Development of Serial Crystallography Methods for Synchrotrons and X-ray Free-Electron Lasers

Ali. A. Ebrahim

A thesis submitted for the degree of PhD in Biochemistry

Department of Life Sciences

University of Essex

February 2020

Abstract

X-ray crystallography of proteins is a well-established method to identify atomic level molecular arrangement, however when exposing crystalline proteins to ionising radiation, damage can occur to their overall molecular structure (global damage), while solvated electrons generated by X-rays can induce changes to metal sites within a protein (site-specific damage) (Garman & Weik, 2015). These changes occur quickly, at doses much lower than required to obtain a complete dataset, meaning there may be many metalloproteins deposited in the Protein Data Bank that are incorrect (Bowman, Bridwell-Rabb, & Drennan, 2016).

The advent of X-ray free-electron lasers (XFELs) that produce femtosecond pulses of extremely high quality (brilliance) X-ray beams, allows data to be collected before radiation damage has the time to occur (Schlichting, 2015). This thesis will describe the development of novel 'chip' based serial data collection and processing strategies, applied at Diamond Light Source microfocus beamline I24, and at BL3 EH2 at the SACLA XFEL, Japan. A technique coined 'multiple serial structures' (MSS) has been developed and used in this thesis to assess how crystalline proteins change as a function of X-ray dose, as enzyme reactivity can be driven in crystals by exploiting X-ray generated solvated electrons to drive redox reactions (Horrell et al., 2016). By performing a near identical data collection strategy at the SACLA XFEL, we have been able to directly compare the effects of accumulated dose in MSS datasets to 'damage free' XFEL structures, using the same target protein. Chip methods have also been examined and developed in this thesis as a tool to assess the 'dark progression' of

radiation damage, a technique we have coined ‘dark progression series’ (DPS). Further, we present a data processing technique that possesses the ability to identify protein-ligand complexes from extremely small subsets of synchrotron and XFEL diffraction data, whereby only a few hundred diffraction images may be needed.

Acknowledgements

First and foremost, I would like to thank my supervisors Dr. Mike Hough and Dr. Robin Owen. Without them none of this would have been possible. Every single day I spend in their presence I learn something new, and will be eternally grateful for the guidance and support they have given me for the entire duration of this PhD project. I have learned so much along the way and have developed the skills to now pursue my own career in science, as something Mike, Robin, and I all have in common is genuine intrigue for crystallography.

Something else we all have in common is our love of Liverpool Football Club. Having spent two out of four of my PhD years at Diamond, one of the tings I'll miss the most (notwithstanding the addition of the biscuit tin, Martin) is arriving on the beamline first thing on a Monday morning; once business it out of the way (or sometimes before its even started) the next logical question is always, "did you watch the game on the weekend?", to which the logical answer is always, "yes". I count myself extremely lucky to have had such immense PhD supervisors, having shared moments of intrigue, such as developing the DPS method, and moments that have made us cry with laughter (see Brad and his umbrella vs a bush, 2019). I'd like to give special mention to Dr. Danny Axford, whom during my two years spent at Diamond has been all but a third supervisor. He has always gone out of his way to help me learn new methods and techniques of data processing, as well as crystallographic theory. All three of you are an absolute wealth of knowledge, and I feel privileged to have spent my formative crystallographic years working with you.

I would like to thank all of the people at the University of Essex and Diamond I've met along the way. Dr. Jonathan Worrall and Dr. Richard Strange at Essex were both insightful and hilarious; beamtimes are always better with you both around (be that at Diamond, or on the other side of the world!). Dr. Darren Sherrell, Martin Appleby, and Brad Davy, all of whom I met at Diamond were all legends in their own way, and testament to the amazing recruitment on I24. Brian, sharing an office with you was a highlight of my time there, you're sorely missed by so many people. There are so many more of you to name, you know who you are.

Finally, my friends and family. I'm genuinely so lucky to have such a group of tight knit individuals; the Orangeballs, Mr Blobby and Friends, Lois & Jasper, and all of the day ones, you're my favourite people. Omar, Fatima, the Norfolk gang, Aunty Sue, and little Wafa, you all mean so much to me and inspire me every day. Dad, you're actually legendary – like, I genuinely think they'll write stories about you in a hundred years' time. Young man, you're one in a billion.

Associated publications

The following is a comprehensive list of publications associated with the work presented in this thesis:

- A subset of the work presented in chapter 4 has been published in Ebrahim, Appleby, et al., (2019).
- A subset of the work presented in chapter 5 has been published in Ebrahim, Moreno-Chicano, et al., (2019).
- A subset of the work presented in chapter 7 has been published in Moreno-Chicano, Ebrahim et al., (2019).
- Data collection and processing methods developed in this thesis were used for a subset of data presented in Davy et al., (2019), with methods used to develop protein crystal samples published in Beale et al., (2019).
- Work presented in chapter 4 and 5 has been presented by the author in a short oral presentation at the British Crystallographic Association Spring Meeting 2019, at the University of Nottingham.

Each of the aforementioned publications can be found in the associated publications section of this thesis.

Author's declaration

This thesis has been written by myself and has not been submitted in any pervious application for any degree. The work in this thesis has been carried out by myself, unless stated otherwise. This thesis is written in accordance with the regulations for the degree of Doctor of Philosophy at the University of Essex.

……

Ali A. Ebrahim
February 2020

Table of contents

Abstract *1*

Acknowledgements *3*

Associated publications *5*

Author's declaration *7*

Table of figures *13*

Table list *17*

Abbreviations *19*

Chapter 1 : Literature review *21*

1.1. Radiation damage in X-ray crystallography **21**

1.1.1. Radiation damage theory 21

1.1.2. Limiting radiation damage 30

1.1.3. Room temperature vs cryo 31

1.2. Data collection methods to reduce radiation damage **32**

1.2.1. Synchrotron methods 33

1.2.2. Serial femtosecond crystallography 35

1.2.3. Multi-crystal delivery methods 37

1.3. Metalloproteins of interest **41**

1.3.1. Achromobacter cycloclastes nitrite reductase 42

1.3.2. DyP-type peroxidase Aa 47

1.4. Aims **51**

Chapter 2 : Crystallographic theory *53*

2.1. Macromolecular crystallisation **53**

2.2. Crystal properties **54**

2.2. Bragg's law **57**

2.3. The Ewald sphere **59**

2.4. Structure factors **59**

2.5. Fourier transform **62**

2.6. Data reduction **62**

2.7. Data quality **63**

Chapter 3 : Materials and methods *67*

3.1. Protein expression and purification **67**

3.1.1. AcNiR 67

3.1.2. DtpAa 68

3.2. Crystallisation **68**

3.2.1. AcNiR 68

3.2.2. DtpAa 70

3.3. Sample preparation 70
3.3.1. Fixed target 'chips' 70
3.3.2. Fixed target loading 73

3.4. Fixed target instrumentation and data collection 76
3.4.1. Sample stages 76
3.4.2. OAV and chip alignment 78
3.4.3. Serial data collection 80

3.5. Data processing and refinement strategies 87
3.5.1. Indexing and integrating SSX data 87
3.5.1. Joint refinement 88
3.5.2. Merging SSX data 89
3.5.3. Indexing and integrating SFX data 91
3.5.4. Merging SFX data 92
3.5.5. Structure solution and refinement 94

3.6. Online UV vis spectrophotometry 95

3.7. Appendices 96
3.7.1. Script process.sh 96
3.7.2. Script process.phil 96
3.7.3. Script refine.phil 96
3.7.4. Script prime.phil 97
3.7.5. Script get_int_file.py 98
3.7.5. Script create-mtz 100

Chapter 4 : New methods for resolving radiation-driven effects in microcrystals *101*

4.1 Introduction 101

4.2 Methods 105
4.2.1 AcNiR purification and microcrystal optimisation 105
4.2.2 Data collection 107

4.3 Results 109
4.3.1 AcNiR purification and microcrystal optimisation 109
4.3.2 MSS data collection, processing, and discrimination between crystal polymorphs 112
4.3.3 Changes to polymorph populations with X-ray dose 122
4.3.4 Influence of dilution on unit cell population 126
4.3.5 AcNiR MSS structures 126

4.4. Discussion 135
4.4.1. Data and structure quality indicators 135
4.4.2. Structural comparison 136
4.4.3. Tracking global radiation damage in MSS data sets 142

4.5. Appendices 145
4.5.1. Example stills_process input parameters 145
4.5.2. Example prime input parameters 146
4.5.1. RADDOSE3D input 148

Chapter 5 : Estimating zero dose structures: SFX vs SSX *151*

5.1. Introduction 151

5.2. Methods 155
5.2.1. Recombinant protein expression, purification, and crystallisation 155
5.2.2. Serial data collection 157

5.2.3. Online UV-vis micro-spectrophotometry 161

5.3. Results 163

5.3.1. SFX Data collection, processing, and refinement 163
5.3.2. DtpAa SFX structure 169
5.3.3. SSX data collection, processing, and refinement 175
5.3.4. DtpAa SSX structures 181
5.3.5. Characterising photoreduction in DtpAa using online UV-vis microspectrophotometry 186

5.4. Discussion 191

5.5. Appendices 198

5.5.1. MSS1 RADDOSE-3D data 198
5.5.2. MSS2 RADDOSE-3D data 200
5.5.3. MSS1-ds1 resolution bin statistics 202
5.5.4. MSS2-ds1 resolution bin statistics 204
5.5.5. UV-vis RADDOSE-3D data 206

Chapter 6 : Dark progression of radiation damage in MSS datasets *209*

6.1. Introduction 209

6.2. Methods 211

6.2.1. Recombinant protein expression, purification, and crystallisation 211
6.2.2. Serial data collection 211

6.3. Results 214

6.3.1. DPS1 SSX data collection, processing, and refinement 215
6.3.2. DtpAa DPS1 structures 220
6.3.3. DPS2 SSX data collection, processing, and refinement 224
6.3.4. DtpAa DPS2 structures 228

6.4. Discussion 232

6.5. Appendices 240

6.5.1. DPS1 RADDOSE 240
6.5.2. DPS2 RADDOSE 242

Chapter 7 : Identification of protein-ligand complexes via SFX and SSX *245*

7.1. Introduction 245

7.2 Methods 251

7.2.1. Protein expression, purification, and crystallisation 251
7.2.2. Serial data collection 251
7.2.3. Generating data subsets 253
7.2.4. Ligand modelling 254

7.3. Results 256

7.3.1. SFX data collection and processing 256
7.3.2. SSX data collection and processing 265
7.3.3. SFX and SSX structures of ligand bound complexes 268
7.3.4. SFX and SSX data subset generation, merging, and refinement 271

7.4. Discussion 281

7.5. Appendices 291

7.5.1. SACLA counter script 291
7.5.2. Pilatus detector geometry 293
7.5.3. SSX silicon spot statistics 294

Chapter 8 : Closing remarks ***295***

Bibliography ***299***

Associated publications ***311***

Table of figures

Figure 1.1. Typical synchrotron endstation setup at beamline I24 Diamond Light Source 23

Figure 1.2. The different primary X-ray scattering processes..... 24

Figure 1.3. Global radiation damage as a loss of diffraction intensity 26

Figure 1.4. Schematic for the decomposition of water leading to the formation of electrons..... 28

Figure 1.5. Schematic representation of the parameters believed to be relevant to radiation damage progression. 29

Figure 1.6. SACLA XFEL, Hyogo, Japan..... 36

Figure 1.7. Liquid jet sample delivery 38

Figure 1.8 Fixed target 'chip' sample delivery method 40

Figure 1.9. Overall and active site structure of AcNiR..... 43

Figure 1.10. Catalytic mechanism of AcNiR 45

Figure 1.11. AcNiR ligand conformations at 190 K and room temperature 46

Figure 1.12. Overall and active site structure of DyP DtpA 48

Figure 1.13. The oxidase/peroxidase cycle of heme peroxidases 50

Figure 2.1. Phase diagram of crystallisation trajectories (Beale et al., 2019)..... 55

Figure 2.2. Unit cell properties of a crystal 56

Figure 2.3. Crystallographic systems of symmetry 58

Figure 2.4. Bragg's law 60

Figure 2.5.The Ewald Sphere..... 61

Figure 3.1 Oxford chip schematics..... 71

Figure 3.2. Sample loading instruments and environment. 74

Figure 3.3. Sample holder and chip cleaning components..... 75

Figure 3.4. SmarAct xyz sample stages..... 77

Figure 3.5. Chip alignment GUI. 79

Figure 3.6. Modified I24 endstation for serial chip experiments 82

Figure 3.7. Formation of dose-resolved datasets 83

Figure 3.8. SACLA BL2 EH3 setup..... 85

Figure 4.1. AcNiR SDS PAGE *110*

Figure 4.2. AcNiR microcrystals *111*

Figure 4.3. AcNiR SSX diffraction *113*

Figure 4.4. AcNiR unit cell distribution for MSS-ds1 crystal subset *115*

Figure 4.5. Number of integrated images used for MSS datasets *118*

Figure 4.6. Decrease in normalised diffracting power as a function of accumulated dose. *124*

Figure 4.7. Change in polymorph unit cell as a function of accumulated dose *125*

Figure 4.8. Unit cell histograms for dilution series *127*

Figure 4.9. Fraction of polymorph as a function of microcrystal suspension concentration. *128*

Figure 4.10. Overall and active site structure of MSS-ds1 all data *130*

Figure 4.11. Distance between copper centres in MSS-ds1 all data *132*

Figure 4.12. MSS-ds1 all data and polymorph 187-193 loop regions *133*

Figure 4.13. X-ray induced active site change in AcNiR *134*

Figure 4.14. PDB percentile ranks of the MSS-ds1 structures *137*

Figure 4.15. Superimposed small and large cell AcNiR 11 kGy structures *138*

Figure 4.16. Superposed AcNiR loop conformations at different temperatures *140*

Figure 5.1. DtpAa microcrystals. *158*

Figure 5.2. Online UV-vis spectrophotometry on I24 ..... *162*

Figure 5.3. DtpAa SFX diffraction *165*

Figure 5.4. DtpAa unit cell distribution showing the refined DtpAa unit cell *166*

Figure 5.5. Detector shift plot. *168*

Figure 5.6. Overall structure of the ferric DtpAa SFX structure. *173*

Figure 5.7. Ferric SFX DtpAa active site water network. *174*

Figure 5.8. Ferric SFX DtpAa heme B *176*

Figure 5.9. DtpAa SSX diffraction from MSS1-ds1 ..... *178*

Figure 5.10. Lengthening of Fe-O bond length as a function of accumulated dose. *185*

Figure 5.11. Superposition of SFX and 313.6 kGy MSS structures heme propionate environment *187*

Figure 5.12. Unit cell volume vs. absorbed dose for MSS1 and MSS2 datasets *188*

Figure 5.13. Single crystal UV-vis spectra of X-ray photoreduction of DtpAa *189*

Figure 5.14. Single crystal UV-vis spectra of X-ray photoreduction of DtpAa at 560 nm.....190

Figure 5.15. R-factor distribution and PDB percentile ranks of the SFX and MSS1-ds1 structures.....193

Figure 5.16. Plot of Fe-O distance as a function of X-ray dose from MSS1 and MSS2.....195

Figure 6.1. Schematic representation of the DPS experiment.....213

Figure 6.2. DtpAa unit cell distribution. 216

Figure 6.3. DtpAa DPS1-ds1 active site water network.....221

Figure 6.4. 90 second time delay Fe-O bond length as a function of accumulated dose. 223

Figure 6.5. DPS1 omit maps. 224

Figure 6.6. Datasets DPS2-ds1-5 Fe-O bond length as a function of accumulated dose.....230

Figure 6.7. Datasets DPS2-ds6-10 Fe-O bond length as a function of accumulated dose.....231

Figure 6.8. DPS2 omit maps. 232

Figure 6.9. Plot of DPS2 Fe-O distance as a function of diffraction weighted dose and time.....234

Figure 6.10. Plot of Fe-O distance as a function of X-ray dose from MSS1, MSS2, and DPS2.....236

Figure 6.11. Normalised global radiation damage indicators vs absorbed dose for DPS2 datasets....238

Figure 7.1. Effects of reducing sample consumption on diffraction hits and crystal orientation using acoustic drop ejection.....249

Figure 7.2. AcNiR SFX diffraction.....258

Figure 7.3. SFX native and ligand bound AcNiR unit cell distributions.....260

Figure 7.4. Detector shift plot. 262

Figure 7.5. AcNiR + NO_2^- SSX diffraction 266

Figure 7.6. SSX ligand bound AcNiR unit cell distribution 267

Figure 7.7. Overall fold and active site regions of native and nitrite bound SFX structures.....270

Figure 7.8. Fo-Fc simulated annealing omit maps from selected data subsets of the SFX AcNiR-nitrite structure. 274

Figure 7.9. Fo-Fc simulated annealing omit maps from selected data subsets of the SSX AcNiR-nitrite structure. 277

Figure 7.10. Fo-Fc data subset omit maps from SFX AcNiR NO_2^- structures versus SFX native AcNiR with the T2Cu bound water molecule included in the model.....279

Figure 7.11. F_o-F_c data subset omit map from SFX AcNiR NO_2^- structures versus SFX native AcNiR with the T2Cu bound water removed......280

Figure 7.12. data quality metrics as a function of the number of merged crystals in data subsets for SFX and SSX data.....283

Figure 7.13. Plot of point group 2, 23, and 222 completeness data from randomly oriented images (Fry et al., 1999)......286

Figure 7.14. Reciprocal lattice plots for the smallest SFX and SSX subsets. 288

Table list

Table 4.1. AcNiR batch microcrystal optimisation *106*

Table 4.2. Number of integrated images used for MSS datasets *117*

Table 4.3. MSS-ds1 all cell statistics *119*

Table 4.4. MSS-ds1 small cell statistics *120*

Table 4.5. MSS-ds1 large cell statistics *121*

Table 4.6. Bragg intensities and normalised Bragg intensities for dose series polymorphs *123*

Table 4.7. Data processing and refinement values for MSS-ds1 datasets *129*

Table 5.1. Chips used to collect ferric DtpAa SFX data *164*

Table 5.2. $CC_{1/2}$ *and* R_{split} *statistics for DtpAa SFX dataset* *170*

Table 5.3. Resolution shell statistics for DtpAa SFX dataset *171*

Table 5.4. DtpAa SFX structure data collection, processing, and refinement statistics. *172*

Table 5.5. Dose increments and number of used integrated files for DtpAa MSS datasets *180*

Table 5.6. Data collection, refinement, and processing statistics for ferric DtpAa SFX and MSS1 datasets *182*

Table 5.7. Data collection, refinement, and processing statistics for MSS2 datasets used for structural comparison *183*

Table 5.8a. Comparison of Fe-His and Fe-O bond lengths in the SFX and MSS1 series *184*

Table 5.8b. Comparison of Fe-His and Fe-O bond lengths in the SFX and MSS2 series *184*

Table 6.1. DPS1 dataset collection and data processing parameters *218*

Table 6.2. Data collection, processing, refinement, and DPI statistics for DPS1-ds1-3 ..... *219*

Table 6.3. Comparison of Fe-O bond lengths in DPS1 datasets *222*

Table 6.4. DPS2 dataset collection and data processing parameters *226*

Table 6.5. Data collection, processing, refinement, and DPI statistics for DPS2-ds1-10 ..... *227*

Table 6.6. Comparison of Fe-O bond lengths in DPS2 datasets *229*

Table 7.1. Chips used to collect AcNiR SFX data with initial approximate hit rate measured by `CHEETAH` *(Barty et al., 2014).* *257*

Table 7.2. Native and NO_2^- bound AcNiR SFX structure data collection, processing, and refinement statistics. *264*

Table 7.3. AcNiR + NO_2^- SSX structure data collection, processing, and refinement statistics. *269*

Table 7.4. AcNiR + NO_2^- SFX subset structure processing and refinement statistics. *273*

Table 7.5. AcNiR + NO_2^- SSX subset structure processing and refinement statistics. *276*

Table 7.6. Data from randomly oriented images of different point group virus crystals (Fry et al., 1999). *285*

Abbreviations

AcNiR – *Achromobacter cycloclastes* nitrite reductase

$AmSO_4$ – Ammonium sulphate

Arg – Arginine

Asp – Aspartic acid

BL – Beamline

Cu – Copper

$CuSO_4$ – Copper sulphate

Cys – Cysteine

DEAE – Diethylaminomethane

DLS – Diamond Light Source

DPS – Dark progression series

DtpA – DyP-type peroxidase A from *Streptomyces lividans*

DtpAa – DyP-type peroxidase Aa from *Streptomyces lividans*

DWD – Diffraction weighted dose

EH – Experimental hutch

FBDD – Fragment based drug design

Fe – Iron

FWHM – Full width half maximum

Glu – Glutamic acid

Gly – Glycine

GUI – Graphical user interface

Gy – Gray

H_2O – Water

HCl – Hydrochloric acid

HEPES – (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)

His – Histidine

HRP – Horseradish peroxidase

IPTG – Isopropyl β-D-1-thiogalactopyranoside

Kan – Kanamycin

kDa – Kilodalton

LB – Lysogeny broth

MSOX – Multiple structures serially obtained from one crystal

MSS – Multiple serial structures

MWCO – Molecular weight cut-off

MX – Macromolecular crystallography

NaCl – Sodium chloride

NO – Nitric oxide

NO_2^- – Nitrite

O – Oxygen

OAV – On axis viewing

OD – Optical density

PDB – Protein Data Bank

PEG – Polyethylene glycol

RMSD – Root mean square deviation

RSZD – Real-space difference density Z-score

RT – Room temperature

SACLA – SPring-8 angstrom compact free electron laser

SDS-PAGE – Sodium dodecyl sulphate polyacrylamide gel electrophoresis

SFX – Serial femtosecond crystallography

SMX – Serial millisecond crystallography

SSX – Serial synchrotron crystallography

T1Cu – Type 1 copper centre

T2Cu – Type 2 copper centre

Tris – tris(hydroxymethyl)aminomethane

XFEL – X-ray free-electron laser

Chapter 1: Literature review

1.1. Radiation damage in X-ray crystallography

The utilisation of X-ray crystallography to understand the atomic level structure of metalloproteins is well-established, and plays a major role in the field of structural biology. In this, X-ray crystallography of metalloproteins provides insight into key biological processes, allowing us to understand how proteins interact with one another, substrates, and drug targets to name a few. The effects of radiation damage are seldom absent in X-ray crystal structures however, with the global effects such as a loss of diffraction spot intensity seen initially in reciprocal space, and site-specific damage subsequently seen in real space models of refined protein structures (Garman & Weik, 2015b; Holton, 2009). Data collection methods have been developed to limit and examine these effects, with the advent of 'damage free' structures collected at X-ray free-electron lasers and their associated methods providing a new outlook on how best crystallographic data should be collected to limit radiation damage. This chapter will characterise radiation damage in terms of theory and its effects, as-well-as examine instances where radiation damage has been studied, the methods used to do this, recent developments in the field, and the background of the radiation sensitive metalloproteins chosen for study in this PhD project.

1.1.1. Radiation damage theory

One of the most common causes of poor-quality data from the X-ray crystallography experiment is the effect of radiation damage on a macromolecular protein crystal

(Garman & Weik, 2015b; Holton, 2009). Energy deposited in the crystal during data collection can lead to chemical transformations that compromise the quality of the final electron density maps by decreasing the intensity and resolution of X-ray diffraction patterns (Zeldin et al., 2013).

In the classical X-ray crystallography experiment, a protein crystal is held in an X-ray beam in a polymer loop on a pin, with the base of the pin magnetically attached to a goniometer allowing the crystal to be rotated in the beam when exposed to X-rays (fig. 1.1). When a protein crystal is irradiated by an X-ray beam, only a small fraction of the radiation will interact with the crystal, with the rest passing straight through it. For example, a 100 µm thick metal-free crystal irradiated with a 13.0 keV X-ray beam (λ of ~0.95Å) will only interact with approximately 3% of the radiation that passes through it; ~6.5% of the interacting X-rays giving rise to diffracted X-rays, whereby an incoming X-ray is elastically scattered with the X-ray wavelength remaining unchanged (Rayleigh scattering) (Garman & Owen, 2006) (fig. 1.2).

The reminder of the radiation is deposited within the crystal and takes two forms: Inelastic (Compton) scattering accounts for ~6.6% of the interacting radiation, where the X-ray scatters from an atomic electron, losing part of its energy to the electron which may then ejected from the atom. However, the majority, ~86.9%, of the energy deposition occurs through the photoelectric effect, whereby the interacting photon transfers all of its energy to an atomic electron; this high-energy photoelectron is then ejected from the atom, leaving a vacant site which is filled with a higher shell electron,

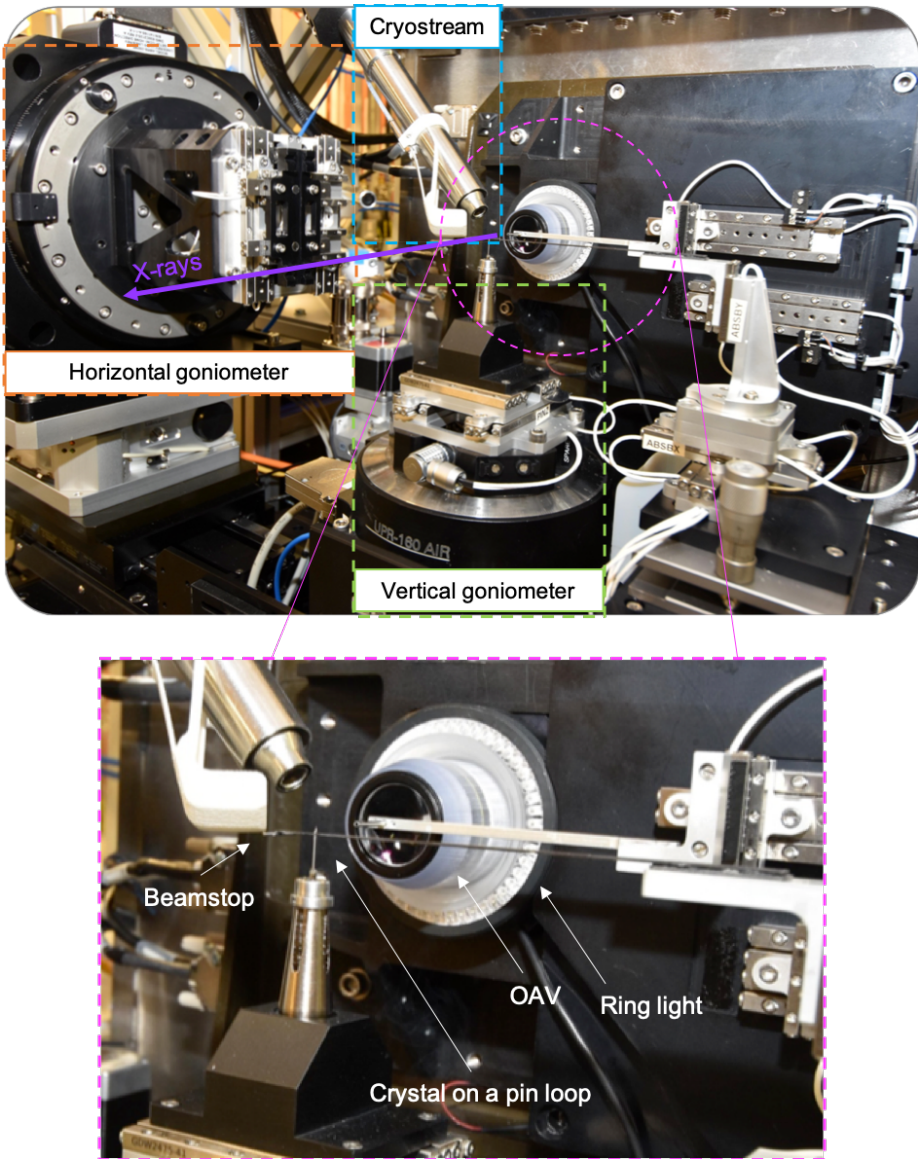


Figure 1.1. Typical synchrotron endstation setup at beamline I24 Diamond Light Source

The endstation setup for the collection of X-ray diffraction from a crystal mounted on a loop on a pin at I24 utilises a vertical goniometer to rotate a crystal during exposure to the X-ray beam, while the horizontal goniometer can be used for collecting data from crystals in trays (in-situ). A cryostream can be used to keep crystals cool in order to delay the onset of radiation damage effects. Samples are viewed using an on-axis viewing camera (OAV) illuminated by a ring light and a backlight (backlight not shown). The beamstop prevents intense, non-diffracted X-rays from damaging the detector.

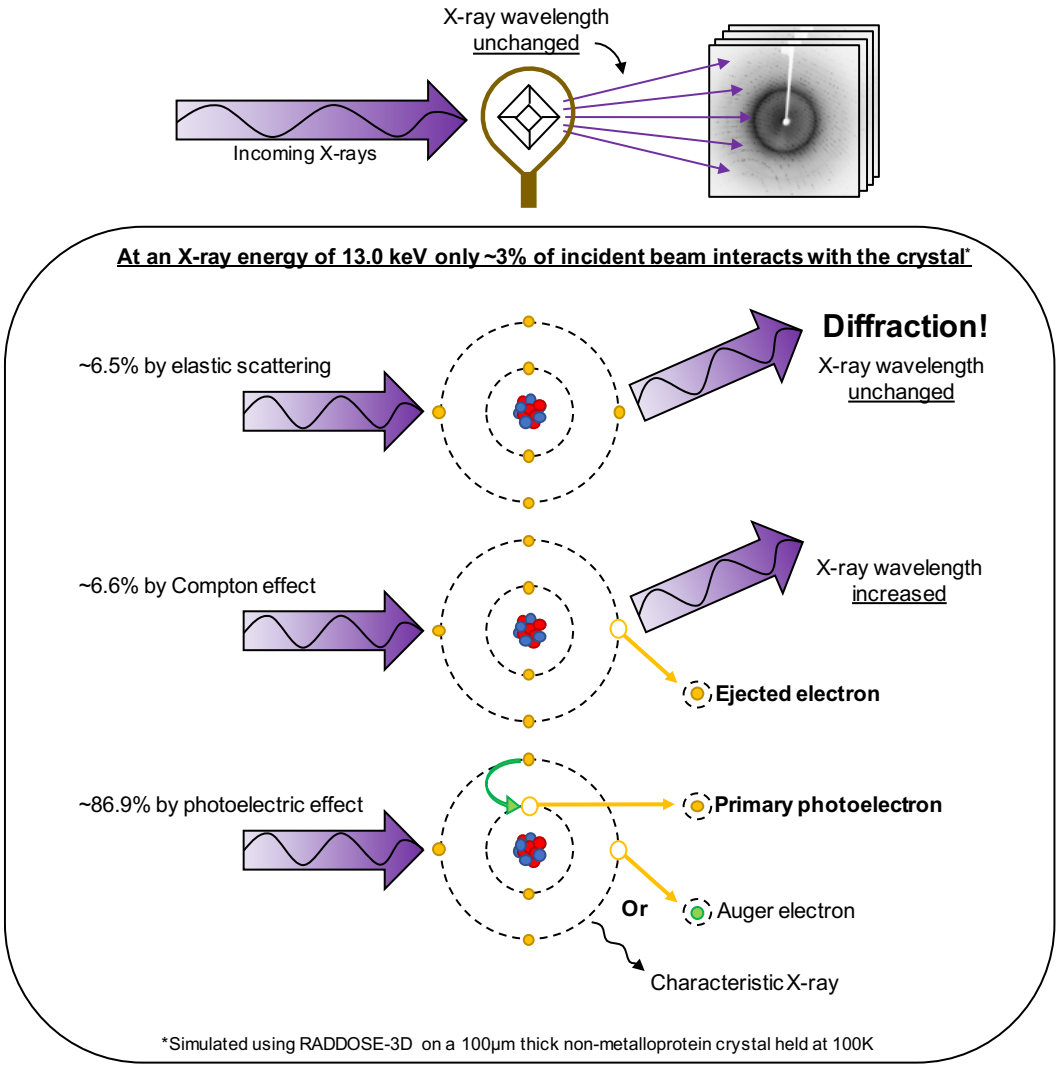


Figure 1.2. The different primary X-ray scattering processes

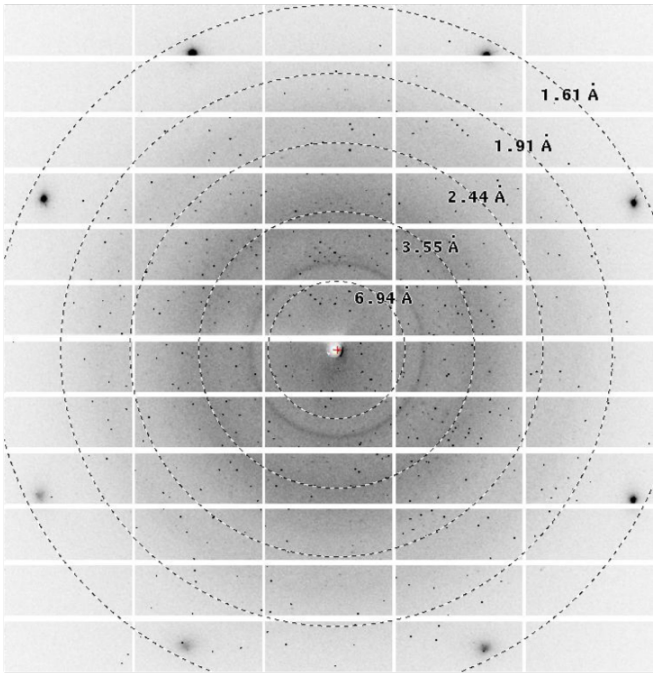
Elastic scattering: ~6.5% of the interacting beam, resulting in the X-ray photon being scattered, causing in diffraction. Compton scattering: ~6.6% of the interacting beam, where the photon loses energy in an atomic electron, resulting in scattering at a longer wavelength (an electron may then be ejected from the atom). Photoelectric absorption: 86.9% of the interacting beam, resulting in the photon transferring all energy to an inner shell electron, which is subsequently ejected from the atom, known as a photoelectron. The resulting orbital vacancy is filled by a higher shell electron, and followed by either fluorescence emission or ejection of a lower energy Auger electron (figure adapted from Taberman, 2018).

simultaneously releasing excess energy in the form of X-ray fluorescence or of a lower energy Auger electron (Garman, 2010; Taberman, 2018) (fig 1.2).

When incident X-ray radiation interacts with an atom in the crystal, the ionisation owing to photoelectric absorption or Compton scattering are referred to as primary radiation damage. Secondary radiation damage effects are those that arise from the formation of up to 500 low-energy secondary electrons per primary absorption event (Garman, 2010). Radiation damage effects manifesting as the overall loss in diffraction intensity, expansion of unit-cell volume, increase in scaling B-factors, rotation of the molecule within the unit cell, and often (but not always) an increase in mosaicity, are known as global effects of radiation damage (fig. 1.3) (Garman, 2010; Garman & Weik, 2015a; Holton, 2009). Global damage is observable in reciprocal space and visible in the crystal as irradiation proceeds as a colour change. Site specific radiation damage is the damage to specific sites and structures within the atomic structure, arising from the formation of photoelectrons which are able to diffuse and induce ~500 further ionisation and excitation events, and is observable in real space (Southworth-Davies et al. 2007; Garman & Weik 2015). Common indicators of secondary radiation damage are the breakage of disulphide bridges and the loss of electron density of the carboxyl groups of acidic residues, with highly exposed carboxyl groups and those in the active site of the enzyme particularly susceptible (Weik et al., 2000).

It has been proposed that electrons can be produced by radiolysis of water within a crystals solvent channels, giving rise to a series of reactions that lead to the formation of hydrogen peroxide, with ionization producing a molecular cation of water

a. Short exposure



b. Long exposure

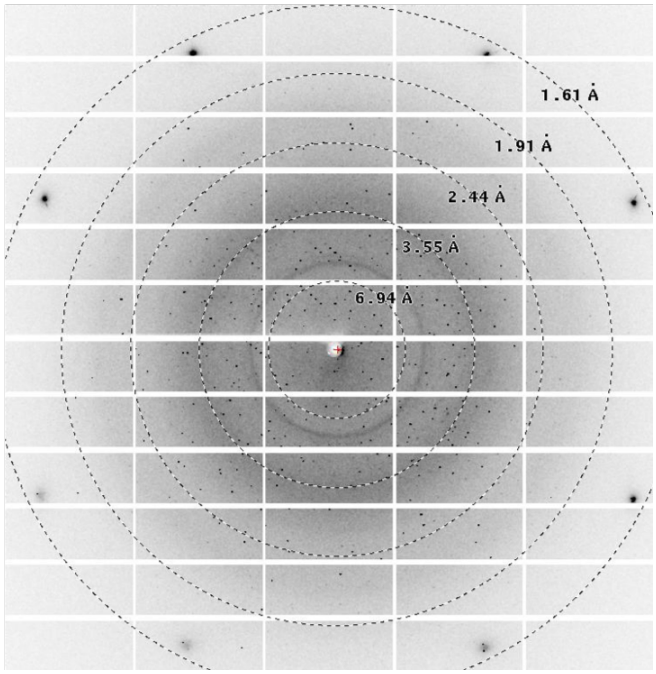


Figure 1.3. Global radiation damage as a loss of diffraction intensity

Loss of intensity in high resolution (> 1.9 Å, in this example) diffraction spots can be seen short and long exposure diffraction patterns (intense diffraction apparent ~1.68 Å are as a result of silicon diffraction silicon diffraction from fixed target silicon wafer used as a sample holder).

and an electron (fig. 1.4) (Hiroki, Pimblott, & Laverne, 2002). Site-specific radiation damage effects are therefore of particular concern in biomolecules containing transition metals, particularly those in high oxidation states. This is due to their high affinity for electrons, meaning they are susceptible toward oxidation state changes at much lower doses than those where global damage is an issue (Beitlich, Kühnel, Schulze-Briese, Shoeman, & Schlichting, 2007). This is due to the abundance of metalloprotein redox centres involved in electron transfer processes (e.g. cytochromes, iron–sulphur cluster proteins, and cupredoxins), with the whole range of reduction potentials in biology represented (Liu et al., 2014).

Radiation damage to crystalline biological samples exposed to an X-ray source is thought to depend on many factors however, including the beam flux, flux density, energy, size and profile, the cryocooling regime, and the physical and chemical properties of the crystal (Garman & Owen, 2006) (fig. 1.5). In all cases, radiation damage can lead to a loss of structural resolution and lead to a lower quality dataset, hindering the interpretation of the studied protein.

X-ray dose is affected by many of the parameters shown in figure 1.5. The dose of X-rays absorbed by a crystal is the energy deposited per unit mass of the sample, and quantified in SI units of Gray (Gy; 1Gy = 1J/Kg $^{-1}$), with dose measurements in MX generally of the order of a million Gray (1 MGy) (Garman, 2010; Paithankar & Garman, 2010). The software program *RADDOSE-3D* allows the macroscopic modelling of an X-ray diffraction experiment for the purpose of better predicting radiation-damage

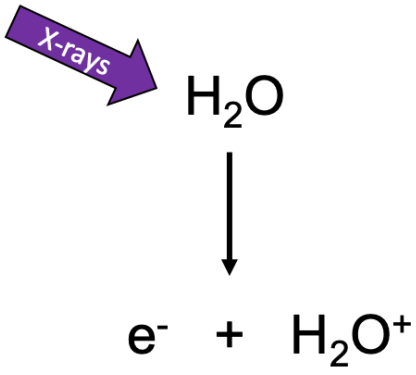


Figure 1.4. Schematic for the decomposition of water leading to the formation of electrons

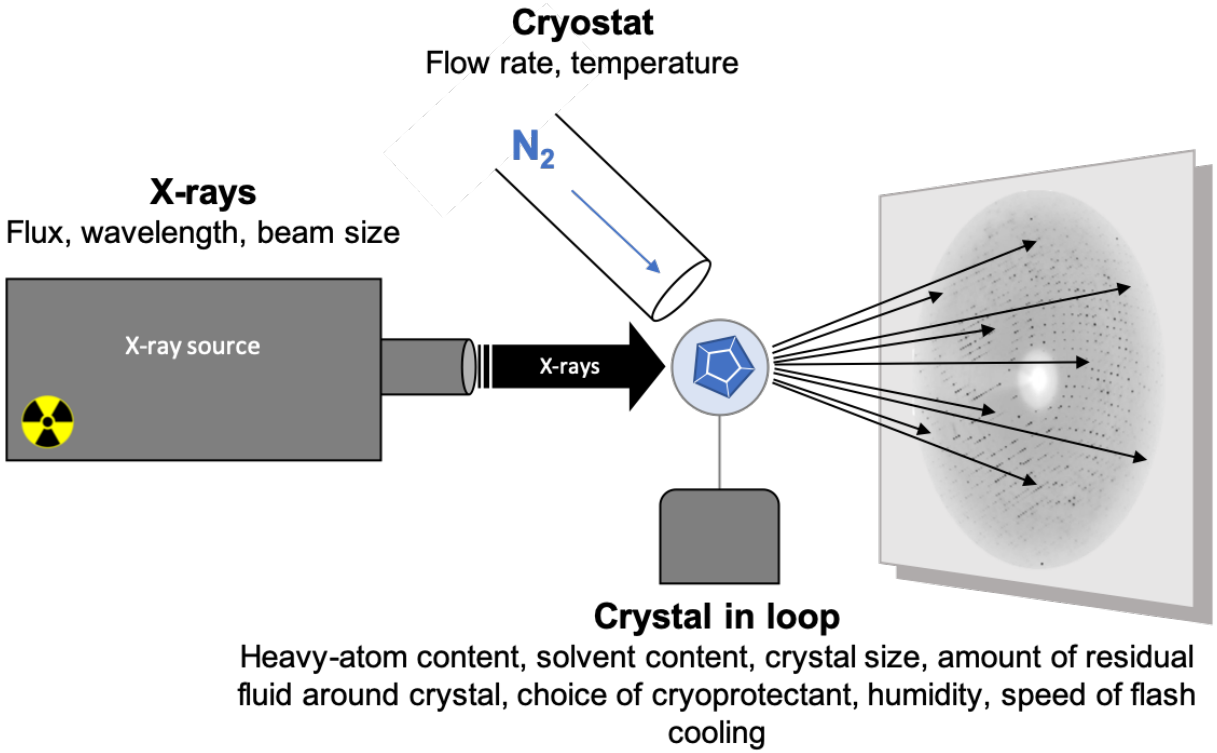


Figure 1.5. Schematic representation of the parameters believed to be relevant to radiation damage progression.

progression, providing a one-dimensional model of absorbed dose in a macromolecular crystal (Paithankar & Garman, 2010; Zeldin et al., 2013). It is possible to run *RADDOSE-3D* via an online server (www.raddo.se), by inputting crystal parameteters such as shape, dimensions, unit cell, number of monomers, heavy atom content, and solvent fraction; beam parameters such as profile, flux, size, energy, and collimation; and rotation parameters such as rotation wedge, and exposure time. Running this provides an estimate of a crystals diffraction weighted dose, the average dose in the diffracting crystal volume from which scattering has occurred (Zeldin et al., 2013).

1.1.2. Limiting radiation damage

With the aim of reducing radiation damage to crystalline proteins, data collection methods have been developed in order to minimise the absorbed dose during structure determination, extending the possibility of collecting data from radiation sensitive crystals. An example of this is utilising 'helical scan' procedures, where start and end points are defined on a crystal, and a single data set is collected as the crystal translates through beam in a helical fashion, so as to move fresh sample into the beam (Flot et al., 2010). The most frequent technique employed to limit radiation damage however is the use of liquid nitrogen to cryocool protein crystals. Cryocooled crystals are then held in a nitrogen stream when X-ray data are being collected; by holding the sample at ~100 K the mobility of free radicals produced by ionising absorption events caused by exposure to ionising radiation is significantly reduced (Garman & Owen, 2006; Nave & Garman, 2005). This results in a longer dose-lifetime of the crystal compared to that at room temperature (RT), allowing higher resolution data

collection, however electrons are still mobile meaning site specific damage, and in particular the reduction of metal centres, is still apparent at 100 K. Typically cryoprotectants such as small polyols, alcohols, sugars, polyethylene glycols, and inorganic salts (list not exhaustive, see Pflugrath, 2015, for a summary of common cryoprotectants) are commonly used to prevent the development of ice crystals in the cooled crystal (Pflugrath, 2015).

Rather than forming crystalline ice when cooled, cryopreserved solutions vitrify into clear, super-cooled glasses, without damaging the protein crystal. *In crystallo* and in solution protein and solvent dynamics show a 'dynamical transition' occurring ~200 K, believed to reflect the coupling of protein and solvent in a glass transition (Ringe & Petsko, 2003; Weik et al., 2001). Interestingly, the radiation sensitivity of protein crystals shows a clear shift in behaviour near ~200 K, with Warkentin and Thorne demonstrating in 2009 the increase in unit cell volume and Wilson *B* factor in thaumatin crystals occurs more rapidly when held in a cryostream above ~200 K, than identical crystals held below ~200 K (Warkentin & Thorne, 2009).

1.1.3. Room temperature vs cryo

The reluctance of the MX community to collect data from crystals at room temperature due to increased rates of radiation damage is well warranted, with the decrease in crystal lifetime at RT around two orders of magnitude compared to protein crystals held at cryogenic temperatures (Garman, 2010; Nave & Garman, 2005). RT data collection introduces advantages, with simple methods based advantages such as eliminating the need to search for cryo-conditions which can be

time and sample intensive (Heymann et al., 2014). More importantly, flash-cooling of crystals has been seen to decrease the heterogeneity of the conformational ensemble also, therefore hiding the identification and interpretation of important functional mechanisms; Fraser et al., 2011, demonstrate in two separate proteins cyclophilin A (CypA) and H-Ras that the reduction in thermal motion associated with cryo-cooling restricts coupled motions in flexible regions otherwise detected by nuclear magnetic resonance (NMR) studies of the same proteins in solution (Fraser et al., 2011).

This has been shown further by the collection of datasets across multiple temperatures; Keedy et al, 2015, present the collection of eight synchrotron datasets spanning 100-310K on the CypA protein, with multi-conformer models showing some alternative conformations in the active site network populated at only 240K and above, yet other alternative conformations populated 180K and below (Keedy et al., 2015). This was deduced to be due to an exchange between rotamer states of a large aromatic ring that resides in the middle of the dynamic active-site network subsequently causing a shift in the conformational ensemble of associated network residues (Keedy et al., 2015), insight only identifiable by collecting at temperatures above 100 K.

1.2. Data collection methods to assess and reduce radiation damage

Even though the diffusion rate of X-ray generated free radical species is reduced significantly at low temperatures, the susceptibility of metal redox states and ligand binding to site specific damage means the solution of low dose structures from metalloprotein crystals is a particularly challenging endeavour. As this can impact the

interpretation of metalloprotein crystal structures, different data collection methods can be used to investigate structural questions such as the influence of X-ray dose on ligand binding, the difference between structural characteristics at room temperature versus 100K, and the prevalence of secondary radiation damage effects. This can be achieved by collecting single or multiple diffraction images from multiple crystals in a serial fashion (i.e. one crystal after another), therefore allowing the researcher to form composite datasets using desired parameters such as time and/or accumulated dose on crystals typically held at room temperature.

1.2.1. Synchrotron methods

Multiple structures from one crystal (MSOX), proposed and developed by Horrell et al. in 2016, presents the collection of a large number of serial datasets at near atomic resolution from the same X-ray exposed region of a single large crystal (Horrell et al., 2016). This methodology allows a series of dose-dependent structures to be collected via the progressive addition of X-ray generated solvated electrons to the crystalline protein, enabling a “structural movie” of catalysis to be achieved. A major advantage of this is the ability to view the transition of structural states as radiolysis drives the enzyme from one state to the next, as opposed to the collection of single datasets that represent an average of the protein state over a particular dose/time, entirely missing subtle changes or classing them as insignificant.

It also is possible to compile a dataset from multiple protein crystals at synchrotron sources, by merging small degree wedges of data to form composite datasets. This remains one of the best approaches to measure structures that are susceptible to

redox reactions (particularly when combined with single-crystal spectroscopy to analyse the redox state), as the X-ray dose individual protein crystals are exposed to is greatly reduced. This technique has been demonstrated frequently: Meharenna et al. (2010) utilised a composite data collection strategy using synchrotron radiation, allowing the Fe-O bond length in the ferryl intermediate species of cytochrome c to be determined as a function of dose, a critically important intermediate in heme metalloenzymes (Meharenna, Doukov, Li, Soltis, & Poulos, 2010). Beitlich et al. (2007) examined cryoradiolytic reduction on multiple crystalline heme proteins via on-line UV-vis spectrophotometry and X-ray crystallography. They reported that the dose required for photoreduction from ferric (Fe^{3+}) to ferrous (Fe^{2+}) iron in myoglobin and chloroperoxidase was accumulated within <0.5 MGy, even though they differ significantly in amino acid structure, reduction potential, and active site chemistry (Beitlich et al., 2007). By measuring the UV-vis spectra of a crystal during a composite data collection, it was possible to assess the damage caused by the X-ray beam to the chromophore of dose limited single crystals.

Other methods are also available at synchrotrons that can be used alongside X-ray crystallography for the purpose of accurate measurement of bond length. These include nuclear magnetic resonance (NMR) and Raman spectroscopy, to name a few, used to identify values of molecular bond lengths and angles with varied precision (Hardcastle & Wachs, 1991; Yannoni et al., 1991). X-ray absorption spectroscopy methods are also commonly utilised, including X-ray absorption near edge structure (XANES) and extended X-ray absorption fine structure (EXAFS), with the latter reliable within an error limit of 0.02 Å (Bianconi et al., 1983; Hennig et al., 2001).

1.2.2. Serial femtosecond crystallography

The aforementioned techniques utilise single large crystals, and therefore have only required classical single crystal delivery methods that encompass a looped crystal on a goniometer. However, the advent of X-ray Free Electron Lasers (XFELs) in order to collect damage free crystallographic data has advanced sample delivery methods to accommodate a large number of very small microcrystals (<30 µm). XFELs are linear accelerator-based X-ray sources (fig. 1.6), producing X-ray pulses nine orders of magnitude more brilliant than synchrotron sources (Schlichting, 2015). When the X-ray pulse interacts with the crystal providing a diffraction pattern, the transferred energy destroys the crystal by Coulomb explosion. Provided the pulse length is short enough, damage caused by chemically reactive species, one of the main causes of radiation damage at synchrotron sources, is mitigated at XFELs as the XFEL pulse terminates before these molecules are created (Chapman, Caleman, & Timneanu, 2014; Nass, 2019). This has been coined ‘diffraction before destruction’ (Neutze et al., 2000). As crystals are destroyed by an XFEL pulse, many thousands of randomly orientated crystals need to be serially presented to the beam in order to collect enough single diffraction patterns to completely sample reciprocal space, a technique known as serial femtosecond crystallography (SFX) (Schlichting, 2015).

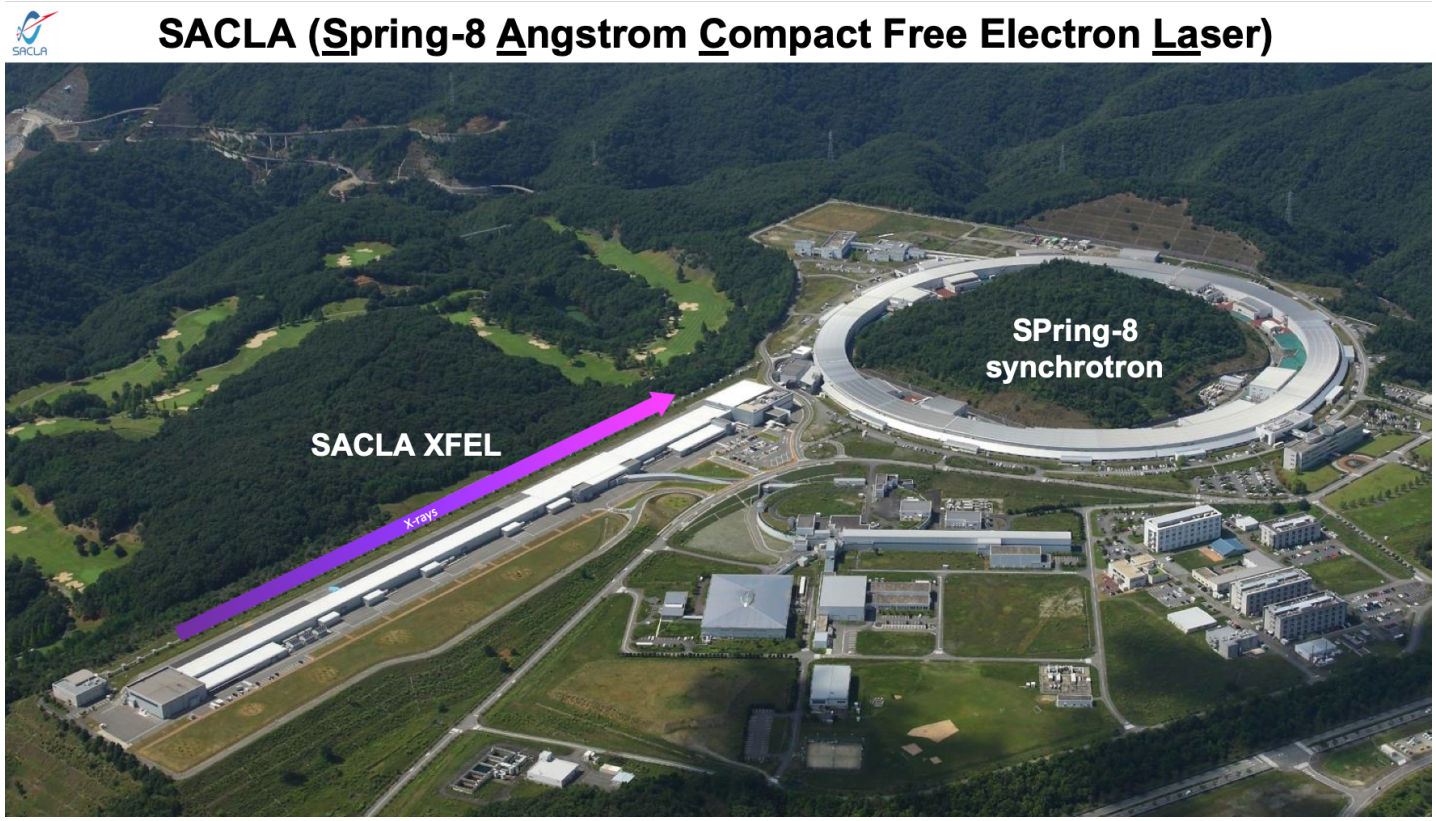


Figure 1.6. SACLA XFEL, Hyogo, Japan

The linear accelerator at the SACLA XFEL next to the Spring-8 synchrotron, with the XFEL experimental hutch located in the building at the tip of the arrow depicting the X-ray direction.

SFX data is highly sought after as it is 'damage free' and on an extremely fast time scale, opening the possibilities for time resolved experiments alongside the main advantage of obtaining structures free of the effects of site-specific damage, provided the pulse length is short enough. Tens of femtosecond pulses are necessary for this as signatures of radiation damage are observable on crystals exposed to pulses as short as 70 fs; a systematic decrease in the intensity of high-resolution reflections is apparent for pulse lengths longer than 100 fs, with indications of damage apparent at 70 fs, due to the presence of a steeper decrease in resolution for high intensity diffraction patterns compared to low intensity diffraction patterns (Lomb et al., 2011; Nass, 2019).

1.2.3. Multi-crystal delivery methods

The destructive nature of an XFEL source has led to the development of many sample delivery methods tailored to accommodate serial crystallography on XFELs. SFX demands that crystals are physically undamaged by sample delivery, whilst being large enough to diffract to a high resolution, produced in a large quantity, and sufficiently isomorphous and randomly oriented to allow merging of diffraction frames and assemble a complete sampling of reciprocal space (Heymann et al., 2014). One of the most widely used serial delivery methods are liquid jet or gas dynamic virtual nozzle (GDVN) technology (DePonte et al., 2009), where crystals are streamed through the X-ray beam via a narrow jet of liquid (Schlichting, 2015) (fig. 1.7). This has been a successful technique to collect SFX data, with a wide array of studies demonstrating this method at XFELs across the world (Aquila et al., 2012; Gisriel et al., 2019; Johansson et al., 2012).

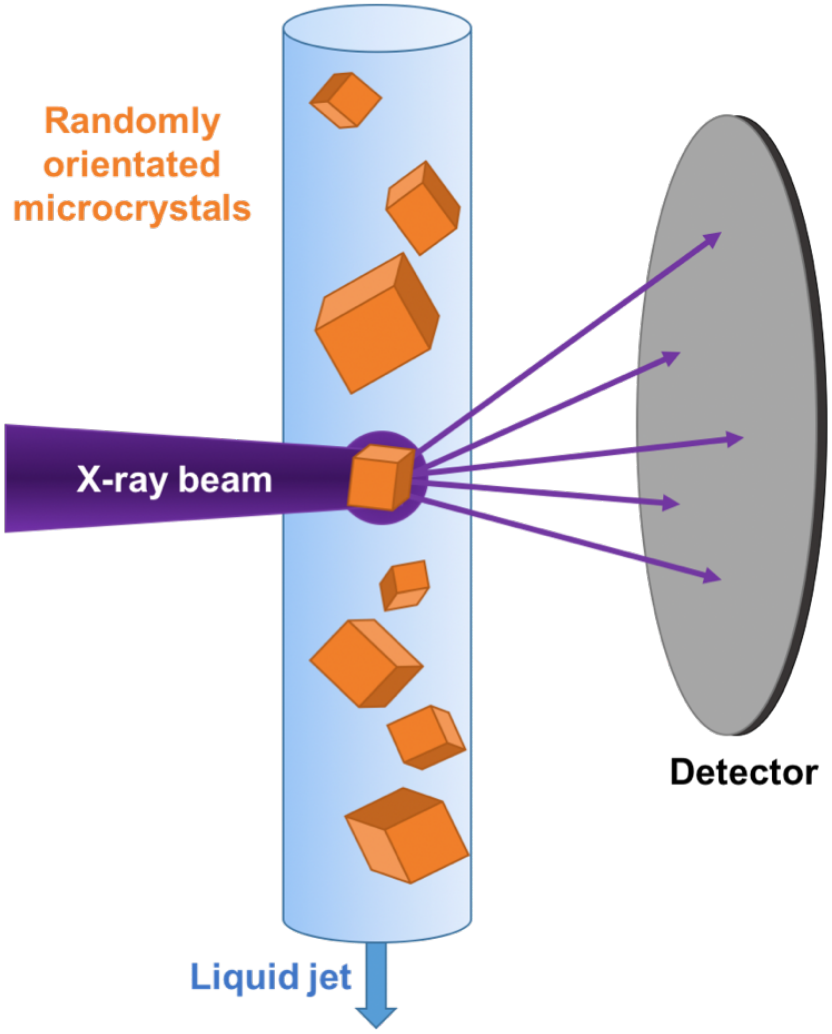


Figure 1.7. Liquid jet sample delivery

Schematic representation of liquid jet GDVN data collection.

A major disadvantage of GDVN however is that due to running the jet continuously at high speed (10 m s^{-1}, displacing 10-30 µl min^{-1}) much of the sample is wasted as it does not intersect the XFEL beam. Methods such as High Viscosity Extrusion jets (also known as "toothpaste jets"), where crystals within a highly viscous media such a lipid cubic phase (LCP) (Weierstall et al., 2014) or hyaluronic acid (Sugahara et al., 2016), are jetted through a stream through X-rays aim to combat rapid sample depletion. The higher viscosity of the media means that the jets can be run slower than a typical GDVN experiment, with typical flow rates between 1-300 nl min^{-1} (Weierstall et al., 2014).

Although originally developed for XFEL data collection, toothpaste jets have also been used at synchrotron beamlines; Nogly et al. demonstrate a synchrotron adaptation of SFX, coined serial millisecond crystallography or serial synchrotron crystallography (SMX or SSX), solving the structure of the light-driven proton-pump bacteriorhodopsin to a resolution of 2.4 Å at a synchrotron microfocus beamline, using this method (Nogly et al., 2015). Toothpaste jets are amenable to synchrotrons as the jet running more slowly than typical GDVN extruders allows the sample to be in the beam for long enough for a diffraction pattern to be obtained.

Crystals can also be presented to the XFEL beam on solid supports, known as fixed targets or 'chips' (fig. 1.10). This method exploits high density sample grids that allow an X-ray beam to raster across many tens of thousands of fixed target positions that contain protein crystals, in order to produce a full dataset from a combination of thousands of single diffraction patterns. This method has been demonstrated widely

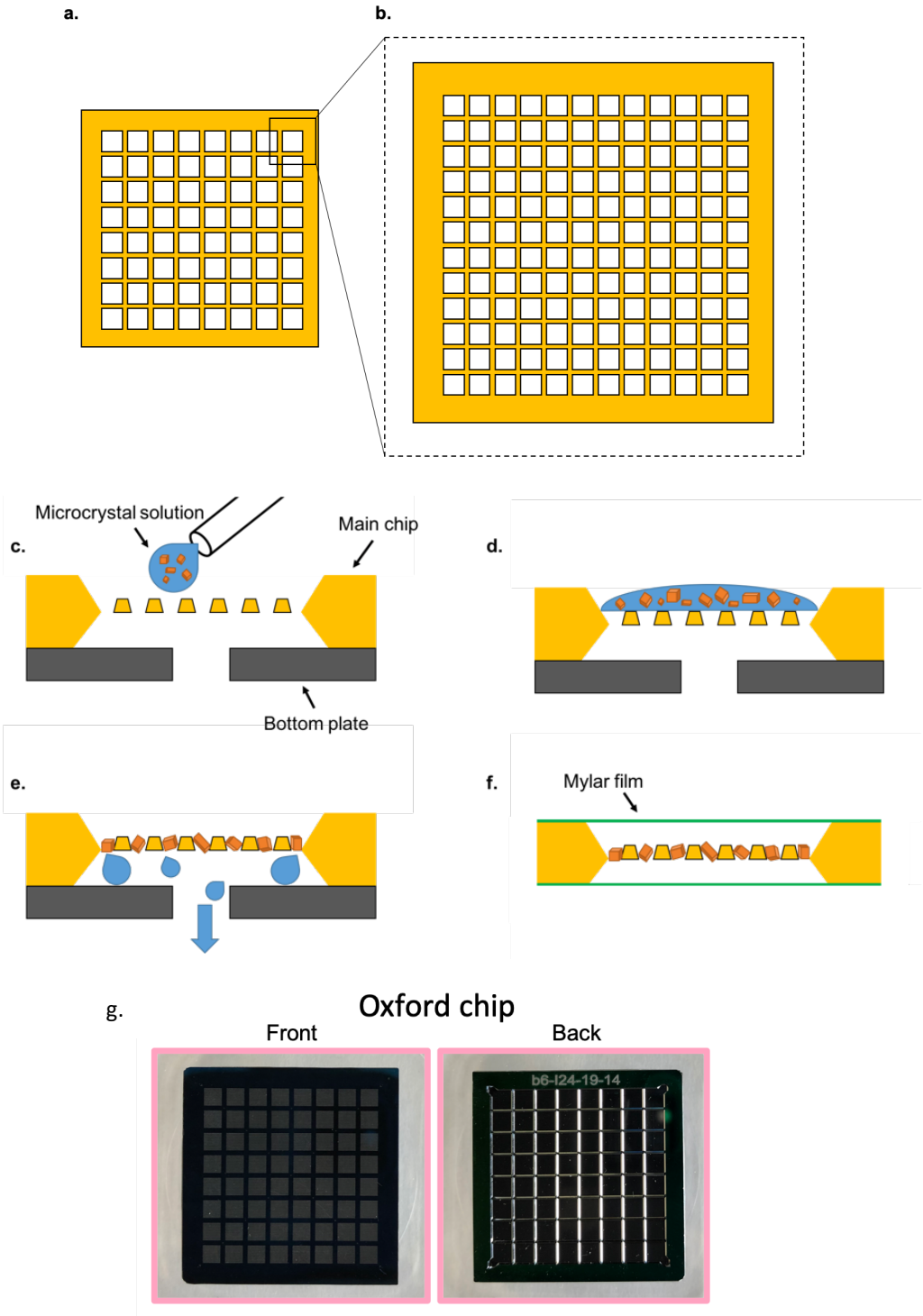


Figure 1.8 Fixed target 'chip' sample delivery method

Schematic of a typical chip, comprising 8 x 8 compartments (a.) with a further 12 x 12 apertures per compartment (b.). Typical chip loading procedure (c., d., e.), whereby a microcrystal solution is drawn through channels, allowing microcrystals to rest in the features and excess crystal solution removed via the bottom plate. (g.) Image of an 'Oxford' chip.

on XFELs using SFX (Frank et al., 2014; Hunter et al., 2014; Schlichting, 2015; Zarrine-Afsar et al., 2012), however shows great promise for the development of microfocus synchrotron radiation techniques, and has been demonstrated numerous times (Baxter et al., 2016; Huang et al., 2015; Oghbaey et al., 2016; Owen et al., 2016). Murray et al. (2015) used hen egg-white lysozyme microcrystals of a 10-15 µm size range loaded into a chip-based support system, yielding a complete, high-resolution (<1.6 Å) dataset from a microfocus synchrotron beamline (Murray et al., 2015).

1.3. Metalloproteins of interest

In order to examine radiation damage effects and develop crystallographic methods, suitable target metalloproteins and metalloprotein-ligand complexes that exhibit unambiguous signs of secondary site-specific radiation damage effects are required. Metalloproteins, a generic term for a protein that contains a metal ion cofactor, are extremely numerous in occurrence, with an estimated 25 – 50 % of all proteins found within organisms containing metal ions (Waldron et al., 2009), and 22 % of protein structures in the protein data bank (PDB) containing biologically relevant transition metals or metallo-cofactors. Metals that are incorporated within enzymes can significantly expand their catalytic repertoire: heme containing proteins have diverse functions including oxygen transport and catalysis, as well as enabling electron transport (Colpa, Fraaije, & Van Bloois, 2014; M. Davies, Hawkins, Pattison, & Rees, 2008; Meharenna et al., 2010; Yasushi Sugano, Muramatsu, Ichiyanagi, Sato, & Shoda, 2007). Copper is also important in metalloproteins, as it is utilised in electron transport and oxidation-reduction reactions (Solomon, 1993; Solomon, Sundaram, & Machonkin, 1996). A copper protein that binds nitrite and a heme peroxidase were

therefore targeted for study in this PhD thesis, due to their useful characteristics and applications, and importantly, their susceptibility toward radiation damage effects.

1.3.1. *Achromobacter cycloclastes* nitrite reductase

Copper nitrite reductase (CuNiR) catalyses the reduction of nitrite (NO_2^-) to nitric oxide (NO), in the microbial respiratory N-oxide reduction denitrification pathway, a process widely studied due to its agronomic and environmental impacts (Hajdu et al., 2000; Horrell et al., 2016; Hough et al., 2008; Murphy, Turley, & Adman, 1997; Nakano et al., 2012; Smith et al., 2012). Nitric oxide, generated by CuNiRs, is converted by nitric oxide reductases (NORs) into N_2O, which has increased by roughly 20 percent in the atmosphere since the 1850s, with the primary driver of the increase being the enhanced microbial production of N_2O in ever expanding fertilised agricultural lands (Smith et al., 2012). N_2O has a 100-year global warming potential 296 times larger than an equal mass of CO_2 (Ehhalt, Prather, & Dentener, 2001), and is therefore a powerful greenhouse gas that is increasingly being recognised as a major player in climate change.

CuNiR is found as a homotrimer in its biological assembly (fig. 1.9a), with each monomer of CuNiR containing a type 1 Cu centre (T1Cu), which has a role in electron transfer, and a catalytic type 2 copper centre (T2Cu) (fig. 1.9b) ~12.6 Å apart, the copper centres are characterised by a low driving force for electron transfer in the absence of nitrite, located close to the surface of the protein. The T1Cu accepts electr-

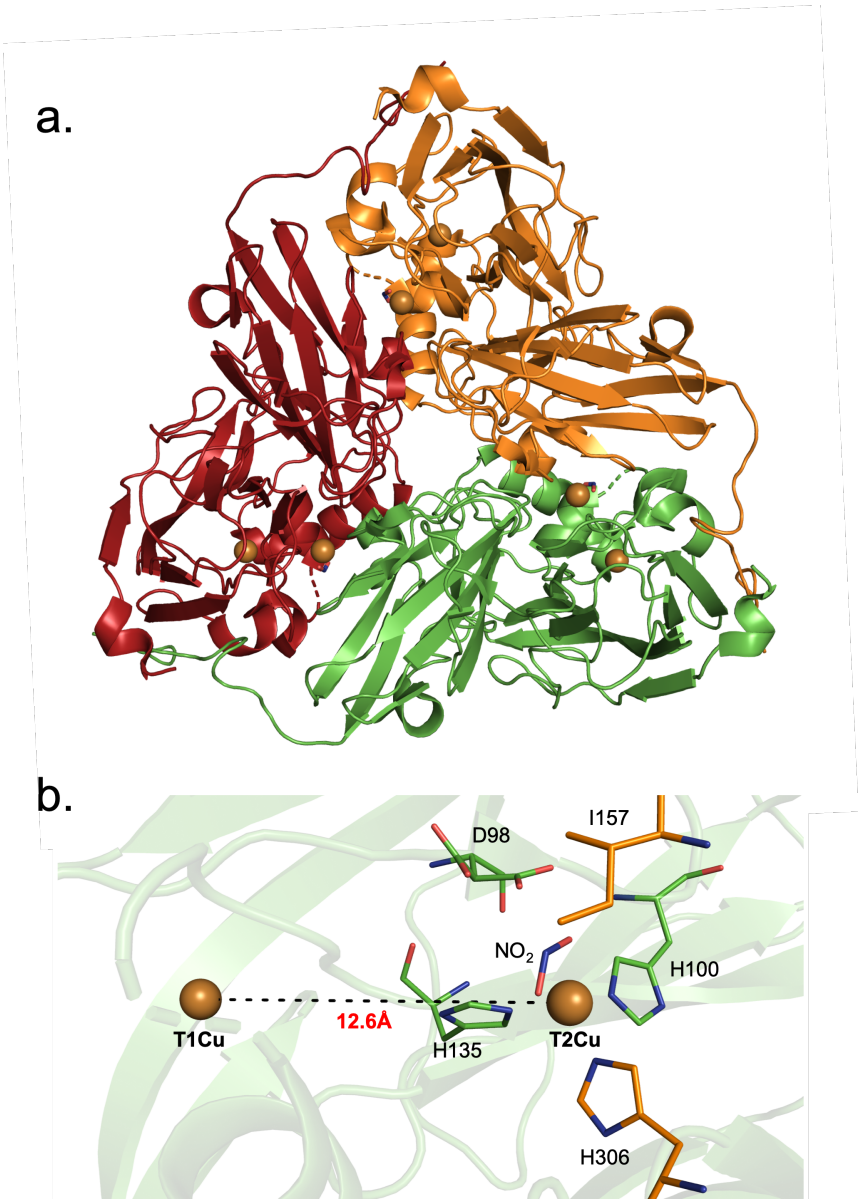


Figure 1.9. Overall and active site structure of AcNiR

a. Overall structure of AcNiR depicted as the biological assembly, a homotrimer, with subunits individually coloured (PDB: 5OFF). b. Distance between type 1 and type 2 copper centres, and conserved residues of interest surrounding the T2Cu are shown.

ons from its natural donor azurin, transferring them to the T2Cu where nitrite is reduced (Horrell et al., 2016; Krzemiński et al., 2011; Kukimoto et al., 1996; Wijma et al., 2006). At the T2Cu nitrite initially binds in a 'top-hat' geometry, before orientating to a 'side-on' geometry. Nitrite is then converted to nitric oxide in a side-on orientation, before a final resting state with a bound water which can then convert to a 'dead-end' state with nothing bound (Krzemiński *et al.*, 2011; Horrell *et al.*, 2016) (fig. 1.10).

Furthermore, as previously mentioned, MSOX was developed and utilised by Horrell et al. to study and further distinguish the catalytic mechanism of the CuNiR from *Achromobacter cycloclastes* (AcNiR). This was achieved by measuring a series of 45 consecutive serial datasets from the same AcNiR crystal held at 100 K with a nitrite ligand bound to the T2Cu site, enabling X-ray driven catalysis to be followed and reaction intermediates to be defined (Horrell et al., 2016) .It was therefore possible to determine ligand turnover in AcNiR at 100 K, with dual occupancy top-hat nitrite converted to single occupancy side-on nitrite after 2.76 MGy, single occupancy side-on nitric oxide after 11.73 MGy exposure, and a single water molecule after 27.6 MGy of exposure. This was elaborated upon in 2018 via the collection of 190 K and RT MSOX series, identifying nitrite to nitric oxide conversion after 4.89 MGy at 190 K, and only 0.18 MGy at RT, while the dead end branch was achieved after 21.65 MGy and 0.30 MGy at 190 K and RT respectively (fig. 1.11) (Horrell et al., 2018).

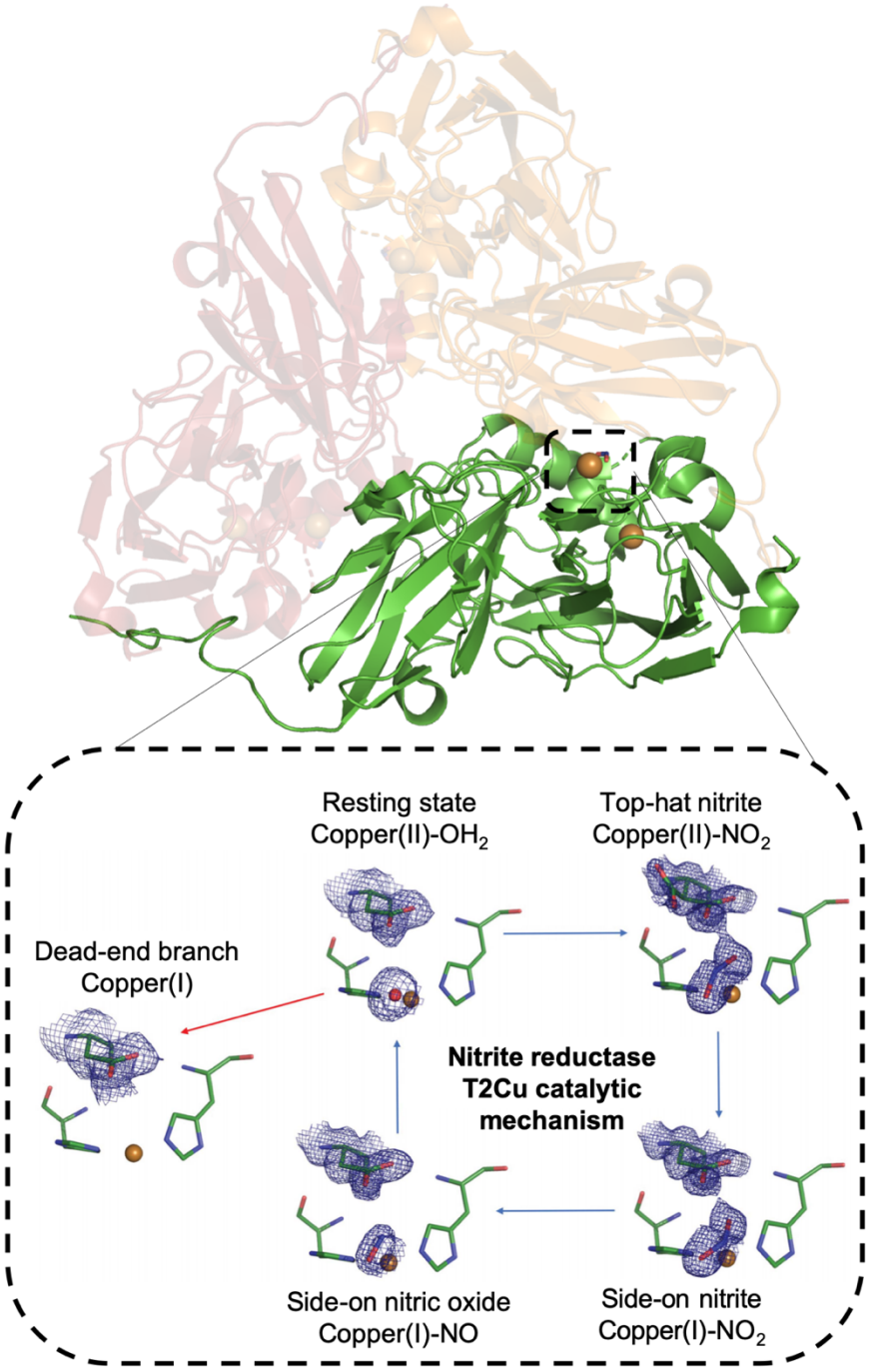


Figure 1.10. Catalytic mechanism of AcNiR

The proposed nitrite reductase mechanism determined using a 190 K MSOX series shows the resting state active site of AcNiR, followed by top-hat bound nitrite at the T2Cu. Following electron transfer a side-on binding mode is apparent of initially nitrite, followed by nitric oxide. The dead end state is indicated by a red arrow (figure adapted with permission from Horrell et al., 2018).

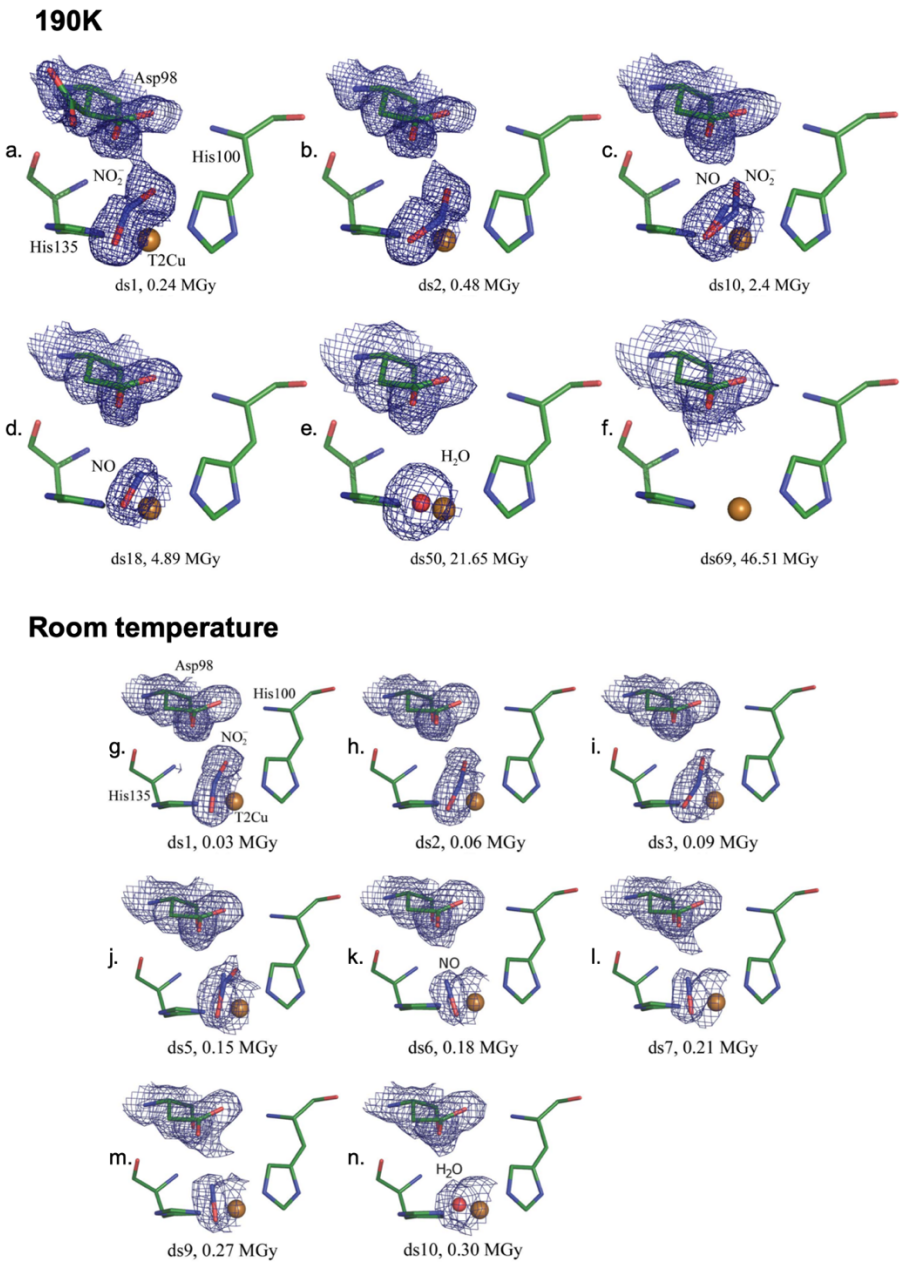


Figure 1.11. AcNiR ligand conformations at 190 K and room temperature

MSOX movies capturing the catalytic cycle of AcNiR at 190 K (a.-f.) and room temperature (g.-n.) indicates the turnover of nitrite to nitric oxide to a final resting state/dead end species of AcNiR occurs at much lower dose at room temperature (figure adapted with permission from *Horrell et al., 2018).*

1.3.2. DyP-type peroxidase Aa

Heme peroxidases (peroxidases containing prosthetic heme groups) catalyse the oxidation of substrates using hydrogen peroxide (H_2O_2) as a final electron acceptor, and have been thoroughly studied (Sugano et al., 2007; Verdín et al., 2006; Wesenberg, Kyriakides, & Agathos, 2003). The dye-decolourising peroxidases (DyPs) are new category of heme peroxidases first recognised only two decades ago, when DyP_{Dec1} was isolated from the fungus *Thenetephorus cucumeris* Dec1 for their ability to degrade a class of commercial dyes that are widely used by the textile industry (known as anthraquinones) (Kim & Shoda, 1999; Roberts et al., 2011). The crystal structures of DyP family metalloproteins reveal a two domain structure with a ferredoxin-like fold, and a penta-coordinated heme iron with a conserved histidine ligand (Zubieta et al., 2007) (fig. 1.12), with an aspartic acid (Asp171) conserved in DyP proteins is proposed as the replacement for the catalytic histidine used by plant peroxidases as a proton donor/acceptor (Sugano et al., 2007).

In 2007, Sugano et al. recognized DyPs not considered members of the plant or animal peroxidase superfamilies, due to their specific primary and tertiary structures and unique reaction characteristics, known as the dye-decolourising peroxidase-type (DyP-type) peroxidase family (Sugano, 2009; Sugano et al., 2007). The most comprehensive overview of the DyP-type peroxidase (Dtp) family can be found on the InterPro database (https://www.ebi.ac.uk/interpro/), this details how this superfamily comprises over 5000 enzymes, of which nearly all 5000 are found in bacteria, with over 100 found in eukaryotes, and more than 10 in archaea (Colpa et al., 2014).

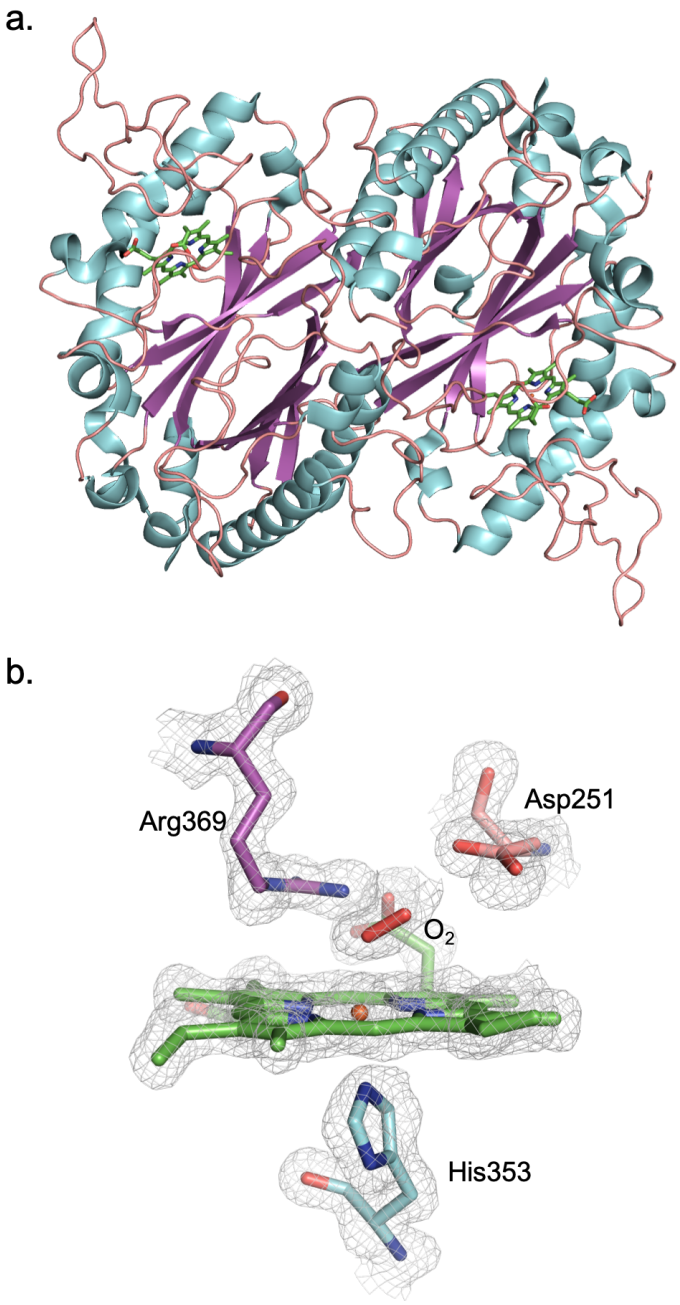


Figure 1.12. Overall and active site structure of DyP DtpA

(a.) Overall view of the dimer found in the asymmetric unit of DtpA, coloured by secondary structure (helix, blue; sheet, purple; loop, pink), with heme sites highlighted in green (PDB: 5MAP). (b.) Active site of 'pure' (oxygen only) oxy-ferrous species of DtpA, with proximal His353, and conserved residues Asp251 and arg369 shown. Electron density contoured at 1σ.

Only a few DyP-type peroxidases of fungal and bacterial derivation have had their biochemical properties investigated thus far, though all characterised proteins feature a non-covalently bound heme (proto heme IX) as a cofactor (Colpa et al., 2014; Li et al., 2012; Sturm et al., 2006; Van Bloois et al., 2010; Zubieta, Joseph, et al., 2007; Zubieta, Krishna, et al., 2007). Potential long range electron transfer between the heme cofactor and enzyme surface has been suggested, due to the fact the DyP-type peroxidases possess the ability to oxidise substrates that are too large to fit in the active site (Roberts et al., 2011).

Peroxidases react via a two-electron redox reaction, reducing H_2O_2 to H_2O and oxidising the enzyme via an iron-oxo intermediate (fig. 1.13) (Nelson et al., 1994; Zubieta, Krishna, et al., 2007). Metal sites in heme peroxidases can be affected X-ray generated photoelectrons, due to the sensitivity of the oxidase and peroxidase cycles to electrons (outlined in figure 1.13). Subsequently, this has led to redox and ligand state changes when collecting crystallographic data on peroxidases, even at 100 K (Kekilli et al., 2014, 2017). The X-ray generated photoreduction of a Dtp from *S. lividans* known as DtpA has been studied previously, using *in situ* single-crystal spectroscopy and diffraction, identifying complete reduction of the ferric heme after only 14.6 kGy (Kekilli et al., 2017). A homologue of DtpA known as DtpAa has also been identified in *Streptomyces lividans* as one of the three genes encoding for DyPs (the third, DtpB is cytoplasmic, while DtpA and DtpAa are extracytoplasmic), with DtpAa identified to have a 100-fold slower compound I formation than DtpA (Lučić et al., 2020).

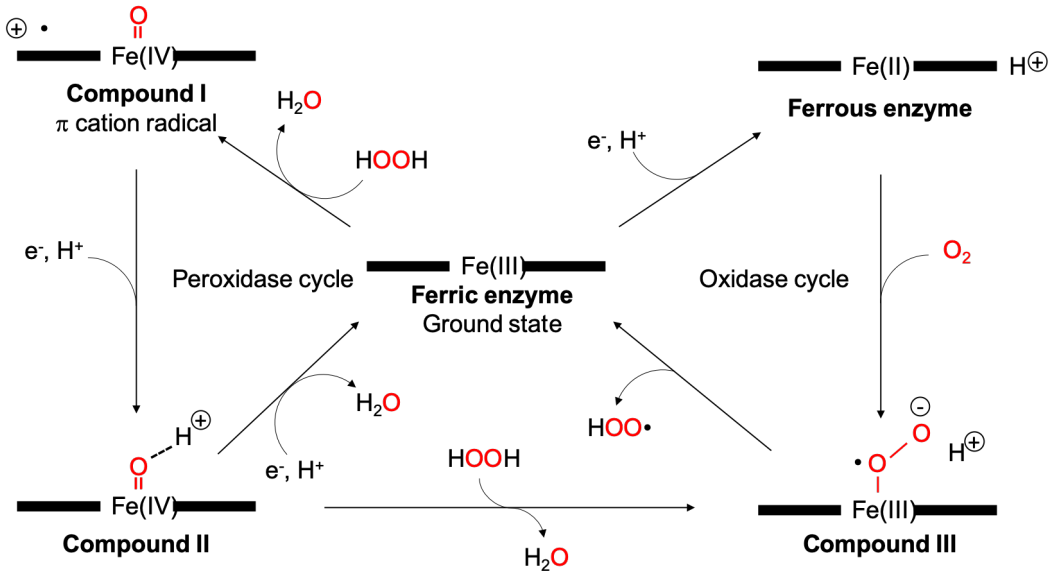


Figure 1.13. The oxidase/peroxidase cycle of heme peroxidases

Schematic representation of H_2O_2 production (oxidase cycle) and H_2O_2 scavenging (peroxidase cycle) by peroxidases. In a peroxidases ground state H_2O_2 scavenging is performed, transferring an oxygen atom from H_2O_2 to the heme group to form compound I. to return to the ground state they need to react with an electron donor (substrate) to form compound II, and then after a second reaction with an electron donor to return to its ground state. In the presence of a strong electron donor the ferrous state enzyme can be formed. After reacting with oxygen compound III can be formed. Compound III can also be formed from compound II if H_2O_2 is present in a high enough concentration. The reaction of compound III with an electron donor allows the formation of H_2O_2, returning the peroxidase to its ground state (Berglund et al., 2002; O'Brien, Daudi, Butt, & Bolwell, 2012).

Interestingly, the reduction of the ferric heme in peroxidases is associated with a linear increase in the Fe-O bond length of a water molecule bound to the Fe atom of the heme: Chreifi et al. utilised a comparative study between synchrotron and XFEL structures of a cytochrome c peroxidase; this was utilised to identify the increase in Fe-O bond length, with a damage free bond length of 1.7 Å, and a 0.53 MGy bond length of 1.9 Å collected at 65 K in (Chreifi et al., 2016; Meharenna et al., 2010). This bond length increase can therefore be used to identify and track the onset and progression of radiation damage in peroxidases.

1.4. Aims

The aims of this PhD project focus on the development of fixed-target methods that aim to reduce, track, and exploit radiation damage, with an emphasis on radiation sensitive proteins and obtaining damage free structures. Two proteins, AcNiR and DtpAa, introduced in this review are the targets and these were used in the development of microcrystal methods that encompass serial synchrotron crystallography, serial femtosecond crystallography, microspectrophotometry, and advanced data processing techniques.

Chapter 2: Crystallographic theory

X-ray crystallography is (as of 2020) the primary technique for the determination of the three-dimensional atomic arrangement of macromolecules, as the diffraction of X-rays from a crystal of identical molecules arranged in a repeating lattice makes it possible to view atomic level structures that cannot be seen by visible light (400 – 700 nm) (Rupp, 2009). The understanding of the theory behind the technique is therefore paramount. The principles of X-ray diffraction are covered extensively in a number of textbooks, and will be discussed in brief here, utilising the textbooks 'Crystallography Made Crystal Clear', Gale Rhodes, 2006, and 'Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology', Bernhard Rupp, 2009 (Rhodes, 2006; Rupp, 2009).

2.1. Macromolecular crystallisation

In order to perform the crystallographic experiment, a researcher first needs to grow a protein crystal. Macromolecular crystal formation occurs in two stages; nucleation and growth, via controlled precipitation of proteins from an aqueous solution of a specific pH, that does not denature the protein. Purified, concentrated protein solution is combined with an equal or similar volume of buffer system (mother liquor). This volume is enclosed in a sealed environment, with an adjacent reservoir of mother liquor at a higher concentration and left to equilibrate via vapour diffusion, causing the protein to become supersaturated and precipitate in solution, instigating nucleation of the macromolecules and then growth of the protein crystal (fig. 2.1a). When large numbers of small crystals are required however batch crystallisation is

utilised, whereby protein is combined rapidly with the buffer system, creating a supersaturated solution of protein that instigates the nucleation zone immediately upon mixing producing many crystals at once (fig. 2.1b). Typically, this is performed on smaller volumes of protein/buffer mixtures and scaled up to satisfy crystal intensive experiments.

2.2. Crystal properties

A crystal is made up from many repeating asymmetric units; where an asymmetric unit contains a representation of the macromolecular molecule with no crystallographic symmetry elements. This is known as the unit cell (though can contain >1 asymmetric unit). Unit cells pack beside and on top of one another, though are not perfectly ordered, and stack in rough alignment with one another, known as the mosaicity of a crystal. The unit cell parameters are designated by six numbers: the lengths of the three cell edges **a**, **b**, and **c**; and their internal angles α, β, and γ, where the angle between **a** and **b** is γ, b and **c** is α, and the angle between **c** and **a** is β (fig. 2.2).

How unit cells stack in specific symmetrical conformations determined by the space group, calculated from combinations of the 32 crystallographic point groups and the 14 possible Bravais lattices, meaning 230 possible space groups exist in crystallography. The geometric conformation of the unit cell exists in 7 conformations, known as crystal systems; cubic ($a = b = c$, $\alpha = \beta = \gamma = 90°$), tetragonal ($a = b \neq c$, $\alpha = \beta = \gamma = 90°$), orthorhombic ($a \neq b \neq c$, $\alpha = \beta = \gamma = 90°$), rhombohedral ($a = b = c$, $\alpha = \beta = \gamma \neq 90°$), hexagonal ($a = b \neq c$, $\alpha = \beta = 90°$ $\gamma = 120°$), monoclinic ($a \neq b \neq c$, $\alpha = \gamma = 90°$ β

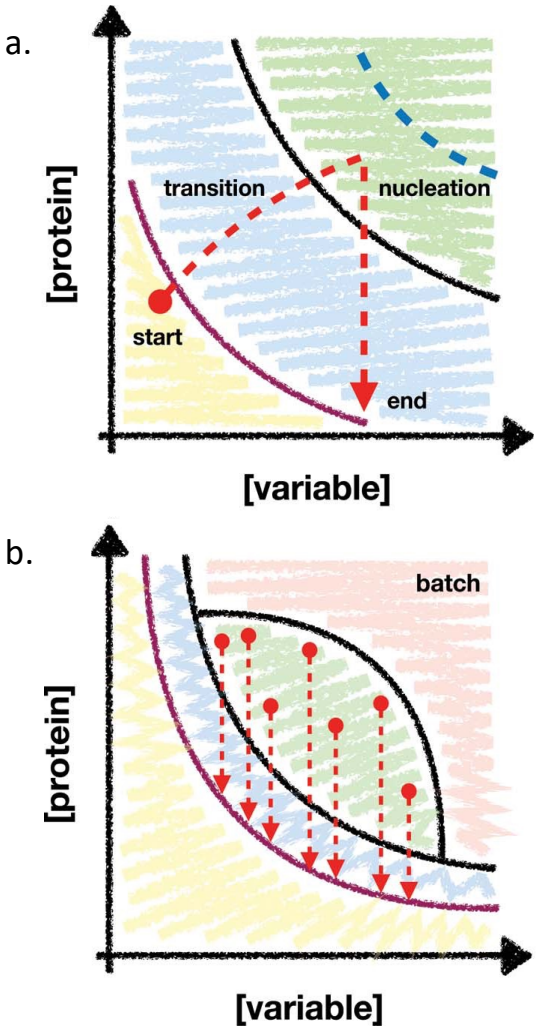


Figure 2.1. Phase diagram of crystallisation trajectories (Beale et al., 2019)

Protein and reservoir component (variable) concentrations are plotted on the y and x axes. The 'variable' is any factor which may influence the crystallization experiment, e.g. PEG, salt or buffer concentration. The purple lines show the boundary of protein supersaturation. Red circles and arrows denote the starting and finishing points of a crystallization experiment; precipitation (pink), nucleation (green), metastable (blue), and undersaturated (yellow) zones. The blue dotted lines show the theoretical limit of nucleation-zone penetration for non-batch methods. (a.) The trajectory of a vapour diffusion experiment: the components of the drop must transition from outside to inside the nucleation zone to crystallise. (b.) Batch experiments are not bound by the nucleation-zone limit and can theoretically exploit every part of the nucleation zone.

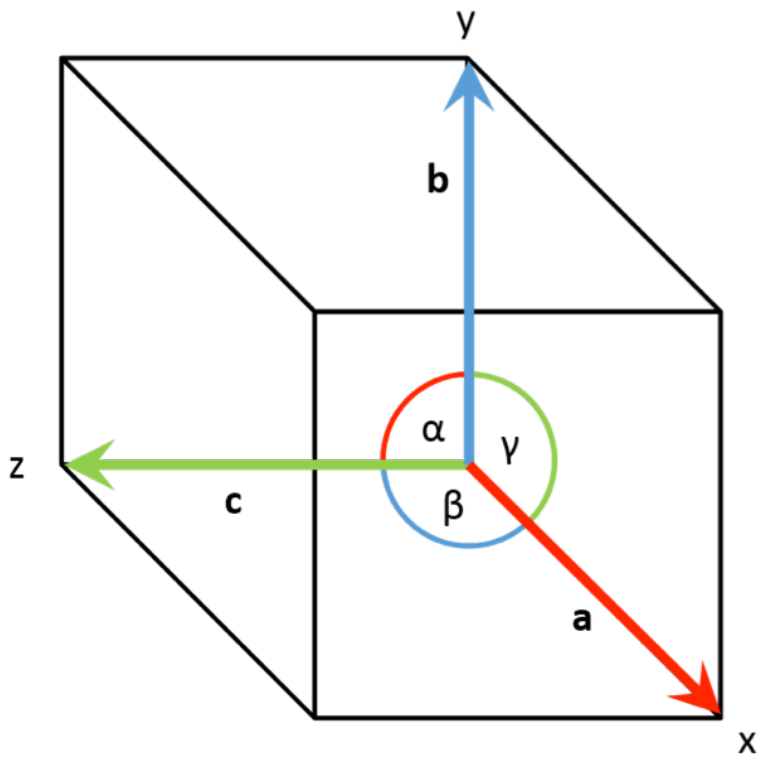


Figure 2.2. Unit cell properties of a crystal

*Example of a unit cell of cubic symmetry, with edges **a**, **b**, **c**, angles α, β, γ, and directions x, y, and z denoted.*

≠ 90°), and triclinic (a ≠ b ≠ c, α ≠ β ≠ γ ≠ 90°) (fig. 2.3). It is possible to draw an infinite number of equivalent parallel planes through these lattice points, as they can all be thought of as potential sources of diffraction. Furthermore, each plane can be designated a set of three indices, known as Miller indices. The three Miller indices *hkl* define the number of times a plane passes through a unit cell edge in a specific direction *x*, *y*, or *z* as coordinates.

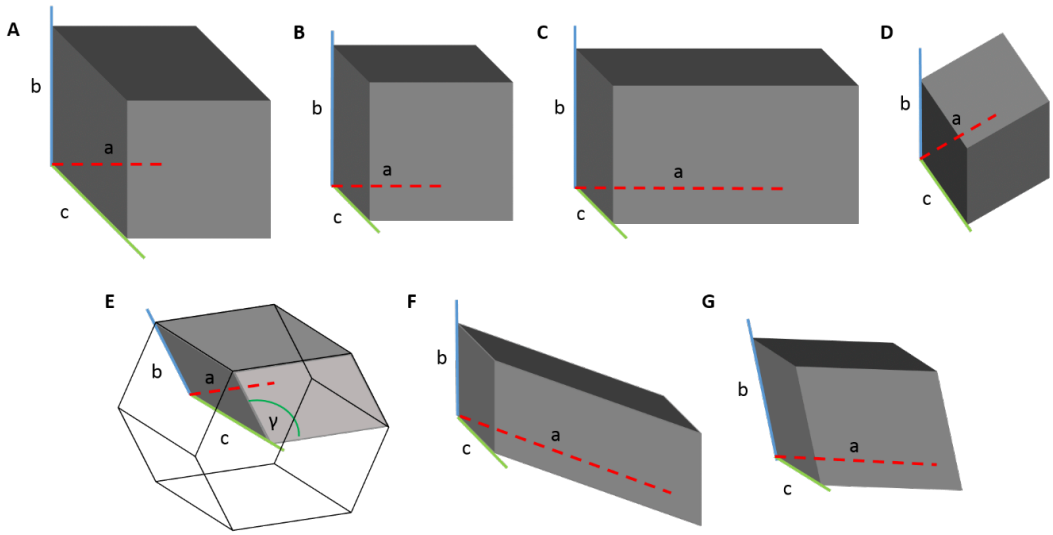
2.2. Bragg's law

When X-rays encounter electrons, they scatter in all directions. When X-rays scatter from different electrons and subsequently travel different distances, they will differ in their relative phases and there will be interference as they add up. We must calculate how far apart the planes must be for the difference in pathlength to be equal to the wavelength of the incoming radiation, so that the scattered rays from the two planes would again be in phase.

By quantifying the angle of incidence (θ), using Bragg's law it is possible to calculate the difference in path length. For a set of planes with interplanar spacing d_{hkl}, a diffracted beam is produced when X-rays of a specific wavelength (λ) act on the planes at angle θ. They are then reflected at the same angle which can be demonstrated when θ meets the condition:

$$2d_{hkl}sin\theta = n\lambda \ .$$

Equation 2.1



	System	Unit-cell dimensions (Å)	Unit-cell angles (°)	Characteristic symmetry
A	Cubic	a = b = c	$\alpha = \beta = \gamma = 90$	Four threefold axes
B	Tetragonal	a = b ≠ c	$\alpha = \beta = \gamma = 90$	One fourfold axis
C	Orthorhombic	a ≠ b ≠ c	$\alpha = \beta = \gamma = 90$	Three perpendicular twofold axes or/and mirrors
D	Trigonal	a = b ≠ c a = b = c	$\alpha = \beta = 90, \gamma = 120$ $\alpha = \beta = \gamma \neq 90$	One threefold axis
E	Hexagonal	a = b ≠ c	$\alpha = \beta = 90, \gamma = 120$	One sixfold axis
F	Monoclinic	a ≠ b ≠ c	$\alpha = \gamma = 90, \beta \neq 90$	Single twofold axis or/and mirror
G	Triclinic	a ≠ b ≠ c	$\alpha \neq \beta \neq \gamma \neq 90$	Only inversion centre possible

Figure 2.3. Crystallographic systems of symmetry

Where this condition is apparent with *n* as an integer of the path length of the wave, intense spots of reflected radiation formulate a diffraction pattern (fig. 2.4). When X-rays scattered from adjacent Bragg planes combine, this is known as constructive interference.

2.3. The Ewald sphere

A geometric construction known as the Ewald sphere helps visualise which Bragg planes are in the correct orientation to diffract. X-rays passing through a crystal lattice (the real space lattice) can diffract in any direction in three dimensions. This means X-ray diffracted from the real space lattice residing at the centre of a sphere with a radius of 1/λ, known as the crystal origin (*0*, *0*, *0*), can exit anywhere at the surface of that sphere, the reciprocal lattice origin (fig. 2.5).

Wherever a reciprocal lattice point intersects with the Ewald sphere, the diffraction condition has been fulfilled; by rotating the crystal along its axes further reciprocal lattice points intersect the Ewald sphere, due to the rotation of the reciprocal lattice. As diffraction from a crystal is confined to points on the reciprocal lattice, rotating the crystal allows the measurement of a complete diffraction dataset.

2.4. Structure factors

A single diffracted X-ray can be described as a structure factor F_{hkl}, consisting of the three terms; frequency, amplitude, and phase. By writing F_{hkl} as a sum of the contributions of each unit cell volume (V), and making V infinitely small in order to precisely equal the correct values of $\rho(x, y, z)$, the structure factor at this point in

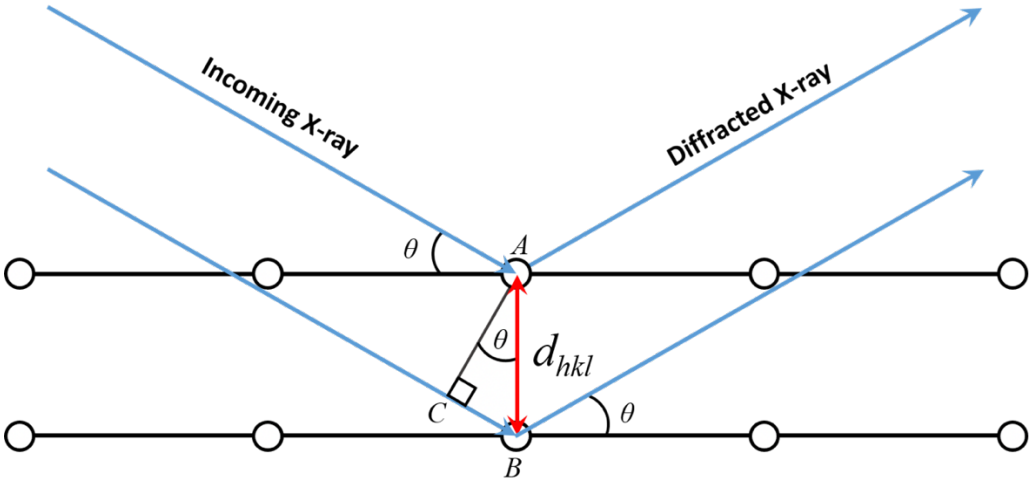


Figure 2.4. Bragg's law

In a crystalline solid, two beams of equal phase and wavelength can reflect at the same angle, combining into a constructive wave, causing high intensity diffraction. The lower wave must travel $2d_{hkl} \sin \theta$ to be in phase with the top beam as BC = AB sin $\theta = d_{hkl} \sin \theta$. If the difference in path length for beams reflected from two successive planes ($2d_{hkl} \sin \theta$) is equal to an integer of the wavelength of the beam ($n\lambda$), then constructive interference will occur.

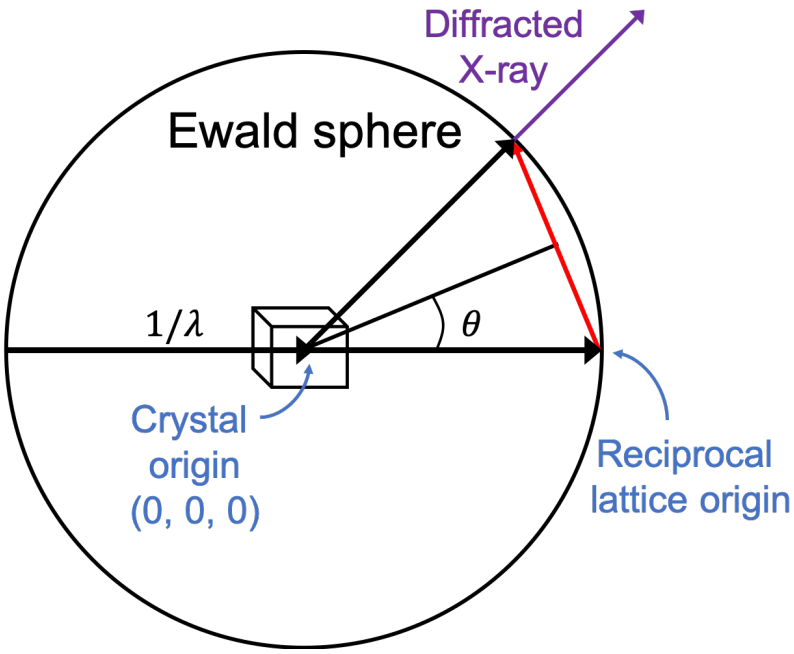


Figure 2.5.The Ewald Sphere

By rotating the crystal, different points of the reciprocal lattice will fulfil the diffraction condition set out by Bragg's law, resulting in constructive interference.

reciprocal space can accurately be described by:

$$F_{hkl} = \int_V \rho(x,y,z)\, e^{2\pi i\,(hx+ky+lz)} dV.$$

Equation 2.2

2.5. Fourier transform

In order to calculate atomic distributions via these scattered X-rays however, the experimentally produced reciprocal lattice data must be converted into real space inside the unit cell. This is achieved using a Fourier transform operation, expressing electron density at any point $\rho(x,y,z)$ in the unit cell as the sum of a sine wave:

$$\rho(x,y,z) = \frac{1}{V}\sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)}.$$

Equation 2.3

This equation represents a simple three-dimensional wave whereby the frequency h in the x-direction, k in the y-direction, and l in the z-direction, has amplitude F_{hkl}, implying the phase α_{hkl} of the three-dimensional wave.

2.6. Data reduction

Crystallographic data reduction first involves the indexing of reflections on diffraction images, as to assign Miller indices (h,k,l) and accurately determine the crystal and experimental geometry. These indexed images are then integrated, whereby reflection intensities, I, and background value, $\sigma(I)$, are determined. Individual integrated images are then scaled, which involves inferring the structure factor from the measured intensity of a reflection by putting the measured intensities on the same scale; this is achieved by making the data internally consistent by adjusting the

scaling model to minimise the difference between symmetry-related observations, producing an estimate of the intensity of each unique reflection, an estimate of the error, and an estimate of the structure amplitude (Evans, 2011). The data are finally merged into a single dataset. Programmes used for data reduction in this thesis will be discussed individually in both the methods and their respective chapters.

2.7. Data quality

A number of data quality metrics are utilised in crystallographic data reduction in order to assess data quality at indexing, integration, scaling, merging and refinement stages. This allows the experimenter to define at what point measured data fails to yield any usable information and make informed decisions about the quality or potential anomalies within their data. Publications reviewing data metrics used to measure the quality of experimental diffraction are numerous (Evans, 2011; Evans & Murshudov, 2013; Karplus & Diederichs, 2012; Wlodawer et al., 2013), therefore will be covered in brief.

The average signal-to-noise ratio, $I/\sigma(I)$, is the intensity of a given reflection, and can be used to define the resolution limit of diffraction when $I/\sigma(I)$ decreases to a value of 2.0. Improving $I/\sigma(I)$ is complemented by a high data redundancy (multiplicity) of measurements, improving the quality of the merged dataset by increasing the number of measured intensities and estimated uncertainties.

R_{split} measures the agreement between sets of intensities by merging odd- and even-numbered patterns from the overall dataset,

$$R_{split} = \frac{1}{2^{1/2}} \frac{\sum| I_{even} - I_{odd}|}{\frac{1}{2}\sum(I_{even} + I_{odd})},$$

Equation 2.4

R_{merge} values measure the spread of *n* independent measurements of the intensity of a reflection, $I_i(hkl)$, around their mean $\bar{I}(hkl)$,

$$R_{merge} = \frac{\sum_{hkl}\sum_{i=1}^{n} | I_i(hkl) - \bar{I}_i(hkl)|}{\sum_{hkl}\sum_{i=1}^{n} \bar{I}_i(hkl)},$$

Equation 2.5

reporting on the agreement between multiple measurements of a given reflection. R_{merge} has subsequently been corrected to be independent of multiplicity, known as R_{meas}, allowing a reliable report on the consistency of individual measurements by adjusting by a factor of $\sqrt{\frac{n}{n-1}}$.

Pearson's correlation coefficients (CC) are also often used for the estimation of resolution limit and data quality, with $CC_{1/2}$ one of the most common; this metric is calculated by separating unmerged data into two randomly selected half sets of unique reflections, and calculating the correlation coefficient between the two data sets, with $CC_{1/2}$ values ranging from 1 for perfect data, to 0 for highly imprecise data. The advantage of this is that $CC_{1/2}$ provides a model-free, empirical measure of the level of discernible signal calculated within each resolution shell with, for good quality values close to 1 observed at low resolution, dropping smoothly to 0 as the measurable signal decreases at higher resolution. This behaviour is advantageous as

deviations in this smooth decrease can be an indicator of an irregularity in the data (Karplus & Diederichs, 2012).

The standard indicator for assessing the agreement between a refined model and the measured data is the crystallographic R value, defined as:

$$R = \frac{\sum_{hkl} |F_{obs}(hkl) - F_{calc}(hkl)|}{\sum_{hkl} F_{obs}(hkl)},$$

Equation 2.6

with $F_{obs}(hkl)$ and $F_{calc}(hkl)$ being the observed and calculated structure factor amplitudes, respectively (Karplus & Diederichs, 2012). R can be 0.0 for perfect agreement and near 0.59 for a random model, with R_{free} a cross validated R based on a small subset of reflections not used during refinement, and R_{work} the larger "working" set of reflections used during refinement, with values R_{free} and R_{work} values of < 0.20 typical in high resolution, high quality refined structures. Further parameters of interest in refinement are B-factors (also referred to as temperature factors), as they are isotropic approximations of the mobility in space for each atom. B-factors are isotropic approximations expressed in Å^2, and range from ~2 to ~100 (Wlodawer et al., 2013).

2.8. Structure refinement

After the initial process of phasing (experimental or otherwise) and model building, the protein model typically contains many imperfections. In order to improve the phases and the interpretation of the electron density map, refinement methods that incorporate multiple 'rounds' of rebuilding and refinement are typically required.

Atomic coordinates are adjusted statistically in order to fit the diffraction data better; R_{work} is used as a measurement of how the calculated amplitudes fit the observed amplitudes, and R_{free} used as refinement validation by using a subset of reflections excluded from refinement, therefore giving an independent measure of the refinement process (Kleywegt & Jones, 1997).

Maximum likelihood, and simulated annealing (SA) refinement are most widely used in X-ray crystallography: maximum likelihood refinement adjusts the phases to minimise the R value, while SA refinement 'heats' the structure to add randomness before slowly 'cooling' the structure during refinement (Adams et al., 1997). Once a round of refinement is completed, any missing atoms, incorrect rotamers, alternate conformations etc. are manually corrected, and the refinement protocol repeated. Refinement constraints are exact mathematical conditions used in structural refinement, whereby one or more parameters are held fixed or determined by the value of one or more refined parameters, e.g. fixing of the x, y, and z coordinates of an atom on an inversion centre (Bourhis et al., 2015). Refinement restraints can also be included, whereby additional information is fixed based on additional observations or pseudo-observations besides the observed structure factors, for example fixing the standard deviation of a chemical bond (Bourhis et al., 2015).

Chapter 3: Materials and methods

3.1. Protein expression and purification

3.1.1. AcNiR

An AcNiR gene, codon optimised for expression in *E. coli* and cloned into a pet26b plasmid was purchased from Genscript previously, and kept in storage at -20 °C at the metalloprotein laboratory at the University of Essex. This vector was transformed into *E. coli* BL21 (DE3) cells and used to produce overnight 10 ml pre-cultures (LB medium; Thermo Fischer scientific). Pre-cultures were successively used to inoculate 1.5 L of LB medium in the presence of 50 µg/ml kanamycin and 2 mM $CuSO_4$ (final concentration), and grown at 37 °C, 180 rpm. At an OD600 of 0.4-0.6, cultures were induced using 2 mM Isopropyl β-d-1-thiogalactopyranoside (IPTG). Flasks were sealed with a sponge bung and tinfoil, and incubated for a further 18 h (overnight) at 18 °C. Cells were harvested via centrifugation (10,000 g, 30 min, 4 °C) and the cell pellet resuspended in 20 mM Tris/HCl, pH 7.5. The resuspended cell suspension was lysed using an EmulsiFlex-C5 cell disrupter (Avestin) followed by centrifugation (18,000g, 20 min, 4°C). The cleared lysate was loaded into 8000 kDa MWCO dialysis tubing (Thermo Fischer Scientific), and dialysed against 5 L 2mM $CuSO_4$, 20 mM Tris pH 7.5 for 2.5 hours, and further dialysed against 5 L 20 mM Tris pH 7.5 for 18 hours (overnight).

The dialysed supernatant was loaded on a 10 ml DEAE sepharose column (GE Healthcare) equilibrated against 20 mM Tris pH 7.5, and eluted by a linear salt gradient using buffers of 50 mM NaCl, 20 mM Tris pH 7.5 (buffer A), and 1000 mM NaCl, 20 mM Tris pH 7.5 (buffer B), with the AcNiR peak fractions eluting at

approximately 30-40 % buffer B. Eluted fractions were run on an SDS PAGE gel to judge purity, pooled and concentrated using a Centricon centrifugal concentrator (VivaSpin) with a 10 kDa cut-off, at 4 °C. This step was followed by application to a disposable PD10 de-salting column in order to remove NaCl that might interfere with subsequent purifications. DEAE and PD10 purification/SDS-PAGE steps were then repeated. Pooled AcNiR fractions were applied to an S200 Sephadex column (GE Healthcare) equilibrated with 20 mM Tris pH 7.5. A major peak eluted consistent with fractions assessed by SDS-PAGE. Fractions were pooled, concentrated to 20 mg/ml, and stored at −80 °C, with concentrations determined via UV–visible spectroscopy (Varian Cary 60 UV–vis spectrophotometer) using an extinction coefficient at 280 nm, measured in units of M^{-1} cm^{-1}.

3.1.2. DtpAa

Purified DtpAa protein was provided by Dr. Jonathan Worrall of the University of Essex in order to perform a collaborative experiment, with the identified expression, purification and crystallisation optimisation methods used by their research group stated below. As purification of DtpAa was not performed by the author full purification details can be found in Ebrahim, Moreno-Chicano, et al., 2019.

3.2. Crystallisation

3.2.1. AcNiR

Crystallisation conditions for single crystals of AcNiR are well established via previous study (Antonyuk et al., 2005; Horrell et al., 2016), with multiple small AcNiR crystals grown in 6 µl drops consisting of 3 µl protein and 3 µl 1.6 M ammonium sulphate, 100

mM sodium citrate pH 4.5, via hanging drop vapour diffusion. These small AcNiR crystals are used as seeds to produce larger (>50 µm) crystals, via transfer into fresh, identical conditions.

In order to complete the planned research for this thesis however, larger volumes of crystallisation solution (>150 µl) containing a high concentration of microcrystals are required, due to the sample intensive nature of serial crystallography (see section 1.2). Trials of conditions to produce microcrystal slurries in batch were performed in order to elucidate optimal large-scale microcrystal conditions; different ratios of protein to buffer, protein concentrations, and crystallisation temperatures were all used to optimise batch crystallisation in 2.5 M Ammonium Sulphate, 0.1 M Sodium Citrate pH 4.5 (results shown in chapter 4.3.1).

From this point forward, batch microcrystals were prepared by rapidly mixing 20 mg/ml AcNiR in 20 mM Tris, pH 7.5 with a solution containing 2.5 M ammonium sulphate, 0.1 M sodium citrate pH 4.5 buffer, in a ratio of 1:3 and mixed by vortexing for 60 seconds. Microcrystals with a diameter of 5-15 µm grew at room temperature over a period of 4-6 days. Microcrystal suspensions were centrifuged at 800 rpm for 30 seconds to sediment crystals, with the crystallisation solution then removed and replaced with a storage buffer comprising 1.6 M ammonium sulphate, 0.1 M sodium citrate pH 4.5, in order to cease crystallisation.

3.2.2. DtpAa

Optimisation of DtpAa crystallisation conditions was performed by the Worrall and Hough groups at the University of Essex, with the material used for data collection in this thesis provided as part of a collaborative project. Optimal conditions to produce DtpAa microcrystals in a total volume of >500 µl were discovered to be a 1:1 ratio of 6.5 mg/ml DtpAa protein solution with a precipitant solution containing 20% PEG 6000, 100 mM HEPES pH 7.0, with crystals growing to typical dimensions of ~15-30 µm.

3.3. Sample preparation

3.3.1. Fixed target 'chips'

Silicon wafer fixed targets, referred to as 'chips', were used primarily for serial data collection from microcrystals in throughout this thesis (fig. 3). Chips of this design, known as an 'Oxford chip', were developed by Diamond Light Source and manufactured by Southampton Nanofabrication Centre at the University of Southampton, via an iterative process of chemical etching. Nanofabrication details are beyond the scope of this thesis, though are readily available via multiple publications (Mueller et al., 2015; Oghbaey et al., 2016; Zarrine-Afsar et al., 2012).

The chip is made out of a single-side polished silicon wafer coated in silicon nitride, measuring 30 mm x 30 mm x 550 µm (fig. 3.1a, b, c). The chip contains a high capacity layout of funnel shaped apertures, split into 8 x 8 compartments known as 'city blocks', which each contain 20 x 20 apertures (fig. 3.1c). The nominal capacity of the chip is therefore 25,600, with a four-fold symmetry. In order to accurately align the

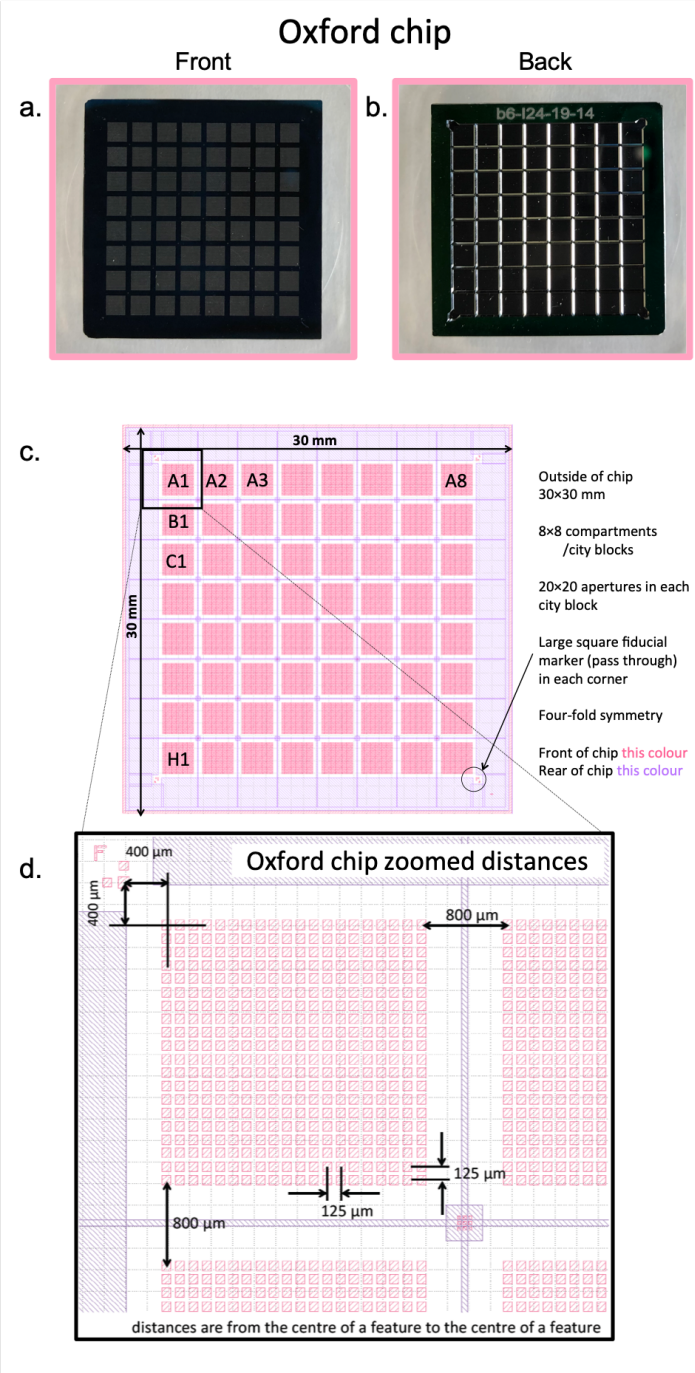


Figure 3.1 Oxford chip schematics.

The front (a.) and back (b.) of the silicon nitride chips. (c.) Schematic of the overall size of the chip and the locations of 'city blocks' of apertures. (d.) Zoomed-in schematic of crystal and fiduciary apertures, as well as the distances separating city blocks (800 µm), and individual apertures from one another (125 µm) and from the fiduciary marker (400 µm). All measurements are taken from the centre of an aperture to the centre of an aperture.

chips to the beam there are fiducial marks 400 µm from the centre of each corner aperture (fig. 3.1d). The distance between city blocks is 800 µm, with the distance between individual apertures 125 µm (all measurements made from the centre of an aperture to the centre of an aperture, see figure 3.1d). Apertures are funnel shaped, with the larger end of the funnel opening on the "front" of the chip where the crystals are loaded onto, so that they can become lodged in the narrower end of the aperture while removing any excess solvent via a weak vacuum from below.

The chip surface was prepared by glow discharge (PELCO easiGlow™ glow discharger) in order to make the silicon more hydrophilic; due to the hydrophobic nature of silicon, when water is placed on the surface of the chip the water will minimise its silicon contact. This results in the crystal solution sitting as large beads on the surface of the chip, making it difficult to evenly spread the crystals across its entire surface. Glow discharging the chips involves depositing negatively charged ions on the silicon, giving a more hydrophilic surface. This is accomplished by placing the chip between a cathode and anode in a partially evacuated chamber. When high voltage is applied between the cathode and anode the electron potential ionises the gas within the chamber, with these negatively charged ions depositing on the chip. Preliminary experiments did not involve glow discharging, however since introducing it to the chip loading protocol the dispersion of the crystal slurry on the chip surface has improved, reducing the volume of crystal slurry required to load a chip, as-well-as improved drawing of liquid through the chip apertures.

3.3.2. Fixed target loading

Chips were loaded within a humidity control enclosure (Solo Containment, Cheshire UK) (fig. 3.2), kept at >80% humidity using a 2.8L ultrasonic humidifier, in order to prevent crystal solutions drying to the chip surface in low humidity. Rubber tubing is used to connect a vacuum pump to a pressure gauge and a sample loader block, via a conical flask and stopcock. Typically, a microcrystal suspension of 100-200 µl is pipetted onto the surface of a glow discharged chip sitting in the sample loading block. A weak vacuum was then applied from beneath the chip to draw microcrystals into individual apertures and remove excess solvent. Development of sample loading procedures used in this thesis can be found in Mueller et al., 2015 and Oghbaey et al., 2016.

In order to securely hold the chip in place when collecting data, the chip is placed in a sample holder (fig. 3.3) and mounted to the stages using a magnetic kinematic mount, allowing rapid, tool-free sample exchange. The sample holder is designed by Diamond and manufactured by Gatehouse engineering, machined from aluminium. The chip holder uses two flat plates with an area slightly larger than the fixed target cut out from the centre of top and bottom half (fig. 3.3a, b, c), and the magnetic kinematic mount (Thorlabs) located on only the bottom half of the holder. Both halves utilise a metal and a rubber O-ring; the metal O-ring is used to clamp mylar film (https://www.spexsampleprep.com/) (fig. 3.3 e, f) on either side of the chip holder in order to prevent the chip from drying out during data collection, while the smaller rubber O-ring is utilised to improve the seal between both halves of the chip holder. The halves are held together with magnets and are secured closed using hex key bolts

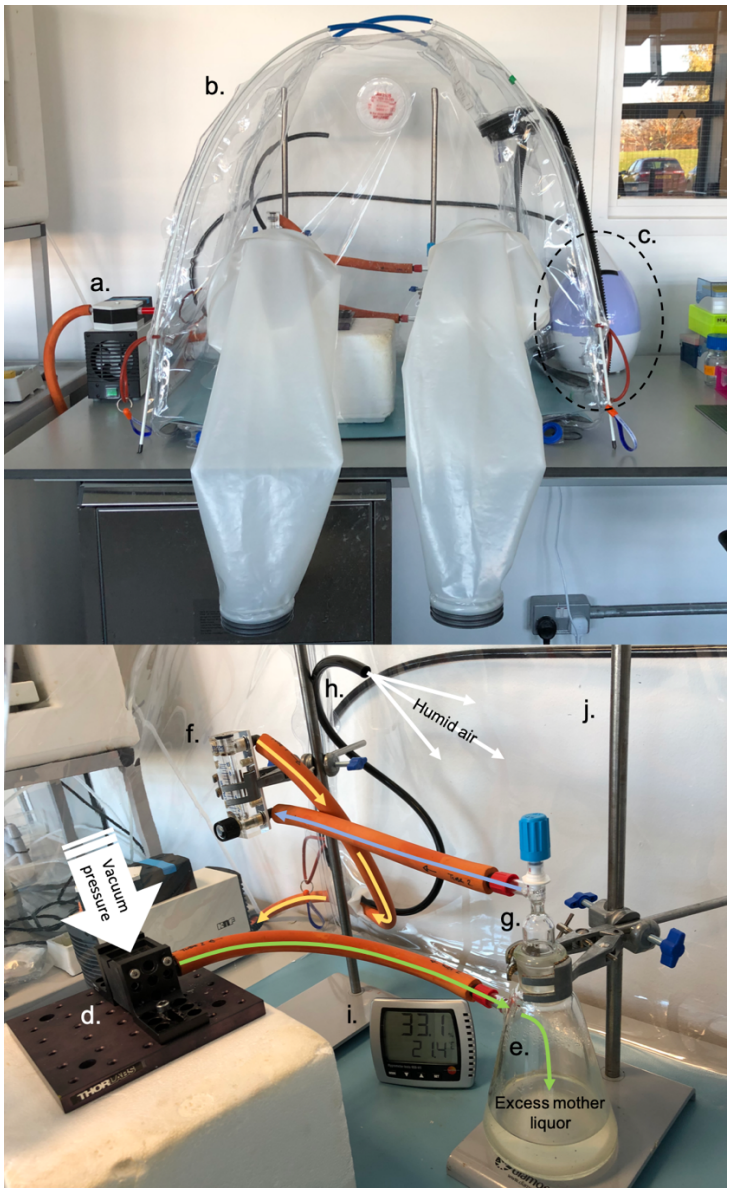


Figure 3.2. Sample loading instruments and environment.

Sample loading equipment in place at Diamond Light Source, consisting of a vacuum pump (a.), glove-box (b.), and humidifier (c.). Within the glove-box vacuum pressure is used to act on a chip loaded with crystal slurry held in a sample block (d.) attached to a Büchner flask (e., green arrow), via a pressure regulator (f., yellow arrow) attached to a stopcock (g., blue arrow). Humid air is pumped into the tent via plastic tubing attached to the humidifier (h.), and measured using a hygrometer (i.). Components were held in place using clamp stands (j.).

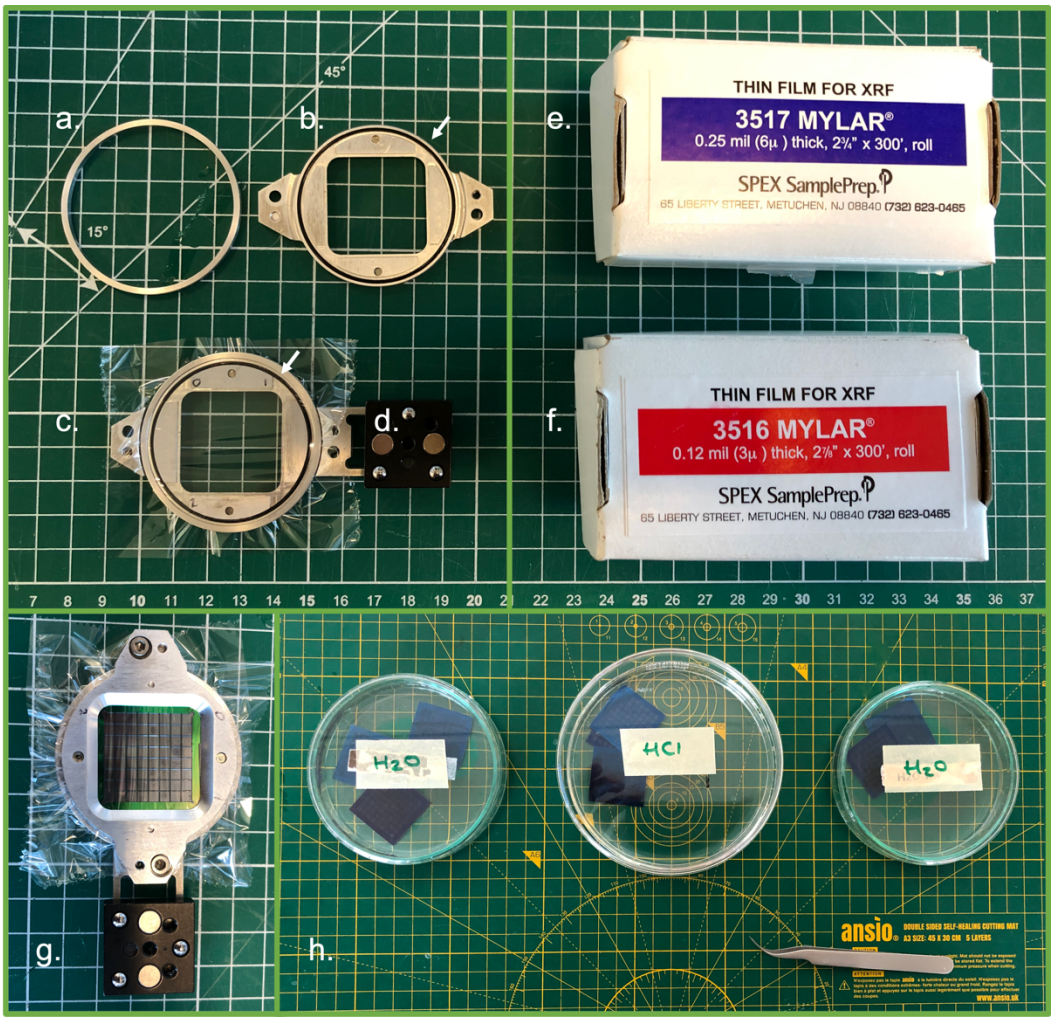


Figure 3.3. Sample holder and chip cleaning components

Sample holders utilise a metal O-ring (a.) to clamp mylar film onto a top (b.) and bottom (c.) half, with the bottom half sporting kinematic mounts (d.) that are used to attach the sample holder to the sample stages. The mylar film (6µm [e.] or 3µm [f.]) as well as rubber O-rings (white arrows) prevent a crystal-loaded chip in a sample holder closed with hex bolts (g.) from quickly drying out. Chips were cleaned using 15-minute baths in H_2O, 1M HCl, and H_2O (h.).

(fig. 3.3g). Chips are cleaned using 15-minute wash in water, 15-minute wash in 1M HCl, and a final 15-minute wash in water (fig. 3.3h).

3.4. Fixed target instrumentation and data collection

Instrumentation for the movement of chips through the X-ray beam were mounted on beamline I24, Diamond Light Source, UK and beamline BL3 EH2, SACLA, Japan. The following will refer to the setup for both environments, unless stated otherwise.

3.4.1. Sample stages

Chips in chip holders were translated through the X-ray beam using three linear positioning stages, movable in *x*, *y*, and *z* directions (fig. 3.4). The *xyz* sample stage is a collaborative design by Diamond Light Source and SmarAct (http://www.smaract.de), and have been custom-built by SmarAct using three linear translation stages, with a kinematic magnetic mount used to attach the chip holder. Each translation stage has 50 mm of travel, precise crossed roller guideways, and an integrated sensor with up to 1 nm resolution, although in practice the resolution is intentionally limited to 100 nm to allow an increased maximum velocity of 20 mm/s. Motion control of the stage setup is accomplished via a DeltaTau (http://www.deltatau.co.uk) Geobrick LV-IMS-II. Sample stages mounted upside-down at DLS only (fig. 3.5).

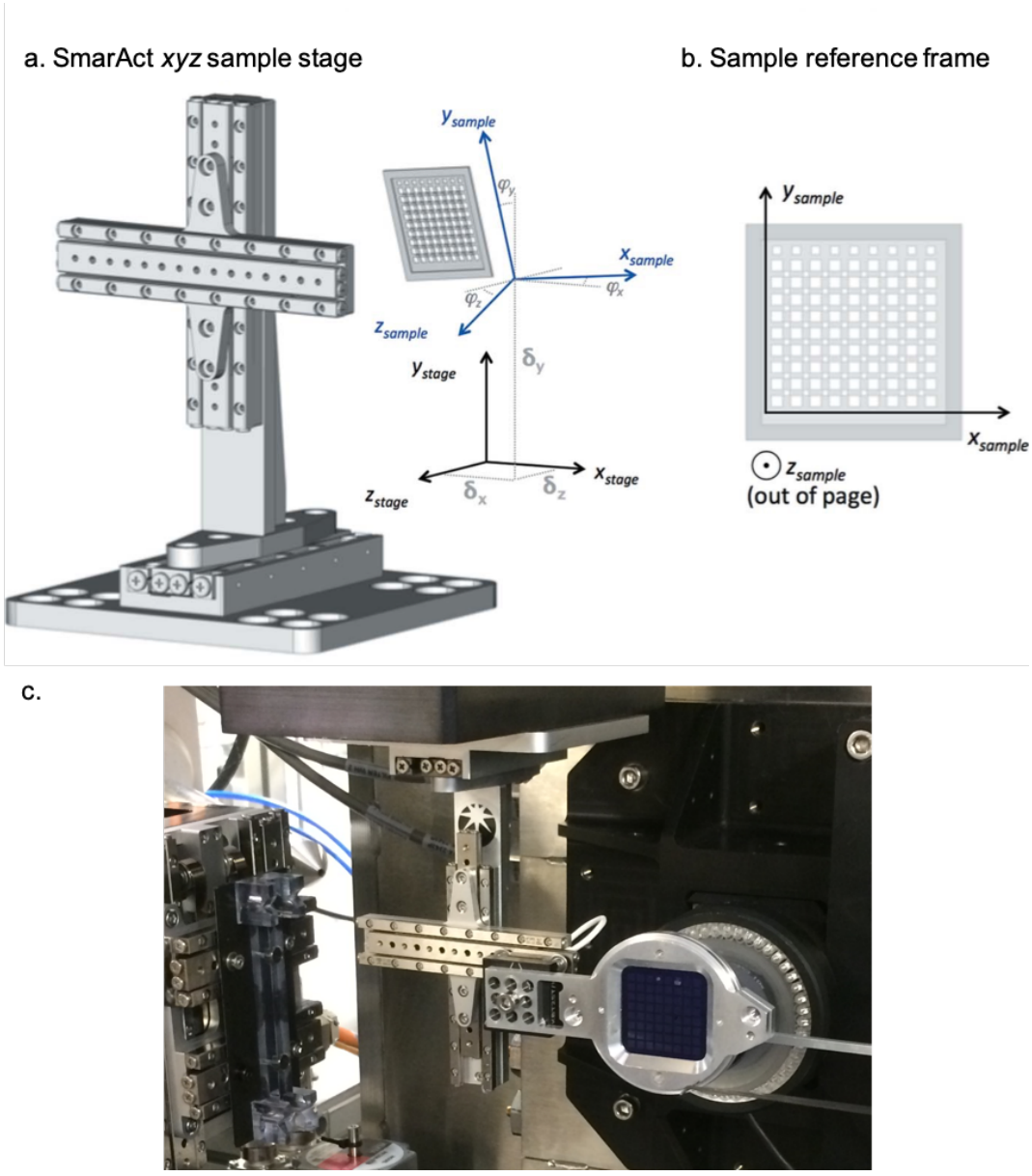


Figure 3.4. SmarAct xyz sample stages

(a.) Schematic representation of stage xyz movement of stages and chip, and (b.) how recording xyz reference positions relates to the movement of the chip. (c.) stages with attached sample holder mounted on the endstation of I24, Diamond Light Source. Figure 3.4a, b, reproduced from Sherrell et al., 2015.

3.4.2. OAV and chip alignment

Alignment of the chip via on-axis viewing is described briefly below. For an in-depth description of the development and implementation of this system refer to Sherrell et al., 2015.

On-axis viewing (OAV) provides simple and direct viewing of the chip on the stages, allowing simplified alignment of the chip in the X-ray beam. The OAV image is viewed and manipulated on the same computer running the stages, with a simple graphical user interface (GUI) utilised to interact with the sample stages (fig. 3.5). The chip alignment system uses a quick coordinate transform calculation to convert the coordinate system of the sample holder and chip to that of the stages, accounting for the rotational misalignment, depth and offset from the origin of the sample holder relative to the OAV.

This allows efficient sample exchange between data collections, despite the fact that no two chips will have the exact same alignment within their respective sample holders; this system allows a particular chip aperture to always have the same coordinates irrespective of how reproducibly the samples were mounted. This is achieved by locating the etched fiducials in three corners of the chip (fig. 3.1c, d). Each fiducial is aligned with on-screen crosshairs using a graphical user interface and brought into the focal plane of the image. The position of the stage in each fiducial position is saved, with the sharpness of the image on the screen provided by the OAV giving a good estimate of the *z* direction translation.

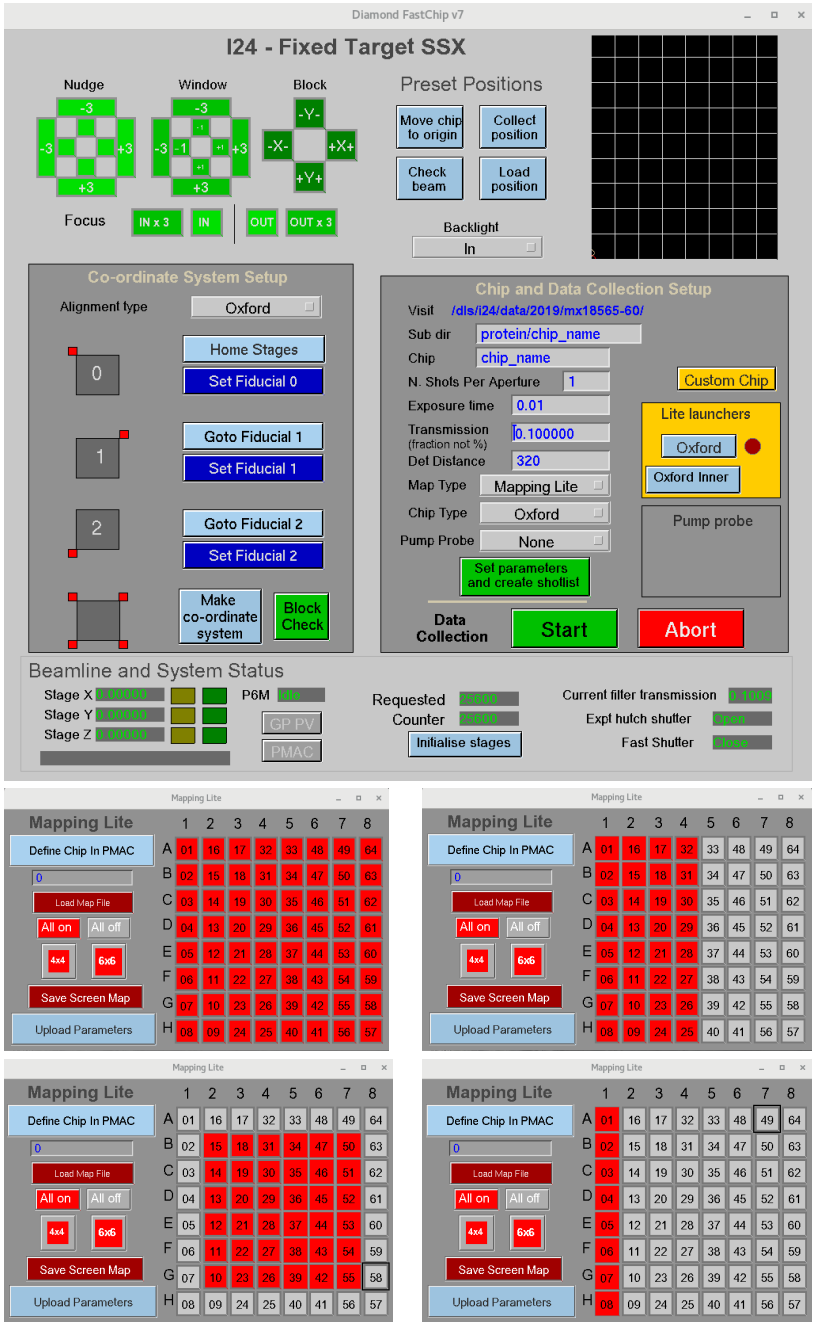


Figure 3.5. Chip alignment GUI.

Screenshot of the GUI used to interact with the stages in order to align chips to the X-ray beam. The series of green buttons (top left) were used to fine tune the location of the chip in reference to its fiducial mark, with the series of blue "Home", "Goto", and "Set" buttons used to make large movements between fiducials, and set their respective locations. Parameters defining where data are written, as-well-as data collection settings are entered within the "Chip and Data Collection Setup" section.

By providing these coordinates the offset and three-dimensional rotation of the target can then be calculated. The coefficients of the matrix A,

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

describe the transformation between the Cartesian coordinate system of the chip against the Cartesian coordinate system of the stages, including any scale factor and offset, considering of all degrees of freedom (i.e. offset, pitch, roll and yaw). Each chip axis is then defined as a combined movement of the stage axes on the PMAC controller, i.e;

$$X_{chip} = a_{11}X_{stage} + a_{12}Y_{stage} + a_{13}Z_{stage},$$

$$Y_{chip} = a_{21}X_{stage} + a_{22}Y_{stage} + a_{23}Z_{stage},$$

$$Z_{chip} = a_{31}X_{stage} + a_{32}Y_{stage} + a_{33}Z_{stage}.$$

ensuring full synchronisation of all three axes during chip motion. All motion of sample holders is then performed via requests to these virtual axes rather than movement of the individual stages.

3.4.3. Serial data collection

3.4.3.1. DLS I24

All data presented in this thesis that were collected at Diamond Light Source were collected at beamline I24 within the macromolecular crystallography village of beamlines. All serial data were collected at an X-ray energy of 12.8 keV, with beam

attenuation specified in respective experimental chapters. A Pilatus3 6M detector was used in shutterless mode in order to collect serial diffraction patterns, with the X-ray shutter left open between apertures on the chip, remaining open for the duration of the experiment. The horizontal and vertical goniometers were retracted in order to allow free movement of sample stages on the I24 endstation (fig. 3.6).

Typically, the time taken to translate the stages between chip apertures was 9 ms. Stages will move between apertures, stop in order to collect a diffraction image at that aperture, and then move to the next feature. This means it is possible to measure all 25,600 positions in under 15 minutes with a typical exposure period of 10ms. Additionally, a rapid data series can be measured from each aperture of a chip, a technique we have coined as Multiple Serial Structure (MSS) data collection. Up to twenty diffraction patterns were recorded (though not an inherent limit) at each aperture prior to translation to the next, fresh aperture. Data can then be selectively sorted into dose-dependent bins and merged, allowing dose dependent structures to be obtained. This movement and binning strategy are shown schematically in figure 3.9.

A typical exposure period of 10 ms is used for each image in an MSS series, meaning a ten-dose series (100 ms X-ray exposure per crystal) can be recorded from an entire chip in ~46 minutes. The image series at each position was individually triggered using a Keysight 33500B signal generator which in turn was triggered by a DeltaTau Geobrick LV-IMS-II stage controller when each crystal position had been reached. The X-ray shutter remained open for the duration of data collection and was not closed

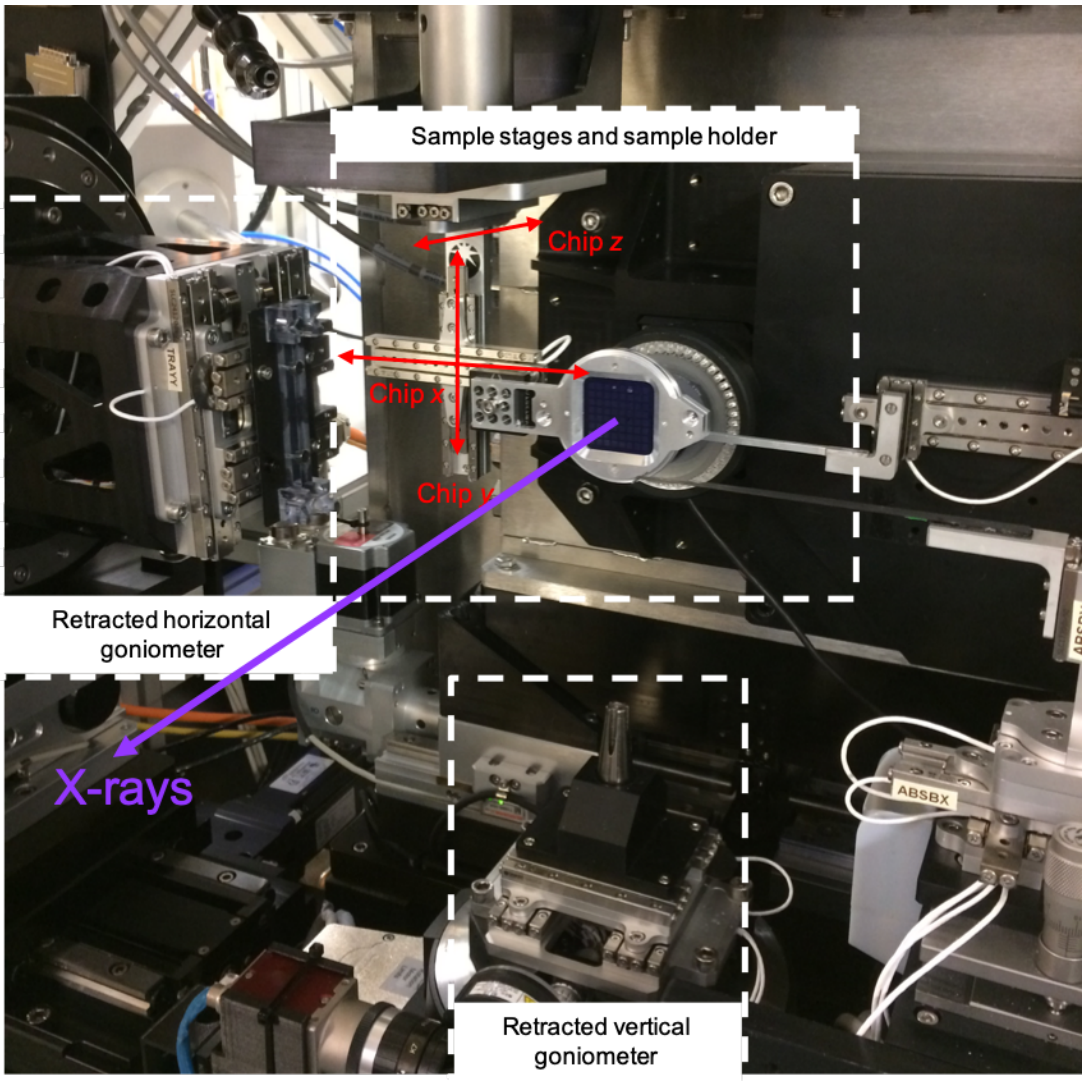


Figure 3.6. Modified I24 endstation for serial chip experiments

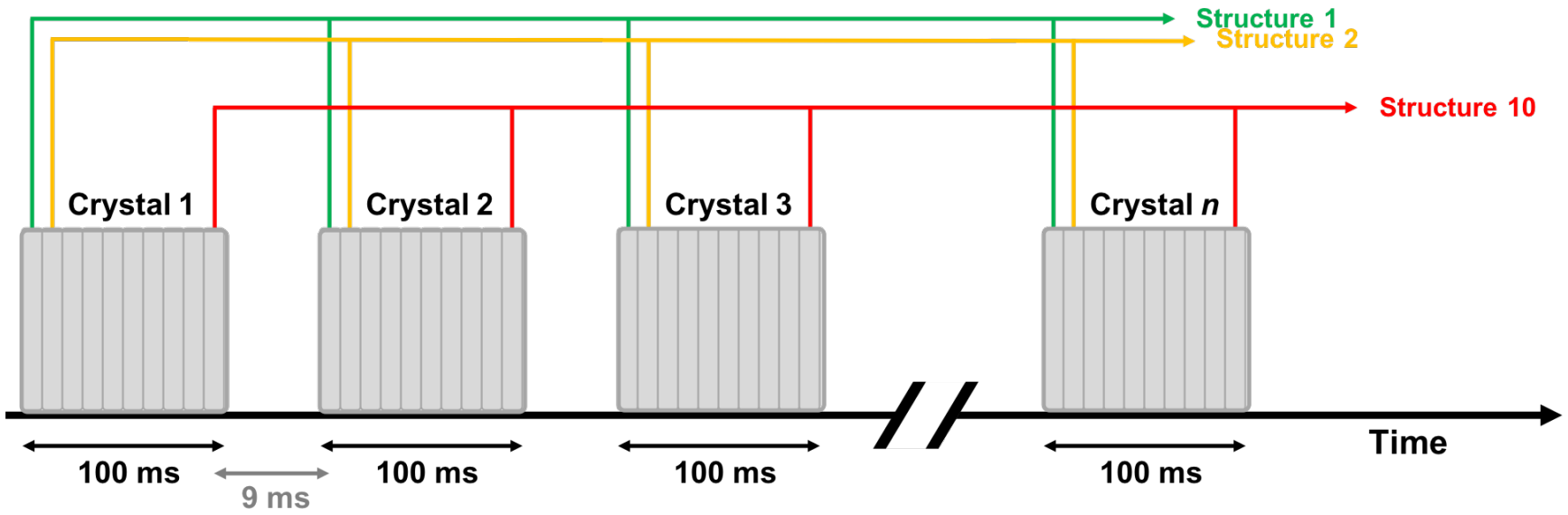


Figure 3.7. Formation of dose-resolved datasets

Multiple images collected at each crystal location allows composite datasets to be formed, achieved by selectively binning data into respective dose-resolved data-series that can be individually processed, merged, and refined.

between apertures on a chip when collecting data at I24, while data collection parameters such as exposure time, detector distance, and beam dimensions were chosen on a sample to sample basis. X-ray fluxes were measured on I24 using a silicon PIN diode as described in Owen et al., 2006, while beamsizes on I24 were measured using a knife-edge scan (Owen et al., 2009). Absorbed doses were estimated using RADDOSE-3D (Paithankar & Garman, 2010).

3.4.3.2. SACLA BL2

The SACLA beamline BL2 EH3 was utilised to collect SFX data presented in this thesis (fig. 3.8a). The presented SFX data were at an X-ray energy of 10.0 keV, using a pulse length of 10 fs and a repetition rate of 30 Hz, with a beam size of 1.2 x 1.3 µm, attenuated to 13% of full flux. Diffraction data were recorded using the SACLA MPCCD detector (Kameshima et al., 2014) (fig. 3.8b).

SFX experiments were performed in a helium chamber (355 x 420 x 400 mm) in order to minimize air scatter (fig. 3.8c). The helium chamber utilises a customised entry port so that chips could be exchanged rapidly (fig 3.8d), minimising the volume of helium gas lost to the atmosphere outside the chamber, as well as keeping a high sustained rate of data collection in order to maximise the use of allotted beamtime. A trigger for the movement of chip stages occurs 15 ms before the 30 Hz X-ray pulse. The detector readout from the X-ray pulse informs the stages translate to the next aperture (~9 ms) to be in position 5 ms before trigger, with a 2 ms safety margin on the wait time (schematic comparison with DLS I24 stage movement shown in figure 3.9).

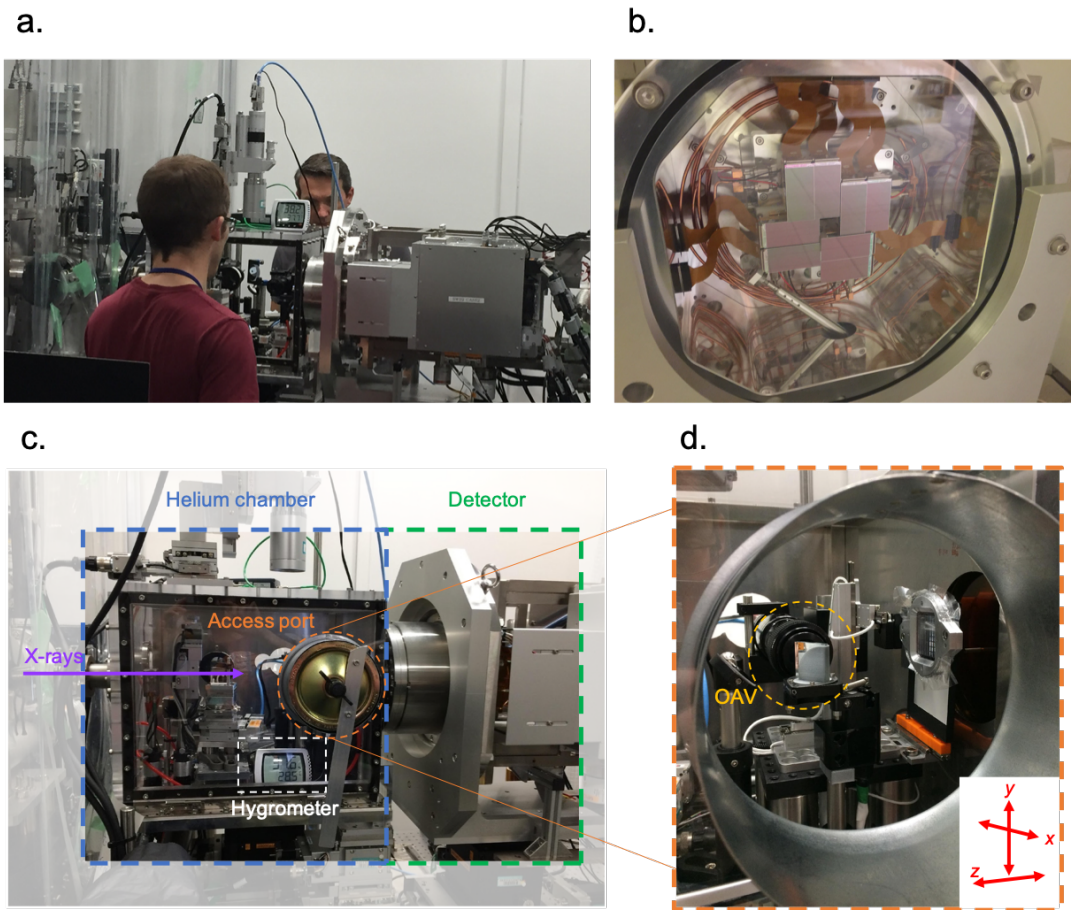


Figure 3.8. SACLA BL2 EH3 setup

(a.) Overview of experimental hutch. (b.) MPCCD detector used for the collection of SFX diffraction data (Kameshima et al., 2014). (c.) Details of helium chamber and detector setup. (d.) View of stages and OAV through access port, used for fast changeover of samples to limit helium lost from the chamber. Movement of chip stages highlighted in inset.

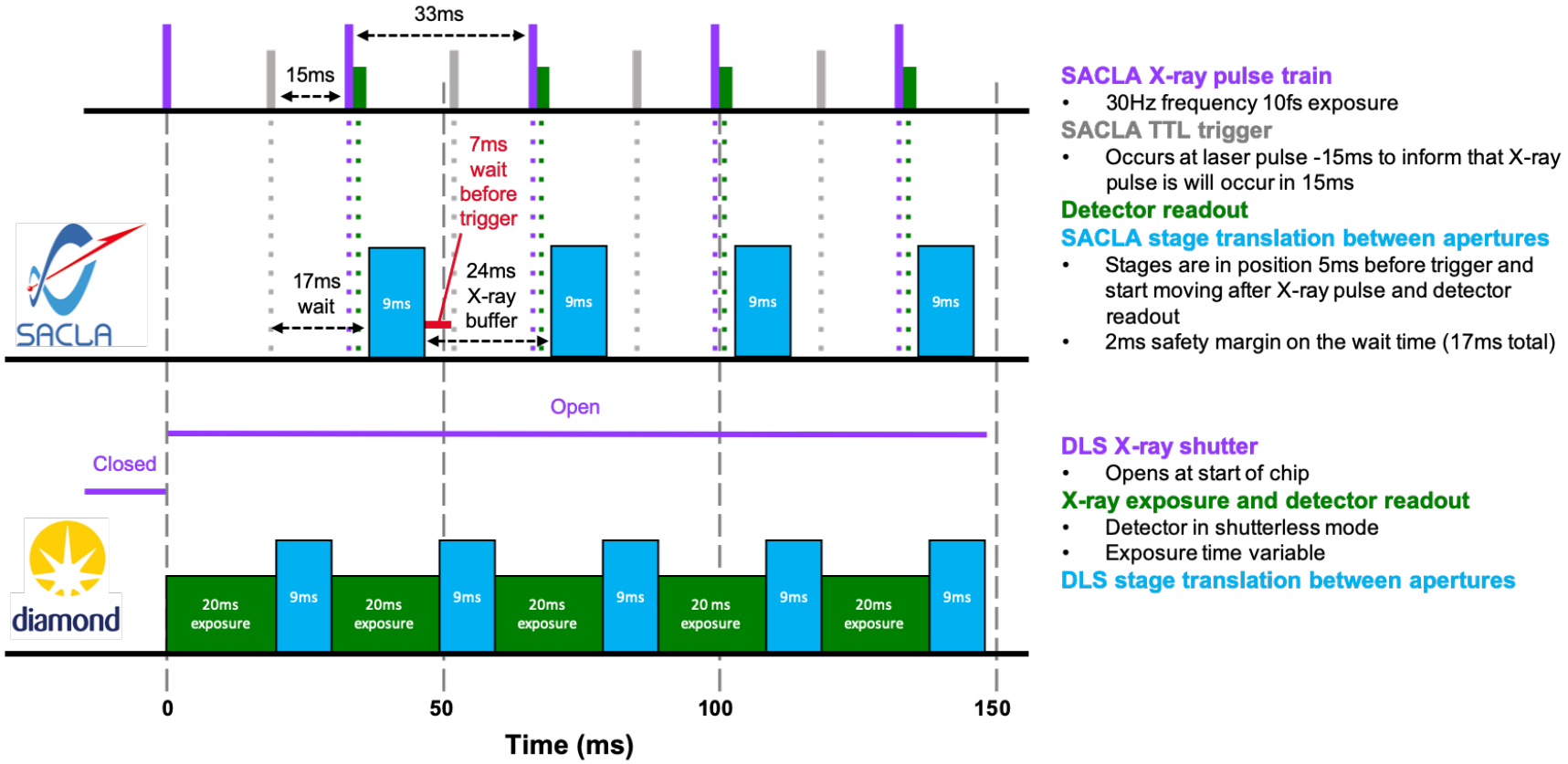


Figure 3.9. Schematic representation of SACLA and DLS stage movement and data collection timings

3.5. Data processing and refinement strategies

Serial data processing was performed using the *DIALS* (Winter et al., 2018), *PRIME* (Uervirojnangkoorn et al., 2015), and *CrystFEL* (White et al., 2012) data processing suites, with *DIALS* and *PRIME* utilised in the processing of SSX data, and *CrystFEL* for SFX data, respectively. The following section will give a general overview of command line data processing, as well as the *PHENIX* (Adams et al., 2010) and *CCP4* (Evans, 2011) suites used in structure refinement, while detailed parameters used to produce individual *mtz* files will be specified within their respective results chapters.

3.5.1. Indexing and integrating SSX data

SSX data took the form of sequentially numbered images in *CBF* format. All images were indexed and integrated using *dials.still_process* in *DIALS* version 1.8.5 using the Diamond computing cluster.

A simple bash shell script (*process.sh*) was used to define the version of DIALS to run, define that *dials.stills_process* should be used, define the path to the CBF files, and define the *process.phil* file (the file used to specify processing parameters). The *process.phil* file specifies the path to a mask used to cover the Si spots on the detector created by the X-ray diffraction through the chip apertures, whether to enable the significance filter and at what I/sigI it is defined, how to handle the known symmetry, the maximum number of lattices to index per diffraction pattern, and the known symmetry details of the crystal. Example of both scripts can be found in appendices sections 3.7.1 and 3.7.2, respectively.

In order to use the cluster the following command was used;

```
$ qsub -cwd -q low.q -pe smp 20 process.sh
```

where *-cwd* sets the current working directory, *-q* assigns the queue that the job will run in on the cluster (*low.q* used primarily, medium.q used for jobs of high importance), while *-pe smp 20* sets the number of processors being used to 20.

3.5.1. Joint refinement

In order to optimise the detector geometry after the indexing/integration for a multi-crystal experiment, a joint refinement is used. A joint refinement combines the data into a single multi-experiment, achieved by refining all of the cells at the same time, allowing the model to better fit the observed data. This can be performed on the full chip or a subsection of the chip using *dials.combine_experiments* on the output files **_refined_experiments.json* and **_indexed.pickle*:

```
$ dials.combine_experiments *_refined_experiments.json
*_indexed.pickle n_subset=2000
reference_from_experiment.average_detector=True
```

This produces the files *combined_experiments.json* and *combined_experiments.pickle*, which can be processed using *dials.refine* against a *refine.phil* file (appendices section 3.7.3).

Command line input:

```
$ dials.refine combined_experiments.json
combined_reflections.pickle refine.phil
```

This produces a *refined_experiments.json* file. This file can be opened in a text editor, and these lines searched for (n.b. numerical values differ between experiments):

```
fast_axis: {1,0.000138539,-5.3528e-05}
slow_axis: {0.000138424,-0.999998,-0.00213172}
origin: {-217.633,226.578,-248.6}
```

These parameters define the refined detector geometry and are added to the top lines of the *process.phil* file in the following format:

```
geometry {
    detector {
        panel {
            fast_axis = 1,0.0001385,-5.35280e-05
            slow_axis = 0.000138424,-0.9999977,-0.00213172
            origin = -217.63276,226.57777,-248.5997
        }
    }
}
```

By re-running *dials.stills_process* with the updated detector geometry it is possible to improve the number of integrated files.

3.5.2. Merging SSX data

Subsequent scaling and merging of SSX data was performed using *PRIME*, using a *.phil* file (*prime.phil*) to take inputs and run with a reference mtz from a rotation dataset in order to correct for indexing ambiguity. The top line of the file specifies the path to the integrated data, while the merging parameters are specified in the final two sections of the script. The *prime.phil* file (example available in appendices

section 3.7.4) is again run using the Diamond computing cluster to produce a *mean_scaled_merge.mtz* that can be used for refinement.

Bash script (*prime.sh*):

```
module load dials
prime.run prime.phil > prime.out
```

For MSS data, data needs to be separated by dose (fig 3.7). It is possible to bin these data into their separate doses within a .dat file, using the script *get_int_file.py* (appendices 3.7.5), written by Martin Appleby, a DLS year in industry student. This is run by using a text file that contains a list of *dials.stills_process* directories that contain the *int*.pickle* files needed for merging, against the arguments *-d* and *-td*; *-d* referring to the dose number, while *-td* refers to the total number of doses:

List file (*list.txt*):

```
/dls/i24/data/path/to/data/chip_1
/dls/i24/data/path/to/data/chip_2
/dls/i24/data/path/to/data/chip_3
…
```

Command line input:

```
$ module load python/ana
$ python get_int_file.py -l list.txt -d 1 -td 10
```

3.5.3. Indexing and integrating SFX data

While SSX data collected at I24 produces single, sequentially numbered *.cbf* files, SFX data collected at SACLA is output in *HDF5* format. In order to handle these data *HDF5* files were indexed and integrated using *indexamajig* in *CrystFEL* version 1.8.5 using the command line.

As with SSX data, simple bash shell scripts were utilised to speed up data processing by sending jobs to the Diamond computing cluster. A bash script was used for indexing (*index.sh*); this defines the version of *CrystFEL* to run, that *indexamajig* should be used, and the the hdf5 file locations (*file.lst*), the geometry file location (**.geom*), the peakfinding algorithm (e.g. *zaef*), the unit cell parameter location (**.cell*), the output file (**.out*), and the number of cores to use on the cluster (*-j 20*). None of the other input parameters were edited, unless stated otherwise in their respective chapters. An example of the shell script can be found below:

Bash script (*index.sh*):

```
$ module load CrystFEL
$ indexamajig -i files.lst --peaks=zaef -threshold=300 --min-
gradient=100000 --min-snr=5 --int-radius=3,4,5 --indexing=asdf
-g *.geom -p *.cell -o *.out -j 20
```

In order to correct for indexing ambiguity *ambigator* was run against the **.out* file from *indexamajig* with a defined point group (e.g. *-y 23*). *--operator* is the symmetry operator for alternate indexing, while *--fg-graph* outputs a table of

correlation coefficients. A **_reindexed.out* file is output after running the following bash script:

Bash script (*ambig.sh*):

```
$ module load CrystFEL
$ ambigator -i *.out -y 23 --operator=k,h,-l --fg-graph=index-
amb-plot -o *_reindexed.out -j 20
```

The *detector_shift* script in *CrystFEL* can be used to refine the detector geometry by updating the beam X, Y, position, using the command:

```
$ ./detector-shift *_reindexed.out 23382-1.geom
```

3.5.4. Merging SFX data

process_hkl in *CrystFEL* was utilised for merging and scaling of SFX data, implementing scaling and post-refinement with no external reference data set. The input file is the **.out* file (*reindexed.out* in *ambigator*), and also uses the same the symmetry, *process.hkl*, and performed initially on the **.out* using a simple bash script *merge.sh*, written by Dr. Danny Axford of DLS beamline I24:

```
$ inp=SFX_10000.out
$ out=SFX_10000.hkl
$ pg="m-3"

$ module load CrystFEL
```

```
$ process_hkl -y $pg --push-res=1.6 --scale --odd-only -i $inp
-o ${out}1 -j 20 &
$ process_hkl -y $pg --push-res=1.6 --scale --even-only -i $inp
-o ${out}2 -j 20 &
$ process_hkl -y $pg --push-res=1.6 --scale -i $inp -o ${out} -
j 20 &
wait
```

Merging via *process_hkl* using *merge.sh* outputs three *hkl* files, **.hkl*, **.hkl1*, and **.hkl2* (**.hkl1* and **.hkl2* each corresponding to half the crystals in the dataset). These data are used to check the merging statistics using `stat.sh`, a script written by Dr. Danny Axford from DLS beamline I24 (example shown for AcNiR):

```
#!/bin/bash

inp=96p4_1p68_ambig.hkl
inp1=96p4_1p68_ambig.hkl1
inp2=96p4_1p68_ambig.hkl2
basename=96p4_1p68_ambig
fom="R1I R2 Rsplit CC CCstar"
pdb="AcNIR96p4.cell"
pg="m-3"
highres="1.68"

module load CrystFEL

if [ ! -d "stat" ]; then
      mkdir stat
fi

for mode in $fom
do
    compare_hkl $inp1 $inp2 -y $pg -p $pdb --fom=$mode --
highres=$highres --nshells=20 --shell-file="stat/${basename}-
$mode".dat 2>>stat/${basename}.log
done
```

```
check_hkl -p $pdb --nshells=20 --highres=$highres -y $pg  --
shell-file="stat/${basename}-shells".dat $inp
2>>stat/${basename}.log
```

An *mtz* file that can be used for refinement, using the *create-mtz* script (see appendices 3.7.6) via the following command:

```
$./create-mtz *_unity.hkl
```

3.5.5. Structure solution and refinement

Structure solution via molecular replacement was handled using *Phaser* within the *PHENIX* suite (Adams et al., 2010), from a published starting model, or a model refined from a rotation dataset. Structures were refined using *phenix.refine* and were rebuilt in *COOT* (Emsley et al., 2010) between rounds of refinement. Validation was performed using tools built within *COOT* (e.g. difference map peaks, check waters, etc.), Molprobity (Chen et al., 2010), the JCSG QCCheck server (https://smb.slac.stanford.edu/jcsg/QC/) and the PDB Validation server (https://validate-rcsb-1.wwpdb.org/). Side chain atoms not supported by electron density were typically deleted from the model unless stated otherwise. Coordinates and structure factors were deposited in the RCSB Protein Data Bank (https://www.rcsb.org/) where appropriate, with accession numbers given within their respective results chapters. Surface areas and volumes were calculated in *3VEE* Volume Assessor (Voss & Gerstein, 2010).

3.6. Online UV-vis spectrophotometry

In situ UV–Vis absorption spectra were collected on beamline I24 at Diamond Light Source using mirror lenses (Bruker) mounted in an off-axis geometry and a deuterium halogen light source (Ocean Optics). Spectra were recorded over the wavelength range of 300 – 800 nm using a Shamrock 303 imaging spectrograph (Andor). Onine UV-vis data presented in this thesis were collected using an X-ray energy of 12.8 keV, an X-ray beamsize of 30x30 µm (FWHM) and a UV-Vis focal spot 40 µm in diameter. An exposure time of 10 ms was utilised for static spectra, with two 10ms accumulations used per kinetic spectrum.

3.7. Appendices

3.7.1. Script process.sh

Bash script `process.sh`:

```
$ module load dials/
$ dials.stills_process /path/to/images*.cbf process.phil
```

3.7.2. Script process.phil

Processing defaults file `process.phil`:

```
# mask to cover Si spots and beamstop appropriate to detector distance
spotfinder.lookup.mask=/dls/i24/data/path/to/mask
integration.lookup.mask=/dls/i24/data/path/to/mask

significance_filter.enable=True
significance_filter.isigi_cutoff=1.0


indexing.stills.refine_candidates_with_known_symmetry=True

# protein name
indexing {
  known_symmetry {
    space_group = P213
    unit_cell = 195.59 195.59 195.59 90 90 90
  }
  #for good quality data with many images indexable up to 3 lattices,
can run for up to 6 max lattices.
  multiple_lattice_search.max_lattices=3
  refinement_protocol.d_min_start=2.5
  basis_vector_combinations.max_refine=5
}
```

3.7.3. Script refine.phil

Refinement defaults file `refine.phil`:

```
refinement {
    parameterisation {
        auto_reduction {
            min_nref_per_parameter = 3
            action = fail fix *remove
        }
        beam {
            fix = *all in_spindle_plane out_spindle_plane wavelength
        }
        detector {
            fix_list = Tau1
        }
    }
    refinery {
            engine = SimpleLBFGS        LBFGScurvs        GaussNewton
LevMar *SparseLevMar
```

```
    }
    reflections {
        outlier {
            algorithm = null auto mcd tukey *sauter_poon
            separate_experiments = False
            separate_panels = True
        }
    }
}
```

3.7.4. Script prime.phil

Merging inputs file `prime.phil`:

```
data = /path/to/data/
run_no = name_of_run
title = TITLE sad anom=true cut 1.9A frames 0-1909
scale {
  d_min = 2.0
  d_max = 6.0
  sigma_min = 1.5
}
indexing_ambiguity {
  flag_on = True
  mode = Auto
  index_basis_in = rotation.mtz
}
postref {
  scale {
    d_min = 2.0
    d_max = 6.0
    sigma_min = 1.5
    partiality_min = 0.1
  }
  crystal_orientation {
    flag_on = True
    d_min = 2.0
    d_max = 45
    sigma_min = 1.5
    partiality_min = 0.1
  }
  reflecting_range {
    flag_on = True
    d_min = 2.0
    d_max = 45
    sigma_min = 1.5
    partiality_min = 0.1
  }
  unit_cell {
    flag_on = True
    d_min = 2.0
    d_max = 45
    sigma_min = 1.5
    partiality_min = 0.1
    uc_tolerance = 3
  }
  allparams {
    flag_on = False
    d_min = 0.1
    d_max = 99
```

```
    sigma_min = 1.5
    partiality_min = 0.1
    uc_tolerance = 3
  }
}
merge {
  d_min = 1.9
  d_max = 45
  sigma_min = -1.5
  partiality_min = 0.1
  uc_tolerance = 3
}
target_anomalous_flag = False
flag_apply_b_by_frame = False
target_unit_cell = 96.4,96.4,96.4,90,90,90
target_space_group = P213
n_residues = 340
pixel_size_mm = 0.172
```

3.7.5. Script get_int_file.py

Sorting int*.pickle files `get_int_file.py`:

```
#MODULES
import os
import numpy as np
import pandas as pd
import argparse


def argarser():
    parser = argparse.ArgumentParser( description="blah")
    parser.add_argument( "-l", "--input_list",
                         help="if you use this flag you must have a
input list file" )
    parser.add_argument( "-i", "--no_of_images", type=int,
                         help="if you want to select a random sample of
of images give number" )
    parser.add_argument( "-s", "--stills_directory", type=str,
                         help="stills processing directory" )
    parser.add_argument("-d", "--dose", type=int,
                         help="dose number in a dose experiment")
    parser.add_argument("-td", "--total_dose", type=int,
                         help="number of doses")
    args=parser.parse_args()
    if args.input_list and args.stills_directory:
       parser.error("cannot have two inputs -l and -s")
    elif args.input_list or args.stills_directory:
       pass
    else:
       parser.error("must have an input: -l or -s")
    if args.dose and args.total_dose:
       pass
    elif args.dose or args.total_dose:
       parser.error("-d and -td are both needed")
    return args

def get_int_pickles( still_dir , dose=None, doses=None):
    # create empty np array for int+pwds
```

```
    pickle_list = np.array( [ [ ] ] )
    # finds all pwd and files in the still dir
    for path, dirs, files in os.walk( still_dir ):
        for name in files:
            # searches for int files
            if name.startswith( "int" ) and name.endswith( ".pickle"
 ):
                if dose:# creates complete int.pickle path
                    integer = int(name.rstrip('.pickle').split('_')[
1])
                    if (integer%doses) +1 == dose:
                        pickle_pwd = os.path.join( path, name )
                        # inputs pickle_pwd into np array
                        pickle_list_1 = np.array( [ [ pickle_pwd ] ]
)
                        pickle_list = np.concatenate(  ( pickle_list,
pickle_list_1 ), axis=1 )
                else:
                    pickle_pwd = os.path.join( path, name )
                    # inputs pickle_pwd into np array
                    pickle_list_1 = np.array( [ [ pickle_pwd ] ] )
                    pickle_list = np.concatenate(    (    pickle_list,
pickle_list_1 ), axis=1 )
    # returns single column np array of int_pwds
    return pickle_list

def main( args ):
    # cols for still_dirs df
    cols = [ "still_dir" ]
    # if -l
    if args.input_list:
        still_dirs = args.input_list
        still_df = pd.read_csv( still_dirs, names=cols )
    elif args.stills_directory:
        still_dir = [ args.stills_directory ]
        still_df = pd.DataFrame( still_dir, columns=cols )
    # empty np array for pickles
    pickle_list = np.array( [ [ ] ] )
    print "dir searching for int files:"
    for still_dir in still_df[ "still_dir" ]:
        print still_dir
        pickle_list_1 =  get_int_pickles(    still_dir,    args.dose,
args.total_dose )
        pickle_list = np.concatenate( ( pickle_list, pickle_list_1 ),
axis=1 )
        print "done"
    pickle_list = np.transpose( pickle_list )
    pickle_len = len( pickle_list )
    if args.no_of_images:
        images = args.no_of_images
    else:
        images = pickle_len
    cols = [ "still_pwd" ]
    pickle_df = pd.DataFrame( pickle_list, columns=cols )
    sample_df = pickle_df.sample( images )
    if args.dose:
       file_name = "prime_input_{0}_images_dose_{1}.dat".format(
images, args.dose )
    else:
        file_name = "prime_input_{0}_images.dat".format( images )
    sample_df.to_csv( file_name, header=False, index=False )
```

```
args = argarser()
main( args )
```

3.7.5. Script create-mtz

```
#!/bin/sh

# When you've edited the relevant parameters, delete this comment and
the following two lines
echo "You need to edit this script first, to set the space group and
cell parameters."
exit 1

OUTFILE=`echo $1 | sed -e 's/\.hkl$/.mtz/'`

echo " Input: $1"
echo "Output: $OUTFILE"
if [ -e $OUTFILE ]; then
        echo "   The output file already exists:"
        echo "   " $OUTFILE
        echo "    To confirm that you want to continue, which will
DESTROY the"
        echo "    current contents of this file, type 'y' and press
enter."
        read conf
        if [ $conf != y ]; then
                echo "Not confirmed."
                exit 1
        else
                echo "Proceeding"
        fi
fi

sed -n '/End\ of\ reflections/q;p' $1 > create-mtz.temp.hkl

echo "Running 'f2mtz'..."
f2mtz HKLIN create-mtz.temp.hkl HKLOUT $OUTFILE > out.html << EOF
TITLE Reflections from CrystFEL
NAME PROJECT wibble CRYSTAL wibble DATASET wibble
CELL 100 100 100  90  90 90
SYMM P1
SKIP 3
LABOUT H K L IMEAN SIGIMEAN
CTYPE  H H H J     Q
FORMAT '(3(F4.0,1X),F10.2,10X,F10.2)'
EOF

if [ $? -ne 0 ]; then echo "Failed."; exit; fi

rm -f create-mtz.temp.hkl
echo "Done."
```

Chapter 4: New methods for resolving radiation-driven effects in microcrystals

4.1 Introduction

The advancement of large crystal room-temperature techniques in X-ray crystallography in recent years allows the determination and subsequent understanding of atomic-level protein structures to be completed with relative ease and confidence. Determining room temperature structures as both 'high-quality' (low R-factors, high $CC_{1/2}$ etc.) and free from radiation damage effects in micro-crystal X-ray crystallography however presents more of a challenge; this is due to crystal-to-crystal variation being more apparent when using a multi-crystal approach, a necessity when considering microcrystals as the primary sample for data collection.

X-ray induced changes are a limiting factor when exploring metalloproteins, where X-ray induced changes can occur rapidly at the active site, with radiation damage to a protein crystal sample resulting from the absorption of photons from the X-ray beam by either the photoelectric effect or Compton scattering (Garman, 2010) (see chapter 1.1.1). Interestingly, the accumulation of radiation damage as the diffraction experiment proceeds can lead to non-isomorphism within the crystal, due in part to the expansion of unit-cell volume, and movement of the protein molecule within the unit-cell (expressed as an increase in B-factors). This effect can be somewhat mitigated via collection at 100K (see chapter 1.1.2, 1.1.3), however in order to reveal protein features or reactions potentially hidden by cryogenic data collection, room

temperature data collection requires that data are merged from multiple crystals due to a significant reduction in comparative crystal lifetimes.

A general increase in error arises when using multi-crystal merging to produce a complete dataset due to non-isomorphism. Initially discovered by Elihard Mitscherlich in 1819 who found that the crystal forms of specific hydrated potassium phosphate and arsenate salts were identical, he described isomorphism as how identical crystals are in shape (Melhado, 1980). The modern definition more accurately denotes isomorphous crystals as crystals with the same space group and unit cell dimensions. Moreover, isomorphism can still exist when the type and position of atoms in both crystals are the same except for a replacement of one (or more) atoms in one structure with different types of atoms in the other, for example, heavy atoms. Crystal polymorphism adds to the non-isomorphism issue: polymorphic crystals occur where crystals of the same molecule exist across two or more crystalline forms; this means that crystals of the same protein can differ in crystal form, space group, and/or unit cell dimensions, and may be caused by differences in crystallisation conditions such as precipitant, buffer, pH value, as well as additives (Bonnefond et al., 2011).

Non-isomorphism within crystals can have effects on data collection and subsequent analysis, being a source of disagreement between symmetry-related reflections. Consequently, datasets collected from more than one crystal of the same form may show differences between them, owing to differences in the structure factors (Borek, Minor, & Otwinowski, 2003). Issues with non-isomorphism have been longstanding; in the late 1950s Crick and Magdoff identified that a small change of 0.5% in unit cell

parameters, can cause an average change in the intensity of structure factors of > 15% (Crick & Magdoff, 1956). Additionally, through exhaustive processing of data taken from lysozyme crystals in the late 1960s, it was recognised that high-resolution measurements had been made from two different types of crystals with essentially identical unit-cell dimensions; the subtly different diffraction patterns, subsequently lead to the identification of type I and type II lysozyme (Blake et al., 2012, 1965).

The aforementioned issues concerning crystal isomorphism mean that robust data collection and processing methods are necessary to identify non-isomorphous crystals in serial crystallography. As many tens of thousands of crystals are utilised to make complete, composite datasets, innovative methods are also needed to identify how these effects can progress via accumulated dose. This chapter aims to combine SSX and MSOX into a method we have coined Multiple Serial Structures (MSS). By collecting multiple low-dose datasets from individual locations on the fixed target, it is possible to produce a series of dose dependent structures, allowing effects of radiation damage to be tracked on a target protein as dose is accumulated. This was achieved by adapting the pre-existing MSOX method of data collection (Horrell et al., 2016; Horrell et al., 2018) to the fixed target data collection method as outlined in Owen et al., 2017, and Sherrell et al., 2015 (Owen et al., 2017; Sherrell et al., 2015).

AcNiR, a copper nitrite reductase from the organism *Achromobacter cycloclastes* (see chapter 1.3.1) was chosen for this series of experiments. AcNiR has been studied by the Hough group at the University of Essex extensively, and used for the multiple structures from one crystal study by Horrell et al. in 2016 due to it undergoing specific

electron driven reaction steps, showing ideal potential of the protein crystals to be durable when collecting multiple dose-dependent datasets (Horrell et al., 2016; Horrell et al., 2018). The reduction of NO^{2-} to NO is the first committed (i.e. irreversible reaction that commits the organism to produce the final product) step in the denitrification pathway (equation 7), and is the process carried out by CuNiRs. Further, this reaction being electron driven makes AcNiR an ideal candidate for low-dose structure study

$$NO_2{}^- + e^- + 2H^+ \rightarrow NO + H_2O$$

Equation 7

4.2 Methods

4.2.1 AcNiR purification and microcrystal optimisation

Recombinant AcNiR was expressed and purified as described in detail in chapter 3. Purified AcNiR in 20 mM Na citrate pH 6 was concentrated to 20 mg/ml^{-1}, quantified by its absorbance at 280 nm. As large volumes of microcrystals are needed for serial crystallography experiments, batch microcrystal optimisation was used to elucidate optimal large-scale microcrystal conditions. This was achieved by applying different ratios of protein to buffer, protein concentrations, and crystallisation temperatures from a starting point of 2.5M ammonium sulphate, 0.1M sodium citrate pH 4.5 (table 4.1).

Batch microcrystals used for experimentation were prepared by rapidly mixing 20mg/ml AcNiR in 20mM Tris pH 7.5 with a solution consisting of 2.5M ammonium sulphate, 0.1M sodium citrate pH 4.5 buffer in a ratio of 1:3 and mixing by vortexing for 30 s. Microcrystals with a diameter of 5-15µm grew at room temperature over a period of 4-6 days. Microcrystal suspensions were centrifuged at 800 rev min^{-1} for 30 seconds to sediment the crystals; the crystallisation buffer was then removed and replaced with a storage buffer consisting of 1.6M ammonium sulphate, 0.1M sodium citrate pH 4.5. Crystals were soaked in a solution of mother liquor supplemented with 100mM sodium nitrite for a duration of 20 minutes prior to loading onto the chip. Serial dilutions were achieved by adding additional storage-buffer solution.

Test #	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Protein concentration (mg/ml)	20	20	20	50	50	50	50	50
Ratio (Protein:Buffer)	1:3	1:1	1:2	1:1	1:1.5	1:1	1:2	1:1.5
Crystallisation temperature (°C)	RT	RT	RT	RT	RT	RT	RT	RT
Vortex time (seconds)	30	30	30	30	30	0	0	0
Test #	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Protein concentration (mg/ml)	50	20	20	20	20	20	50	50
Ratio (Protein:Buffer)	1:1.25	1:1	1:1.25	1:1	1:1	1:1	1:1	1:1.25
Crystallisation temperature (°C)	RT	4°C	4°C	4°C	4°C	18°C	18°C	18°C
Vortex time (seconds)	0	0	0	0	0	0	0	0

Table 4.1. AcNiR batch microcrystal optimisation

4.2.2 Data collection

Data in this chapter were collected using the fixed-target serial crystallography methods as described in chapter 3 sections 3.3, 3.4, and 3.5, and therefore only the specific parameters to the experiments performed within this chapter will be referred to in detail.

4.2.2.1 Sample loading, beamline instrumentation and parameters

Fixed target chips were loaded as described in chapter 3.3.2, with instrumentation used for the movement of chips through the X-ray beam on beamline I24 also described in detail in chapter 3.4. For this experiment a beamsize of 8 x 8 µm (full width half maximum; FWHM) was used, with all data collected at 12.8 keV using a PILATUS3 6M detector with a crystal-to-detector distance of 310 mm. Beam flux was 3 x 10^{12} photons s^{-1}, measured immediately prior to the experiments using a silicon PIN diode (Owen et al., 2009), and attenuated tenfold for the described data collections (3 x 10^{11} photons s^{-1}). The diffraction weighted dose absorbed by each crystal was estimated using *RADDOSE-3D* (Paithankar & Garman, 2010).

4.2.2.2 Data collection strategies

A multiple serial structure (MSS) method of data collection was developed and utilised in this chapter. Collection of multiple serial structures differs from the typical fixed target methodology – rather than a single diffraction image being collected from each aperture position prior to translation to a fresh aperture, multiple diffraction images are recorded at each position. An exposure period of 20 ms (corresponding to an absorbed dose of ~11 kGy) was utilised for each image meaning a twenty-image series

totalled 400 ms (~220 kGy) X-ray exposure per crystal. The number of images to be collected at each position was selected using a GUI (see chapter 3, figure 3.5), with the image series at each position triggered using a Keysight 33500B signal generator, which in turn was triggered by a DeltaTau Geobrick LV-IMS-II stage controller when each crystal position had been reached. The X-ray shutter remained open during the entire duration of the data collection (shutterless mode).

4.2.2.3 Data processing, structure solution and refinement

Data were written to disk in the form of sequentially numbered images in *CBF* format, with all images indexed using *dials.stills_process* in *DIALS v1.8.5* (Winter et al., 2018). Data from each position were selectively binned by dose using the script *get_int_file.py* (appendices 3.7.5), and subsequently merged into dose-dependent structures, shown schematically in figure 3.7. Subsequent scaling and merging steps were performed using *PRIME* (Uervirojnangkoorn et al., 2015). Indexing ambiguity in space group $P2_13$ was resolved using a reference dataset, with structures refined from the same starting model of AcNiR-NO_2^- from which water and nitrite had been removed (PDB: 5I6L; Horrell et al., 2016) using the *PHENIX* crystallography suite (Adams et al., 2010), with structures rebuilt between iterative rounds of refinement using *COOT* (Emsley et al., 2010). Structure validation was performed using the JCSG Quality Control Check server (https://smb.slac.stanford.edu/jcsg/QC/) and the PDB validation server (https://validate-rcsb-1.wwpdb.org/). Coordinates and structure factors were deposited in the RCSB Protein Data Bank. Surface areas and volumes were calculated in *3VEE* (Voss and Gerstein, 2010).

4.3 Results

4.3.1 AcNiR purification and microcrystal optimisation

The expression and purification of AcNiR has been explored and refined through previous studies (Antonyuk et al., 2005; Horrell et al., 2016), making the process of expression straightforward for this body of work. SDS PAGE was utilised in order to clarify the level of contamination in the sample after DEAE ion exchange chromatography (fig. 4.1a, revealing that further purification was required. The samples were then recombined and concentrated to under 10mL in order to be passed through a Sephadex G-75 size exclusion chromatography column, which was effective in removing the remaining contamination (fig. 4.1b).

This purified protein was used to elucidate microcrystal conditions in batch. Out of the conditions evaluated in table 4.1, only conditions 1 (20mg/ml AcNiR:2.5 M Ammonium Sulphate and 0.1 M Sodium Citrate pH 4.5 [1:3]) and 3 (20mg/ml AcNiR:2.5 M Ammonium Sulphate and 0.1 M Sodium Citrate pH 4.5 [1:2]) gave appropriate batch microcrystal slurries, producing crystals ranging from 10-20 µm at a concentration of 3.5×10^6 after 4 days for condition 1 (fig. 4.2), and 6.4×10^5 after 6 days for condition 3, respectively. As condition 1 gave a greater quantity of microcrystals in the slurry this was selected as the condition to repeat and scale up. Repeats of condition 1 gave identical results at both 200µL and 4 mL total volumes.

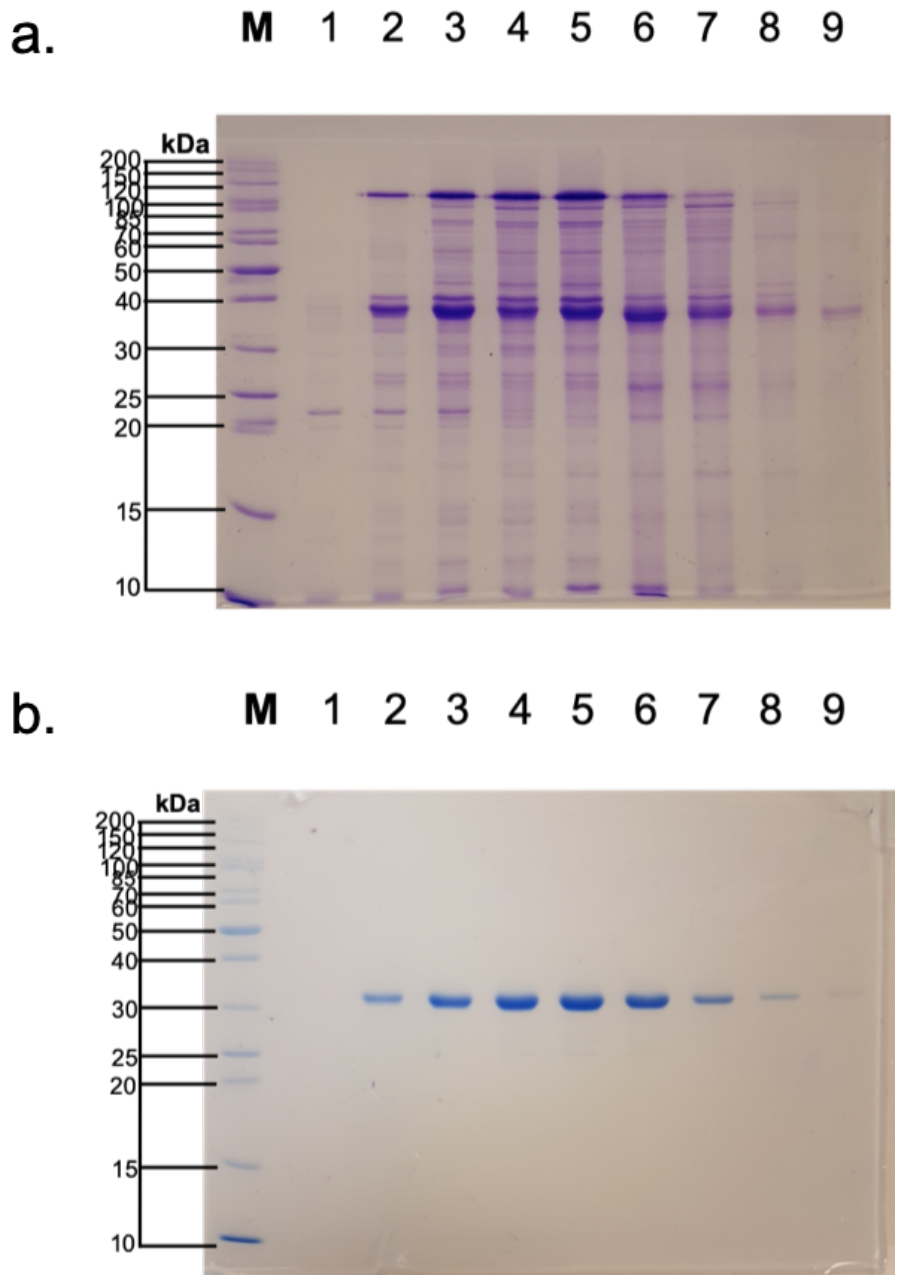


Figure 4.1. AcNiR SDS PAGE

(a.) Ion exchange chromatography via DEAE sepharose shows strong bands for monomer (~37 kDa) and trimer (~117 kDa) AcNiR, though also indicate that further purification is required due to a large number of contaminants. (b.) Size exclusion chromatography via Sephadex G-75 show strong bands for AcNiR with extremely low levels of contamination.

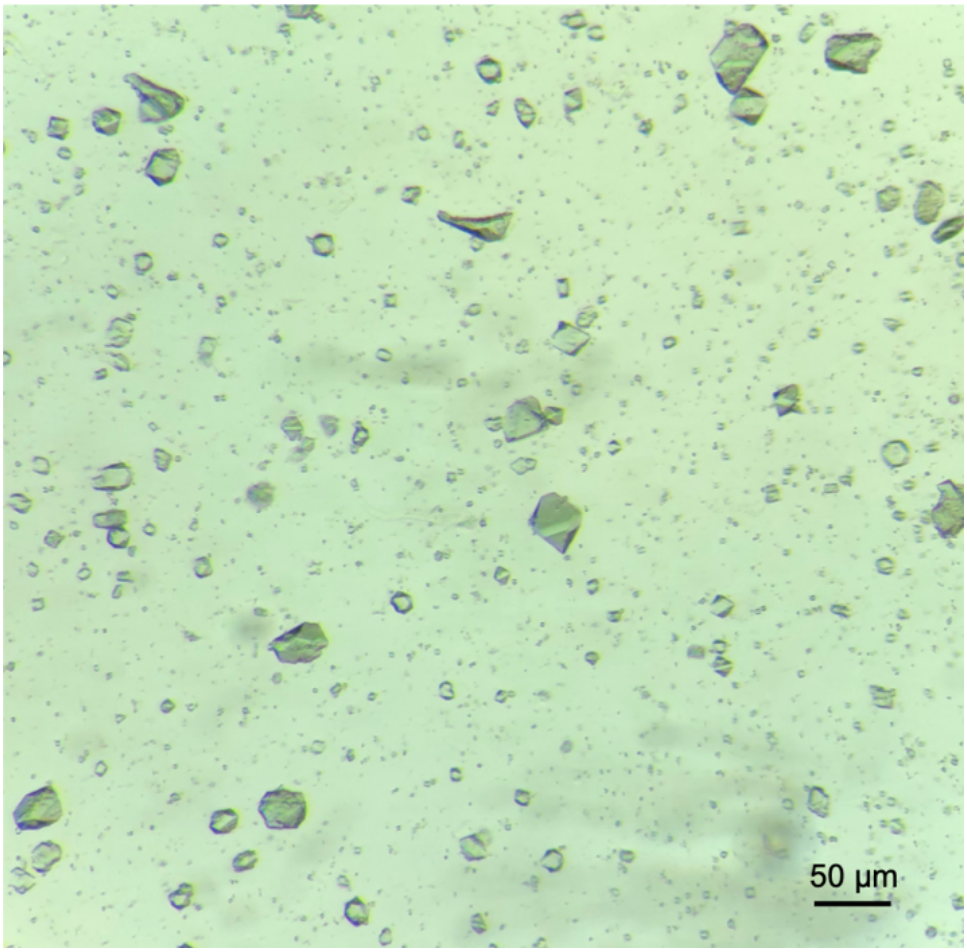


Figure 4.2. AcNiR microcrystals

Batch grown AcNiR protein crystals vary in size, with cubic crystals between 5 – 30 µm commonly seen, with the majority of crystals ~ 8 µm.

4.3.2 MSS data collection, processing, and discrimination between crystal polymorphs

By using an MSS method of data collection, it was possible to collect a total of 20 sequential images of 20 ms exposure at each aperture of the chip. A single chip loaded with AcNiR crystal slurry was used to collect the entire 20 dose-point dataset. The total data therefore comprised 512,000 images, and was possible to collect in less than 3 hours, with initial diffraction measured to 1.6 Å toward the corners of the detector (fig. 4.3). Data were indexed and integrated using *dials.stills_process* in *DIALS v1.8.5*, and the script *index.sh* (detailed outline of this processing method can be found in methods 3.5, though important parameters shall be described in this section).

For the 20 dose MSS series a joint refinement method was utilised in order to optimise the number of hits by updating the detector geometry. This was performed on a subset of images, allowing an updated geometry for a 310 mm detector distance by adding the following to the *index.sh* file:

```
fast_axis = 0.9999978683114457, 1.1160941591376779e-06, -
0.002064793287198249
      slow_axis = -2.304596132502108e-06, -0.9999986277164502,
      -0.0016566713329461246
            origin = -216.04450994806226, 226.762983964268, -
            308.3022872184386
```

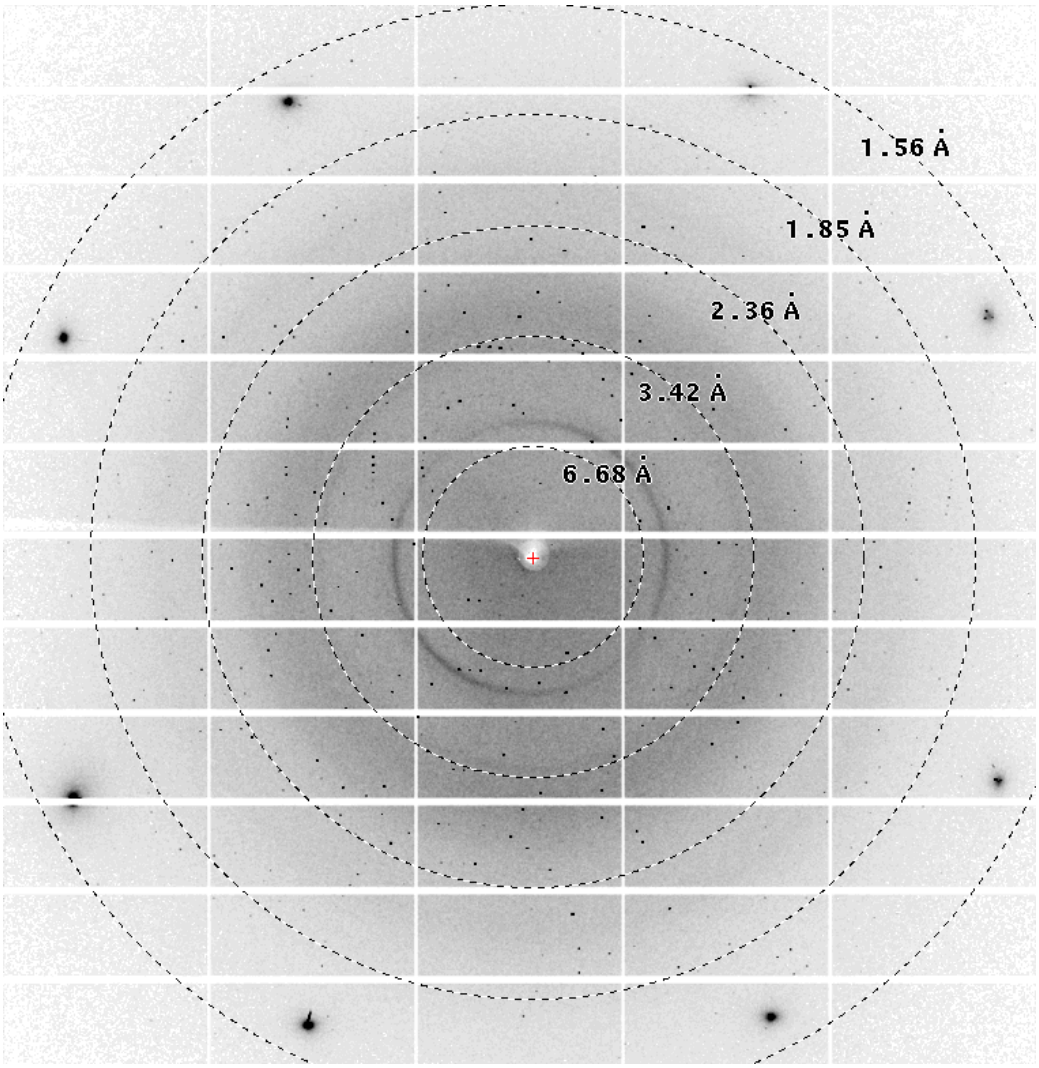


Figure 4.3. AcNiR SSX diffraction

Example image of AcNiR microcrystal diffraction pattern from MSS-ds1. Reflections were seen to greater than 1.6 Å resolution. Intense silicon diffraction spots were masked (methods section 3.5.1) to avoid introducing systematic errors into data processing.

Indexing and integration were performed again using the updated detector geometry, with indexable diffraction patterns identified on 332,272 images in total, a 64.9% hit rate. It was possible however to index 589,403 patterns, owing to multi-lattice indexing of up to three patterns per image. Multi-lattice indexing provided percentages of single, double and triple lattices of 36, 36, and 28%, respectively. The space group was found to be the same as that of previous AcNiR structures, $P2_13$.

CBF image files were binned into individual dose-dependent datasets using the script *get_int_file.py* (appendices 3.7.5), meaning individual diffraction images (stills) from the first image measured from each aperture would make the first dose-dependent dataset, known as MSS-dataset1 (MSS-ds1). It was possible to plot a histogram derived from a subset of crystals in order to accurately determine how isomorphous the unit cell is within the crystal slurry using the following command within *DIALS*:

```
$ cctbx.xfel.plot_uc_cloud_from_experiments
*_refined_experiments.json
```

The unit cell histogram revealed a bimodal distribution of unit cell parameters within the full MSS-ds1 dataset; unit cell dimensions gave a range of a = b = c = 97.8 +/- 0.7 Å, and two distinct peaks at 96.5 +/- 0.2 Å and 97.8 +/- 0.2 Å (fig. 4.4).

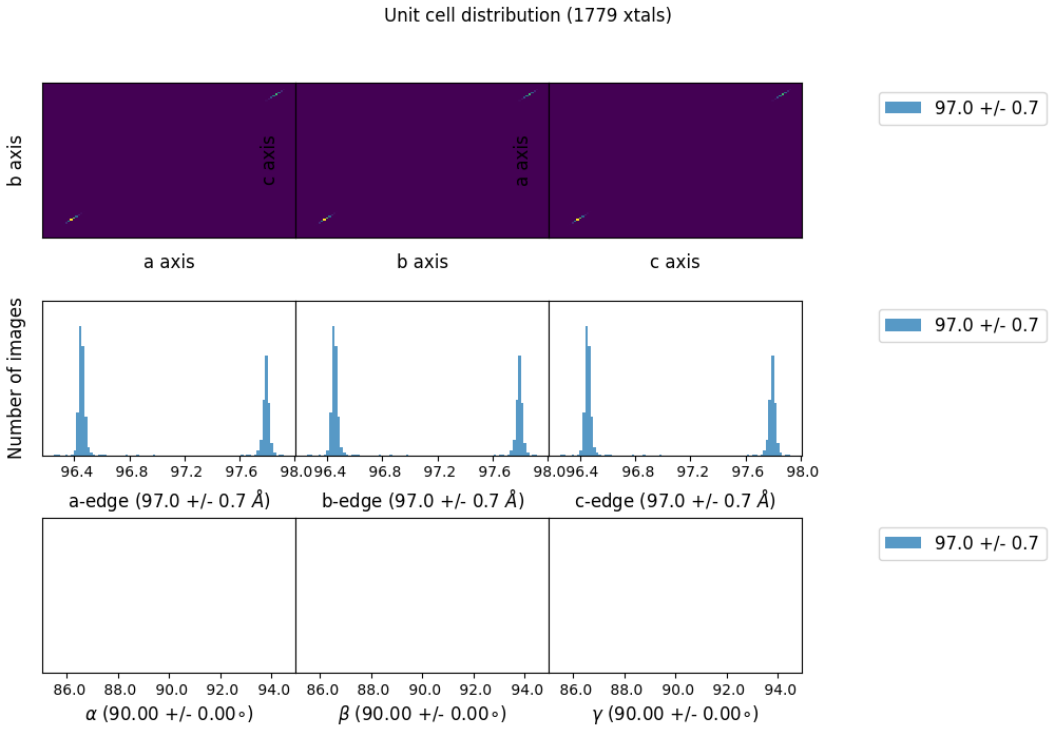


Figure 4.4. AcNiR unit cell distribution for MSS-ds1 crystal subset

The command `cctbx.xfel.plot_uc_cloud_from_experiments` *within* DIALS *shows multiple unit cells within a single crystal population. This plot was used to interpret all data unit cell dimensions of a = b = c = 97.0 ± 0.7 Å, α = β = γ = 90.00 °, small cell unit cell dimensions of a = b = c = 96.5 ± 0.2 Å, α = β = γ = 90.00 °, and large cell unit cell dimensions of a = b = c = 97.8 ± 0.2 Å, α = β = γ = 90.00 °, respectively.*

All patterns in MSS-ds1 were initially merged into a single MSS dataset. These data were then reprocessed and binned into two groups above and below a mid-peak cut-off of 97.25 Å, in order to define two distinct datasets containing only a single unit cell, either large or small. MSS-ds1 AcNiR-large were therefore processed and refined against unit cell dimensions of a = b = c = 97.75 Å, with MSS-ds1 AcNiR-small processed and refined against a = b = c = 96.38 Å. The MSS-ds1 all data AcNiR structure comprised 38,908 integrated frames, with 24,976 integrated frames used for MSS-ds1 small structure, while the number of integrated frames used for MSS-ds1 large structure comprised 13,932, respectively (table 4.2, fig. 4.5).

It was possible to merge dose 1 structures with good quality statistics to a resolution of 1.48 Å, despite a nominal resolution of 1.7 Å imposed by the crystal to detector distance. This was made possible due to the extremely high multiplicity obtained using the fixed target serial method, with a dose 1 redundancy of 927.03 (301.80) for the all cell structure, 548.16 (206.50) in the small cell structure and 357.91 (137.03) in the large cell structure (tables 4.3-4.5). R_{split} and $CC_{1/2}$ were also used to assess the quality of the data and at what resolution to cut the data during the merging process, with respective metrics in dose 1 structures of 5.15 (54.54) and 99.70 (55.42) using all data, 5.71 (87.73) and 99.60 (72.60) for small cell and 7.43 (81.12) and 99.46 (48.66) for large cell (tables 4.3-4.5). These resolution cut-offs were chosen as $CC_{1/2}$ approached 50% in the all data structure, and subsequently used as the cut-off for both the small and large cell structures.

Dose	kGy	Number of integrated images			Percentage of population	
		All data	Small cell	Large cell	% small	% large
1	11	38908	24367	13481	62.63	34.65
2	22	38584	24506	13102	63.52	33.96
3	33	38284	24020	13398	62.75	35
4	44	37803	22622	14383	59.85	38.05
5	55	35424	19100	15643	53.92	44.16
6	66	32944	14907	17427	45.25	52.9
7	77	31032	11209	19420	36.13	62.59
8	88	30009	8605	20854	28.68	69.5
9	99	29392	6934	21950	23.6	74.69
10	110	28726	5808	22437	20.22	78.11
11	121	28082	5043	22609	17.96	80.52
12	132	27372	4428	22546	16.18	82.37
13	143	26641	4005	22265	15.04	83.58
14	154	25873	3666	21844	14.17	84.43
15	165	25269	3386	21569	13.4	85.36
16	176	24521	3091	21117	12.61	86.12
17	187	23723	2850	20568	12.02	86.71
18	198	23054	2717	20055	11.79	87
19	209	22272	2536	19460	11.39	87.38
20	220	21490	2399	18824	11.17	87.6
Total		**589403**	**196199**	**382952**		

Table 4.2. Number of integrated images used for MSS datasets

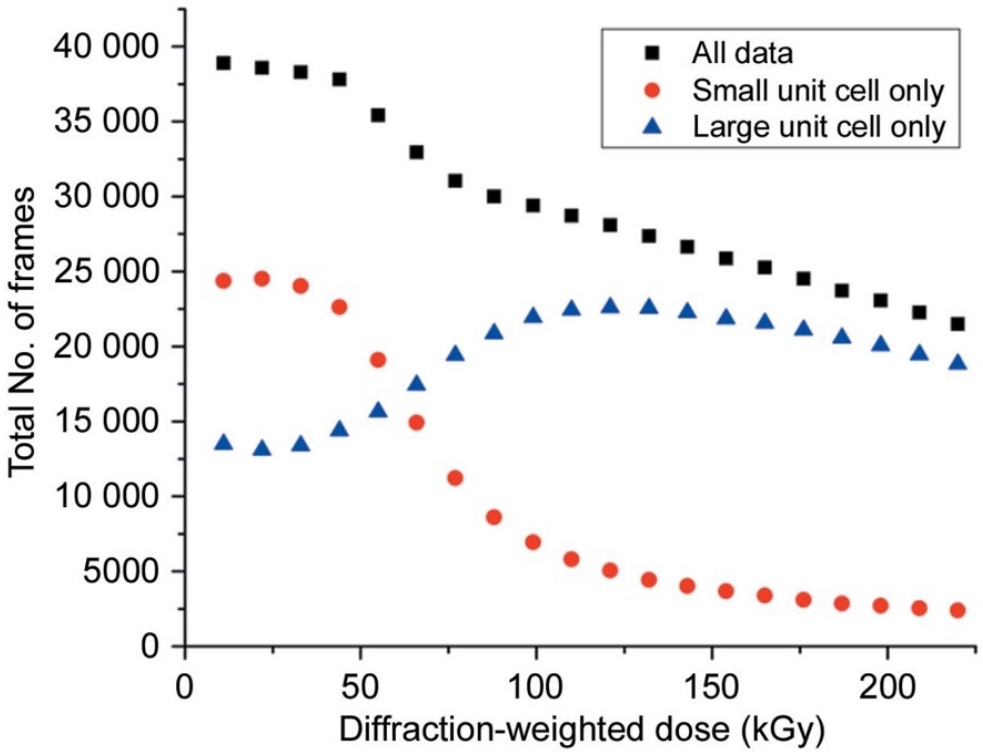


Figure 4.5. Number of integrated images used for MSS datasets

The number of all data, small cell, and large cell integrated images used for selective integration into small and large cell datasets, representing the entire dose series. Figures produced using `OriginLab` *graphing and data analysis software (www.originlab.com).*

Dose 1 all cells									
Bin	**Resolution range**	**Completeness**	**<N_obs>**	**Rmerge**	**Rsplit**	**CC1/2**	**<I/sigI>**	**<I>**	**<I**2>**
1	29.21 - 4.01	100	1777.42	98.7	4.01	99.49	11.25	575.7	1.81
2	4.01 - 3.19	100	1330.95	91.99	4.3	99.41	7.78	350.6	1.78
3	3.19 - 2.78	99.9	1208.03	90.11	4.63	99.18	3.81	154.8	1.62
4	2.78 - 2.53	100	1128.9	89.08	4.86	99.1	2.5	89.3	1.57
5	2.53 - 2.35	100	1073.15	88.59	5.4	98.92	2.02	65.6	1.55
6	2.35 - 2.21	100	1033.33	88.38	5.57	98.8	1.77	55.2	1.52
7	2.21 - 2.10	100	1029.13	88.21	5.72	98.85	1.51	46.3	1.52
8	2.10 - 2.01	100	990.23	88.26	6.09	98.77	1.2	35.1	1.55
9	2.01 - 1.93	100	994.49	89.02	6.9	98.32	0.99	27.3	1.52
10	1.93 - 1.86	100	952.4	89.8	7.48	98.22	0.79	21	1.52
11	1.86 - 1.81	100	963.75	90.89	8.64	97.4	0.62	15.9	1.48
12	1.81 - 1.75	100	939.07	92.38	10.59	96.65	0.48	11.8	1.51
13	1.75 - 1.71	100	921.81	93.55	12.57	95.29	0.39	9.3	1.52
14	1.71 - 1.67	100	883.15	94.48	14.44	94.44	0.33	7.9	1.54
15	1.67 - 1.63	100	739.9	95.46	18.05	91.57	0.28	6.7	1.55
16	1.63 - 1.59	100	646.07	95.99	21.42	88.77	0.25	5.8	1.59
17	1.59 - 1.56	100	599.59	96.41	25.24	84.96	0.22	5	1.58
18	1.56 - 1.53	100	520.31	96.86	28.46	79.97	0.2	4.4	1.58
19	1.53 - 1.51	100	427.73	97.23	35.19	72.39	0.18	3.9	1.6
20	1.51 - 1.48	100	301.79	97.75	54.54	55.42	0.15	3.2	1.83
	TOTAL	**100**	**927.03**	**94.45**	**5.15**	**99.7**	**1.88**	**76.8**	**8.06**

Table 4.3. MSS-ds1 all cell statistics

Dose 1 small cell									
Bin	**Resolution range**	**Completeness**	**<N_obs>**	**Rmerge**	**Rsplit**	**CC1/2**	**<I/sigI>**	**<I>**	**<I**2>**
1	29.07 - 4.01	100	1052.82	93.82	4.63	99.3	9.84	519.9	2.01
2	4.01 - 3.19	100	786.27	83.29	4.42	99.42	7.01	326.5	2.05
3	3.19 - 2.78	100	714.87	80.59	4.67	99.48	3.53	146.9	1.97
4	2.78 - 2.53	100	666.66	80.07	5.36	99.27	2.38	86.7	1.98
5	2.53 - 2.35	100	636.74	80.2	5.81	99.13	1.92	63.3	1.94
6	2.35 - 2.21	100	608.51	79.94	6.14	99.03	1.73	54.6	1.9
7	2.21 - 2.10	100	610.11	80.07	6.16	99.23	1.52	47	1.93
8	2.10 - 2.01	100	584.13	81.32	6.84	99.05	1.21	35.3	1.95
9	2.01 - 1.93	100	588.07	82.66	7.37	98.96	0.98	27.2	1.97
10	1.93 - 1.86	100	559.72	83.76	8.19	98.71	0.83	21.9	1.94
11	1.86 - 1.81	100	568.48	85.88	9.36	98.23	0.65	16.5	1.85
12	1.81 - 1.75	100	553.58	88.35	11.42	97.64	0.5	12.3	1.89
13	1.75 - 1.71	100	543.02	90.02	13.74	96.83	0.41	9.9	1.95
14	1.71 - 1.67	100	522.79	91.35	15.5	96.1	0.36	8.5	1.92
15	1.67 - 1.63	100	428.48	92.49	19.3	95.02	0.31	7.2	2.06
16	1.63 - 1.59	100	376.73	93.23	22.38	92.83	0.28	6.4	2.07
17	1.59 - 1.56	100	355.53	94.34	26.12	89.97	0.24	5.3	2
18	1.56 - 1.53	100	304.81	94.77	30.64	87.55	0.22	4.9	2.09
19	1.53 - 1.51	100	249.02	95.29	37.07	82.69	0.21	4.4	2.08
20	1.51 - 1.48	100	206.5	96.34	49.81	72.6	0.17	3.5	2.21
	TOTAL	**100**	**548.16**	**87.73**	**5.71**	**99.6**	**1.75**	**72.3**	**8.63**

Table 4.4. MSS-ds1 small cell statistics

Dose 1 large cell									
Bin	**Resolution range**	**Completeness**	**<N_obs>**	**Rmerge**	**Rsplit**	**CC1/2**	**<I/sigI>**	**<I>**	**<I**2>**
1	29.48 - 4.01	100	680.57	87.98	5.44	99.13	9.7	499.1	1.97
2	4.01 - 3.19	100	510.96	81.61	5.72	99.13	6.57	302.9	1.98
3	3.19 - 2.78	100	463.46	80.81	6.2	99.01	3.14	131.9	2.01
4	2.78 - 2.53	100	433.88	81.7	7.16	98.72	2.06	76.2	1.92
5	2.53 - 2.35	100	414.2	82.61	8.05	98.31	1.62	54.5	1.93
6	2.35 - 2.21	100	396.77	82.99	8.3	98.45	1.41	45.6	1.88
7	2.21 - 2.10	100	397.55	83.54	8.58	98.35	1.24	39.4	1.83
8	2.10 - 2.01	100	381.27	85.34	9.88	97.82	0.95	28.7	1.82
9	2.01 - 1.93	100	384.7	86.63	10.76	97.66	0.8	22.8	1.81
10	1.93 - 1.86	100	366.39	88.31	12.74	96.83	0.63	17.2	1.86
11	1.86 - 1.81	100	372.2	90.06	14.44	96.04	0.51	13.4	1.86
12	1.81 - 1.75	100	362.7	92.3	18.87	93.69	0.38	9.7	1.88
13	1.75 - 1.71	100	356.61	93.33	21.55	91.32	0.33	8.1	1.81
14	1.71 - 1.67	100	343.8	94.61	25.37	89.58	0.27	6.5	1.91
15	1.67 - 1.63	100	281.27	95.32	31.98	85.73	0.24	5.7	1.96
16	1.63 - 1.59	100	247.34	95.96	36.96	81.44	0.21	4.9	2.03
17	1.59 - 1.56	100	232.84	96.61	43.76	76.6	0.18	4.1	2.11
18	1.56 - 1.53	100	199.83	96.9	51.7	68.19	0.16	3.7	2.08
19	1.53 - 1.51	100	163.89	97.23	62.06	59.94	0.15	3.3	2.17
20	1.51 - 1.48	100	137.03	97.74	81.12	48.66	0.12	2.8	2.47
	TOTAL	**100**	**357.91**	**85.7**	**7.43**	**99.46**	**1.57**	**65.7**	**9.03**

Table 4.5. MSS-ds1 large cell statistics

4.3.3 Changes to polymorph populations with X-ray dose

The diffraction-weighted dose was calculated using *RADDOSE-3D* (Paithankar & Garman, 2010), with a crystal size of 8 x 8 x 8µm selected to represent the smallest cubic crystals exposed to the beam (full input parameters and complete output data can be found in appendices section 4.5.2). The cumulative dose was calculated to be 11 kGy per 20 ms dataset, equating to a dose rate of 1.1 MGy s^{-1}. In order to quantify global radiation damage in MSS datasets total Bragg intensity was normalised for each dataset against the Bragg intensity for dose 1 in both small and large cell polymorphs (table 4.6). Global radiation damage was evident in the form of a rapid decrease in normalised diffracting power with accumulated dose in both polymorphs (fig. 4.6).

By assessing the unit cell population change throughout the dose series, it is possible to show how the mean cell dimension undergoes a small progressive increase in each polymorph, consistent with many previous studies showing unit-cell expansion with accumulated X-ray dose. However, in this data series an interchange between the polymorph cell populations is also evident; the AcNiR-small population rapidly decreases, with a simultaneous increase in the AcNiR-large cell population (fig. 4.6). As represented in figure 4.7, the number of overall integrated images decreases in the all data dose series as expected, due to the accumulation of dose and subsequent global radiation damage effects. By plotting the selectively integrated small (a = b = c = 96.83 Å) and large cell (a = b = c = 97.25 Å) data in the MSS series, it is possible to see that there is a lack of overlap between the small and large cell structures throughout the MSS series. Further, in figure 4.7 it is possible to show that the increase in cell begins immediately upon irradiation, and the switch from small to

Dose	kGy	<I> small	<I> small normalised	<I> large	<I> large normalised
1	11	72.3	1	65.7	1
2	22	73.6	1.018	67.1	1.0214
3	33	65.1	0.9005	66	1.0046
4	44	50.5	0.6985	57.5	0.8752
5	55	37.4	0.5173	46	0.7002
6	66	29.1	0.4025	37.2	0.5663
7	77	24.4	0.3375	30.5	0.4643
8	88	21.4	0.296	27.9	0.4247
9	99	19	0.2628	23.8	0.3623
10	110	17.4	0.2407	21.9	0.3334
11	121	16	0.2214	20.2	0.3075
12	132	14.7	0.2034	19.1	0.2908
13	143	13.6	0.1882	17.9	0.2725
14	154	12.8	0.1771	16.9	0.2573
15	165	12.3	0.1702	15.8	0.2405
16	176	11.5	0.1591	15.1	0.2299
17	187	10.9	0.1508	14.5	0.2208
18	198	10.3	0.1425	13.7	0.2086
19	209	9.8	0.1356	13	0.1979
20	220	9.4	0.1301	12.6	0.1918

Table 4.6. Bragg intensities and normalised Bragg intensities for dose series polymorphs

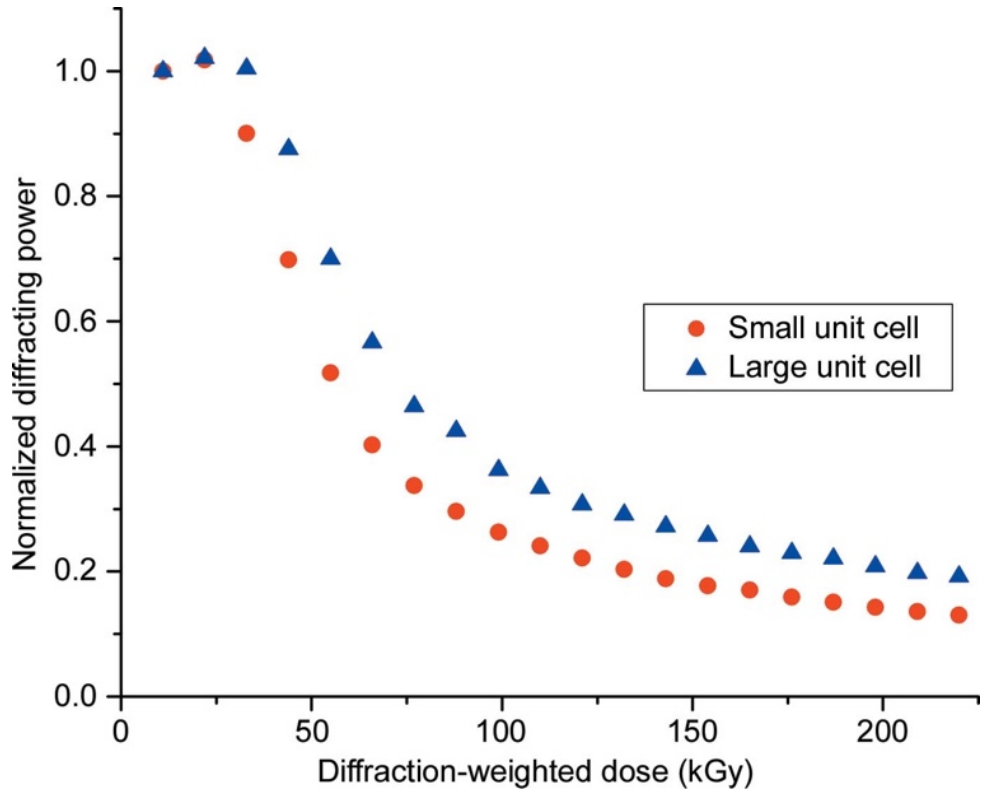


Figure 4.6. Decrease in normalised diffracting power as a function of accumulated dose.

Total diffracting power for each polymorph dose point was calculated by dividing the total Bragg intensity (I) for each dataset as reported in `PRIME` *after integration against the total Bragg intensity for the small or large cell ds1 dataset. It can be seen that although there is a very slight increase in intensity in small and large cell ds1 and large cell ds2 datasets, there is an exponential decay in diffraction intensity as dose is accumulated.*

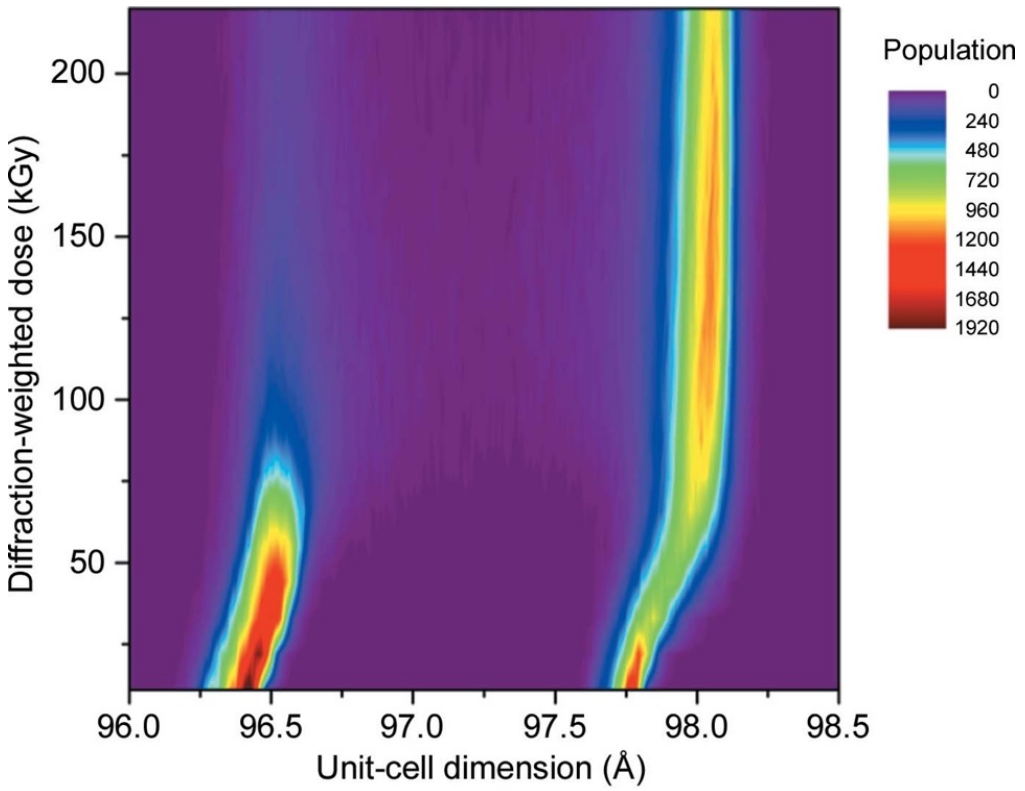


Figure 4.7. Change in polymorph unit cell as a function of accumulated dose

Change in unit cell dimension (a = b = c, space group $P2_13$) as a function of accumulated dose shown using a two-dimensional histogram, with an extremely high population of all indexed images at low dose exhibiting either a unit cell of 96.38 or 97.75 Å (small and large cell, respectively), with respective polymorph population cell dimensions increasing to 96.56 and 98.04 Å by 220 kGy. The number of small cell indexed images has decreased vastly by the end of the data collection, whilst the number of large cell has increased.

large cell represents a damaged AcNiR-large cell, rather than a switch to the AcNiR-large polymorph that is identified in the original population at dose 1.

4.3.4 Influence of dilution on unit cell population

To provide insight into what concentration of microcrystals would give the best hit rate with limited multiple lattices, a serial dilution was performed on a single stock of AcNiR batch produced microcrystals (concentration not measured prior, concentration of similar batches ~ 3 x 10^6 crystals ml^{-1}). Data were collected on only 4 city blocks (1600 apertures) per concentration, with 50 µl of slurry loaded to the chip using the same vacuum loading method as used previously. Intriguingly, data processing revealed that unit cell populations are affected depending on the dilution of microcrystals (fig. 4.8, 4.9). The undiluted crystals had an average a = b = c = 96.5 +/- 0.2 Å. This unit cell value increases across the serial dilution with a 2x dilution exhibiting a unit cell of 96.9 +/- 0.5 Å, 4x exhibiting a unit cell of 97.4 +/- 0.6 Å, 8x exhibiting a unit cell of 97.6 +/- 0.5 Å, and a 10x dilution exhibiting a unit cell of 97.7 +/- 0.4 Å, respectively. A final dilution of 15x exhibited a unit cell of 97.8 +/- 0.0 Å.

4.3.5 AcNiR MSS structures

Selected structures were refined in order to investigate the polymorphic unit cell and the influence of accumulated dose on the active site throughout the dose series. It was possible to refine all MSS-ds1 structures to 1.48 Å, with an R_{work}/R_{free} of 0.235/0.276 in the all data structure, and 0.186/0.217 and 0.205/0.228 in the small and large cell structures, respectively (table 4.7). In the all data MSS-ds1 structure, electron density reveals the characteristic homotrimer seen in CuNiRs (fig. 4.10a),

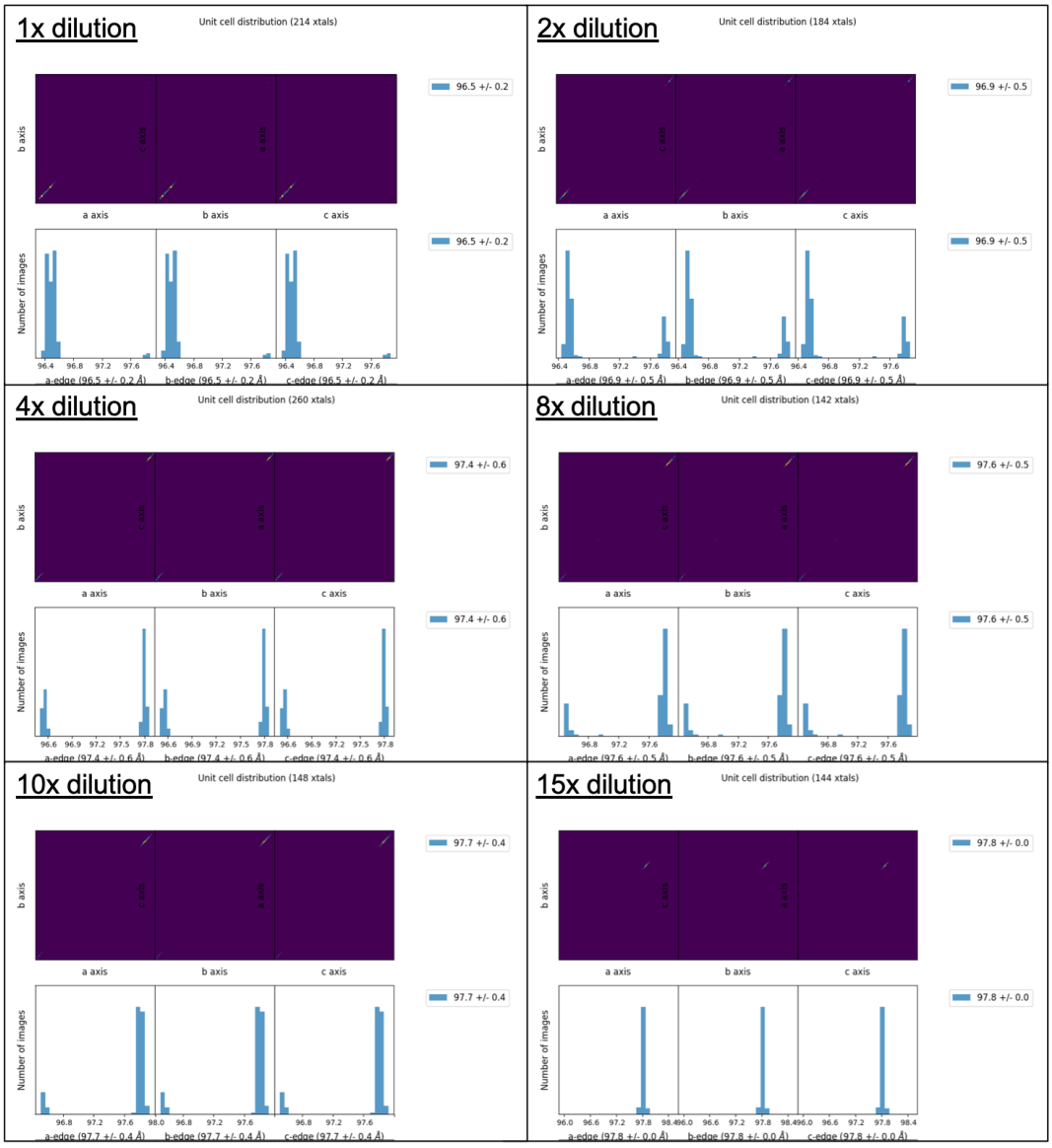


Figure 4.8. Unit cell histograms for dilution series

Unit cell histograms for serial dilutions from a single AcNiR microcrystal batch. As the concentration of crystals decreases within the slurry volume the initial unit cell of a = b = c = 96.5 +/- 0.2 Å shifts toward a larger cell, with the final 15x dilution presenting a unit cell of a = b = c = 97.8 Å. Histograms calculated and produced using `DIALS v1.8`.

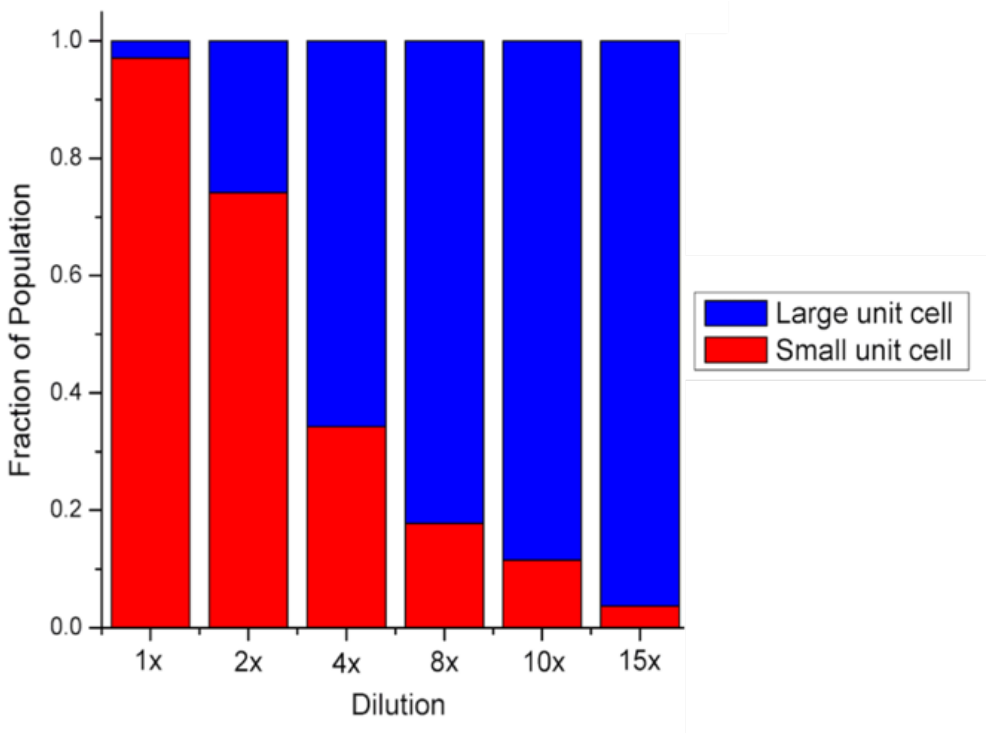


Figure 4.9. Fraction of polymorph as a function of microcrystal suspension concentration.

Data collected on multiple crystal dilutions from 1600 apertures indicate as the concentration of microcrystals in a suspension decreases, the fraction of integrated images that exhibit a small unit cell also decreases. Dilution on the X axis is shown as a serial dilution, as microcrystal concentration measurements were not calculated prior to data collection.

	Small Cell dose1	Large Cell dose1	All Data dose1
Data Collection			
Cumulative dose (AcNiR)	11	11	11
Number of integrated frames	24,976	13,932	38,908
Number of images used	24,367	13,481	38,798
Data processing			
Space Group	$P2_13$		
Cell dimensions (Å)	96.38	97.75	96.87
Resolution (Å)	29.21-1.48 (1.51-1.48)	29.21-1.48 (1.51-1.48)	29.21-1.48 (1.51-1.48)
R_{split} (%)	5.71 (87.73)	7.43 (81.12)	5.15 (54.54)
$CC_{1/2}$	99.60 (72.60)	99.46 (48.66)	99.70 (55.42)
<I>	72.3 (3.5)	65.7 (2.8)	76.8 (3.2)
I/σ(I)	1.75 (0.17)	1.57 (0.12)	1.88 (0.15)
Multiplicity	548.16 (206.50)	357.91 (137.03)	927.03 (301.8)
Completeness (%)	100 (100)	100 (100)	100 (100)
Refinement			
Number of reflections	49894	49430	50595
R_{work}/R_{free}	0.186 / 0.217	0.205 / 0.228	0.235 / 0.276
RMSD bond lengths (Å)	0.013	0.012	0.012
RMSD bond angles (deg)	1.63	1.62	1.58
Ramachandran plot:			
Most favoured (%)	97.3	96.7	97.5
Allowed (%)	2.7	3.3	2.5
PDB accession code	6GB8	6GBB	6GBY

Table 4.7. Data processing and refinement values for MSS-ds1 datasets

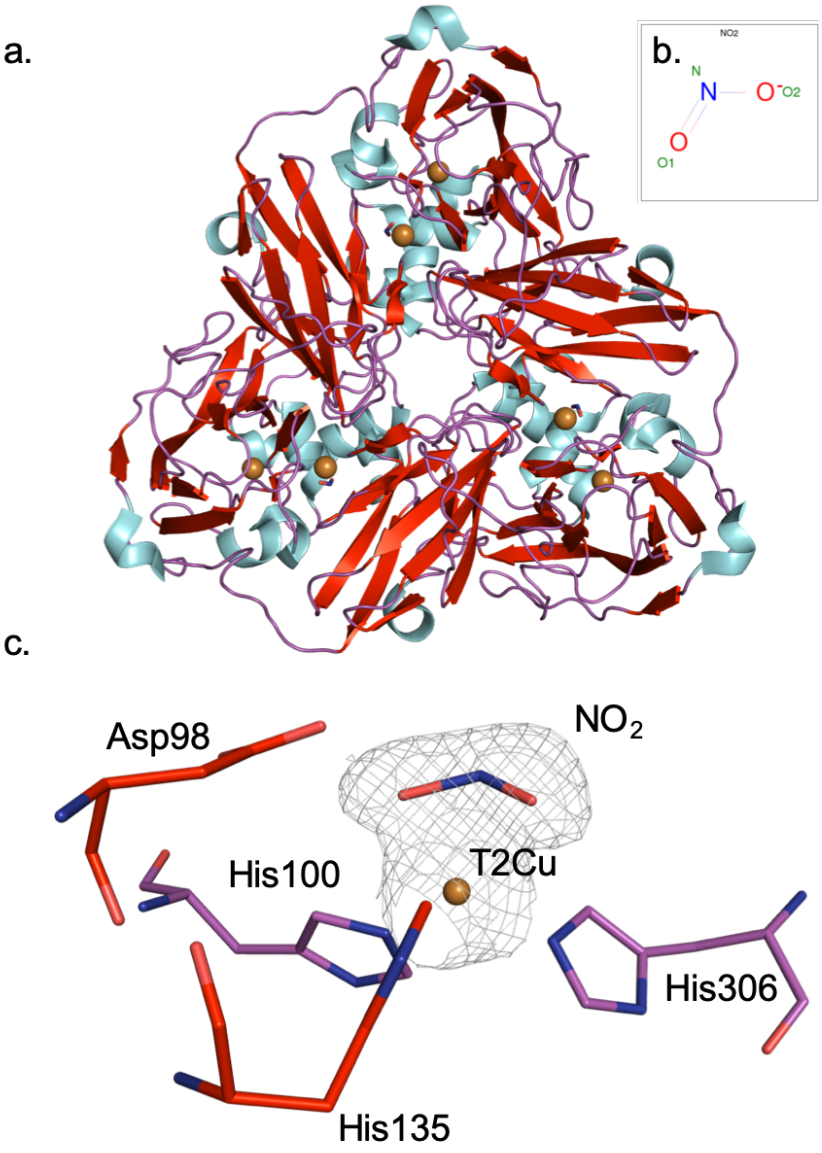


Figure 4.10. Overall and active site structure of MSS-ds1 all data

(a.) Overall homotrimeric structure of AcNiR all data dataset MSS-ds1, shown as a cartoon and coloured by secondary structure (blue = alpha helices, red = beta sheets, purple = loop). (b.) Chemical structure of nitrite. (c.) Active site arrangement of MSS-ds1 all data structure, including Asp98 in the proximal position, and histidine residues His100, His135, and His306 that coordinate the copper atom at the T2Cu (electron density omitted around these residues for clarity). Electron density contoured at 1σ.

with each monomer containing a type 1 copper centre (T1Cu) and type 2 copper centre (T2Cu). A full occupancy NO_2^- is bound to the type 2 copper centre (fig. 4.10b, c) in a 'top hat' conformation, while the T1Cu and T2Cu are ~12.7 Å apart (fig. 4.11), as seen in previously (Horrell et al., 2016). The T2Cu is coordinated by three histidines, His100, 135, and 306, with Asp98 in the proximal conformation (Horrell et al., 2018).

Structurally, the MSS-ds1 all data and polymorph structures differ in a loop region, consisting of residues 187-193 and 201-205, existing in a dual conformation in the all data structure, and in only one or the other conformation in the polymorphs (fig. 4.12). The active site remains unchanged in the dose series until ~44 AcNiR (ds4) where NO_2^- is converted to NO and an H_2O (fig. 4.13), with the high-resolution limit decreasing to 1.53 Å. This active site state remains unchanged until ~66 AcNiR (ds6), whereby the water molecule remains and a second water molecule is bound in place of the NO, with the high-resolution limit decreasing to 1.63 Å. The active site remains unchanged at higher doses, however past MSS-ds11 it becomes difficult to confidently determine the active site due to limited resolution of the data.

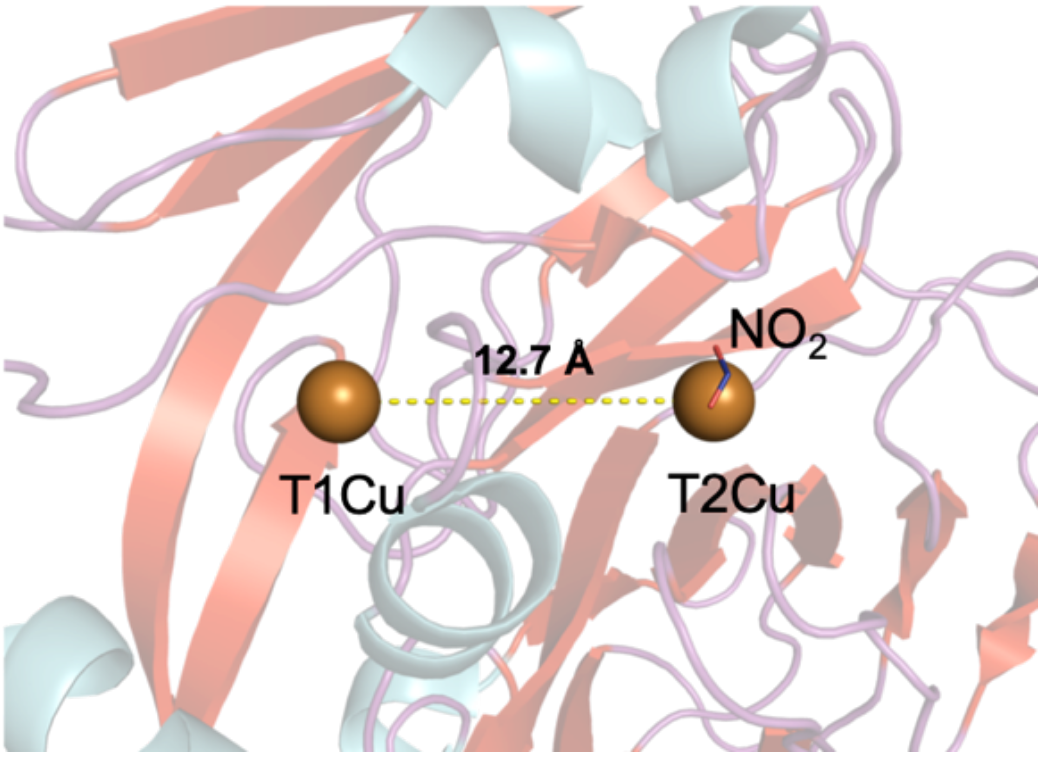


Figure 4.11. Distance between copper centres in MSS-ds1 all data

Distance measured between the T1Cu and the T2Cu is ~12.7 Å in MSS-ds1 all data.

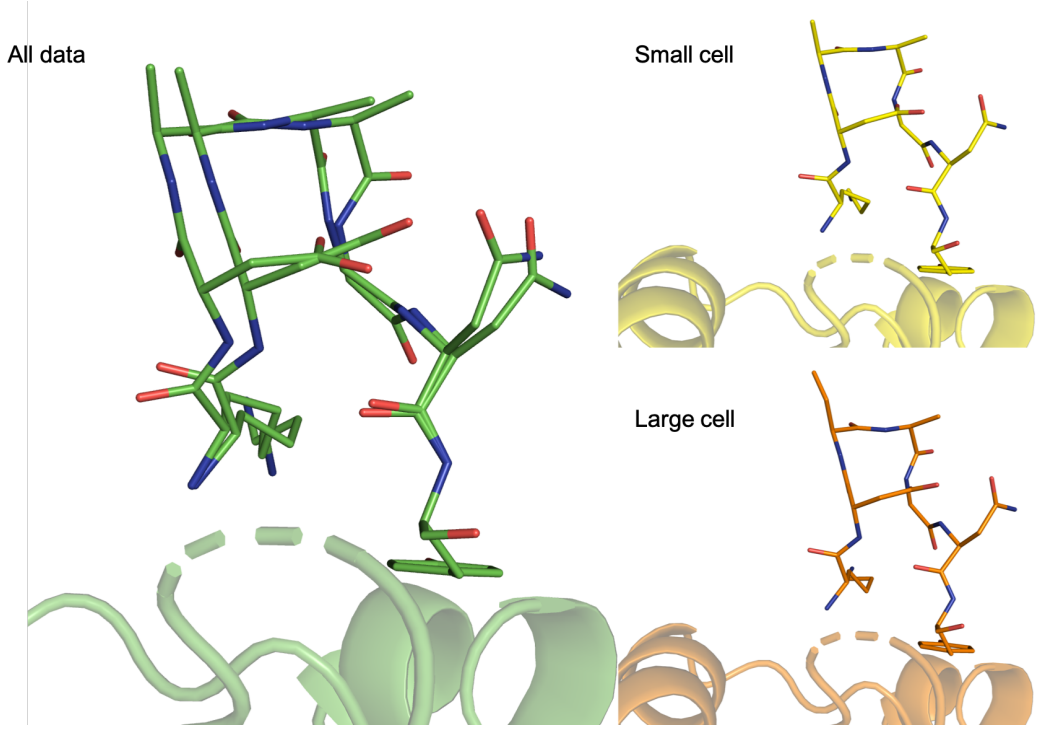


Figure 4.12. MSS-ds1 all data and polymorph 187-193 loop regions

MSS-ds1 all data and polymorph structures differ in a loop region, consisting of residues 187-193, existing in a dual conformation in the all data structure and a single loop in the small and large cell structures.

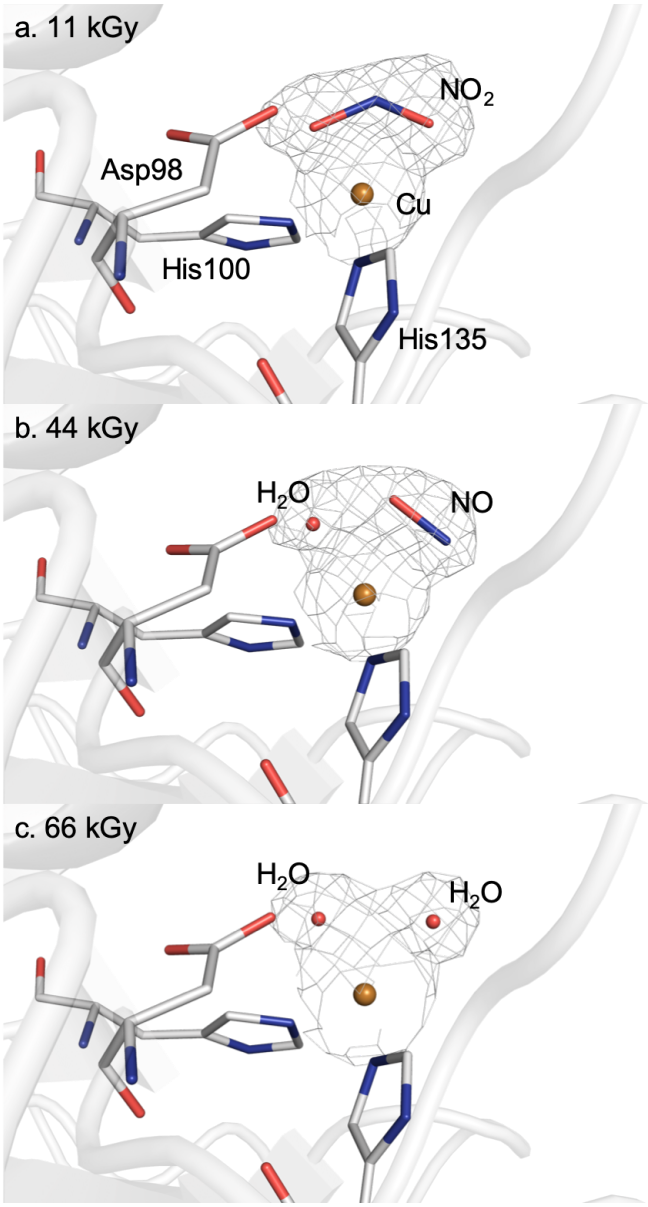


Figure 4.13. X-ray induced active site change in AcNiR

Dose series structures for small cell MSS-ds1, MSS-ds4, and MSS-ds6 (11, 44, and 66 kGy, respectively) reveal X-ray induced ligand turnover in AcNiR, showing the conversion of nitrite to nitric oxide to water. (a.) Active site arrangement after an initial exposure of 11 kGy for the small cell structure reveals a single occupancy NO_2 bound in the top hat position above the T2Cu, with Asp98 in the proximal position. (b.) After an accumulated dose of 44 kGy NO is bound to the T2Cu with a water molecule, (c.) and after a further 22 kGy nitric oxide is converted to water. Maps contoured at 1σ, with electron density around histidine and aspartic acid omitted for clarity.

4.4. Discussion

4.4.1. Data and structure quality indicators

The polymorphic nature of AcNiR crystals was initially discovered using the native unit cell histogram plot in *DIALS*. This is not obviously reflected in the all data merging and refinement parameters for the all data structure, with very good $CC_{1/2}$ and multiplicity for all data MSS-ds1 (99.70 [55.42], 927.03 [301.8]) (see table 4.7). However, once data were separated, $CC_{1/2}$ improved for small cell (99.60 [72.60]), with very good $CC_{1/2}$ values for large cell (99.46 [48.66]), with very good multiplicity for both small and large polymorphs (548.16 [206.50], 357.91 [137.03]) (see table 4.7). Although fewer images were used in MSS-ds1 small vs all (23,467 vs 38,798), as-well-as presenting lower multiplicity statistics, the outer shell CC1/1 is vastly improved (72.60 vs 55.42). High resolution structures with an outer shell resolution of 1.51-1.48Å were achieved despite a nominal resolution limit of 1.7 Å, imposed by crystal to detector distance (inscribed circle on the detector surface). This was due to the multiplicity of the data; the high multiplicity associated with serial data collections is apparent here, despite the data being collected from only a single chip, allowing complete reflections from the corners of the detector to be used in data processing.

Utilisation of careful, stepwise data processing allowed the identification of a microcrystal slurry that exhibits multiple unit cell populations within it. Separation of unit cell polymorphs via data processing methods led to an improvement in data quality, with the all data dose 1 structure exhibiting R_{work}/R_{free} values of 0.235/0.276 compared to the small and large cell dose 1 exhibiting an R_{work}/R_{free} of 0.186/0.216 and 0.205/0.227, respectively. As all data were cut to the same resolution limit of

1.48Å, the difference between the all cell and single cell R values suggest that this analysis step prior to merging is beneficial over simply merging all data regardless of unit cell parameter. Further, percentile ranks, used to compare structures deposited to the PDB against all deposited structures as-well-as against X-ray structures of a similar resolution (111,664 total PDB entries at the time of deposition, 3964 at a resolution range of 1.50 – 1.46 Å), place the R_{free} values for both the small and large cell MSS-ds1 structures favourably, with both structures residing in the upper percentile ranks for R_{free} (fig. 4.14). The all data MSS-ds1 structure however performs markedly worse in comparison, with R_{free} ranking in the bottom 30% when compared to all deposited structures, and R_{free} ranking against depositions in the 1.50 – 1.46 Å resolution range in the bottom 10% of structures (fig. 4.14).

4.4.2. Structural comparison

When superimposing the MSS-ds1 large and small structures (fig. 4.15a), an r.m.s.d. value of 0.16 Å is apparent. While structures are extremely similar, there are structural differences observed at the N- and C-termini, and in the loop structure around residues 187-193 and 201-205 (fig. 4.15a, b). Further, when viewing the structure as a symmetry generated trimer, the biological assembly of AcNiR, it is possible to reveal a difference between the overall volume between the polymorphs; the volume of the AcNiR large trimer was 157399 $Å^3$, an increase of 1901 $Å^3$ over the AcNiR small volume of 155498 $Å^3$. The corresponding increase in surface area was 659 $Å^2$, with AcNiR large exhibiting a surface area of 22998 $Å^2$ and AcNiR small exhibiting a surface area of 22339 $Å^2$. Volume and surface area measured using *3vee* (3vee.molmovdb.org; Voss & Gerstein, 2010). Without this polymorph separation

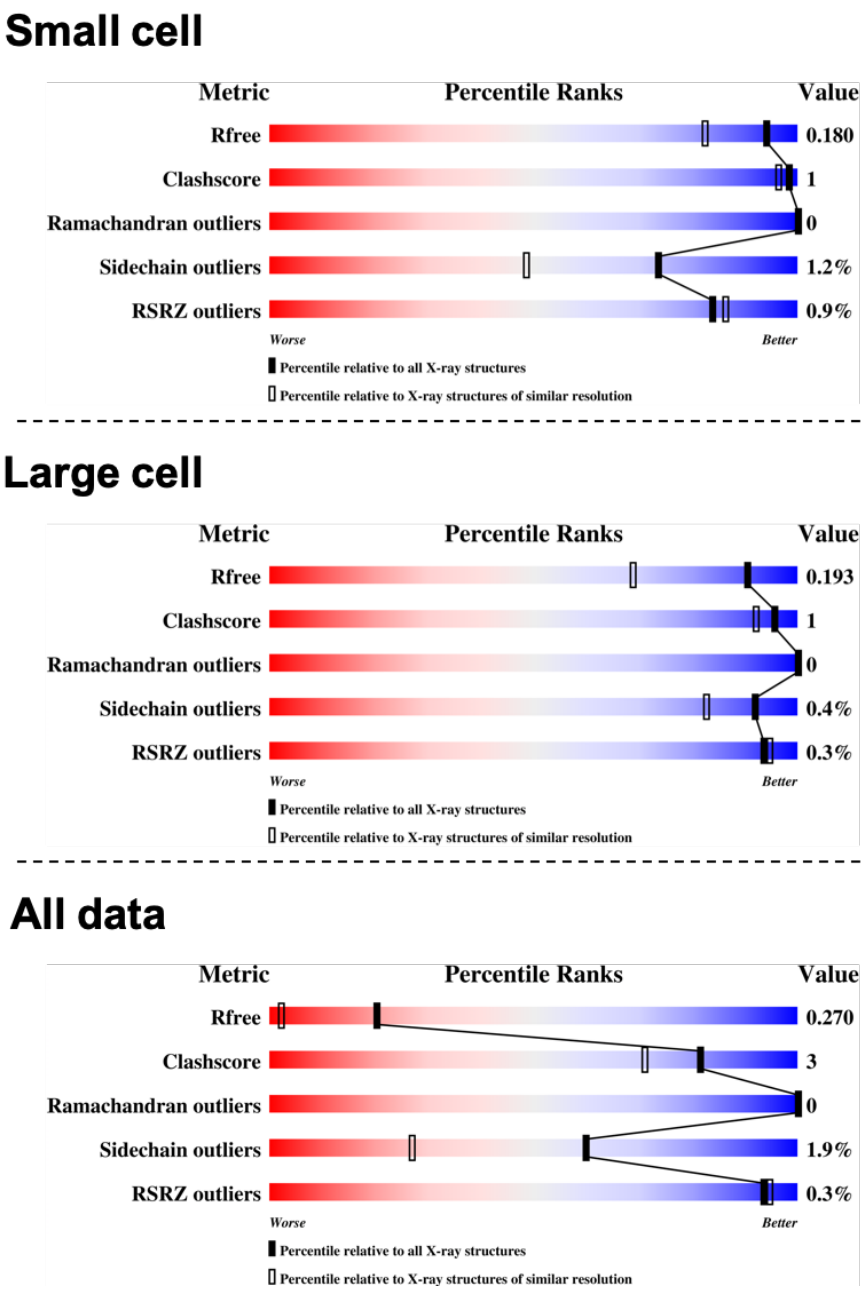


Figure 4.14. PDB percentile ranks of the MSS-ds1 structures

Deposited small cell (6GB8), large cell (6GBB), and all data (6GBY) structures also fall within better than average percentile ranks for factors including clashscore, Ramachandran outliers, sidechain outliers, as-well-as real-space R-value Z-score (RSRZ score) which measures the quality of fit between a part of an atomic model and the data in real space. Figures were generated via PDB Validation Report. Both small cell and large cell percentiles are an improvement over those of the all data structure, indicating the strength of using a robust data processing method when collecting serial data.

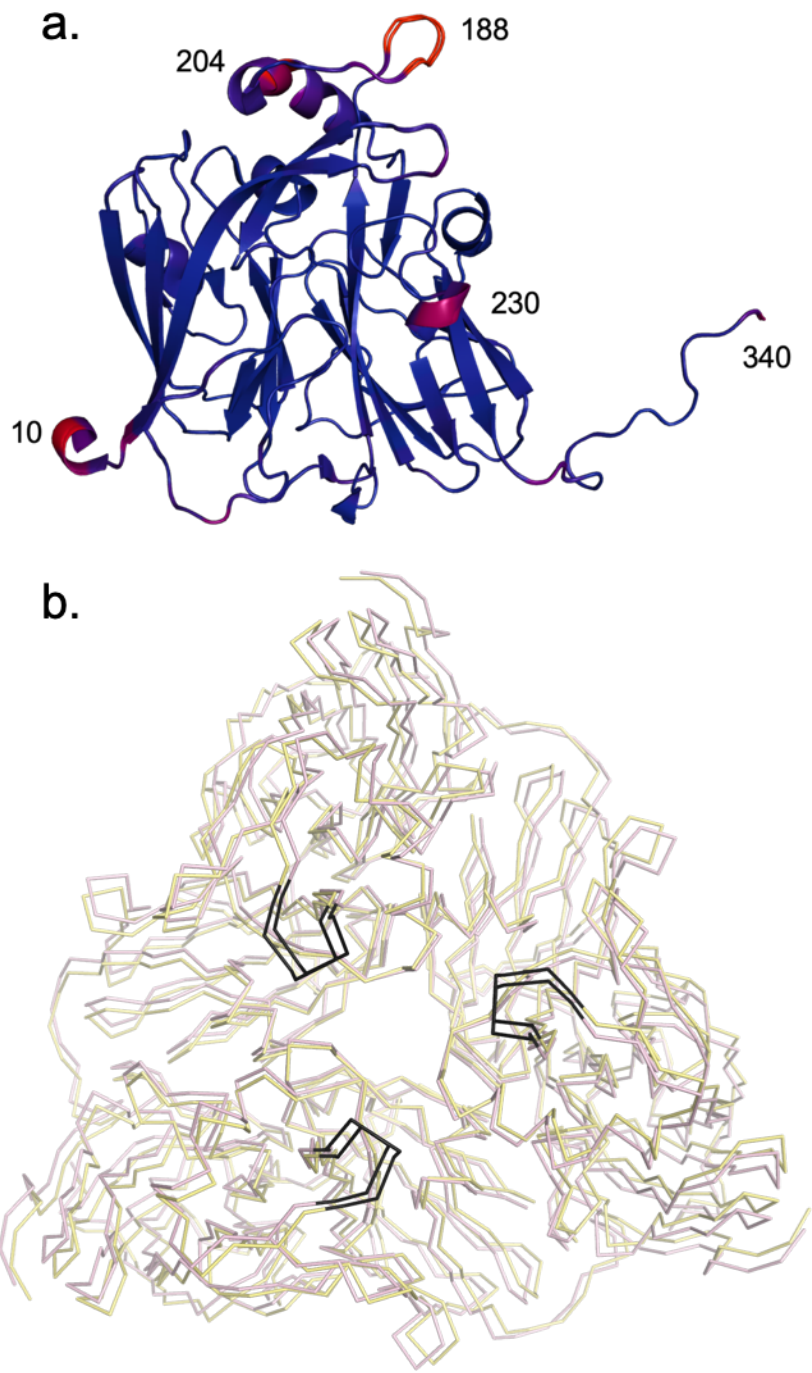


Figure 4.15. Superimposed small and large cell AcNiR 11 kGy structures

(a.) Superimposed MSS-ds1 small and large structures coloured by r.m.s.d. between structures, with blue indicating low values and red indicating high values. The highest deviation between structures can be seen in the loop region around Asp188, residues 187-193. (b.) Superposition of aligned small (pink) and large (yellow) structures (biological assembly) viewed from down the threefold axis (residues 186-195 shown in dark grey).

procedure, the MSS-ds1 all data structure exists in a dual conformation, with a shift between the loop region residues 187-193 and 201-205 representing both large and small polymorph structures. Interestingly, this polymorphic loop region is also apparent in AcNiR structures that have been deposited in the PDB; Horrell et al. (2016) deposited 6 serial structures of AcNiR from a multiple structure using one crystal (MSOX) experiment performed at 100 K that all also exhibit the same dual conformation of residues 187-193 (PDB codes: 5I6K, 5I6L, 5I6M, 5I6N, 5I6O, 5I6P).

Due to the structural differences in the loop region around Asp188 are observed between polymorphs, further comparison was undertaken to seek to understand the different loop conformations; multiple single crystal structures are available for AcNiR, measured at a variety of temperatures though primarily data have been collected at 100K. Superposition of this loop region in AcNiR structures determined at different temperatures reveals a progressive shift of the loop from 100 K (cell length 95.41 Å) to 240 K (cell length 96.13 Å; PDB entry 5N8F) and room temperature (cell length 96.23 Å; PDB 5OFF), with the latter being very similar in structure to the AcNiR small polymorph (fig. 4.16). A further shift then occurs to the AcNiR large polymorph (98.21 Å).

When comparing conversion at the active site with a published dose series of AcNiR at room temperature, a similar progression of NO_2^- to NO to H_2O is seen. Horrell et al published a ten frame MSOX series at room temperature, with the same top-hat orientation of the bound NO_2 in the 30 kGy dose 1 structure (Horrell et al., 2018); NO_2^- is converted to NO after 180 kGy in their MSOX series, compared to just 44 kGy in

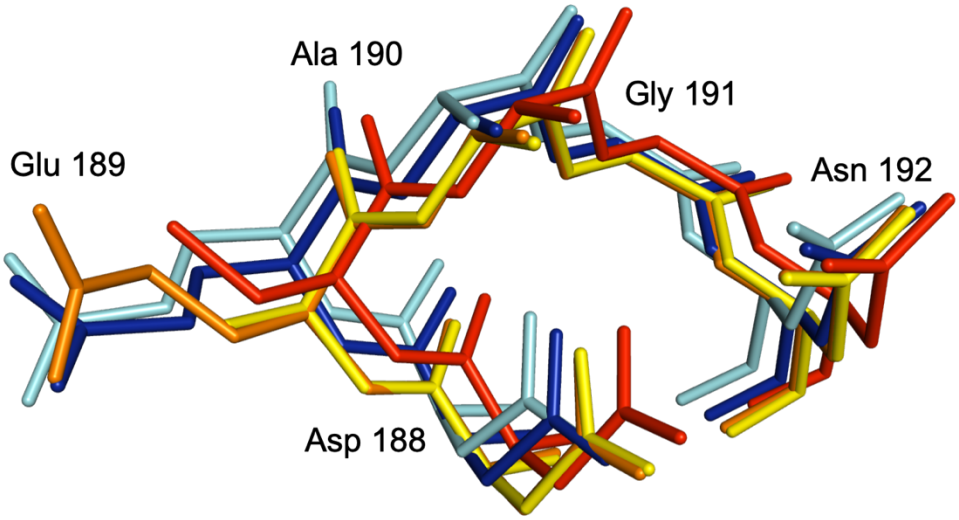


Figure 4.16. Superposed AcNiR loop conformations at different temperatures

Asp 188 loop region from single crystal AcNiR structures at 100K (cyan; PDB: 2BW4), 240K (blue; PDB: 5N8F), and RT (orange; PDB: 5OFF), aligned against the multi crystal small cell (yellow), and large cell (red) structures. Interestingly there is a progressive shift as temperature increases between the single crystal structures, to the point at where the single crystal room temperature loop region is very closely aligned to that of the small cell structure.

MSS-ds4, with the NO replaced with a water by 300 kGy compared to only 66 kGy in MSS-ds6. These studies differ in key sample and data collection parameters however, which goes towards explaining why conversion happens at a much faster rate in an MSS experiment. Crystals used for the Horrell et al. MSOX study were approximately 300 x 200 x 200 µm, with a 30 x 30 µm beam, and an X-ray flux of 1.6 x 10^{11} photons s^{-1} utilised, a dose rate of 300 kGy s^{-1}. This is compared to the average 8 x 8 x 8 µm crystals used in this fixed target experiment, an 8 x 8 µm beam, and a higher X-ray flux of 3 x 10^{11} photons s^{-1}, a comparative dose rate of 550 kGy s^{-1}. Although it has been demonstrated experimentally that radiation damage in protein crystals is mitigated when using smaller X-ray beams (Sanishvili et al., 2011), there is an increase in X-ray exposure per unit sample volume when using microcrystals with microbeams compared to an MSOX large crystal method. This leads to photoelectron trajectories distributing through a greater percentage of total crystal volume in microcrystals, inducing ligand turnover via secondary radiation damage events throughout a high proportion of the crystal, explaining ligand turnover occurring at lower dose in this MSS method when compared to an MSOX method on a larger AcNiR crystal.

It is apparent from figure 4.7 that the switch from small to large cell for any particular microcrystal yields a cell that is consistent with the 'damaged large cell' of a particular dose rather than the large cell at dose point 1. The increase in unit cell for both polymorphs begins immediately upon irradiation, while the switching of polymorphs is minimal within the first 100 ms before proceeding rapidly. The lack of overlap between the two populations implies a specific structural change between polymorphs and could suggest that expansion of the small cell acts as a trigger/seed

leading to subsequent polymorph swapping. The precise mechanism by which unit cell expansion between AcNiR-small and AcNiR-big remains unclear, though insights may be gained from known site-specific radiation damage phenomena; as decarboxylation of aspartic acid residues has been well characterised (Burmeister, 2000; Holton, 2009; Garman, 2010), the dual conformation loop region containing residues Asp188 and Glu189 is of note.

When comparing dose points 1 and 15, election density of Asp188 remains clear at both dose points, while Glu189 was disordered at both dose points. This similarity between both points may point toward solvent rather than being involved in crystal contacts. Other possible explanations for the dose driven polymorph exchange may be related to hydration or thermodynamic factors, which could arise from heating of the microcrystal and surrounding mother liquor in the beam, or the generation of gases by radiolysis (Meents et al., 2010). An interesting follow-up experiment to this would be performing the same experiment at cryogenic temperature. 'Mini chips' that comprised of only a single city block can be mounted on a pin and then cryocooled. can be used at cryo as the entire city block can be covered by the nitrogen stream at the endstation. By collecting and processing data from multiple mini chips, it would be possible to check whether the polymorphic unit cell is still apparent and the comparative dose rate at which the unit cell flip takes place.

4.4.3. Tracking global radiation damage in MSS data sets

As seen in figure 4.7 there is an initial plateau region or lag phase spanning the first 35 kGy of exposure, before the expected exponential decay in diffracting power seen

throughout the rest of the data. This intensity decay could arise from a number of factors: it is possible that this is an artefact of exceeding the count rate limit of a single photon-counting detector, such as the PILATUS3, resulting in a plateau, though care was taken however to ensure count rates were below the maximum of 10 x 10^6 counts per second per pixel, the maximum that PILATUS3 detectors are capable of accurately recording. In this instance, the maximum observed counts in a Bragg spot were ~8000, or 0.4 x 10^6 counts per second per pixel, only ~4% of the maximum count rate, with the more typical count rates observed ~<4000, or less than 2% of the maximum count rate. Furthermore, crystals are not rotated at all during data collection, meaning count rates should be steady throughout the time taken to record an image and the count-rate correction applied by the detector is extremely accurate. A lag phase can result from the outrunning of global radiation damage effects, resulting from the effects of phenomena such as beam induced heating. When site specific damage dominates in the initial period of data collection, the global intensity decay will deviate from an exponential decay (Owen et al., 2014; Sygusch & Allaire, 1988). It has also been theorised that Gaussian non-tophat profile beams can affect intensity decay, with a non-uniform beam profile contributing to a non-exponential dose response, arising from non-uniform illumination of crystals producing non-uniform damage; this can lead to initial increases or plateaus of integrated intensities with dose, imitating the effect of a delayed onset of damage (Warkentin et al. 2017). Additionally, any timing error in chip motion coupling with detector triggering, could significantly reduce the diffracting power of the first image. In this experiment, as with all chip-based experiments, the beam size is approximately equal to the aperture size, due to the availability of different chips that have different aperture sizes and the

tuneable beam at I24, that can be used to change the beam size to accommodate larger apertures/crystals. Any timing error in chip motion and/or detector triggering would result in the beam profile exceeding the aperture limits, therefore reducing the intensity incident on the aperture

In terms of the horizontal beam profile, chip aperture and motion only, 68% of the beam (FWHM 8 μm) falls within the chip 7 μm aperture assuming the beam and centre of the aperture are perfectly aligned and co-centred. However, if the chip aperture is offset by only a few μm there can be a significant drop in intensity; if the chip aperture is offset by 1, 2, or 3 μm the intensity incident on the aperture falls by 3, 11, or 23%, respectively. Consequently, small errors in chip aperture positioning during the initial alignment stages of the experiment can result in large errors in the recorded intensity at the end of an experiment. Furthermore, as this is evident at the start of the data collection, this may be an artefact of chip movement; as the chip decelerates into position co-centred with the beam this deceleration may lead to small vibrations the induce unwanted movement in the chip. This movement may cause an initial fluctuation in incident on the chip aperture due to slight unwanted vibratory movement of the chip aperture in and out of the beam. Even though extensive care is taken before, between and in the development of experiments to tune the stages as to eliminate this as a source of systematic error, it cannot be discounted as a cause of this initial lag phase.

4.5. Appendices

4.5.1. Example stills_process input parameters

```
spotfinder.lookup.mask=mask.pickle
#spotfinder.filter.min_spot_size=2
integration.lookup.mask=mask.pickle
#significance_filter.enable=True #being tested, check with Danny first
#significance_filter.isigi_cutoff=1.0
#refinement.parameterisation.detector.fix = None


geometry {
  detector {
    panel {
      fast_axis = 0.9999983227108509, 0.0, -0.001831550022519978
      slow_axis  =  -3.4478783685469325e-06,  -0.9999982281102588,  -
0.0018824888990550894
      origin    =    -216.01488427460455,    226.7536585966384,    -
308.42815247014596
      }
    }
  }


indexing.stills.refine_candidates_with_known_symmetry=True

indexing {
  known_symmetry {
    space_group = P213
    unit_cell = 96.5 96.5 96.5 90 90 90
  }

  refinement_protocol.n_macro_cycles = 1
  refinement_protocol.d_min_start=2.5
  basis_vector_combinations.max_refine=5
  stills.indexer=stills
  stills.method_list= ftt1d real_space_grid_search
  multiple_lattice_search.max_lattices=3
}

verbosity=10
integration {
  integrator=stills
  profile.fitting=False

  background {
    simple {
      outlier {
        algorithm = null
      }
    }
  }
}
profile {
  gaussian_rs {
    min_spots.overall = 0
  }
}
```

```
refinement {
  parameterisation {
    beam.fix=all
    detector.hierarchy_level=0
    auto_reduction {
      action=fix
      min_nref_per_parameter=1
    }
    treat_single_image_as_still=True
  }
  reflections {
    outlier.algorithm=null
    weighting_strategy.override=stills
    weighting_strategy.delpsi_constant=1000000
  }
}
```

4.5.2. Example prime input parameters

```
data = ../weezer/int_dats/prime_input_38908_images_dose_1.dat
run_no = dose_1_all_cells
title = weezer_dose01
icering {
  flag_on = False
  d_upper = 3.9
  d_lower = 3.85
}
scale {
  d_min = 1.8
  d_max = 6.0
  sigma_min = 1.5
}
postref {
  residual_threshold = 5
  residual_threshold_xy = 5
  scale {
    d_min = 1.8
    d_max = 6.0
    sigma_min = 1.5
    partiality_min = 0.1
  }
  crystal_orientation {
    flag_on = True
    d_min = 2.0
    d_max = 45
    sigma_min = 1.5
    partiality_min = 0.1
  }
  reflecting_range {
    flag_on = True
    d_min = 2.0
    d_max = 45
    sigma_min = 1.5
    partiality_min = 0.1
  }
  unit_cell {
    flag_on = True
    d_min = 2.0
```

```
    d_max = 45
    sigma_min = 1.5
    partiality_min = 0.1
    uc_tolerance = 3
  }
  allparams {
    flag_on = False
    d_min = 0.1
    d_max = 99
    sigma_min = 1.5
    partiality_min = 0.1
    uc_tolerance = 3
  }
}
merge {
  d_min = 1.48
  d_max = 30
  sigma_min = -3
  partiality_min = 0.1
  uc_tolerance = 3
}
target_unit_cell = 96.9,96.9,96.9,90,90,90
flag_override_unit_cell = False
target_space_group = P213
target_anomalous_flag = False
flag_weak_anomalous = False
target_crystal_system = None
n_residues = 360
indexing_ambiguity {
  mode = Auto
  index_basis_in            =            /dls/i24/data/2017/nt14493-
63/processing/merged/CuNIR_cryo_rotation_free.mtz
  assigned_basis = None
  d_min = 3.0
  d_max = 10.0
  sigma_min = 1.5
  n_sample_frames = 300
  n_selected_frames = 100
}
hklisoin                =                /dls/i24/data/2017/nt14493-
63/processing/merged/CuNIR_cryo_rotation_free.mtz
hklrefin = None
flag_plot = False
flag_plot_expert = False
n_postref_cycle = 0
n_postref_sub_cycle = 1
n_rejection_cycle = 1
sigma_rejection = 3
n_bins = 20
pixel_size_mm = 0.172
frame_accept_min_cc = 0.25
flag_apply_b_by_frame = False
flag_monte_carlo = False
b_refine_d_min = 99
partiality_model = Lorentzian
flag_LP_correction = True
flag_volume_correction = True
flag_beam_divergence = False
n_processors = 20
gamma_e = 0.003
voigt_nu = 0.5
```

```
polarization_horizontal_fraction = 1.0
flag_output_verbose = False
flag_replace_sigI = False
percent_cone_fraction = 5.0
isoform_name = None
flag_hush = False
timeout_seconds = 300
queue {
  mode = None
  qname = psanaq
  n_nodes = 12
}
isoform_cluster {
  n_clusters = 2
  isorefin = None
  d_min = 3.0
  d_max = 10.0
  sigma_min = 1.5
  n_sample_frames = 300
  n_selected_frames = 100
}
```

4.5.1. RADDOSE3D input

```
Crystal

Type Cuboid
Dimensions 8 8 8
PixelsPerMicron 0.5
AbsCoefCalc  RD3D
UnitCell  96.8 96.8 96.8 90 90 90
NumMonomers  12
NumResidues  340
ProteinHeavyAtoms Cu 2 S 10
#SolventHeavyConc P 425
SolventFraction 0.4
Beam

Type Gaussian
Flux 3.0e11
FWHM 8 8
Energy 12.8
Collimation Rectangular 24 24
Wedge 0 0
ExposureTime 0.02
# AngularResolution 2
```

RADDOSE3D Output Data

```
Cuboid (Polyhedron) crystal of size [8, 8, 8] um [x, y, z] at a
resolution of 2.00 microns per voxel edge.
Simple DDM.
Gaussian beam, 24.0x24.0 um with 8.00 by 8.00 FWHM (x by y) and
3.0e+11 photons per second at 12.80 keV.
Wedge 1:
Collecting data for a total of 0.0s from phi = 0.0 to 0.0 deg.

Crystal coefficients calculated with RADDOSE-3D.
Photelectric Coefficient: 2.34e-04 /um.
Inelastic Coefficient: 2.06e-05 /um.
```

```
Elastic Coefficient: 1.92e-05 /um.
Attenuation Coefficient: 2.73e-04 /um.
Density: 1.24 g/ml.

Average Diffraction Weighted Dose         : 0.011456 MGy
Elastic Yield                             : 5.15e+05 photons
Diffraction Efficiency (Elastic Yield/DWD): 4.50e+07 photons/MGy
Average Dose (Whole Crystal)              : 0.020273 MGy
Average Dose (Exposed Region)             : 0.020273 MGy
Max Dose                                  : 0.032066 MGy
Average Dose (95.0 % of total absorbed energy threshold (0.01 MGy)):
0.022010 MGy
Dose Contrast (Max/Threshold Av.)         : 1.46
Used Volume                               : 100.0%
Absorbed Energy (this Wedge)              : 1.31e-08 J.
Dose Inefficiency (Max Dose/mJ Absorbed)  : 2439.8 1/g
Dose Inefficiency PE (Max Dose/mJ Deposited): 2493.3 1/g
Final Dose Histogram:
Bin  1,  0.0 to  0.1 MGy: 100.0 %
Bin  2,  0.1 to  3.4 MGy:  0.0 %
Bin  3,  3.4 to  6.7 MGy:  0.0 %
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 %
```

Chapter 5: Estimating zero dose structures: SFX vs SSX

5.1. Introduction

Obtaining a functionally relevant three-dimensional model when collecting X-ray data from protein crystals is a necessity that underpins X-ray crystallography. These structures provide insight into the function of proteins on the atomic level, as well as developing our understanding of how they interact with other proteins or with small molecules. These biological interpretations are proposed purely on the assumption that the structure derived from diffraction data is the true and accurate representation of the protein of interest, representative of its native (or ligand bound) state and free from artefacts. As covered in detail in section 1.1, global radiation damage effects can limit crystal lifetimes when collecting room temperature data (Garman, 2010), therefore macromolecular crystallography is typically carried out at cryogenic temperatures to mitigate radiation-damage-induced structural perturbation (Garman & Owen, 2006; Holton, 2009).

Despite the much faster onset of global radiation damage, there is an increasing recognition of the importance of carrying out crystallographic studies at room temperature in order to determine a more accurate depiction of structure/function relationship, as well as protein dynamics, in-vivo. This was demonstrated in a study by Fischer et al. in 2015, whereby data were collected from crystals in a fragment-based ligand discovery experiment at both cryo and room temperature, revealing transient binding sites at room temperature that would be abolished at cryogenic temperatures (Fischer, Shoichet, & Fraser, 2015). The importance of collecting room

temperature data is further underlined in a recent publication from Keedy et al., where use of a multitemperature data collection method allowed the solution of an allosteric network by revealing 'hidden' low-occupancy conformation states for protein and ligands (Keedy et al., 2018).

Although the benefits associated with collecting room temperature data are pronounced, structures of peroxidases present a challenge in terms of obtaining an intact, undamaged, structure due to their sensitivity to reduction in synchrotron experiments (Kekilli et al., 2017). Heme peroxidases catalyse many biologically relevant/important reactions, therefore understanding their structural mechanism is of high interest, prompting extensive efforts to obtain 'intact' peroxidase structures at high resolution. However, the Fe III (resting) and Fe IV (intermediate) states are phenomenally sensitive to reduction; the absorption of X-ray photons results in the ejection of photoelectrons, leading to a cascade of reactions as photoelectrons propagate through the crystal, creating hundreds of additional electrons and cations (Sutton et al., 2013). When these are absorbed by metal cofactors, a change in oxidation state affecting conformation and coordination can be apparent much lower than typically used in structure solution (Beitlich et al., 2007).

The high redox potential of peroxidases enables them to utilise Fe(IV) as a potent oxidant (Meharenna et al., 2010), however the reduction of the ferryl heme is associated with a linear increase in the Fe-O bond length of the water molecule bound to the Fe atom, within the extensive H-bonded network associated with the active site dynamics of heme peroxidases. It may be possible to mitigate these radiation damage

effects via serial femtosecond crystallography at XFELs: detailed in chapter 1.2, the ability to collect data before secondary radiation damage effects have time to occur is made possible by collecting X-ray data on the femtosecond timescale. It is therefore of interest to compare 'damage free' data against low dose data collected at a synchrotron light source, due to the longer time scales of synchrotron data collection techniques and the associated conformational changes of biomolecules (Levantino et al., 2015). This has been investigated previously by Keedy et al., 2015: in order to map the conformational landscape of enzymes, SFX data were compared against multitemperature synchrotron datasets, with XFEL data used to confirm that heterogeneity between multitemperature datasets in room temperature synchrotron data is not due to radiation damage (Keedy et al., 2015).

The comparison of synchrotron and XFEL data has been performed previously on peroxidase enzymes; this has been utilised to identify the aforementioned difference in Fe-O bond length, with a damage free bond length of 1.7 Å, and a 0.525 MGy bond length of 1.9 Å collected at 65 K in cytochrome c peroxidase (Meharenna et al., 2010). This high dose dataset was the final dataset of a dose series data collection, comprising 15 total datasets; the first 13 datasets were used to assess how the Fe-O bond length changes as a function of X-ray dose in a preliminary study to their XFEL data collection. It was possible for Meharenna et al. to plot these bond distances and fit a linear function to extrapolate to zero dose, giving an estimated damage free bond length of 1.72 Å.

As explained in chapter 1.3.2, DtpAa is a DyP family heme peroxidase found in the bacterium *Streptomyces lividans*, and has been studied thoroughly at the University of Essex by the Worrall group due to their role in the maturation of copper oxidases (Petrus et al., 2016). Due to this being a heme peroxidase, the rapid active site alterations associated with radiation damage mentioned above make DtpAa an ideal target to investigate methods of collecting damage free datasets at XFELs, and comparing this data with synchrotron techniques that aim to estimate radiation damage effects, namely whether specific bond lengths could be estimated in SFX datasets using SSX. The factor of differentiation in the study compared to that of Meharenna et al. however is the use of identical crystal preparations, sample delivery method for both synchrotron and XFEL data collections, and completing both at room temperature vs 65K (Meharenna et al., 2010). This was achieved by SFX, combined with the MSS technique used previously (see chapter 4.2, 4.3). The technique and results were further explored by online UV-vis spectrophotometry quantify reduction of the heme.

5.2. Methods

Expression, purification of DtpAa were performed and optimised by Dr. Jonathan Worrall and Dr. Amanda Chaplin of the Worrall group at the University of Essex, with Dr. Tadeo Moreno-Chicano from the Hough group at the University of Essex involved in the optimisation of crystallisation. Crystals were obtained by the author and subjected to experimental data collection, and subsequent in-depth data processing and analysis. Methods of sample preparation are explained below.

5.2.1. Recombinant protein expression, purification, and crystallisation

5.2.1.1. Molecular biology and recombinant protein expression

The SLI_2601 gene encoding DtpAa was amplified using polymerase chain reaction from the genomic DNA of *Streptomyces lividans* strain 1326 (*S. lividans* stock number 1326, John Innes Centre). The gene was cloned into the NdeI and HindIII sites of a pET28a vector (Novagen) to create an N-terminal His_6-tagged construct (pET2602) for overexpression in *E. coli* BL21 (DE3) cells.

1.4 L of high-salt LB medium was used to express DtpAa at 37°C, 180 rev min^{-1} to an OD_{600} of 1.0-1.2. At this point 0.25 mM 5-aminolevulinic acid and 100 µM iron citrate (final concentrations) were added for use as a heme precursor and iron supplement, respectively. 0.5 mM (final concentration) β-D-thiogalactopyranoside (Melford) was added to induce the cultures, and carbon monoxide gas was bubbled through the cultures for 30-60 seconds to help stabilise the heme during over-expression. Flasks were sealed and incubated at 30°C and 100 rev min^{-1} for a further 18 hours. Cells were harvested via centrifugation at 10000g for 10 minutes at 4°C.

5.2.1.2. DtpAa purification

In preparation for purification via chromatography, harvested cells were resuspended in 50 mM Tris-HCl, 500 mM NaCl (Fisher) and 20 mM imidazole (Sigma) pH 8 (buffer A), and lysed using an Emulsiflex-C5 cell disruptor (Avestin). Cell debris were cleared via centrifugation at 22000g for 30 mins at 4°C, and the clarified supernatant loaded onto a 5 mL nickel-nitrilotriacetic acid-sepharose column (GE Healthcare) equilibrated with buffer A. A 50 mM Tris-HCl, 500 mM NaCl with 500 mM imidazole buffer (buffer B) was used to elute the bound protein via a linear gradient, with the DtpAa peak eluting at approximately 30-40% buffer B.

DtpAa was pooled and concentrated using a Centricon (VivaSpin) centrifugal concentrator with a 10 kDa cut-off at 4°C, and further purified via size exclusion chromatography using an S200 sephadex (GE Healthcare) size exclusion column, equilibrated with 20 mM NaPi, 100 mM NaCl, pH 7. A major peak eluted with fractions analysed via SDS-PAGE. DtpAa was concentrated, determined by UV-vis spectroscopy (Varian Cary 60 UV-vis spectrophotometer) using an extinction coefficient at 280 nm of 46075 M^{-1} cm^{-1}, and stored at -20°C.

5.2.1.3. Crystallisation and crystal size optimisation

Microcrystal conditions were developed and optimised by Dr. Tadeo Moreno-Chicano at the University of Essex by searching potential conditions from initial hits arising from commercial crystallisation screens. Hits were identified in the commercial screen pHClear in condition D4: 20% PEG 6000, 100mM HEPES pH 7.0. This condition was transferred to small volume batches (20 – 50 µl) at first, with microcrystals of an

average size of 20 x 20 x 20 µm grown for use in this research (fig. 5.1), in batch volumes of 0.4-0.5 ml total volume. This was achieved by mixing a 1:1 ratio of 6.5 mg/ml DtpAa with a crystallisation buffer containing 20% PEG 6000, 100 mM HEPES pH 7.0.

5.2.2. Serial data collection

5.2.2.1. Sample loading

DtpAa microcrystals were loaded onto silicon chips at both the Spring-8 Ångstrom Free-Electron Laser (SACLA) and Diamond Light Source (DLS) using an identical method as explained in detail in section 3.4, and chapter 4.

5.2.2.2. SACLA beamline instrumentation, methods, and parameters

Instrumentation for the movement of chips through the X-ray beam was mounted on beamline BL2 EH3 at SACLA as described in detail in chapter 3.4. Serial femtosecond crystallography (SFX) data were measured at SACLA beamline BL2 EH3 using an X-ray energy of 10.0 keV. A pulse length of 10 fs with a 1.25 x 1.34 µm beam, and a pulse energy of 298 µJ pulse $^{-1}$ with the beam attenuated to 13% of full flux. The translation of the chip between apertures happened during the 33 ms separating the 30 Hz XFEL pulses, making it possible to record a single diffraction image from the centre of each of the 25,600 apertures on the chip in under 14 minutes, whilst ensuring the chip was stationary at the time of data collection. The sample stages, viewing system and sample-loaded chip are within a sealed helium chamber in order to minimise air scatter, with the SACLA MPCCD detector (Kameshima et al., 2014) outside the chamber behind a mylar window (fig. 3.8). A custom entry port allowed rapid exchan-

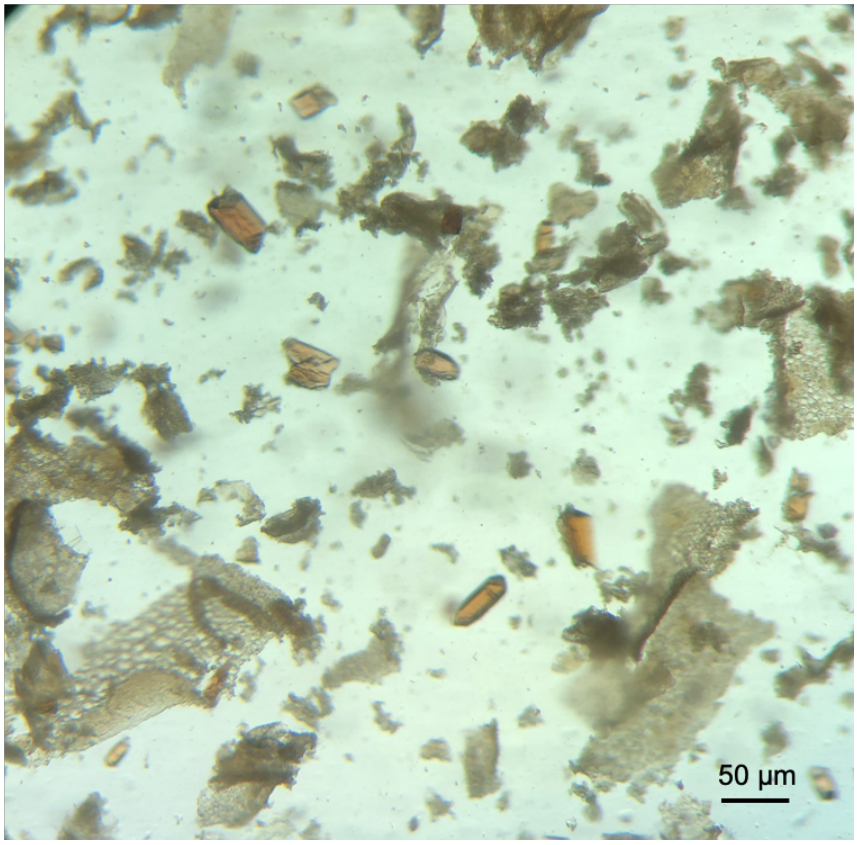


Figure 5.1. DtpAa microcrystals.

Batch grown DtpAa protein crystals vary in size, with cubic and cuboid crystals between 10 – 50 µm in diameter commonly seen.

ge of chips, with minimal helium loss. The program *CHEETAH* (Barty et al., 2014) was utilised to estimate hit rates on the beamline during data collection.

5.2.2.3. DLS beamline instrumentation, methods, and parameters

Serial synchrotron crystallography (SSX) data were measured at DLS beamline I24 using an X-ray energy of 12.8 keV, with no attenuation of the beam. A Pilatus3 6M detector was utilised for the collection of diffraction images while in shutterless mode, meaning the X-ray shutter was not closed between apertures on the chip, remaining open for the duration of the experimental data collection. An MSS method (see chapters 3 and 4) was utilised in order to collect serial datasets from chip apertures. This MSS method was performed by collecting sequential diffraction patterns at each aperture position using an exposure time of 10 ms, with data binned into one dataset per dose interval. This method allowed the collection of two dose-dependent data series, the first of 5 (MSS1) sequential structures, and a second series of 10 dose points (MSS2) after preliminary analysis of MSS1 data suggested that 5 data points were insufficient to establish clear trends. The series of exposures at each aperture position were triggered via a Keysight 33500B signal generator which itself was triggered by a DeltaTau Geobrick LV-IMS-II stage controller. Beam fluxes for MSS1 and MSS2 were 3.2 x 10^{12} and 3.0 x 10^{12} photons s^{-1}, measured using a silicon PIN diode (Owen et al., 2006), with corresponding beam sizes of 7 x 7 and 9 x 8 μm (measured using a knife edge scan), respectively.

5.2.2.4. Data processing, structure solution and refinement

Initial hit finding for data measured at SACLA was carried out at the beamline using *CHEETAH* (Barty et al., 2014), with peak-finding, integration and merging all performed in *CrystFEL* (White et al., 2012) thereafter. DLS data were processed in *DIALS*, with *dials.stills_process* used for indexing and integration with subsequent scaling and merging completed using *PRIME* (Winter et al., 2018).

With data collected from both SACLA and DLS, resolution limits were assessed using $CC_{1/2}$ and R_{split} parameters. Structures were solved using molecular replacement using a starting model obtained from small wedges of data using larger DtpAa crystals mounted between two 6 µm layers of mylar film (not reported in this thesis). In all cases water molecules were removed from the starting model prior to refinement to avoid biasing electron density in areas of interest during refinement.

Structures were refined in *PHENIX*, using *phenix.refine* (Adams et al., 2010) and rebuilt between rounds of refinement using *COOT* (Paul Emsley & Cowtan, 2004), with atoms not well supported by electron density deleted from the model. Validation was performed using JCSG QCCheck server (https://smb.slac.stanford.edu/jcsg/QC/), as well as tools within *PHENIX* including the built-in *MolProbity* (Chen et al., 2010) functionality. Error in bond lengths were estimated using the online diffraction precision indicator (DPI) server (http://cluster.physics.iisc.ernet.in/dpi/; Kumar et al., 2015) and calculating bond error between 2 atoms using equation 8:

$$Bond\ length\ error = \sqrt{DPI\ atom\ a^2 + DPI\ atom\ b^2}$$

Equation 8.

RADDOSE-3D (Paithankar & Garman, 2010) was used to estimate absorbed doses, with dose increments corresponding to the total dose accumulated within the exposure time of the first image.

5.2.3. Online UV-vis micro-spectrophotometry

In situ UV–Vis absorption spectra were collected on beamline I24 at Diamond Light Source. Spectra were collected using mirror lenses (Bruker) mounted in an off-axis geometry (fig. 5.2) and a deuterium halogen light source (Ocean Optics). Spectra were recorded over the wavelength range of 300 – 800 nm using a Shamrock 303 imaging spectrograph (Andor). Data were collected using an X-ray energy of 12.8 keV, an X-ray beamsize of 30x30 µm (FWHM) and a UV–vis focal spot 40 µm in diameter. Spectra were collected using an exposure time of 10 ms, with 2 accumulations per spectrum. During the experiment, data were viewed in *Andor Solis* spectrograph software, with data of interest analysed in *OriginPro* (https://www.originlab.com).

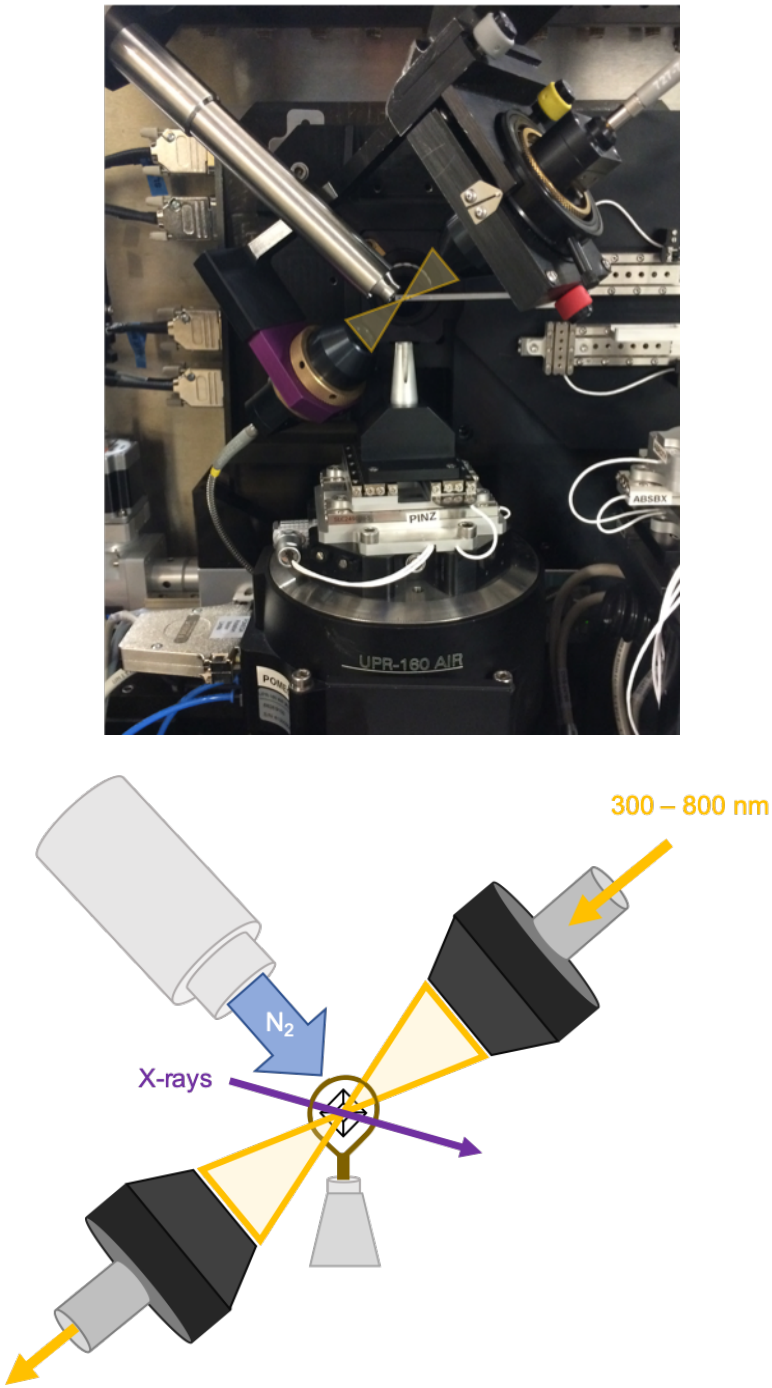


Figure 5.2. Online UV-vis spectrophotometry on I24

(a.) UV-vis mirror lenses mounted in an off-axis geometry on beamline I24 at DLS, with light-sample interaction highlighted in yellow. (b.) Schematic depiction. A cryostream was not used in this study, with collections performed at room temperature (294 K).

5.3. Results

5.3.1. SFX Data collection, processing, and refinement

Chips were loaded using the methods as described in chapter 3. Data were collected from 10 chips (table 5.1) using BL2 EH3 at SACLA, in order to obtain a high enough redundancy to account for the partiality of stills data. Initial estimation of hit rate using *CHEETAH* gave an approximate hit rate of 22.8%, with 64,092 indexed patterns from a possible 281,600, with diffraction to higher than 2 Å on initial viewing of *HDF5* files (fig. 5.3). The *CrystFEL* software suite was used for SFX data reduction in this thesis. Although this is the first SFX structure described in this body of work, a detailed explanation of the method of data processing is available in chapter 3.5, therefore only relevant parameters and analysis will be referred to in this section.

Indexing and integration were handled by the *indexamajig* programme, using parameters that were optimised using an initial AcNiR dataset (not reported in this thesis), an initial detector geometry refined for this beamtime, and unit cell parameters that were defined using a room temperature data collection at I24 prior to SFX data collection. This was run against SACLA data (an *HDF5* file) for each individual chip. The resulting *DtpAa.out* file was used to define accurate SFX unit cell parameters using the *cell_explorer* program within *CrystFEL*, providing a distribution for the unit cell dimensions of DtpAa, to which a Gaussian function can be fitted in order to define the mean values. Updated unit cell dimensions of a = 72.72, b = 68.18, c = 74.62, and β = 105.58 Å, space group $P2_1$ (fig. 5.4), were added to a new unit cell file named *DtpAa_refined.cell*.

Run number	Chip name	Approximate hit rate (%)	Exposure events
23382	Lisbon	12.96	25700
23383	Maniwa	6.5	25700
23384	Nagoya	5.43	25700
23392	Tehran	9.23	25700*
23393	Tehran1	11.56	4830
23394	Urumqi	5.9	25700
23443	Zagreb	60.9	25700
23444	Aleppo	31.3	25700
23445	Banjul	27.1	25700*
23447	Banjul1	15.83	1205
23448	Cuenca	5.83	25700
23449	Durban	90.5	25700

Table 5.1. Chips used to collect ferric DtpAa SFX data

** Incomplete runs due to loss of beam. Data from missed chip areas were collected on the following run. 25700 was the default number of exposure events rather than 25600 (the number of apertures on a chip) to allow for translation between chip city blocks (~33 ms).*

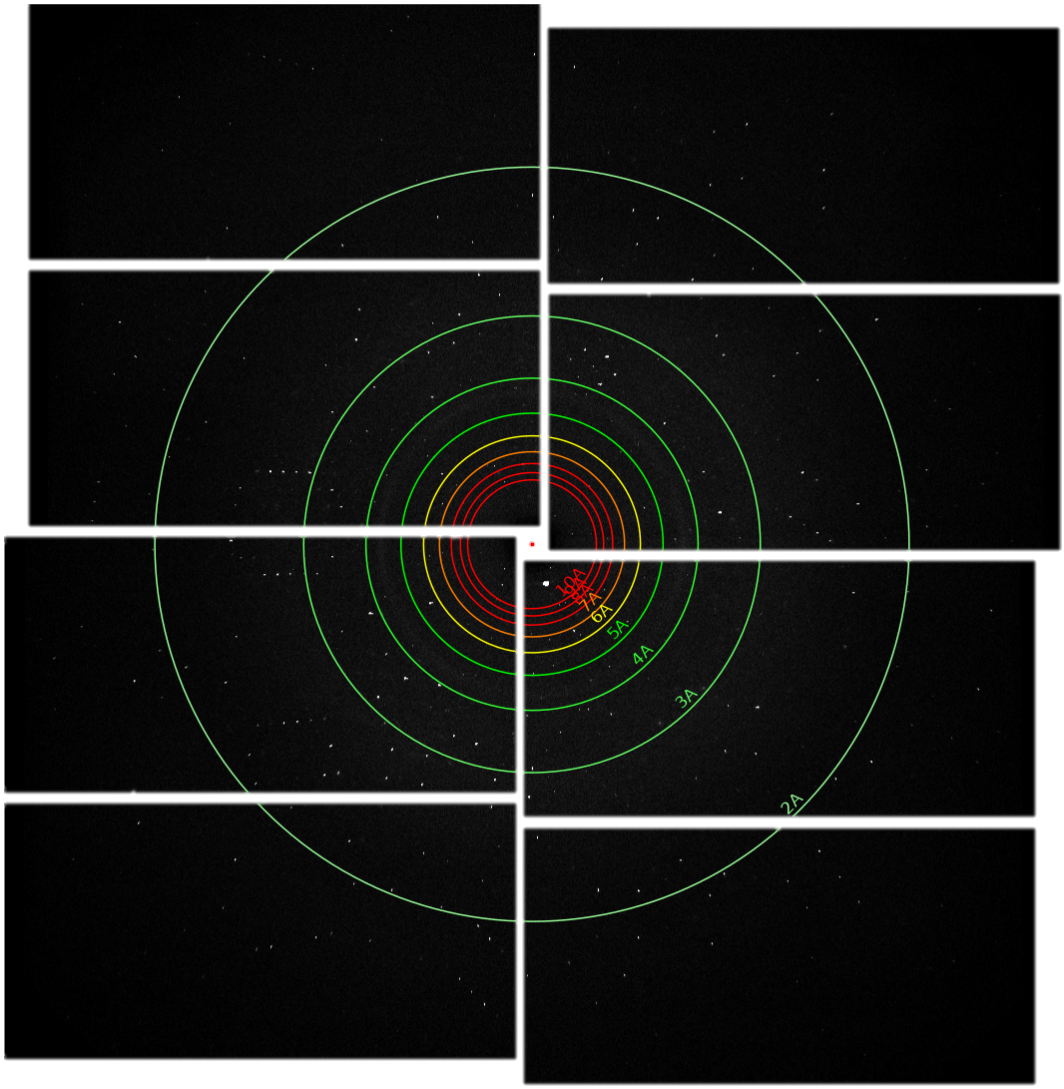


Figure 5.3. DtpAa SFX diffraction

Example image of DtpAa microcrystal diffraction pattern from the initial DtpAa dataset (run number 23383), produced using `hdfsee` *in the* `CrystFEL` *suite. Reflections were seen to greater than 2 Å resolution, with single lattices indicating that chips were not overloaded.*

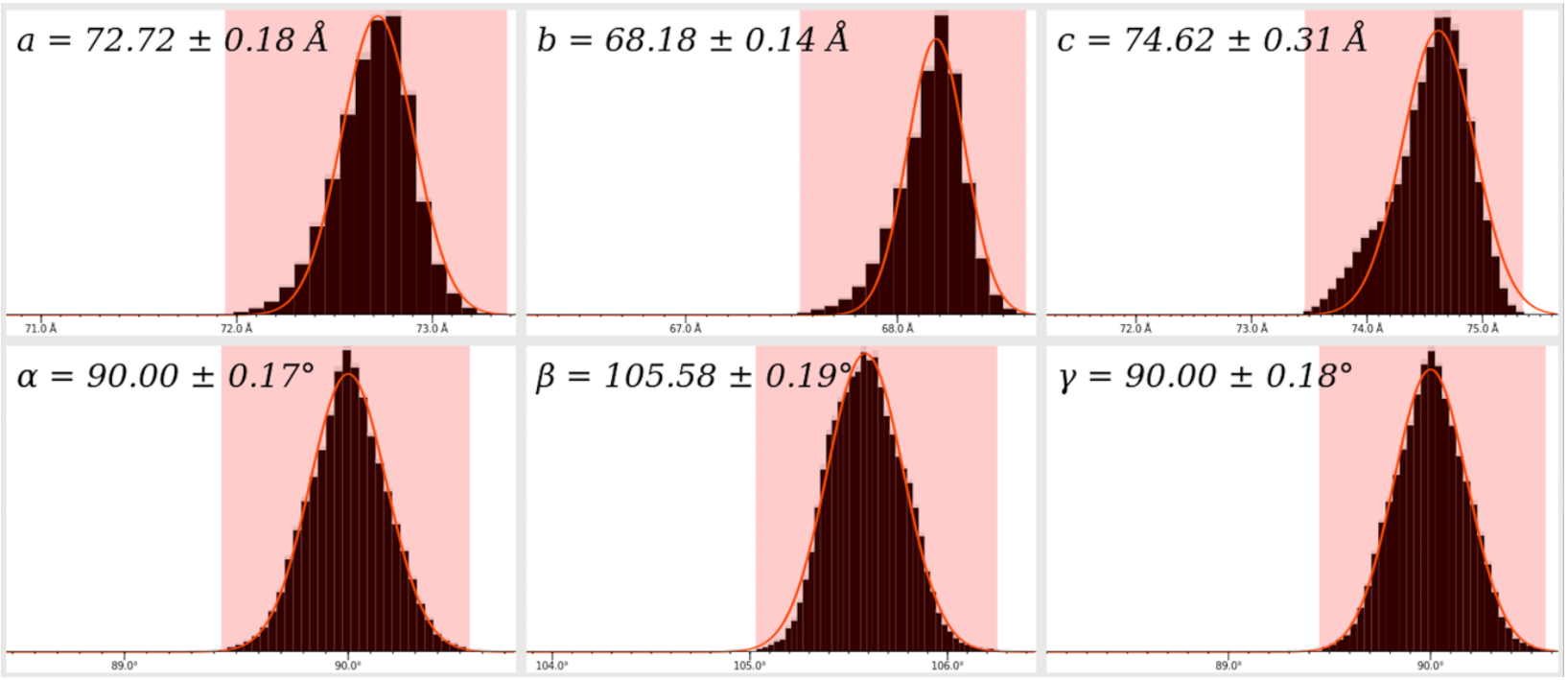


Figure 5.4. DtpAa unit cell distribution showing the refined DtpAa unit cell

The `cell_explorer` *program in the* `CrystFEL` *suite shows an even unit cell distribution of a single unit cell population, with the mean and standard deviation of these dimension measured by fitting a Gaussian function to the data. These updated unit cell dimensions of a = 72.72 ±0.18, b = 68.18 ±0.14, c = 74.62 ±0.31 Å, α = 90.00 ± 0.17, β = 105.58 ±0.19, and γ = 90.00 ±0.18 °, were used to update unit cell parameters used in indexing and integration in order to improve indexing.*

The *detector_shift* script was next used to refine the detector geometry by updating the beam X, Y, position, using the command:

```
$ ./detector-shift updated_DtpAa.out 23382-1.geom
```

This provided an updated geometry file named *23383-1-predrefine.geom*, as well as a scatter plot detailing the X, Y detector shifts (equivalent to the required beam shift) for each pattern (fig. 5.5).

The updated geometry file was added to the *indexamajig* parameters and the process run again for each chip. Indexed reflection data from each chip were combined using the command:

```
$ cat *chip_name*.out *chip_name*.out *chip_name*.out > combined.out
```

From this, it was possible to integrate a total of 73,281 frames. These data were subsequently merged in *CrystFEL* using *process_hkl*, by feeding *combined.out* into a simple bash script *merge.sh* (see section 3.5.4). This script produces three output files, a *combined.hkl* file containing all of the merged reflections, as well as *combined.hkl1* and *combined.hkl2*, each containing half-datasets. Output files *combined.hkl1* and *combined.hkl2* were used for the generation of data quality statistics, using the script stat.sh (see section 3.5.4).

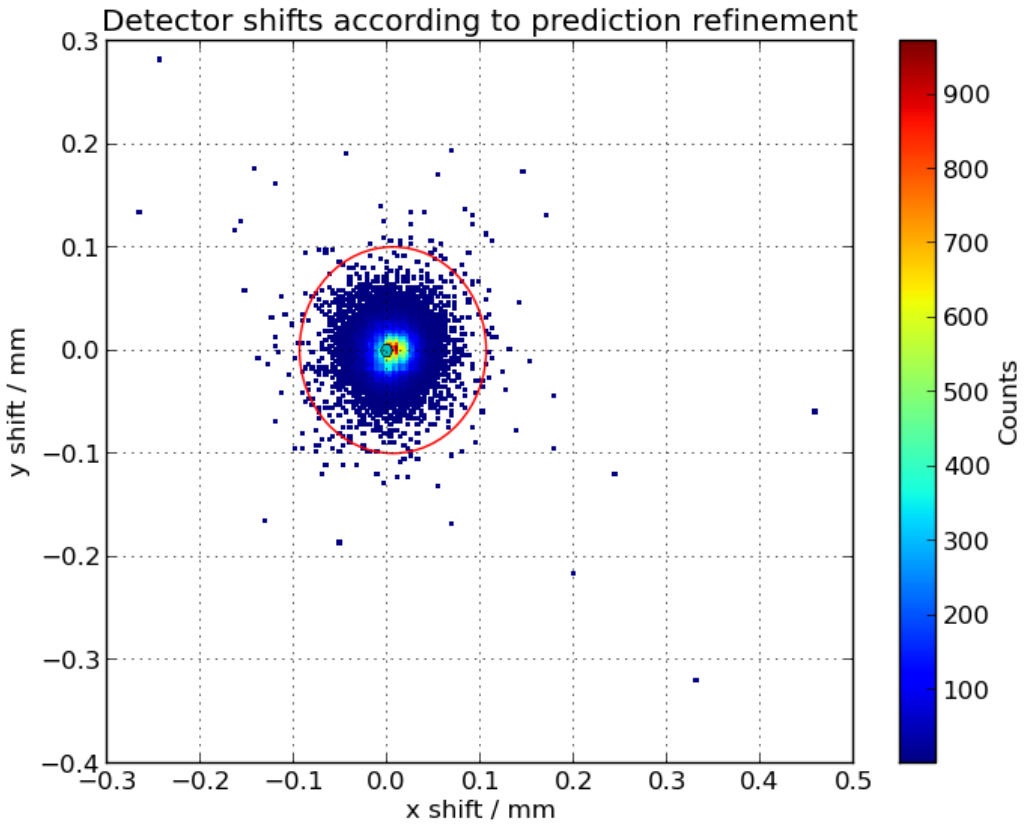


Figure 5.5. Detector shift plot.

Plot generated by the program detector-shift indicating the distribution of measurements for every DtpAa crystal.

The completeness of the data in different resolution shells could be inspected within the shells.dat output file from stat.sh, and was used alongside $CC_{1/2}$ and R_{split} statistics to estimate at what particular resolution limit these data should be cut. In this, it can be shown that when using $CC_{1/2}$ limit of > 0.5 with an R_{split} limit ~50%, these data should be cut around 1.88 Å (tables 5.2, 5.3). This resolution limit was added to the merge.sh script, and run again against the *combined.out* file, with the new *combined.hkl* output converted to an *mtz* using the *create-mtz CrystFEL* executable for subsequent structure refinement.

PhaserMR was utilised for molecular replacement within the *PHENIX* crystallography suite, using a previously solved DtpAa microcrystal structure as the molecular replacement model (data not shown). The structure was refined using *phenix.refine* to an R_{work} and R_{free} of 13.2% and 16.7%, respectively (table 5.4).

5.3.2. DtpAa SFX structure

The crystal structure of DtpAa reveals two monomers in the asymmetric unit, with a ferredoxin-like fold typical of dye decolourising peroxidases (Sugano et al., 2007), represented by a mix of α-helix and β-sheet in the secondary structure motif (fig. 5.6a), with the active site in monomers A and B containing a six-coordinate heme group (fig. 5.6b). As shown in figure 5.7a, at the heme in monomer A residue His326 acts as the proximal ligand with and Fe – N bond length of 2.19 Å. The distal heme coordination site is occupied by a full occupancy water molecule (W1) bound to the heme Fe, with a bond length of 2.40 Å. A second water, W2, is hydrogen bonded to

Shell	$CC_{1/2}$	R_{split} (%)
10.09	0.9866866	6.52
5.17	0.9891377	5.96
4.32	0.9868919	6.1
3.86	0.9903483	5.87
3.54	0.9909499	5.79
3.31	0.9892009	6.27
3.13	0.9871	6.76
2.99	0.9881138	6.72
2.86	0.9877805	7.04
2.76	0.9865419	7.39
2.67	0.9876691	7.74
2.59	0.9857062	8.45
2.52	0.984854	8.96
2.45	0.9829514	9.77
2.4	0.9791849	10.54
2.34	0.9771465	11.05
2.29	0.9753589	11.69
2.25	0.9715545	12.91
2.21	0.9684627	13.21
2.17	0.9630742	15.3
2.13	0.954574	16.84
2.1	0.9510256	17.78
2.07	0.9389044	20
2.04	0.9150496	23.54
2.01	0.8948828	27.42
1.98	0.8946519	27.98
1.96	0.8398976	34.29
1.94	0.8230546	41.71
1.91	0.6779935	52.44
1.89	0.7222474	58.29

Table 5.2. $CC_{1/2}$ and R_{split} statistics for DtpAa SFX dataset

# of reflections	Completeness (%)	Measurements	Redundancy	Shell (Å)
2979	99.97	2927912	982.9	9.16
2889	100	1861184	644.2	4.51
2892	100	1585919	548.4	3.78
2874	100	1361591	473.8	3.37
2874	100	1155615	402.1	3.09
2853	100	1099497	385.4	2.89
2877	100	1082545	376.3	2.74
2866	100	1037389	362	2.61
2879	100	1008215	350.2	2.5
2857	100	970617	339.7	2.41
2834	100	928805	327.7	2.33
2880	100	906468	314.7	2.26
2841	100	854447	300.8	2.2
2832	100	815446	287.9	2.14
2852	100	789879	277	2.09
2839	100	751745	264.8	2.05
2881	100	730944	253.7	2
2849	100	690651	242.4	1.97
2831	100	657737	232.3	1.93
2872	100	610724	212.6	1.9

Table 5.3. Resolution shell statistics for DtpAa SFX dataset

SFX Ferric DtpAa	
Data Collection	
Cumulative dose (kGy)	N/A
Number of images used	72615
Data processing	
Space Group	$P2_1$
Cell dimensions (Å, °)	a=72.72, b=68.18, c=74.62, beta=105.58
Resolution (Å)	41.46-1.88 (1.91-1.88)
R_{split} (%)	7.22 (58.29)
$CC_{1/2}$	99.27 (0.722)
Multiplicity	380.6 (212.6)
Completeness (%)	100 (100)
Refinement	
Number of reflections	57312
Rwork/Rfree	0.132 / 0.167
RMSD bond lengths (Å)	0.009
RMSD bond angles (°)	0.92
Ramachandran plot	
Most favoured (%)	98.5
Allowed (%)	1.50
PDB accession code	6I43

Table 5.4. DtpAa SFX structure data collection, processing, and refinement statistics.

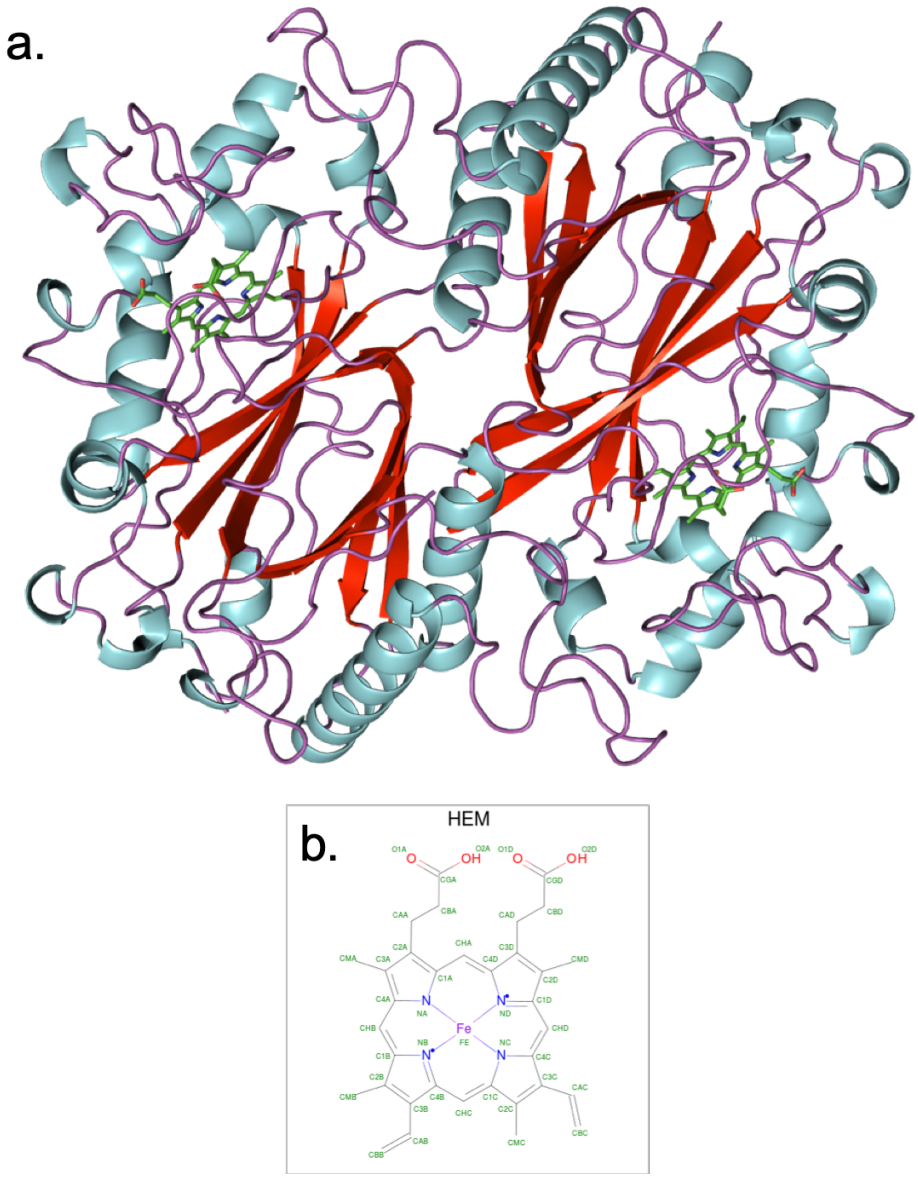


Figure 5.6. Overall structure of the ferric DtpAa SFX structure.

DtpAa is found as a dimer in the asymmetric unit of the SFX structure. (a.) The overall DtpAa SFX structure is shown as cartoon, with each monomer coloured by secondary structure (blue = alpha helices, red = beta sheets, purple = loop) and individual heme groups represented in green. (b.) The chemical structure of the heme group(s).

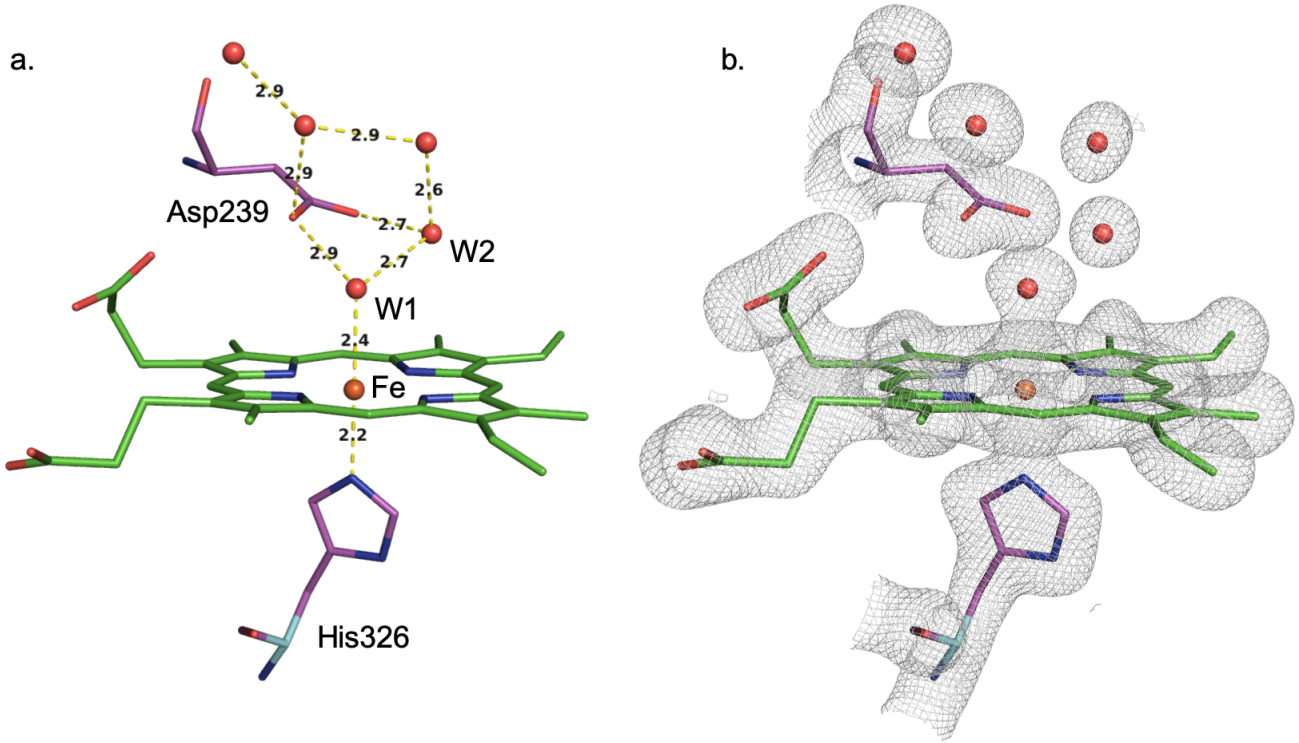


Figure 5.7. Ferric SFX DtpAa active site water network.

(a.) Bond lengths (Å) for the water network in monomer A, including the heme coordinating proximal His326 residue and Asp239. (b.) $2F_o - F_c$ electron density map for the chain of ordered waters found at the active site (monomer A) of the 1.88 Å resolution DtpAa SFX structure, contoured at 1σ. Electron density map shows clearly resolved waters and their interaction with Asp239.

W1 at a distance of 2.68 Å and also interacts with the charged side chains of Asp239 (2.92 Å) and Arg342 (2.74 Å), conserved residues within the DyP family of enzymes. These waters at the distal site are accompanied by further waters that form a proton exchange network, well supported by electron density (fig. 5.7b). Monomer B forms a different active site arrangement however, with an extended electron density at the W1 site above the heme (fig. 5.8). In this instance it has been modelled as two water molecules, and therefore not included in the study of the movement of active site water molecules, as is the focus of this chapter.

5.3.3. SSX data collection, processing, and refinement

Chips were loaded as specified in section 3.3. Serial synchrotron structures from DtpAa were collected from beamline I24 at DLS using the same sample delivery system used for SFX at SACLA, as discussed in 5.3.1 and 5.3.2. Further, a similar MSS data collection protocol as established in chapter 4 was utilised in order to collect sequential datasets from apertures of the chip.

Two separate MSS data collections were performed, with both series were carried out at an energy of 12.8 KeV using an unattenuated X-ray beam. MSS1 comprised 5 sequential datasets, while MSS2 comprised 20 sequential datasets. As with dose calculations presented in previous chapters, absorbed dose was estimated using *RADDOSE-3D* (ZELDIN 2013), with associated dose increments corresponding to the total dose accumulated within the exposure time of the first image.

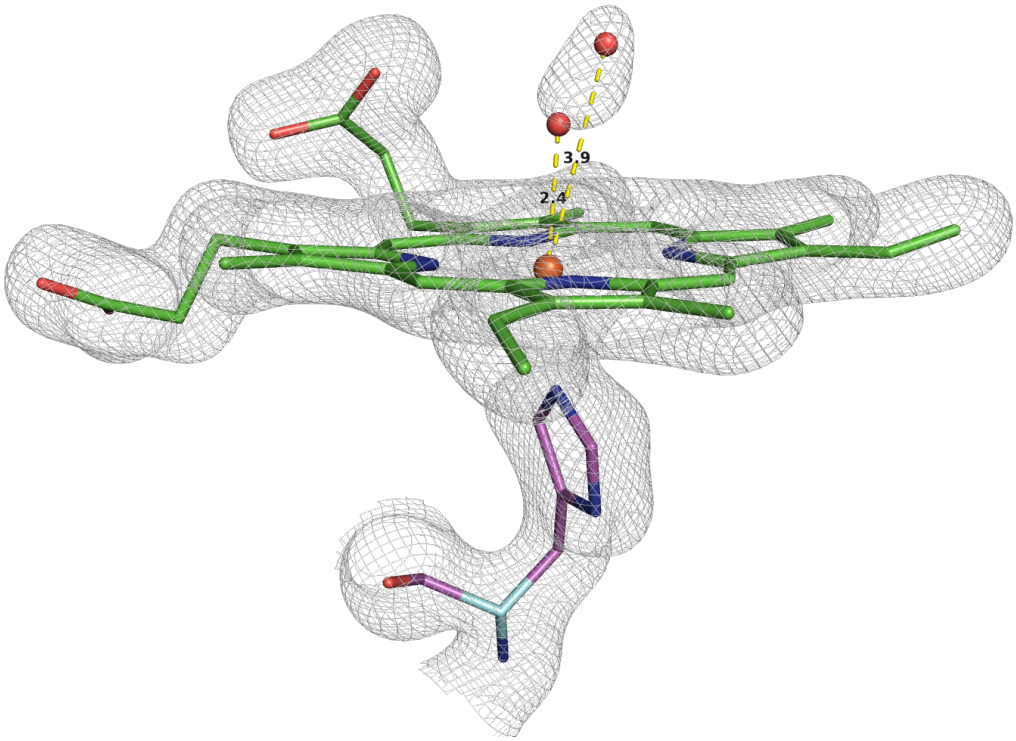


Figure 5.8. Ferric SFX DtpAa heme B.

$2F_o - F_c$ electron density map for waters found at the heme monomer B of the 1.88 Å resolution DtpAa SFX structure, contoured at 1σ. Electron density map is ambiguous for bound water in this monomer, therefore excluded from study form this point forward.

Using the PILATUS3 detector of I24 in shutterless mode allowed twenty successive 10 ms diffraction patterns to be obtained from a single crystal in an aperture in 200 ms (for a twenty dose MSS experiment). After data collection from a crystal, the stages translated the chip to the next aperture and the process was repeated, allowing dose-dependent data to be collected from all crystals on a chip in under 45 minutes. Data were indexed and integrated using *dials.stills_process* in *DIALS v1.8.5* on the Diamond computing cluster. A detailed outline of this processing method can be found in methods section 3.5, though important parameters shall be described in this section.

For both MSS series a joint refinement method (methods 3.5.1) was utilised in order to optimise the number of hits by updating the detector geometry. This was performed on a subset of images, giving an updated geometry for a 310 mm detector distance of:

```
fast_axis = 0.9999909062487288, 0.0, -0.001423903618435921
      slow_axis = -5.358189844691172e-06,  -0.9999929197833288,
      -0.00376302464818867
            origin = -216.72821789476495, 226.31169680726825, -
            307.399222175199
```

Indexing and integration was next performed using the updated detector geometry, with diffraction seen past 1.8 Å in MSS1-ds1 (5 dose series) (fig. 5.9). Data for the 5-dose series were collected from a single chip only, while data for the 20-dose series, data were collected from 3 chips to improve data redundancy. For both the 5- and 20-dose MSS series, *cbf* files were binned using the *get_int_file.py* script (see appen-

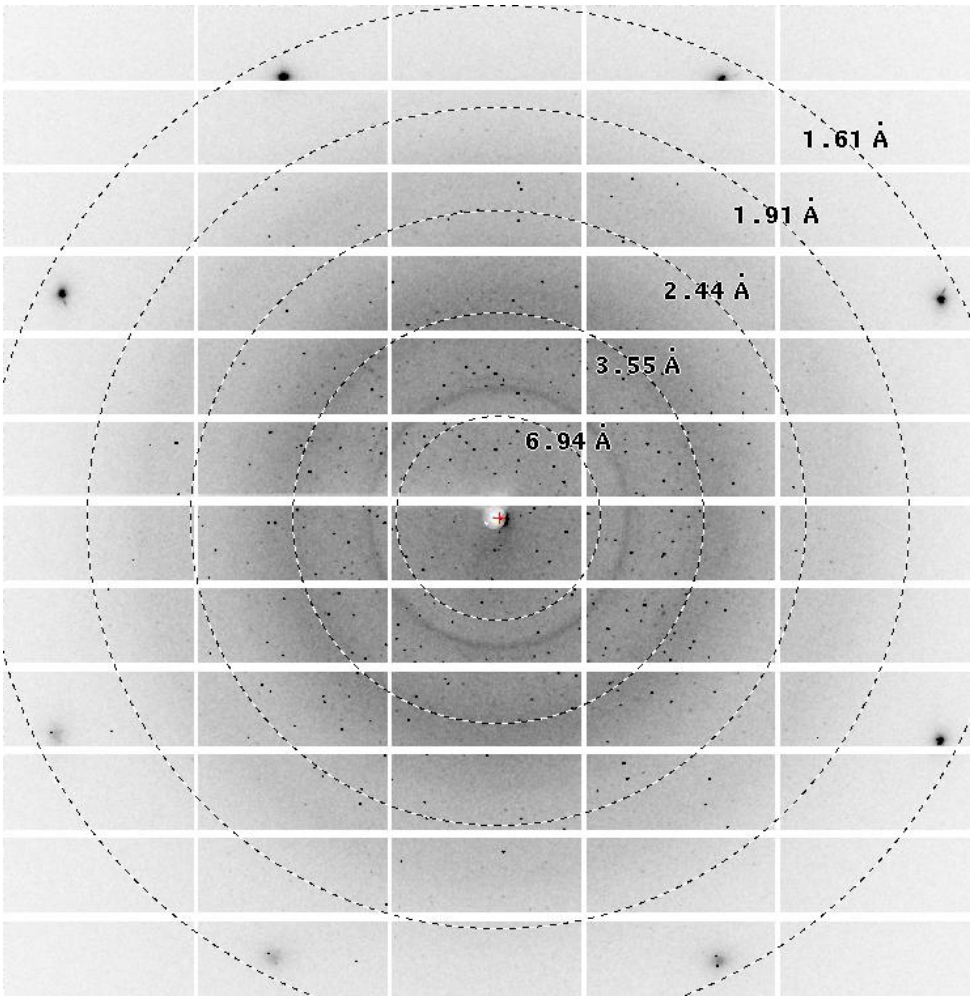


Figure 5.9. DtpAa SSX diffraction from MSS1-ds1

Example image of DtpAa microcrystal diffraction pattern from the initial DtpAa MSS1-ds1 dataset, produced using ADXV*. Reflections were seen to greater than 1.8 Å resolution.*

dices 3.7.5). Numbers of integrated images in dose bins for the 5- and 20-dose series are shown in table 5.5. Only the initial 10 dose bins gave enough indexed images to allow data processing with acceptable data quality metrics, with fewer than 1500 integrated files available for MSS2-ds11 – ds20 (table 5.5). The 5-dose DtpAa MSS experiment (MSS1) was measured with a 32.8 kGy per increment, while the first 10 exposures used for the 20-dose series were measured with a slightly increased incremental dose of 39.2 kGy measured using *RADDOSE-3D* (input parameters for dose calculations can be found in appendices section 5.5.1, 5.5.2).

Data within each dose bin were then scaled and merged using *PRIME* to generate *mtz* files for each dose dependent dataset. This merging process was performed initially in order to assess relevant data quality metrics including R_{split}, $CC_{1/2}$, N_obs (multiplicity), and completeness, in order to determine a resolution cut-off for each dose point. Upon investigation of datasets MSS2-ds9 and MSS2-ds10, the quality of the maps around the active site of interest became too low in resolution due to radiation damage to accurately place waters with confidence, meaning only the first 8 datasets of MSS2 were analysed form this point forwards. A $CC_{1/2}$ of > 0.5, completeness of > 97%, and N_obs of > 10 were used to determine the resolution cut-off, with the resolution range between MSS1-ds1 and MSS2-ds8 of 1.78 – 2.18 Å, with examples of pre- and post-cut data statistics for MSS1-ds1 and MSS2-ds1 in appendices tables 5.5.3, 5.5.4.

Data were then refined using the methods mentioned previously (chapter 3.5.5), though all waters were removed from the molecular replacement model and re-appl-

Dataset	Dose (kGy)	Number of integrated files
MSS1-ds1	32.8	9751
MSS1-ds2	65.6	9833
MSS1-ds3	98.4	10002
MSS1-ds4	131.2	9801
MSS1-ds5	164	9243
MSS2-ds1	39.2	15294
MSS2-ds2	78.4	14719
MSS2-ds3	117.6	14032
MSS2-ds4	156.8	13307
MSS2-ds5	196	12464
MSS2-ds6	235.2	11524
MSS2-ds7	274.4	10058
MSS2-ds8	313.6	8492
MSS2-ds9	352.8	7339
MSS2-ds10	392	6318
MSS2-ds11	431.2	1425
MSS2-ds12	470.4	1299
MSS2-ds13	509.6	1132
MSS2-ds14	548.8	1051
MSS2-ds15	588	939
MSS2-ds16	627.2	833
MSS2-ds17	666.4	780
MSS2-ds18	705.6	692
MSS2-ds19	744.8	679
MSS2-ds20	784	663

Table 5.5. Dose increments and number of used integrated files for DtpAa MSS datasets

ied using the find waters tool in *COOT* during rounds of refinement, in order to remove any experimenter bias in water placement (data reduction and refinement statistics for both series are given in tables 5.6 and 5.7). The initial MSS1 resolution was 1.78 Å, with only a limited loss of resolution during the 50 ms data collection for each microcrystal (table 5.6). higher resolution was seen in MSS2-ds1 at 1.70 Å, with the series ending at MSS2-ds8 resolution of 2.18 Å (table 5.7).

5.3.4. DtpAa SSX structures

In MSS1-ds1, electron density reveals as in the SFX structure a six-coordinate heme, however the distal heme coordination site occupied by W1 has a bond length to the heme Fe of 2.48 Å. As the series continues and dose is accumulated, this trend of an increasing Fe-O bond length continues throughout the series (table 5.8a), with the final dose at 164.0 kGy presenting an Fe-O bond length of 2.97 Å.

The initial dataset in the MSS2 series was refined to a resolution of 1.70 Å, with an Fe-O distance of 2.5 Å. Although 20 data points were collected in this series, only 8 were used in analysis as the electron density became too noisy around the W1 site and the heme pocket, therefore making the placement of water molecules either unreliable, or non-existent. A similar trend of bond length increasing as a function of accumulated dose was seen in the MSS2 series (table 5.8b), with an Fe – H2O bond distance of 3.76 Å in the final dose point of MSS2, with an accumulated dose of 313.6 kGy (fig. 5.10).

Dataset	SACLA SFX	MSS1-ds1	MSS1-ds2	MSS1-ds3	MSS1-ds4	MSS1-ds5
Data Collection						
Cumulative dose / kGy	N/A	32.8	65.6	98.4	131.2	164.0
Integrated frames	73281	9751	9833	10002	9801	9243
Number of images used*	72615	8596	8608	8700	8342	7787
Data processing						
Space Group	P21					
Cell dimensions / Å/deg.	a=72.72, b=68.18, c=74.62, beta=105.58	a = 72.95 (0.33) b = 68.30 (0.13) c = 74.78 (0.43) β = 105.65 (0.21)	a = 72.96 (0.35) b = 68.30 (0.14) c = 74.79 (0.45) β = 105.68 (0.22)	a = 72.97 (0.36) b = 68.29 (0.15) c = 74.82 (0.46) β = 105.72 (0.25)	a = 72.99 (0.39) b = 68.29 (0.17) c = 74.86 (0.49) β = 105.76 (0.28)	a = 73.01 (0.42) b = 68.28 (0.20) c = 74.92 (0.54) β = 105.79 (0.34)
Resolution / Å	41.46-1.88	44.64 - 1.78 (1.81 - 1.78)	44.63 - 1.78 (1.81 - 1.78)	44.62 - 1.83 (1.86 - 1.83)	44.63 - 1.90 (1.93 - 1.90)	44.64 - 1.98 (2.01 - 1.98)
Rsplit	7.22 (58.29)	18.92 (47.72)	18.03 (51.62)	16.97 (56.25)	16.40 (61.71)	16.73 (62.72)
CC1/2	99.27 (0.722)	94.43 (67.71)	95.34 (67.20)	95.99 (66.77)	96.63 (59.07)	96.94 (60.18)
Multiplicity	380.6 (212.6)	27.12 (11.36)	28.36 (10.48)	32.21 (10.79)	34.48 (10.19)	35.48 (9.39)
Completeness (%)	100 (100)	100 (99.9)	100 (99.9)	100 (99.9)	100 (99.8)	100 (99.8)
Refinement						
Number of reflections	57312	67891	67896	62551	55953	49512
Rwork/Rfree	0.132 / 0.167	0.1647 / 0.2032	0.1662 / 0.2114	0.1668 / 0.2142	0.1670 / 0.2120	0.1720 / 0.2322
RMSD bond lengths (Å)	0.009	0.006	0.006	0.006	0.007	0.007
RMSD bond angles (deg)	0.92	0.81	0.82	0.84	0.87	0.89
Ramachandran plot						
Most favoured (%)	98.5	98.3	98.2	98.1	97.5	96.9
Allowed (%)	1.50	1.67	1.81	1.81	2.50	3.06
PDB accession code	6I43	6I7Z	6I8E	6I8I	6I8J	6I8K

* Number of merged patterns in CrystFEL

Table 5.6. Data collection, refinement, and processing statistics for ferric DtpAa SFX and MSS1 datasets

Dataset	MSS2-1	MSS2-2	MSS2-3	MSS2-4	MSS2-5	MSS2-6	MSS2-7	MSS2-8
Data Collection								
Cumulative dose / kGy	39.2	78.4	117.6	156.8	196	235.2	274.4	313.6
Integrated frames	15294	14719	14032	13307	12464	11524	10058	8492
Number of images used	13024	12513	11889	11121	10210	9070	7678	6221
Data processing								
Cell dimensions / Å	a = 72.99 (0.54) b = 68.36 (0.31) c = 74.95 (0.61) β = 105.63 (0.43)	a = 73.01 (0.55) b = 68.36 (0.31) c = 74.94 (0.61) β = 105.67 (0.44)	a = 73.03 (0.55) b = 68.36 (0.32) c = 74.94 (0.62) β = 105.72 (0.44)	a = 73.04 (0.56) b = 68.37 (0.33) c = 74.97 (0.63) β = 105.76 (0.47)	a = 73.06 (0.59) b = 68.37 (0.35) c = 75.03 (0.66) β = 105.81 (0.52)	a = 73.07 (0.63) b = 68.38 (0.37) c = 75.08 (0.70) β = 105.84 (0.54)	a = 73.07 (0.67) b = 68.38 (0.40) c = 75.10 (0.74) β = 105.85 (0.60)	a = 73.07 (0.73) b = 68.39 (0.45) c = 75.06 (0.80) β = 105.82 (0.65)
Resolution / Å	44.71 - 1.70 (1.73 - 1.70)	44.69 - 1.73 (1.76 - 1.73)	44.67 - 1.74 (1.77 - 1.74)	44.68 - 1.78 (1.81 - 1.78)	44.70 - 1.82 (1.85 - 1.82)	44.71 - 1.93 (1.96 - 1.93)	44.72 - 2.03 (2.07 - 2.03)	44.71 - 2.18 (2.22 - 2.18)
Number of reflections	78148	74214	72957	68910	64459	54263	46140	37330
Rsplit	18.24 (73.81)	18.31 (71.41)	18.15 (73.88)	18.05 (71.35)	17.38 (74.98)	17.17 (79.11)	17.90 (78.30)	18.79 (74.51)
CC1/2	95.78 (49.99)	95.87 (49.15)	95.09 (51.04)	95.59 (52.06)	96.56 (50.64)	96.63 (54.40)	96.92 (50.62)	96.87 (53.87)
Multiplicity	52.99 (12.44)	51.36 (13.38)	48.90 (11.86)	47.42 (11.65)	45.97 (10.87)	46.51 (12.73)	44.54 (11.42)	41.36 (11.78)
Completeness (%)	100 (99.9)	100 (100)	100 (99.8)	100 (99.8)	100 (99.5)	100 (99.8)	100 (99.6)	100 (99.5)
Refinement								
Number of reflections	78148	74203	72955	68181	63833	53667	46136	37329
Rwork/Rfree	0.1805 / 0.2346	0.1774 / 0.2337	0.1780 / 0.2304	0.1784 / 0.2321	0.1798 / 0.2359	0.1803 / 0.2355	0.1816 / 0.2492	0.1778 / 0.2555
RMSD bond lengths (Å)	0.007	0.007	0.007	0.007	0.007	0.007	0.008	0.008
RMSD bond angles (deg)	0.85	0.84	0.85	0.85	0.86	0.91	0.94	0.98
Ramachandran plot								
Most favoured (%)	97.91	98.05	98.05	97.08	97.22	96.66	96.52	94.44
Allowed (%)	2.09	1.95	1.95	2.78	2.78	3.34	3.34	5.15
PDB accession code	6I8O	6I8P	6I8Q	6Q31	6Q34	6Q3D	6Q3E	6IBN

Table 5.7. Data collection, refinement, and processing statistics for MSS2 datasets used for structural comparison

Structure	SFX	MSS1-1	MSS1-2	MSS1-3	MSS1-4	MSS1-5
Resolution (Å)	1.88	1.78	1.78	1.83	1.90	1.98
Fe-His (Å)	2.19	2.24	2.25	2.29	2.25	2.23
Fe-O (Å)	2.40	2.48	2.70	2.77	2.71	2.97
PDB accession code	6I43	6I7Z	6I8E	6I8I	6I8J	6I8K

Table 5.8a. Comparison of Fe-His and Fe-O bond lengths in the SFX and MSS1 series

Structure	SFX	MSS2-1	MSS2-2	MSS2-3	MSS2-4	MSS2-5	MSS2-6	MSS2-7	MSS2-8
Resolution (Å)	1.88	1.70	1.73	1.74	1.78	1.82	1.93	2.03	2.18
Fe-His (Å)	2.19	2.27	2.27	2.31	2.30	2.28	2.22	2.13	2.23
Fe-W1 (Å)	2.40	2.50	2.67	2.64	2.93	2.91	3.16	3.37	3.76
PDB accession code	6I43	6I8O	6I8P	6I8Q	6Q31	6Q34	6Q3D	6Q3E	6IBN

Table 5.8b. Comparison of Fe-His and Fe-O bond lengths in the SFX and MSS2 series

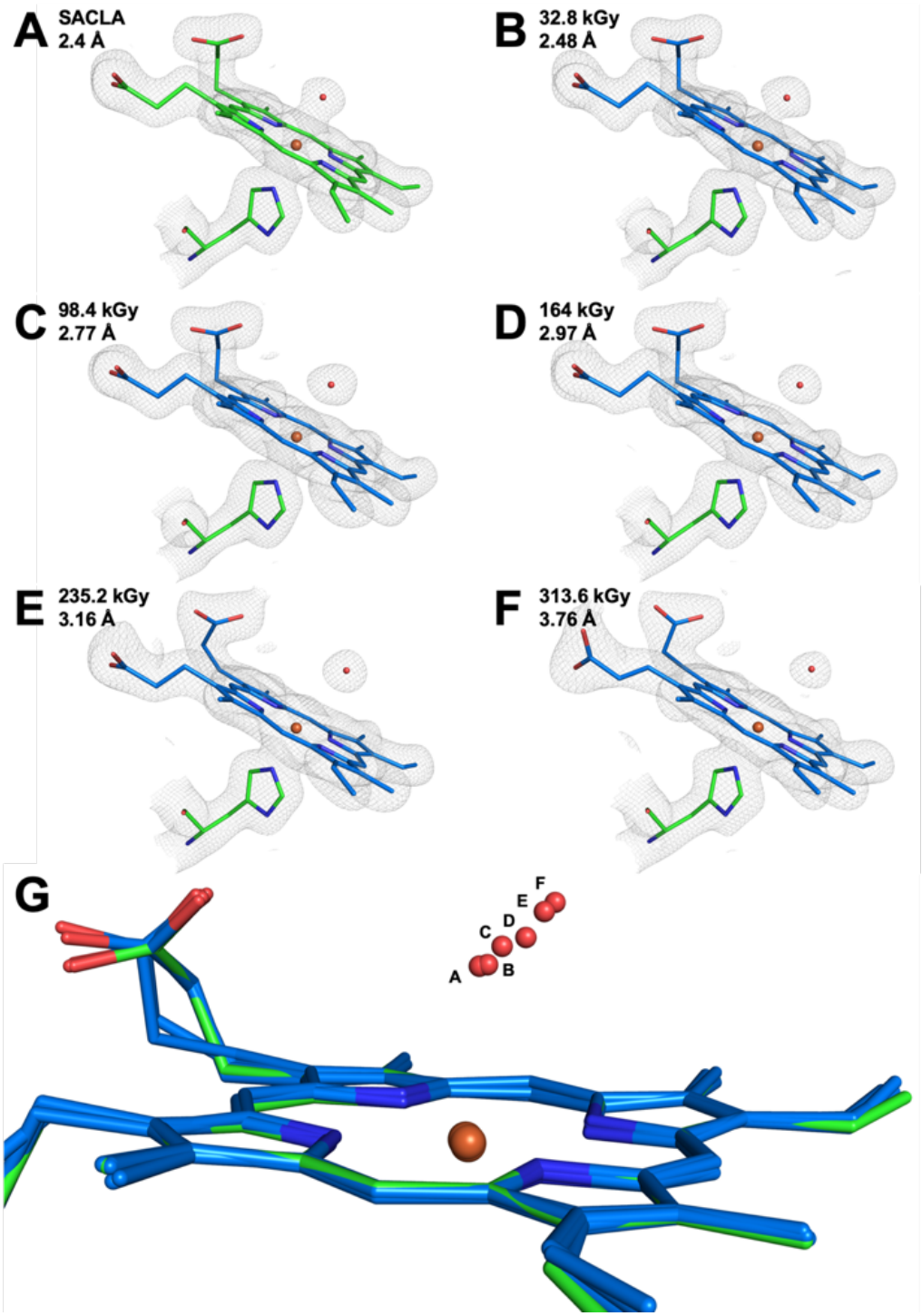


Figure 5.10. Lengthening of Fe-O bond length as a function of accumulated dose.

$2F_o$-F_c electron-density maps contoured at 1σ for the heme environment of (a.) the SFX structure (green) and (b. – f.) selected structures from both MSS1 and MSS2 series (blue). (g.) Superposition of these selected MSS1 and MSS2 structures reveal a dose-dependent migration of the proximal W1 water molecule away from the heme Fe.

Further structural changes were observed in the heme pocket in the highest dose datasets MSS2-ds7 and MSS2-ds8 (274.4 and 313.6 kGy, respectively): a rearrangement of nearby waters in the final dose points are shown with a concurrent flip of the adjacent heme propionate (fig. 5.11). Further, as seen in section 4.4 concerning AcNiR, we again note an increase in unit cell volume across the dose series, indicative of the onset of global radiation damage resulting from disorder of the crystal lattice as dose is accumulated (fig. 5.12).

5.3.5. Characterising photoreduction in DtpAa using online UV-vis microspectrophotometry

Kinetic UV-vis spectra were collected on DtpAa crystals produced in the same batch as those used for SACLA and DLS X-ray diffraction experiments detailed above, with average dimensions of 20 x 20 x 20 µm. *RADDOSE-3D* (Paithankar & Garman, 2010) was used to estimate the absorbed dose, with an average diffraction weighted dose of 10.8 kGy per 20 ms exposure. Spectral changes at 560 nm were investigated, with a peak developing rapidly upon irradiation by the X-ray beam, with photoreduction of ferric DtpAa (fig. 5.13, dark blue) occurring within the first 40 ms of exposure (fig. 5.14), or within 21.6 kGy. This indicates photoreduction occurs extremely rapidly and within the first dataset of both the MSS1 and MSS2 series (32.8, and 39.2 kGy, respectively).

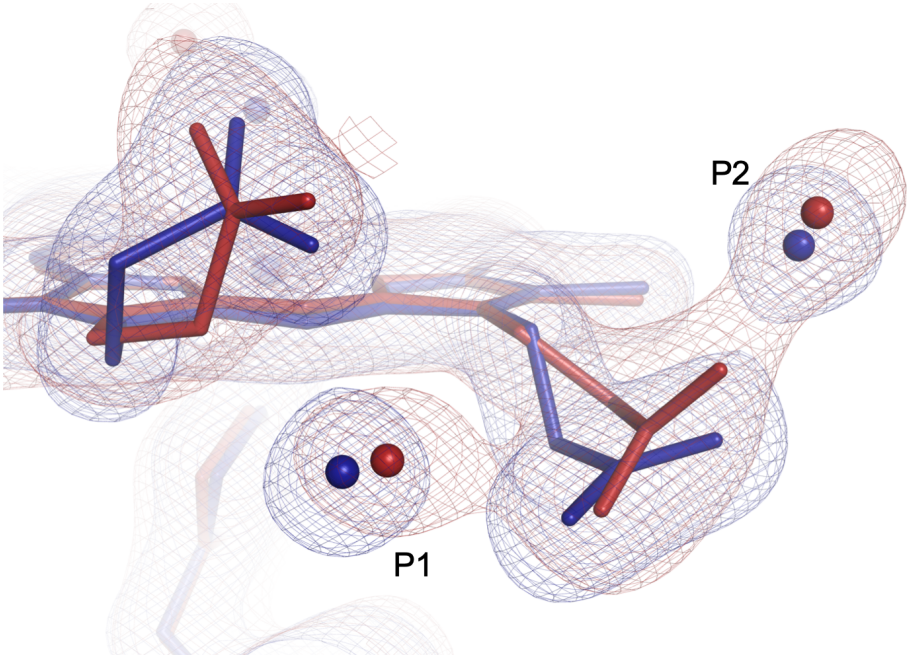


Figure 5.11. Superposition of SFX and 313.6 kGy MSS structures heme propionate environment

$2F_o$-F_c electron density maps contoured at 1σ for the heme propionate environment of DtpAa in the SFX dataset (blue) and from the MSS2-ds8 (313.6 kGy, red). Structures show a shift in corresponding waters located in the propionate environment (P1, P2) between the 'damage free' and highest dose structure, with rearrangement of water molecules possibly driven by a flip in propionate arms, or vice versa.

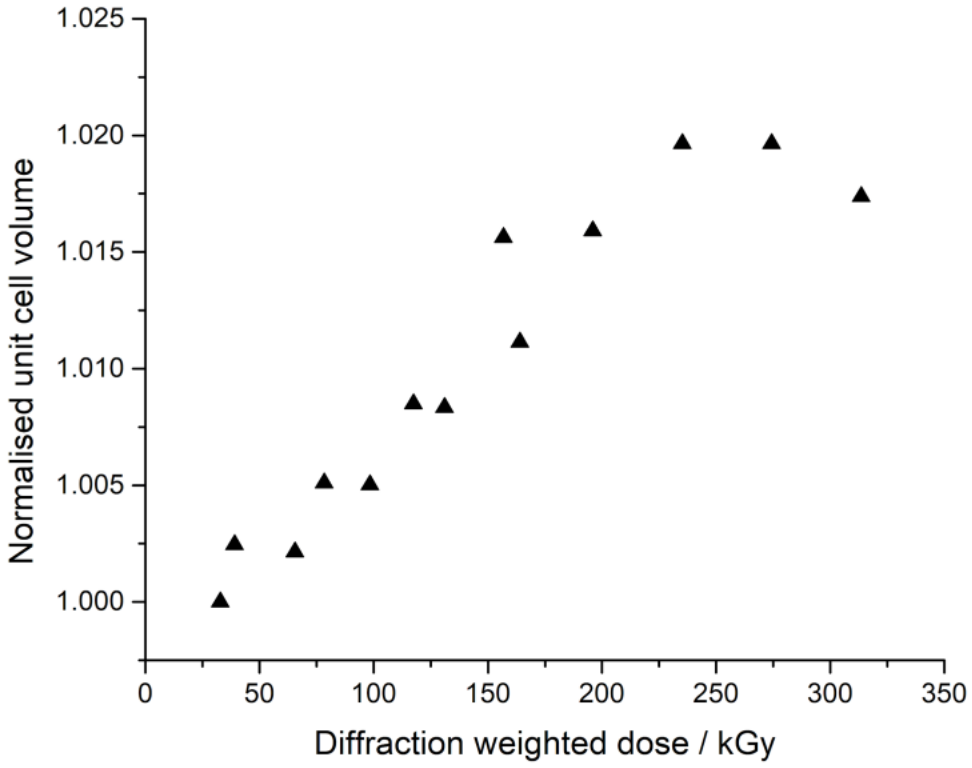


Figure 5.12. Unit cell volume vs. absorbed dose for MSS1 and MSS2 datasets

Increase in unit cell volume is evident as a function of absorbed dose, indicative of the onset of global radiation damage effects.

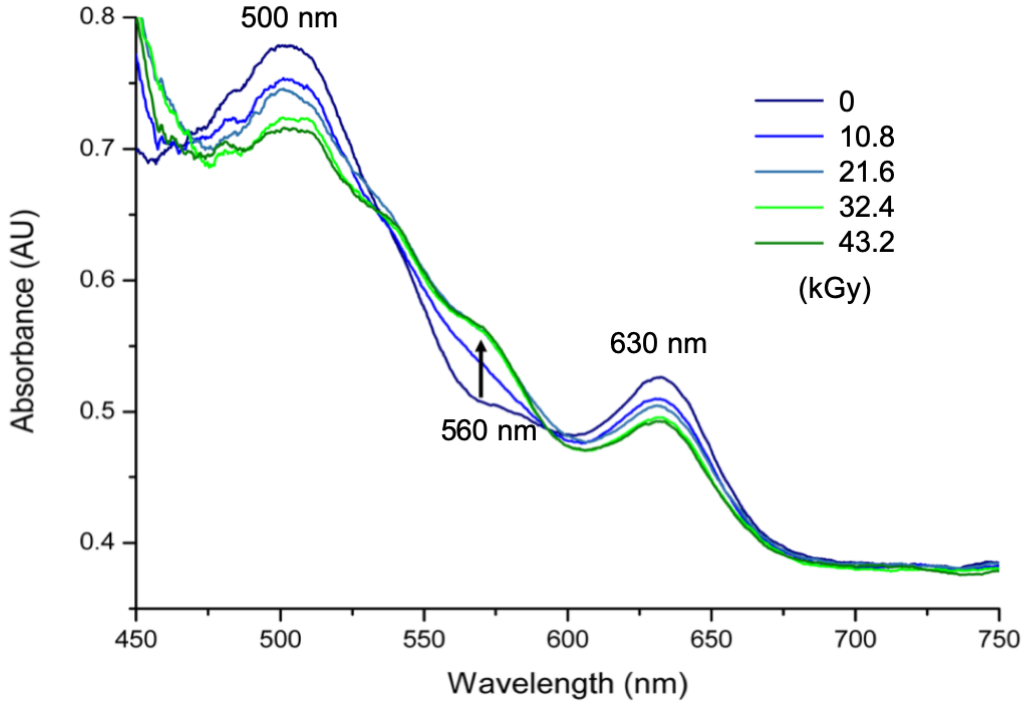


Figure 5.13. Single crystal UV-vis spectra of X-ray photoreduction of DtpAa

Single crystal UV-vis spectra showing the photoreduction of ferric DtpAa (dark blue) with measurements recorded every 20 ms (accumulation of 2 x 10ms exposures). Arrow indicates the change in spectra for 560 nm wavelength, with no change in this feature recorded after 21.6 kGy (40 ms, teal), as shown in timepoints at 32.4 kGy (60 ms, light green) and 43.2 kGy (80 ms, dark green).

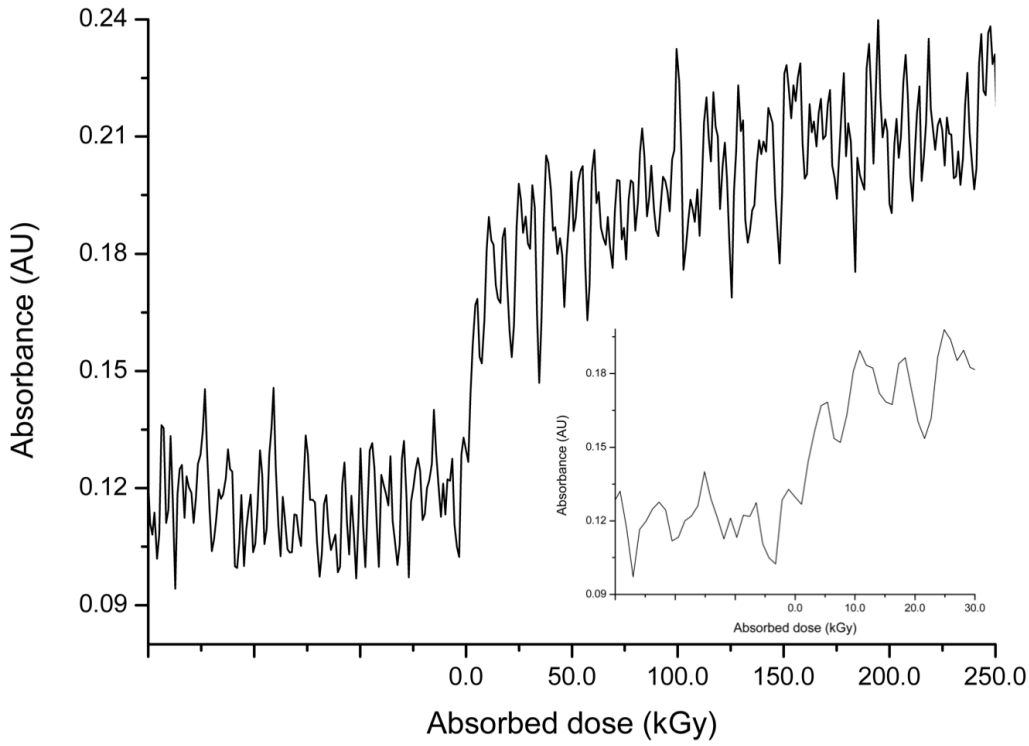


Figure 5.14. Single crystal UV-vis spectra of X-ray photoreduction of DtpAa at 560 nm

Increase in absorbance spectra for 560 nm wavelength. Photoreduction is complete within the first 30 – 35 ms of exposure (inset).

5.4. Discussion

To our knowledge, this is the first reported method that uses an identical sample delivery method and microcrystal sample preparation to compare SFX and dose-resolved SSX structures, at the same temperature, though MSS1 data was initially presented and briefly analysed by Dr. Tadeo Moreno-Chicano (Moreno-Chicano, 2018, PhD thesis). A similar study has been performed, with dose resolved crystallographic data collected at 65 K by Meharenna et al., 2010 compared to the SFX crystal structure of the same cytochrome *c* peroxidase enzyme collected at cryogenic temperature by Chreifi et al., 2016 (Chreifi et al., 2016; Meharenna et al., 2010). There are discrepancies between data collection methods however, such as a large variation in crystal size between crystals used for SFX data collection, ranging from 150 μm – 1 mm, while crystals of 'similar size' to one another were used for synchrotron structures (exact sizes not reported). Further, a rotation method utilised for synchrotron structures, while a helical data collection mode was employed for SFX (Chreifi et al., 2016; Meharenna et al., 2010). Though these differences don't completely hinder the presented comparison, the lack of synergy introduces a risk of increased error, due to having a greater number of differences between experimental variables. Further, the use of room temperature data rather than cryo allows the experimenter to view and understand the mechanism of a protein under close-to biological temperature. Advantageously, it is therefore possible to draw more accurate conclusions when the only differing factor between data collections is the light source.

Although the SFX structure collected from SACLA was solved to a resolution limit of 1.88 Å while MSS-ds1 was solved to a slightly higher resolution of 1.78 Å, it is still appropriate to draw conclusions when comparing structures due to identical crystal batch preparations being used for both data series. Clearly resolved waters are identified in these structures, which is important in recognising the effects of accumulated X-ray dose at the active site of peroxidases. In the SFX structure the Fe-O distance is 0.08 Å shorter than that of MSS1-ds1 (2.40 vs 2.48 Å), indicating that even at extremely low doses, in this case 32.8 kGy, radiation damage effects can be seen at functionally relevant sites. Although this signifies the importance of XFELs as tools that can be used to collect highly accurate, "damage free" structural data, such a small difference in bond length also highlights the ability of SSX to collect incredibly accurate data when low dose data collections at synchrotrons are used in lieu of XFELs. We show with this it is still possible to collect data at a high resolution at a synchrotron, with over a factor of 7 fewer integrated frames (SFX=73281 integrated frames; MSS1-ds1=9751 integrated frames).

Within the rest of the structure, the SFX and MSS1-ds1 models are essentially identical, bar an additional glycine residue at the end of monomer B in the SFX structure, and a 0.7 Å shift in Lys228 toward the surface of the protein in the SSX structure. Both structures exhibit lower than average R-factors when compared to structures in the PDB, when using resolutions bins of 1.9 Å and 1.8 Å, respectively (fig. 5.15). The active site region, which includes heme A, the proximal ligand His326, and conserved dyp side chains Asp239 and Arg342, also remains in the same conformation when comparing both structures.

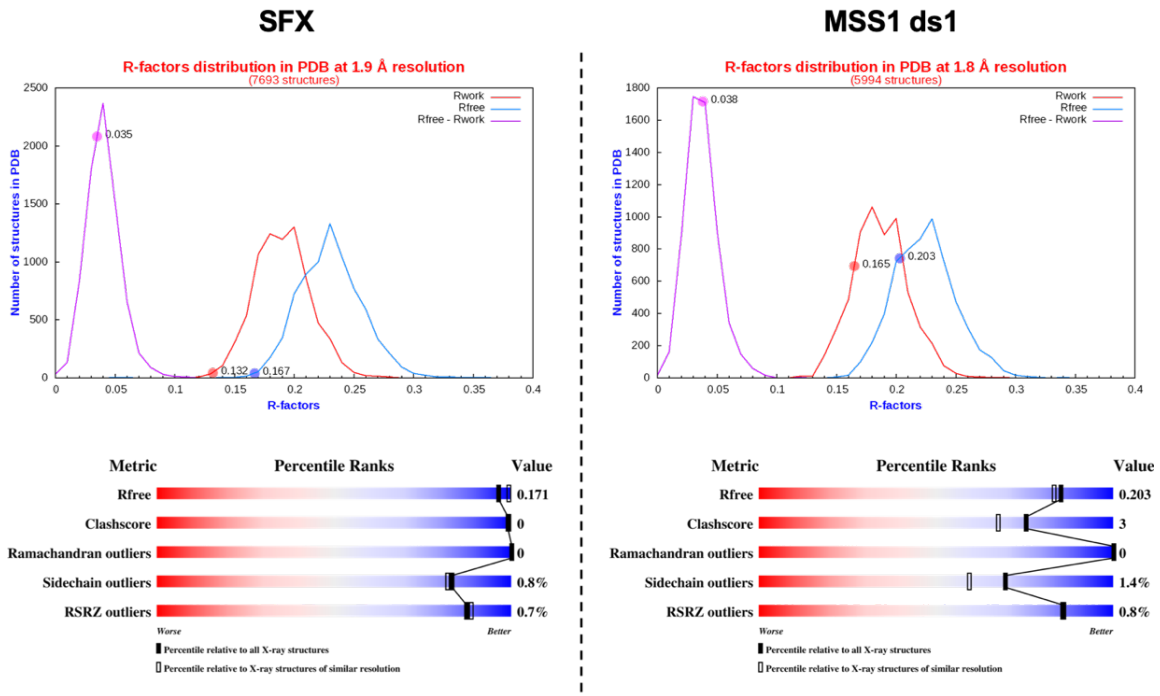


Figure 5.15. R-factor distribution and PDB percentile ranks of the SFX and MSS1-ds1 structures

Both the deposited SFX structure (PDB: 6I43) and deposited MSS1-ds1 structure (PDB: 6I7Z) exhibit lower than average R-factors when compared to other structures deposited in the PDB at the same nominal resolution. Both structures also fall within better than average percentile ranks for factors including clashscore, Ramachandran outliers, sidechain outliers, RSRZ score. Figures were generated via `QC check v3.1` *and PDB Validation Report, respectively.*

The MSS1 5 dose series showed the hypothesised trend of an increase in Fe-O bond length with accumulated dose, though it was unknown whether the distal water had stabilised after the final dose. For this reason, a second MSS dose series (MSS2) of 20 datasets was collected, allowing a greater range of SSX dose points to draw conclusions from. The measured Fe-O distance in the SFX structure is the shortest compared to all of the collected SSX structures. Although a relatively low dose was used in the MSS1-ds1 dataset, its structure is not identical to the SFX structure collected at SACLA. The Fe-O bond length increases as dose is accumulated, as seen in previous publications on the water network around the heme of peroxidase enzymes. A plot of diffraction weighted dose vs Fe-O distance for both datasets shows that this bond length increases almost linearly with accumulated dose (fig. 5.16), with the elongation of this bond length well fitted by a linear function. The extrapolation of this function to the y intercept gives a value of 2.37 ($\pm$0.05) Å, a value extremely close to the 2.40 ($\pm$0.13) Å of the SFX DtpAa structure. This falls within the estimated standard uncertainty in bond length obtained from the DPI value of the Fe and W1 in the SFX structure. Additionally, although relatively low dose is used for each initial MSS data point, we can assume the heme is reduced within the first MSS exposure, as shown by the 21.6 kGy needed to fully reduce ferric DtpAa in the UV-vis data in figures 5.13, 5.14. Intriguingly, it is unknown whether the lengthening of the Fe-O bond is due to a repulsion event caused by this reduction, meaning a follow-up experiment using greater time intervals between MSS datasets could further our understanding on the behaviour of peroxidase active site mechanics post X-ray irradiation.

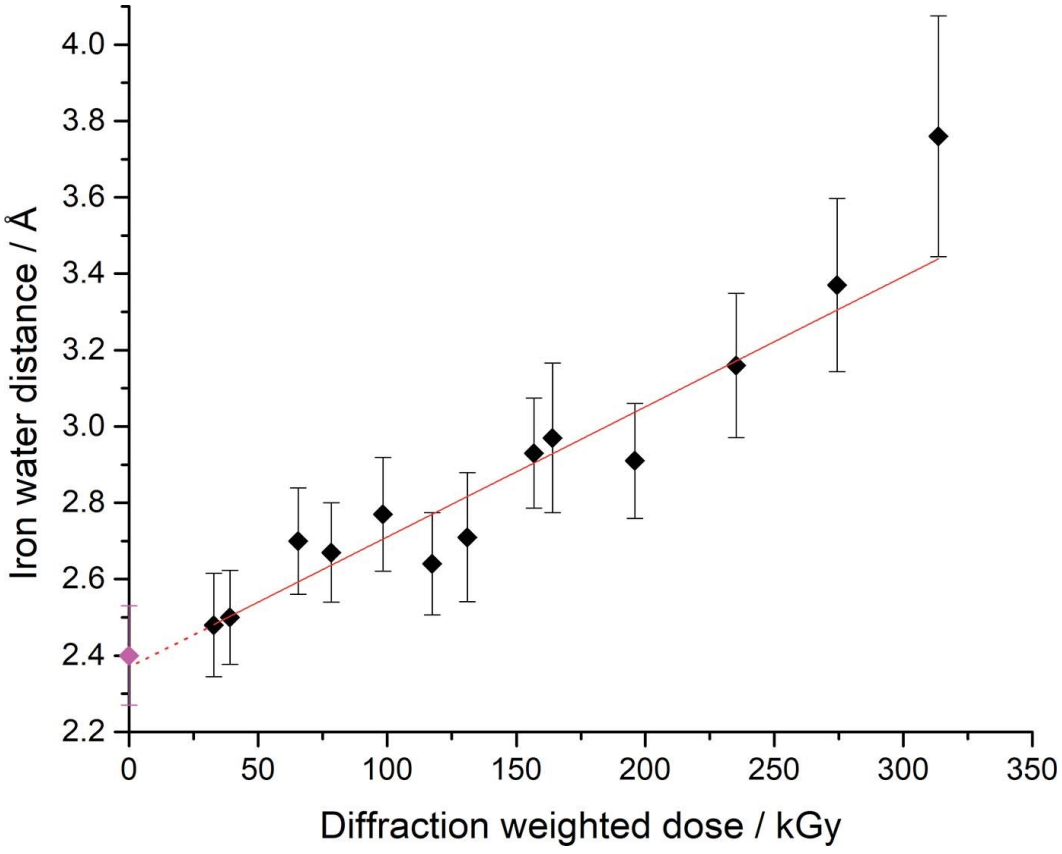


Figure 5.16. Plot of Fe-O distance as a function of X-ray dose from MSS1 and MSS2

The SFX DtpAa bond length is plotted as the zero-dose point (magenta), with both MSS1 and MSS2 bond lengths plotted in black. A linear function is used to fit the elongation of the Fe-O bond (red line) and can be extrapolated back to zero-dose (red line, dashed). This extrapolation to zero-dose gives a value of 2.37 (±0.05) Å, within error of the Fe-O bond length of the SFX structure (2.40 [±0.13] Å). Error bars shown in this plot were produced using the Diffraction Precision Index, and signify estimated standard uncertainty in bond length.

The ability to estimate the SFX Fe-O bond length using an MSS method shows that MSS using synchrotron radiation is extremely effective if XFEL data collections are out of reach. The scarcity of XFEL beamtime is a major limiting factor when the need to produce a structure free from the artefacts of radiation damage is of vital importance. On a practical level, an MSS method can be used to extract functionally relevant features of SFX structures, where MSS data can be used to compare to XFEL structures, and/or used as preliminary data for an XFEL data collection application. Moreover, although it has become possible to outrun most radiation damage processes via the advent of XFEL crystallography, this might not always be the case as explored by several experimental studies; Inoue et al., 2016, used thin diamond films to investigate the temporal evolution of structural damage in crystalline structures, demonstrating a decrease in the intensity of reflections after a time delay of 20 fs, an indication of global damage (Inoue et al., 2016). Site specific damage has also been shown in XFEL data. Nass et al. used an XFEL pulse duration of 80 fs with a photon energy above the absorption edge of iron (the metal atom in the active site) when collecting data from ferredoxin crystals; they showed that one of the two 4FE-4S clusters in the structure was more damaged than the other, indicating that the local environment may play a role in the dynamics of radiation damage at XFELs (Nass, 2019).

When concerning the hardware on synchrotron beamlines, the scope of MSS at synchrotrons will be enhanced by the advancement in hybrid photon counting (HPC) detector technology. New HPC detectors such as the EIGER2, as-well-as charge integrating HPC detectors such as the JUNGFRAU, both from DECTRIS

(www.dectris.com), allow diffraction to be detected within shorter time intervals at synchrotrons due to higher frame rates compared to the PILATUS3 6m (Casanas et al., 2016; Leonarski et al., 2018). Ultimately, this, as well as higher flux densities at new/upgraded synchrotrons, will open further possibilities for time- and/or dose-resolved studies using chips at synchrotron facilities, allowing us to further probe the active site mechanics of DtpAa using MSS data collection methods.

5.5. Appendices

5.5.1. MSS1 RADDOSE-3D data

MSS1 RADDOSE-3D dose calculation input data

```
Crystal

Type Cuboid
Dimensions 20 20 20
PixelsPerMicron 0.5
AbsCoefCalc  RD3D
UnitCell  72.57 68.14 74.8 90 105.5 90
NumMonomers  4
NumResidues  360
ProteinHeavyAtoms Fe 1 S 5
#SolventHeavyConc P 425
SolventFraction 0.45


Beam

Type Gaussian
Flux 3.0e12
FWHM 9 8
Energy 12.8
Collimation Rectangular 24 24
Wedge 0 0                =
ExposureTime 0.01
# AngularResolution 2
```

MSS1 RADDOSE-3D dose calculation output data

```
Cuboid (Polyhedron) crystal of size [20, 20, 20] um [x, y, z] at a
resolution of 2.00 microns per voxel edge.
Simple DDM.
Gaussian beam, 24.0x24.0 um with 9.00 by 8.00 FWHM (x by y) and
3.0e+12 photons per second at 12.80 keV.
Wedge 1:
Collecting data for a total of 0.0s from phi = 0.0 to 0.0 deg.

Crystal coefficients calculated with RADDOSE-3D.
Photelectric Coefficient: 2.02e-04 /um.
Inelastic Coefficient: 2.00e-05 /um.
Elastic Coefficient: 1.85e-05 /um.
Attenuation Coefficient: 2.40e-04 /um.
Density: 1.20 g/ml.

Average Diffraction Weighted Dose         : 0.032872 MGy
Elastic Yield                             : 1.10e+07 photons
Diffraction Efficiency (Elastic Yield/DWD): 3.33e+08 photons/MGy
Average Dose (Whole Crystal)              : 0.025547 MGy
Average Dose (Exposed Region)             : 0.025547 MGy
Max Dose                                  : 0.127374 MGy
Average Dose (95.0 % of total absorbed energy threshold (0.01 MGy)):
0.042880 MGy
Dose Contrast (Max/Threshold Av.)         : 2.97
Used Volume                               : 100.0%
Absorbed Energy (this Wedge)              : 2.50e-07 J.
Dose Inefficiency (Max Dose/mJ Absorbed)  : 508.8 1/g
Dose Inefficiency PE (Max Dose/mJ Deposited): 519.8 1/g
Final Dose Histogram:
Bin  1,  0.0 to  0.1 MGy: 95.0 %
Bin  2,  0.1 to  3.4 MGy:  5.0 %
Bin  3,  3.4 to  6.7 MGy:  0.0 %
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 %
```

5.5.2. MSS2 RADDOSE-3D data

MSS2 RADDOSE-3D dose calculation intput data

```
Crystal

Type Cuboid
Dimensions 20 20 20
PixelsPerMicron 0.5
AbsCoefCalc  RD3D
UnitCell  72.57 68.14 74.8 90 105.5 90
NumMonomers  4
NumResidues  360
ProteinHeavyAtoms Fe 1 S 5
#SolventHeavyConc P 425
SolventFraction 0.45


Beam

Type Gaussian
Flux 3.2e12
FWHM 8 8
Energy 12.8
Collimation Rectangular 24 24
Wedge 0 0
ExposureTime 0.01
# AngularResolution 2
```

MSS2 RADDOSE-3D dose calculation output data

```
Cuboid (Polyhedron) crystal of size [20, 20, 20] um [x, y, z] at a
resolution of 2.00 microns per voxel edge.
Simple DDM.
Gaussian beam, 24.0x24.0 um with 8.00 by 8.00 FWHM (x by y) and
3.2e+12 photons per second at 12.80 keV.
Wedge 1:
Collecting data for a total of 0.0s from phi = 0.0 to 0.0 deg.

Crystal coefficients calculated with RADDOSE-3D.
Photelectric Coefficient: 2.02e-04 /um.
Inelastic Coefficient: 2.00e-05 /um.
Elastic Coefficient: 1.85e-05 /um.
Attenuation Coefficient: 2.40e-04 /um.
Density: 1.20 g/ml.

Average Diffraction Weighted Dose         : 0.039158 MGy
Elastic Yield                             : 1.18e+07 photons
Diffraction Efficiency (Elastic Yield/DWD): 3.00e+08 photons/MGy
Average Dose (Whole Crystal)              : 0.027389 MGy
Average Dose (Exposed Region)             : 0.027389 MGy
Max Dose                                  : 0.152653 MGy
Average Dose (95.0 % of total absorbed energy threshold (0.01 MGy)):
0.049846 MGy
Dose Contrast (Max/Threshold Av.)         : 3.06
Used Volume                               : 100.0%
Absorbed Energy (this Wedge)              : 2.68e-07 J.
Dose Inefficiency (Max Dose/mJ Absorbed)  : 568.8 1/g
Dose Inefficiency PE (Max Dose/mJ Deposited): 581.1 1/g
Final Dose Histogram:
Bin  1,  0.0 to  0.1 MGy: 91.0 %
Bin  2,  0.1 to  3.4 MGy:  9.0 %
Bin  3,  3.4 to  6.7 MGy:  0.0 %
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 %
```

5.5.3. MSS1-ds1 resolution bin statistics

MSS1-ds1 before resolution cut

Bin	Resolution range		Completeness			N_obs	Rmerge	Rsplit	CC1/2	N_ind	I/sigI	I	sigI	I**2
1	44.64	4.61	100	4020	4020	57.41	87.37	17.05	91.69	4020	10.51	369.5	30.6	2.1
2	4.61	3.66	100	3947	3947	42.36	72.53	16.62	90.98	3947	10.42	342.4	28.6	2.04
3	3.66	3.2	100	3897	3898	37.91	70.44	16.28	92.56	3897	6.75	199.2	26.5	2.06
4	3.2	2.91	100	3927	3927	36.27	70.22	17.26	92.21	3927	4.1	104.5	23.6	2.04
5	2.91	2.7	99.9	3886	3888	32.16	69.71	18.92	90.41	3886	3.21	72.3	21.2	1.94
6	2.7	2.54	100	3905	3906	30.63	70.05	20.31	89.9	3905	2.51	51.6	19.5	1.95
7	2.54	2.41	100	3888	3888	28.3	69.41	21.61	89.33	3888	2.22	43	18.4	2.04
8	2.41	2.31	100	3898	3898	27.85	69.63	22.08	88.71	3898	2.01	37.5	17.8	1.96
9	2.31	2.22	100	3890	3890	25.21	69.8	23.5	88.09	3890	1.83	32.7	17.1	1.97
10	2.22	2.14	100	3870	3870	24.53	70.14	24.96	85.88	3870	1.67	29.2	16.8	1.95
11	2.14	2.07	100	3863	3863	22.82	69.72	25.25	86.65	3863	1.53	25.5	16.1	1.96
12	2.07	2.02	100	3886	3887	20.79	70.77	27.97	85.47	3886	1.36	21	15	2.01
13	2.02	1.96	100	3894	3894	19.68	71.13	29.3	84.87	3894	1.26	18.1	13.9	2.02
14	1.96	1.91	100	3867	3867	18.12	72.21	32.63	82.82	3867	1.12	15	13	2.07
15	1.91	1.87	100	3878	3878	15.56	73.15	37.82	77.3	3878	1	12.4	12.1	2.07
16	1.87	1.83	100	3894	3894	14.18	74.09	41.97	75.79	3894	0.93	10.9	11.4	2.18
17	1.83	1.79	100	3882	3883	12.57	75.26	45.45	69.89	3882	0.84	9.3	10.8	2.12
18	1.79	1.76	99.9	3865	3869	10.16	74.54	51.41	68.46	3865	0.83	8.5	10.1	2.27
19	1.76	1.73	99.6	3854	3868	7.63	74.25	58.54	64.4	3854	0.78	7.4	9.5	2.42
20	1.73	1.7	98.7	3796	3845	5.65	73.41	67.23	60.11	3796	0.75	6.8	9.1	2.61
TOTAL	**44.64**	**1.7**	**99.9**	**77807**	**77880**	**24.61**	**77.34**	**19.43**	**94.53**	**77807**	**2.81**	**71.7**	**17.1**	**6.6**

MSS1-ds1 after resolution cut

Bin	Resolution range		Completeness			N_obs	Rmerge	Rsplit	CC1/2	N_ind	I/sigI	I	sigI	I**2
1	44.64	4.83	100	3504	3504	59	88.94	17.02	91.65	3504	10.28	363	30.7	2.11
2	4.83	3.83	100	3437	3437	43.38	73.04	16.59	91.01	3437	10.95	364.9	28.9	2.04
3	3.83	3.35	100	3430	3431	39.42	70.85	16.29	92.32	3430	8.07	249.3	27.5	2.05
4	3.35	3.04	100	3403	3403	36.72	70.34	17.59	91.3	3403	4.93	133.6	24.8	2.05
5	3.04	2.83	100	3401	3401	34.35	69.99	17.49	91.73	3401	3.69	88.7	22.5	1.95
6	2.83	2.66	100	3403	3404	32.03	70	19.12	90.54	3403	2.97	65.2	20.7	1.97
7	2.66	2.53	100	3394	3395	29.97	69.91	20.33	90.19	3394	2.43	49.2	19.3	1.94
8	2.53	2.42	100	3358	3358	28.35	69.5	21.9	88.99	3358	2.2	42.7	18.4	2.05
9	2.42	2.32	100	3431	3431	28.04	69.75	21.9	88.99	3431	2	37.3	17.9	1.95
10	2.32	2.24	100	3378	3378	25.62	69.69	23.32	87.89	3378	1.89	34	17.2	1.99
11	2.24	2.17	100	3358	3358	24.95	70.11	24.32	87.17	3358	1.72	30.5	17	1.97
12	2.17	2.11	100	3414	3414	23.65	69.65	25.14	85.57	3414	1.62	27.7	16.4	1.93
13	2.11	2.05	100	3379	3379	21.75	70.36	26.79	86.16	3379	1.44	23.4	15.7	1.99
14	2.05	2	100	3372	3373	20.62	70.45	27.69	85.59	3372	1.35	20.4	14.6	1.98
15	2	1.96	100	3366	3366	19.51	71.42	29.9	84.31	3366	1.23	17.5	13.7	2
16	1.96	1.92	100	3363	3363	18.19	72.21	32.62	82.82	3363	1.13	15	13	2.09
17	1.92	1.88	100	3401	3401	15.67	73.2	37.47	77.48	3401	1	12.5	12.2	2.07
18	1.88	1.84	100	3351	3351	14.63	73.92	40.87	76.34	3351	0.94	11.1	11.6	2.14
19	1.84	1.81	100	3374	3375	13.24	74.6	43.76	73.94	3374	0.87	9.9	11	2.18
20	1.81	1.78	99.9	3374	3376	11.36	74.84	47.72	67.71	3374	0.86	9.2	10.5	2.18
TOTAL	**44.64**	**1.78**	**100**	**67891**	**67898**	**27.12**	**77.35**	**18.92**	**94.43**	**67891**	**3.11**	**81.2**	**18.2**	**5.9**

5.5.4. MSS2-ds1 resolution bin statistics

MSS2-ds1 before resolution cut

Bin	Resolution range		Completeness			N_obs	Rmerge	Rsplit	CC1/2	N_ind	I/sigI	I	sigI	I**2
1	44.71	4.34	99.9	4824	4827	116.41	94.33	14.34	93.97	4824	10.19	413.6	36.1	2.06
2	4.34	3.45	100	4727	4727	91.59	88.92	15.42	92.74	4727	8.17	312	34.7	2.06
3	3.45	3.01	100	4712	4714	83.65	88.73	16.77	92.31	4712	4.16	143.2	32.3	2.05
4	3.01	2.74	100	4683	4683	80	88.99	17.94	91.76	4683	2.54	76.4	28.9	1.97
5	2.74	2.54	100	4700	4702	68.92	87.36	19.3	90.47	4700	1.97	52.2	25.6	1.9
6	2.54	2.39	100	4691	4691	66.27	88.49	21.23	89.64	4691	1.57	39.1	24.1	1.97
7	2.39	2.27	100	4654	4654	56.56	87.83	23.25	87.56	4654	1.45	34.1	22.7	1.93
8	2.27	2.17	100	4714	4714	55.24	88.18	24.72	85.9	4714	1.26	28.7	22.3	1.95
9	2.17	2.09	100	4666	4667	48.95	87.69	25.78	85.89	4666	1.16	25	21.1	1.92
10	2.09	2.02	100	4661	4661	44.3	88.79	29.47	83.25	4661	0.98	19.4	19.5	1.95
11	2.02	1.95	100	4678	4678	39.31	89.02	31.81	80.3	4678	0.92	16.4	17.4	2
12	1.95	1.9	100	4667	4667	33.92	90.15	38.6	76.39	4667	0.78	12.9	16.2	2.1
13	1.9	1.85	100	4648	4649	28.03	90.03	43.15	74.01	4648	0.71	10.5	14.4	2.18
14	1.85	1.8	100	4712	4713	24.06	89.9	47.15	68.85	4712	0.68	9.2	13.4	2.1
15	1.8	1.76	100	4638	4638	19.14	90.05	55.23	64.21	4638	0.64	7.8	12.1	2.28
16	1.76	1.72	100	4648	4649	15.48	90.1	64.68	54.8	4648	0.6	6.8	11.3	2.32
17	1.72	1.69	99.9	4670	4673	12.37	90.3	75.15	47.87	4670	0.57	6.1	10.6	2.48
18	1.69	1.66	99.7	4641	4654	9.56	90.43	94.25	43.21	4641	0.5	5.1	10.2	3.16
19	1.66	1.63	99.1	4578	4621	6.78	120.82	96.78	62.22	4578	1.08	19.9	10.8	79.45
20	1.63	1.6	96.2	4531	4712	4.79	89.79	153.38	38.02	4531	0.39	3.7	9.8	7.9
TOTAL	**44.71**	**1.6**	**99.7**	**93443**	**93694**	**45.57**	**91.22**	**20.15**	**92.16**	**93443**	**2.04**	**63**	**19.8**	**8.49**

MSS2-ds1 after resolution cut

Bin	Resolution range		Completeness			N_obs	Rmerge	Rsplit	CC1/2	N_ind	I/sigI	I	sigI	I**2
1	44.71	4.61	99.9	4031	4034	120.34	95.1	14.3	94.05	4031	10.11	413.6	36.3	2.1
2	4.61	3.66	100	3957	3957	94.21	89.07	15.03	92.88	3957	9.31	359.1	34.7	2
3	3.66	3.2	99.9	3928	3930	86.99	88.67	16.12	92.56	3928	5.51	200.2	33.8	2.03
4	3.2	2.91	100	3922	3922	80.7	88.96	17.77	91.39	3922	3.19	103.7	30.9	2.02
5	2.91	2.7	100	3893	3894	78.49	88.61	17.98	91.55	3893	2.38	69.3	28.1	1.93
6	2.7	2.54	100	3927	3928	67.85	87.34	19.65	90.04	3927	1.91	50.2	25.4	1.89
7	2.54	2.41	100	3912	3912	67.46	88.48	20.99	89.82	3912	1.58	39.7	24.2	1.98
8	2.41	2.31	100	3904	3904	57.2	87.82	22.79	88.03	3904	1.51	35.7	22.9	1.95
9	2.31	2.22	100	3912	3912	56.42	88.12	24.23	86	3912	1.3	30.1	22.6	1.92
10	2.22	2.14	100	3872	3872	52.81	87.87	24.74	86.35	3872	1.24	27.7	21.7	1.91
11	2.14	2.07	100	3899	3900	47.37	88.06	27.08	84.87	3899	1.09	23.3	20.9	1.92
12	2.07	2.02	100	3895	3895	44.23	88.85	29.76	82.87	3895	0.97	19.1	19.3	1.95
13	2.02	1.96	100	3899	3899	39.53	88.91	31.37	81.1	3899	0.94	16.8	17.5	2.01
14	1.96	1.91	100	3879	3879	36.33	90.16	36.71	76.58	3879	0.8	13.6	16.6	2.04
15	1.91	1.87	100	3903	3903	29.12	89.76	40.64	75.73	3903	0.75	11.5	15	2.12
16	1.87	1.83	100	3886	3887	27	90.08	45.86	71.79	3886	0.69	9.8	14.1	2.2
17	1.83	1.79	100	3888	3888	22.53	90.01	49.08	66.26	3888	0.66	8.7	13	2.08
18	1.79	1.76	100	3903	3903	18.77	90.06	56.4	63.29	3903	0.63	7.7	12	2.31
19	1.76	1.73	100	3903	3904	15.66	90.1	64.66	54.48	3903	0.6	6.9	11.4	2.32
20	1.73	1.7	99.9	3835	3839	12.44	90.04	73.81	49.99	3835	0.58	6.3	10.7	2.53
TOTAL	**44.71**	**1.7**	**100**	**78148**	**78162**	**52.99**	**91.18**	**18.24**	**95.78**	**78148**	**2.31**	**73.6**	**21.6**	**7.09**

5.5.5. UV-vis RADDOSE-3D data

UV-vis RADDOSE-3D dose calculation intput data

```
Crystal

Type Cuboid
Dimensions 20 20 20
PixelsPerMicron 0.5
AbsCoefCalc  RD3D
UnitCell  72.57 68.14 74.8 90 105.5 90
NumMonomers  4
NumResidues  360
ProteinHeavyAtoms Fe 1 S 5
#SolventHeavyConc P 425
SolventFraction 0.45

Beam

Type Gaussian
Flux 3.79e12
FWHM 30 30
Energy 12.8
Collimation Rectangular 90 90
Wedge 0 0
ExposureTime 0.02
# AngularResolution 2
```

UV-vis RADDOSE-3D dose calculation output data

```
Cuboid (Polyhedron) crystal of size [20, 20, 20] um [x, y, z] at a
resolution of 2.00 microns per voxel edge.
Simple DDM.
Gaussian beam, 90.0x90.0 um with 30.00 by 30.00 FWHM (x by y) and
3.8e+12 photons per second at 12.80 keV.
Wedge 1:
Collecting data for a total of 0.0s from phi = 0.0 to 0.0 deg.

Crystal coefficients calculated with RADDOSE-3D.
Photelectric Coefficient: 2.02e-04 /um.
Inelastic Coefficient: 2.00e-05 /um.
Elastic Coefficient: 1.85e-05 /um.
Attenuation Coefficient: 2.40e-04 /um.
Density: 1.20 g/ml.

Average Diffraction Weighted Dose         : 0.010893 MGy
Elastic Yield                             : 9.01e+06 photons
Diffraction Efficiency (Elastic Yield/DWD): 8.27e+08 photons/MGy
Average Dose (Whole Crystal)              : 0.020998 MGy
Average Dose (Exposed Region)             : 0.020998 MGy
Max Dose                                  : 0.025714 MGy
Average Dose (95.0 % of total absorbed energy threshold (0.02 MGy)):
0.021449 MGy
Dose Contrast (Max/Threshold Av.)         : 1.20
Used Volume                               : 100.0%
Absorbed Energy (this Wedge)              : 2.06e-07 J.
Dose Inefficiency (Max Dose/mJ Absorbed)  : 125.0 1/g
Dose Inefficiency PE (Max Dose/mJ Deposited): 127.7 1/g
Final Dose Histogram:
Bin  1,  0.0 to  0.1 MGy: 100.0 %
Bin  2,  0.1 to  3.4 MGy:  0.0 %
Bin  3,  3.4 to  6.7 MGy:  0.0 %
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 %
```

Chapter 6: Dark progression of radiation damage in MSS datasets

6.1. Introduction

The data collected from DtpAa crystals using an MSS approach in the previous chapter establish that the Fe-O bond distance between the heme Fe and the oxygen of a bound water (W1) increases approximately linearly as a function of accumulated dose. However, it is unknown whether the increase in bond length a function of accumulated dose alone, or a function of both absorbed dose and elapsed time. This question is examined in this chapter by increasing the time between MSS datasets from tens of milliseconds to many tens of seconds, observing whether the bond length still increases with accumulated dose, and how similar these bond lengths are at similar dose intervals to those observed in chapter 5, figure 5.10.

It has been shown previously that the 'dark progression' of radiation damage in crystals is apparent when X-rays are turned off, at temperatures between 180 and 240 K (Warkentin et al., 2011), while no dark progression of damage has been observed at cryogenic temperatures (Holton, 2009). Further, the examples of dark progression presented in Warkentin et al., 2011, were observed as changes in global radiation damage metrics only, namely as a rise in relative B-factor as a function of time: each time the X-rays were shut off (shown as a time interval between data points), a rise in B-factor is evident; at each temperature, larger time intervals lead to more damage, with evidence of a saturation of damage with dark interval time seen

at 240 K (Warkentin et al., 2011). The aim of this preliminary results chapter is that if dark progression of radiation damage is apparent when collecting MSS datasets, to examine it at both global and site-specific levels. In doing so, the preliminary results presented in this chapter examine the dose/time dependence of Fe-O bond length change, as to guide future experiments.

6.2. Methods

6.2.1. Recombinant protein expression, purification, and crystallisation

Identical DtpAa crystals to those subject to experimentation in chapter 5 were obtained by the author via a collaboration between the Hough and Worrall groups at the University of Essex, and subjected to experimental data collection, and subsequent in-depth data processing and analysis. Methods of sample preparation including expression, purification, and crystallisation of DtpAa are explained in detail in chapter 5 section 2. Microcrystals of an average size of 20 x 20 x 20 µm were used in this for the research presented in this chapter (see chapter 5, section 2). Microcrystal slurries were achieved by mixing a 1:1 ratio of 6.5 mg/ml DtpAa with a crystallisation buffer containing 20% PEG 6000, 100 mM HEPES pH 7.0.

6.2.2. Serial data collection

6.2.2.1. Sample loading

DtpAa microcrystals were loaded onto silicon chips at Diamond Light Source (DLS) using an identical method as explained in detail in section 3.3.2.

6.2.2.2. DLS beamline instrumentation, methods, and parameters

Serial synchrotron crystallography (SSX) data were measured at DLS beamline I24 using an X-ray energy of 12.8 keV, with no attenuation of the beam. An MSS method was used to collect sequential datasets from chip apertures, performed by collecting sequential diffraction patterns at each aperture position and binning data into one dataset per dose intervals (described previously, in detail, in chapters 3.4 and 3.5).

The Pilatus3 6M detector at DLS beamline I24 was used in shutterless mode for the collection of MSS diffraction images, and set at a detector distance of 320 mm for all datasets presented in this chapter. The X-ray shutter was not closed between apertures on the chip, remaining open for the duration of the experimental data collection, though was closed during long delays (90+ second) between datasets.

3-dose dark progression series (DPS1) was collected from DtpAa crystals with a ~90 second delay between datasets. This was achieved by collecting single 10ms exposures from a single column on a chip, with ~60 seconds needed for data collection from all 3,200 apertures of a single column, and ~30 seconds needed to manually re-initiate the data collection on the GUI for the second and third passes; this was repeated for columns 2-8, with the total data collection performed in < 1 hour (fig 6.1a). In order to build on the interesting but unclear results of the DPS1 datasets, a second DPS series was collected (DPS2), with 5 sequential 10ms exposures collected at each aperture, followed by a ~120 second delay, and then another 5 sequential 10ms exposures collected from the same apertures. The delay between the first and second groups of 5 exposures was again achieved manually using the GUI, with ~90 seconds needed to collect an initial 5-dose series from a half of a chip column, ~30 seconds needed to re-initiate the data collection on the GUI, with the second 5-dose series then collected from the same apertures (fig 6.1b). The series of five 10ms exposures at each aperture position were triggered via a Keysight 33500B signal generator which itself was triggered by a DeltaTau Geobrick LV-IMS-II stage controller. Beam fluxes for DPS1 and DPS2 were 3.97 x 10^{12} and 3.94 x 10^{12} photons s^{-1}, measured using a silicon PIN diode (Owen et al., 2009), with corresponding beam sizes of 8 x 8

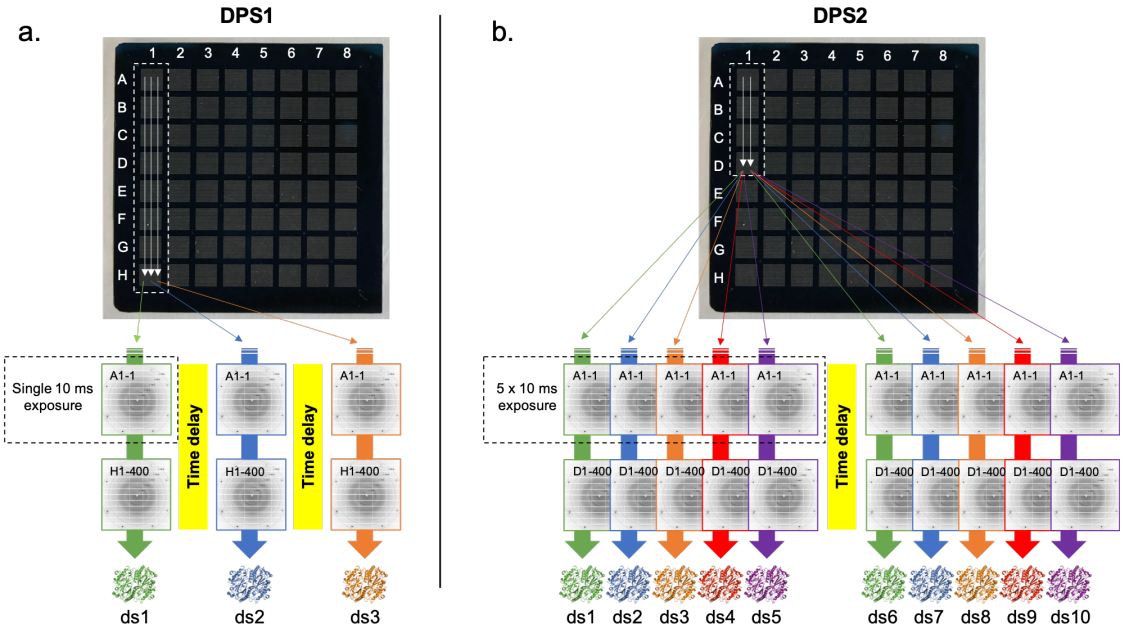


Figure 6.1. Schematic representation of the DPS experiment

(a.) White arrows represent data collection from an entire column of city blocks (aperture A1-1 to aperture H1-400). A single 10 ms diffraction pattern is collected from each aperture in a column taking ~60 seconds. The entire column is the subject to a second 10 ms exposure, with a third collected once the column is complete. With time taken to restart the collection (~30 seconds), the time delays between dose dependent structures (green, blue, orange) are ~90 seconds in DPS1. (b.) White arrows represent data collection from half of a column of city blocks (A1-1 to D1-400). Five sequential 10 ms exposures are taken at each aperture, with entire collection from a half column taking ~90 seconds. A second set of five sequential 10 ms exposures are then taken from the same half column of apertures. By including time taken to restart the collection, the time between the first 5 and second 5 exposures is ~120 seconds.

and 9 x 9 µm used for data collections DPS1 and DPS2 (measured using a knife edge scan), respectively.

6.2.2.3. Data processing, structure solution and refinement

Data were processed in *DIALS*, indexing and integration was handled using *dials.stills_process* version 1.8.5, with subsequent scaling and merging completed using *PRIME* (Winter et al., 2018). Resolution limits were assessed using completeness, $CC_{1/2}$ and R_{split} parameters. Structures were solved via molecular replacement using *phaserMR* in the *PHENIX* crystallographic software suite. The 32.8 kGy room temperature DtpAa SSX structure collected in chapter 2 and subsequently deposited to the PDB (PDB:6I7Z) was used as the search model in molecular replacement. *AutoBuild* wizard in *PHENIX* was used to provide an automated system for model building, and used as an initial process for all structures presented in this chapter. Structures were further refined using *phenix.refine* (Adams et al., 2010), and rebuilt between rounds of refinement using *COOT* (Paul Emsley & Cowtan, 2004). Atoms not well supported by electron density were deleted from the model, with validation performed using the built-in *MolProbity* (Chen et al., 2010) functionality in *PHENIX*. Error in bond lengths were estimated using the online diffraction precision indicator (DPI) server (http://cluster.physics.iisc.ernet.in/dpi/; Kumar et al., 2015). *RADDOSE-3D* was used to estimate absorbed doses, with dose increments corresponding to the total dose accumulated within the exposure time of the first image (Paithankar & Garman, 2010).

6.3. Results

6.3.1. DPS1 SSX data collection, processing, and refinement

DPS1 comprised 3 sequential datasets, with dose 1 comprising single 10ms exposures collected from all apertures in a single column of a chip (aperture A1-1 – H1-400). After data collection from a column was complete the stages translated the chip back to the first aperture, allowing data collection from the same chip column to be repeated. This allowed collection of a second and third dose dataset from each aperture. Data collection from a single column took ~58 seconds, with a further ~30 seconds required to restart the data collection, giving an estimated delay time between datasets of ~1.5 minutes. Data collection from the entire chip therefore took ~36 minutes. The diffraction weighted dose was estimated to be 48.6 kGy per dataset (see appendices for *RADDOSE-3D* input parameters and output file).

Data were indexed and integrated using *dials.stills_process* in *DIALS* on the Diamond computing cluster. A detailed outline of this processing method can be found in chapter 3.5, though important parameters shall be described in this section. For this dose series, data were collected from 2 chips, with 2083 indexed images from a possible 51200 for dose 1, equating to a hit rate of 5.5%. *cctbx.xfel.plot_UC_cloud_from_experiments* was used on a subset of 871 crystals in order to asses unit cell parameters before processing the entirety of the data, with DtpAa was found to be in space group $P2_1$, with unit cell dimensions of a = 72.8 $\pm$ 0.3, b = 68.3 $\pm$ 0.1, c = 74.2 $\pm$ 0.4 Å, β = 105.56 $\pm$ 0.23 ° (fig. 6.2).

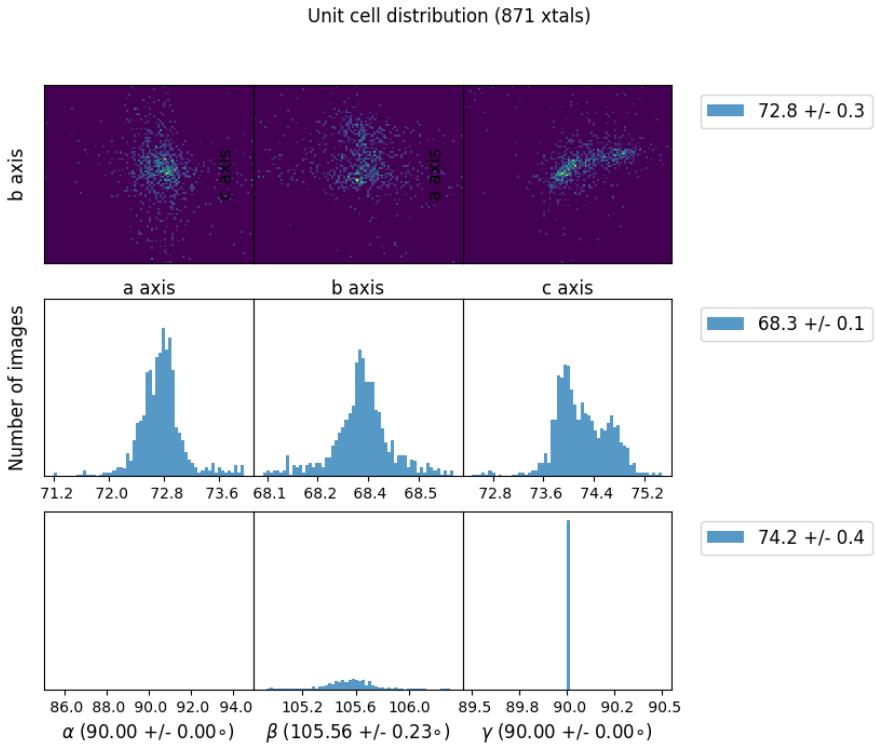


Figure 6.2. DtpAa unit cell distribution.

`DIALS cctbx.xfel.plot_UC_cloud_from_experiments` *shows the unit cell distribution from a subset of DtpAa crystals, used to optimise the unit cell for subsequent data processing. The unit cell was found in* $P2_1$*, dimensions of a = 72.8* $\pm$*0.3, b = 68.3* $\pm$*0.1, c = 74.2* $\pm$*0.4 Å,* β *= 105.56* $\pm$*0.23 °.*

Integrated images were binned via dose using the script *get-int-file.py* (appendices 3.7.5), with subsequent datasets within each dose bin scaled and merged using *PRIME* to generate an *mtz* file corresponding to each dose dependent dataset. Each dataset was initially merged to 1.6 Å in order to assess relevant data quality metrics, including R_{split}, $CC_{1/2}$, N_obs (multiplicity), and completeness, allowing the determination of a resolution cut-off for each dose point. As in previous chapters, outer shell metrics of $CC_{1/2}$ of > 0.5, completeness of > 97%, and N_obs of > 10 were used as thresholds in determining an appropriate resolution cut-off; using this, DPS1-ds1 was cut and merged at a resolution of 2 Å, with doses 2 and 3 merged to a resolution of 2.1 and 2.3 Å, respectively (tables 6.1, 6.2). *AutoBuild* was utilised for initial structure building, with rounds of refinement subsequently performed in *phenix.refine* using automatic weighting, and *COOT*. Structures were refined to an R_{work}/R_{free} of 0.171/0.228 for DPS1-ds1, 0.173/0.230 for DPS1-ds2, and 0.177/0.253 for DPS1-ds3 (table 6.2).

Dark Progression Series (DPS)	Exposure time	Delay between structures	Number of integrated files	Dose (kGy)	Resolution cut-off (Å)
DPS1-ds1	10 ms	~90 seconds	2803	48.6	2.00
DPS1-ds2			2398	97.2	2.10
DPS1-ds3			1511	145.8	2.30

Table 6.1. DPS1 dataset collection and data processing parameters

Dataset	DPS1-ds1	DPS1-ds2	DPS1-ds3
Data Collection			
Cumulative dose (kGy)	48.6	97.2	145.8
Number of integrated frames	2803	2398	1511
Exposure time (milliseconds)	10	10	10
Time delay between MSS (seconds)	n/a	~90	~90
Data processing			
Space Group	P21		
Cell dimensions (Å, deg)	a = 72.79 (0.44) b = 68.35 (0.27) c = 74.25 (0.52) β = 105.55 (0.35)	a = 72.76 (0.48) b = 68.34 (0.30) c = 74.31 (0.58) β = 105.60 (0.40)	a = 72.75 (0.55) b = 68.34 (0.35) c = 74.41 (0.66) β = 105.62 (0.50)
Resolution (Å)	44.48-2.00 (2.03-2.00)	44.48-2.10 (2.14-2.10)	44.49-2.30 (2.34-2.30)
Rsplit (%)	22.90 (56.65)	23.69 (60.14)	25.15 (64.26)
CC1/2 (%)	94.76 (59.81)	94.43 (60.90)	93.78 (58.15)
N_obs	25.04 (8.28)	22.67 (7.59)	19.61 (6.83)
Completeness (%)	100 (99.7)	99.9 (99.2)	99.9 (98.9)
Refinement			
Number of reflections	47602	41168	31394
Rwork/Rfree	0.171 / 0.228	0.173 / 0.230	0.177 / 0.253
RMSD bond lengths (Å)	0.007	0.007	0.008
RMSD bond angles (deg)	0.872	0.886	0.921
Ramachandran plot			
Most favoured (%)	98.75	97.92	97.49
Allowed (%)	1.22	2.16	2.50
Diffraction precision index (DPI)			
W1 DPI error (Å)	0.163	0.177	0.386
Fe DPI error (Å)	0.149	0.178	0.347
W1-Fe bond length error (Å)	0.221	0.251	0.519

Table 6.2. Data collection, processing, refinement, and DPI statistics for DPS1-ds1-3

6.3.2. DtpAa DPS1 structures

As seen in previous chapters and DtpAa structures, the active site in monomer A contains a six-coordinate heme group with a proton exchange network formed by waters supported by electron density (fig 6.3a, b); His326 acts as the proximal ligand with the bond length of Fe – N 2.28 Å (fig. 6.3c), while electron density reveals the distal heme coordination site occupied by a water (W1). The bond length of this water to the heme Fe is 2.48 Å. A second water, W2, interacts with the conserved DyP family residue Asp239 at a distance of 2.49 Å.

As in chapter 5, the focus of this experiment is assessing the movement of W1 away from the Fe in the heme of monomer A as a function of accumulated dose. By using a longer time delay between datasets than typically used in MSS experiments (seconds rather than milliseconds [chapter 4,5]) however, it should be possible to infer whether the movement of W1 is exclusively a function of accumulated dose, or a function of both absorbed dose and elapsed time (see chapter 5, section 4). After a time delay between dose 1 and dose 2 of ~90 seconds, the DPS1-ds2 bond length was not observed to increase; the Fe-O bond length for DPS1-ds2 is shorter, at 2.42 ± 0.25 Å. After a further ~90 second time delay, the bond returns to a length similar to that of the DPS1-ds1 structure, at 2.49 ± 0.52 Å away from the heme Fe (table 6.3, figure 6.4, omit maps fig. 6.5).

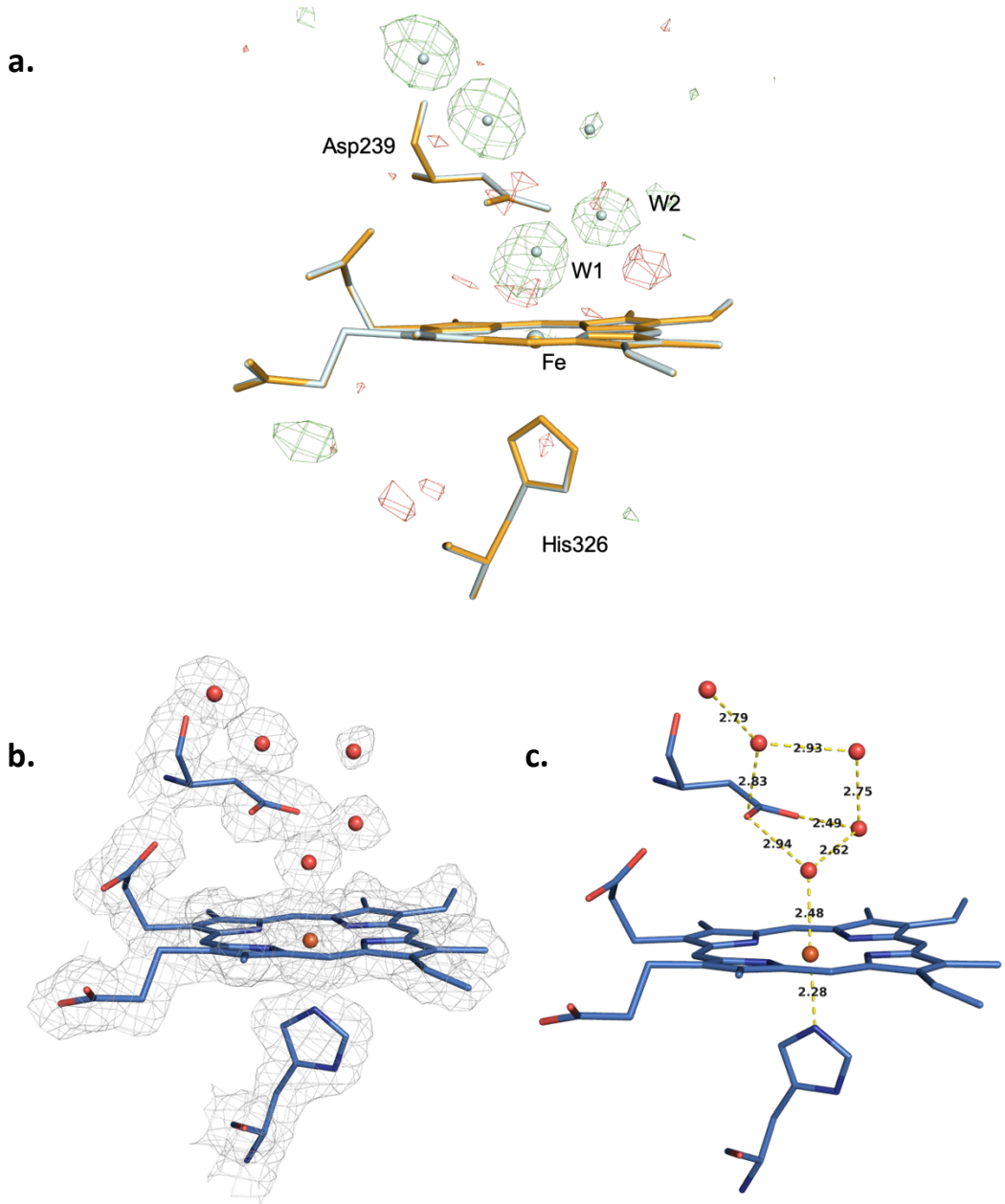


Figure 6.3. DtpAa DPS1-ds1 active site water network

(a.) $F_o - F_c$ electron density omit map depicts clear evidence of positive electron density for a water network, contoured at 3σ. The structure used for omit refinement is shown in light orange, with DPS1-ds1 superimposed in pale blue. (b.) $2F_o - F_c$ electron density map for the chain of ordered waters found at the active site of monomer A. Electron density of the 2 Å resolution MSS1-ds1 DtpAa structure is contoured at 1σ. Electron density map shows clearly resolved waters and their interaction with Asp239. (c.) Bond lengths for the water network, including the heme coordinating proximal His326 residue and Asp239. The His326-Fe distance is 2.28 Å, while the Fe-O distance is 2.48 Å. Interaction of W2 with Asp239 can be seen, with a bond distance of 2.49 Å.

Structure	DPS1-1	DPS1-2	DPS1-3
Dose (kGy)	48.6	97.2	145.8
Resolution (Å)	2.00	2.10	2.30
Fe-O (Å)	2.48 ± 0.221	2.42 ± 0.251	2.49 ± 0.519

Table 6.3. Comparison of Fe-O bond lengths in DPS1 datasets

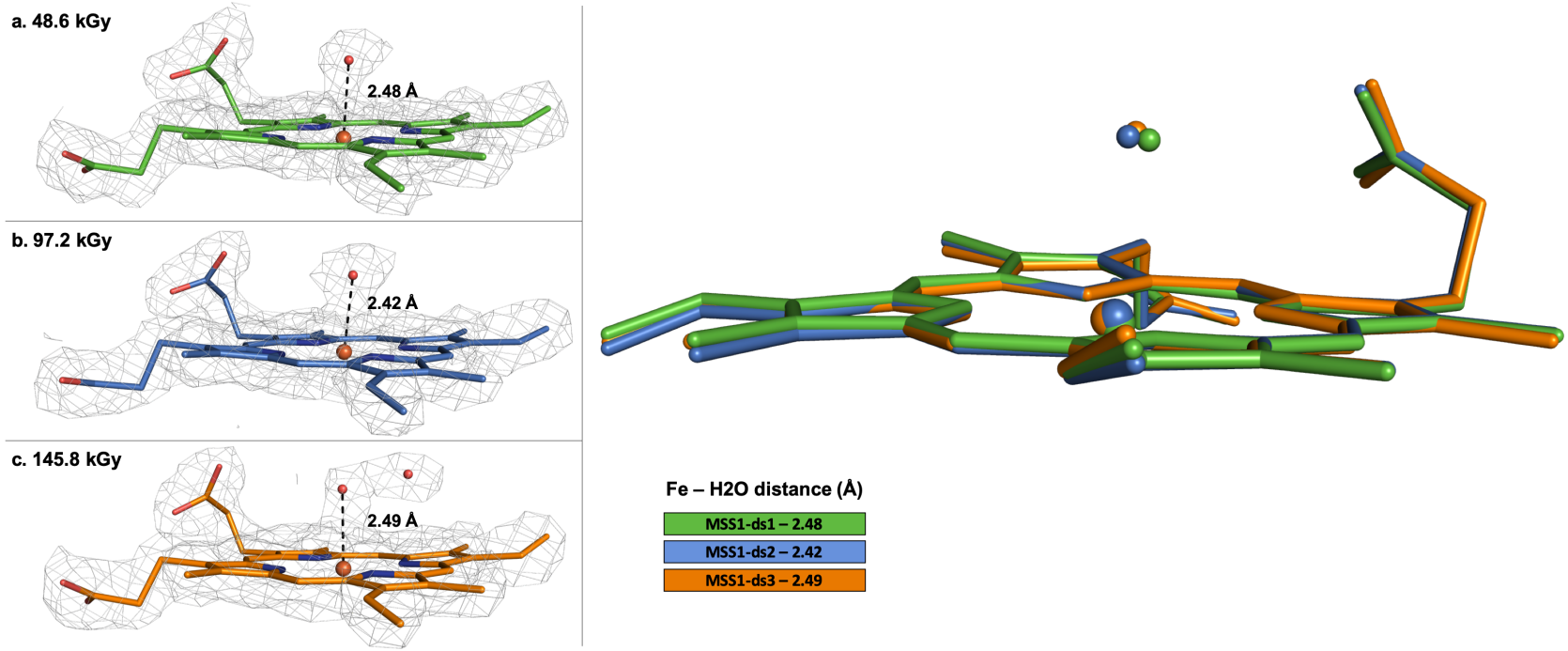


Figure 6.4. 90 second time delay Fe-O bond length as a function of accumulated dose.

$2F_o$-F_c electron-density maps contoured at 1σ for the heme environment of (a.) the initial 10 ms exposure 48.6 kGy DPS1-ds1 structure (green), (b.) the 97.2 kGy DPS1-ds2 structure (blue), and (c.) the 145.8 kGy DPS1-ds3 structure (orange). (d.) Superposition of DPS1-ds1, DPS1-ds2, and DPS1-ds3 90 second time delay structures reveals little migration of the proximal W1 water molecule away from the heme Fe with accumulated dose.

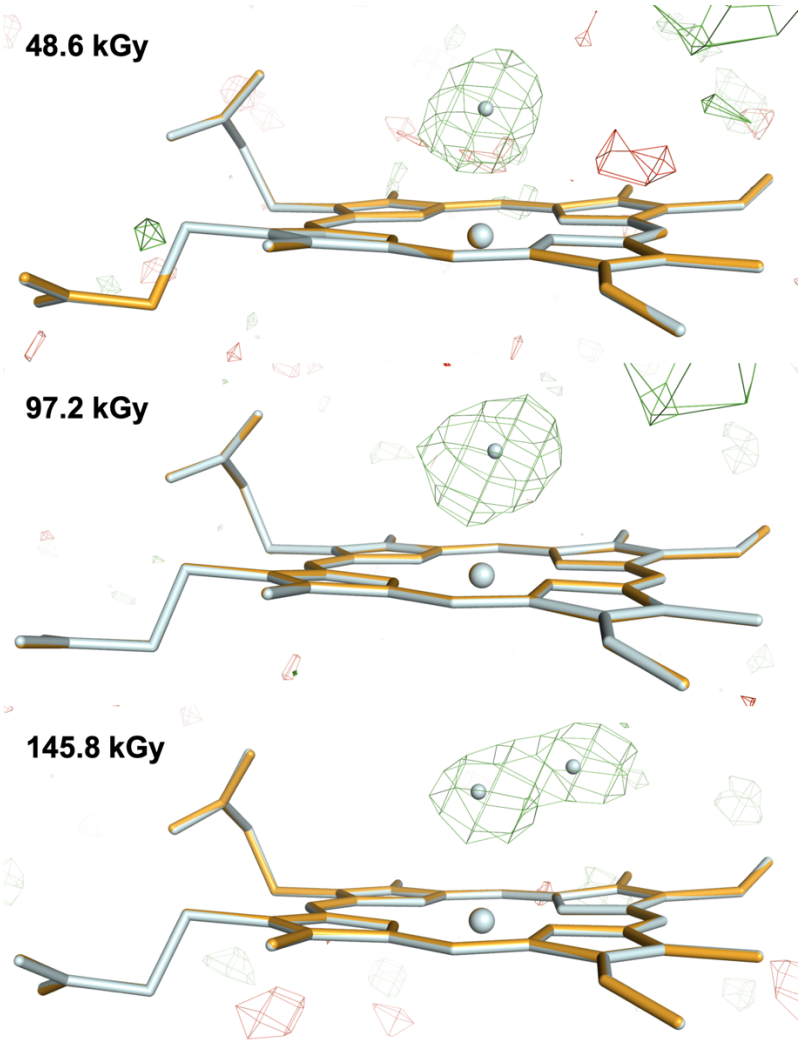


Figure 6.5. DPS1 omit maps.

F_o-F_c electron-density omit maps contoured at 3σ for the heme environment for DPS1. The structures used for omit refinement are shown in light orange, with structures utilised for the elucidation of Fe-O bond length superimposed in pale blue. Clear positive electron density is evident for a single water at dose points of 48.6 and 97.2 kGy, while a second water is evident within 145.8 kGy of X-ray exposure.

6.3.3. DPS2 SSX data collection, processing, and refinement

DPS2 comprised 10 sequential 10ms exposures, with a ~120 second delay between exposures 5 and 6 used in order to investigate whether the Fe-W1 bond length recovers over time when a delay separates an MSS method identical to that of chapter 5.3.4. The full data collection for one chip took < 1 hour, with a second chip collected in order to improve the dataset redundancy. Diffraction weighted dose was estimated to be 38.6 kGy per dataset (see appendices for *RADDOSE-3D* input parameters and output file). Data were again indexed and integrated using *dials.stills_process* on the Diamond computing cluster. For this dose series, data were collected from 2 chips, with 13178 indexed images from a possible 51200 for dose 1, equating to a hit rate of 25.7%, with data binned into dose dependent datasets using the *get-int-file.py* python script (appendices section 3.7.5).

As before, each dataset was initially merged to 1.6 Å in order to assess relevant data quality metrics (Rsplit, CC1/2, N_obs, and completeness), to determine a distinct resolution cut-off for each dose point. DPS2-ds1, ds2, and ds3 presented good merging statistics at 1.6 Å, though subsequent datasets were cut and merged at resolutions with 'good' corresponding merging statistics (table 6.4, 6.5). AutoBuild was again utilised for initial structure building. Rounds of refinement were performed using phenix.refine and COOT in order to accurately place W1 and the surrounding water network in the DPS2 datasets. Structures were refined to R_{work}/R_{free} values found in table 6.5.

Dark Progression Series (DPS)	Exposure time	Delay between structures	Number of integrated files	Dose (kGy)	Resolution cut-off (Å)
DPS2-ds1	10 ms	10 ms	13178	38.6	1.60
DPS2-ds2			12768	77.2	1.60
DPS2-ds3			12260	115.8	1.60
DPS2-ds4			11681	154.4	1.70
DPS2-ds5			10752	193	1.70
DPS2-ds6		10 ms*	6753	231.6	1.85
DPS2-ds7			6065	270.2	1.90
DPS2-ds8			5501	308.8	2.00
DPS2-ds9			5163	347.4	2.05
DPS2-ds10			4886	386	2.10
* DPS2-ds6-10 datasets collected ~120 seconds after DPS2-ds1-5					

Table 6.4. DPS2 dataset collection and data processing parameters

Dataset	DPS2-ds1	DPS2-ds2	DPS2-ds3	DPS2-ds4	DPS2-ds5	DPS2-ds6	DPS2-ds7	DPS2-ds8	DPS2-ds9	DPS2-ds10
Data Collection										
Cumulative dose (kGy)	38.6	77.2	115.8	154.4	193	231.6	270.2	308.8	347.4	386
Number of integrated frames	13178	12768	12260	11681	10752	6753	6065	5501	5163	4886
Exposure time (milliseconds)	10									
Time MSS collected (seconds)	0.01	0.02	0.03	0.04	0.05	120.01	120.02	120.03	120.04	120.05
Data processing										
Space Group	P21									
Cell dimensions / Å/deg.	a = 73.03 (0.33) b = 68.39 (0.22) c = 75.14 (0.33) β = 105.66 (0.32)	a = 73.04 (0.34) b = 68.38 (0.23) c = 75.14 (0.34) β = 105.71 (0.32)	a = 73.06 (0.37) b = 68.38 (0.25) c = 75.17 (0.37) β = 105.76 (0.35)	a = 73.09 (0.41) b = 68.37 (0.28) c = 75.22 (0.42) β = 105.81 (0.40)	a = 73.12 (0.44) b = 68.37 (0.30) c = 75.29 (0.46) β = 105.87 (0.44)	a = 73.14 (0.53) b = 68.38 (0.32) c = 75.28 (0.53) β = 105.82 (0.51)	a = 73.14 (0.56) b = 68.39 (0.35) c = 75.29 (0.54) β = 105.85 (0.53)	a = 73.16 (0.58) b = 68.39 (0.37) c = 75.31 (0.56) β = 105.87 (0.56)	a = 73.18 (0.59) b = 68.38 (0.36) c = 75.31 (0.58) β = 105.86 (0.54)	a = 73.18 (0.59) b = 68.38 (0.35) c = 75.32 (0.58) β = 105.87 (0.56)
Resolution / Å	44.76-1.60 (1.63-1.60)	44.75-1.60 (1.63-1.60)	44.76-1.60 (1.63-1.60)	44.76-1.70 (1.73-1.70)	44.76-1.70 (1.73-1.70)	44.76-1.85 (1.88-1.85)	44.77-1.90 (1.93-1.90)	44.78-2.00 (2.03-2.00)	44.78-2.05 (2.09-2.05)	44.78-2.10 (2.14-2.10)
Rsplit (%)	17.28 (63.23)	17.15 (68.45)	17.07 (72.38)	16.56 (63.04)	16.71 (70.40)	20.60 (77.40)	21.88 (80.33)	22.51 (71.13)	23.01 (76.50)	22.95 (70.51)
CC1/2 (%)	96.59 (57.21)	96.89 (56.00)	97.10 (52.74)	97.18 (59.66)	97.68 (58.15)	97.07 (56.64)	96.63 (55.04)	96.34 (62.10)	96.30 (55.79)	96.69 (57.72)
N_obs	49.56 (7.69)	47.53 (6.67)	45.05 (5.08)	48.74 (9.73)	44.16 (7.33)	35.42 (8.07)	33.90 (6.91)	33.21 (8.14)	32.40 (8.66)	31.29 (8.93)
Completeness (%)	99.9 (99.1)	99.9 (98.6)	99.8 (97.1)	99.9 (99.1)	99.8 (96.9)	99.7 (95.8)	99.6 (94.9)	99.7 (96.3)	99.7 (96.1)	99.7 (97.3)
Refinement										
Number of reflections	93925	93918	93849	78470	78417	60923	56213	48354	44932	41798
Rwork/Rfree	0.169 / 0.201	0.171 / 0.196	0.175 / 0.198	0.171 / 0.214	0.179 / 0.218	0.1836 / 0.224	0.186 / 0.234	0.186 / 0.234	0.187 / 0.249	0.189 / 0.251
RMSD bond lengths (Å)	0.006	0.006	0.006	0.006	0.006	0.007	0.007	0.008	0.007	0.007
RMSD bond angles (deg)	0.853	0.840	0.877	0.846	0.852	0.850	0.862	0.887	0.879	0.892
Ramachandran plot										
Most favoured (%)	98.33	98.06	98.47	98.47	98.47	98.47	97.92	98.19	97.08	97.36
Allowed (%)	1.54	1.92	1.48	1.47	1.47	1.47	2.04	1.79	2.89	2.56
Diffraction precision index (DPI)										
W1 DPI error (Å)	0.074	0.078	0.085	0.095	0.104	0.130	0.144	0.168	0.214	0.205
Fe DPI error (Å)	0.061	0.063	0.067	0.079	0.084	0.118	0.135	0.158	0.191	0.226
W1-Fe bond length error (Å)	0.096	0.100	0.108	0.124	0.134	0.176	0.197	0.231	0.287	0.305

Table 6.5. Data collection, processing, refinement, and DPI statistics for DPS2-ds1-10

6.3.4. DtpAa DPS2 structures

The Fe-O bond length for DPS2-ds1 is similar to that of DPS1-ds1, at 2.46 $\pm$ 0.096 Å, however, when collecting a series of 5 sequential dose points a trend becomes clear, with the Fe-O bond length increasing from 2.46 $\pm$ 0.096 to 2.82 $\pm$ 0.134 Å (table 6.6, fig. 6.6). After a time delay between dose point 5 and dose point 6 of ~120 seconds, the bond length recovers to 2.48 $\pm$ 0.176 Å. The bond length is then observed to again lengthen to 2.64 $\pm$ 0.287 Å by dose 9, similar to the increase in bond length seen from doses 1 to 5. Interestingly DPS2-ds10 Fe-O bond length is shorter than that of DPS2-ds9, however the position of W1 may be influenced by W2 in the DtpAa active site water network (fig. 6.7). Omit maps for DPS2 are shown in figure 6.8.

Structure	DPS2-ds1	DPS2-ds2	DPS2-ds3	DPS2-ds4	DPS2-ds5	DPS2-ds6	DPS2-ds7	DPS2-ds8	DPS2-ds9	DPS2-ds10
Dose (kGy)	38.6	77.2	115.8	154.4	193	231.6	270.2	308.8	347.4	386
Resolution (Å)	1.60	1.60	1.60	1.70	1.70	1.85	1.90	2.00	2.05	2.10
Fe-O (Å)	2.46 ± 0.096	2.56 ± 0.100	2.54 ± 0.108	2.68 ± 0.124	2.82 ± 0.134	2.48 ± 0.176	2.49 ± 0.197	2.51 ± 0.231	2.64 ± 0.287	2.55 ± 0.305

Table 6.6. Comparison of Fe-O bond lengths in DPS2 datasets

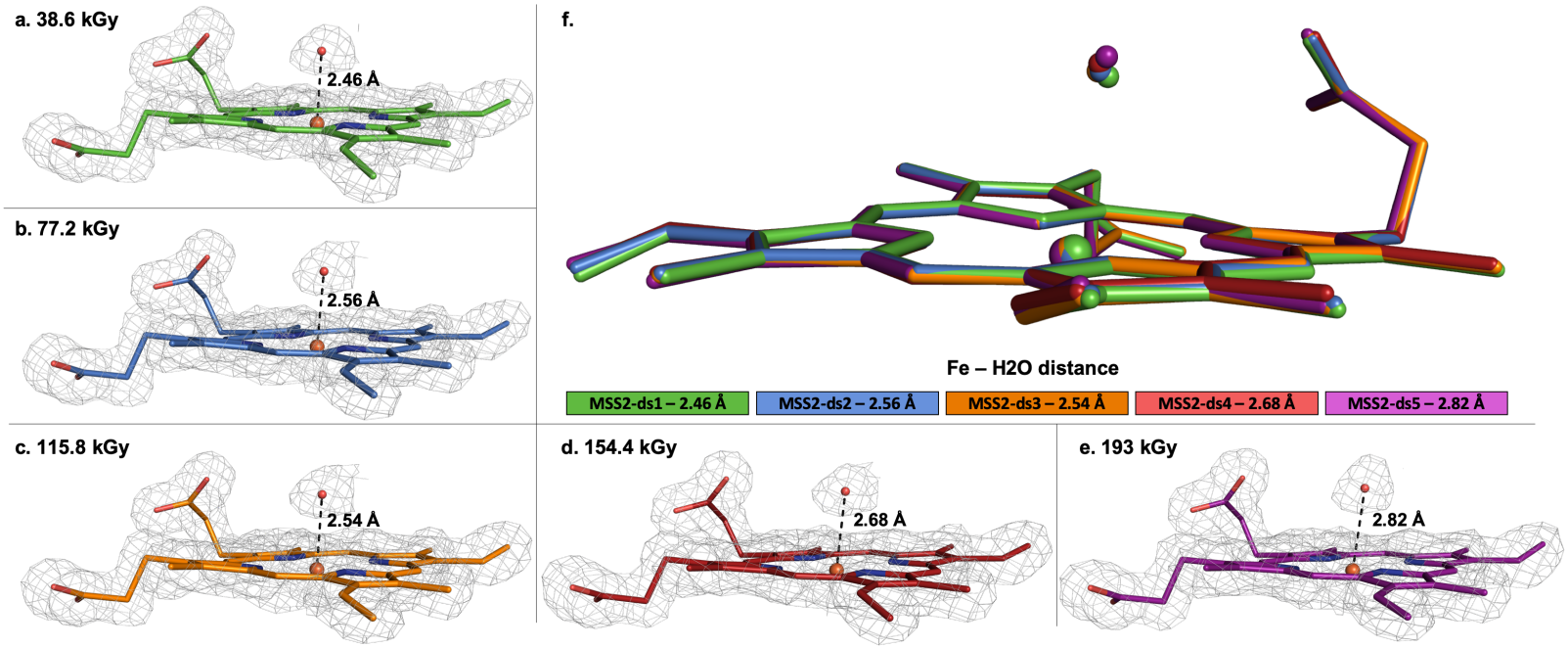


Figure 6.6. Datasets DPS2-ds1-5 Fe-O bond length as a function of accumulated dose.

$2F_o$-F_c electron-density maps contoured at 1σ for the heme environment of (a. – e.) the initial MSS2-ds1-5 10 ms exposure structures (38.6, 77.2, 115.8, 154.4, 193 kGy; green, blue, orange, red, purple), with their associated Fe-O bond lengths shown. (f.) Superposition of MSS2-ds1-5 reveals a progressive migration of the proximal W1 water molecule away from the heme Fe with accumulated dose, with little change in the conformation of the heme.

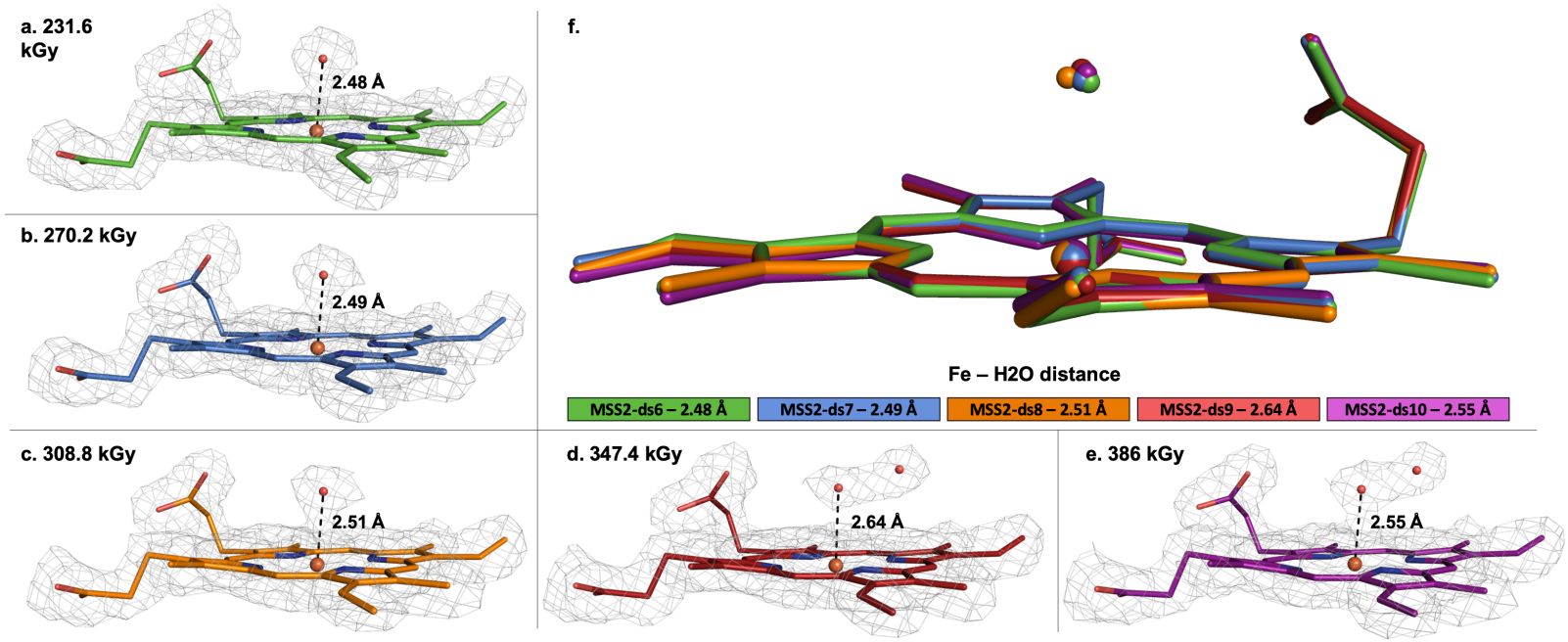


Figure 6.7. Datasets DPS2-ds6-10 Fe-O bond length as a function of accumulated dose

$2F_o$-F_c electron-density maps contoured at 1σ for the heme environment of (a. – e.) MSS2-ds6-10 10 ms exposure structures (231.6, 270.2, 308.8, 347.4, 386 kGy; green, blue, orange, red, purple), collected after a ~120 second delay after the collection of MSS2-ds5, with associated Fe-O bond lengths shown. (f.) Superposition of MSS2-ds6-10 again reveals a progressive migration of the proximal W1 water molecule away from the heme Fe with accumulated dose. The MSS2-ds10 Fe-O bond length is seen to be shorter than the bond in the previous dose dependent structure, though the water distance relative to the heme Fe may be influenced by the water molecule (W2) in close proximity to W1.

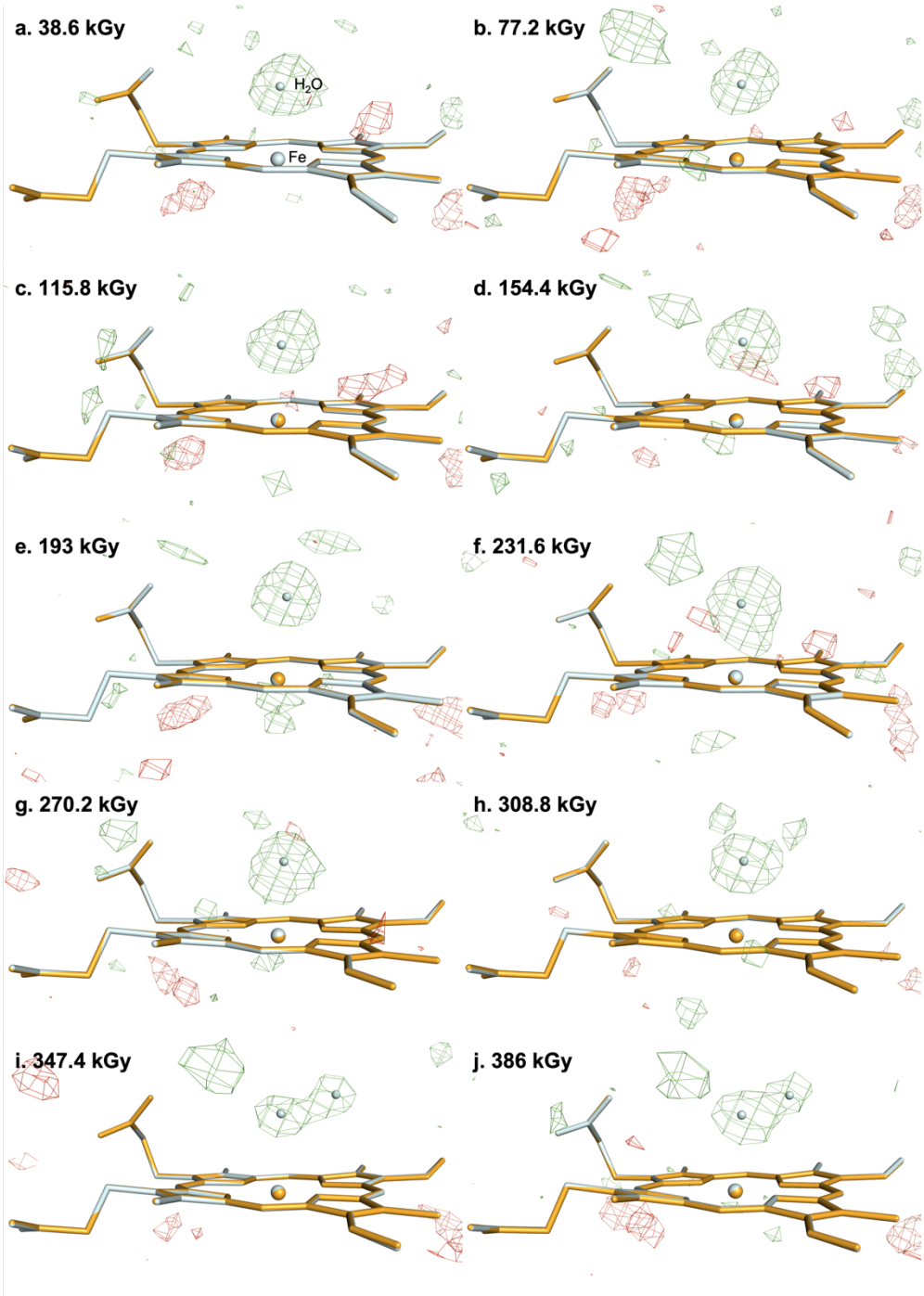


Figure 6.8. DPS2 omit maps.

F_o-F_c electron-density omit maps contoured at 3σ for the heme environment for DPS2. The structures used for omit refinement are shown in light orange, with structures utilised for the elucidation of Fe-O bond length superimposed in pale blue. Positive features for bound water electron density are evident throughout the dose series. Datasets f. – j. were collected ~120 seconds after those shown in figures a. – e.

6.4. Discussion

By performing an MSS series sub-divided by larger time intervals, we have been able to demonstrate a new data collection technique that builds on MSS, a new serial method we have coined 'dark progression series' (DPS). This DPS experiment has been successfully used to add context to the results presented in chapter 5, section 3: by using long time intervals of 90+ seconds between multiple datasets, it can be shown that the Fe-O bond length returns to a distance similar to that of only a single 10 ms exposure (fig. 6.4). 10 ms exposures were used for MSS1 and MSS2 series in section 5.3.4, corresponding to bond lengths of 2.48 and 2.50 Å in MSS1-ds1 and MSS2-ds1 (chapter 5 section 3.4), compared to 2.48, 2.42, and 2.49 Å in DPS-ds1-3 from 10 ms exposure structures that have ~90 second intervals between their collection. Additionally, by assessing bond lengths of structures from DPS2 (consisting of two 5-dose MSS series separated by ~120 seconds), the Fe-O distance returns to 2.48 Å in DPS2-ds6 from a length of 2.82 Å in DPS2-ds5, before it again lengthens to 2.55 Å by dose 10 (figure 6.9). DPS2 structures clearly show that there is a recovery event happening at the heme, where the coordination of the heme is returning to a 'low damage' state, referred to from this point onward as 'sawtooth' recovery due to the shape of the plot shown in figure 6.9.

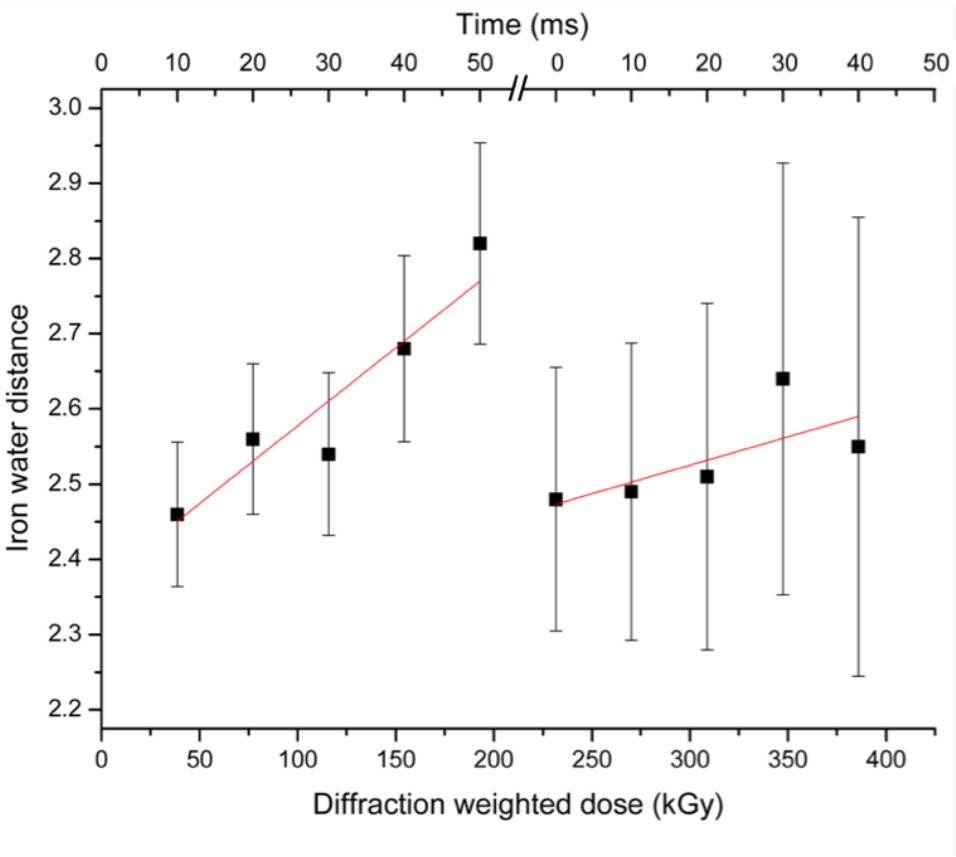


Figure 6.9. Plot of DPS2 Fe-O distance as a function of diffraction weighted dose and time

Both DPS2-ds1-5 and DPS2-ds6-10 are plotted in black, with linear functions fitted to the individual MSS series. The separation in the time axis signifies a delay period of ~120 seconds. Extrapolating the linear function of DPS2-ds1-5 to zero-dose gives a bond length of 2.371 ± 0.107, while the same for the function fitted to doses 6-10 is 2.298 ± 0.566. The slope of the linear function fitted to DPS-ds1-5 is 0.00207, while that of DPS-ds6-10 is 0.00076. Error bars in this plot were produced using the Diffraction Precision Index, signifying estimated standard uncertainty in bond length.

When comparing the first five dose points of the sawtooth dataset with those of MSS1 and MSS2 from chapter 5, the iron water distances for DPS2-ds1-5 fall within error of the bond lengths of the first five dose points of MSS2 (fig. 6.10). Both MSS data series were collected at dose intervals very similar to that of DPS2 (MSS2 = 39.2 kGy per dataset; DPS2 = 38.6 kGy per dataset). Furthermore, the *Y* intercept for the linear function of all MSS1 and MSS2 dose points plotted in chapter 5 (fig. 5.16) is strikingly similar to that of DPS2-ds1-5, with the *Y* intercept for points plotted in figure 6.10 2.369 $\pm$ 0.050 Å almost identical to the DPS2-ds1-5 *Y* intercept of 2.371 $\pm$ 0.107. The similar intercept value that presented in chapter 5 strengthens the effectiveness of our method of zero-dose extrapolation being used to estimate damage free data, as data for chapters 5 and 6 were collected and processed completely separately from one another, validating MSS as both accurate and reliably reproducible. The *Y* intercept for DPS2-ds6-10 is 2.298 $\pm$ 0.566. Interestingly the error values for DPS2-ds1-5 are similar to those of MSS1 and MSS2, though those for DPS2-ds6-10 are higher after the time delay. The poor error values for DPS1 can be attributed to low resolution and N_obs due to the low numbers of diffraction patterns for those datasets (~2000 vs ~13000). Again, where there is a significant jump in DPI between DPS2-ds5 and DPS2-ds6, there is also a large reduction in diffraction patterns (~10700 vs ~6700), N_obs (44 vs 35), and resolution (1.70 vs 1.85 Å).

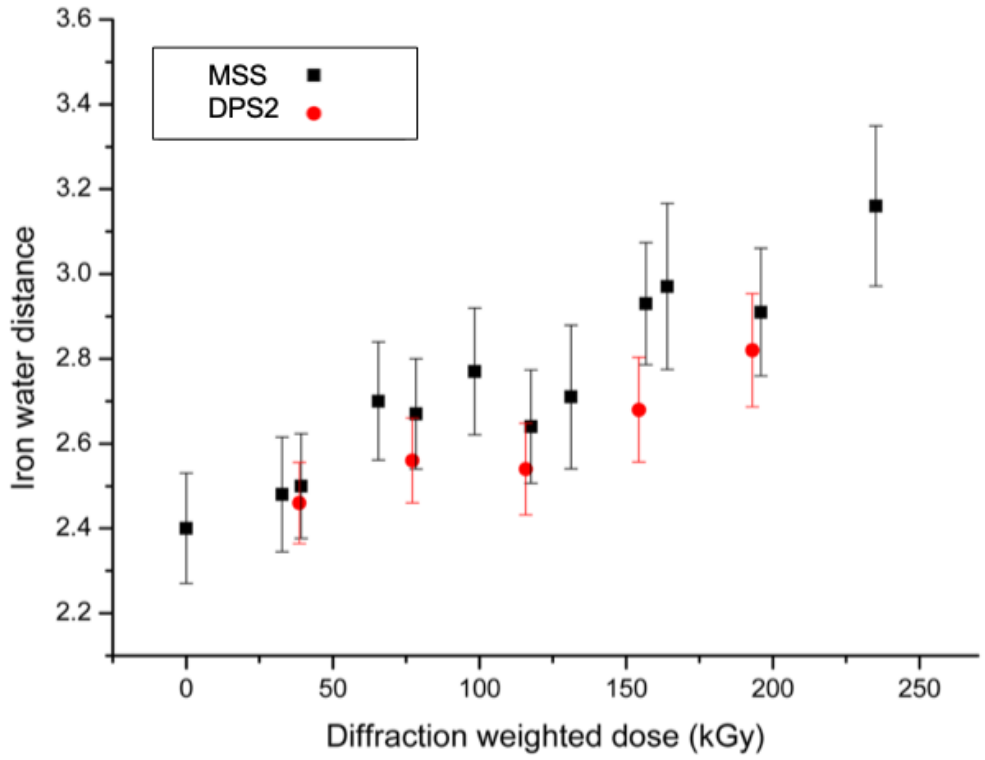


Figure 6.10. Plot of Fe-O distance as a function of X-ray dose from MSS1, MSS2, and DPS2

MSS1 and MSS2 bond distances (black), and DPS2-ds1-5 (red) can be seen to overlap when plot on the same axis. The DPS2 plot points fall within error of corresponding MSS plot points at similar dose.

It is however still possible to show that dark progression of radiation damage is apparent: by plotting normalised diffracting power (I_n / I_1) for DPS2 datasets processed to the same resolution as DPS2-ds1, 1.6 Å, there is a large jump in loss of overall diffraction intensity between DPS2-ds5 and DPS2-ds6 (53.1 vs 39.8) (fig. 6.11a). Unit cell volume increases with accumulated dose throughout the DPS2 dataset as expected (Ravelli & McSweeney, 2000), though not between DPS2-ds5 and DPS2-ds6 where the X-rays were turned off (fig. 6.11b). Dark progression of radiation damage may have an effect on intensities by increasing the non-isomorphism within the crystal via cascades of photoelectrons released during secondary radiation damage events, though as unit cell increase is not affected by dark progression in this instance, this may indicate that it is more due to thermal expansion of the crystal lattice when exposed to X-rays than an increase in non-isomorphism.

There are a number of interesting experiments that could build on the data collected in this preliminary chapter. Establishing whether the redox state of DtpAa recovers after a dark period would be a logical next experiment, establishing whether the movement of the water away from the heme Fe is directly related to its reduction by the X-ray beam (chapter 5, figures 5.13, 5.14). This could be achieved by using an MSOX (Horrell 2016; 2019) or MSS method on DtpAa crystals coupled with online UV-vis microspectrophotometry, answering whether the bond distance 'recovery' is a result of reoxidation at the heme. Additionally, using collecting DPS datasets with progressively shorter time delays would allow quantification of the shortest amount of time needed for the bond length to recover. This could again be coupled with an MSOX method coupled with UV-vis online microspectrophotometry. Using an MSOX

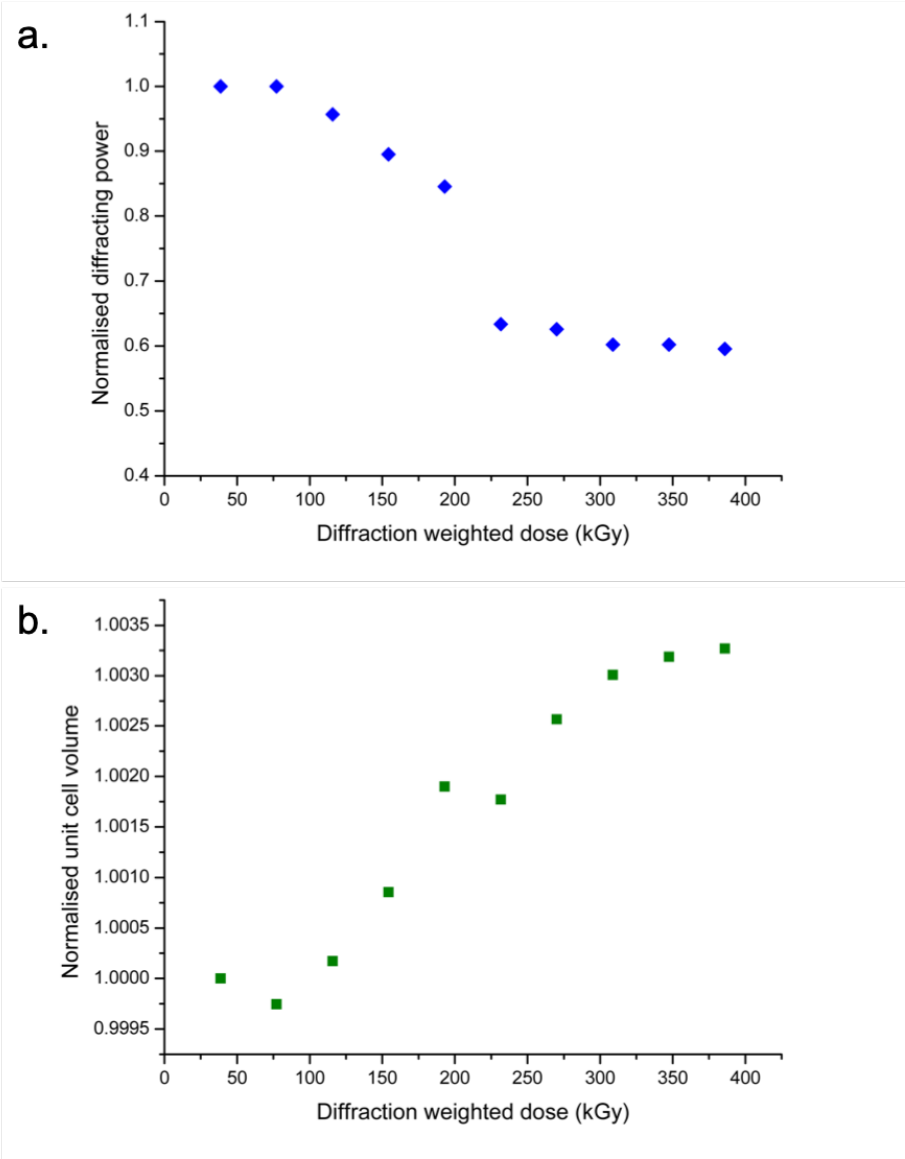


Figure 6.11. Normalised global radiation damage indicators vs absorbed dose for DPS2 datasets

(a.) Decrease in normalised diffracting power can be shown throughout DPS2 datasets when processed to the same 1.6 Å resolution cut-off. The distinct drop in diffracting power between points 5 and 6, collected ~120 seconds apart, is greater than that seen between any dose points separated by 10 ms intervals, indicative of the dark progression of radiation damage. Normalised unit cell increases across DPS2, though points 5 and 6 only vary slightly, indicating that unit cell expansion may not be driven during dark progression periods. The increase in unit cell volume plateaus toward the end of the data collection, with an overall expansion of 3.3 %.

method typically requires very large (> 400 µm) crystals to be grown. Doing this same DPS experiment on large crystals, and an MSOX experiment on both large crystals and microcrystals, would allow insight into the influence of crystal volume on the dark progression of radiation damage, by identifying whether bond length increase happens at the same rate in crystals of different size. Importantly, in order to improve the ease of performing this experiment is the automation of data collection. Similar to how scripts and a GUI were developed to run standard chip and MSS experiments, achieving this for the DPS method would allow exact time delays to be established between DPS collections. The data collected in this thesis relied on manually resetting and running the data collections via the GUI, leading to slightly different time delays between DPS1 and DPS2, as data series were collected from either a whole, or half of a column.

6.5. Appendices

6.5.1. DPS1 RADDOSE

Input Data

```
Crystal

Type Cuboid
Dimensions 20 20 20
PixelsPerMicron 0.5
AbsCoefCalc  RD3D
UnitCell  72.6 68.1 74.8 90 105.5 90
NumMonomers  4
NumResidues  360
ProteinHeavyAtoms Fe 1 S 5
#SolventHeavyConc P 425
SolventFraction 0.45

Beam

Type Gaussian
Flux 3.97e12
FWHM 8 8
Energy 12.8
Collimation Rectangular 24 24

Wedge 0 0
ExposureTime 0.01
```

Output Data

Cuboid (Polyhedron) crystal of size [20, 20, 20] um [x, y, z] at a resolution of 2.00 microns per voxel edge.
Simple DDM.
Gaussian beam, 24.0x24.0 um with 8.00 by 8.00 FWHM (x by y) and 4.0e+12 photons per second at 12.80 keV.
Wedge 1:
Collecting data for a total of 0.0s from phi = 0.0 to 0.0 deg.

Crystal coefficients calculated with RADDOSE-3D.
Photelectric Coefficient: 2.02e-04 /um.
Inelastic Coefficient: 2.00e-05 /um.
Elastic Coefficient: 1.86e-05 /um.
Attenuation Coefficient: 2.40e-04 /um.
Density: 1.20 g/ml.

Average Diffraction Weighted Dose : 0.048580 MGy
Elastic Yield : 1.46e+07 photons
Diffraction Efficiency (Elastic Yield/DWD): 3.00e+08 photons/MGy
Average Dose (Whole Crystal) : 0.033979 MGy
Average Dose (Exposed Region) : 0.033979 MGy
Max Dose : 0.189382 MGy
Average Dose (95.0 % of total absorbed energy threshold (0.01 MGy)): 0.061839 MGy
Dose Contrast (Max/Threshold Av.) : 3.06
Used Volume : 100.0%
Absorbed Energy (this Wedge) : 3.33e-07 J.
Dose Inefficiency (Max Dose/mJ Absorbed) : 568.7 1/g
Dose Inefficiency PE (Max Dose/mJ Deposited): 581.1 1/g
Final Dose Histogram:
Bin 1, 0.0 to 0.1 MGy: 91.0 %
Bin 2, 0.1 to 3.4 MGy: 9.0 %
Bin 3, 3.4 to 6.7 MGy: 0.0 %
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 %

6.5.2. DPS2 RADDOSE

Input Data

```
Crystal

Type Cuboid
Dimensions 20 20 20
PixelsPerMicron 0.5
AbsCoefCalc  RD3D
UnitCell  73.2 68.4 75.3 90 105.8 90
NumMonomers  4
NumResidues  360
ProteinHeavyAtoms Fe 1 S 5
#SolventHeavyConc P 425
SolventFraction 0.45

Beam

Type Gaussian
Flux 3.94e12
FWHM 9 9
Energy 12.8
Collimation Rectangular 27 27

Wedge 0 0
       # Start and End rotational angle of the crystal with Start
< End

ExposureTime 0.01
```

Output Data

Cuboid (Polyhedron) crystal of size [20, 20, 20] um [x, y, z] at a resolution of 2.00 microns per voxel edge.
Simple DDM.
Gaussian beam, 27.0x27.0 um with 9.00 by 9.00 FWHM (x by y) and 3.9e+12 photons per second at 12.80 keV.
Wedge 1:
Collecting data for a total of 0.0s from phi = 0.0 to 0.0 deg.

Crystal coefficients calculated with RADDOSE-3D.
Photelectric Coefficient: 2.00e-04 /um.
Inelastic Coefficient: 1.98e-05 /um.
Elastic Coefficient: 1.84e-05 /um.
Attenuation Coefficient: 2.38e-04 /um.
Density: 1.19 g/ml.

Average Diffraction Weighted Dose : 0.038626 MGy
Elastic Yield : 1.41e+07 photons
Diffraction Efficiency (Elastic Yield/DWD): 3.66e+08 photons/MGy
Average Dose (Whole Crystal) : 0.033355 MGy
Average Dose (Exposed Region) : 0.033355 MGy
Max Dose : 0.148761 MGy
Average Dose (95.0 % of total absorbed energy threshold (0.01 MGy)): 0.051946 MGy
Dose Contrast (Max/Threshold Av.) : 2.86
Used Volume : 100.0%
Absorbed Energy (this Wedge) : 3.23e-07 J.
Dose Inefficiency (Max Dose/mJ Absorbed) : 460.2 1/g
Dose Inefficiency PE (Max Dose/mJ Deposited): 470.2 1/g
Final Dose Histogram:
Bin 1, 0.0 to 0.1 MGy: 91.0 %
Bin 2, 0.1 to 3.4 MGy: 9.0 %
Bin 3, 3.4 to 6.7 MGy: 0.0 %
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 %

Chapter 7: Identification of protein-ligand complexes via SFX and SSX

7.1. Introduction

Advances in crystallographic screening and collections methods have led to the development of what we now know as 'high-throughput' crystallography. For example, sparse matrix screens coupled with laboratory robots make it possible to explore many different crystallisation parameters such as pH, precipitant concentration, additives etc. whilst exercising relatively minimal effort and skill on the part of the researcher. It is now therefore possible to screen more than two orders of magnitude more crystallisation conditions from the same amount of sample than was possible 25 years ago (McPherson & Gavira, 2014). This has allowed crystallography to be widely used as a high-throughput method for industrial drug fragment screening purposes, with many industrial laboratories incorporating fragment-based approaches from the early 2000s (Davies & Tickle, 2012).

Fragment-based drug discovery (FBDD) entails screening small-molecule libraries against a target protein to identify bioactive molecules, with biophysical methods such as X-ray crystallography utilised to identify weakly binding fragments (Patel, Bauman, & Arnold, 2014). Its popularity is due in part to X-ray crystallography remaining one of the most sensitive biophysical techniques within the practical constraints of a typical fragment-screening experiment, despite false negatives being a possibility due to ligand binding not being tolerated within the crystal environment (Collins et al., 2018; Davies & Tickle, 2012). X-ray crystallography has the overarching

advantage to being able to see ligand binding interactions, allowing structure-based drug design as an outcome from FBDD (Anderson, 2003). This has led to dedicated facilities at 3rd generation light sources, such as the XChem facility at Diamond Light Source, employing machine-assisted crystal harvesting, automated X-ray diffraction data collection and streamlined data processing (Collins et al., 2018). The XChem suite is the first of its kind to enable such an experiment, generating thousands of protein-fragment complexes for academic and industrial collaborators alike, making the technique accessible to the non-expert user (Delbart et al., 2018; Harding et al., 2017; Keedy et al., 2018; McIntyre et al., 2017).

However, FBDD is a technique typically performed at cryogenic temperatures at synchrotrons, introducing the risk of hiding structural ensembles in protein crystals by eliminating the packing defects needed for the modelling of functional motions (Fraser et al., 2011). Although the only current approach to avoid structure-altering radiation damage at room temperature is SFX, an advantage when accurately determining the active site of a protein-ligand complex, ligand binding studies at XFELs have received little attention. A limited number of studies have sought to address the challenge of obtaining damage-free room temperature crystal structures of protein-ligand complexes in a manner that is both time and sample efficient. Bublitz et al. examined the feasibility of SFX for structure determination of ligand complexes of P-type ATPase in membrane proteins at LCLS; although the resolution limits of SFX data were slightly lower than those collected at synchrotron sources (3 Å vs 2.5 Å, due potentially to microcrystal optimisation being on an experimentally shorter time scale), the identification of ligands and their binding sites could still be accomplished.

Further, even at low redundancy of 4.2 (outer shell) due to the difficulty in obtaining large volumes of membrane protein microcrystals, combined with a poor outer shell metrics such as an R_{split} of 538, $CC_{1/2}$ of 0.03, and signal to noise ratio ($I/\sigma(I)$) of 0.39, the identification of ligands could still be demonstrated (Bublitz et al., 2015). More recently, Naitow et al. demonstrated effective SSX protein-ligand complex solution at SACLA, using both oil- and water-based crystal carriers to present crystal samples successfully to the beam, solving the structure of apo and ligand bound thermolysin at 2.1 Å and 2.0 Å, respectively (Naitow et al., 2017). Both Bublitz et al. and Naitow et al. suffer from high sample consumption however, with respective crystal slurries depleted at rates of 13-30 µl min^{-1} and 0.48 µl min^{-1}, with the former utilising multiple 10 minute data collections to form datasets (data collection time not reported in the publication from Naitow et al., 2017) (Bublitz et al., 2015; Naitow et al., 2017).

Regrettably, high throughput ligand screening at XFELs has suffered as a consequence of high sample consumption typically used in an SFX experiment. Sample delivery techniques such as jetting are also not well suited to fast sample changes necessary for ligand binding experiments. Furthermore, XFEL beamtime is scarce, with fewer than 10 operational XFELs (typically with less than 4 beamlines per facility) available out of 50 light sources that are operational or under construction (as of February 2020). As mentioned in previous chapters, fixed target methods aim to combat this by using low volumes of crystal suspension on a solid support. This technique also permits fast data collections, with SSX data collected in a short amount of time (< 15 min) from single chip loaded with < 200 µl of crystal slurry enough to produce highly redundant datasets (see chapter 4 sections 3, 4). Limiting SSX crystal consumption on

chips has been explored previously by using acoustic waves to dispense crystals onto chips, known as acoustic drop ejection (ADE): Davy et al., 2019 compared the conventional pipette chip loading method presented in this thesis with chips loaded using ADE with a commercially available PolyPico pico-litre dispenser (http://www.polypico.com); it was possible to demonstrate an overall reduction in sample consumption coupled with an increase in diffraction hits per unit volume of dispensed crystal slurry (fig. 7.1a). Further, it was shown that crystals do not assume a preferred orientation with either loading method, with data collected from multiple protein crystal samples (fig. 7.1b) (Davy et al., 2019).

In this chapter, a time and sample efficient data collection and processing method is demonstrated. This provides an effective method of reducing the volume of diffraction data which has been used to explore how much data is truly needed to unambiguously identify a bound ligand in XFEL data. Moreover, it is possible to demonstrate unambiguous ligand discrimination from low redundancy crystal datasets using widely available automated ligand-finding methods. This is compared to SSX data subject to an identical method of data processing, in order to establish the potential difference between necessary data volumes of stills data for ligand identification at both XFELs and synchrotrons.

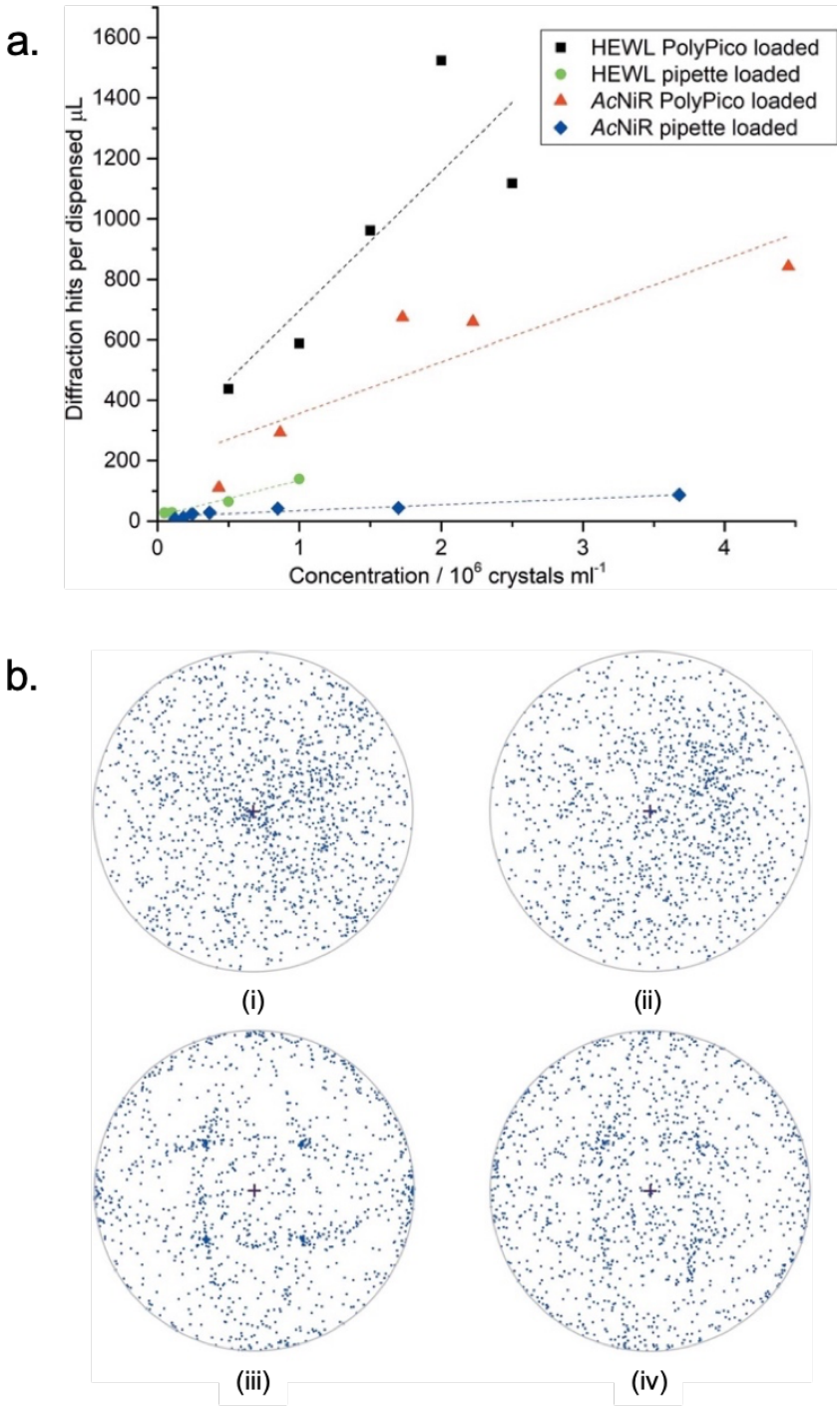


Figure 7.1. Effects of reducing sample consumption on diffraction hits and crystal orientation using acoustic drop ejection

(a.) ADE using a PolyPico results in a greater than fivefold increase in diffraction hits per unit volume of crystal slurry consumed at all concentrations for lysozyme crystals (HEWL). A tenfold increase in diffraction hits per unit volume is evident in AcNiR. (b.) Stereographic projections illustrating the crystal orientation of 1000 randomly selected crystals for each loading method

and crystal type: (i) pipette-loaded AcNiR, (ii) PolyPico-loaded AcNiR, (iii) pipette-loaded HEWL and (iv) PolyPico-loaded HEWL. The beam direction (z) is shown as the central '+' in each plot, with each spot representing the relative direction of the 001 hkl of each crystal. 12 o'clock on the circular projection represents a 90° rotation of the crystal around the x axis, whereas the point at 3 o'clock represents a 90° rotation around y. Despite the randomness of the spots indicating no preferred orientation is evident for the majority of the data used for the projection, indication of systematic orientation in pipette loaded HEWL (iii) is evident, though is likely to depend on sample density on chip, crystal size and morphology (reproduced with permission from Davy et al., 2019).

7.2 Methods

7.2.1. Protein expression, purification, and crystallisation

Recombinant AcNiR was expressed and purified as described in detail in section 3.1.1. Batch microcrystals were prepared as mentioned in section 4.2.1. Microcrystals at a concentration of 3.7 x 10^6 crystals ml^{-1} and sized between 5-15 µm were used for both SFX and SSX experimentation. Microcrystal batches at a volume of 150 µl were soaked in 100 mM sodium nitrite for approximately 20 minutes prior to loading on the chip.

7.2.2. Serial data collection

7.2.2.1. Sample loading

Data collected for this chapter utilised the fixed target 'chip' method used in chapters 4 and 5, and is described in detail in chapter 3. AcNiR microcrystals were loaded onto silicon chips at both the Spring-8 Ångstrom Free-Electron Laser (SACLA) and Diamond Light Source (DLS) using an identical method as explained in detail in section 3.2.2.

7.2.2.2. SACLA beamline instrumentation and parameters

Sample stages for the movement of chips through the X-ray beam were mounted on beamline BL2 EH3 at SACLA as described previously in detail in chapter 3.4. SFX data were measured at SACLA beamline BL2/EH3 with a photon energy of 10.0 keV, repeat rate 30 Hz and pulse length 10 fs. The beam, of 1.25 x 1.34 µm size and pulse energy of 289 µJ $pulse^{-1}$ was attenuated to 13% of full flux to minimise detector overloads. The crystal loaded chip was translated between X-ray pulses so that each aperture was exposed only once, with the translation between apertures occurring in the 30ms delay between the 30 Hz XFEL pulses, ensuring the chip was stationary whilst the

diffraction image was recorded. Measurement from all 25,600 positions on a chip was achieved in ~14 minutes. As with previous data collections, the stages and SACLA MPCCD detector (Yabashi, Tanaka, & Ishikawa, 2015) was sealed within a helium environment in order to minimise air scatter (see chapter 3.4.3.2), assisting in removing as much background as possible from diffraction images.

7.2.2.3. DLS beamline instrumentation, methods, and parameters

Data collected on beamline I24 at Diamond light source for this experiment used a beamsize of 8 x 8 µm (FWHM). All data were collected at 12.8 keV using a PILATUS3 6M detector with a crystal-to-detector distance of 310 mm. Beam flux was 3 x 10^{12} photons s^{-1}, measured immediately prior to the experiments using a silicon PIN diode (Owen et al., 2009), and attenuated tenfold for the described data collections (3 x 10^{11} photons s^{-1}). The diffraction weighted dose absorbed by each crystal was estimated using *RADDOSE-3D* (Paithankar & Garman, 2010).

7.2.2.4. Data processing, structure solution, and refinement

Initial hit finding for data measured during the SACLA experiment was carried out at the beamline using *CHEETAH* (Barty et al., 2014), with *CrystFEL* (White et al., 2012) utilised for peak-finding, integration, resolution of indexing ambiguity, and merging thereafter. SSX data collected at Diamond was also processed in the same way using *CrystFEL* in this chapter. This is in contrast to previous chapters in order to allow consistency between SSX and SFX data processing methods. Native and ligand bound structure data were assessed using typical data quality metrics such as $CC_{1/2}$ and R_{split} in order to define appropriate resolution limit cut-offs for merging of structures.

Refinement was performed using the *PHENIX* suite (Liebschner et al., 2019), with all structures validated using *Molprobity* (Chen et al., 2010), the JCSG QCcheck server (https://smb.slac.stanford.edu/jcsg/QC/) and tools within *PHENIX* and *COOT* (Emsley, Lohkamp, Scott, & Cowtan, 2010; Emsley & Cowtan, 2004).

7.2.3. Generating data subsets

In order to explore the minimum number of images needed to determine nitrite ligand binding accurately in SFX and SSX AcNiR data, randomly selected images from the indexed data file (**.out* in *CrystFEL*) were used to form further datasets with varied, defined numbers of images. The *script sacla_counter_paper.py* (written by Martin Appleby, see appendices 7.5.1) was used to generate subsets of SFX data from the **.out* file using the following command line input;

```
$ python sacla_counter_paper.py -i chipname.out -o 10000.out -n
10000
```

where *-i* defines the input **.out* file from which subsets are generated, *-o* defines the output file, and *-n* defines the number of images chosen to produce the subset. These subsets were scaled and merged in the same manner as the datasets containing all images. Subsets were subsequently used in refinement versus the AcNiR model determined using all data, from which the ligand had been removed. *process_hkl* was used for merging these data within *CrystFEL* using the following command line input;

```
$ module load CrystFEL
$ process_hkl -y m-3 --push-res=1.6 --scale -i 10000.out -
o{out}1 -j 20
```

The output file from merging was then converted to an *mtz* file using *create-mtz* in *CrystFEL* (see chapter 3.5.4). Rfree column and F's were added using *truncate* in the *CCP4i* suite. Omit maps were generated using torsion based simulated-annealing refinement in *phenix.refine* in order to minimise model bias. As a final validation step, selected subsets were refined against the structure of native AcNiR (H_2O bound to the active site in place of nitrite), using the same procedure described above. As an additional validation step, selected subsets were refined against the structure of the native enzymes (where the ligands were not present) using the procedure described above.

7.2.4. Ligand modelling

The NO_2^- ligand was initially modelled into the all-image datasets based on the *mFo-DFc* difference electron density maps, performed using the '*find ligands*' tool inbuilt in the *COOT* software. Ligand restraints were produced using *ACEDRG* (Long et al., 2017). In order to avoid model bias within data subsets, these data were refined by two parallel approaches; firstly, the data were refined against the 'all images' structures from which the ligands had been removed, using simulated annealing (SA) in phenix.refine. As a secondary test that bias was not present, selected structures were refined against the native AcNiR structures where the NO_2^- ligand was not present, replaced by a single water molecule in the active site. SA omit difference maps were then generated. As the position of the ligand is known in the 'all-images'

models, these were then compared to the difference density map generated from specific subsets, in order to examine the difference in electron density between the all-data models and subset models. In all cases, when using `COOT` the ligand was loaded and the `find ligands` feature, searching the *Fo-Fc* simulated annealing omit map for suitable hits.

7.3. Results

7.3.1. SFX data collection and processing

Data were collected from 4 chips for the resting state structure, and 2 chips for the nitrite bound structure (table 7.1). Initial estimation of hit rate using *CHEETAH* gave approximate hit rates seen in table 7.1, with diffraction to higher than 2 Å on initial viewing of *HDF5* files (fig. 7.2). Indexing and integration were handled by *indexamajig* in *CrystFEL*, using a detector geometry refined for this beamtime, and unit cell parameters that were defined using a room temperature data collection at I24 prior to SFX data collection (see chapter 3). The bash script *index.sh* was used for indexing and integration of SFX data (chapter 3.5), and run against the SACLA *hdf5* file for each individual chip to produce a **chip_name*.out* file. Chips were initially processed individually, with the resulting **chip_name*.out* files combined using the command:

```
$ cat berlin.out canton.out quetta.out rimini.out > BCQR.out

$ cat kansas.out urasoe.out > KU.out
```

A total of 82918 frames were integrated for the complete native AcNiR dataset (*BCQR.out*) and 41024 frames for the complete ligand bound dataset (*KU.out*). This was used to define accurate SFX unit cell parameters using the *cell_explorer* program within *CrystFEL*, which generated a unit cell distribution plot for AcNiR that allows Gaussian function to be fit in order to define mean unit cell values. Unlike in chapter 6, the entirety of the data were used to define unit cell parameters due to the

Run number	Sample	Chip name	Approximate hit rate (%)	Exposure events
23368	AcNiR	Berlin	50	25700
23369	AcNiR	Canton	60	25700
23412	AcNiR + NO_2^-	Kansas	60	25700
23432	AcNiR	Quetta	70	25700
23436	AcNiR + NO_2^-	Urasoe	80	25700
23464	AcNiR	Rimini	80	25700

Table 7.1. Chips used to collect AcNiR SFX data with initial approximate hit rate measured by CHEETAH *(Barty et al., 2014).*

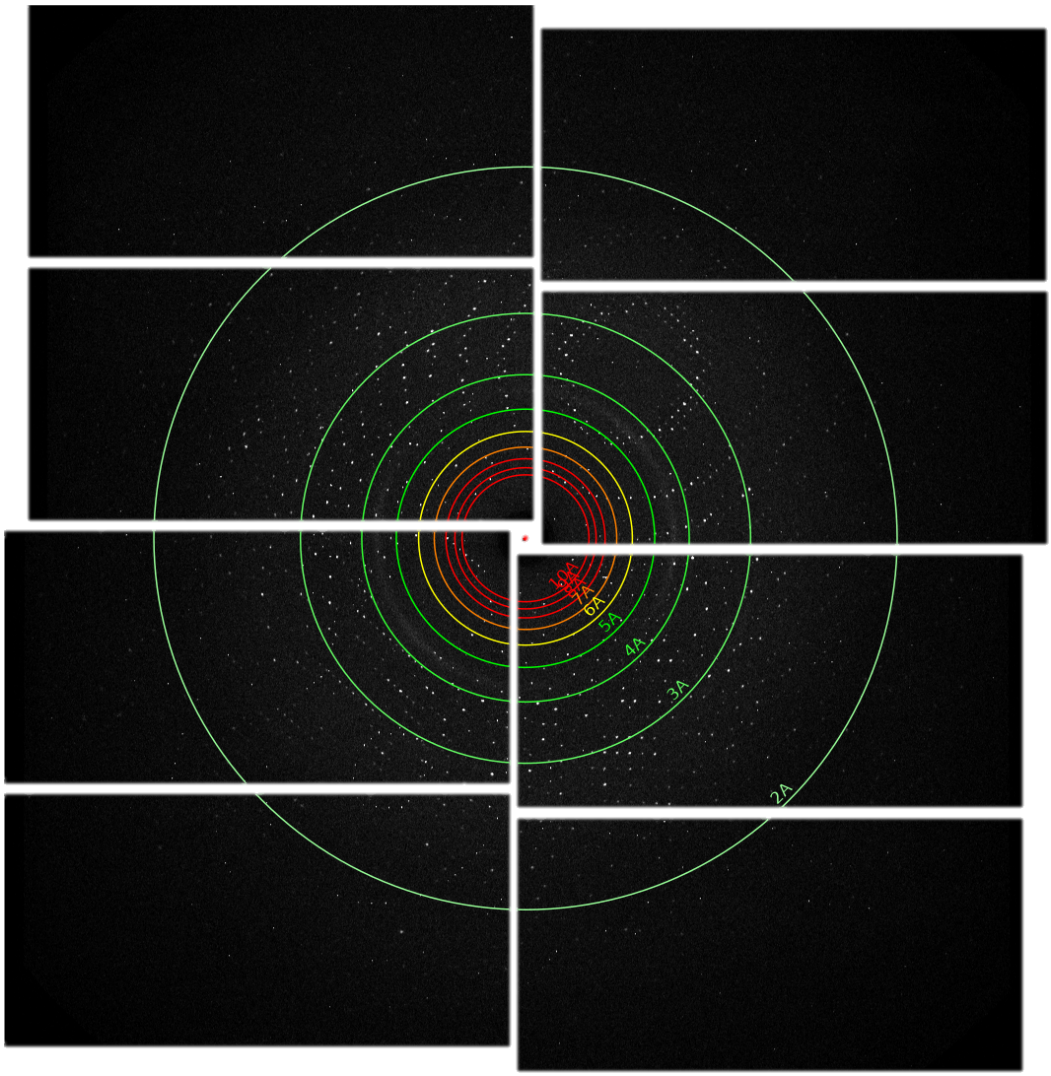


Figure 7.2. AcNiR SFX diffraction

Example image of AcNiR microcrystal diffraction pattern from Berlin dataset (run number 23368), produced using `hdfsee` *in the* `CrystFEL` *suite. Reflections were seen to greater than 2 Å resolution.*

prevalence of polymorphic unit cell of AcNiR microcrystals. As in chapter 3, a polymorphic unit cell population is evident for both native and ligand-bound samples (fig. 7.3), with data again found to be in space group $P2_13$. Only a single unit cell variant was used for structures in order to prevent ambiguous electron density around the active site. The large cell variant was used for AcNiR + NO_2^- datasets as there was more data available, with the large cell variant selected for native data in order to keep unit cell variation to a minimum. Updated unit cell dimensions of a = b = c = 97.6 Å, and $\alpha = \beta = \gamma = 90°$ were used for the ligand bound large unit cell data, with unit cell dimensions of a = b = c = 97.7 Å, and $\alpha = \beta = \gamma = 90°$ used for the native large unit cell data. The unit cell was set using specific *AcNiR.cell* files for the *indexamajig* program (chapter 3.5.3). Unit cell tolerance was limited, defined in the input syntax under:

```
$ --tolerance=1,1,1,1.5
```

over-riding the default tolerances for reciprocal space axis length and reciprocal space angles of 5 and 1.5 percent, respectively.

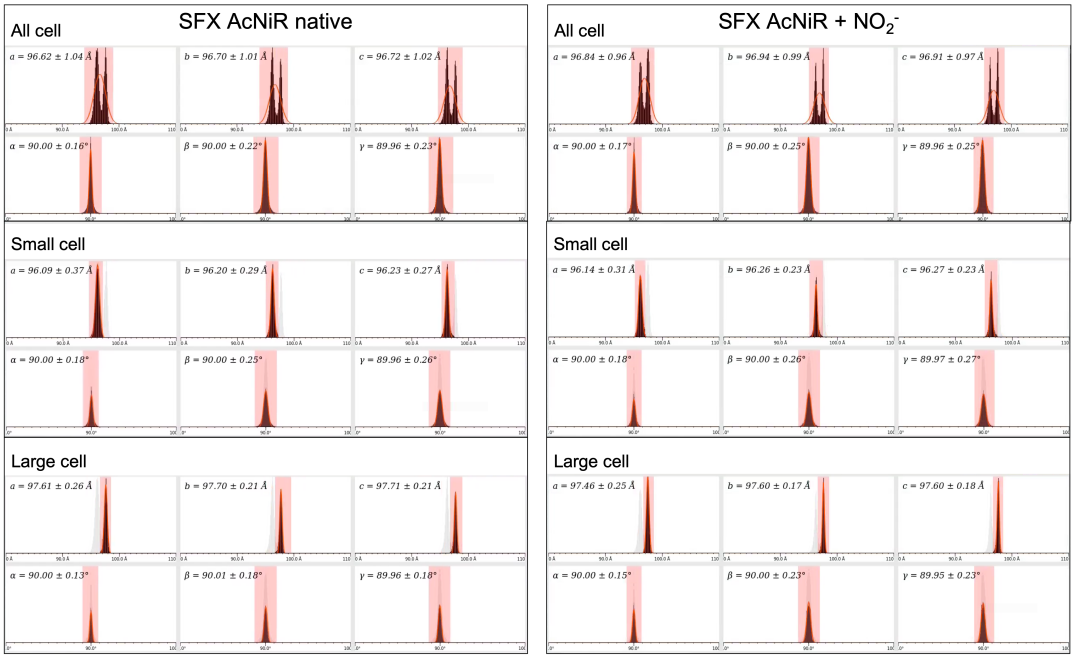


Figure 7.3. SFX native and ligand bound AcNiR unit cell distributions

The `cell_explorer` *program in the* `CrystFEL` *suite indicates a polymorphic unit cell is present within the crystal populations for native and ligand bound AcNiR crystals. The distribution in unit cell length is ~1 Å for both native and* NO_2^- *data prior to splitting of unit cell population. The respective large unit cell parameters used for AcNiR native and AcNiR* NO_2^- *were a = b = c = 97.7 Å,* $\alpha = \beta = \gamma = 90.00°$*, and a = b = c = 97.6 Å,* $\alpha = \beta = \gamma = 90.00°$*.*

The *detector_shift* script (chapter 3.5.3) was used in order to check and optimise the detector position, using the command:

```
$ ./detector-shift BCQR.out 23368-1.geom

$ ./detector-shift KU.out 23412-1.geom
```

This provided an updated geometry file named *23368-1-predrefine.geom* for native AcNiR and *23412-1-predrefine.geom* for NO_2^- bound AcNiR, as well as a scatter plots detailing the X, Y detector shifts (equivalent to the required beam shift) for each pattern in each respective dataset (fig.7.4). The *indexamajig* step was repeated for both native and ligand bound AcNiR, producing files called *BCQR_all.out* and *KU_AcNiR_large.out*. It was possible to integrate a total of 83028 frames for the native AcNiR dataset, and 16586 frames for the ligand bound AcNiR large cell dataset.

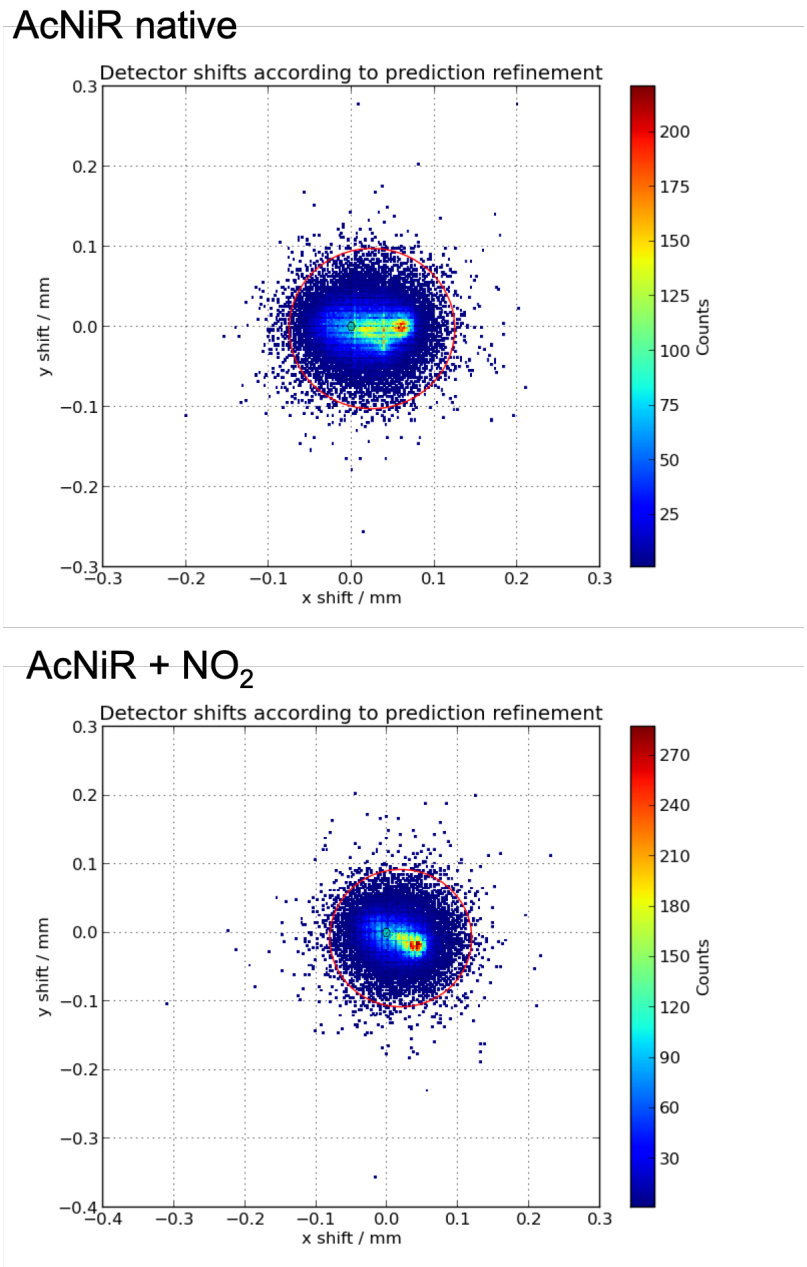


Figure 7.4. Detector shift plot.

Plot generated by the program detector-shift indicating the distribution of measurements for every AcNiR crystal.

In order to correct for indexing ambiguity in the space group $P2_13$, the *CrystFEL* program *Ambigator* was run against both **.out* files using the indexing ambiguity operator k, h, -l, and the bash script *ambig.sh* (see section 3.5.3). These data were subsequently merged in *CrystFEL* using *process_hkl*, using the bash script *merge.sh* (see section 3.5.4). Merged output files *BCQR_large.hkl* and *KU_large.hkl* (all merged reflections) were produced, alongside *BCQR_large.hkl1*, *KU_large.hkl1*, *and BCQR_large.hkl2*, *KU_large.hkl2* (half-datasets for the generation of data quality statistics, using the script *stat.sh* [see section 3.5.4]). Dataset completeness was inspected within the shells.dat output file from *stat.sh*; this was used alongside $CC_{1/2}$ and R_{split} statistics to estimate at what particular resolution limit these data should be merged. Based on this these data were cut at 1.9 Å for the ligand bound structure, and 1.80 Å for the native structure (table 7.2). This resolution limit was added to the *merge.sh* script, and run again against each **.out* file, allowing the production of all data mtz files for both structures using the *create-mtz CrystFEL* executable. As in chapters 4, 5, and 6 molecular replacement was performed using *PhaserMR* within the *PHENIX* crystallography suite. Refinement was performed using *PHENIX.refine*, with a final R_{work} and R_{free} of 14.3 % and 16.3 % for the resting state structure and 13.9 % and 17.4 % for the ligand bound structure, respectively (table 7.2).

	SFX AcNiR + NO_2^-	SFX Native AcNiR
Data Collection		
Cumulative dose (kGy)	N/A	N/A
Chips used	2	4
Images collected	51200	102400
Number of images used	16586	25119
Data processing		
Space Group	$P2_13$	$P2_13$
Cell dimensions (Å, °)	a = b = c = 97.6, $\alpha = \beta = \gamma = 90°$	a = b = c = 97.7, $\alpha = \beta = \gamma = 90$
Resolution (Å)	43.7-1.90 (1.93-1.90)	43.69-1.80 (1.83-1.80)
R_{split} (%)	9.73 (0.59)	8.2 (75.18)
$CC_{1/2}$	0.99 (0.63)	0.99 (0.56)
Multiplicity	3281.4 (2299.1)	3764.72 (1806.7)
Completeness (%)	100 (100)	100 (100)
Refinement		
Number of reflections	24729	29106
R_{work}/R_{free} (%)	13.9/17.4	14.3/16.3
RMSD bond lengths (Å)	0.007	0.006
RMSD bond angles (°)	0.9	0.887
Ramachandran plot		
Most favoured (%)	98.8	98.49
Allowed (%)	1.1	1.5
PDB accession code	6QWG	N/A

Table 7.2. Native and NO_2^- bound AcNiR SFX structure data collection, processing, and refinement statistics.

7.3.2. SSX data collection and processing

Chips were loaded as specified in chapter 3. A single 20 ms exposure was measured at each aperture of the chip. A single chip loaded with 100 mM NO_2^- soaked AcNiR crystal slurry was used to collect the dataset, with the total dataset comprising 25600 images and collected in less than 15 minutes. Diffraction was measured to > 1.6 Å toward the corners of the detector (fig. 7.5). Indexing and integration of cbf files was handled by `indexamajig` in `CrystFEL` (rather than `DIALS` as used for SSX data in chapters 4 and 5) in order to minimise differences between SSX and SFX data processing. A geometry file for the PILATUS3 6M was created by Dr. Danny Axford from beamline I24 at Diamond Light Source, refined for specifically for this beamtime and detector position (appendices 7.5.1). As before a `*.out` file was produced for using the script `index.sh` (chapter 3.5.3). Again, a unit cell polymorph was evident, so data were restricted to a single unit cell that had the greatest number of integrated stills available, in this case the small cell variant of AcNiR (fig. 7.6).

Indexing and integration were repeated with an updated cell, restricted to a = b = c = 96.4 Å, and $\alpha = \beta = \gamma = 90°$ (space group $P2_13$). `AcNiR_SSX.out` was produced using a total of 10386 patterns for a 1.7 Å small cell dataset (~41% hit rate). These data were merged in CrystFEL using `process_hkl`, using the bash script `merge.sh` (chapter 3.5.4). Merged output files containing all merged reflections `AcNiR_SSX.hkl` were produced alongside half datasets `AcNiR_SSX.hkl1` and `AcNiR_SSX.hkl2`, for the generation of data quality statistics using the script stat.sh. Data were cut at 1.7 Å due to the presence of silicon diffraction spots at ~1.69 Å. These are typically masked in the processing of SSX data using `dials.stills_process`, however this was not possi-

Figure 7.5. AcNiR + NO_2^- SSX diffraction

Clear diffraction can be seen past 2 Å, together with the presence of extremely intense silicon spots at ~1.69 Å.

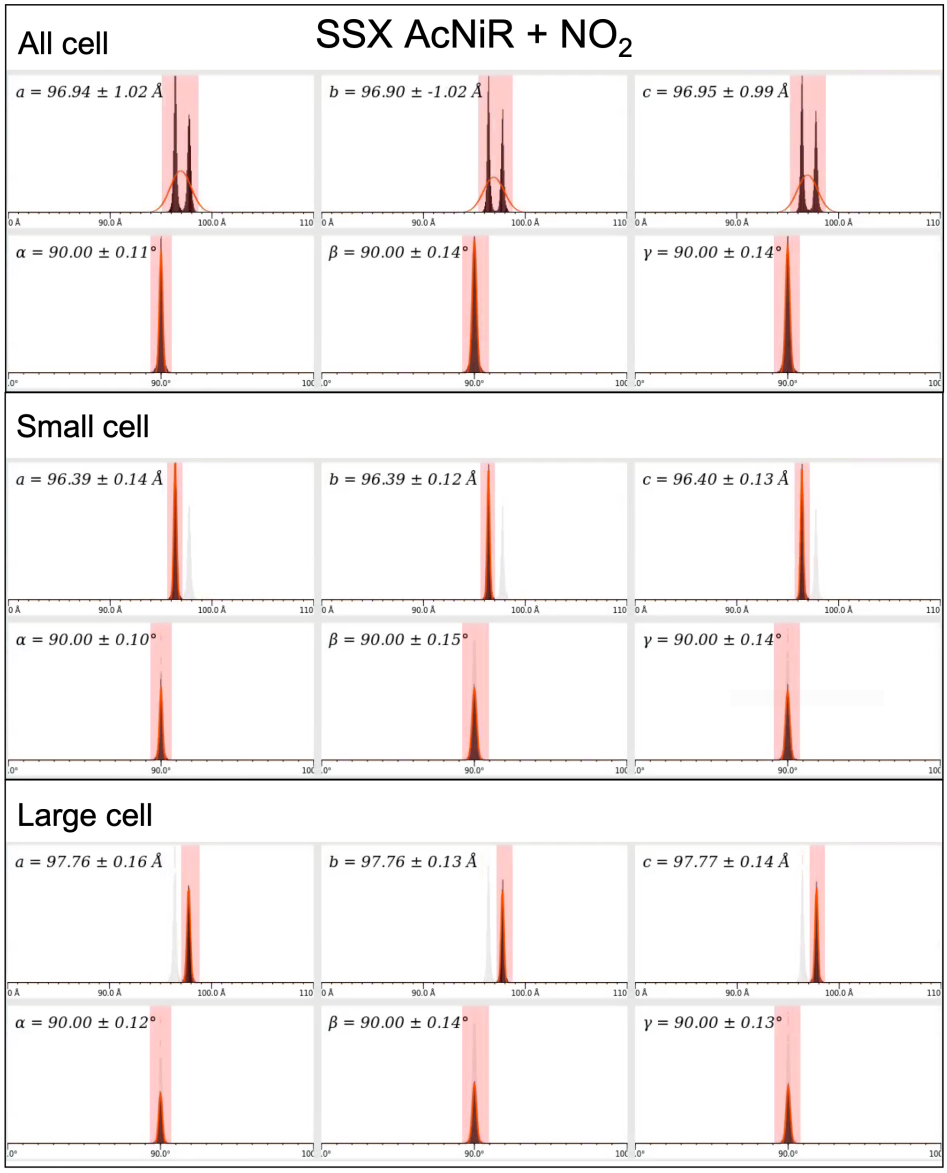


Figure 7.6. SSX ligand bound AcNiR unit cell distribution

The `cell_explorer` *program in the* `CrystFEL` *suite indicates a polymorphic unit cell is again present within the crystal populations for native AcNiR crystals. The distribution in unit cell length is ~1 Å prior to splitting of unit cell population. The unit cell parameters used for AcNiR + NO_2^- processing were a = b = c = 96.4 Å, $\alpha = \beta = \gamma = 90.00°$.*

ble in *CrystFEL* (see discussion). *stat.sh* was again used to assess data quality, using the shells.dat output file from stat.sh alongside $CC_{1/2}$ and R_{split} statistics (table 7.3). A resolution limit of 1.7 Å was added to the merge.sh script, and run again against each **.out* file, allowing the production of all data mtz files for both structures using the *create-mtz CrystFEL* executable. As in previous chapters 4 and 5 molecular replacement was performed using *PhaserMR* within the *PHENIX* crystallography suite. Refinement was performed using *PHENIX.refine*, with a final R_{work} and R_{free} of 14.1 % and 17.5 % for the ligand bound SSX structure (table 7.3).

7.3.3. SFX and SSX structures of ligand bound complexes

The SFX structure of AcNiR in complex with nitrite was determined to 1.90 Å resolution, while the native structure was determined to 1.80 Å. Both structures share the same overall fold of alpha helices and beta sheets, shown in figure 7.7a in its homotrimeric biological assembly. The position of the ligand-binding site within the protein fold of the biological assembly is shown in figure 7.7b. The ligand binding site (the type 2 Cu site) displays clear electron density for a bound nitrite molecule in the ligand soaked AcNiR crystal structure (fig. 7.7c). A bidentate O-binding geometry is apparent as previously described in multiple publications depicting AcNiR structures at ambient temperatures (Meyder et al., 2017; Horrell et al., 2016; Horrell et al., 2019) and in chapter 4. The native SFX structure exhibits a similar conformation of key residues Asp98, His100, His135, and His306 around the T2Cu as the nitrite bound SFX structure; however, where nitrite is typically bound in NO_2^- soaked crystals, native AcNiR the has a water bound (fig. 7.7d).

SSX AcNiR + NO_2^-	
Data Collection	
Cumulative dose (kGy)	11
Chips used	1
Images collected	25600
Number of images used	10386
Data processing	
Space Group	$P2_13$
Cell dimensions (Å, °)	a = b = c = 96.4, $\alpha = \beta = \gamma = 90$
Resolution (Å)	34.08-1.70 (1.73-1.70)
Rsplit (%)	9.66 (41.8)
CC1/2	0.99 (0.78)
Multiplicity	876.95
Completeness (%)	100 (100)
Refinement	
Number of reflections	33109
Rwork/Rfree (%)	14.1/17.5
RMSD bond lengths (Å)	0.006
RMSD bond angles (°)	0.908
Ramachandran plot	
Most favoured (%)	98.49
Allowed (%)	1.5

Table 7.3. AcNiR + NO_2^- SSX structure data collection, processing, and refinement statistics.

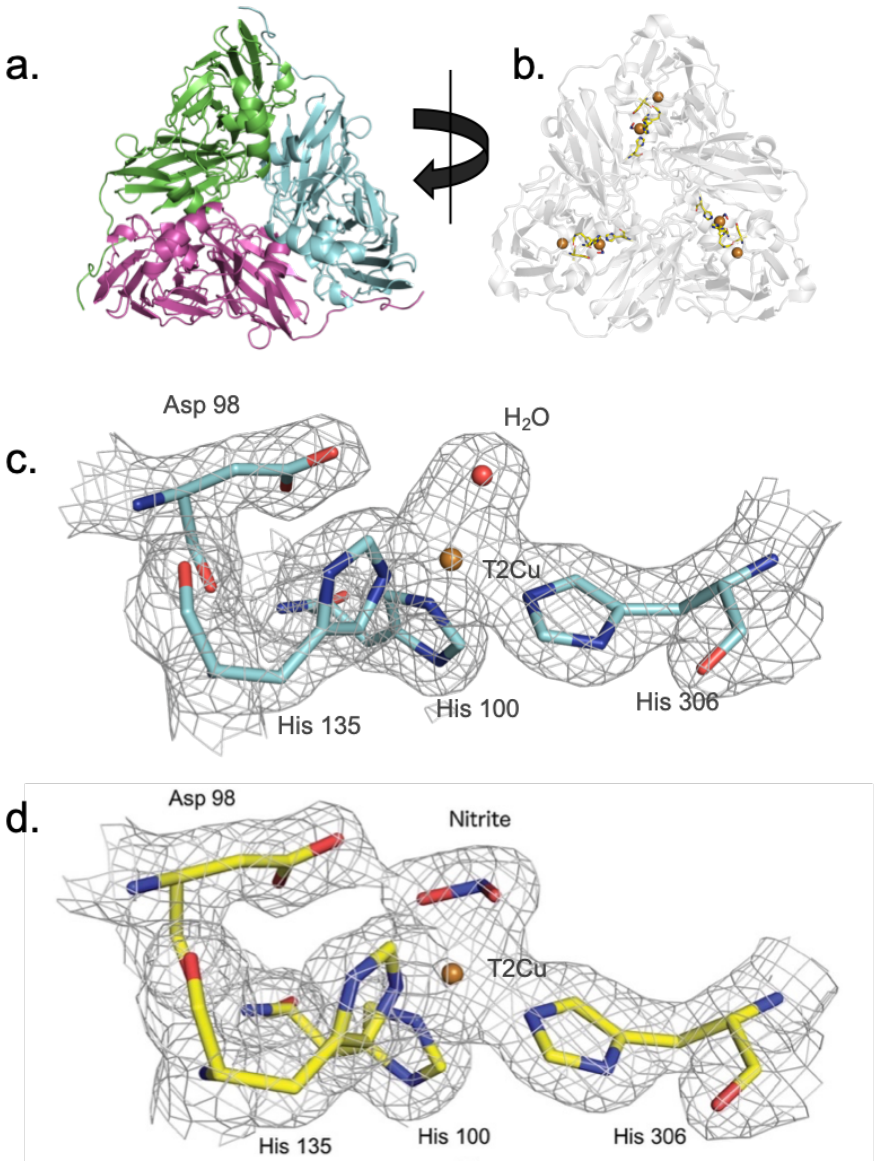


Figure 7.7. Overall fold and active site regions of native and nitrite bound SFX structures

(a.) Overall fold of native AcNiR represented as a cartoon, with monomers that form the biological assembly individually coloured green, cyan, and magenta. Transparent cartoon representing the overall fold of nitrite bound structure is shown from the opposite face of the protein in order to more easily distinguish the active site. (b.) 2Fo-Fc representation of the active site conformation of native SFX AcNiR shows water bound at the T2Cu, while the (c.) active site conformation of NO_2^- bound SFX AcNiR shows distinct electron density for bound nitrite. Maps contoured at 1σ.

7.3.4. SFX and SSX data subset generation, merging, and refinement

To test the lower limits of the number of diffraction patterns that would allow the identification of bound ligands in high-throughput chip experiments, the data were partitioned into subsets of decreasing size to produce independent merged data sets containing progressively fewer diffraction patterns. This was achieved using the script `sacla_counter_paper.py`. by using the following command was used to generate subsets:

```
$ python sacla_counter_paper.py -i KU_AcNiR_large.out -o
SFX_10000.out -n 10000
```

From the 16586 images in the all data SFX dataset, subsets of 10000, 5000, 2000, 1000, 800, 600, and 500-100 were produced. Merging of subsets was handled by *process.hkl*, and performed initially on the *SFX_10000.out* subset using a simple bash script *merge.sh*;

```
$ inp=SFX_10000.out
$ out=SFX_10000.hkl
$ pg="m-3"

$ module load CrystFEL

$ process_hkl -y $pg --push-res=1.6 --scale --odd-only -i $inp
-o ${out}1 -j 20 &
$ process_hkl -y $pg --push-res=1.6 --scale --even-only -i $inp
-o ${out}2 -j 20 &
$ process_hkl -y $pg --push-res=1.6 --scale -i $inp -o ${out} -
j 20 &
wait
```

Merging symmetry m-3 was used in this instance in order to merge Friedel pairs. As before these data were split into half datasets (*sfx_10000.hkl1*; *SFX_10000.hkl2*) when merged in order to assess where resolution limits should be set, using a $CC_{1/2}$ of >0.5 as the cut-off. This was performed for all SFX subsets giving the resolution limits found in table 7.4. However, as interpretation of merging statistics including $CC_{1/2}$ and R_{split} were unreliable with the small image datasets, the 1000 image subset resolution cut-off of 2.11 Å was used as the resolution limit for all subsets below 1000 images. Due to SSX data being cut at 1.7 Å due to the issue with silicon spot diffraction, despite SSX diffraction to higher resolutions, all SSX subsets were cut to 1.7 Å. All merged subset *.hkl* files were successfully converted to an mtz using the *create-mtz CrystFEL* executable (appendices 3.7.5), giving a total of 19 subset mtzs used (11 SFX, 8 SSX). An R_{free} column was added and intensities converted to structure factor amplitudes using the program *truncate* in *CCP4i*.

A no ligand-occupancy simulated-annealing (SA) refinement in *PHENIX* using the settings stated in section 7.2.4 was performed for SFX subsets, using the all data structure with NO_2^- atoms omitted from the initial model was used to generate mF_o-DF_c SA omit maps for each subset. Weightings for atomic displacement parameters and stereochemistry were determined automatically in refinement, with weightings auto-optimised and hydrogens added automatically. Despite exhibiting merging statistics that would typically be considered poor (table 7.4), features of positive electron density were evident even in the *SFX_100.mtz* subset, featuring only 98 merged images (fig. 7.8). The ability of *COOT* to auto find NO_2^- at the T2Cu active site of AcNiR was used as a metric of the information content of electron density maps

SFX AcNiR + NO_2^- subsets												
Structure	**All data**	**10000**	**5000**	**2000**	**1000**	**800**	**600**	**500**	**400**	**300**	**200**	**100**
Scaling and merging												
Images merged	16586	9988	4993	1997	996	795	597	497	398	298	199	98
Unique reflections	24729	24001	22575	20649	18162	18161	18160	18160	18160	18160	18159	18101
Resolution (Å)	43.7-1.90	43.7-1.92	43.7-1.96	43.7-2.02	43.7-2.11	43.7-2.11	39.9-2.11*	39.9-2.11*	39.9-2.11*	39.9-2.11*	39.9-2.11*	39.9-2.11*
Outer shell (Å)	1.93-1.90	1.99-1.92	2.03-1.96	2.09-2.02	2.15-2.11	2.15-2.11	2.15-2.11	2.15-2.11	2.15-2.11	2.15-2.11	2.15-2.11	2.15-2.11
Completeness (%)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	99.6 (99.6)
Multiplicity	3281.4 (2299.1)	2198.4 (1461.9)	1009.4 (713.3)	391.7 (276.0)	181.8 (128.7)	136.4 (96.6)	95.3 (67.3)	77.6 (54.8)	61.0 (43.0)	42.7 (30.1)	27.1 (19.2)	13.0 (9.3)
CC1/2	0.99 (0.63)	0.98 (0.62)	0.97 (0.56)	0.92 (0.55)	0.92 (0.59)	0.84 (0.55)	0.82 (0.45)	0.81 (0.48)	0.77 (0.40)	0.71 (0.40)	0.61 (0.38)	0.4 (0.07)
Rsplit (%)	9.73 (0.59)	12.1 (62.5)	15.8 (61.8)	23.2 (70.1)	30.8 (68.9)	33.4 (73.5)	36.9 (87.7)	39.5 (91.0)	43.5 (105.1)	48.9 (110.0)	59.7 (129.5)	78.3 (160.1)
Refinement												
Resolution range (Å)	43.66-1.90	43.66-1.92	43.66-1.96	43.66-2.02	43.66-2.11	43.66-2.11	34.5-2.11	34.5-2.11	34.5-2.11	34.5-2.11	34.5-2.11	34.5-2.11
Outer shell (Å)	2.00-1.90	1.98-1.92	2.05-1.96	2.13-2.02	2.22-2.11 2	2.22-2.11 2	2.22-2.11 2	2.22-2.11 2	2.22-2.11 2	2.22-2.11 2	2.22-2.11 2	2.22-2.11 2
Rwork (%)	13.9 (23.9)	14.7 (26.4)	15.4 (24.3)	16.3 (24.2)	17.8 (24.1)	18.4 (24.8)	19.2 (26.1)	19.9 (27.0)	20.7 (27.3)	22.4 (28.3)	24.3 (29.1)	28.6 (32.8)
Rfree (%)	17.4 (32.0)	18.7 (32)	19.0 (28.2)	20.7 (25.9)	23.3 (28.1)	23.3 (27.4)	25.1 (29.8)	25.5 (31.5)	26.5 (33.4)	28.3 (32.8)	30.2 (33.8)	35.1 (40.0)
RMSD bond (Å)	0.007	0.009	0.004	0.005	0.005	0.004	0.004	0.004	0.005	0.004	0.006	0.006
RMSD angle (°)	0.9	0.93	0.76	0.79	0.81	0.77	0.77	0.8	0.8	0.76	0.81	0.84
Ramachandran favoured (%)	98.8	98.8	98.8	99.8	98.5	98.5	98.8	97.9	97.6	97.6	97.9	95.8

Table 7.4. AcNiR + NO_2^- SFX subset structure processing and refinement statistics.

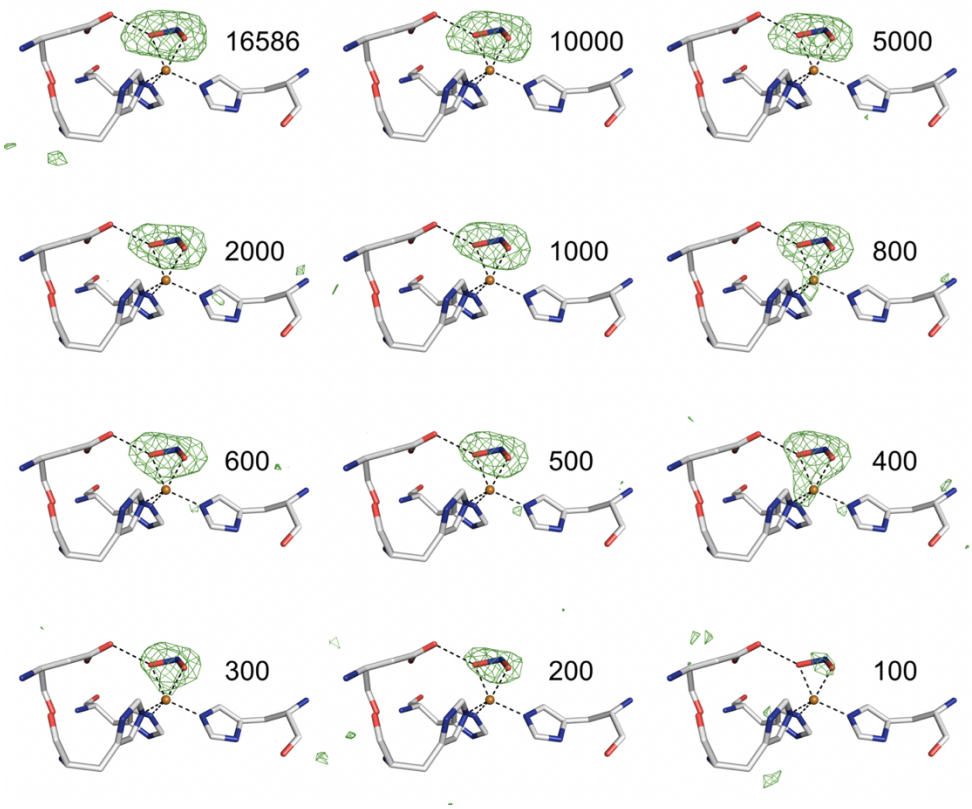


Figure 7.8. Fo-Fc *simulated annealing omit maps from selected data subsets of the SFX* Ac*NiR-nitrite structure.*

The type 2 Cu active site region coordinates are shown from the SFX all data structure, where the nitrite ligand occupancy was 1. Features of positive electron density are present throughout the subsets where the NO_2^- *ligand typically binds. Positive density also seen for the 100-image subset, despite the* `find ligands` *feature in* `COOT` *unable to successfully locate a ligand at this position in this subset. Maps were contoured at* 3σ.

obtained. The lowest number SFX subset this was achievable with was the *SFX_200.mtz* subset, despite the aforementioned positive density in the *SFX_100.mtz* subset shown in figure 7.8. Merging and refinement statistics for all SFX structures can be found in table 7.4.

The same SA omit refinement was repeated in *PHENIX* against the SSX subsets. Features of positive electron density were evident again in the smallest subset *SSX_100.mtz*, with 98 images merged. In contrast to the SFX data it was possible to successfully auto find NO_2^- at the active site using *find ligands* in the *SSX_300.mtz* subset (merging and refinement statistics for all SSX structures can be found in table 7.5). Again, despite exhibiting merging statistics that would typically be considered extremely poor, very clear *Fo-Fc* simulated annealing omit map features for the ligand were evident in subsets of small numbers of diffraction patterns. *COOT* successfully located NO_2^- binding at the type 2 Cu active site in subsets of very few crystals. Although noisy, positive electron density is again evident in the lowest number datasets (fig. 7.9), despite *COOT find ligands* unsuccessful attempt at determining nitrite in subsets 200 and 100.

SSX AcNiR + NO_2^- subsets									
Structure	**All data**	**10000**	**5000**	**1000**	**500**	**400**	**300**	**200**	**100**
Scaling and merging									
Images merged	10386	9998	4999	999	499	399	298	198	99
Unique reflections	33109	33109	33109	33109	33108	33108	33106	33088	32291
Resolution (Å)	34.08-1.70	34.08-1.70	34.08-1.70	34.08-1.70	34.08-1.70	34.08-1.70	34.08-1.70	34.08-1.70	34.08-1.70
Outer shell (Å)	1.73-1.70	1.73-1.70	1.73-1.70	1.73-1.70	1.73-1.70	1.73-1.70	1.73-1.70	1.73-1.70	1.73-1.70
Completeness (%)	100 (100)	100 (100)	100 (100)	100 (100)	100 (99.9)	100 (99.9)	100 (99.9)	99.9 (99.8)	97.5 (93.4)
Multiplicity	876.95	841.62	400.21	81.18 (50.4)	46.02 (28.8)	37.02 (23.5)	28.37 (18.1)	16.41 (10.6)	7.43 (4.8)
CC1/2	0.99 (0.78)	0.98 (0.78)	0.97 (0.70)	0.85 (0.33)	0.75 (0.23)	0.71 (0.19)	0.61 (0.16)	0.51 (0.10)	0.31 (0.16)
Rsplit (%)	9.66 (41.80)	9.83 (42.59)	14.07 (51.90)	32.73 (106.98)	45.96 (144.30)	51.69 (160.13)	59.50 (170.14)	73.63 (184.31)	97.63 (206.9)
Refinement									
Resolution range (Å)	34.08-1.70	34.08-1.70	34.08-1.70	34.08-1.70	34.08-1.70	34.08-1.70	34.08-1.70	34.08-1.70	34.08-1.70
Outer shell (Å)	1.75-1.70	1.75-1.70	1.75-1.70	1.75-1.70	1.75-1.70	1.75-1.70	1.75-1.70	1.75-1.70	1.75-1.70
Rwork (%)	14.14 (20.62)	13.33 (18.52)	14.06 (19.24)	18.25 (27.96)	21.44 (30.89)	22.51 (31.84)	23.89 (32.21)	29.94 (33.47)	33.62 (34.52)
Rfree (%)	17.51 (27.24)	15.74 (22.27)	16.98 (22.85)	21.23 (28.34)	25.34 (33.89)	26.45 (36.80)	28.16 (36.92)	34.79 (35.00)	38.15 (39.35)
RMSD bond (Å)	0.006	0.012	0.012	0.012	0.014	0.013	0.013	0.017	0.018
RMSD angle (°)	0.908	1.168	1.307	1.239	1.253	1.307	1.203	1.760	1.659
Ramachandran favoured (%)	98.49	98.19	97.89	98.19	98.19	97.58	98.19	92.45	96.37

Table 7.5. AcNiR + NO_2^- SSX subset structure processing and refinement statistics.

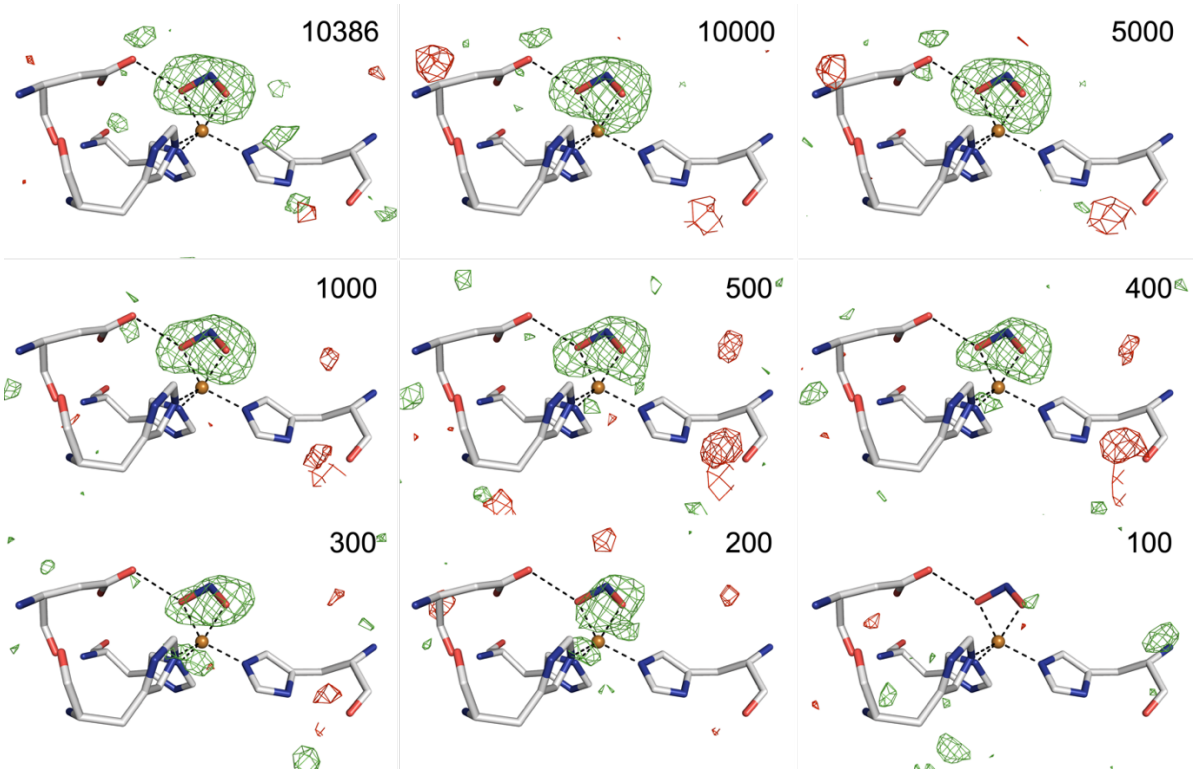


Figure 7.9. Fo-Fc *simulated annealing omit maps from selected data subsets of the SSX* Ac*NiR-nitrite structure.*

The type 2 Cu active site region coordinates are shown from the SFX all data structure consisting of 10386 images, where the nitrite ligand occupancy was 1. Features of positive electron density are present throughout the subsets where the NO_2^- *ligand typically binds, though almost entirely absent in the 100-image subset. Maps were contoured at 3*σ*.*

In order to validate the electron density in SFX ligand bound structures, selected SFX subsets were refined against the corresponding native structures obtained by SFX where NO_2^- was not present in sample preparation. The validity of this data processing approach is scrutinised when using AcNiR-nitrite complex as the test sample, due to NO_2^- being such a small ligand at only 3 atoms. This particular case is especially challenging as NO_2^- displaces the water found bound at the active site in native AcNiR, which is seen in the native AcNiR SFX structure presented in results section 7.3.3 (Antonyuk et al., 2005; Kekilli et al., 2014), with the presence of a second water molecule also a possibility (chapter 4, section 3). Refinement of AcNiR SFX data and subsets versus the native AcNiR SFX structure with water bound at the T2Cu produced clear positive difference map features at the type 2 copper centre, where nitrite is typically bound in either a top-hat or side-on conformation (fig. 7.10). This was repeated with the bound T2Cu water completely removed, again indicating positive electron density that is indicative of a molecule that is large that a single bound water atom (fig. 7.11).

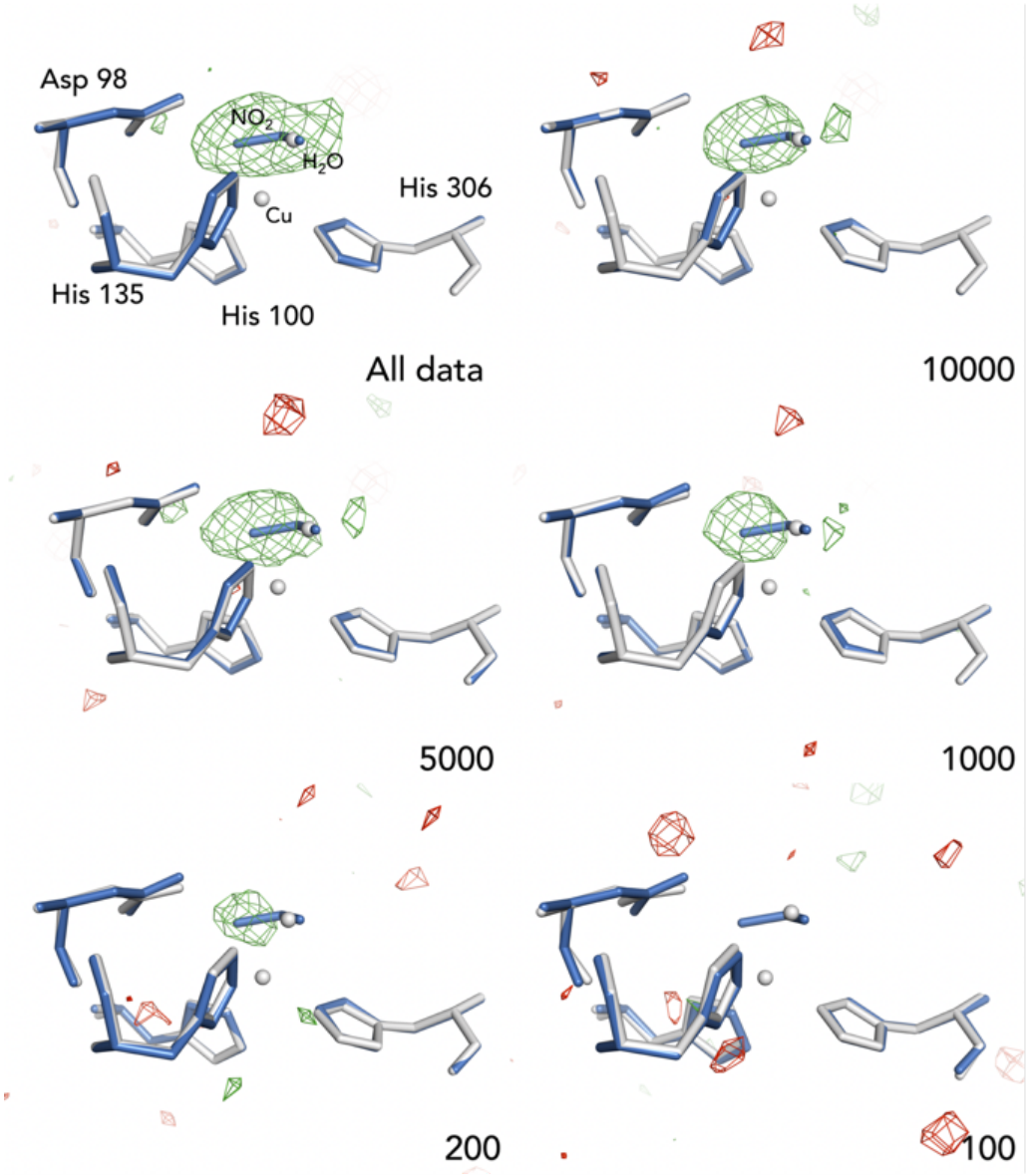


Figure 7.10. Fo-Fc data subset omit maps from SFX AcNiR NO_2^- structures versus SFX native AcNiR with the T2Cu bound water molecule included in the model.

The structures from omit refinement are shown in grey, with the superimposed all-data structure of the AcNiR-nitrite complex shown in blue. Positive features of electron density are present despite water bound to the T2Cu in the native AcNiR SFX structure, indicating an NO_2^- ligand is present rather than a bound water.

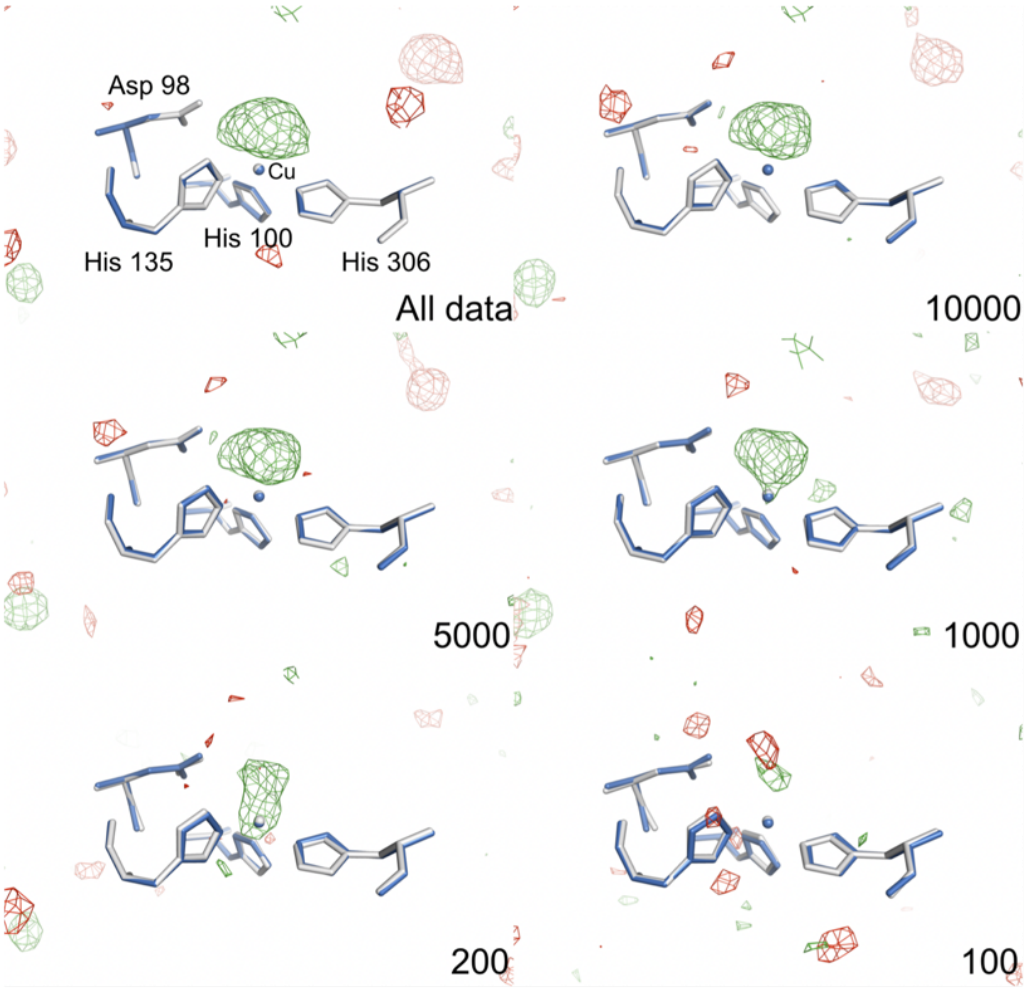


Figure 7.11. F_o-F_c data subset omit map from SFX AcNiR NO_2^- structures versus SFX native Ac*NiR with the T2Cu bound water removed.*

The structures from omit refinement are shown in grey, with the superimposed all-data structure of the AcNiR-nitrite complex shown in blue. Clear and distinct features of electron density are present in the maps. F_o-F_c electron density maps become especially noisy at 200 and fewer images with positive electron density covering the T2Cu as well as part of the binding area of NO_2^-. The 100-image dataset shows positive electron density at the binding site, however with the presence of negative electron density in the same vicinity, it is difficult to confidently distinguish this as a bound ligand rather than noise.

7.4. Discussion

By using a fixed target method at XFELs, our results indicate that it is possible to achieve high-quality SFX crystal structures that allow unambiguous ligand identification. Furthermore, this is possible using extremely small quantities of protein microcrystals compared to typical jetting methods (Schlichting, 2015). As well as this, with the high throughput nature of using silicon chips allowing a rapid switchover between individual samples, it was possible to collect all of the SFX data needed for this experiment in around 90 minutes of an allotted 36 hours of SACLA beamtime, with < 15 minutes needed to collect data from each of the 6 chips. When concerning SSX beamtime, the extremely high redundancy dataset used in this experiment was collected in only 15 mins of beamtime at DLS. Although all diffraction data used in this experiment were measured in a combined time of < 2 hours, it was still possible to confidently model ligands in all data structures as well as in subsets unambiguously. The subset data presented in this chapter clearly demonstrate that only a fraction of diffraction images collected within a total 36-hour XFEL experiment is necessary to confidently locate ligands, demonstrated by the automated location of NO_2^- in the 200 image SFX SA omit map. This 200-image subset is only ~1.2% of the total number of images in the full 16586 data set, sufficient to appropriately model NO_2^- whilst ensuring the possibility of model bias is prohibited. In this, it should be possible to optimise allotted XFEL beamtime to screen ligands, with data collected from a single chip at a 50% hit rate (12800 integrated diffraction patterns) more than enough to assess ligand binding, opening opportunities for FBDD experiments at XFELs.

The overall completeness of AcNiR SFX datasets stays at 100% all the way to the 200-image subset, and 99.9% in the 100-image subset (table 7.4, figure 7.12a), with outer shell completeness of 100% and 99.6%, respectively. SSX data presented similar statistics, though 100% overall completeness started to decrease after the 400-image subset, to 99.9% in subsets 300 and 200, and 100-image subsets at 97.5% (table 7.5, figure 7.12b). Data presented in this chapter strongly suggests that completeness of serial datasets is a key metric for assessing the suitability of merged ligand bound datasets for further investigation. Metrics of data quality such as $CC_{1/2}$ and R_{split} are extremely poor for SFX and SSX subsets, yet successful ligand characterization is possible provided that the data are complete. For AcNiR, with cubic symmetry, the data remained essentially complete in all of the subset sizes analysed, with density for the nitrite ligand remaining apparent down to <200 indexed images in SFX datasets and <400 indexed images in SSX datasets.

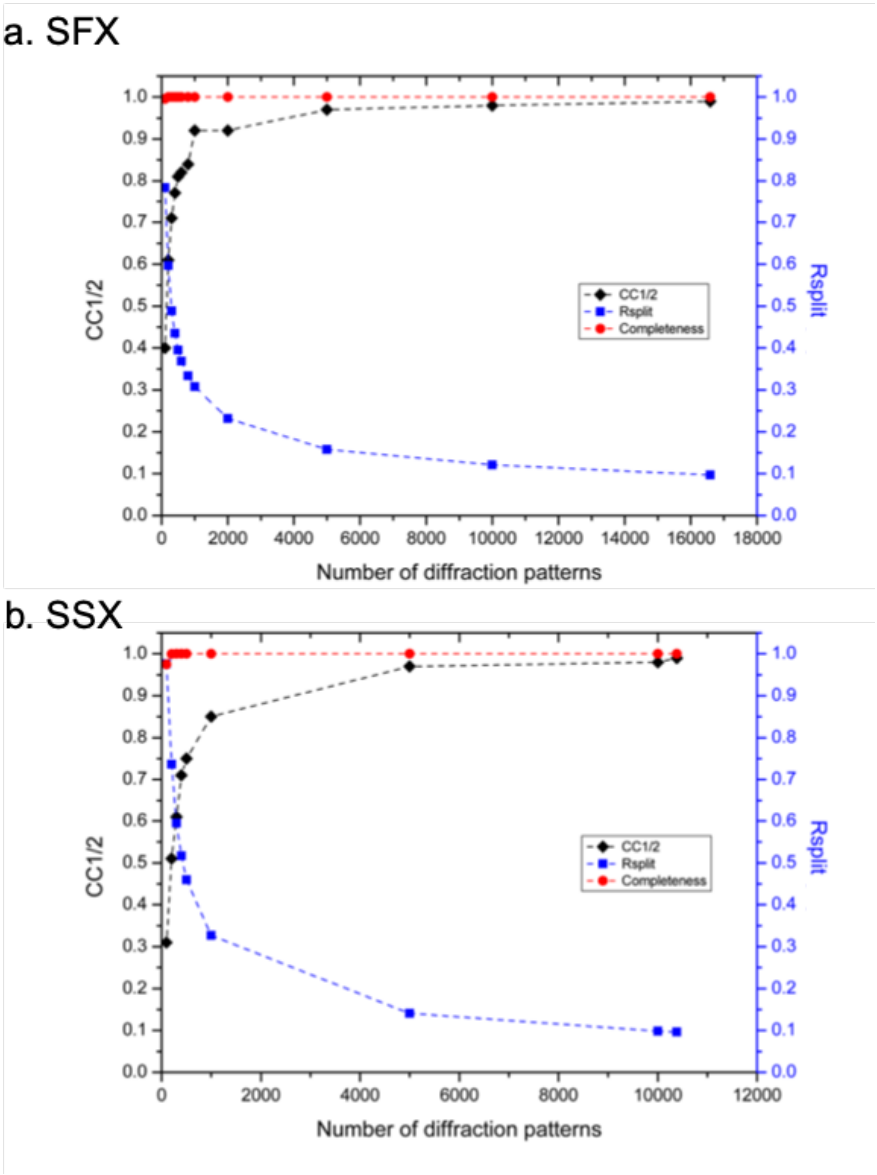


Figure 7.12. data quality metrics as a function of the number of merged crystals in data subsets for SFX and SSX data

(a.) SFX $CC_{1/2}$, R_{split}, and completeness data plotted against number of diffraction patterns. (b.) SSX $CC_{1/2}$, R_{split}, and completeness data plotted against number of diffraction patterns. In both plots it is possible to see that despite the drop off of $CC_{1/2}$ and R_{split} metrics, completeness stays at 100% for nearly all subsets.

Despite NO_2^- being such a small ligand, crystallisation in $P2_13$, a high symmetry space group, results in less data being needed for a complete dataset due to the high redundancy of data collected (table 7.4, 7.5). The bandwidth of the XFEL beam allows complete data to be obtained from fewer crystals than the monochromatic beam available at Diamond; typical bandwidths of XFEL pulses are $\Delta E/E \approx 2 \times 10^{-3}$, compared to $\Delta E/E \approx 10^{-4}$ at monochromatic synchrotron beamlines, about 1/20th of the bandwidth of XFEL pulses. This effectively allows small degree wedges of diffraction to be collected from XFEL pulses compared to monochromatic synchrotron diffraction. Though typical XFEL serial experiments utilise tens to hundreds of thousands of microcrystals (Schlichting, 2015) the high completeness of small data sets parallels the success in forming complete data sets from multiple thin wedges in virus crystallography. Fry et al., 1999, show the completeness of the final data set is a function of the number of wedges collected and the point group of the crystals used (table 7.6) (Fry, Grimes, & Stuart, 1999). In this, it possible to collect almost complete (98.6%) data in the point group 23 with only 36° of data, with nearly twice as much data needed in point group 2 (fig. 7.13).

	Completeness (%)					
Point group	**Number of images (oscillation range)**					
	5 (2°)	**10 (4°)**	**20 (8°)**	**40 (16°)**	**60 (24°)**	**90 (36°)**
1	2.2	4.4	8.5	16.3	23.4	33
2	4.3	8.4	16.2	29.4	40.8	54.4
222	8.2	15.7	29	48.9	63.5	77.8
4	8.4	16.1	29.6	50.3	64.9	79.2
422	15.1	28	48.4	72.6	85.5	94.2
23	22.7	40.2	64	86.2	94.6	98.6
432	37.6	60.9	83.8	96.8	99.2	99.8
3	6.5	12.6	23.5	41.3	54.8	69.5
321	12.3	22.8	40.5	63.5	77.8	89.4
6	12.5	23.1	41	64.7	78.9	90.2
622	22.4	39.1	62.6	84.6	93.5	98.1

Table 7.6. Data from randomly oriented images of different point group virus crystals

The table contains data from Fry et al., 1999, and displays percentage completeness to 3 Å resolution, for unique subset for each point group. Data were merged, duplicates removed, and the percentage completeness calculated. It is possible to see from this that almost complete data can be produced from 30° of crystal oscillation, equating to 90 images at an oscillation range of 0.4° (Fry et al., 1999).

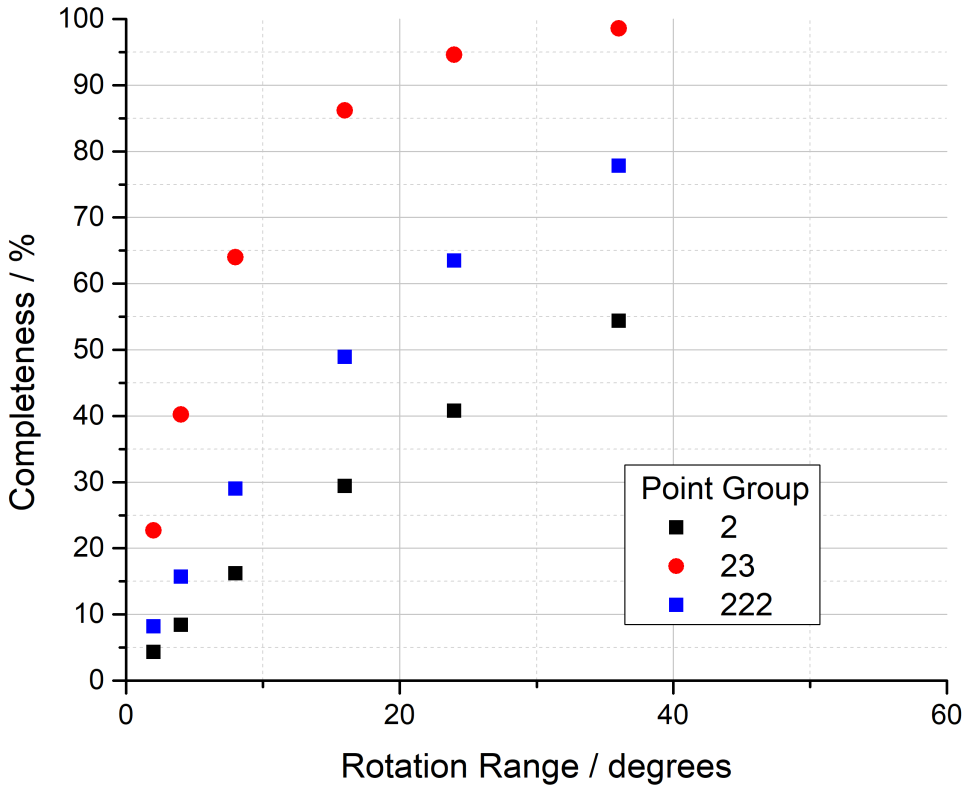


Figure 7.13. Plot of point group 2, 23, and 222 completeness data from randomly oriented images (Fry et al., 1999).

Plot of data from table 7.6 collected by Fry et al., 1999 shows the rotation range necessary to collect complete data in point groups 2, 23, and 222, in degrees. Almost double the amount of data is needed to collect complete data in point group 2 compare to point group 23. The point groups chosen for this plot correspond to AcNiR (23) and DtpAa (2; chapter 5, 6).

For small subsets, or numbers of serial diffraction images to provide complete data crystals must be randomly orientated when presented to the X-ray beam, in order to fully sample the complete reciprocal lattice. By using *ViewHKL* in the *CCP4i* software package (Krissinel & Evans, 2012) it is possible to make a comparison between *hk0* intensity-weighted reciprocal lattices of SSX and SFX subsets 400-100. By running the mtzs from these subsets though *ViewHKL* it can be seen that although intensities become more sparse as the number of images is reduced, notably toward the edge of the plot as a function of resolution, the symmetry remains the same within SSX subsets and SFX subsets (fig. 7.14). If crystals were to assume a preferred orientation plots would not be symmetrical, as the same areas of the reciprocal lattice would be repeatedly sampled leading to groupings of intensities in random sections of the plot (Axford et al., 2014).

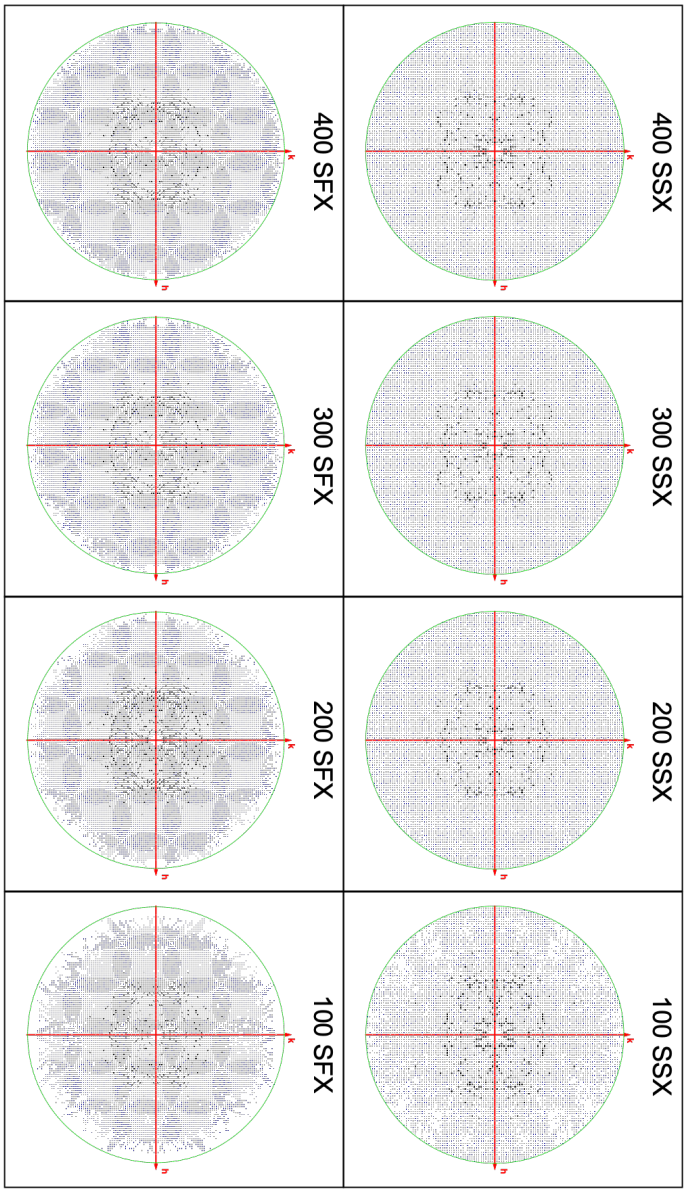


Figure 7.14. Reciprocal lattice plots for the smallest SFX and SSX subsets.

hk0 intensity-weighted reciprocal lattices of SSX and SFX subsets 400-100 indicate preferred orientation of crystals is not apparent in the lowest crystal subsets, shown via symmetrical HK0 indices throughout the plots. Preferred orientation would lead to highly-intense non-symmetrical regions within the plots as a result of repeated collection from the same reciprocal lattice points. Indices can be seen to become less abundant as the datasets are cut to fewer images. The green border in SFX plots represents the resolution limit of 2.11 Å, while the same green border in SSX plots depicts the 1.7 Å resolution limit. Indices become less frequent at higher resolution as a function of dataset size more noticeably in SFX data. Images were generated by `ViewHKL` *(CCP4; Krissinel & Evans, 2012).*

Importantly, this method is adaptable to both synchrotron and XFEL beamtime, meaning ligand binding experiments such as those explained in this chapter can be performed at synchrotrons as a test- or proof-of-theory experiment. These results can then go on to support applications for XFEL experimental beamtime, which can be difficult to obtain but is advantageous due to the absence of secondary radiation damage effects at room temperature. Difficulties arose in this experiment however when processing cbf files produced at Diamond in *CrystFEL*, a program that is predominantly designed to handle hdf5 files. SSX data that could have been indexed and merged to a higher resolution, shown in $CC_{1/2}$ and R_{split} statistics presented in table 7.5 remaining favourable despite no resolution cut being applied. This was not the case in the merging of SFX subsets, with resolution cuts having to be applied (2.11 Å in SFX subset 1000; 1.70 Å in SFX subset 1000). This was due to the current inability of *CrystFEL* to mask the highly intense silicon diffraction spots located in each diffraction pattern at ~1.69 Å (fig. 7.5), leading to fluctuations in merging statistics throughout resolution bins indicating that data in both high- and low-resolution shells were unreliable (appendices 7.5.3). A repeat of this study using DIALS would therefore be interesting as a comparison, allowing insight into the role resolution has in confidently auto-finding ligands in low numbers of diffraction patterns.

More broadly, our data clearly show that substantial information content is present in noisy and apparently low-quality data sets derived from small numbers of merged diffraction patterns with very poor merging and refinement statistics. Importantly, refinement of data subsets against native structures unambiguously showed clear density for ligands, providing definite evidence that positive electron density

described is not owing to model bias from prior knowledge of the binding mode. An interesting follow-up experiment would be using PolyPico ADE technology to load chips, as discussed in Davy et al., 2019. This could allow specified numbers of crystals to be loaded directly onto chips, further decreasing the volume of crystal slurry needed for an XFEL experiment. Despite the extremely low numbers of crystals used to confidently find ligands in SFX and SSX subsets, SFX and monochromatic SSX still require many more crystals than might be required in a wide-bandpass Laue experiment, where datasets can be produced from ~50 crystals using polychromatic Laue radiation (Meents et al., 2017). An attractive experiment would therefore be repeating the experiment performed in this chapter using Laue radiation. In doing this it would be possible to discover how few crystals are needed to identify bound ligands in data from light sources that provide much wider comparative bandpasses to XFELs and Diamond beamline I24.

7.5. Appendices

7.5.1. SACLA counter script

```
import re
import subprocess
import sys
import argparse

def print_flush(string):
    sys.stdout.write('\r%s' % string)
    sys.stdout.flush()

def file_len(fname):
    p=subprocess.Popen(['wc',       '-l',       fname],       stdout=subprocess.PIPE,
stderr=subprocess.PIPE)
    result,err = p.communicate()
    if p.returncode:
       raise IOError(err)
    return int(result.strip().split()[0])

def argparser(argv=None):
    parser = argparse.ArgumentParser(description = 'Unit Cell Counter\n\
                                        Creates .dat file contianing the name/path and
cctbx data of\
                                        each int file in a set of directories.\n\
                                        EXAMPLE\n               ./unit_cell_counter.py
dir=/dls/i24/data/2017/nt14493-63/ \
                                    p_name=CuNIR chip_name=waylinCD proctype=auto\n
')
    parser.add_argument("-i","--input_file", type=str, required=True,
                            help="input    file    e.g.    /dls/x02-1/data/2017/mx15722-
8/processing/acnir/ali/BCQR/BCQR_reindexed.out", default=None)
    parser.add_argument("-o","--output_file", type=str,
                            help="output    file    e.g.    /dls/x02-1/data/2017/mx15722-
8/processing/acnir/ali/BCQR/BCQR_reindexed.out", default='subset.out')
    parser.add_argument("-n","--number_of_crystals", type=int,
                    help="number of crystals in subset", default=1000)
    parser.add_argument("-s","--splitter_size", type=int,
                    help="number of crystals in subset", default=1000)
    parser.add_argument("-sf","--splitter_flag", type=bool, default=False,
                    help="split dataset into subsets of a set size")
    parser.add_argument("-c","--counter", type=bool,
                        help="count number of crystals in file", default=False)
    parser.add_argument("-v", "--verbosity", type=int, choices=[0, 1, 2],
                            help="increase output verbosity")
    args = parser.parse_args()
    return args

#----- Begin geometry file -----
geometry_file_start=re.compile(r'-----\s*Begin\s*geometry\s*file\s*-----')
geometry_file_end=re.compile(r'-----\s*End\s*geometry\s*file\s*-----')
unit_cell_start=re.compile(r'-----\s*Begin\s*unit\s*cell -----')
unit_cell_end=re.compile(r'-----\s*End\s*unit\s*cell -----')
crystal_start=re.compile(r'---\s*Begin\s*crystal')
crystal_end=re.compile(r'---\s*End\s*crystal')
p=                                                                  re.compile(r'Cell
parameters\s*(\d+\.\d+)\s*(\d+\.\d+)\s*(\d+\.\d+)\s*nm,\s*(\d+\.\d+)\s*(\d+\.\d+)\s*(\
d+\.\d+)\s*deg')

def check_line(line, out_list, counter):
    cryst_search = crystal_end.search(line)
    if cryst_search:
        counter +=1
    out_list.append(line)
    return out_list, counter

def crystal_counter(args):
    input_file_name=args.input_file
    out_list = []
    counter = 0
    with open(input_file_name, 'r') as input_file:
        print('seraching for crystals')
        for line in input_file:
```

```
            out_list, counter = check_line(line, out_list, counter)
            s = 'number of crystals = %04d' % (counter)
            print_flush(s)
    print('number_of_crystals = %04d'%counter)


def file_writer(out_list,output_file_name, meta_list=None):
    print('outputting data')
    with open(output_file_name, 'w') as out_file:
        if meta_list:
           for item in meta_list:
               out_file.write("%s" % item)
        for item in out_list:
           out_file.write("%s" % item)

def dataset_splitter(args):

    number_of_crystals = args.number_of_crystals
    subset_size = args.splitter_size
    input_file_name=args.input_file
    output_file_name=args.output_file
    counter = 0
    out_list = []
    meta_list = []
    i=0
    with open(input_file_name, 'r') as input_file:
        print('searching for crystals')
        for line in input_file:
            line_check = line
            meta_list.append(line)
            end = unit_cell_end.search(line_check)
            if end:
               break
        for line in input_file:
            out_list, counter = check_line(line, out_list, counter)
            s   =   '%04d:%04d   %02d%%      '   %   (i+counter,   number_of_crystals,
100*float(i+counter)/number_of_crystals)
            print_flush(s)
            if  counter == subset_size:
               file_name=str(i)+'-'+str(i+counter)+output_file_name
               file_writer(out_list, file_name, meta_list=meta_list)
               i +=counter
               if i == number_of_crystals:
                  break
               counter = 0
               out_list = []

def main(args):
    number_of_crystals = args.number_of_crystals
    input_file_name=args.input_file
    output_file_name=args.output_file
    counter = 0
    out_list = []

    with open(input_file_name, 'r') as input_file:
        print('seraching for crystals')
        for line in input_file:
            out_list, counter = check_line(line, out_list, counter)
            s   =   '%04d:%04d   %02d%%      '   %   (counter,   number_of_crystals,
100*float(counter)/number_of_crystals)
            print_flush(s)
            if counter == number_of_crystals:
               break

    print('outputting data')
    with open(output_file_name, 'w') as out_file:
        for item in out_list:
           out_file.write("%s" % item)

if __name__ == '__main__':
   args=argparser()
   if args.counter:
      crystal_counter(args)
   elif args.splitter_flag:
      dataset_splitter(args)
   else:
       main(args)
```

7.5.2. Pilatus detector geometry

```
photon_energy = 12800
adu_per_eV = 0.0001
clen = 0.310
coffset = -0.000640
res = 5814.0  ; 172 micron pixel size

0/min_fs = 0
0/max_fs = 2462
0/min_ss = 0
0/max_ss = 2526
0/corner_x = -1256.22
0/corner_y = -1318.18
0/fs = x
0/ss = y

bad_beamstop/min_x = -50
bad_beamstop/max_x = 50
bad_beamstop/min_y = -50
bad_beamstop/max_y = 50

bad_v1/min_fs = 487
bad_v1/max_fs = 493
bad_v1/min_ss = 0
bad_v1/max_ss = 2526

bad_v2/min_fs = 981
bad_v2/max_fs = 987
bad_v2/min_ss = 0
bad_v2/max_ss = 2526

bad_v3/min_fs = 1475
bad_v3/max_fs = 1481
bad_v3/min_ss = 0
bad_v3/max_ss = 2526

bad_v4/min_fs = 1969
bad_v4/max_fs = 1975
bad_v4/min_ss = 0
bad_v4/max_ss = 2526

bad_h1/min_ss = 195
bad_h1/max_ss = 211
bad_h1/min_fs = 0
bad_h1/max_fs = 2462

bad_h2/min_ss = 407
bad_h2/max_ss = 423
bad_h2/min_fs = 0
bad_h2/max_fs = 2462

bad_h3/min_ss = 619
bad_h3/max_ss = 635
bad_h3/min_fs = 0
bad_h3/max_fs = 2462

bad_h4/min_ss = 831
bad_h4/max_ss = 847
bad_h4/min_fs = 0
bad_h4/max_fs = 2462

bad_h5/min_ss = 1043
bad_h5/max_ss = 1059
bad_h5/min_fs = 0
bad_h5/max_fs = 2462

bad_h6/min_ss = 1255
bad_h6/max_ss = 1271
bad_h6/min_fs = 0
bad_h6/max_fs = 2462

bad_h7/min_ss = 1467
bad_h7/max_ss = 1483
bad_h7/min_fs = 0
bad_h7/max_fs = 2462
```

```
bad_h8/min_ss = 1679
bad_h8/max_ss = 1695
bad_h8/min_fs = 0
bad_h8/max_fs = 2462

bad_h9/min_ss = 1891
bad_h9/max_ss = 1907
bad_h9/min_fs = 0
bad_h9/max_fs = 2462

bad_ha/min_ss = 2103
bad_ha/max_ss = 2119
bad_ha/min_fs = 0
bad_ha/max_fs = 2462

bad_hb/min_ss = 2315
bad_hb/max_ss = 2331
bad_hb/min_fs = 0
bad_hb/max_fs = 2462
```

7.5.3. SSX silicon spot statistics

1/d centre	CC	nref	d / A	Min 1/nm	Max 1/nm
1.375	0.9618779	2548	7.27	0.293	2.457
2.776	0.9613766	2452	3.60	2.457	3.095
3.319	0.9628433	2433	3.01	3.095	3.543
3.721	0.9573324	2398	2.69	3.543	3.899
4.050	0.9514235	2422	2.47	3.899	4.200
4.332	0.9547936	2398	2.31	4.200	4.463
4.581	0.9478593	2393	2.18	4.463	4.698
4.805	0.9353092	2372	2.08	4.698	4.912
5.011	0.9289672	2404	2.00	4.912	5.109
5.200	0.9244878	2386	1.92	5.109	5.291
5.377	0.8971831	2399	1.86	5.291	5.462
5.543	0.8476797	2383	1.80	5.462	5.623
5.699	0.8106871	2392	1.75	5.623	5.775
5.847	0.7918163	2362	1.71	5.775	5.919
5.988	0.4943828	2380	1.67	5.919	6.057
6.123	-0.0094089	2358	1.63	6.057	6.189
6.252	0.5424889	2396	1.60	6.189	6.315
6.376	0.4233668	2370	1.57	6.315	6.437
6.495	0.3339694	2386	1.54	6.437	6.554
6.610	0.2322363	2349	1.51	6.554	6.667

1/d centre	Rsplit/%	nref	d / A	Min 1/nm	Max 1/nm
1.375	9.85	2548	7.27	0.293	2.457
2.776	9.71	2452	3.60	2.457	3.095
3.319	10.08	2433	3.01	3.095	3.543
3.721	10.75	2398	2.69	3.543	3.899
4.050	11.34	2422	2.47	3.899	4.200
4.332	11.32	2398	2.31	4.200	4.463
4.581	12.29	2393	2.18	4.463	4.698
4.805	13.73	2372	2.08	4.698	4.912
5.011	14.54	2404	2.00	4.912	5.109
5.200	16.06	2386	1.92	5.109	5.291
5.377	18.65	2399	1.86	5.291	5.462
5.543	22.67	2383	1.80	5.462	5.623
5.699	26.98	2392	1.75	5.623	5.775
5.847	30.33	2362	1.71	5.775	5.919
5.988	51.08	2380	1.67	5.919	6.057
6.123	1895.53	2358	1.63	6.057	6.189
6.252	58.63	2396	1.60	6.189	6.315
6.376	64.16	2370	1.57	6.315	6.437
6.495	73.93	2386	1.54	6.437	6.554
6.610	97.58	2349	1.51	6.554	6.667

Fluctuations in $CC_{1/2}$ and R_{split} statistics can be seen in shells 1.67 and 1.63 Å due to the highly intense silicon diffraction spots. These could not be masked therefore data were cut at 1.7 Å.

Chapter 8: Closing remarks

In this thesis, we have demonstrated the capability of SSX and SFX methods to obtain room-temperature dose-dependent structures using microcrystals, in a time and sample efficient manner. The measurement of room-temperature data is of high value due to these structures presenting more accurate structural conformations to what would be expected in a biological context. In this we have been able to develop and publish a new data collection method using chips, coined multiple serial structures (MSS). MSS has allowed us to explore how our target proteins AcNiR and DtpAa behave under accumulated X-ray irradiation, by taking snapshots of the structure at dose intervals. Interestingly, in chapter 4, although it the catalytic cycle of AcNiR was the original focus of the experiment, we were able to show how MSS can be used to identify polymorphic unit cells of batch grown microcrystals. The data processing methods explored in this thesis permitted the separation of polymorphic crystal populations; we were able to show an improvement in data quality, indicated by an improvement in crystallographic data metrics. This is in contrast to single crystal methods, where polymorphs may only be visible in weak electron density and/or dual conformations.

As shown in chapter 5, a major advantage of the chip data collection method is that it can be used at either synchrotrons or XFELs, allowing near identical experimental conditions where the only difference is the X-ray source. In this, we have shown that by comparing the MSS technique to data collected from chips at the SACLA XFEL in Japan, MSS can be used to extract functionally relevant features of damage free SFX

structures. The scarcity of XFEL beamtime means using MSS may be used as estimation in lieu of XFEL beamtime, with the potential for using high value results to enhance XFEL experimental beamtime applications.

The diversity of experiments that become available when using chips and an MSS method can also be shown in this thesis. In chapter 6 a new technique coined ‘dark progression series’ (DPS) allowed us to interrogate questions raised at the end of chapter 5, namely whether the elongation of the Fe-O bond in monomer A of DtpAa extends as a function of accumulated dose, time, or both. By developing the DPS method, we were able to demonstrate that at long enough time intervals this bond length ‘recovers’ to a distance seen in low dose DtpAa structures, indicating that the elongation is dose dependent, and previously unexplored in peroxidases. The further development of this technique is of high interest, as by shortening the time intervals between DPS structures we would be able to discover at what point the bond length no longer recovers. By coupling this with online UV-vis microspectrophotometry we hope to relate what we see in our DPS data collections to peroxidase active site chemistry.

A major advantage of these serial methods is the high redundancy of crystallographic data. A high redundancy is prerequisite in serial crystallography, as in order to gain high quality crystallographic data metrics, many diffraction patterns are needed due to the partiality of still data. Despite this, we have been able to show in chapter 7 that only a fraction of the data collected in a chip experiment at synchrotrons and XFELs is sufficient for unambiguous ligand identification. This demonstrates the feasibility of

high-throughput structure determination of protein-ligand complexes at XFELs, and sets up an exciting future for XFEL based research into fragment-based drug design.

Although chapters 4, 5, and 6 establish new and exciting techniques, there are many caveats that apply when measuring bond distances and/or determining catalytic intermediate structures using (serial) crystallography, meaning care and attention must be paid during data collection, processing, refinement, and interpretation. For example, bond length may be influenced by the experimenters chosen refinement strategy, such as using simulated annealing refinement only versus a typical maximum-likelihood refinement, or whether specific atoms have been restrained during the refinement process. Further, the number of serial diffraction images used can also influence the interpretation of electron density; as redundancy decreases and accumulated dose increases throughout an MSS series, electron density maps increase in noise and present statistics that are more favourable towards poorer resolution shells (see section 5.3.3, discussed in section 5.4). Incorporating methods that accurately measure bond length such as EXAFS and XANES would be very beneficial to assess the accuracy of the data presented in this thesis. Further, investigating how bond lengths are interpreted when different refinement strategies are applied would be extremely beneficial to the field, and could potentially be developed using the data presented in this thesis.

In summary, we have shown that chip techniques involving microcrystals are extremely versatile, and the utilisation of serial crystallography methods are pushing the boundaries of macromolecular crystallography. The techniques demonstrated in

this thesis make a valuable addition to the crystallographer's toolbox, with the future of macromolecular crystallography extremely bright.

Bibliography

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., … Zwart, P. H. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, *66*(2), 213–221. https://doi.org/10.1107/S0907444909052925

Adams, P. D., Pannu, N. S., Read, R. J., & Brünger, A. T. (1997). Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(10), 5018–5023. https://doi.org/10.1073/pnas.94.10.5018

Anderson, A. C. (2003). The process of structure-based drug design. *Chemistry and Biology*. https://doi.org/10.1016/j.chembiol.2003.09.002

Antonyuk, S. V., Strange, R. W., Sawers, G., Eady, R. R., & Hasnain, S. S. (2005). Atomic resolution structures of resting-state, substrate- and product-complexed Cu-nitrite reductase provide insight into catalytic mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(34), 12041–12046. https://doi.org/10.1073/pnas.0504207102

Aquila, A., Hunter, M. S., Doak, R. B., Kirian, R. A., Fromme, P., White, T. A., … Chapman, H. N. (2012). Time-resolved protein nanocrystallography using an X-ray free-electron laser. *Optics Express*, *20*(3), 2706. https://doi.org/10.1364/oe.20.002706

Axford, D., Ji, X., Stuart, D. I., & Sutton, G. (2014). In cellulo structure determination of a novel cypovirus polyhedrin. *Acta Crystallographica Section D: Biological Crystallography*, *70*(5), 1435–1441. https://doi.org/10.1107/S1399004714004714

Barty, A., Kirian, R. A., Maia, F. R. N. C., Hantke, M., Yoon, C. H., White, T. A., & Chapman, H. (2014). Cheetah: Software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *Journal of Applied Crystallography*, *47*(3), 1118–1131. https://doi.org/10.1107/S1600576714007626

Baxter, E. L., Aguila, L., Alonso-Mori, R., Barnes, C. O., Bonagura, C. A., Brehmer, W., … Cohen, A. E. (2016). High-density grids for efficient data collection from multiple crystals. *Acta Crystallographica. Section D, Structural Biology*, *72*(Pt 1), 2–11. https://doi.org/10.1107/S2059798315020847

Beale, J. H., Bolton, R., Marshall, S. A., Beale, E. V., Carr, S. B., Ebrahim, A., … Owen, R. L. (2019). Successful sample preparation for serial crystallography experiments. *Journal of Applied Crystallography*, *52*, 1385–1396. https://doi.org/10.1107/S1600576719013517

Beitlich, T., Kühnel, K., Schulze-Briese, C., Shoeman, R. L., & Schlichting, I. (2007). Cryoradiolytic reduction of crystalline heme proteins: analysis by UV-Vis spectroscopy and X-ray crystallography. *Journal of Synchrotron Radiation*, *14*(1), 11–23. https://doi.org/10.1107/S0909049506049806

Berglund, G. I., Carlsson, G. H., Smith, A. T., Szöke, H., Henriksen, A., & Hajdu, J. (2002). The catalytic pathway of horseradish peroxidase at high resolution. *Nature*, *417*(6887), 463–468. https://doi.org/10.1038/417463a

Bianconi, A., Dell'Ariccia, M., Gargano, A., & Natoli, C. (1983). Bond Length

Determination Using XANES, 57–61. https://doi.org/10.1007/978-3-642-50098-5_8

Blake, C. C. F., Fenn, R. H., Johnson, L. N., Koenig, D. F., Mair, G. A., North, A. C. T., … Vernon, C. A. (2012). PART 25 . A HISTORICAL PERSPECTIVE Chapter 25 . 1 . How the structure of lysozyme was actually determined, 845–872.

Blake, C. C. F., Koneing, D. F., Mair, G. A., North, A. C. T., Phillips, D. C., & Sarma, V. R. (1965). Structure of Hen Egg-White Lysozyme: A Three-dimensional Fourier Synthesis at 2 Å Resolution. *Nature*, *206*(4986), 757–761. https://doi.org/10.1038/206757a0

Bonnefond, L., Schellenberger, P., Basquin, J., Demangeat, G., Ritzenthaler, C., Chênevert, R., … Sauter, C. (2011). Exploiting protein engineering and crystal polymorphism for successful X-ray structure determination. *Crystal Growth and Design*, *11*(10), 4334–4343. https://doi.org/10.1021/cg101468p

Borek, D., Minor, W., & Otwinowski, Z. (2003). Measurement errors and their consequences in protein crystallography. In *Acta Crystallographica - Section D Biological Crystallography* (Vol. 59, pp. 2031–2038). International Union of Crystallography. https://doi.org/10.1107/S0907444903020924

Bourhis, L. J., Dolomanov, O. V., Gildea, R. J., Howard, J. A. K., & Puschmann, H. (2015). The anatomy of a comprehensive constrained, restrained refinement program for the modern computing environment - Olex2 dissected. *Acta Crystallographica Section A: Foundations of Crystallography*, *71*(1), 59–75. https://doi.org/10.1107/S2053273314022207

Bowman, S. E. J., Bridwell-Rabb, J., & Drennan, C. L. (2016). Metalloprotein Crystallography: More than a Structure. *Accounts of Chemical Research*, *49*(4), 695–702. https://doi.org/10.1021/acs.accounts.5b00538

Bublitz, M., Nass, K., Drachmann, N. D., Markvardsen, A. J., Gutmann, M. J., Barends, T. R. M., … Schlichting, I. (2015). Structural studies of P-type ATPase-ligand complexes using an X-ray free-electron laser. *IUCrJ*, *2*, 409–420. https://doi.org/10.1107/S2052252515008969

Casanas, A., Warshamanage, R., Finke, A. D., Panepucci, E., Olieric, V., Nöll, A., … Wang, M. (2016). EIGER detector: Application in macromolecular crystallography: *Acta Crystallographica Section D: Structural Biology*, *72*(9), 1036–1048. https://doi.org/10.1107/S2059798316012304

Chapman, H. N., Caleman, C., & Timneanu, N. (2014). Diffraction before destruction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1647), 20130313–20130313. https://doi.org/10.1098/rstb.2013.0313

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., … Richardson, D. C. (2010). MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, *66*(1), 12–21. https://doi.org/10.1107/S0907444909042073

Chreifi, G., Baxter, E. L., Doukov, T., Cohen, A. E., McPhillips, S. E., Song, J., … Poulos, T. L. (2016). Crystal structure of the pristine peroxidase ferryl center and its relevance to proton-coupled electron transfer. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(5), 1226–1231. https://doi.org/10.1073/pnas.1521664113

Collins, P. M., Douangamath, A., Talon, R., Dias, A., Brandão-Neto, J., Krojer, T., & Delft, F. von. (2018). Achieving a Good Crystal System for Crystallographic X-ray

Fragment Screening, *610*, 251–264. https://doi.org/10.20944/PREPRINTS201809.0383.V1

Colpa, D. I., Fraaije, M. W., & Van Bloois, E. (2014). DyP-type peroxidases: A promising and versatile class of enzymes. *Journal of Industrial Microbiology and Biotechnology*, *41*(1), 1–7. https://doi.org/10.1007/s10295-013-1371-6

Crick, F. H. C., & Magdoff, B. S. (1956). The theory of the method of isomorphous replacement for protein crystals. I. *Acta Crystallographica*, *9*(11), 901–908. https://doi.org/10.1107/s0365110x56002552

Davies, M., Hawkins, C., Pattison, D., & Rees, M. (2008). Mammalian Heme Peroxidases: From Molecular Mechanisms to Health Implications. *Antioxidants & Redox Signaling*, *10*(7), 1199–1234. https://doi.org/10.1089/ars.2007.1927

Davies, T. G., & Tickle, I. J. (2012). Fragment screening using X-ray crystallography. *Topics in Current Chemistry*, *317*, 33–59. https://doi.org/10.1007/128-2011-179

Davy, B., Axford, D., Beale, J. H., Butryn, A., Docker, P., Ebrahim, A., … Owen, L. (2019). short communications Reducing sample consumption for serial crystallography using acoustic drop ejection short communications, 1820–1825. https://doi.org/10.1107/S1600577519009329

Delbart, F., Brams, M., Gruss, F., Noppen, S., Peigneur, S., Boland, S., … Ulens, C. (2018). An allosteric binding site of the α7 nicotinic acetylcholine receptor revealed in a humanized acetylcholine-binding protein. *Journal of Biological Chemistry*, *293*(7), 2534–2545. https://doi.org/10.1074/jbc.M117.815316

DePonte, D. P., Doak, R. B., Hunter, M., Liu, Z., Weierstall, U., & Spence, J. C. H. (2009). SEM imaging of liquid jets. *Micron*, *40*(4), 507–509. https://doi.org/10.1016/j.micron.2008.12.009

Ebrahim, A., Appleby, M. V., Axford, D., Beale, J., Moreno-Chicano, T., Sherrell, D. A., … Owen, R. L. (2019). Resolving polymorphs and radiation-driven effects in microcrystals using fixed-target serial synchrotron crystallography. *Acta Crystallographica Section D: Structural Biology*, *75*, 151–159. https://doi.org/10.1107/S2059798318010240

Ebrahim, A., Moreno-Chicano, T., Appleby, M. V., Chaplin, A. K., Beale, J. H., Sherrell, D. A., … Hough, M. A. (2019). Dose-resolved serial synchrotron and XFEL structures of radiation-sensitive metalloproteins. *IUCrJ*, *6*, 543–551. https://doi.org/10.1107/S2052252519003956

Ehhalt, D., Prather, M., & Dentener, F. (2001). Atmospheric chemistry and greenhouse gases. In *IPCC Third Assessment Report Climate Change 2001: The Scientific Basis* (pp. 239–288). Retrieved from http://www.osti.gov/scitech/biblio/901482

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, *66*(4), 486–501. https://doi.org/10.1107/S0907444910007493

Emsley, Paul, & Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, *60*(12 I), 2126–2132. https://doi.org/10.1107/S0907444904019158

Evans, P. R. (2011). An introduction to data reduction: Space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D: Biological Crystallography*, *67*(4), 282–292. https://doi.org/10.1107/S090744491003982X

Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography*, *69*(7),

1204–1214. https://doi.org/10.1107/S0907444913000061
Fischer, M., Shoichet, B. K., & Fraser, J. S. (2015). One Crystal, Two Temperatures: Cryocooling Penalties Alter Ligand Binding to Transient Protein Sites. *ChemBioChem*, *16*(11), 1560–1564. https://doi.org/10.1002/cbic.201500196
Flot, D., Mairs, T., Giraud, T., Guijarro, M., Lesourd, M., Rey, V., … Mitchell, E. (2010). The ID23-2 structural biology microfocus beamline at the ESRF. *Journal of Synchrotron Radiation*, *17*(1), 107–118. https://doi.org/10.1107/S0909049509041168
Frank, M., Carlson, D. B., Hunter, M. S., Williams, G. J., Messerschmidt, M., Zatsepin, N. A., … Evans, J. E. (2014). Femtosecond X-ray diffraction from two-dimensional protein crystals. *IUCrJ*, *1*, 95–100. https://doi.org/10.1107/S2052252514001444
Fraser, J. S., van den Bedem, H., Samelson, A. J., Lang, P. T., Holton, J. M., Echols, N., & Alber, T. (2011). Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proceedings of the National Academy of Sciences*, *108*(39), 16247–16252. https://doi.org/10.1073/pnas.1111325108
Fry, E. E., Grimes, J., & Stuart, D. I. (1999). Virus Crystallography, *12*.
Garman, E. F. (2010). Radiation damage in macromolecular crystallography: What is it and why do we care? *Acta Crystallographica Section D*, 69–77. https://doi.org/10.1007/978-94-007-6232-9-7
Garman, E. F., & Owen, R. L. (2006). Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, *62*(1), 32–47. https://doi.org/10.1107/S0907444905034207
Garman, E. F., & Weik, M. (2015a). Radiation damage to macromolecules: Kill or cure? *Journal of Synchrotron Radiation*, *22*(2), 195–200. https://doi.org/10.1107/S160057751500380X
Garman, E. F., & Weik, M. (2015b). Radiation damage to macromolecules: Kill or cure? *Journal of Synchrotron Radiation*, *22*(2), 195–200. https://doi.org/10.1107/S160057751500380X
Gisriel, C., Coe, J., Letrun, R., Yefanov, O. M., Luna-Chavez, C., Stander, N. E., … Zatsepin, N. A. (2019). Membrane protein megahertz crystallography at the European XFEL. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-12955-3
Hajdu, J., Neutze, R., Sjögren, T., Edman, K., Szöke, A., Wilmouth, R. C., & Wilmot, C. M. (2000). Analyzing protein functions in four dimensions. *Nature Structural Biology*, *7*(11), 1006–1012. https://doi.org/10.1038/80911
Hardcastle, F. D., & Wachs, I. E. (1991). Determination of niobium-oxygen bond distances and bond orders by Raman spectroscopy. *Solid State Ionics*, *45*(3–4), 201–213. https://doi.org/10.1016/0167-2738(91)90153-3
Harding, R. J., Ferreira De Freitas, R., Collins, P., Franzoni, I., Ravichandran, M., Ouyang, H., … Arrowsmith, C. H. (2017). Small Molecule Antagonists of the Interaction between the Histone Deacetylase 6 Zinc-Finger Domain and Ubiquitin. *Journal of Medicinal Chemistry*, *60*(21), 9090–9096. https://doi.org/10.1021/acs.jmedchem.7b00933
Hennig, C., Reich, T., Funke, H., Rossberg, A., Rutsch, M., & Bernhard, G. (2001). EXAFS as a tool for bond-length determination in the environment of heavy atomso. *Journal of Synchrotron Radiation*, *8*(2), 695–697. https://doi.org/10.1107/S0909049500016186

Heymann, M., Opthalage, A., Wierman, J. L., Akella, S., Szebenyi, D. M. E., Gruner, S. M., & Fraden, S. (2014). Room-temperature serial crystallography using a kinetically optimized microfluidic device for protein crystallization and on-chip X-ray diffraction. *IUCrJ*, *1*(2014), 349–360. https://doi.org/10.1107/S2052252514016960

Hiroki, A., Pimblott, S. M., & Laverne, J. A. (2002). Hydrogen peroxide production in the radiolysis of water with high radical scavenger concentrations. *Journal of Physical Chemistry A*, *106*(40), 9352–9358. https://doi.org/10.1021/jp0207578

Holton, J. M. (2009). A beginner's guide to radiation damage. *Journal of Synchrotron Radiation*, *16*(2), 133–142. https://doi.org/10.1107/S0909049509004361

Horrell, S., Antonyuk, S., Eady, R., Hasnain, S., Hough, M., & Strange, R. (2016). Serial crystallography captures enzyme catalysis in copper nitrite reductase at atomic resolution from one crystal. *IUCrJ*, *3*, 271–281. https://doi.org/10.1107/S205225251600823X

Horrell, S., Antonyuk, S. V., Eady, R. R., Hasnain, S. S., Hough, M. A., Strange, R. W., … G., Z. W. (2016). Serial crystallography captures enzyme catalysis in copper nitrite reductase at atomic resolution from one crystal. *IUCrJ*, *3*(4), 778–784. https://doi.org/10.1107/S205225251600823X

Horrell, S., Kekilli, D., Sen, K., Owen, R. L., Dworkowski, F. S. N., Antonyuk, S. V, … Hough, M. A. (2018). Enzyme catalysis captured using multiple structures from one crystal at varying temperatures. *IUCrJ*, *5*(3), 283–292. https://doi.org/10.1107/S205225251800386X

Hough, M., Antonyuk, S., Strange, R., Eady, R., & Hasnain, S. (2008). Crystallography with Online Optical and X-ray Absorption Spectroscopies Demonstrates an Ordered Mechanism in Copper Nitrite Reductase. *Journal of Molecular Biology*, *378*(2), 353–361. https://doi.org/10.1016/j.jmb.2008.01.097

Huang, C. Y., Olieric, V., Ma, P., Panepucci, E., Diederichs, K., Wang, M., & Caffrey, M. (2015). In meso in situ serial X-ray crystallography of soluble and membrane proteins. *Acta Crystallographica Section D: Biological Crystallography*, *71*, 1238–1256. https://doi.org/10.1107/S1399004715005210

Hunter, M. S., Segelke, B., Messerschmidt, M., Williams, G. J., Zatsepin, N. A., Barty, A., … Frank, M. (2014). Fixed-target protein serial microcrystallography with an X-ray free electron laser. *Scientific Reports*, *4*, 1–5. https://doi.org/10.1038/srep06026

Inoue, I., Inubushi, Y., Sato, T., Tono, K., Katayama, T., Kameshima, T., … Yabashi, M. (2016). Observation of femtosecond X-ray interactions with matter using an X-ray-X-ray pump-probe scheme. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(6), 1492–1497. https://doi.org/10.1073/pnas.1516426113

Johansson, L. C., Arnlund, D., White, T. A., Katona, G., Deponte, D. P., Weierstall, U., … Neutze, R. (2012). Lipidic phase membrane protein serial femtosecond crystallography. *Nature Methods*, *9*(3), 263–265. https://doi.org/10.1038/nmeth.1867

Kameshima, T., Ono, S., Kudo, T., Ozaki, K., Kirihara, Y., Kobayashi, K., … Hatsui, T. (2014). Development of an X-ray pixel detector with multi-port charge-coupled device for X-ray free-electron laser experiments. *Review of Scientific Instruments*, *85*(3). https://doi.org/10.1063/1.4867668

Karplus, P. A., & Diederichs, K. (2012). Linking Crystallographic Model and Data Quality. *Science*, *336*(6084), 1030–1033. https://doi.org/10.1126/science.1218231

Keedy, D. A., Hill, Z. B., Biel, J. T., Kang, E., Pearce, N. M., Rettenmaier, T. J., … Delft, F. Von. (2018). An expanded allosteric network in PTP1B by multitemperature crystallography , fragment screening , and covalent tethering, 1–36.

Keedy, D. A., Hill, Z. B., Biel, J. T., Kang, E., Rettenmaier, T. J., Brandão-Neto, J., … Fraser, J. S. (2018). An expanded allosteric network in PTP1B by multitemperature crystallography, fragment screening, and covalent tethering. *ELife*, *7*, 1–36. https://doi.org/10.7554/eLife.36307

Keedy, D. A., Kenner, L. R., Warkentin, M., Woldeyes, R. A., Hopkins, J. B., Thompson, M. C., … Fraser, J. S. (2015). Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. *ELife*, *4*, 1–26. https://doi.org/10.7554/elife.07574

Kekilli, D., Dworkowski, F. S. N., Pompidor, G., Fuchs, M. R., Andrew, C. R., Antonyuk, S., … Hough, M. A. (2014). Fingerprinting redox and ligand states in haemprotein crystal structures using resonance Raman spectroscopy. *Acta Crystallographica Section D: Biological Crystallography*, *70*(5), 1289–1296. https://doi.org/10.1107/S1399004714004039

Kekilli, D., Moreno-Chicano, T., Chaplin, A. K., Horrell, S., Dworkowski, F. S. N., Worrall, J. A. R., … Hough, M. A. (2017). Photoreduction and validation of haem-ligand intermediate states in protein crystals by in situ single-crystal spectroscopy and diffraction. *IUCrJ*, *4*, 263–270. https://doi.org/10.1107/S2052252517002159

Kim, S. J. U. N., & Shoda, M. (1999). Purification and Characterization of a Novel Peroxidase from Geotrichum candidum Dec 1 Involved in Decolorization of Dyes These include : Purification and Characterization of a Novel Peroxidase from Geotrichum candidum Dec 1 Involved in Decolorization of , *65*(3), 1029–1035.

Kleywegt, G. J., & Jones, T. A. (1997). Model building and refinement practice. *Methods in Enzymology*, *277*(1995), 208–230. https://doi.org/10.1016/S0076-6879(97)77013-7

Krzemiński, Ł., Ndamba, L., Canters, G. W., Aartsma, T. J., Evans, S. D., & Jeuken, L. J. C. (2011). Spectroelectrochemical investigation of intramolecular and interfacial electron-transfer rates reveals differences between nitrite reductase at rest and during turnover. *Journal of the American Chemical Society*, *133*(38), 15085–15093. https://doi.org/10.1021/ja204891v

Kukimoto, M., Nishiyama, M., Tanokura, M., Murphy, M. E. P., Adman, E. T., & Horinouchi, S. (1996). Site-directed mutagenesis of azurin from Pseudomonas aeruginosa enhances the formation of an electron-transfer complex with a copper-containing nitrite reductase from Alcaligenes faecalis S-6. *FEBS Letters*, *394*(1), 87–90. https://doi.org/10.1016/0014-5793(96)00934-9

Kumar, K. S. D., Gurusaran, M., Satheesh, S. N., Radha, P., Pavithra, S., Thulaa Tharshan, K. P. S., … Sekar, K. (2015). Online-DPI: A web server to calculate the diffraction precision index for a protein structure. *Journal of Applied Crystallography*, *48*, 939–942. https://doi.org/10.1107/S1600576715006287

Leonarski, F., Redford, S., Mozzanica, A., Lopez-Cuenca, C., Panepucci, E., Nass, K., … Wang, M. (2018). Fast and accurate data collection for macromolecular crystallography using the JUNGFRAU detector. *Nature Methods*, *15*(10), 799–

804. https://doi.org/10.1038/s41592-018-0143-7
Levantino, M., Yorke, B. A., Monteiro, D. C. F., Cammarata, M., & Pearson, A. R. (2015). ScienceDirect Using synchrotrons and XFELs for time-resolved X-ray crystallography and solution scattering experiments on biomolecules, 41–48.
Li, J., Liu, C., Li, B., Yuan, H., Yang, J., & Zheng, B. (2012). Identification and molecular characterization of a novel DyP-type peroxidase from Pseudomonas aeruginosa PKE117. *Applied Biochemistry and Biotechnology*, *166*(3), 774–785. https://doi.org/10.1007/s12010-011-9466-x
Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., … Adams, P. D. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology*, *75*, 861–877. https://doi.org/10.1107/S2059798319011471
Liu, J., Chakraborty, S., Hosseinzadeh, P., Yu, Y., Tian, S., Petrik, I., … Lu, Y. (2014). Metalloproteins containing cytochrome, iron-sulfur, or copper redox centers. *Chemical Reviews*, *114*(8), 4366–4369. https://doi.org/10.1021/cr400479b
Lomb, L., Barends, T. R. M., Kassemeyer, S., Aquila, A., Epp, S. W., Erk, B., … Schlichting, I. (2011). Radiation damage in protein serial femtosecond crystallography using an x-ray free-electron laser. *Physical Review B - Condensed Matter and Materials Physics*, *84*(21), 1–6. https://doi.org/10.1103/PhysRevB.84.214111
Long, F., Nicholls, R. A., Emsley, P., Gražulis, S., Merkys, A., Vaitkus, A., & Murshudov, G. N. (2017). AceDRG: A stereochemical description generator for ligands. *Acta Crystallographica Section D: Structural Biology*, *73*(2), 112–122. https://doi.org/10.1107/S2059798317000067
Lučić, M., Chaplin, A. K., Moreno-Chicano, T., Dworkowski, F. S. N., Wilson, M. T., Svistunenko, D. A., … Worrall, J. A. R. (2020). A subtle structural change in the distal haem pocket has a remarkable effect on tuning hydrogen peroxide reactivity in dye decolourising peroxidases from Streptomyces lividans. *Dalton Transactions*, *49*(5), 1620–1636. https://doi.org/10.1039/C9DT04583J
McIntyre, P. J., Collins, P. M., Vrzal, L., Birchall, K., Arnold, L. H., Mpamhanga, C., … Bayliss, R. (2017). Characterization of Three Druggable Hot-Spots in the Aurora-A/TPX2 Interaction Using Biochemical, Biophysical, and Fragment-Based Approaches. *ACS Chemical Biology*, *12*(11), 2906–2914. https://doi.org/10.1021/acschembio.7b00537
McPherson, A., & Gavira, J. A. (2014). Introduction to protein crystallization. *Acta Crystallographica Section F:Structural Biology Communications*, *70*(1), 2–20. https://doi.org/10.1107/S2053230X13033141
Meents, A., Wiedorn, M. O., Srajer, V., Henning, R., Sarrou, I., Bergtholdt, J., … Chapman, H. N. (2017). Pink-beam serial crystallography. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-01417-3
Meents, Alke, Gutmann, S., Wagner, A., & Schulze-Briese, C. (2010). Origin and temperature dependence of radiation damage in biological samples at cryogenic temperatures. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(3), 1094–1099. https://doi.org/10.1073/pnas.0905481107
Meharenna, Y., Doukov, T., Li, H., Soltis, S. M., & Poulos, T. (2010). Crystallographic and single-crystal spectral analysis of the peroxidase ferryl intermediate.

Biochemistry, *49*(14), 2984–2986. https://doi.org/10.1021/bi100238r

Melhado, E. M. (1980). Mitscherlich's Discovery of Isomorphism. *Historical Studies in the Physical Sciences*, *11*(1), 87–123. https://doi.org/10.2307/27757472

Moreno-Chicano, T., Ebrahim, A., Axford, D., Appleby, M. V., Beale, J. H., Chaplin, A. K., … Hough, M. A. (2019). High-throughput structures of protein-ligand complexes at room temperature using serial femtosecond crystallography. *IUCrJ*, *6*, 1074–1085. https://doi.org/10.1107/S2052252519011655

Mueller, C., Marx, A., Epp, S. W., Zhong, Y., Kuo, A., Balo, A. R., … Dwayne Miller, R. J. (2015). Fixed target matrix for femtosecond time-resolved and in situ serial micro-crystallography. *Structural Dynamics (Melville, N.Y.)*, *2*(5), 054302. https://doi.org/10.1063/1.4928706

Murphy, M. E. P., Turley, S., & Adman, E. T. (1997). PROTEIN CHEMISTRY AND STRUCTURE : Structure of Nitrite Bound to Copper-containing Nitrite Reductase from Alcaligenes faecalis : MECHANISTIC IMPLICATIONS Structure of Nitrite Bound to Copper-containing Nitrite Reductase from Alcaligenes faecalis, *272*(45), 28455–28460.

Murray, T. D., Lyubimov, A. Y., Ogata, C. M., Vo, H., Uervirojnangkoorn, M., Brunger, A. T., & Berger, J. M. (2015). A high-transparency, micro-patternable chip for X-ray diffraction analysis of microcrystals under native growth conditions. *Acta Crystallographica Section D: Biological Crystallography*, *71*, 1987–1997. https://doi.org/10.1107/S1399004715015011

Naitow, H., Matsuura, Y., Tono, K., Joti, Y., Kameshima, T., Hatsui, T., … Kunishima, N. (2017). Protein-ligand complex structure from serial femtosecond crystallography using soaked thermolysin microcrystals and comparison with structures from synchrotron radiation. *Acta Crystallographica Section D: Structural Biology*, *73*(8), 702–709. https://doi.org/10.1107/S2059798317008919

Nakano, S., Takahashi, M., Sakamoto, A., Morikawa, H., & Katayanagi, K. (2012). Structure-function relationship of assimilatory nitrite reductases from the leaf and root of tobacco based on high-resolution structures. *Protein Science*, *21*(3), 383–395. https://doi.org/10.1002/pro.2025

Nass, K. (2019). Radiation damage in protein crystallography at X-ray free-electron lasers research papers, 211–218. https://doi.org/10.1107/S2059798319000317

Nave, C., & Garman, E. F. (2005). Towards an understanding of radiation damage in cryocooled macromolecular crystals. *Journal of Synchrotron Radiation*, *12*(3), 257–260. https://doi.org/10.1107/S0909049505007132

Nelson, R. E., Fessler, L. I., Takagi, Y., Blumberg, B., Keene, D. R., Olson, P. F., … Fessler, J. H. (1994). Peroxidasin: a novel enzyme-matrix protein of Drosophila development. *The EMBO Journal*, *13*(15), 3438–3447. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=395246&tool=pmcentrez&rendertype=abstract

Neutze, R., Wouts, R., van der Spoel, D., Weckert, E., & Hajdu, J. (2000). Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature*, *406*(6797), 752–757. https://doi.org/10.1038/35021099

Nogly, P., James, D., Wang, D., White, T. A., Zatsepin, N., Shilova, A., … Weierstall, U. (2015). Lipidic cubic phase serial millisecond crystallography using synchrotron radiation. *IUCrJ*, *2*, 168–176. https://doi.org/10.1107/S2052252514026487

O'Brien, J. A., Daudi, A., Butt, V. S., & Bolwell, G. P. (2012). Reactive oxygen species and their role in plant defence and cell wall metabolism. *Planta*, *236*(3), 765–779. https://doi.org/10.1007/s00425-012-1696-9

Oghbaey, S., Sarracini, A., Ginn, H. M., Pare-Labrosse, O., Kuo, A., Marx, A., … Miller, R. J. D. (2016). Fixed target combined with spectral mapping: approaching 100% hit rates for serial crystallography. *Acta Crystallographica Section D Structural Biology*, *72*(8), 944–955. https://doi.org/10.1107/S2059798316010834

Owen, R. L., Axford, D., Sherrell, D. A., Kuo, A., Ernst, O. P., Schulz, E. C., … Mueller-Werkmeister, H. M. (2017). Low-dose fixed-target serial synchrotron crystallography. *Acta Crystallographica Section D Structural Biology*, *73*(4), 373–378. https://doi.org/10.1107/S2059798317002996

Owen, R. L., Holton, J. M., Schulze-Briese, C., & Garman, E. F. (2009). Determination of X-ray flux using silicon pin diodes. *Journal of Synchrotron Radiation*, *16*(2), 143–151. https://doi.org/10.1107/S0909049508040429

Owen, R. L., Juanhuix, J., & Fuchs, M. (2016). Current advances in synchrotron radiation instrumentation for MX experiments. *Archives of Biochemistry and Biophysics*, *602*, 21–31. https://doi.org/10.1016/j.abb.2016.03.021

Owen, R. L., Paterson, N., Axford, D., Aishima, J., Schulze-Briese, C., Ren, J., … Evans, G. (2014). Exploiting fast detectors to enter a new dimension in room-temperature crystallography. *Acta Crystallographica Section D: Biological Crystallography*, *70*(5), 1248–1256. https://doi.org/10.1107/S1399004714005379

Paithankar, K. S., & Garman, E. F. (2010). Know your dose: RADDOSE. *Acta Crystallographica Section D: Biological Crystallography*, *66*(4), 381–388. https://doi.org/10.1107/S0907444910006724

Patel, D., Bauman, J. D., & Arnold, E. (2014). Advantages of crystallographic fragment screening: Functional and mechanistic insights from a powerful platform for efficient drug discovery. *Progress in Biophysics and Molecular Biology*, *116*(2–3), 92–100. https://doi.org/10.1016/j.pbiomolbio.2014.08.004

Petrus, M. L. C., Vijgenboom, E., Chaplin, A. K., Worrall, J. A. R., van Wezel, G. P., & Claessen, D. (2016). The DyP-type peroxidase DtpA is a Tat-substrate required for GlxA maturation and morphogenesis in Streptomyces. *Open Biology*, *6*(1), 150149-. https://doi.org/10.1098/rsob.150149

Pflugrath, J. W. (2015). IYCr crystallization series Practical macromolecular cryocrystallography IYCr crystallization series, 622–642. https://doi.org/10.1107/S2053230X15008304

Ravelli, R. B. G., & McSweeney, S. M. (2000). The "fingerprint" that X-rays can leave on structures. *Structure*, *8*(3), 315–328. https://doi.org/10.1016/S0969-2126(00)00109-X

Rhodes, G. (2006). *Crystallography Made Crystal Clear* (3rd ed.). Academic Press.

Ringe, D., & Petsko, G. A. (2003). The 'glass transition' in protein dynamics: what it is, why it occurs, and how to exploit it. *Biophysical Chemistry*, *105*(2–3), 667–680. https://doi.org/10.1016/S0301-4622(03)00096-6

Roberts, J. N., Singh, R., Grigg, J. C., Murphy, M. E. P., Bugg, T. D. H., & Eltis, L. D. (2011). Characterization of Dye-Decolorizing Peroxidases from Rhodococcus jostii RHA1. *Biochemistry*, *50*(23), 5108–5119. https://doi.org/10.1021/bi200427h

Rupp, B. (2009). *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology* (1st ed.). Garland Science.

Sanishvili, R., Yoder, D. W., Pothineni, S. B., Rosenbaum, G., Xu, S., Vogt, S., … Fischetti, R. F. (2011). Radiation damage in protein crystals is reduced with a micron-sized X-ray beam. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(15), 6127–6132. https://doi.org/10.1073/pnas.1017701108

Schlichting, I. (2015). Serial femtosecond crystallography: The first five years. *IUCrJ*, *2*(2013), 246–255. https://doi.org/10.1107/S205225251402702X

Sherrell, D. A., Foster, A. J., Hudson, L., Nutter, B., O'Hea, J., Nelson, S., … Owen, R. L. (2015). A modular and compact portable mini-endstation for high-precision, high-speed fixed target serial crystallography at FEL and synchrotron sources. *Journal of Synchrotron Radiation*, *22*(6), 1372–1378. https://doi.org/10.1107/s1600577515016938

Smith, K. a, Mosier, A. R., Crutzen, P. J., & Winiwarter, W. (2012). The role of N2O derived from crop-based biofuels, and from agriculture in general, in Earth's climate. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1593), 1169–1174. https://doi.org/10.1098/rstb.2011.0313

Solomon, E. I. (1993). Electronic structures of active sites in copper proteins: Contributions to reactivity. *Journal of Inorganic Biochemistry*, *51*(1–2), 450. https://doi.org/10.1016/0162-0134(93)85478-q

Solomon, E. I., Sundaram, U. M., & Machonkin, T. E. (1996). Multicopper Oxidases and Oxygenases. *Chemical Reviews*, *96*(7), 2563–2606. https://doi.org/10.1021/cr950046o

Southworth-Davies, R. J., Medina, M. A., Carmichael, I., & Garman, E. F. (2007). Observation of Decreased Radiation Damage at Higher Dose Rates in Room Temperature Protein Crystallography. *Structure*, *15*(12), 1531–1541. https://doi.org/10.1016/j.str.2007.10.013

Sturm, A., Schierhorn, A., Lindenstrauss, U., Lilie, H., & Brüser, T. (2006). YcdB from Escherichia coli reveals a novel class of Tat-dependently translocated hemoproteins. *Journal of Biological Chemistry*, *281*(20), 13972–13978. https://doi.org/10.1074/jbc.M511891200

Sugahara, M., Song, C., Suzuki, M., Masuda, T., Inoue, S., Nakane, T., … Iwata, S. (2016). Oil-free hyaluronic acid matrix for serial femtosecond crystallography. *Scientific Reports*, *6*, 4–9. https://doi.org/10.1038/srep24484

Sugano, Y. (2009). DyP-type peroxidases comprise a novel heme peroxidase family. *Cellular and Molecular Life Sciences*, *66*(8), 1387–1403. https://doi.org/10.1007/s00018-008-8651-8

Sugano, Yasushi, Muramatsu, R., Ichiyanagi, A., Sato, T., & Shoda, M. (2007). DyP, a unique dye-decolorizing peroxidase, represents a novel heme peroxidase family: ASP171 replaces the distal histidine of classical peroxidases. *Journal of Biological Chemistry*, *282*(50), 36652–36658. https://doi.org/10.1074/jbc.M706996200

Sutton, K. A., Black, P. J., Mercer, K. R., Garman, E. F., Owen, R. L., Snell, E. H., & Bernhard, W. A. (2013). Insights into the mechanism of X-ray-induced disulfide-bond cleavage in lysozyme crystals based on EPR, optical absorption and X-ray diffraction studies. *Acta Crystallographica Section D: Biological Crystallography*,

69(12), 2381–2394. https://doi.org/10.1107/S0907444913022117

Sygusch, J., & Allaire, M. (1988). Sequential radiation damage in protein crystallography. *Acta Crystallographica Section A*, *44*(4), 443–448. https://doi.org/10.1107/S0108767388001394

Taberman, H. (2018). Radiation damage in macromolecular crystallography —An experimentalist's view. *Crystals*, *8*(4), 5–9. https://doi.org/10.3390/cryst8040157

Uervirojnangkoorn, M., Zeldin, O. B., Lyubimov, A. Y., Hattne, J., Brewster, A. S., Sauter, N. K., … Weis, W. I. (2015). Enabling X-ray free electron laser crystallography for challenging biological systems from a limited number of crystals. *ELife*, *2015*(4), 1–29. https://doi.org/10.7554/eLife.05421

Van Bloois, E., Torres Pazmiño, D. E., Winter, R. T., & Fraaije, M. W. (2010). A robust and extracellular heme-containing peroxidase from Thermobifida fusca as prototype of a bacterial peroxidase superfamily. *Applied Microbiology and Biotechnology*, *86*(5), 1419–1430. https://doi.org/10.1007/s00253-009-2369-x

Verdín, J., Pogni, R., Baeza, A., Baratto, M. C., Basosi, R., & Vázquez-Duhalt, R. (2006). Mechanism of versatile peroxidase inactivation by Ca2+ depletion. *Biophysical Chemistry*, *121*(3), 163–170. https://doi.org/10.1016/j.bpc.2006.01.007

Voss, N. R., & Gerstein, M. (2010). 3V: Cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Research*, *38*(SUPPL. 2), 555–562. https://doi.org/10.1093/nar/gkq395

Waldron, K. J., Rutherford, J. C., Ford, D., & Robinson, N. J. (2009). Metalloproteins and metal sensing. *Nature*, *460*(7257), 823–830. Retrieved from http://dx.doi.org/10.1038/nature08300

Warkentin, M., Hopkins, J., & Thorne, R. E. (2011). research papers Dark progression reveals slow timescales for radiation damage between T = 180 and 240 K research papers, 792–803. https://doi.org/10.1107/S0907444911027600

Warkentin, M., & Thorne, R. E. (2009). Slow cooling of protein crystals. *Journal of Applied Crystallography*, *42*(5), 944–952. https://doi.org/10.1107/S0021889809023553

Weierstall, U., James, D., Wang, C., White, T. A., Wang, D., Liu, W., … Cherezov, V. (2014). Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms4309

Weik, M., Ravelli, R. B. G., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., … Sussman, J. L. (2000). Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proceedings of the National Academy of Sciences*, *97*(2), 623–628. https://doi.org/10.1073/pnas.97.2.623

Weik, M., Ravelli, R. B. G., Silman, I., Sussman, J. L., Gros, P., & Kroon, J. (2001). Specific protein dynamics near the solvent glass transition assayed by radiation-induced structural changes. *Protein Science*, *10*(10), 1953–1961. https://doi.org/10.1110/ps.09801

Wesenberg, D., Kyriakides, I., & Agathos, S. N. (2003). White-rot fungi and their enzymes for the treatment of industrial dye effluents. *Biotechnology Advances*, *22*(1–2), 161–187. https://doi.org/10.1016/j.biotechadv.2003.08.011

White, T. A., Kirian, R. A., Martin, A. V., Aquila, A., Nass, K., Barty, A., & Chapman, H. N. (2012). CrystFEL: A software suite for snapshot serial crystallography. *Journal*

of Applied Crystallography, *45*(2), 335–341. https://doi.org/10.1107/S0021889812002312

Wijma, H. J., Jeuken, L. J. C., Verbeet, M. P., Armstrong, F. A., & Canters, G. W. (2006). A random-sequential mechanism for nitrite binding and active site reduction in copper-containing nitrite reductase. *Journal of Biological Chemistry*, *281*(24), 16340–16346. https://doi.org/10.1074/jbc.M601610200

Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., … Evans, G. (2018). *DIALS* : implementation and evaluation of a new integration package. *Acta Crystallographica Section D Structural Biology*, *74*(2), 85–97. https://doi.org/10.1107/S2059798317017235

Wlodawer, A., Minor, W., Dauter, Z., & Jaskolski, M. (2013). Protien crystallography for aspiring crystallographers. *Febs J.*, *280*(22), 5705–5736. https://doi.org/10.1111/febs.12495.

Yabashi, M., Tanaka, H., & Ishikawa, T. (2015). Overview of the SACLA facility. *Journal of Synchrotron Radiation*, *22*, 477–484. https://doi.org/10.1107/S1600577515004658

Yannoni, C., Bernier, P., Bethune, D., Meijer, G., & Salem, J. (1991). NMR Determination of the Bond Lengths in C60. *Journal of the American Chemical Society*, *113*(8), 3190–3192. https://doi.org/10.1021/ja00008a068

Zarrine-Afsar, A., Barends, T. R. M., Müller, C., Fuchs, M. R., Lomb, L., Schlichting, I., & Miller, R. J. D. (2012). Crystallography on a chip. *Acta Crystallographica Section D: Biological Crystallography*, *68*(3), 321–323. https://doi.org/10.1107/S0907444911055296

Zeldin, O. B., Brockhauser, S., Bremridge, J., Holton, J. M., & Garman, E. F. (2013). Predicting the X-ray lifetime of protein crystals, (11). https://doi.org/10.1073/pnas.1315879110

Zubieta, C., Joseph, R., Krishna, S. S., McMullan, D., Kapoor, M., Axelrod, H. L., … Wilson, I. A. (2007). Identification and structural characterization of heme binding in a novel dye-decolorizing peroxidase, TyrA. *Proteins: Structure, Function and Genetics*, *69*(2), 234–243. https://doi.org/10.1002/prot.21673

Zubieta, C., Krishna, S. S., Kapoor, M., Kozbial, P., McMullan, D., Axelrod, H. L., … Wilson, I. A. (2007). Crystal structures of two novel dye-decolorizing peroxidases reveal a ??-barrel fold with a conserved heme-binding motif. *Proteins: Structure, Function and Genetics*, *69*(2), 223–233. https://doi.org/10.1002/prot.21550

Associated publications

The following is a comprehensive list of publications associated with the work presented in this thesis:

- A subset of the work presented in chapter 4 has been published in Ebrahim, Appleby, et al., (2019).
- A subset of the work presented in chapter 5 has been published in Ebrahim, Moreno-Chicano, et al., (2019).
- A subset of the work presented in chapter 7 has been published in Moreno-Chicano, Ebrahim et al., (2019).
- Data collection and processing methods developed in this thesis were used for a subset of data presented in Davy et al., (2019), with methods used to develop protein crystal samples published in Beale et al., (2019).

STRUCTURAL BIOLOGY
ISSN 2059-7983

Resolving polymorphs and radiation-driven effects in microcrystals using fixed-target serial synchrotron crystallography

Ali Ebrahim,[a,b] Martin V. Appleby,[b] Danny Axford,[b] John Beale,[b] Tadeo Moreno-Chicano,[a] Darren A. Sherrell,[b] Richard W. Strange,[a] Michael A. Hough[a]* and Robin L. Owen[b]*

[a]School of Biological Sciences, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, England, and [b]Diamond Light Source, Harwell Science and Innovation Campus, Didcot OX11 0DE, England. *Correspondence e-mail: mahough@essex.ac.uk, robin.owen@diamond.ac.uk

Received 23 April 2018
Accepted 16 July 2018

Keywords: radiation damage; fixed-target serial crystallography; metalloproteins; polymorphism; room temperature.

PDB references: copper nitrite reductase, small-cell polymorph data set 1, 6gb8; large-cell polymorph data set 1, 6gbb; non-polymorph-separated data set 1, 6gby; large-cell polymorph data set 15, 6gcg

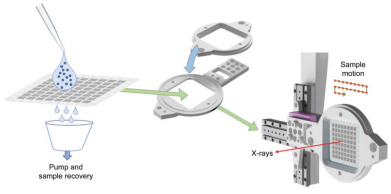
Supporting information: this article has supporting information at journals.iucr.org/d

The ability to determine high-quality, artefact-free structures is a challenge in micro-crystallography, and the rapid onset of radiation damage and requirement for a high-brilliance X-ray beam mean that a multi-crystal approach is essential. However, the combination of crystal-to-crystal variation and X-ray-induced changes can make the formation of a final complete data set challenging; this is particularly true in the case of metalloproteins, where X-ray-induced changes occur rapidly and at the active site. An approach is described that allows the resolution, separation and structure determination of crystal polymorphs, and the tracking of radiation damage in microcrystals. Within the microcrystal population of copper nitrite reductase, two polymorphs with different unit-cell sizes were successfully separated to determine two independent structures, and an X-ray-driven change between these polymorphs was followed. This was achieved through the determination of multiple serial structures from microcrystals using a high-throughput high-speed fixed-target approach coupled with robust data processing.

1. Introduction

X-ray crystallography using synchrotron radiation is at the core of structural biology, providing atomic-level insight into key biological processes. However, it has long been recognized that the X-rays that are used to determine structures also cause changes to the crystal lattice and protein structure, a phenomenon known as radiation damage. Polymorphism and non-isomorphism of crystals have been a challenge in protein crystallography since the earliest days of the field, with an early example of non-isomorphism being the separation of lysozyme into type I and type II in the 1960s (described in detail in Blake *et al.*, 2012) to allow the structure determination of (type II) lysozyme. The rise of cryo-crystallography in the 1990s (reviewed by Garman, 1999) made single-crystal structure determination routine, but the use of more intense X-ray beams, and the desire to determine structures from ever smaller crystals, has made multi-crystal structure determination the norm once more.

Polymorphism can be present even between crystals harvested from the same crystallization drop, with variations in unit-cell parameters or even space group being observed. These differences may be owing to some external process such as heavy-atom derivatization, dehydration or cryocooling, where the resulting, unwanted, changes in cell dimensions can cause structure determination to fail (Crick & Magdoff, 1956). Differences can also arise from structural variation, where



OPEN ACCESS

small changes in loops (Yogavel *et al.*, 2010) or conformational flexibility (Redinbo *et al.*, 1999) can result in significant changes in cell dimensions and space group. Non-isomorphism can also take more subtle forms deriving from, for example, weakly bound ligands, only becoming apparent in electron-density maps following careful cross-comparison of many data sets (Pearce *et al.*, 2017). In the case of multi-crystal and serial micro-crystallography data-set formation, however, larger scale differences are often used as the basis for the formation of a final data set through brute-force merging or a more refined approach such as hierarchical cluster analysis (Foadi *et al.*, 2013; Santoni *et al.*, 2017). Alternatively, statistical approaches such as the use of a genetic algorithm to optimize data-quality metrics such as the R value or $\langle I/\sigma(I)\rangle$ can also be used to obtain a single high-quality data set from many crystals (Zander *et al.*, 2016), although in the future this could also be used to identify and separate non-isomorphous groups.

Radiation damage results from energy deposited in crystals by X-rays and is manifested in two ways. Firstly, global radiation damage results in changes to the unit cell, increased disorder and a loss of diffracting power and consequently resolution [for comprehensive reviews of radiation damage in macromolecular crystallography, see Holton (2009) and Garman (2010)]. Secondly, site-specific radiation damage, which is most commonly observed in the form of disulfide reduction, the decarboxylation of side chains, and the reduction of metals and other redox centres. These changes occur on different dose scales and are temperature-dependent. Cryocooled (100 K) crystals are considered to no longer give useful diffraction beyond absorbed doses of 30 MGy (the Garman limit; Owen *et al.*, 2006). Despite the protection that cryocooling confers, site-specific changes occur at significantly lower doses than this, with the reduction of redox centres occurring at doses as low as 10 kGy in some cases (Kekilli *et al.*, 2017), some 3000 times lower than the Garman limit, a dose that can be achieved in a few milliseconds at modern synchrotron beamlines (Owen & Sherrell, 2016).

The vast majority of protein structures have been determined at 100 K in order to mitigate the global effects of radiation damage. The use of low temperatures, and an 'as low a dose as practicable per data set' strategy, mitigates damage and allows experiments such as multiple structures from one crystal (MSOX) to be performed in which electron-driven catalysis outruns electron-driven damage processes (Horrell *et al.*, 2016). At elevated temperatures both 'normal' data collection and experiments such as MSOX become considerably more challenging as the rates of damage increase.

The drawback of increased rates of damage must be weighed against the benefits that data collection at higher temperatures provides: decreased viscosity and increased thermal motion that allow more functionally relevant changes to be observed. Protein dynamics and reactivity are considerable within the crystal lattice, but are partly suppressed in the highly viscous, glassy solvent environment of crystals cooled to 100 K. Increases in mosaicity from cooling may be avoided by working at a temperature close to that at which the protein was crystallized. Increased reactivity within crystals, additional conformations of side chains and differences in ligand binding are observed when working at room temperature (RT; Fraser *et al.*, 2011; Fischer *et al.*, 2015), and there is a considerable incentive to determine structures at 'close to physiological' temperatures. RT structures may also be more directly relatable to solution kinetics experiments.

In terms of radiation damage, variable-temperature experiments have shown that most of the dose-lifetime extension gained at 100 K remains present at temperatures as high as 200 K (Warkentin *et al.*, 2012, 2013). Elevated-temperature crystallography remains a considerable challenge, however, particularly for small weakly diffracting crystals, which require a tightly focused intense beam. Despite these experimental challenges, MSOX series have been successfully determined at 190 K (where much of the advantage for crystal lifetime of data collection at 100 K is maintained but dynamic freedom is increased) and at RT using macrocrystals (Horrell *et al.*, 2018), revealing the considerable benefit of working at higher temperatures in that more of the reaction may be observed in the lifetime of the crystal owing to the higher dynamic freedom of the crystalline enzyme.

Here, we describe the development of a modified MSOX approach applied to microcrystals at RT. This approach is no longer multiple structures from one crystal, but multiple serial structures from many crystals (MSS). In order to effectively produce MSS series from microcrystals there are significant technical and methodological challenges to overcome: approaches for serial crystallography at synchrotron and XFELs provide a way forward.

Serial synchrotron crystallography (SSX) is an emerging area allowing structure determination by recording single images from many thousands of crystals (Diederichs & Wang, 2017). Typically, the data obtained are merged to provide a single structure. A number of sample-delivery methods to realize SSX have been developed, but a particular advantage of fixed-target approaches is that they can be readily modified to facilitate the determination of multiple structures.

MSS experiments require the ability to record data from a large number of crystals at a large number of dose points. The dose is controlled either by varying the beam intensity or the time that the crystal spends in the beam, and should be as similar as possible for all of the crystals studied, for example by matching aperture and crystal sizes and using a sample in which the microcrystals are relatively homogeneous in size and morphology. Other relevant factors include accurate and consistent stage movement and exposure timing. In order to follow a reaction, a number of well spaced dose points are required. In this work, we have examined the feasibility of measuring multiple data sets in close succession from microcrystals using fixed-target SSX. Using this approach, tens of dose-dependent data sets may be obtained highly efficiently, with each microcrystal exposed for a total of only a few hundred milliseconds.

We describe the approach with a detailed case study of MSS experiments on microcrystals of copper nitrite reductase from *Achromobacter cycloclastes* (*Ac*NiR). The nature of the data gained and the global and site-specific radiation-induced

changes to the structure are presented. We also describe a practical approach to separating polymorphs within a population of microcrystals and obtaining separate structures of each form, leading to two separate MSS series collected from a single batch of crystals.

2. Materials and methods

2.1. Crystallization and crystal loading to fixed targets

Recombinant *Ac*NiR was expressed and purified as described previously (Horrell *et al.*, 2016). Batch microcrystals were prepared by rapidly mixing 20 mg ml^{-1} *Ac*NiR in 20 m*M* Tris pH 7.5 with a solution consisting of 2.5 *M* ammonium sulfate, 0.1 *M* sodium citrate pH 4.5 buffer in a ratio of 1:3 and mixing by vortexing for 60 s. Microcrystals with a diameter of 5–15 µm grew at room temperature over a period of 4–6 d. Microcrystal suspensions were centrifuged at 800 rev min^{-1} for 30 s to sediment the crystals; the crystallization buffer was then removed and replaced with a storage buffer consisting of 1.6 *M* ammonium sulfate, 0.1 *M* sodium citrate pH 4.5. Crystals were soaked in a solution of mother liquor supplemented with 100 m*M* sodium nitrite for a duration of 20 min prior to loading onto the chip. Serial dilutions were achieved by adding additional storage-buffer solution.

Silicon nitride fixed targets or 'chips' of a new design, but based on those described previously (Mueller *et al.*, 2015; Oghbaey *et al.*, 2016), were used for the experiments described here. These chips follow the funnel-like design of previous chips but utilize a new higher capacity layout while retaining approximately the same dimensions (30 × 30 mm). Each chip comprises 8 × 8 individual 'city blocks', with each city block containing 20 × 20 apertures; the nominal capacity of each chip is therefore 25 600 (Fig. 1). The apertures are funnel-shaped, with the smaller end being 7 × 7 µm and the larger end being 85 × 85 µm.

Chips were prepared by glow-discharge cleaning in a similar manner to the cleaning of cryo-EM grids. Glow discharging improved the dispersion of the crystal slurry on the chip surface, reducing the volume of crystal slurry required to load a chip, and also resulted in improved drawing of liquid through the chip apertures. Chips were loaded within a humidity-control enclosure (Solo Containment, Cheshire, England). Typically, 100–200 µl of a microcrystal suspension was pipetted onto the surface of the chip, after which gentle suction was applied from below to draw microcrystals into the individual wells. The chip was then sealed between two layers of 6 µm thick Mylar in the sample holder before transfer to the beamline (Fig. 1).

2.2. Moving chips/instrumentation and beamline parameters

Instrumentation for the movement of chips through the X-ray beam was mounted on beamline I24 at Diamond Light Source as described previously (Owen *et al.*, 2017). An X-ray beam size of 8 × 8 µm (full width at half maximum; FWHM) was used. All data were measured at 12.8 keV using a PILATUS3 6M detector with a crystal-to-detector distance of

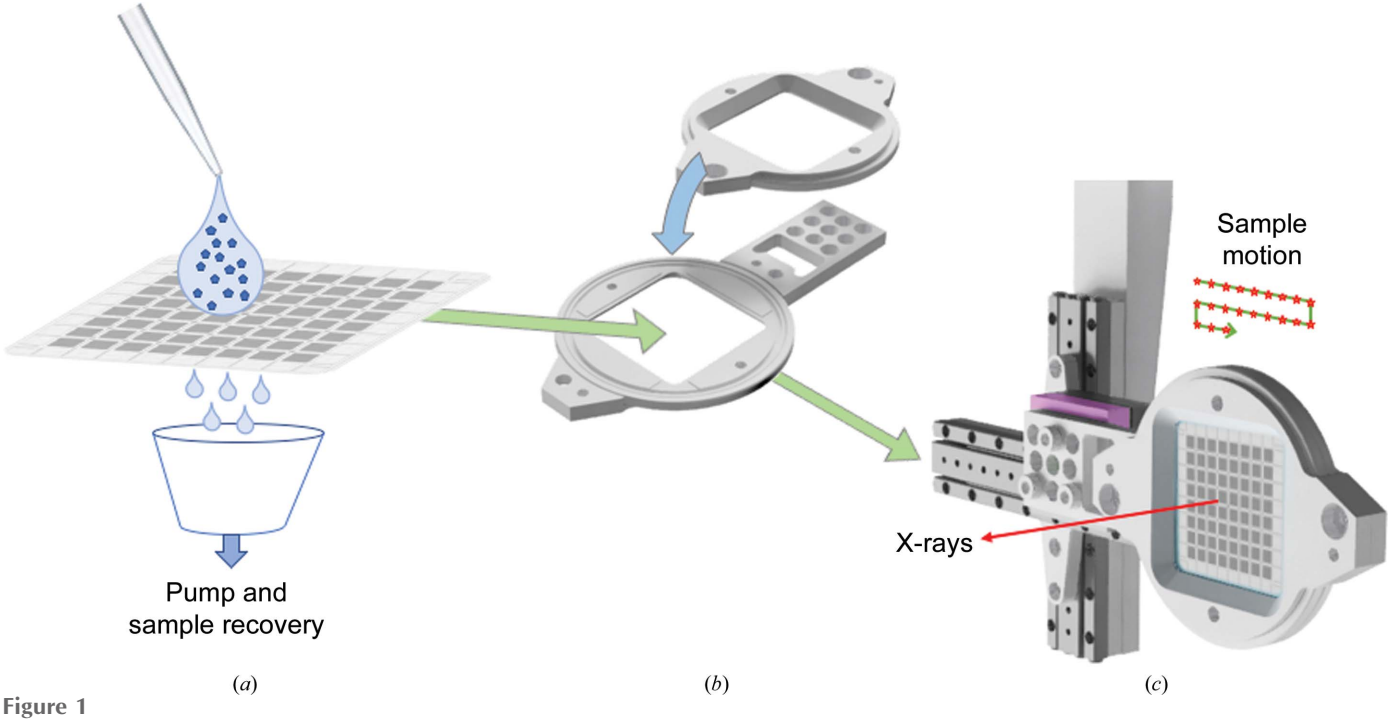


Figure 1
Schematic of chip-loading procedure. (*a*) The microcrystal suspension is pipetted onto the surface of the glow-discharged chip, with excess liquid being removed by the application of suction to the opposite surface. (*b*) Placing chips into the holder: a thin film of Mylar held in place by O-rings seals the chip and prevents drying out. (*c*) Loading of the chip and holder assembly onto the beamline sample stage during sample exchange: a kinematic mount (magenta) holds the chip in place in a precise and reproducible position, with subsequent alignment carried out using fiducial markings on each chip. The direction of the X-ray beam is indicated as a red arrow, while a schematic of chip movement is shown in green and red stars indicate positions where the chip is stationary and data are collected.

Table 1
Data-collection, processing and refinement statistics for selected *Ac*NiR structures used for polymorph separation and dose-series data.

Data were either processed together or binned into small-cell and large-cell subsets prior to scaling and merging. Data were collected from a single chip (25 600 positions) with 20 sequential images each of 20 ms per position. The beam size was 8 × 8 µm, with an incident flux of 3×10^{11} photons s^{-1} at a wavelength of 0.9686 Å. The space group for all data was $P2_13$. The diffraction-weighted dose per data set was 11 kGy.

	Small cell, dose 1	Large cell, dose 1	All data, dose 1	Large cell, dose 15
Cumulative dose (kGy)	11	11	11	165
No. of integrated frames	24976	13932	38908	21836
No. of images used	23467	13481	38798	21569
Cell dimension (Å)	96.38 (0.06)	97.75 (0.05)	96.87 (0.66)	97.99 (0.14)
Resolution (Å)	29.21–1.48 (1.51–1.48)	29.21–1.48 (1.51–1.48)	29.21–1.48 (1.51–1.48)	29.54–1.80 (1.83–1.80)
R_{merge} (%)	87.73 (96.34)	85.70 (97.74)	94.45 (97.75)	79.52 (99.55)
R_{split} (%)	5.71 (87.73)	7.43 (81.12)	5.15 (54.54)	5.83 (69.07)
$CC_{1/2}$	99.60 (72.60)	99.46 (48.66)	99.70 (55.42)	99.85 (50.16)
Mean intensity $\langle I \rangle$	72.3 (3.5)	65.7 (2.8)	76.8 (3.2)	28.2 (1.4)
Signal-to-noise ratio $I/\sigma(I)$	1.75 (0.17)	1.57 (0.12)	1.88 (0.15)	0.72 (0.06)
Multiplicity	548.16 (206.50)	357.91 (137.03)	927.03 (301.80)	978.36 (799.19)
Completeness (%)	100 (100)	100 (100)	100 (100)	100 (100)
No. of reflections	49882	51981	50616	29321
R/R_{free}	0.186/0.216	0.205/0.227	0.235/0.276	0.167/0.205
R.m.s.d., bond lengths (Å)	0.013	0.012	0.012	0.006
R.m.s.d., bond angles (°)	1.64	1.62	1.58	0.88
Ramachandran plot				
Most favoured (%)	97.3	96.4	97.5	99.4
Allowed (%)	2.7	3.6	2.5	0.6
PDB code	6gb8	6gbb	6gby	6gcg

310 mm. The beam flux of 3.0×10^{12} photons s^{-1} was measured immediately prior to the experiments using a silicon PIN diode as described previously (Owen *et al.*, 2009) and was attenuated tenfold for the data collections described below. The dose absorbed by each crystal was estimated using *RADDOSE-3D* (Zeldin, Gerstel *et al.*, 2013). Note that for the beam parameters described here, a diffraction-weighted dose (Zeldin, Brockhauser *et al.*, 2013) of 11 kGy corresponds to a maximum dose (as reported by older versions of *RADDOSE*) of 31 kGy. Data-collection parameters are shown in Table 1. As with all serial experiments, there will be some crystal-to-crystal variation in absorbed dose and so the calculated dose represents an average value. The largest difference in absorbed dose is likely to arise in crystals which are only partially exposed to X-rays if, for example, they are not centred in the apertures of the chips. We sought to minimize the variation in absorbed dose by ensuring that the beam size and intensity remained unchanged over the duration of the experiment, matching the beam size and chip aperture, and using crystals of a single morphology and of similar size.

2.3. Data-collection strategies

Rapid data series were measured from each aperture of a chip, with up to 20 diffraction images recorded at each position prior to translation to a fresh aperture. Data from each short series could then be sorted into dose bins and selectively merged, allowing dose-dependent structures to be obtained. This movement and dose-binning strategy is shown schematically in Fig. 2. Typically, an exposure period of 10 ms is used

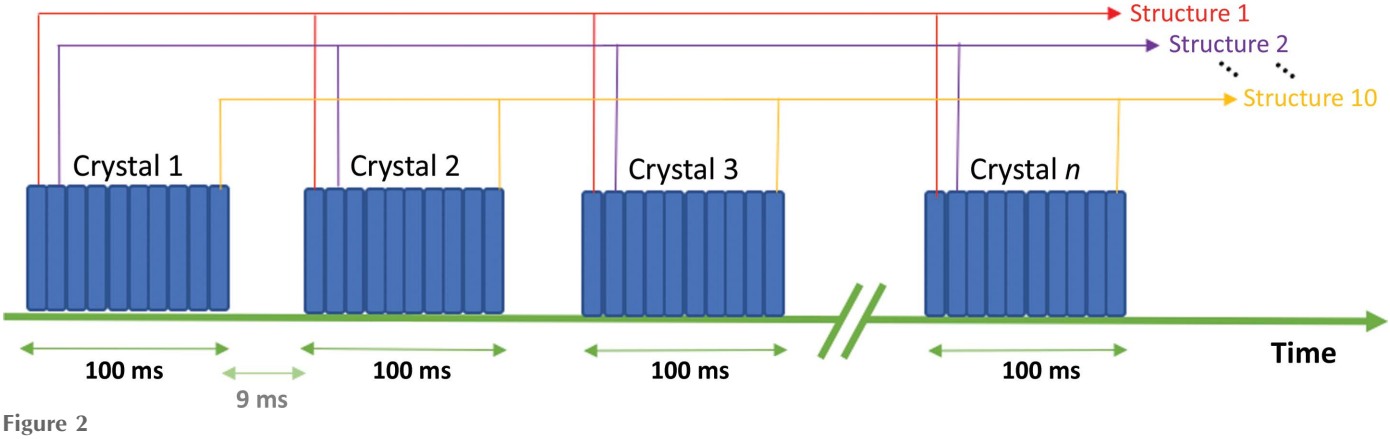


Figure 2
Schematic of MSS data collection and formation of dose series. At each position of the fixed target, multiple images are measured in shutterless mode using the PILATUS3 6M. In the example shown here, ten sequential exposures of 10 ms are recorded at each position and the translation time between positions is 9 ms. Images from each dose series are then grouped together in time (dose) bins, allowing dose-dependent structures to be obtained.

for each image, meaning that a ten-frame series (100 ms X-ray exposure per crystal) can be recorded from an entire chip (25 600 positions) in 46 min. The image series at each position was individually triggered using a Keysight 33500B signal generator, which in turn was triggered by a DeltaTau Geobrick LV-IMS-II stage controller when each crystal position had been reached. The X-ray shutter remained open for the duration of data collection and was not closed between apertures on a chip.

2.4. Data processing, structure solution and refinement

Data took the form of sequentially numbered images in CBF format. All images were indexed using *dials.still_process* in *DIALS* v.1.8.5 (Winter *et al.*, 2018) with subsequent scaling and merging performed using *prime* (Uervirojnangkoorn *et al.*, 2015). As an example of typical data volumes, throughputs and hit rates, the set of 20 dose points described below comprised some 500 000 images collected in less than 3 h. Bragg peaks were observed on 332 272 images, and 589 403 patterns were indexed (owing to multi-lattice indexing of up to three patterns per image). Of the indexed patterns, the percentages with single, double and triple lattices were 36, 36 and 28%, respectively.

Data were binned into the different dose points to produce final MTZ files for each MSS data set. The indexing ambiguity in space group $P2_13$ was resolved by the use of a reference data set collected from a single *Ac*NiR crystal at 100 K. Owing to the manner in which serial crystallography data are collected, 'traditional' metrics such as $R_{\rm merge}$ which compare individual measurements do not reflect the quality of the data set. The quantities $R_{\rm split}$ (White *et al.*, 2016), which compares separately merged halves of the data, and $CC_{1/2}$ (Karplus & Diederichs, 2012), which reports on the precision of merged measurements, were therefore used to assess the quality and resolution of each data set (Table 1). Both the mean intensity $\langle I\rangle$ and signal-to-noise ratio $I/\sigma(I)$ of each data set are also quoted; comparatively low values of $I/\sigma(I)$ reflect the challenge of accurately estimating the error of intensities in monochromatic serial data.

Structures were refined from a starting model of the room-temperature *Ac*NiR–nitrite complex, from which water and ligands had been removed (PDB entry 5i6l; Horrell *et al.*, 2016), using *REFMAC*5 (Murshudov *et al.*, 2011) in the *CCP*4*i* interface and/or *PHENIX* (Adams *et al.*, 2010). Structures were rebuilt in *Coot* (Emsley *et al.*, 2010) between rounds of refinement, and validation was performed using tools within *Coot*, *MolProbity* (Chen *et al.*, 2010), the *JCSG Quality Control Check* server (https://smb.slac.stanford.edu/jcsg/QC/) and the PDB validation server (https://validate-rcsb-1.wwpdb.org). Side-chain atoms that were not supported by electron density were deleted from the model. Coordinates and structure factors were deposited in the RCSB Protein Data Bank with the accession numbers given in Table 1. Surface areas and volumes were calculated in the 3*V* volume assessor (Voss & Gerstein, 2010).

3. Results and discussion

3.1. Discrimination between *Ac*NiR crystal polymorphs in the microcrystal population

A total of 20 successive images of 20 ms exposure each were measured at each aperture position of a chip, providing 20 room-temperature data sets at different dose points. Indexing of individual diffraction patterns (stills) from the first image measured from each *Ac*NiR microcrystal revealed a bimodal distribution of unit-cell parameters (Fig. 3). As a consequence, images were binned into two groups above and below a mid-peak cutoff of 97.25 Å, which we henceforth refer to as *Ac*NiR-big ($a = b = c = 97.75$ Å) and *Ac*NiR-small ($a = b = c =$ 96.38 Å). Data-processing statistics for each group (for data set 1) compared with those obtained by merging all patterns are given in Table 1. Separation of the unit-cell polymorphs

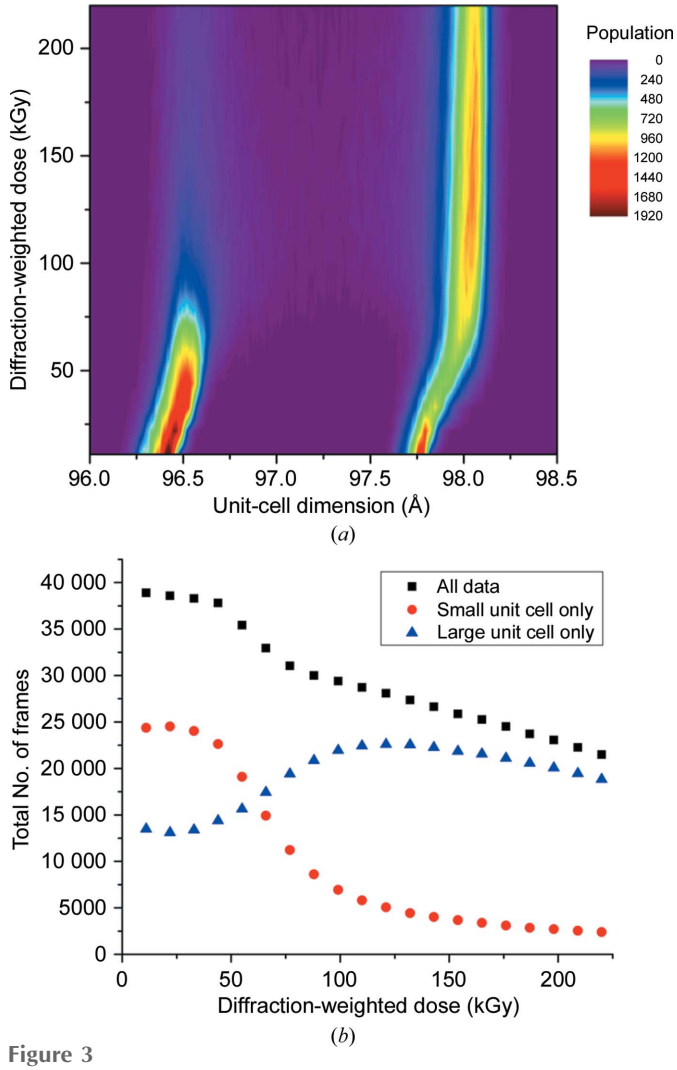


Figure 3
(*a*) Two-dimensional histogram showing changes in unit-cell dimension ($a = b = c$ in space group $P2_13$) and population across a chip as a function of absorbed dose. Populations are the numbers of indexed images in unit-cell bins of width 0.01 Å. The starting unit-cell dimensions of the polymorphs are 96.38 and 97.75 Å; these increase to 96.56 and 98.04 Å at 220 kGy. (*b*) Number of integrated images of each polymorph on the chip as a function of absorbed dose. Images were selectively integrated into 'small' or 'large' unit-cell groups (details are given in the text).

led to an improvement in data quality, suggesting that this analysis step is beneficial over simply merging all data regardless of unit-cell parameter. Refinement of the structure arising from each polymorph revealed two different structures, as shown in Fig. 4. The *Ac*NiR-big and *Ac*NiR-small structures were superimposable with an r.m.s.d. of 0.16 Å (Supplementary Fig. S1). While the overall *Ac*NiR-big and *Ac*NiR-small structures were very similar, significant structural differences were observed at the N- and C-termini and in the loop structure around residues 187–193 and 201–205 (Fig. 4*a*). Examination of the symmetry-generated *Ac*NiR trimer (the biological assembly) revealed further differences related to the unit-cell polymorphs (Fig. 4*b*), whereby the volume of the *Ac*NiR-big trimer was 157 399 Å^3, an increase of some 1901 Å^3 over the *Ac*NiR-small volume of 155 498 Å^3. The corresponding increase in surface area was 659 Å^2: from 22 998 Å^2 for *Ac*NiR-big to 22 339 Å^2 for *Ac*NiR small.

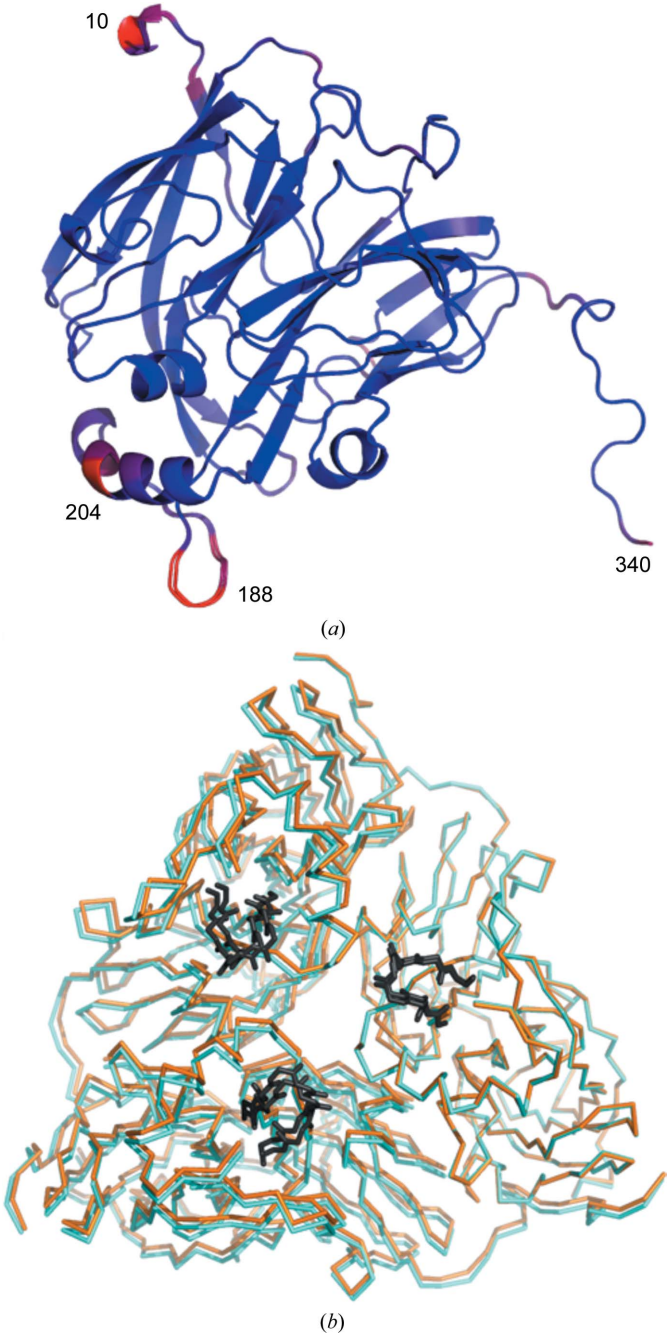


Figure 4
(*a*) Superposition of the data set 1 structures of *Ac*NiR-big and *Ac*NiR-small coloured by r.m.s.d. between the structures from low values (blue) to high values (red). (*b*) Superposition (using monomers *A*) of trimers generated by crystal symmetry for data set 1 of *Ac*NiR-small (orange) and *Ac*NiR-big (cyan). The view is down the threefold axis. While the structures of individual monomers are very similar, structural changes across the trimer are larger and are related to the change in unit-cell volume. The loop regions (residues 186–193) are shown in dark grey.

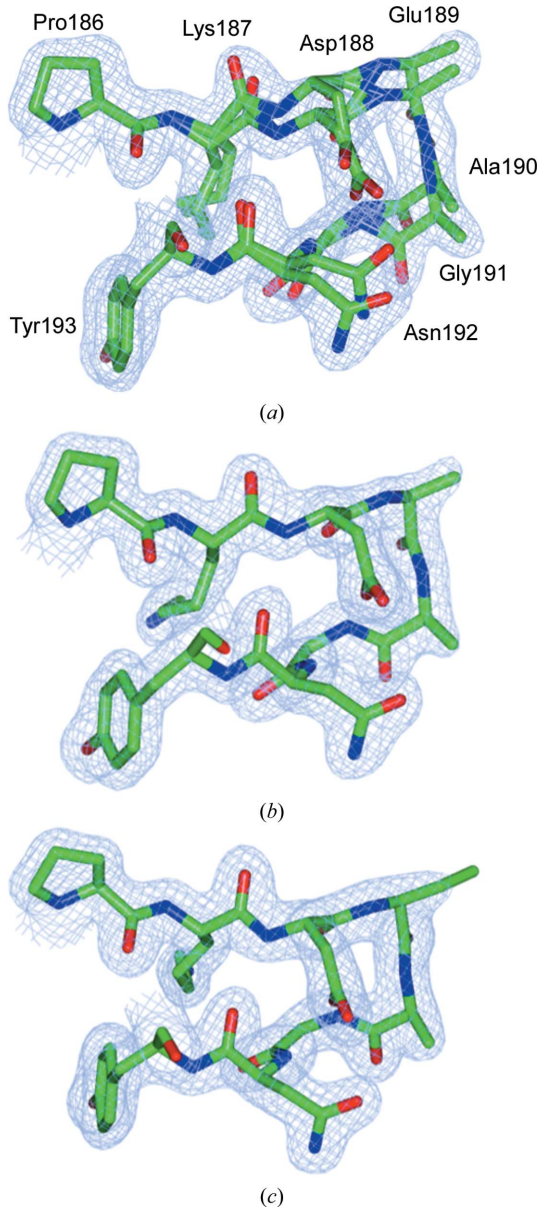


Figure 5
$2F_o - F_c$ electron-density maps for the *Ac*NiR data set 1 structures derived from (*a*) all images, (*b*) the small unit cell only and (*c*) the large unit cell only. The maps are contoured at 0.311 e$^-$ Å^{-3} (all data), 0.368 e$^-$ Å^{-3} (*Ac*NiR-small) and 0.348 e$^-$ Å^{-3} (*Ac*NiR-big). Note the dual conformations in (*a*) with occupancies (0.3 and 0.7) consistent with the proportion of large-cell and small-cell images within data set 1. This figure was prepared using *CCP4mg* (McNicholas *et al.*, 2011).

Without the polymorph-separation procedure, the single combined data set produced an electron-density map with dual conformations representing the two polymorph structures (Fig. 5). Comparison of these structures with earlier 100 K crystal structures (for example PDB entry 2bw4, where the unit-cell parameter was 95.41 Å; Antonyuk *et al.*, 2005) revealed a pattern of changes that affect the crystal contacts between *Ac*NiR monomers. These structural differences between the two polymorphs are explored further in §3.3.

The reason why two different polymorphs exist within a single sample of batch-grown *Ac*NiR microcrystals and the factors determining the proportion of each that is present remain unclear. Interestingly, different batch crystallizations of similar age exhibited different proportions of *Ac*NiR-big and *Ac*NiR-small (data not shown). To investigate possible causes of the polymorph distribution, a second chip of *Ac*NiR microcrystals was used in which different dilutions of the starting microcrystal stock were applied to different 'city blocks' of the chip. A progressive shift with dilution was observed, from an almost entirely small-cell population in the undiluted sample to a predominantly large-cell population at dilutions greater than 15× (Supplementary Fig. S2). The mechanism by which dilution influences the cell polymorph remains unclear. We note that as the dilution was achieved by the addition of further crystal-storage buffer to the microcrystal suspension, there were no changes to the pH or the precipitant concentration in this process.

3.2. Global radiation damage in MSS data sets

As dose was accumulated at a relatively high dose rate of 1.1 MGy s^{-1}, evidence of global radiation damage was found in the form of a rapid decrease in diffracting power (Fig. 6). An initial plateau region or lag phase is apparent spanning the first three data sets, corresponding to ∼20–35 kGy and 20–30 ms exposure time. The subsequent fall in diffracting power follows an exponential decay, with a somewhat faster fall-off with dose for the *Ac*NiR-small population than the *Ac*NiR-big population. The half-doses for the two polymorphs were 55 and 75 kGy, respectively (as estimated from Fig. 6). The half-dose for the large-cell polymorph is larger than that for the small cell owing to the convolution of the decay in diffracting power with the radiation-driven polymorph switch (Fig. 3*b*).

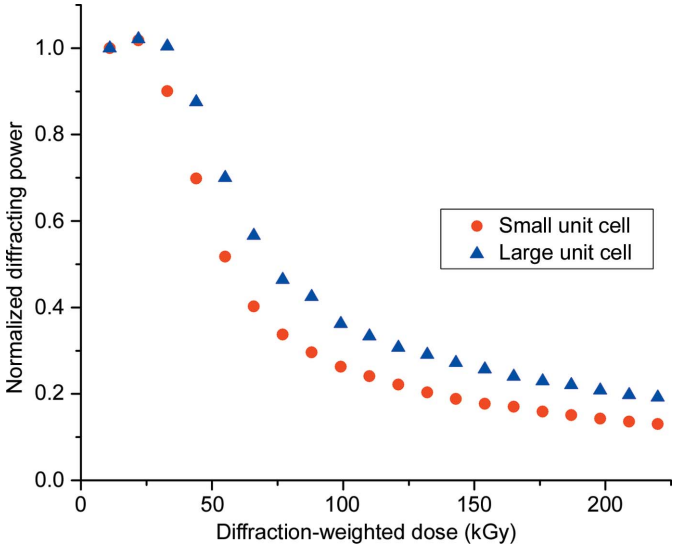


Figure 6
Total diffracting power of crystals as a function of accumulated X-ray dose. Diffracting power was defined as the total Bragg intensity for each dose point as reported by *prime* after integration and scaling over the resolution range 30–1.48 Å.

Such an intensity decay can arise from a number of sources. Firstly, exceeding the count-rate limit of a single photon-counting detector such as the PILATUS3 could result in a plateau. Care was taken to ensure that count rates were well below the maximum of 10×10^6 counts per second per pixel that PILATUS3 detectors are capable of accurately recording. In the experiments described here the maximum observed counts in a Bragg spot was ∼8000, corresponding to a count-rate of 0.4×10^6 counts per second per pixel or 4% of the maximum count rate. More typical maximum counts in a Bragg spot were <4000, or less than 2% of the maximum. Further, as the crystals are not rotated at all during data collection, count rates are likely to be steady throughout the duration over which an image is recorded: a key assumption in any count-rate correction that is made during detector readout. Secondly, a lag phase can result from the outrunning of global effects such as beam-induced heating. In an initial period when site-specific damage dominates, the global decay in intensity will deviate from an exponential decay (Sygusch & Allaire, 1988, Owen *et al.*, 2014). Thirdly, as the beam size is approximately equal to the aperture size, any temporal error in chip motion and detector triggering would mean that the diffracting power observed in the first image would be significantly reduced. Considering the horizontal beam profile, chip aperture and motion only, 68% of the beam intensity (beam FWHM 8 µm) falls within the 7 µm chip aperture assuming that they are perfectly co-centred. If the chip aperture is offset by 1, 2 or 3 µm the intensity incident on the aperture falls by 3, 11 or 23%, respectively. Thus, small errors in positioning can result in large differences in the recorded intensity. While extreme care was taken to tune the stages to eliminate this as a source of systematic error, it cannot be discounted. Fourthly, the Gaussian non-tophat profile of the beam can affect intensity decay. Warkentin and coworkers examined the effect of beam profile and dose rates on the rates of global radiation damage at room temperature in thaumatin and lysozyme crystals, showing that a non-uniform beam profile can result in a non-exponential dose response. All of these factors may contribute to the intensity-decay profile observed (Warkentin *et al.*, 2017).

Individual MSS data sets in the series of 20 showed good-quality merging statistics to a resolution of 1.48 Å (Table 1), despite a nominal resolution limit of 1.7 Å imposed by the crystal-to-detector distance (inscribed circle on the detector surface). This is owing to the extremely high redundancy achieved with our high hit-rate chip-based serial data collection. Note that in the case of small-cell *Ac*NiR, in addition to a global decrease in diffracting power, the decrease in resolution

is significantly affected by the reduction in population resulting from X-ray-driven transfer from the small-cell to the large-cell *Ac*NiR polymorph (Table 1). The resulting decrease in multiplicity significantly impacts the resolution to which acceptable merging statistics are obtained.

3.3. Changes to polymorph populations with X-ray dose

Monitoring of the mean cell dimension in each polymorph together with the population distribution reveals intriguing changes as X-ray dose is accumulated (Fig. 3). The mean cell dimension within each population undergoes a small, progressive increase (Supplementary Fig. S3) consistent with many previous studies showing unit-cell expansion with dose. While often observed, unit-cell expansion is generally not regarded as a reliable metric of radiation damage owing to the lack of reproducibility of the effect between different crystals of the same protein (Murray & Garman, 2002). Remarkably, however, in our MSS data an interchange between the unit-cell polymorph populations throughout the dose series is also evident (Fig. 3*b*). The *Ac*NiR-small population rapidly decreases, with a concomitant increase in the population of *Ac*NiR-big. It is apparent from Fig. 3(*a*) that (i) the switch from small to large cell for any particular microcrystal yields a cell that is consistent with the 'damaged large cell' of a particular dose rather than the large cell at dose point 1 and (ii) the lack of overlap between the two populations implies a specific structural change between polymorphs. Interestingly, the increase in unit cell for both polymorphs begins immediately upon irradiation (Fig. 3, Supplementary Fig. S3), while the switching of polymorphs is minimal within the first 100 ms before proceeding rapidly. While not conclusive, this could suggest that expansion of the small cell acts as a trigger or seed leading to subsequent polymorph swapping.

As structural differences in the loop region around Asp188 were observed between polymorphs, further comparison was undertaken to seek to understand the mechanism behind the switching between populations. Multiple single-crystal structures are available for resting-state AcNiR, primarily measured at 100 K but also at elevated cryogenic temperatures (Sen *et al.*, 2017). Superposition of this loop region in *Ac*NiR structures determined at different temperatures (Supplementary Fig. S4) reveals a progressive shift in the position of the loop from 100 K (cell length 95.41 Å) to 240 K (PDB entry 5n8f; 96.13 Å) and RT (PDB entry 5off; 96.23 Å). The latter is very similar in structure to the *Ac*NiR-small polymorph (96.38 Å). A further shift then occurs to the *Ac*NiR-big polymorph (98.21 Å). While further work is required to prove a link, the apparent correlation of loop position with cell length suggests that this loop reorganization may indeed be related to the switching of polymorphs.

The precise mechanism by which the large unit-cell expansion from *Ac*NiR-small to *Ac*NiR-big occurs upon irradiation remains unclear, but insights may be gained from known site-specific radiation-damage phenomena. Decarboxylation of aspartic acid and glutamic acid residues has been well characterized (Burmeister, 2000; Holton, 2009; Garman, 2010), and notably the dual conformation of the loop region contains two such residues: Asp188 and Glu189. To examine this possible cause of the shift between polymorphs, the structure of data set 15 with the large-cell polymorph was examined (Table 1, Supplementary Fig. S5). The electron density for Asp188 remained clear, with no evidence of site-specific radiation damage, while Glu189 was disordered at both dose points 1 and 15, and appears to point towards solvent rather than being involved in crystal contacts. Indeed, the loop structures and density are highly similar in data set 1 and data set 15 (Supplementary Fig. S6). Notably, by this point in the dose series the majority of the images arise from crystals that began with the small cell and subsequently switched to the large cell. Other possible explanations for the dose-driven polymorph exchange are related to changes in hydration or thermodynamic factors, which could arise from heating of the microcrystal and surrounding mother liquor in the beam or from the generation of gases by radiolysis.

4. Conclusions and outlook

We have demonstrated the capability to obtain room-temperature dose-dependent structures from microcrystals in silicon nitride fixed-target chips in a highly sample- and time-efficient manner. This approach is termed multiple serial structures from many crystals (MSS). An important advantage of the MSS approach is that each data set/structure within the dose series may be improved by simply repeating the measurement on additional chips and increasing the number of merged stills contributing to each dose point. This is in contrast to single-crystal experiments, in which improvements in resolution/redundancy must typically be gained at the cost of a higher dose per data set. The observed resolution of 1.48 Å is comparable to the resolution of 1.40 Å achieved using a single large room-temperature crystal at a comparable dose (Horrell *et al.*, 2018). Through the use of a serial approach, we were able to obtain serial structures at lower dose points from significantly smaller crystals using a beam with a flux density more than an order of magnitude greater.

A further advantage is that polymorphs within a batch crystal population may be separated based on unit-cell binning, leading to an improvement in data quality together with the ability to refine structures of the polymorphs independently. This is in contrast to single-crystal experiments, where polymorphs may only be visible in the form of weak additional density or dual conformations. In the future, this approach could be expanded to exploit more complex forms of grouping data, allowing polymorphs with similar unit cells to be separated, and to complement computational approaches for revealing structural heterogeneity in proteins (Lang *et al.*, 2014).

The ability to obtain dose series in a sample- and time-efficient manner at room temperature opens the door to the routine production of MSS movies of redox-enzyme function. The approach remains straightforward even when only microcrystals are available. The ability to collect data and obtain dose series using the same experimental setup as used

for zero-dose SFX data collection will be particularly powerful as fine slicing of dose may allow the extrapolation of progressive X-ray-induced changes back to zero dose, which could then be verified by comparison with SFX structures determined under near-identical conditions. The approach is also well suited to time-resolved applications, as each crystal could be optically excited upon reaching the beam position and the changes then tracked over the duration of an image series. The ability, provided by a fixed-target approach, to finely control the dose and the time that each crystal spends in the X-ray beam while recording multiple slices of data is a valuable addition to the serial crystallographer's toolbox.

Acknowledgements

All data for this paper were measured on beamline I24 at Diamond Light Source under proposal NT14493. Sam Horrell and Demet Kekilli of the University of Essex are thanked for their contributions to the early stages of this work.

Funding information

TMC was supported by Leverhulme Trust award RPG-2014-355 to MAH and RWS. RWS and MAH were supported by BBSRC awards BB/M020924/1 and BB/M022714/1, respectively.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Antonyuk, S. V., Strange, R. W., Sawers, G., Eady, R. R. & Hasnain, S. S. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 12041–12046.
Blake, C. C. F., Fenn, R. H., Johnson, L. N., Koenig, D. F., Mair, G. A., North, A. C. T., Oldham, J. W. H., Phillips, D. C., Poljak, R. J., Sarma, V. R. & Vernon, C. A. (2012). *International Tables for Crystallography*, Vol. F, edited by E. Arnold, D. M. Himmel & M. G. Rossmann, pp. 845–872. Chester: International Union of Crystallography.
Burmeister, W. P. (2000). *Acta Cryst.* D**56**, 328–341.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Crick, F. H. C. & Magdoff, B. S. (1956). *Acta Cryst.* **9**, 901–908.
Diederichs, K. & Wang, M. (2017). *Methods Mol. Biol.* **1607**, 239–272.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Fischer, M., Shoichet, B. K. & Fraser, J. S. (2015). *Chembiochem*, **16**, 1560–1564.
Foadi, J., Aller, P., Alguel, Y., Cameron, A., Axford, D., Owen, R. L., Armour, W., Waterman, D. G., Iwata, S. & Evans, G. (2013). *Acta Cryst.* D**69**, 1617–1632.
Fraser, J. S., van den Bedem, H., Samelson, A. J., Lang, P. T., Holton, J. M., Echols, N. & Alber, T. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 16247–16252.
Garman, E. (1999). *Acta Cryst.* D**55**, 1641–1653.
Garman, E. F. (2010). *Acta Cryst.* D**66**, 339–351.
Holton, J. M. (2009). *J. Synchrotron Rad.* **16**, 133–142.
Horrell, S., Antonyuk, S. V., Eady, R. R., Hasnain, S. S., Hough, M. A. & Strange, R. W. (2016). *IUCrJ*, **3**, 271–281.
Horrell, S., Kekilli, D., Sen, K., Owen, R. L., Dworkowski, F. S. N., Antonyuk, S. V., Keal, T. W., Yong, C. W., Eady, R. R., Hasnain, S. S., Strange, R. W. & Hough, M. A. (2018). *IUCrJ*, **5**, 283–292.
Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
Kekilli, D., Moreno-Chicano, T., Chaplin, A. K., Horrell, S., Dworkowski, F. S. N., Worrall, J. A. R., Strange, R. W. & Hough, M. A. (2017). *IUCrJ*, **4**, 263–270.
Lang, P. T., Holton, J. M., Fraser, J. S. & Alber, T. (2014). *Proc. Natl Acad. Sci. USA*, **111**, 237–242.
McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. (2011). *Acta Cryst.* D**67**, 386–394.
Mueller, C. *et al.* (2015). *Struct. Dyn.* **2**, 054302.
Murray, J. & Garman, E. (2002). *J. Synchrotron Rad.* **9**, 347–354.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Oghbaey, S. *et al.* (2016). *Acta Cryst.* D**72**, 944–955.
Owen, R. L., Axford, D., Sherrell, D. A., Kuo, A., Ernst, O. P., Schulz, E. C., Miller, R. J. D. & Mueller-Werkmeister, H. M. (2017). *Acta Cryst.* D**73**, 373–378.
Owen, R. L., Holton, J. M., Schulze-Briese, C. & Garman, E. F. (2009). *J. Synchrotron Rad.* **16**, 143–151.
Owen, R. L., Paterson, N., Axford, D., Aishima, J., Schulze-Briese, C., Ren, J., Fry, E. E., Stuart, D. I. & Evans, G. (2014). *Acta Cryst.* D**70**, 1248–1256.
Owen, R. L., Rudiño-Piñera, E. & Garman, E. F. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 4912–4917.
Owen, R. L. & Sherrell, D. A. (2016). *Acta Cryst.* D**72**, 388–394.
Pearce, N. M., Krojer, T., Bradley, A. R., Collins, P., Nowak, R. P., Talon, R., Marsden, B. D., Kelm, S., Shi, J., Deane, C. M. & von Delft, F. (2017). *Nature Commun.* **8**, 15123.
Redinbo, M. R., Stewart, L., Champoux, J. J. & Hol, W. G. J. (1999). *J. Mol. Biol.* **292**, 685–696.
Santoni, G., Zander, U., Mueller-Dieckmann, C., Leonard, G. & Popov, A. (2017). *J. Appl. Cryst.* **50**, 1844–1851.
Sen, K., Horrell, S., Kekilli, D., Yong, C. W., Keal, T. W., Atakisi, H., Moreau, D. W., Thorne, R. E., Hough, M. A. & Strange, R. W. (2017). *IUCrJ*, **4**, 495–505.
Sygusch, J. & Allaire, M. (1988). *Acta Cryst.* A**44**, 443–448.
Uervirojnangkoorn, M., Zeldin, O. B., Lyubimov, A. Y., Hattne, J., Brewster, A. S., Sauter, N. K., Brunger, A. T. & Weis, W. I. (2015). *Elife*, **4**, e05421.
Voss, N. R. & Gerstein, M. (2010). *Nucleic Acids Res.* **38**, W555–W562.
Warkentin, M. A., Atakisi, H., Hopkins, J. B., Walko, D. & Thorne, R. E. (2017). *IUCrJ*, **4**, 785–794.
Warkentin, M., Badeau, R., Hopkins, J. B., Mulichak, A. M., Keefe, L. J. & Thorne, R. E. (2012). *Acta Cryst.* D**68**, 124–133.
Warkentin, M., Hopkins, J. B., Badeau, R., Mulichak, A. M., Keefe, L. J. & Thorne, R. E. (2013). *J. Synchrotron Rad.* **20**, 7–13.
White, T. A., Mariani, V., Brehm, W., Yefanov, O., Barty, A., Beyerlein, K. R., Chervinskii, F., Galli, L., Gati, C., Nakane, T., Tolstikova, A., Yamashita, K., Yoon, C. H., Diederichs, K. & Chapman, H. N. (2016). *J. Appl. Cryst.* **49**, 680–689.
Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K. & Evans, G. (2018). *Acta Cryst.* D**74**, 85–97.
Yogavel, M., Khan, S., Bhatt, T. K. & Sharma, A. (2010). *Acta Cryst.* D**66**, 584–592.
Zander, U., Cianci, M., Foos, N., Silva, C. S., Mazzei, L., Zubieta, C., de Maria, A. & Nanao, M. H. (2016). *Acta Cryst.* D**72**, 1026–1035.
Zeldin, O. B., Brockhauser, S., Bremridge, J., Holton, J. M. & Garman, E. F. (2013). *Proc. Natl Acad. Sci. USA*, **110**, 20551–20556.
Zeldin, O. B., Gerstel, M. & Garman, E. F. (2013). *J. Appl. Cryst.* **46**, 1225–1230.

IUCrJ
ISSN 2052-2525
BIOLOGY | MEDICINE

Dose-resolved serial synchrotron and XFEL structures of radiation-sensitive metalloproteins

Ali Ebrahim,[a,b] Tadeo Moreno-Chicano,[a]‡ Martin V. Appleby,[b] Amanda K. Chaplin,[a]§ John H. Beale,[b] Darren A. Sherrell,[b] Helen M. E. Duyvesteyn,[b,c] Shigeki Owada,[d,e] Kensuke Tono,[d,e] Hiroshi Sugimoto,[e] Richard W. Strange,[a] Jonathan A. R. Worrall,[a] Danny Axford,[b] Robin L. Owen[b]* and Michael A. Hough[a]*

[a]School of Biological Sciences, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK, [b]Diamond Light Source, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0DE, UK, [c]Division of Structural Biology (STRUBI), The Henry Wellcome Building for Genomic Medicine, University of Oxford, Roosevelt Drive, Oxford, Oxfordshire OX3 7BN, UK, [d]RIKEN SPring-8 Center, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan, and [e]Japan Synchrotron Radiation Research Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5198, Japan. *Correspondence e-mail: robin.owen@diamond.ac.uk, mahough@essex.ac.uk

Received 6 February 2019
Accepted 22 March 2019

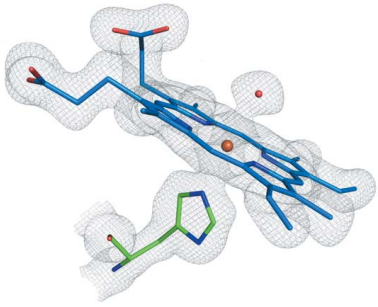
Edited by E. E. Lattman, University at Buffalo, USA

‡ Current address for TMC: Department of Biomolecular Mechanisms, Max Planck Institute for Medical Research, Jahnstrasse 29, 69120 Heidelberg, Germany.
§ Current address for AKC: Department of Biochemistry, Sanger Building, 80 Tennis Court Road, Cambridge CB2 1GA, UK.

Keywords: XFELs; microcrystals; serial femtosecond crystallography; serial synchrotron crystallography; serial millisecond crystallography; fixed targets; heme peroxidase; metalloproteins; radiation damage.

PDB references: Dye-type peroxidase Aa from *Streptomyces lividans*, MSS 78.4 kGy structure, 6i8p; 274.4 kGy structure, 6q3e; 235.2 kGy structure, 6q3d; 164 kGy structure, 6i8k; 98.4 kGy structure, 6i8i; 32.8 kGy structure, 6i7z; 39.2 kGy structure, 6i8o; 196 kGy structure, 6q34; 117.6 kGy structure, 6i8q; 65.6 kGy structure, 6i8e; 156.8 kGy structure, 6q31; 131.2 kGy structure, 6i8j; 313.6 kGy structure, 6ibn; SFX structure of damage-free ferric state of dye-type peroxidase Aa from *Streptomyces lividans*, 6i43

Supporting information: this article has supporting information at www.iucrj.org



An approach is demonstrated to obtain, in a sample- and time-efficient manner, multiple dose-resolved crystal structures from room-temperature protein microcrystals using identical fixed-target supports at both synchrotrons and X-ray free-electron lasers (XFELs). This approach allows direct comparison of dose-resolved serial synchrotron and damage-free XFEL serial femtosecond crystallography structures of radiation-sensitive proteins. Specifically, serial synchrotron structures of a heme peroxidase enzyme reveal that X-ray induced changes occur at far lower doses than those at which diffraction quality is compromised (the Garman limit), consistent with previous studies on the reduction of heme proteins by low X-ray doses. In these structures, a functionally relevant bond length is shown to vary rapidly as a function of absorbed dose, with all room-temperature synchrotron structures exhibiting linear deformation of the active site compared with the XFEL structure. It is demonstrated that extrapolation of dose-dependent synchrotron structures to zero dose can closely approximate the damage-free XFEL structure. This approach is widely applicable to any protein where the crystal structure is altered by the synchrotron X-ray beam and provides a solution to the urgent requirement to determine intact structures of such proteins in a high-throughput and accessible manner.

1. Introduction

Enzymology and structural biology are highly dependent on the accurate three-dimensional models obtained by X-ray crystallography. Such structures provide insight into function and can form a basis for understanding how proteins interact with each other or with small molecules. Fundamentally, the structure obtained should be representative of the native state of the protein. However, macromolecular crystallography is typically carried out at cryogenic temperatures (100 K) to minimize radiation-damage-induced structural perturbation (Garman & Weik, 2017; Holton, 2009). There is an increasing recognition of the importance of determining structures at ambient or 'room' temperature so as to be more representative of the structures and dynamics adopted by proteins *in vivo* at physiological temperature (Keedy *et al.*, 2018, 2015; Fischer *et al.*, 2015; Weik & Colletier, 2010). A major challenge in conventional synchrotron-based X-ray crystallography, particularly at room temperature, is the extremely rapid onset of radiation damage, *i.e.* changes to the structure of the protein caused by the ionizing effects of the X-ray beam (Garman &

Weik, 2017; Holton, 2009). The rapidity of site-specific radiation damage means it is present in differing levels of severity in almost all crystallographic datasets determined using synchrotron radiation, even if great care to avoid it is taken during data collection by minimizing the absorbed dose; the challenge becomes even greater if only microcrystals are available. When collecting data from microcrystals, microbeams of increased brilliance are required for optimal data collection, though the use of such beams comes with a concomitant increase in the rate of X-ray induced changes. It is critical that the site-specific effects of radiolysis are understood in detail and minimized in order to produce structures that are accurately representative of radiation-sensitive proteins *in vivo*. We note however that in some protein crystals that do not contain metals or other redox centres, radiation damage, while present, may cause little change to the observed structure.

It is estimated that approximately one-third of all proteins require a metal ion, with around half of all enzymes utilizing a metal for catalytic function (Waldron *et al.*, 2009). Heme enzymes catalyse many essential reactions in biology and understanding their structures throughout their reaction cycles is of high interest, prompting extensive efforts made to obtain 'intact' structures at high resolution (Chreifi *et al.*, 2016; Casadei *et al.*, 2014; Gumiero *et al.*, 2011; Moody & Raven, 2018). A major challenge is to obtain the higher valence states of these proteins, for example peroxidases (Fe^{III} resting state and Fe^{IV} intermediate states), as these states are phenomenally sensitive to reduction in synchrotron experiments caused by the presence of large numbers of solvated electrons or other radiolytically produced species, generated by the interaction of X-rays with the crystal (Kekilli *et al.*, 2017; Beitlich *et al.*, 2007; Denisov *et al.*, 2007). This site-specific damage is known to occur at doses much lower than those typically required to collect a dataset (Garman & Weik, 2017; Holton, 2009). Structures of peroxidases and other redox-sensitive metalloproteins obtained from synchrotron X-ray crystallography, even at 100 K, are therefore likely to represent a superposition of resting and damaged states. This site-specific damage is an extremely pressing problem if mechanistic conclusions are to be drawn from the structures obtained.

In contrast, X-ray free-electron lasers (XFELs) promise damage and artefact-free crystallography, provided the pulse duration is short enough (Schlichting, 2015; Nass *et al.*, 2015; Lomb *et al.*, 2011). For serial femtosecond crystallography (SFX), data collection from each microcrystal can be completed before site-specific and global radiation damage occurs, but at the expense of longer term crystal destruction, such that a new crystal must be presented to every pulse. Radiation-induced changes have been detected in SFX data with pulse durations as short as 40 fs (Nass *et al.*, 2015) or 70 fs (Lomb *et al.*, 2011). In contrast, data measured with pulse durations of 10 fs or shorter are considered to be free of typical radiation-induced site-specific radiation damage (Halsted *et al.*, 2018; Andersson *et al.*, 2017). Direct comparison of XFEL and synchrotron structures of the same protein presents many challenges because of differences between these methods, such as crystal size, mosaicity, temperature, cryoprotection, crystallization conditions and resolution. The epitome of this being that almost all XFEL structures are obtained from tens of thousands of room-temperature microcrystals while most synchrotron structures are obtained from a single crystal held at 100 K.

We report a new method based on a highly efficient fixed-target silicon nitride chip system (Oghbaey *et al.*, 2016; Mueller *et al.*, 2015). This system allows for data to be measured at room temperature from microcrystals in the same manner using synchrotron or XFEL radiation. Our fixed-target approach enables time- and sample-efficient data collection by both SFX and serial synchrotron crystallography (SSX), and simultaneously minimizes any differences in structure by eliminating the experimental variables outlined above. It is also well suited to tracking functionally relevant changes in redox enzymes as X-ray generated solvated electrons drive the enzyme along the catalytic pathway in the crystal when exposed to the X-ray beam. Multiple serial structures (MSS) can be obtained from a set of crystals on a single fixed target, as sequential exposure events to each crystal are binned and processed as individual dose-dependent datasets. This is analogous to the multiple structures from one crystal (MSOX) approach (Horrell *et al.*, 2016, 2018), previously applied to the measurement of repeated datasets from the same exposed region of a single large crystal to produce a dose series. In comparison, our new approach exposes each crystal to the X-ray beam for only a few tens of milliseconds and is well suited for high-throughput structure solution from microcrystals held at room temperature, using XFEL or synchrotron radiation sources.

Herein, we have chosen to use an extracellular dye-type heme peroxidase found in *Streptomyces lividans* and referred to as DtpAa. Using DtpAa as the exemplar, we describe the application of our method of combined SFX and MSS experiments, though the method can be used for any redox-sensitive system. Starting in the catalytic resting state, our approach reveals multiple well resolved structural states of the enzyme, with a low-dose synchrotron MSS structural series showing clearly resolved changes to the active-site region of the enzyme within tens of milliseconds. Extrapolation of varying structural parameters to zero dose produced a close match to the damage-free structure determined using SFX. Thus, a low-dose series of synchrotron MSS is anchored by a damage-free SFX structure, both being determined using the same fixed-target serial sample-delivery system.

We present this approach as a general method to efficiently collect both SFX and SSX data under near-identical conditions, characterize subtle site-specific changes caused by X-rays in proteins and allow direct comparison of, and extrapolation to, damage-free XFEL structures from low-dose synchrotron models.

2. Materials and methods

2.1. Sample preparation

The SLI_2602 gene encoding DtpAa was amplified from the genomic DNA of *S. lividans* strain 1326 (*S. lividans* stock

number 1326, John Innes Centre) by polymerase chain reaction (the primers used for amplification are reported in the Supporting information). The gene was subsequently cloned into the NdeI and HindIII sites of a pET28a vector (Novagen) to create an N-terminal His_6-tagged construct (pET2602) for overexpression in *Escherichia coli*. The pET2602 vector was transformed into *E. coli* BL21 (DE3) cells. Overnight precultures [low-salt Luria–Bertani (LB) medium; Melford] were successively used to inoculate 1.4 l of high-salt LB medium with 50 mg ml^{-1} kanamycin and were grown at 37°C, 180 rev min^{-1}. At an OD_{600} of 1.0–1.2, 5-aminolevulinic acid (0.25 m*M* final concentration) and iron citrate (100 µ*M* final concentration) were added consecutively for their use as a heme precursor and iron supplement, respectively. Cultures were then induced by adding isopropyl *β*-D-thiogalactopyranoside (Melford) to a final concentration of 0.5 m*M*, and carbon monoxide gas was bubbled through the culture for 30–60 s. Flasks were then sealed and incubated for a further 18 h at 30°C and 100 rev min^{-1}. Cells were harvested via centrifugation (10 000*g*, 10 min, 4°C) and the cell pellet resuspended in 50 m*M* Tris–HCl, 500 m*M* NaCl (Fisher) and 20 m*M* imidazole (Sigma) at pH 8 (buffer A). The resuspended cell suspension was lysed using an EmulsiFlex-C5 cell disrupter (Avestin) followed by centrifugation (22 000*g*, 30 min, 4°C). The clarified supernatant was loaded onto a 5 ml nickel–nitrilotriacetic acid–sepharose column (GE Healthcare) equilibrated with buffer A and eluted by a linear imidazole gradient using buffer B (buffer A with 500 m*M* imidazole). The DtpAa peak eluting at approximately 30–40% buffer B was pooled and concentrated using a Centricon (VivaSpin) with a 10 kDa cut-off at 4°C followed by application to an S200 Sephadex column (GE Healthcare) equilibrated with 20 m*M* NaPi, 100 m*M* NaCl, pH 7. A major peak eluted consistent with a monomeric species with fractions assessed by SDS–PAGE then concentrated and stored at −20°C. DtpAa concentrations were determined by UV–vis spectroscopy (Varian Cary 60 UV–vis spectrophotometer) using an extinction coefficient at 280 nm of 46 075 M^{-1} cm^{-1}.

Microcrystals were grown in batch (typically 0.4–0.5 ml total volume) by mixing in a 1:1 ratio a 6.5 mg ml^{-1} DtpAa protein solution with a precipitant solution containing 20% PEG 6000, 100 m*M* HEPES pH 7.0, to typical dimensions of 15 µm. Silicon nitride fixed-target chips with either 7, 12, 14 or 37 µm apertures at their narrowest opening and a nominal capacity of 25 600 crystals were loaded as previously described (Ebrahim *et al.*, 2019), with an identical loading protocol used both at Diamond and the SPring-8 Ångstrom Free Electron Laser (SACLA) XFEL. In brief, chips were loaded with 100–200 µl of microcrystal suspension within a humidity enclosure (Solo Containment, Cheshire, England) and sealed between two layers of 6 µm thick Mylar.

2.2. Data collection and fixed-target motion

SFX data were measured at SACLA beamline BL2 EH3 using an X-ray energy of 10.0 keV, a pulse length 10 fs and a repetition rate of 30 Hz, with the beam attenuated to 13% of full flux. Chips were translated within the interval between X-ray pulses, ensuring that the chip had stopped at the centre of each crystal position (the centre of the aperture) and was exposed only once to X-rays, before moving to the next position during the next pulse interval. Data were typically collected from all 25 600 positions on a chip in 14 min using the SACLA MPCCD detector (Kameshima *et al.*, 2014), with experiments performed in a helium chamber to minimize air scatter. A modified custom entry port to the helium chamber permitted rapid exchange of chips, meaning that measurement from all positions with subsequent sample exchange and alignment interval of <5 min between data collections allowed a sustained data-collection rate of just over 3 chips per hour. While sufficient data for structure solution and refinement were obtained from crystals mounted on only 2 chips (*ca* 13 000 hits), for the structure described here data were collected from a total of 11 chips, still in under 4 h of beam time, in order to increase the redundancy of the data and the quality of the maps obtained.

Data collection at beamline I24, Diamond Light Source was carried out using an unattenuated X-ray beam of energy 12.8 keV and a Pilatus3 6M detector in shutterless mode. To form a dose-dependent series of DtpAa structures, 5 (MSS-1) and 10 (MSS-2) sequential diffraction patterns were measured at each crystal position each with an exposure time of 10 ms and subsequently binned into one dataset per dose interval (Fig. 1). The series of exposures at each chip position was individually triggered via a Keysight 33500B signal generator which was itself triggered by a DeltaTau Geobrick LV-IMS-II stage controller when each desired crystal position had been attained. The X-ray shutter was not closed between apertures on a chip and remained open for the duration of the experiment. X-ray fluxes were measured using a silicon PIN diode as previously described (Owen *et al.*, 2006) and were 3.2×10^{12} and 3.0×10^{12} photons s^{-1} for MSS-1 and MSS-2, respectively. The corresponding beamsizes (measured using a knife-edge scan) were 7×7 and 9×8 µm, respectively. Absorbed doses were estimated using *RADDOSE-3D* (Zeldin *et al.*, 2013), with dose increments corresponding to the total dose accumulated within the exposure time of the first image, and are detailed in Table S1 in the Supporting information. We note that crystals will be subjected to a small additional absorbed dose during deceleration of the stages prior to the time when the detector starts recording the first diffraction image. While challenging to accurately determine, we estimate that an upper bound for this dose is ∼3 kGy. These experiments were carried out using the same fixed-target chips and translation system as used at SACLA. Details of the datasets are given in Table S1.

2.3. Data analysis

For data measured at SACLA, initial hit finding at the beamline was carried out in *CHEETAH* (Barty *et al.*, 2014). Peak-finding, integration and merging were all performed in *CrystFEL* (White *et al.*, 2016). Data from Diamond beamline I24 were indexed using *DIALS* (version 1.8.5) (Winter *et al.*,

2018) with subsequent scaling and merging performed using *PRIME* (Uervirojnangkoorn *et al.*, 2015). MSS data from beamline I24 consisted of cbf image files numbered sequentially. These were binned into dose points using a simple partitioning script. Multiple lattices were allowed during indexing.

In both cases resolution limits were assessed using $CC_{1/2}$ and R_{split} parameters (White *et al.*, 2013, 2016) together with behaviour in refinement. Structures were solved by molecular replacement using a starting model obtained from a small number of larger DtpAa crystals mounted between two layers of thin film (Axford *et al.*, 2012; Doak *et al.*, 2018) and used to obtain rotation wedges. Water molecules were removed from this model prior to refinement. Structures were refined initially using *REFMAC*5 (Murshudov *et al.*, 2011) within the *CCP*4 suite (Winn *et al.*, 2011) and later in *PHENIX* (Adams *et al.*, 2010) and rebuilt between refinement cycles using *Coot* (Emsley *et al.*, 2010). Atoms not well supported by electron density (primarily surface side chains) were deleted from the model. Validation was performed using *MolProbity* (Richardson *et al.*, 2018), *QCCheck* and tools within *Coot* and *PHENIX*. Estimates of bond-length error were calculated from the coordinate diffraction precision index as described (Gurusaran *et al.*, 2014) using the online diffraction precision indicator (DPI) server (Kumar *et al.*, 2015).

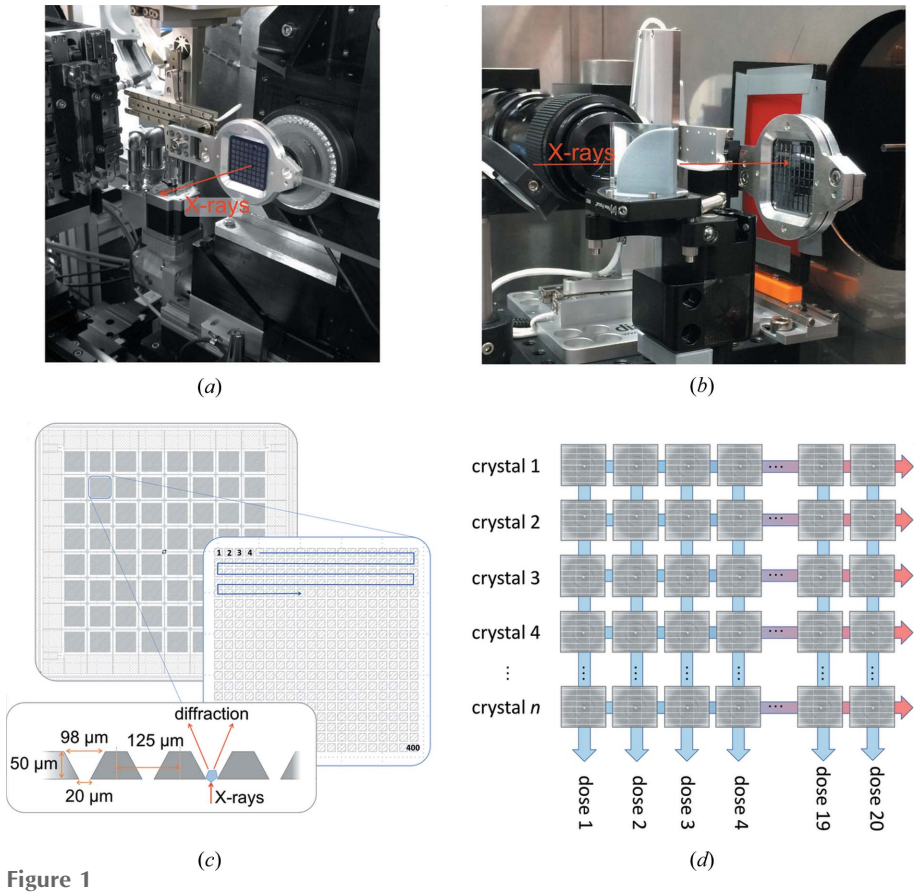


Figure 1
Fixed-target instrumentation in place at (*a*) beamline I24, Diamond Light Source and (*b*) beamline BL2 EH3, SACLA. (*c*) Schematic of fixed target used showing layout of 8 × 8 'city blocks', each comprising 20 × 20 apertures. Shown is a zoomed-in view of a single city block with motion path followed and chip cross-section. (*d*) Formation of dose-resolved datasets by collecting multiple images at each chip aperture. For XFEL data collection, only a single dose point is recorded at each position.

3. Results

3.1. Sample- and time-efficient serial data collection at synchrotron microfocus and XFEL beamlines using silicon nitride fixed-target chips

We used high-capacity silicon nitride fixed targets or 'chips' each containing 25 600 apertures based on those described previously (Oghbaey *et al.*, 2016) to hold the microcrystals used to determine room-temperature serial crystallography structures of DtpAa. Importantly, this sample-delivery system was used in a near-identical manner for both the SFX and the MSS experiments, (Fig. 1), allowing for a direct comparison of the resulting structures. Typically hit rates (we define hit rate as the percentage of frames collected that could be indexed) of ∼30% were achieved on each chip allowing structures to be determined in a highly time- and sample-efficient manner. The volume of microcrystal suspension required per chip was typically 100–200 µl. A schematic of the chip setup and methodological approach is shown in Fig. 1.

3.2. Damage-free DtpAa structure using serial femtosecond crystallography

To produce an 'anchor' structure of DtpAa, *i.e.* resting state ferric, free of any effects of the X-ray beam on the structure, we used the SACLA XFEL (Ishikawa *et al.*, 2012) beamline BL2 EH3 to perform SFX with an X-ray energy of 10 keV, a pulse length of 10 fs with a 1.25 × 1.34 µm beam and a pulse energy of 289 µJ pulse^{-1}. The chip was translated between apertures in the 33 ms separating the 30 Hz XFEL pulses, with a single image recorded at each position. The SFX structure was determined to a resolution of 1.88 Å from a total of 72 615 indexed and merged diffraction patterns [Fig. 2(*a*), Table S1]. The overall structure reveals a ferredoxin-like fold typical of dye decolourizing peroxidases (Sugano, 2009) with two DtpAa monomers in the crystallographic asymmetric unit. The structure was of high quality (Table S1) and refined to an R_{work} and R_{free} of 13.2% and 16.7%, respectively. The refined model exhibited a mean-determined *B* factor of 34.7 Å^2.

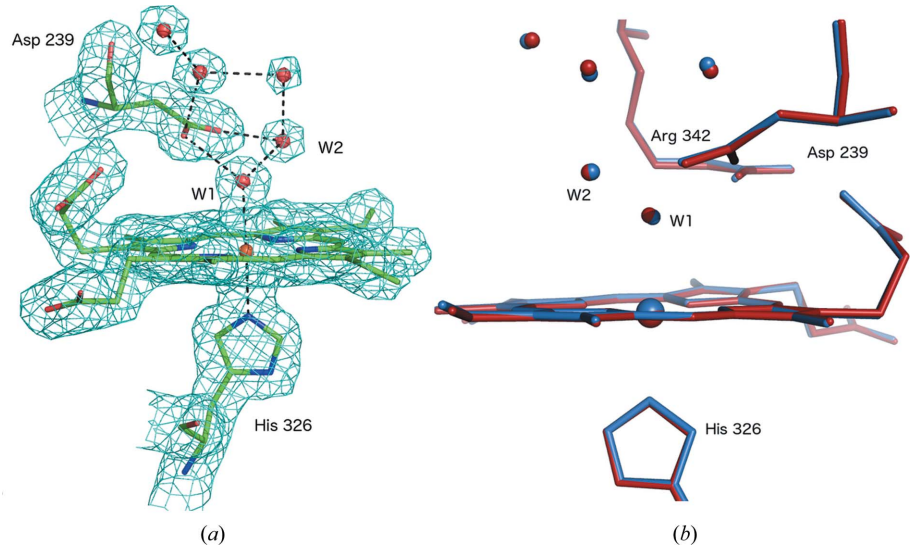


Figure 2
(*a*) $2F_o$–F_c electron-density map contoured at 1σ for the damage-free SFX structure of DtpAa at 1.88 Å resolution, showing the clear and well resolved water network within the heme pocket. Water molecules interact extensively with the pocket residue Asp239 as well as with Arg369 (omitted for clarity). (*b*) Superposition of the SFX structure (blue) with the 32.8 kGy SSX structure (red). Small changes to the heme-pocket water network are apparent even at this low dose.

The heme Fe is six-coordinate with residue His326 acting as the proximal ligand with an Fe—N bond length of 2.19 Å (we note here that monomer B appears to be inactive and is consequently not discussed further). The distal heme coordination site is occupied by a well defined, full occupancy water molecule (W1), bound to the Fe at a distance of 2.40 Å. A number of further, well defined water molecules occupy the remainder of the heme distal pocket [Fig. 2(*a*)]. W1 is hydrogen bonded to a second water, W2, at a distance of 2.68 Å and also interacts with the charged side chains of Asp239 (2.92 Å) and Arg342 (2.74 Å). Interestingly, the side chains of these two amino acids are only 3.13 Å apart (Arg $N_{\eta 1}$ to Asp $O_{\delta 2}$) suggesting a charge-based interaction.

3.3. DtpAa structures from serial synchrotron crystallography

Serial synchrotron crystallography was carried out at Diamond Light Source beamline I24 at an X-ray energy of 12.8 keV using the same chip and translation system as used for SFX at SACLA. The beam size and flux were measured immediately prior to each experiment, see Materials and methods for details, with approximate values of 7 × 7 µm and 3.1×10^{12} photons s^{-1}. Following each translation of the chip to bring a fresh aperture/crystal into the beam, a series of 10 ms exposures were recorded using the PILATUS3 detector in shutterless mode. This allowed multiple successive snapshots of the same microcrystal within 100 ms. Following exposure of a crystal to the X-ray beam, the chip was translated to the next aperture position and the process repeated [shown schematically in Fig. 1(*d*)]. Using this approach, the total experimental time per fully loaded chip for ten dose points is 45 min, but the total exposure (and hence the absorbed dose) of any individual microcrystal is low and multiple time- (dose-) resolved structures are obtained from a single fixed target. We note here that the 10 ms minimum exposure time was imposed by the maximum frame rate of the current detector available (PILATUS3 6M) and not by limitations arising from the fixed-target movement or synchronization of the target and the X-ray beam. Diffraction images were indexed and integrated independently using *DIALS* (`dials.stills_process`) (Winter *et al.*, 2018) with a simple image-binning procedure used to assign the resulting data to dose bins [Fig. 1(*d*)]. Data within each dose bin were then scaled and merged together using *PRIME* (Uervirojnangkoorn *et al.*, 2015) to form dose-resolved datasets. Using this approach, a complete dataset for each X-ray dose was formed and the corresponding structure refined using the methods described above. The scaling and refinement statistics for each structure are given in Table S1. We first describe an MSS experiment series comprising five dose points with a dose increment of 32.8 kGy (MSS1). An increase in unit-cell volume and trends in scaling statistics clearly indicate the onset of global radiation damage resulting from disorder in the crystalline lattice as dose is accumulated (Fig. S2). The initial resolution was 1.78 Å with only a limited loss of diffracting power/resolution during the 50 ms of total exposure for each microcrystal. Dataset 1 of this series (MSS1-ds1, 32.8 kGy) reveals a six-coordinate heme with a slightly lengthened Fe—O bond at 2.48 Å compared with the SFX structure [Fig. 2(*b*)]. A superposition of MSS1-ds1 with the SFX structure is shown in Fig. 2(*b*). With increasing dose, distinct changes occur around the heme pocket consistent with reduction of the heme iron by X-ray generated solvated photoelectrons (Beitlich *et al.*, 2007; Kekilli *et al.*, 2017). In MSS1-ds2, the Fe—O bond is 2.70 Å with this continuing to lengthen until the last dataset (MSS1-ds5, 164.0 kGy) where it reaches a value of 2.97 Å (Table S3).

In order to provide additional dose points and obtain higher dose SSX structures, a second MSS series was measured with an increased incremental dose value (MSS2). In this case, 10 × 10 ms exposures were measured per crystal position (Table S1 and Fig. S3) with a dose interval of 39.2 kGy. The initial dataset was refined to a resolution of 1.70 Å with the resolution remaining as high as 1.93 Å by dose point 6. After this point (60 ms exposure) the resolution limit decayed, with structure refinement only carried out to 2.18 Å resolution by the last dataset (MSS2-ds8). For comparison, dataset 10 reached only 2.7 Å resolution. In this series the first dataset (MSS2-ds1), associated with a dose of 39.2 kGy, exhibited a Fe–W1 distance of 2.50 Å. This distance increased in successive dose point structures, reaching 2.64 Å in ds3, 2.91 Å in MSS2-ds5 and 3.76 Å in MSS2-ds8 (Fig. 3, Tables S1 and S2).

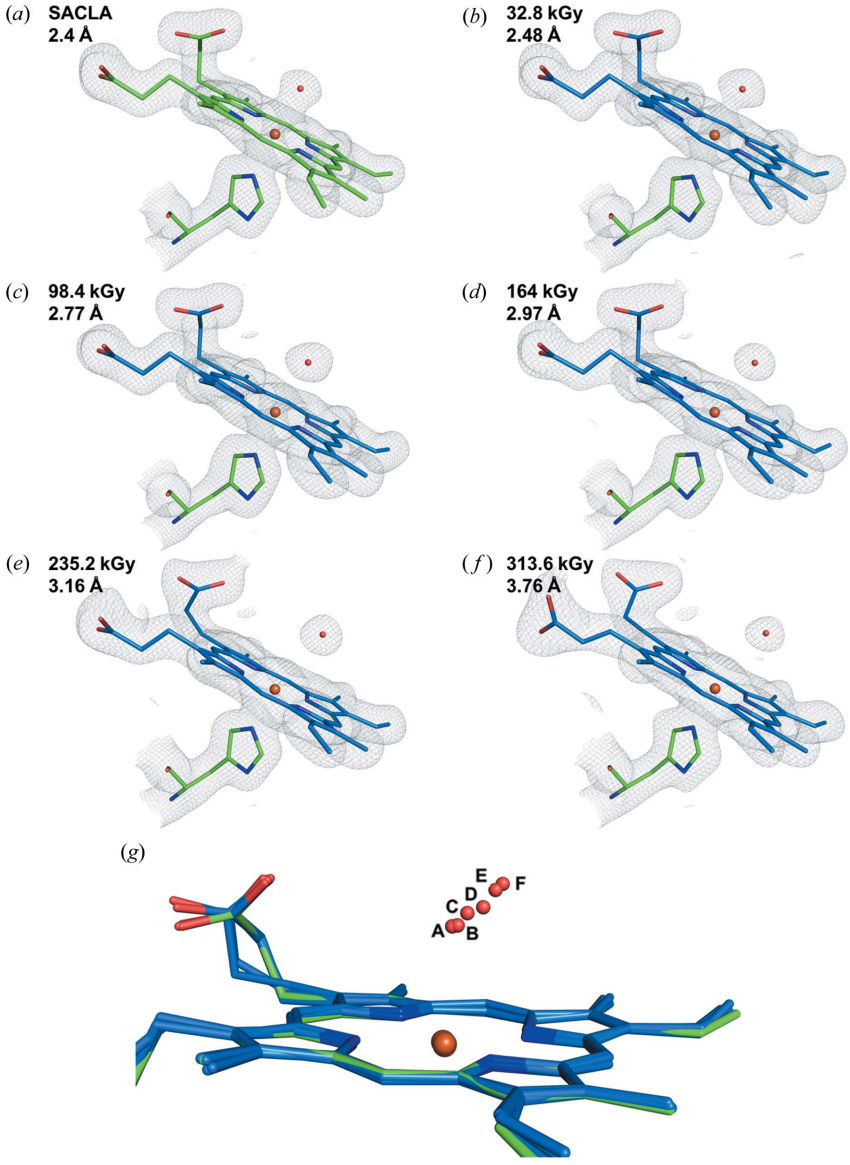


Figure 3
$2F_o$–F_c electron-density maps contoured at 1σ for the heme environment of DtpAa in (*a*) the SFX dataset from SACLA and (*b*–*f*) selected structures from the two MSS series. (*g*) Superposition of selected structures revealing the dose-dependent migration of the water molecule W1 away from the heme Fe. The SFX structure is shown in green with MSS in blue.

Additional structural changes were evident in the heme pocket, with rearrangement of water structures and a flip of a heme propionate as dose was accumulated (Figs. S4 and S5).

The Fe–O distance in all structures from both MSS series is plotted in Fig. 4 and migration of the water away from the heme Fe is shown in Fig. 3(*g*).

It is of considerable interest to compare low-dose synchrotron structures and damage-free XFEL structures determined under near-identical experimental conditions, and to explore if dose-series data may be used to extrapolate back to the 'native' state present prior to X-ray exposure, a so-called 'zero-dose extrapolation'. This approach is analogous to the zero-dose extrapolation of diffraction intensities within conventional single-crystal datasets that has been described previously (Diederichs *et al.*, 2003; Diederichs, 2006; Diederichs & Junk, 2009). In this way, the SFX structure provides a starting point from which synchrotron datasets (inevitably incurring radiation damage and consequent structural change) may be interpreted. A vivid example is shown in a plot of the Fe–W1 distance in the SFX 'anchor' structure and both MSS series, Fig. 4. A near-linear relationship is observed, demonstrating that water migration away from the Fe is dose-dependent under the conditions used. A linear fit to the data yields an intercept (*i.e.* extrapolated to zero dose) of 2.37 Å, which is very close to the value in the SFX structure (2.40 Å) at comparable resolution and within the experimental error for this bond length in the room-temperature structures. The SFX and MSS datasets were deposited in the Protein Data Bank with accession codes as indicated in the supplementary tables in the Supporting information.

4. Discussion

To our knowledge, this is the first reported method for directly comparing dose-resolved serial synchrotron and XFEL structures of radiation-sensitive metalloproteins using the same microcrystal preparations and sample-delivery system. The resolutions achieved with each X-ray source are comparable (1.88 Å SFX and 1.70 Å SSX) allowing the direct comparison of structural features. The SSX data collection allowed sequences of 5–10 MSS dose points to be measured in a time of tens to hundreds of milliseconds per microcrystal. The effective resolution remained high for a substantial proportion of each series. Interestingly, the MSS structures of DtpAa showed well resolved water molecules (Figs. 2 and 3) in all structures, indicating that the progression of reactions within the exposed crystal volume is relatively uniform.

Determining a sequence of dose-dependent structures from the same microcrystals allows subtle and relatively rapidly occurring structural changes to be resolved. In the case of DtpAa, an elongation of the Fe—water bond and eventual bond breakage were observed during tens of milliseconds of exposure to an intense microfocus X-ray beam. By obtaining MSS throughout the process, sufficient data points were recorded in order to be able to fit a function with confidence and allow a zero-dose extrapolation to be made. This provided a close approximation to the structure determined by SFX, providing an alternative approach to obtain a good approx-

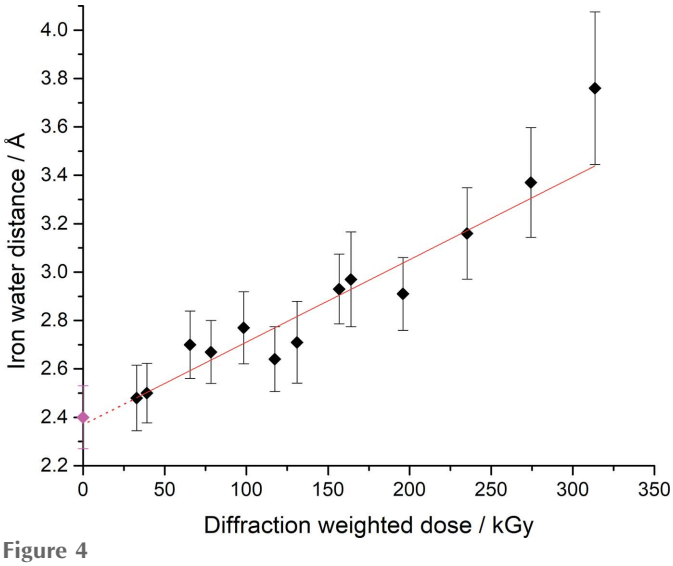


Figure 4
Plot of Fe–W1 distance as a function of X-ray dose from the two measured MSS series. The SFX structure determined using SACLA is plotted as the zero-dose point (magenta). The elongation of the bond length with dose is well fitted by a linear function (red line). The deviation at higher doses is associated with a dissociation of W1 from the immediate vicinity of the heme Fe. The extrapolation to zero dose (dashed red line) gives a value of 2.37 (±0.05) Å which is very close to the 2.40 (±0.13) Å value of this parameter in the SFX structure. Error bars shown are the estimated standard uncertainty in bond length obtained from the DPI value of the Fe and W1 atoms (see Materials and methods).

imation of the 'gold standard' damage-free structure obtained using an XFEL.

4.1. How close can we get to the damage-free enzyme structure using synchrotron radiation?

Despite a relatively low absorbed dose of 32.8 kGy in the MSS1-ds1 dataset, the structure is not identical to that determined by SFX [Fig. 2(*b*)]. Notably, the iron—water bond in MSS1-ds1, the shortest out of all the SSX structures, is elongated compared with the SFX structure. A simple linear fit of the plot of iron—water bond length as a function of dose allowed an extrapolation of the SSX data to zero dose (*y*-axis intercept), yielding a comparable distance to that observed in the SFX structure (Fig. 4).

While not the main focus of this report, the MSS series we present reveal a number of structural states populated during the initial response of DtpAa to X-rays. The elongation of the iron—water bond is consistent with Fe^{III} to Fe^{II} reduction. This reduction is consistent with the generation of solvated electrons by the interaction of X-rays with solvent molecules in the crystal (Kwon *et al.*, 2017; Moody & Raven, 2018). In contrast to the situation at 100 K, where X-ray generated radicals are largely immobilized, room-temperature reactions that involve mass transport may allow such radicals to contribute to the structural changes that our methods allow to be resolved. The relevance of these structures to the function of this class of enzymes will be explored in detail elsewhere.

An additional advantage of our approach is that the same methodology and sample-delivery system is used at synchrotron and XFEL sources/beamlines. This allows for an effective comparison of the structures produced by each X-ray source, allowing the use of comparable crystal sizes, temperatures and sample-delivery methods, factors that might otherwise cause structural heterogeneity.

In summary, we have shown that microcrystals loaded into fixed-target silicon nitride chips can be efficiently employed for data collection at both synchrotron and XFEL sources, allowing near-identical conditions for experiments. Using this technology, we have characterized subtle site-specific changes caused by X-rays in proteins, and directly compared low-dose synchrotron models with, and extrapolation to, damage-free SFX structures. Our method has the potential to be applied to a wide range of enzymes and other proteins especially those that are highly sensitive to radiation damage, including the characterization of electron-driven mechanistic steps in detail through a dose series such as redox reactions in redox metalloenzymes. On a practical level, our approach can be used to extract functionally relevant features of damage-free SFX structures (which require access to scarce beam time at XFELs), reconstructed from extrapolation of MSS determined at multiple low-dose points. Notably, the time interval per MSS structure will be reduced by at least an order of magnitude with upcoming advances in detectors and synchrotron brilliance.

Acknowledgements

XFEL experiments were performed at BL2 EH3 of SACLA with the approval of the Japan Synchrotron Radiation Research Institute (JASRI) (Proposal No. 2017B8014). We thank the SACLA beamline staff for technical assistance. We acknowledge the contributions of Drs Minoru Kubo and Takashi Nomura (University of Hyogo) and Takehiko Tosha (RIKEN). The authors thank the beamline staff of BL3 at SACLA. Synchrotron data were measured on beamline I24 at Diamond Light Source under proposal NT14493. Support for travel from the UK XFEL Hub at Diamond Light Source is gratefully acknowledged. We thank Thomas White for helpful discussions regarding the processing of SFX data in *CrystFEL*. R. L. Owen, J. A. R. Worrall, R. W. Strange, D. Axford and M. A. Hough designed the experiment. A. Ebrahim, T. Moreno-Chicano, A. K. Chaplin and J. A. R. Worrall were involved in sample preparation. A. Ebrahim, T. Moreno-Chicano, R. W. Strange, J. A. R. Worrall, M. V. Appleby, J. H. Beale, D. Axford, D. A. Sherrell, H. M. E. Duyvesteyn, H. Sugimoto, K. Tono, S. Owada, R. L. Owen and M. A. Hough participated in data collection. A. Ebrahim, T. Moreno-Chicano, H. M. E. Duyvesteyn, M. V. Appleby, D. Axford, R. L. Owen and M. A. Hough analyzed the data. A. Ebrahim, R. L. Owen and M. A. Hough wrote the manuscript with contributions from all authors.

Funding information

T. Moreno-Chicano was supported by the Leverhulme Trust award RPG-2014-355 to M. A. Hough and R. W. Strange. M. A. Hough acknowledges support from BBSRC award BB/

M022714/1 and J. A. R. Worrall, D. A. Sherrell, D. Axford, R. L. Owen, M. A. Hough and R. W. Strange acknowledge support from a BBSRC Japan–UK International Partnering Award, BB/R021015/1.

References

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. & Zwart, P. H. (2010). *Acta Cryst.* D**66**, 213–221.

Andersson, R., Safari, C., Dods, R., Nango, E., Tanaka, R., Yamashita, A., Nakane, T., Tono, K., Joti, Y., Båth, P., Dunevall, E., Bosman, R., Nureki, O., Iwata, S., Neutze, R. & Brändén, G. (2017). *Sci. Rep.* **7**, 4518.

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.

Barty, A., Kirian, R. A., Maia, F. R. N. C., Hantke, M., Yoon, C. H., White, T. A. & Chapman, H. (2014). *J. Appl. Cryst.* **47**, 1118–1131.

Beitlich, T., Kühnel, K., Schulze-Briese, C., Shoeman, R. L. & Schlichting, I. (2007). *J. Synchrotron Rad.* **14**, 11–23.

Casadei, C. M., Gumiero, A., Metcalfe, C. L., Murphy, E. J., Basran, J., Concilio, M. G., Teixeira, S. C., Schrader, T. E., Fielding, A. J., Ostermann, A., Blakeley, M. P., Raven, E. L. & Moody, P. C. (2014). *Science*, **345**, 193–197.

Chreifi, G., Baxter, E. L., Doukov, T., Cohen, A. E., McPhillips, S. E., Song, J., Meharenna, Y. T., Soltis, S. M. & Poulos, T. L. (2016). *Proc. Natl Acad. Sci. USA*, **113**, 1226–1231.

Denisov, I. G., Victoria, D. C. & Sligar, S. G. (2007). *Radiat. Phys. Chem.*, **76**, 714–721.

Diederichs, K. (2006). *Acta Cryst.* D**62**, 96–101.

Diederichs, K. & Junk, M. (2009). *J. Appl. Cryst.* **42**, 48–57.

Diederichs, K., McSweeney, S. & Ravelli, R. B. G. (2003). *Acta Cryst.* D**59**, 903–909.

Doak, R. B., Nass Kovacs, G., Gorel, A., Foucar, L., Barends, T. R. M., Grünbein, M. L., Hilpert, M., Kloos, M., Roome, C. M., Shoeman, R. L., Stricker, M., Tono, K., You, D., Ueda, K., Sherrell, D. A., Owen, R. L. & Schlichting, I. (2018). *Acta Cryst.* D**74**, 1000–1007.

Ebrahim, A., Appleby, M. V., Axford, D., Beale, J., Moreno-Chicano, T., Sherrell, D. A., Strange, R. W., Hough, M. A. & Owen, R. L. (2019). *Acta Cryst.* D**75**, 151–159.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Fischer, M., Shoichet, B. K. & Fraser, J. S. (2015). *ChemBioChem*, **16**, 1560–1564.

Garman, E. F. & Weik, M. (2017). *Methods Mol. Biol.* **1607**, 467–489.

Gumiero, A., Metcalfe, C. L., Pearson, A. R., Raven, E. L. & Moody, P. C. (2011). *J. Biol. Chem.* **286**, 1260–1268.

Gurusaran, M., Shankar, M., Nagarajan, R., Helliwell, J. R. & Sekar, K. (2014). *IUCrJ*, **1**, 74–81.

Halsted, T. P., Yamashita, K., Hirata, K., Ago, H., Ueno, G., Tosha, T., Eady, R. R., Antonyuk, S. V., Yamamoto, M. & Hasnain, S. S. (2018). *IUCrJ*, **5**, 22–31.

Holton, J. M. (2009). *J. Synchrotron Rad.* **16**, 133–142.

Horrell, S., Antonyuk, S. V., Eady, R. R., Hasnain, S. S., Hough, M. A. & Strange, R. W. (2016). *IUCrJ*, **3**, 271–281.

Horrell, S., Kekilli, D., Sen, K., Owen, R. L., Dworkowski, F. S. N., Antonyuk, S. V., Keal, T. W., Yong, C. W., Eady, R. R., Hasnain, S. S., Strange, R. W. & Hough, M. A. (2018). *IUCrJ*, **5**, 283–292.

Ishikawa, T., Aoyagi, H., Asaka, T., Asano, Y., Azumi, N., Bizen, T., Ego, H., Fukami, K., Fukui, T., Furukawa, Y., Goto, S., Hanaki, H., Hara, T., Hasegawa, T., Hatsui, T., Higashiya, A., Hirono, T., Hosoda, N., Ishii, M., Inagaki, T., Inubushi, Y., Itoga, T., Joti, Y., Kago, M., Kameshima, T., Kimura, H., Kirihara, Y., Kiyomichi, A., Kobayashi, T., Kondo, C., Kudo, T., Maesaka, H., Marechal, X. M., Masuda, T., Matsubara, S., Matsumoto, T., Matsushita, T., Matsui, S., Nagasono, M., Nariyama, N., Ohashi, H., Ohata, T., Ohshima, T., Ono, S., Otake, Y., Saji, C., Sakurai, T., Sato, T., Sawada, K., Seike, T., Shirasawa, K., Sugimoto, T., Suzuki, S., Takahashi, S., Takebe, H., Takeshita, K., Tamasaku, K., Tanaka, H., Tanaka, R., Tanaka, T., Togashi, T., Togawa, K., Tokuhisa, A., Tomizawa, H., Tono, K., Wu, S. K., Yabashi, M., Yamaga, M., Yamashita, A., Yanagida, K., Zhang, C., Shintake, T., Kitamura, H. & Kumagai, N. (2012). *Nat. Photonics* **6**, 540–544.

Kameshima, T., Ono, S., Kudo, T., Ozaki, K., Kirihara, Y., Kobayashi, K., Inubushi, Y., Yabashi, M., Horigome, T., Holland, A., Holland, K., Burt, D., Murao, H. & Hatsui, T. (2014). *Rev. Sci. Instrum.* **85**, 033110.

Keedy, D. A., Hill, Z. B., Biel, J. T., Kang, E., Rettenmaier, T. J., Brandao-Neto, J., Pearce, N. M., von Delft, F., Wells, J. A. & Fraser, J. S. (2018). *Elife*, **7**, e36307.

Keedy, D. A., Kenner, L. R., Warkentin, M., Woldeyes, R. A., Hopkins, J. B., Thompson, M. C., Brewster, A. S., Van Benschoten, A. H., Baxter, E. L., Uervirojnangkoorn, M., McPhillips, S. E., Song, J., Alonso-Mori, R., Holton, J. M., Weis, W. I., Brunger, A. T., Soltis, S. M., Lemke, H., Gonzalez, A., Sauter, N. K., Cohen, A. E., van den Bedem, H., Thorne, R. E. & Fraser, J. S. (2015). *Elife*, **4**, e07574.

Kekilli, D., Moreno-Chicano, T., Chaplin, A. K., Horrell, S., Dworkowski, F. S. N., Worrall, J. A. R., Strange, R. W. & Hough, M. A. (2017). *IUCrJ*, **4**, 263–270.

Kumar, K. S. D., Gurusaran, M., Satheesh, S. N., Radha, P., Pavithra, S., Thulaa Tharshan, K. P. S., Helliwell, J. R. & Sekar, K. (2015). *J. Appl. Cryst.* **48**, 939–942.

Kwon, H., Smith, O., Raven, E. L. & Moody, P. C. E. (2017). *Acta Cryst.* D**73**, 141–147.

Lomb, L., Barends, T. R. M., Kassemeyer, S., Aquila, A., Epp, S. W., Erk, B., Foucar, L., Hartmann, R., Rudek, B., Rolles, D., Rudenko, A., Shoeman, R. L., Andreasson, J., Bajt, S., Barthelmess, M., Barty, A., Bogan, M. J., Bostedt, C., Bozek, J. D., Caleman, C., Coffee, R., Coppola, N., DePonte, D. P., Doak, R. B., Ekeberg, T., Fleckenstein, H., Fromme, P., Gebhardt, M., Graafsma, H., Gumprecht, L., Hampton, C. Y., Hartmann, A., Hauser, G., Hirsemann, H., Holl, P., Holton, J. M., Hunter, M. S., Kabsch, W., Kimmel, N., Kirian, R. A., Liang, M. N., Maia, F. R. N. C., Meinhart, A., Marchesini, S., Martin, A. V., Nass, K., Reich, C., Schulz, J., Seibert, M. M., Sierra, R., Soltau, H., Spence, J. C. H., Steinbrener, J., Stellato, F., Stern, S., Timneanu, N., Wang, X. Y., Weidenspointner, G., Weierstall, U., White, T. A., Wunderer, C., Chapman, H. N., Ullrich, J., Struder, L. & Schlichting, I. (2011). *Phys. Rev. B*, **84**, 214111.

Moody, P. C. E. & Raven, E. L. (2018). *Acc. Chem. Res.* **51**, 427–435.

Mueller, C., Marx, A., Epp, S. W., Zhong, Y., Kuo, A., Balo, A. R., Soman, J., Schotte, F., Lemke, H. T., Owen, R. L., Pai, E. F., Pearson, A. R., Olson, J. S., Anfinrud, P. A., Ernst, O. P. & Dwayne Miller, R. J. (2015). *Struct. Dyn.* **2**, 054302.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Nass, K., Foucar, L., Barends, T. R. M., Hartmann, E., Botha, S., Shoeman, R. L., Doak, R. B., Alonso-Mori, R., Aquila, A., Bajt, S., Barty, A., Bean, R., Beyerlein, K. R., Bublitz, M., Drachmann, N., Gregersen, J., Jönsson, H. O., Kabsch, W., Kassemeyer, S., Koglin, J. E., Krumrey, M., Mattle, D., Messerschmidt, M., Nissen, P., Reinhard, L., Sitsel, O., Sokaras, D., Williams, G. J., Hau-Riege, S., Timneanu, N., Caleman, C., Chapman, H. N., Boutet, S. & Schlichting, I. (2015). *J. Synchrotron Rad.* **22**, 225–238.

Oghbaey, S., Sarracini, A., Ginn, H. M., Pare-Labrosse, O., Kuo, A., Marx, A., Epp, S. W., Sherrell, D. A., Eger, B. T., Zhong, Y., Loch, R., Mariani, V., Alonso-Mori, R., Nelson, S., Lemke, H. T., Owen, R. L., Pearson, A. R., Stuart, D. I., Ernst, O. P., Mueller-Werkmeister, H. M. & Miller, R. J. D. (2016). *Acta Cryst.* D**72**, 944–955.

Owen, R. L., Rudiño-Piñera, E. & Garman, E. F. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 4912–4917.
Richardson, J. S., Williams, C. J., Hintze, B. J., Chen, V. B., Prisant, M. G., Videau, L. L. & Richardson, D. C. (2018). *Acta Cryst.* D**74**, 132–142.
Schlichting, I. (2015). *IUCrJ*, **2**, 246–255.
Sugano, Y. (2009). *Cell. Mol. Life Sci.* **66**, 1387–1403.
Uervirojnangkoorn, M., Zeldin, O. B., Lyubimov, A. Y., Hattne, J., Brewster, A. S., Sauter, N. K., Brunger, A. T. & Weis, W. I. (2015). *Elife*, **4**, e05421
Waldron, K. J., Rutherford, J. C., Ford, D. & Robinson, N. J. (2009). *Nature*, **460**, 823–830.
Weik, M. & Colletier, J.-P. (2010). *Acta Cryst.* D**66**, 437–446.
White, T. A., Barty, A., Stellato, F., Holton, J. M., Kirian, R. A., Zatsepin, N. A. & Chapman, H. N. (2013). *Acta Cryst.* D**69**, 1231–1240.
White, T. A., Mariani, V., Brehm, W., Yefanov, O., Barty, A., Beyerlein, K. R., Chervinskii, F., Galli, L., Gati, C., Nakane, T., Tolstikova, A., Yamashita, K., Yoon, C. H., Diederichs, K. & Chapman, H. N. (2016). *J. Appl. Cryst.* **49**, 680–689.
Winn *et al.* (2011). *Acta Cryst.* D**67**, 235–242
Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K. & Evans, G. (2018). *Acta Cryst.* D**74**, 85–97.
Zeldin, O. B., Gerstel, M. & Garman, E. F. (2013). *J. Synchrotron Rad.* **20**, 49–57.

IUCrJ
ISSN 2052-2525
BIOLOGY | MEDICINE

High-throughput structures of protein–ligand complexes at room temperature using serial femtosecond crystallography

Tadeo Moreno-Chicano,[a]‡ Ali Ebrahim,[a,b] Danny Axford,[b] Martin V. Appleby,[b] John H. Beale,[b] Amanda K. Chaplin,[a]§ Helen M. E. Duyvesteyn,[b,c] Reza A. Ghiladi,[d] Shigeki Owada,[e,f] Darren A. Sherrell,[b] Richard W. Strange,[a] Hiroshi Sugimoto,[e] Kensuke Tono,[e,f] Jonathan A. R. Worrall,[a] Robin L. Owen[b]* and Michael A. Hough[a]*

Received 8 May 2019
Accepted 21 August 2019

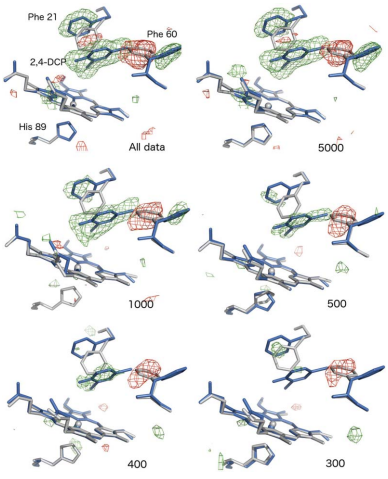
Edited by K. Moffat, University of Chicago, USA

‡ Current address: Institut de Biologie Structurale, 71 Avenue des Martyrs, 38000 Grenoble, France.
§ Current address: Department of Biochemistry, Sanger Building, 80 Tennis Court Road, Cambridge CB2 1GA, England.

Keywords: serial femtosecond crystallography; ligand binding; high throughput; X-ray crystallography; damage-free structures; X-ray free-electron lasers.

PDB references: dehaloperoxidase B, complex with 2,4-dichlorophenol, 6i7f; complex with 5-bromoindole, 6i6g; dye-type peroxidase Aa, complex with imidazole, 6i7c; copper nitrite reductase, complex with nitrite, 6qwg

Supporting information: this article has supporting information at www.iucrj.org



[a]School of Life Sciences, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, England, [b]Diamond Light Source, Harwell Science and Innovation Campus, Didcot OX11 0DE, England, [c]Division of Structural Biology (STRUBI), University of Oxford, The Henry Wellcome Building for Genomic Medicine, Roosevelt Drive, Oxford OX3 7BN, England, [d]Department of Chemistry, North Carolina State University, Raleigh, NC 27695-8204, USA, [e]RIKEN SPring-8 Center, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan, and [f]Japan Synchrotron Radiation Research Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5198, Japan. *Correspondence e-mail: robin.owen@diamond.ac.uk, mahough@essex.ac.uk

High-throughput X-ray crystal structures of protein–ligand complexes are critical to pharmaceutical drug development. However, cryocooling of crystals and X-ray radiation damage may distort the observed ligand binding. Serial femtosecond crystallography (SFX) using X-ray free-electron lasers (XFELs) can produce radiation-damage-free room-temperature structures. Ligand-binding studies using SFX have received only modest attention, partly owing to limited beamtime availability and the large quantity of sample that is required per structure determination. Here, a high-throughput approach to determine room-temperature damage-free structures with excellent sample and time efficiency is demonstrated, allowing complexes to be characterized rapidly and without prohibitive sample requirements. This yields high-quality difference density maps allowing unambiguous ligand placement. Crucially, it is demonstrated that ligands similar in size or smaller than those used in fragment-based drug design may be clearly identified in data sets obtained from <1000 diffraction images. This efficiency in both sample and XFEL beamtime opens the door to true high-throughput screening of protein–ligand complexes using SFX.

1. Introduction

The accurate determination of the structures of protein–ligand complexes is essential for drug discovery, enzymology and biotechnology. Developments in the automation of protein crystallization, ligand soaking, harvesting, structure determination, ligand modelling and structural refinement have allowed the high-throughput screening of soaked crystals at synchrotron X-ray beamlines (Collins *et al.*, 2018; Pearce, Krojer, Bradley *et al.*, 2017; Pearce, Krojer & von Delft, 2017). For important classes of proteins, the binding of ligands may be affected by X-ray-driven changes either in the oxidation state of redox centres within the protein or to amino-acid side chains involved in protein–ligand interactions. In these cases, there is a premium on structure determination using low-dose methods. Prime examples of this are heme enzymes, where the iron centre in resting iron(III) and high-valent iron(IV) states is exquisitely prone to reduction by solvated photoelectrons generated by the interaction of synchrotron X-rays with solvent in the crystal (see, for example, Beitlich *et al.*, 2007; Kekilli *et al.*, 2017). Heme enzymes, such as the cytochrome

P450s, are involved in the metabolism/breakdown of approximately 90% of small-molecule drugs, and are more generally themselves drug targets in yeast, fungi and tuberculosis infections (McLean & Munro, 2017; Guengerich *et al.*, 2016; Rendic & Guengerich, 2015). Importantly, the determination of protein–ligand complexes at room temperature is likely to better reflect *in vivo* conditions than crystals cryogenically cooled to 100 K (for an interesting example, see Fischer *et al.*, 2015). Indeed, significant differences in binding have been observed at room temperature (RT) compared with 100 K (Keedy *et al.*, 2018). Furthermore, ligand soaking into microcrystals (1–20 µm) has the theoretical potential to be more effective than soaking into larger crystals (>50 µm) (McPherson, 2019). The distance that a ligand needs to penetrate into the crystal to reach its centre is proportionately shorter for smaller crystals, potentially leading to higher occupancy rates.

These issues in combination place a high value on protein–ligand complexes determined from microcrystals at RT that are free of observable effects of radiation damage. The only current approach that can deliver this is serial femtosecond crystallography (SFX) at X-ray free-electron lasers (XFELs; Schlichting, 2015) using short (<20 fs) X-ray pulses (Inoue *et al.*, 2016; Lomb *et al.*, 2011; Nass *et al.*, 2015; Nass, 2019). Ligand-binding studies using SFX have received little attention, largely owing to the scarcity of beamtime and high sample requirements in typical sample-delivery systems such as the gas dynamic virtual nozzle (GDVN) injectors (Schlichting, 2015). The drive to obtain damage-free, RT structures is balanced against the strong practical driver to minimize sample consumption per obtained structure, and the desire to collect data from multiple candidate ligands in a short time period.

A limited number of studies have sought to address the challenge of obtaining damage-free, RT crystal structures of protein–ligand complexes in a manner that is efficient both in sample and in data-collection time. An early study examined ligand binding to a P-type ATPase membrane protein in microcrystals delivered to the beam using a liquid microjet injector (Bublitz *et al.*, 2015). This work demonstrated the applicability of SFX to ligand-binding studies, showing that ligands could be clearly resolved even if the high-resolution data collected are weak and statistically poor. A more recent study (Naitow *et al.*, 2017) explored the feasibility of SFX ligand-binding studies using microcrystals of the model system thermolysin delivered by a high-viscosity water- or oil-based injector. The small-molecule ligand was readily resolved in electron-density maps, with clear differences in binding modes observed between the room-temperature SFX and 100 K synchrotron-radiation (SR) structures.

Here, we describe the rapid determination of protein–ligand complexes at RT. Microcrystals were mounted in silicon fixed targets or 'chips' at the SPring-8 Ångstrom Free Electron Laser (SACLA), Hyogo, Japan (Ishikawa *et al.*, 2012). The fixed-target sample-delivery approach minimizes sample consumption, provides high hit rates and allows multiple high-quality data sets to be measured in a very short time, an important advantage given the limited availability of XFEL beamtime. The chip system also allows rapid switching experiments in which crystals of different targets are soaked with different ligands. Moreover, the short time between soaking, chip loading and the completion of data collection reduces the need for long-term protein–ligand crystal stability that is required for a typical injector experiment. This also ensures that crystals are exposed to the soaked ligand for a similar length of time.

We have applied this approach to crystals of two heme peroxidase enzymes: a multifunctional dehaloperoxidase from the marine annelid *Amphitrite ornata* (DHP-B; Barrios *et al.*, 2014; Franzen *et al.*, 2012; McCombs, Moreno-Chicano, *et al.*, 2017; McCombs, Smirnova *et al.*, 2017) and a dye-decolourizing peroxidase (Sugano, 2009) of industrial relevance (Colpa *et al.*, 2014) from *Streptomyces lividans* (DtpAa). We also examine the challenging case of detecting nitrite binding to copper nitrite reductase from *Achromobacter cycloclastes* (*Ac*NiR; Horrell *et al.*, 2017), where the ligand displaces a water molecule bound in the active site. The enzyme and crystal systems used are of cubic (high), orthorhombic (medium) and monoclinic (low) symmetry space groups, as well as exhibiting full to partial ligand occupancies within the same crystallographic asymmetric units. The complexes investigated include ligands directly binding to the heme, together with those occupying a binding pocket but not bound to the iron, with ligand sizes of 3–10 non-H atoms (Supplementary Fig. S1). We note that the typical molecular weight of the fragments used in fragment-based drug design is approximately 150–250 Da, with a typical size of 200 Da (Price *et al.*, 2017).

We explore the potential of this approach for rapid SFX screening of ligands/drug candidates, examining the minimum number of merged diffraction patterns required to reliably detect ligand binding and the future potential of this approach at current and planned XFEL beamlines. We assess several metrics for ligand fit to electron density with the data sets presented in the light of the recent debate around ligand validation (Smart *et al.*, 2018). Remarkably, data sets comprising of <1000 merged diffraction patterns allowed clear and unambiguous identification of ligand-binding modes, despite extremely poor merging and refinement statistics. The number of crystals required for complete data is lowered by the bandwidth of the XFEL beam. Our work thus demonstrates that high-throughput screening is eminently practicable using SFX, with modest requirements for sample quantity and experimental time.

2. Materials and methods

2.1. Protein production and crystallization

Dye-type peroxidase Aa (DtpAa) from *S. lividans* was expressed and purified as described previously (Ebrahim, Moreno-Chicano *et al.*, 2019). Crystals were grown in batch using a modification of the crystallization conditions used for growing large single crystals, consisting of 25%(*w*/*v*) PEG

1500 and 100 m*M* MIB buffer (Hampton Research, comprising MES, boric acid and imidazole pH 8). The final protein concentration in the batches ranged from 6.5 to 2.1 mg ml^{-1}. Crystals grew in 1–2 days to approximate dimensions of 20–30 µm and were transported to SACLA at ambient temperature (the crystals were transported in hand luggage without cooling). Dehaloperoxidase B (DHP) from *A. ornata* was expressed and purified as described previously (McCombs, Moreno-Chicano *et al.*, 2017). Batch microcrystallization was used, mixing 30 mg ml^{-1} DHP in 20 m*M* MES pH 6.0 with 40%(*w*/*v*) PEG 4000, 200 m*M* ammonium sulfate in a 1:4 ratio in a total volume of 250–500 µl. DHP microcrystals grew in 3–5 days at 4°C to typical dimensions of 20–30 µm and were transported to SACLA at 4°C. 5-Bromoindole (5BR) and 2,4-dichlorophenol (DCP) (Sigma) were each dissolved in 100% DMSO and 20 µl of the resulting solution was added to a 200 µl crystal suspension to yield final ligand concentrations of 5 m*M* DCP and 50 m*M* 5BR. Microcrystals were soaked in batches for 3–5 min immediately prior to loading onto the silicon chip. *Ac*NiR microcrystals were grown as described previously (Ebrahim, Appleby *et al.*, 2019) and were soaked in 100 m*M* potassium nitrite for approximately 20 min prior to loading onto the chip.

2.2. Data collection and processing

Microcrystals were loaded into fixed-target chips as described previously (Ebrahim, Appleby *et al.*, 2019; Oghbaey *et al.*, 2016). The chips were fabricated commercially (Southampton Nanofabrication Centre; https://www.southampton-nanofab.com) using a method based on that described previously (Oghbaey *et al.*, 2016). Typically, 100–200 µl of microcrystal suspension was loaded onto a chip containing 25 600 apertures and excess liquid was removed using a weak vacuum applied to the underside of the chip surface. For DHP microcrystals around 1.5 mg of protein was loaded in each chip, requiring around 4.5 mg for a complete data set (three or four chips), while *Ac*NiR microcrystals were loaded in quantities of around 2 mg for a complete data set (two chips at 1 mg per chip). In the case of DtpAa even less protein was needed: only 0.45–6.0 mg per chip and around 1.80 mg for a complete data set. SFX data were measured on SACLA (Ishikawa *et al.*, 2012) beamline BL2 EH3 with a photon energy of 10.0 keV, a repeat rate of 30 Hz and a pulse length of 10 fs. The beam, with a 1.25 × 1.34 µm spot size (FWHM) and a pulse energy of 289 µJ per pulse (pre-attenuation), was attenuated to 13% of full flux to minimize detector overloads. The SACLA beam was in SASE mode, with FWHM bandwidth ∼70 eV. The fixed-target chip was translated between X-ray pulses such that each crystal position was exposed only once, and the measurement of all 25 600 positions on a chip took 14 min. The hit rate during data collection was monitored using *Cheetah* (Barty *et al.*, 2014), while peak finding, indexing and merging of data were performed using *CrystFEL* v.0.6.4 (White *et al.*, 2016). Structures were refined using starting models of ligand-free structures from which water and other solvent molecules had been removed. Refinement was initially carried out in *REFMAC*5 (Murshudov *et al.*, 2011) within the *CCP*4 suite and completed in *PHENIX* (Adams *et al.*, 2010). All structures were validated using *MolProbity* (Williams *et al.*, 2018), the *JCSG Quality Control Check* server and tools within *PHENIX* (Adams *et al.*, 2010) and *Coot* (Emsley *et al.*, 2010).

To explore the limits of ligand identification in SFX data sets, randomly selected images from the indexed data (`*.stream` files from *CrystFEL*) formed data subsets with defined, variable numbers of images. These were scaled and merged in the same manner as the data sets containing all images and were used in refinement versus the model for the appropriate complex determined using all data, from which the ligand had been removed. OMIT maps were generated using torsion-based simulated-annealing refinement in *phenix.refine* (Adams *et al.*, 2010) in order to minimize model bias. As an additional validation step, selected subsets were refined against the structure of the native enzymes (where the ligands were not present) using the procedure described above.

2.3. Ligand modelling

Ligands were initially modelled into the all-image data sets based on the $mF_o - DF_c$ difference electron-density maps. In all cases, ligand density was unambiguous and ligands were modelled with near-full occupancy in one of the two subunits of the homodimeric enzymes (for DHP and DtpAa) or in the single subunit of *Ac*NiR in the crystallographic asymmetric unit. The ligands were straightforwardly located in an automated manner using the 'Find Ligands' feature of *Coot*. The second monomer in the DHP asymmetric unit contained a lower occupancy ligand (5BR) or very weak ligand density (DCP), while in DtpAa the second monomer did not show a bound exogenous ligand in the active site. Restraints for nonstandard ligands were produced using *ACEDRG* (Long *et al.*, 2017). For the data subsets, the data were refined by two parallel approaches to avoid model bias. Firstly, the data were refined against the 'all-images' structures, from which the ligands had been removed, using simulated annealing in *phenix.refine* to remove bias. As an additional test that bias was not present, selected structures were refined against the native, ligand-free structures of the enzymes and simulated-annealing (SA) OMIT difference maps were generated. The known position of the ligand from the 'all-images' models was then compared with the difference density map generated from that subset. The quality of the fit of modelled ligands to the electron-density maps was determined using *EDIAscorer* (Meyder *et al.*, 2017). The 'Find Ligands' feature of *Coot* was also used for each subset, in this case searching the $mF_o - DF_c$ SA OMIT map for suitable hits.

$F_o - F_o$ isomorphous difference maps between the DHP–5BR and DHP–DCP data sets were generated in *PHENIX* with the native ligand-free DHP structure (see below) used to phase the data sets (although near-identical results were generated if either of the above ligand-bound structures were used for phasing).

Table 1
Data-collection, processing and refinement statistics for full SFX data sets for enzyme–ligand complexes.

Values in parentheses are for the outermost resolution shell.

Structure	DHP–DCP	DHP–5BR	DtpAa–imidazole	*Ac*NiR–nitrite
Data collection and processing				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_1$	$P2_13$
Unit-cell parameters (Å, °)	a = 60.9, b = 67.2, c = 68.7, $\alpha = \beta = \gamma$ = 90	a = 61.0, b = 67.3, c = 68.8, $\alpha = \beta = \gamma$ = 90	a = 72.5, b = 68.0, c = 73.5, $\alpha = \gamma$ = 90, β = 105.6	a = 97.6, b = 97.6, c = 97.6, $\alpha = \beta = \gamma$ = 90
Chips used	3	4	4	2
Images collected	76800	102800	102800	51200
Indexed images merged	32618	53662	20316	16586
Unique reflections	24749	24840	56220	24729
Resolution (Å)	37.7–1.85 (1.90–1.85)	45.6–1.85 (1.90–1.85)	70.8–1.88 (1.93–1.88)	43.7–1.90 (1.93–1.90)
Completeness (%)	100 (100)	100 (100)	100 (100)	100 (100)
Multiplicity	579 (340)	907.7 (524.0)	101.6 (64.2)	3281.4 (2299.1)
$CC_{1/2}$	0.99 (0.66)	1.00 (0.65)	0.96 (0.60)	0.99 (0.63)
R_{split} (%)	6.6 (61.9)	5.5 (66.6)	15.8 (63.9)	9.73 (58.61)
Refinement				
Resolution range (Å)	34.4–1.85	45.6–1.85	35.3–1.88	43.7–1.90
R_{work} (%)	16.8	16.7	13.9	13.7
R_{free} (%)	19.9	18.9	17.6	17.2
R.m.s.d., bond lengths (Å)	0.010	0.005	0.010	0.006
R.m.s.d., bond angles (°)	1.23	0.96	0.87	0.90
Ramachandran most favoured (%)	98.2	98.9	98.5	98.8
PDB code	6i7f	6i6g	6i7c	6qwg

3. Results

3.1. Determination of protein–ligand complex structures by SFX in a time- and sample-efficient manner

SFX structures for each enzyme–ligand complex were determined from data measured from either two (*Ac*NiR), three (DHP–DCP) or four (DHP–5BR and DtpAa–imidazole) chips. This took approximately 14 min of data collection and ∼16 min of beamtime per chip (sample-change, hutch-search and alignment time are included). In each case, structure solution was by molecular replacement and the resolution and data quality were sufficient to clearly define essentially all main-chain and most side-chain atoms together with well defined networks of water molecules. The quality of the data sets and structures is given in Table 1. For each structure, clear positive difference density was evident for the ligands, which were unambiguously located. The chemical structures of the ligands used in this study are shown in Supplementary Fig. S1.

3.2. SFX structures of ligand-bound complexes

In each DHP structure, clear $F_o - F_c$ electron density was apparent in the heme pocket consistent with a high-occupancy bound ligand in one monomer of the dimer and a second lower occupancy binding site in the other. This difference in occupancy is consistent with previous single-crystal structures of DHP complexes with a range of different ligands in this space group (see, for example, McCombs, Moreno-Chicano *et al.*, 2017). DCP exhibited a binding site that was virtually identical to those previously observed for the guaiacol substrates 4-bromoguaiacol (PDB entry 6cke), 4-nitroguaiacol (PDB entry 6ch5) and 4-methoxyguaiacol (PDB entry 6ch6) (McGuire *et al.*, 2018), while the 5BR complex was consistent with a computationally hypothesized binding site (Barrios *et al.*, 2014), with both results together demonstrating that SFX provides accurate substrate-binding orientations.

The details of the binding modes themselves are beyond the scope of this manuscript and will be described in detail in a separate publication. Strong electron-density peaks were present for the two Cl atoms of DCP and the single Br atom of 5BR, allowing the ligand orientation to be easily confirmed, although it is important to note that the electron density was well defined for all atoms of the ligands. For both DHP–ligand structures one monomer had near-full occupancy, but significantly lower occupancy (as refined in *PHENIX*; Adams *et al.*, 2010) was observed in the second monomer of the homodimer [Figs. 1(*a*) and 1(*b*)]. This feature allowed us to examine the effect of ligand occupancy on ligand detectability in maps derived from SFX data (see below).

The SFX structure of DtpAa was determined in space group $P2_1$ to 1.88 Å resolution (Table 1). The overall structure of the enzyme homodimer was highly similar to that of ferric DtpAa crystallized in a condition that did not contain imidazole (Ebrahim, Moreno-Chicano *et al.*, 2019). The examination of $F_o - F_c$ difference maps indicated that an imidazole ligand was coordinated *via* an N atom to the distal position of the heme iron in one monomer of the DtpAa dimer with full occupancy. The Fe—N (imidazole) bond was 2.2 Å, while imidazole also formed two hydrogen bonds (2.7 and 2.9 Å) to Asp239 [Fig. 1(*c*)], and the heme pocket also contains several well ordered water molecules. Comparison with the ligand-free ferric DtpAa structure also obtained by SFX (Ebrahim, Moreno-Chicano *et al.*, 2019) revealed that the imidazole displaces the distal water molecule from the heme and induces a number of modest structural rearrangements in the heme pocket (Supplementary Fig. S2). A second imidazole molecule is bound to the protein away from the heme pocket, forming a 2.7 Å bond to Thr351 and interacting via a bridging water with Glu283. In contrast, for monomer *A* no imidazole ligand was observed in the distal heme pocket and instead a water molecule is bound at a distance of 2.4 Å in a similar manner to

that in the ferric DtpAa structure (Ebrahim, Moreno-Chicano *et al.*, 2019).

The structure of *Ac*NiR in complex with nitrite was determined to 1.90 Å resolution (Table 1). The type 2 Cu site, which is the site of ligand binding, displayed clear electron density for a bound nitrite molecule with a bidentate O-binding geometry as previously described in multiple 100 K and room-temperature structures obtained from single crystals (Meyder *et al.*, 2017; Horrell *et al.*, 2016) [Fig. 1(*d*)]. The positions of the ligand-binding sites within the protein fold for each complex are shown in Supplementary Fig. S3.

3.3. Ligand density features as a function of the number of diffraction patterns in a data set

As described above, all three ligands were unambiguously identified in maps derived from the full data sets, demonstrating that ample diffraction patterns had been included in the merged data sets, which had good data-quality metrics (Table 1). To test the lower limits of the number of diffraction patterns that would allow us to identify bound ligands in high-throughput SFX experiments, the data were partitioned into subsets of decreasing size to produce independent merged data sets containing progressively fewer diffraction patterns, (Supplementary Tables S1–S4). OMIT difference maps were generated by simulated-annealing refinement in *PHENIX* (Adams *et al.*, 2010) using the all-data structure with ligand atoms omitted in the initial model. As expected, the merging and refinement statistics, and consequently the resolution cut-off, progressively deteriorated as the number of merged patterns was reduced (Fig. 2, Supplementary Tables S1–S4, Supplementary Fig. S4). The OMIT map quality was, as might be reasonably expected, proportional to the number of images in the data set. As a simple practical test to emulate a typical crystallographic workflow, the 'Find Ligands' feature of *Coot* (Debreczeni & Emsley, 2012; Emsley, 2017) was used to test whether each ligand could be correctly fitted into the simulated-annealing $F_o - F_c$ map without bias from the experimentalist's prior knowledge of the correct pose.

3.4. 2,4-Dichlorophenol–DHP complex

We first examined the effect of the number of crystals included in a data set on the resulting electron-density maps for the complex between DHP and DCP. Very clear $F_o - F_c$ simulated-annealing OMIT map features for the ligand were evident in subsets considerably smaller than the 'full' data sets. For example, a subset of 5000 crystals showed merging statistics that would still be considered acceptable by standard assessments [R_{split} = 0.17 (0.73), $CC_{1/2}$ = 0.95 (0.56) to 1.95 Å

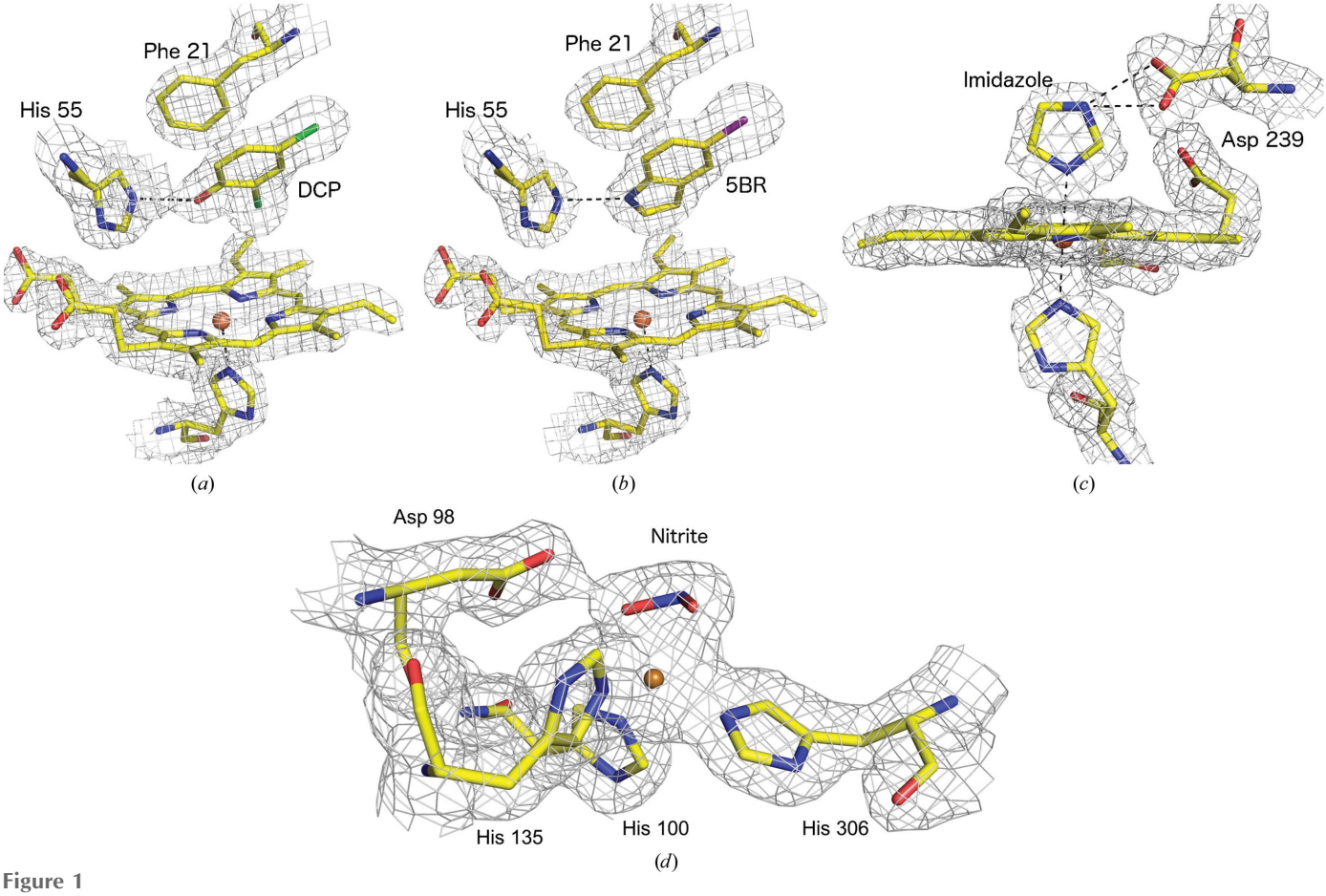


Figure 1
$2F_o - F_c$ electron-density maps, contoured at 1σ, showing the complexes of DHP with (*a*) DCP with the Cl atoms shown in green and (*b*) 5BR with the Br atom shown in purple, (*c*) the complex of DtpAa with imidazole and (*d*) the complex of *Ac*NiR with nitrite. In each case, the active site of the monomer with the highest ligand occupancy is shown. The maps in (*a*)–(*d*) were generated using the all-image data sets.

resolution] and unsurprisingly ligand finding was straightforward. When the data set was reduced to containing only 1000 crystals the merging statistics were poor [R_{split} = 0.39 (0.65), $CC_{1/2}$ = 0.72 (0.57) to 2.2 Å resolution], and with 500 images these metrics indicated very poor or even meaningless data quality [R_{split} = 0.56 (0.92), $CC_{1/2}$ = 0.57 (0.42) to 2.2 Å resolution]. The refinement statistics also deteriorated with decreasing data-set size (Supplementary Table S1).

Remarkably, data sets comprising far fewer than 1000 indexed patterns displayed very clear features in simulated-annealing OMIT maps of the distal pocket, covering all atoms of the best-ordered DCP ligand (in monomer *B*). Examples are shown in Supplementary Fig. S5, where the $F_o - F_c$ OMIT map allowed all atoms of the ligand to be unambiguously modelled, even when the merging statistics were very poor and refinement *R* factors were high (Supplementary Table S1). Because of the poor merging statistics with <1000 images, it was not possible to use these metrics to assess the resolution limit in merging for these data; however, refinement using the same resolution limit as the 1000-image set still allowed straightforward ligand placement.

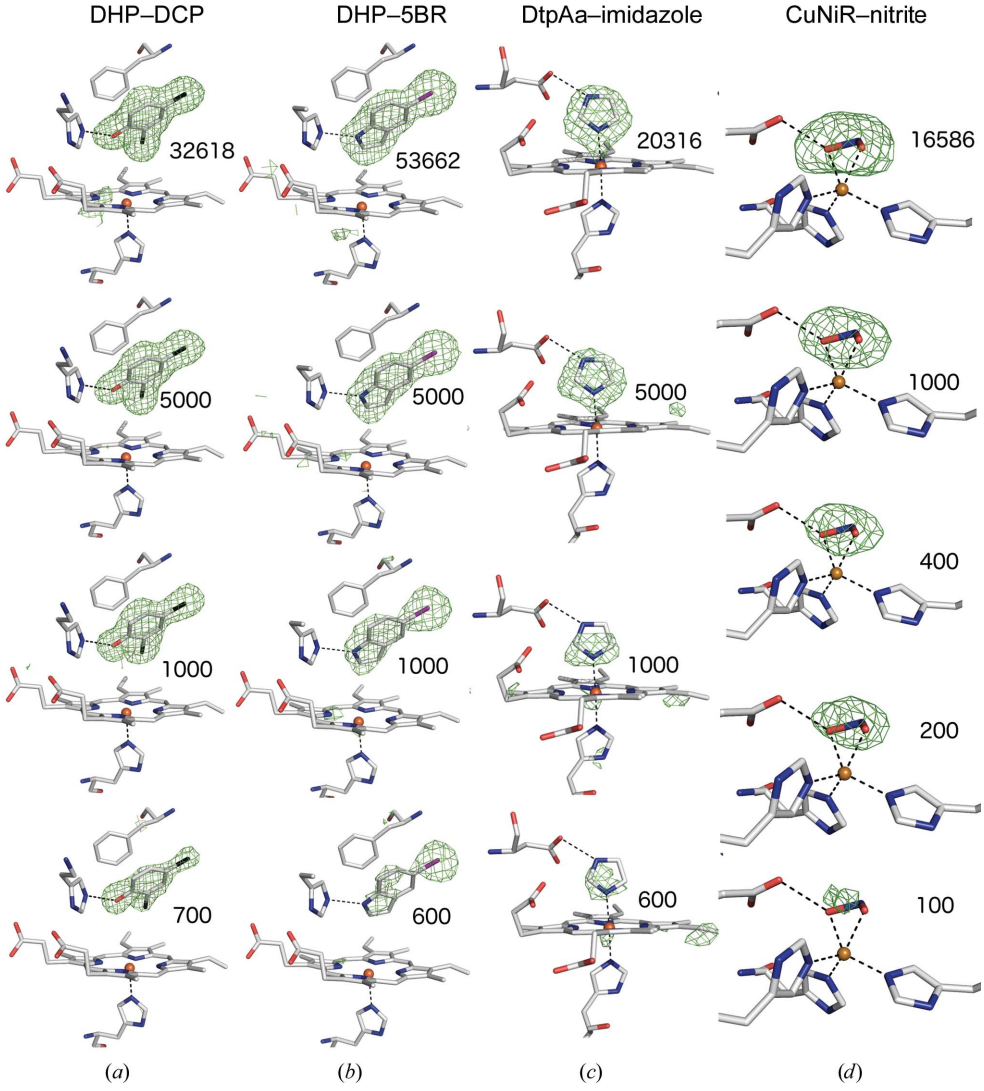


Figure 2
$F_o - F_c$ simulated-annealing OMIT maps contoured at 3σ for the heme region from selected data subsets for (*a*) DHP–DCP, (*b*) DHP–5BR, (*c*) DtpAa–imidazole and (*d*) *Ac*NiR–nitrite, each superposed on the refined structure from all data. For (*a*) and (*b*) the highest occupancy ligand monomer of the homodimer is shown. Additional subsets are shown in Supplementary Figs. S5, S6 and S8–S11.

For data sets produced from <400 crystals, difference map quality rapidly deteriorated (Supplementary Fig. S5). This deterioration of the maps appears to approximately coincide with a loss of data completeness and redundancy in these data sets. The lower occupancy ligand present in the second monomer failed to be located in data-subset OMIT maps of decreasing size more rapidly than was the case for the fully occupied ligand (Supplementary Fig. S6). *EDIAscorer* (Meyder *et al.*, 2017) electron-density analysis is shown in Fig. 3 and Supplementary Fig. S7, showing the excellent quality of the difference map for all ligand atoms down to very low image numbers.

3.5. 5-Bromoindole–DHP complex

Data and map quality followed a similar pattern with reducing crystal numbers to that described above for DCP (Table 1, Fig. 2, Supplementary Figs. S8 and S9). In this case, the lower occupancy of the two 5BR ligands was automatically found in *Coot* with a data set from 2000 images, but this step failed with 1000 images. For the higher occupancy 5BR ligand, the correct pose was found down to a data set of 800 images, while in data sets comprising 500, 600 or 700 images an incorrect pose was found by *Coot*, although manual re-orientation was straightforward based on the $F_o - F_c$ map. A remarkable observation is that even in a data set comprising only 200 images (with 75.3% data completeness) the heavier Br atom of the ligand was clearly identified, with OMIT map peaks of 6.8σ (1.83 $e^- Å^{-3}$) in monomer *B* and 4.3σ (1.16 $e^- Å^{-3}$) in monomer *A* at its position (Supplementary Figs. S8 and S9).

3.6. Imidazole complex of DtpAa

The imidazole ligand provided a more challenging case owing to its smaller size in comparison to DCP and 5BR and

because of the lower symmetry space group ($P2_1$) of the DtpAa crystals. For the latter reason, the merging statistics deteriorated more rapidly than for DHP (Supplementary Table S3). In particular, data completeness began to deteriorate, with the 2000-image data set being essentially complete, while this was not the case for the 1000-crystal data set. With subsets of 5000 images or larger, *Coot* was able to successfully locate both the heme-coordinated imidazole and the second imidazole ligand located in the inter-monomer cleft [Supplementary Fig. S3(*c*)]. With smaller subsets, the latter ligand was not found, although the heme-bound imidazole was located in data sets of as few as 800 images. In these very small data sets the imidazole ring as positioned by *Coot* was sometimes rotated around its normal axis while still fitting the symmetrical electron-density feature well, but this was readily corrected by applying simple chemical knowledge, *i.e.* that N atoms rather than C atoms should be forming the coordination bond to the Fe atom and be oriented towards the Asp residue in the heme pocket. Simulated-annealing OMIT maps for the DtpAa–imidazole complex are shown in Fig. 2 and Supplementary Fig. S10, with electron-density statistics in Fig. 3 and Supplementary Fig. S7. We note that for all three ligands, even when automated ligand finding failed, significant ligand density was present that could allow manual identification in cases where the binding pocket was known in advance.

3.7. *Ac*NiR complex with nitrite

Although nitrite is the smallest ligand of interest used in this study, *Ac*NiR has the inherent characteristic of crystallizing in a high-symmetry space group ($P2_13$), resulting in fewer data being required for a complete data set owing to the high redundancy of the data collected (Table 1). Again, very clear $F_o - F_c$ simulated-annealing OMIT map features for the ligand were evident in subsets of small numbers of diffraction patterns, despite exhibiting merging statistics that would typically be considered rather poor (Figs. 2 and 3, Supplementary Table S4 and Supplementary Fig. S11). *Coot* successfully located nitrite binding at the type 2 Cu active site in subsets of very few crystals, with 200 being the lowest number of crystals that were needed to successfully auto-find the nitrite ligand. Although *Coot* was unsuccessful at determining the ligand in the lowest crystal subset of only 100 crystals, positive electron density is still identifiable at the site where ligand binding is expected, although this did not allow for reasonable modelling of a ligand.

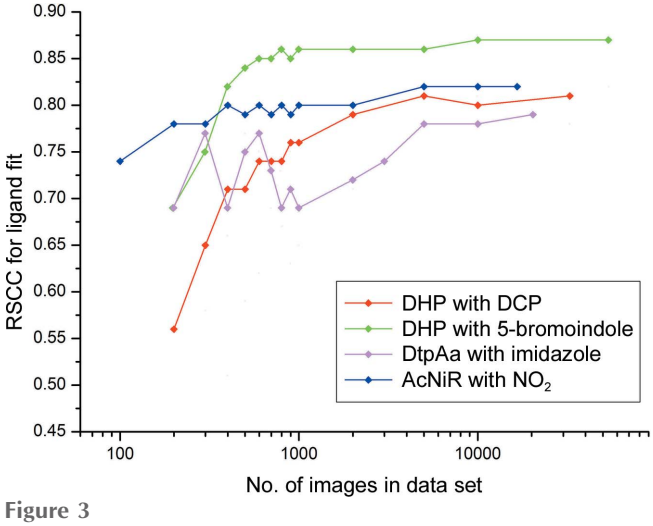


Figure 3
Real-space correlation coefficient (RSCC) values from *EDIAscorer* (Meyder *et al.*, 2017) as a function of the number of images per subset. Data are shown for the highest occupancy binding site for each complex. A plot including values for additional binding sites is shown in Supplementary Fig. S7.

3.8. OMIT maps from simulated-annealing refinement against ligand-free structures of the native enzymes

Although simulated-annealing refinement as described above would reasonably be expected to remove all model bias, as an additional validation step selected subsets were refined against the corresponding native structures obtained by SFX (Ebrahim, Moreno-Chicano *et al.*, 2019; Moreno-Chicano *et al.*, manuscript in preparation), where the ligands were not present. OMIT map generation followed an identical procedure to that described above, with the exception of the input coordinate file used. The resulting OMIT maps are shown in Fig. 4 and Supplementary Figs. S12, S13 and S14) for data subsets of differing sizes. The results of this process corresponded well with the previously described OMIT maps, suggesting that model bias was not significant in the previous procedure for any of the complexes. Notably, for the two DHP ligand structures, in addition to very clear OMIT map density for the ligands themselves the map features clearly define the movements of heme-pocket residues that are necessary to accommodate the ligand (Fig. 4 and Supplementary Fig. S12). This provides further evidence of the information content within these data sets, despite the low numbers of diffraction patterns and extremely poor statistics. Importantly, we used *Ac*NiR–nitrite as a very challenging case to test the limitations of our approach as the nitrite ligand contains only three atoms and also displaces a water molecule upon binding (Antonyuk *et al.*, 2005). In addition, the water density in the native structure is disordered, with the presence of a second water molecule a possibility. Notably, refinement of *Ac*NiR data and subsets versus the native *Ac*NiR SFX structure produced clear positive difference map features for the nitrite atoms that are separated from the water molecule present in the native structure (Supplementary Fig. S14). For comparison, SA OMIT maps produced from refinement of the same subsets against the native *Ac*NiR SFX model with the copper-coordinated water molecule deleted are shown in Supplementary Fig. S15.

3.9. Detection of differences between ligands from $F_o - F_o$ isomorphous difference maps

For the DHP case, in which two different ligands bind in a similar binding pose to the same enzyme pocket, we tested the ability to distinguish between these ligands using $F_o - F_o$ isomorphous difference maps. For the full data sets, an F_o(DHP–5BR) $- F_o$(DHP–DCP) map is shown in Fig. 5. Strong positive density (a 32σ peak) is present where the Br

atom of 5BR occupies a similar position to a Cl atom of DCP, consistent with the larger number of electrons on the Br atom. A negative feature is present over the second Cl atom of DCP, consistent with a C atom occupying a similar position in 5BR. Finally, a positive peak is present for the C5 atom of 5BR where no equivalent atom is present in DCP. As the number of crystals in a data subset decreases, the map features become less prominent, with the C5 feature disappearing in subsets of 1000 crystals or smaller. However, the features corresponding to the Br and Cl atoms are remarkably still evident, albeit much weaker, in subsets comprised of as few as 200 crystals (Fig. 5).

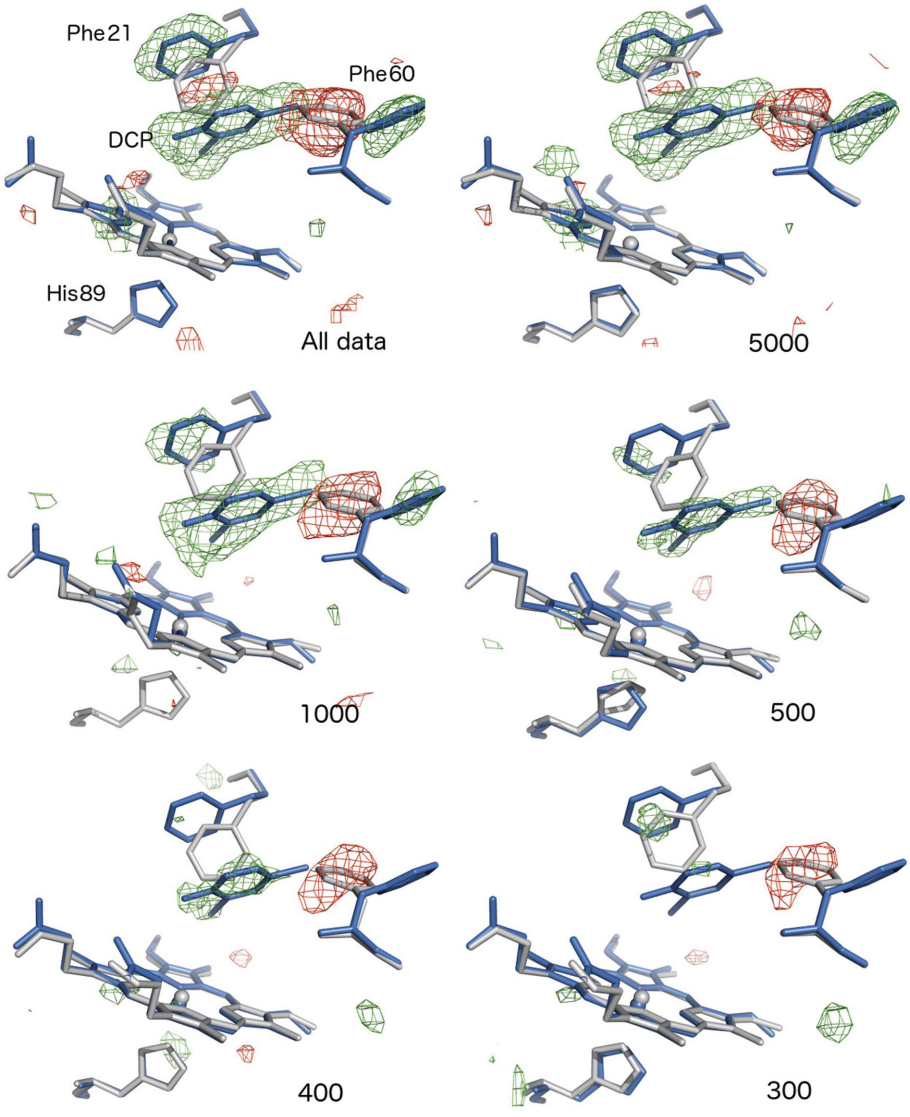


Figure 4
Difference map features produced by simulated-annealing refinement against ligand-free native structures clearly reveal ligand binding and active-site rearrangements in the absence of the risk of model bias. $F_o - F_c$ OMIT maps, contoured at 3σ, are shown for DCP data subsets refined versus the native DHP structure. In each case, the native DHP structure from OMIT refinement versus a particular subset is shown in grey, while the superimposed structure of the ligand complex is shown in blue. Positive difference map features are shown in green, with negative features in red. Note that the flips of Phe21 and Phe60 to accommodate ligand binding, together with the ligand density itself, are very clearly defined in the data set obtained from all data and this is maintained in the 5000-image subset. Clear OMIT map features are apparent for Phe60 and DCP in data sets with as few as 400 images, while this was no longer the case in the 300-image subset.

4. Discussion

4.1. High-throughput determination of ligand-bound SFX structures using fixed targets

Our results indicate that high-quality SFX crystal structures that allow unambiguous ligand identification may be achieved using our fixed-target approach. This can be achieved using a small quantity of enzyme sample with high throughput and rapid switchover between different proteins and ligands. Typical data-collection times for complexes were 30–60 min using all data measured, and using these data sets ligand modelling was clear and unambiguous. In comparison to previously published data for ligand-binding experiments (Naitow *et al.*, 2017; Bublitz *et al.*, 2015), the fixed-target approach allows the high-throughput production of multiple intact enzyme–ligand complex structures. In addition, the soaking and data-collection strategy can be easily adapted and optimized for a synchrotron beamline using the same sample-loading and mounting system.

Our results cover several different ligand-binding scenarios, such as coordinate-bond formation to a heme iron (imidazole) or copper (nitrite) and noncoordinate ligand binding in a pocket (DHP ligands), with the latter being highly relevant to the binding of ligands to pharmacologically important proteins such as cytochromes P450. In each structure, binding sites are present with different occupancies, allowing a further test of the ability of the method to locate high- or low-occupancy ligands.

The limits of our ability to identify ligand binding were tested using the small ligands nitrite (46 Da) and imidazole (68 Da). Both of these are much smaller than the fragments used in fragment-based drug design (FBDD), where 200 Da is a typical molecular weight (Price *et al.*, 2017). In the case of *Ac*NiR, a particular challenge was that nitrite displaces a water molecule on binding. In *Ac*NiR structures determined from single crystals, distinguishing between the electron-density features of active-site waters and nitrite is challenging and requires high-

resolution data (Antonyuk *et al.*, 2005; Horrell *et al.*, 2018). Nonetheless, our method allowed the identification of these ligands in subsets comprising a small fraction of the full data sets. Our results are therefore strongly indicative that the ligands used in FBDD will be readily detected using our approach.

4.2. What is the minimal quantity of data required to identify ligand-binding modes?

Analysis of simulated-annealing OMIT maps generated from data subsets containing only a subset of merged diffraction patterns clearly demonstrates that only a small fraction of the total data-collection time that we used is in fact necessary to locate ligands in the correct binding pose. For example, for the 5BR complex of DHP a subset of just 800 indexed images (∼1.5% of the total number of images in the full data set) was sufficient to correctly model the ligand using a careful strategy to preclude the possibility of model bias. A conservative approach of measuring several times this minimal number would still require only a small proportion of the 25 600 crystal positions on each chip. Our data also show that useful information is contained in data sets obtained from extremely small numbers of microcrystals; for example, the Br atom of 5BR was identified in a data set of only 200 crystals (<0.4% of the total data set).

In a lower symmetry space group (DtpAa; $P2_1$), the ability to detect ligand binding in data sets of <2000 images was compromised by a lack of data completeness at higher resolution, although ligand finding was still achieved with 800 images. Notably, for the DHP structures in space group $P2_12_12_1$ data completeness remained good in very small data sets; for example, for the DCP complex the 400-image data set retained >90% completeness in the highest resolution shell.

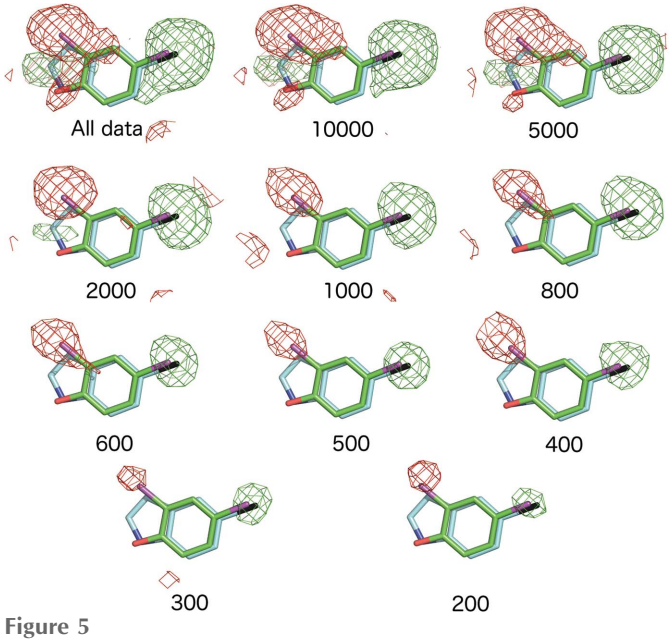


Figure 5
$F_o - F_o$ isomorphous difference maps comparing the 5BR and DCP ligand complexes of DHP. Maps are F_o(5BR) − F_o(DCP) contoured at 3σ. With all data included, the map shows a clear positive peak near to the position of the Br atom of 5BR (black) and one of the Cl atoms of DCP (magenta), consistent with the greater number of electrons in bromine. A negative peak is present at the position of the second Cl atom of DCP, where the closest atom of 5BR is a carbon. An additional but weaker positive peak is present close to the C5 atoms of 5BR where no atom is present in DCP.

The high completeness of data sets formed from (relatively) small numbers of crystals parallels the success in forming complete data sets from multiple thin wedges in virus crystallography (Fry *et al.*, 1999). The completeness of the final data set is a function of the number of wedges collected and the point group of the crystals used, with the prerequisite for each approach being that the crystals must be randomly orientated. The completeness of the data obtained from small numbers of crystals here illustrates that this is the case for DtpAa, DHP and *Ac*NiR crystals on silicon chips. The bandwidth of the XFEL beam allows complete data to be obtained from fewer crystals than would be the case with a more monochromatic beam, yet still requires many more crystals than might be required in a wide-bandpass Laue experiment (Meents *et al.*, 2017). Our data strongly suggest that data completeness is the key metric for assessing the suitability of data sets for ligand-binding studies and that very poor values of other typically used metrics of data quality (for example $CC_{1/2}$ and R_{split}) still allow successful ligand characterization provided that the data are complete. For *Ac*NiR, with cubic symmetry, the data remained essentially complete in all of the subset sizes analysed, with density for the nitrite ligand remaining apparent down to <200 indexed images. We note that substantially more diffraction patterns would be required to obtain complete data on a monochromatic beamline.

More broadly, our data clearly show that substantial information content is present in noisy and apparently low-quality data sets derived from small numbers of merged diffraction patterns with very poor merging and refinement statistics. For example, a data set formed of 200 patterns revealed a very clear peak for the Br atom of the 5BR ligand (outer shell completeness 70.9% in DCP). Importantly, refinement of data subsets against native structures unambiguously showed not only clear density for ligands, but also any movements of the active-site residues needed to accommodate ligand binding (Fig. 4 and Supplementary Fig. S12). This provides conclusive evidence that the ligand density that we describe is not owing to model bias from prior knowledge of the binding mode.

4.3. Future potential of the 'chip-soak' approach for high-throughput structure determination of protein–ligand complexes

In this work, SFX structures were recorded from two (*Ac*NiR), three (DHP–DCP) or four (DHP–5BR and DtpAa–imidazole) chips, aiming for 1–2 structures per hour. The number of chips used for a single structure was subsequently seen to err significantly on the side of caution, as in all cases sufficient data for unambiguous ligand identification were available from significantly less than half a chip. Crucially, careful data analysis demonstrated that data sets comprising of no more than a few hundred to a few thousand indexed

images are sufficient to correctly model ligands into clear difference density features. Thus, without modification of the approach or changes to the experimental conditions, an approximately 4–5-fold increased throughput of multiple protein–ligand structures per hour could easily be realized. Rapid on-site data analysis should allow on-the-fly decision making as to whether sufficient data have been collected for a particular soak and if a ligand is indeed bound. A key advantage of the fixed-target sample-delivery method is that switching between samples of different protein–ligand soaks is no more time-consuming than continuing with a chip of the same sample. With typical loading rates of approximately 30%, multiple ligand soaks could be carried out on a single redesigned chip, again drastically increasing throughput. As a further example, for systems where approximately 1000 hits would be sufficient, at the latter hit rate some eight ligand complexes could be characterized on a single chip.

The sample quantity required for our approach (in the range of 1.35–6.0 mg protein per data set) is less than required in liquid-jet approaches, although higher than has been reported for high-viscosity (LCP) injection systems at XFEL (Weierstall *et al.*, 2014) and synchrotron (Weinert *et al.*, 2017) beamlines. An additional factor is ligand consumption. In our case, without optimization to minimize sample consumption, the typical ligand quantities used were in the range 4–40 µmol.

Our system of work is applicable at other current and future XFEL sources, such as PAL (60 Hz repetition rate), SwissFEL (100 Hz) and LCLS (120 Hz), as well as SACLA (30 Hz). However, XFEL sources with very high repetition rates or complex pulse patterns (for example EuXFEL and LCLS-II) may require a modified or different approach. We have demonstrated that at a source with a modest repetition rate sufficient data for multiple, unrelated, protein–ligand structures may be obtained within a couple of hours. Increasing this level of throughput to ∼5–20 structures per hour at higher repetition-rate sources, or collecting fewer images per complex (see above) as is practical, would allow, for example, >200 structures to be determined in a single 12 h shift, similar to dedicated synchrotron beamlines. Fixed targets are also well suited to time-resolved crystallography of, for example, protein–ligand complexes using laser pump–probe methods (Schulz *et al.*, 2018) and it is important to note that in time-resolved experiments significantly more data may be required as crystals may contain a mix of states.

Another key advantage is that the chip approach allows us to test soaking protocols at synchrotron beamlines under identical conditions to those used at the XFEL in order to ensure that soaking does not damage crystals and also that ligands are bound, albeit in a radiation-damaged structure. At such high rates of sample delivery, automation of chip loading and robotic sample exchange will of course become increasingly important. Our work demonstrates the feasibility of high-throughput room-temperature ligand screening by SFX using microcrystals and is highly applicable to drug-discovery efforts, including in fragment-based drug design. Our approach would be of particular importance in cases where only small weakly diffracting crystals are obtained or when the enzyme–ligand complexes are radiation-sensitive. We have demonstrated the ability to identify ligand binding by our high-throughput approach using 'conventional' approaches to both refinement and ligand finding. Further data-analysis improvements to the ability to identify in particular low-occupancy ligands in FBDD could be achieved using a multi-data-set approach, for example in *PanDDa*, with subtraction of the ligand-free ground state (Pearce, Krojer, Bradley *et al.*, 2017) and with refinement against a composite of the ligand-free and ligand-bound structures (Pearce, Krojer & von Delft, 2017).

In conclusion, we demonstrate (i) a method to rapidly measure SFX data sets from protein–ligand complexes and to rapidly switch between ligands during beamtime, (ii) that data sets comprised of hundreds to a few thousands of diffraction patterns can be sufficient for unambiguous ligand identification and (iii) that even ligands smaller than those used in fragment-based drug design may be located using our approach. These data demonstrate the feasibility of high-throughput structure determination of protein–ligand complexes at XFEL sources.

Acknowledgements

XFEL experiments were performed on BL2 EH3 at SACLA with the approval of the Japan Synchrotron Radiation Research Institute (JASRI; Proposal No. 2017B8014). We acknowledge the contributions of Drs Minoru Kubo and Takashi Nomura (University of Hyogo) and Dr Takehiko Tosha (RIKEN). Support for travel from the UK XFEL Hub at Diamond Light Source is gratefully acknowledged. We are grateful to an anonymous reviewer of the manuscript for the suggestion of generating isomorphous difference maps for the two DHP ligand-bound structures. Author contributions are as follows. RLO, JARW, RWS, RAG and MAH designed the experiment. AE, TM-C, AKC and JARW were involved in sample preparation. AE, TM-C, RWS, JARW, MVA, JHB, DA, DAS, HMED, HS, KT, SO, RLO and MAH participated in data collection. AE, TM-C, HMED, MVA, DA, RLO and MAH analysed the data. AE, TM-C, RLO and MAH wrote the manuscript with contributions from all authors.

Funding information

AE was supported by a joint Diamond–University of Essex PhD studentship. TM-C was supported by Leverhulme Trust award RPG-2014-355 to MAH and RWS. MH acknowledges support from BBSRC award BB/M022714/1 and JARW, DAS, DA, RLO, MAH and RWS acknowledge support from BBSRC Japan–UK International Partnering Award BB/R021015/1. HMED was supported by The Wellcome Trust (award ALR00750). RAG was supported by National Science Foundation Award CHE-1609446. The University of Essex and Diamond Light Source contributed equally to this work.

References

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. & Zwart, P. H. (2010). *Acta Cryst.* D**66**, 213–221.
Antonyuk, S. V., Strange, R. W., Sawers, G., Eady, R. R. & Hasnain, S. S. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 12041–12046.
Barrios, D. A., D'Antonio, J., McCombs, N. L., Zhao, J., Franzen, S., Schmidt, A. C., Sombers, L. A. & Ghiladi, R. A. (2014). *J. Am. Chem. Soc.* **136**, 7914–7925.
Barty, A., Kirian, R. A., Maia, F. R. N. C., Hantke, M., Yoon, C. H., White, T. A. & Chapman, H. (2014). *J. Appl. Cryst.* **47**, 1118–1131.
Beitlich, T., Kühnel, K., Schulze-Briese, C., Shoeman, R. L. & Schlichting, I. (2007). *J. Synchrotron Rad.* **14**, 11–23.
Bublitz, M., Nass, K., Drachmann, N. D., Markvardsen, A. J., Gutmann, M. J., Barends, T. R. M., Mattle, D., Shoeman, R. L., Doak, R. B., Boutet, S., Messerschmidt, M., Seibert, M. M., Williams, G. J., Foucar, L., Reinhard, L., Sitsel, O., Gregersen, J. L., Clausen, J. D., Boesen, T., Gotfryd, K., Wang, K.-T., Olesen, C., Møller, J. V., Nissen, P. & Schlichting, I. (2015). *IUCrJ*, **2**, 409–420.
Collins, P. M., Douangamath, A., Talon, R., Dias, A., Brandao-Neto, J., Krojer, T. & von Delft, F. (2018). *Methods Enzymol.* **610**, 251–264.
Colpa, D. I., Fraaije, M. W. & van Bloois, E. (2014). *J. Ind. Microbiol. Biotechnol.* **41**, 1–7.
Debreczeni, J. É. & Emsley, P. (2012). *Acta Cryst.* D**68**, 425–430.
Ebrahim, A., Appleby, M. V., Axford, D., Beale, J., Moreno-Chicano, T., Sherrell, D. A., Strange, R. W., Hough, M. A. & Owen, R. L. (2019). *Acta Cryst.* D**75**, 151–159.
Ebrahim, A., Moreno-Chicano, T., Appleby, M. V., Chaplin, A. K., Beale, J. H., Sherrell, D. A., Duyvesteyn, H. M. E., Owada, S., Tono, K., Sugimoto, H., Strange, R. W., Worrall, J. A. R., Axford, D., Owen, R. L. & Hough, M. A. (2019). *IUCrJ*, **6**, 543–551.
Emsley, P. (2017). *Acta Cryst.* D**73**, 203–210.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Fischer, M., Shoichet, B. K. & Fraser, J. S. (2015). *ChemBioChem*, **16**, 1560–1564.
Franzen, S., Thompson, M. K. & Ghiladi, R. A. (2012). *Biochim. Biophys. Acta*, **1824**, 578–588.
Fry, E. E., Grimes, J. & Stuart, D. I. (1999). *Mol. Biotechnol.* **12**, 13–23.
Guengerich, F. P., Waterman, M. R. & Egli, M. (2016). *Trends Pharmacol. Sci.* **37**, 625–640.
Horrell, S., Antonyuk, S. V., Eady, R. R., Hasnain, S. S., Hough, M. A. & Strange, R. W. (2016). *IUCrJ*, **3**, 271–281.
Horrell, S., Kekilli, D., Sen, K., Owen, R. L., Dworkowski, F. S. N., Antonyuk, S. V., Keal, T. W., Yong, C. W., Eady, R. R., Hasnain, S. S., Strange, R. W. & Hough, M. A. (2018). *IUCrJ*, **5**, 283–292.
Horrell, S., Kekilli, D., Strange, R. W. & Hough, M. A. (2017). *Metallomics*, **9**, 1470–1482.
Inoue, I., Inubushi, Y., Sato, T., Tono, K., Katayama, T., Kameshima, T., Ogawa, K., Togashi, T., Owada, S., Amemiya, Y., Tanaka, T., Hara, T. & Yabashi, M. (2016). *Proc. Natl Acad. Sci. USA*, **113**, 1492–1497.
Ishikawa, T., Aoyagi, H., Asaka, T., Asano, Y., Azumi, N., Bizen, T., Ego, H., Fukami, K., Fukui, T., Furukawa, Y., Goto, S., Hanaki, H., Hara, T., Hasegawa, T., Hatsui, T., Higashiya, A., Hirono, T., Hosoda, N., Ishii, M., Inagaki, T., Inubushi, Y., Itoga, T., Joti, Y., Kago, M., Kameshima, T., Kimura, H., Kirihara, Y., Kiyomichi, A., Kobayashi, T., Kondo, C., Kudo, T., Maesaka, H., Marechal, X. M., Masuda, T., Matsubara, S., Matsumoto, T., Matsushita, T., Matsui, S., Nagasono, M., Nariyama, N., Ohashi, H., Ohata, T., Ohshima, T., Ono, S., Otake, Y., Saji, C., Sakurai, T., Sato, T., Sawada, K., Seike, T., Shirasawa, K., Sugimoto, T., Suzuki, S., Takahashi, S., Takebe, H., Takeshita, K., Tamasaku, K., Tanaka, H., Tanaka, R., Tanaka, T., Togashi, T., Togawa, K., Tokuhisa, A., Tomizawa, H., Tono, K., Wu, S. K., Yabashi, M., Yamaga, M., Yamashita, A., Yanagida, K., Zhang, C., Shintake, T., Kitamura, H. & Kumagai, N. (2012). *Nat. Photonics*, **6**, 540–544.
Keedy, D. A., Hill, Z. B., Biel, J. T., Kang, E., Rettenmaier, T. J., Brandao-Neto, J., Pearce, N. M., von Delft, F., Wells, J. A. & Fraser, J. S. (2018). *Elife*, **7**, e36307.
Kekilli, D., Moreno-Chicano, T., Chaplin, A. K., Horrell, S., Dworkowski, F. S. N., Worrall, J. A. R., Strange, R. W. & Hough, M. A. (2017). *IUCrJ*, **4**, 263–270.
Lomb, L., Barends, T. R. M., Kassemeyer, S., Aquila, A., Epp, S. W., Erk, B., Foucar, L., Hartmann, R., Rudek, B., Rolles, D., Rudenko, A., Shoeman, R. L., Andreasson, J., Bajt, S., Barthelmess, M., Barty, A., Bogan, M. J., Bostedt, C., Bozek, J. D., Caleman, C., Coffee, R., Coppola, N., DePonte, D. P., Doak, R. B., Ekeberg, T., Fleckenstein, H., Fromme, P., Gebhardt, M., Graafsma, H., Gumprecht, L., Hampton, C. Y., Hartmann, A., Hauser, G., Hirsemann, H., Holl, P., Holton, J. M., Hunter, M. S., Kabsch, W., Kimmel, N., Kirian, R. A., Liang, M. N., Maia, F. R. N. C., Meinhart, A., Marchesini, S., Martin, A. V., Nass, K., Reich, C., Schulz, J., Seibert, M. M., Sierra, R., Soltau, H., Spence, J. C. H., Steinbrener, J., Stellato, F., Stern, S., Timneanu, N., Wang, X. Y., Weidenspointner, G., Weierstall, U., White, T. A., Wunderer, C., Chapman, H. N., Ullrich, J., Strüder, L. & Schlichting, I. (2011). *Phys. Rev. B*, **84**, 214111.
Long, F., Nicholls, R. A., Emsley, P., Gražulis, S., Merkys, A., Vaitkus, A. & Murshudov, G. N. (2017). *Acta Cryst.* D**73**, 112–122.
McCombs, N. L., Moreno-Chicano, T., Carey, L. M., Franzen, S., Hough, M. A. & Ghiladi, R. A. (2017). *Biochemistry*, **56**, 2294–2303.
McCombs, N. L., Smirnova, T. & Ghiladi, R. A. (2017). *Catal. Sci. Technol.* **7**, 3104–3118.
McGuire, A. H., Carey, L. M., de Serrano, V., Dali, S. & Ghiladi, R. A. (2018). *Biochemistry*, **57**, 4455–4468.
McLean, K. J. & Munro, A. W. (2017). *Drug Discov. Today*, **22**, 566–575.
McPherson, A. (2019). *Acta Cryst.* F**75**, 132–140.
Meents, A., Wiedorn, M. O., Srajer, V., Henning, R., Sarrou, I., Bergtholdt, J., Barthelmess, M., Reinke, P. Y. A., Dierksmeyer, D., Tolstikova, A., Schaible, S., Messerschmidt, M., Ogata, C. M., Kissick, D. J., Taft, M. H., Manstein, D. J., Lieske, J., Oberthuer, D., Fischetti, R. F. & Chapman, H. N. (2017). *Nat. Commun.* **8**, 1281.
Meyder, A., Nittinger, E., Lange, G., Klein, R. & Rarey, M. (2017). *J. Chem. Inf. Model.* **57**, 2437–2447.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Naitow, H., Matsuura, Y., Tono, K., Joti, Y., Kameshima, T., Hatsui, T., Yabashi, M., Tanaka, R., Tanaka, T., Sugahara, M., Kobayashi, J., Nango, E., Iwata, S. & Kunishima, N. (2017). *Acta Cryst.* D**73**, 702–709.
Nass, K. (2019). *Acta Cryst.* D**75**, 211–218.
Nass, K., Foucar, L., Barends, T. R. M., Hartmann, E., Botha, S., Shoeman, R. L., Doak, R. B., Alonso-Mori, R., Aquila, A., Bajt, S., Barty, A., Bean, R., Beyerlein, K. R., Bublitz, M., Drachmann, N., Gregersen, J., Jönsson, H. O., Kabsch, W., Kassemeyer, S., Koglin, J. E., Krumrey, M., Mattle, D., Messerschmidt, M., Nissen, P., Reinhard, L., Sitsel, O., Sokaras, D., Williams, G. J., Hau-Riege, S., Timneanu, N., Caleman, C., Chapman, H. N., Boutet, S. & Schlichting, I. (2015). *J. Synchrotron Rad.* **22**, 225–238.
Oghbaey, S., Sarracini, A., Ginn, H. M., Pare-Labrosse, O., Kuo, A., Marx, A., Epp, S. W., Sherrell, D. A., Eger, B. T., Zhong, Y., Loch, R., Mariani, V., Alonso-Mori, R., Nelson, S., Lemke, H. T., Owen, R. L., Pearson, A. R., Stuart, D. I., Ernst, O. P., Mueller-Werkmeister, H. M. & Miller, R. J. D. (2016). *Acta Cryst.* D**72**, 944–955.
Pearce, N. M., Krojer, T., Bradley, A. R., Collins, P., Nowak, R. P., Talon, R., Marsden, B. D., Kelm, S., Shi, J., Deane, C. M. & von Delft, F. (2017). *Nat. Commun.* **8**, 15123.
Pearce, N. M., Krojer, T. & von Delft, F. (2017). *Acta Cryst.* D**73**, 256–266.
Price, A. J., Howard, S. & Cons, B. D. (2017). *Essays Biochem.* **61**, 475–484.

Rendic, S. & Guengerich, F. P. (2015). *Chem. Res. Toxicol.* **28**, 38–42.
Schlichting, I. (2015). *IUCrJ*, **2**, 246–255.
Schulz, E. C., Mehrabi, P., Müller-Werkmeister, H. M., Tellkamp, F., Jha, A., Stuart, W., Persch, E., De Gasparo, R., Diederich, F., Pai, E. F. & Miller, R. J. D. (2018). *Nat. Methods*, **15**, 901–904.
Smart, O. S., Horský, V., Gore, S., Svobodová Vařeková, R., Bendová, V., Kleywegt, G. J. & Velankar, S. (2018). *Acta Cryst.* D**74**, 228–236.
Sugano, Y. (2009). *Cell. Mol. Life Sci.* **66**, 1387–1403.
Weierstall, U., James, D., Wang, C., White, T. A., Wang, D., Liu, W., Spence, J. C. H., Doak, R. B., Nelson, G., Fromme, P., Fromme, R., Grotjohann, I., Kupitz, C., Zatsepin, N. A., Liu, H., Basu, S., Wacker, D., Han, G. W., Katritch, V., Boutet, S., Messerschmidt, M., Williams, G. J., Koglin, J. E., Seibert, M. M., Klinker, M., Gati, C., Shoeman, R. L., Barty, A., Chapman, H. N., Kirian, R. A., Beyerlein, K. R., Stevens, R. C., Li, D., Shah, S. T., Howe, N., Caffrey, M. & Cherezov, V. (2014). *Nat. Commun.* **5**, 3309.
Weinert, T., Olieric, N., Cheng, R., Brünle, S., James, D., Ozerov, D., Gashi, D., Vera, L., Marsh, M., Jaeger, K., Dworkowski, F., Panepucci, E., Basu, S., Skopintsev, P., Doré, A. S., Geng, T., Cooke, R. M., Liang, M., Prota, A. E., Panneels, V., Nogly, P., Ermler, U., Schertler, G., Hennig, M., Steinmetz, M. O., Wang, M. & Standfuss, J. (2017). *Nat. Commun.* **8**, 542.
White, T. A., Mariani, V., Brehm, W., Yefanov, O., Barty, A., Beyerlein, K. R., Chervinskii, F., Galli, L., Gati, C., Nakane, T., Tolstikova, A., Yamashita, K., Yoon, C. H., Diederichs, K. & Chapman, H. N. (2016). *J. Appl. Cryst.* **49**, 680–689.
Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S. & Richardson, J. S. (2018). *Protein Sci.* **27**, 293–315.

JOURNAL OF SYNCHROTRON RADIATION

ISSN 1600-5775

Reducing sample consumption for serial crystallography using acoustic drop ejection

Bradley Davy,[a] Danny Axford,[a] John H. Beale,[a] Agata Butryn,[a] Peter Docker,[a] Ali Ebrahim,[b,a] Gabriel Leen,[c,d] Allen M. Orville,[a,e] Robin L. Owen[a]* and Pierre Aller[a]*

[a]Diamond Light Source, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0DE, UK, [b]School of Biological Sciences, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK, [c]PolyPico Technologies Ltd, Unit 10, Airways Technology Park, Rathmacullig Wes, Ballygarvan, Cork T12 DY95, Ireland, [d]Department of Electronic and Computer Engineering, University of Limerick, Ireland, and [e]Research Complex at Harwell, Rutherford Appleton Laboratory, Didcot OX11 0FA, UK. *Correspondence e-mail: robin.owen@diamond.ac.uk, pierre.aller@diamond.ac.uk

Received 22 March 2019
Accepted 28 June 2019

Edited by G. Grübel, HASYLAB at DESY, Germany

Keywords: serial crystallography; sample delivery; fixed targets; acoustic dispensing.

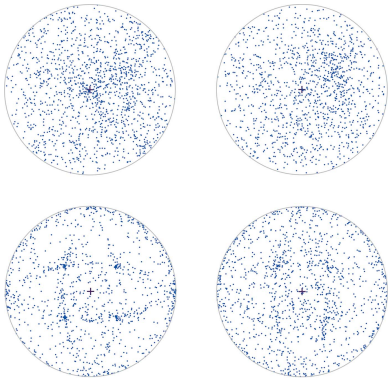
Efficient sample delivery is an essential aspect of serial crystallography at both synchrotrons and X-ray free-electron lasers. Rastering fixed target chips through the X-ray beam is an efficient method for serial delivery from the perspectives of both sample consumption and beam time usage. Here, an approach for loading fixed targets using acoustic drop ejection is presented that does not compromise crystal quality, can reduce sample consumption by more than an order of magnitude and allows serial diffraction to be collected from a larger proportion of the crystals in the slurry.

1. Introduction

Serial femtosecond crystallography (SFX) at X-ray free-electron lasers (XFELs) has become an important facet of the crystallographers' toolbox, allowing both time-resolved and ground-state measurements of X-ray sensitive samples (Schlichting, 2015). However, the opportunities that the high peak brilliance and femtosecond duration of XFEL pulses provide come with a challenge: the need to provide new samples at the repetition rate of the X-ray source or detector. Several approaches have been developed to meet this challenge ranging from liquid jets to high viscosity extruders, on-demand droplet injectors coupled to a tape drive, and fixed targets (Grünbein & Kovacs, 2019; Martiel *et al.*, 2019). This complementary range of delivery solutions means that an approach can be chosen and tailored to best suit the experiment at hand.

The success and impact of SFX has inspired the development and implementation of serial synchrotron crystallography (SSX), where many of the same sample delivery techniques are used (Diederichs & Wang, 2017). The subsequent success of SSX has now driven the development of synchrotron beamlines dedicated to serial crystallography, such as P14.EH2 at PETRA III (http://www.embl-hamburg.de/services/mx/P14_EH2/index.html), and this illustrates the desire of structural biologists to exploit serial approaches.

A challenge common to many serial approaches is sample consumption. The volume of sample consumed is many orders of magnitude greater than that required for traditional synchrotron approaches when a complete dataset may be obtained from a single crystal held at 100 K. Indeed, the sample requirements for a serial experiment are often an unwelcome surprise for the first-time user of SFX or SSX, as usually only a single 'still' image is collected from each crystal.



OPEN ACCESS

This is reflected by developments to reduce sample consumption for serial experiments such as flow-focusing for in-flow SSX (Monteiro *et al.*, 2019).

Recently, we have developed fixed target sample delivery as a serial approach that works well at both synchrotrons and XFELs. These are based on thin films (Doak *et al.*, 2018) and, predominantly for us, silicon nitride 'chips' (Ebrahim *et al.*, 2019). Typically, to load a silicon nitride chip, ~100–200 µl of crystal slurry is required. This slurry is pipetted over a chip and crystals are drawn to the apertures through use of a weak vacuum. Sufficient data for structure solution can typically be obtained from a single chip.

Acoustic dispensing is a technique that uses high-frequency acoustic waves to dispense small volumes of liquid. The ejected droplets may contain protein crystals (Soares *et al.*, 2011; Roessler *et al.*, 2016; Fuller *et al.*, 2017), live cells (Demirci & Montesano, 2007) or indeed almost any small molecule (Teplitsky *et al.*, 2015). Commonly referred to as acoustic drop ejection (ADE), here we use a variant that makes use of disposable dispensing cartridges allowing rapid switching between samples (Leen, 2016). Using the commercially available PolyPico pico-litre dispenser (https://www.polypico.com) synchronized with compact, high-precision *xyz* stages (http://www.smaract.com), we demonstrate the use of ADE to dramatically reduce sample consumption for fixed target serial crystallography.

2. Methods

The loading of fixed targets using acoustic dispensing is a two-step process with a calibration step required prior to chip loading. For convenience we physically separate these steps [Figs. 1(*a*) and 1(*b*)]. The same PolyPico head dispenser was used for both aspects of the experiment and was mounted on kinematic mounts allowing transfer between calibration and loading in a few seconds.

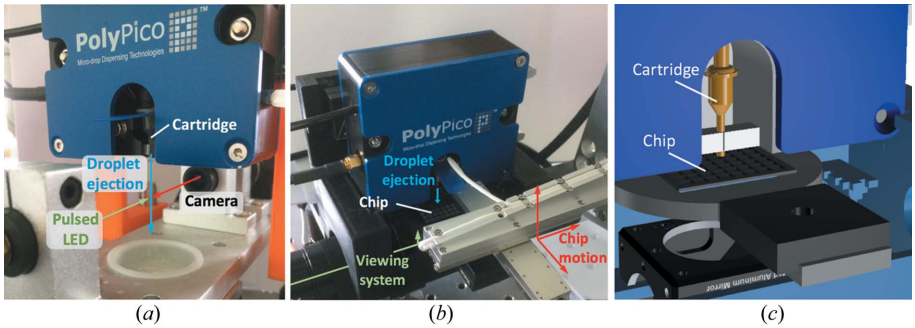
2.1. Drop calibration

For optimal loading of chips, the volume of droplets ejected by the PolyPico dispenser should be calibrated for each crystal slurry. Using a pipette and a tip-like adapter, the crystal slurry is loaded into a cartridge which has a dispensing aperture ranging from 30 µm to 150 µm in diameter. For the experiments described here, we loaded 10–20 µl of slurry into cartridges with any slurry not used easily recovered after the experiment using the same pipette and adapter. The cartridge aperture size is chosen based on the typical size of the crystals in the slurry. In practice, we find that an aperture diameter approximately twice the size of the crystals used works well as a compromise between minimizing drop size and avoiding clogging if larger crystals are present. The width, amplitude and frequency of the acoustic wave applied to the cartridge base must be tuned until stable droplets are ejected from the crystal slurry. Ejected droplets are visualized using a high-resolution camera and stroboscopic LED [Fig. 1(*a*)] with image recognition software allowing real-time readback of the average droplet volume. Typically, when using a 1 kHz acoustic wave and a cartridge aperture of 100 µm, 80–100 pl (approximate diameter 60 µm) droplets can be obtained. Once the optimal parameters for ADE of crystal slurry have been determined, chips can be loaded.

2.2. Chip loading

The setup for ADE loading of fixed targets is shown in Figs. 1(*b*) and 1(*c*). Chips are mounted on a three-axis stage and can be viewed through a high-resolution camera which allows viewing of both fixed targets and droplets ejected by the dispensing head. The tip of the dispensing head is within 0.5 mm of the surface of the chip. Following alignment of chip fiducials, chips can be moved as previously described (Sherrell *et al.*, 2015). In this case the stages act as the 'master', sending a TTL pulse to the dispensing head with droplets ejected on demand when each aperture is reached. Following the ejection of a user-defined number of droplets at 1 kHz, the stages move to the next aperture on the chip. The loading of a chip with 25 600 positions takes less than 4 min and consumes less than 4 µl of slurry. To avoid dehydration, the chip and dispensing head are enclosed in a high-relative-humidity environment (>90%) [Fig. 1(*b*)]. Following loading, the chips are sealed with a thin film (typically 6 µm) of mylar. Chips with a funnel-shaped aperture (size of the small end of the funnel: 7 µm) were used; the volume of each aperture was ~160 pl, and apertures are spaced by 125 µm (centre-of-aperture to centre-of-aperture distance).

In order to conserve sample and also minimize the beam time required for X-ray data collection, only the central area of chips was acoustically loaded (6 × 6 'city blocks', 14 400 apertures) in the experiments described here. In this case, the time required to load a chip was 2 min 15 s. In total, the complete acoustic loading process including alignment and loading takes approximately 5 min (full chip), and throughput is equal to or faster than X-ray data



(*a*) (*b*) (*c*)

Figure 1
Experimental setup for (*a*) calibration of ejected droplets and (*b*) chip loading; in each, the direction of droplet ejection is shown in blue. (*c*) Schematic of chip loading from a similar viewpoint to (*b*) with translation stages hidden and the cartridge highlighted in yellow.

collection. Chips were loaded by both using the ADE approach described above and also, for comparison, manually using a pipette.

2.3. Sample preparation

Microcrystals of chicken egg-white lysozyme (HEWL) were prepared using an adaptation of a previously described protocol. In brief, high-purity lysozyme powder (Sigma–Aldrich L6876-5 G) was resuspended in 100 m*M* sodium acetate pH 3.0 to a final concentration of 25 mg ml^{-1} and mixed with an equal volume of crystallization buffer (16.8% *w*/*v* sodium chloride, 4.8% *w*/*v* PEG 6000 and 0.06 *M* sodium acetate pH 3.0) at room temperature. The mixture was vortexed for 10 s and then left for an hour until crystal growth saturation. Using this method, we obtained homogeneous rectangular crystals with an average size of 10 µm × 10 µm × 15 µm. Microcrystals of copper nitrite reductase from *Achromobacter cycloclastes* (*Ac*NiR) ranging in size from 15 µm to 70 µm were grown using a protocol described previously (Ebrahim *et al.*, 2019). The concentration of crystals in each slurry was estimated using a Hemocytometer cell counter.

2.4. X-ray data collection

Following loading, chips were transferred to Beamline I24, Diamond Light Source. Diffraction data were collected as previously described (Owen *et al.*, 2017) though only from the central region of chips loaded by the PolyPico (14 400 apertures), with data collection taking 4 min 20 s. Data were collected using an X-ray energy of 12.8 keV, a beam size of 7 µm × 6 µm, 10 ms exposures and a flux attenuated to 8×10^{11} photons s^{-1}.

Hit-rates were obtained using dials.stills_process (Winter *et al.*, 2018; Brewster *et al.*, 2016, 2018) with up to ten lattices per image indexed. Subsequent scaling and merging of data was performed using *PRIME* (Uervirojnangkoorn *et al.*, 2015). In all cases the majority of indexed images contained a single lattice with the percentage of single lattice images being 77% (HEWL, PolyPico loaded), 81% (HEWL, pipette loaded), 85% (*Ac*NiR PolyPico loaded) and 66% (*Ac*NiR, pipette loaded). In the following, we define the diffraction hit-rate as the total number of indexed patterns divided by the number of collected images.

3. Results

In preparatory experiments, we varied and defined the optimal number of acoustically ejected droplets. When dispensing two drops per single chip aperture, we observed higher hit-rates than when using a single droplet. The dispensing of three or more drops overflowed the apertures resulting in excess liquid on the surface of the chip. Therefore, all of the results presented here were obtained using two droplets per aperture.

Diffraction hit-rates for HEWL crystals loaded manually and using acoustic dispensing as a function of crystal concentration are shown in Table 1. As might be expected, in both cases diffraction hit-rates increase with crystal concentration. Also, for a given concentration, higher diffraction hit-rates are obtained using pipette loading. However, acoustic loading requires a significantly lower volume crystal slurry to achieve these, as illustrated by the number of diffraction hits obtained per dispensed microlitre of crystal slurry (Fig. 2).

Table 1
HEWL loading parameters and hit-rates.

The diffraction hit-rate (DHR) and absolute hit-rate (AHR) are defined in the text. Volumes dispensed (*V*) for the PolyPico are an upper bound: droplet volumes vary from drop to drop so a conservative average value is used.

Loading method	Crystal concentration (crystals ml^{-1})	*V* (µl)	Calculated crystals dispensed	Indexed patterns	DHR (%)	AHR (%)
Pipette	5×10^4	75	3750	2053	14.3	55
Pipette	1×10^5	75	7500	2129	14.8	28
Pipette	5×10^5	75	37500	4850	33.7	13
Pipette	1×10^6	75	75000	10462	72.6	14
PolyPico	5×10^5	3	1500	1311	9.1	87
PolyPico	1×10^6	3	3000	1763	12.2	59
PolyPico	1.5×10^6	3	4500	2883	20.0	64
PolyPico	2×10^6	3	6000	4573	31.8	76
PolyPico	2.5×10^6	3	7500	3355	23.3	45

Similar trends are seen for *Ac*NiR crystals (Fig. 2, Table 2), which significantly differ from HEWL crystals both in shape and chemical composition of the crystallization conditions, with an increasing hit-rate for both pipette and acoustically loaded chips as a function of slurry concentration. Higher diffraction hit-rates are also seen for pipette loaded chips at the expense of increased sample consumption.

Using the crystal concentration measured as described above, the number of crystals used in each experiment, and hence the fraction from which diffraction was recorded, can be estimated. We refer to this quantity as the absolute hit-rate

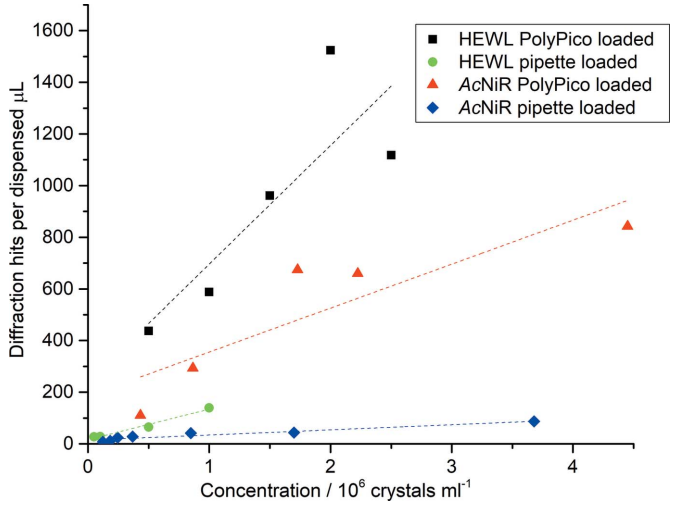


Figure 2
Diffraction hits per unit volume of dispensed crystal slurry. Acoustic dispensing results in more than a fivefold increase in hits per unit volume of slurry consumed at all concentrations for HEWL and a tenfold increase for *Ac*NiR.

Table 2
*Ac*NiR loading parameters and hit-rates.

The diffraction hit-rate (DHR) and absolute hit-rate (AHR) are defined in the text; V is the volume dispensed. Note that for pipette-loaded *Ac*NiR, diffraction data were collected from a full chip (25 600 apertures) in contrast to all other data which were collected from 14 400 apertures.

Loading method	Crystal concentration (crystals ml^{-1})	V (µl)	Calculated crystals dispensed	Indexed patterns	DHR (%)	AHR (%)
Pipette	1.2×10^5	150	18000	1185	4.6	6.6
Pipette	1.8×10^5	150	27000	1687	6.6	6.2
Pipette	2.5×10^5	150	37500	3574	14.0	9.5
Pipette	3.7×10^5	150	55500	4145	16.2	7.5
Pipette	8.5×10^5	150	127500	6223	24.3	4.9
Pipette	1.7×10^6	150	255000	6587	25.7	2.6
Pipette	3.7×10^6	150	555000	13061	51.0	2.4
PolyPico	4.3×10^5	3	1290	332	2.3	25.7
PolyPico	8.6×10^5	3	2580	878	6.1	34.0
PolyPico	1.7×10^6	3	5100	2023	14.1	39.7
PolyPico	2.2×10^6	3	6600	1979	13.7	30.0
PolyPico	4.4×10^6	3	13200	2526	17.5	19.1

and it is given for HEWL and *Ac*NiR in Tables 1 and 2, respectively. For pipette loading, it can be seen that, although higher diffraction hit-rates are achieved by increasing the crystal slurry concentration (this could also be achieved by simply increasing the volume of slurry loaded onto the chip), this is at the expense of the absolute hit-rate, with diffraction recorded from a decreasing proportion of the crystals used in the loading process. Although the diffraction hit-rate may be lower for acoustically loaded fixed targets, a larger proportion of the crystals grown produce a diffraction pattern.

Importantly, the loading method does not significantly affect the quality of diffraction observed (Fig. 3). Both HEWL and *Ac*NiR crystals exhibit similar R_{split} and $CC_{1/2}$ for both acoustic and pipette loading, and in all cases data quality is high. Differences in quality observed are of the same order as chip-to-chip variation when using the same loading approach, thus diffraction quality is not compromised by acoustic loading.

For both pipette and acoustically loaded chips crystals are observed to be predominantly randomly orientated on the chips, illustrated by the stereographic projections in Fig. 4. To generate these plots data were reindexed in *P*1 so no symmetry equivalents are plotted. For more heavily loaded chips we do see some indication of systematic orientations, and this starts to become apparent in the case of pipette loaded HEWL [Fig. 4(*c*)]. Two orthogonal ellipses with a width of 70° at the centre of the projection become visible. These are consistent with loaded crystals lying on the internal walls of the chip apertures which are chemically etched along the silicon 111 crystal planes 54.74° from the surface of the chip (35.26° to the beam direction), with the ellipses reminiscent of stereographic projections of silicon etch planes as illustrated by Seidel *et al.* (1990). The degree of observed systematic orientation is likely to be dependent on the density of sample on the chip, crystal size and morphology, and also loading method, with acoustic loading less likely to yield systematic orientations.

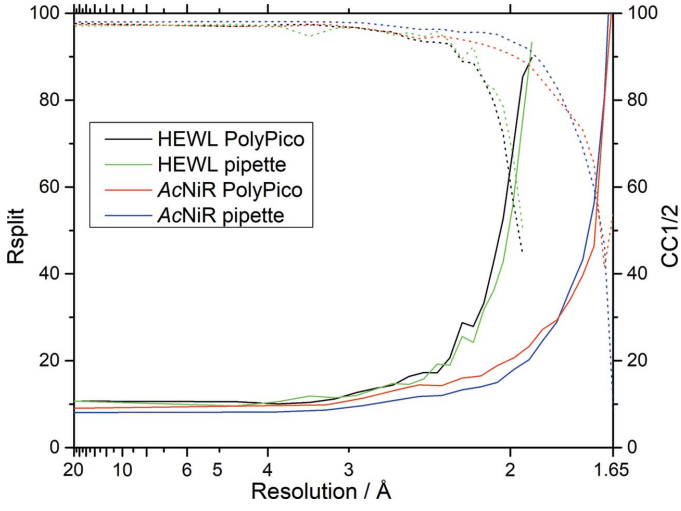


Figure 3
Data merging statistics. R_{split} (solid line) and $CC_{1/2}$ (dotted line) are shown for HEWL (black and green) and *Ac*NiR (red and blue) crystals loaded on the chip with either PolyPico or a pipette. Scaling and merging have been performed using *PRIME* on 12 546 and 13 563 integrated images for HEWL and AcNiR, respectively.

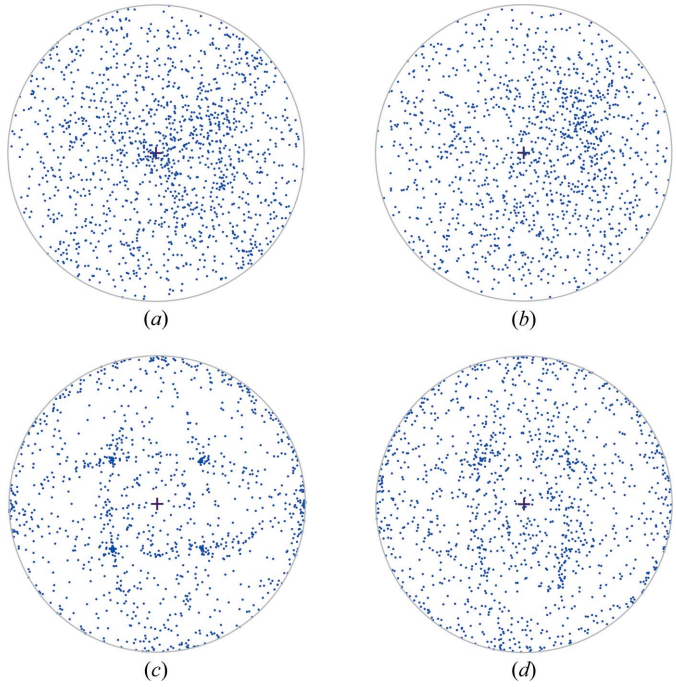


Figure 4
Stereographic projections illustrating the crystal orientation of 1000 randomly selected crystals for each loading method and crystal type: (*a*) pipette-loaded *Ac*NiR, (*b*) PolyPico-loaded *Ac*NiR, (*c*) pipette-loaded HEWL and (*d*) PolyPico-loaded HEWL. The plots represent the direction of the 001 *hkl* of each crystal (reindexed in *P*1) relative to the beam direction (*z*) which is shown as the central '+' into the page. A point at 12 o'clock on the circular projection represents a 90° rotation of the crystal around *x* whereas the point at 3 o'clock represents a 90° rotation around *y*. Plots were produced using the module dials.stereographic_projection (Winter *et al.*, 2018).

4. Discussion

Acoustic dispensing provides a means of reducing sample consumption for serial crystallography without compromising crystal quality with high-quality diffraction observed using both loading approaches. Acoustic dispensing has been previously exploited in the context of sample delivery, whereas here it is used for loading fixed targets that are subsequently passed to the beamline. This decoupling of acoustic ejection and X-ray data collection is advantageous as time taken to optimize drop ejection, which varies with the composition of the crystal slurry, does not impact the beam time efficiency.

Optimal loading is obtained with crystals less than ~50 µm in size using cartridges with a 100 µm aperture. Increased hit-rates are obtained as the crystal slurry concentration increases, though settling of larger crystals or clumping can cause the PolyPico aperture to clog with time. This may explain why diffraction hit-rates do not increase as much as expected at the highest slurry concentration (Fig. 2). At lower crystal concentrations, we observe that the ejection process visibly disturbs the crystal slurry within the cartridge, slowing any settling process and multiple chips can be loaded from the same cartridge. Any long-term crystal settling can also be addressed by removing and reinserting the cartridge to resuspend the crystal slurry. In order to minimize any potential settling for high slurry concentrations, future loading setups will either make use of a rocking system or cartridges will be fed through a capillary fed by a syringe mounted on a rocker as used by Fuller *et al.* (2017).

While higher diffraction hit-rates can be obtained using traditional pipette loading, this is at the expense of increased sample consumption and the proportion of prepared crystals from which diffraction data are collected (*i.e.* the absolute hit-rate) falls. To obtain a similar number of indexed images, acoustic dispensing consumes tenfold less crystal slurry (*Ac*NiR, Fig. 2) than traditional pipette loading at the same sample concentration. Acoustic loading has the additional benefit that an increased fraction of the crystals produced for, and consumed by, the experiment result in diffraction. The success of acoustic loading is dependent on the chemical composition and viscosity of the crystal slurry and the parameters of the acoustic wave need to be optimized for each sample. As more viscous media may not be suitable for acoustic dispensing and the effect of crystal morphology is as yet unclear, acoustic loading of fixed targets is very much a complementary technique to pipette loading. We have demonstrated, however, that if samples are scarce, acoustic loading can help ensure a larger fraction of crystals see the X-ray beam and reduce the volume of sample required.

Funding information

Diffraction data were measured on beamline I24 at Diamond Light Source (proposal Nos. NT14493 and MX19458). This work was supported by Diamond Light Source, which receives funding from the UK Government and the Wellcome Trust. PA, AB and AMO were also partially supported by funding from the Biotechnology and Biological Sciences Research Council and a Strategic Award from the Wellcome Trust (102593); AMO was also partially supported by a Royal Society Wolfson Fellowship (RSWF\R2\182017) and a Wellcome Investigator Award in Science (210734/Z/18/Z). GL is Founder, Director and CTO of PolyPico Technologies Ltd (Cork, Ireland).

References

Brewster, A. S., Waterman, D., Parkhurst, J. M., Gildea, R. J., Michels-Clark, T., Young, I. D., Bernstein, H. J., Winter, G., Evans, G. & Sauter, N. K. (2016). *Comput. Crystallogr. Newsl.* **7**, 32–53.

Brewster, A. S., Waterman, D. G., Parkhurst, J. M., Gildea, R. J., Young, I. D., O'Riordan, L. J., Yano, J., Winter, G., Evans, G. & Sauter, N. K. (2018). *Acta Cryst.* D**74**, 877–894.

Demirci, U. & Montesano, G. (2007). *Lab Chip*, **7**, 1139–1145.

Diederichs, K. & Wang, M. (2017). *Protein Crystallography: Methods and Protocols*, edited by A. Wlodawer, Z. Dauter & M. Jaskolski, pp. 239–272. New York: Springer.

Doak, R. B., Nass Kovacs, G., Gorel, A., Foucar, L., Barends, T. R. M., Grünbein, M. L., Hilpert, M., Kloos, M., Roome, C. M., Shoeman, R. L., Stricker, M., Tono, K., You, D., Ueda, K., Sherrell, D. A., Owen, R. L. & Schlichting, I. (2018). *Acta Cryst.* D**74**, 1000–1007.

Ebrahim, A., Appleby, M. V., Axford, D., Beale, J., Moreno-Chicano, T., Sherrell, D. A., Strange, R. W., Hough, M. A. & Owen, R. L. (2019). *Acta Cryst.* D**75**, 151–159.

Fuller, F. D., Gul, S., Chatterjee, R., Burgie, E. S., Young, I. D., Lebrette, H., Srinivas, V., Brewster, A. S., Michels-Clark, T., Clinger, J. A., Andi, B., Ibrahim, M., Pastor, E., de Lichtenberg, C., Hussein, R., Pollock, C. J., Zhang, M., Stan, C. A., Kroll, T., Fransson, T., Weninger, C., Kubin, M., Aller, P., Lassalle, L., Bräuer, P., Miller, M. D., Amin, M., Koroidov, S., Roessler, C. G., Allaire, M., Sierra, R. G., Docker, P. T., Glownia, J. M., Nelson, S., Koglin, J. E., Zhu, D., Chollet, M., Song, S., Lemke, H., Liang, M., Sokaras, D., Alonso-Mori, R., Zouni, A., Messinger, J., Bergmann, U., Boal, A. K., Bollinger, J. M. Jr, Krebs, C., Högbom, M., Phillips, G. N. Jr, Vierstra, R. D., Sauter, N. K., Orville, A. M., Kern, J., Yachandra, V. K. & Yano, J. (2017). *Nat. Methods*, **14**, 443–449.

Grünbein, M. L. & Nass Kovacs, G. (2019). *Acta Cryst.* D**75**, 178–191.

Leen, G. (2016). *Biosens. J.* **5**, 139.

Martiel, I., Müller-Werkmeister, H. M. & Cohen, A. E. (2019). *Acta Cryst.* D**75**, 160–177.

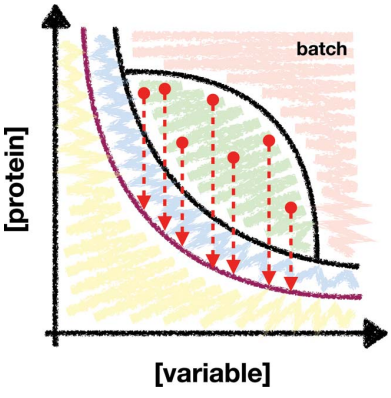
Monteiro, D. C. F., Vakili, M., Harich, J., Sztucki, M., Meier, S. M., Horrell, S., Josts, I. & Trebbin, M. (2019). *J. Synchrotron Rad.* **26**, 406–412.

Owen, R. L., Axford, D., Sherrell, D. A., Kuo, A., Ernst, O. P., Schulz, E. C., Miller, R. J. D. & Mueller-Werkmeister, H. M. (2017). *Acta Cryst.* D**73**, 373–378.

Roessler, C. G., Agarwal, R., Allaire, M., Alonso-Mori, R., Andi, B., Bachega, J. F. R., Bommer, M., Brewster, A. S., Browne, M. C., Chatterjee, R., Cho, E., Cohen, A. E., Cowan, M., Datwani, S., Davidson, V. L., Defever, J., Eaton, B., Ellson, R., Feng, Y., Ghislain, L. P., Glownia, J. M., Han, G., Hattne, J., Hellmich, J., Héroux, A., Ibrahim, M., Kern, J., Kuczewski, A., Lemke, H. T., Liu, P., Majlof, L., McClintock, W. M., Myers, S., Nelsen, S., Olechno, J., Orville, A. M., Sauter, N. K., Soares, A. S., Soltis, S. M., Song, H., Stearns, R. G., Tran, R., Tsai, Y., Uervirojnangkoorn, M., Wilmot, C. M., Yachandra, V., Yano, J., Yukl, E. T., Zhu, D. & Zouni, A. (2016). *Structure*, **24**, 631–640.

Schlichting, I. (2015). *IUCrJ*, **2**, 246–255.
Seidel, H., Csepregi, L., Heuberger, A. & Baumgartel, H. (1990). *J. Electrochem. Soc.* **137**, 3612–3626.
Sherrell, D. A., Foster, A. J., Hudson, L., Nutter, B., O'Hea, J., Nelson, S., Paré-Labrosse, O., Oghbaey, S., Miller, R. J. D. & Owen, R. L. (2015). *J. Synchrotron Rad.* **22**, 1372–1378.
Soares, A. S., Engel, M. A., Stearns, R., Datwani, S., Olechno, J., Ellson, R., Skinner, J. M., Allaire, M. & Orville, A. M. (2011). *Biochemistry*, **50**, 4399–4401.
Teplitsky, E., Joshi, K., Ericson, D. L., Scalia, A., Mullen, J. D., Sweet, R. M. & Soares, A. S. (2015). *J. Struct. Biol.* **191**, 49–58.
Uervirojnangkoorn, M., Zeldin, O. B., Lyubimov, A. Y., Hattne, J., Brewster, A. S., Sauter, N. K., Brunger, A. T. & Weis, W. I. (2015). *eLife*, **4**, e05421.
Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K. & Evans, G. (2018). *Acta Cryst.* D**74**, 85–97.

research paper



JOURNAL OF APPLIED CRYSTALLOGRAPHY

ISSN 1600-5767

Received 14 June 2019
Accepted 2 October 2019

Edited by J. Hajdu, Uppsala University, Sweden and The European Extreme LIght Infrastucture, Czech Republic

Keywords: serial macromolecular crystallography; XFELs; batch crystallization; vapour diffusion; micro-crystallization.

Supporting information: this article has supporting information at journals.iucr.org/j

Successful sample preparation for serial crystallography experiments

John H. Beale,[a]* Rachel Bolton,[a,b] Stephen A. Marshall,[c] Emma V. Beale,[a] Stephen B. Carr,[d] Ali Ebrahim,[a,e] Tadeo Moreno-Chicano,[f] Michael A. Hough,[e] Jonathan A. R. Worrall,[e] Ivo Tews[b] and Robin L. Owen[a]

[a]Diamond Light Source Ltd, Harwell Science and Innovation Campus, Fermi Avenue, Didcot OX11 0DE, UK, [b]Institute for Life Sciences, University of Southampton, Southampton SO17 1BJ, UK, [c]Manchester Institute of Biotechnology, The University of Manchester, Princess Street, Manchester M1 7DN, UK, [d]Research Complex at Harwell, Rutherford Appleton Laboratory, Harwell Oxford, Didcot OX11 0FA, UK, [e]School of Life Sciences, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK, and [f]Institute de Biologie Structurale, 71 Avenue des Martyrs, 38000 Grenoble, France. *Correspondence e-mail: john.beale@psi.ch

Serial crystallography, at both synchrotron and X-ray free-electron laser light sources, is becoming increasingly popular. However, the tools in the majority of crystallization laboratories are focused on producing large single crystals by vapour diffusion that fit the cryo-cooled paradigm of modern synchrotron crystallography. This paper presents several case studies and some ideas and strategies on how to perform the conversion from a single crystal grown by vapour diffusion to the many thousands of micro-crystals required for modern serial crystallography grown by batch crystallization. These case studies aim to show (i) how vapour diffusion conditions can be converted into batch by optimizing the length of time crystals take to appear; (ii) how an understanding of the crystallization phase diagram can act as a guide when designing batch crystallization protocols; and (iii) an accessible methodology when attempting to scale batch conditions to larger volumes. These methods are needed to minimize the sample preparation gap between standard rotation crystallography and dedicated serial laboratories, ultimately making serial crystallography more accessible to all crystallographers.

1. Introduction

1.1. Modern serial crystallography

Serial macromolecular crystallography (SMX), the collection and merging of data from multiple crystals, is not new. Prior to the widespread adoption of cryo-cooling methods in the early 1990s, data sets derived from many crystals were the norm. For certain types of protein crystal, particularly those of viral capsid proteins, cryo-cooling is not possible and the merging of multiple small wedge rotations is a necessary and effective way of acquiring a complete data set (Fry *et al.*, 1999). The availability of crystals of limited size may also require the use of a microfocus beamline and a similar multi-crystal–multi-wedge approach (Evans *et al.*, 2011). However, since the development of X-ray free-electron laser (XFEL) radiation sources, the number of Protein Data Bank (PDB; https://www.rcsb.org/) depositions from SMX methods has increased [Fig. 1(*a*)]. The XFEL beam destroys the sample upon interaction (Neutze *et al.*, 2000), precluding wedged data collection, and ultimately takes serial data collection to its logical extreme, *i.e.* one image per crystal. This necessitates the need for the delivery of a steady stream of hundreds or thousands of micro-crystals into the path of the X-ray beam in order to sample reciprocal space appropriately.

The sample requirements of modern SMX delivery approaches are, therefore, radically different from those of the single-crystal or conventional multi-crystal experiments, and so are the delivery approaches that have been devised to handle them. Broadly, four sample-delivery methods exist for SMX at XFELs and synchrotrons: jets (DePonte *et al.*, 2008; Sierra *et al.*, 2016; Weierstall *et al.*, 2012; Oberthuer *et al.*, 2017), extruders (Weierstall *et al.*, 2014; Botha *et al.*, 2015; Martin-Garcia *et al.*, 2017; Weinert *et al.*, 2017), acoustic drop ejectors (ADE) (Roessler *et al.*, 2013, 2016; Fuller *et al.*, 2017) and fixed targets (Frank *et al.*, 2014; Feld *et al.*, 2015; Hunter *et al.*, 2015; Murray *et al.*, 2015; Sherrell *et al.*, 2015; Roedig *et al.*, 2017). These categories are both broad and rapidly evolving due to the relative youth of modern SMX. This means there is a lack of standardization across facilities and laboratories, presenting a confusing picture to crystallographers wanting to practise SMX. This lack of standardization also makes direct comparisons challenging [see Grünbein & Nass Kovacs (2019) for a thorough overview]. However, all have different ideal sample requirements. The aim of the experiment should dictate the type of approach used. Therefore, this will also dictate the sample requirements. The delivery method and sample should then be combined with the optimum source to ensure acceptable hit rates for the experiment to be completed within the allocated beamtime. For example, if the investigation is a time-resolved study of a light-activated enzyme–substrate complex, a fixed-target approach could be used at a low-repetition-rate source, *e.g.* SACLA, Japan (Ishikawa *et al.*, 2012) or SwissFEL, Switzerland (Milne *et al.*, 2017). The fixed targets developed at Diamond Light Source, UK, are best loaded with 10–30 µm crystals at a concentration of 5–10 $\times$ 10^5 crystals ml^{-1} and require 100–150 µl of slurry per load (Davy *et al.*, 2019), but how can such a sample be created? What is the total sample volume that will be required during the experiment? The investigator wanting to perform this, or any, SMX experiment must grapple with these sample requirements, and it is these requirements that remain a serious impediment to the broader application of serial methods.

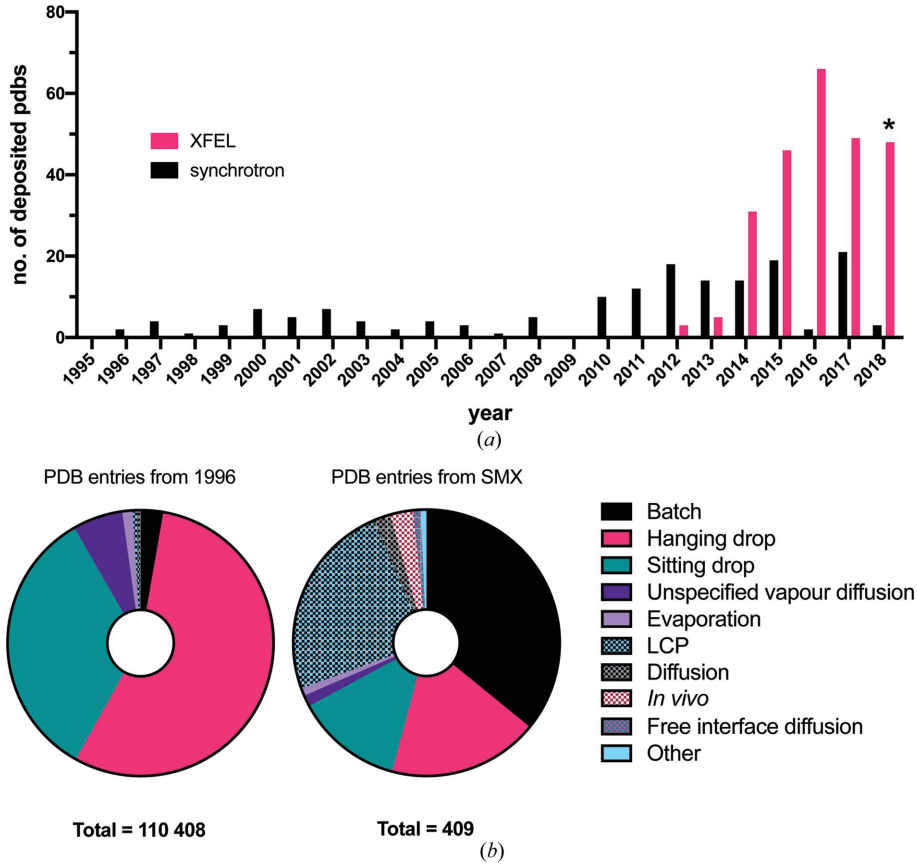


Figure 1
A summary of PDB depositions and crystallization methods from SMX experiments. (*a*) The frequency, plotted by year, of PDB depositions from serial experiments collected at XFEL and synchrotron light sources. PDB entries for this figure were selected on the basis of the number of reported crystals (>10), the reported radiation source and the indexing software used. The asterisk (*) indicates that the data from 2018 are not complete. (*b*) A comparison of the crystallization methods used in the PDB as a whole (left) with the serial experiments identified in panel (*a*) (right) over the same time period.

1.2. The re-emergence of batch methods

The large volumes of micro-crystalline samples required for SMX experiments also dictate the type of crystallization method to be used. Fig. 1(*b*) compares the relative abundance of different crystallization strategies over the same period for single-crystal crystallography and SMX. Vapour diffusion methods are significantly less popular for SMX than for single-crystal methods. SMX studies still use vapour diffusion methods but at a reduced frequency. Their place has principally been filled by batch methods, but also lipid cubic phase (LCP) and *in vivo* methods. The reason for the dominance of batch methods is perhaps not surprising, given an understanding of the crystallization process. The crystallization phase diagram [see Reis-Kautt & Ducruix (1992) and Rupp (2015) for in-depth descriptions of the kinetics and thermodynamics] highlights the problem with methods such as vapour diffusion [see Fig. 2(*a*)]. All crystallization methods apart from the batch approach rely upon a transition phase where the crystallization component concentrations must be 'driven' to the nucleation region by some process [Fig. 2(*b*)], *e.g.* drop equilibration (vapour diffusion).

This transition phase has several disadvantages, best exemplified by considering a vapour diffusion experiment. Firstly, the exact trajectory of the experiment is difficult to

ascertain. The starting point (protein and reservoir concentrations) and finishing point (appearance of crystals) can be inferred, but not the journey between the two, *i.e.* the exact conditions that gave rise to nucleation and subsequent crystal growth are not easy to determine. Secondly, as the component concentrations within the drop mixture have to 'move' into the nucleation zone, it can be difficult, though not impossible,[1] to penetrate the nucleation zone deeply [see blue dotted lines in Figs. 2(*a*) and 2(*b*)]. Transitionary phase micro-crystallization therefore requires a high rate of nucleation at the edge of the nucleation region. Finally, a successful condition in a small volume can be difficult to scale to a large volume. The exact kinetics within the drop might be essential for successful crystallization. Therefore, when scaling the experiment up to larger volumes, one must consider the additional challenge of maintaining the respective volumes of the reservoir, drop and space between.

In contrast with vapour diffusion, a batch experiment attempts to hit the nucleation zone immediately upon mixing of the protein and reservoir solutions (McPherson, 1982). The combination should create a supersaturated solution of protein which nucleates immediately. Possible batch crystallization trajectories are plotted in Fig. 2(*c*). Unlike vapour diffusion, the entire nucleation zone can be exploited in the experiment, potentially resulting in more nucleation. Scaling of the experiment is also simpler, since larger volumes of the reservoir and protein solution should produce similar results when mixed. A variant of the batch method, here called 'seeded batch', uses seeds (see Appendix *A*1 in the supporting information for a discussion of different types of seeds) as nucleants [Fig. 2(*d*)]. If the phase diagram is known, different regions of the metastable zone can be targeted to achieve different results. There are still questions as to the exact conditions that give rise to crystals in a batch experiment, such as how the protein and reservoir components interact in the pre-mixing time. However, these micro-scale effects will most likely be protein-condition specific and resolved naturally during the process of optimization.

The literature is not devoid of micro-crystallization examples, but a complete description of a method to make the transition from vapour diffusion to batch crystallization is currently lacking. Several papers have described techniques to identify micro-crystallization conditions using vapour diffusion. Luft *et al.* (2015) and Lee *et al.* (2018) both showed how nonlinear optics could be used to identify conditions which favour micro- (and nano-)crystalline growth in 96-well sitting-drop plates. Lee *et al.* (2018) also showed how adapting the vapour diffusion protocol using a 'controlled evaporation' approach increases the propensity for micro-crystallization. Both of these studies effectively focused on re-screening crystallization cocktails to find new conditions which yielded micro-crystals but did not suggest how then to scale these conditions for practical SMX. Other studies have focused on how to scale methods once a suitable condition has been identified. Ibrahim *et al.* (2015), using the case of Photosystem II, showed how different protein seed preparations and an understanding of the phase diagram could be used to find an optimum seeding protocol, whereas Kupitz *et al.* (2014) described practical large-scale methods, such as batch techniques and a novel adaptation of free-interface diffusion (FID). Darmanin *et al.* (2016) demonstrated how dynamic light scattering and powder diffraction can help test crystals prior to SMX beamtime and help ensure the sample is well optimized for the technique. However, a complete description of a method to make the transition from an initial vapour diffusion crystallization condition to a large-scale batch crystallization condition is still lacking.

This paper endeavours to shed light on how to perform this transition from nanolitre vapour diffusion crystallization to

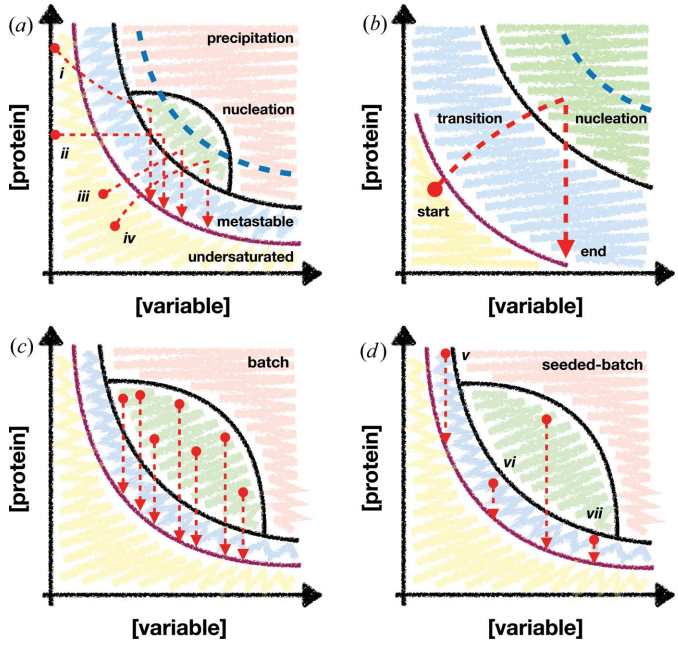


Figure 2
Examples of crystallization trajectories plotted onto phase diagrams. Protein concentration and a reservoir component 'variable' concentration are plotted on the *y* and *x* axes, respectively. The 'variable' could be any factor which may influence the crystallization experiment, *e.g.* PEG, salt or buffer concentration. The purple lines show the boundary of protein supersaturation [adapted from Chayen *et al.* (1992)]. The red circles and arrows denote the starting and finishing points of a crystallization experiment. The regions of the diagram are labelled in panel (*a*): precipitation, nucleation, metastable and undersaturated, and these are highlighted in pink, green, blue and yellow, respectively. The blue dotted lines show the theoretical limit of nucleation-zone penetration for non-batch methods. Potential crystallization trajectories for the transitionary phase methods of free-interface diffusion (i), dialysis (ii), evaporation (iii) and vapour diffusion (iv) are highlighted. (*b*) Highlighting the trajectory of a vapour diffusion experiment. The components of the drop must transition from outside to inside the nucleation zone through some process. (*c*), (*d*) More diverse examples of batch and seeded-batch experiments, respectively. Batch experiments [panel (*c*)] are not bound by the nucleation-zone limit and can, therefore, theoretically reach every part of the region. The trajectories v, vi and vii in panel (*d*) show potential trajectories for growing large single crystals, micro-crystals and micro-crystals from a less-concentrated sample, respectively.

[1] It is potentially possible to cross the nucleation limit line using a 'controlled evaporation' approach (Lee *et al.*, 2018) that rapidly shifts the crystallization drop deeper into the nucleation zone, and can thereby increase the nucleation rate.

large-scale batch crystallization. This task is split into three stages: (i) optimizing crystals grown using vapour diffusion methods towards conditions appropriate for batch crystallization by finding the nucleation zone, (ii) identifying promising batch crystallization strategies by plotting a phase diagram and, finally, (iii) demonstrating a practical approach to scaling batch conditions to create the large volumes (>100 µl) of micro-crystalline slurries often needed for SMX experiments. Frequently observed problems during scaling and other crystallization tips are presented in the supporting information.

2. Methods

2.1. PDB analysis

2.1.1. Data gathering. The PDB analysis was conducted using data gathered on 24 July 2019. Experimental crystallization conditions were extracted from the PDB archive online. Of the 134 321 PDB entries based on crystal diffraction (X-ray, electron and neutron), 110 858 included information about how the protein was crystallized. Manual inspection of the method types led to the division of these methods into 18 broad types: vapour diffusion (sitting and hanging drop), batch, evaporation, LCP, diffusion, dialysis, counter-diffusion, *in vivo*, temperature change, FID, spontaneous growth, dilution, concentration, connected bilayer, lyophilization, centrifugal crystallization and gel acupuncture. In the few cases where the method was completely ambiguous, the crystallization method was taken from the associated publication.

2.1.2. SMX analysis. A list of PDB IDs was created by selecting SMX indicators from information contained within the PDB header. These indicators were (i) the number of reported crystals used in the experiment (>10 was used as an arbitrary indication of a serial experiment), (ii) the radiation source, *e.g.* SACLA or FREE ELECTRON LASER, and (iii) the indexing software used, *e.g. CrystFEL* (White *et al.*, 2016) or *cctbx.xfel* (Brewster *et al.*, 2018). Any PDB entry which fulfilled one or more of these conditions was considered an SMX experiment. These criteria gave a data set of 409 PDB IDs, consisting of 248 and 161 from XFEL and synchrotron light sources, respectively.

2.1.3. Precipitant equilibration time analysis. Precipitant concentration data were extracted from PDB experimental crystallization conditions for the precipitants polyethylene glycol (PEG) 8000, PEG 1000, PEG 400, 2-methyl-2,4-pentanediol (MPD), NaCl and $(NH_4)_2SO_4$, comprising 5259, 1421, 10 013, 3087, 9049 and 5020 data points, respectively. Concentrations of <5% *w*/*v* or *v*/*v* and <0.5 *M* were considered likely to be only additives rather than primarily precipitants and were, therefore, excluded from the analysis. To estimate the equilibration times (90% of initial reservoir concentration at 293 K) for the different precipitant concentrations, single-phase exponential decay curves (*Prism 8*; GraphPad Software, San Diego, California, USA) were fitted to the data presented by Forsythe *et al.* (2002). Equilibration times for different precipitants were then extrapolated from the decay curves.

2.2. Protein preparation

2.2.1. UbiX. UbiX protein was produced as previously described (White *et al.*, 2015). Briefly, BL21 (DE3) *Escherichia coli* cells (NEB) transformed with pNic28-Bsa4 containing *Pseudomonas aeruginosa* UbiX, codon-optimized for *E. coli*, were grown at 310 K in 22 l of Terrific Broth in a fermenter with constant aeration. The cells were induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) at $OD_{600} \simeq 0.8$, at which point the temperature was reduced to 291 K for 18 h. Cells were harvested by centrifugation at 6000*g* for 10 min. A mass of 200 g of cells was resuspended in 50 m*M* Tris pH 8.0, 0.5 *M* NaCl, supplemented with 0.1 mg ml^{-1} DNase, 0.1 mg ml^{-1} RNase and cOmplete protease inhibitor (Sigma–Aldrich), before homogenization by French Press at 20 kpsi (1 psi $\simeq$ 6893 Pa). The resultant lysate was clarified by ultracentrifugation at 125 000*g* for 1 h before being loaded onto 50 ml of Ni-NTA agarose (Qiagen) in a gravity flow column. The resin was washed 2 × 4 times with lysate buffer containing 10 m*M* imidazole and then 40 m*M* imidazole. Bound UbiX was then eluted from the resin using 50 m*M* Tris pH 8.0, 0.5 *M* NaCl, 0.25 *M* imidazole, before desalting into 20 m*M* Tris pH 8.0, 0.2 *M* NaCl on P-6DG resin (BioRad).

2.2.2. FutA. The FutA gene from *Prochlorococcus MED4* was inserted into a pET-24*b*(+) vector, transformed into *E. coli* BL21 (DE3) cells (NEB) and grown at 310 K in 1 l of lysogeny broth. At $OD_{600} \simeq 0.4$ the temperature was reduced to 291 K, and then at $OD_{600} \simeq 0.6$ cells were induced with IPTG and incubated for 18 h. Cells were harvested by two rounds of centrifugation at 5000*g*.

A mass of 2–4 g of cells was resuspended in IBB buffer (0.1 *M* Tris, 0.5 *M* NaCl, 1% Triton-X, 5 m*M* $MgCl_2$, 10 m*M* β-mercaptoethanol). Cells were lysed by incubation with 50 mg of lysozyme and sonication, and then the inclusion bodies were washed by three cycles of 20 ml IBB buffer and centrifugation (40 min at 125 000*g* and 277 K). The inclusion bodies were dissolved in 20 ml of 0.2 *M* Tris pH 9.0, 6 *M* urea and 10 m*M* β-mercaptoethanol, incubated for 1 h at 277 K, and harvested by centrifugation for 40 min at 125 000*g* and 277 K.

FutA was refolded by rapidly diluting the supernatant into 2 l of 0.2 *M* Tris pH 9.0, 0.2 *M* NaCl, 0.4 *M* L-Arginine, 0.1 m*M* $NH_4Fe(SO_4)_2$ and incubating at 277 K for 48 h. The refold solution was concentrated to 150 ml using an Amicon stirred cell (Merck) and dialysed overnight in 2 l of 100 m*M* Tris pH 9.0, 50 m*M* NaCl. The dialysed solution was loaded onto a 5 ml HiTrap SP XL column (GE Healthcare) equilibrated in 0.1 *M* Tris pH 9.0, 50 m*M* NaCl. The protein was eluted by the addition of 0.1 *M* Tris pH 9.0, 1 *M* NaCl and the resulting fractions containing FutA were concentrated to 80 mg ml^{-1}.

2.3. Protein crystallization

2.3.1. UbiX. Initial crystallization trials of UbiX used 96-well three-drop SWISSCI plates, with protein at 30, 20 and 10 mg ml^{-1} supplemented with 0.2 m*M* flavin mononucleotide (FMN). UbiX was mixed in a 1:1 ratio with precipitant, in 600 nl drops. Crystals were grown at 294 K. Multiple

conditions were found to produce cubic crystals from sparse-matrix screening of UbiX; of these, 0.1 *M* MES pH 6.5, 1.6 *M* ammonium sulfate was chosen for optimization.

A phase diagram was made over two 96-well three-drop SWISSCI plates, varying the ammonium sulfate concentration on the horizontal axis from 0.1 to 3.0 *M* with constant 0.1 *M* MES pH 6.5. The UbiX concentration was varied along the vertical axis and split over the two plates, starting from 5 mg ml^{-1} and increasing to 80 mg ml^{-1} in 5 mg ml^{-1} increments. Each concentration of UbiX was supplemented with 0.2 m*M* FMN prior to crystallization. Two 300 nl drops per well were set up, one drop containing a 1:1 protein-to-precipitant ratio and the other containing a 3:2:1 ratio of protein to precipitant to seeds. The seed stock was made from the initial condition identified in the sparse-matrix screen; crystals from five drops were added to 50 µl of reservoir solution and crushed using a Hampton Seed Bead, with 90 s of vortexing.

2.3.2. FutA. To grow seed crystals of FutA, 52 mg ml^{-1} FutA solution was crystallized in 24-well XRL plates (Molecular Dimensions) containing 0.2 *M* NaSCN and varying concentrations of PEG 3350 from 10 to 20%(*w*/*v*). FutA and precipitant were mixed in a 1:1 ratio in 1 µl drops and the plate incubated at 294 K. FutA seed stocks were made by pooling ten 1 µl drops, adding 40 µl of 20% PEG 3350 and vortexing the solution with a Hampton Seed Bead for 180 s. A phase diagram was created as described in Section 2.4. The FutA and precipitant concentrations were varied between 18.75 and 80.00 mg ml^{-1} in eight steps, and between 5 and 40%(*w*/*v*) in 12 steps, respectively, with a constant concentration of 0.2 *M* NaSCN applied to all reservoir solutions.

For batch crystallization, FutA (52 mg ml^{-1}), FutA seed stock and crystallization buffer were mixed in a 1:1.5:1.5 ratio. Crystallization buffer [38%(*w*/*v*) PEG 3350, 0.25 *M* Tris pH 7.1] was mixed with FutA solution and vortexed for 3 s. FutA seeds, diluted 1:100 in 20%(*w*/*v*) PEG 3350, were added to the crystallization solution, which was then vortexed for 10 s. This mixture was incubated at 294 K for approximately 1–2 h and the micro-crystals were used fresh for any subsequent experiments.

2.4. Phase diagram crystallization experiments

With the exception of UbiX, all phase diagrams were generated from Greiner 96-well CrystalQuick X plates by varying the protein and precipitant concentrations over the vertical and horizontal axes, respectively. Each well contained 30 µl of the reservoir solution. Two drops of 300 nl were set up within each well, one drop containing only protein and precipitant (1:1 ratio) and the other containing protein, precipitant and seeds in a 3:2:1 ratio. The plates were incubated at 293 K in a ROCK IMAGER (Formulatrix) and imaged every 3 h for 24 h.

3. Transitioning from vapour diffusion to batch

Modern serial crystallography projects focus predominantly on proteins where a crystal structure of the protein of interest is already known [though there are notable exceptions, such as Sawaya *et al.* (2014) and Colletier *et al.* (2016)]. Therefore, the vast majority of SMX projects are likely to evolve from work in which crystals can already be grown and most probably in vapour diffusion plates. This paper will focus on the process of transitioning from a small-scale (<0.2–2.0 µl) vapour diffusion experiment to a large-scale (≥100 µl) batch protocol. Techniques such as second-order nonlinear imaging of chiral crystals (SONICC) (Luft *et al.*, 2015; Lee *et al.*, 2018) and dynamic light scattering (Abdallah *et al.*, 2015), although extremely useful in identifying conditions with micro-crystals, are not yet in the standard crystallographers' toolbox and have, therefore, been avoided here. The tools that are described herein were chosen for either their widespread adoption or their relatively low cost, in the hope that the methods proposed are translatable to the majority of crystallization laboratories.

3.1. Identifying a batch-like crystallization process in a vapour diffusion crystallization condition

The equilibration time of a sitting-drop experiment is dependent upon the composition of both the drop and reservoir volumes and on the volume of air in the well (Luft *et al.*, 1996; Forsythe *et al.*, 2002; Martins *et al.*, 2008). An understanding of the effect that drop components have on the drop equilibration time and knowledge of when crystals appear give an insight into the major crystallization 'force', *i.e.* the process that is driving crystallization, within the drop. Does protein crystallization require the equilibration of the drop components to find the nucleation zone (vapour diffusion), or is the nucleation zone found simply by mixing the drop components, with crystallization beginning immediately (batch)?

Fig. 3(*a*) shows the principal precipitant concentrations for all vapour diffusion experiments that were reported and could be extracted from PDB entries (for example, https://www.rcsb.org/pdb/explore/materialsAndMethods.do?structureId=100d) using either PEG (400, 1000 or 8000) and/or salt-based [NaCl and $(NH_4)_2SO_4$] precipitants. Calculated equilibration times [extrapolated from principal precipitant concentrations using values calculated by Forsythe *et al.* (2002)] are shown in Fig. 3(*b*). Although these equilibration times are based upon mono-component solutions where equilibration has been shown to be longer than in more complex mixtures (Luft & DeTitta, 1995), the broad trend is still applicable. The fact that longer equilibration times are observed for PEG precipitants means that, if crystals appear rapidly (within the first 12–24 h of a vapour diffusion experiment), then although the drop equilibrium will already be shifting, the crystallization 'force' is still more likely to be 'batch like' than pure vapour diffusion. A batch-like process may also be true for rapidly appearing crystals under salt-based conditions; however, if crystals appear after 4–5 days, the drop equilibration is probably complete, meaning that, again, the crystallization force is more likely to be batch like.

Knowledge of how crystallization time and drop equilibration intersect has two implications. Firstly, by limiting (in the

case of PEG precipitants) or lengthening (generally, in the case of salt precipitants) the time horizon of a vapour diffusion experiment, vapour diffusion crystallization conditions can be screened for batch-like conditions. Secondly, and very practically, the hunt for batch-like conditions can be done in small-volume (200 nl) 96-well sitting-drop plates, which are already widely used and integrated into most crystallization facilities.

At this point, it is also worth mentioning microbatch methods (Chayen *et al.*, 1990, 1992), which were initially designed to make batch crystallization more compatible with robotic methods. This paper focuses on using vapour diffusion tools to make the conversion into batch as these are generally more widely used than microbatch, but the conversion could also be made using microbatch techniques instead (Chayen, 1998). However, successfully growing crystals in microbatch plates is not necessarily a marker of a batch-like condition, *i.e.* hitting the nucleation zone immediately upon mixing protein and precipitant. This is because evaporation occurs through the oil covering the microbatch drop, changing the concentration of crystallization solution components (Chayen, 1998). Indeed, this evaporation can even be exploited to aid crystallization by tailoring the mixture of mineral oils used to cover the crystallization drop to allow more evaporation (D'Arcy *et al.*, 2003). Ultimately, this evaporation process means that crystals grown in a microbatch experiment may suffer the same transitionary phase problems as described for vapour diffusion crystallization, making it difficult to pinpoint the nucleation zone and the exact concentration of components in the condition required for crystal nucleation. Nevertheless, crystallization time in microbatch, like in vapour diffusion, could very likely act as a guide to help identify the nucleation zone, but it might add a step in the process of transitioning to true batch crystallization.

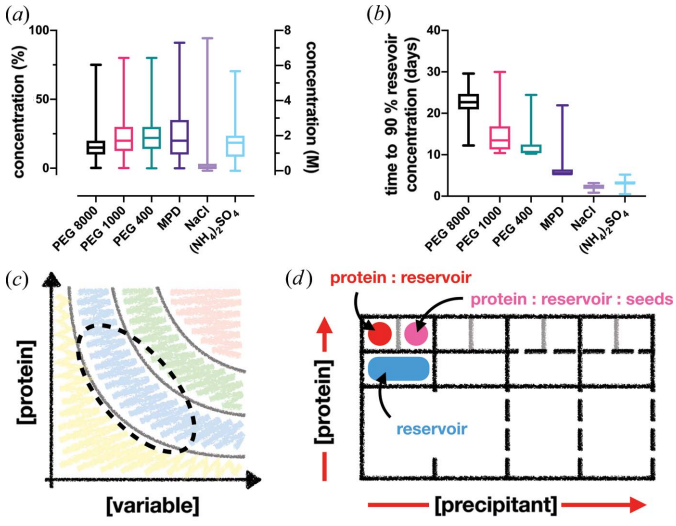


Figure 3
Manipulating vapour diffusion crystallization conditions into batch. (*a*), (*b*) Box-and-whisker plots of the submitted PDB precipitant concentrations from vapour diffusion crystallization experiments and their extrapolated equilibration times (time to 90% reservoir concentration), respectively. The diffusion times were calculated from data given by Forsythe *et al.* (2002). (*c*) The archetypal phase diagram, showing the likely area where the majority of vapour diffusion crystallization experiments begin (dotted line). (*d*) A design of a crystallization experiment in a two-drop 96-well sitting-drop plate to determine the phase diagram of the protein–precipitant mixture. One drop contains only protein and reservoir solution and the other contains protein, reservoir and seed solution, allowing the plotting of the nucleation and metastable zones, respectively.

3.2. Optimizing for batch crystallization

Upon examination of the crystallization time, if the protein of interest already crystallizes in a batch-like process, the nucleation and metastable regions of the condition can be explored (see Section 4). If the crystallization condition is not already batch like, the crystallization time can act as a rough guide as to how far a given condition is from the nucleation region. Therefore, by varying drop component concentrations and using either a shorter (PEG-based conditions) or a longer (salt-based conditions) crystallization time as the optimization metric, a batch-like condition can be discovered.

In theory, a true vapour diffusion experiment could start anywhere in the phase diagram. However, given the PEG and protein concentrations typically used in sparse-matrix screening, the most likely starting region is as highlighted in Fig. 3(*c*). A simple test to assess whether a vapour diffusion condition begins in the metastable region is to add seeds to the crystallization experiment. The addition of seeds to a supersaturated protein solution should produce crystals rapidly and can therefore act as a further guide in optimization. Some other potential paths are listed here and an example of the steps taken to move from vapour diffusion to a batch-like process is shown in Appendix *A*2 in the supporting information.

(i) *Multivariate experimental design.* Essentially, instead of limiting crystallization optimization to a two-dimensional approach, it is better to explore a wider region of 'crystallization space' by varying all components of the crystallization drop simultaneously [for a full description see Shaw Stewart & Mueller-Dieckmann (2014)]. The *XSTEP* package, from Douglas Instruments, is available to do this.

(ii) *Changing the ratio of protein to reservoir volume in the drop.* Most crystallization screening starts at a 1:1 protein-to-reservoir volume ratio. However, changing this will shift the starting point on the phase diagram diagonally, exploring different areas of the diagram.

(iii) *Sparse-matrix micro-seeding.* If the current condition is not yielding anything positive, the researcher can look for new crystallization conditions using seeds as random nucleants (Ireton & Stoddard, 2004; D'Arcy *et al.*, 2007). This method can identify novel reservoir conditions which may have a more batch-like propensity.

4. Exploring the metastable and nucleation regions

Once a batch condition has been discovered, a point in the nucleation zone has also been discovered. This condition can then be used as an anchoring point to discover the size and shape of the nucleation and metastable regions of the phase diagram. Knowledge of these regions is of great utility when

attempting to scale to larger volumes, since parameters such as protein concentration, crystal size and nucleation rate can be factored into the scaling arithmetic, ultimately leading to better outcomes.

4.1. Designing a phase diagram experiment

Once the parameters of a batch-like experiment have been identified, it becomes straightforward to generate a phase diagram. This can be done by taking the precipitant and protein and varying their concentration to form the *x* and *y* axes of the plot. A two-drop-per-well experiment can be particularly effective [Fig. 3(*d*)]. The first drop should comprise the protein and reservoir mixture, while the second should contain a mix of protein, reservoir and seeds; a 3:2:1 ratio is a good place to start (Ireton & Stoddard, 2004) (see Section 2.4). The results from the first drop will effectively plot the nucleation region, as only protein and precipitant concentrations that hit the nucleation zone will give rise to crystals and be observed. In the second drop, drops in the nucleation and metastable region should both yield crystals, as the seeds will act as nucleants and allow crystal growth. A comparison between the two drops should allow all four regions of the phase diagram to be determined.

4.2. Phase diagram examples

FutA, a periplasmic iron-binding protein associated with an Fe^{3+} uptake ABC transporter from *Prochlorococcus MED4* (Polyviou *et al.*, 2018), and UbiX, a flavin prenyltransferase

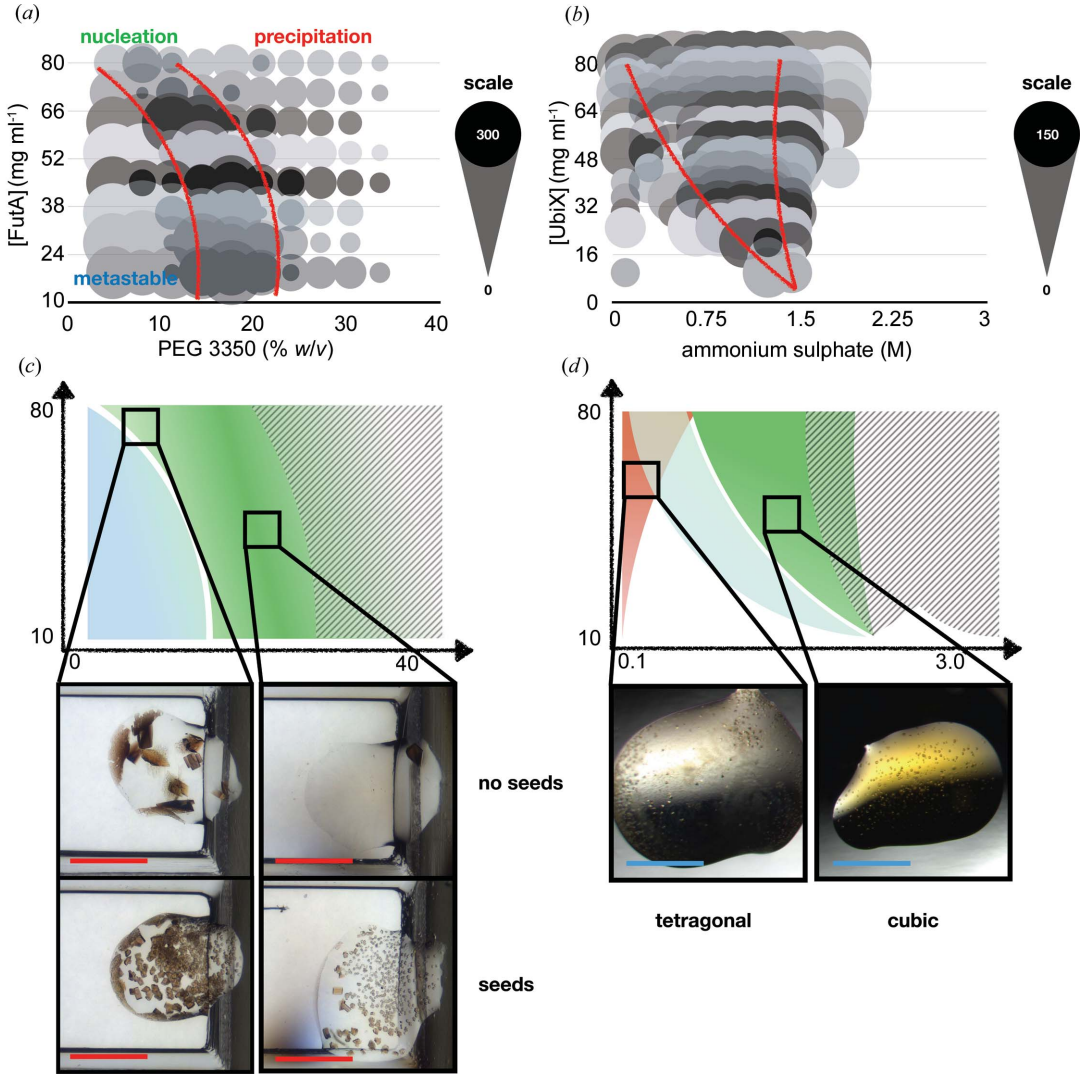


Figure 4
Phase diagrams for FutA and UbiX. The raw plots for *Prochlorococcus MED4* FutA and *P. aeruginosa* UbiX are shown in panels (*a*) and (*b*), respectively. The plots are based on two vapour diffusion crystallization experiments, with and without protein crystal seeds (see Section 4.1). The size of each circle corresponds to the approximate number of crystals observed in the crystallization drop. The opaque and shadowed circles show the number of crystals present from drops with no seeds and seeds, respectively. The red lines refer to the approximate boundaries between the different zones of the diagram. (*c*), (*d*) Representations of the plots shown in panels (*a*) and (*b*), respectively: darker shading indicates regions of higher nucleation, grey hatching shows drops where precipitation was visible, and the pink shading in the UbiX plot [panel (*d*)] highlights the region where a tetragonal crystal form appears. The crystallization drop images in panel (*c*) show the different levels of nucleation observed in both the seeded and un-seeded conditions. The images in panel (*d*) show the two different crystal forms of UbiX. The red and blue scale bars in the images denote 600 and 300 µm, respectively.

from *P. aeruginosa* involved in ubiquinone biosynthesis (White *et al.*, 2015), make interesting case studies of experimentally determined phase diagrams (two further phase diagrams are presented in Appendix *A*3 in the supporting information). The FutA phase diagram [Figs. 4(*a*) and 4(*c*)], when crystallized in 0.2 *M* NaSCN and PEG 3350, is striking, because the nucleation zone does not have the expected bow shape, illustrating the importance of experimental determination of the phase boundaries. The nucleation rate was somewhat proportional to both protein and precipitant concentrations. However, protein precipitation was observed when the precipitant was further increased. The basal nucleation rate was relatively low, so a seeded-batch protocol was developed (see Section 5.1).

UbiX, when crystallized in ammonium sulfate, produced two different crystal forms as confirmed by X-ray diffraction: cubic and tetragonal (data are not shown). The tetragonal form was associated with poorer quality (lower resolution) diffraction, so the cubic form was the goal of the crystallization experiment. Fortunately, the phase diagram showed that the tetragonal and cubic crystal forms were created from relatively distinct regions of the phase diagram [Figs. 4(*b*) and 4(*d*)]. Tetragonal crystals only appeared at very low precipitant concentrations [pink shaded area in Fig. 4(*d*)], whereas the cubic form was favoured at higher precipitant concentrations. The barrier between protein precipitation and the nucleation region was relatively clearly defined: drops contained either crystals or precipitation, with both rarely occurring together. Like FutA, the nucleation rate could be influenced by precipitant concentration, but not greatly, again suggesting that perhaps a seeded-batch protocol would be more appropriate. A description of the scaling of UbiX batch crystallization to larger volumes is given in Appendix *A*4.

5. Scaling batch conditions to larger volumes

Once an appropriate condition or conditions have been identified, the next task is to attempt to scale these batch or seeded-batch conditions, aiming for an eventual final volume of >50 µl but really as large as is feasible and appropriate. Scaling can be a daunting and frustrating prospect and not without reason. Protein volumes and therefore sample consumption will increase greatly. This paper cannot present any hard and fast rules, only a collection of ideas and suggestions. Like a cliff diver, at some point you have to take the plunge.

5.1. Optimizing crystal size and concentration

Creating a protocol where the final size of the micro-crystals can be systematically changed is a huge advantage (Dods *et al.*, 2017). Crystal size can be optimized to the sample-delivery approach and other experimental factors, such as the required diffusion time for a ligand or the light penetration depth. Crystal concentration (crystals per millilitre) will ultimately be determined by the nucleation rate and is inversely proportional to crystal size. That is to say, the greater the level of nucleation, the greater the number of crystals that must grow from the finite amount of protein in the batch condition, so the smaller the crystals will be. However, whereas crystal concentration can be manipulated by the removal or addition of buffer after completion of the crystallization experiment, size homogeneity has to be tailored at the crystallization step. Therefore, although crystal concentration is an important consideration due to its relationship to crystal size, ultimately crystal size and size homogeneity should be the key heuristics in the scaling process as these cannot be changed (that said, see Table 2 in the supporting information for some limited advice concerning crystal crushing).

A hemocytometer [a small particle counter – Fig. 10(*e*) in the supporting information] allows the experimenter to assess a representative sample of the micro-crystals from a given crystallization experiment, allowing both their size range and the concentration to be estimated. Fig. 5 shows how this can be performed using FutA as an exemplar.

The process is as follows. During a large-scale (>20 µl) batch experiment, take regular aliquots (2.5–5.0 µl) of the crystallization experiment and view in a hemocytometer [Fig. 5(*a*)]. Ensure the batch crystallization experiment is homogeneous before taking an aliquot, and make a note of the number of crystals and their size distribution [Figs. 5(*b*) and 5(*c*)]. These data can then be used to compare different batch conditions and iterate towards an ideal protocol for a given sample-delivery approach, *e.g.* probing alterations in precipitant and/or protein concentrations or optimizing the ratios of components in the crystallization solution. It should also be noted that it is theoretically possible that the taking of these aliquots could hinder protein crystallization. However, if such effects from collecting these aliquots do occur, they have yet to be observed.

The power of this technique is shown in the case of FutA. From the initial phase diagram, 52 mg ml^{-1} of FutA solution, mixed in a 1:1 ratio with 0.2 *M* NaSCN, 12.5%(*w*/*v*) PEG 3350, was selected as a starting point for a seeded-batch experiment. However, as can be seen from Fig. 5 this was not ideal as the crystals were not sufficiently homogeneous in size. Although the eventual crystal concentration and size were acceptable [Figs. 5(*a*), top panel, 5(*b*) and 5(*c*)] for an SMX experiment (data are not shown), many large crystals (>40 µm) were formed early (1–2 h) in the experiment. It was only after 3 h that showers of micro-crystals were observed. This delayed start created an asymmetric size distribution [Fig. 5(*d*)], with two crystal-size populations being observed. Altering the PEG concentration did not appear to improve the homogeneity in the crystal size, but the addition of a neutral buffer did. This change was prompted by the wish to improve the durability of FutA crystals during ligand-soaking experiments. The NaSCN was exchanged for 0.1 *M* Tris pH 7.1 in the crystallization buffer because the FutA crystals dissolved in the presence of ligand and NaSCN. The exchange improved the crystal stability and also reduced the tendency for the crystals to clump together. In the presence of Tris, the propensity of the FutA to precipitate at higher PEG concentrations was also reduced. The PEG concentration could then be increased

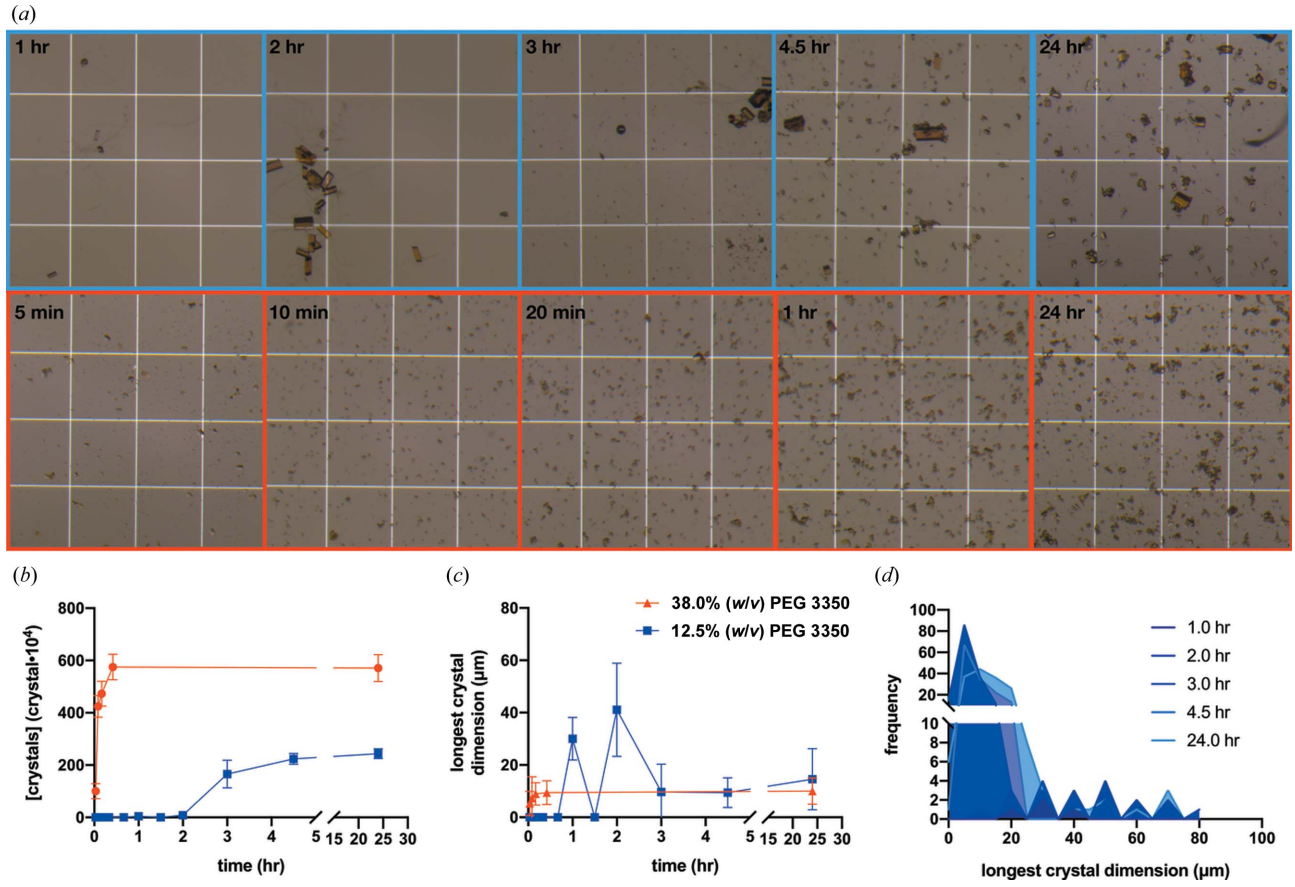


Figure 5
Observing a 100 µl FutA batch crystallization over 24 h. (*a*) The growth of two FutA batch crystallization experiments, the top (blue) in 0.2 *M* NaSCN, 12.5%(*w*/*v*) PEG 3350 and the bottom (red) in 0.1 *M* Tris pH 7.1, 38.0%(*w*/*v*) PEG 3350. The pictures show aliquots viewed in a hemocytometer. The white boxes in the images have dimensions of 250 × 250 µm. (*b*), (*c*) Demonstrations of how the mean number of crystals and longest dimension change over time. (*d*) A histogram of crystal size over 24 h for the 12.5%(*w*/*v*) PEG 3350 condition.

from 12.5 to 38.0%(*w*/*v*) PEG 3350. These changes reduced the size and increased the concentration of the FutA crystals obtained from the seeded-batch crystallization [Figs. 5(*a*), bottom panel, 5(*b*) and 5(*c*)].

5.2. Scaling up in volume

The proposed sample-delivery mode in the SMX experiment can also dictate the final volume of the batch crystallization experiment. Some ADE and extruder delivery systems require only 20 µl of sample per load. Therefore, a final experimental volume of 100 µl, assuming that a 'reasonable' crystal concentration can be achieved, should be perfectly adequate for these delivery approaches. If larger volumes are required, pooling of multiple 100 µl experiments is also possible. This being the case, a step-wise volume increase from 200 nl to an approximate final volume of 100 µl could prove safest. If larger volumes of sample are required, multiple batches of 100 µl can be set up concurrently and pooled together. However, if a step-wise scale in crystallization volume has proved successful, larger volumes of 1 ml or more could also be attempted if applicable, feasible and necessary. An example of such a scaling protocol is described below. At each step, the user should assess the number of crystals and range of sizes. If these change, slight alterations should be attempted in component concentrations and/or ratios.

(i) *Increase the volume in robot-compatible plates.* Liquid-handling robots for 96-well experiments, such as the Mosquito (TTP Labtech), can aspirate volumes of up to 1.2 µl, giving an effective limit of 2.4 µl on the drop size, assuming a 1:1 protein-to-reservoir ratio. This drop size can be accommodated in some 96-well sitting-drop plates, such as the Greiner CrystalQuick [Fig. 10(*a*) in the supporting information] or the SWISSCI MRC 48-well plates. An under-oil experiment at these volumes could also be attempted, perhaps using SWISSCI under-oil or Terizaki plates [Figs. 10(*b*) and 10(*c*), respectively], the former having a maximum volume of 4 µl. The advantage of using such plates is that most are still compatible with commercially available crystallization robots and storage hotels, thus simplifying standardization and monitoring.

(ii) *Increase the drop volume to 10–20 µl.* This entails moving from robot-compatible plates into either 24-well hanging- or sitting-drop plates, PCR tubes or 0.5 ml centrifuge tubes. The crystallization experiment should be monitored in the drop or tube over 1–7 days, taking note of the crystal number and size.

(iii) *Increase the drop volume to 20–100* µl. This is achieved by moving into 0.5 ml centrifuge tubes or 96-well chimney-well plates [Fig. 10(*d*)]. Aliquots are taken every 3–4 h to measure the crystal number and size using a hemocytometer [Fig. 10(*e*), and described in Section 5.1]. Gentle or even vigorous agitation may now be required, depending on the current vessel; potential mixers are shown in Figs. 10(*f*), 10(*g*) and 10(*h*).

(iv) *Increase the drop volume to 0.5–1.0 ml (if required)*. If all the preceding steps are consistent, the user could try to move to 1.5 ml centrifuge tubes.

(v) *Increase the volume to 5–10 ml (if required)*. The user should only attempt this if the protein can be easily produced and the delivery approach requires large (>1 ml) volumes.

5.3. Other tips and ideas

Table 1 in the supporting information shows some recurrent problems that have been encountered when scaling several different proteins to large-volume batch crystallization. Some potential solutions to these problems are suggested in the table; these are by no means perfect or exhaustive but might be helpful. Other crystallization tips are listed in Table 2 in the supporting information.

6. Conclusions

The aim of this paper was to suggest methods and ideas to aid in converting a vapour diffusion crystallization experiment into a larger-scale batch experiment. Given what can seem like the somewhat arbitrary whims of protein crystallization, the creation and subsequent understanding of a crystallization phase diagram is perhaps the surest way to approach these tasks. Vapour diffusion crystallization experiments can be converted into batch crystallization by understanding the role the precipitant is playing in the crystallization process and looking at the timescale of crystal nucleation and growth. Optimizing a vapour diffusion experiment in this manner allows the nucleation zone to be found, and hence the conditions for batch crystallization. Once a batch condition has been found, a phase diagram can be created. From the information in the phase diagram, batch or seeded-batch protocols can be gradually scaled to test the condition in larger volumes. This approach may ease the burden on the required protein volume and make the process of transitioning to batch crystallization more efficient. Ultimately, protein crystallization is fickle and should be assumed to fail randomly. Given this capricious tendency, the more time spent understanding the crystallization process, the greater the chance that good quality crystals will be obtained when they are required on a beamline.

7. Related literature

The following additional literature is cited in the supporting information: Bergfors (2003), Berrow *et al.* (2007), Bunker *et al.* (2012), Chayen *et al.* (2006), de la Cruz *et al.* (2017), Ino *et al.* (2011), Luft & DeTitta (1999), McPherson & Shlichta (1988), Nanev *et al.* (2017), Stura & Wilson (1992), Zhang *et al.* (1996).

Acknowledgements

This work would not have been possible without numerous informal conversations with open crystallographers who were happy to share their experiences with batch crystallization and SMX methods in general, in particular Danny Axford, Agata Butryn, Pierre Aller and Halina Mikolajek. We also thank the Research Complex at Harwell for the use of their crystallization facility and laboratory space. Louise Bird, formerly of the Protein Production UK, kindly provided the eGFP-pOPINF plasmid.

Funding information

Research regarding UbiX was funded by the BBSRC grant BB/P000622/1 to Professor David Leys.

References

Abdallah, B. G., Zatsepin, N. A., Roy-Chowdhury, S., Coe, J., Conrad, C. E., Dörner, K., Sierra, R. G., Stevenson, H. P., Camacho-Alanis, F., Grant, T. D., Nelson, G., James, D., Calero, G., Wachter, R. M., Spence, J. C. H., Weierstall, U., Fromme, P. & Ros, A. (2015). *Struct. Dyn.* **2**, 041719.

Bergfors, T. (2003). *J. Struct. Biol.* **142**, 66–76.

Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D. I. & Owens, R. J. (2007). *Nucleic Acids Res.* **35**, e45.

Botha, S., Nass, K., Barends, T. R. M., Kabsch, W., Latz, B., Dworkowski, F., Foucar, L., Panepucci, E., Wang, M., Shoeman, R. L., Schlichting, I. & Doak, R. B. (2015). *Acta Cryst.* D**71**, 387–397.

Brewster, A. S., Waterman, D. G., Parkhurst, J. M., Gildea, R. J., Young, I. D., O'Riordan, L. J., Yano, J., Winter, G., Evans, G. & Sauter, N. K. (2018). *Acta Cryst.* D**74**, 877–894.

Bunker, R. D., Dickson, J. M. J., Caradoc-Davies, T. T., Loomes, K. M. & Baker, E. N. (2012). *Acta Cryst.* F**68**, 1259–1262.

Chayen, N. E. (1998). *Acta Cryst.* D**54**, 8–15.

Chayen, N. E., Saridakis, E. & Sear, R. P. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 597–601.

Chayen, N. E., Shaw Stewart, P. D. & Blow, D. M. (1992). *J. Cryst. Growth*, **122**, 176–180.

Chayen, N. E., Shaw Stewart, P. D., Maeder, D. L. & Blow, D. M. (1990). *J. Appl. Cryst.* **23**, 297–302.

Colletier, J.-P., Sawaya, M. R., Gingery, M., Rodriguez, J. A., Cascio, D., Brewster, A. S., Michels-Clark, T., Hice, R. H., Coquelle, N., Boutet, S., Williams, G. J., Messerschmidt, M., DePonte, D. P., Sierra, R. G., Laksmono, H., Koglin, J. E., Hunter, M. S., Park, H. W., Uervirojnangkoorn, M., Bideshi, D. K., Brunger, A. T., Federici, B. A., Sauter, N. K. & Eisenberg, D. S. (2016). *Nature*, **539**, 43–47.

Cruz, M. J. de la, Hattne, J., Shi, D., Seidler, P., Rodriguez, J., Reyes, F. E., Sawaya, M. R., Cascio, D., Weiss, S. C., Kim, S. K., Hinck, C. S., Hinck, A. P., Calero, G., Eisenberg, D. & Gonen, T. (2017). *Nat. Methods*, **14**, 399–402.

D'Arcy, A., Mac Sweeney, A., Stihle, M. & Haber, A. (2003). *Acta Cryst.* D**59**, 396–399.

D'Arcy, A., Villard, F. & Marsh, M. (2007). *Acta Cryst.* D**63**, 550–554.

Darmanin, C., Strachan, J., Adda, C. G., Ve, T., Kobe, B. & Abbey, B. (2016). *Sci. Rep.* **6**, 25345.

Davy, B., Axford, D., Beale, J. H., Butryn, A., Docker, P., Ebrahim, A., Leen, G., Orville, A. M., Owen, R. L. & Aller, P. (2019). *J. Synchrotron Rad.* **26**, 1820–1825.
DePonte, D. P., Weierstall, U., Schmidt, K., Warner, J., Starodub, D., Spence, J. C. H. & Doak, R. B. (2008). *J. Phys. D Appl. Phys.* **41**, 195505.
Dods, R., Båth, P., Arnlund, D., Beyerlein, K. R., Nelson, G., Liang, M., Harimoorthy, R., Berntsen, P., Malmerberg, E., Johansson, L., Andersson, R., Bosman, R., Carbajo, S., Claesson, E., Conrad, C. E., Dahl, P., Hammarin, G., Hunter, M. S., Li, C., Lisova, S., Milathianaki, D., Robinson, J., Safari, C., Sharma, A., Williams, G., Wickstrand, C., Yefanov, O., Davidsson, J., DePonte, D. P., Barty, A., Brändén, G. & Neutze, R. (2017). *Structure*, **25**, 1461–1468.
Evans, G., Axford, D., Waterman, D. & Owen, R. L. (2011). *Crystallogr. Rev.* **17**, 105–142.
Feld, G. K., Heymann, M., Benner, W. H., Pardini, T., Tsai, C.-J., Boutet, S., Coleman, M. A., Hunter, M. S., Li, X., Messerschmidt, M., Opathalage, A., Pedrini, B., Williams, G. J., Krantz, B. A., Fraden, S., Hau-Riege, S., Evans, J. E., Segelke, B. W. & Frank, M. (2015). *J. Appl. Cryst.* **48**, 1072–1079.
Forsythe, E. L., Maxwell, D. L. & Pusey, M. (2002). *Acta Cryst.* D**58**, 1601–1605.
Frank, M., Carlson, D. B., Hunter, M. S., Williams, G. J., Messerschmidt, M., Zatsepin, N. A., Barty, A., Benner, W. H., Chu, K., Graf, A. T., Hau-Riege, S. P., Kirian, R. A., Padeste, C., Pardini, T., Pedrini, B., Segelke, B., Seibert, M. M., Spence, J. C. H., Tsai, C.-J., Lane, S. M., Li, X.-D., Schertler, G., Boutet, S., Coleman, M. & Evans, J. E. (2014). *IUCrJ*, **1**, 95–100.
Fry, E. E., Grimes, J. & Stuart, D. I. (1999). *Mol. Biotechnol.* **12**, 13–23.
Fuller, F. D., Gul, S., Chatterjee, R., Burgie, E. S., Young, I. D., Lebrette, H., Srinivas, V., Brewster, A. S., Michels-Clark, T., Clinger, J. A., Andi, B., Ibrahim, M., Pastor, E., de Lichtenberg, C., Hussein, R., Pollock, C. J., Zhang, M., Stan, C. A., Kroll, T., Fransson, T., Weninger, C., Kubin, M., Aller, P., Lassalle, L., Bräuer, P., Miller, M. D., Amin, M., Koroidov, S., Roessler, C. G., Allaire, M., Sierra, R. G., Docker, P. T., Glownia, J. M., Nelson, S., Koglin, J. E., Zhu, D., Chollet, M., Song, S., Lemke, H., Liang, M., Sokaras, D., Alonso-Mori, R., Zouni, A., Messinger, J., Bergmann, U., Boal, A. K., Bollinger, J. M., Krebs, C., Högbom, M., Phillips, G. N., Vierstra, R. D., Sauter, N. K., Orville, A. M., Kern, J., Yachandra, V. K. & Yano, J. (2017). *Nat. Methods*, **14**, 443–449.
Grünbein, M. L. & Nass Kovacs, G. (2019). *Acta Cryst.* D**75**, 178–191.
Hunter, M. S., Segelke, B., Messerschmidt, M., Williams, G. J., Zatsepin, N. A., Barty, A., Benner, W. H., Carlson, D. B., Coleman, M., Graf, A., Hau-Riege, S. P., Pardini, T., Seibert, M. M., Evans, J., Boutet, S. & Frank, M. (2015). *Sci. Rep.* **4**, 6026.
Ibrahim, M., Chatterjee, R., Hellmich, J., Tran, R., Bommer, M., Yachandra, V. K., Yano, J., Kern, J. & Zouni, A. (2015). *Struct. Dyn.* **2**, 041705.
Ino, K., Udagawa, I., Iwabata, K., Takakusagi, Y., Kubota, M., Kurosaka, K., Arai, K., Seki, Y., Nogawa, M., Tsunoda, T., Mizukami, F., Taguchi, H. & Sakaguchi, K. (2011). *PLoS One*, **6**, e22582.
Ireton, G. C. & Stoddard, B. L. (2004). *Acta Cryst.* D**60**, 601–605.
Ishikawa, T., Aoyagi, H., Asaka, T., Asano, Y., Azumi, N., Bizen, T., Ego, H., Fukami, K., Fukui, T., Furukawa, Y., Goto, S., Hanaki, H., Hara, T., Hasegawa, T., Hatsui, T., Higashiya, A., Hirono, T., Hosoda, N., Ishii, M., Inagaki, T., Inubushi, Y., Itoga, T., Joti, Y., Kago, M., Kameshima, T., Kimura, H., Kirihara, Y., Kiyomichi, A., Kobayashi, T., Kondo, C., Kudo, T., Maesaka, H., Maréchal, X. M., Masuda, T., Matsubara, S., Matsumoto, T., Matsushita, T., Matsui, S., Nagasono, M., Nariyama, N., Ohashi, H., Ohata, T., Ohshima, T., Ono, S., Otake, Y., Saji, C., Sakurai, T., Sato, T., Sawada, K., Seike, T., Shirasawa, K., Sugimoto, T., Suzuki, S., Takahashi, S., Takebe, H., Takeshita, K., Tamasaku, K., Tanaka, H., Tanaka, R., Tanaka, T., Togashi, T., Togawa, K., Tokuhisa, A., Tomizawa, H., Tono, K., Wu, S., Yabashi, M., Yamaga, M., Yamashita, A., Yanagida, K., Zhang, C., Shintake, T., Kitamura, H. & Kumagai, N. (2012). *Nat. Photon.* **6**, 540–544.
Kupitz, C., Grotjohann, I., Conrad, C. E., Roy-Chowdhury, S., Fromme, R. & Fromme, P. (2014). *Philos. Trans. R. Soc. B*, **369**, 20130316.
Lee, D. B., Kim, J.-M., Seok, J. H., Lee, J.-H., Jo, J. D., Mun, J. Y., Conrad, C., Coe, J., Nelson, G., Hogue, B., White, T. A., Zatsepin, N., Weierstall, U., Barty, A., Chapman, H., Fromme, P., Spence, J., Chung, M. S., Oh, C.-H. & Kim, K. H. (2018). *Sci. Rep.* **8**, 2541.
Luft, J. R., Albright, D. T., Baird, J. K. & DeTitta, G. T. (1996). *Acta Cryst.* D**52**, 1098–1106.
Luft, J. R. & DeTitta, G. T. (1995). *Acta Cryst.* D**51**, 780–785.
Luft, J. R. & DeTitta, G. T. (1999). *Acta Cryst.* D**55**, 988–993.
Luft, J. R., Wolfley, J. R., Franks, E. C., Lauricella, A. M., Gualtieri, E. J., Snell, E. H., Xiao, R., Everett, J. K. & Montelione, G. T. (2015). *Struct. Dyn.* **2**, 041710.
Martin-Garcia, J. M., Conrad, C. E., Nelson, G., Stander, N., Zatsepin, N. A., Zook, J., Zhu, L., Geiger, J., Chun, E., Kissick, D., Hilgart, M. C., Ogata, C., Ishchenko, A., Nagaratnam, N., Roy-Chowdhury, S., Coe, J., Subramanian, G., Schaffer, A., James, D., Ketwala, G., Venugopalan, N., Xu, S., Corcoran, S., Ferguson, D., Weierstall, U., Spence, J. C. H., Cherezov, V., Fromme, P., Fischetti, R. F. & Liu, W. (2017). *IUCrJ*, **4**, 439–454.
Martins, P. M., Rocha, F. & Damas, A. M. (2008). *PLoS One*, **3**, e1998.
McPherson, A. (1982). *Preparation and Analysis of Protein Crystals*, 2nd ed. New York: R. E. Krieger Publishing Co.
McPherson, A. & Shlichta, P. (1988). *J. Cryst. Growth*, **90**, 47–50.
Milne, C., Schietinger, T., Aiba, M., Alarcon, A., Alex, J., Anghel, A., Arsov, V., Beard, C., Beaud, P., Bettoni, S., Bopp, M., Brands, H., Brönnimann, M., Brunnenkant, I., Calvi, M., Citterio, A., Craievich, P., Csatari Divall, M., Dällenbach, M., D'Amico, M., Dax, A., Deng, Y., Dietrich, A., Dinapoli, R., Divall, E., Dordevic, S., Ebner, S., Erny, C., Fitze, H., Flechsig, U., Follath, R., Frei, F., Gärtner, F., Ganter, R., Garvey, T., Geng, Z., Gorgisyan, I., Gough, C., Hauff, A., Hauri, C., Hiller, N., Humar, T., Hunziker, S., Ingold, G., Ischebeck, R., Janousch, M., Juranić, P., Jurcevic, M., Kaiser, M., Kalantari, B., Kalt, R., Keil, B., Kittel, C., Knopp, G., Koprek, W., Lemke, H., Lippuner, T., Llorente Sancho, D., Löhl, F., Lopez-Cuenca, C., Märki, F., Marcellini, F., Marinkovic, G., Martiel, I., Menzel, R., Mozzanica, A., Nass, K., Orlandi, G., Ozkan Loch, C., Panepucci, E., Paraliev, M., Patterson, B., Pedrini, B., Pedrozzi, M., Pollet, P., Pradervand, C., Prat, E., Radi, P., Raguin, J., Redford, S., Rehanek, J., Réhault, J., Reiche, S., Ringele, M., Rittmann, J., Rivkin, L., Romann, A., Ruat, M., Ruder, C., Sala, L., Schebacher, L., Schilcher, T., Schlott, V., Schmidt, T., Schmitt, B., Shi, X., Stadler, M., Stingelin, L., Sturzenegger, W., Szlachetko, J., Thattil, D., Treyer, D., Trisorio, A., Tron, W., Vetter, S., Vicario, C., Voulot, D., Wang, M., Zamofing, T., Zellweger, C., Zennaro, R., Zimoch, E., Abela, R., Patthey, L. & Braun, H. (2017). *Appl. Sci.* **7**, 720.
Murray, T. D., Lyubimov, A. Y., Ogata, C. M., Vo, H., Uervirojnangkoorn, M., Brunger, A. T. & Berger, J. M. (2015). *Acta Cryst.* D**71**, 1987–1997.
Nanev, C. N., Saridakis, E. & Chayen, N. E. (2017). *Sci. Rep.* **7**, 35821.
Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. (2000). *Nature*, **406**, 752–757.
Oberthuer, D., Knoška, J., Wiedorn, M. O., Beyerlein, K. R., Bushnell, D. A., Kovaleva, E. G., Heymann, M., Gumprecht, L., Kirian, R. A., Barty, A., Mariani, V., Tolstikova, A., Adriano, L., Awel, S., Barthelmess, M., Dörner, K., Xavier, P. L., Yefanov, O., James, D. R., Nelson, G., Wang, D., Calvey, G., Chen, Y., Schmidt, A., Szczepek, M., Frielingsdorf, S., Lenz, O., Snell, E., Robinson, P. J., Šarler, B., Belšak, G., Maček, M., Wilde, F., Aquila, A., Boutet, S., Liang, M., Hunter, M. S., Scheerer, P., Lipscomb, J. D., Weierstall, U., Kornberg, R. D., Spence, J. C. H., Pollack, L., Chapman, H. N. & Bajt, S. (2017). *Sci. Rep.* **7**, 44628.
Polyviou, D., Machelett, M. M., Hitchcock, A., Baylay, A. J., MacMillan, F., Moore, C. M., Bibby, T. S. & Tews, I. (2018). *J. Biol. Chem.* **293**, 18099–18109.

Reis-Kautt, M. & Ducruix, A. (1992). *Crystallization of Nucleic Acids and Proteins*, 1st ed., edited by A. Ducruix & R. Giegé, ch. 9, pp. 195–218. Oxford University Press.

Roedig, P., Ginn, H. M., Pakendorf, T., Sutton, G., Harlos, K., Walter, T. S., Meyer, J., Fischer, P., Duman, R., Vartiainen, I., Reime, B., Warmer, M., Brewster, A. S., Young, I. D., Michels-Clark, T., Sauter, N. K., Kotecha, A., Kelly, J., Rowlands, D. J., Sikorsky, M., Nelson, S., Damiani, D. S., Alonso-Mori, R., Ren, J., Fry, E. E., David, C., Stuart, D. I., Wagner, A. & Meents, A. (2017). *Nat. Methods*, **14**, 805–810.

Roessler, C. G., Agarwal, R., Allaire, M., Alonso-Mori, R., Andi, B., Bachega, J. F. R., Bommer, M., Brewster, A. S., Browne, M. C., Chatterjee, R., Cho, E., Cohen, A. E., Cowan, M., Datwani, S., Davidson, V. L., Defever, J., Eaton, B., Ellson, R., Feng, Y., Ghislain, L. P., Glownia, J. M., Han, G., Hattne, J., Hellmich, J., Héroux, A., Ibrahim, M., Kern, J., Kuczewski, A., Lemke, H. T., Liu, P., Majlof, L., McClintock, W. M., Myers, S., Nelsen, S., Olechno, J., Orville, A. M., Sauter, N. K., Soares, A. S., Soltis, S. M., Song, H., Stearns, R. G., Tran, R., Tsai, Y., Uervirojnangkoorn, M., Wilmot, C. M., Yachandra, V., Yano, J., Yukl, E. T., Zhu, D. & Zouni, A. (2016). *Structure*, **24**, 631–640.

Roessler, C. G., Kuczewski, A., Stearns, R., Ellson, R., Olechno, J., Orville, A. M., Allaire, M., Soares, A. S. & Héroux, A. (2013). *J. Synchrotron Rad.* **20**, 805–808.

Rupp, B. (2015). *Acta Cryst.* F**71**, 247–260.

Sawaya, M. R., Cascio, D., Gingery, M., Rodriguez, J., Goldschmidt, L., Colletier, J.-P., Messerschmidt, M. M., Boutet, S., Koglin, J. E., Williams, G. J., Brewster, A. S., Nass, K., Hattne, J., Botha, S., Doak, R. B., Shoeman, R. L., DePonte, D. P., Park, H.-W., Federici, B. A., Sauter, N. K., Schlichting, I. & Eisenberg, D. S. (2014). *Proc. Natl Acad. Sci. USA*, **111**, 12769–12774.

Shaw Stewart, P. & Mueller-Dieckmann, J. (2014). *Acta Cryst.* F**70**, 686–696.

Sherrell, D. A., Foster, A. J., Hudson, L., Nutter, B., O'Hea, J., Nelson, S., Paré-Labrosse, O., Oghbaey, S., Miller, R. J. D. & Owen, R. L. (2015). *J. Synchrotron Rad.* **22**, 1372–1378.

Sierra, R. G., Gati, C., Laksmono, H., Dao, E. H., Gul, S., Fuller, F., Kern, J., Chatterjee, R., Ibrahim, M., Brewster, A. S., Young, I. D., Michels-Clark, T., Aquila, A., Liang, M., Hunter, M. S., Koglin, J. E., Boutet, S., Junco, E. A., Hayes, B., Bogan, M. J., Hampton, C. Y., Puglisi, E. V., Sauter, N. K., Stan, C. A., Zouni, A., Yano, J., Yachandra, V. K., Soltis, S. M., Puglisi, J. D. & DeMirci, H. (2016). *Nat. Methods*, **13**, 59–62.

Stura, E. A. & Wilson, I. (1992). *Crystallization of Nucleic Acids and Proteins*, 1st ed., edited by A. Ducruix & R. Giegé, ch. 5, pp. 99–126. Oxford University Press.

Weierstall, U., James, D., Wang, C., White, T. A., Wang, D., Liu, W., Spence, J. C. H., Bruce Doak, R., Nelson, G., Fromme, P., Fromme, R., Grotjohann, I., Kupitz, C., Zatsepin, N. A., Liu, H., Basu, S., Wacker, D., Won Han, G., Katritch, V., Boutet, S., Messerschmidt, M., Williams, G. J., Koglin, J. E., Marvin Seibert, M., Klinker, M., Gati, C., Shoeman, R. L., Barty, A., Chapman, H. N., Kirian, R. A., Beyerlein, K. R., Stevens, R. C., Li, D., Shah, S. T. A., Howe, N., Caffrey, M. & Cherezov, V. (2014). *Nat. Commun.* **5**, 3309.

Weierstall, U., Spence, J. C. H. & Doak, R. B. (2012). *Rev. Sci. Instrum.* **83**, 035108.

Weinert, T., Olieric, N., Cheng, R., Brünle, S., James, D., Ozerov, D., Gashi, D., Vera, L., Marsh, M., Jaeger, K., Dworkowski, F., Panepucci, E., Basu, S., Skopintsev, P., Doré, A. S., Geng, T., Cooke, R. M., Liang, M., Prota, A. E., Panneels, V., Nogly, P., Ermler, U., Schertler, G., Hennig, M., Steinmetz, M. O., Wang, M. & Standfuss, J. (2017). *Nat. Commun.* **8**, 542.

White, M. D., Payne, K. A. P., Fisher, K., Marshall, S. A., Parker, D., Rattray, N. J. W., Trivedi, D. K., Goodacre, R., Rigby, S. E. J., Scrutton, N. S., Hay, S. & Leys, D. (2015). *Nature*, **522**, 502–506.

White, T. A., Mariani, V., Brehm, W., Yefanov, O., Barty, A., Beyerlein, K. R., Chervinskii, F., Galli, L., Gati, C., Nakane, T., Tolstikova, A., Yamashita, K., Yoon, C. H., Diederichs, K. & Chapman, H. N. (2016). *J. Appl. Cryst.* **49**, 680–689.

Zhang, G., Gurtu, V. & Kain, S. R. (1996). *Biochem. Biophys. Res. Commun.* **227**, 707–711.