, especially with smaller beams. This comes at the cost of a very large number of wedges, and is thus only considered practical with the more cuboid shaped crystals, as discussed in the text.

Crystal Shape	Beam FWHM [μm ^{2}]	Translational/Standard	Translational/Helical
Cuboid	10×10	6.85	2.75
Plate	20×20	6.71	1.39
Short Needle	20×20	6.73	1.75
Long Needle	75×75	6.84	1.25

Table 5.2: Relative improvement in Diffraction Inefficiency for a Translational collection limited to 10 wedges. Although always providing a major advantage over a Standard collection, this is only greater than a factor of two for the Cuboid crystal, where this may be a good strategy for radiation sensitive crystals.

5.2 The Offset strategy

In a conventional data collection, the crystal volume that is at the intersection of the rotation and beam axes is illuminated for the whole of the exposure. This leads to a high maximum dose for the Standard data collection, as described in Chapter 3. If these axes do not intersect – easily achieved by displacing the goniometer axis by a few μm– then this hot-spot is spread into a circular segment drawn out by the beam at a distance equal to the offset from the rotation axis. Figure 5.2*a* shows a slice map in the plane of the beam, perpendicular to the rotation axis of the dose distribution, for a 360° data collection with a Gaussian beam of FWHM 20 μm. The high dose hot-spot is clearly visible where the rotation axis cuts the plane. Figure 5.2*b*, plotted with the same colourscale, shows the same simulation, except with the beam offset by 1 FWHM (20 μm) from the rotation axis. The toroidal region of maximum dose, which is at a dose lower than the maximum shown in figure 5.2*b*, is clearly visible around the rotation axis. Dose iso-surfaces for these two conditions are shown in figures 5.2*c* and 5.2*d*. For comparison, the maximum dose for the standard strategy is 24.8 MGy, and 9.0 MGy for the Offset strategy.



(a) Slice map for Standard.

(b) Slice map for Offset.

(c) Dose isosurface map for Standard data collection.

(d) Dose isosurface map for Offset.

Figure 5.2: Comparison of dose distribution with and without offset. In *(c)* and *(d)*, note the larger light blue wire-frame 8 MGy isosurface in the Offset strategy, and the absence of any regions in the Offset map above the 20 MGy isosurface, shown as a red isosurface red in *(c)*

5.2.1 Simulations

In order to assess if and when an offset rotation axis is desirable, simulations were performed with a beam of equal x and y FWHM, for wedges of total angular range 0°-720° in steps of 45°, with beam offsets of 0-40 μm (0-2 times the beam FWHM), in steps of 5 μm (0.25 FWHM). The simulations were then run for cubic crystals with edges of length 20 to 100 μm (1 FWHM to 5 FWHM), in steps of 20 μm. These 765 simulations, shown in figure 5.3, were run with a nominal beam size of 20×20 μm^2 FWHM, collimated to 40×40 μm^2, with a crystal resolution of 2 μm per pixel edge. Results, presented as fractions of beam FWHM, can then be applied to any general beam-crystal size combination. Default absorption and attenuation coefficients based on an average crystal composition at 12.4 keV were used, along with a nominal flux of 5×10^{11} ph/s.

5.2.2 Discussion and results

Considering the set of simulations presented in figure 5.3, it is immediately clear that an increased offset does not necessarily lead to an improvement in Diffraction Efficiency. This is most striking for the smaller crystal sizes. The cause of this is twofold: firstly, beyond a certain point, the dose spreading has a smaller effect on DWD, and so diminishing gains are achieved. Secondly, when the offset is larger than half the crystal edge length, parts or all of the taurus of maximum dose will miss the crystal, meaning that, to achieve the same number of diffracted photons, the edges where the beam centroid 'catches' the crystal must be highly exposed. An additional complication

presented by excessive offset is that with the beam at least partially missing the crystal (the taurus of high dose not being fully within the crystal volume), there will be higher exposure in the solvent surrounding the crystal, increasing the background per unit diffraction.

In order to confirm that the decrease in Diffraction Efficiency for large offsets were caused by the beam 'clipping' the edges, and to further explore the diminishing gains achieved through increased offsetting, two further crystal sizes were simulated. Simulations for crystals with edges of 200 μm and 400 μm are shown in figure 5.4.

As expected, the drop in Diffraction Efficiency seen in figure 5.3 is not observed, since, at these crystal dimensions, the bulk of the beam remains within the crystal for all offsets. The trend for a bigger effect at larger oscillation angles is very clear, and offsetting appears to offer diminishing returns beyond 1 FWHM.

Based on all the simulations presented here, the following recommendations can be made:

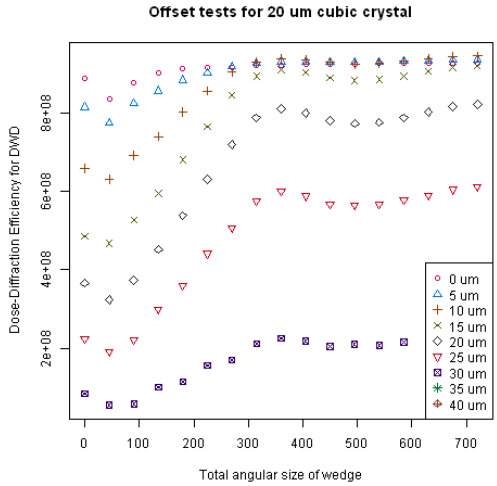
1. For crystals with size of the same order as the beam, there is little to be gained from offsetting, but it is worth noting that a larger rotation angle > $360°$ does improve dose distribution.

2. For crystals with edges greater than $3\times$ the beam FWHM, offsetting the crystal by one beam FWHM, and collecting for $360°$ or more can offer a ~20% improvement in Diffraction Efficiency, with larger improvements possible for larger crystals.

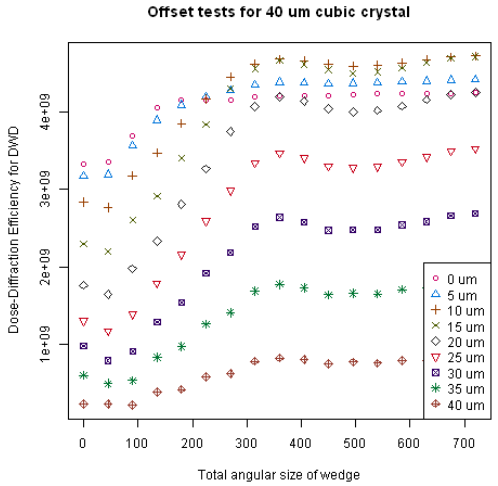
3. For crystals much bigger than the beam (edges more than $10\times$ the beam FWHM), offsetting by up to 2 FWHM can lead to an improvement of ~40%

These recommendations are for a perfect Gaussian beam, so, before employing them as routine at synchrotron beamlines, it is worth performing similar simulations with actual beam profiles in RADDOSE-3D. This would allow a better estimation of suitable guidelines, taking into account the non-Gaussian shape of experimental X-ray beams.

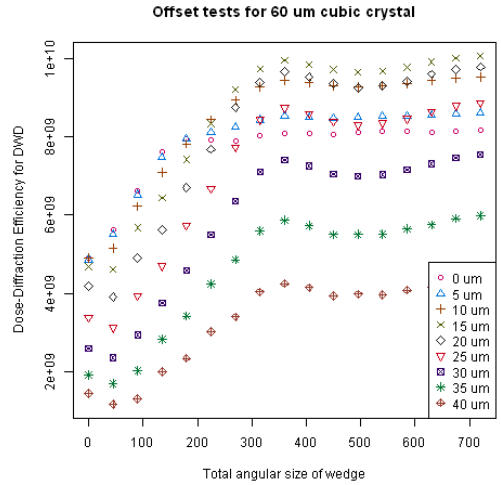
Finally, although the improvements for the Offset strategy are smaller than for the Helical or Translational strategies, it is important to note that they achieve this without using any additional crystal volume. In effect, the Diffraction Efficiency improvements achieved through Offset come at no cost in terms of using available crystal volume, and so are considered highly worthwhile.



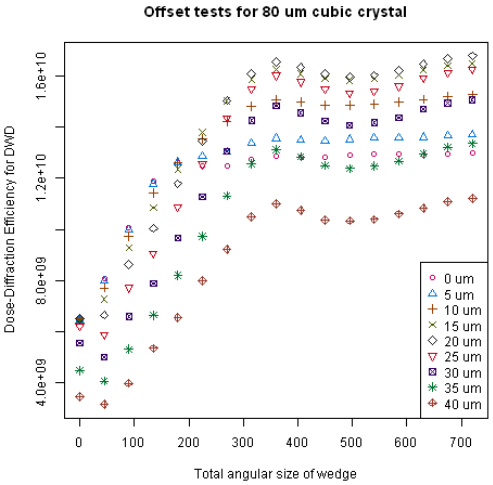
(a) Offset series for a 20 μm crystal. Many of these offsets are clearly suboptimal, since the peak intensity of the beam is completely outside of the crystal from 15 μm or greater offsets ($> 10\sqrt{2}$ μm).



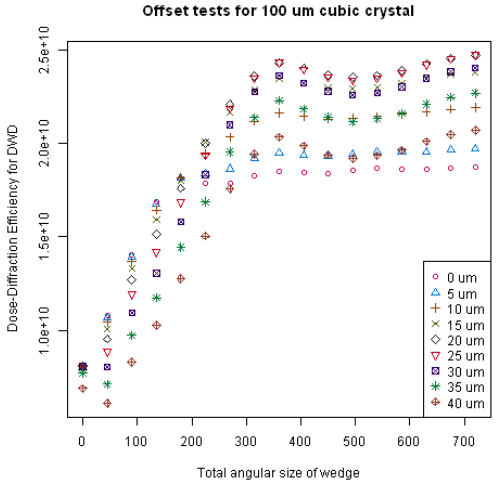
(b) Offset series for a 40 μm crystal. Some of these offsets are clearly suboptimal, since the peak intensity of the beam is completely outside of the crystal from 30 μm or greater offsets ($> 20\sqrt{2}$ μm).



(c) Offset series for a 60 μm crystal.

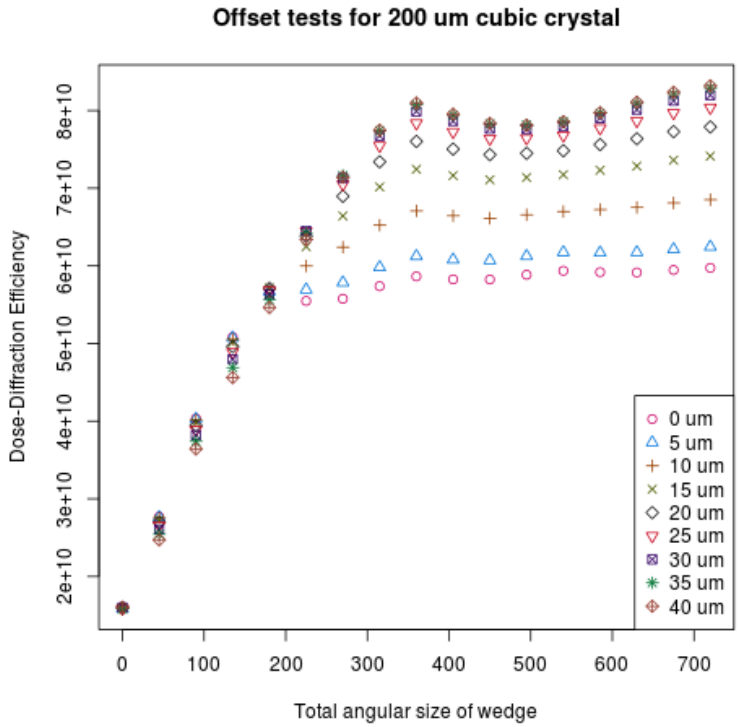


(d) Offset series for a 70 μm crystal.

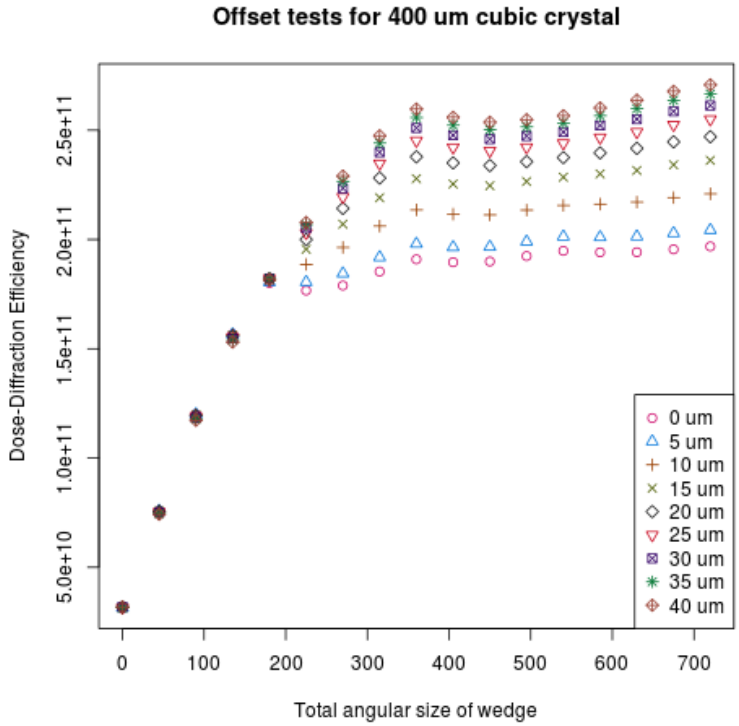


(e) Offset series for a 100 μm crystal.

Figure 5.3: Offset data for crystals up to 100 μm. In these cases, a bigger offset can lead to a drop in Diffraction Efficiency, due to the crystal being unevenly exposed with relatively high damage around the edges.



(a) Offset series for a 200 μm crystal.



(b) Offset series for a 400 μm crystal.

Figure 5.4: Offset data for larger crystals, showing the general trends more clearly. In both cases, significant improvements are only achieved for rotations above $\sim 360^{\circ}$, and there are diminishing gains for offsetting the crystal beyond 1 FWHM (here 20 μm).

Chapter 6

Development of High Throughput micron beam Proton Induced X-ray Emmission (HT microPIXE)

6.1 The role of metals in structural biology

Metals play a wide variety of roles in protein function, often in key sites, where they can change oxidation states in order to facilitate catalysis, or play vital structural or regulatory roles (Degtyarenko, 2000). It is thus highly important to correctly identify and characterise the metals in metalloproteins. Heavy atom analysis – here defined as any atoms heavier than the organic elements C, N, O, and H – can reveal information about protein function (e.g. the metal affected by an uptake regulator), confirm putative metal identification made on the basis of electron density maps, reveal if oligonucleotides are bound (by looking at the phosphorous signal), confirm correct crystallisation of proteins with known bound metals, or confirm the extent of Se-Met incorporation prior to attempting MAD phasing, especially for recombinant proteins grown in mamalian expression systems, where Se-met uptake can be too limited for successful phasing to be achieved. With about one third of all proteins having an associated metal (Shi and Chance, 2011), there is a vital need for powerful tools to allow the routine and reliable characterisation of metalloproteins. Existing techniques either require large sample volumes, or do not provide an unambiguous and comprehensive characterisation of the samples. The wide-spectrum atomic fluorescence stimulated by high energy protons in the microPIXE technique is ideally suited to the task of characterising metalloproteins, and this chapter charts efforts to adapt this technique for the high-throughput analysis of protein samples, as well as a preliminary investigation into the use of the technique for other types of biological samples.

6.1.1 Alternative techniques

Five principal types of rival techniques exist for metal analysis of proteins:

1. Wet lab assays that rely on detecting chemical compounds bound to metal in samples. These require large volumes, typically around 100μl (microPIXE <1μl), and only test for one species at a time. They are not currently used for metalloproteomics.

2. Circumstantial evidence, such as coordination factors or B factors from electron density maps. These are not definitive, and, crucially, require a crystal structure before they can be employed.

3. Mass spectrometry, although powerful is often not unambiguous due to effects such as partial occupancy, glycosylation, or covalent modification.

4. X-ray Absorption Spectroscopy (XAS) is a powerful technique, but has high experimental requirements (Ascone and Strange, 2009), and has a much larger sample volume requirement than microPIXE. It does, however, provide information on electronic states.

5. Synchrotron Radiation X-Ray Fluorescence (SR-XRF) has a much lower limit of detection than microPIXE (parts per billion for SR-XRF *vs* parts per million for microPIXE), but only excites a limited range of elements at a time. This is due to the narrower energy-dependent fluorescence cross sections for X-rays compared to protons, so that for SR-XRF, the energy must be tuned to just above the elemental absorption edge for optimum sensitivity. Furthermore, SR-XRF requires beamtime at a synchrotron, which is more scarce than proton

microprobe time.

MicroPIXE, described in detail below, has two major advantages over rival techniques: firstly, due to the broader energy range of fluorescence cross-sections that atoms present to protons *vs* X-rays, it excites all relevant biological elements simultaneously, eliminating the risk of missing a 'hidden' element. Secondly, as a consequence of this broader range of excitation, it is possible to take the ratios of elements. Through the sulphur normalisation technique described below, this allows the direct measurement of stoichiometry i.e. it is possible to say "this protein has 3 irons per monomer" as opposed to just "this protein contains iron".

6.2 X-ray fluorescence as a metalloprotein assay

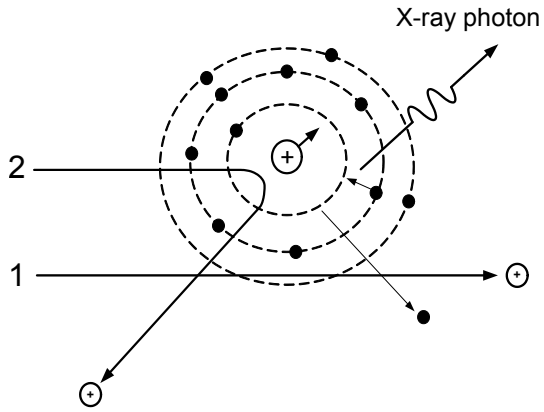


Figure 6.1: The principle of PIXE: 1. A proton passes through the atom without interacting. 2. An inelastically scattered proton ejects an electron then, as another electron falls into the (lower energy) hole, a photon is emitted with an energy that is unique to the element from which the proton scatters. These photons are used to identify and quantify elements present. Adapted from Garman and Grime (2005)

MicroPIXE operates by using the inelastic scattering from a high energy (typically 2.5 MeV), highly focused (<2μm) proton beam to eject inner shell electrons from the atoms in a target; see fig.6.1. The higher energy outer electrons will fall into the lower energy inner shell, and the atom

then emits an X-ray with exactly the energy of the difference between the two levels. These energies are a function of the atomic number and so are unique to each element. The presence of elements heavier than neon[1] can thus be detected to a very low Limit of Detection (LoD), and the relative intensities of elemental peaks can be used to calculate the precise stoichiometry of these elements. If the absolute concentration of one of the elements is also known, it is possible to infer accurate concentrations for all the others. The standard approach to this is to dope the sample with a known quantity of an element that would not otherwise be present (often yttrium or indium) and to use this for normalisation. There are however drawbacks to this technique, principally quantifying the dopant concentration into dried mg/cm^2.

The ideal situation is to use an internal normaliser: an element that is already present in the sample and is of known concentration. The sulphur in protein samples is ideal for this, since the number of sulphur atoms per molecule is known from the number of cysteine and methionine residues in the primary sequence (Garman, 1999; Garman and Grime, 2005). This means that it is possible to infer the number of metal atoms per protein molecule independently of the absolute protein concentration, using the formula:

$$\mathrm{N}_y = \frac{\text{Areal density of element y}}{\text{Areal density of sulphur}} \times \frac{\text{Atomic mass of sulphur}}{\text{Atomic mass of y}} \times \mathrm{N}_S$$

where N_x is the number of atoms of element 'y' per protein molecule, and N_S is the number of sulphur atoms per protein molecule.

[1] Due to the 130 μm beryllium detector window (between the evacuated sample chamber and the detector), which absorbs lower energy photons.

During the experiments, two data sets are recorded simultaneously: the energy spectrum of Rutherford Back Scattered protons (RBS) and the microPIXE energy spectrum. The RBS data are a measure of the energy spectrum of protons that have been elastically scattered by the nucleus of target atoms at close to 180° (see fig.6.2 for an example energy spectrum). This provides information on the sample thickness and gross composition (i.e. the ratio of C, N, H & O which make up the bulk of the sample, and which emit X-rays that are too low in energy to be seen in the PIXE spectrum). The sample composition is revealed by the presence of characteristic proton energies, which have a fixed upper limit at the maximum momentum transfer (*i.e.* scattered at $180°$). The sample thickness can be found from the energy-widths of the peaks, where a thicker sample will slow the exiting protons more, leading to a broadening of the energy profile. This information is used by the PIXE spectrum fitting algorithm to calibrate for the differential absorption of X-rays at different energies before they exit the target (Grime, 1996). RBS spectrum fitting treats the film and sample as two separate layers, and assumes a constant sample thickness (homogeneous slab approximation), which can lead to issues when there is a non-uniform thickness across the measured area (for example the spectrum of a 2D raster map of a drop of dried fluid shown in fig.6.2). Maps are created by placing a software bin around the energy associated with an atomic fluorescence peak in the X-ray detector, and then plotting counts in that energy range against the known x and y beam position when each count is recorded. Because there is no background subtraction, they are a qualitative tool best used for localisation of elements. An example of a fitted PIXE spectrum for a population of cells is also shown in fig.6.3, demonstrating the distribution of

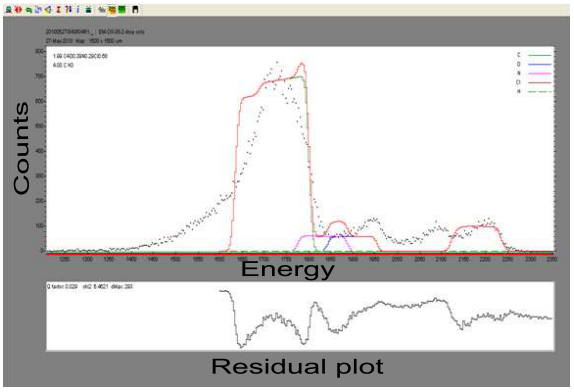


Figure 6.2: Example of an RBS spectrum and fit for the sample of mouse cerbral spinal fluid in fig.6.9. The position of the high energy (right hand side) edge of elemental peaks indicates the species, and the depth of the peak indicates the sample thickness. Note the low energy end of the (red) fit which is trying to fit a single thickness to a sample of varying depth.

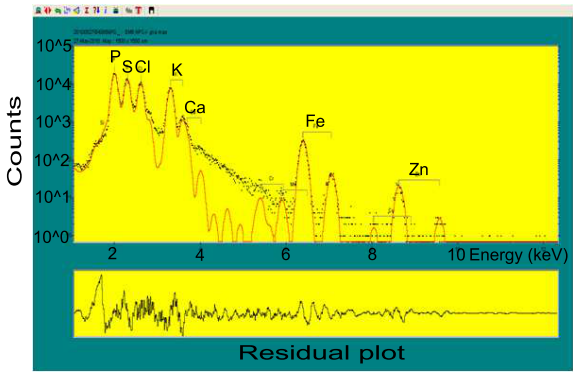


Figure 6.3: Example of a fitted PIXE spectrum for a sample of Niemann-Pick infected astroglia cells. Note the logarithmic count scale and simultaneous measurement of all elements heavier than neon.

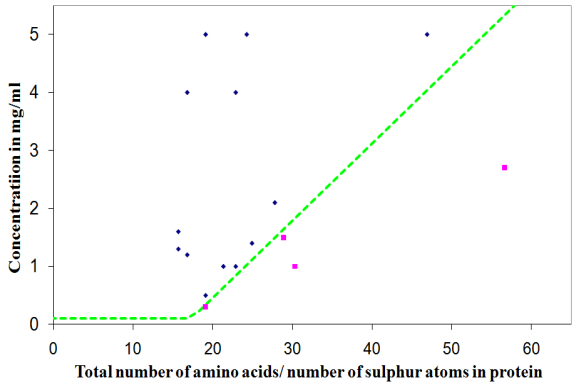


Figure 6.4: The Y-G plot empirically relating liquid protein concentration to protein size/number of S-containing residues. Blue diamonds: measurement possible. Pink squares: measurement not possible. The green line is an empirical model currently used as a bound between possible and impossible samples. Adapted from Garman and Grime (2005)

heavy elements.

The most common reason for the failure of microPIXE experiments on proteins is an insufficient signal to Limit Of Detection (LoD) ratio for the sulphur. This has been empirically linked to a minimum protein concentration through the Yates-Garman plot (Y-G plot, figure 6.4) which provides a guideline for the minimum concentration (mg/ml) required for a given liquid protein sample based on the ratio of total amino acids/protein to S-containing amino acids (Garman and Grime, 2005).

The form of this plot was found by comparing possible choices of axis until a separation of the two populations was observed between samples that were successfully analysed, and those where the elemental content could not be determined. It is thus fully empirical. The axes of the original plot were chosen so as to be easily interpretable for biological samples. The mg/ml is milligrams of protein sample per ml of solution and the total number of Amino Acids / S-containing Amino Acids is a measure of mass of protein present per sulphur atom (since the mass of any amino acid is approximately 110 Daltons).

If we consider that there is some line to be drawn that will separate possible and impossible samples in fig.6.4, we can look at this in terms of the limiting case of Signal/Limit of Detection where a possible sample is defined as being some multiple (usually between 2.5 and 10) of the LoD. The ratio will affect the errors, so a minimum value should be defined in light of the accuracy needed for the measurement. The LoD has been defined by Maxwell *et al.* (1989):

$$LoD = 3 \times \sqrt{BG + OL} + 0.01 \times OL \tag{6.1}$$

where BG is the background within one FWHM of the X-ray peak centroid and OL is the overlap area with neighbouring peaks. The elemental peaks in a PIXE spectrum are fitted as Gaussians and it can easily be shown that the overlap of two Gaussians with constant FWHM will scale linearly with exposure time. Since the bremsstrahlung background[2] and signal are only functions of total charge, they will also scale with time (under constant current). This means that the overall sig-

[2] The X-ray background due to the deceleration of the electrons that are ejected from atoms by the proton beam.

nal/LoD will scale approximately as $t^{1/2}$ for $BG >> OL$.

The signal can be modelled analytically as a Gaussian with area $\propto Q[S]MY_{sulphur}$ where Q is the total charge, $[S]$ is the sulphur mass fraction, M is the total mass of sample in the beam and $Y_{sulphur}$ is the PIXE yield for sulphur. Overlap is simply a function of neighbouring elemental peaks, which can be modelled in the same way. We note here that the mg/ml used in the Y-G plot ignores the potentially significant contribution the buffer material can make to the total mass of dry sample (5mg/ml for a typical 20mM KBr, 20mM Tris buffer). The bremsstrahlung background is known to be made up of four components (Ishii, 2006): nuclear bremsstrahlung (NB), atomic bremsstrahlung (AB), secondary-electron bremsstrahlung (SEB), and quasi-free-electron bremsstrahlung (QFEB).

6.3 Experimental set-up

The experiments reported here were all performed on the microbeam at the University of Surrey Ion Beam Centre (IBC) (Simon *et al.*, 2004). The beam from the 2MV Tandem Accelerator was focused using a quadrupole lens system to ~2μm, and then passed into the evacuated sample chamber. The layout of the sample chamber is shown in figure 6.5. Data were acquired either by holding the beam at a single point for a time (point measurement) or by rastering across a pre-defined region (map or mask measurement). The basic protocol for microPIXE analysis of liquid protein samples is:

1. Prepare liquid samples at a sufficient concentration with no contaminating sulphur, and low

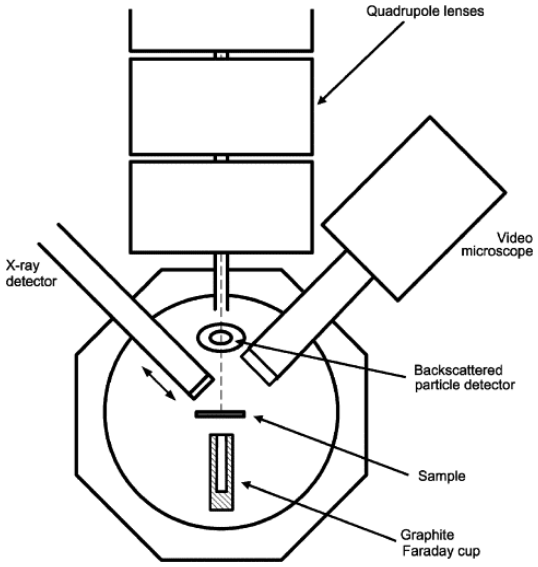


Figure 6.5: Schematic of the sample chamber. The light microscope is used to approximately position the sample and the Faraday cup absorbs the protons that pass through the sample. Adapted from Garman and Grime (2005)

chlorine (since this overlaps the sulphur peak, increasing the LoD, and thus giving larger errors than necesarry on the resulting stoichiometry).

2. Transfer samples to sample-holder, by pipette in 0.2 μm drops for the low throughput method. Cover to prevent dust contamination, and allow to air-dry.

3. Locate sample, and acquire data at the IBC using the OMDAQ control software (Grime and Dawson, 1994).

4. Analyse data using the OMDAQ interface, which uses GUPIX (Maxwell *et al.*, 1989) for PIXE analysis, with built-in RBS normalisation (Grime and Dawson, 1994).

Efforts to develop this as a high throughput technique must thus automate and streamline these basic steps so that analyses of many samples can be performed in one block of time, with minimal user intervention.

6.4 Other samples

During the early phases of the project, preliminary work on extending the range of biological samples that can be quantified was performed. In collaboration with Dr. Emyr Llyod-Evans and Oscar Chen from Prof. Fran Platt's group in the Department of Pharmacology, University of Oxford, measurements on a variety of samples relating to Niemann-Pick disease (NPC) were performed. NPC is a lysosomal storage disorder, and it is hypothesised that zinc and iron might have important roles in the disease mechanism. Attempts to establish methods that will allow accurate quantification of what, if any, changes in metal ion levels occur with diseased and non-diseased phenotypes were thus made.

6.4.1 Whole cells

For protein samples, point measurements are normally used in low throuput microPIXE. This has the advantage of an increased flux density, since for a map scan, a significant portion of the beam time will be spent off the sample. However, for protein samples, the 2 μm beam was already several orders of magnitude larger than the protein molecule being studied. For the case of human astroglia cells, which are of the order of 20 μm, the opposite was true. It was therefore more appropriate to average over masked areas containing only cells. This has the advantage of effectively measuring millions of cells simultaneously. Because a microbeam technique was being used, data at each (x,y) coordinate were being recorded, so it was possible to remove areas where potential impurities affected the overall average, for example the calcium hot-spot shown in figure 6.6. Since the cells

were intact, this was considered an acceptable manipulation. For normalisation, the measurements on two different elements were tested: firstly sulphur, which could be approximated to a measure of total protein content, for which the mean was assumed constant across phenotypes. This was not ideal since there is evidence of S-Fe clusters in the cytoplasm (Ye and Rouault, 2010) and the assumption of the sulphur content being constant and independent of disease type is hard to validate. A second, possible method would be to use phosphorous. In analogy to the technique used for proteins with sulphur, the amount of DNA in non-dividing cells is constant. In order to estimate if this is the dominant source of cellular phosphorous, an approximate calculation was made. P atoms / cell:

$$6 \times 10^{9} bases / 6.022 \times 10^{23} = 9.96 \times 10^{-15} mol/cell$$

Dividing by an average cellular volume (10^3 μm^3 = $10^{-12} l$):

$$\frac{9.96 \times 10^{-15}}{10^{-12}} \approx 10 mM$$

If this is much larger than cytosolic levels, it would be a good estimator. Furthermore, it would be possible to validate this by mapping single cells using a micro-beam and directly comparing the nuclear and cytosolic P levels.

The measured ratios of WT/NPC cell iron and zinc content presented in table 6.1 were consistent with the biological hypothesis put forward by our collaborators in the Pharmacology Department:

Fe is depleted in NPC cells and Zn is unchanged. Although the wet lab techniques used were less accurate (Ferene-S assay for Iron, Zn Staining for zinc), the agreement with the general trend was promising.

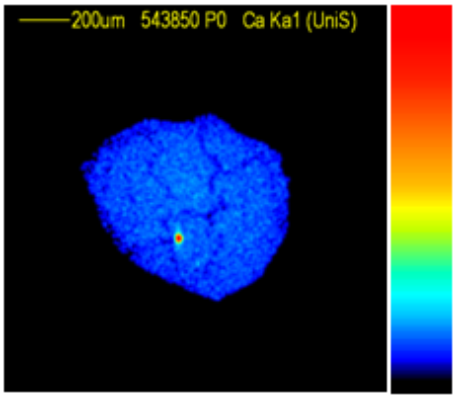


Figure 6.6: Example of a problematic Ca hot spot in a map of a population of NPC astroglial cells. The obvious high intensity spot of Ca (several orders of magnitude higher than the rest) was removed before quantification. Analysis indicated that it was probably a particle of calcium carbonate. Colour bar is in arbitrary X-ray count units.

6.4.2 Homogenates

WT and NPC brain homogenates from mice were analysed. In this case, the question of whether to measure points or areas and, if areas, which areas to measure became more complex. Figures 6.7 and 6.8 illustrate some of these issues. The differential precipitation of the Fe shown in figure 6.8 could easily lead to an erroneous measurement if an unrepresentative region was selected. It was thus decided to mask the drops, so that the beam scanned across a pre-determined region

Element	S normalised	P normalised
Fe	1.12 ± 0.12	1.16 ± 0.12
Cu	0.87 ± 0.52	0.90 ± 0.53
Zn	1.04 ± 0.11	1.08 ± 0.12

Table 6.1: WT/NPC concentration ratios for three metal ions normalised with both S and P. Note the consistency across normalisation protocols: the relative amount of Fe, Cu and Zn is the same (within error) for both S or P normalisation, indicating that these may offer a possible method for the normalisation of metal levels in whole cell samples.

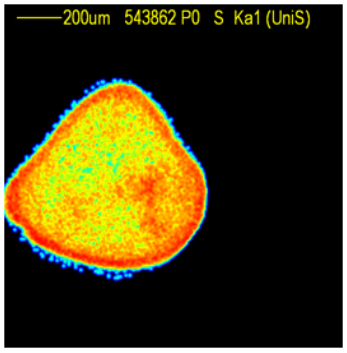


Figure 6.7: Slight differential precipitation in a sulphur map of a drop of mouse brain homogenate. Relatively homogenous, but evidently more concentrated at the edges. Colour bar as for figure 6.6.

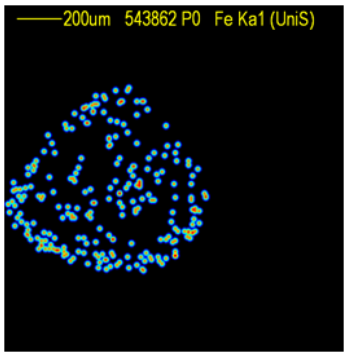


Figure 6.8: Iron map of the same sample as in fig.6.7 showing strong differential precipitation. This sample was more inhomogeneous so average levels could be taken across the whole drop. Colour bar as for figure 6.6.

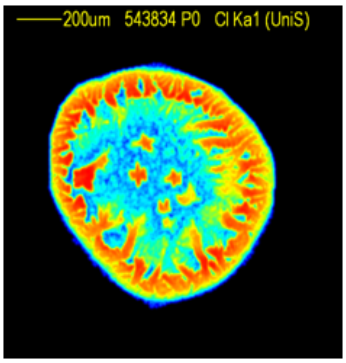


Figure 6.9: Cl map of Cerebro-Spinal Fluid showing the differential precipitation problems associated with mixed solutions leading to both concentration and depth variation. See fig.6.2 for RBS. Colour bar as for figure 6.6.

the same shape as the drop, thus not wasting any beamtime, and take measurements of the whole sample. In terms of normalisation, there was no obvious internal independent parameter against which to normalise. Y or In doping could have been used, but that would not normalise the intrinsic noise that is present due to the variation between molecular composition of the same cell type in a population at all, and would therefore have required many samples to be measured: highly impractical for a routine technique on difficult to obtain samples. S could be used to normalise, but this would need to be carefully validated either by an independent method or by Y/In doping across a population.

6.4.3 Fluids: Cerebral Spinal Fluid (CSF), serum and others

These samples, which can provide important data regarding metal regulation in NPC affected mice, had very low sulphur levels that were difficult to quantify accurately. In addition CSF in particular is

very high in Cl, completely overwhelming any S signal. A whole-sample measurement with In or Y doping to normalise would therefore be the next step. Normalisations to S were thus not deemed reliable until independently validated using alternative experimental techniques.

6.5 The need for HT microPIXE

Although low throughput microPIXE is a very powerful technique, its applicability is limited, due to a combination of slow turnaround times, high costs, and the labour intensive nature of sample preparation and analysis. As a consequence, the kinds of analysis that can be carried out are limited to 'final piece of the puzzle' type experiments, where microPIXE serves to confirm a hypothesis, or to identify a metal in special cases.

By increasing the throughput by an order of magnitude using high throughput techniques, it would be possible to:

- Use microPIXE as a routine characterisation tool for newly purified proteins.
- Identify the impact on metal binding of point mutations, or of different loading and purification protocols.
- Reduce the cost of analysis when limited to 10 samples for a day of beam-time is prohibitively high (~£300 per sample).
- Increase the turnaround rate from 4-6 weeks to less than a week, making the whole process much more attractive to fast moving biological projects.

The basic principle for HT microPIXE is to prepare an array of closely spaced samples, then to sequentially collect data in an automated fashion from each sample in the array. The data could then be analysed offline in an efficient manner.

6.6 Feasibility study of HT microPIXE

Before commencing work on developing high throughput methods for microPIXE, a feasibility study was carried out by performing measurements on closely placed drops of protein, prepared using the low throughput protocol. Samples of cytochrome 550 (C_{550}) containing iron, and pseudoazurin, containing copper, prepared in a S-free buffer as described in Garman and Grime (2005), were pipetted in 0.2 μl drops onto a standard low throughput sample holder made up of 4 μm thick poly-propylene[3] film glued to aluminium holders with a 1 cm hole across which the film was placed. Samples were then allowed to dry at room temperature before being transported to the IBC. An elemental map is shown in figure 6.10, and individual point measurements revealed no detectable Fe in the pseudoazurin sample, and no detectable Cu in the C_{550}, indicating no cross talk. This was an important experiment, since it showed that there are no intrinsic limitations, such as a halo on the proton beam, that would prevent the analysis of closely spaced samples: a vital pre-requisite for high-throughput analysis.

[3] $(C_3H_6)_n$. Purchased from Fluxana GmbH & Co. Germany. www.fluxana.com.

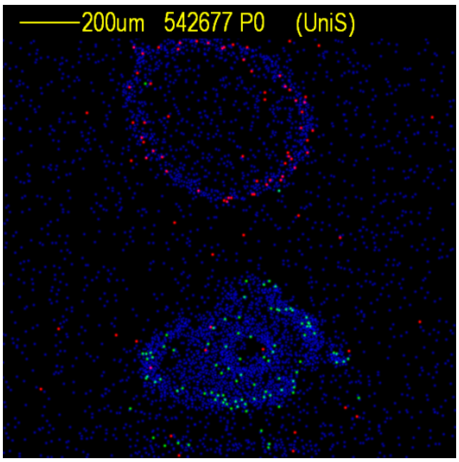


Figure 6.10: Elemental map showing absence of cross talk between two closely spaced (~200 μm separation) 0.2μl drops of C_{550} (top) and pseudoazurin (bottom). Blue: S, Red: Fe and Green: Cu. Colour bar as for figure 6.6.

6.7 Requirements of HT microPIXE

To implement High Throughput micro-PIXE (HT-microPIXE), new requirements must be met at each of the four stages of the sample analysis:

1. Sample Preparation:
   (a) Sample holders must be prepared such that all the samples for an overnight run can be inserted into the vacuum chamber simultaneously so that the chamber need not be opened during the run.
   (b) Samples must be well separated so that there is no cross talk due to either sample merging or any potential beam halo.
   (c) Samples must be arranged in a known configuration to allow automated data acquisition.
   (d) The identity of each sample must be known and passed on to the data acquisition software.
2. Data Acquisition:
   (a) Set-up time must be much shorter than the run time (typically under 1 hour set-up for a 15 hour run).
   (b) Effective localisation of samples within the array must be possible.
   (c) Protocols must be optimised to maximise the fraction of beam-time spent on the sample, rather than on the backing foil.
   (d) Sample ID must be tracked and passed on to the analysis stage.
   (e) Software must be optimised for a multi-run workflow.
   (f) Must be able to accurately and reliably collect data from the samples, either by moving the sample stage or by using the beam focusing system to steer the beam across each sample.

(g) Automated end conditions based on time or data parameters (e.g. Peak counts for element X) must be available.

3. Data Analysis:

(a) Analysis for a full run (100s of samples) must be possible in one day.
(b) Analysis workflow must be flexible to deal with different numbers of repeats per sample and varied elements of interest.
(c) Amalgamation of repeated measurements for each sample must be automated.
(d) Significance of each elemental result must be automatically determined.
(e) Results must be able to be viewed and interpreted by non-specialist users.

6.8 Sample Preparation

Before developing detailed sample preparation protocols, it was necessary to estimate the number of samples that can be measured in an average run, and the approximate space on the mounting film that was available to print these samples. Typical focused micro-beam (1-3 μm diameter) currents at the IBC range from 0.5-1.5 nA. Based on this, experience has shown that it takes approximately 5-10 minutes to collect enough data from one point for a typical protein sample. Given this time range, an overnight run (6pm-9am) of 14 hours would allow us to measure between 84 and 168 samples.

There are two possible ways of acquiring data from an array of samples on a backing film: either by magnetically deflecting the beam (maximum linear range 2-3 mm depending on the beam energy) or by moving the sample stage (range of order of cm). For maximum reliability, it was originally decided to aim to make the arrays containing all the samples approximately 2 mm x 2 mm in size so that the whole array could fit into the beam deflection range, thus not relying on the integrity of the sample stage movement.

A non-contact system for making the sample arrays was highly desirable, since any contact based system would require much larger drop spacing, would increase chances of cross-contamination from the tip of the dispensing needle, and would require extremely careful depth calibration due to the unfavourable mechanical properties of a disc of 4 μm polypropylene film (rather like a very small trampoline). If non-contact printing did not prove viable, it would have been possible to explore the use of a Mosquito liquid dispensing robot[4] with modified protocols as an alternative.

With these parameters set, it was decided to attempt array preparation using an ArrayJet Marathon non-contact printer[5] which was available in the Dunn School of Pathology thanks to a collaborator, Dr. Nigel Saunders. The ArrayJet is normally used to prepare microarrays, and is designed for high-throughput, precise printing of low-volume samples. It was thus, in principle, an ideal solution to our needs.

The machine takes samples from 384 well plates, with each well containing a minimum volume of 10 μl for standard plates, down to 5 μl for special low volume plates (which required a machine recalibration). The samples are aspirated from the 384 well plate by a 'spyder' with either 12 or 36 nozzles/legs (12 spyder or 36 spyder) into the printer head, which has 36 nozzles, and then printed onto a glass slide row by row, with 12 (3 nozzles per spot) or 36 spots in each row. Array geometries other than 12 $\times$ n or 36 $\times$ n are achieved by loading the wells corresponding to the undesired columns of each row with blanks made up of the system buffer.

Due to the design of the system, it is important always to load the 384 well plates with full array-

[4] http://www.ttplabtech.com/products/mosquito/crystallography.html
[5] www.arrayjet.co.uk

rows (12 or 32) of samples so that the spyder never introduces air into the system. The complex engineering of the spyder-print head microfluidics makes the mapping between the 384 well plate and the printed array highly non-linear and samples must be appropriately loaded into the plate to ensure the desired array is printed (see table 6.2 for the 12-spyder mapping). For this reason, a record of both the plate and array coordinates is saved.

Fitted with a low volume 12 sample head (JetSpyder 12 LV), the machine aspirates 0.7 μl from each sample well and uses this to print multiple drops onto the glass slide. Drop size is fixed due to the machine geometry at approximately 60 μm diameter, and drop spacing is variable, with 200 μm a typical spacing between drop centres. By filling the wells in the 384 plate that map to columns 11 and 12 of the final array with blanks, it is possible to print 100 samples (10×10) with 10 μm clearance at the outermost drop edges into a 2.08^2 mm^2 array, thus allowing all the samples to fit inside the 2.4^2 mm^2 beam deflection range.

384 well	A	B	C	D	E	F
1	a10	c10	a6	c6	a2	c2
2	b10	d10	b6	d6	b2	d2
3	a4	c4	a12	c12	a8	c8
4	b4	d4	b12	d12	b8	d8
5	a5	c5	a1	c1	a9	c9
6	b5	d5	b1	d1	b9	d9
7	a11	c11	a7	c7	a3	c3
8	b11	d11	b7	d7	b3	d3

Table 6.2: Mapping of a 6 by 8 section of a 384 well plate into final array printed onto the film for the ArrayJet printer. Rows in the final array are labelled a-d and columns 1-12. Any integer number of rows may be loaded but for each row all 12 wells associated with the columns of this row must have either sample or a blank present to prevent the Spyder aspirating air. The row sequence a-d may be repeated in the region of the plate bounded by I-P 1-6 for array rows e-h, in the plate region A-H 7-12 for array rows i-l and in the plate region I-P 7-12 for array rows m-p leading to a final square array of up to 12x12 distinct samples. Multiple rows of the same sample may be printed without loading extra wells of the plate by printing row 'a' N times, then row 'b' N times etc. leading to an array with N times more rows than if each sample had only been printed once.

6.9 ArrayJet printing proof of principle

A proof of principle experiment was carried out on the ArrayJet printer by placing a 4 μm polypropylene film over a standard size glass slide (held in place by the surface electrostatic interactions between the film and the slide), and printing a set of protein samples that were concurrently being measured for collaborators by low-throughput (LT) microPIXE directly onto the slide. After printing, a LT sample holder with glue applied around the rim was carefully dropped over the array; not a long term solution! A light microscope image taken within an hour of printing is shown in figure 6.11.

This early experiment showed missing samples (half columns) and uneven drop shapes. An array of printer buffer[6] printed at the same time (figure 6.12) showed highly homogeneous drops and no missing rows. There was no observed correlation between the missing samples and the sample composition or concentration. It was therefore decided to examine the printing protocol and sample additives in order to address the problem of the missing samples.

MicroPIXE analysis of the buffer array shown in figure 6.12 showed no contamination with any detectable (i.e. non organic) elements above background. This was a very important control result, satisfying the prerequisite of not introducing any impurities into the sample during printing. When analysed in the beam, the sample array shown in figure 6.12 showed additional missing columns. A long run (11h 40m) was performed on a mask containing the individual spots that were detectable and also the missing regions. Optical microscopy of the sample after data col-

[6] 10% glycerol, 60% ethylene glycol, 0.3% triton X (detergent).

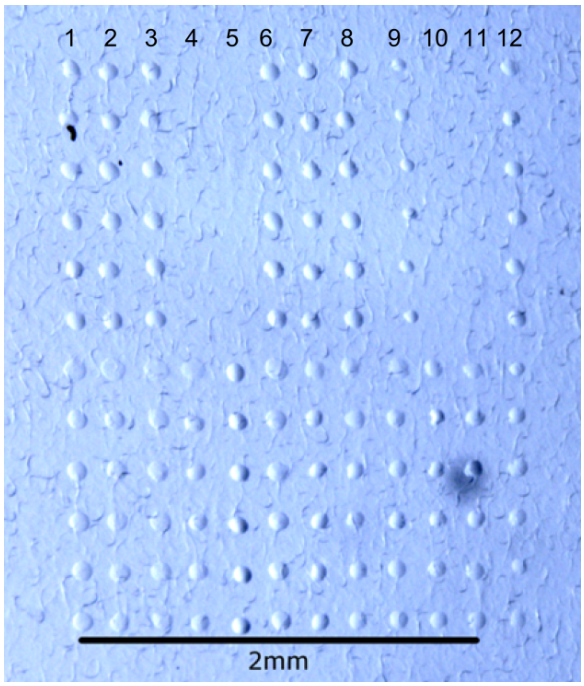


Figure 6.11: Proof of principle print run, November 2010. 24 protein samples were printed, each six times (12 in the top 6 rows, 12 in the bottom 6 rows), making up two sets of 6 rows. The missing half columns in the top half of positions 4, 5, 10, and 11 were an obvious problem and this was subsequently addressed by changing the protocol (see text).

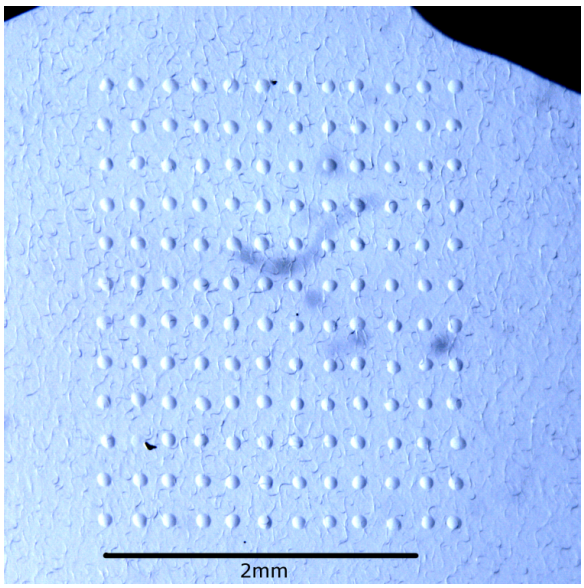


Figure 6.12: Image of print buffer array showing no missing rows. This indicated that the polyprolene film itself was not the cause of the missing rows seen in figure 6.11. The dark region at the top of the image is glue on the edge of the aluminium LT sample holder.

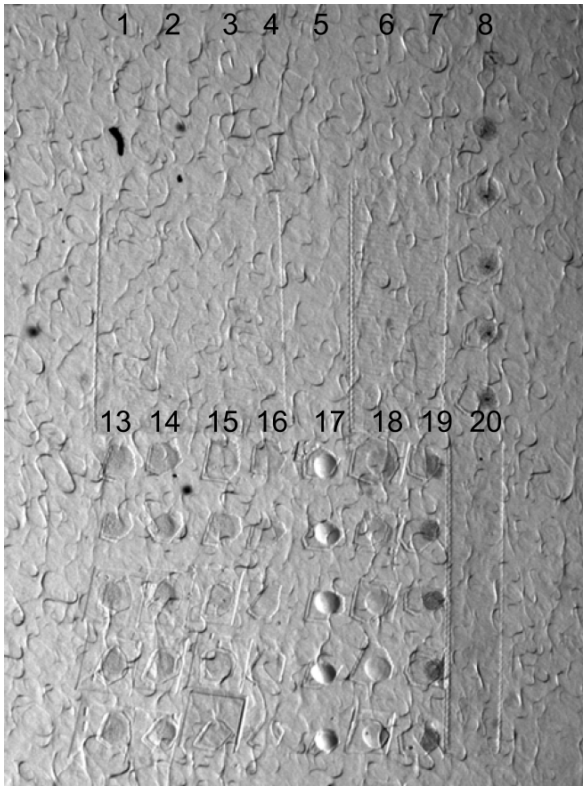


Figure 6.13: Optical image of the test array from figure 6.11 taken after data collection. The rows that are not visible in the elemental map, figure 6.14, are not present here either, indicating either complete evaporation or that the sample was never correctly dispensed (i.e. only buffer was printed). The vertical and horizontal scanning pattern of the beam path is visible where it has etched away some of the polyprolene film.

lection is shown in figure 6.13. Additional missing half columns are observed in positions 1- 3, 6, 7 and 9. Rows 20-24 only contained MilliQ water and so were not expected to remain. This is consistent with the elemental map (figure 6.14). These two figures are linearly transformed and overlaid in figure 6.15, clearly showing the missing sample problem. Before investigating a solution to this, a new sample holding system had to be devised so that it would be possible to create a consistent substrate for printing that was fully compatible with the ArrayJet printer.

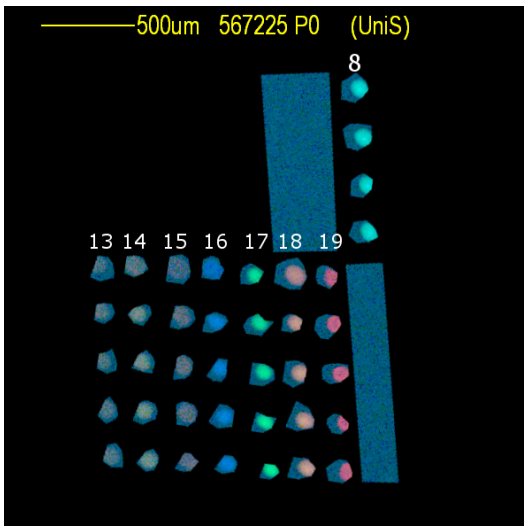


Figure 6.14: Elemental map of the array shown in figure 6.11. Sulphur: green, bromine: red and phosphorous: blue. Sulphur is an indicator of protein being present and bromine and phosphorous were key buffer components for many of the samples. The rectangular regions are scans showing that there was no elementally detectable sample in these regions (they show background, stronger in sulphur because there is more bremsstrahlung there). The results for the samples present are consistent with those expected given the sample buffers and protein concentrations. Columns 9-12 and 21-24 were not scanned due to beam time constraints.

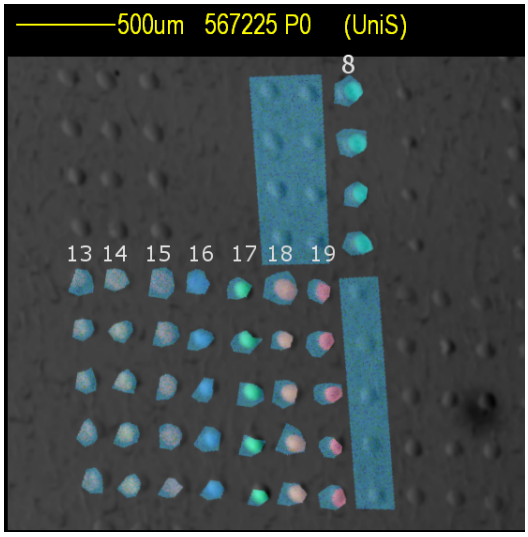


Figure 6.15: Overlay of figure 6.13 and figure 6.14 (linearly transformed due to different geometry of PIXE detector and optical microscope). This clearly shows the additional missing samples to the left of sample 8 and to the right of sample 19. Columns 9-12 and 21-24 were not scanned due to beam time constraints.

6.10 Development of new sample holders

Having proved the concept of printing protein arrays using a high throughput non-contact printer onto the polyprolene film, the next step was to develop an integrated solution allowing a workflow that could go from a set of samples in tubes to a printed array that could be placed with high precision directly into the sample chamber. This necessitated two new pieces of hardware:

1. A new sample holder onto which the film is glued which is both compatible with the ArrayJet printer and the IBC sample chamber.

2. A new sample stick (for the beamline) into which the sample holder (1) can be placed.

6.10.1 Sample holder design parameters

The ArrayJet printer defines the outer dimensions of the aluminium sample holder to be those of a standard glass slide (75 mm × 25 mm × 1 mm) and requires an extremely flat surface. It was decided that trying to pack a film tensioning mechanism into the 1 mm depth would not be feasible, so a jig that pre-tensioned the film over the sample holder was designed to facilitate in the reliable and reproducible gluing of the 4 μm thick polyprolene to the holder. The new sample stick for the microPIXE beamline required a 2 mm section along each edge to secure the sample holder in place, which did not present any design problems. In order to give sufficient surface area for good adhesion, 5 mm was set as the minimum dimension of any surface onto which the film had to be attached.

It was thus decided to use the design in figure 6.16. The large 8 mm × 8 mm windows are

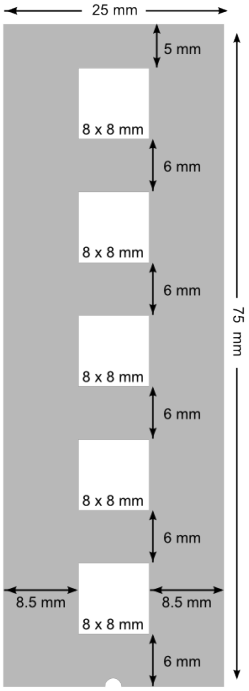


Figure 6.16: Design of new aluminium sample holder. With five 8 mm × 8 mm windows, the slide design allows flexibility for multiple sample dispensing protocols. The notch guarantees the correct orientation when placed into the beamline sample stick.

of sufficient size to pipette up to 4 drops by hand, or to use another drop delivery method (e.g. contact based printing) if required for other applications. Based on a 10×10 array of drops, this sample holder could hold up to 500 samples, allowing for a great deal of flexibility in data collection protocols. The notch at the bottom of the slide guarantees its correct orientation in the sample stick, and aids with alignment of the holder.

6.10.2 Sample jig

In order to ensure that the polypropylene was sufficiently flat for reliable printing and to allow the sample holders to be prepared in a reproducible manner, a jig was designed to hold the film

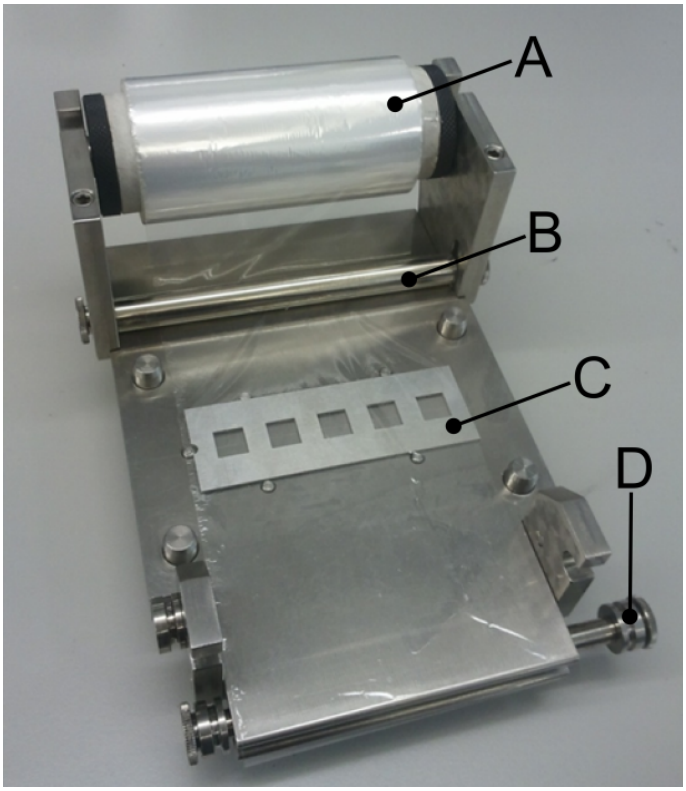


Figure 6.17: Jig for assembly of sample holders. The film is kept clean by being retained on a hop (A). The first tensioning roller (B) holds one end of the film taut, the sample holder (C) is secured in the middle and the second tensioning roller (D) adds light tension to the film while the glue dries. The experimenter is then free to gently smooth wrinkles out of the film a few times during drying, ensuring a consistent result.

taut over the sample holder during drying, shown in figure 6.17. The design was finalized with, and manufactured by, the Physics Department workshop at Oxford University. It clamps the film (from a roll) at one end and tensions it at the other. By using a non-instant contact glue (e.g. Revell Contacta Professional), the user can periodically smooth out any small ripples that appear in the film, thus allowing it to dry smoothly over a period of hours. Using the new holder, figure 6.16, and the film dispensing jig, figure 6.17, it was possible to reproducibly fabricate elementally clean sample holders with similar levels of film tension. This allowed us to move on to optimising the printing protocols.

6.10.3 Sample stick

For automated data acquisition, the sample holder must be reliably held perpendicular to the beam and in the plane of the x and y axes of the sample stage motor. This was not possible with the old sample stick due to it only clamping the sample holder along the vertical axis, requiring foil shims to secure the sample holder, and having a jaw for the sample holder which was too wide horizontally since it was only 25 mm along this axis (i.e. the jaw overlapped into the sample windows). In addition, excessive sample drift was observed during early overnight runs due to differential thermal contraction of the sample stick relative to the chamber.

By making the new stick out of stainless steel (similar to the rest of the sample chamber) rather than aluminium (as was the case previously), this differential thermal contraction was reduced to within an acceptable tolerance of less than 10 μm. The new sample stick was designed in collaboration with Dr. Geoff Grime and Dr. Vladimir Palitsin from the IBC, and was then manufactured to secure the sample holder rigidly and integrate the control samples directly into the mount (Pb Glass for calibration, YAG fluorescence screen for centring the beam and Cu wire grid for determining the beam size). The final design is shown in figures 6.18 and 6.19, and represents a significant improvement in experimental capabilities for users at the IBC.

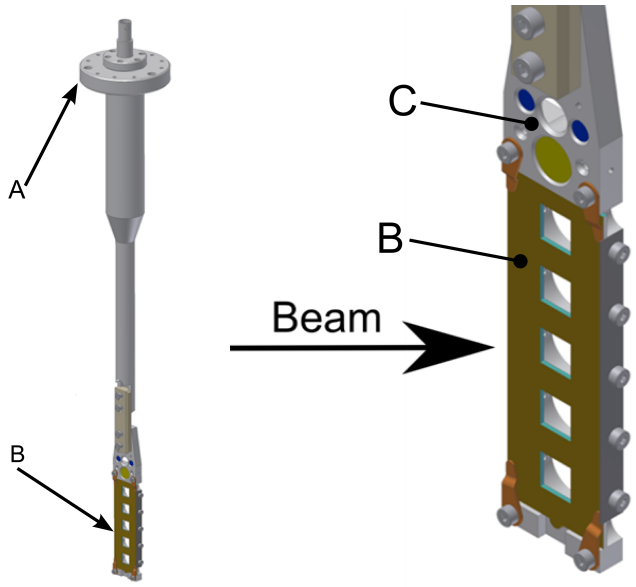


Figure 6.18: 3D representations of the new sample stick design showing the mounting point where this stick forms a seal with the sample chamber flange (A), the sample holder (B) and the in-built calibration standard samples (C). Adapted from figures kindly provided by Vladimir Palitsin, University of Surrey.

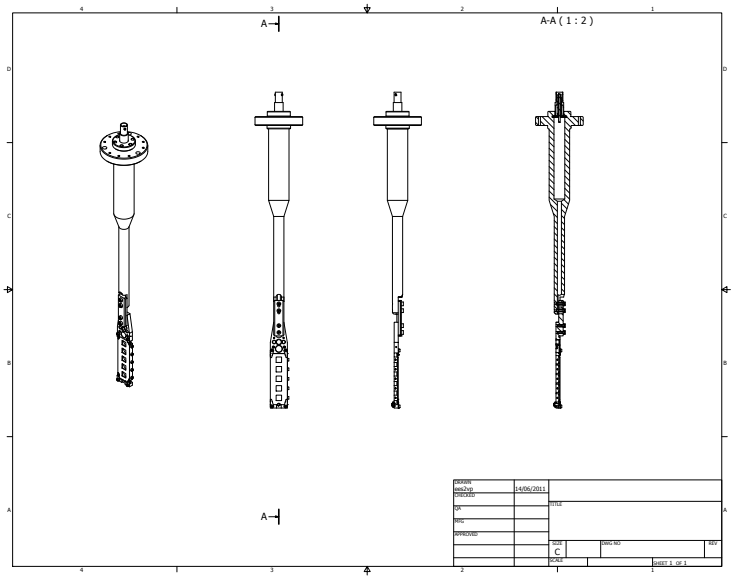


Figure 6.19: Engineering drawing of the new sample stick assembly designed to securely hold new sample holders and eliminate thermal drift. Figure kindly provided by Vladimir Palitsin, University of Surrey.

6.11 Optimisation of printing protocols

6.11.1 Verification that there is no cross-talk between drops

Having successfully printed micro-arrays of both protein solutions and salt standards onto the new sample holders (array shown in figure 6.20), it was extremely important to prove that there was no cross talk between drops caused either by the printing head not being completely clean, by printing splatter, or by any halo on the ion beam itself. Because the ArrayJet printer head collects and prints a whole row of the array at a time, it was important to check the elemental composition of adjacent drops both from a single row and from a single column:

- Drops from a single row are printed out from adjacent printer heads at a high velocity. Checking this verifies that there is no splatter or bleeding/merging of the drops as they form.

- Drops from a single column are printed from the same head of the ArrayJet device. By verifying these, we are also checking that there is no residue from the previous sample left in the print head chamber which could contaminate subsequent drops.

This was achieved by printing an array of the various salt samples detailed in table 6.3, such that there were highly contrasting samples adjacent to each other both in rows and in columns, as shown in table 6.4. Two overnight runs were then performed, each of which allowed 30 data sets per line to be collected along a line between two drops: one set between two rows, figure 6.21, and one set between two columns: figure 6.22.

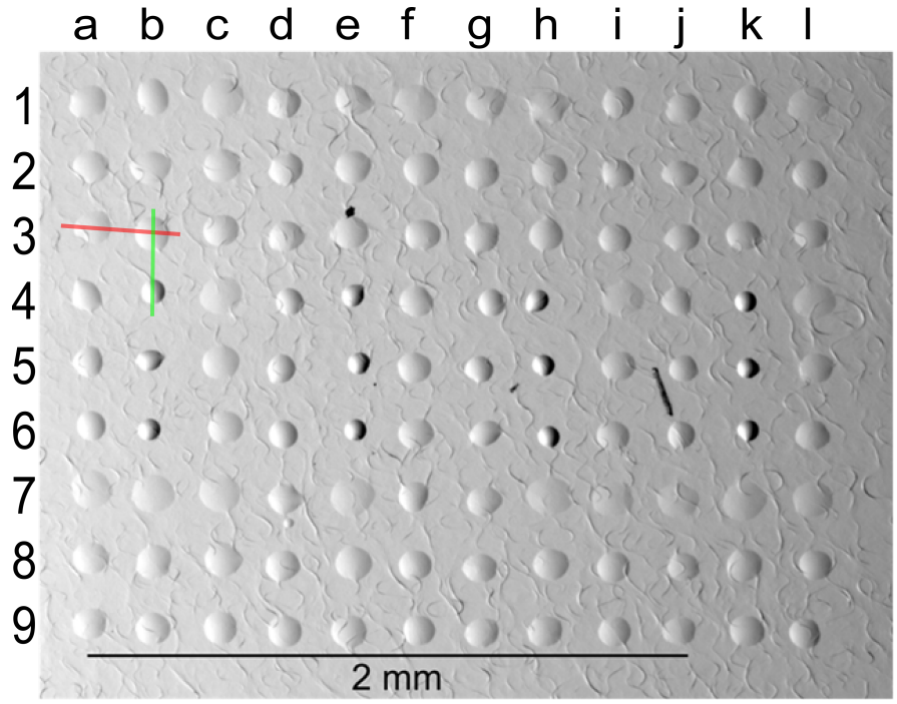


Figure 6.20: Microscope image of standard salt array showing the successful printing of multiple samples onto the polypropylene film. This demonstrates that the film can be fixed flat enough onto the mount for good spatial resolution and that the surface of the film is suitable for drop adhesion. The green line shows where the data for figure 6.21 (adjacent row verification) were collected and the red line from where those for figure 6.22 were collected (adjacent column verification).

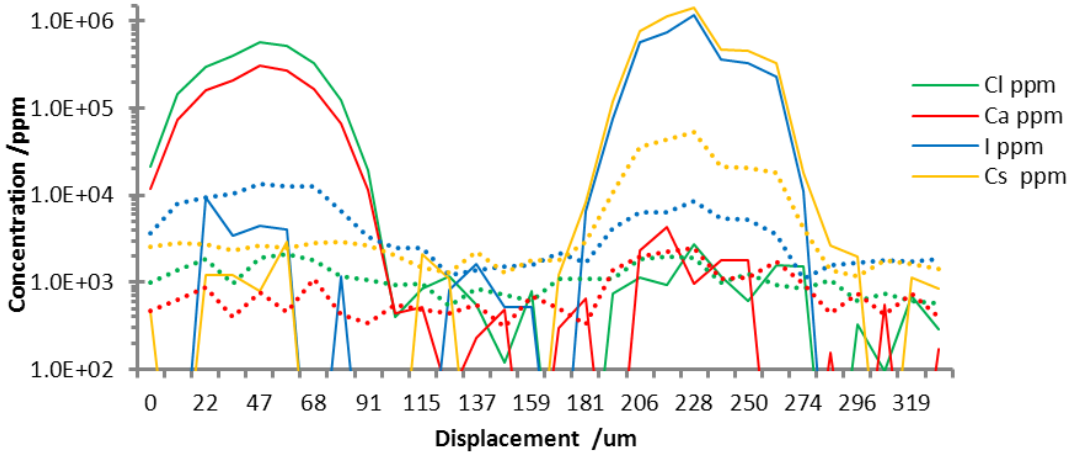


Figure 6.21: Elemental data from the analysis of 30 PIXE spectra collected across two drops (CaCl and CsI) of our standard salt array. The clear absence of any calcium or chlorine in the CsI drop and *vice versa* demonstrates the absence of any cross talk. Errors in peak regions (above LoDs) are within the thickness of the line – note the logarithmic y scale – and errors below the LoD are meaningless, since the signal is of the same order as the background. The dotted lines indicate the LoDs for each element.

Sample	Contents
A	$Ba(C_2H_3O_2)$
B	$CaCl_2.2H_20$
C	$FeSO_4.7H_20$
D	$NaSO_4.6H_20$
E	CsI
F	ZnBr
G	LiCl
H	K_2HPO_4
I	RbCl

Table 6.3: Details of salt samples used for the cross-talk experiment, the array for which is shown in figure 6.20. All samples were prepared at 100 mM and in 10% glycerol.

	a	b	c	d	e	f	g	h	i
1	A	B	C	A	B	C	A	B	C
2	A	B	C	A	B	C	A	B	C
3	A	B	C	A	B	C	A	B	C
4	D	E	F	D	E	F	D	E	F
5	D	E	F	D	E	F	D	E	F
6	D	E	F	D	E	F	D	E	F
7	G	H	I	G	H	I	G	H	I
8	G	H	I	G	H	I	G	H	I
9	G	H	I	G	H	I	G	H	I

Table 6.4: Configuration of the array show in figure 6.20, using the samples as detailed in table 6.3.

Based on the data shown in figure 6.21 and figure 6.22, it can be concluded that there is no cross talk between drops in either adjacent rows or adjacent columns. This result provides validation that the sample preparation technique is clean, providing the grounding for the veracity of subsequent results that are of biological interest.

6.11.2 Sample additives for better print quality

In order to improve the print quality, 10% glycerol was added to the salt array in figure 6.20. This should approximately recreate the viscosity of the print buffer, which helps drop formation

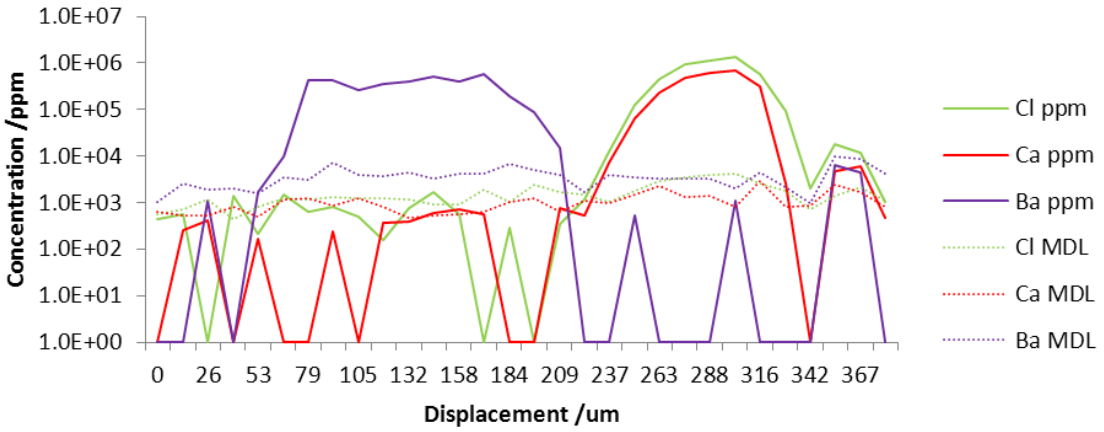


Figure 6.22: Elemental data from the analysis of 30 PIXE spectra collected across two drops ($Ba(C_2H_3O_2)$ and CaCl) of our standard salt array. Although the barium acetate drop has spread a little, probably due to the effect of the acetate on the viscosity, this again shows that there is no cross talk at all above the minimum detectable limit between the drops. Errors in peak regions (above LoDs) are within the thickness of the line and errors below the LoD are meaningless since the signal is of the same order as the background. Note the logarithmic y scale. The dotted lines indicate the Minimum Detectable Limit (same as LoD) for each element.

by minimising pressure differentials in the fluidics of the ArrayJet system. Based on the promising results both in terms of drop quality, and the absence of any missing half-columns of the array in figure 6.20, the effect of adding a small amount of glycerol to biological samples before printing was investigated.

Because it is important to minimize the diluting effect of adding a volume of glycerol to a sample, an experiment was carried out to determine the minimum quantity of glycerol that would have the required effect. A highly concentrated sample of cytochrome C oxidase kindly provided by a collaborator (Dr. Florian Hilbers, MPI of Biophysics, Frankfurt) was therfore used. The sample consisted of 23 μl of 97 mg/ml protein, allowing for a serial dilution in seven steps from 13.4 mg/ml to 0 mg/ml across the 7 rows of the array as shown in table 6.5.

Samples were prepared by adding 8.9 μl of the protein stock to each relevant well in the ArrayJet compatible Genetix 384 well plates, and using the Oxford LMB Mosquito high throughput pos-

itive displacement liquid dispensing system to add 1.1 μl of glycerol solution (to handle the low volume and extreme viscosity). Because the Mosquito always dispenses solutions 8 at a time and must dispense the same volume for all 8 samples, the glycerol was pre-diluted in the test protein's sample buffer so that, for all samples, adding 1.1 μl of solution from the appropriate 96 well stock plate provided the correct final concentration (see table 6.6).

Final array row:	a	b	c	d	e	f	g
Final concentration (mg/ml)	0	0.5	0.9	2.2	4.5	8.9	13.4

Table 6.5: Final printed array protein concentrations after addition of glycerol samples from table 6.6 for the cyctochrome C oxidase protein sample.

Final array column	Final glycerol concentration	Glycerol concentration in 1.1μl aliquot
1	0%	0%
2	1%	9.1%
3	3%	27%
4	5%	45%
5	7%	63%
6	9%	82%
7	11%	100%

Table 6.6: Pre-dilution of 1.1μl glycerol aliquots for the Mosquito system that led to the correct final concentrations of protein samples in the sample array.

As an illustration of the mapping shown in table 6.2, the configurations of the 96 well plate containing the glycerol reservoirs, table 6.7, and the 384 well plate both before, table 6.8, and after, table 6.9, the addition of the glycerol by the Mosquito device are shown.

This combination of custom programming of the Mosquito, and selective pre-dilutions of solutions, allowed accurate preparation of the array. Two arrays were printed using the ArrayJet: a 1$\times$ array, with one drop per array coordinate, and a 3$\times$ array, with three drops per coordinate position.

96 WELL PLATE	1	2	3
A	X	6	2
B	4	X	X
C	5	1	X
D	X	7	3
E	X	6	2
F	4	X	X
G	5	1	X
H	X	7	3

Table 6.7: Set-up of the 96 well plate containing the glycerol reservoirs with sample concentrations as indicated in table 6.6. Because the pitch of the Mosquito pipettes is twice that of the 384 well plates, glycerol was aliquoted four times from each column of the 96 well plate to make up the final configuration shown in table 6.9. For example the contents of column 1 was aliquoted to the 384 well plate into C1, E1, K1, M1 then into D1, F1, L1, N1, into C2, E2, K2, M2 and finally into D2, F2, L2, N2. 'x' indicates an empty well.

Figure 6.23a is a microscope image of the 1x array taken after 2 hours of drying time in the humidity controlled ArrayJet chamber and shows a clear, high quality array. An attempt to take another image 2 days after printing proved impossible with the laboratory microscope set-up due to lack of contrast (sample normal to lens axis). It was, however, possible to identify the array in the ion beam chamber thanks to the microscope being at $45°$ to the sample and having a different illumination geometry. This optical visualisation challenge was also evident with the $3\times$ array, so for future experiments, an optically opaque 'landing light' was placed in array coordinate A1 for localising the array, since sample spacing is known so that visualisation of the drops would not then be required when the samples were mounted in the beamline chamber.

Figures 6.23b and 6.23c show, respectively, the two key elemental maps for the $1\times$ array: the sulphur map, a proxy for protein content (figure 6.23b) and the potassium map, the most elementally visible element in the samples due to its high concentration in the buffer (figure 6.23c). Due to the limited beam time on each sample, the sulphur map only shows the sample drying structure

	Protein Contents					
384 well plate	**1**	**2**	**3**	**4**	**5**	**6**
A	x	x	a	c	a	c
B	x	x	b	d	b	d
C	a	c	x	x	x	x
D	b	d	x	x	x	x
E	a	c	a	c	x	x
F	b	d	b	d	x	x
G	x	x	a	c	a	c
H	x	x	b	d	b	d
I	x	x	e	g	e	g
J	x	x	f	x	f	x
K	e	g	x	x	x	x
L	f	x	x	x	x	x
M	e	g	e	g	x	x
N	f	x	f	x	x	x
O	x	x	e	g	e	g
P	x	x	f	x	f	x

Table 6.8: 384 well set-up for the protein/glycerol array based on the mapping shown in table 6.2. The wells that are colour coded in red and with 'x' were filled with print buffer only. The yellow wells were left empty since these would correspond to the 8th row on the printed array, which was not used. When printed, this would produce an array with 7 spots of sample 'a' and 5 blanks in the first row, 7 spots of sample 'b' and 5 blanks in the second row etc.

			Final Contents			
384 well	1	2	3	4	5	6
A	x	x	a6	c6	a2	c2
B	x	x	b6	d6	b2	d2
C	a4	c4	x	x	x	x
D	b4	d4	x	x	x	x
E	a5	c5	a1	c1	x	x
F	b5	d5	b1	d1	x	x
G	x	x	a7	c7	a3	c3
H	x	x	b7	d7	b3	d3
I	x	x	e6	g6	e2	g2
J	x	x	f6	x6	f2	x2
K	e4	g4	x	x	x	x
L	f4	x4	x	x	x	x
M	e5	g5	e1	g1	x	x
N	f5	x5	f1	x1	x	x
O	x	x	e7	g7	e3	g3
P	x	x	f7	x7	f3	x3

Table 6.9: Final contents of the 384 well plate before printing. The glycerol contents are as described in table 6.6 and the protein concentrations as in table 6.5. This results in an array with rows a,b,c...g and columns 1-7 giving a two dimensional gradient in both protein and glycerol concentration. Column Coding as in table 6.8.

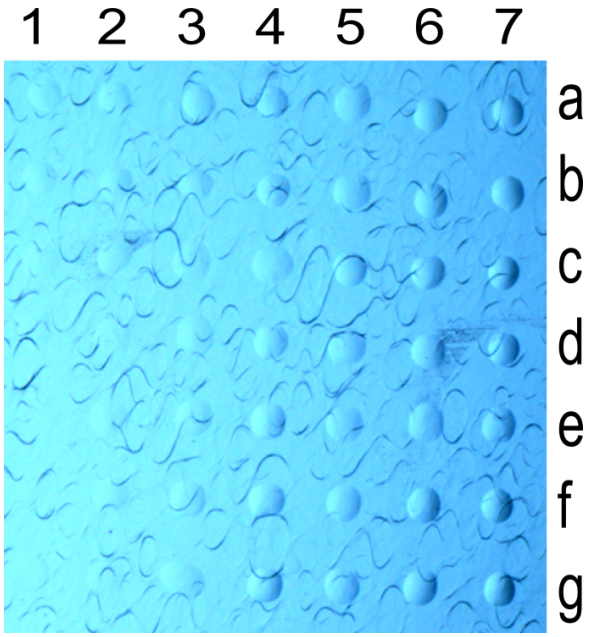
for the last two rows containing 8.9 mg/ml and 13.4 mg/ml of protein respectively. The potassium map, which in general provides better contrast, shows slightly more compact drops starting from column three, or at 3% and higher glycerol concentrations. We thus adopted a minimum of 3% glycerol as a recommendation for 1× (1 drop per position) arrays.

Figure 6.24a, the 3× array (3 drops per position), shows a much stronger trend in the glycerol gradient with well formed drops only visible at 11% glycerol. The extent of merging can be seen to increase as glycerol decreases in concentration. This is also shown in the elemental maps, figures 6.24b and 6.24c, where an increase in the level of drop merging is visible right down to the lowest glycerol levels.

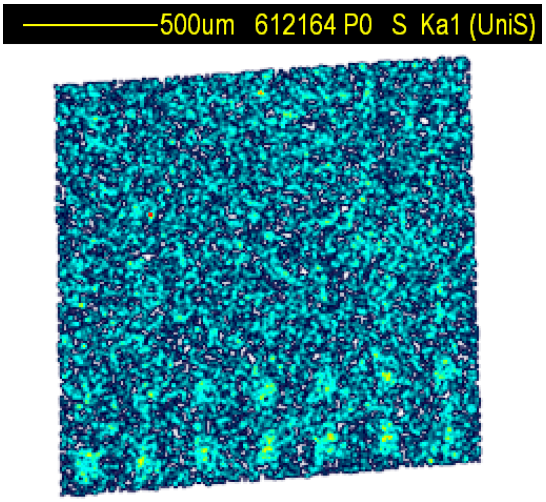
A key observation in comparing either of the microscope images with its associated elemental map is that a sample not being visible under the microscope does not necessarily imply that it is not present on the film. This is not so unexpected for samples that are of the order of 1 μm in thickness. The improved visibility of the high glycerol concentration samples in the microscope after 2 hours as seen in both arrays is probably due the slower drying of glycerol containing samples.

6.11.3 Drop multiplicity

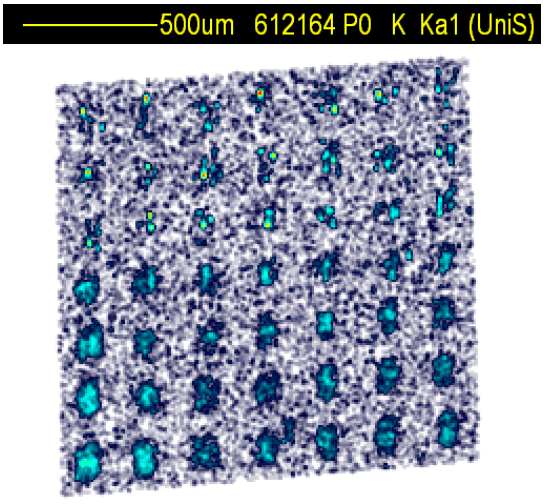
Printing multiple drops will improve the final areal density, but will also increase the likelihood of drop merging, and reduce the quality of the array. In order to determine the extent to which printing multiple drops should be attempted, an estimation of the improvement to final areal density of protein was made.



(a) Microscope image taken after two hours drying time of the glycerol/protein concentration gradient array for single drops (1× array). The glycerol concentration was, in columns 1-7 from left: 0, 1%, 3%, 5%, 7%, 9%, and 11% respectively. The protein concentration was, in rows from top to bottom: 0 mg/ml, 0.44 mg/m, 0.89 mg/ml, 2.22 mg/ml, 4.45 mg/ml, 8.9 mg/ml and 13.4 mg/ml respectively. This clearly showed that glycerol concentration has an effect on the visual clarity, roundness and diameter of the drops and that protein concentration has no visible effect.

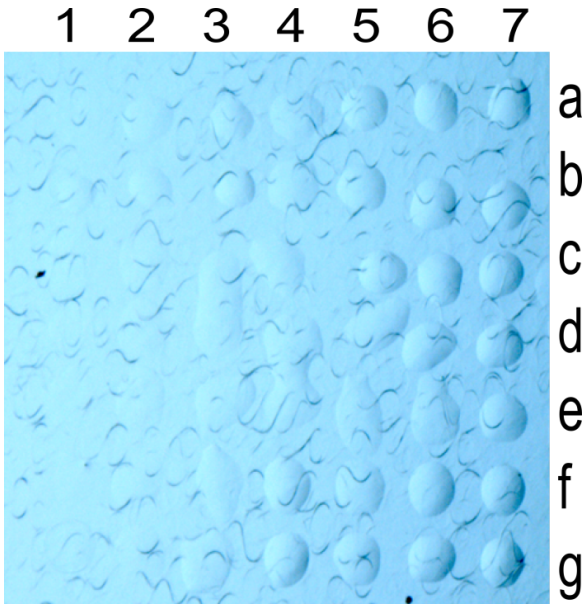


(b) Sulphur map of the 1× array shown in figure 6.23a collected over 12 hours. Very low signal is present due to minimal beam time on the actual sample (see section 6.12.1). This does however show the higher levels of protein concentration in rows F and G, consistent with the printing scheme. The scale bar is shown at the top of the image for all elemental maps of arrays.

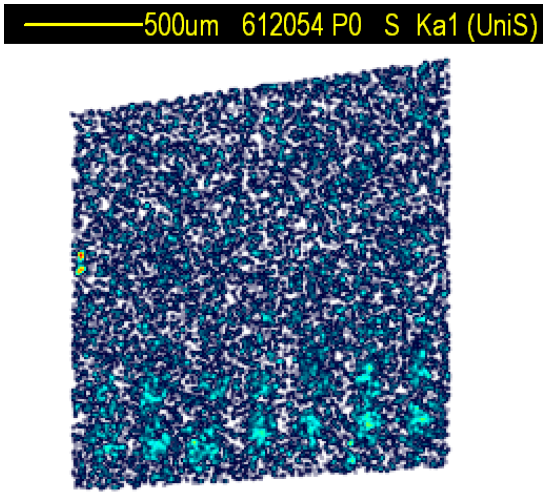


(c) Potassium (main buffer salt) map of the 1× array shown in figure 6.23a. Low signal is present due to minimal beam time on the actual sample (see section 6.12.1). This does however show that the whole periodic array is very clearly present, and that the drops are well separated.

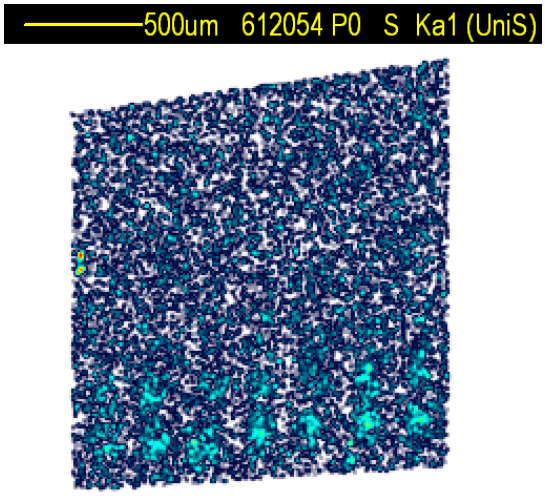
Figure 6.23: Microscope image (a), sulphur map (b), and potassium map (c) for the 1× glycerol and protein concentration gradient experiment.



(a) Microscope image taken after two hours drying time of glycerol/protein concentration gradient array for triple drops in each position (3× array). It can be seen that significantly larger drops are present, that at glycerol concentrations of less than 9% there is significant merging of drops and that the general row and column order is less good than in figure 6.23a.The glycerol concentrations are, in columns from left: 0, 1% ,3%, 5%, 7%, 9%, 11% respectively. The protein concentrations are, in rows from top to bottom: 0 mg/ml, 0.44 mg/m, 0.89 mg/ml, 2.22 mg/ml, 4.45 mg/ml, 8.9 mg/ml and 13.4 mg/ml respectively.



(b) Sulphur map of the 3× array shown in figure 6.24a, collected over 12 hours. Very low signal is present due to minimal beam time on the actual sample (see section 6.12.1). This does however show higher levels of protein in rows F and G, consistent with the printing scheme.



(c) Potassium (main buffer salt) map of the 3x array shown in figure 6.24a. Low signal is present due to minimal beam time on actual sample (see section 6.12.1). This does however show the periodic array, as well as some of the drops merging.

Figure 6.24: Microscope image (a), sulphur map (b), and potassium map (c) for the 3× glycrerol and protein concentration gradient experiment.

Assuming a hemispherical drop shape, the volume and area for one drop is

$$V_1 = \frac{2}{3}\pi r_1^3 \; ; \; A_1 = \pi r_1^2$$

giving a surface density of

$$\frac{V_1}{A_1} = \frac{2}{3} r_1 = \frac{2}{3}\left(\frac{3V_1}{2\pi}\right)^{\frac{1}{3}}$$

thus, for a doubling of volume (2 drops), there is only an increase in the areal density of $2^{1/3} \cong 1.26$ and for three drops, $3^{1/3} \cong 1.44$. This rough calculation indicates that there is an advantage to doubling the sample volume, but that it is far from a simple doubling of sample thickness and so is not worth pursuing at all costs. It was decided that for the 1.4 improvement in areal density estimated for 3× drops, the loss of array quality shown in figure 6.24a was not worthwhile.

A subsequent experiment with a series of samples that formed part of our collaboration with Novartis (See section 6.14) printing 1× and 2× arrays (both at 10% glycerol) did not show a decrease in array quality for two drops, and so a 2× printing protocol was adopted as a standard protocol offering a way of increasing areal density without sacrificing sample integrity[7].

6.12 HT data acquisition methods

Automated data acquisition requires the ability to cue up large numbers of runs, correctly tagging each of these with the appropriate sample information. In order to do this reliably, a Control File

[7] Arrays subsequently printed (June 2013) at the ArrayJet headquarters in Edinburgh have shown excellent spatial resolution up to 3×. This is likely due to improved tuning of the ArrayJet in-house machine compared to the one previously used.

that contains all the relevant fields for data collection and analysis was developed. This feeds the information on the analysis parameters to the control module of OMDAQ and also stores the sample information relevant for interpretation, thus keeping all the required data in one place. This control file is a Comma Separated Value (CSV) file with the following column headers, which allows the user to fully automate data collection:

1. Sample information (to be filled in by the sample owner):
   (a) Label: Micro-tube label before dispensing into array plate.
   (b) Owner: The group who prepared the sample.
   (c) Type: general sample types. Must be one of:
      i. Protein
      ii. Serum
      iii. Crystal (not for HT)
      iv. Cell (for future development)
      v. Homogenate (for future development)
      vi. Other
   (d) Description: The caption that OMDAQ will use. Similar to 'label' but can be longer.
   (e) Sequence Length: For proteins only.
   (f) Number of Cysteine residues: For proteins only.
   (g) Number of Methionine residues: For proteins only.
   (h) Elements of interest: Main elements that are of interest.
   (i) Sample Volume: Total volume of sample sent for analysis.
   (j) Sample Concentration: Nominal concentration of sample.
   (k) Current Buffer: Should include chemical formulae and concentrations. For crystals, this should be for both the mother liquor and the buffer.
   (l) Prep Notes: Any buffer exchanges, purification protocols involving metal (e.g. Ni) columns or other relevant notes.
   (m) Storage instructions: how to store the sample pre-printing (temperature).
2. Acquisition information (to be filled in by operator before array printing and data collection):
   (a) Plate coords: Coordinates in 384 well plate where sample is to be placed before printing.
   (b) Array coords: Final array coordinates of the sample after printing.
   (c) stageX, stageY: the x and y coordinates of the centre of the sample stage relative to the centre of the top left drop in mm (not needed for beam-deflection data collection).
   (d) Scan: 1 = point measurement, 2 = map, 3 = mask

(e) Size: if a map or a mask, this is the size in microns of one side of the map/size of the box that contains the mask.

(f) End conditions[8]: Syntax is: StopCon[n] = value; where StopCon is a case insensitive string selected from:

i. Element symbol: Value is the required statistical precision in %. = 100/sqrt(background subtracted peak counts). The X-ray peak used is the Kα (Z < 40) or Lα (Z >= 40)

ii. RBS: Value is the required number of counts in the RBS spectrum integrated between 200 keV and the beam energy.

iii. T: The upper limit on run time in seconds.

iv. Q: The upper limit on accumulated charge in nanoCoulombs.

This syntax was then added as an input module to the OMDAQ code by Dr. Geoff Grime, who is the author of the code, and who has been developing it independently for over a decade. This forms the basis of of how high throughput data collections can be controlled.

6.12.1 Spatial localisation of the spots in the sample chamber.

In order to collect data from multiple samples it is necessary to be able to automatically move to different areas of the film. It is also important to maximize the proportion of time the beam spends on the actual sample. By way of illustration, collecting a map of a typical entire array (10×10 samples, 200 μm centre-to-centre spacing, 60 μm diameter drops) has a sample area to total scan area ratio of:

$$Scan\ Area = (10 \times 0.2 + 0.08)^2 = 4.32\ \text{mm}^2$$

[8] In the end conditions, 'element' and 'RBS' conditions are AND conditions (required to be true simultaneously to trigger a stop). 'T' and 'Q' conditions are upper limits (stops whenever one of them is exceeded). 'n' is an optional spectrum number (0 - 7). If there is only one PIXE or RBS spectrum, this can be omitted. If 'n' is omitted and there are multiple PIXE or RBS detectors then the condition is triggered when the first spectrum meets the end criterion. There is no space between 'StopCon' and 'n' (e.g `StopCon3`). 'value' is a real number. This must start with a digit (0 - 9) and negative numbers are not accepted. Values less than 1 must start with 0 (i.e. 0.005, not .005). There can be any number of spaces surrounding the = sign. e.g. `Zn = 5, S = 5, Q = 100, T = 300` Will stop either when the statistical errors on both Zn and S are below 5%, the total charge reaches 100 nC or when the time reaches 300s, whichever happens first.

$$Sample\ Area = 10 \times 10 \times \pi \times (0.03)^2 = 0.28\ \text{mm}^2$$

$$\frac{Sample\ Area}{Scan\ Area} = 6.5\%$$

Thus only 6.5 % of the beam time is spent on the samples. For a typical 14h run, this equates to 31s per sample, which is far short of the minimum target of 5 min. It is thus essential to predefine the sample region so that satisfactory statistics can be obtained and the beam time is not wasted. The experimental set-up allows two ways of defining the irradiated region:

1. Deflecting the beam: once a map has been taken from an area on the sample film, it is possible to select sub-regions (masks), or a series of individual points from which to collect data. This has the advantage of higher accuracy and reproducibility, but the disadvantages of fixing the maximum region across which data can be collected (2-3 mm depending on beam energy) and of increasing the pixel size[9]. It also requires a ~1 hour map of the whole array to be collected before setting the HT run up so that samples can be identified from their elemental spectra.

2. Moving the sample stage: this has the advantage of having the possibility of moving across large distances (cm), and can be pre-programmed in the control file. However pre-programming relies on extremely high standards of array quality so that the sample spot positions are entirely predictable and also that the motors have limited accuracy (~10 μm error). A solution to this is to collect a circular mask at each predicted pre-programmed location and to apply

[9] Pixel size = scan size/512

a threshold filter to the elemental data to select the sample region.

In the first stages of development, the chosen method was to print enough samples within the beam deflection scan region (100 can easily be achieved in a 2 mm × 2 mm array), collect a long exposure map of the array, and then assign 4-5 point spectra per spot based on this; the extra set-up time required by the long exposure map acquisition is compensated for by the improved efficiency of being able to collect data overnight without intervention. This is termed Map And Points (MAP) data collection.

In the later stages of development (2013), the motor on the sample stage was upgraded, allowing much better accuracy and reproducibility. This led to the final protocol described in section 6.13.

6.12.2 Data analysis: choice of metrics

When dealing with large data sets, it is important to be able to condense the thousands of data into a manageable and a relevant set of summary statistics. With this in mind, the following metrics were chosen for each sample when using the MAP data collection protocol:

1. **Estimated elemental stoichiometry (number of atoms of element of interest per protein molecule):** the mean of the sulphur normalized elemental Parts Per Million (PPM), weighted by the inverse square of the statistical errors, multiplied by the total number of Cys and Met residues:

$$Estimated\ Stoichiometry = (N_{cys} + N_{met}) \times \frac{\Sigma_{replicates} \frac{S\ normalised\ elemental\ ratio}{statistical\ error^2}}{\Sigma_{replicates} \frac{1}{statistical\ error^2}}$$

2. **Standard Deviation of the replicates:** the standard deviation for the weighted mean of the S normalized elemental stoichiometries.

3. **Mean elemental PPM over mean LoD PPM:** an average measure of how significant the result is as a function of the LoD. If this is less than one, the element is considered not to be present, 1-3 potentially present and >3 significantly present, as suggested in the GUPIX (Maxwell *et al.*, 1989) user guide.

4. **Range/Mean:** The maximum value of the sulphur ratios minus the minimum value divided by the weighted mean. This provides a measure of the maximum spread of the data. This is particularly instructive for samples with a poor sulphur peak (either due to insufficient concentration or due to high levels of overlapping chlorine or phosphorous from the buffer).

These metrics are offered for every element analysed, for every sample. For a large run, this can still represent a great deal of data. The user can therefore automatically be presented with only elements that reach a certain standard significance threshold (3× LoD or 1× LoD), simplifying the output significantly over a several thousand cell spreadsheet.

6.13 Final protocol and method for HT micro-PIXE

The final MAP protocol used for the high throughput microPIXE data collection is presented in figure 6.25 as well as detailed below in text format:

Sample preparation:

1. Samples prepared in sulphur free, low chlorine and low phosporous buffer to minimum concentration as specified from the Y-G plot (see figure 6.4). Sample holder assembled using jig as described above.
2. Control File for all samples prepared.
3. Samples transferred to 384 well plate (as per control file) and 10% v/v glycerol added using adapted Mosquito protocol.
4. Samples printed onto sample holder in ArrayJet non-contact printer. Left to dry in ArrayJet chamber for at least 2 hours to minimise dust contamination.

Data Acquisition:

1. Collect calibration reading on Pb Glass, profile beam on Cu grid[10] and check that there is no scattering from the Faraday cup by taking spectra with no sample in the beam.
2. Locate array and collect long map to localise samples (1 hour).
3. Assign points to each sample in the Control File.
4. Automatic data acquisition (14 hours).

Analysis:

1. Analyse Data in OMDAQ.
2. Combine Analysis CSV from OMDAQ with Control File CSV.
3. Export to summary spreadsheets.

Although useful, this protocol is limited by the fact that it requires a long map, and the selection of multiple points on a map can introduce a spread of concentration values for which it is hard to account. Following an upgrade of the sample stage mechanism at the IBC, a stage-moving based protocol is currently under development:

1. Localise the array in the sample chamber, and record the (x, y) coordinates of the corners interactively in OMDAQ. By additionally providing the number of rows and columns in the array, the software interpolates predicted spot positions.

[10] A centring grid built in to the new sample holder that provides a standard edge profile suitable for focusing the beam.

2. The stage is moved to each of these pre-programmed spots, and a small ($\sim 150^2$ μm^2) map is collected at each of these points. The software attempts to find a circular region (the drop) in the sulphur map in real time.
3. Once a good fit is made (1-2 mins), a masked collection is performed on the circular region until the end condition specified by the input file is met. During this masked collection, no beamtime is wasted on regions outside of the sample.
4. The data from the two runs (map and mask) are combined and analysed.

This protocol has the advantage of analysing a whole drop, and, because it records spatial information[11], it is possible to perform statistical analyses to look for hot spots, or to cross-validate various segmentations of a mask (e.g. break the pixels in the image into two or more sets, and correlate the results for these.). This highly promising method of validation is, however, out of the scope of this thesis, and it is hoped that it will provide an exciting project for a future student.

A feasibility run in June 2013 using this method has been performed, and the sample localisation and data acquisition was highly successful. All 35 samples were successfully automatically identified by the circle fitting system within 30s, and the processed data from the masked collections was of a higher statistical quality (more counts) than comparable data based on multiple single point data collections.

6.14 Measurements in collaboration

In addition to the development work carried out on the HT method, a number of experiments were performed as collaborations with other groups, providing access to relevant samples, and also a valuable insight into potential applications of the the technique.

[11] The software produces a listmode file that records the x, y, beam position, time, and energy of each X-ray event in the detector.

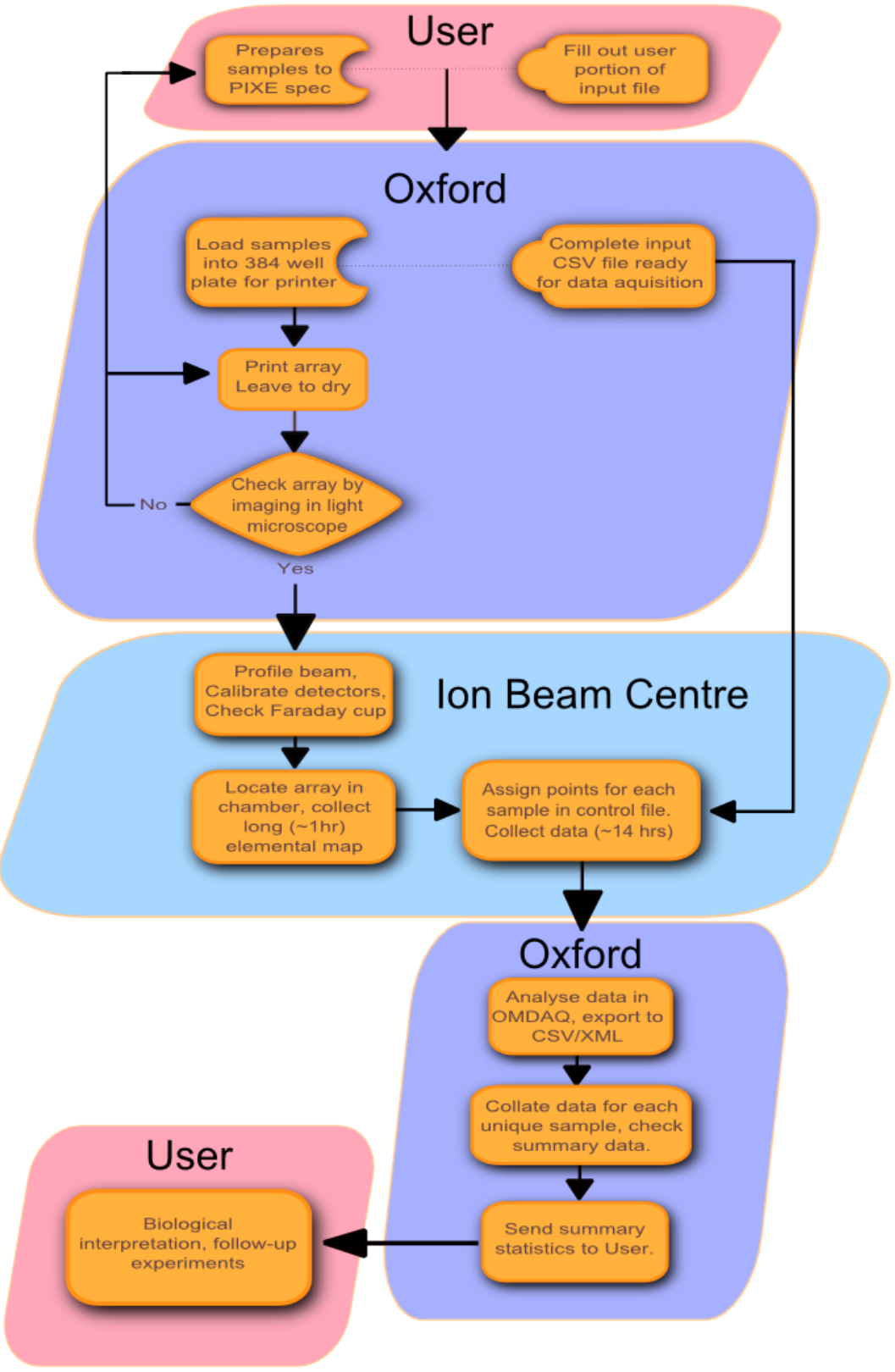


Figure 6.25: flow-chart of MAP PIXE protocol. Different physical locations are colour coded: the user's institution in red, Oxford in blue and the Ion Beam Centre in Guildford in pale blue.

1. With Shee Yong and Prof. Ben Berks (Biochemistry, Uni. of Oxford): analysis of the alkaline phosphatase PhoX to identify metal binding of two Fe and three Ca ions in the X-ray structure.

2. With Meindert Lamers (MRC, Uni. of Cambridge): Analysis of engineered cytochrome proteins. Stoichiometry was successfully determined.

3. With James Lillington and Prof. Susan Lea (Pathology, Uni. of Oxford): PhoD alkaline phosphatase crystals to identify putative bound Fe and Ca. Samples exhibited impurities and could not be analysed.

4. With Oscar Chen and Fran Platt (Pharmacology, Oxford): analysis and method development for a wide variety of non-protein biological samples, discussed above.

5. With Emma Dixon and Prof. Nicole Zitzmann (Biochemistry, Uni. of Oxford): Analysis of membrane protein complex P7 for metal binding and multimer stoichiometry. Removal of excess metal before analysis was unsuccessful.

6. With Dr. Robert Gilbert (STRUBI, Uni. of Oxford): Analysis of TIF32 protein for metal identification. Samples were of insufficient concentration.

7. With Novartis, (Cambridge, USA). Analysis of the effect of different preparation protocols on metal binding. More than 25 proteins successfully analysed.

8. With Dr. Michael Demydchuk and Prof. Randy Read (LMB, Uni. of Cambridge). Mea-

surements of IDS Iduronate-2-sulfatase, a lysosomal sulfatase associated with Hunter syndrome. Verified correct identification of metals as phosphorus and calcium.

9. With Dr. Lauren Jackson and Prof. David Owen (CIMR, Uni. of Cambridge). Analysing bound metals in fragments of VARP, a protein involved in vesicle fusion. Zinc identified.

10. With Dr. Arwen Pearson (Astbury Centre, Uni. of Leeds). Identification of active site stoichiometry in mutant series from several engineered enzymes.

11. With Dr. Helge Dorfmueller and Prof. Daan van Aalten (Dundee Uni.). Identification of bound metals in mutants of NodC, an acetylglucosaminyltransferase. Sample concentration was at first insufficient, then removal of Cl from the buffer was not successful, leading to failure to accurately quantify the sulphur signal.

12. With Sarah Rollauer and Prof. Susan Lea. Biophysical characterisation of protein NirDL, a fusion protein involved in heme biosynthesis, to identify putative Fe content. Fe found, but only at sub-stoichiometric levels (0.1 atoms/protein molecule).

13. With Dr. Steven Hardwick and Prof. Ben Luisi (Biochemistry, Uni. of Cambridge). Analysis of CcFLAG-Degradosome, a multi protein complex, with putative iron and zinc. Analysis identified multiple metals, but was inconclusive with regards to stoichiometry due to low sample concentration, and the large size of the complex (~5000 residues).

14. With Dr. Ehmke Polh (Durham Uni.). Analysis of ferric uptake regulator mutants for iron stoichiometry. Stoichiometry identified successfully.

Chapter 7

Conclusions

7.1 Radiation Damage

The work presented in this thesis represents a major step forwards in making MX radiation damage research directly applicable to routine data collection on challenging samples. It has addressed the problems associated with not taking into account the experimental beam and crystal profiles, and the inability of current systems to model complex strategies involving translation (Krojer *et al.*, 2013).

7.1.1 Comparing dose metrics

The dose metrics I_n/I_1 and B_{rel} have been investigated under carefully controlled conditions and over a variety of dose contrast conditions, and I_n/I_1 has been shown to exhibit significantly more reproducible behaviour than B_{rel}, with respect to the dose metrics proposed here. This is attributed to I_n/I_1 being more robust to small fluctuations in the weak high resolution data, and less sensitive to the particulars of the scaling step during data processing.

7.1.2 Using RADDOSE-3D and DWD in routine MX

The result that Diffraction Weighted Dose (DWD) faithfully predicts overall intensity decay in a strategy independent fashion is extremely promising, and has potential impacts before, during, and after data collection:

1. **Optimising dose strategy.** RADDOSE-3D, which outputs Diffraction Efficiency, can be used to create general guidelines for optimal dose strategy (see Chapter 5), providing a valuable

reference for routine experiments. In extreme cases where well diffracting crystal volume is at a premium, it can be used to generate and test a bespoke strategy which makes the best use of all the diffraction volume.

2. **Predicting crystal lifetime.** Given a 3D exposure strategy, RADDOSE-3D with DWD can be used to assess the maximum tolerable dose for a sample, using either a poorly diffracting crystal volume to characterise the decay, or global guidelines such as those found in Owen *et al.* (2006). Although RADDOSE-3D can create a reliable I_n/I_1-DWD plot, it is not yet clear how to best apply this to estimations of dose lifetimes. As discussed in section 4.4.4, the current guidelines for a predictive 'cut-off' of I_n/I_1 are somewhat empirical, and would benefit from further work made possible by the use of RADDOSE-3D. One attractive approach is to evaluate the resolution cut-off, as defined by $CC_{1/2}$=0.5, as a function of DWD, thus investigating a parameter that is unambiguously of key biological importance against a quantity that can be modelled.

3. **Data processing.** Potentially the most exciting possible application of RADDOSE-3D and DWD is their application to data processing. Because RADDOSE-3D can output an image-by-image DWD which takes into account crystal thickness and dose history, it can serve as a powerful additional source of input to scaling, and offers real promise in tackling the difficult, but high reward problem of zero-dose extrapolation (Diederichs *et al.*, 2003). Additionally, since RADDOSE-3D models both diffraction decay, and crystal thickness for each image, it provides estimates of two of the major contributors to scaling. Comparison between a 'simulated

scaling' approach using RADDOSE-3D, and standard intensity scaling could prove instructive, and potentially lead to better error models in scaling. A pre-requisite for these approaches would be a better understanding of the form of the diffraction decay model η, and this is actively under research by a project student in the Garman Group.

7.1.3 Working with challenging crystals

Processing data from subsequent damage sets with difficult crystals is not trivial. Significant challenges arise as a consequence of balancing the need for a consistent treatment (so that datasets can be compared effectively) with processing parameters that generate the best statistics. Crystals change with damage, and if we are trying to observe those changes in the data processing statistics, it is not desirable to allow the integration and scaling steps to 'correct out' these reductions in diffraction quality. On the other hand, with difficult crystals such as $PepT_{St}$, successful processing is not possible without allowing some of these 'fudge factors' to change.

Although in principle integration parameters should not affect the output intensities, standard radiation damage data processing protocols used in the group call for a constant set of input parameters. This approach fails much earlier with difficult protein samples, and so more work is warranted to explore the effects of an unrestrained integration for sequential damage datasets. The work in section 4.4 provides valuable guidance for future experiments seeking to optimise the use of RADDOSE-3D in an on-line environment.

7.1.4 Photoelectrons, and what RADDOSE-3D does not do

A fundamental assumption of the work presented here is that damage is localised to the approximate region where the primary interaction occurs. Since photoelectron path lengths are known to be of the order of ~4 μm at E_{X-ray} = 18.5 keV (Sanishvili *et al.*, 2011), the assumption only holds for beams and crystals significantly larger than this path length. Although the regime considered here encloses the majority of modern synchrotron based data collection set-ups, it is important to note that it is not comprehensive, and that extensions to the code-base must be made to take account for photoelectrons escape from the beam area, or from the crystal itself.

7.2 MicroPIXE

The work presented here has taken microPIXE from a low throughput, 10 sample per day technique, to a high throughput technique, which can handle an order of magnitude more samples in the same timeframe. Following the proof of concepts presented here, the next step is to develop automated statistical validation tools, and to open up microPIXE to a wide variety of biological users. The bulk of the science has been done, and these following stages are primarily commercial and logistical in nature.

Appendix 1: RADDOSE-3D command reference

The final version of this command reference was written by both Markus Gerstel and OBZ, based on a draft created by OBZ.

RADDOSE-3D can take input from one or more files and/or from standard input (STDIN). Any input will be processed by the `InputParser` class and the RADDOSE-3D ANTLR parser. This section describes the syntax of accepted input. Advanced users of RADDOSE-3D can create their own input method that need not rely on the `InputParser` class or the RADDOSE-3D ANTLR parser. This feature will not be covered in detail in this reference. The simplest use case of RADDOSE-3D will involve only one file describing the entire experiment. In some instances it may be desired to split up the input into a number of files, e.g. one file describing the crystal, one automatically updated file describing the current beam on the beamline, and one file chosen from a set of possible wedge strategies. Each file can contain an arbitrary number (including none) of `Crystal`, `Beam` and `Wedge` block (henceforth called *blocks*). Splitting up blocks across multiple files is however not allowed. The parser will read the input sequentially, and, when multiple sources are given, one source after the other in the specified order. While the parser may accept `Crystal`, `Beam` and `Wedge` blocks in any order, the exposure of a wedge can only take place if both the crystal and the beam have been set either in an earlier file or before the `Wedge` block within the same file.

8.1 General syntax considerations

Any keywords specified below are case-insensitive. Upper (**`CRYSTAL`**), lower (**`crystal`**) and mixed case (**`CrYsTaL`**) are equivalent.

The characters **#**, **!** and the character sequence **//** denote the start of a comment. Any text from that position until the end of the current line is ignored.

Tabular and newline characters are treated as white space. They can therefore by freely used to format the file for increased readability.

The order of statements within a `Crystal`, `Beam` and `Wedge` block generally is not relevant. There are two exceptions to this rule: The leading keyword (**CRYSTAL**, **BEAM**, **WEDGE**) must be the first keyword of the block. If a keyword is repeated within the same block, then the latter will always override the former. Every block must be self-contained, e.g. the energy set for the previous `Beam` is not remembered when setting up the following `Beam`, and must be repeated.// Numeric values can be given in scientific notation ($2.0\text{e}2 = 2e + 2 = 200$), negative values may not have a space between the sign ('-') and the value ($-1.9\text{e} - 1 = -.19 = -0.19$).

8.2 `Crystal` block

A `Crystal` block must begin with the keyword **CRYSTAL**. At least the **TYPE** and **DIMENSION** must be specified. Depending on the chosen **TYPE** further declarations may be required.

8.2.1 **TYPE**

With the keyword **TYPE** the underlying crystal implementation is chosen. Currently two distinct crystal implementations exist:

TYPE CUBOID defines a solid crystal with a cuboid shape.

TYPE SPHERICAL defines a solid crystal with a spherical shape.

8.2.2 DIMENSION

DIMENSION specifies the size of the crystal. Dimensions are given in micrometers (μm). The keyword **DIMENSION** can take either one or three parameters:

DIMENSION ***D*** with a single number (see section 8.1) as parameter is used for specifying the crystal dimensions for spherical crystals. The parameter sets the crystal diameter. This syntax cannot be used for cuboid crystals.

DIMENSION ***X Y Z*** with three numbers as parameters *X*, *Y* and *Z* is used to set the dimensions for cuboid crystals (**TYPE CUBOID**). *X* defines the length of the crystal orthogonal to both the beam and the goniometer at L=P=0, (see below) *Y* defines the length along the goniometer axis at L=P=0 and *Z* defines the length along the beam axis.

If three parameters are given for a spherical crystal (**TYPE SPHERICAL**) the value for *X* sets the diameter of the crystal while the values of *Y* and *Z* are ignored.

8.2.3 DDM / DIFFRACTIONDECAYMODEL

`DDM SIMPLE`

`DDM LINEAR`

`DIFFRACTIONDECAYMODEL LINEAR`

`DIFFRACTIONDECAYMODEL SIMPLE`

How voxel diffraction intensity changes with dose. Simple: no change, Linear: linear decrease in resolution-integrated intensity, with half-intensity at 43 MGy. Default is Simple.

8.2.4 PIXELSPERMICRON

`PIXELSPERMICRON` ***F***

specifies the resolution of the voxel grid used to represent the crystal in voxels/μm. Defaults to 0.5 voxels/μm.

8.2.5 ANGLEP

ANGLEP ***F*** sets the angle in the plane of the loop between the crystal *Y* axis and the goniometer axis. The angle is to be given in degrees, but without the degree symbol (°). The default *P* ('plane') angle is $0°$.

The rotation angle to be applied to the crystal in the plane of the loop (right handed rotation about *Z* axis applied to all voxels, as shown in figure 8.1).

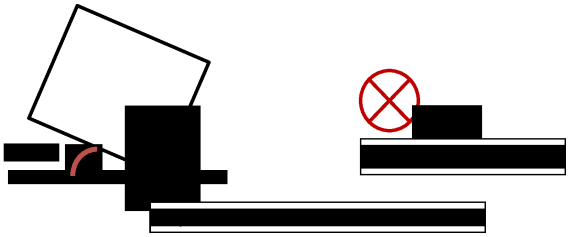


Figure 8.1: Schematic of **ANGLEP**.

8.2.6 ANGLEL

ANGLEL ***F*** sets the loop angle between the plane of the crystal loop and the goniometer axis. The angle is to be given in degrees, but without the degree symbol (°). The default *L* ('loop') angle is 0°.

The rotation angle to be applied to the angle of the crystal in the loop (right handed rotation about *X* axis applied to all voxels, as shown in figure 8.2).

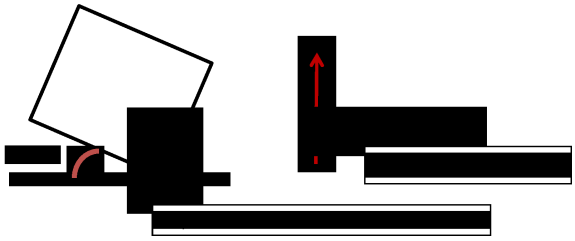


Figure 8.2: Schematic of **ANGLEL**.

8.2.7 ABSCOEFCALC

This keyword specifies whether the program should use average absorption and attenuation coefficients, or whether it should calculate them from input crystal parameters.

ABSCOEFCALC DUMMY will cause RADDOSE-3D to assume an absorption coefficient of 0.237 mm^{-1}

and an attenuation coefficient of 0.281 mm^{-1}. These values are representative of an average crystal at an incident energy of 12.4 keV. Crystal composition keywords will have no effect.

ABSCOEFCALC RD

ABSCOEFCALC RDV2 or

ABSCOEFCALC RDV3

These three commands are equivalent. RADDOSE-3D will call a previous version of RADDOSE to estimate absorption and attenuation coefficients.

The composition of the crystal has to be described using the keywords **UNITCELL**, **NUMMONOMERS**, **NUMRESIDUES**, **NUMRNA**, **NUMDNA**, **PROTEINHEAVYATOMS**, **SOLVENTHEAVYCONC** and **SOLVENTFRACTION**. The use of these keywords is described in the sections 8.2.8–8.2.15 below.

8.2.8 UNITCELL

This keyword only has an effect when the absorption and attenuation coefficients are estimated using a legacy version of RADDOSE (see section 8.2.7).

UNITCELL ***A B C***

UNITCELL ***A B C*** α β γ

Dimensions and angles of the unit cell a b c α β γ

The first three numbers specify the unit cell size in Angstroms. The second three numbers optionally specify the unit cell angles alpha, beta and gamma.

The (optional) angles are to be given in degrees, but without the degree symbol (°). If no angles are specified RADDOSE-3D assumes default angles of $90°$.

8.2.9 NUMMONOMERS

This keyword only has an effect when the absorption and attenuation coefficients are estimated using a legacy version of RADDOSE (see section 8.2.7).

NUMMONOMERS *I* specifies the number of monomers in the unit cell. Only integer numbers *I* should be used. This number should not be confused with the number of monomers in the asymmetric unit.

8.2.10 NUMRESIDUES

This keyword only has an effect when the absorption and attenuation coefficients are estimated using a legacy version of RADDOSE (see section 8.2.7).

NUMRESIDUES *I* specifies the number of amino acid residues per monomer. Only integer numbers *I* should be used. Using this keyword the number and types of atoms are calculated according to

the formula

$$\text{amino acid} = 5C + 1.35N + 1.5O + 8H$$

Sulfur atoms, e.g. from CYS and MET residues, should be added explicitly with the **PROTEINHEAVYATOMS** keyword.

The default value for *I* is 0.

8.2.11 NUMRNA

This keyword only has an effect when the absorption and attenuation coefficients are estimated using a legacy version of RADDOSE (see section 8.2.7).

NUMRNA ***I*** specifies the number of RNA nucleotides per monomer. Only integer numbers *I* should be used. Using this keyword the number and types of atoms are calculated assuming an average nucleotide content defined as

$$\text{mean nucleotide} = 11.25H + 9.5C + 3.75N + 7O + 1P$$

If a more accurate estimate is required, individual atoms may be entered explicitly with the **PROTEINHEAVYATOMS** keyword.

The default value for *I* is 0.

8.2.12 `NUMDNA`

This keyword only has an effect when the absorption and attenuation coefficients are estimated using a legacy version of RADDOSE (see section 8.2.7).

`NUMDNA` *I*

specifies the number of DNA deoxynucleotides per monomer. Only integer numbers *I* should be used. Using this keyword the number and types of atoms are calculated assuming an average deoxynucleotide content defined as

$$\text{mean nucleotide} = 11.75H + 9.75C + 4N + 6O + 1P$$

If a more accurate estimate is required, individual atoms may be entered explicitly with the `PROTEINHEAVYATOMS` keyword.

The default value for *I* is 0.

8.2.13 `PROTEINHEAVYATOMS`

This keyword only has an effect when the absorption and attenuation coefficients are estimated using a legacy version of RADDOSE (see section 8.2.7).

`PROTEINHEAVYATOMS` *El I (El I (El I ..))* defines a list of atoms to add to the protein part

of the absorption. Each species is defined by a two character string *El* for the elemental symbol, and an integer number *I* of atoms of that species per monomer.

The command **`PROTEINHEAVYATOMS S 10 Se 2`** would add 10 sulphur and 2 selenium atoms per monomer.

8.2.14 `SOLVENTHEAVYCONC`

This keyword only has an effect when the absorption and attenuation coefficients are estimated using a legacy version of RADDOSE (see section 8.2.7).

`SOLVENTHEAVYCONC` ***`El I (El I (El I ..))`*** defines the concentration of elements (not including water) in the solvent in millimoles per litre. Oxygen and lighter elements should not be specified.

The command **`SOLVENTHEAVYCONC Na 1000 Cl 1000`** specifies 1M sodium chloride in the solvent.

8.2.15 `SOLVENTFRACTION`

This keyword only has an effect when the absorption and attenuation coefficients are estimated using a legacy version of RADDOSE (see section 8.2.7).

`SOLVENTFRACTION` ***`F`***

The fraction of the unit cell that is occupied by solvent. If not given explicitly, this value is estimated from NRES, NRNA and NDNA using 1.35 g/ml for protein, and 2.0 g/ml for RNA and DNA.

8.3 `Beam` block

A `Beam` block must begin with the keyword **`BEAM`**. At least the **`TYPE`** must be specified. Depending on the chosen **`TYPE`**, further declarations may be required.

8.3.1 `TYPE`

With the keyword **`TYPE`**, the underlying beam implementation is chosen. Currently two distinct beam implementations exist:

`TYPE TOPHAT` defines a beam with uniform flux.

`TYPE GAUSSIAN` defines a beam with a 2-dimensional Gaussian flux profile. The full-width half-maximum must be specified with the **`FWHM`** keyword (see section 8.3.3).

8.3.2 FLUX

FLUX ***F*** specifies the total beam flux in photons per second. The flux parameter *F* can be specified in scientific notation (e.g. **1.3e12**).

8.3.3 FWHM

FWHM ***X Y***

The FWHM of the beam (vertical), (horizontal). Not needed if a Top-Hat beam is used. This defines the *X* and *Y* FWHM of the beam respectively in the RADDOSE coordinate system.

8.3.4 ENERGY

ENERGY ***F***

ENERGY ***F*** **KEV**

specifies the incident photon energy in keV. The optional keyword **KEV** can be appended for human readability of the input file.

8.3.5 COLLIMATION

COLLIMATION RECTANGULAR ***X Y***

specifies the horizontal and vertical collimation of the beam. Delimits where the beam has non-zero intensity. This is defined by the slits. For an uncollimated Gaussian beam, set to $\approx 3\times$ FWHM.

8.4 `Wedge` block

A `Wedge` block must begin with the keyword **`WEDGE`**.

`WEDGE` ***A B***

A and *B* define the start and end angle of the rotation in degrees (°), assuming the front face of the crystal (*X*-*Y* plane) is normal to the beam. Rotation is right handed about the *Y* axis.

8.4.1 `EXPOSURETIME`

`EXPOSURETIME` ***F***

specifies the total exposure time for this wedge in seconds.

8.4.2 `ANGULARRESOLUTION`

`ANGULARRESOLUTION` ***F***

specifies the angular step size used for wedge iterations in degrees (°). Defaults to 2°.

8.4.3 `STARTOFFSET`

`STARTOFFSET` ***X Y Z***

offset translation in µm applied to the crystal relative to the origin (defined as the intersection of the beam and the aligned goniometer axis) for the starting position of the wedge. Defaults to 0 0 0.

8.4.4 TRANSLATEPERDEGREE

`TRANSLATEPERDEGREE` *X Y Z*

translation of the goniometer during exposure in µm/° for helical scanning, leading to improvements in dose distribution. Defaults to 0 0 0.

8.4.5 ROTAXBEAMOFFSET

`ROTAXBEAMOFFSET` *F*

the offset in µm along *X* (vertical in most set-ups) between the beam axis and the rotation axis. Used to create 'offset' scanning for improvements in dose distribution. Defaults to 0 µm.

Appendix 2: publications derived from the work presented in this thesis.

Journal of
Synchrotron Radiation

ISSN 0909-0495

Received 27 July 2012
Accepted 29 October 2012

Optimizing the spatial distribution of dose in X-ray macromolecular crystallography

Oliver B. Zeldin, Markus Gerstel and Elspeth F. Garman*

Laboratory of Molecular Biophysics, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. E-mail: elspeth.garman@bioch.ox.ac.uk

X-ray data collection for macromolecular crystallography can lead to highly inhomogeneous distributions of dose within the crystal volume for cases when the crystal is larger than the beam or when the beam is non-uniform (Gaussian-like), particularly when crystal rotation is fully taken into account. Here the spatial distribution of dose is quantitatively modelled in order to compare the effectiveness of two dose-spreading data-collection protocols: helical scanning and translational collection. Their effectiveness in reducing the peak dose per unit diffraction is investigated *via* simulations for four common crystal shapes (cube, plate, long and short needles) and beams with a wide range of full width half maximum values. By inspection of the chosen metric, it is concluded that the optimum strategy is always to use as flat (top-hat) a beam as possible and to either match the beam size in both dimensions to the crystal, or to perform a helical scan with a beam which is narrow along the rotation axis and matched to the crystal size along the perpendicular axis. For crystal shapes where this is not possible, the reduction in peak dose per unit diffraction achieved through dose spreading is quantified and tabulated as a reference for experimenters.

© 2013 International Union of Crystallography
Printed in Singapore – all rights reserved

Keywords: radiation damage; dose maps; data-collection simulations; *RADDOSE-3D*.

1. Introduction

Radiation damage is widely recognized as one of the key bottlenecks in macromolecular structure solution by crystallography at both room temperature and 100 K (Garman, 2010; Holton, 2009). There are two main parameters that affect the extent of radiation damage in a crystal: the absorption sensitivity of the crystal to incident photons, and how many photons are incident on the crystal. These are considered together in terms of the dose, the amount of energy absorbed per unit mass of crystal (Gy = J kg^{-1}). The dose absorbed by a crystal fully immersed in a top-hat beam (uniform intensity) is given by

$$\text{Dose} = I_{\text{Incident}}(1 - \text{e}^{-\mu_{\text{abs}} z})/C_{\text{mass}}, \qquad (1)$$

where I_{Incident} is the total incident energy flux on the crystal, and μ_{abs}, z and C_{mass} are the absorption coefficient, the thickness and the total irradiated mass of the crystal, respectively. The sensitivity of the crystal to damage is pivotally affected by the absorption coefficient, itself a function of the crystal composition and beam energy. The physical origin of this and how it is calculated is well described in previous work (Murray *et al.*, 2005). Strategies to minimize the crystal sensitivity through the selection of weakly absorbing cryo-buffers and mother liquors are emerging (Murray *et al.*, 2004; Holton, 2009). Work on the effect of free-radical scavengers on the radiation sensitivity of crystals has shown some promise [see Allan *et al.* (2013) for a survey], but, since there are still conflicting results and many unresolved questions, the field remains very active. A significant factor in the conflicting results from radiation studies in general is the natural inhomogeneity of protein crystals grown in the same drop and given nominally the same treatment (*e.g.* Nowak *et al.*, 2009).

The software program *RADDOSE* (Murray *et al.*, 2004; Paithankar *et al.*, 2009; Paithankar & Garman, 2010) is widely used to calculate the absorbed dose for a macromolecular crystal. It requires the beam properties (size, profile in x and y, flux and energy), the unit-cell parameters, the number of amino acids in the protein, and the mother-liquor/cryobuffer constituents as inputs, and uses them to estimate the overall absorption coefficient of the crystal. This calculation takes into account all physical phenomena relevant in the incident X-ray energy ranges used in macromolecular crystallography (MX) (7–20 keV or 1.6–0.56 Å): the photoelectric effect in *RADDOSE* V1 (version 1) (Murray *et al.*, 2004), as well as fluorescence escape after photoelectric absorption in *RADDOSE* V2 (Paithankar *et al.*, 2009) and the Compton effect in *RADDOSE* V3 (Paithankar & Garman, 2010).

As described in previous work (Garman, 2010), for a typical 100 µm-thick macromolecular crystal uniformly irradiated by 12.4 keV photons (1 Å), only 2% of the incident photons will interact with the crystal (estimated from $1 - \text{e}^{-\mu_{\text{att}} z}$ with z = 100 µm, where μ_{att} is the attenuation coefficient of the beam in the crystal). Under these conditions, of the 2% of interacting

photons, 84% undergo photoelectric absorption, 8% are subject to the Compton effect and 8% are Rayleigh scattered (elastic, so interfere coherently to give diffraction). Some fraction (dependent on the atomic number of the element) of the photoelectrically absorbed photons will lead to fluorescence (as opposed to emission of an Auger electron), and there is a probability of these fluorescent photons escaping the crystal without further interactions, leading to a lower effective dose absorbed by the crystal. An estimation of the fraction of energy originating from the photoelectric effect escaping from the crystal was introduced in the update from *RADDOSE* V1 to V2. Fluorescence is only significant for heavier elements: *e.g.* for an iron atom there is a 34% probability of fluorescence production as opposed to emission of an Auger electron.

For the simple geometrical case described above (uniform X-ray exposure throughout the crystal volume), the *RADDOSE* treatment is accurate and offers robust guidance regarding the uniform dose (30 MGy) that a crystal can withstand before biological fidelity is compromised at 100 K (Owen *et al.*, 2006) if no electron scavengers are present (Allan *et al.*, 2013). Diffraction from the crystal may not be sustained until this 30 MGy experimental limit is reached, but it is unlikely to survive beyond it.

Considering the crystal as a single static unit has several drawbacks:

(i) For cases where the beam is smaller than the crystal, rotation will have a significant effect on the size of the irradiated region and on the profile of the dose distribution. This can only be dealt with through a three-dimensional treatment of the problem as presented here using a newly developed and developing code, *RADDOSE-3D*.

(ii) For Gaussian beam profiles, it is not possible to assign a single dose, since different regions will have different doses. Versions 1–3 of *RADDOSE* address this problem by giving a worst-case dose estimate based on the peak flux of the Gaussian beam profile.

(iii) It approximates the intensity drop-off of the beam within the crystal to be solely due to absorption (μ_{abs}), leaving out the small additional attenuation of the beam due to diffracted photons (μ_{att}).

Since the release of the original *RADDOSE* code in 2004, data collection using microbeams (between 1 and 30 µm in size) has grown dramatically in availability and popularity (*e.g.* Riekel *et al.*, 2005; Axford *et al.*, 2012). As a consequence, the experimental reality has drifted far from the uniform-dose approximation described above. Dose is a scalar field within a crystal: the idea of a single dose for the whole crystal has little meaning for cases where the dose can vary by orders of magnitude between differently exposed regions of the crystal. This paper first introduces the concept of dose distributions in MX using a typical data-collection strategy as an example. We then quantify the effect of two dose contrast mitigation strategies that are already in use at synchrotrons around the world, analysing these strategies with a view to optimizing the diffraction lifetime of macromolecular crystals.

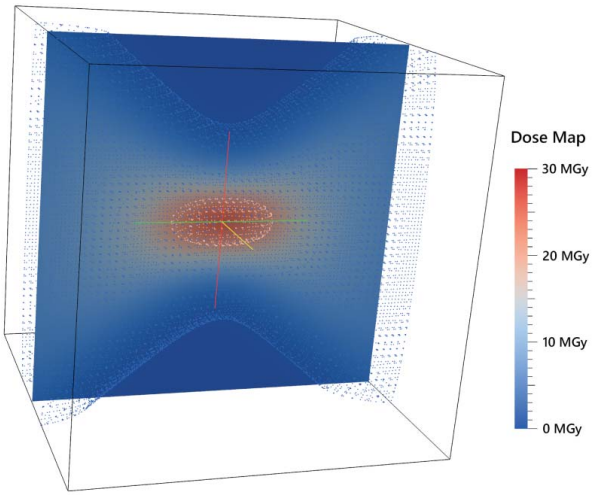


Figure 1
Dose map of the Standard data collection described in §1.1 (100 µm × 100 µm × 100 µm crystal, 20 µm × 20 µm FWHM beam, 90° wedge). The dark lines around the outside represent the edges of the crystal and the rotation axis is along the thin yellow line coming out of the plane of the page. The blue dotted isosurface represents the surface enclosing 95% of the absorbed dose (0.61 MGy), and the red dotted isosurface represents the Henderson limit of 20 MGy (Henderson, 1990). For this example, the region in the middle suffers a dose of 26.8 MGy, which is just below the 30 MGy experimental dose limit (Owen *et al.*, 2006). The colour bar for the *y*-axis slice is shown to the right.

1.1. A typical synchrotron MX experiment

Most MX experiments involve collecting data from one or more locations on a crystal, while rotating the crystal in a Gaussian (like) beam through a defined angular range. Experimentally determined beam profiles are often non-perfect Gaussians, and can sometimes be very far from such (Krojer & von Delft, 2011). Fig. 1 shows the colour-map dose profile for a 100 µm × 100 µm × 100 µm crystal at 2 µm^3 voxel resolution exposed over a 90° rotation range in a Gaussian-shaped beam with a full width at half-maximum (FWHM) in *x* and *y* of 20 µm × 20 µm. The beam has energy 12.4 keV (1 Å) and flux 5×10^{11} photons s^{-1}, and the total exposure time for the wedge is 60 s. The dose isomap is additionally presented as a rotating object movie in the supplementary video.[1] A dose distribution, such as the one shown in Fig. 1, displays the dose values evaluated at each voxel coordinate (number of voxels: 125000). An instructive way of inspecting these results is to divide the voxel population up into dose bins, as detailed in Table 1. This illustrates clearly how much of the crystal is weakly exposed, and the high levels of dose contrast across the crystal, with a peak dose of 25.6 MGy and an average dose for the whole crystal volume of 1.2 MGy.

1.2. Aggregate dose metrics

In order to facilitate comparison of various data-collection strategies, aggregate metrics that indicate the dose absorbed

[1] Supplementary data for this paper are available from the IUCr electronic archives (Reference: XH5033). Services for accessing these data are described at the back of the journal.

Table 1
Dose distribution by voxel population.

A very large fraction of the voxels are in the low-dose regime, but there are still voxels present at much higher doses.

Dose range (MGy)	Voxels in this range (%)
0–0.1	33.3
0.1–3.4	45.7
3.4–6.8	11.7
6.8–10.1	5.3
10.1–13.4	2.0
13.4–16.8	1.0
16.8–20.1	0.6
20.1–23.4	0.3
23.4–26.8	0.1

across the crystal would be more appropriate than the standard and often ambiguous 'average dose' currently in common usage. The set suggested here are:

Average Dose, Whole Crystal (AD-WC). This is the average of the dose values for the whole crystal volume, *i.e.* the total absorbed energy over the total mass of the crystal. It is a crude metric that does not reflect the fact that the vast majority of the data may have been collected from one highly damaged region. Changes in beam shape or collection strategy for the same exposure time, beam energy and crystal composition should not result in large effects on this value since the total absorbed energy will be approximately constant. It is only meaningful if the crystal is smaller than the beam, and is then a good metric if the beam has a top-hat profile.

Average Dose, Exposed Region (AD-3FWHM). This is the average dose calculated with the same absorbed energy as for the AD-WC, but with the mass in the denominator of the dose equation calculated from the volume swept out by the beam. For the Gaussian beams described here, this is defined as the volume swept out by a rectangular mask of dimensions 3 × FWHM of the beam profile in x and y (which for a true Gaussian will include >99% of the flux and thus >99% of the absorbed energy).

95% Threshold Average Dose (TAD-95). This is the average dose for the region defined by the isodose surface which encloses 95% of the absorbed energy. This has the advantage of leaving out the very weakly exposed regions arising from the tails of the Gaussian beam profile, but the absolute value is clearly sensitive to the percentage chosen for the absorbed energy threshold.

Comparing these three average dose metrics, it is noted that there is a constant ratio between the numerators of the equations defining them, and all have similar values (1:0.99:0.95). The main difference between them is in the denominator of the dose equation (the mass term). For AD-WC, this is only a function of crystal size and density; for AD-3FWHM it is also a function of the number of FWHMs of the beam profile used for the threshold, and of any rotation or translation of the crystal during exposure; for TAD-95, it is a function of threshold chosen, the crystal size and density, the beam FWHM, flux and energy, and the rotation and translations of the crystal during the data collection.

Table 2
Aggregate results for the simulation described in Table 1.

The definitions of the parameters are given in §1.2. Note the very large spread over the three dose metrics: 1:4.1:20.7 [Average (Whole Crystal):95% Threshold Average:Maximum Dose].

Average Dose (Whole Crystal)	1.5 MGy
95% Threshold (1.11 MGy) Average Dose	6.0 MGy
Maximum Dose	30.3 MGy
Dose Contrast (Maximum/Threshold Average)	5.05
Total absorbed energy	1.86 mJ
Dose Inefficiency (Maximum Dose/Absorbed Energy)	1.6×10^4 kg^{-1}

Maximum Dose. This is the highest value of the dose reached anywhere in the crystal (*i.e.* voxel with highest dose). It is a useful metric since very high maximum doses can have a large negative impact on data quality and can significantly increase crystal non-isomorphism (Ravelli & McSweeney, 2000; Ravelli *et al.*, 2002; Murray & Garman, 2002).

Dose Contrast. This is the ratio of the Maximum Dose to TAD-95 and is a useful quick metric which indicates the range of the dose across the region from which most of the diffraction has originated. Thus a lower rather than a high number implies that most of the data have come from a more homogeneously damaged region.

Dose Inefficiency. Minimizing the maximum dose is not a sufficient criterion by which to judge the effectiveness of a given data-collection strategy, since lowering the dose can trivially be achieved by reducing the exposure. A better metric to consider is the maximum dose per unit diffraction. Since, within the narrow energy range used in MX, the cross section for diffraction and the cross section for absorption events (and thus the absorbed energy) maintain a proportional relationship, the absorbed energy can be used as a proxy for the diffraction yield of the crystal. A 'Dose Inefficiency' metric can thus be defined as the maximum dose over the total absorbed energy (*i.e.* over the total diffracted intensity). The SI unit for this metric is kg^{-1}, which reflects the fact that scaling a given crystal-strategy system up (larger crystal volume, higher total flux, unchanged flux density) will increase the amount of diffraction for the same maximum dose. This is intuitive since a bigger crystal illuminated with a bigger beam would be expected to diffract more than if both were smaller. A lower Dose Inefficiency implies that for each mJ absorbed the peak dose in the crystal is lower, and so the strategy is deemed better. Therefore halving the Dose Inefficiency means that the maximum dose is halved for a given amount of absorbed energy (and thus to first order, also for the same amount of diffraction).

For the typical experimental case described in §1.1, the results for these six metrics are presented in Table 2. The range of values obtained clearly illustrates how no single dose metric can describe the full picture of the effects induced by the dose distribution in the crystal. Different dose regions have different levels of unit-cell expansion (Ravelli & McSweeney, 2000; Murray & Garman, 2002; Ravelli *et al.*, 2002), and when these regions are adjacent in a crystal (as in the case of a Gaussian beam) this will lead to strain along the crystal axes, which would in turn be expected to affect

diffraction quality. It can thus be justifiably assumed that relatively lower Dose Contrast and Dose Inefficiency are indicators of better data-collection strategies. To achieve homogeneous dose distributions and thus minimize stresses building up within the crystal, the ideal beam from a radiation damage minimization point of view is a top-hat beam.

Having highlighted the inadequate nature of a single average or peak dose for describing the absorbed energy profile for data collection performed with a Gaussian-shaped beam, the focus of what follows will be on using the above metrics to compare a number of methods for improving data quality by spreading the dose as evenly as possible throughout a crystal.

2. Methods

The simulation results presented in this paper were generated using Java code developed as part of an on-going major upgrade of the *RADDOSE* software code to give a new version: *RADDOSE-3D* (manuscript in preparation). This code effects a full voxel-based three-dimensional simulation of the dose profile of an MX experiment, and reports both three-dimensional dose fields and summary statistics. This is achieved by first creating a virtual representation of the crystal consisting of evenly spaced voxels at the specified resolution then and for each wedge: (i) determining, by rotation and translation operations, which of the crystal voxels intersect with the X-ray beam, (ii) for each affected voxel, calculating the specific X-ray flux at this point, as determined by the beam flux profile and the attenuation of the beam as it travels through parts of the crystal to reach the voxel, and (iii) calculating the absorbed energy at each voxel using this specific X-ray flux. The code finally generates summary statistics from the voxel map. For all the cases considered here, attenuation and absorption coefficients of $\mu_{\rm att}$ = 2.81 mm^{-1} and $\mu_{\rm abs}$ = 2.37 mm^{-1}, respectively, and a density of ρ = 1.2 g cm^{-3} were used. The μ values are based on an average empirical formulae taken from a survey of the Protein Data Bank (Berman *et al.*, 2002) as used in a previous paper by Holton & Frankel (2010) and are for an incident X-ray energy of 12.4 keV. The density value is constituted from a combination of 50% each of a typical protein density (1.4 g cm^{-3}) (Fischer *et al.*, 2004) and of solvent [low-density amorphous ice, 0.94 g cm^{-3} (Mishima *et al.*, 1985)]. All simulations were performed with 2 µm × 2 µm × 2 µm voxels and at an angular resolution of 1° per iteration, using an uncollimated Gaussian beam with various FWHM values. The details of the code and the method by which the calculations have been expedited will be presented elsewhere. A Visual Basic script was used to generate the large number of input data required to sample the parameter space, and the comparative plots were prepared using *R* (http://www.r-project.org/). The dose profiles were visualized using the *Paraview* software package (http://www.paraview.org/).

A total of 84 simulations were performed, using seven Gaussian beam FWHM sizes, four typical crystal shapes, and three established data-collection strategies. The crystal shapes were chosen to represent some of the more common crystal morphologies encountered by experimenters. Their dimensions were defined to give a constant volume across all shapes (Cube, 100 µm × 100 µm × 100 µm; Plate, 200 µm × 200 µm × 25 µm; Short Needle, 70 µm × 200 µm × 70 µm; Long Needle, 35.4 µm × 800 µm × 35.4 µm). Seven beams having equal total flux but different FWHM values in x and y (FWHM: 3 µm × 3 µm, 5 µm × 5 µm, 10 µm × 10 µm, 20 µm × 20 µm, 50 µm × 50 µm, 75 µm × 75 µm, 100 µm × 100 µm) were used in order to compare data-collection strategies for different crystal shapes. The three strategies compared were as follows:

Standard. 90° wedge collected symmetrically across one edge of the crystal with the beam centred for all dimensions of the crystal.

Helical. As for Standard, but translating the crystal along the goniometer axis during data collection. The translation was defined so that at the start and end positions the beam centre was half of its FWHM from the crystal edge. For example, the centre of a 10 µm × 10 µm beam on a Plate (200 µm × 200 µm × 25 µm) was translated from −95 µm to +95 µm (centre of crystal side being defined as 0 µm) along the goniometer axis, as shown in Fig. 2.

Translate. As for Standard, except that the total exposure was broken up into several individual wedges at different positions on the crystal along the goniometer axis. The first wedge was collected half a beam profile FWHM from the edge of the crystal, then the crystal was translated by one beam profile FWHM for each subsequent wedge until the beam centre was within half a FWHM of the other edge of the crystal, as shown in Fig. 3 for a Short Needle and a 20 µm × 20 µm beam for the simulations described above. A summary of the number of wedges for each beam size and crystal shape is included in the supplementary information as Table S1.

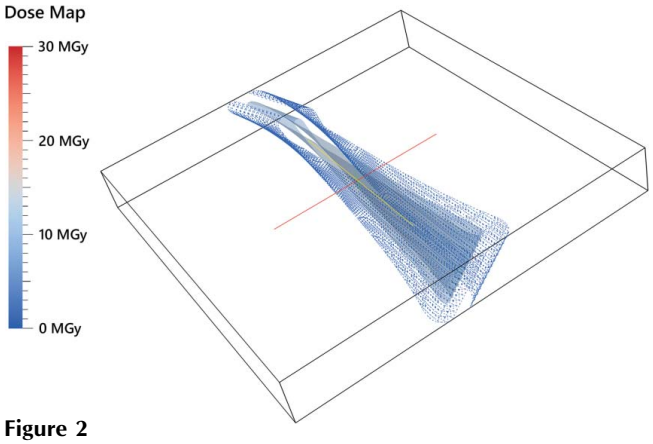


Figure 2
Dose map for a Helical collection for a Plate crystal with a 10 µm × 10 µm FWHM beam. The outer isosurface (points) is a 0.71 MGy surface that encloses 95% of the absorbed energy. The inner surface is a 5 MGy surface for visual clarity.

3. Results

Before analysing the results obtained for different data-collection strategies with each crystal type, the metrics must be

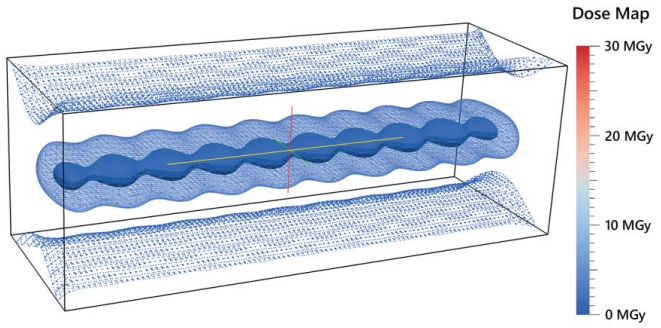


Figure 3
Dose map of a Translational collection strategy for a Short Needle crystal. The outer isosurface (points) is a 0.11 MGy surface that encloses 95% of the absorbed energy. The intermediate and inner surfaces are presented for visual clarity at iso-values of 1.8 MGy and 2.5 MGy, respectively.

compared in order to establish their relative merits, utility and relevance in evaluating the experimental protocols.

Figs. 4 and 5 show all the metrics plotted against the FWHM area of the beam profile for a Cube-shaped crystal subjected to a Standard strategy, and for a Long Needle with a Helical strategy, respectively. The beam FWHM area was chosen as an independent variable since, all other variables remaining constant, increasing the beam FWHM will have the effect of spreading the dose more evenly through the crystal and because, for a Gaussian beam, the peak intensity at constant flux decreases with increasing $(\text{FWHM})^2$. Looking first at the two most intuitive metrics, AD-WC and Maximum Dose, the behaviour is as expected. Since the AD-WC is only a function of absorbed energy, it is constant until the beam profile FWHM starts to be comparable with the crystal size, and then it begins to drop for both cases as more of the photons in the tails of the distribution are lost since they miss the crystal. For both of these strategies, the maximum dose is a proportional function of the peak intensity of the beam; it would thus be expected to fall as the beam peak intensity falls. As mentioned above, peak intensity goes as $(\text{FWHM})^{-2}$ for constant flux, and so on a log–log plot this follows a straight line with negative gradient.

Considering next the two other dose metrics for a Cube-shaped crystal, TAD-95 and AD-3FWHM, similar behaviour is observed for both the Standard–Cube and Helical–Long-Needle strategy–crystal-type conditions. They both track downwards as the mass in the dose equation increases with beam size, and then converge as the limit of including the whole crystal is approached. For the Standard–Cube case, the convergence occurs at around the 50 µm × 50 µm FWHM beam and, for the Helical–Long-Needle case, at around the 20 µm × 20 µm FWHM beam. This discrepancy is expected since the Long Needle has dimensions of 35.4 µm × 35.4 µm in the plane normal to the goniometer axis, and the Cube has dimensions of 100 µm × 100 µm in this plane. Thus the whole crystal will be irradiated by a 3 × FWHM beam envelope for a smaller value of FWHM for the Long Needle, since the translation is performed along the whole length of the crystal during a Helical scan, making the length along the rotation axis irrespective to when full illumination is achieved.

Having considered all the dose metrics, TAD-95 was chosen as the preferred metric for use in evaluating strategies, since it has the advantage of being an average, and thus representative of more than just the peak dose. It also defines the mass in the dose equation intelligently, using an isosurface that encloses 95% of the total absorbed energy.

As a way of comparing the Maximum Dose 'worst-case' scenario returned by *RADDOSE* versions 1, 2 and 3 with the more representative TAD-95, Fig. 6 shows these metrics plotted against each other for a Cube-shaped crystal. This is

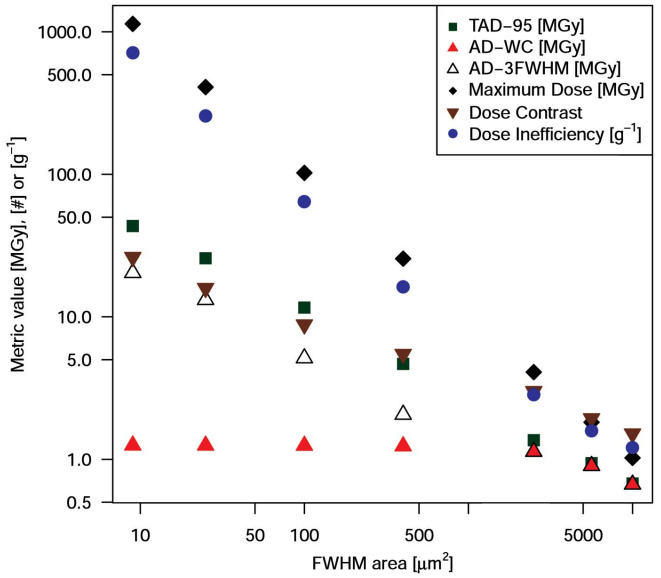


Figure 4
Metric comparison for a Cube crystal with a Standard data collection. The *y* axis is shared by three units as indicated in the key. Note the log–log scale.

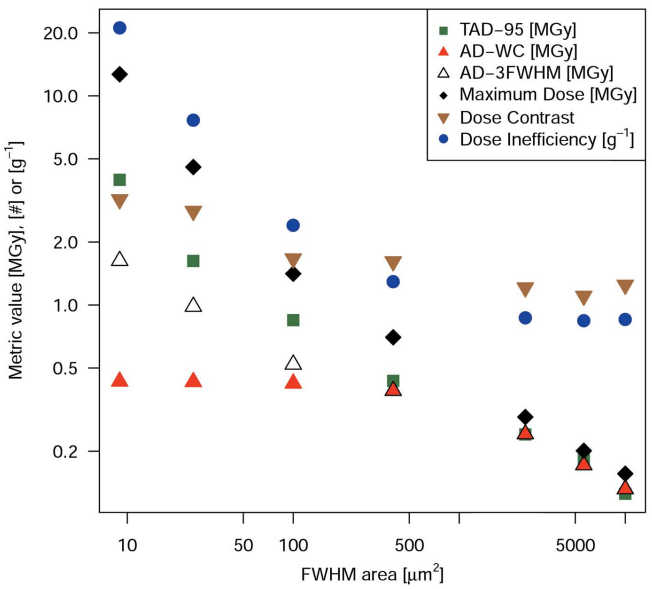


Figure 5
Metric comparison for a Long Needle with a Helical strategy. The *y* axis is shared by three units as indicated in the key. Note the log–log scale.

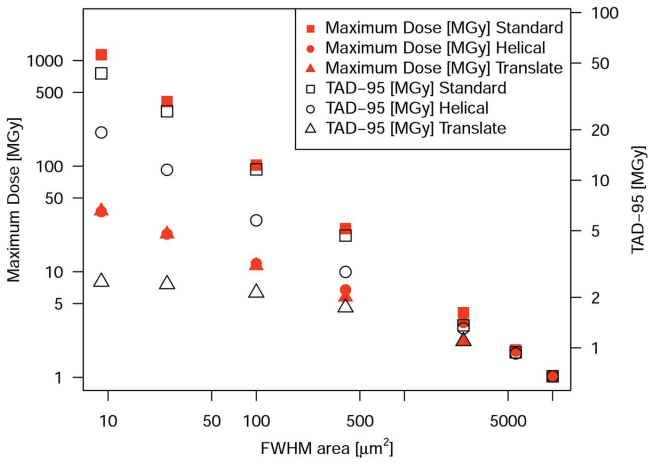


Figure 6
Comparison of Maximum Dose and TAD-95 for a Cube crystal. The Maximum Dose appears not to differentiate between Helical and Translational strategies, although the TAD-95 is clearly lower for the Translational method. Note the log–log scale.

particularly instructive since Maximum Dose suggests very similar performance for Helical and Translational strategies, but, looking a little further, it is clear that a Translational strategy results in a lower dose than the Helical one in terms of TAD-95 for beams smaller than 75 μm × 75 μm FWHM. Similar trends are observed for the Plate, Short Needle and Long Needle, and are presented in the supplementary material (Figs. S1–S3).

Whereas the dose metrics discussed above are informative concerning the average dose in different segmentations of the crystal, Dose Contrast and Dose Inefficiency offer a way of examining the dose distribution which is independent of the total exposure suffered by the crystal. Thus, doubling the exposure time in a collection protocol for a crystal would double the values of the dose metrics, but leave the Dose Contrast and Dose Inefficiency unchanged.

For both crystal-strategy pairs shown in Figs. 4 and 5, the Dose Inefficiency (Maximum Dose/total absorbed energy) drops and then begins to level off for larger beams. This is caused by the absorbed-energy term being constant, and then dropping as a larger proportion of the beam is lost at the sides of the crystal (as seen in the AD-WC points). Consequentially, the tailing off is more pronounced for the Long Needle crystal which has smaller dimensions in the plane perpendicular to the goniometer axis. Dose Contrast also uses Maximum Dose as the numerator, but has TAD-95 as its denominator. Its behaviour for large beams is thus less affected by losing beam off the side of the crystals.

Although Dose Contrast offers an intuitive way of examining the spread of dose in a crystal, Dose Inefficiency is a more informative and complete metric, since it reports on the highest dose a crystal will experience for a given amount of diffraction: an immediately experimentally useful number.

Plots of the TAD-95 and Dose Inefficiency against beam FWHM for each of the crystal shapes (Cube, Plate, Short and Long Needles, respectively) are shown in Fig. 7. Looking first at the Cube, it can be noted that TAD-95 for the Translational strategy outperforms the Helical strategy, and both Translational and Helical outperform the Standard strategy. The conclusions relating to the Helical *versus* Standard data collection presented here are in agreement with experimental results to be found in the literature (Flot *et al.*, 2010). The authors concluded that a helical data collection leads to a flatter profile in the inter-frame *B*-factors across the φ rotation range used (270°) relative to a Standard-type data collection. Their results were obtained for a 1.7 Å resolution dataset collected from a 10 μm × 300 μm × 10 μm crystal of the feruloyl esterase module of xylanase 10B from *Clostridium thermocellum*, using a beam with a FWHM of approximately 7.5 μm × 5 μm.

The Dose Inefficiency shows very similar performance for both the Translational and Helical dose-spreading strategies. This is because Dose Inefficiency reports on how well the Maximum Dose is minimized per unit diffraction, whereas TAD-95 reports on average dose. Thus both strategies are similarly effective in reducing the maximum dose per unit diffraction, but a translational scan spreads the bulk of the dose over a larger region, leading to a lower TAD-95. A similar effect, but with a smaller difference between Translational and Helical strategies, is observed for the Plate crystal.

For the Long and Short Needles, the same trend is observed in the TAD-95s, but the Dose Inefficiencies additionally begin to diverge, showing lower values for the Translational strategy. This is a small effect for the Short Needle, with a Helical: Translate ratio of 1.3 for a 3 μm × 3 μm FWHM beam. For the Long Needle the effect becomes much more pronounced with a ratio of 2.8 under the same conditions. This spread in the Dose Inefficiency can be understood as being due to the larger length of the crystal along the goniometer axis compared with the beam size. This results in the effect of the extra low-dose regions in a Translational strategy over a Helical strategy being compounded, leading to a lower maximum dose at a given total exposure for the Translational strategy.

It is instructive to note that for all four crystal shapes the Dose Contrast shows the opposite trend: a lower dose contrast for a Helical scan (Cube shown in Fig. 8, and the other three crystal shapes shown as Figs. S4–S6 in the supplementary information). This is because a translation of the crystal divides the peak dose by the number of possible steps which will avoid excessive overlap of the beam positions: the more steps, the greater the reduction. In terms of peak dose, a Helical strategy is the $N \rightarrow \infty$ limit of this. This serves as a good illustration of why Dose Inefficiency is a superior metric to Dose Contrast: the 'better' dose contrast for a helical scan is simply caused by the higher maximum dose at N points in the crystal and does not reflect the improvement found in other regions through using a Translational strategy.

4. Discussion

The central question raised by these results is how to interpret them in terms of day-to-day experimental practice. Current guidelines exist in the form of the 20 MGy dose limit predicted for MX by Henderson (1990) and a 30 MGy experimental

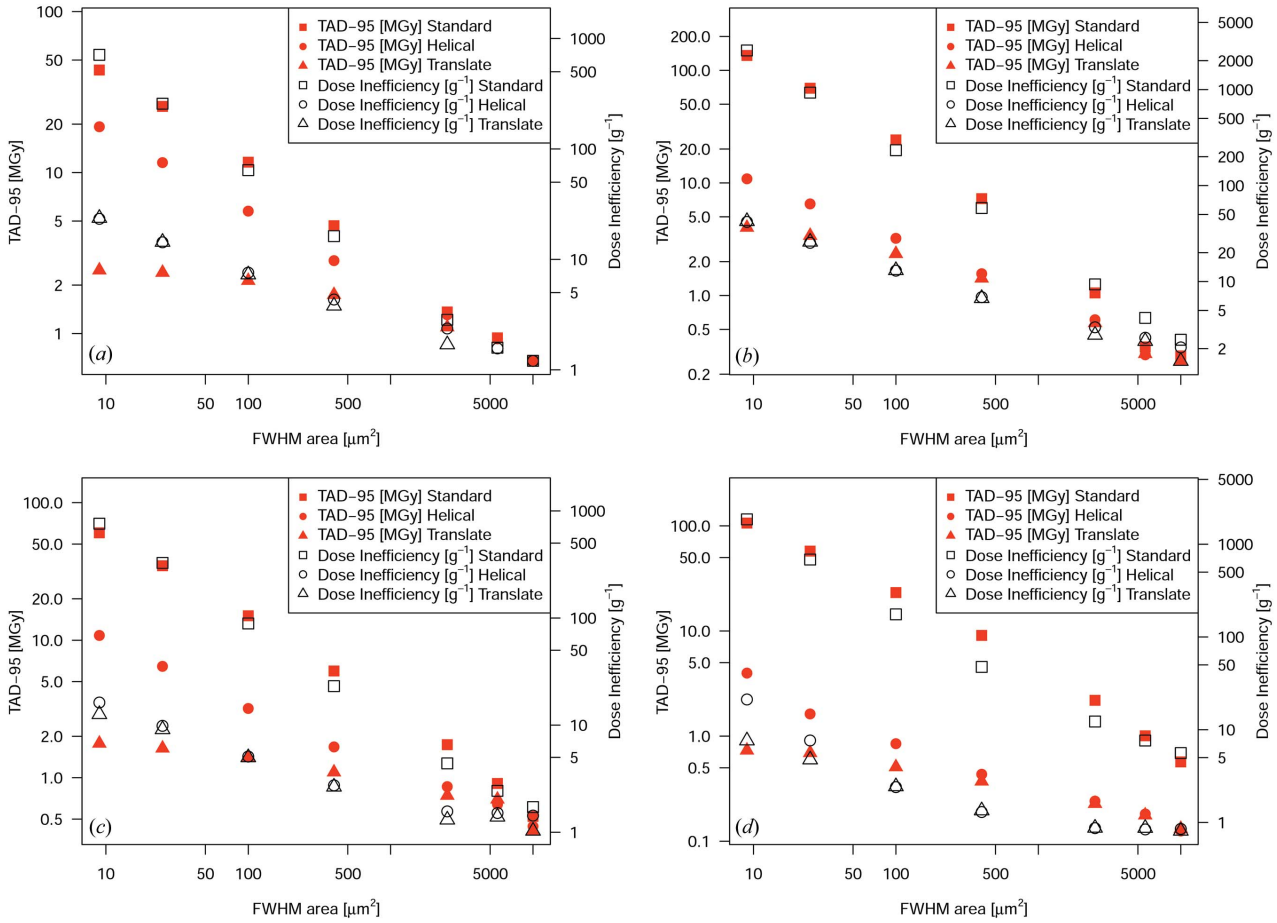


Figure 7
(*a*) TAD-95/Dose Inefficiency plots for the four crystal shapes: (*a*) Cube, (*b*) Plate, (*c*) Short Needle and (*d*) Long Needle. In all four cases the Helical and Translational methods converge for both metrics at larger beam sizes. The crystals (*c*) and (*d*) have a longer dimension along the goniometer axis and exhibit a better Dose Inefficiency for smaller beams. Note log–log axes.

dose limit which provides a recommendation regarding the dose beyond which a crystal is likely to give data of questionable biological fidelity (Owen *et al.*, 2006). This limit was measured under experimental conditions carefully optimized to be as close to the even-dose case as possible. Indeed, researchers working to understand radiation damage in MX make significant efforts to ensure that these conditions are well emulated in their experiments. For instance, this is achieved through experimental protocols such as matching the crystal size to the beam and ensuring a top-hat beam profile for data collection (Murray & Garman, 2002; Owen *et al.*, 2006), or by limiting the oscillation of the crystals to small wedges (Kmetko *et al.*, 2006), so that the irradiated region does not suffer from an uneven dose distribution such as the one shown in Fig. 1.

In order to interpret a dose map in the context of existing guidelines, the physics behind diffraction pattern formation must be considered. Each elastically scattered photon will have originated from a region of the crystal in some state of damage ranging from a zero-dose state to a highly damaged state which will have absorbed a dose well above 30 MGy. It can be hypothesized that the final diffraction pattern is a linear superposition of the individual patterns that would originate

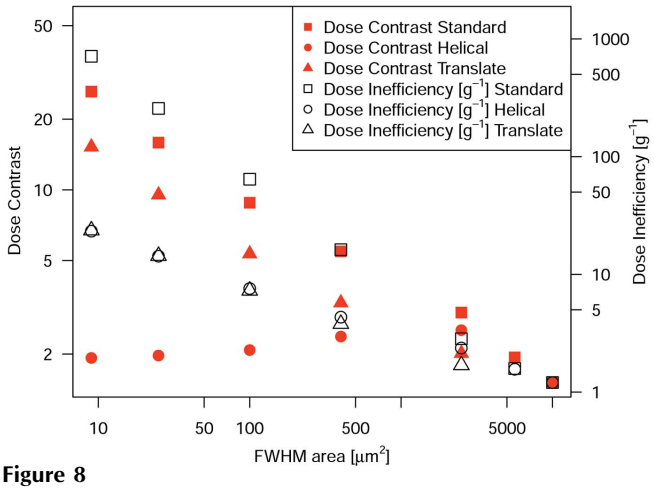


Figure 8
Comparison of Dose Contrast and Dose Inefficiency metrics. Contrast is lower for Helical than for Translational strategies, despite similar Dose Inefficiencies, illustrating the shortcomings of using Dose Contrast as the main metric. Note log–log axes.

from each of these damage states. The experimental dose limit indicates that a region of crystal that has suffered a dose above 30 MGy may have compromised biological fidelity. Experi-

mentalists should therefore aim to minimize or, ideally, eliminate any regions at or above that limit while simultaneously extracting the maximum amount of diffraction from a crystal. The Dose Inefficiency metric is a way of describing how well a proposed strategy achieves this goal. TAD-95, on the other hand, is a metric which indicates the average final dose state of the crystal. Strategies can be optimized to minimize Dose Inefficiency and experiments can be monitored *via* TAD-95. These then give useful practical dose metrics, since the dose field from each simulation result is interpretable as a single number which can be compared across different experimental scenarios.

The most pronounced result that can be surmised from Fig. 7 is that the largest beam size (100 μm × 100 μm FWHM) is the most effective at minimizing Dose Inefficiency for all the cases considered, and also produces a lower TAD-95 for the same total flux. Across the board, matching the beam size to the crystal is the most efficient way of exposing a crystal evenly. Given a crystal shape and beam size (or relative beam size in the case of crystals that are not of similar size to the ones used as examples here), Fig. 7 can be used to evaluate the comparative advantages of a Helical or a Translational strategy over a Standard data collection. Fig. 7 can thus be used as a guide to evaluate whether the advantage is worth the extra complexity of these data-collection protocols.

An additional strategy that has been reported is to use a beam size that matches the crystal in the vertical axis and to then perform a helical scan. An additional simulation was carried out for a 100 μm × 10 μm beam using a Helical strategy on the Cube crystal. When compared with a 100 μm × 100 μm beam, the Dose Inefficiencies were almost identical: 1.2 g^{-1} for the Standard 100 μm × 100 μm beam and 1.1 g^{-1} for the Helical 100 μm × 10 μm beam. Since the crystals in the two simulations absorbed different amounts of energy (and hence diffracted different numbers of photons), for comparison the TAD-95 was also normalized. This again gave very similar results: 0.80 g^{-1} for the Standard 100 μm × 100 μm beam and 0.84 g^{-1} for the Helical 100 μm × 10 μm beam. These simulations show that dose distributions can be optimized by matching only the vertical axis of the beam to the crystal size and employing a Helical strategy.

Clearly, for a full simulation of real-life MX data collections, the photon flux density and beam shape are required. Importantly, it will also be necessary to have accurate information on the crystal topology, since crystals do not necessarily grow in the shapes modelled above. The ability to parameterize crystal shapes *in situ* on the beamline is an active area of research (see, for instance, Brockhauser *et al.* 2008; Khan *et al.*, 2012). A further issue still to be addressed is *RADDOSE* versions and also *RADDOSE-3D* do not take into account the possibility of the escape of the primary photoelectrons from the crystal surface. Their range for 15.1 and 18.5 keV incident X-rays has been found to be around 4 μm (Sanishvili *et al.*, 2011) experimentally. Simulations show that this phenomenon can reduce the dose suffered by microcrystals, especially if using beams with energies above 25 keV (Nave & Hill, 2005; Cowan & Nave, 2008). For *RADDOSE-3D* to have better applicability for microcrystals and nanocrystals [such as those used in serial femtosecond MX crystallography (*e.g.* Chapman *et al.*, 2011)], the photoelectron escape is clearly an important addition to be made to the dose description.

5. Conclusions

The relative effectiveness of two dose-spreading strategies compared with a non-translational data-collection strategy has been evaluated here using two new metrics, the Threshold Average Dose and the Dose Inefficiency. Our simulations show an order-of-magnitude reduction in Dose Inefficiency for either a Helical or Translational strategy, and the results can be used by the experimenter to judge when either of these strategies would be worth employing on-line.

The profile of the beam has a pivotal effect on the dose distribution within an irradiated crystal, and top-hat-shaped beams will always result in more uniform dose distributions than Gaussian beams.

In terms of minimizing the maximum dose per unit diffraction, the most efficient way of collecting data with a Gaussian beam is to closely match the beam size to the crystal in order to obtain the most even dose profile possible while not excessively exposing the non-crystal regions of the loop (cryobuffer/mother-liquor). As already discussed, this effect can also be achieved by matching only the vertical (perpendicular to the rotation axis) size of the beam to the crystal dimensions and performing a Helical scan. This has an important implication for beamline design, since it only requires collimation in one dimension to optimize dose distribution. Relative to collimating in two dimensions, this would result in a linear reduction in the amount of beam area lost to the collimator for a given Cube crystal smaller than the vertical beam dimensions.

When this beam shape is not available, either a Helical or a Translational strategy provides an improvement in both the TAD-95 and the Dose Inefficiency. Judged by these metrics, a Translational strategy is more effective than a Helical one, and this is more pronounced the smaller the beam is relative to the crystal. Since these results are limited to the metrics discussed here, other effects, such as the implications for scaling data collected using a Helical or Translational strategy, could affect these conclusions in experimental reality, and must thus also be considered for a more complete treatment.

This paper has highlighted the large improvement in dose homogeneity that could be obtained by optimizing the data-collection protocols for a particular beam and crystal size. Clearly, there is a need for experimental investigations into the impact of strategies with different Dose Inefficiencies on final data quality and this will be a focus of future work.

We gratefully acknowledge the UK's Engineering and Physical Sciences Research Council (EPSRC) for funding, through studentships at the Life Science Interface (OZ) and the Systems Biology (MG) programmes at the University of Oxford's Doctoral Training Centre. We also thank James

Holton and Robert Fischetti for valuable and insightful conversations on the subject of dose metrics and distributions. Finally, we would like to thank the referees for their highly constructive comments which have benefited the final manuscript.

References

Allan, E. G., Kanderb, M. C., Carmichael, I. & Garman, E. F. (2013). *J. Synchrotron Rad.* **20**, 23–36.
Axford, D., Owen, R. L., Aishima, J., Foadi, J., Morgan, A. W., Robinson, J. I., Nettleship, J. E., Owens, R. J., Moraes, I., Fry, E. E., Grimes, J. M., Harlos, K., Kotecha, A., Ren, J., Sutton, G., Walter, T. S., Stuart, D. I. & Evans, G. (2012). *Acta Cryst.* D**68**, 592–600.
Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J. D. & Zardecki, C. (2002). *Acta Cryst.* D**58**, 899–907.
Brockhauser, S., Di Michiel, M., McGeehan, J. E., McCarthy, A. A. & Ravelli, R. B. G. (2008). *J. Appl. Cryst.* **41**, 1057–1066.
Chapman, H. N. *et al.* (2011). *Nature (London)*, **470**, 73–77.
Cowan, J. A. & Nave, C. (2008). *J. Synchrotron Rad.* **15**, 458–462.
Fischer, H., Polikarpov, I. & Craievich, A. F. (2004). *Protein Sci.* **13**, 2825–2828.
Flot, D., Mairs, T., Giraud, T., Guijarro, M., Lesourd, M., Rey, V., van Brussel, D., Morawe, C., Borel, C., Hignette, O., Chavanne, J., Nurizzo, D., McSweeney, S. & Mitchell, E. (2010). *J. Synchrotron Rad.* **17**, 107–118.
Garman, E. F. (2010). *Acta Cryst.* D**66**, 339–351.
Henderson, R. (1990). *Proc. R. Soc. B*, **241**, 6–8.
Holton, J. M. (2009). *J. Synchrotron Rad.* **16**, 133–142.
Holton, J. M. & Frankel, K. A. (2010). *Acta Cryst.* D**66**, 393–408.
Khan, I., Gillilan, R., Kriksunov, I., Williams, R., Zipfel, W. R. & Englich, U. (2012). *J. Appl. Cryst.* **45**, 936–943.
Kmetko, J., Husseini, N. S., Naides, M., Kalinin, Y. & Thorne, R. E. (2006). *Acta Cryst.* D**62**, 1030–1038.
Krojer, T. & von Delft, F. (2011). *J. Synchrotron Rad.* **18**, 387–397.
Mishima, O., Calvert, L. D. & Whalley, E. (1985). *Nature (London)*, **314**, 76–78.
Murray, J. & Garman, E. (2002). *J. Synchrotron Rad.* **9**, 347–354.
Murray, J. W., Garman, E. F. & Ravelli, R. B. G. (2004). *J. Appl. Cryst.* **37**, 513–522.
Murray, J. W., Rudiño-Piñera, E., Owen, R. L., Grininger, M., Ravelli, R. B. G. & Garman, E. F. (2005). *J. Synchrotron Rad.* **12**, 268–275.
Nave, C. & Hill, M. A. (2005). *J. Synchrotron Rad.* **12**, 299–303.
Nowak, E., Brzuszkiewicz, A., Dauter, M., Dauter, Z. & Rosenbaum, G. (2009). *Acta Cryst.* D**65**, 1004–1006.
Owen, R. L., Rudiño-Piñera, E. & Garman, E. F. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 4912–4917.
Paithankar, K. S. & Garman, E. F. (2010). *Acta Cryst.* D**66**, 381–388.
Paithankar, K. S., Owen, R. L. & Garman, E. F. (2009). *J. Synchrotron Rad.* **16**, 152–162.
Ravelli, R. B. G. & McSweeney, S. M. (2000). *Structure*, **8**, 315–328.
Ravelli, R. B. G., Theveneau, P., McSweeney, S. & Caffrey, M. (2002). *J. Synchrotron Rad.* **9**, 355–360.
Riekel, C., Burghammer, M. & Schertler, G. (2005). *Curr. Opin. Struct. Biol.* **15**, 556–562.
Sanishvili, R., Yoder, D. W., Pothineni, S. B., Rosenbaum, G., Xu, S., Vogt, S., Stepanov, S., Makarov, O. A., Corcoran, S., Benn, R., Nagarajan, V., Smith, J. L. & Fischetti, R. F. (2011). *Proc. Natl Acad. Sci.* **108**, 6127–6132.

CrossMark

electronic reprint

Journal of
Applied Crystallography
ISSN 0021-8898

RADDOSE-3D: time- and space-resolved modelling of dose in macromolecular crystallography

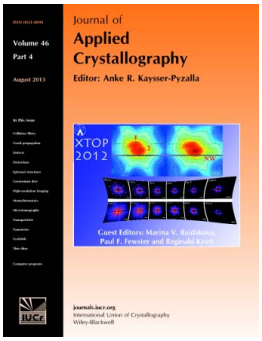
Oliver B. Zeldin, Markus Gerstel and Elspeth F. Garman

J. Appl. Cryst. (2013). **46**, 1225–1230

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site or institutional repository provided that this cover page is retained. Republication of this article or its storage in electronic databases other than as specified above is not permitted without prior permission in writing from the IUCr.

For further information see http://journals.iucr.org/services/authorrights.html



Journal of Applied Crystallography covers a wide range of crystallographic topics from the viewpoints of both techniques and theory. The journal presents papers on the application of crystallographic techniques and on the related apparatus and computer software. For many years, the *Journal of Applied Crystallography* has been the main vehicle for the publication of small-angle scattering papers and powder diffraction techniques. The journal is the primary place where crystallographic computer program information is published.

Crystallography Journals Online is available from **journals.iucr.org**

Journal of
Applied Crystallography
ISSN 0021-8898

Received 22 January 2013
Accepted 26 April 2013

RADDOSE-3D: time- and space-resolved modelling of dose in macromolecular crystallography

Oliver B. Zeldin, Markus Gerstel and Elspeth F. Garman*

Laboratory of Molecular Biophysics, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. Correspondence e-mail: elspeth.garman@bioch.ox.ac.uk

RADDOSE-3D allows the macroscopic modelling of an X-ray diffraction experiment for the purpose of better predicting radiation-damage progression. The distribution of dose within the crystal volume is calculated for a number of iterations in small angular steps across one or more data collection wedges, providing a time-resolved picture of the dose state of the crystal. The code is highly modular so that future contributions from the community can be easily integrated into it, in particular to incorporate online methods for determining the shape of macromolecular crystals and better protocols for imaging real experimental X-ray beam profiles.

© 2013 International Union of Crystallography
Printed in Singapore – all rights reserved

1. Introduction

Radiation damage is one of the most common causes for the failure of X-ray macromolecular crystallography (MX) experiments (Holton, 2009; Garman, 2010). This damage is caused by absorption of some of the incident X-ray photons by the crystal. The energy deposited in the crystal leads to chemical transformations which eventually lead to a decrease in the intensity of the diffraction pattern and ultimately compromise the quality of the final calculated electron density. In some cases, damage can be so severe that it is not possible to solve the structure and obtain reasonable electron density using the diffraction images collected while the damage progresses. A number of metrics have been proposed to quantify the progression of radiation damage to protein crystals, and since the earliest days of MX radiation-damage research, dose (D, energy deposited per unit mass) has been accepted as the appropriate calculated metric against which to evaluate these radiation-damage indicators (Blake & Phillips, 1962).

The program *RADDOSE* has evolved through three iterations: v1 (Murray *et al.*, 2004), v2 (Paithankar *et al.*, 2009) and v3 (Paithankar & Garman, 2010). It provides a one-dimensional model of absorbed dose in a non-rotating macromolecular crystal, and allows predictive estimates of crystal lifetime to be made through comparison of the results with an experimentally determined dose limit (Owen *et al.*, 2006). For cases where the crystal is smaller than the beam, and where the beam has a uniform profile, this provides a valuable piece of information and input for use by strategy optimization tools such as *BEST* (Bourenkov & Popov, 2010).

However, modern advances in synchrotron technology, in particular data collection using micro-beams between 1 and 30 µm in size (*e.g.* Riekel *et al.*, 2005; Axford *et al.*, 2012), have led to the dose distribution in crystals often being highly inhomogeneous (Zeldin *et al.*, 2013). This paper presents a new predictive software tool, *RADDOSE-3D*, which allows full three-dimensional modelling of dose distributions for MX.

1.1. Interactions of X-rays with macromolecular crystals: *RADDOSE* v1 to v3

At the energies relevant for standard MX experiments (5–20 keV), there are three physical processes that contribute to the attenuation of the X-ray beam within the crystal: elastic scattering, which gives rise to the diffraction pattern observed at the detector, the photoelectric effect, which dominates the absorption cross section, and Compton scattering, which has a low probability at these energies but becomes important at energies above 20 keV, rising to 67% of all the interaction events at 40 keV (Paithankar & Garman, 2010). Each atom type from which the crystal is constituted has an individual energy-dependent cross section for these phenomena. Heavier atoms generally have larger absorption coefficients and contribute disproportionately to the total energy absorbed from the X-ray beam by a crystal.

In order to calculate an overall X-ray absorption cross section for the crystal, the unit-cell composition must be known. *RADDOSE* calculates this from the unit-cell size, the number of monomers, the monomer primary sequence length (protein or DNA), the number of heavy atoms in the monomer (sulfur from cysteines and methionines, or selenium from seleno-methionine substitutions, and/or known bound metal ions), the millimolar concentration of heavy atoms in the solvent and, optionally, the solvent content. If no solvent content is provided, the program calculates this from steric and volume considerations and fills the remaining volume of the unit cell with water.

The first version of *RADDOSE* only took into account photoelectric absorption, which is by far the dominant effect at the relevant incident X-ray energies. Photoelectron ejection leaves the interacting atom in an excited state, from which it can decay *via* one of two paths: Auger electron emission or fluorescent photon emission. The probability of fluorescent escape (as compared to Auger emission) is small for the atoms that constitute typical proteins and thus was not considered in the first version of *RADDOSE*. However, for larger Z atoms, fluorescent escape can occur at a significant probability (*e.g.* 30% of K-shell de-excitations for an iron atom are by fluorescent decay and 70% by Auger emission). The second version of *RADDOSE* (Paithankar *et al.*, 2009) calculated both the fluorescent emission probability and the probability that these fluorescent photons might escape the crystal rather than be reabsorbed, making the simplifying assumption that all the fluorescent photons are produced in the centre of the crystal. The energy loss in the crystal due to the Compton effect (inelastic X-ray scattering) was included in version 3 of *RADDOSE* with the appropriate angular distribution and self-absorption treatment (Paithankar & Garman, 2010). This version was made available on request to groups working with high-energy incident beams ($E_{\text{X-ray}} > 20$ keV) for MX.

electronic reprint

RADDOSE v2 has become a widely used tool for estimating the absorbed dose in MX experiments, owing to its convenient use of biologically relevant parameters as inputs and its detailed consideration of the relevant physics. An alternative approach (Warkentin *et al.*, 2012), yielding results within 5% of those output by *RADDOSE*, is to compile the crystal composition manually and use mass energy-absorption coefficients from the NIST tables (http://www.nist.gov/pml/data/xray_gammaray.cfm).

To calculate dose, *RADDOSE* takes the total absorption coefficient (μ_{abs}) and calculates the energy absorbed by the crystal using a simple exponential absorption function of the crystal thickness, z:

$$\text{Absorbed energy} = E_{\text{Incident}}[1 - \exp(-\mu_{\text{abs}} z)]. \quad (1)$$

This treatment is correct for a crystal that is completely immersed in a beam with a flat top-hat profile. In the case of a Gaussian beam, *RADDOSE* uses the incident photon flux and beam FWHM x and y values to calculate the peak beam intensity, and then uses this flux to provide a maximum dose: a worst case scenario for damage. This dose is correct for the central part of the irradiated crystal which is subjected to the beam peak intensity for the entire rotation range. However, this approximation has obvious limitations in an era when many beams are smaller than the crystals being irradiated and when the beams can have highly featured profiles (Krojer & Von Delft, 2011). One solution to this problem is to use a sacrificial crystal to experimentally characterize radiation-damage sensitivity before determining an optimal data collection strategy. This approach is one of the options implemented in the program *BEST* (Bourenkov & Popov, 2010). Experimental characterization has the obvious disadvantage of requiring a spare crystal of the desired type to be available and can add a time burden since a data collection and burn series must be performed before the desired data collection can proceed.

1.2. Limitations of current predictive models of radiation damage

A simple one-dimensional single-body treatment of the crystal as a non-rotating object presents four major limitations for the optimal prediction of crystal lifetimes. (i) The photoelectron may escape from the crystal surface and the accompanying dose reduction is not taken into account. (ii) The range of doses present within the diffracting volume of a crystal, as a consequence either of rotation or of non-uniform beam profiles, is not taken into account. (iii) Realistic crystal shapes and/or beam profiles cannot be incorporated into the calculations. (iv) If the crystal is larger than the beam, new volumes of the crystal rotate into the beam during the experiment.

Previous Monte Carlo simulations and experimental work have shown that photoelectrons produced by the absorption of 18.5 keV X-rays in a protein crystal cause damage up to about 4 µm from the site of ionization (Nave & Hill, 2005; Sanishvili *et al.*, 2011). As a consequence, for very small crystals, a significant proportion of the initially absorbed energy may escape the crystal surface and thus reduce the effective dose. Nave & Hill (2005) calculated that for a 4 × 4 × 4 µm crystal, approximately seven-eighths of the energy would escape the crystal volume. Although this is an issue for platforms using micro- and nano-crystals such as the X-ray free-electron laser (X-FEL) (Chapman *et al.*, 2011), for crystals of the order of 100 µm in size or more, these effects are small. The probability of photoelectron escape is not included in the work reported here; however, the full three-dimensional treatment of a rotating crystal described below provides the framework necessary for such considerations to be realistically modelled in the future.

The consequence of not being able to model factors (ii)–(iv) above for systematic radiation-damage studies has been that crystals must always be chosen that are smaller than the beam and that experimental design must be carefully optimized to create an even dose profile (Kmetko *et al.*, 2006; Owen *et al.*, 2006). Thus, existing guidelines for conducting experiments that take into account the effect of dose on data quality are largely derived for even-dose cases. Examples of such guidelines for 100 K data collection include the experimental dose limit of 30 MGy, at which the diffraction intensity will have dropped to 70% or less of its initial value (Owen *et al.*, 2006), and the coefficient of sensitivity, which gives a value for the expected rate of increase in relative B value, B_{rel}, as radiation damage

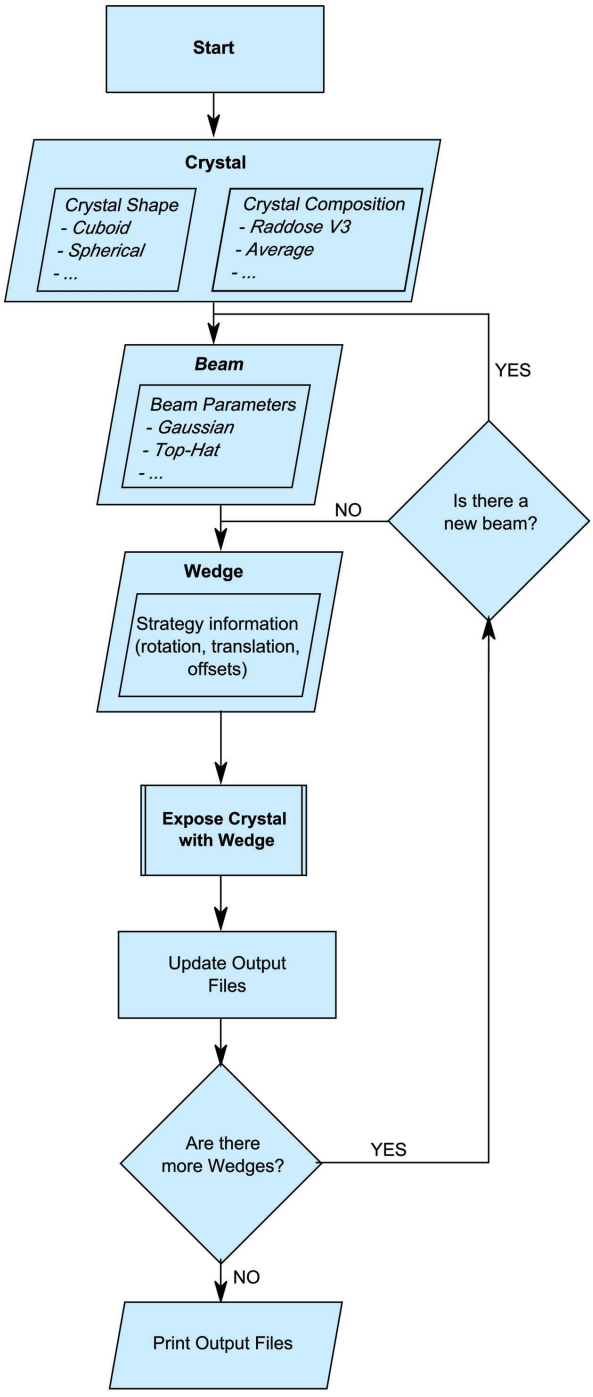


Figure 1
Flow chart illustrating the structure of the code.

electronic reprint

progresses: $S_{AD} = \Delta B_{rel}/8\pi^2\Delta D \simeq 0.014$ Å^2 MGy^{-1} (Kmetko *et al.*, 2006).

For the general application of dose-lifetime guidelines to the crystallographic determination of novel structures, this one-dimensional model is appropriate within the regime for which it was designed. However, it offers limited predictive power for cases where dose contrast is high. For instance, for the example of the use of the program presented in §5, the dose at the centre of the crystal is ∼11 times that of the average for the whole crystal. In cases such as this, a one-dimensional model is not appropriate, and there is a clear need for a more sophisticated three-dimensional treatment as described here.

Additionally, a minor improvement to the accuracy of the simulation has been made by distinguishing between the absorption (μ_{abs}) and the attenuation coefficient (μ_{att}) when calculating the dose at each depth point in the crystal. Previous versions of *RADDOSE* have treated μ_{abs} and μ_{att} as equal since the difference is very small for the conditions typically used in MX where 2% absorption of the beam is typical (Garman, 2010). Because *RADDOSE-3D* calculates a full three-dimensional dose field, the new program treats these two coefficients separately and thus corrects this approximation.

2. Logical structure of the code

The design principle behind *RADDOSE-3D* was to create an open-source package that not only provides a full three-dimensional simulation of MX experiments for the calculation of dose profiles but also creates a general framework that could be extended to treat other macroscopic phenomena in MX, such as absorption corrections for low-energy incident photons. In its most general sense, this package takes a voxel field representation of a crystal and a computational model of a beam, and then simulates the interaction of the two *via* a strategy here described by a wedge which specifies the angular range and translations associated with a single exposure within an MX experiment. The crystal has various properties associated with each voxel: position in three-dimensional space, dose and fluence (photons per unit area). For a given small volume element in the crystal, the dose is a function of the energy absorbed in that element. This is calculated for each step using the intensity at the leading edge of the volume[1] and the proportion of the beam absorbed by that volume element: parameterized by the absorption coefficient. The beam is described by the photon energy, an intensity profile and the total flux.

The logical flow of the code, as shown in Fig. 1, starts by using the parameters defined in the input file or through the online interface available at http://www.raddo.se to set up the three components (crystal, beam and wedge) described above. Input files describe a single crystal and one or more beams and wedges of contiguous data collection; an example is shown in Fig. 2. The full grammar for the input files is specified in detail in the supplementary material.[2] A single run of the program thus always acts on a single crystal but can simulate several wedges (*e.g.* in different places on the crystal), each wedge being associated with the most recently defined beam. For instance, different beams might describe changes in energy for multiple-wavelength anomalous diffraction data collection, or changes in flux due to a change in transmission or an electron storage ring refill. The code works its way through the simulation in the order described in the input file. Pseudo-code for the calculation performed during the exposure routine is shown in Fig. 3. The output files are

```
# Complicated input file for RADDOSE-3D
Crystal
Type Cuboid
Dimensions 50 300 50
AngleP 45
AngleL 0
PixelsPerMicron 0.48

CoefCalc RDV3
unitcell 78 78 78
NumResidues 51
nummonomers 24
ProteinHeavyAtoms S 6 Zn 2
SolventHeavyConc P 0.425

Beam
Type Gaussian
Flux 1e12
FWHM 20 20
Energy 12.4
Collimation Rectangular 60 60

Wedge 0 360
ExposureTime 60
AngularResolution 1
StartOffset -75 -75

Wedge 0 360
ExposureTime 60
StartOffset -45 -45
TranslatePerDegree 0.125 0.125

Beam
Type Gaussian
Flux 4e11
FWHM 15 15
Energy 12.4
Collimation Rectangular 15 15

Wedge 0 360
ExposureTime 60
StartOffset 50 50
RotAxBeamOffset 0

Wedge 0 360
ExposureTime 60
StartOffset 75 75
RotAxBeamOffset 20
```

Figure 2
Structure of a *RADDOSE-3D* input file. The crystal is first described by a 'Crystal' block (salmon coloured in the electronic version of the journal), which remains active throughout. Next, a beam (the green block), followed by one or more wedges (blue), is described. As illustrated in Fig. 1, each wedge defines how to simulate an exposure of the previously specified beam with the crystal. Here the first two wedges will act with the first beam and the last two with the second beam. The full input syntax is described in the supplementary material.

```
for each angle within range and resolution set in Wedge
  for each voxel in Crystal
    let coords be voxel coordinates of voxel
    Rotate coords by angle
    Translate coords by angle*(transX, transY)
    Calculate fluence at coords using Beam
    Find depth of coords along beam to upstream face of crystal
    Calculate dose and fluence at coords
    Update dose and fluence for voxel
  end
end
Calculate aggregate metrics
Output Wedge aggregate metrics
```

Figure 3
Pseudo-code showing how the dose for each wedge is calculated by *RADDOSE-3D*.

[1] Itself calculated by taking the intensity at the face of the crystal and attenuating this for the depth of the volume element within the crystal.
[2] Supplementary material is available from the IUCr electronic archives (Reference: RR5041). Services for accessing this material are described at the back of the journal.

electronic reprint

Table 1
Default output files produced by *RADDOSE-3D*.

File	Contents	Class
`output-Summary.txt` and STDOUT	Human readable description of experiment, output metrics and histograms of final dose state for each wedge.	`OutputSummaryText` implements `Output` and `ExposeObserver`
`output-Summary.csv`	Easily parsable summary of output metrics for each wedge.	`OutputSummaryCSV` implements `Output`
`output-DoseState.csv`	x, y, z, dose, fluence data for the final state of the crystal. For plotting dose maps.	`OutputFinalDoseStateCSV` implements `Output`

Table 2
Metrics output in the summary file of *RADDOSE-3D*.

Further discussion of the applicability of each of these metrics to different circumstances is presented by Zeldin *et al.* (2013).

Average dose–whole crystal, AD–WC	Total absorbed energy (J) divided by the mass of the whole crystal.
Average dose–exposed region, AD–ER	Total absorbed energy (J) divided by the mass of a region defined as any voxel having nonzero dose.
Threshold average dose, TAD	The average dose inside a volume bounded by the isosurface calculated to enclose a fraction of the total absorbed energy: the default is 95%.
Maximum dose, max dose	The dose in the voxel with the highest dose value anywhere in the crystal.
Dose contrast	Maximum dose divided by threshold average dose: a measure of the contrast in the dose distribution.
Used volume	The fraction of the crystal voxels with a nonzero dose.
Absorbed energy	The sum of the absorbed energy for all the voxels in the crystal.
Final dose histogram	A histogram showing the fraction of voxels in each dose range.

updated after calculating each wedge, so that a step-by-step report of the dose state of the crystal can be provided at the end of the simulation. This step-by-step output allows information to be obtained on the time course of the dose accumulation.

3. Technical details of the code

RADDOSE-3D is written in Java and follows an object-oriented programming architecture. In this programming paradigm, concepts are described by classes which contain program code and a description of attributes. Objects are instances of such classes, wherein program code and associated data are combined into a consistent unit with an internal state. Objects can be used to represent a part of the actual experiment, such as a specific crystal in an experiment. The overall program structure is documented in Figs. S1–S6 in the supplementary material. Major elements of the diffraction experiment such as the crystal, the X-ray beam and wedges of X-ray exposure are represented by objects. These objects are created from configuration files, or standard input (STDIN), as illustrated in Fig. S2. This generic input structure can easily be extended to accommodate other sources, such as a future graphical user interface or automated streaming input from other programs in a beamline environment (*e.g.* beam profile and crystal topography).

`Crystal` objects contain a representation of the physical crystal volume as a set of evenly spaced voxel coordinates, which can be rotated and translated using functions supplied by the object. For each voxel the associated dose and fluence values are stored. The size of the voxels can be user defined. The recommended voxel size for convergence of the dose metrics is approximately one-tenth of the beam FWHM (minimum of x and y), as demonstrated by the simulation results shown in Fig. S7 of the supplementary material.

For any crystal type, the attenuation and absorption coefficients can be calculated from the unit-cell parameters using a previous version of *RADDOSE* (v1, 2 or 3), or alternatively average values of the absorption coefficient of 0.237 mm^{-1} and attenuation coefficient of 0.281 mm^{-1}, which are appropriate for an incident beam energy of 12.4 keV (1 Å), can be used. These values were calculated (Holton & Frankel, 2010) using an average protein composition from a survey of the Protein Data Bank (Berman *et al.*, 2002). When using these averages, the crystal density is taken as 1.2 g ml^{-1} with an assumed 50% solvent content with a solvent mass of ∼1 g ml^{-1} and a protein mass of 1.41 g ml^{-1} (Fischer *et al.*, 2004).

The `Beam` interface facilitates access to information describing the beam profile, energy and flux (Fig. S4). `Wedge` objects contain all the information pertaining to the data collection strategy: rotational range, any offset either in the starting orientation and position of the crystal or between the rotation axis and the beam axis, and any translation that is to take place during the wedge exposure.

The output of *RADDOSE-3D* is created by modules following the observer design pattern (Gamma *et al.*, 1994). This program structure allows easy creation of new output modules that can produce any required statistic in any format. Depending on the user's needs, different output modules can be selected. The default output module selection is shown in Table 1. A diagram of all the `Output` classes is given in Fig. S6.

4. Program output

The raw output of the three-dimensional dose simulation is a scalar field describing the dose state of the crystal at every voxel and at every time step. Because each wedge is broken into time steps corresponding to a defined angular resolution (default 2°), a fully time-resolved picture yields a dose field for each angular step computed during the simulation. Since it is neither practical nor very illuminating to manually look at several hundred (*e.g.* 180 dose fields for a single 360° wedge at default resolution) four-dimensional arrays, each with hundreds of thousands of entries (typical number of voxels ∼125 000), it is vital to compute aggregate metrics.

Because there is, as of yet, no experimental consensus regarding what appropriate dose-based metric should be used to describe uneven dose distributions (Zeldin *et al.*, 2013), the code currently produces several metrics, as summarized in Table 2. They range from the highly intuitive average dose–whole crystal (AD–WC), a simple measure of the average dose in the whole crystal volume, or max dose, the highest dose value at any point in the crystal volume, to the more sophisticated metrics, such as threshold average dose (TAD) or average dose–exposed volume (AD–EV), which provide potential routes to better describe the dose field of a crystal. Contrast metrics, such as dose contrast (max dose divided by TAD), are also output as a way to assess the approximate pattern of dose within the diffracting volume.

electronic reprint

Consider a fully, evenly exposed crystal and a crystal with only one section highly damaged: both of these cases can have absorbed the same total number of photons and so have the same AD–WC. However it is clear that they will produce very different diffraction patterns. In the case of TAD, a better estimate is achieved by only including the volume that encloses 95% of the absorbed energy, so that unexposed or weakly exposed regions that are associated with the very weak tails of a typical Gaussian beam are thus discounted, leading to a more informative dose estimate. A full discussion of the various metrics listed in Table 2 is presented by Zeldin *et al.* (2013).

Because wedges can overlap each other, all the metrics except for absorbed energy are calculated for the whole crystal volume at the end of each wedge; they represent the dose state at that time point during the experiment. However, absorbed energy for a given wedge thus refers to the energy absorbed only during the specified wedges, not during earlier wedges (*i.e.* it is not cumulative).

In addition to a summary file containing these metrics which are calculated after each wedge, the code can currently output two additional files in comma separated value (CSV) format, as detailed in Table 1. Firstly it provides an `output-summary.csv` file, which contains the same information as the summary file but with one column per metric and one row per wedge. This is an easy to parse file for convenient use in batch processing and scripting. The second additional file is an `output-DoseField.csv` file which is in the format *X*, *Y*, *Z* coordinate, dose, and finally fluence, with one row for each voxel in the crystal. This provides a full map of the final dose field, after all the wedges have been exposed. It can be used to create three-dimensional figures and to visualize the state of the crystal using software such as *ParaView* by KitWare (http://www.paraview.org/) or the R programming language.

5. Example of the use of *RADDOSE-3D*

To illustrate the workflow of using the software, a simple 90° wedge data collection is simulated. The input file for this is described in Fig. 4(*a*). As described above, the crystal block states that the crystal is a cuboid of size 100 × 100 × 100 µm, and the absence of any explicit crystal composition instructions will cause the program to default to average absorption and attenuation coefficients. Pixel resolution, defined in pixels per micrometre, is a particularly important parameter to be specified. As detailed in §3, the resolution should not be smaller than one-tenth of the beam FWHM, which for a 20 × 40 µm FWHM beam such as the one considered here is 2 µm per voxel, hence 0.5 voxels µm^{-1}. Setting this value too low will lead to errors in dose estimation, particularly for max dose, and too high a value will increase the runtime needlessly. The beam block describes a Gaussian profile with a vertical FWHM of 40 µm and a horizontal FWHM of 20 µm, with 10^{12} 12.1 keV photons per second and an outer beam collimation of 80 × 40 µm (vertical by horizontal). This collimation refers to the size of the slits used to collimate the beam, but for uncollimated Gaussians, setting this to twice the beam FWHM will speed up the runtime of the code, while only introducing an error that is much smaller than the likely fidelity of the true beam profile to a perfect Gaussian. Finally, the wedge block defines a 90° data collection over 30 s that will be processed in angular steps of 2°. It is only worth changing the 2° default angular resolution in the case of very small wedges (*e.g.* 5° wedges used to probe relative *B* factors). Fig. 4(*b*) shows the summary output file for this simulation and Fig. 5 a dose map created from the `DoseField.csv` file using *Paraview*. As can be seen both visually from the dose map and through the variation in the dose metrics in the summary file, this is a highly inefficient use of the crystal volume, with a maximum dose of 12.9 MGy and an AD–WC of 1.2 MGy. The TAD of 3.7 MGy represents the average dose within the region that has absorbed 95% of the energy, and so offers an average that effectively weights the weakly diffracting tails of the Gaussian to zero.

6. Discussion

Thanks to the highly extensible approach chosen when designing the *RADDOSE-3D* software, there is tremendous scope to take into account experimentally determined beamline and crystal parameters. New `Beam` classes can be written to take an experimental profile measurement as input; `Crystal` classes could model accurate crystal shapes using experimental data from imaging techniques such as tomography (Brockhauser *et al.*, 2008) or online confocal microscopy (Khan *et al.*, 2012). The main dose calculation function has been

```
Crystal
Type Cuboid
Dimensions 100 100 100
PixelsPerMicron 0.5
Beam
Type Gaussian
Flux 1e12
FWHM 40 20
Energy 12.1
Collimation Rectangular 80 40
Wedge 0 90
ExposureTime 30
AngularResolution 2
```

(*a*)

```
Cuboid Crystal of size [100, 100, 100] um [x, y, z] at a resolution
of 2.00 microns per voxel edge.
Simple Dose Decay Model used. Diffraction intensity does not decay
with increasing dose.
Gaussian beam, 80.0x40.0 um with 40.00 by 20.00 FWHM (x by y) and
1.0e+12 photons per second at 12.10 keV.
Wedge 1:
Collecting data for a total of 30.0s from phi = 0.0 to 90.0 deg.
Dummy crystal coefficients used.
Absorption coefficient: 2.37e-04 /um.
Attenuation coefficient: 2.81e-04 /um.
Elastic coefficient: 1.80e-05 /um.
Density: 1.20 g/ml.
Average Dose (Whole Crystal)      : 1.17 MGy
Average Dose (Exposed Region)     : 2.89 MGy
Max Dose                          : 12.865 MGy
Average Dose (95.0 % of total absorbed energy threshold (0.60 MGy)):
3.65 MGy
Dose Contrast (Max/Threshold Av.) : 3.53
Used Volume                       : 40.4%
Absorbed Energy (this Wedge)      : 1.49e-03 J.
Dose Inefficiency (Max Dose/mJ Absorbed) : 8.7 1/g
Final Dose Histogram:
Bin  1,  0.0 to  0.1 MGy:  6.8 %
Bin  2,  0.1 to  3.4 MGy: 58.9 %
Bin  3,  3.4 to  6.7 MGy: 23.0 %
Bin  4,  6.7 to 10.1 MGy:  9.0 %
Bin  5, 10.1 to 13.4 MGy:  2.2 %
Bin  6, 13.4 to 16.7 MGy:  0.0 %
Bin  7, 16.7 to 20.0 MGy:  0.0 %
Bin  8, 20.0 to 23.4 MGy:  0.0 %
Bin  9, 23.4 to 26.7 MGy:  0.0 %
Bin 10, 26.7 to 30.0 MGy:  0.0 %
Bin 11, 30.0 MGy upwards:  0.0 %
```

(*b*)

Figure 4
(*a*) Simple input file describing a 90° wedge of data, for which the final dose field is visualized in Fig. 5. (*b*) Summary output file for the input shown in (*a*). Note the large spread between the maximum dose (12.9 MGy) and the average dose–whole crystal (1.2 MGy), and the inefficient use of the crystal volume: only 40% has a nonzero dose, and, as can be seen from the dose histogram, most of this is within the 0.1–3.4 MGy bin, so is weakly exposed.

electronic reprint

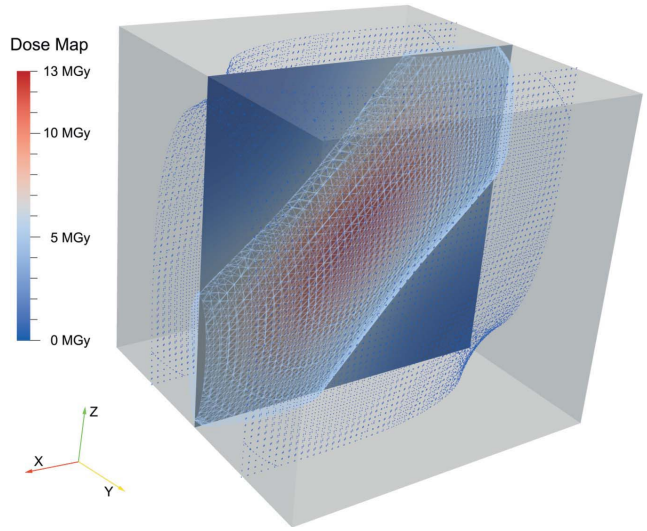


Figure 5
Dose map for the input configuration described in Fig. 4. Isosurfaces are shown at 0.05 and 5 MGy, and a slice map is shown in the plane perpendicular to the rotation axis and along the beam axis. The crystal has been rotated about the *y* axis, with the beam starting down the *z* axis and ending along the *x* axis, as labelled in the figure. Visualization produced using *Paraview*.

implemented such that it could be adapted to monitor a real-time feed of the beamline status (shutter status and goniometer position), allowing users to monitor dose during data collection and thus to make more informed decisions relating to the collection of multiple wedges from one crystal. An immediate example of this is that scientists at synchrotron light sources could implement their own `Beam` classes to simulate beam profiles from the experimental source at that particular beamline. Because these additions are encapsulated within broader interfaces, the way they interact with the rest of the code is pre-determined, and so they can be added without re-testing the broader functionality of the code.

More generally, the entire framework could be repurposed for the calculation of absorption corrections in cases where this becomes an important factor (*e.g.* for long-wavelength beamlines; Armin Wagner, personal communication).

We envisage this code being used on multiple levels: firstly as a routine online application for optimizing experimental protocols for radiation-sensitive crystals such as membrane proteins or multi-domain complexes with tenuous crystal contacts. Secondly, it will allow a more accurate reporting of the damage state of solved crystal structures, and thirdly, it provides a powerful characterization tool for radiation-damage studies, solving one of the major challenges in collecting systematic and statistically significant data for radiation-damage investigations (Allan *et al.*, 2013). It thus allows a broader range of protocols to be used and much more accurate estimation of dose to be implemented in the analysis.

To allow the general use of *RADDOSE-3D*, a free web service is provided at http://www.raddo.se. Users can submit configuration files as individual jobs which are then run from a queue. Once finished, the users are notified by e-mail and can then download the associated result files. Submitted files and results are only visible for the submitting user and are only retained for a limited number of days.

As the web service is a shared system, submitted jobs might be queued for some time depending on concurrent demand and estimated job runtime. Users who require the immediate processing of their job requests, have a large number of job requests or want to extend the operation of *RADDOSE-3D* with their customized output modules can obtain a copy of *RADDOSE-3D* by contacting the corresponding author.

7. Conclusions

We have presented a tool that takes the crystal, beam and experimental parameters of a macromolecular crystallography diffraction experiment and generates a step-by-step picture of the dose profile within the crystal as the data collection proceeds. This basic output can then be used to describe the damage profile within a crystal through the use of appropriate summary metrics, which serve either as a routine tool for predictively evaluating the damage state of a given macromolecular crystal or as a more accurate method of estimating dose for radiation-damage studies.

Given the heterogeneous nature of dose fields, it is vital that future work focuses on experimentally characterizing the metrics that can be extracted from a dose field and establishing those which best describe the damage state of the crystal, as observed in the diffraction pattern. Current research efforts focus on answering this question by performing large-scale analyses in test proteins using strategies that result in a wide range of dose profiles.

We gratefully acknowledge the Engineering and Physical Sciences Research Council (EPSRC) in the UK for funding through studentships at the Life Science Interface (OZ) and the Systems Biology (MG) programmes of the University of Oxford's Doctoral Training Centre. We also thank James Holton for valuable discussions during the development of this software and Jean-Paul Ebejer for the development of the online web interface.

References

Allan, E. G., Kander, M. C., Carmichael, I. & Garman, E. F. (2013). *J. Synchrotron Rad.* **20**, 23–36.
Axford, D. *et al.* (2012). *Acta Cryst.* D**68**, 592–600.
Berman, H. M. *et al.* (2002). *Acta Cryst.* D**58**, 899–907.
Blake, C. & Phillips, D. C. (1962). *Proceedings of the Symposium on the Biological Effects of Ionizing Radiation at the Molecular Level*, pp. 183–191. Vienna: International Atomic Energy Agency.
Bourenkov, G. P. & Popov, A. N. (2010). *Acta Cryst.* D**66**, 409–419.
Brockhauser, S., Di Michiel, M., McGeehan, J. E., McCarthy, A. A. & Ravelli, R. B. G. (2008). *J. Appl. Cryst.* **41**, 1057–1066.
Chapman, H. N. *et al.* (2011). *Nature*, **470**, 73–77.
Fischer, H., Polikarpov, I. & Craievich, A. F. (2004). *Protein Sci.* **13**, 2825–2828.
Gamma, E., Helm, R., Johnson, R. & Vlissides, J. (1994). *Design Patterns: Elements of Reusable Object-Oriented Software.* Reading: Addison Wesley.
Garman, E. F. (2010). *Acta Cryst.* D**66**, 339–351.
Holton, J. M. (2009). *J. Synchrotron Rad.* **16**, 133–142.
Holton, J. M. & Frankel, K. A. (2010). *Acta Cryst.* D**66**, 393–408.
Khan, I., Gillilan, R., Kriksunov, I., Williams, R., Zipfel, W. R. & Englich, U. (2012). *J. Appl. Cryst.* **45**, 936–943.
Kmetko, J., Husseini, N. S., Naides, M., Kalinin, Y. & Thorne, R. E. (2006). *Acta Cryst.* D**62**, 1030–1038.
Krojer, T. & von Delft, F. (2011). *J. Synchrotron Rad.* **18**, 387–397.
Murray, J. W., Garman, E. F. & Ravelli, R. B. G. (2004). *J. Appl. Cryst.* **37**, 513–522.
Nave, C. & Hill, M. A. (2005). *J. Synchrotron Rad.* **12**, 299–303.
Owen, R. L., Rudiño-Piñera, E. & Garman, E. F. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 4912–4917.
Paithankar, K. S. & Garman, E. F. (2010). *Acta Cryst.* D**66**, 381–388.
Paithankar, K. S., Owen, R. L. & Garman, E. F. (2009). *J. Synchrotron Rad.* **16**, 152–162.
Riekel, C., Burghammer, M. & Schertler, G. (2005). *Curr. Opin. Struct. Biol.* **15**, 556–562.
Sanishvili, R., Yoder, D. W., Pothineni, S. B., Rosenbaum, G., Xu, S., Vogt, S., Stepanov, S., Makarov, O. A., Corcoran, S., Benn, R., Nagarajan, V., Smith, J. L. & Fischetti, R. F. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 6127–6132.
Warkentin, M., Badeau, R., Hopkins, J. B., Mulichak, A. M., Keefe, L. J. & Thorne, R. E. (2012). *Acta Cryst.* D**68**, 124–133.
Zeldin, O. B., Gerstel, M. & Garman, E. F. (2013). *J. Synchrotron Rad.* **20**, 49–57.

electronic reprint

Chapter 8
Elemental Analysis of Proteins by Proton Induced X-ray Emission (microPIXE)

Elspeth F. Garman and Oliver B. Zeldin

Abstract The identification and quantification of metals bound to proteins is a crucial problem to be solved in structural biology. This chapter will describe the technique of proton induced X-ray emission with a microfocused proton beam (microPIXE) as a tool for analysing the elemental composition of liquid and crystalline protein samples. The proton beam induces characteristic X-ray emission from all elements in the protein, which can be interpreted in terms of the metal content of the protein molecule with a relative accuracy of between 10 and 20 %. The compelling advantage of this method is that the sulphur atoms in the methionines and cysteines of the protein provide an internal calibration of the number of protein molecules present so that systematic errors are minimised and the technique is entirely internally self-consistent. This is achieved by the simultaneous measurement of the energy of backscattered protons (Rutherford backscattering), to enable the matrix composition and thickness to be determined, and so correct the PIXE data for the self-absorption of X-rays in the sample. The technical and experimental procedures of the technique will be outlined, and examples of recent measurements given which have informed a range of investigations in structural biology. The use of the technique is increasing and we are in the final stages of developing it to be a routine high-throughput method.

Keywords microPIXE • Metalloproteins • Trace element analysis • Structural genomics

E.F. Garman (✉) • O.B. Zeldin
Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK
e-mail: elspeth.garman@bioch.ox.ac.uk

R. Read et al. (eds.), *Advancing Methods for Biomolecular Crystallography*,
NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-007-6232-9_8, © Springer Science+Business Media Dordrecht 2013

8.1 Introduction

The characterisation of metal atoms in proteins and other large bio-molecules is a crucial problem to be solved in structural biology. The structure–function relationship of molecules is often pivotally dependent on bound metal ions or co-factors, so that understanding function requires knowledge of their identity and concentration. In X-ray crystallography, the electron density commonly indicates the presence of a metal ion, but at present there is no optimum technique for assigning a unique atomic number to that density. In addition, there are a variety of other questions that can be answered by knowing the elemental composition of the molecule, such as whether or not DNA is bound, how much selenium–methionine has been substituted for sulphur–methionine during expression, and whether or not a protein subjected to site directed mutagenesis binds a particular metal ion. For structure determination by NMR techniques, the presence of paramagnetic atoms can make measurement impossible, and it is problematic to identify them uniquely so that they can be removed. For any of the above reasons, performing a trace element analysis on our samples may be necessary.

Metal detection and quantification in proteins imposes stringent requirements on the analytical techniques employed. Three key performance indicators need to be considered: *sensitivity* (minimum detectable limit, MDL, or limit of detection, LoD), *spatial selectivity*, and *quantitative accuracy and precision*.

The sensitivity is constrained by the mass fraction of the metal in the protein, i.e. typically one metal atom of mass around 50 Da in a molecule of around 100 kDa, or 500 $\mu g\ g^{-1}$, and so any method must be able to provide adequate analytical precision at these concentration levels.

Spatial resolution is necessary in order to distinguish small crystals or dried protein precipitate from the surrounding buffer material. The small size of typical samples requires a spatial resolution of less than 10 μm. It is also necessary to have a mapping capability in order to identify the sample or regions of contamination.

The quantitative accuracy and precision must be sufficient to give an unambiguous determination of the number of metal atoms present in a protein molecule. This determination can be facilitated by forming ratios to the sulphur present in the protein (or phosphorus in nucleic acids) and so the accuracy requirements are relaxed. Nonetheless, determining the ratio M_Z/M_S (where M_Z and M_S are the mass of the unknown metal and sulphur, respectively) to an accuracy and precision of better than 10 %, as is typically required for reliable determination of stoichiometries, can be difficult to achieve.

It is additionally highly desirable for an assay to have the ability to identify unexpected elements (i.e. with no pre-selection of analytes): multi-elemental capability is essential when trying to determine the identity of unknown metals. Speed and convenience are also very important factors in the selection of appropriate analytical methods.

The technique which forms the subject of this chapter, microbeam proton induced X-ray emission (microPIXE) [6], is the only option readily available at

present which satisfies all the above constraints. PIXE relies on the interaction of high energy (2–4 MeV) protons with the electron shells of the target atoms to stimulate the emission of characteristic X-rays which can be detected to identify and quantify the atoms present [9]. This is in many ways analogous to energy dispersive X-ray analysis (EDX) using an electron microprobe, but the crucial advantage for analytical purposes is that the relatively high mass of the primary particles compared with the electrons with which they are interacting means that the continuous background of bremsstrahlung X-rays, which seriously degrades the MDL of EDX, is essentially absent. This gives PIXE the potential to achieve detection limits in the region of $\mu g\ g^{-1}$, as is required for the determination of metalloprotein stoichiometries. MeV proton beams can be focused to micrometre dimensions and this provides the basis of a high performance scanning microanalytical instrument, the nuclear microprobe [7].

The application of the microPIXE technique to protein analysis has been developed over the last 15 years. The systematics of the technique have been established [3, 4] and the method has been well proven through measurements carried out on well over 150 different proteins in crystalline or liquid form, with multiple studies being carried out on a number of them.

The compelling advantage of proton beam trace element analysis compared to other currently available techniques for proteins is the internal normalisation. Since every element heavier than fluorine can be observed simultaneously, the sulphur from the methionines and the cysteines in the protein can be used as an internal standard for the amount of protein present. Concentrations of all elements of interest can be simultaneously measured relative to the sulphur signal, and since the primary sequence of the protein is known, the number of atoms of sulphur per molecule is also determined.

The measured ratio can thus be trivially computed into a stoichiometric ratio for the number of atoms of the element of interest per protein molecule, $N_{\mathbf{X}}$, and can then be determined in a straightforward manner using the relationship:

$$N_X = \frac{C_X}{C_S} \times \frac{M_S}{M_X} \times N_S \tag{8.1}$$

where N_x = number of atoms of X per protein molecule,
C_X = PIXE measured concentration of element X,
C_S = PIXE measured concentration of sulphur,
M_S/M_X = Mass ratio of sulphur and element X, and
N_s = Number of sulphur atoms per protein molecule, already known from the primary amino acid sequence.

Samples containing DNA offer an alternative internal standard, as the phosphorus content is proportional to the length of the DNA strand. In DNA-protein complexes, this can be used to determine the number of DNA molecules per protein molecule [1], or in place of the sulphur standard, as an internal normalisation.

The internal normalisation aspect of the method applied to proteins makes any absolute measurements unnecessary, thus minimising the systematic errors. The initial protein concentration does not have to be known, as long as it is above a minimum required concentration (see below). The internal standard provided by nature allows an accuracy of between ±10 and ±20 % to be achieved: this level of error would be very difficult to reach if absolute measurements were necessary.

8.2 Implementation of the Technique

Our microPIXE analyses are carried out with 2.5 MeV protons focused to a diameter of 2–3 μm using the University of Surrey Ion Beam Centre's microbeam facility (beamline described as installed at Oxford in Grime et al. [8] and subsequently moved to Surrey). Characteristic X-rays are detected using an 80 mm^2 solid state Si(Li) detector (giving high energy resolution), and the detector is fitted with a 130 μm Be filter. During exposure to the proton beam, samples are held under vacuum on a thin organic film. The film is mounted on an aluminium holder, which is held in a 'ladder' capable of holding up to four samples, within the vacuum chamber. Both liquid and crystal samples can be mounted in this way.

For a measurement, first a coarse scan (150×150 μm up to 2.0×2.0 mm^2 depending on the sample size) is collected over the protein drop or crystal by scanning the proton beam spatially in *X* and *Y*. A software window is placed round the X-ray peak in the spectrum associated with a particular element of interest, and the counts are sorted into an *XY* grid to give individual elemental maps.

Simultaneous detection of Rutherford backscattered (RBS) protons (e.g. Fig. 8.1) allows the thickness and matrix composition of the sample to be accurately determined [5]. Quantitative information is obtained by collecting spectra (e.g. Fig. 8.1) at three or four points on the sample and also on the backing film alone for 3–10 min each. These spectra are analysed using the program OMDAQ to fit the RBS spectra, and which also provides an interface to GUPIX [2] for processing the PIXE spectra. The number of atoms of each element of interest per protein molecule can then be computed from Eq. (8.1). In order to avoid the generation of spurious signals from the material of the target chamber, the beam is stopped in a Faraday cup fabricated from spectroscopically pure graphite. Beam currents are in the range 0.1–1 nA.

8.3 Experimental Considerations

Both liquid and crystalline protein samples can be investigated using microPIXE and its application to proteins is now well established [3, 4].

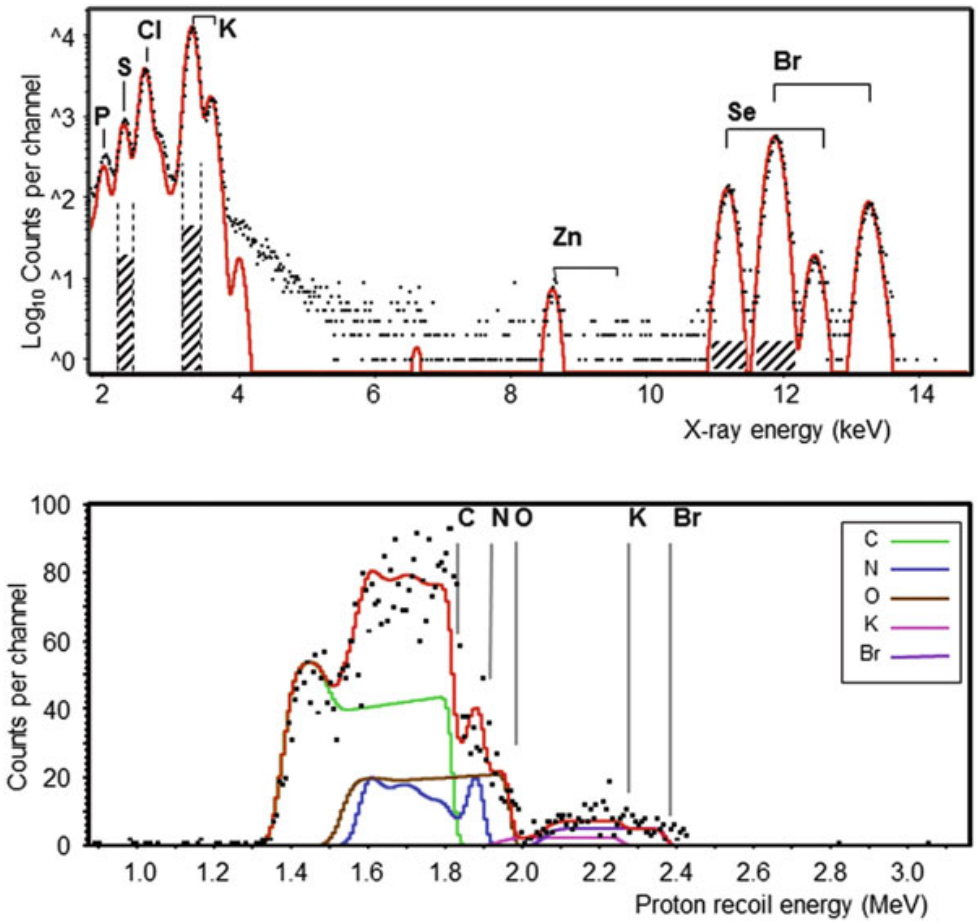


Fig. 8.1 PIXE (*top*) and RBS spectra (*bottom*) from a 2 μ diameter point on a thin liquid protein sample from which the S/Se ratio was sought, and which had been buffer exchanged from KCl into KBr to avoid overlap of the sulphur and chlorine peaks. The spectra were recorded simultaneously in 5 min using a beam of 2.5 MeV protons at a current of 100 pA. In each spectrum, the recorded data are shown as *black points* and the theoretical fit generated by the spectrum processing software is shown as a *solid red line*. In the PIXE spectrum, the identity of the major peaks is marked: note the log *y* scale of this spectrum. The *shaded regions* indicate the energy limits used to generate elemental maps such as those shown in Fig. 8.3. The RBS spectrum is modelled as a 6.15 μ thick layer of material with chemical formula $C_5N_{1.15}O_{2.28}K_{0.21}Br_{0.13}$ mounted on a 1.85 μ polymer film with composition $C_5O_{1.8}$. *Coloured lines* indicate the partial spectra corresponding to each element in the sample structure, and energy of the protons interacting with the surface layer of each element is indicated by *vertical lines* (Color figure online)

8.3.1 Required Concentration of Protein

The concentration of a liquid protein sample does not have to be known accurately for microPIXE, because the internal sulphur normalisation provides the standard. However, there is an MDL for an individual element of approximately 1–10 ppm of the dry weight, corresponding to a minimum concentration of liquid protein samples necessary for reliable quantitation by microPIXE. This minimum has been

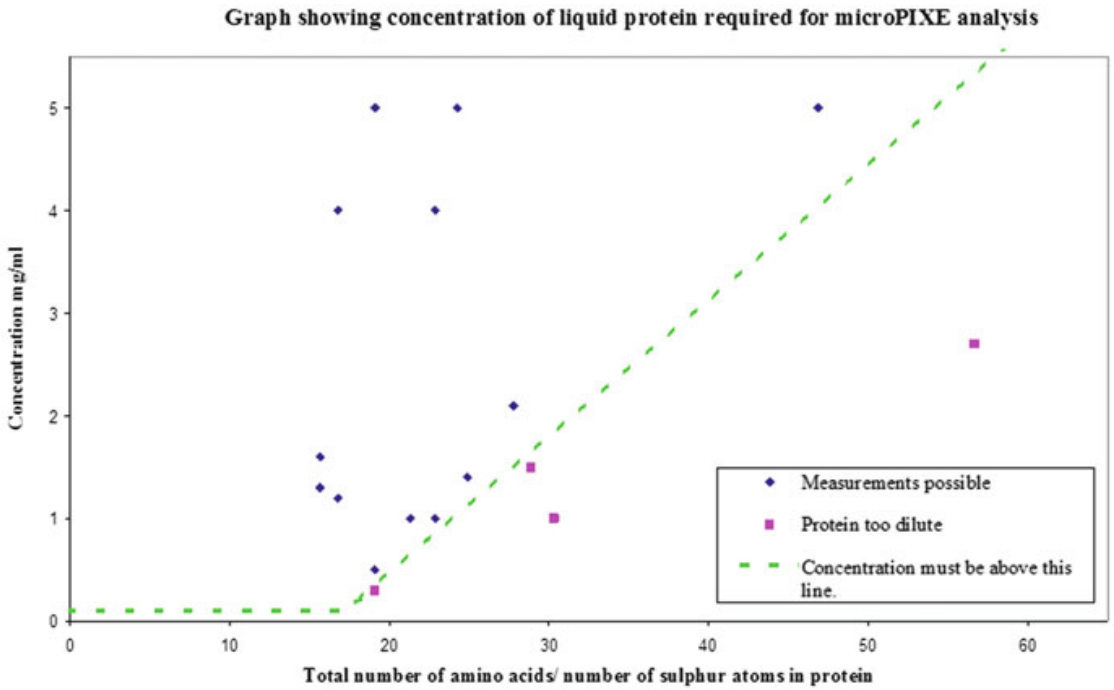


Fig. 8.2 Graph showing the minimum concentration of dried liquid protein residues (mg ml^{-1}) required for analysis by microPIXE as a function of the number of amino acids per sulphur atom in the protein molecule. The *points* represent actual experimental measurements and the *dotted line* represents an empirical guide as to whether or not analysis by microPIXE is feasible. Crystalline samples have concentrations of 100–1,000 mg ml^{-1} and are thus all of sufficient concentration for PIXE analysis

carefully investigated over the last few years, and an empirical limit has been derived. Figure 8.2 shows a graph of measurements performed on liquid samples over a range of fairly low concentrations, defined for convenience in terms of mg ml^{-1}. The concentration has been plotted against the number of amino acids per sulphur atom in the protein, known from the primary sequence information.

The sulphur signal from the samples lying below the dotted line on the plot was too weak to be measured with any confidence. All the samples lying above the line were successfully analysed and reliable quantitative results obtained. It is thus recommended that proteins are concentrated so that they definitely lie above the line before microPIXE analysis is attempted.

8.3.2 Protein Buffer and Crystal Mother Liquor

A major issue that must be addressed for every sample prior to the experimental preparations is that of the constituents of the protein buffer (for liquid samples) or the mother liquor (for crystals). Any buffers containing sulphur will affect the internal normalisation standard and thus result in an unreliable analysis. Many additives used in protein crystallisation include one or more sulphur atoms in their

composition: for instance HEPES, BES and MES. This makes them undesirable buffers for PIXE measurements. This problem is exemplified in the following case study from one of our samples.

A protein was supplied for measurement at a concentration of 1.3 mg ml^{-1} in 10mM HEPES. The HEPES corresponded to a concentration of 2.38 mg ml^{-1} and thus a sulphur concentration of 0.32 mg ml^{-1}. In contrast, the 15 kDa protein contained 8 cysteines and 1 methionine, so the protein sample had a sulphur concentration of 0.025 mg ml^{-1}, 13 times lower than that of the HEPES. A reliable measurement was thus impossible, and in these circumstances buffer exchange into a non-sulphur containing solution would be essential.

Protein solutions and mother liquors also often contain salts. Most of these do not interfere with the microPIXE analysis, a notable exception being sodium chloride, which for liquid samples can give a chlorine peak stronger than the sulphur peak. Since they have consecutive atomic numbers (16 and 17), they give X-ray peaks close in energy. The low energy tail of an intense chlorine peak can affect the accuracy with which the sulphur peak can be fitted by GUPIX and quantified. In a number of cases, proteins have been successfully exchanged into potassium (or sodium) bromide. Although the chlorine peak is still large, it is no longer a major problem. Both the potassium and bromine X-ray peaks are well separated from other peaks of interest, and they can also be used to check for internal consistency, since they should be detected in a 1:1 atomic ratio.

Given a free choice, ammonium acetate is the perfect buffer, since it contains only organic elements which produce X-rays too low in energy to be detected in these experiments.

8.4 Sample Preparation

The samples are dried and then for the experiments, they are held under vacuum on thin organic polypropylene (C_3H_6) film (4 μm thick), stretched and mounted with adhesive across a 1 cm diameter hole in a 1 mm thick aluminium holder.

For liquid protein samples, a volume of between 0.1 and 0.3 μl of the necessary minimum concentration (see above) is required. This is pipetted very gently onto the film under a microscope and allowed to dry at room temperature in a covered environment to prevent dust contamination.

If heat is applied to accelerate drying, we find that the dried drop is more likely to peel off the film either before or during the measurement. Usually, three approximately 0.15 μl drops are deposited onto the film to allow for some redundancy and multiple measurements if required. The drops often dry as a ring, or the protein concentrates at the edge of the drop and the buffer crystallises out in the central part.

For crystalline samples, the crystals are deposited gently onto the film using a cryo-loop and allowed to dry. Both unwashed and washed crystals are usually investigated to check the effect of the mother liquor on the measurements. Washing

is carried out by moving the crystal in a cryo-loop from its growth drop to a microbridge containing 10 μl of milliQ water (tap water contains chlorine which contaminates the samples). It is moved around the drop of water in the loop, and then transferred to a second drop of fresh water, and so on five or six times. This often results in disintegration of the crystal, but this does not matter because microPIXE does not require a crystal of diffracting quality, as it is sensitive only to the atomic species present.

8.5 Examples

During the last 15 years, we have accumulated a body of results from the analysis of more than 150 different proteins by microPIXE. Proteins have been quantitatively analysed for a range of elements, including magnesium, phosphorus from DNA, potassium, calcium, manganese, iron, cobalt, nickel, copper, zinc, selenium, iodine, barium, platinum, gold, mercury, and tungsten. We describe just one case below in order to illustrate the use of microPIXE analysis and how it has informed various aspects of structural studies on particular proteins. Many more examples can be found in Garman and Grime [4].

MicroPIXE analysis was carried out of the third KH domain of hnRNP K in complex with 'single-stranded' DNA. hnRNP K is one of the major proteins found in hnRNP particles which are ribonucleoprotein complexes containing protein and pre-messenger RNA. hnRNP K contains hnRNP K homology (KH) domains which bind both Cytosine-Thymine rich single-stranded DNA (ssDNA) and Cytosine-Uracil rich ssRNA. Co-crystallization of the third KH domain of human hnRNP K with a 15-mer ssDNA gave both rod shaped and square plate crystals in the same crystallization drop. The square plates were similar in morphology to native crystals obtained previously.

Using X-ray diffraction data from the rod shaped crystals, initial molecular-replacement trials using the structure of the native protein alone as a search model failed to allow the structure of the putative protein-DNA complex to be solved. No heavy-atom derivatives could be obtained to enable phase determination. MicroPIXE experiments showed that these crystals contained large amounts of phosphorus (Fig. 8.3a, b).

The phosphorus to sulphur atomic ratio was measured to be 4.4 and 4.7 in two different rod shaped crystals. This agreed well with a theoretical ratio of 4.67, corresponding to a stochiometry of three KH3 domains (protein molecules) per 15-mer ssDNA, taking into account the fact that the synthetic DNA has only 14 phosphates. For the square plate crystal form (Fig. 8.3c, d), the microPIXE measurements gave a maximum phosphorus to sulphur atomic ratio of below the LoD at 1.1×10^{-2}, showing that DNA was not bound.

Knowledge of the protein to DNA molecular ratio from the microPIXE measurements forced a re-examination of the X-ray data and of the self-rotation calculations. This added further evidence for a crystal asymmetric unit containing three KH3

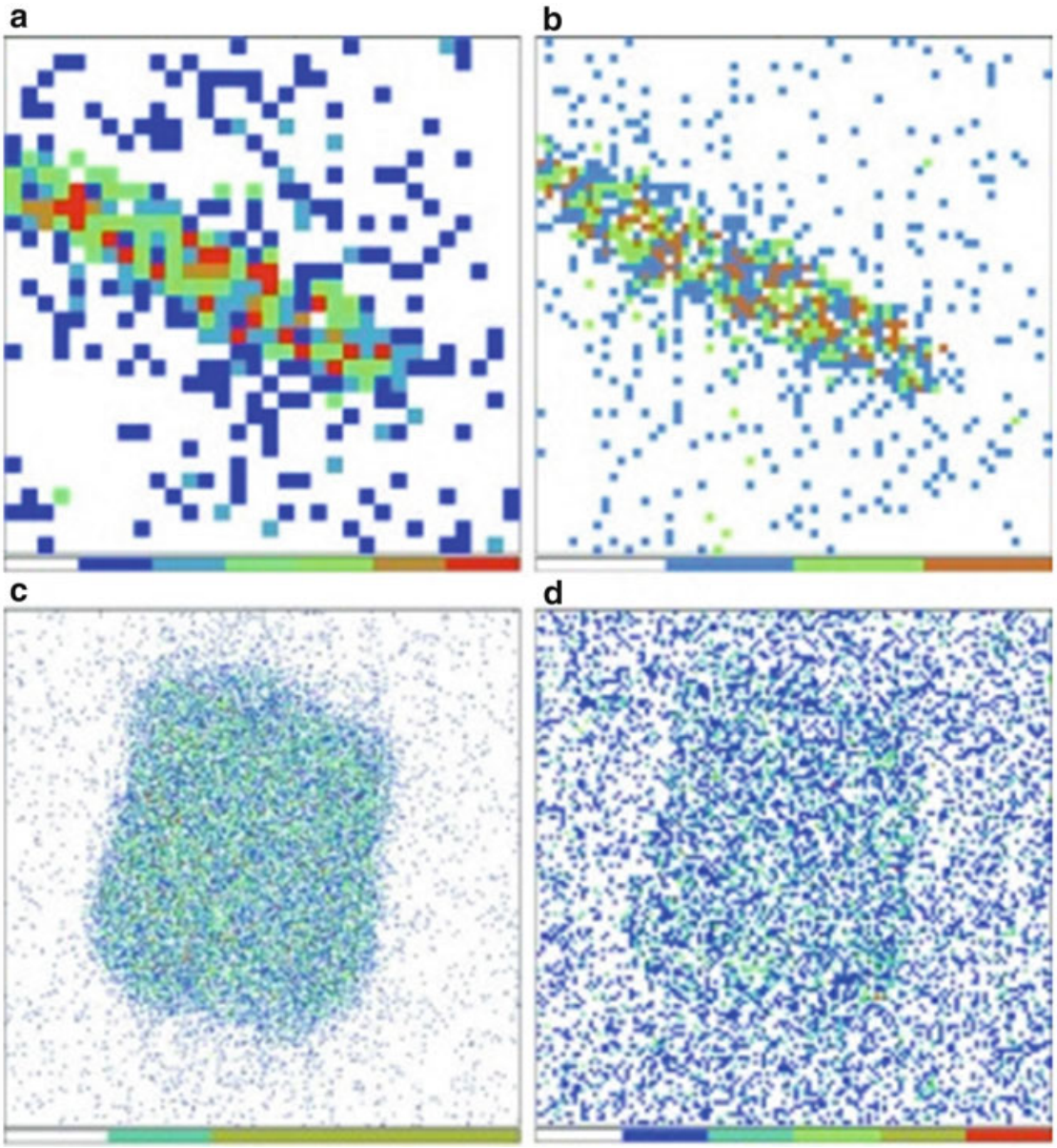


Fig. 8.3 (**a**) Sulphur and (**b**) Phosphorous elemental maps (150×150 μm^2) of ssDNA binding crystal (*rod shape*): DNA bound. Result: 4.4 phosphorus/sulphur atom (1 methionine per protein molecule) and thus 3 protein molecules per 15mer ssDNA. (**c**) Sulphur and (**d**) Phosphorus elemental maps (250×250 μm^2) of a ssDNA crystal (*plate shape*): DNA is not bound: the faint crystal outline is due to an increase in the background bremsstrahlung when scanning over the crystal (Reproduced courtesy of the IUCr from Backe et al. [1]. http://dx.doi.org/10.1107/S0907444904002628)

domains per 15-mer of DNA in the rod shaped crystals. Subsequent molecular-replacement trials using just one copy of the KH3 domain gave a cross-rotation function plot revealing three peaks, in good agreement with the self-rotation function. An electron density omit map was then calculated and gave clear difference density for the DNA, showing that there were two KH3 domains in contact with each DNA molecule and one KH3 domain positioned away from it (Fig. 8.4). Thus the structure of the KH3-DNA complex was solved with the aid of microPIXE analysis [1].

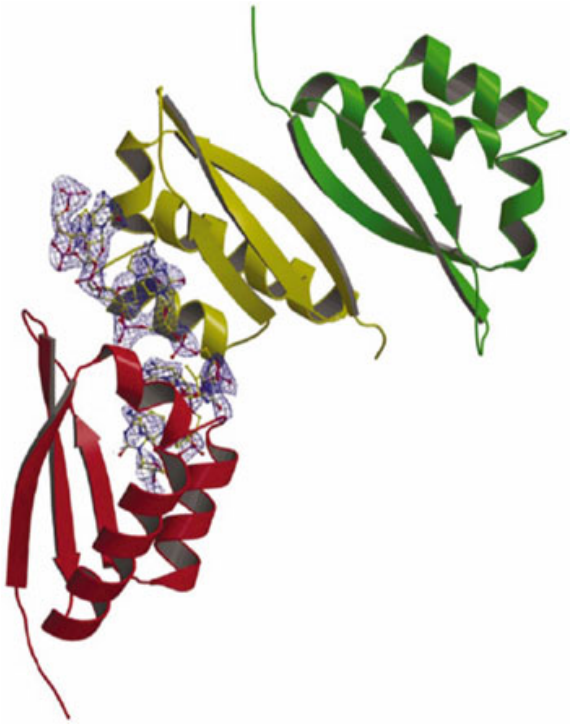


Fig. 8.4 Ribbon representation of the three KH3 domains in the asymmetric unit (coloured *red*, *yellow* and *green*) and an omit map showing the electron density around the single stranded DNA drawn at the 1σ level. Only two of the KH3 domains (*red*, *yellow*) interact with the DNA. The third KH3 domain (*green*) interacts with the second (*yellow*) via the strand β1 edges of their respective β1 sheets, an interaction also seen in crystals of the uncomplexed protein [1] (Reproduced courtesy of the IUCr from Backe et al. [1]. http://dx.doi.org/10.1107/S0907444904002628) (Color figure online)

8.6 Current Developments

The main bottlenecks presented by the current generation of microPIXE assays are in turnaround time and access to beamline facilities. These can be effectively addressed by a high-throughput approach. Current work is focusing on creating a fully integrated pipeline with the goal of making the technique available as a routine assay to laboratories worldwide. Samples are printed into micro-arrays using a non-contact printer, and data collected using automated protocols on the beamline. Data processing will be automated, allowing a large number of samples to be handled in any single 24 h run. Current proof of principle experiments indicate that it is realistic to increase the throughput from ∼10 samples/day to ∼200 samples/24 h. Such a large increase in throughput opens up the possibility of applying the technique to new areas of research, or as a routine assay within a structural genomics pipeline.

Acknowledgments The developments described above would not have been possible without our longstanding collaboration with Dr. Geoff Grime and the Ion Beam Centre at the University of Surrey. We are very grateful for their continuing support of this project. OBZ is funded by an

EPSRC Life Science Interface Doctoral Training Centre PhD studentship, and the high throughput methods development is supported by a grant from the John Fell Fund of the University of Oxford.

References

1. Backe PH, Ravelli RBG, Garman E, Cusack S (2004) Crystallization, microPIXE and preliminary crystallographic analysis of the complex between the third KH domain of hnRNP K and single-stranded DNA. Acta Crystallogr D 60:784–787
2. Campbell JL, Hopman TL, Maxwell JA, Nejedly Z (2000) The Guelph PIXE software package III: alternative proton database. Nucl Instrum Methods B 170:193–204
3. Garman E (1999) Leaving no element of doubt: analysis of proteins using microPIXE. Structure 7:R291–R299
4. Garman EF, Grime GW (2005) Elemental analysis of proteins by microPIXE. Prog Biophys Mol Biol 89:173–205
5. Grime GW (1996) The "Q factor" method: quantitative microPIXE analysis using RBS normalisation. Nucl Instrum Methods B 109:170–174
6. Grime GW (1999) High energy ion beam analysis methods (and background). In: Lindon JC, Tranter GE, Holmes JL (eds) Encyclopedia of spectroscopy and spectrometry. Academic, Chichester, pp 750–760
7. Grime GW (1999) Proton microprobe (method and background). In: Lindon JC, Tranter GE, Holmes JL (eds) Encyclopedia of spectroscopy and spectrometry. Academic, Chichester, pp 1901–1905
8. Grime GW, Dawson M, Marsh M, McArthur IC, Watt F (1991) The Oxford submicron nuclear facility. Nucl Instrum Methods B 54:52–53
9. Johansson SAE, Campbell JL, Malqvist KG (1995) Particle induced X-ray emission spectrometry (PIXE). In: Johansson SAE, Campbell JL, Malmqvist KG (eds) Chemical analysis: a series of monographs on analytical chemistry and its applications. Wiley, New York, p 167

Bibliography

Adams, P. D. , Afonine, P. V. , Bunkóczi, G. , Chen, V. B. , Davis, I. W. , Echols, N. , Headd, J. J. , Hung, L.-W. , Kapral, G. J. , Grosse-Kunstleve, R. W. , McCoy, A. J. , Moriarty, N. W. , Oeffner, R. , Read, R. J. , Richardson, D. C. , Richardson, J. S. , Terwilliger, T. C. , and Zwart, P. H. . PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological Crystallography*, 66(Pt 2):213–21, 2010. ISSN 1399-0047. doi: 10.1107/S0907444909052925.

Allan, E. G. , Kander, M. C. , Carmichael, I. , and Garman, E. F. . To scavenge or not to scavenge, that is STILL the question. *Journal of Synchrotron Radiation*, 20(Pt 1):23–36, 2013. ISSN 1600-5775. doi: 10.1107/S0909049512046237.

Arndt, U. W. and Wonacott, A. J. , editors. *The rotation method in crystallography*. North-Holland, 1977.

Ascone, I. and Strange, R. . Biological X-ray absorption spectroscopy and metalloproteomics. *Journal of Synchrotron Radiation*, 16(Pt 3):413–21, 2009. ISSN 1600-5775. doi: 10.1107/S0909049509010425.

Axford, D. , Owen, R. L. , Aishima, J. , Foadi, J. , Morgan, A. W. , Robinson, J. I. , Nettleship, J. E. , Owens, R. J. , Moraes, I. , Fry, E. E. , Grimes, J. M. , Harlos, K. , Kotecha, A. , Ren, J. , Sutton, G. , Walter, T. S. , Stuart, D. I. , and Evans, G. . In situ macromolecular crystallography using microbeams. *Acta crystallographica. Section D, Biological Crystallography*, 68(Pt 5):592–600, 2012. ISSN 1399-0047. doi: 10.1107/S0907444912006749.

Barker, A. I. , Southworth-Davies, R. J. , Paithankar, K. S. , Carmichael, I. , and Garman, E. F. . Room-temperature scavengers for macromolecular crystallography: increased lifetimes and modified dose dependence of the intensity decay. *Journal of Synchrotron Radiation*, 16(Pt 2):205–16, 2009. ISSN 0909-0495. doi: 10.1107/S0909049509003343.

Berglund, G. I. , Carlsson, G. H. , Smith, A. T. , Szöke, H. , Henriksen, A. , and Hajdu, J. . The catalytic pathway of horseradish peroxidase at high resolution. *Nature*, 417(6887): 463–8, 2002. ISSN 0028-0836. doi: 10.1038/417463a.

Berman, H. M. , Battistuz, T. , Bhat, T. N. , Bluhm, W. F. , Bourne, P. E. , Burkhardt, K. , Feng, Z. , Gilliland, G. L. , Iype, L. , Jain, S. , Fagan, P. , Marvin, J. , Padilla, D. , Ravichandran, V. , Schneider, B. , Thanki, N. , Weissig, H. , Westbrook, J. D. , and Zardecki, C. . The Protein Data Bank. *Acta Crystallographica Section D Biological Crystallography*, 58(6): 899–907, 2002. ISSN 09074449. doi: 10.1107/S0907444902003451.

Bijvoet, J. M. . Structure of Optically Active Compounds in the Solid State. *Nature*, 173(4411): 888, 1954.

Blake, C. and Phillips, D. . 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.

Blundell, T. L. and Johnson, L. . *Protein Crystallography (Molecular Biology Series)*. Academic Press, 1976.

Bourenkov, G. P. and Popov, A. N. . A quantitative approach to data-collection strategies. *Acta crystallographica. Section D, Biological Crystallography*, 62(Pt 1):58–64, 2006. ISSN 0907-4449. doi: 10.1107/S0907444905033998.

Bourenkov, G. P. and Popov, A. N. . Optimization of data collection taking radiation damage into account. *Acta crystallographica. Section D, Biological Crystallography*, 66(Pt 4):409–19, 2010. ISSN 1399-0047. doi: 10.1107/S0907444909054961.

Boutet, S. , Lomb, L. , Williams, G. J. , Barends, T. R. M. , Aquila, A. , Doak, R. B. , Weierstall, U. , Deponte, D. P. , Steinbrener, J. , Shoeman, R. L. , Messerschmidt, M. , Barty, A. , White, T. a. , Kassemeyer, S. , Kirian, R. a. , Seibert, M. M. , Montanez, P. a. , Kenney, C. , Herbst, R. , Hart, P. , Pines, J. , Haller, G. , Gruner, S. M. , Philipp, H. T. , Tate, M. W. , Hromalik, M. , Koerner, L. J. , van Bakel, N. , Morse, J. , Ghonsalves, W. , Arnlund, D. , Bogan, M. J. , Caleman, C. , Fromme, R. , Hampton, C. Y. , Hunter, M. S. , Johansson, L. , Katona, G. , Kupitz, C. , Liang, M. , Martin, A. V. , Nass, K. , Redecke, L. , Stellato, F. , Timneanu, N. , Wang, D. , Zatsepin, N. a. , Schafer, D. , Defever, J. , Neutze, R. , Fromme, P. , Spence, J. C. H. , Chapman, H. N. , and Schlichting, I. . High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography. *Science (New York, N.Y.)*, 362, 2012. ISSN 1095-9203. doi: 10.1126/science.1217737.

Bravais, A. . *Mémoire sur les systèmes formés par les points distribués régulièrement sur un plan ou dans l'espace*. J. Ecole Polytech., Paris, 1850.

Brockhauser, S. , Di Michiel, M. , McGeehan, J. E. , McCarthy, A. a. , and Ravelli, R. B. G. . X-ray tomographic reconstruction of macromolecular samples. *Journal of Applied Crystallography*, 41(6):1057–1066, 2008. ISSN 0021-8898. doi: 10.1107/S002188980802935X.

Brockhauser, S. , White, K. I. , McCarthy, A. A. , and Ravelli, R. B. G. . Translation calibration of inverse-kappa goniometers in macromolecular crystallography. *Acta crystallographica. Section A, Foundations of crystallography*, 67(Pt 3):219–28, 2011. ISSN 1600-5724.

Burley, S. , Petsko, G. , and Ringe, D. . EFFECTS OF X-IRRADIATION OF SINGLE CRYSTALS OF RIBONUCLEASE A. *Acta Crystallographica Section A*, 40(2):41, 1984.

Burmeister, W. . Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallographica Section D Biological Crystallography*, 56(3):328–341, 2000. ISSN 0907-4449. doi: 10.1107/S0907444999016261.

Burnley, B. T. , Afonine, P. V. , Adams, P. D. , and Gros, P. . Modelling dynamics in protein crystal structures by ensemble refinement. *eLife*, 1:e00311, 2012. ISSN 2050-084X. doi: 10.7554/eLife.00311.

Chapman, H. N. , Fromme, P. , Barty, A. , White, T. a. , Kirian, R. a. , Aquila, A. , Hunter, M. S. , Schulz, J. , DePonte, D. P. , Weierstall, U. , Doak, R. B. , Maia, F. R. N. C. , Martin, A. V. , Schlichting, I. , Lomb, L. , Coppola, N. , Shoeman, R. L. , Epp, S. W. , Hartmann, R. , Rolles, D. , Rudenko, A. , Foucar, L. , Kimmel, N. , Weidenspointner, G. , Holl, P. , Liang, M. , Barthelmess, M. , Caleman, C. , Boutet, S. , Bogan, M. J. , Krzywinski, J. , Bostedt, C. , Bajt, S. , Gumprecht, L. , Rudek, B. , Erk, B. , Schmidt, C. , Hömke, A. , Reich, C. , Pietschner, D. , Strüder, L. , Hauser, G. , Gorke, H. , Ullrich, J. , Herrmann, S. , Schaller, G. , Schopper, F. , Soltau, H. , Kühnel, K.-U. , Messerschmidt, M. , Bozek, J. D. , Hau-Riege, S. P. , Frank, M. , Hampton, C. Y. , Sierra, R. G. , Starodub, D. , Williams, G. J. , Hajdu, J. , Timneanu, N. , Seibert, M. M. , Andreasson, J. , Rocker, A. , Jönsson, O. , Svenda, M. , Stern, S. , Nass, K. , Andritschke, R. , Schröter, C.-D. , Krasniqi, F. , Bott, M. , Schmidt, K. E. , Wang, X. , Grotjohann, I. , Holton, J. M. , Barends, T. R. M. , Neutze, R. , Marchesini, S. , Fromme, R. , Schorb, S. , Rupp, D. ,

Adolph, M. , Gorkhover, T. , Andersson, I. , Hirsemann, H. , Potdevin, G. , Graafsma, H. , Nilsson, B. , and Spence, J. C. H. . Femtosecond X-ray protein nanocrystallography. *Nature*, 470(7332):73–7, 2011. ISSN 1476-4687. doi: 10.1038/nature09750.

Corbett, M. C. , Latimer, M. J. , Poulos, T. L. , Sevrioukova, I. F. , Hodgson, K. O. , and Hedman, B. . Photoreduction of the active site of the metalloprotein putidaredoxin by synchrotron radiation. *Acta Crystallographica Section D, Biological Crystallography*, 63(9): 951–960, 2007. doi: 10.1107/S0907444907035160.

Cowan, J. and Nave, C. . The optimum conditions to collect X-ray data from very small samples. *Journal of Synchrotron Radiation*, 15(Pt 5):458–62, 2008. ISSN 0909-0495. doi: 10.1107/S0909049508014623.

Crick, F. H. C. and Magdoff, B. S. . The theory of the method of isomorphous replacement for protein crystals. I. *Acta Crystallographica*, 9(11):901–908, 1956. ISSN 0365-110X. doi: 10.1107/S0365110X56002552.

Dauter, Z. and Wilson, K. S. . Chapter 9 . 1 . Principles of monochromatic data collection. In Arnold, E. , Himmel, D. M. , and Rossmann, M. G. , editors, *International Tables for Crystallography Volume F: Crystallography of biologicalmacromolecules*, pages 211–230. International Union of Crystallography, 2012.

Dauter, Z. . Efficient use of synchrotron radiation for macromolecular diffraction data collection. *Progress in biophysics and molecular biology*, 89(2):153–72, 2005. ISSN 0079-6107. doi: 10.1016/j.pbiomolbio.2004.09.004.

Debreczeni, J. , Bunkoczi, G. , Ma, Q. , Blaser, H. , and Sheldrick, G. M. . research papers In-house measurement of the sulfur anomalous signal and its use for phasing research papers. *Acta Crystallographica Section D Biological Crystallography*, 59:688–696, 2003.

Degtyarenko, K. . Bioinorganic motifs: towards functional classification of metalloproteins. *Bioinformatics (Oxford, England)*, 16(10):851–64, 2000. ISSN 1367-4803.

Diederichs, K. , Mcsweeney, S. , and Ravelli, R. B. G. . Zero-dose extrapolation as part of macromolecular synchrotron data reduction research papers Zero-dose extrapolation as part of macromolecular synchrotron data reduction. *Acta Crystallographica Section D Biological Crystallography*, 59:903–909, 2003.

Drenth, J. . Chapter 2.1. Introduction to basic crystallography. In Arnold, E. , Himmel, D. M. , and Rossmann, M. G. , editors, *International Tables for Crystallography Volume F: Crystallography of biologicalmacromolecules*, pages 45–63. International Union of Crystallography, 2012.

Emsley, P. , Lohkamp, B. , Scott, W. G. , and Cowtan, K. . Features and development of Coot. *Acta Crystallographica. Section D, Biological Crystallography*, 66(Pt 4):486–501, 2010. ISSN 1399-0047. doi: 10.1107/S0907444910007493.

Evans, P. . Scaling and assessment of data quality. *Acta crystallographica. Section D, Biological Crystallography*, 62(Pt 1):72–82, 2006. ISSN 0907-4449. doi: 10.1107/S0907444905036693.

Ferraz Leal, R. M. , Burenkov, G. , Russi, S. , and Popov, A. N. . 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. ISSN 0909-0495. doi: 10.1107/S0909049512049114.

Ferraz Leal, R. M. , Burenkov, G. , Russi, S. , and Popov, A. N. . 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, 2013. ISSN 0909-0495. doi: 10.1107/S0909049512049114.

Fioravanti, E. , Vellieux, F. M. D. , Amara, P. , Madern, D. , and Weik, M. . Specific radiation damage to acidic residues and its relation to their chemical and structural environment. *Journal of Synchrotron Radiation*, 14(1):84–91, 2007. doi: 10.1107/S0909049506038623.

Fischer, H. , Polikarpov, I. , and Craievich, A. F. . Average protein density is a molecular-weight-dependent function. *Protein science*, 13(10):2825–8, 2004. ISSN 0961-8368.

Fletterick, R. , Sygusch, J. , Murray, N. , Madsen, N. , and Johnson, L. . Low-resolution structure of the glycogen phosphorylase a monomer and comparison with phosphorylase b. *Journal of Molecular Biology*, 103(1):1–13, 1976. ISSN 00222836.

Gabadinho, J. , Beteva, A. , Guijarro, M. , Rey-Bakaikoa, V. , Spruce, D. , Bowler, M. W. , Brockhauser, S. , Flot, D. , Gordon, E. J. , Hall, D. R. , Lavault, B. , McCarthy, A. A. , McCarthy, J. , Mitchell, E. , Monaco, S. , Mueller-Dieckmann, C. , Nurizzo, D. , Ravelli, R. B. G. , Thibault, X. , Walsh, M. A. , Leonard, G. A. , and McSweeney, S. M. . MxCuBE: a synchrotron beamline control environment customized for macromolecular crystallography experiments. *Journal of Synchrotron Radiation*, 17(5):700–7, 2010. ISSN 1600-5775.

Gamma, E. , Helm, R. , Johnson, R. , and Vlissides, J. . *Design patterns : elements of reusable object-oriented software*. Addison Wesley, 1994.

Garman, E. F. . Leaving no element of doubt: analysis of proteins using microPIXE. *Structure*, 7 (12):R291–9, 1999. ISSN 0969-2126.

Garman, E. F. and Schneider, T. R. . Macromolecular Cryocrystallography. *Journal of Applied Crystallography*, 30(3):211–237, 1997. ISSN 0021-8898. doi: 10.1107/S0021889897002677.

Garman, E. F. . Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallographica Section D Biological Crystallography*, 66(Pt 4):339–51, 2010. ISSN 1399-0047. doi: 10.1107/S0907444910008656.

Garman, E. F. and Grime, G. W. . Elemental analysis of proteins by microPIXE. *Progress in Biophysics and Molecular Biology*, 89(2):173–205, 2005. ISSN 0079-6107. doi: 10.1016/j.pbiomolbio.2004.09.005.

Gonzalez, A. and Nave, C. . Radiation damage in protein crystals at low temperature. *Acta crystallographica. Section D, Biological Crystallography*, 50(Pt 6):874–7, 1994. ISSN 0907-4449.

Gonzalez, A. , Thompson, A. W. , and Nave, C. . Cryo-protection of protein crystals in intense x-ray beams. *Review of Scientific Instruments*, 63(1):1177, 1992. ISSN 00346748.

Grime, G. . The " Q factor" method: quantitative microPIXE analysis using RBS normalisation. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 109-110:170–174, 1996. ISSN 0168583X. doi: 10.1016/0168-583X(95)00901-9.

Grime, G. W. and Dawson, M. . A PC-based data acquisition package for nuclear microbeam systems. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 89(1-4):223–228, 1994. ISSN 0168583X.

Haas, D. J. and Rossmann, M. G. . Crystallographic studies on lactate dehydrogenase at −75°C. *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry*, 26(7):998–1004, 1970. ISSN 05677408.

Henderson, R. . Cryo-Protection of Protein Crystals against Radiation Damage in Electron and X-Ray Diffraction. *Proc. R. Soc. Lond. B 1990*, 1990. doi: 10.1098/rspb.1990.0057.

Hendrickson, W. A. and Ogata, C. M. . Phase determination from multiwavelength anomalous diffraction measurements. *Methods in Enzymology*, 276(1993), 1997.

Hendrickson, W. A. . Radiation damage in protein crystallography. *Journal of Molecular Biology*, 106(3):889–893, 1976. ISSN 00222836.

Hendrickson, W. A. , Love, W. E. , and Karle, J. . Crystal structure analysis of sea lamprey hemoglobin at 2 Åresolution. *Journal of Molecular Biology*, 74(3):331–361, 1973. ISSN 00222836.

Hilgart, M. C. , Sanishvili, R. , Ogata, C. M. , Becker, M. , Venugopalan, N. , Stepanov, S. , Makarov, O. , Smith, J. L. , and Fischetti, R. F. . Automated sample-scanning methods for radiation damage mitigation and diffraction-based centering of macromolecular crystals. *Journal of Synchrotron Radiation*, 18(Pt 5):717–22, 2011. ISSN 1600-5775. doi: 10.1107/S0909049511029918.

Holton, J. M. . A beginner's guide to radiation damage. *Journal of Synchrotron Radiation*, 16(Pt 2):133–42, 2009. ISSN 0909-0495. doi: 10.1107/S0909049509004361.

Holton, J. M. and Frankel, K. . The minimum crystal size needed for a complete diffraction data set. *Acta Crystallographica Section D Biological Crystallography*, 66(Pt 4):393–408, 2010. ISSN 1399-0047. doi: 10.1107/S0907444910007262.

Hope, H. . Cryocrystallography of biological macromolecules: a generally applicable method, February 1988. ISSN urn:issn:0108-7681.

Howells, M. R. , Beetz, T. , Chapman, H. N. , Cui, C. , Holton, J. M. , Jacobsen, C. J. , Kirz, J. , Lima, E. , Marchesini, S. , Miao, H. , Sayre, D. , Shapiro, D. a. , Spence, J. C. H. , and Starodub, D. . An assessment of the resolution limitation due to radiation-damage in x-ray diffraction microscopy. *Journal of electron spectroscopy and related phenomena*, 170(1-3): 4–12, 2009. ISSN 1873-2526. doi: 10.1016/j.elspec.2008.10.008.

Ishii, K. . Continuous X-rays produced in light-ion–atom collisions. *Radiation Physics and Chemistry*, 75(10):1135–1163, 2006. ISSN 0969806X.

Kabsch, W. . Xds. *Acta crystallographica. Section D, Biological Crystallography*, 66(Pt 2):125–32, 2010. ISSN 1399-0047. doi: 10.1107/S0907444909047337.

Kahn, R. , Fourme, R. , Bosshard, R. , Chiadmi, M. , Risler, J. , Dideberg, O. , and Wery, J. . Crystal structure study of Opsanus tau parvalbumin by multiwavelength anomalous diffraction. *FEBS Letters*, 179(1):133–137, 1985. ISSN 00145793.

Khan, I. , Gillilan, R. , Kriksunov, I. , Williams, R. , Zipfel, W. R. , and Englich, U. . Confocal microscopy on the beamline: novel three-dimensional imaging and sample positioning. *Journal of applied crystallography*, 45(Pt 5):936–943, 2012a. ISSN 0021-8898. doi: 10.1107/S002188981203470X.

Khan, I. , Gillilan, R. , Kriksunov, I. , Williams, R. , Zipfel, W. R. , and Englich, U. . Confocal microscopy on the beamline: novel three-dimensional imaging and sample positioning. *Journal of Applied Crystallography*, 45(5):936–943, 2012b. ISSN 0021-8898. doi: 10.1107/S002188981203470X.

Kmetko, J. , Husseini, N. S. , Naides, M. , Kalinin, Y. , and Thorne, R. E. . Quantifying X-ray radiation damage in protein crystals at cryogenic temperatures. *Acta Crystallographica Section D Biological Crystallography*, 62(Pt 9):1030–8, 2006. ISSN 0907-4449. doi: 10.1107/S0907444906023869.

Kmetko, J. , Warkentin, M. , Englich, U. , and Thorne, R. E. . Can radiation damage to protein crystals be reduced using small-molecule compounds? *Acta Crystallographica Section D Biological Crystallography*, 67(10):881–893, 2011. ISSN 0907-4449. doi: 10.1107/S0907444911032835.

Krojer, T. and von Delft, F. . 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(Pt 3):387–97, 2011. ISSN 1600-5775. doi: 10.1107/S0909049511008235.

Krojer, T. , Pike, A. C. W. , and von Delft, F. . Squeezing the most from every crystal: the fine details of data collection. *Acta Crystallographica Section D Biological Crystallography*, 69(7): 1303–1313, 2013. ISSN 0907-4449. doi: 10.1107/S0907444913013280.

Leal, R. M. F. , Bourenkov, G. P. , Svensson, O. , Spruce, D. , Guijarro, M. , and Popov, A. N. . Experimental procedure for the characterization of radiation damage in macromolecular crystals. *Journal of Synchrotron Radiation*, 18(Pt 3):381–6, 2011. ISSN 1600-5775. doi: 10.1107/S0909049511002251.

Leiros, H. K. S. , Timmins, J. , Ravelli, R. B. G. , and McSweeney, S. M. . Is radiation damage dependent on the dose rate used during macromolecular crystallography data collection? *Acta crystallographica. Section D, Biological Crystallography*, 62(Pt 2):125–32, 2006. ISSN 0907-4449. doi: 10.1107/S0907444905033627.

Leiros, H.-K. S. d. , McSweeney, S. M. , and Smalå s, A. O. . Atomic resolution structures of trypsin provide insight into structural radiation damage. *Acta Crystallographica Section D Biological Crystallography*, 57(4):488–497, 2001. ISSN 0907-4449. doi: 10.1107/S0907444901000646.

Leslie, A. and Powell, H. . Processing diffraction data with mosflm. In Read, R. and Sussman, J. , editors, *Evolving Methods for Macromolecular Crystallography*, volume 245 of *NATO Science Series*, pages 41–51. Springer Netherlands, 2007. ISBN 978-1-4020-6314-5. doi: 10.1007/978-1-4020-6316-9_4.

Lomb, L. , Barends, T. , Kassemeyer, S. , Aquila, A. , Epp, S. , Erk, B. , Foucar, L. , Hartmann, R. , Rudek, B. , Rolles, D. , Rudenko, A. , Shoeman, R. , Andreasson, J. , Bajt, S. , Barthelmess, M. , Barty, A. , Bogan, M. J. , Bostedt, C. , Bozek, J. , Caleman, C. , Coffee, R. , Coppola, N. , DePonte, D. , Doak, R. B. , Ekeberg, T. , Fleckenstein, H. , Fromme, P. , Gebhardt, M. , Graafsma, H. , Gumprecht, L. , Hampton, C. , Hartmann, A. , Hauser, G. , Hirsemann, H. , Holl, P. , Holton, J. M. , Hunter, M. , Kabsch, W. , Kimmel, N. , Kirian, R. , Liang, M. , Maia, F. R. N. , Meinhart, A. , Marchesini, S. , Martin, A. , Nass, K. , Reich, C. , Schulz, J. , Seibert, M. M. , Sierra, R. , Soltau, H. , Spence, J. C. , Steinbrener, J. , Stellato, F. , Stern, S. , Timneanu, N. , Wang, X. , Weidenspointner, G. , Weierstall, U. , White, T. , Wunderer, C. , Chapman, H. , Ullrich, J. , Strüder, L. , and Schlichting, I. . Radiation damage in protein serial femtosecond crystallography using an x-ray free-electron laser. *Physical Review B*, 84(21):1–6, 2011. ISSN 1098-0121. doi: 10.1103/PhysRevB.84.214111.

Maxwell, J. , Camvell, J. , and Teesdale, W. . The Guelph PIXE software package. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 43(2):218–230, 1989. ISSN 0168583X. doi: 10.1016/0168-583X(89)90042-6.

McCarthy, A. A. , Brockhauser, S. , Nurizzo, D. , Theveneau, P. , Mairs, T. , Spruce, D. , Guijarro, M. , Lesourd, M. , Ravelli, R. B. G. , and McSweeney, S. . A decade of user operation on the macromolecular crystallography MAD beamline ID14-4 at the ESRF. *Journal of Synchrotron Radiation*, 16(Pt 6):803–12, 2009. ISSN 1600-5775.

McCoy, A. J. , Grosse-Kunstleve, R. W. , Adams, P. D. , Winn, M. D. , Storoni, L. C. , and Read, R. J. . Phaser crystallographic software. *Journal of applied crystallography*, 40(Pt 4): 658–674, 2007. ISSN 0021-8898. doi: 10.1107/S0021889807021206.

Meents, A. , Wagner, a. , Schneider, R. , Pradervand, C. , Pohl, E. , and Schulze-Briese, C. . Reduction of X-ray-induced radiation damage of macromolecular crystals by data collection at 15 K: a systematic study. *Acta crystallographica. Section D, Biological Crystallography*, 63 (Pt 3):302–9, 2007. ISSN 0907-4449. doi: 10.1107/S0907444906053261.

Meents, A. , Dittrich, B. , and Gutmann, S. . A new aspect of specific radiation damage: hydrogen abstraction from organic molecules. *Journal of Synchrotron Radiation*, 16(Pt 2):183–90, 2009.

Murray, J. W. and Garman, E. F. . Investigation of possible free-radical scavengers and metrics for radiation damage in protein cryocrystallography. *Journal of Synchrotron Radiation*, pages 347–354, 2002.

Murray, J. W. , Garman, E. F. , and Ravelli, R. B. G. . 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.

Murshudov, G. N. , Skubák, P. , Lebedev, A. a. , Pannu, N. S. , Steiner, R. a. , Nicholls, R. a. , Winn, M. D. , Long, F. , and Vagin, A. a. . REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological Crystallography*, 67(Pt 4): 355–67, 2011. ISSN 1399-0047. doi: 10.1107/S0907444911001314.

Nanao, M. H. , Sheldrick, G. M. , and Ravelli, R. B. G. . Improving radiation-damage substructures for RIP. **Acta crystallographica. Section D, Biological Crystallography**, 61(Pt 9): 1227–37, 2005. ISSN 0907-4449. doi: 10.1107/S0907444905019360.

Nave, C. and Hill, M. . Will reduced radiation damage occur with very small crystals? *Journal of Synchrotron Radiation*, 12(Pt 3):299–303, 2005.

Nave, C. and Garman, E. F. . Towards an understanding of radiation damage in cryocooled macromolecular crystals. *Journal of Synchrotron Radiation*, 12(Pt 3):257–60, 2005. ISSN 0909-0495. doi: 10.1107/S0909049505007132.

Newstead, S. , Drew, D. , Cameron, A. D. , Postis, V. L. G. , Xia, X. , Fowler, P. W. , Ingram, J. C. , Carpenter, E. P. , Sansom, M. S. P. , McPherson, M. J. , Baldwin, S. A. , and Iwata, S. . Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. *The EMBO journal*, 30(2):417–26, 2011. ISSN 1460-2075.

Nowak, E. , Brzuszkiewicz, A. , Dauter, M. , Dauter, Z. , and Rosenbaum, G. . To scavenge or not to scavenge: that is the question. *Acta Crystallographica Section D*, 65(9):1004–1006, 2009. doi: 10.1107/S0907444909026821.

Owen, R. L. , Yorke, B. a. , Gowdy, J. a. , and Pearson, A. R. . Revealing low-dose radiation damage using single-crystal spectroscopy. *Journal of Synchrotron Radiation*, 18(Pt 3):367–73, 2011. ISSN 1600-5775. doi: 10.1107/S0909049511004250.

Owen, R. L. , Rudiño Piñera, E. , and Garman, E. F. . 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):4912–7, 2006. ISSN 0027-8424. doi: 10.1073/pnas.0600973103.

Owen, R. L. , Holton, J. M. , Schulze-Briese, C. , and Garman, E. F. . Determination of X-ray flux using silicon pin diodes. *Journal of Synchrotron Radiation*, 16(Pt 2):143–51, 2009. ISSN 0909-0495. doi: 10.1107/S0909049508040429.

Owen, R. L. , Axford, D. , Nettleship, J. E. , Owens, R. J. , Robinson, J. I. , Morgan, A. W. , Doré, A. S. , Lebon, G. , Tate, C. G. , Fry, E. E. , Ren, J. , Stuart, D. I. , and Evans, G. . Outrunning free radicals in room-temperature macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography*, 68(7):810–818, 2012. ISSN 0907-4449. doi: 10.1107/S0907444912012553.

Paithankar, K. S. and Garman, E. F. . Know your dose: RADDOSE. *Acta Crystallographica Section D Biological Crystallography*, 66(Pt 4):381–8, 2010.

Paithankar, K. S. , Owen, R. L. , and Garman, E. F. . Absorbed dose calculations for macromolecular crystals: improvements to RADDOSE. *Journal of Synchrotron Radiation*, 16(Pt 2): 152–62, 2009.

Popov, A. N. and Bourenkov, G. P. . Choice of data-collection parameters based on statistic modelling. *Acta Crystallographica Section D Biological Crystallography*, 59(7):1145–1153, 2003. ISSN 0907-4449. doi: 10.1107/S0907444903008163.

Pozharski, E. . On the variability of experimental data in macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography*, 68(Pt 9):1077–87, 2012. ISSN 1399-0047. doi: 10.1107/S0907444912020100.

Ramagopal, U. , Dauter, Z. , Thirumuruhan, R. , Fedorov, E. , and Almo, S. C. . Radiation-induced site-specific damage of mercury derivatives: phasing and implications. *Acta Crystallographica. Section D, Biological Crystallography*, 61(Pt 9):1289–98, 2005. ISSN 0907-4449. doi: 10.1107/S0907444905022316.

Ravelli, R. B. G. and McSweeney, S. M. . The 'fingerprint' that X-rays can leave on structures. *Structure (London, England : 1993)*, 8(3):315–28, 2000. ISSN 0969-2126.

Ravelli, R. B. G. , Theveneau, P. , McSweeney, S. , and Caffrey, M. . Unit-cell volume change as a metric of radiation damage in crystals of macromoleculesPresented at the 'Second International Workshop on Radiation Damage to Crystalline Biological Samples' held at Advanced Photon Source, Chicago, USA, in December 2001. *Journal of Synchrotron Radiation*, 9 (6):355–360, 2002. ISSN 0909-0495.

Ravelli, R. B. G. , Leiros, H.-k. S. d. , Pan, B. , Caffrey, M. , Mcsweeney, S. , Horowitz, J. , Cedex, F. G. , Avenue, W. , and Cedex, F.-G. . Specific Radiation Damage Can Be Used to Solve Macromolecular Crystal Structures. *Structure*, 11(03):217–224, 2003.

Riekel, C. , Burghammer, M. , and Schertler, G. . Protein crystallography microdiffraction. *Current opinion in structural biology*, 15(5):556–62, 2005. ISSN 0959-440X. doi: 10.1016/j.sbi.2005.08.013.

Rossmann, M. G. . 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. ISSN 0021-8898.

Sanishvili, R. , Yoder, D. W. , Pothineni, S. B. , Rosenbaum, G. , Xu, S. , Vogt, S. , Stepanov, S. , Makarov, O. a. , Corcoran, S. , Benn, R. , Nagarajan, V. , Smith, J. L. , and Fischetti, R. F. . Radiation damage in protein crystals is reduced with a micron-sized X-ray beam. *Proceedings of the National Academy of Sciences*, 108(15):6127–6132, 2011. ISSN 0027-8424. doi: 10.1073/pnas.1017701108.

Schiltz, M. , Dumas, P. , Ennifar, E. , Flensburg, C. , Paciorek, W. , Vonrhein, C. , and Bricogne, G. . Phasing in the presence of severe site-specific radiation damage through dose-dependent modelling of heavy atoms. *Acta crystallographica. Section D, Biological Crystallography*, 60(Pt 6):1024–31, 2004. ISSN 0907-4449. doi: 10.1107/S0907444904006377.

Shi, W. and Chance, M. R. . Metalloproteomics: forward and reverse approaches in metalloprotein structural and functional characterization. *Current Opinion in Chemical Biology*, 15(1): 144–8, 2011. ISSN 1879-0402. doi: 10.1016/j.cbpa.2010.11.004.

Simon, A. , Jeynes, C. , Webb, R. , Finnis, R. , Tabatabaian, Z. , Sellin, P. , Breese, M. , Fellows, D. , Vandenbroek, R. , and Gwilliam, R. . The new Surrey ion beam analysis facility. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 219-220:405–409, 2004. ISSN 0168583X. doi: 10.1016/j.nimb.2004.01.091.

Sliz, P. , Harrison, S. C. , and Rosenbaum, G. . How does Radiation Damage in Protein Crystals Depend on X-Ray Dose ? Notes from the Bench. *Structure*, 11(02):13–19, 2003.

Solcan, N. , Kwok, J. , Fowler, P. W. , Cameron, A. D. , Drew, D. , Iwata, S. , and Newstead, S. . Alternating access mechanism in the POT family of oligopeptide transporters. *The EMBO journal*, 31(16):3411–21, 2012. ISSN 1460-2075. doi: 10.1038/emboj.2012.157.

Southworth-Davies, R. J. , Medina, M. a. , Carmichael, I. , and Garman, E. F. . Observation of decreased radiation damage at higher dose rates in room temperature protein crystallography. *Structure*, 15(12):1531–41, 2007. ISSN 0969-2126. doi: 10.1016/j.str.2007.10.013.

Sygusch, J. and Allaire, M. . Sequential radiation damage in protein crystallography. *Acta Crystallographica Section A Foundations of Crystallography*, 44(4):443–448, 1988. ISSN 0108-7673.

Teng, T. Y. and Moffat, K. . Primary radiation damage of protein crystals by an intense synchrotron X-ray beam. *Journal of Synchrotron Radiation*, 7(Pt 5):313–7, 2000. ISSN 0909-0495. doi: 10.1107/S0909049500008694.

Teng, T.-Y. and Moffat, K. . Radiation damage of protein crystals at cryogenic temperatures between 40□K and 150□K. *Journal of Synchrotron Radiation*, 9(4):198–201, 2002. ISSN 0909-0495. doi: 10.1107/S0909049502008579.

Wang, B. C. . Resolution of phase ambiguity in macromolecular crystallography. *Methods in enzymology*, 115:90–112, 1985. ISSN 0076-6879.

Warkentin, M. and Thorne, R. E. . Glass transition in thaumatin crystals revealed through temperature-dependent radiation-sensitivity measurements. *Acta Crystallographica Section D Biological Crystallography*, 66(10):1092–1100, 2010. ISSN 0907-4449. doi: 10.1107/S0907444910035523.

Warkentin, M. , Badeau, R. , Hopkins, J. , and Thorne, R. E. . Dark progression reveals slow timescales for radiation damage between T = 180 and 240□K. *Acta Crystallographica Section D Biological Crystallography*, 67(9):792–803, 2011. ISSN 0907-4449. doi: 10.1107/S0907444911027600.

Warkentin, M. , Mulichak, A. M. , Lisa, J. , and Thorne, R. E. . Global radiation damage at 300 and 260 K with dose rates approaching 1 MGy s. *Acta Crystallographica Section D Biological Crystallography*, 68:124–133, 2012. doi: 10.1107/S0907444911052085.

Warkentin, M. , Hopkins, J. B. , Badeau, R. , Mulichak, A. M. , Keefe, L. J. , and Thorne, R. E. . Global radiation damage: temperature dependence, time dependence and how to outrun it. *Journal of Synchrotron Radiation*, 20(1):2007–2013, 2013. ISSN 0909-0495. doi: 10.1107/S0909049512048303.

Weik, M. , Ravelli, R. B. , Silman, I. , Sussman, J. L. , Gros, P. , and Kroon, J. . Specific protein dynamics near the solvent glass transition assayed by radiation-induced structural changes. *Protein science : a publication of the Protein Society*, 10(10):1953–61, 2001. ISSN 0961-8368.

Weik, M. , Ravelli, R. B. G. , Kryger, G. , McSweeney, S. , Raves, M. L. , Harel, M. , Gros, P. , Silman, I. , Kroon, J. , and Sussman, J. L. . Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proceedings of the National Academy of Sciences*, 97(2):623–62, 2000. ISSN 00278424.

Weik, M. , Bergès, J. , Raves, M. L. , Gros, P. , McSweeney, S. , Silman, I. , Sussman, J. L. , Houée-Levin, C. , and Ravelli, R. B. G. . Evidence for the formation of disulfide radicals in protein crystals upon X-ray irradiation. *Journal of Synchrotron Radiation*, 9(Pt 6):342–6, 2002. ISSN 0909-0495.

Winn, M. D. , Ballard, C. C. , Cowtan, K. D. , Dodson, E. J. , Emsley, P. , Evans, P. R. , Keegan, R. M. , Krissinel, E. B. , Leslie, A. G. W. , McCoy, A. , McNicholas, S. J. , Murshudov, G. N. , Pannu, N. S. , Potterton, E. a. , Powell, H. R. , Read, R. J. , Vagin, A. , and Wilson, K. S. . Overview of the CCP4 suite and current developments. *Acta Crystallographica, Section D, Biological Crystallography*, 67(Pt 4):235–42, 2011. ISSN 1399-0047. doi: 10.1107/S0907444910045749.

Ye, H. and Rouault, T. a. . Human iron-sulfur cluster assembly, cellular iron homeostasis, and disease. *Biochemistry*, 49(24):4945–56, 2010. ISSN 1520-4995. doi: 10.1021/bi1004798.

Zeldin, O. B. , Gerstel, M. , and Garman, E. F. . Optimising the spatial distribution of dose in X-ray macromolecular crystallography. *Journal of Synchrotron Radiation*, 20(1):49–57, 2013a. doi: 10.1107/S0909049512044706.

Zeldin, O. B. , Gerstel, M. , and Garman, E. F. . RADDOSE-3D: time- and space-resolved modelling of dose in macromolecular crystallography. *Journal of Applied Crystallography*, 46, 2013b.