

Molecular Imaging using Serial Femtosecond Crystallography

Submitted by
Susannah Holmes
BSc (Hons), BNano

A thesis submitted in total fulfilment
of the requirement for the degree of

Doctor of Philosophy

Department of Chemistry & Physics, School of Molecular Sciences,
La Trobe Institute for Molecular Science
College of Science, Health and Engineering

La Trobe University
Victoria, Australia

December 2021

Contents

Abstract	xiv
Statement of Authorship	xv
Acknowledgements	xvi
List of Publications	xviii
Abbreviations	xix
1 X-ray Crystallography	1
1.1 History of X-ray Crystallography	1
1.2 X-ray Diffraction	2
1.3 Crystal Packing & Symmetry	3
1.4 X-ray Data Collection	6
1.5 Structural analysis	7
1.5.1 Laue	7
1.5.2 Ewald construction	8
1.5.3 Partiality	10
1.5.4 Structure Factors	10
1.5.5 Electron Density	11
1.5.6 Temperature factor	12
1.5.7 Phase problem	12
1.6 Protein Crystallography	13

1.6.1	Crystallisation	13
1.6.2	Conventional protein crystallography	18
1.7	X-ray sources	18
1.7.1	Synchrotrons	18
1.7.2	XFELs	23
2	Serial Crystallography	26
2.1	Sample delivery	27
2.1.1	Moving target delivery systems	28
2.1.2	Fixed-target delivery systems	30
2.2	Data Collection & Detectors	31
2.3	Data Analysis	33
2.3.1	Hit Finding	33
2.3.2	Indexing	38
2.3.3	Integration	43
2.3.4	Phasing Methods	44
2.3.5	Model building, metrics and refinement	49
2.4	Research and Development	52
2.4.1	Synchrotron developments	52
2.4.2	XFEL developments	53
3	Multi-Hit Serial Femtosecond Crystallography	56
3.1	Introduction	56
3.2	Lysozyme	57
3.2.1	Biological background	57
3.2.2	Preparation	57
3.3	Motivation	58
3.4	Publication	60
3.5	Summary	91

4	A Comparison of Serial Crystallography Delivery Systems	93
4.1	Myeloid Differentiation Primary-Response 88	93
4.2	Introduction	95
4.3	Materials and Methods	97
4.3.1	MyD88/MAL Crystallisation	97
4.3.2	Microfluidic Silicon Chip	98
4.3.3	LCP Injector	101
4.3.4	GDVN Injector	102
4.4	Results	103
4.4.1	Microfluidic Silicon Chip	103
4.4.2	LCP Injector	110
4.4.3	GDVN Injector	115
4.4.4	Comparison	122
4.5	Discussion	122
4.6	Conclusion	131
5	The Structural Determination of MyD88	134
5.1	Introduction	134
5.2	Publication	136
5.3	Summary	151
6	Conclusions and Future Work	152
6.1	Summary of thesis	152
6.2	Future Work	155
A	Chapter 3 Supplementary Information	158
B	Chapter 5 Supplementary Information	165

List of Figures

1.1	A conceptual diagram of Bragg diffraction, where X-rays interact with a periodic array of atoms. The incident and diffracted X-rays are shown as blue arrows, atoms are shown as solid blue circles, d refers to the d-spacing and n refers to the diffraction order.	3
1.2	The 14 Bravais Lattice Types. A crystal family, or lattice system is assigned based on the combination of their axis lengths and angles. The lattice systems are further classified into their Bravais Lattice types based on their centring. For example, in an orthorhombic body-centred Bravais lattice, all three axial lengths (a , b & c) are unique, while all three angles (α , β & γ) are identical, assigning it to the orthorhombic lattice system, the crystal has lattice points at the eight corners of the unit cell, but also an additional lattice point at the centre of the cell. This figure is an edited version from Mendez 2019 [5]. . . .	5
1.3	Conventional crystallography setup showing a crystal loop mounted on a goniometer.	6
1.4	Some examples of a crystal with six different Miller indices depicted. The distance between parallel planes is known as d_{hkl} , the plane spacing. This figure is an edited version from Valtonen 2002 [8].	8

- 1.5 A schematic of crystal diffraction showing how the crystal in direct space relates to the reciprocal space using a mathematical formula. As the crystal rotates within the X-ray beam the possible reflections associated with the lattice planes which make up the crystal move through reciprocal space. When a reflection intersects the Ewald sphere the Bragg condition for that reflection is satisfied and a Bragg spot is produced which may be recorded on the detector. The Ewald sphere of radius $1/\lambda$. The incident wave vector \mathbf{k}_0 diffracts off the crystal in real space, where the crystal is set at the centre of the Ewald sphere. The intersection of \mathbf{k}_0 and the Ewald sphere is notated as $O^*(hkl)$, the origin of the reciprocal lattice. The intersection of the diffracted wave vector \mathbf{k} and the Ewald sphere is the reciprocal lattice point $P^*(hkl)$, where $OP^*(hkl)$ gives $1/d_{hkl}$, where d_{hkl} is the d-spacing. 9
- 1.6 Schematic illustration of a protein crystallisation phase diagram. The y-axis is protein concentration, while the x-axis is adjustable, and can be used for variables including precipitant concentration, pH, and temperature. Four crystallisation methods and their different routes to reach the nucleation and metastable zones are represented (i. batch method, ii. vapour diffusion, iii. dialysis, iv. free interface diffusion), for protein concentration versus precipitant concentration. The black dots show optimal starting conditions for each of these methods, with alternative starting points shown for free interface diffusion, and dialysis due to the undersaturated protein solution containing either pure protein or protein mixed with a low concentration of precipitating agent. The solubility curve is where the concentration of protein in the solute is in equilibrium with the crystals. The supersolubility curve is the line separating conditions for spontaneous nucleation from conditions where the crystallisation solution remains clear if undisturbed. This figure is reprinted with permission from Chayen 2004 [30]. Copyright 2004 Elsevier Ltd. All rights reserved. 15

1.7	Crystallisation via vapour diffusion. A hanging drop setup is shown on the left, where the droplet (represented as a blue circle) containing the protein, buffer and precipitant, is placed on the underside of the cover-slip, while the sitting drop setup is shown on the right, where the droplet is instead placed on a platform (represented in black), that sits slightly raised above the reservoir (represented in blue). Both setups are sealed to prevent moisture loss.	16
1.8	A plot showing the approximate year each new generation of X-ray producing light sources came online, as well as the Log Average Brightness and Log Peak Brightness of those light sources.	19
1.9	A simple diagram showing a typical third generation synchrotron. The main components of a third generation synchrotron include an electron gun and linear accelerator (linac), booster ring, storage ring composed of bending magnets and insertion devices, and the beamlines.	20
1.10	A diagram showing a wiggler and undulator, with an alternating static magnetic field (shown in red and blue), the electrons (e^-) entering the field, causing them to deflect and emit photons. The intensity from a wiggler is proportional to the number of electrons (N_e) multiplied by the number of poles (N_p), whereas the intensity from an undulator is proportional to the number of electrons (N_e) multiplied by the square of the number of poles (N_p^2).	22
1.11	Self amplified spontaneous emission (SASE) occurs when electrons are passed through an undulator. Initially electrons are bunched together with random phases, such that their contributions sum incoherently. After going through an undulator, the electron phases align, therefore producing extremely short and intense pulses of radiation. This figure is reprinted from Zhukovsky 2016 [34]. CC BY 4.0.	24

2.1	A diagram depicting a gas-focused dynamic virtual nozzle (GDVN) setup, and a double flow focusing nozzle (DFFN). The arrows on both images indicate the gas flow surrounding the capillary. This figure is reprinted with permission from Zhao 2019 [53]. Copyright 2019 Federation of European Biochemical Societies.	29
2.2	A diagram depicting the layout of a microfluidic chip fixed-target delivery system. The chip, with crystals sitting within the micropores, is attached to goniometer system. This figure is reprinted with permission from Zhao 2019 [53]. Copyright 2019 Federation of European Biochemical Societies.	31
2.3	A decision tree showing which of the delivery methods listed would be best, when accounting for sample viscosity, sample supply, SX technique, sample stability (whether minimal handling is a priority), if it is high-throughput, and whether the experiment is being performed at MHz rates. The delivery methods include: GDVN = gas-focused dynamic virtual nozzle; LCP = lipidic cubic phase injector (more broadly HVI) ; FT = fixed target systems; and DOD = drop-on-demand. This figure is reprinted from Cheng 2020 [69]. CC BY 4.0.	32
2.4	An example of a Ramachandran plot, showing the residues that are falling into the energetically favoured region (green), the energetically allowed region (yellow), and the high-energy unfavoured region (red) for the phi and psi torsion angles.	52
3.1	The European XFEL GDVN setup, where a 3D-printed nozzle is used to deliver the sample to the beam. The X-ray trains are delivered at a repetition rate of 10 Hz, with a 1.1 MHz pulse repetition rate within the trains. This figure is reprinted from Wiedorn 2018 [46]. CC BY 4.0.	59

-
- 4.1 For the particular case of the TLR4 MyD88-dependent pathway (right pathway), the MAL adaptor protein is used to recruit MyD88 in order to activate three pathways; the NF- κ B (nuclear factor- κ B) pathway, p38 (mitogen-activated protein kinase) MAPK pathway, and the JNK (JUN N-terminal kinase) MAPK pathway. This figure is reprinted with permission from O'Neill 2007 [177]. Copyright 1969, Nature Publishing Group. 95
- 4.2 Negative stained TEM image of MyD88^{TIR} needle-like crystals. The crystals were stained with 1% uranyl acetate and viewed with a JEOL JEM-2010 TEM. Image supplied by Dr. Connie Darmanin. 98
- 4.3 Sample loading procedure: A droplet of crystal suspension is pipetted onto the front side of the chip (A). The mother liquor then fills the pores, forming a meniscus on the lower side due to capillary action (B). The mother liquor is removed by touching the underside of the chip with filter paper, thereby drawing the excess liquid through the pores and soaking up the excess solution (C). Crystals that are larger than the pores are retained on the upper side of the chip, and arrange themselves in a periodic way according to the pore structure (D). The wet filter paper is then removed (E). The chip is then flash-frozen in liquid nitrogen (F). During A-E, the chip is under a continuous stream humidified air in order to control evaporation. This figure is reprinted from Roedig 2015 [185]. CC BY 4.0. 99
- 4.4 Static data collection setup of MyD88 crystals at Petra III, P11 beamline. (a) The goniometer setup with a silicon chip mounted under a cryogenic nozzle, and with the capillary beam stop in-place, (b) a close up view of a region on the chip showing the pores with crystals on the surface as indicated by the dark regions. The data was collected using a grid scan mode under liquid nitrogen conditions. 100

-
- 4.5 An image captured from the in-line microscope at the Petra III P11 beamline showing the distribution of crystals on the microfluidic silicon chip. Darker areas indicate regions where the sample is densely packed and single crystals cannot be distinguished, while lighter areas indicate regions of no or little sample, and where single, needle-like crystals can be distinguished. 104
- 4.6 37,746 detector images obtained using a microfluidic silicon chip system at Petra III were analysed for peaks. After hit finding, all the peaks from detector images that contained more than 5 peaks (10,754 images) were merged to generate this powder diffraction image. The image shows enough data was collected to represent the complete rotation of the crystal up to 8 Å, with weak diffraction extending beyond 4 Å. 106
- 4.7 CrystFEL indexing results showing the unit cell distributions found from the Petra III microfluidic silicon chip diffraction data. The unit cell parameters identified by indexing deviated greatly, and showed broad, undefined unit cell distributions. 107
- 4.8 Detector images showing crystal diffraction observed at the P11 beamline at Petra III using the microfluidic silicon chip. (a) Extremely sparse Bragg peaks were observed in a large proportion of the detector images, while the images that had more peaks (b) showed Bragg peaks located close together, indicative of multi-crystal diffraction. 108
- 4.9 CrystFEL indexing solution obtained from the simulated diffraction data generated from the published MyD88 structure (PDB code 4EO7) showing the unit cell distribution from the simulated single-crystal diffraction. 109
- 4.10 CrystFEL indexing solution obtained from the simulated diffraction data generated from the published MyD88 structure (PDB code 4EO7) showing the unit cell distribution from the simulated multi-crystal diffraction. 110

- 4.11 An image showing the HVI nozzle (top cross) with a stream of MyD88 crystals in LCP. The lower cross on the image indicates the location at which the sample and X-ray beam interact. The ripples of the sample (seen below the lower cross) demonstrate the effect of the X-ray beam on the sample after exposure. 111
- 4.12 Detector images collected from MyD88 crystals in LCP, on the CXI beamline at the LCLS. (a) An example of a detector image obtained. The peaks can be clearly identified, but there are also two regions where crystal diffraction is hard to distinguish from background signal; a centre region of high intensity due to the LCP, and a less intense ring region, caused by the solvent. (b) The same image as shown in (a) after a detector mask has been applied (blue) to exclude the LCP, solvent, bad pixels and detector edges, and after hit finding and indexing. Locations of peaks identified from hit finding correspond to the cyan rings and predicted peak locations from an indexing solution correspond to the pink rings. 112
- 4.13 Another example of a detector image collected from MyD88 crystals in LCP, on the CXI beamline at the LCLS, with peaks identified from hit finding circled by cyan rings. (a) A detector image after hit finding, where multiple distinct solvent rings can be seen in the image. In an attempt to minimise the region excluded from analysis due to masking, two thinner annulus shaped masks were tested (b). 113
- 4.14 A 2D powder pattern generated using the data collected from MyD88 crystals in LCP at the CXI beamline at the LCLS. The powder pattern was generated by summing all the peaks identified during hit finding together into one image in order to visually estimate the resolution of our data. 114
- 4.15 Indexing results for the MyD88 crystals in LCP. DirAx, MOSFLM and XGANDALF algorithms were used for indexing within the Psocake GUI. Unit cell distributions show broad, undefined unit cell parameters. 115
- 4.16 The tip of the GDVN nozzle (top) with the jet stream shown in this figure flowing from the nozzle vertically down. 116

-
- 4.17 A GDVN MyD88 crystal detector image obtained from the CXI beamline at the LCLS, at a 111 mm detector distance. Initial hit finding included a mask (red) that was applied to all detector images to exclude bad pixels, detector edges, the jet streak region, and the detector shadowing region. Detector shadowing was only observed in data obtained at a 111 mm detector distance, and therefore the detector shadowing region was only masked for that data subset. Blue circles indicate locations where hit finding has identified Bragg peaks. 117
- 4.18 GDVN MyD88 crystal diffraction patterns obtained from the CXI beamline at the LCLS after hit finding, showing narrow, minimal jet streaking (a) and no jet streaking (b). 118
- 4.19 A 2D powder pattern generated from data obtained from GDVN delivered MyD88 crystals, using the CXI beamline at the LCLS. The powder pattern was generated by summing all the peaks identified during hit finding together into one image in order to visually estimate the resolution of our data. . . . 119
- 4.20 Indexing solutions obtained from data collected at 181 mm and 106.1 mm detector distances using GDVN delivery at the CXI beamline at the LCLS. The unit cell distributions have narrow and well defined peaks. 120
- 4.21 A comparison showing the normalised integrated intensity 1D plots for the GDVN and microfluidic silicon chip data, with an insert showing a clearer view of the area within the rectangle. The dotted line shown, is the common peak that was used to normalise both data sets. 122

List of Tables

1.1	A comparison of current and future Hard X-ray Facilities	25
3.1	Candidate and co-author contributions	61
4.1	Crystallography data collection and analysis statistics. MyD88 crystal diffraction was obtained at the P11 beamline at the Petra III synchrotron using a microfluidic silicon chip system. MyD88 crystal diffraction was also obtained at the CXI beamline at the LCLS XFEL using the GDVN and HVI sample delivery systems. From these experiments, data analysis was performed, and the results of hit finding and indexing are presented in this table.	121
4.2	Sample delivery efficiency comparison. Based on a sample volume of 2.5 μL .	130
5.1	Candidate and co-author contributions	136

Abstract

Serial femtosecond crystallography (SFX) is a technique that has allowed scientists to determine the structure of proteins to an atomic resolution using sub-micron sized crystals. With the development of SFX, the structural determination of crystals too small to be measured using conventional crystallography, could be solved. Using SFX, a crystal can be probed, and diffraction obtained prior to significant damage occurring to the crystal lattice. The process of obtaining crystallography data prior to the crystal being destroyed is typically referred to as "diffraction before destruction". However, as we show in this thesis, it is possible to obtain SFX data from the same crystal more than once using consecutive X-ray pulses. I investigate this phenomenon, referred to as "multi-hit" SFX using lysozyme diffraction data obtained at the European XFEL. Using the diffraction patterns obtained from only those lysozyme crystals that had previously been probed, a second structure was solved, and after a detailed comparison it was determined that no significant radiation damage occurred. The ability to obtain a second diffraction pattern from the same crystal opens up the possibility of performing novel experiments using time-resolved SFX (tr-SFX).

SFX is still a relatively new field, with continuous developments occurring. A number of different options for sample delivery exist, and part of the work in this thesis is to compare and contrast three of these sample delivery systems. This comparison used the biologically significant Myeloid differentiation primary-response 88 (MyD88) protein. The crystals were needle-like and were prone to aggregation which was taken into account when choosing which delivery systems were compared. Using SFX with the gas-focused dynamic virtual nozzle (GDVN), the structure of MyD88 was determined to a resolution of 2.3Å and compared to the structure of MyD88 determined via microcrystal electron diffraction (microED).

Statement of Authorship

This thesis includes work by the author that has been published or accepted for publication as described in the text. Except where reference is made in the text of the thesis, this thesis contains no other material published elsewhere or extracted in whole or in part from a thesis accepted for the award of any other degree or diploma. No other person's work has been used without due acknowledgement in the main text of the thesis. This thesis has not been submitted for the award of any degree or diploma in any other tertiary institution.

Susannah Holmes

9 December 2021

Acknowledgements

I would like to express my gratitude to my primary supervisors Dr. Connie Darmanin for her constant support, encouragement, accountability, and feedback, and Prof. Brian Abbey for his encouragement, support, and feedback. Dr Connie Darmanin supported and guided me from a perspective of biology, biochemistry, and chemistry, generously sharing her time and expertise, while Prof. Brian Abbey supported, and shared his expertise with me from a chemistry and physics perspective. I would also like to express my gratitude to my secondary supervisors Dr. Andrew Martin for his theoretical physics expertise and thesis feedback, and Dr. Nadia Zatsepin for sharing her physics and data analysis expertise with me, helping me to optimise my approach to crystallography data analysis, particularly with respect to the structural determination of MyD88.

My gratitude also goes out to Assoc. Prof. Narelle Brack for her continuous support and encouragement during not only my PhD candidature, but also through my time at La Trobe university as an undergraduate. I would also like to thank my fellow students for their encouragement, brainstorming sessions, understanding and support, particularly Dr Nicholas Anthony, Dr Robert Sikos, and soon to be Dr. Ruqaya Maliki.

My research was a collaboration of scientists that was made capable by the Australian Research Council Centre of Excellence in Advanced Molecular Imaging. This work was supported by an Australian Government Research Training Program Scholarship and a La Trobe University Postgraduate Research Scholarship, with additional scholarship support from the Australian Research Council (ARC) Centre of Excellence for Advanced Molecular Imaging.

Lastly, I would like to thank Stephanie Marsell for her constant encouragement and verbal

pushes to keep me on track, and my mother Lana Holmes for her love and support.

List of Publications

- Holmes, S, et al. *Megahertz pulse trains enable multi-hit serial crystallography experiments at XFELs*. (submitted to Nature Communications)
- Holmes, S, et al. *MyD88 TIR domain higher-order assembly interactions revealed by microcrystal electron diffraction and serial femtosecond crystallography*. Nature Communications **12**, 2578 (2021)
- Anthony, N, Darmanin, C, Bleackley, M.R, Cadenazzi, G, Holmes, S, Nugent, K.A & Abbey, B. *Ptychographic Imaging of NaD1 Induced Yeast Cell Death*. Biomedical Optics Express **10**, 4964 (2019)
- Weidorn, M.O, et al. *Megahertz serial crystallography*. Nature Communications **9**, 4025 (2018)

Abbreviations

FT = Fourier transform

FFT = Fast Fourier Transform

SIR = Single Isomorphous Replacement

SIRAS = Single Isomorphous Replacement with Anomalous Scattering

MIR = Multi-Isomorphous Replacement

MIRAS = Multi-Isomorphous Replacement with Anomalous Scattering

SAD = Single-wavelength Anomalous Diffraction

MAD = Multi-wavelength Anomalous Diffraction

MR = Molecular Replacement

XFEL = X-ray Free Electron Laser

FID = Free Interface Diffusion

ESRF = European Synchrotron Radiation Facility

NSLS II = National Synchrotron Light Source II

DLS = Diamond Light Source

RF = Radio Frequency

K-B = Kirkpatrick-Baez

SASE = Self-Amplified Spontaneous Emission

LCLS = Linear Coherent Light Source

SACLA = Spring-8 Angstrom Compact Free Electron Laser

PAL-XFEL = Pohang Accelerator Laboratory X-ray Free Electron Laser

EuXFEL = European X-ray Free Electron Laser

SHINE = Shanghai High Repetition Rate XFEL and Extreme Light Facility

SX = Serial Crystallography
SFX = Serial Femtosecond Crystallography
GDVN = Gas-focused Dynamic Virtual Nozzle
DFFN = Double Flow Focusing Nozzle
HVI = High Viscosity Injector
SMX = Serial Millisecond Crystallography
SSX = Serial Synchrotron Crystallography
tr-SFX = time-resolved SFX
LCP = Lipidic Cubic Phase
DOD = Drop-On-Demand
ADE = Acoustic Droplet Ejection
CSPAD = Cornell-SLAC Pixel Array Detector
SNR = Signal-to-Noise Ratio
ADC = Analogue-to-Digital-Converter
NCS = Non-Crystallographic Symmetry
SPI = Single Particle Imaging
AGIPD = Adaptive Gain Integrating Pixel Detector
PDB = Protein Data Bank
HPLC = High Pressure Liquid Chromatography
TLR = Toll-Like Receptor
PRR = Pattern Recognition Receptor
PAMP = Pathogen Associated Molecular Pattern
DAMP = Danger Associated Molecular Pattern
MyD88 = Myeloid Differentiation primary-response 88
MAL = MyD88-Adaptor-Like
TIR = Toll/Interleukin-1 Receptor

Chapter 1

X-ray Crystallography

1.1 History of X-ray Crystallography

Scientists have always had questions and hypotheses about the underlying order and structure of crystals, but many of these hypotheses were unable to be tested until the discovery of X-rays in 1895 by Wilhelm Roentgen. The discovery of X-rays led to new level of scientific understanding of crystal structures, and in 1912, the first diffraction pattern was obtained by Max Laue, Walter Friedrich, and Paul Knipping. With this first diffraction pattern, obtained by passing X-rays through a copper sulfate crystal and recording the results on photographic plates, the dual wave-particle nature of X-rays was demonstrated [1]. In the same year, William Lawrence Bragg showed how an observed diffraction pattern could be used to determine the position of atoms within a crystal. It was with this experiment that the field of X-ray crystallography arguably began. William Lawrence Bragg, working with his father, William Henry Bragg, then went on to identify the first ever crystal structures; that of salt [2] and diamond [3].

With X-ray crystallography comes the ability to identify the structure of both inorganic and organic materials. From a non-biological perspective, understanding the underlying structure of inorganic materials means that we can delve into what causes a material to exhibit specific properties. From this fundamental understanding, new and improved materials

can be created. X-ray crystallography also allows us to develop our understanding of biological structures, mechanisms, and interactions via the structural determination of protein crystals, with the first protein structure, that of Myoglobin, solved in 1958 by John Kendrew [4]. The ability to determine the structure of proteins allows us to understand their function within our body and how that structure changes in response to different stimuli. Knowing how these proteins look when they are in their active/inactive state can tell us why, for example, they may be causing disease. This knowledge allows us to develop an understanding of how specific proteins interact with one another, providing us with the information to either increase or block their mode of action. This step can be of great importance in knowing how to design and develop therapeutics that can cure disease/illnesses or alleviate symptoms. Since the first protein structure solved in 1958, over 100,000 protein structures have now been solved by X-ray crystallography, and with continuous developments in the field of X-ray crystallography, this number will continue to grow.

1.2 X-ray Diffraction

There are a number of different ways in which X-rays interact with matter. These different types of interactions are dependent on the wavelength of the X-rays, and include coherent scattering, Compton scattering, the photoelectric effect, pair production, and photodisintegration. X-ray scattering is modelled on a particular case of coherent scattering, called Thomson scattering, which describes the scattering of X-rays as they interact with free charged particles:

$$I_{th} = I_0 \frac{e^4}{m_e^2 r^2 c^4} \sin^2 \theta, \quad (1.1)$$

where I_{th} is the intensity of the scattered wave, I_0 is the intensity of the incoming wave, e is the charge of the electron, m_e is the mass of the electron, r is the classical radius of the electron, c is the speed of light, and θ is the angle between the electron's direction of acceleration and direction of observation. The special case of this relationship, referred to as X-ray diffraction, occurs when X-rays interact with matter in crystal form, and is described by Bragg's Law:

$$n\lambda = 2d\sin\theta. \quad (1.2)$$

Due to the regularly spaced repeating units that make up a crystal, when light of similar wavelength to the inter-atomic (d) spacing interacts with a crystal, a diffraction pattern can be obtained. Bragg's law indicates that only constructive interference ($n\lambda$) will produce visible diffraction peaks, called Bragg peaks or reflections, observed at specific locations. As can be seen in Figure 1.1, the distance travelled by X-rays that have interacted with atoms and have then been diffracted, needs to be an integer multiple, called the diffraction order (n) of the X-ray wavelength (λ) to maintain the same phase so that they interact constructively. If the X-rays travel a non-integer multiple of the wavelength, it will be out of phase, leading to destructive interference. This equation is fundamental to enabling the determination of the structure of a protein from the diffraction patterns which contain the Bragg peaks.

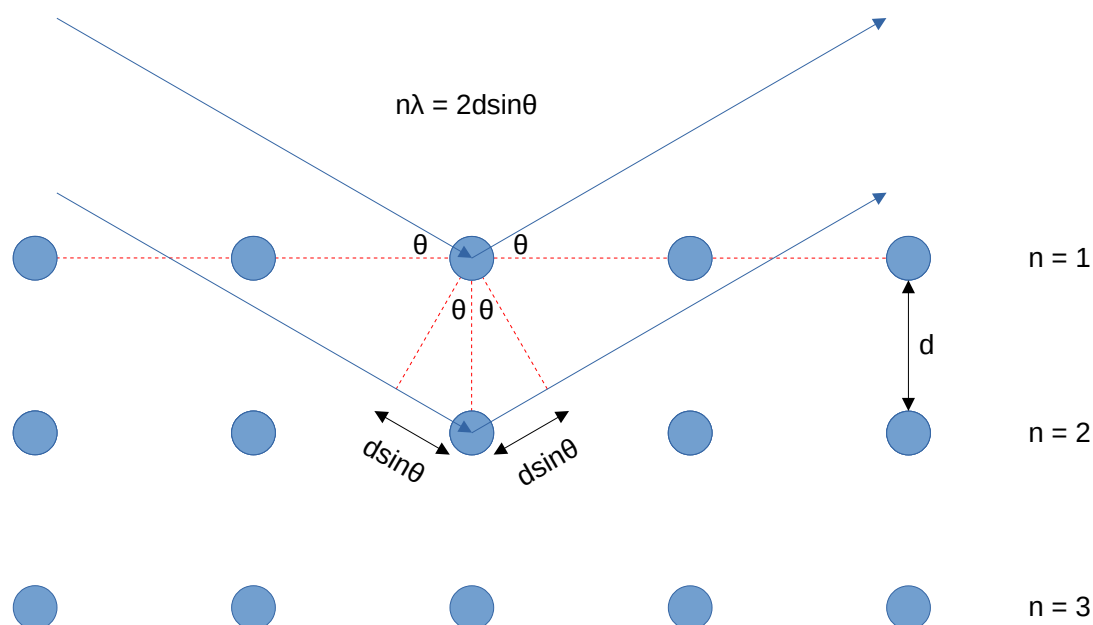


FIGURE 1.1: A conceptual diagram of Bragg diffraction, where X-rays interact with a periodic array of atoms. The incident and diffracted X-rays are shown as blue arrows, atoms are shown as solid blue circles, d refers to the d-spacing and n refers to the diffraction order.

1.3 Crystal Packing & Symmetry

X-ray crystallography is a widely accepted technique used for the structural determination of crystalline materials. A crystal is characterised as an ordered, periodically repeating array of

atoms or molecules. This repeating array can be described by a unit cell, a representation of the smallest repeating unit of a crystal that has axes defined as a , b , and c and angles defined as α , β , and γ . The unit cell is further classified into different crystal lattice systems by the differing combinations of lengths and angles that are possible. There are 7 lattice systems: triclinic, monoclinic, orthorhombic, tetragonal, rhombohedral, hexagonal, and cubic. These crystal systems are then further classified by their centring: primitive (P), base centred (S), body centred (I), and face centred (F), which results in 14 different known classifications, referred to as Bravais lattices (Fig. 1.2).

Crystals are further grouped by their symmetry or symmetry operations which describe how the molecule is ordered within the crystal array. It is these symmetry relationships, identified through the systematic presence or absence of Bragg peaks, that allow us to characterise the whole crystal from X-ray diffraction. There are different symmetries that can be associated with molecules within the crystal. Rotational symmetry, where the orientations of the molecules within a crystal are related by an angular function, translational symmetry, where the molecules within a crystal are related by an x-y-z translation, and combinations of these two symmetry types which are referred to as screw axes. There are also different levels of rotational symmetry, depending on how often within a 360° rotation the molecule is identically positioned when compared to the initial reference point (0°). A 1-fold rotation means that there is no point of rotation in which the unit cell is identical. A 2-fold rotation, describes when the molecule is identical twice within a 360° rotation, at 0° and at 180° . Up to a 6-fold rotation is possible in crystallography.

These symmetry operations are grouped together in what are referred to as Point groups. There are 32 possible crystallographic point groups in 3-dimensional space. From the various combinations of point groups and Bravais lattices, 230 space groups can be defined. Not all of these 230 space groups are possible though for biological molecules, since space groups containing mirror or inversion symmetry are not present in naturally occurring proteins [6]. The space groups are described in detail in the International Tables for Crystallography Volume A: Space-group symmetry [7].

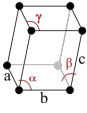
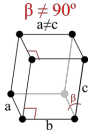
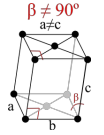
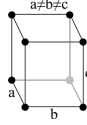
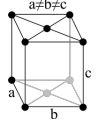
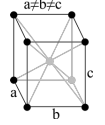
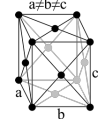
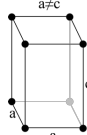
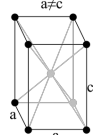
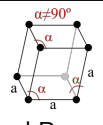
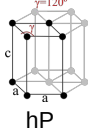
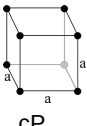
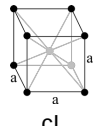
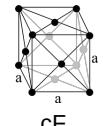
Crystal Family	Lattice System	14 Bravais Lattices			
		Primitive (P)	Base centred (S)	Body centred (I)	Face centred (F)
Triclinic (a)					
		aP			
Monoclinic (m)					
		mP	mS		
Orthorhombic (o)					
		oP	oS	oI	oF
Tetragonal (t)					
		tP		tI	
Hexagonal (h)	Rhombohedral				
	Hexagonal				
Cubic (c)					
		cP		cI	cF

FIGURE 1.2: The 14 Bravais Lattice Types. A crystal family, or lattice system is assigned based on the combination of their axis lengths and angles. The lattice systems are further classified into their Bravais Lattice types based on their centring. For example, in an orthorhombic body-centred Bravais lattice, all three axial lengths (a , b & c) are unique, while all three angles (α , β & γ) are identical, assigning it to the orthorhombic lattice system, the crystal has lattice points at the eight corners of the unit cell, but also an additional lattice point at the centre of the cell. This figure is an edited version from Mendez 2019 [5].

1.4 X-ray Data Collection

The first step required for the analysis of a crystal is X-ray diffraction data. The data is collected by placing the sample in the X-ray beam region. Different methods are used to deliver and manipulate the crystals within this region and are dependent on the nature of the sample and which experimental technique is used. These techniques include conventional crystallography, powder diffraction and serial crystallography. In conventional crystallography, rotation of the crystal is a necessary step so that data for all orientations of the sample can be collected. Therefore, a single crystal is placed within a loop or capillary system and then mounted onto a goniometer (Fig. 1.3). For powder diffraction techniques, rotation is not necessary as the sample is in a powder or micro-crystalline form that is assumed to cover all orientations. Therefore, a basic sample holder consisting of a well or capillary, where that sample can be placed, is all that is required.

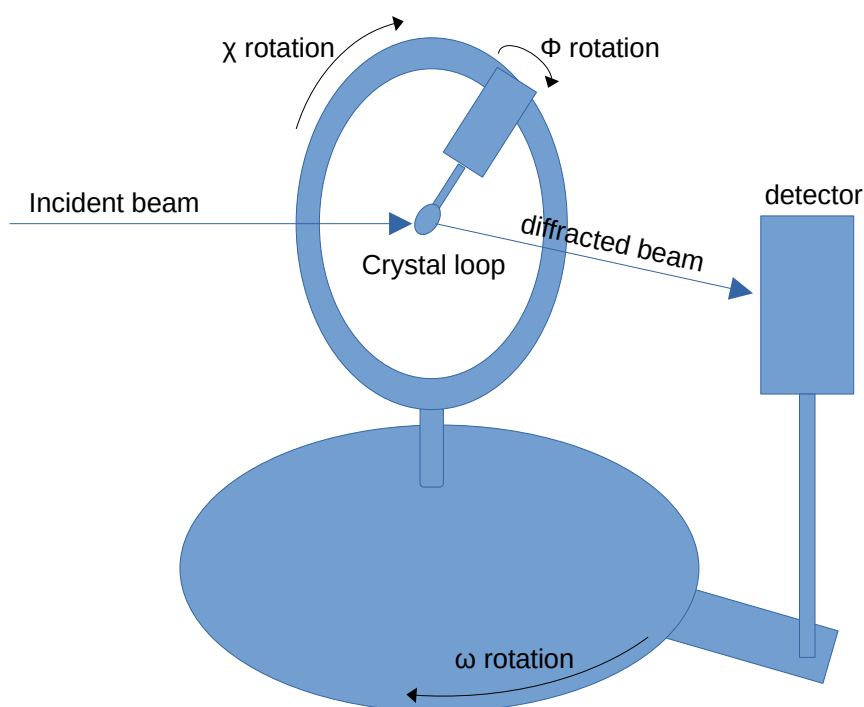


FIGURE 1.3: Conventional crystallography setup showing a crystal loop mounted on a goniometer.

After the X-rays have interacted with the sample, diffraction is recorded using a detector. The quality of the detector is a significant factor in determining what type of experiments can be carried out (see subsection 2.2 for further details).

1.5 Structural analysis

Following diffraction, the data, in the form of discrete Bragg peaks cannot be directly used to describe the "real-space" crystal lattice. These Bragg peaks instead describe the crystal's lattice in what is referred to as "reciprocal space", "momentum space" or "Fourier space". This reciprocal space is the visual representation of the Fourier transform of the crystal lattice. The reciprocal space positions of the Bragg peaks can be transformed into their real or direct space positions via the Laue equations, Ewald sphere, structure factors, electron density reconstruction, temperature factors and phase recovery all which are outlined in the following subsections.

1.5.1 Laue

The Laue equations (Eq. 1.3), give the conditions necessary for the incident wave to be diffracted by the crystal. If we have a crystal with the three basis vectors \mathbf{a} , \mathbf{b} , \mathbf{c} , where atoms are located at positions $\mathbf{x} = p\mathbf{a} + q\mathbf{b} + r\mathbf{c}$, an incident wave vector \mathbf{k}_0 and diffracted wave vector \mathbf{k} , then the three conditions that need to be met are:

$$\begin{aligned}\mathbf{a} \cdot \Delta\mathbf{k} &= 2\pi h \\ \mathbf{b} \cdot \Delta\mathbf{k} &= 2\pi k \\ \mathbf{c} \cdot \Delta\mathbf{k} &= 2\pi l,\end{aligned}\tag{1.3}$$

where $\Delta\mathbf{k}$ is $\mathbf{k} - \mathbf{k}_0$, and h , k , and l are integers. If all conditions are met, the incoming wave is reflected onto a set of lattice planes, a mathematical construct described by Miller indices (Fig. 1.4) defined as h/n , k/n , l/n , where n is the integer $hp + kq + lr$. It is the Miller indices which provide a description of the orientation of a plane or set of parallel planes.

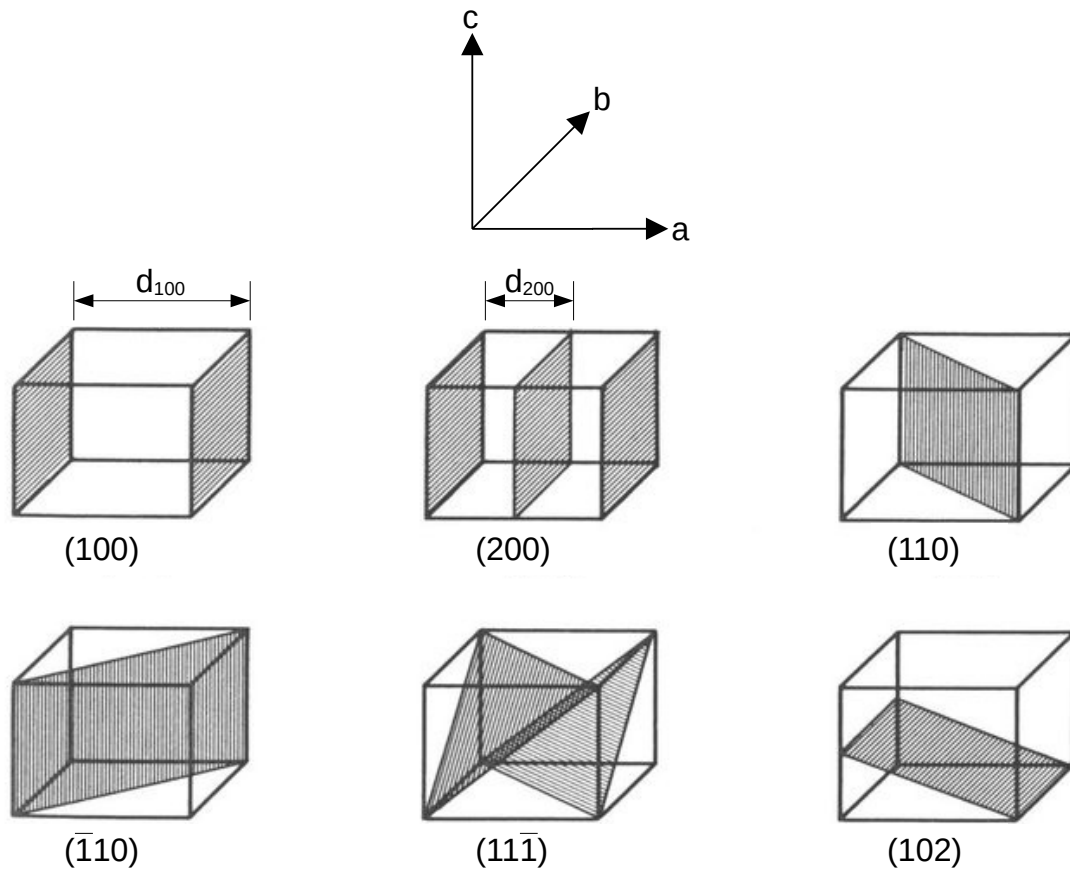


FIGURE 1.4: Some examples of a crystal with six different Miller indices depicted. The distance between parallel planes is known as d_{hkl} , the plane spacing. This figure is an edited version from Valtonen 2002 [8].

1.5.2 Ewald construction

The Ewald sphere is a geometric construction that depicts the relationship between the wave vector of an incident X-ray beam (\mathbf{k}_0), the diffracted X-ray beam (\mathbf{k}), the diffraction angle of a reflection, and the reciprocal lattice of the crystal [9, 10], with the assumption that the beam is fully coherent (Fig. 1.5). It depicts only those directions, under which the diffracted waves appear due to constructive interference and leads to the understanding of what we refer to as the reciprocal lattice. The Ewald sphere also helps to illustrate the idea of partiality; whether a reflection is a "full" or "partial" reflection.

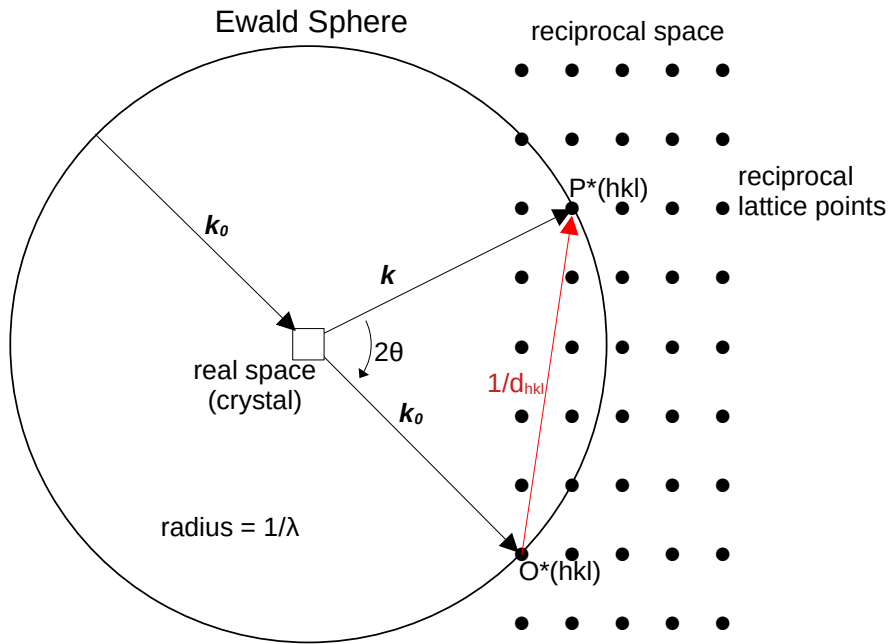


FIGURE 1.5: A schematic of crystal diffraction showing how the crystal in direct space relates to the reciprocal space using a mathematical formula. As the crystal rotates within the X-ray beam the possible reflections associated with the lattice planes which make up the crystal move through reciprocal space. When a reflection intersects the Ewald sphere the Bragg condition for that reflection is satisfied and a Bragg spot is produced which may be recorded on the detector. The Ewald sphere of radius $1/\lambda$. The incident wave vector \mathbf{k}_0 diffracts off the crystal in real space, where the crystal is set at the centre of the Ewald sphere. The intersection of \mathbf{k}_0 and the Ewald sphere is notated as $O^*(hkl)$, the origin of the reciprocal lattice. The intersection of the diffracted wave vector \mathbf{k} and the Ewald sphere is the reciprocal lattice point $P^*(hkl)$, where $OP^*(hkl)$ gives $1/d_{hkl}$, where d_{hkl} is the d-spacing.

If we assume a fully coherent beam and define partiality as whether or not the centre of the Bragg peak (maximum) lies along the Ewald sphere, then it is possible to obtain a "full reflection" from a single snapshot. But, if instead a full reflection is defined as the complete Bragg peak, multiple diffraction angles may be required. Sampling a complete Bragg peak is possible in conventional crystallography where the crystal can be slowly rotated through the X-ray beam, but when performing crystallography experiments with an incoherent or partially coherent beam the Ewald sphere depiction is not as accurate. For an incoherent or partially coherent beam, a single Ewald sphere is incorrect, as one Ewald sphere corresponds

to one wavelength (where the radius of an Ewald sphere = $1/\lambda$). More accurately, an amalgamation of Ewald spheres into an "Ewald shell" is closer to depicting the reality of diffraction from an incoherent or partially coherent source.

1.5.3 Partiality

Partiality is an important factor to consider in crystallography. As described above, in conventional crystallography we can optimise the experimental conditions, adjusting the degree of rotation so that the full reflections are measured. In serial crystallography, because of the uncontrolled, random orientation of the crystal, the angle at which the incident X-rays diffract from the crystal results in partial measurements, called partial reflections. These reflections need to be considered when performing serial crystallography experiments, as there are limited options for minimising these partial reflections. Therefore, initial indexing of serial crystallography data requires some tolerance in identification of peak centres, and serious consideration is needed to then accurately integrate these partial reflections.

A way that serial crystallography overcomes these partial reflections is via the Monte Carlo technique, where the full reflection intensities are estimated by averaging over all of the data [11]. While this seems quite straightforward, large data redundancies are necessary in order to ensure that the averages of the reflections calculated are accurate.

Post-refinement methods are used in order to improve the accuracy when estimating full reflections. These methods involve an iterative approach that estimates the initial fully integrated intensities by initial diffraction geometry estimates [11]. The geometry parameters are then refined for each pattern to allow increasing agreement between geometry estimates and the calculated intensities.

1.5.4 Structure Factors

Structure factors \mathbf{F}_{hkl} represent the final diffracted waves scattered by all atoms in the unit cell for each set of miller planes (hkl):

$$\mathbf{F}_{hkl} = |\mathbf{F}_{hkl}| \exp(i\alpha_{hkl}) = \sum_{j=1}^N f_j \exp(2\pi i[hx_j + ky_j + lz_j]), \quad (1.4)$$

where N is the number of atoms in the unit cell, and f_j is the scattering contribution for each atom. The intensity of each reflection I_{hkl} that ends up on the detector is proportional to the amplitude of the structure factor:

$$I_{hkl} \propto |\mathbf{F}_{hkl}|^2. \quad (1.5)$$

While we are now able to calculate the amplitude contribution, the phase information from the data is lost and this is what is commonly known as the "phase problem" (see subsection 1.5.7).

1.5.5 Electron Density

In order to solve the structure of a crystal, we must first generate an electron density map. A mathematical relationship between the structure factors and electron density ρ_{xyz} , allows us to calculate these density maps, where V is the volume of the unit cell:

$$\rho_{xyz} = \frac{1}{V} \sum_h \sum_k \sum_l |\mathbf{F}_{hkl}| \exp(-2\pi i[hx + ky + lz - \phi_{hkl}]). \quad (1.6)$$

As can be seen in Equation 1.6, every reflection contributes to the electron density at each point within the unit cell. The relationship between the electron density and structure factors above are described by a Fourier transform (FT):

$$\rho_{xyz} = FT^{-1}(\mathbf{F}_{hkl}). \quad (1.7)$$

While the electron density function (Eq. 1.6) requires knowledge of the phase, the Patterson function:

$$P_{uvw} = \frac{1}{V} \sum_h \sum_k \sum_l |\mathbf{F}_{hkl}|^2 \cos 2\pi[hu + kv + lw], \quad (1.8)$$

derived from the electron density function uses the amplitude of the structure factor $|\mathbf{F}_{hkl}|$, which is related to the intensity, rather than using the structure factor \mathbf{F}_{hkl} , thereby allowing what is referred to as a Patterson map to be calculated.

1.5.6 Temperature factor

The temperature factor, also known as the Debye-Waller factor or B-factor, is the parameter used to describe the attenuation of X-rays due to thermal motion [12]:

$$B = 8\pi^2 U_i^2, \quad (1.9)$$

where U_i^2 is the mean squared displacement of the atom. Thermal motion causes the atoms to be displaced from their mean atomic position and it is this displacement that causes diminished scattering intensities, this needs to be accounted for, in order to accurately calculate electron density maps. The higher the temperature, the more these thermal motions cause "smearing" of the Bragg peaks, therefore many experiments try to minimise this by utilising cryo-cooling. The thermal contribution is incorporated into the atomic scattering factor (f_T):

$$f_T = f_0 \exp\left(\frac{-B \sin^2 \theta}{\lambda^2}\right), \quad (1.10)$$

which is proportional to the amplitude of the wave.

1.5.7 Phase problem

The phase problem exists in crystallography as it is only possible to measure the amplitude of the Bragg peaks from experimental data and not the phase [13] Various methods have been developed in order to overcome this problem.

Phase retrieval methods include direct methods [14, 15], Single Isomorphous Replacement (SIR) [16, 17], Single Isomorphous Replacement with Anomalous Scattering (SIRAS), Multi-Isomorphous Replacement (MIR), Multi-Isomorphous Replacement with Anomalous Scattering (MIRAS), Single-wavelength Anomalous Diffraction (SAD) [18, 19], Multi-wavelength Anomalous Diffraction (MAD) [20] and Molecular Replacement (MR) [14], as well phase improvement via density modification [21–24] as a secondary step.

Direct methods [14, 15] work on the assumption that the crystal is made up of similarly shaped atoms, and also that all atoms have positive electron density. With these two assumptions, there exist statistical relationships between sets of structure factors, which can then be used to estimate possible phase values. As the structure becomes larger and more

complex, the statistical relationship is diminished, and because of this, direct methods can only be used on small molecules and cannot be used on large molecules such as proteins. Therefore it is these other phase retrieval methods, previously mentioned, that are typically relied on in protein crystallography, and these methods are described in further detail in subsection 2.3.4.

1.6 Protein Crystallography

Protein crystallography is the study of the atomic structure of proteins that are crystallised and grown to sufficient size that they generate detectable Bragg peaks when illuminated by a hard X-ray source. There are a large number of steps involved in the structural determination of proteins, with many pitfalls requiring the user to back-track and find new paths forward. The process can be categorised into three main steps that includes protein crystallisation, experimental method, and data analysis.

1.6.1 Crystallisation

In order to perform protein crystallography experiments, obtaining high quality crystals of the protein of interest is a necessary first step. Crystallisation involves combining protein molecules in such a way that a repeating, ordered array (the unit cell), of molecules is produced. This repeating array is held together via non-covalent bonds, hydrogen bonds and salt bridges. The process of crystallisation is thermodynamic in nature, with the protein molecules arranging themselves into a low-energy solid state nucleating out from the solvent state in which they were in. Most proteins do not form crystals naturally and therefore the crystallisation process is induced by the addition of crystallisation buffers.

The process of crystallisation can require very specific sets of conditions to be met for successful crystal nucleation and growth to occur. Depending on the X-ray source, the crystal size requirement can vary. A crystal of at least 20 μm is needed for diffraction using conventional crystallography with lab-based sources. While for synchrotrons, crystals of at least 2 μm are required, which (for smaller crystals) requires specialised micro/nano

focus synchrotron beamlines [25–27]. Furthermore, even sub-micron crystals, as well as single particles can be investigated using X-ray Free Electron Laser (XFEL) sources, with structural determination of particles with diameters in the 10^2 nm range now possible [28, 29].

The crystallisation process requires protein to be mixed with buffer solutions. Conditions associated with the buffer solution, such as salt concentration, pH and ionic strength, temperature and protein concentration all influence the success or failure of protein crystallisation. Phase diagrams (Fig. 1.6), showing protein concentration for these different variables can be used to determine what set of conditions is best and include an undersaturated region in which crystals dissolve, and a supersaturated region, where crystal growth occurs. This supersaturated region is split into three zones; the metastable zone, the nucleation zone (also referred to as the labile or crystallisation zone), and the precipitation zone [30, 31].

For crystallisation to be successful, nucleation zone conditions are necessary as this is where spontaneous nucleation occurs. Furthermore, for sustained growth, and for the growth of large, well-ordered crystals, movement from the nucleation zone into the metastable zone is required [30]. If the nucleation zone is sustained, only tiny crystals will be produced, so when determining optimal conditions during crystallisation screening, the aim is to push the solution into these two zones.

Furthermore, these conditions are unique for each protein, with no "one-size-fits-all" combination of conditions leading to successful crystallisation. Therefore, it is necessary to screen many combinations of crystallisation conditions in order to determine the optimal conditions for that specific protein. The optimal starting conditions can also vary depending on what crystallisation method is utilised, as can be seen in Figure 1.6, where batch method, vapour diffusion, dialysis, and Free Interface Diffusion (FID) all have different optimal starting conditions, before the solution is guided into the nucleation and metastable zones.

One of the ways to perform crystallisation trials is via vapour diffusion. Vapour diffusion utilises the evaporation and diffusion of water between a small droplet (0.5–10 μL), and a reservoir [32]. The droplet contains the protein, buffer, and precipitant while the reservoir contains a higher concentration of both buffer and precipitant, and contains no protein

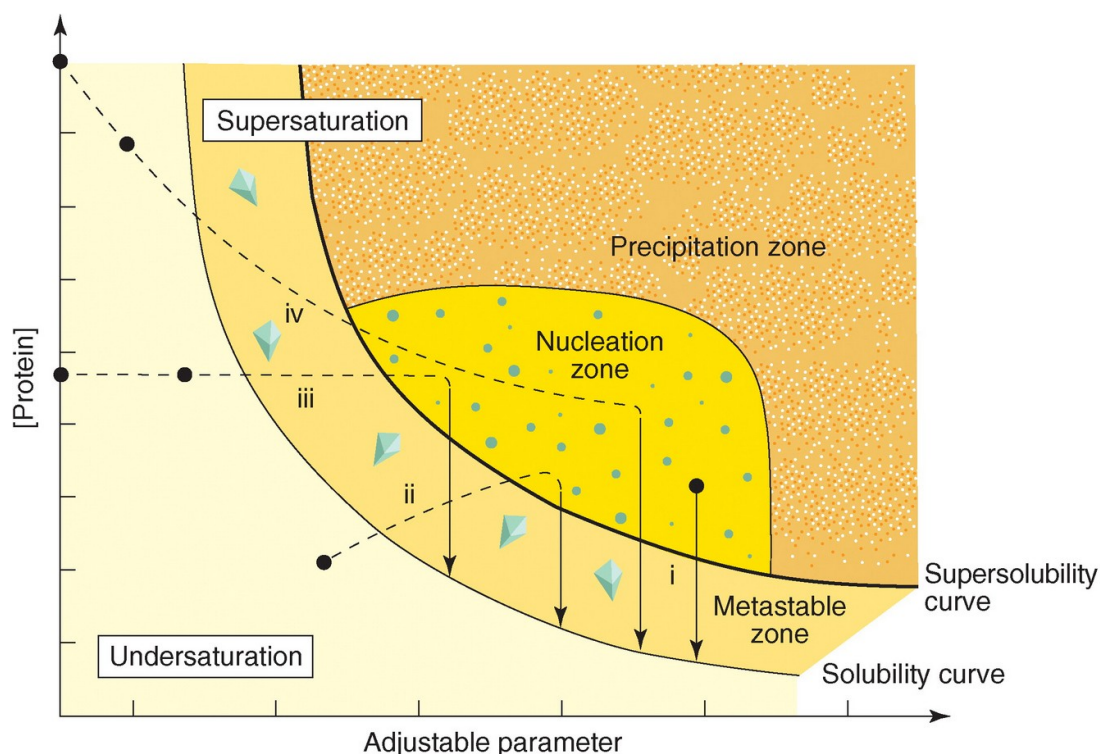


FIGURE 1.6: Schematic illustration of a protein crystallisation phase diagram. The y-axis is protein concentration, while the x-axis is adjustable, and can be used for variables including precipitant concentration, pH, and temperature. Four crystallisation methods and their different routes to reach the nucleation and metastable zones are represented (i. batch method, ii. vapour diffusion, iii. dialysis, iv. free interface diffusion), for protein concentration versus precipitant concentration. The black dots show optimal starting conditions for each of these methods, with alternative starting points shown for free interface diffusion, and dialysis due to the undersaturated protein solution containing either pure protein or protein mixed with a low concentration of precipitating agent. The solubility curve is where the concentration of protein in the solute is in equilibrium with the crystals. The supersolubility curve is the line separating conditions for spontaneous nucleation from conditions where the crystallisation solution remains clear if undisturbed. This figure is reprinted with permission from Chayen 2004 [30]. Copyright 2004 Elsevier Ltd. All rights reserved.

[32]. The two volumes are sealed together within a well, and the two volumes equilibrate, leading to a slow increase of both the protein and precipitant concentration in the drop, allowing conditions for crystal growth to be induced [32]. Setting up various combinations of conditions for screening can be quite tedious, therefore crystallisation robots are usually employed to automate this screening process. Traditionally vapour diffusion used a "hanging drop" method (Fig. 1.7), where the drop was manually placed on the bottom side of a cover

slip that is then used to seal the well. With robots, a "sitting drop" method (Fig. 1.7) was developed, where the drop is delivered to a platform above the reservoir. This was more easily automated than the hanging drop method, and while some robotic systems can now perform both the hanging drop, and sitting drop methods, the sitting drop method is still predominantly utilised by robots.

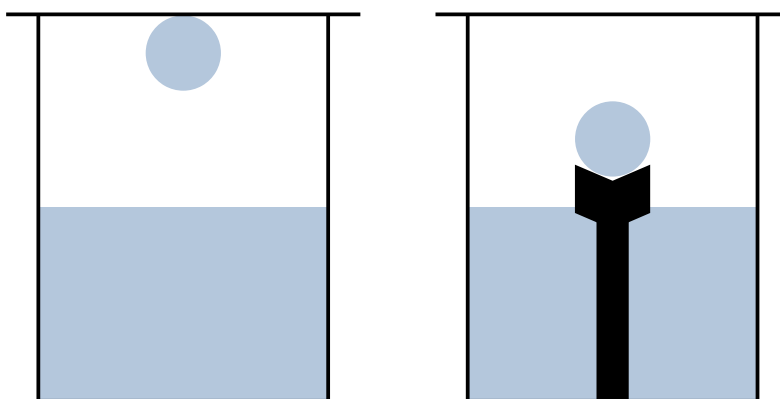


FIGURE 1.7: Crystallisation via vapour diffusion. A hanging drop setup is shown on the left, where the droplet (represented as a blue circle) containing the protein, buffer and precipitant, is placed on the underside of the cover-slip, while the sitting drop setup is shown on the right, where the droplet is instead placed on a platform (represented in black), that sits slightly raised above the reservoir (represented in blue). Both setups are sealed to prevent moisture loss.

While vapour diffusion allows for fast, cost effective screening, with minimal protein consumption, and is the predominant method used for initial screening, or where only small volumes of crystal are required, other methods are used when larger volumes are crystal are needed. One of these methods is the batch method, where supersaturation is achieved by directly mixing the protein, buffer, and precipitant together into one supersaturated, homogeneous solution [32]. This solution is then left sealed and undisturbed so that crystal growth can occur [32]. Because the solution is left undisturbed, it is during the solution preparation phase, when the protein and precipitant solutions first come into contact that nucleation must occur, as after the mixing step, the solution should be within the metastable zone [32].

The dialysis method is another method used for crystallisation. Dialysis utilises a semi-permeable membrane to cause diffusion and equilibration of a precipitant concentration, allowing a slow and steady change in concentration until crystallisation conditions are achieved. A droplet containing the protein solution is placed on what is referred to as a button, and it is then sealed with the dialysis membrane. The button is then immersed in a reservoir containing the precipitant solution, and diffusion and equilibration of the precipitant across the membrane occurs. Diffusion across the dialysis membrane necessitates that the precipitant be a small molecule, such as a salt or alcohol. Dialysis is advantageous for crystallisation trials, as the button can be removed from reservoir and re-immersed into a reservoir with different conditions, allowing for the same protein sample to be reused until crystallisation conditions are achieved [32].

Lastly, the method known as FID, can also be used for protein crystallisation, although it's main advantage is in the crystallisation screening step. Utilising a "liquid-liquid" FID method, where the protein and precipitant solutions are layered on top of one another in a sealed capillary, and then left to mix. While difficult to setup and requiring large sample volumes, the creation of a physical manifestation of a phase diagram is produced. The concentration gradient created via this method allows produces a non-crystal solution area, as well as areas of crystals with varying sizes, which demonstrates all the various zones present within a phase diagram [30, 32].

With the majority of initial crystal screening occurring with robots, companies have created protein screening kits, where 96-well plates can be bought pre-filled with 96 different sets of conditions for testing [33]. Further optimisation after initial robotic screening is still required, as the conditions found with these screening tests may need slight modifications for optimal crystallisation. Depending on the size of the crystals and the amount produced, different X-ray facilities and delivery methods can be used to obtain diffraction data. At most synchrotron beamlines, crystals need to be a minimum of 5 μm , with some micro-focus beamlines capable of imaging crystals as small as 2 μm .

1.6.2 Conventional protein crystallography

Conventional protein crystallography generally involves taking a single crystal and flash freezing it for data collection. A whole data set can be generated from a single crystal that is rotated through the X-ray beam by 180° in order to obtain the full picture, although the degree of rotation is highly dependent on the crystal space group and for some, a 60° rotation is acceptable. In this case only basic spot finding is necessary, as it is easy to double check manually that all Bragg peaks have been identified and selected. Furthermore, because the starting angle and rotation angle of the crystal is known in conventional crystallography, two diffraction patterns from known angles (e.g. 0° and 90°) can be used to index the crystal, allowing any indexing ambiguities to be eliminated. Around the world synchrotron facilities have beamlines dedicated to these protein crystallography experiments, where data collection and structure analysis is a streamlined process that uses robotic systems to mount frozen crystals that have been pre-loaded into cassettes, as well as automated indexing programs. Remote access to these beamlines is also available so it is no longer required that an on-site presence is necessary, and therefore conventional crystallography has become a routine method.

1.7 X-ray sources

Original protein crystallography work was performed with lab-based X-ray diffraction, utilising X-ray tubes. These weak sources necessitated large, high quality crystals in order to obtain good diffraction patterns and often required weeks of data collection for a complete data set to be obtained. With the development of Synchrotrons and XFELs, the stringent crystal requirements required for successful lab based X-ray diffraction were overcome and further advancements in the field of protein crystallography were made.

1.7.1 Synchrotrons

While synchrotron radiation was first observed at General Electric in 1947, the first generation synchrotron light sources were developed in the 1970s for the primary purpose of

nuclear physics research, not for experiments utilising the radiation they produced. The synchrotron radiation observed was incoherent, broadband radiation produced by utilising simple bending magnets to guide the electrons. While not specifically designed for what we now think of as synchrotron research, scientists, with their parasitic usage of this type of incoherent radiation, were able to successfully perform the first dedicated synchrotron experiments (Fig. 1.8).

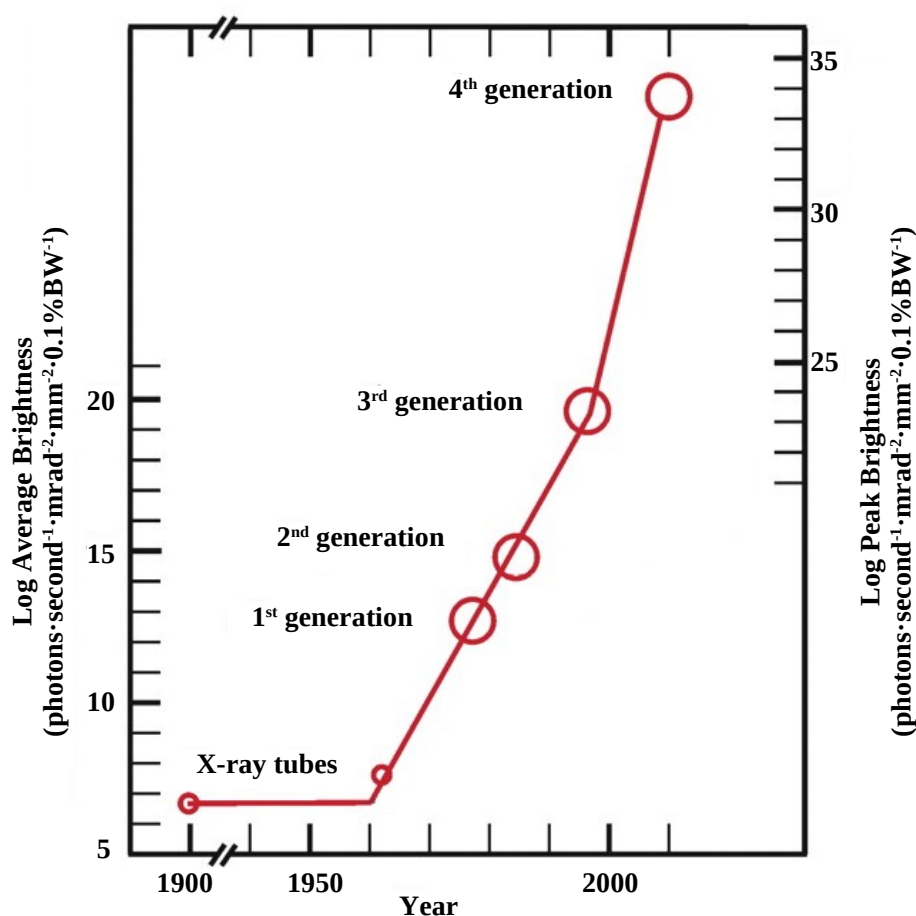


FIGURE 1.8: A plot showing the approximate year each new generation of X-ray producing light sources came online, as well as the Log Average Brightness and Log Peak Brightness of those light sources.

Second generation light sources were then created in the 1980s, with dedicated electron storage rings and with the primary purpose of producing synchrotron radiation. While these facilities still employed bending magnets and did not have the ability to produce the highly

coherent radiation that can be produced today, these dedicated light source were a step forward in the development of radiation facilities capable of protein crystallography.

Third generation light sources are what are considered the first synchrotrons that could produce highly coherent radiation due to the incorporation of insertion devices such as undulators and wigglers in the 1990s, with the European Synchrotron Radiation Facility (ESRF) being the first to open its doors in 1994. Synchrotron facilities considered third generation light sources are still in use today and include the ESRF, Australian Synchrotron, National Synchrotron Light Source (NSLS) II, Petra III, and Diamond Light Source (DLS). The main components of a third generation synchrotron include an electron gun, linear accelerator (linac), booster ring, storage ring and beamline (Fig. 1.9).

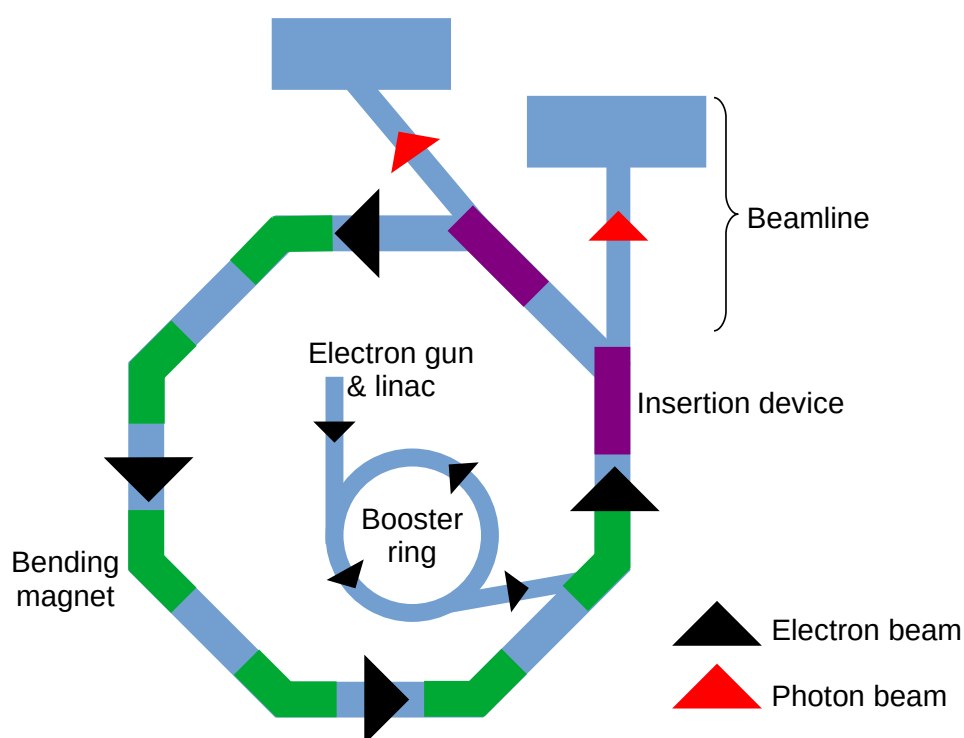


FIGURE 1.9: A simple diagram showing a typical third generation synchrotron. The main components of a third generation synchrotron include an electron gun and linear accelerator (linac), booster ring, storage ring composed of bending magnets and insertion devices, and the beamlines.

Synchrotrons utilise electrons from a source such as the tungsten element in an electron gun. They then enter the linear accelerator where the electrons are bunched and accelerated through a series of Radio Frequency (RF) systems. The electrons are then moved to an

evacuated booster ring which employs combined function electromagnets, used for steering and focusing, and RF cavities for further acceleration. The electrons are then injected into a storage ring composed of bending magnets, insertion devices which include wigglers, undulators, and RF systems.

Bending magnets (dipole magnets) are used to steer the electrons in a closed circular path, keeping them circulating within the storage ring. RF systems are composed of klystrons, waveguides and RF cavities. The RF system is what maintains the energy of the electrons within the storage ring, as the electrons are continuously losing energy as they expel synchrotron radiation. The klystron, which contains an electron gun that produces electrons, is then intersected by a low energy microwave signal. This microwave signal divides the electrons into bunches therefore producing a pulsed beam. This pulsed beam is then passed through the waveguide where a high energy microwave signal is generated. These high energy microwaves are then passed to the RF cavities where they interact with the electrons in the storage ring, increasing their energy. Quadrupole magnets are used for focusing the electrons and help to mitigate the Coulomb repulsion present between electrons. Sextupole magnets are then used to correct for any chromatic aberrations that may have occurred due to the quadrupole magnet focusing. The inclusion of insertion devices into third generation synchrotrons is how more brilliant synchrotron sources were developed. Insertion devices, which refer to both undulators and wigglers are arrays of dipole magnets that are set out in a periodic arrangement so as to generate an alternating static magnetic field (Fig. 1.10). It is this alternating magnetic field that causes the electrons to be deflected sinusoidally, causing them to emit the radiation which is then utilised by synchrotron beamlines. What distinguishes an undulator and wiggler, is not the physical devices, but what level of magnetic field is generated.

Wigglers, are high field devices used to increase flux levels and have a deflection parameter of $K \gg 1$, while undulators are lower field devices that are utilised to narrow the frequency of the radiation produced to a single frequency and its harmonics and has a deflection parameter of $K \ll 1$ where K can be determined by:

$$K = \frac{eB\lambda_u}{2\pi m_e c}, \quad (1.11)$$

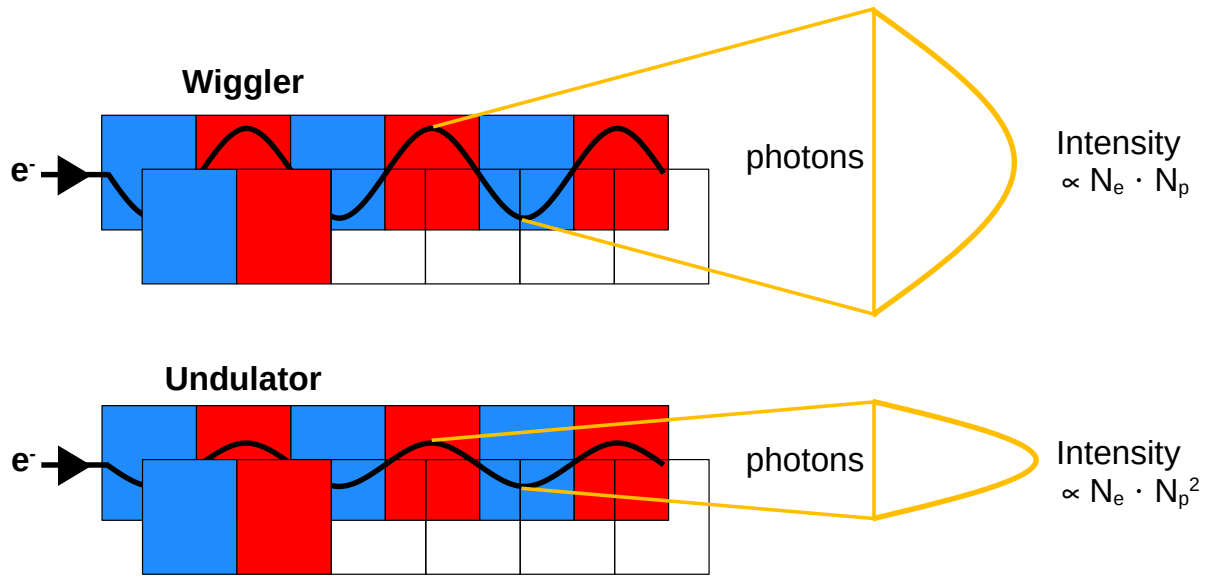


FIGURE 1.10: A diagram showing a wiggler and undulator, with an alternating static magnetic field (shown in red and blue), the electrons (e^-) entering the field, causing them to deflect and emit photons. The intensity from a wiggler is proportional to the number of electrons (N_e) multiplied by the number of poles (N_p), whereas the intensity from an undulator is proportional to the number of electrons (N_e) multiplied by the square of the number of poles (N_p^2).

where e is the electron charge, B is the magnetic field, λ_u is the spatial period of the magnets, m_e is the electron rest mass, and c is the speed of light.

Once the X-ray radiation has been produced it is focused via beamline optics located in the optics hutch before travelling to the experimental hutch for further micro-focusing and sample interaction. The basic optics components of the beamline include filters, slits, photon shutters, monochromators and mirrors. The filters are critical in determining a lower limit for the beamlines energy range, while the slits are used to adjust the beam size horizontally and vertically, as well as to eliminate any beam that has been scattered by the optics. Photon shutters are utilised in order to block the beam, so that scientists can still access the beamline and their experiments while the facility is operational. Monochromators are used to select for a narrow band of radiation, while X-ray mirror systems are used for a variety of purposes including as high-energy filters for elimination of unwanted higher order harmonics. The mirror systems are also used for focusing, with modern focusing mirrors e.g. Kirkpatrick-Baez (K-B), enabling micron or even sub-micron sized focal spots.

Synchrotron sources are now the crystallography workhorses, with hundreds operational around the world including the NSLS II in the US, and Petra III in Germany where I have undertaken experiments. These sources are continually upgraded to develop even more highly focused and intense X-ray beams. These synchrotron-based developments have led to advancements in the creation of micro and nano-focused beamlines, allowing for successful diffraction from ever smaller ($<2\text{ }\mu\text{m}$) crystals. Such beamlines include the ID23-2 beamline at the European Synchrotron Radiation Facility (ESRF) [25], the PXI beamline at the Swiss Light Source (SLS), the 17-ID-2 beamline at the National Synchrotron Light Source II (NSLS II) [26], and the P11 beamline at Petra III [27]. More recently these advances and developments have resulted in the birth of the next generation of X-ray sources, the X-ray Free Electron Laser (XFEL) (see subsection 1.7.2). Due to the small number of XFELs in the world, beamtime at these facilities is rare and highly prized, and therefore, improvements in synchrotron technology are still highly relevant.

1.7.2 XFELs

X-ray Free Electron Lasers (XFELs) are often described as a "fourth generation light source" that takes advantage of the phenomenon of Self-Amplified Spontaneous Emission (SASE) which results in the production of highly coherent and intense light. An XFEL consists of a linear accelerator in which electrons are accelerated through hundreds of resonators. These resonators contain oscillating microwaves that transfer energy to the electrons. The electrons are then passed through undulators causing them to emit radiation. Because the radiation emitted is travelling faster than the electrons, the radiation overtakes, interacts and speeds up or slows down the electrons resulting in microbunching (Fig. 1.11). The process of SASE allows these electron microbunches to emit light in-phase and produce extremely short intense pulses of radiation with an average brightness of 10^{25} photons radiated per second per solid unit angle, that otherwise wouldn't be possible.

The first XFEL, the Linear Coherent Light Source (LCLS), came online in 2009. As of 2021, there are five XFEL sources operating in the hard X-ray regime accepting users around the world. These facilities include the LCLS in the United States [35, 36], the Swiss

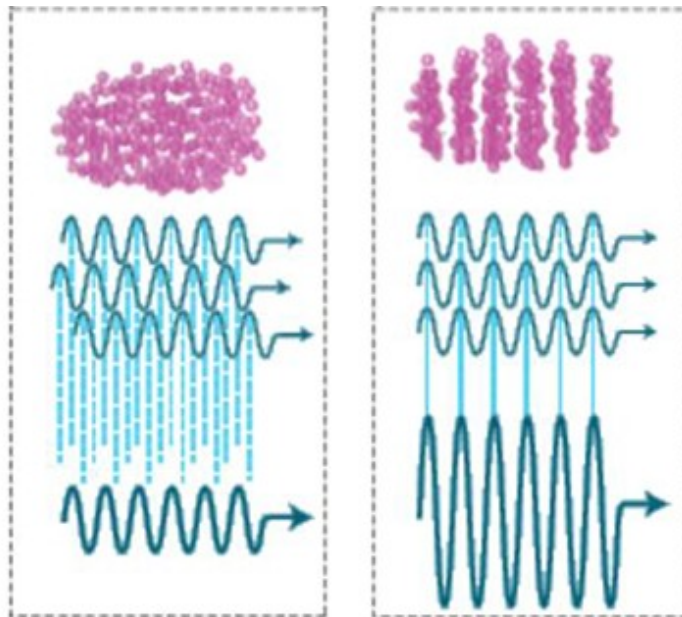


FIGURE 1.11: Self amplified spontaneous emission (SASE) occurs when electrons are passed through an undulator. Initially electrons are bunched together with random phases, such that their contributions sum incoherently. After going through an undulator, the electron phases align, therefore producing extremely short and intense pulses of radiation. This figure is reprinted from Zhukovsky 2016 [34]. CC BY 4.0.

Free Electron Laser (SwissFEL) in Switzerland [37], the Spring-8 Angstrom Compact Free Electron Laser (SACLA) in Japan [38], the Pohang Accelerator Laboratory X-ray Free Electron Laser (PAL-XFEL) in South Korea [39], and the European X-ray Free Electron Laser (EuXFEL) in Germany [40]. Other hard X-ray facilities in various stages of development include, LCLS-II in the United States, and the Shanghai High Repetition Rate XFEL and Extreme Light Facility (SHINE) in China. A comparison of these facilities is provided in Table 1.1.

Since the establishment of the first XFEL in 2009, further developments in XFEL science [41–44] have meant that ever shorter pulses can be produced, as is now possible at the EuXFEL in Germany, where first user experiments performed in September, 2017 [40, 45, 46]. At the EuXFEL, the resonators are made from niobium, which is cooled to -271°C allowing superconductivity and therefore nearly all the energy can be transferred to the

TABLE 1.1: A comparison of current and future Hard X-ray Facilities

Facility	First Users	Accelerator Technology ²	Pulses Per Second	Minimum Wavelength (nm)	Maximum Electron Energy (GeV)	Peak Brilliance
LCLS	2009	NC	120	0.15	14.3	2×10^{33}
SACLA	2011	NC	60	0.08	8.5	1×10^{33}
PAL-XFEL	2017	NC	60	0.06	10	1.3×10^{33}
EuXFEL	2017	SC	27,000	0.05	17.5	5×10^{33}
SwissFEL	2019	NC	100	0.1	5.8	1×10^{33}
LCLS-II, CuRF	2022 ¹	NC	120	0.05	15	2×10^{33}
LCLS-II, SCRF	2022 ¹	SC	1,000,000	0.25	5	1×10^{32}
SHINE	2025 ¹	SC	1,000,000	0.05	8	1×10^{33}

¹ Expected to begin operation

² NC = Normal Conducting, SC = Super Conducting

electrons [47], which allows for extremely short femtosecond pulses of high intensity radiation to be emitted.

With the advent of XFELs came the need to branch out from traditional crystallography methods, as single protein crystals could no longer be arbitrarily rotated through the beam. Thousands and thousands of crystals were instead required in order to obtain enough single crystal "snapshots" for structural determination. This method, called serial crystallography came with a whole new set of unique challenges, which included that these thousands of crystal snapshots had unknown orientations and that partial reflections were no longer avoidable. These issues, as well as the differences between conventional and serial crystallography will be discussed in Chapter 2.

Chapter 2

Serial Crystallography

Over the past few decades, protein crystallography has seen a migration from lab-based sources to synchrotrons, and now to XFELs, made possible by advancements in X-ray light sources that have allowed facilities to produce higher intensities, shorter exposure times, and faster repetition rates. The need to expose crystals to milliseconds of radiation with conventional crystallography at a synchrotron meant that there were limitations on how small a crystal could be, and still produce useful data before radiation damage impacted the data quality [48]. As early as 1986, it was suggested that if this diffraction data could be acquired with sufficient speed, that the issue of radiation dose limits could be overcome as the diffraction could be collected before damage became apparent [49].

This became a reality with the advent of XFELs, where femtosecond exposure times meant that diffraction data could be collected prior to radiation damage, using a technique termed "diffraction before destruction". With advanced XFEL sources now capable of detecting diffraction from micron-sized crystals before these crystals are damaged by radiation, structures that were previously "unsolvable" due to crystal size limits can now be solved [50–52].

The development of crystallography at XFELs came with its own set of challenges due to the increased source intensity. While the increase in intensity meant that higher resolution diffraction from smaller crystals was now possible, "diffraction before destruction" resulted in crystals being destroyed after a single pulse, and therefore the conventional method of

rotating a crystal through the beam to fully sample reciprocal space, was no longer viable. The new technique of Serial Crystallography (SX), or more specifically with reference to XFELs, Serial Femtosecond Crystallography (SFX) was therefore developed.

SX involves delivering a large amount of crystals to the X-ray interaction region and obtaining a very large number of single diffraction patterns, with one diffraction pattern collected per crystal. While conventional crystallography methods involve suspending a crystal on a static loop or other static rotatable system (see subsection 1.4), SX requires a continuous supply of crystals for diffraction and therefore new delivery systems had to be designed to accommodate for this. Two key modes of delivery have arisen from the development of SX; moving target systems, such as the Gas-focused Dynamic Virtual Nozzle (GDVN), Double Flow Focusing Nozzle (DFFN) and High Viscosity Injector (HVI), and fixed-target systems, such as the nylon loop, microfluidic chip, and conveyor belt system [53–55].

Improvements and innovations in sample delivery methods, optics, detectors, and data analysis pipelines due to the advent of XFELs and SX are consistently being published, allowing higher resolution structures, faster data collection speeds, and new techniques to be utilised (see section 2.4). While SX was initially developed as SFX and used exclusively at XFELs, serial crystallography can also be used for synchrotron experiments, where it is referred to as Serial Millisecond Crystallography (SMX), or Serial Synchrotron Crystallography (SSX). Advances in the field of SX, relevant to both SFX and SMX, are discussed in this chapter.

2.1 Sample delivery

In this section, the development of both moving target systems, and fixed-target systems for SX delivery will be discussed.

2.1.1 Moving target delivery systems

The first delivery system used for SX was the gas-focused dynamic virtual nozzle (GDVN), a moving target delivery system [56]. Still one of the most commonly used systems today, it can be used not only for SX experiments, but also for time-resolved SFX (tr-SFX) experiments. The GDVN design (Fig. 2.1) incorporates a capillary within a capillary. The inner capillary typically consists of a 50 or 100 μm diameter tapered nozzle and is the capillary that the sample is injected through to produce a sample stream. A constant flow of gas is produced in the outer capillary so that the gas surrounds the sample stream [54, 56]. Adjustments to the gas flow allows us to focus the sample stream into a narrower stream of approximately 1-5 μm in diameter.

Successful delivery of a narrow stream via the GDVN requires a minimum flow rate and velocity of 10 $\mu\text{L}/\text{min}$ and 10 m/s respectively, to maintain stability [54, 57]. This leads to significant sample consumption (ml) which can be an issue for samples that are hard to produce in large volumes. Furthermore, the design of the GDVN leads to clogging of the outer capillary as a result of either the sample dehydrating as it exits, or due to the crystals salting out of buffers. This clogging issue can lead to a loss of beamtime while the capillaries are cleaned or replaced.

To address the requirement for large sample volumes in this type of gas-focused delivery system, a newer system called the double-flow focusing nozzle (DFFN) system has been developed [58]. The DFFN works in a similar way to the GDVN, with one major difference - instead of one liquid output (GDVN), there are two liquid outputs (DFFN). Firstly, the sample liquid (primary liquid), inside the inner-most capillary, is focused by a sheath liquid (secondary liquid). The sheath liquid, generally a solvent, replaces some of the sample volume so that less sample is consumed. The sheath and sample liquids are then focused by a surrounding gas flow, which results in a very thin sample stream.

With this double-flow focusing method, sample streams with diameters of $< 1 \mu\text{m}$ can be produced (Fig. 2.1). Furthermore, utilising a solvent such as ethanol as the sheath liquid has been observed to minimise clogging of the crystals at the nozzle tip, as it acts as a barrier for the sample, preventing evaporation or salting out. With the advantages of the DFFN,

this newer method seems likely to quickly supplant the GDVN [53, 58].

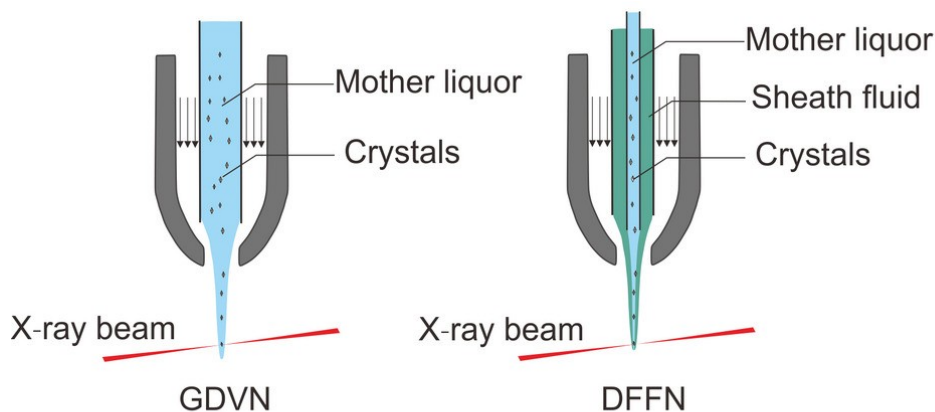


FIGURE 2.1: A diagram depicting a gas-focused dynamic virtual nozzle (GDVN) setup, and a double flow focusing nozzle (DFFN). The arrows on both images indicate the gas flow surrounding the capillary. This figure is reprinted with permission from Zhao 2019 [53]. Copyright 2019 Federation of European Biochemical Societies.

Another delivery method that has been developed for the purpose of minimising sample consumption is the High Viscosity Injector (HVI), which utilises highly viscous substances as sample delivery media [54]. Initially, the HVI was developed for membrane protein structural studies with Monoolein, which forms a Lipidic Cubic Phase (LCP), and was used as both the in-situ growth and delivery medium. LCP is a biomembrane mimetic, meaning that it imitates the conditions present at the cell membrane, and allows for a stable environment for membrane protein crystallisation [59].

HVI systems have become increasingly popular and can now be used with crystals that are grown in solution and mixed post-crystallisation with an HV medium for delivery. HVI systems have therefore expanded to include other HV substances such as mineral-oil based grease, petroleum jelly, and cellulose [60–64]. This method makes it possible to produce extremely slow flow rates (0.001–0.3 $\mu\text{L}/\text{min}$), decreasing the volume of crystals required for SX down to approximately 100 μL [54, 57, 64]. HVI is advantageous for use in synchrotron-based serial crystallography experiments [65], as well as low repetition rate XFELs, as it can produce a stable jet at much slower speeds than a GDVN or DFFN, which means that minimal sample is wasted [57, 64, 66–68].

One other moving target delivery system in development is the Drop-On-Demand (DOD)

system. Moving target systems consume much more sample than is necessary due to the need for the delivery of a stable and continuous stream of sample to the beam. Drop-on-demand aims to eliminate sample wastage by delivering sample to the beam intermittently at a frequency that is in-sync with the pulse repetition rate [69]. One specific method that falls under this drop-on-demand umbrella, is Acoustic Droplet Ejection (ADE) [70]. ADE works by using focused high frequency acoustic waves, propagated through the crystal solution, to eject picolitre and nanolitre droplet trains with a high temporal and spatial accuracy [53, 69, 71–73]. This delivery system has been used as a stand-alone delivery method, where sample droplets were delivered to the interaction region [74], and alongside the conveyor belt fixed-delivery system (described below), where droplets were placed on a conveyor belt system for enzyme catalysis studies [75].

2.1.2 Fixed-target delivery systems

The majority of developments in SX sample delivery have been focused on moving target delivery systems, largely because there is no need for sample alignment, a particular challenge when working with the smaller crystals utilised in SX. However fixed-target delivery systems have evolved significantly, with various systems available, and capable of being used for both synchrotrons and XFEL SX experiments [53, 76]. The nylon loop is one such system, and has been previously used for conventional crystallography, but now also has a place in SX. The crystal solution is suspended on the nylon loop, which is then mounted to a goniometer for raster or helical scanning of the crystals. Using helical scanning with long crystals allows for X-ray exposure to be optimally spaced in order to maximise the number of undamaged diffraction patterns that can be obtained per crystal [53].

The microfluidic chip is another fixed-target system whose design predominantly involves the use of micropores in thin films or silicon wafers. The crystal solution is either placed or directly grown on the device before the excess liquor is drawn off so that as much background signal as possible can be eliminated. Microfluidic chips are designed so that they can be attached to goniometer systems via brass pins, and magnetic bases (as in conventional crystallography setups) (Fig. 2.2) so that the chip can be translated horizontally and

vertically to obtain crystal diffraction [53]. This method allows for extremely low sample consumption, but requires more arduous sample preparation and mounting [53, 54].

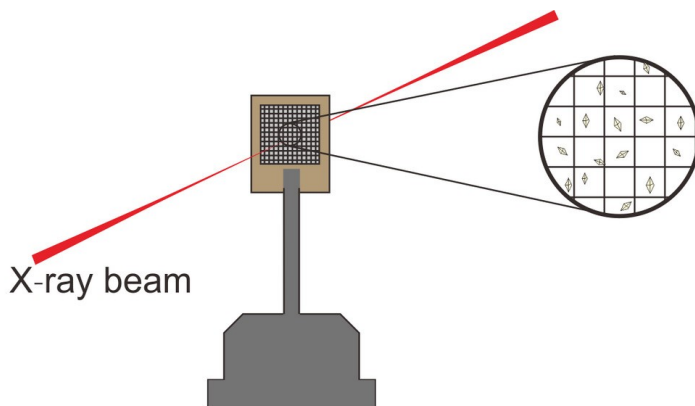


FIGURE 2.2: A diagram depicting the layout of a microfluidic chip fixed-target delivery system. The chip, with crystals sitting within the micropores, is attached to goniometer system. This figure is reprinted with permission from Zhao 2019 [53]. Copyright 2019 Federation of European Biochemical Societies.

The conveyor belt system is another fixed-target delivery method that is being used for SX. Conveyor belts, as the name suggests, involve crystals in solution being deposited on a Kapton film, either by spray or injection methods [53, 69, 73, 75, 77]. This method is mainly used for time-resolved studies [73, 77, 78], as it is overly complicated for standard SX [53, 69].

As SX delivery systems develop, the advantages and disadvantages of each system may change. For example, microfluidic chip designs were previously only capable of being used at synchrotron facilities, but with new research published by Roedig et al. [79], chip systems have now been successfully developed for use at XFELs. A useful decision tree and guide for what delivery system to choose, has been published in Cheng et al., in 2020 [69], and can be seen in Figure 2.3.

2.2 Data Collection & Detectors

SX development has also led to the development of new detector technology both at XFELs and synchrotrons. There is a multitude of different parameters that are required for optimal

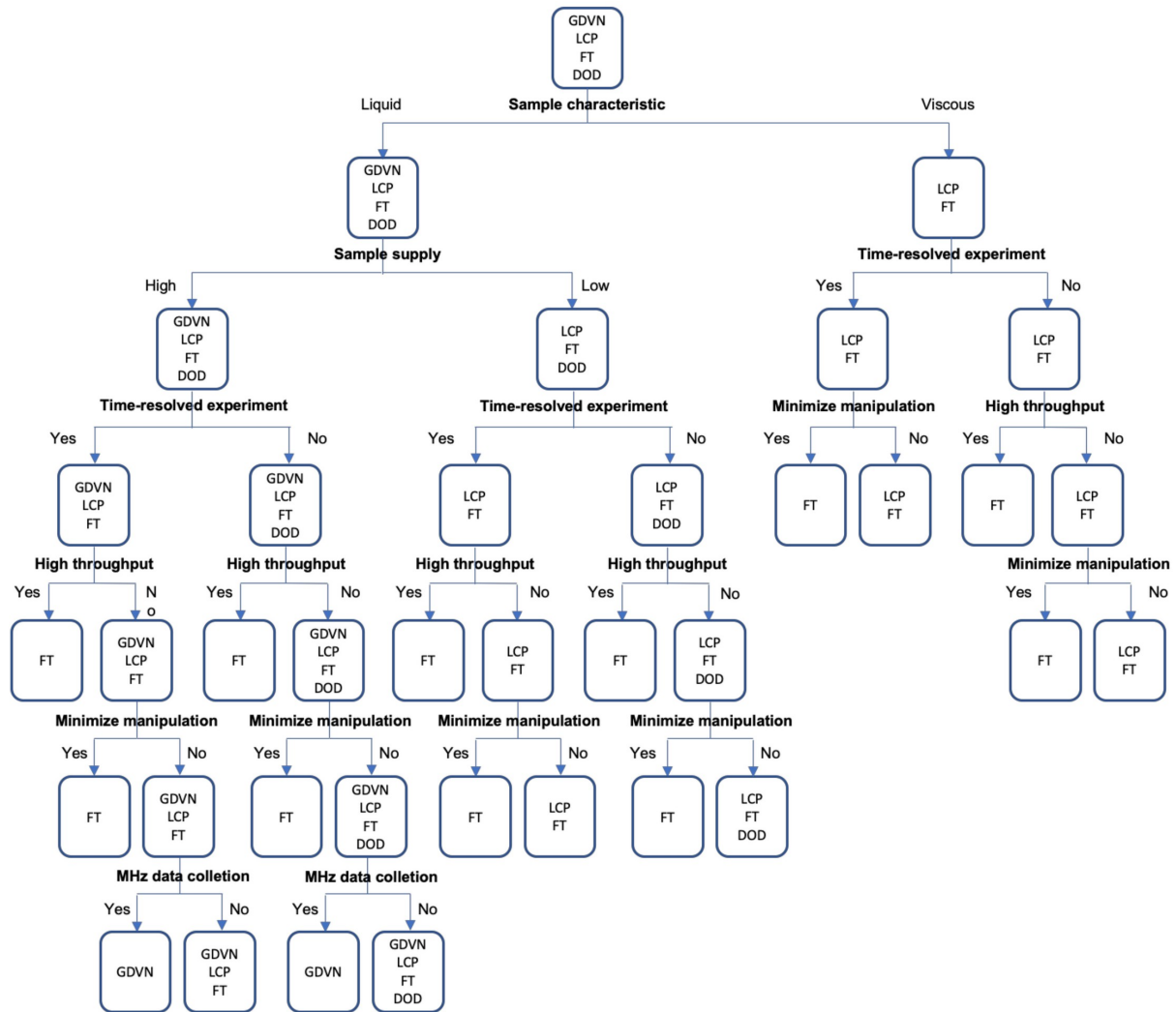


FIGURE 2.3: A decision tree showing which of the delivery methods listed would be best, when accounting for sample viscosity, sample supply, SX technique, sample stability (whether minimal handling is a priority), if it is high-throughput, and whether the experiment is being performed at MHz rates. The delivery methods include: GDVN = gas-focused dynamic virtual nozzle; LCP = lipidic cubic phase injector (more broadly HVI) ; FT = fixed target systems; and DOD = drop-on-demand. This figure is reprinted from Cheng 2020 [69]. CC BY 4.0.

detection and recording of diffraction patterns for SX. These parameters include photon sensitivity, at both low counts and high counts (dynamic range), and the ability to keep up with the repetition rate of the X-ray pulse during the internal storage phase (where information is saved in the detector) and in the read-out phase. The improvement in read-out speed is a key requirement for XFEL detector development, as it is the read-out speed that is often the limiting factor for data collection. For example XFEL pulse repetition

rates operate between 120 Hz to 4.5 MHz, and therefore the detector read-out speeds had to cope with these fast repetition rates. The first detector capable of these XFEL repetition rates was the Cornell-SLAC Pixel Array Detector (CSPAD) at the LCLS [80]. This was specifically developed so that the detector could match up to the 120 Hz source repetition rate. While solely focusing on improving the read-out rate capabilities of detectors has been sufficient in the past, the pulse-train repetition rates now available at some XFELs, such as the EuXFEL with its 4.5 MHz pulse repetition rate, the development of a second storage step was necessary to ensure all data could be successfully read-out. This second storage step involves each pixel in the detector having 352 analogue memory cells for internal storage, before the data needs to be read out, this means that the data can be successfully read-out in between the trains. This necessary multi-step storage was incorporated into the design of the AGIPD, the detector specifically developed for the EuXFEL [81].

As well as the read-out time of the detector, the photon sensitivity of the pixels, can also be a major bottleneck to the success of an experiment. A high dynamic range for photon sensitivity can greatly impact the quality of experimental data, allowing for the low intensity data to be collected, as well as maintaining variation in high intensity data without over-saturation of the detector occurring. The advances in the detector technology has helped drive the development of SFX and tr-SFX, and also SMX, with detector development for synchrotrons including the Pilatus and Eiger series of detectors produced by DECTRIS.

2.3 Data Analysis

SX data analysis follows the same steps as conventional crystallography, but with slight changes that will be discussed in the sections below.

2.3.1 Hit Finding

After data collection, the first challenge is the identification of Bragg peaks on the detector images, referred to as "spot finding" in conventional crystallography, or "hit finding" in SX. In conventional crystallography, only a small number of diffraction patterns are required to

fully sample reciprocal space. Using SX, where large numbers of diffraction patterns are necessary to fully sample reciprocal space, basic hit finding algorithms, and manual peak selections are no longer viable. Also, due to the way the data is collected, the issue of partial reflections is significant. The different requirements for SX have led to a large push towards developing new algorithms for SX, which are outlined in the following sections. Hit finding is the first step in the long data analysis pipeline for the structural determination of a protein. After the initial X-ray diffraction experiment has been completed and detector images have been obtained, it is vital that Bragg peaks are identified and their position accurately determined. Identifying Bragg peaks in a detector image is what is referred to as "hit finding". Due to how the crystals are situated within the sample stream, not every sample and X-ray interaction will be classified as a "hit" that contains Bragg peaks. This is because there is no way to align a crystal within the stream to ensure that a crystal is hit every time. Therefore, much of the data collected is classified as "non-hits", with hit rates ranging from 4.5 - 50% for GDVN experiments, 0.45 - 13% for LCP experiments and 50 - 100% for fixed target methods [72]. Classification of data has led to the development of a multitude of new hit finding algorithms that aim to improve the quality of peak identification in detector images that have: low Signal-to-Noise Ratio (SNR) issues; radial variations in SNR ; uneven SNR distributions across the detector modules; and image-to-image changes in SNR that occur during experiments.

Automated hit finding is an essential step in SX as current experiments can obtain hundreds of thousands of images that need to be analysed for Bragg peaks. Original peak-search algorithms from the 1980s, included basic peak finding that found peaks based solely on whether a pixel, located at the coordinates (R, S) was of a higher value than those pixels located at coordinates $(R+d, S)$, $(R-d, S)$, $(R, S-d)$ and $(R, S+d)$, where d was a user selected value of raster steps [82]. From this original search algorithm, a multitude of peak finding algorithms have been created that deal with the various problems that arise from variations in background noise within an image and background variation between detector images. Popular hit finding algorithms include Zaef [83], PeakFinder8 [84], PeakFinder9 [85], and most recently RPF [86]. These hit finding algorithms are discussed in detail below.

Zaef peak finding algorithm There are many different hit finding algorithms available within CrystFEL, a crystallography software suite, one of which is called "Zaef" named after Stefan Zaefferer, whose research the algorithm is based on [83]. When utilising Zaef as the hit finding algorithm, a gradient search procedure is used in order to find what is referred to as a "foot point" or outer edge of a peak within the diffraction pattern. The gradient:

$$\left[\frac{\partial i(x, y)}{\partial x} \right]^2 + \left[\frac{\partial i(x, y)}{\partial y} \right]^2, \quad (2.1)$$

where $i(x, y)$ is the image intensity at position (x, y) , is checked at each point within the detector image against a predefined minimum gradient. If the gradient found is larger than this minimum, a possible peak edge or "foot point" has been found. From the foot points, a 5×5 pixel mask is applied and successively moved until centred on the assumed maximum intensity pixel.

The distance between this maximum intensity point and the foot point is then compared to a predefined radius and if it is within that radius, a possible peak has been found. A second selection criteria is added on top of the original work by Zaefferer, where no two peaks can be identified within two times the predefined radius. This extra criteria allows for filtering of image artefacts or Bragg peaks that may have an elongated shape that would otherwise have been predicted to be two individual peaks. While this method can be optimised for some detector images, it can be insufficient for images with poor SNR or low peak signal intensity [83]. This hit finding algorithm, while quite simplistic compared to other hit finding algorithms available, can still work well with diffraction data that has minimal variation in the SNR, and because of its simplicity is faster to run, requiring less computing power.

Peakfinder8 & Peakfinder9 Peakfinder8 is another hit finding algorithm that has subsequently been incorporated into CrystFEL. Originally developed as the hit finding algorithm within Cheetah, an SX hit finding, indexing and integration program [84], it works by computing and outputting two arrays; one for the average intensity $\mu(r)$ and the other for the standard deviation $\sigma(r)$, both as a function of radius (r). It then computes the radius dependent thresholds ($\text{thresh}(r)$):

$$\text{thresh}(r) = \mu(r) + \text{SNR} * \sigma(r), \quad (2.2)$$

where a minimum SNR is predefined by the user. This computation is looped 5 times with each iteration excluding pixels that are above that iteration's threshold. This allows the radial average and standard deviations to exclude outliers such as Bragg peaks, therefore resulting in a more accurate background-only radial average and standard deviation. The final threshold ($\text{thresh}(r)$) array is then filtered against a predefined minimum Analogue-to-Digital-Converter (ADC) criterion:

$$\text{thresh}(r) = \text{MAX}(\text{thresh}(r), \text{ADC}_{\min}). \quad (2.3)$$

Each detector module is then scanned for regions that have a predefined number of connected pixels that are all above the computed threshold, outputting a list of potential peaks. Each potential peak is then integrated to find its centre position and ADC_{\max} value.

A cross-check using a three-ring local background approach is then performed on the potential peaks to ensure that the peaks comply with the predefined minimum SNR and have not been falsely identified due to a noisy region. Lastly, an optional filter can be applied to exclude the weaker of two peaks that are within a predefined distance of one another [84].

While Peakfinder8 takes into account background signal variation that occurs radially with approximate circular symmetry, the algorithm can't handle more localised background variation such as variation due to shadowing. Furthermore, the background threshold estimation utilised in Peakfinder8 is computationally intensive due to the iterative approach it adopts. With these issues in mind, an improved version called Peakfinder9 has been developed [85].

Peakfinder9 differs from PeakFinder8 by the way in which it estimates the background noise. The background is estimated locally based on border pixels within a specific radius, and includes the following steps:

1. The maximum intensity pixel identified as a peak ($I_{x,y}$) must be of a higher intensity than the border pixels (I_{Border}), in addition to a constant (c_1) for increased accuracy:

$$I_{x,y} > I_{\text{Border}} + c_1. \quad (2.4)$$

2. The maximum intensity pixel must also be larger than any of it's neighbouring pixels:

$$I_{x,y} > I_{x\pm 1, y\pm 1}. \quad (2.5)$$

3. As is the case with PeakFinder8, the maximum intensity pixel must be above a threshold calculated using the average intensity μ and the standard deviation σ of the border pixels:

$$I_{x,y} > thresh(r). \quad (2.6)$$

4. All pixels connected to the peak pixel and above the threshold, are summed together and are required to be above a second threshold:

$$\sum I_{a,b} > thresh_2(r). \quad (2.7)$$

Peakfinder8 allows much better background estimation than original hit finding algorithms such as Zaef, due to the ability to calculate the radial background variation. Peakfinder9 improves on background estimates even further by employing a localised approach to estimating noise, allowing local effects such as background artefacts and shadowing to be taken into account, and allowing the issue of radial variation to be bypassed.

RPF A new hit finding algorithm called Robust Peak Finding (RPF), developed by Marjan Hadian-Jazi [86], also employs a local background signal approach by utilising robust statistics. The use of robust statistics allows for better peak identification in images with low or varying SNR, while also severely limiting the number of false positive peaks that are identified.

RPF works by first dividing the detector image into small circular or quadrilateral patches for individual searches. One at a time, in order of descending intensity, each pixel within a patch is selected and a window around that "main" pixel is set. Within this window, a Gaussian profile is fit and the median of the data is found. A scale parameter is calculated using the Gaussian information and utilising Modified Selective Statistical Estimation (MSSE). MSSE is advantageous for peak finding as it assumes that some of the data points will always belong to the peak under investigation.

From the MSSE output, a peak threshold is estimated and is then used to determine the first iteration of the patch threshold, which is originally set to 0. The Peak-to-Average-Power-Ratio (PAPR) is then calculated. If the PAPR is considered acceptable, pixels above the threshold that are adjacent to the main pixel or pixels that are adjacent to those "main-adjacent" pixels previously identified are recorded. The SNR is then calculated for these pixels and if it is acceptable, their peak info is stored. Utilising both SNR and PAPR as statistical quality measures means that we can have much finer control over what is counted as a peak. PAPR differs from SNR as it is a measure of the degree the main pixel differs from the rest of the data within the window, whereas SNR is a measure of how much all pixels considered as a "peak" differ from the rest of the data within the window.

If either the PAPR or SNR are not considered acceptable, the pixel under investigation is masked from the list and the next pixel above the patch threshold is investigated as above. When no pixels above the patch threshold remain, the algorithm moves onto the next patch. When no patches remain, the algorithm lists all peaks found.

A key advantage of this hit finding algorithm is that it is not reliant on the geometry constraints of the detector, and therefore is capable of performing hit finding on all the "patches" within a detector image at the same time, running the patch search in parallel for faster hit finding detection. The SNR is also calculated independently between images and therefore allows for more accurate detection of the peaks compared to hit finding algorithms that utilise a single background noise threshold for the whole data set.

2.3.2 Indexing

After hit finding has identified all Bragg peaks, the diffraction patterns are filtered so that only those with a predefined minimum number of peaks are passed on to the indexing stage. It is this indexing stage that a major difference is seen between conventional crystallography and SX. In conventional crystallography, the orientation of the crystal within the X-ray beam is known relative to a starting orientation. Therefore, these extra known crystal orientations can be used as back-up, confirming that the correct unit cell has been selected and eliminating any indexing ambiguity that may have arisen from determining the unit cell based solely

on the first diffraction pattern. However, in SX the crystal orientations are all unknown, therefore a second diffraction pattern with a known orientation relative to the first cannot be used to solve any indexing ambiguities. This means that for a unit cell to be correctly identified with SX, more Bragg peaks per pattern are required so that there are no indexing ambiguities present.

Whether for conventional crystallography or SX, the indexing process involves calculating the lattice vectors from the Bragg peaks identified (previously described in section 1.2), until an orientation is correctly assigned. The specifics of how this is done varies between indexing methods, and descriptions of this process can be found within the specific indexing methods described below. Generally, from the Bragg peaks identified, their 2D detector coordinates are projected onto 3D reciprocal space vectors. These vectors are then utilised in slightly different ways depending on which indexing algorithm is selected. Some algorithms work within reciprocal space, while other algorithms utilise a Fourier transform in order to work in real space. The indexing stage is extremely important in determining the correct unit cell via accurate assignment of Miller indices to the Bragg peaks. The need for correct identification is necessary so that in the next step - Integration - the intensities can be accurately determined, and this accuracy is essential in order for the electron density of the lattice to accurately model the shape of the protein structure.

Unlike in conventional crystallography, where a single crystal is imaged at a time, multiple crystals can be present within the interaction region of the beam with SX and therefore can be an issue for conventional indexing programs. Various indexing methods have been developed to address these issues, utilising different strategies to sort or filter which of the reflections within an image belong to which crystal (see specific indexing methods for details). Many of the indexing methods available require a minimum number of peaks to successfully index an image. For most indexing methods, this minimum number of peaks is approximately 20, but some newer indexing methods can successfully index with as little as 5 peaks per image, which is promising as fewer Bragg peaks are obtained when performing SX on small crystals, particularly when using a synchrotron source.

As mentioned previously, many indexing methods have been developed, all having advantages and disadvantages depending on the type of X-ray diffraction data being analysed,

whether the data has a limited number of peaks per image, or if there are multiple crystals diffracting within a single image. Original methods such as MOSFLM [87], DirAx [88], and XDS [89, 90], and newer methods such as *felix* [91], *taketwo* [92], XGANDALF [93], and SPIND [94] have different starting requirements, differing advantages, and therefore may be of use for either initial indexing or secondary indexing stages.

MOSFLM Published under the paper titled "The Rossmann Fourier autoindexing algorithm in MOSFLM", this indexing method was originally designed to analyse diffraction patterns captured on mosaic film [87]. An indexing method that is based on the 1D Fast Fourier Transform (FFT) approach, this approach initially converts the 2D detector coordinates obtained directly from the detector image into 3D reciprocal space coordinates, where the coordinates are known to be on the surface of the Ewald sphere, otherwise the peak would not exist.

These 3D reciprocal space coordinates are then converted into reciprocal space vectors, where the vector origin is at the reciprocal space origin, and vector end is at the 3D reciprocal space coordinates of the peaks. MOSFLM then creates ~ 7300 vectors originating from the reciprocal space origin and ending at ~ 7300 periodically spaced points on the northern hemisphere of the Ewald sphere. These vectors are then used to create ~ 7300 1D projections of the reciprocal space vectors, which are plotted on histograms, where the x axis is the magnitude of the 1D reciprocal space vector projections \AA^{-1} , and the counts are how many of these vector projections are contributing. Fast Fourier Transforms are then performed on each histogram to obtain the direct lattice spacing $d(\text{\AA})$.

From these ~ 7300 FFTs, the 30 with the largest maxima then have their vector parameters further refined until its maximum value is obtained.

Combinations of three vectors out of the set of 30 are then used to obtain basis sets for the direct space primitive unit cell. These basis sets give the $[\mathbf{A}]$ or $[\mathbf{UB}]$ matrix.

The different combinations of basis sets are then used to determine which set has the miller indices with the lowest deviation from integer values:

$$\vec{h} = [\mathbf{A}]^{-1}[\Phi]^{-1}\vec{x}, \quad (2.8)$$

where $[\mathbf{A}]$ is the 9 components of the direct space basis vectors $(\vec{a}_1, \vec{a}_2, \vec{a}_3)$, \vec{h} is the approximate Miller indices for a peak with reciprocal space lattice constants \vec{x} , and $[\Phi]$ is the matrix of rotation of the crystal. The different Bravais lattice routines are then called, and a solution is found that takes into account any higher order symmetry present.

As can be surmised from the description of this method, the fewer peaks present in the diffraction pattern, the more ambiguity there is when determining the correct cell parameters. This is even more evident in SX compared to conventional crystallography, as an incorrectly indexed pattern cannot be checked or corrected against another pattern where the orientation is known. Therefore the limiting part of this methodology is the 1D Fourier transform projection that requires approximately 20 Bragg peaks within a diffraction pattern for favourable indexing results. An alternative indexing method is DirAx, which also relies on 1D Fourier transforms but with a slightly different approach.

DirAx Published under the paper titled "Indexing in single-crystal diffractometry with an obstinate list of reflections", DirAx indexing unlike MOSFLM doesn't use 7300 randomly chosen projections as its starting point for obtaining the 1D projections of the reciprocal space vectors [88]. Instead, it iterates through all the 3D reciprocal space coordinate combinations of three (including the origin coordinate) and determines the direct vector (the normal to the triplet plane). If all three points selected were from the same reciprocal lattice, the vector normal will be a direct lattice vector and a periodic signal will be apparent on the histogram. Each of the line projections is checked for which projection has the shortest period and largest maximum, as this corresponds to the reciprocal distance vector \vec{d}^* . A direct lattice vector \vec{t} is obtained from this, where \vec{t} is in the direction of \vec{d}^* , and has length of $\frac{1}{|\vec{d}^*|}$. Each potential direct lattice vector (\vec{t}) is saved alongside the number of reflections that contributed to the largest maximum (n_F). All the potential direct lattice vectors are then reduced to a unique set, where those with the same length and the same or opposite vector direction are averaged. The three shortest potential direct lattice vectors with the highest n_F are then chosen as the direct lattice basis set $(\vec{a}_1, \vec{a}_2, \vec{a}_3)$

Based on this basis set, all identified peaks in the diffraction pattern are assigned their Miller indices, and those that are within a specified error range of an integer value are

accepted as belonging to the crystal, while those outside of the error range are eliminated from further analysis. As with MOSFLM, the different Bravais lattice routines are called, and a solution is found that takes into account any higher order symmetry present.

XGANDALF E(X)tended (G)r(A)die(N)t (D)escent (A)lgorithm for (L)attice (F)inding, or XGANDALF [93], unlike some of the older indexing methods was specifically developed for SX data and its specific challenges in mind. Diffraction patterns that include only small numbers of Bragg peaks, large background noise, or multiple crystals diffracting in a single image can all lead to difficulties in indexing that many of the indexing methods initially developed for conventional crystallography can struggle with. Unlike many of the other indexing methods that have been developed specifically for SX, XGANDALF does not require any prior unit cell information. Based on the idea of a Fourier based indexing algorithm, this method employs a similar, but non-Fourier transform approach that utilises periodic basis functions and a non-linear weighting scheme and allows for increased noise tolerance and improved hit finding [93].

FELIX FELIX is an algorithm designed specifically for indexing multiple crystals in X-ray free-electron laser snapshot detector images [91]. Due to the delivery methods and micron-sized crystals utilised in SX, overlapping of crystals within the beam interaction region can frequently occur. Therefore it is important that indexing methods develop specific strategies that allow them to successfully extract information from these multi-crystal diffraction patterns and FELIX addresses this issue. Whereas some indexing methods require no prior cell information, the FELIX algorithm has been designed to untangle the problem of multi-crystal diffraction data and therefore requires prior knowledge of the unit cell as a starting point for indexing. Because this indexing method relies on prior unit cell information, this algorithm while not useful as a stand-alone indexing method, is extremely useful as a secondary indexing step so that more information can be extracted from the SX data set.

FELIX describes multi-crystal diffraction data in a 3D misorientation space called Rodriguez–Frank (RF) space. Using this misorientation space, FELIX can theoretically identify more than ten crystals per image. Where other indexing methods have trouble indexing

diffraction patterns where hkl families are closely positioned or overlapping, FELIX can untangle these reflections. Due to the increased number of crystals that can be indexed in a single image, crystal concentrations do not need to be limited for the sake of indexing. The ability to utilise higher crystal concentrations also means that there will be less empty frames, meaning a higher hit rate can be achieved. Therefore fewer images, and shorter beamtimes are required to achieve the same data quality compared to other indexing methods in which only one crystal is indexed per image.

TakeTwo TakeTwo is an indexing method specifically designed to tackle the near-2D reciprocal space information obtained from SX experiments [92]. It is described as near-2D, as the only 3D information that is obtained from these snapshots is due to the slight curvature of the Ewald sphere, which is essentially zero at low resolutions compared to conventional crystallography experiments where the crystal is rotated, therefore providing 3D information [92]. This indexing algorithm requires the unit cell dimensions and space group, and therefore can be used advantageously as a secondary indexer, like that of the FELIX algorithm.

SPIND Sparse-Pattern Indexing (SPIND), is an indexing method developed for diffraction patterns which contain a low number of Bragg peaks [94]. SPIND can successfully index diffraction patterns with as few as five Bragg peaks, but does require prior unit cell information and so therefore, like FELIX and TakeTwo, is useful as a secondary indexer utilised for extracting further information out of a data set [94]. This method can significantly help in reducing the amount of sample and beamtime required by better utilisation of diffraction data, particularly in circumstances where there are enough diffraction patterns to determine a unit cell, but where the detector images obtained are predominantly images with a low number of Bragg peaks [94].

2.3.3 Integration

As mentioned in the indexing step, it is vital to accurately determine the unit cell and orientation in order to assign the correct miller indices to each peak position. This is a necessity because the Miller indices of each peak position must be known in order to integrate the

intensities at all peak positions. Integration involves estimating the intensity at 2D detector coordinates that have been identified or predicted as Bragg peaks. The intensity estimate includes a background intensity subtraction [95]. Integration occurs at each identified and predicted Bragg peak coordinate that is related to the successfully indexed unit cell. This allows the integrated intensities for both the strong, weak, and absent reflections - all equally important - to be obtained and therefore relies on the accurate determination of the unit cell.

The integrated intensities from all images are then merged and scaled so that their intensities are normalised. This is an important step in serial crystallography, due to the random orientation of the crystals, both partial and full reflections are collected. In order to correctly merge data that includes partial peaks, the data is averaged to obtain accurate peak intensities, this means that large redundancies are needed to obtain the most accurate results. These merged intensities allow us to calculate the first information about the electron density of the lattice. Because phase information is lost in the process of X-ray diffraction, the integration step only allows for a segment of the information to be obtained. Further steps are required in order to deal with the phase problem in order to accurately determine the complete structure. These further steps can involve heavy atom replacement (heavy atom method) or, if a similar structure is already known, the method of molecular replacement (MR) can be used. Additional methods are outlined in subsection 1.5.7.

2.3.4 Phasing Methods

The phase problem (described in Chapter 1.5.7) can be overcome via various methods that include single isomorphous replacement (SIR) [16, 17], single isomorphous replacement with anomalous scattering (SIRAS), multi-isomorphous replacement (MIR), multi-isomorphous replacement with anomalous scattering (MIRAS), single wavelength anomalous diffraction (SAD) [18, 19], multi-wavelength anomalous diffraction (MAD) [20] and Molecular Replacement (MR) [14]. These methods can also be further improved via density modification methods [21–24] such as solvent flattening, histogram matching and non-crystallographic averaging. The most common method of phasing for serial crystallography has been MR,

but this requires a known homologue as a starting model, and therefore cannot be applied to novel structures that have no known homologues. In these cases, the use of isomorphous replacement and anomalous diffraction methods have been successfully performed [96, 97]. These more commonly used methods are described in more detail below.

Isomorphous Replacement Methods

Heavy atom isomorphous replacement methods work by substituting heavy atoms into the structure being investigated, all the while maintaining the same space group, unit cell dimensions and atomic positioning. This is achieved by the use of heavy atoms of a similar size to the atom being substituted, so that the impact to the structure is as minimal as possible. In the scheme of heavy atom replacement, a heavy atom is one that has more electrons when compared to atoms that are naturally found in proteins, such as hydrogen, carbon, nitrogen, oxygen and sulfur [98].

The first experiments that utilised isomorphous replacement, specifically with proteins were in 1956 and 1958, and involved protein crystals that were soaked in heavy atom solutions in order to create isomorphous crystals [4, 99]. After the process of creating isomorphous crystals, X-ray diffraction is performed on both the native crystal, and the heavy atom derived isomorphous crystals. The measurable intensity differences between the crystals are then used to determine the positions of the heavy atoms.

For the case of single isomorphous replacement (SIR), a single heavy atom element is substituted into the crystal structure. The differences in the amplitudes of a reflection in the native crystal ($|F_P|$), compared to the heavy atom derived crystal ($|F_{PH}|$) can be used as an estimate of the heavy atom structure factor amplitude ($|F_H|$):

$$|F_H| \simeq |F_{PH}| - |F_P|. \quad (2.9)$$

From this estimate, the heavy atom positions can be determined using either the Patterson map or direct methods. From the heavy atom positions determined, a more accurate $|F_H|$ can be calculated along with the corresponding phase (α_H). Using the cosine rule, the native crystal phase (α_P) can then be calculated, but due to the centrosymmetric nature, multiple solutions are present. Furthermore, measurement errors resulting in inaccurate

structure factor estimates and inaccurate heavy atom positions, mean that it is seldom simple to determine the corresponding phase. This phase ambiguity, if present in a small crystal, may be solved via phase probability, where the assumption is made that the calculated and refined F_{PH} contains all errors, and that those errors follow a Gaussian distribution [16].

Proteins are complex molecules containing thousands of atoms, and therefore in order to obtain the structure of proteins via isomorphous replacement, more than one derivative is usually necessary [100–104]. The multi-isomorphous replacement (MIR) method utilises multiple derivatives of the native crystal, which means that ambiguities in phases may be eliminated. Even then, the use of phase enhancement via density modification (see paragraph below titled Density Modification) is usually needed in order to increase the accuracy to a point that allows for high quality interpretable electron density maps to be produced.

Anomalous Diffraction Methods

Another method utilised to overcome the phase problem is to utilise anomalous diffraction. Anomalous scattering, also referred to as resonant scattering is when the energy of the incident X-ray is similar to the energy required to induce an excited state in the chosen atom, therefore causing strong X-ray absorbance. When this occurs a rapid change is observed in the phase of the diffracted X-ray that is unique compared to the other elements present within the crystal. Carbon, nitrogen, oxygen, and hydrogen atoms are not contributors to anomalous scattering at the wavelengths used for X-ray crystallography, so in order to obtain anomalous scattering a heavier atom must be present, either naturally, or via isomorphous replacement techniques [97, 101, 103–105].

Similar to the requirement for multiple crystal derivatives to solve for the phase of proteins using isomorphous replacement techniques, single wavelength anomalous diffraction (SAD) without isomorphous replacement, is also lacking enough information to fully solve any phase ambiguity if no heavy atoms are naturally present within the structure, and so multiple wavelength anomalous diffraction (MAD) is required [97, 101, 103–106]. These wavelengths are chosen based on the absorption edge of heavier elements present within the protein crystal, and therefore the protein under investigation must have enough heavy atoms to directly utilise this method [101]. If this is not the case, a combination of isomorphous

replacement and anomalous diffraction techniques can be utilised.

While elements with atomic numbers of roughly 20 to 40, and >60 , have accessible absorption edges [107], recently native sulfur and chlorine atoms have been successfully used with MAD, for successful phasing leading to the structural determination of lysozyme via serial femtosecond crystallography [108].

Isomorphous Replacement with Anomalous Scattering It can prove difficult to produce enough stable isomorphous crystal derivatives for use with MIR techniques, and therefore a combination of isomorphous replacement and anomalous diffraction can be performed. SIR or MIR, in combination with SAD or MAD (SIRAS/MIRAS) allows us to limit the need to produce stable derivatives, while allowing proteins that may not have enough naturally occurring heavy atoms for use with MAD, to still be successfully phased. The different wavelengths chosen are determined based on the heavy atom replacements used, where a wavelength close to the absorption edge of that atom is utilised in order to maximise anomalous scattering [101–104].

Molecular Replacement

Molecular replacement (MR) is a widely used phasing method, but unlike the other phasing methods described, MR requires a priori information in the form of a structurally known homologue. The homologue, referred to as the 'starting model' provides a good starting point for the phasing of the X-ray data. To be used as a starting model, the homologue usually requires a sequence identity of $>25\%$ in common with the structure under investigation, as well as $<2.0 \text{ \AA}$ r.m.s. deviation with respect to the $C\alpha$ atoms [109].

The first steps involve the calculation of two Patterson maps via the Patterson function (Eq. 1.8), one from the intensities of the experimental data obtained, and one from the atomic coordinates of the starting model. A Patterson map is a self-convolution of the electron density, and has peaks at inter-atomic vectors, not absolute atomic positions [109]. After the two Patterson maps are calculated, they are aligned (rotational and translational alignment) with one another to obtain the correct orientation of the starting model [109]. This correctly aligned phasing model, can then be used with the electron density function

(Eq. 1.6), in order to generate an electron density map. From this initial electron density map, the model can be altered and rebuilt (see subsection 2.3.5) where necessary until the structure is successfully determined [46, 63, 109, 110].

Density Modification Methods

Density modification techniques are used to help visualise the structure under investigation. These methods include solvent flattening, histogram matching, and non-crystallographic symmetry averaging [23].

Solvent flattening works in two ways, by removing areas of negative electron density, and by altering the value of electron density in solvent regions to that of a typical solvent value rather than typical protein value to allow for better accuracy with initial refinement iterations [23]. Programs have been developed that can automatically define the protein–solvent boundary, and therefore automatically apply solvent flattening [23]. Further refinements and variations have also been developed, such as the "solvent-flipping method", by Abrahams and Leslie, in 1996 [111].

Another density modification technique is that of histogram matching [23]. It is a complementary technique to solvent flattening, and works by altering the values of electron-density points so that they fall within the bounds of an expected distribution [23]. This technique works because biological macromolecules, particularly polypeptide structures are broadly similar in their atomic composition, and atomic bonds [23]. The overall protein structural similarities can be used, to estimate the electron density values for proteins, even when individual atomic positions aren't known [23].

Lastly, Non-Crystallographic Symmetry (NCS) averaging is helpful when there is more than one copy of a molecule present within a unit cell, and the molecules are related to one another via a symmetry operation [23, 112]. NCS averaging works by averaging the density among those NCS related molecules, reducing the noise and increasing the constraints on the phase, thereby improving the accuracy of the electron density, and is a staple that is used in both conventional and serial crystallography [112].

2.3.5 Model building, metrics and refinement

Using MR methods, and having obtained an initial model and electron density map, an iterative process of building and refining the model is performed to produce a more accurate version of the structure. The building and refinement phases can initially be performed automatically, leading to rough improvements in model fitting, after which manual adjustments are necessary. Different programs have been developed that perform auto-building [113] starting off with ARP/wARP in 1999 [114], with current pipelines including Buccaneer [115], PHENIX AutoBuild [116], SHELXE [117, 118], as well as an updated version of ARP/wARP [119].

These auto-building programs work iteratively, with multiple steps per iteration that include processes such as locating candidate C-alpha positions, growing C α atoms into chain fragments, joining and merging fragments, linking nearby N and C termini, sequencing the chains, correcting any insertions or deletions, filtering regions of poor density, NCS rebuilding in order to complete NCS copies of chains, pruning any remaining clashing chains, and rebuilding the side chains [113, 115].

With manual building a similar process is done, but with the addition of a graphic interface for visualisation during the building phase. Developments in molecular graphics has given us the tools to expand our understanding of molecules and molecular dynamics, and a variety of programs are available [120]. Popular programs used for visualisation and model building include COOT [121], PyMOL [122], Chimera [123] and VMD [124].

A range of different refinement programs that are included in these autobuild pipelines and can also be utilised during manual refinement include phenix.refine [125], REFMAC5 [126] and SHELXL [118]. To determine if the building step has actually improved the accuracy of the model, during this refinement step, R-factor statistics including the R_{work} and R_{free} are calculated:

$$R = \frac{\sum_{h,k,l} ||\mathbf{F}_{obs}(h, k, l)| - |\mathbf{F}_{calc}(h, k, l)||}{\sum_{h,k,l} |\mathbf{F}_{obs}(h, k, l)|}, \quad (2.10)$$

and these R-factors are regularly reported as part of the process of submitting a new structure [127]. R-factors are a measure of the agreement between the experimental structure

factor amplitudes $|\mathbf{F}_{obs}|$, and the calculated structure factor amplitudes $|\mathbf{F}_{calc}|$ based on the structural model.

While the R_{work} alone gives a good indication of how well the model is fitting the experimental data, over-fitting of the data can lead to inaccurate modelling, and so a second R-factor is utilised, the R_{free} . The R_{free} is calculated with the same R-factor equation (Eq. 2.10), but uses a subset of the experimental data that has been excluded from the model building and refinement steps. This subset is usually 5-10% of the data and is used to determine if over-fitting and user-introduced bias is taking place. A quality model will not only have a low R_{work} , but also a low R_{free} , to show that the model is both accurate to the data and is not over-fitted [127].

Other statistics and validation techniques also play a vital role in obtaining the most accurate representation of the protein structure and this is where the two crystallography methods differ. While conventional crystallography has well developed gold standards in the field, this is still not the case for SX. In the case of SX data, variations in which statistics are reported, do still occur. Statistics that are used in SX include the Pearson correlation coefficients, $CC_{\frac{1}{2}}$, and CC^* , which are used as data quality indicators. The Pearson correlation coefficient, $(CC_{\frac{1}{2}})$:

$$CC_{\frac{1}{2}} = \frac{\sum(a_i - \langle a \rangle)(b_i - \langle b \rangle)}{\sqrt{\sum(a_i - \langle a \rangle)^2 \sum(b_i - \langle b \rangle)^2}}, \quad (2.11)$$

is used primarily for selecting the best high-resolution data cut-off for a structure [128]. It is calculated by splitting the experimental data into two data subsets ($a \& b$), where each subset is randomly assigned half of the measurements of each unique reflection [129, 130], and where a_i and b_i are defined as the intensities of the unique reflections, and $\langle a \rangle$ and $\langle b \rangle$ represent the averages [129, 130].

CC^* is calculated from the $CC_{1/2}$:

$$CC^* = \sqrt{\frac{2CC_{\frac{1}{2}}}{1 + CC_{\frac{1}{2}}}}, \quad (2.12)$$

and is another quality indicator that can be used to determine whether there is over-fitting of the model, or whether the data quality is limiting model improvement [128, 129]. Comparing CC^* to the CC_{work} and CC_{free} are also valuable for checking data quality. CC_{work} and CC_{free}

are both calculated using the same equation (Eq. 2.11), but with different data subsets. For CC_{work} , the two subsets are comparing the experimental intensities with the intensities calculated from the molecular model, and for the CC_{free} , the two subsets are comparing the subset of the experimental data that was not used for modelling and refinement, with the subset of the intensities calculated from the molecular model [128, 129].

When comparing the CC^* to the CC_{work} and CC_{free} , a CC_{work} that is larger than CC^* implies that data over-fitting has occurred, while a CC_{free} smaller than CC^* implies that the model is not accounting for all the data [129]. Furthermore, a CC_{free} that is close to the same value of CC^* , implies that it is the data quality that is the limiting factor [129].

Other metrics that are important in SX and are also reported in conventional crystallography include $I/\sigma(I)$, B-factors and Ramachandran outliers. $I/\sigma(I)$ is the average intensity of a reflection over the standard deviation and is a good metric for assessing the quality of low resolution data [128]. The B-factors of the structure (see subsection 1.5.6), are a measure of attenuation of X-rays due to thermal motion resulting in atoms that are displaced from their mean atomic positions. The smaller the B-factor value, the more accurate the atomic positioning is within the structure. Three overall B-factors are examined, that of the structure as a whole, the main chain B-factor, and the side chain B-factor. A high side chain B-factor (>50) indicates the likelihood that the side chains are from sections with increased mobility and flexibility, which suggests that alternate side-chain conformations could be present that need to be correctly modelled. Therefore, it is important that these B-factors are small, to show that the model only has minimal uncertainty.

One other important metric to look at when determining the accuracy of the molecular structure is the Ramachandran plot. The Ramachandran plot depicts three different regions for the amino acid sides chains within the model, an energetically favoured region, an allowed region, and a high-energy unfavoured region, for the φ and ψ torsional angles of the main chain $C\alpha$ bonds [131] (Fig. 2.4). φ refers to the torsion angle $C\alpha - N$, while ψ refers to $C\alpha - C$ bond. This additional information helps to predict the most likely φ and ψ bond angle combinations that are energetically favoured for the individual amino acids, while also fitting the electron density present, allowing for a more accurate identification of the position of the amino acid residues within the model [131].

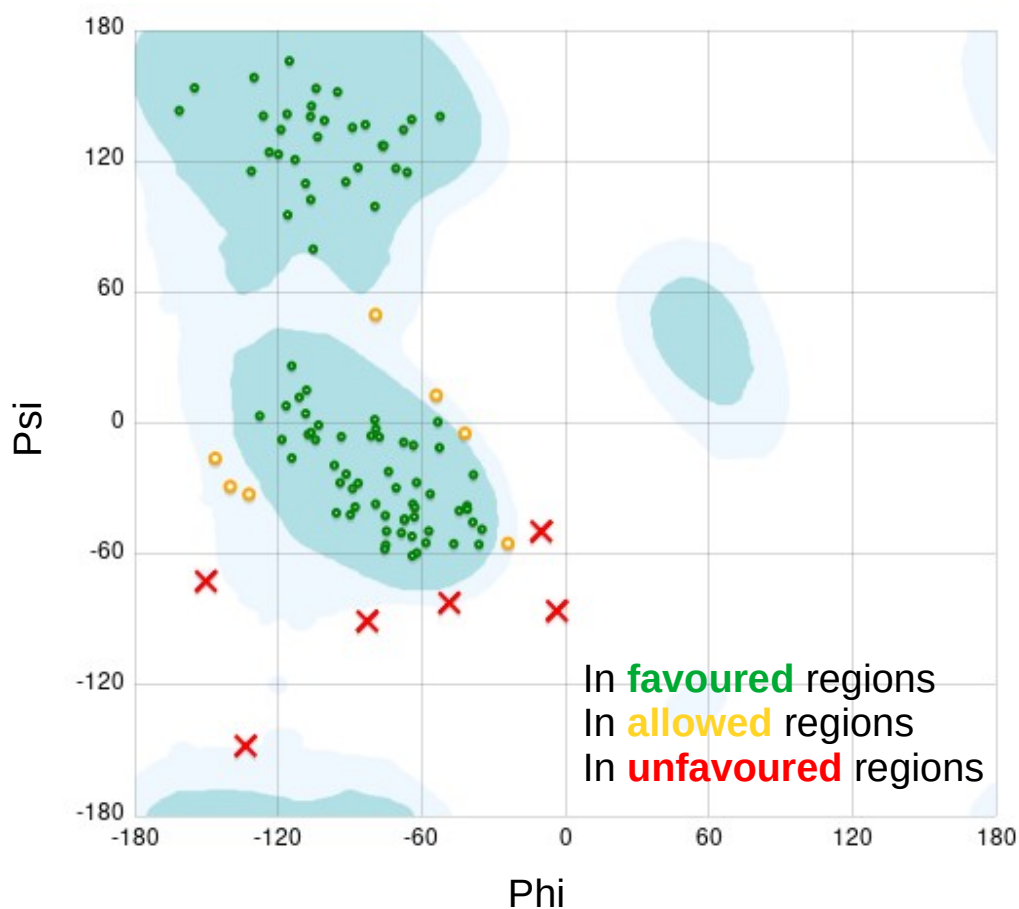


FIGURE 2.4: An example of a Ramachandran plot, showing the residues that are falling into the energetically favoured region (green), the energetically allowed region (yellow), and the high-energy unfavoured region (red) for the phi and psi torsion angles.

SX is still a developing field and further work in developing highly rigorous statistical reporting methodologies that can be consistently utilised for every structure, is one area in particular that still needs to be thoroughly addressed [132].

2.4 Research and Development

2.4.1 Synchrotron developments

With the improvements in instrumentation, technique development and data analysis methodology, synchrotrons have come a long way since the first generation synchrotron source came

online in the 1970's. The exposure times needed to obtain diffraction using a synchrotron source are in the millisecond range compared to the femtosecond regime achieved by an XFEL source, but there are currently far more synchrotron beamlines available compared to XFELs [133–135] which means they are more accessible to users and this is what drives the instrument development at synchrotrons.

A specific technique that has been developed for SX at synchrotrons is pink beam serial crystallography. Pink beam SX is when a polychromatic beam ($\Delta E/E \approx 5 \times 10^{-2}$) is used, which provides a much higher flux (100x) than a monochromatic beam and allows for a larger slice of reciprocal space to be obtained per diffraction pattern. As a higher flux can be obtained, much shorter exposure times (100 picoseconds) are now necessary, which therefore makes it possible to perform time-resolved serial crystallography experiments that have previously been possible only at XFELs [136, 137]. Pink beam SX, means that instead of obtaining preliminary data at synchrotrons and then having to apply for limited XFEL beamtime, that both the preliminary SX, and tr-SX can both be performed at a synchrotron source [136].

2.4.2 XFEL developments

There are many areas of XFEL based research, including expanding our fundamental understanding of XFEL physics. Such research includes characterising the temporal and spatial coherence of the beam [138, 139], understating all aspects of beam-sample interactions including the mechanisms involved in radiation damage at extreme dose rates [50, 140–142], beam induced shock-waves [143–148], as well as the characterisation of pump-probe jitter [149].

There is much cross-over in the area of SX sample delivery development for use at both synchrotrons and XFELs (see section 2.1). This includes upgrades to goniometer instrumentation at XFEL facilities so that the fixed-target methods [150] originally developed for SMX can also be used for SFX [53, 69, 79]. Other research has focused on improvements to the GDVN, with improvements to the process of nozzle fabrication [151]; as well as newer designs such as the DFFN, and segmented flow generation using a microfluidic device before

GDVN injection [152].

Technique development is another major area of XFEL research. The original XFEL protein crystallography technique was that of SFX [51, 153], developed as a way to overcome the radiation damage limits of conventional crystallography, and allowing us to obtain diffraction data from micro and nano-sized crystals. Following on from the success of SFX, was the development of time-resolved serial femtosecond crystallography (tr-SFX). tr-SFX expands our ability to understand biological systems, making it possible to observe the molecular dynamics of proteins on time-scales as small as picoseconds, by creating a "stop-motion" molecular movie. This molecular movie is created by capturing the system under investigation at a specific time interval after a change in conformation is induced, so that diffraction data can be obtained when the system is in an intermediate form. Conformational changes can be induced by light or chemical activation and are tailored to the biological system under investigation [140, 154–159]. Time-resolved SFX studies that have been performed so far, have included biological systems such as Bacteriorhodopsin [160, 161], Rhodopsin I and II [162, 163], Photosystem II [78, 164], and Phytochromes [165, 166].

A technique still in its infancy, is Single Particle Imaging (SPI). Unlike with other X-ray methods for determining the structure of molecules, SPI aims to eliminate the need for protein crystallisation, and directly image the structure of molecules [28, 167–169]. Initial 3D reconstructions of biological samples at low resolution have been achieved, including the Mimivirus [170], Melbourne virus [171], Rice dwarf virus [172], and PR772 virus [173]. A large amount of diffraction data is required in order to successfully reconstruct single particles, but it is extremely difficult to isolate and transport single particles of <40 nm into the interaction region of the beam, and it is also difficult to minimise background scatter from the samples delivery medium and account for the variation in background scatter between shots [167]. Therefore developments in delivery methods are extremely important for the future of SPI, with one such method currently in development that is using liquid cells as a delivery medium, designed so that they don't restrict incident or diffracted X-rays at either low or high tilt angles [174].

XFEL research and development, when compared to synchrotrons, is still in its infancy, with ongoing research into the fundamental understanding of the physics behind the XFEL

beam structure, as well as continual development of sample delivery methods and data analysis steps. The following chapters contain the research performed during my PhD candidature, aiming to provide insights into how multiple hits of a single crystal delivered by GDVN to the XFEL are possible (Chapter 3), the advantages, disadvantages, and difficulties that arise when delivering specific crystals for SX (Chapter 4), and how SFX is used to solve the structure of a key biological target, MyD88, involved in the immune response (Chapter 5).

Chapter 3

Multi-Hit Serial Femtosecond Crystallography

3.1 Introduction

The European XFEL (EuXFEL) is one of the world's most powerful XFEL source and was the first facility to operate in the megahertz regime. In this chapter I present an analysis of data obtained at the SPB/SFX beamline at the EuXFEL. During the initial data collection and subsequent data analysis for this experiment a number of technical challenges were addressed. These challenges included the development of a data acquisition pipeline using the newly built Adaptive Gain Integrating Pixel Detector (AGIPD), which included 16 detector modules that are effectively independent detector systems in their own right [81]. The AGIPD detector was specifically developed for the EuXFEL, as no other detector available at the time was capable of keeping up with the MHz pulse repetition rate. The EuXFEL can generate up to 2,700 pulses per train, with a train repetition rate of 10 Hz and a 220 ns separation between pulse trains; a pulse rate of 4.5 MHz is possible. Development of a detector for the EuXFEL was critical, as a poorly performing detector could be a fundamental bottleneck, resulting in experiments that are unable to fully utilise the EuXFELs MHz repetition rate.

The EuXFEL became operational in 2017, and the first ever user experiments were

carried out in September, 2017 [45, 46]. One of these first experiments, performed at the SPB/SFX beamline, involved a large open collaboration that included over 100 scientists from all over the world. My role on this experiment was to participate in the analysis of MHz SFX data. This experiment demonstrated that megahertz serial crystallography was possible, addressing questions about whether the short pulse intervals would allow jet recovery, or whether the shock wave induced from the first X-ray pulse would travel upstream and destroy the crystals before they reached the interaction region. While these issues were observed at jet speeds of 25 m/s, at speeds above 50 m/s the jet recovered in time for the next sample-X-ray interaction [46]. At the fast sample delivery speeds utilised during this experiment in combination with the beam size, a crystal may not have time to exit the beam interaction region prior to the next X-ray pulse. The key question of whether we can observe a single crystal having a second interaction with the EuXFEL beam is addressed in this chapter.

3.2 Lysozyme

3.2.1 Biological background

Lysozyme is a protein with a crystal system that is commonly used as a control in experiments that are focused on method or facility development. It is commonly used because it is a stable and well-characterised protein that is cheap, commercially available and easy to grow in large quantities. It is also robust and comparatively radiation tolerant with many structures already published and deposited to the Protein Data Bank (PDB) [46, 52].

3.2.2 Preparation

The HEWL crystals were grown using the rapid-mixing batch method, and crystals ranging in diameter from 6-8 μm were obtained [46]. The 6-8 μm crystal size range was achieved through the following steps. A 3:1 ratio of precipitant (1M NaCl, 40%(v/v) ethylene glycol, 15%(w/v) PEG 4000, 50 mM acetate buffer, pH 3.5 and filtered through a 0.45 μm filter) to HEWL (Sigma-Aldrich; 126 mg/ml in a 50 mM acetate buffer, pH 3.5 and filtered through

a 0.10 μm filter) at 1°C was used [46]. The sample was immediately vortexed, and then incubated at 1°C for 30 minutes [46]. At this point the crystals were stable for sample injection and delivery.

During the experiment, a GDVN delivery system was used, and included a 50 μm diameter 3D printed nozzle tip and a co-propagating helium gas flow for focusing and accelerating the stream of crystals into the interaction region. The sample flow rate of the GDVN was controlled by a High Pressure Liquid Chromatography (HPLC) pump, where a water line connected to the HPLC was used to apply pressure to a plunger that was connected to the sample reservoir, resulting in a continuous sample stream. The flow rate of the surrounding gas, was then adjusted to optimise for a small and stable sample stream. Different combinations of sample and gas flow rates were tested during the experiment so a range of velocities could be obtained. Some of these flow rate combinations were tested in a lab setting utilising water and crystal samples in order to determine these velocities. Not all flow rate combinations were lab tested, and therefore the velocities of some of these flow rate combinations have been theoretically calculated and presented in the publication below.

One of the major challenges during this experiment, was due to the new charge integrating detector, built specifically to address the megahertz capabilities of the EuXFEL. As mentioned above, the AGIPD is capable of recording at MHz repetition rates. The maximum capabilities of the EuXFEL involve trains with 2700 pulse at a repetition rate of 4.5 MHz. While the AGIPD isn't capable of reading out all 2700 pulses in a train, each pixel in the AGIPD was created to contain 352 analogue memory cells, therefore allowing measurement of 352 pulses at the 4.5 MHz rates. The AGIPD also uses a dynamic gain switching technique, storing 3 voltage levels, making the AGIPD capable of detecting a high dynamic range ($1 - 10^4$ photons at 12 keV).

3.3 Motivation

The first publication from the SPB/SFX EuXFEL beamline resulted in the successful application of SFX at a MHz XFEL facility [46]. A 1.1 MHz rate was used, and this pulse train format can be seen in Figure 3.1.

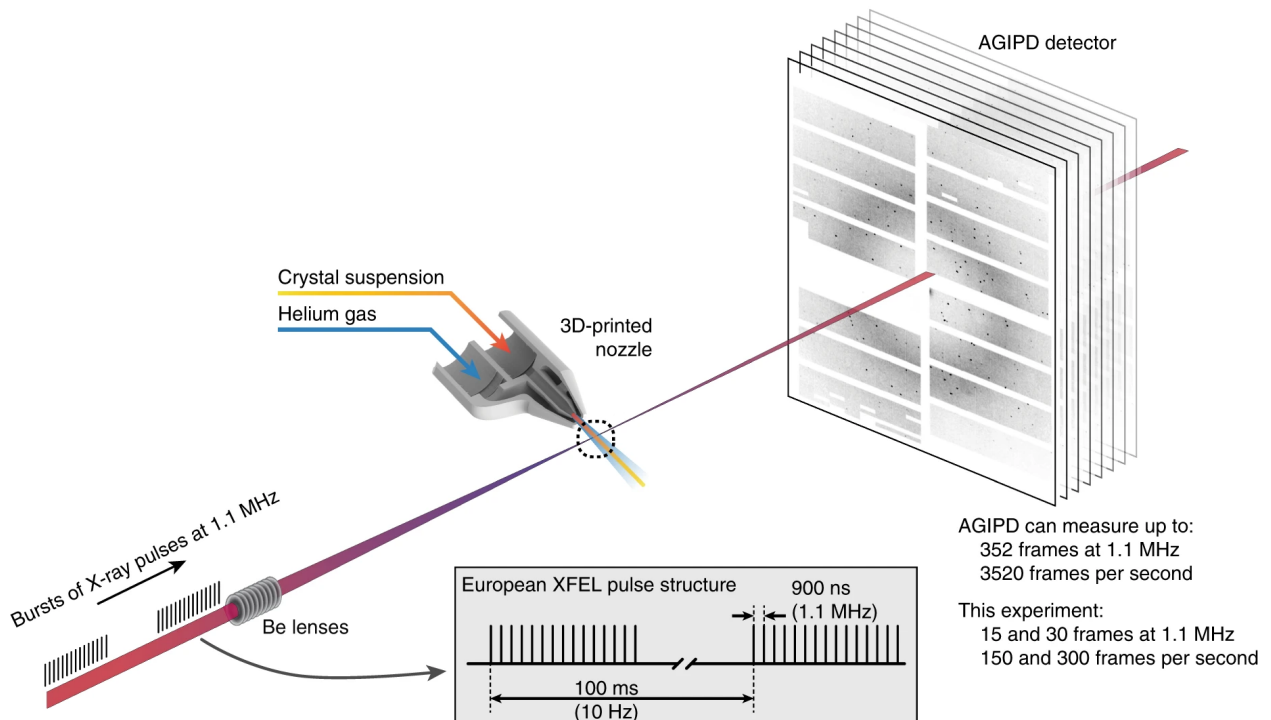


FIGURE 3.1: The European XFEL GDVN setup, where a 3D-printed nozzle is used to deliver the sample to the beam. The X-ray trains are delivered at a repetition rate of 10 Hz, with a 1.1 MHz pulse repetition rate within the trains. This figure is reprinted from Wiedorn 2018 [46]. CC BY 4.0.

A closer look at indexed and integrated results from the HEWL crystal diffraction patterns revealed occurrences of single crystals being hit twice with two consecutive pulses. This phenomenon was termed "multi-hit" SFX. This observation demonstrated for the first time at a MHz XFEL source, that crystals were capable of surviving more than one pulse. We have demonstrated in our analysis of these multi-hits, that even at jet speeds of 102 m/s, that multi-hits were still being observed.

This observation indicates that the beam profile must be larger than the initial analysis suggested in the first MHz SFX paper [46]. Therefore, in order to determine a more accurate beam-size that was in agreement with our experimental observations, the original YAG images were re-analysed, and further in-depth analysis was performed. Based on this analysis the beam size was confirmed to be larger than originally thought, supported by the presence of multi-hit crystals at the 102 m/s jet speed (taking into account errors in the jet speeds).

During the re-analysis of the YAG images, it was difficult to determine an accurate threshold that distinguished the background from the actual beam. Therefore, it was necessary to optimise the threshold of the incident beam taking into account the standard deviation of the local intensity as well as the possible error range for the jet speeds. A 5% error in jet speed was included in all calculations, and is described in more detail in the publication below.

During the original beamtime, a total of 749,874 detector images were collected during HEWL data collection. Of those images, 25,193 images were identified as hits, using Cheetah hit finding. 24,733 of those hits, were then successfully indexed using CrystFEL, which gave 25,531 indexed crystal lattices that were then used to determine the 1.8 Å structure of HEWL, published in the paper by Wiedorn et al. 2018 [46]. It was demonstrated in this paper, and a paper by Yefanov et al. 2019 [175], that crystals were equally probable for all pulses within a train, and that they were also equally likely to be indexed.

From these observations, further investigations were performed to determine whether these multi-hit crystals produced undamaged and equally high resolution diffraction patterns when comparing the diffraction obtained from the first and second pulses. The stream files for the HEWL crystal analysis were obtained from Wiedorn et al. [46], and were used to identify, the diffraction patterns resulting from single crystals hit by more than one X-ray pulse. To determine which of the diffraction patterns were due to this multi-hit phenomenon, a Python script was written that incorporated Whirligig, a script available within the CrystFEL suite [176]. Whirligig identifies multi-hits by calculating whether two consecutive diffraction patterns have similar orientations but is not designed for the pulse-train format of the EuXFEL. A more detailed explanation of how multi-hits were accurately identified can be found in the publication below.

3.4 Publication

TABLE 3.1: Candidate and co-author contributions

Manuscript Title	“Megahertz pulse trains enable multi-hit serial crystallography experiments at XFELs”
Journal	Nature Communications
Publication Status	Submitted
Authors	<p>Susannah Holmes, Henry J. Kirkwood, Richard Bean, Klaus Giewekemeyer, Hugh Marman, Andrew V. Martin, Marjan Hadian-Jazi, Max O. Wiedorn, Dominik Oberthur, Luigi Adriano, Nasser Al-Qudami, Saša Bajt, Imrich Barák, Sadia Bari, Johan Bielecki, Sandor Brockhauser, Mathew A. Coleman, Francisco Cruz-Mazo, Cyril Danilevski, Katerina Dörner, Alfonso M. Gañán-Calvo, Rita Graceffa, Michael Heymann, Hans Fangohr, Matthias Frank, Steffen Hauf, Alexander Kaukher, Yoonhee Kim, Bostjan Kobe, Juraj Knoška, Torsten Laurus, Romain Letrun, Luis Maia, Marc Messerschmidt, Markus Metz, Thomas Michelat, Grant Mills, Serguei Molodtsov, Diana C.F. Monteiro, Astrid Münnich, Gisel E. Peña Murillo, Gianpietro Previtali, Adam Round, Tokushi Sato, Robin Schubert, Joachim Schulz, Megan Shelby, Carolin Seuring, Jonas A. Sellberg, Marcin Sikorski, Alessandro Silenzi, Stephan Stern, Jola Sztuk-Dambietz, Janusz Szuba, Martin Trebbin, Patrick Vagovic, Thomas Ve, Britta Weinhausen, Krzysztof Wrona, P. Lourdu Xavier, Chen Xu, Oleksandr Yefanov, Keith A. Nugent, Henry Chapman, Adrian Mancuso, Anton Barty, Connie Darmanin & Brian Abbey</p>
Nature and extent of candidates contribution	Seventy percent including the multi-hit data extraction and analysis, beam-diameter analysis, interpretation of results and lysozyme structural determination.
Nature and extent of co-authors contributions	Thirty percent including data that was provided and published in Wiedorn et al. [46], radiation dose calculations, lysozyme structural determination, discussion of ideas, and critical review.

Megahertz pulse trains enable multi-hit serial crystallography experiments at XFELs

Susannah Holmes¹, Henry J. Kirkwood², Richard Bean², Klaus Giewekemeyer², Hugh Marman¹, Andrew V. Martin³, M. Hadian-Jazi^{1,2,24}, Max O. Wiedorn⁴, Dominik Oberthur⁴, , Luigi Adriano⁵, Nasser Al-Qudami², Saša Bajt^{5,11}, Imrich Barák⁶, Sadia Bari⁵, Johan Bielecki², Sandor Brockhauser^{2,22}, Mathew A. Coleman⁷, Francisco Cruz-Mazo^{8,23}, Cyril Danilevski², Katerina Dörner², Alfonso M. Gañán-Calvo⁸, Rita Graceffa², Hans Fangor², Michael Heymann⁹, Matthias Frank⁷, Alexander Kaukher², Yoonhee Kim², Bostjan Kobe¹⁰, Juraj Knoška^{4,18}, Torsten Laurus⁵, Romain Letrun², Luis Maia², Marc Messerschmidt²⁰, Markus Metz⁴, Thomas Michelat², Grant Mills², Serguei Molodtsov^{2,25,26}, Diana C. F. Monteiro^{11,12}, Astrid Münnich², Gisel Peña Murillo⁴, Gianpietro Previtali², Adam Round^{2,21}, Tokushi Sato^{2,5}, Robin Schubert², Joachim Schulz², Megan Shelby⁷, Carolin Seuring^{4,11}, Jonas A. Sellberg¹³, Marcin Sikorski², Alessandro Silenzi², Stephan Stern², Jola Sztuk-Dambietz², Janusz Szuba², Martin Trebbin^{14,15}, Patrick Vagovic², Thomas Ve¹⁶, Britta Weinhausen², Krzysztof Wrona², P. Lourdu Xavier^{4,19}, Chen Xu², Oleksandr Yefanov⁴, Keith A. Nugent^{1,17}, Henry Chapman^{4,11,18}, Adrian Mancuso², Anton Barty⁴, Brian Abbey^{1*}, Connie Darmanin^{1*}

¹Australian Research Council (ARC) Centre of Excellence in Advanced Molecular Imaging, Department of Chemistry and Physics, La Trobe Institute for Molecular Sciences, La Trobe University, Bundoora, VIC 3086, Australia. ²European XFEL, Holzkoppel 4, 22869 Schenefeld, Germany. ³School of Science, RMIT University, Melbourne, VIC 3000, Australia. ⁴Center for Free-Electron Laser Science, Deutsches Elektronen-Synchrotron DESY, Notkestrasse 85, 22607 Hamburg, Germany. ⁵DESY, Notkestrasse 85, Hamburg, Germany. ⁶Institute of Molecular Biology, SAS, Dubravská cesta 21, 845 51 Bratislava, Slovakia. ⁷Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA. ⁸Dept. de Ingeniería Aeroespacial y Mecánica de Fluidos, ETSI, Universidad de Sevilla, 41092 Sevilla, Spain. ⁹Institute of Biomaterials and Biomolecular Systems, University of Stuttgart, Am Pfaffenwaldring 57, 70569 Stuttgart, Germany, ¹⁰School of Chemistry and Molecular Biosciences, Institute for Molecular Bioscience and Australian Infectious Diseases Research Centre, University of Queensland, Brisbane, QLD 4072, Australia. ¹¹The

Hamburg Centre for Ultrafast Imaging, Universität Hamburg, Luruper Chaussee 149, Hamburg 22761, Germany. ¹²Hauptman-Woodward Medical Research Institute, 700 Ellicott St., Buffalo, NY 14203, USA, ¹³Biomedical and X-ray Physics, Department of Applied Physics, AlbaNova University Center. KTH Royal Institute of Technology, SE-106 91 Stockholm, Sweden. ¹⁴Department of Chemistry, State University of New York at Buffalo, 760 Natural Sciences Complex, Buffalo, New York 14260, USA. ¹⁵Hauptman-Woodward Medical Research Institute, 700 Ellicott Street, Buffalo, New York 14203, USA. ¹⁶Institute for Glycomics, Griffith University, Southport, QLD 4222, Australia. ¹⁷Laser Physics Centre, Research School of Physics Australian National University, ACT, 2601 Australia. ¹⁸Department of Physics, Universität Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany. ¹⁹Max-Planck Institute for the Structure and Dynamics of Matter, Luruper Chaussee 175, 22761 Hamburg, Germany. ²⁰School of Molecular Science, Arizona State University, Tempe, AZ, 85281, USA. ²¹Max Planck Institute of Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany. ²²Biological Research Center, Szeged (BRC), Temesvari krt. 62., 6726 Szeged, Hungary University of Szeged, Arpad ter 2, 6720 Szeged, Hungary, ²³Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ 08544, USA, ²⁴ Australian Nuclear Science and Technology Organisation (ANSTO), NSW 2234, Australia, ²⁵Institute of Experimental Physics, TU Bergakademie Freiberg, Leipziger, Str. 23, 09599 Freiberg, Germany, ²⁶ITMO University, Kronverksky pr. 49, St. Petersburg, 197101, Russia.

*Correspondence to: c.darmanin@latrobe.edu.au and b.abbey@latrobe.edu.au

Abstract

The pulse structure of the European X-ray Free Electron Laser (EuXFEL) and Linac coherent Light Source (LCLS) II enables Serial Femtosecond Crystallography (SFX) data to be collected at megahertz (MHz) repetition rates. Previous work has shown that it is possible to use consecutive X-ray pulses within a single ‘pulse train’ to collect diffraction patterns from individual crystals. However, it has not previously been established whether structural information can be obtained from the same crystal hit multiple times as it traverses the X-ray beam within a liquid jet. Here, we exploit the MHz pulse structure of the EuXFEL to obtain two complete datasets from the well-known model system lysozyme. One from the crystal hit once and the second from the same single crystal hit again before it exits the beam. The two datasets from the same protein crystal are separated in time by less than a microsecond and yield two separate structures with a resolution up to 2.1 Å. Comparisons, using difference electron density maps, between the first and second hit structure reveal no significant indications of radiation damage visible surrounding the di-sulphide bonds in either of the lysozyme structures or significant changes within the active site. This is consistent with dose estimates

calculated using RADDOSE-3D XFEL. These results pave the way towards MHz SFX being used as a tool for tracking sub-microsecond structural changes in individual single crystals, a technique which we here refer to as ‘multi-hit serial crystallography’.

Introduction

The advent of Serial Femtosecond Crystallography (SFX) has created new opportunities via which macromolecular structures can be probed and their dynamics investigated. SFX is ideally suited to studying the molecular dynamics of molecules undergoing irreversible processes which cannot be measured using conventional synchrotron or lab-based X-ray sources¹⁻⁵. One practical hurdle to implementing SFX is obtaining a large enough data set for high-resolution 3D structure determination which typically comes at the cost of high sample consumption. The first generation of XFEL facilities typically had pulse repetition rates of the order of 120 Hz or less meaning that obtaining a large enough data set for structure retrieval is an inefficient process both in terms of the amount of sample required but also in terms of the amount of XFEL beamtime needed⁶. With the development of new high-repetition rate sources like the EuXFEL⁷ both data collection times and sample consumption are significantly reduced.^{1,8} Another avenue of research open to MegaHertz (MHz) XFEL facilities is the potential to use the unique pulse structure of these sources to perform novel time-resolved experiments^{7,9,10}. Typical time-resolved SFX (tr-SFX) experiments are performed either by optical pump/X-ray probe; reactions initiated by a laser or ‘mix-and-inject experiments; where reactions are initiated in the crystal via solvent diffusion.¹¹ A split and delay method also enables timing regimes of 20fs – 100fs to be accessed in order to probe molecular dynamics¹². The ultra-short pulse duration of the XFEL supports the study of in-situ molecular dynamics on sub-picosecond timescales by measuring a large number of crystals at various pump-probe delay times or mixing times¹³⁻¹⁶. In addition, to probing ultra-fast molecular dynamics, these measurements normally result in diffraction before destruction, where each crystal is measured once before being destroyed.

The time delay between two consecutive pulses at MHz XFEL sources is normally within the micro to nanosecond range. This creates an opportunity to explore a range of molecular dynamics occurring on sub-microsecond timescales by using consecutive pulses in the train (which are typically on the order of 10’s fs) to capture multiple diffraction patterns from the *same* crystal as it traverses the X-ray beam. This technique, which we here refer to as ‘multi-hit tr-SFX’ can be thought of as a bridge between SFX and conventional crystallography; in SFX typically thousands of crystals are hit once and destroyed by the XFEL beam, whereas in conventional crystallography one crystal is continuously rotated, and multiple diffraction patterns are collected prior to significant radiation

damage occurring. Previously, multiple hits of the same single crystal using XFELs have only been detected using high-viscosity jet streams and static chip system. This is because, unlike liquid jets, high-viscosity jets can flow at speeds slow enough that even at lower (e.g. non-MHz) repetition rate XFEL facilities and synchrotrons can observe multiple hits on the same crystal¹⁷⁻²⁰. The current paper thus represents the first experimental investigation comparing XFEL derived protein structures, from the same crystal, measured in consecutive pulses using continuous liquid jets.

Multi-hit tr-SFX makes it possible to study novel time-resolved molecular dynamics within the same crystal enabling us to differentiate between two crystal states within the one crystal. For example, when using an optical pump/X-ray probe system, the ‘dark’ state and optically activated ‘pumped’ state can be measured on the same crystal consecutively enabling the exact correlation of molecular information at a precise time interval. This provides the potential to detect a protein movement within the same crystal on a submicron time scale. In chemical ‘mix-and-inject’ experiments in which the distance along the jet of the X-ray beam from the mixing region determines the initial time point, the timing between consecutive pulses determines the timescale of the molecular dynamics that can be probed within a single crystal. The advantage of using multi-hit tr-SFX over conventional tr-SFX is that at least two diffraction patterns can be collected from a single crystal at different time points, within microsecond to sub-microsecond time regime, making it easier to track/correlate any changes in the molecular structure. These time scales allow us to probe the molecular dynamics of proteins, more specifically ligand binding, domain folding, transition states (i.e. switching between the active and inactive form)²¹⁻²⁵, helical motion²⁶ and side-chain rotations.²⁷ The combination of femtosecond intra-pulse and sub-microsecond inter-pulse timing resolution is something which is specific to MHz XFEL sources. The feasibility of performing experiments that exploit both characteristics is the topic of the present paper.

An example problem that could benefit from multi-hit tr-SFX is understanding the mechanism of the *Bacillus subtilis* response regulator, SpoOF, which is involved in sporulation. SpoOF can induce a shift in protein conformation on picosecond to millisecond time-scales^{24,28}. Previous NMR studies have demonstrated the secondary structure dynamics of the protein involves a complex series of movements including rotation of bonds in methyl groups (thought to occur within nanoseconds) and side chain flipping of buried residues (which occurs on the scale of seconds). Critical protein-protein contacts between SpoOF and its binding partners meanwhile take place on millisecond time-scales²⁴. Therefore, multi-hit tr-SFX may be useful in probing the intermediate states between the protein-

protein interactions in SpoOF and the flipping of residues to accommodate it. Other example applications which will benefit from multi-hit tr-SFX include studying protein-protein interactions i.e. determining the initial binding interactions involved in the adaptor protein oligomerisation of myeloid differentiation primary response gene88, which would aid in understanding its role in immunity²⁹. In addition, studying the binding of co-factors such as ATP or NADP, which induce extensive conformational changes within proteins and result in the transition between active to inactive forms of the enzyme, is known to occur on millisecond timescales and will also be a target for multi-hit tr-SFX. Multi-hit tr-SFX could also be used to capture the intermediate helical/sheet conformational changes in the secondary structures of proteins prior to side chain flipping events. Thus, the successful realisation of multi-hit tr-SFX will enable the capture of intermediate states in-situ based on the inter-pulse timing, which can be varied within a single pulse train³⁰.

Currently, the most common method for delivering samples to the XFEL beam is via a liquid jet, formed using a flow focusing nozzle or Gas-focused Dynamic Virtual Nozzle (GDVN)³¹⁻³³, in which the sample is continuously replenished. Typically, no single crystal contributes more than one diffraction pattern to the dataset, either because it is destroyed during interaction with the XFEL or because it has moved out of the interaction region prior to the next X-ray pulse arriving³⁴⁻³⁶. However, for crystals which are micron-sized or larger it is possible to obtain high-resolution, time-resolved, structural data at an XFEL using an X-ray beam which is much larger than the dimensions of the crystal³⁶. Crystals of this size are typically used for tr-SFX experiments where the primary motivation behind using the XFEL is the short pulse duration which enables snapshots to be taken of dynamically evolving molecular structures¹¹. The combination of femtosecond intra-pulse and sub-microsecond inter-pulse timing which is now accessible at the EuXFEL thus creates a new opportunity to explore molecular dynamics on sub-microsecond timescales by varying the inter-pulse spacing. An analysis of EuXFEL data collected at different liquid jet flow rates, ranging from 40-102 m/s, demonstrates that it is possible to obtain separate high-resolution datasets from the first and second hits on the *same* single crystal, hit twice, in a MHz SFX experiment. This lays the foundation for being able to collect femtosecond snapshots of molecular dynamics which occur on millisecond timescales. Further, both an analysis of the measured diffraction intensities from the two datasets and the resulting structure does not reveal any signs of radiation damage. This is consistent with RADDOSE-3D (XFEL version)³⁷ calculations showing that, under the experimental conditions used here, the crystal received less than half the dose (0.106 MGy) required to induce measurable radiation damage in lysozyme crystals at room temperature (RT) (0.38 MGy)³⁸. We have also established that even larger beam sizes are readily achievable at EuXFEL whilst providing, for micron-sized crystals, enough photon

flux to generate high-resolution diffraction data. Hence, it is possible, to further optimise the experimental setup to significantly increase the number of single crystals which are hit multiple times. In addition, depending on the breakup of the jet, it may even be possible to collect more than two diffraction patterns per single crystal prior to the sample either being damaged or exiting the beam (as indicated in our model) if the experimental setup (crystal size, sample flow rate, and beam size) is optimised. We note here, that the "multi-hit diffraction" scenario is clearly distinct from "multi-crystal diffraction", in which two or more crystals arrive at the X-ray beam at the same time, and which can be identified in the diffraction pattern and by using multi-crystal indexing options³⁹⁻⁴¹.

It should be emphasised, however, that in multi-hit tr-SFX the dose received by the crystal from the first hit should be below the RT damage limit. This can be estimated prior to the experiment (e.g. using programs such as RADDOSE) and confirmed during the experiment by comparing, on the fly, the diffraction data collected during the first and second hit. Whilst this mode of data collection likely precludes the use of nanocrystals due to their weak diffraction, it does allow for multi-hit, femtosecond diffraction experiments to be performed on microcrystals which are commonly used for tr-SFX¹⁴. In addition, considering that the experimental setup for multi-hit tr-SFX can be easily and quickly realised on the SFX/SBP beamline at the EuXFEL, multi-hit tr-SFX is an option available to all users. The key questions addressed in this paper are:

- How often are multi-hits (in this case double hits) observed in MHz SFX experiments?
- Why do multi-hits occur and what is the influence of the experimental geometry?
- How does the data quality and resolution of the first and second hits compare?
- Under what experimental conditions will multi-hit tr-SFX occur and how can this information be used to tailor future MHz XFEL experiments?

Results

Identifying consecutive hits for a single lysozyme crystal

Assessment of the frequency of multi-hit single crystal used X-ray data corresponding to three different jet speeds (where most of the data was collected); the statistics are summarised in Table 1. The number of images collected is compared to the number of images with hits detected by Cheetah⁴² and indexed by CrystFEL^{42,43}, as well as the percentage of those patterns indexed where crystals were hit twice as they passed through the beam. A crystal was determined a 'multi-hit' if it satisfied the following criteria: i) the two hits were from consecutive pulses and ii) the patterns showed very similar crystallographic orientations (the angle between each pair of basis vectors was $< 5^\circ$ and the lengths

of the basis vectors were in agreement to within 10%, determined using CrystFEL's Whirligig program^{20,42,43}. The data shows that, as expected, when the jet speed increases the percentage of crystals which were hit twice by the beam decreases. For the slowest (42 m/s), intermediate (78 m/s), and fastest (102 m/s) jet speeds the corresponding multi-hit percentages were 6.4%, 0.9% and 0.3%, respectively.

Table 1 Summary of jet speeds, experimental conditions, and statistics for lysozyme crystals¹. This table shows the flow rates of the gas and liquid used as well the experimentally and theoretically determined jet speeds for the current analysis. The data statistics were calculated utilising the CrystFEL software suite^{42,43}.

Target jet speed	50 m/s	75 m/s	110 m/s*
Liquid flow (μl/min)	15	13	13
Gas flow (mg/min)	23	50	85
Experimental jet speed (m/s)	42 ± 2.1	78 ± 3.9	-
Theoretical jet speed (m/s)	-	-	102 ± 5.1
Total no. frames	440,000	60,000	240,000
No. of hits	10,726 (2.4%)	1,638 (2.7%)	3,733 (1.6%)
No. of indexed frames	9,970 (93%)	1,509 (92.1%)	3,474 (93.1%)
No. double hits	1,190	28	20
No. single hits	8,780	1,481	3,454

* A '100 m/s' jet speed was determined experimentally for a 13 μl/min liquid and 80 mg/min gas flow rate giving a speed of 105 m/s, which is similar to the theoretically calculated 102 m/s jet speed.

Crystal rotation and Bragg peak analysis

To confirm that our method for determining which crystals were hit twice is reliable for the three jets speeds in Table 1, the change in crystal rotation for all the diffraction data collected from consecutive hits in the pulse trains was calculated. Rotation of the crystal as it 'tumbles' in the liquid jet between

consecutive X-ray pulses results in a small change in the position of the Bragg peak as illustrated in Figure 1. A significant increase in the number of crystals with $< 5^\circ$ change in orientation between consecutive hits was observed for the 42 m/s jet speed (Figure 1 and Supplementary Figures S1 and S2). This indicates that many of the crystals that apparently have only a very small change in their orientation are in fact the *same* crystal hit twice by the XFEL beam. By contrast, the number of crystals with a change in orientation of $> 5^\circ$ between consecutive hits does not vary significantly, consistent with crystals arriving at the X-ray interaction region with a random orientation. These results verify the predictions of the *whirligig* program; the same approach was also used to determine the number of multi-hit crystals that occurred for the two faster jet speeds. Due to the lower hit rates for the two faster jets speeds (78 and 102 m/s) the number of double hit diffraction patterns is much less than for the 42 m/s jet speed (see Table 1).

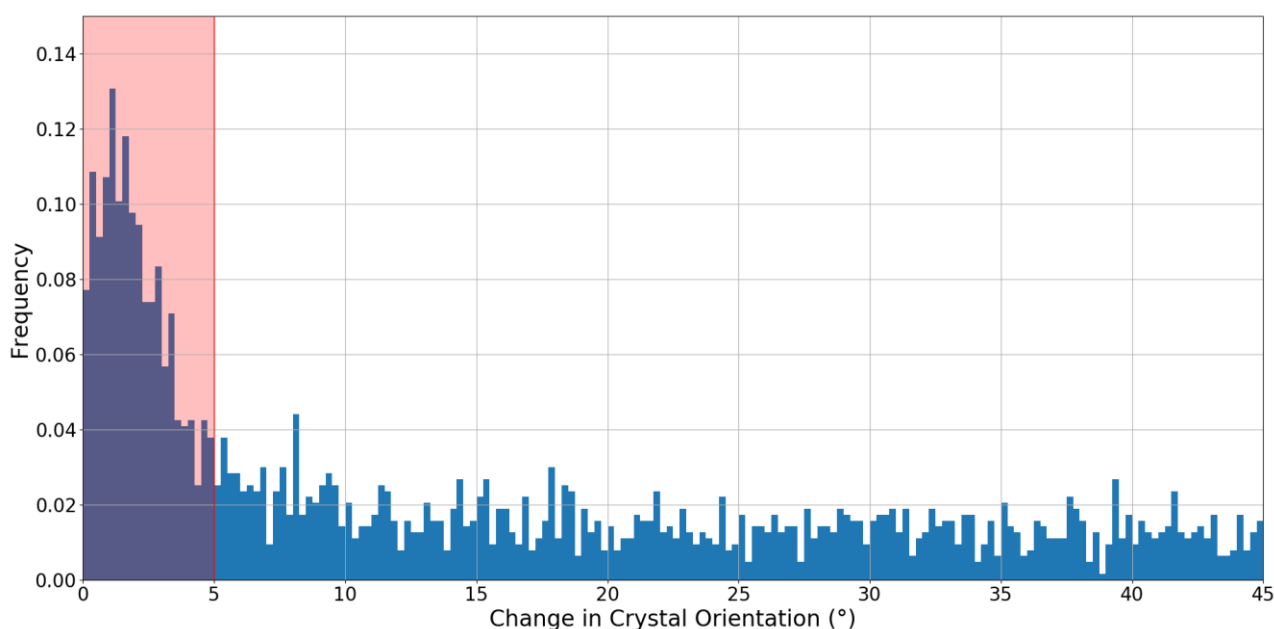
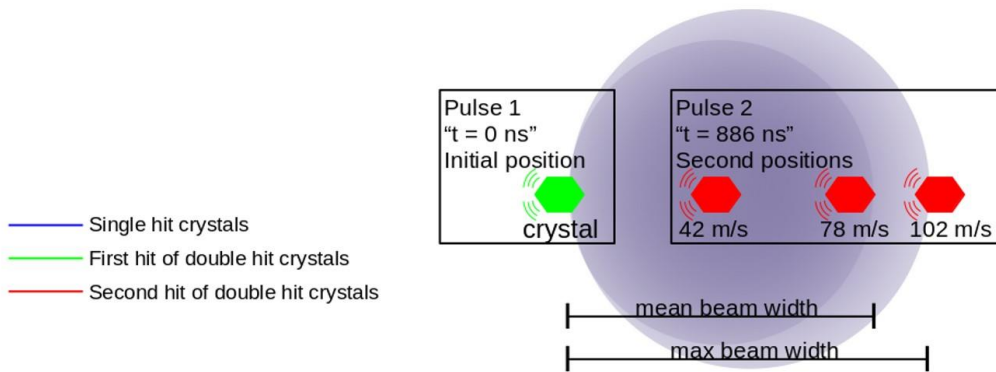


Figure 1. Frequency of diffraction patterns as a function of the relative change in crystal orientation. The change in crystal orientation was characterized by the reciprocal space vector \vec{a} , between consecutive diffraction measurements (separated in time by 886 ns) within the X-ray pulse train. The liquid jet flow rate was 42 m/s. An increase in frequency above 0.04 for consecutive images with a change in orientation of less than 5 degrees, indicated by the region shaded in red, can be observed and were classified as double hit crystals.

Understanding multi-hit tr-SFX as a function of jet speed

Irrespective of crystal concentration, multiple hits on a single crystal are primarily determined by the jet speed, beam size, and pulse repetition rate. To illustrate how multiple hits occur, a graphical representation of the crystal path through the beam based on the experimental data is presented in Figure 2. For the slowest jet speed (42 m/s) the crystal can be initially hit within the tail (lowest intensity) region of the X-ray beam followed by a second hit within the FWHM of the beam. The exposure of the same crystal to a second XFEL pulse within the most intense part of the X-ray beam means that the crystal may absorb sufficient dose that it is either destroyed or no longer diffracts to high enough resolution to generate a third hit, however, this is dependent on the beam properties and the dose the crystals receive during the first two hits. Theoretically, if the dose on the crystal, when placed centrally within the FWHM, is below the dose threshold and the crystal is travelling is slow enough and the beam is large enough, there is a possibility that a third hit could occur. However, this needs to be further investigated as it would only be possibly for jet speeds slower than 42 m/s and would depend on how radiation sensitive the crystal was. As the jet speed increases to 78 m/s, multiple crystal hits are only possible within the tail regions of the X-ray beam (Figure 2). At the fastest jet speed (102 m/s), multiple crystal hits can only occur in the extreme tail regions of the beam where the incident intensity is low compared to the central region (Figure 2). Therefore, for the two faster jets it limits the hit number per crystal to two.

A



B

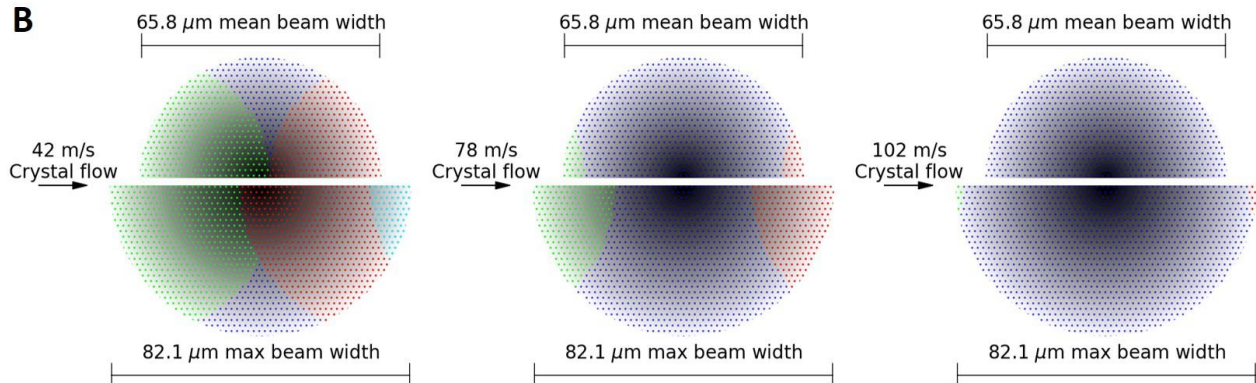


Figure 2. Model for how multi-hits occur for a single crystal. **A.** Schematic diagram (not to scale) illustrating the minimum distances travelled by an 8 μm crystal for the three different jet speeds overlaid with the average beam Full Width (FW, dark purple shaded region) and the maximum beam FW (light purple shaded region). The green crystal depicts the initial position, and the red crystal illustrates how far the crystal travels after the first hit for 42 m/s, 78 m/s, and 102 m/s jet speeds. **B.** Schematic representation of the crystal path through the X-ray beam for each of the three jet speeds for the mean beam FW (upper half) and maximum beam FW (lower half). The beam profile (shaded grey) is overlaid with the regions that the crystal travels through for the single hits (blue) to occur as well as the first (green) and second (red) hits of the double hit crystal. For the 42 m/s jet speed it also shows a possibility of the crystals being hit a third time (aqua) if the crystal and beam conditions were optimal. Note, for 42 m/s and 78 m/s, regions where no hits occur are possible.

Crystal transit through the X-ray beam

Assuming that at least 1 μm of the crystal needs to interact with the X-ray beam to generate a diffraction pattern and a maximum crystal size of 8 μm , the minimum distance travelled by a single crystal hit twice by the XFEL beam in this experiment was 35.4 μm , 56.7 μm , and 79.9 μm , for jet speeds of 42 m/s, 78 m/s, and 102 m/s respectively. Figure 2 presents a schematic showing how far the crystal travels between consecutive pulses for each jet speed. While Figure 3 provides the characteristic beam profile, based on YAG images, showing the full width (FW) and full width half maxima (FWHM) of the X-ray beam during the experiment. The mean, minimum, and maximum, FWs were calculated to be 65.8 μm , 41.7 μm and 82.1 μm , respectively. While the FWHMs were calculated to be 18.7 μm , 11.9 μm and, 23.3 μm , respectively. This analysis confirms that for each of the three jet speeds, the X-ray beam diameter was sufficiently large that crystals are hit at least twice by the X-ray beam, which is consistent with the experimental observations.

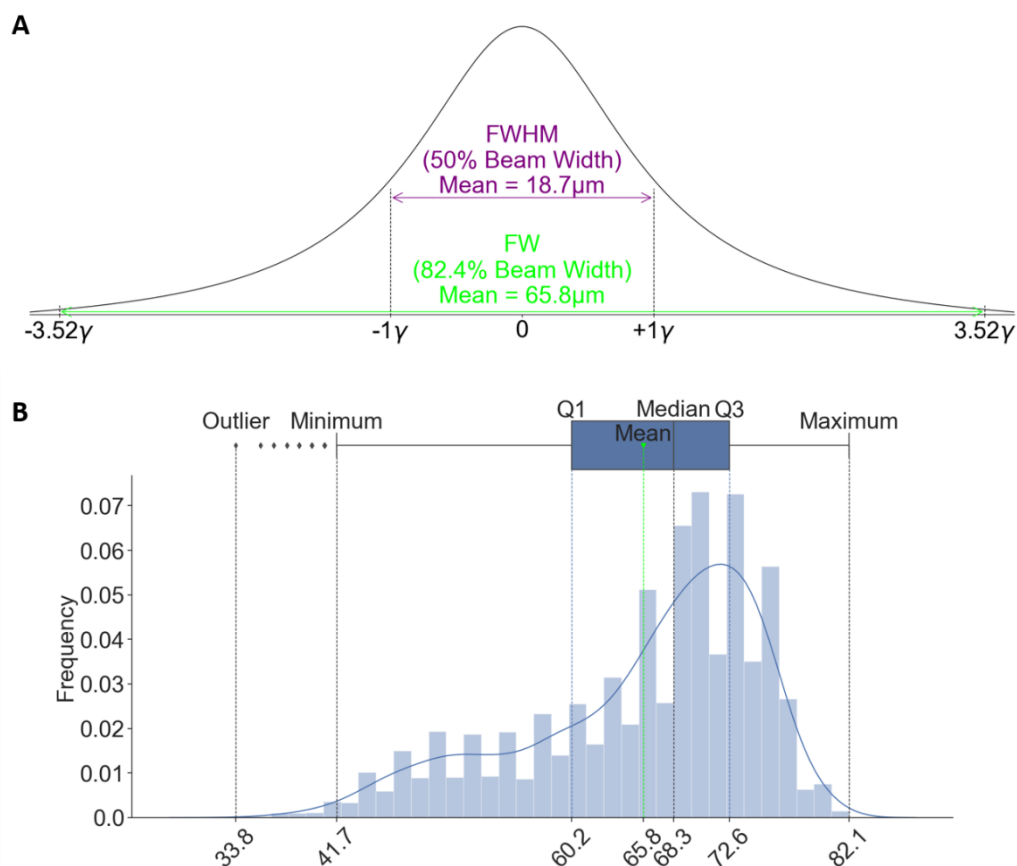


Figure 3. Characteristic beam profile. **A.** A histogram showing the X-ray beam profile. The beam profile was modelled using a Lorentzian distribution with a Full Width Half Maximum (FWHM) = 2γ (50% of beam) and Full Width (FW) = 7.04γ (82.4% of beam). **B.** The Lorentzian distribution used to determine the FWHM and FW from 6773 YAG images. The furthest outlier, minimum, Q1 (25th percentile), mean, median, Q3 (75th percentile), and maximum have been indicated. To obtain the FWHM and FW for each YAG image, a 7.5% noise threshold was applied to the image combined with a 3×3 median filter to account for the noise.

Data quality check

In addition to calculations of the absorbed dose, the data quality was carefully analysed to confirm that there was no degradation of the diffraction patterns collected from the second hit with respect to the first. Initially, the Bragg peak intensities and resolution of the diffraction data collected from the first and second hits was compared. This data was used to generate three independent powder plots (first hits, second hits, and single hits) for the three jet speeds (42, 78 and 102 m/s). For the slowest jet speed (42 m/s), we observed that the integrated intensity of the first hit was less than the second hit (note that the intensity of the second hit was, as expected, similar to that of the ‘single hit’ crystals, see Figure 4a). As the jet speed increases, the diffraction intensity profiles between the first and

second hit become similar (see Fig. 4b and 4c) but substantially lower when compared to the single hit crystals. This is consistent with our interpretation of how double hits occur (Figure 1) since our model predicts that for the faster jet speeds both first and second hits occur only within the tail region of the X-ray beam.

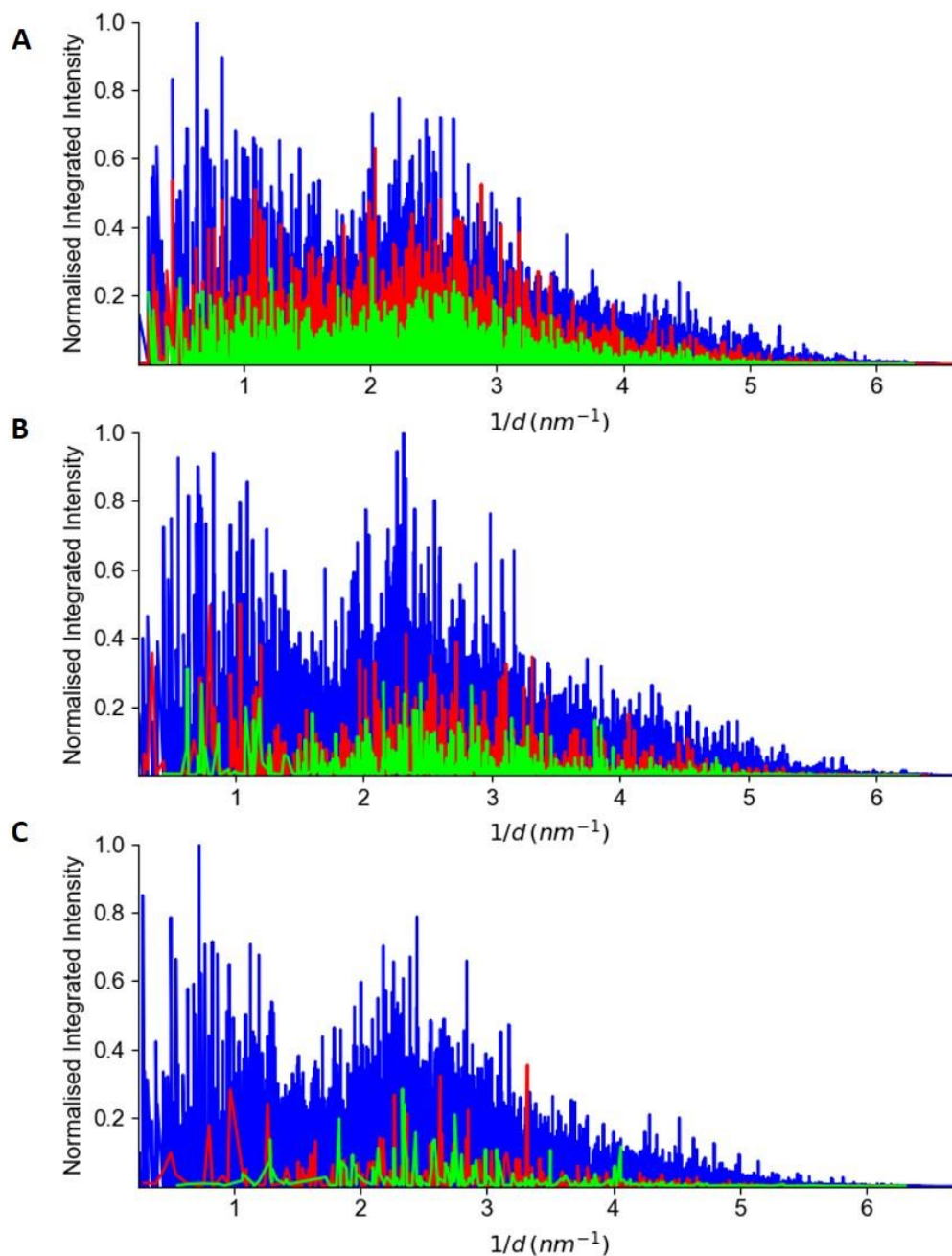


Figure 4. Normalized integrated intensity plots. Integrated intensities were extracted from the data, normalized and plotted against $1/d$ (where d is the lattice spacing) for the **A.** 42 m/s jet data; **B.** 78 m/s jet data; and **C.** 102 m/s jet data. Blue represents data for single hit crystals only; green represents the first hit of the double hit crystal; red represents the second hit of the double hit crystal. A threshold of $I/\text{sig}(I) > 2$ was applied to the analysis.

The final check was to perform a complete structure retrieval based on the independent data sets collected from the first and second hits and analyse the resulting electron density maps. A comparison of the difference electron density (DED) maps from the first and second hits is presented in Figure 5. The site specifically in the local vicinity of the active pocket of lysozyme, known to be sensitive to the effects of radiation damage, is shown in Figure 5A, B and C and Supplementary Figure S6. Superposition of the two structures, generated independently for the first and second hit data sets, does not reveal any significant differences in the backbone (rmsd of 0.3893 Å). The time-resolved DED maps (i.e. 886 ns apart) calculated the difference between the first hit and second hit structure factors and indicated that no significant changes could be attributed to the time-delay between the first and second hit on the same crystal (Figure 5A, Supplementary Figure S6). Crucially, examination of the di-sulfide bonds (Figure 5D and Supplementary Figure S6), which are very prone to radiation damage, did not result in presence of DED maps surrounding the atoms. Hence, based on these maps we conclude that there is no evidence of radiation damage occurring between first and second hits, meaning that any significant changes that did occur in a multi-hit tr-SFX experiment performed under these conditions could be attributed to actual molecular dynamics event.

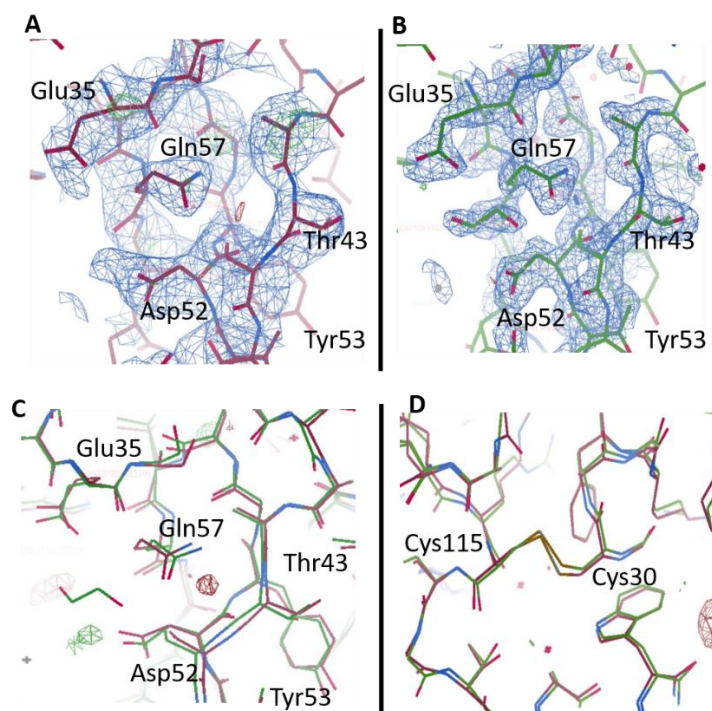


Figure 5. Lysozyme structural maps showing the active site pocket. The electron density map with the omit map displayed for the active site region of lysozyme in **A.** the first hit structure and **B.** the second hit structure. **C.** Shows the first (red) and second (green) hit structures superimposed and overlaid with the difference electron density (DED) map for the active site and **D.** shows the DED maps for a representative di-sulfide bond Cys115-Cys30. No differences density is detected between

the first and second hit structures. **E.** The 2Fo-Fc map at 1 sigma is shown in blue and difference maps at 3 sigma are shown in green (positive) and red (negative) density.

A statistical comparison between the three data sets is shown in Table 2 and 3. The Wilson plots and CC* for all data sets are included in the supplementary information (Supplementary Figure S4). Interestingly, the second hit data set quality is slightly better compared to the first hit data set which is reflected in its overall structural resolution, number of unique reflections and the CC values. Why the first hit data set is of poorer quality, is likely due to the non-uniform beam shape. The more the crystal is exposed within the beam the more intense the peaks are and this can be considered proportional to the resolution (where the most intense peaks are at lower resolution and the higher resolution are often less intense). Therefore, if the beam has a longer weaker tail at one end, diffraction from the crystal would be much weaker, with peaks below the SNR level remaining undetected, therefore affecting the number of unique reflections identified in the highest resolution shell. This can be seen in the YAG images (Supplementary Figure S3B) where the tail extends further out on the first hit side having less intensity compared to the tail region on the opposing side where the second hit region is located. As the stream files were obtained after hit-finding and indexing, the detection of first hits was not factored into the SNR threshold, and this may explain why we observe a reduced number of unique reflections in the first hit data set (2,072) compared to the second hit data set (7,263). The CC* is a commonly used metric to assess data quality for structure determination.⁴⁴ A comparison between the CC* for all three data sets reveal that the data quality varies between the different jet speeds which is reflected in the resolution. The first hit data had the lowest CC* value with a decline in data quality observed above 3.0 Å resolution (Supplementary Figure S4D), therefore a 3.1 Å resolution cut off was used for the structure refinement. With this cut off, the CC* values for the first and second hits were comparable; 0.873 and 0.764, respectively (Table 2). CrystFEL was also used to compare the structure factor of the first and second hit data sets to compare their quality. For the quality comparison, the first and second hit data sets were split in half; the half dataset for the second hits was then compared to the other half dataset from the first hits and vice versa and this resulted in stable CC* values up to a resolution of 2.5 Å (Supplementary Figure S4D, green line), confirming there is good agreement between the two datasets. Further to this, the single hit data set (i.e. data from crystals only hit once) was also split and compared to the first and second hit data showing a similar degree of consistency in terms of data quality (Supplementary Figure S4D, purple line). Hence, aside from the increase in the number of reflections in the second hit data – resulting in higher resolution data due to the higher incident intensity, the data quality between first, second, and single hits was consistent.

Table 2 SFX data collection and processing statistics for lysozyme single crystal data as well as the multi-hit (first and second) data sets. Values for the outer shell are given in parentheses.

Data Set	Single Hit	First Hit	Second Hit
Diffraction source	EuXFEL	EuXFEL	EuXFEL
Photon Energy (mean value, eV)	9232	9232	9232
Pulse energy at sample (assuming 50% beamline transmission, μJ)	290	290	290
Wavelength (\AA)	1.3	1.3	1.3
Temperature (K)	293	293	293
Detector	1-megapixel AGIPD	1-megapixel AGIPD	1-megapixel AGIPD
Pulse length (fs)	50	50	50
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$
a, b, c (\AA)	79.30, 79.30, 37.73	79.30, 79.30, 37.73	79.30, 79.30, 37.73
α, β, γ ($^\circ$)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution range (\AA)	21.66-2.10 (2.155-2.10)	35.49-3.20 (3.28-3.20)	21.66-2.10 (2.155-2.10)
Indexed	10,106	962	962
No. of unique reflections	7,418 (535)	2072 (403)	7,263 (494)
Completeness (%)	99.84 (100)	93.6 (90.79)	97.75 (92.34)
Redundancy	47.24 (28.34)	4.55 (3.8)	6.66 (3.98)
$\langle I/\sigma(I) \rangle$	5.1 (4.2)	4.36 (8.3)	2.9 (3.1)
$\text{CC}_{1/2}$	0.906 (0.796)	0.412 (0.638)	0.615 (0.413)
CC^*	0.975 (0.942)	0.764 (0.882)	0.873 (0.764)
Overall B factor from Wilson plot (\AA^2)	19.58	32.8	20.4

Table 3 SFX and refinement statistics for lysozyme single hit crystal data and the multi-hit (first hit and second hit) data sets. Values for the outer shell are given in parentheses.

Data Set	Single Hit	First Hit	Second Hit
Resolution range (Å)	21.66-2.10 (2.155-2.10)	35.489–3.20 (3.285–3.20)	21.66-2.10 (2.155-2.10)
Completeness (%)	99.84 (100)	93.6 (90.79)	97.75 (92.34)
No. of reflections, working set	6681 (737)	1850 (129)	6547 (716)
No. of reflections, test set	479 (56)	207 (9)	445 (49)
Final R_{cryst}	0.152 (0.114)	0.314 (0.323)	0.249 (0.244)
Final R_{free}	0.216 (0.183)	0.426 (0.406)	0.299 (0.355)
R.m.s. deviations			
Bonds (Å)	0.0007	0.005	0.004
Angles (°)	1.453	1.447	1.254
Average B factors (Å ²)	19.58	32.8	20.4
Protein	20.61	11.44	21.44
Ligands	37.73	27.46	37.24
Ions	22.99	33.37	40.32
Waters	29.96	28.35	28.01
Ramachandran plot			
Most favoured (%)	98.43	88.98	96.85
Allowed (%)	1.57	11.02	3.15

Discussion

The use of different jet speeds during the experiment enabled an analysis of the speed of recovery of the jet due to the jet explosion. The reliable recovery of the liquid jet between consecutive pulses at

the jet speeds used in this experiment has been previously reported in the literature^{1,10} and won't be discussed in detail here. Briefly, the gap size used in this experiment is much smaller than the beam diameter. Thus, if the crystal is hit within the tail region of the beam (which is the case for the first hits), the gap will not reach the crystal before the next pulse arrives. Hence, the results presented here confirm previous reports that under standard experimental conditions, the pulse power, beam size, jet diameter, and jet speed, can be chosen to avoid any interaction of the crystal with the expanding front of the opening gap formed by a previous pulse^{1,10,45}.

The results from this experiment confirm that quality structural data can be collected from crystals which are hit more than once by a MHz XFEL pulse using a flow focusing injector. In fact, the second hit data set could be refined to a higher resolution (2.1 Å) compared to the first hit data set (3.1 Å). This allows for the optimisation of multi-hit tr-SFX experiments since the height of the beam profile can be widely tuned to match the distance travelled by the crystals between consecutive X-ray pulses. The multi-hit structures from the first and second hits in this experiment demonstrate that high-resolution data (e.g. < 2.5 Å), can be collected from micro-crystals with per pulse dose rates less than that generally required to induce radiation damage at room temperature. For the second hit data analysed, the data was of high quality, comparable to that of the single hit crystals demonstrating that radiation damage did not affect the structure. However, the first hit data could not be solved to the equivalent resolution, having a lower percentage completeness (33%) in the 2.1 Å resolution shell which can be explained by the asymmetry of the extreme tails when the beam was at its largest size. Given that hit finding and indexing did not consider optimisation of these weaker peaks and the aim of this proposal was to generate a structure, the peak finding algorithm was not optimised for identifying weaker peaks. Therefore by optimising the experiment, one may pick up more peaks in the higher resolution shell for the first hit data and possibly even identify third hits for the slowest jet speeds and improve these statistics, however, for the sake of this proof of concept we have demonstrated that analysis of first hit and second hit data is possible here. The same experiment is not currently possible at lower repetition rate sources using flow focusing injectors but as more high-repetition rate XFEL sources come online the opportunity to compare time-resolved data collected from the same crystal with femtosecond intra-pulse and sub-microsecond inter-pulse timing opens new possibilities for studying molecular dynamics. In 2019, the focusing optics at the SPB/SFX beamline at the EuXFEL were upgraded enabling the production of a long vertical line-focus aligned to the liquid jet providing better setup optimisation for this type of experiment.

The time interval between consecutive pulses at the EuXFEL is extremely short when compared to other non-MHz XFEL sources, however sub-microsecond pulse spacing is clearly still long on the time-scale of radiation damage⁴⁶. Two methods were used to independently determine the dose absorbed by the crystal during a single pulse. The RADDOSE-3D version 4 (X-FEL)³⁷ gave an estimated dose of 0.2 MGy, taking into account the photo electron escape and the FW of the beam. The second method gave an estimate of 0.165 MGy maximum dose per pulse received by the crystal within the beam using the approach of Marman *et.al.*⁴⁷ which includes photo electron escape and in addition specifically the absorbed dose as a function of the crystal position within the incident beam (Supplementary Figure S5). Both methods produced dose estimates similar to those published under the same experimental conditions, (0.5 MGy reported by Weideron *et. al.*¹). We note that the primary reason the absorbed dose per pulse is lower here is the fact that the FW beam size is larger in the present case to provide a large enough X-ray interaction region to generate multiple hits. At the XFEL it has previously been reported, due to the extremely short pulse duration, the per-pulse radiation damage limit, even at room temperature, could increase to the range of 30-150 MGy⁴⁸⁻⁵⁰ as it can outrun the slower process contributing to radiation damage. However, the effects of radiation damage at the XFEL as a function of pulse duration is very much an active and ongoing area of research^{49,51-53} and depends both on the dose rates as well as the molecular details of the sample. It is also known that under cryo-conditions protein crystals can typically withstand a radiation dose of 30 MGy^{54,55}, at RT (which we have assumed here) the radiation dose limit decreases by approximately two orders of magnitude, as a result of the diffusion of free radicals^{38,56,57}. Hence in the analysis presented here, we have adopted the most conservative view that the per pulse absorbed dose must be below 0.2 MGy⁵⁴ in line with room temperature experiments conducted at the synchrotron before radiation damage is to occur. These values fall within the dose estimates obtained for this crystal system where the dose of the crystal in the tail regions can be as low as 0.04 MGy (Supplementary Figure S5) and as high as 0.2 MGy central to the beam.

Our model for how a single crystal is able to produce multiple diffraction patterns has been validated by the observed changes in the measured integrated intensity profiles and explains why, for example, the second hit for the 42 m/s jet has a similar intensity profile to the single hit crystals. It also accounts for the fact that for the faster jet speeds the intensity profiles of the first and second hits are similar. Another interesting observation was that even within the extreme tails of the XFEL beam, where the absorbed dose is around 2 orders of magnitude lower than in the beam centre, there is still sufficient intensity to generate high-resolution diffraction data from crystals < 10 μm in diameter. This is consistent with previous published reported from the LCLS⁴⁸.

Based on our model for how multiple hits occur we can develop a set of parameters for users that can be employed to either maximise or minimise multiple crystal hits. The critical parameters for multi-hit tr-SFX are the beam size, jet speed, and pulse structure. The inter-pulse spacing used for the current experiment was 886ns (1.125 MHz), however, we can use our model to predict how multi-hit tr-SFX would work at 4.5 MHz. A comparison of the two regimes for the EuXFEL operating at 1.1 MHz and 4.5 MHz is provided in Figure 6.

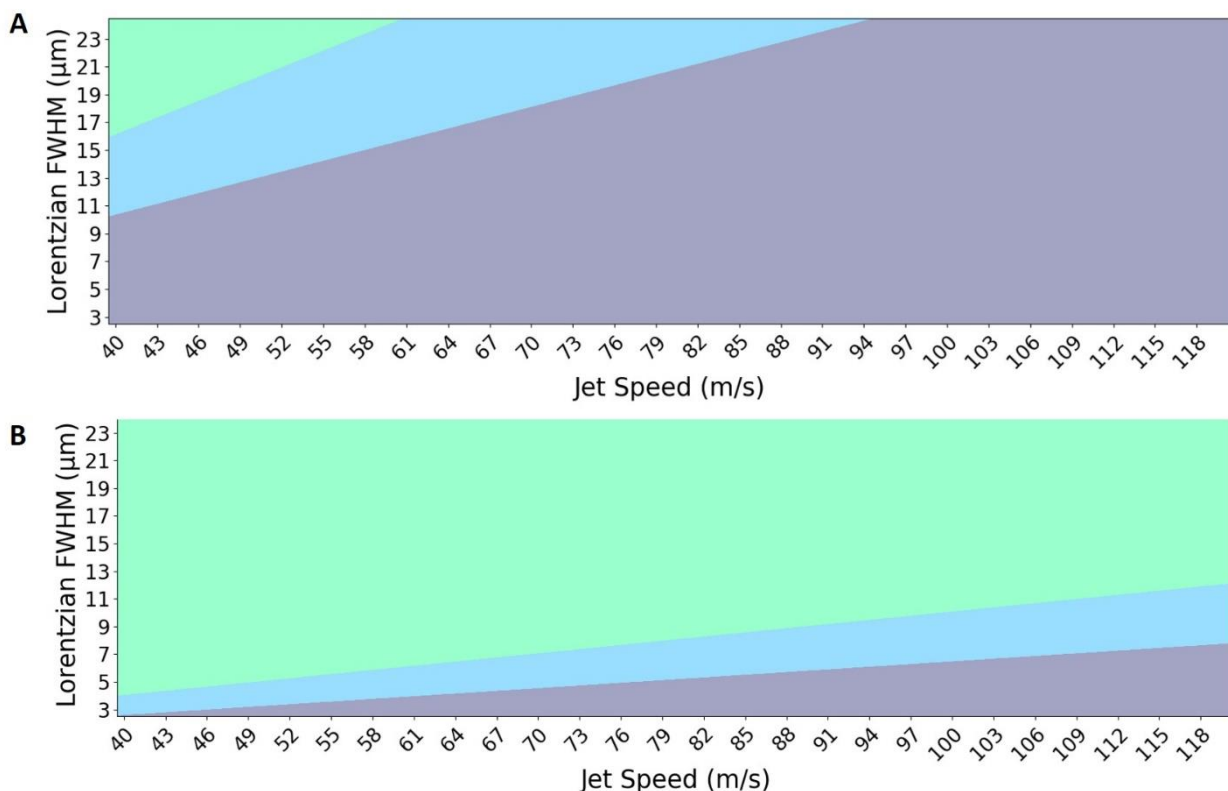


Figure 6. Parameters for optimizing the collection of double-hit data at the EuXFEL. **A.** EuXFEL repetition rate of 1.1 MHz during this experiment and **B.** EuXFEL repetition rate of 4.5 MHz. The green shaded area indicates parameter combinations that will result in double-hits that allow the second hit to occur within the horizontal FWHM of the beam; the blue shaded area indicates parameter combinations that will result in double-hits that allow a second hit to occur within in the tail region of the beam, and the grey shaded area indicates parameter combinations that will result in only single hits. This analysis is independent of crystal size (i.e. crystal centre-to-crystal centre hits).

In addition to the integrated intensity, we also studied the resolution of the diffraction data from both single and double hits. A comparison of the second hit and single hit patterns shows a difference in the resolution of 1 Å which can be explain be the beam profile and SNR detection levels. This is also supported by the structural analysis where both the first and second hit structures do not exhibit any

noticeable signs of radiation damage. Hence, the key finding by this collaboration is that the diffraction data, statistics, structural analysis, and radiation damage calculations all point to the same conclusion: that the crystal does not appear to experience measurable radiation damage in the first hit by the XFEL beam for this crystal system.

The availability of MHz XFEL sources has created many new opportunities for exploiting their unique pulse structure⁵⁸. The approach described and demonstrated here, measuring structure from the same single crystal in-flight with consecutive MHz XFEL pulses, opens up the possibility of correlating and analysing structural dynamics on timescales ranging from 222 ns (4.5 MHz) to 1.8 μ s (approximated based on a crystal travelling at 42 m/s through an interaction region with a mean beam width of 65.8 μ m). This covers a range of molecular dynamics of interest to structural biology including helix motions, side chain rotations, and protein folding and unfolding. The experimental conditions used for the current experiment are readily achievable at the EuXFEL and have been used for a recent publications involving tr-SFX⁵⁹ demonstrating that multi-hit tr-SFX is a viable option for current and future users of this facility. Two types of experiments have been established by the multi-hit tr-SFX collaboration as benefitting from this mode of operation, the first is chemically triggered SFX where a ‘mix-and-inject’ set up is used on the beamline. Using multi-hit tr-SFX would allow the dynamics of each single crystal to be probed at a minimum of two timepoints with sub-microseconds separation. The second type of experiment to benefit from multi-hit tr-SFX is optical pump/X-ray probe experiments where two active structures (having been activated by the optical pump) can be obtained from the same crystal. While the experiment performed here was not tr-SFX, we note that the second hit data set collected here had virtually identical quality to the single hit data. These results therefore show that multi-hit tr-SFX is possible, as two undamaged structures can be obtained from the same crystal. Whilst it should be noted that for proteins that are particularly radiation sensitive, multi-hit tr-SFX may not be an option, the fact that the two independent structures from the first and second hit do not exhibit any detectable signs of radiation damage is encouraging. Based on the criteria we have established here to observe multiple hits using MHz XFEL sources and via discussions with the EuXFEL staff, the current work paves the way to multi-hit tr-SFX becoming an established mode of operation for users.

Online Method

The SPB/SFX beam was focused using compound refractive lenses (CRLs). Lysozyme (HEWL) crystals were delivered to a 1.125 MHz XFEL beam using a 3D printed gas-focused dynamic virtual

nozzle (GDVN)^{31,32,60} nozzle with 60 μm gas and 50 μm liquid orifices. Various sample delivery jet speeds ranging between 40-102 m/s, were tested during the experiment. All the data generated from these speeds were merged to generate the first and second hit data sets. The majority of the data was generated from three speeds, 42 m/s, 78 m/s and 102 m/s, and these were used to generate statistics for the beam profile. The two slowest jet speeds were checked experimentally in lab by dual-pulse imaging^{1,61}, whilst the fastest jet speed (102 m/s) was theoretically determined (see supplementary data). A summary of the liquid and gas flow rates and their equivalent jet speeds are shown in Table 1, alongside the theoretically calculated 102 m/s jet speed. The lysozyme crystal size varied between 6 x 6 x 6 μm and 8 x 8 x 8 μm as characterised via an optical microscope. All of the data used for this paper (run numbers r0066-r0087, r0145-r0150, r0153) is available from CXIDB-ID80 (www.cxidb.org/id-80.html).

Identification of multi-hits in the data

The CrystFEL stream files with the indexed crystal data¹ were used to identify the crystals classified as multi-hits for the different injector speeds. The stream files were analysed for multi-hits using the *whirligig* script from the *CrystFEL* crystallography suite⁴³, which defines crystals hit twice as those that had similar crystallographic orientations ($< 5^\circ$ change in angle between each pair of basis vectors) in consecutive frames. Based on this analysis we confirmed that a proportion of single crystals were hit twice, classifying them as ‘multi-hit crystals’. In addition, the lengths of the basis vectors had to be within 10% agreement to qualify as a multi-hit. Filtering using the *whirligig* program based on these criteria alone does not account for the pulse train format of the EuXFEL. Therefore, python code was developed to filter out any adjacent frames that were not from the same pulse train. This data was further sorted into three separate categories: data from crystals hit just once, data from the first hit of crystals hit twice (designated ‘first hit’), and data from the second hit of crystals hit twice (designated ‘second hit’). These three stream files were then used as the input files for further analysis in Python, using multiple parameters (integrated intensity, a^* , b^* , c^* , h , k , l) within the stream files to calculate $1/d$ versus normalised integrated intensity plots.

Python code was also used to analyse the *CrystFEL* stream files, where the change in orientation between all consecutive images was calculated in an identical manner to the *whirligig* program. The output was plotted as a histogram, showing the change in angle (degrees) for consecutive images, to confirm those consecutive hits that were selected as the double hit data set were truly from the same crystal.

Analysis of data quality

Merging and scaling of the Bragg peaks were performed using Partialator in the *CrystFEL suite version 8*^{43,62}. Figures of merit were calculated using `compare_hkl` (Rsplit, $CC_{1/2}$, CC^*) and `check_hkl` (SNR, multiplicity, completeness), that are also part of the *CrystFEL suite*^{43,62}. To generate a complete dataset to compare the first hit with the second hit on the same crystal, data collected from all jet speeds (40-102 m/s) were merged to form the ‘multi-hit’ structural data sets. Data from a single speed (42 m/s) was used to analyse and solve the single hit structure. The statistics for the single hits, first hits and second hit data sets are presented in Tables 2 and 3 and the Wilson plots for all the data sets are shown in the supplementary section (Supplementary, Figure S4).

Structure refinement was performed in Collaborative Computational Project 4 interactive version 2 (CCP4i2) using the MTZ output from CrystFEL. A solvent free version of lysozyme (PDB code 6FTR) was used as the initial starting model for molecular replacement in Phaser⁶³ and the R_{free} flags were generated (utilizing 10% of the data) followed by iterative cycles of Refmac5⁶⁴ refinement and rebuilding of the model in Coot⁶⁵. The MolProbity⁶⁶ and Xtriage (Phenix)⁶⁷ tools were used to validate the model. To compare the first and second hit structures, difference electron density maps (DED)^{7,68-70}, were generated in CCP4i2. The two data sets were scaled, and the difference amplitudes determined by subtracting the observed structure factor amplitudes of the first hit data set from those of the second hit data set. The DED maps were then calculated by using the difference amplitudes and phases from the first hit data. The DED maps generated positive electron density (green, indicating the presence of increased density in the second hit data compared to the first hit data) and negative electron density (red, indicating decreased density in the second hit data compared to the first hit data) areas scaled to $\pm 3\sigma$ contour levels. Figure 5A and C highlights the quality of the electron density maps surrounding the active site and a di-sulphide bond in lysozyme (which typically are more sensitive to the effects of radiation damage).

X-ray beam profile analysis

The nominal energy for the SFX data collection was 9.232 keV. The beam size and beam profile were estimated based on 6773 individual YAG images collected using the in-line microscope positioned within in the chamber. The optical images were generated from single shots using a 15 μm thick Ce:YAG screen (see Supplementary Figure S3 and Methods). The point spread function (PSF) of the YAG was determined, based on published estimates⁷¹ (see Supplementary Figure S3), to be 2 μm . Using an Edmund optics standard, the optical microscope resolution was determined to be 8 μm . A

Lorentzian distribution was fitted to the optical images using Python code. The actual beam size was determined via a convolution, taking into account both the PSF for the YAG and optical microscope. A FWHM ($\text{FWHM} = 2\gamma$) for each image was determined from a Lorentzian fit to the beam profile using a 3×3 median filter and a 7.5% noise threshold applied to the data. A Full Width ($\text{FW} = 7.04\gamma$) value of 82.38% was also determined for each image, and the overall mean, minimum and maximum FW calculated for the complete data set. The mean, minimum, and maximum FWHMs and FW were calculated. Note that the experimentally determined beam size based on analysis of the occurrence of multi-hits as a function of jet speed (Figure 1) was consistent with the X-ray beam size determined optically.

Crystal transit through the X-ray beam

The distance travelled by the crystal through the X-ray beam can be calculated based on the jet speed (see Table 1) and beam size. Using the upper limit on the measured size of the lysozyme crystals (i.e. $8 \times 8 \times 8 \mu\text{m}$) and the lowest possible jet speed (accounting for a 5% uncertainty – see Table 1) gives the minimum distance a single crystal could travel whilst still interacting with two consecutive X-ray pulses, spaced 886 ns apart. Assuming that at least $1 \mu\text{m}$ of the crystal needs to interact with the X-ray beam to generate a diffraction pattern, the minimum distance travelled by a single crystal hit twice by the XFEL beam was calculated. For the fastest jets, multi-hits were only possible in the presence of the maximum FW of the beam.

Radiation dose calculations

To determine the dose the crystal receives during its interaction with the X-ray pulse two independent methods were used. The first method is based on monte-carlo modelling of the primary photoelectron trajectories, taking in account any photo electron escape from the crystal that might occur subsequent to the crystal interacting with the X-ray pulse. This approach to determining the dose absorbed by the crystal was adapted from a discrete simulation of radiation damage model based on Marman *et. al.*⁴⁷. The X-ray beam was modelled as a symmetric 2-dimensional Lorentzian distribution with a full width of $65.5 \mu\text{m}$ (accounting for 99% of the X-ray flux). The total X-ray interaction area accounted for in the model was $100 \mu\text{m}$ in diameter allowing the crystal to be simulated prior to entering and thus only partially exposed to the X-ray beam. The spatial resolution of the model was $0.1 \mu\text{m}$. This was sufficient to allow us to determine the dose received by the crystal at different points within the beam (i.e. within the tail region or within the central portion of the beam) as it travels through the X-ray interaction region. The Lorentzian flux distribution and the appropriate X-ray cross-sections for photo-ionisation, elastic scattering, and Compton scattering at this X-ray energy were used to calculate the relevant interaction rates at any given point within the beam.⁷² The crystal was modelled as a

square ($8\ \mu\text{m} \times 8\ \mu\text{m}$); and was assumed to pass through the centre of the beam travelling perpendicular to the y-axis. The absorbed dose was calculated at 33 independent positions along the x-axis. These positions corresponded to the crystal moving from the edge of the beam to its centre in $1\ \mu\text{m}$ steps; therefore 7 of the 33 positions corresponded to a ‘partial’ hit, where part of the crystal remained unexposed and the remaining 26 positions corresponded to full exposure at varying incident flux densities. At each position, the dose absorbed by the exposed portion of the crystal was calculated. A summary of this approach to calculating the radiation damage is presented in Supplementary Figure S5. We note that in the tail regions of the beam the dose received by the crystal can reduce by as much as two orders of magnitude. As an independent confirmation of the first approach, a second method, based on RADDOSE-3D version 4^{37,73} using the XFEL option, assuming a Gaussian distribution for the beam (no collimation) and maximum full width of beam was used and the crystal was placed central to the beam. In summary, using these two independent methods, the dose absorbed by the crystal during both the first and second hits was calculated.

Acknowledgements

We acknowledge European XFEL in Schenefeld, Germany, for provision of X-ray free electron laser beam-time at Scientific Instrument SPB/SFX and use of the XBI biological sample preparation laboratory, enabled by the XBI User Consortium and would like to thank the instrument group and facility staff for their assistance and everyone who attended the first experiment. We would also like to acknowledge the funding support for B.A., C.D, K.A.N and S.H from Australian Research Council Center of Excellence in Advanced Molecular Imaging (CE140100011), www.imagingcoe.org and the Australian Nuclear Science and Technology Organisation (ANSTO); the International Synchrotron Access Program (ISAP) managed by the Australian Synchrotron, part of ANSTO, and funded by the Australian Government. M.L.S., M.A.C. and M.F. performed, in part, under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. M.L.S., M.A.C. and M.F. were supported by the NIH grant 1R01GM117342-01. I.B acknowledges support by Ministry of Education, Science, Research and Sport of the Slovak Republic and by grant from the Slovak Research and Development Agency under contract APVV-14-0181. M.L.S., M.A.C. and M.F. performed, in part, under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. M.L.S., M.A.C. and M.F. were supported by the NIH grant 1R01GM117342-01. T.V is a recipient of ARC DECRA (DE170100783). B.K work is funded by National Health and Medical Research Council (NHMRC) grants 1107804 and 1160570 to BK and TV; BK is Australian Research Council (ARC) Laureate Fellow (FL180100109). F.C-M and A.M.G-C acknowledge Ministerio de Economía y Competitividad (Spain), Plan Estatal 2013-2016 Retos, project DPI2016-78887-C3-1-R. S.B.

acknowledged the Helmholtz Initiative and Networking Fund through the Young Investigators Program and the Deutsche Forschungsgemeinschaft SFB755/B03.

Author Contributions

The experiment was conceived by H.C., A.B. and A.P.M. as an open experiment for the EuXFEL user community. A.B. coordinated and led the collaboration. B.A. and C.D conceived the idea of studying the multi-hit phenomena and edited the manuscript. S.H. developed the model for double hit scenario and wrote the manuscript. H.K, A.V.M and K.G. contributed to the beam profile data collection and analysis. H.M analysed dose measurements. M.O.W, A.B., A.P.M., K.A.N. were involved in discussions. R.B. was the instrument scientist responsible for this experiment. A.P.M. led development of the SPB/SFX instrument. M.O.W and S.B. led the sample injection team. C.D., D.O., R.S., I.B., M.A.C., M.F., M.S., T.V. B.K., K.D., K.O, R.L., G.P., S.H., M.Me., G.P., P.L.X., and C.S. contributed to sample preparation. S.B, A.B., D.C.FM., M.T., M.O.W., J.S., M.H., J.K., J.B., R.G., L.G., and L.A., contributed to sample injector delivery. A.B, H.C., M.S., O.Y., and M.J-H. contributed to data processing. A.P.M., R.B., J.S., S.S., K.G., Y.K., M.M., G.M., A.R., P.V., B.W., and J.S., developed and operated the SPB/SFX instrument at EuXFEL. F.C-M., A.M.G-C., S.M., J.A.S., B.A. M.A.C., M.F., and M.S., participated in the XFEL 2012 experiment. T.L., K.W., D.B., L.M., G.P., N.A-Q., S.B., T.M., C.Dan., A.S., C.X., S.H., developed XFEL DAQ and controls. J.S-D., T.L., and A.K. developed and responsible for AGIPD. The manuscript was written by S.H., C.D. and B.A. with input from all authors.

Competing Interests statement

The authors declare no competing interest.

References

- 1 Wiedorn, M. O. *et al.* Megahertz serial crystallography. *Nature Communications* **9** (2018).
- 2 Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73-U81, doi:10.1038/nature09750 (2011).
- 3 Suga, M. *et al.* Native structure of photosystem II at 1.95 angstrom resolution viewed by femtosecond X-ray pulses. *Nature* **517**, 99-U265, doi:10.1038/nature13991 (2015).
- 4 Boutet, S. *et al.* High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography. *Science* **337**, 362-364, doi:10.1126/science.1217737 (2012).
- 5 Gati, C. *et al.* Atomic structure of granulin determined from native nanocrystalline granulovirus using an X-ray free-electron laser. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 2247-2252 (2017). doi:10.3390/ijms20143421 (2019).
- 7 Pandey, S. *et al.* Time-resolved serial femtosecond crystallography at the European XFEL. *Nature*

Methods **17**, 73–, doi:10.1038/s41592-019-0628-z (2020).

- 8 Grunbein, M. L. *et al.* MHz data collection of a microcrystalline mixture of different jack bean proteins. *Scientific Data* **6**, doi:ARTN 18, 10.1038/s41597-019-0010-0 (2019).
- 9 Gisriel, C. *et al.* Membrane protein megahertz crystallography at the European XFEL. *Nature Communications* **10**, doi:10.1038/s41467-019-12955-3 (2019).
- 10 Yefanov, O. *et al.* Evaluation of serial crystallographic structure determination within megahertz pulse trains. *Structural Dynamics-Us* **6**, 064702, doi:A10.1063/1.5124387 (2019).
- 11 Orville, A. M. Recent results in time resolved serial femtosecond crystallography at XFELs. *Current Opinion in Structural Biology* **65**, 193-208, doi:10.1016/j.sbi.2020.08.011 (2020).
- 12 Nass, K. *et al.* Structural dynamics in proteins induced by and probed with X-ray free-electron laser pulses. *Nature Communications* **11**, 1814, doi:10.1038/s41467-020-15610-4 (2020).
- 13 Aquila, A. *et al.* Time-resolved protein nanocrystallography using an X-ray free-electron laser. *Optics Express* **20**, 2706-2716 (2012).
- 14 Tenboer, J. *et al.* Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science* **346**, 1242-1246 (2014).
- 15 Orville, A. M. Entering an era of dynamic structural biology. . . *Bmc Biology* **16**, 55 doi:10.1186/s12915-018-0533-4 (2018).
- 16 Schmidt, M. Time-Resolved Macromolecular Crystallography at Pulsed X-ray Sources. *Int J Mol Sci* **20**, 1401, doi:10.3390/ijms20061401 (2019).
- 17 Liu, W., Wacker, D., Wang, C., Abola, E. & Cherezov, V. Femtosecond crystallography of membrane proteins in the lipidic cubic phase. *Philosophical Transactions of the Royal Society B-Biological Sciences* **369**, 20130314, doi:10.1098/Rstb.2013.0314 (2014).
- 18 Weierstall, U. *et al.* Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nature Communications* **5** (2014).
- 19 Nogly, P. *et al.* Lipidic cubic phase serial millisecond crystallography using synchrotron radiation. *Iucrj* **2**, 168-176 (2015).
- 20 Hadian-Jazi, M., Berntsen, P., Marman, H., Abbey, B. & Darmanin, C. Analysis of Multi-Hit Crystals in Serial Synchrotron Crystallography Experiments Using High-Viscosity Injectors. *Crystals* **11**, 49, doi:10.3390/cryst11010049 (2021).
- 21 Grunbein, M. L. *et al.* Megahertz data collection from protein microcrystals at an X-ray free-electron laser. *Nature Communications* **9**, 3487, doi:10.1038/s41467-018-05953-4 (2018).
- 22 Volkman, B. F., Lipson, D., Wemmer, D. E. & Kern, D. Two-state allosteric behavior in a single-domain signaling protein. *Science* **291**, 2429-2433, doi:DOI 10.1126/science.291.5512.2429 (2001).
- 23 Moy, F. J. *et al.* Assignments, Secondary Structure, Global Fold, and Dynamics of Chemotaxis-Y Protein Using 3-Dimensional and 4-Dimensional Heteronuclear (C-13,N-15) Nmr-Spectroscopy. *Biochemistry-Us* **33**, 10731-10742, doi:10.1021/bi00201a022 (1994).
- 24 Feher, V. A. & Cavanagh, J. Millisecond-timescale motions contribute to the function of the bacterial response regulator protein Spo0F. *Nature* **400**, 289-293 (1999).

- 25 Nicolai, A., Delarue, P. & Senet, P. Decipher the Mechanisms of Protein Conformational Changes Induced by Nucleotide Binding through Free-Energy Landscape Analysis: ATP Binding to Hsp70. *Plos Comput Biol* **9**, e1003379, doi:10.1371/journal.pcbi.1003379 (2013).
- 26 Xie, M. Z. *et al.* Functional protein dynamics on uncharted time scales detected by nanoparticle-assisted NMR spin relaxation. *Sci Adv* **5**, doi:ARTN eaax5560, 10.1126/sciadv.aax5560 (2019).
- 27 Hansen, A. L. & Kay, L. E. Quantifying millisecond time-scale exchange in proteins by CPMG relaxation dispersion NMR spectroscopy of side-chain carbonyl groups. *J Biomol Nmr* **50**, 347-355, doi:10.1007/s10858-011-9520-6 (2011).
- 28 Feher, G. & Okamura, M. Y. The primary and secondary acceptors in bacterial photosynthesis: II. The structure of the Fe²⁺-Q(-) complex. *Applied Magnetic Resonance* **16**, 63-100 (1999).
- 29 Clabbers, M. T. B. *et al.* MyD88 TIR domain higher-order assembly interactions revealed by microcrystal electron diffraction and serial femtosecond crystallography. *Nature Communications* **12**, 2578, doi:10.1038/s41467-021-22590-6 (2021).
- 30 Kirkwood, H. J. *et al.* Initial observations of the femtosecond timing jitter at the European XFEL. *Optics Letters* **44**, 1650-1653, doi:10.1364/Ol.44.001650 (2019).
- 31 Oberthuer, D. *et al.* Double-flow focused liquid injector for efficient serial femtosecond crystallography (vol 7, 44628, 2017). *Scientific Reports* **7**, doi:10.1038/srep46846 (2017).
- 32 Ganan-Calvo, A. M. Generation of steady liquid microthreads and micron-sized monodisperse sprays in gas streams. *Physical Review Letters* **80**, 285-288, doi:DOI 10.1103/PhysRevLett.80.285 (1998).
- 33 DePonte, D. P. *et al.* Gas dynamic virtual nozzle for generation of microscopic droplet streams. *J Phys D Appl Phys* **41**, 195505, doi:10.1088/0022-3727/41/19/195505 (2008).
- 34 Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752-757, doi:Doi 10.1038/35021099 (2000).
- 35 Stan, C. A. *et al.* Liquid explosions induced by X-ray laser pulses. *Nat Phys* **12**, 966-971, doi:10.1038/Nphys3779 (2016).
- 36 Gisriel, C. *et al.* Membrane protein megahertz crystallography at the European XFEL *Nature Communications* **10**, 5021 (2019).
- 37 Dickerson, J. L., McCubbin, P. T. N. & Garman, E. F. RADDOS-XFEL: femtosecond time-resolved dose estimates for macromolecular X-ray free-electron laser experiments. *Journal of Applied Crystallography* **53**, 549-560, doi:10.1107/S1600576720000643 (2020).
- 38 de la Mora, E. *et al.* Radiation damage and dose limits in serial synchrotron crystallography at cryo- and room temperatures. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 4142-4151, doi:10.1073/pnas.1821522117 (2020).
- 39 Beyerlein, K. R. *et al.* FELIX: an algorithm for indexing multiple crystallites in X-ray free-electron laser snapshot diffraction images. *Journal of Applied Crystallography* **50**, 1075-1083, doi:10.1107/s1600576717007506 (2017).
- 40 Ginn, H. M. *et al.* TakeTwo: an indexing algorithm suited to still images with known crystal parameters. *Acta Crystallographica Section D-Structural Biology* **72**, 956-965,

doi:10.1107/S2059798316010706 (2016).

- 41 Gildea, R. J. *et al.* New methods for indexing multi-lattice diffraction data. *Acta Crystallographica Section D-Structural Biology* **70**, 2652-2666, doi:10.1107/S1399004714017039 (2014).
- 42 Barty, A. *et al.* Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *Journal of Applied Crystallography* **47**, 1118-1131, doi:10.1107/S1600576714007626 (2014).
- 43 White, T. A. *et al.* CrystFEL: a software suite for snapshot serial crystallography. *Journal of Applied Crystallography* **45**, 335-341 (2012).
- 44 Karplus, P. A. & Diederichs, K. Linking Crystallographic Model and Data Quality. *Science* **336**, 1030-1033, doi:10.1126/science.1218231 (2012).
- 45 Ganan-Calvo, A. M. Scaling Laws of an Exploding Liquid Column under an Intense Ultrashort X-Ray Pulse. *Physical Review Letters* **123**, 064501 doi:10.1103/PhysRevLett.123.064501 (2019).
- 46 Holton, J. M. A beginner's guide to radiation damage. *Journal of Synchrotron Radiation* **16**, 133-142, doi:10.1107/S09090495090004361 (2009).
- 47 Marman, H., Darmanin, C. & Abbey, B. The Influence of Photoelectron Escape in Radiation Damage Simulations of Protein Micro-Crystallography. *Crystals* **8**, doi:ARTN 267 10.3390/cryst8070267 (2018).
- 48 Boutet, S. *et al.* High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography. *Science* **337**, 362-364 (2012).
- 49 Nass, K. Radiation damage in protein crystallography at X-ray free-electron lasers. *Acta Crystallographica Section D-Structural Biology* **75**, 211-218, doi:10.1107/S2059798319000317 (2019).
- 50 Lomb, L. *et al.* Radiation damage in protein serial femtosecond crystallography using an x-ray free-electron laser. *Physical Review B* **84** (2011).
- 51 Abbey, B. *et al.* X-ray laser-induced electron dynamics observed by femtosecond diffraction from nanocrystals of Buckminsterfullerene. *Sci Adv* **2**, doi:UNSP e1601186 10.1126/sciadv.1601186 (2016).
- 52 Amin, M., Badawi, A. & Obayya, S. S. Radiation Damage in XFEL: Case study from the oxygen-evolving complex of Photosystem II. *Scientific Reports* **6**, 36492 doi: 10.1038/srep36492 (2016).
- 53 Standfuss, J. Membrane protein dynamics studied by X-ray lasers - or why only time will tell. *Current Opinion in Structural Biology* **57**, 63-71, doi:10.1016/j.sbi.2019.02.001 (2019).
- 54 Owen, R. L. *et al.* Outrunning free radicals in room-temperature macromolecular crystallography. *Acta Crystallographica Section D-Structural Biology* **68**, 810-818, doi:10.1107/S0907444912012553 (2012).
- 55 Owen, R. L., Rudino-Pinera, E. & Garman, E. F. Experimental determination of the radiation dose limit for cryocooled protein crystals. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 4912-4917, doi:10.1073/pnas.0600973103 (2006).
- 56 Ebrahim, A. *et al.* Dose-resolved serial synchrotron and XFEL structures of radiation-sensitive

- metalloproteins. *Iucrj* **6**, 543-551, doi:10.1107/S2052252519003956 (2019).
- 57 Gotthard, G. *et al.* Specific radiation damage is a lesser concern at room temperature. *Iucrj* **6**, 665-680, doi:10.1107/S205225251900616x (2019).
- 58 Mills, G., Bean, R. & Mancuso, A. P. First Experiments in Structural Biology at the European X-ray Free-Electron Laser. *Appl Sci-Basel* **10**, 3642 doi:10.3390/app10103642 (2020).
- 59 Echelmeier, A. *et al.* Segmented flow generator for serial crystallography at the European X-ray free electron laser. *Nature Communications* **11**, 4511, doi:10.1038/s41467-020-18156-7 (2020).
- 60 DePonte, D. P., Nass, K., Stellato, F., Liang, M. & Chapman, H. N. in *Advances in X-Ray Free-Electron Lasers: Radiation Schemes, X-Ray Optics, and Instrumentation* Vol. 8078 *Proceedings of SPIE* (eds T. Tschentscher & D. Cocco) (2011).
- 61 Knoska, J. *et al.* Ultracompact 3D microfluidics for time-resolved structural biology. *Nature Communications* **11**, doi:10.1038/s41467-020-14434-6 (2020).
- 62 White, T. A. *et al.* Recent developments in CrystFEL. *Journal of Applied Crystallography* **49**, 680-689, doi:10.1107/s1600576716004751 (2016).
- 63 McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).
- 64 Murshudov, G. N. Low resolution refinement in the program - REFMAC. *Acta Crystallographica a-Foundation and Advances* **67**, C134-C134, doi:Doi 10.1107/S0108767311096711 (2011).
- 65 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).
- 66 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D-Structural Biology* **66**, 12-21, doi:10.1107/S0907444909042073 (2010).
- 67 Afonine, P. V. *et al.* phenix.model_vs_data: a high-level tool for the calculation of crystallographic model and data statistics. *Journal of Applied Crystallography* **43**, 669-676, doi:10.1107/S0021889810015608 (2010).
- 68 Tenboer, J. *et al.* Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science* **346**, 1242-1246, doi:10.1126/science.1259357 (2014).
- 69 Olmos, J. L., Jr. *et al.* Enzyme intermediates captured "on the fly" by mix-and-inject serial crystallography. *Bmc Biology* **16**, doi:10.1186/s12915-018-0524-5 (2018).
- 70 Claesson, E. *et al.* The primary structural photoresponse of phytochrome proteins captured by a femtosecond X-ray laser. *Elife* **9**, e53514 doi:10.7554/eLife.53514 (2020).
- 71 Koch, A., Raven, C., Spanne, P. & Snigirev, A. X-ray imaging with submicrometer resolution employing transparent luminescent screens. *Journal of the Optical Society of America a-Optics Image Science and Vision* **15**, 1940-1951, doi:Doi 10.1364/Josaa.15.001940 (1998).
- 72 Berger, M. J. *et al.* NIST Standard Reference Database 8 (XGAM). doi:https://dx.doi.org/10.18434/T48G6X (2010).
- 73 Bury, C. S., Brooks-Bartlett, J. C., Walsh, S. P. & Garman, E. F. Estimate your dose: RADDose-3D. *Protein Sci* **27**, 217-228, doi:10.1002/pro.3302 (2018).

3.5 Summary

The indexed data was filtered in order to identify the presence of crystals consecutively hit by multiple X-ray pulses. The supplementary information associated with this paper can be found in Appendix A. This filtering was performed by creating a Python script. This Python script incorporated a CrystFEL program called Whirligig, that worked to identify consecutive indexed frames that were of similar orientation. After this initial step, further checks and filters were applied in order to handle the additional data structure output caused by the unique pulse-train structure of the EuXFEL.

From this first analysis, we concluded that there were crystals that had survived the first pulse and had been hit a second time by the X-ray beam. Based on this conclusion, it was determined that a complete analysis of the beam diameter should be undertaken in order to provide a more detailed picture. The YAG images of the beam were therefore analysed, and the minimum, maximum, and mean FW and FWHM were calculated from those 6773 images. This beam diameter analysis allowed us to model the crystal trajectory through the beam so that we could determine where within the beam the crystals were capable of being hit more than once. The FW and FWHM calculations were also used to estimate the different radiation doses received by those crystals hit once or twice by the beam.

In order to eliminate any doubt that filtering and identification of multi-hit crystals was valid, the change in orientation was calculated between all of the consecutive indexed diffraction patterns. A histogram was created to display the frequency of these changes in orientation in order to determine if the correct level of filtering was applied when assigning a diffraction pattern as being from a crystal hit once, or multiple times. If, multi-hits had been falsely identified, this would have been clearly displayed on the histogram as an even frequency distribution across all orientation changes. The histogram instead displayed a peak corresponding the presence of multi-hits, confirming that the correct level of filtering was used to identify occurrences of multi-hits.

After filtering the diffraction patterns into three data subsets, depending on whether the diffraction pattern was from a crystal hit once or twice, and whether that pattern was a result of the first or second X-ray pulse exposure, the structure of HEWL was solved for

each of these three data subsets. The three solved structures were then compared in order to determine if the structures showed any signs of radiation damage. No visible structural changes were present, leading us to believe minimal radiation induced damage occurred prior to a second diffraction pattern being obtained. Whilst this experiment confirmed the presence of multi-hit SFX data enabling the retrieval of structures from consecutive hits, additional optimisation of the experimental parameters is possible to increase the likelihood that multiple hits are observed. With an increase in the fraction of multi-hits it will be possible to confirm whether, at the highest structural resolutions, radiation induced damage can be observed.

In order to optimise the experimental parameters to increase the chances of multi-hits, two diagrams were created that show what combination of beam size and jet speed maximise the occurrence of multi-hits. These diagrams can be used to help future experiments in identifying the best combination of parameters that could be used to obtain the highest multi-hit diffraction rate.

In this chapter I have demonstrated that it is possible to obtain multiple diffraction patterns from a single crystal at an XFEL. This work lays the foundation for follow up studies that will investigate how multi-hits could be exploited for novel tr-SFX experiments as well as determining if there could be a possible advantage to including multi-hits in terms of sample consumption.

Chapter 4

A Comparison of Serial Crystallography Delivery Systems

4.1 Myeloid Differentiation Primary-Response 88

Understanding how the innate immune system works is extremely important in helping to understand and prevent diseases and illnesses. Knowing the mechanisms and pathways involved in our immune response to invading pathogens will allow us to understand and develop therapeutics for a range of infectious and inflammatory diseases [177–179].

One class of proteins that play a major role in the innate immune response are the Toll-like receptor (TLR) proteins. Ten TLR proteins have been identified in humans, labelled TLR1 to TLR10. TLRs are potential targets for therapeutics, including for the treatment of viral, bacterial, mycobacterial, and fungal diseases, sepsis, chronic inflammation, autoimmune diseases, vaccine development, cancer, and atherosclerosis [179]. These TLR proteins span the cell membrane and are usually expressed on sentinel cells, a class of cells that act as a first line of defence and include macrophage and dendritic cells. Acting as "Pattern Recognition Receptors" (PRRs), the TLR proteins recognise and bind directly to "Pathogen Associated Molecular Patterns" (PAMPs) or, they bind to "Danger Associated Molecular Patterns" (DAMPs) that are released by dying or damaged cells as a warning [179–181]. This in turn activates cytoplasmic signalling, and recruitment of a subgroup of proteins

known as TLR adaptor proteins that include Myeloid Differentiation primary-response 88 (MyD88), MyD88-Adaptor-Like (MAL/TIRAP), TIR domain-containing adaptor protein-inducing interferon β (TRIF/TICAM-1), TRIF-related adaptor molecule (TRAM/TICAM-2), and Sterile-alpha and Armadillo motif containing protein (SARM) [177, 179–181]. Recruitment of TLR adaptors to TLR4 occurs via the Toll/interleukin-1 (IL-1) receptor (TIR). These interactions initialise the downstream signalling pathways within the cell and lead to the induction of pro-inflammatory genes.

Although, TLR pathways have been well characterised, information on the molecular dynamics of the signalling proteins is limited, impeding the development of therapeutic strategies and our understanding of the effects of polymorphic variants on human disease. Therefore, any structural information that helps us to deepen our understanding of the signalling processes will have a major impact to the field.

Of particular interest in this adaptor protein subgroup is the MyD88 protein, a key signalling adaptor for TLR1-3 and TLR4-10, as well as for the Interleukin-1 and Interleukin-18 receptors (IL-1R & IL-18R) [177]. MyD88 is recruited in different ways depending on which TLR is involved. Here we investigate the recruitment of MyD88 specifically with respect to the TLR4 pathway (Fig. 4.1). The initial triggering of TLR4 results in the recruitment of TIR/IL-1 receptors which then results in recruitment of MAL. The MAL protein then recruits MyD88 and this recruitment initiates MyD88 oligomerisation. This oligomerisation is what causes a cascade of reaction within the cell resulting in the induction of pro-inflammatory genes, leading to an immune response.

The structure of MyD88 has previously been successfully solved via solution NMR [182], and X-ray crystallography to a resolution of 1.45 Å [183], but these structures were crystallised in a self-contained solution, where none of the other partner proteins involved in TLR4 signalling were present alongside MyD88 within the crystallisation solution. It is the interaction between MAL and MyD88 which we are interested in characterising and understanding. In this chapter, using a structural biology approach, we aim to investigate and develop the best experimental methods for obtaining X-ray diffraction data, particularly with respect to MyD88, co-crystallised with MAL. Our conclusions about how MAL recruits MyD88 in the initial phase of TLR4 signalling are discussed in detail in Chapter 5.

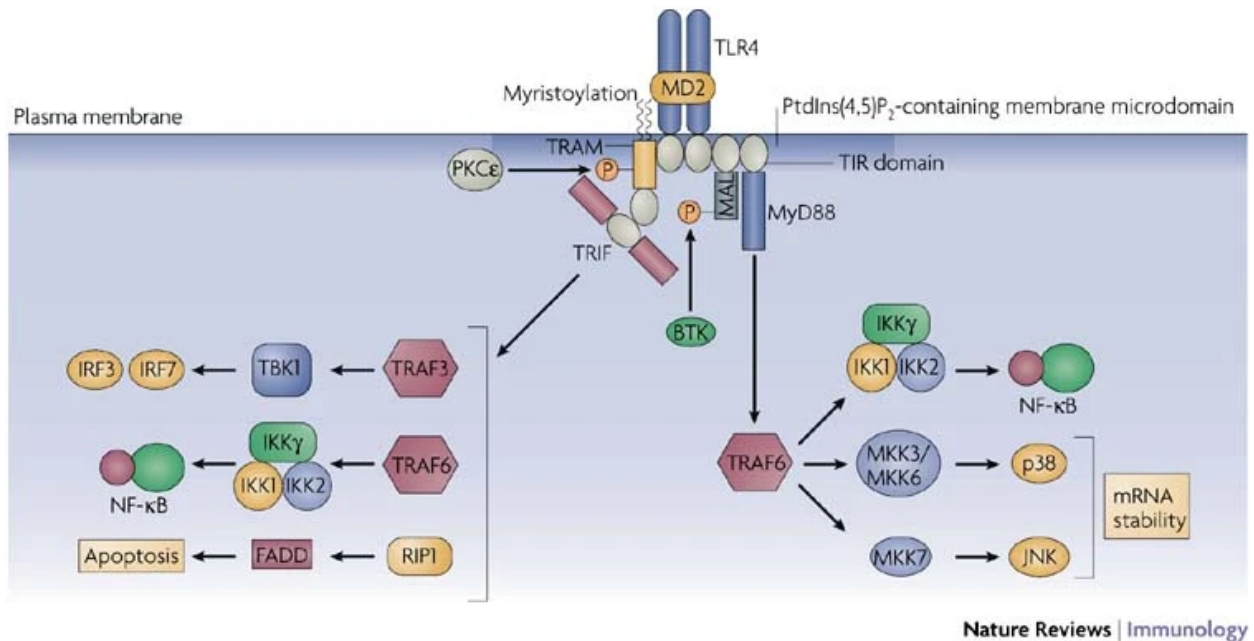


FIGURE 4.1: For the particular case of the TLR4 MyD88-dependent pathway (right pathway), the MAL adaptor protein is used to recruit MyD88 in order to activate three pathways; the NF- κ B (nuclear factor- κ B) pathway, p38 (mitogen-activated protein kinase) MAPK pathway, and the JNK (JUN N-terminal kinase) MAPK pathway. This figure is reprinted with permission from O'Neill 2007 [177]. Copyright 1969, Nature Publishing Group.

We have co-crystallised MyD88 with MAL in order to more accurately mimic a biologically relevant setting so that we can identify the process of MyD88 oligomerisation, and how MAL is involved in this process. Using SX methods we have tested and optimised sample delivery conditions in order to obtain high quality X-ray diffraction data on this crystal system.

4.2 Introduction

As discussed in Chapter 2, serial crystallography (SX) has been developing rapidly over the last decade. In the context of XFELs this technique is termed serial femtosecond crystallography (SFX), and was developed to enable crystallography experiments at XFELs. The highly intense nature of the XFEL beam, and the fast pulse regime capabilities resulted in the achievement of femtosecond data collection rates which also generally results in destruction of the crystal after a single X-ray pulse, leading to the term "diffraction before destruction".

It is not possible to perform conventional crystallography at an XFEL, as a single crystal is not capable of surviving the number of beam exposures that would be required in order to fully sample reciprocal space. Following on from the development of SFX, Serial Millisecond Crystallography (SMX) was developed, a technique that uses the same process as SFX, but performed at synchrotron facilities where millisecond exposure times are typically used to collect diffraction data.

SMX has several advantages compared to conventional crystallography at a synchrotron, because it can be carried out at room temperature and does not require growth of large crystals ($>5\text{ }\mu\text{m}$). Therefore, it can be used for crystals down to approximately $1\text{ }\mu\text{m}$ in diameter and can be used for the high-throughput delivery of crystals to the beam. Crystallisation for many proteins, particularly membrane-associated proteins, can be extremely difficult and can result in small crystals that are not large enough for conventional crystallography. Therefore both SFX and SMX provide an opportunity to collect structural data from these proteins, many of which are high impact drug targets [184]. While the beam intensity and focal size of the synchrotron source cannot compare to an XFEL source, access is significantly easier to synchrotrons compared to XFELs, and therefore utilising SMX where practicable can reduce the pressure on XFEL sources and help guide SFX experiments.

There are a number of delivery methods that have been developed for SFX and SMX. Developments in these delivery methods need to consider a variety of aspects including how to minimise sample handling as well as any other mechanical or chemical stresses involved, so that the best possible crystal quality can be maintained. Other considerations include choosing sample mounting materials and crystal delivery mediums that produce minimal unwanted diffraction, as well as how to best maintain a stable environment during all phases of the experiment including in the preparation and sample delivery phases. Development of various SX delivery methods can be divided into two main areas; moving target delivery systems, and fixed-target delivery systems.

Moving target delivery systems involve a continuously replenished sample stream that is delivered to the beam interaction region. The gas-focused dynamic virtual nozzle (GDVN), is one such moving target delivery system. The GDVN uses an outer gas stream to focus the inner sample stream into a smaller-diameter, faster-flowing stream (Fig. 2.1), and

is described in detail in Subsection 2.1.1 [56]. Another moving target delivery system is the High Viscosity Injector (HVI) (also referred to as high viscosity extrusion (HVE)), of which currently the most commonly used HVI system is the Arizona Lipidic Cubic Phase (LCP) injector system [66], which was specifically designed for crystals grown in LCP. This injector system allows for much slower sample injection, involving flow rates of 0.001–0.3 $\mu\text{L}/\text{min}$, compared to GDVN flow rates that require a minimum of 10 $\mu\text{L}/\text{min}$ and 10 m/s to produce and maintain a stable stream [57]. Fixed-target delivery systems have also been developed, and include nylon loops, conveyor belts, and microfluidic chips. A silicon version of a microfluidic chip system [185, 186] has been tested and discussed in this chapter.

The aim of this chapter is to investigate and compare three different delivery methods used for the collection of X-ray diffraction data from the MyD88 needle-like crystals that have formed when co-crystallised with MAL [187]. Due to their limited size they are not suitable for conventional crystallography, and therefore three different SX delivery systems, GDVN, HVI with LCP, and microfluidic silicon chips have been tested. In this chapter the advantages and disadvantages of these different delivery systems are discussed with a focus on how they perform in the context of the MyD88 crystals.

4.3 Materials and Methods

All materials were purchased from Sigma-Aldrich unless otherwise specified. The initial microfluidic chips were purchased from Suna-Precision (GmbH, Germany) and then custom made pore sizes were fabricated at the Melbourne Centre for Nanofabrication (MCN, Melbourne).

4.3.1 MyD88/MAL Crystallisation

MyD88 and MAL proteins were provided by Prof. Bostjan Kobe (University of Queensland) and Dr. Thomas Ve (Griffith University). The two proteins were co-crystallised using the batch method as outlined in Darmanin et. al [187]. MAL induced MyD88^{TIR} crystals were produced by incubating MAL^{TIR} (0.5–3 μM) with MyD88^{TIR} (60–100 μM) in 10 mM

HEPES pH 7.5, 150 mM NaCl at 30°C for 60 - 120 minutes. The crystals were needle-like in appearance, with dimensions of 0.5 - 10 μm (in the longest length) x 0.2 x 0.05 μm (Fig. 4.2).

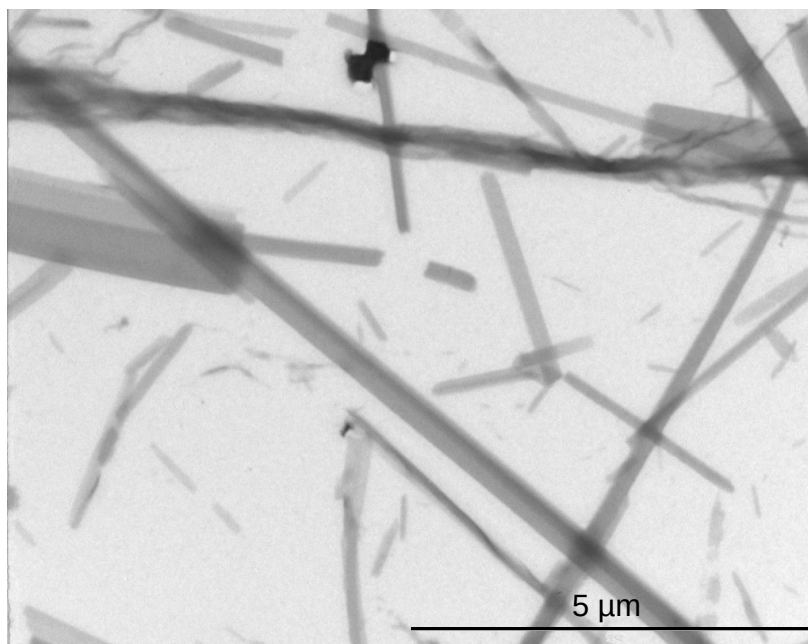


FIGURE 4.2: Negative stained TEM image of MyD88^{TIR} needle-like crystals. The crystals were stained with 1% uranyl acetate and viewed with a JEOL JEM-2010 TEM. Image supplied by Dr. Connie Darmanin.

4.3.2 Microfluidic Silicon Chip

Sample Delivery

Based on designs by Roedig et.al [185, 186], 2.5 μL of MyD88 crystals in solution was deposited on a silicon chip with pores of various diameters (1, 2, 5 & 10 μm) in a humidified environment (Fig. 4.3). Various MyD88-to-MAL ratios were tested for crystallisation optimisation, including 16:1 and 25:1 ratios. Crystal-to-buffer ratios tested for sample deposition onto the chip included 1:1 and 3:1 ratios, as well as a neat solution (where the crystal solution was centrifuged, the mother liquor removed and exchanged with the equivalent volume of buffer). The buffer solutions that were tested included 16% and 40% glycerol concentrations, and was included as a cryo-protectant.

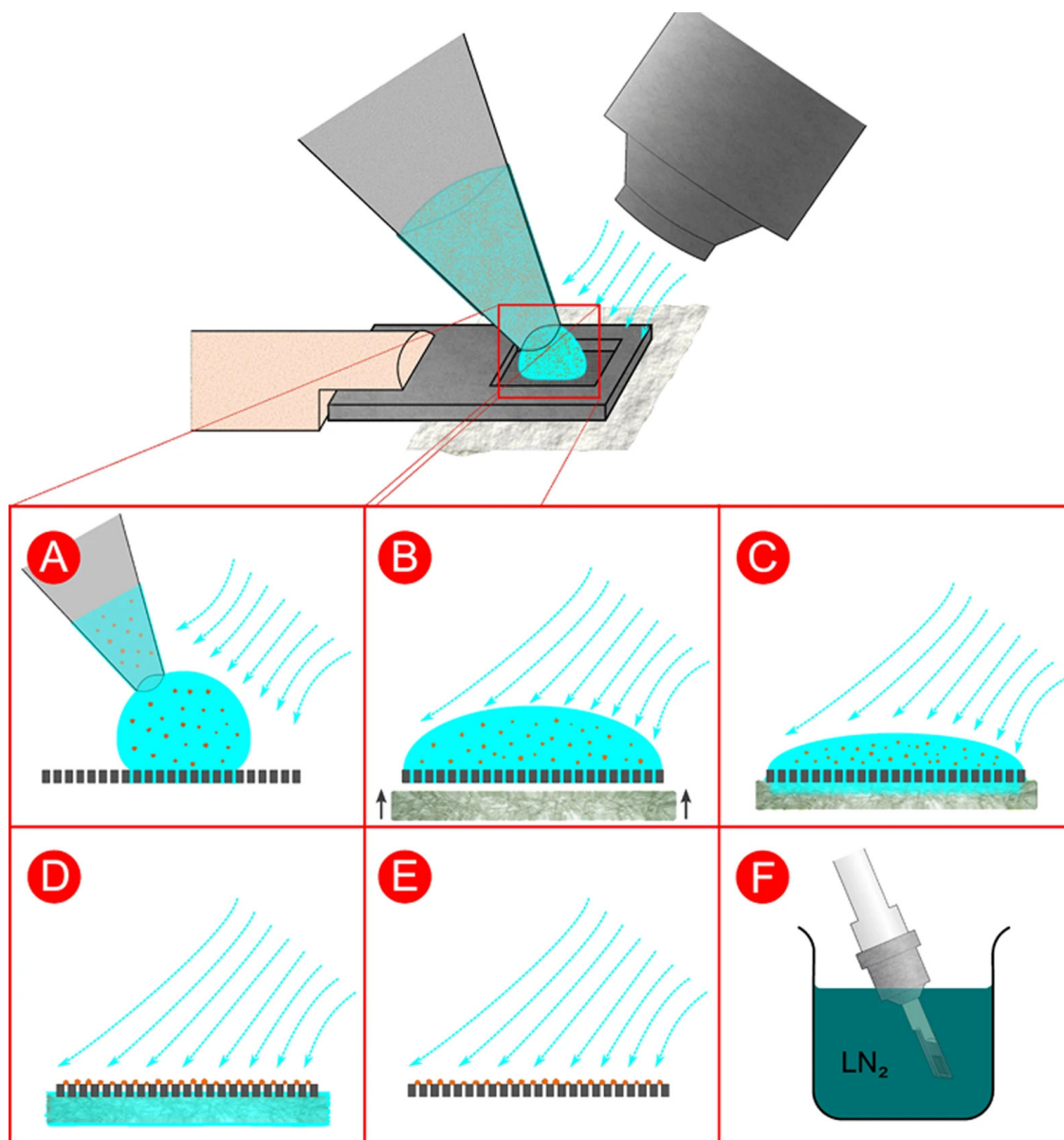


FIGURE 4.3: Sample loading procedure: A droplet of crystal suspension is pipetted onto the front side of the chip (A). The mother liquor then fills the pores, forming a meniscus on the lower side due to capillary action (B). The mother liquor is removed by touching the underside of the chip with filter paper, thereby drawing the excess liquid through the pores and soaking up the excess solution (C). Crystals that are larger than the pores are retained on the upper side of the chip, and arrange themselves in a periodic way according to the pore structure (D). The wet filter paper is then removed (E). The chip is then flash-frozen in liquid nitrogen (F). During A-E, the chip is under a continuous stream humidified air in order to control evaporation. This figure is reprinted from Roedig 2015 [185]. CC BY 4.0.

A 2.5 μL aliquot of the crystal solution was deposited onto the silicon chip, and then the excess liquid was removed by drawing the solution through the pores via the underside of

the chip. This allowed the crystals to settle on the chip and, if done correctly in a uniform fashion. The chip was then immediately flash frozen in liquid nitrogen and mounted onto a standard crystallography goniometer system at the P11 beamline at Petra III synchrotron, Hamburg, Germany [27]. Figure 4.4 shows details of the microfluidic silicon chip experiment at the P11 beamline, where Figure 4.4a shows the setup, and Figure 4.4b shows a close-up of the pores within the chip, with the darker regions indicating the presence of crystals.

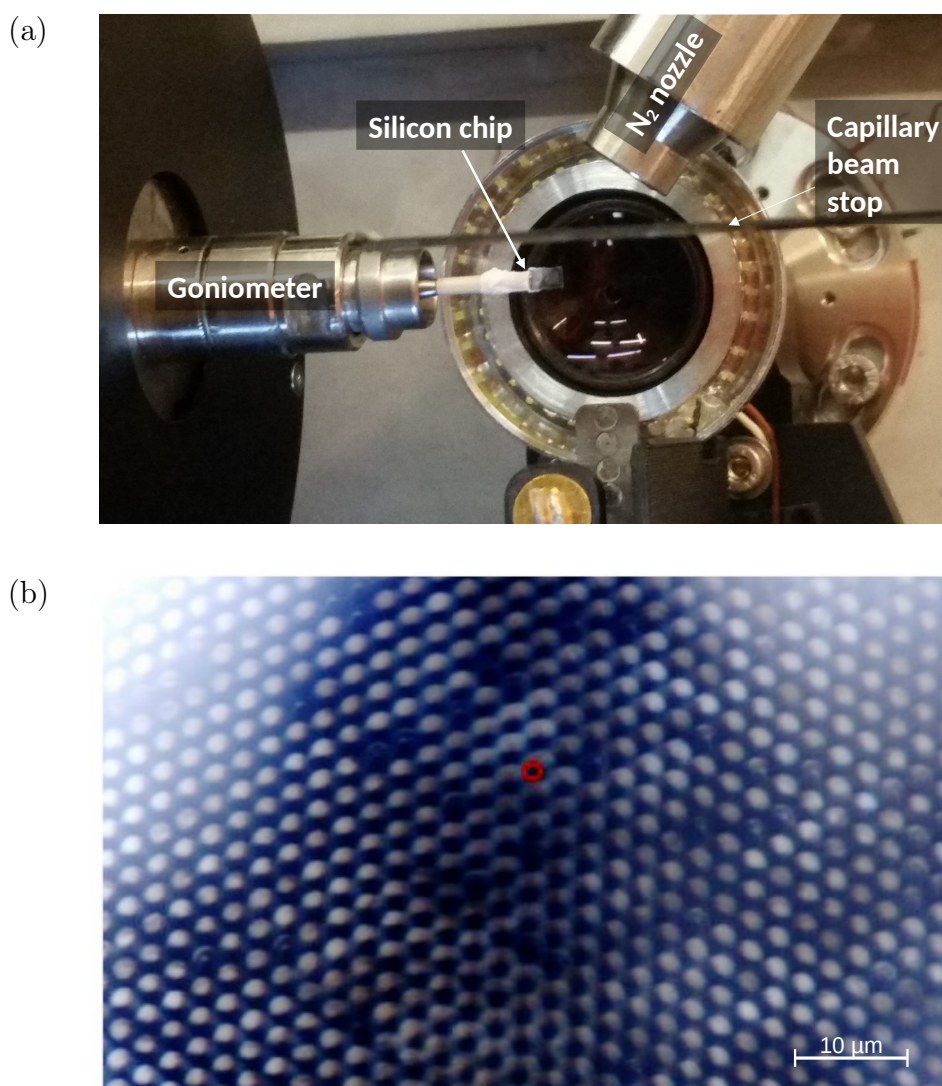


FIGURE 4.4: Static data collection setup of MyD88 crystals at Petra III, P11 beamline. (a) The goniometer setup with a silicon chip mounted under a cryogenic nozzle, and with the capillary beam stop in-place, (b) a close up view of a region on the chip showing the pores with crystals on the surface as indicated by the dark regions. The data was collected using a grid scan mode under liquid nitrogen conditions.

Data Collection & Analysis

SMX data was collected at the Petra III synchrotron using the P11 beamline during two experimental beamtimes [27]. Both beamtimes used an identical experimental setup, that included a 12 keV beam, a beam size of 2x2 μm and flux levels of 1.7×10^{11} ph/s & 6.4×10^{11} ph/s. A 100 μm pinhole, and a detector distance of 587.8 mm. A 100 μm capillary beam-stop made by the beamline scientists was used for this experiment to help reduce the background scatter on the detector. Data was collected via "grid scan" mode, where the goniometer was moved left-to-right and top-to-bottom, with a step size of 10 μm , an oscillation of 0.1° at each step, obtaining two frames/step, with a total exposure time/step of 2000 ms. The detector used for these experiments was the Pilatus 6M-F (25 Hz) detector.

The Petra III synchrotron data was collected in CBF format. OffDA, an offline version of OnDA [188] was used for hit finding. From OffDA, HDF5 files were output with the peak information, and indexing and integration was then run with CrystFEL version 0.6.2 using the Indexamajig program [176, 189, 190].

4.3.3 LCP Injector

Sample Delivery

A 25:1 ratio of MyD88:MAL protein was used for crystallisation. The crystals were filtered using a syringe connected to a 10 μm frit filter (BioRad). The initial crystal concentration was 1×10^8 crystals/mL. Monoolein (MO) and MyD88 crystals were mixed at a 60:40 ratio of lipid:crystal to maintain the lipidic cubic phase (LCP). The samples were mixed using the standard LCP method with 2 Hamilton syringes [191, 192]. Briefly, molten MO was added to syringe one and MyD88 was added to syringe two. The volume of the crystals added was calculated so that the final percentage of crystal solution mixed with the MO was 40%. The two syringes were coupled together and mixed within the syringe system to allow for the cubic phase to form. To prevent the formation of lamellar phase that can occur due to the temperature drop (sample $< 18^\circ$) of the sample as it is injected into a vacuum chamber [66], 2 μL of 7.9 MAG was added to the sample. This addition was performed via syringe mixing, where 7.9 MAG in a syringe was coupled with the sample syringe, and they were gently

mixed together. The sample was then loaded into a 40 μL sample reservoir and attached to the LCP injector supplied by the LCLS. Glass capillary nozzle tips which had a diameter of 50 μm at the interaction region were kindly supplied by Dan Du Pointe (LCLS, US).

Data Collection & Analysis

Room temperature XFEL data was collected at Linac Coherent Light Source (LCLS), USA, at the Coherent X-ray Imaging (CXI) end-station using a CSPAD detector [80, 193]. An optimal flow rate of 3 $\mu\text{L}/\text{min}$ was used, with a gas pressure of 160 psi, and a beam diameter of 3.5 μm at 9.5 keV with 10% transmission. The beamtime was running in parasitic mode, which meant we did not have the opportunity to change the beam parameters and were reliant on the experimental parameters chosen by the main users. The repetition rate was 120 Hz with a pulse duration of 36 fs. Detector images were collected at a detector distance of 168 mm, relative to the sample.

The SLAC national accelerator laboratory servers were used for analysis of the LCP data. A Psocake GUI, provided on the servers was used for hit finding, indexing and integration [194, 195].

4.3.4 GDVN Injector

Sample Delivery

A 25:1 ratio of MyD88:MAL protein was used for crystallisation. Several different concentrations of crystal:buffer were tested, which included a 2x concentration, 2x dilution, and 4x dilution in order to determine the optimal crystal concentration that resulted in the highest hit rate. The starting crystal suspension had 1.5×10^9 crystals/mL before it was diluted or concentrated. A more concentrated sample was prepared by centrifuging the sample to pellet the crystals, the buffer solution was removed and then replaced with half the volume of buffer, resulting in a 2-fold increase in the crystal concentration. Dilutions of the sample was prepared by taking an aliquot of crystals from the neat stock and mixing it with a volume of the buffer (10 mM HEPES pH 7.5, 150 mM NaCl) to reach the appropriate dilution.

Prior to placing the sample into the GDVN sample holder, the sample was manually

filtered using a syringe connected to a 20 μm frit filter (BioRad) in order to ensure that only crystals below 20 μm in size were present. This was done as a preventative measure, in order to minimise crystals clogging the injector. The filtered sample was then loaded into a 1.2 mL sample holder ready for delivery using standard GDVN protocols. A suspension of MyD88 crystals were delivered to the interaction region using a GDVN. Glass capillary nozzle tips with a diameter of 50 μm were kindly supplied by Prof. John Spence's group (Arizona State University, US). A 20 μm stainless steel in-line filter was used. Every 15 minutes during data collection the samples were vortexed to prevent crystals from settling, after which, the injector lines were flushed with water.

Data Collection & Analysis

The XFEL GDVN data was collected at room temperature at the CXI end-station at the LCLS, using the CSPAD detector [80, 193]. The GDVN optimal settings included a flow rate of 20 $\mu\text{L}/\text{min}$ and a gas pressure of 160 psi. The experimental setup included a beam diameter of 1 μm at 9.5 keV with 100% transmission. The LCLS repetition rate was 120 Hz with a pulse duration of 45 fs. Diffraction was collected at three detector distances relative to sample, 286 mm, 186 mm, and 111 mm.

Hit finding was initially performed using the Psocake GUI, but later analysis was performed utilising the Peakfinder8 algorithm in Cheetah [84], both provided on the SLAC national accelerator laboratory servers. Indexamajig, in the CrystFEL software suite was then used for indexing and integration. Post refinement processing utilised the Partialator program also available in the CrystFEL software suite [176, 189, 190].

4.4 Results

4.4.1 Microfluidic Silicon Chip

Sample Preparation

The crystal solution was placed on the silicon chip and the excess liquid was drawn off under a humidified environment, as mentioned previously (see Subsection 4.3.2). While this step

was performed under humidified conditions in order to eliminate evaporation of the crystal solution from the chip, maintaining a stable humidity even for the small amount of time needed to prepare the chip proved extremely difficult. Therefore flash freezing the chip in liquid nitrogen was performed as quickly as possible after the excess liquid was drawn off. Freezing of the samples required optimisation of an appropriate freezing solution. A neat crystal solution, with 40% glycerol in buffer was chosen as the optimal concentration.

The microfluidic chip, glued to a standard brass pin was mounted onto the goniometer via strong magnets. These magnets caused a sudden pull on the chip/brass pin, snapping it into place and occasionally breaking the chip off the pin with the force of the snap (see Fig. 4.4a). After the microfluidic silicon chip was mounted onto the goniometer setup, the in-line microscope was used to align a section of the chip for scanning. The distribution of crystals on the chip was uneven, as can be seen in Figure 4.5, where darker regions indicate regions where the sample is too densely packed and single crystals are not able to be distinguished, and lighter regions where single, needle-like crystals are able to be distinguished.

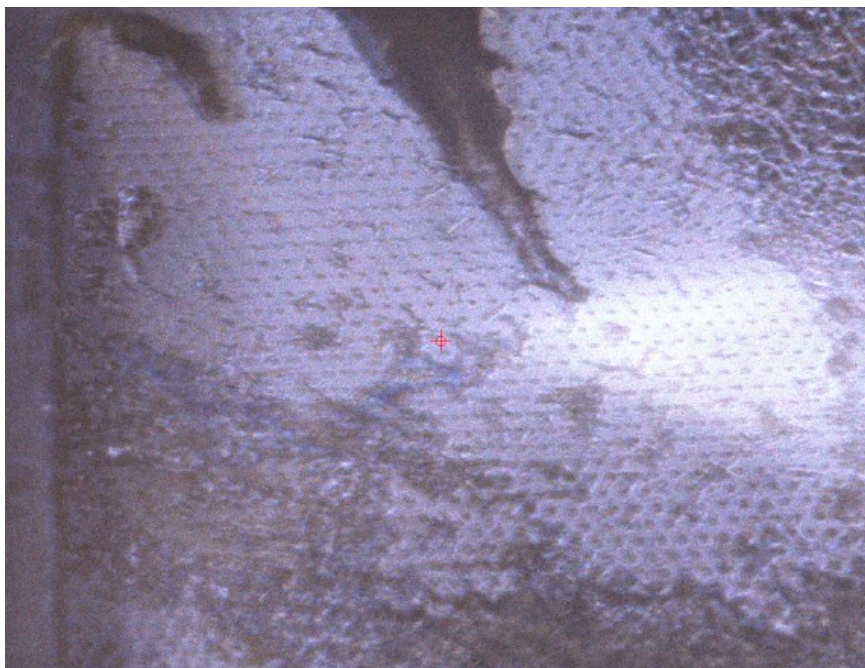


FIGURE 4.5: An image captured from the in-line microscope at the Petra III P11 beamline showing the distribution of crystals on the microfluidic silicon chip. Darker areas indicate regions where the sample is densely packed and single crystals cannot be distinguished, while lighter areas indicate regions of no or little sample, and where single, needle-like crystals can be distinguished.

Data Collection & Analysis

The first step of data analysis was to assess and identify the peaks in the detector images. OffDA hit finding parameters were optimised using the following parameters. In order for an image to be counted as a hit, 5 peaks needed to be identified in a single image, otherwise the image was classified as a miss and was not analysed further. The signal-to-noise ratio (SNR) was optimised to a value of 2.25, and the analog-to-digital converter count (ADC) threshold was optimised to a value of 11, where the ADC threshold is the minimum threshold value for the entire image. This number was lower than expected due to the combination of very low signal due to the removal of the excess buffer solution, and the use of a glass capillary beam-stop that was designed by P11 beamline scientists to significantly reduced background noise. The parameters surrounding what was considered to be peak size were also set. To correctly distinguish whether the ADC counts picked up by the detector are from noise or from a peak, a peak needed to consist of a minimum of 5 pixels, a maximum of 200 pixels, with a radius of 4 pixels for the local background determination. To mitigate the identification of false peak, a detector mask that included the solvent ring region, the beam-stop region, and the detector edge pixels was applied to the detector images.

Dirax and MOSFLM indexing algorithms, supplied through the CrystFEL software suite were used to index the data. Both indexing algorithms were run with no prior cell information as input ("-nolatt-nocell"). Both indexing algorithms had the "retry" parameter enabled, meaning that if initial indexing failed, a second attempt would be made after the weakest 10% of peaks were removed. Both indexing algorithms had the "multi" parameter disabled. If this setting is enabled, once a indexing solution has been found, those peaks that were part of the solution are removed before a second round of indexing is performed in order to identify instances of multiple crystals within a single image. For the majority of the diffraction patterns, the number of peaks that had been identified per pattern, were already at the lower range required for successful indexing, and therefore it was extremely unlikely that enough peaks would be remain after the first round of indexing for any additional lattice to be identified. The "refine" parameter was enabled for both indexing algorithms. This enabled prediction refinement steps that included the use of crystal parameter refinement,

detector geometry refinement, and outlier rejection [189]. The integration radii was optimised and set to 2, 5, and 7 pixels.

From the 37,746 diffraction patterns analysed, 10,754 patterns were found to have hits (28.48%). From these 10,754 hits, only 2,025 were indexed (18.83%). The statistics are shown in Table 4.2. The Petra III data showed weak crystal diffraction, with strong crystal diffraction seen to a resolution of approximately 8 Å, and some weaker diffraction observed at resolutions beyond 4 Å based on the diffraction rings present from a summation of the images that had been identified as hits (Fig. 4.6).

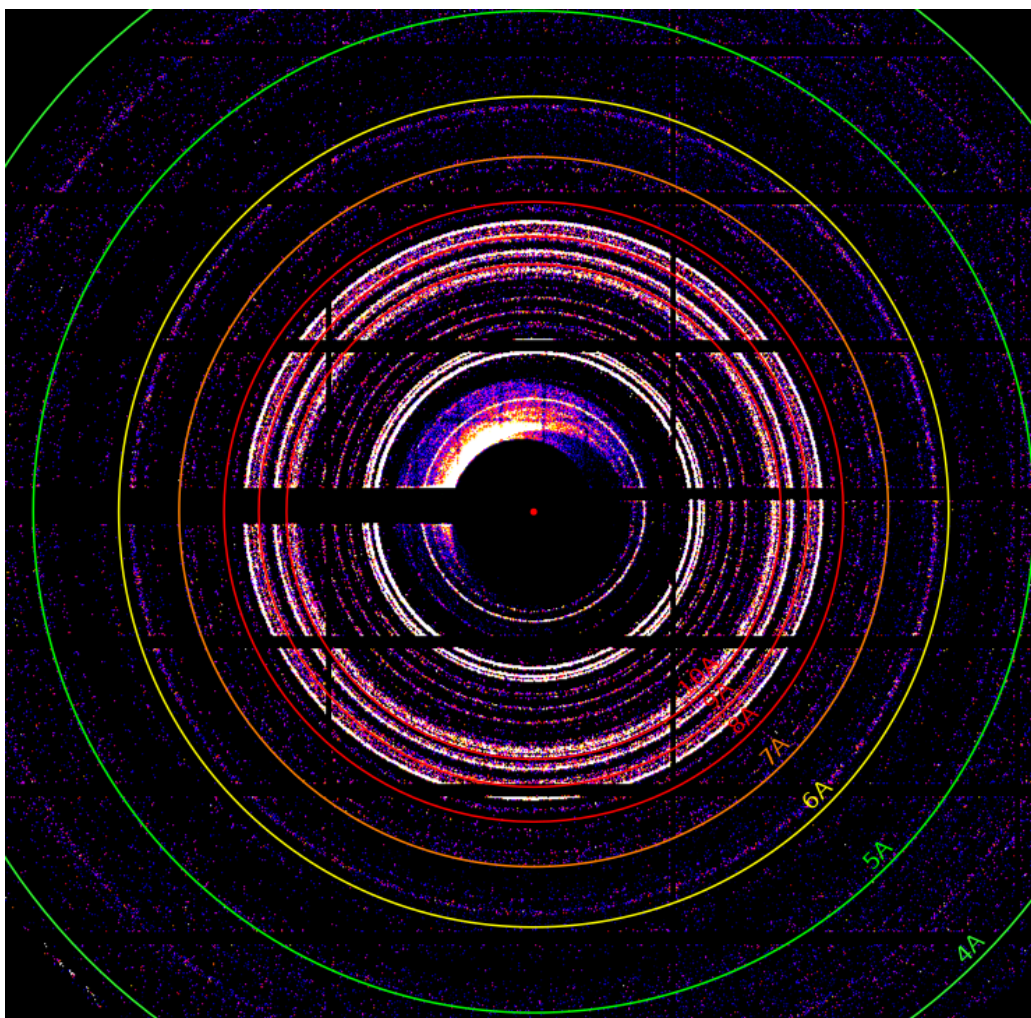


FIGURE 4.6: 37,746 detector images obtained using a microfluidic silicon chip system at Petra III were analysed for peaks. After hit finding, all the peaks from detector images that contained more than 5 peaks (10,754 images) were merged to generate this powder diffraction image. The image shows enough data was collected to represent the complete rotation of the crystal up to 8 Å, with weak diffraction extending beyond 4 Å.

The indexing profiles obtained from the chip data (Fig. 4.7) show large distributions in the lattice parameters, indicating large errors associated with the predicted unit cell.

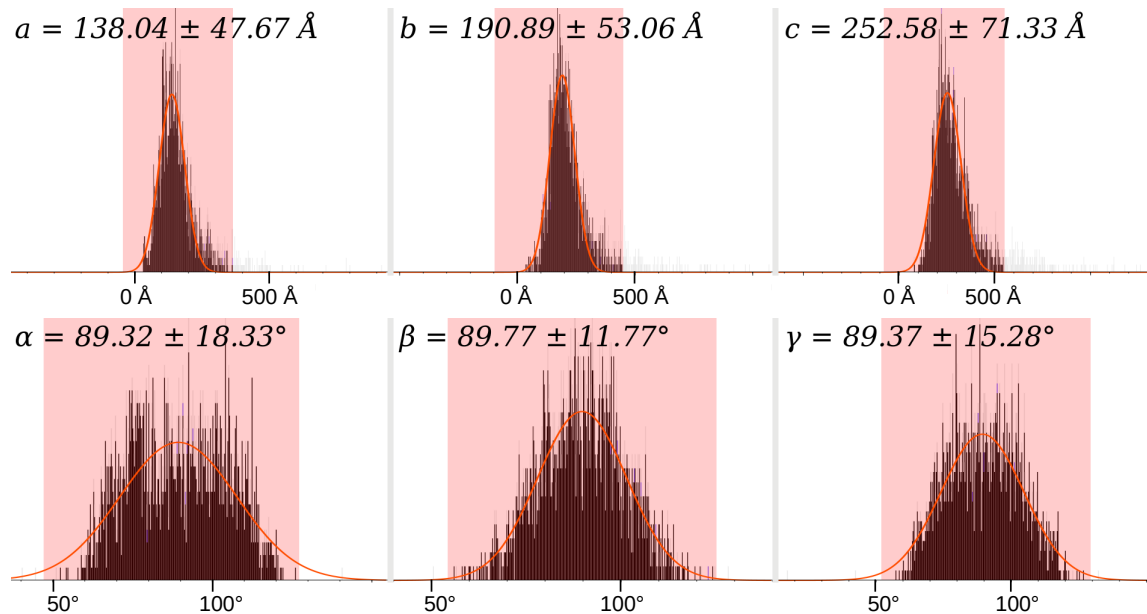


FIGURE 4.7: CrystFEL indexing results showing the unit cell distributions found from the Petra III microfluidic silicon chip diffraction data. The unit cell parameters identified by indexing deviated greatly, and showed broad, undefined unit cell distributions.

An inspection of the diffraction patterns after hit finding, revealed that a large proportion of the patterns, consisted of a very sparse number of Bragg peaks, with 54% of the patterns classified as a hit, containing <20 peaks (Fig. 4.8a). Visual observations also highlighted the fact that many of the diffraction patterns showed peaks that were located very close to one another (Fig. 4.8b).

With the indexing results providing broad and undefined unit cell distributions and having observed the issue of crystal clumping on the chip surface, it was hypothesised that diffraction from a single crystal was too weak to obtain enough Bragg peaks for indexing. It was therefore only due to diffraction from multiple crystals in a single detector image that a strong enough signal was produced, allowing for enough Bragg peaks to be detected that indexing could provide a solution, and therefore, the broad and undefined unit cell distributions obtained from indexing were predominantly due to multi-crystal diffraction patterns that had been incorrectly indexed.

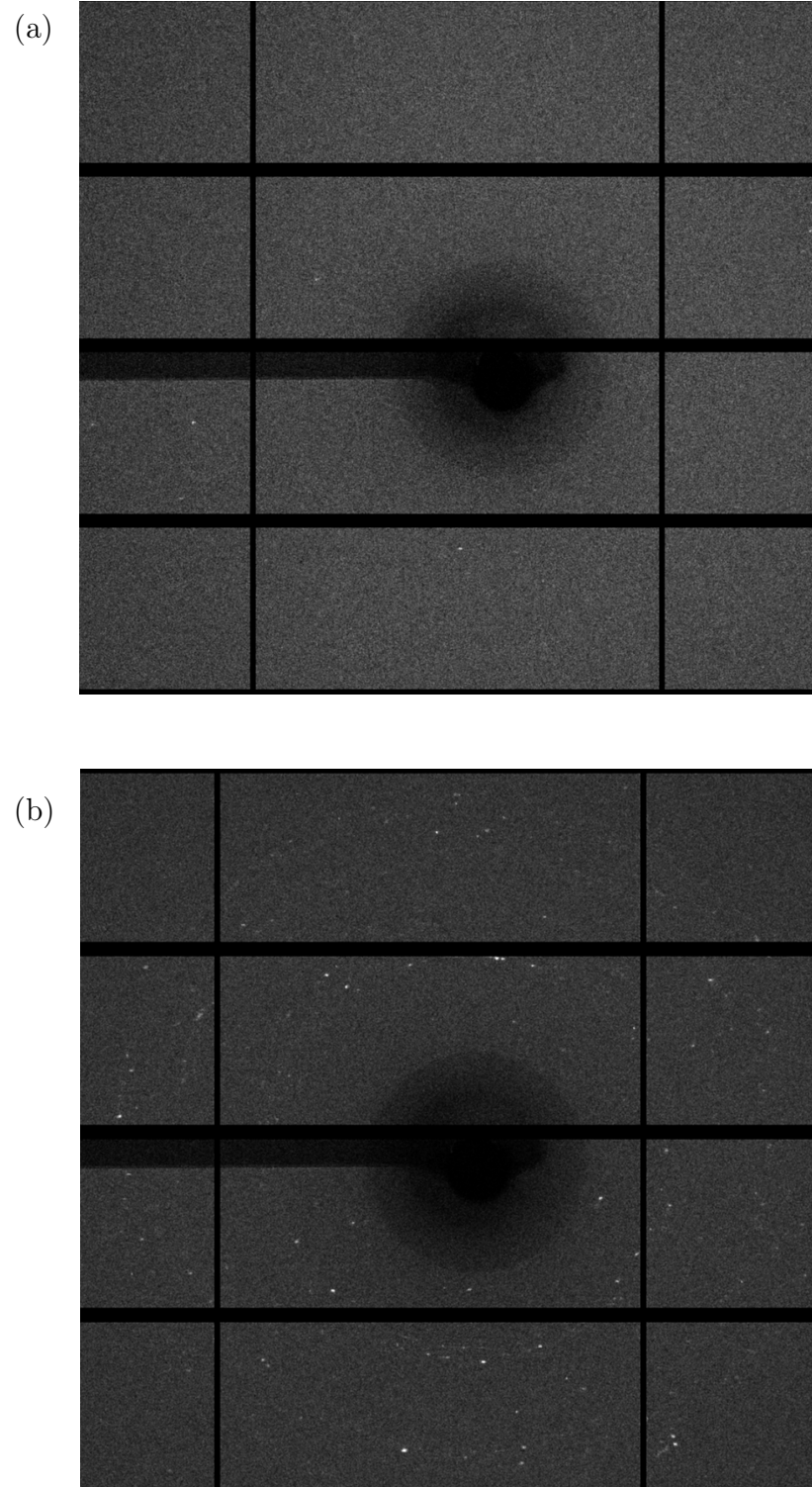


FIGURE 4.8: Detector images showing crystal diffraction observed at the P11 beamline at Petra III using the microfluidic silicon chip. (a) Extremely sparse Bragg peaks were observed in a large proportion of the detector images, while the images that had more peaks (b) showed Bragg peaks located close together, indicative of multi-crystal diffraction.

To test this hypothesis, a simulation involving the generation of single and multi-crystal diffraction patterns based on previously solved MyD88 structure (PDB code 4EO7) was performed, and from this simulation single-crystal partial diffraction patterns were generated. These partial diffraction patterns were then combined, so that two partial diffraction patterns were merged into a single pattern, simulating a multi-crystal diffraction pattern. The diffraction patterns simulated for both single-crystal diffraction, and multi-crystal diffraction, were then analysed in CrystFEL. Both data sets were run through the same hit finding and indexing pipeline as per the analysis of the experimental Petra III chip data.

Indexing profiles were obtained for both the single-crystal simulated diffraction (Fig. 4.9), and the multi-crystal simulated diffraction (Fig. 4.10). The indexing results obtained from the single-crystal simulated diffraction show narrow distributions of the unit cell parameters, with low error values in the range typically seen from a correctly identified unit cell. The indexing results from multi-crystal simulated diffraction resulted in a much broader unit cell distribution, similar to what was seen with the experimental Petra III chip data and showed similar error estimates (Fig. 4.7).

Therefore it is most likely that the Petra III diffraction patterns that contained enough Bragg peaks for the indexing algorithms to provide a solution, were predominantly multi-crystal diffraction patterns.

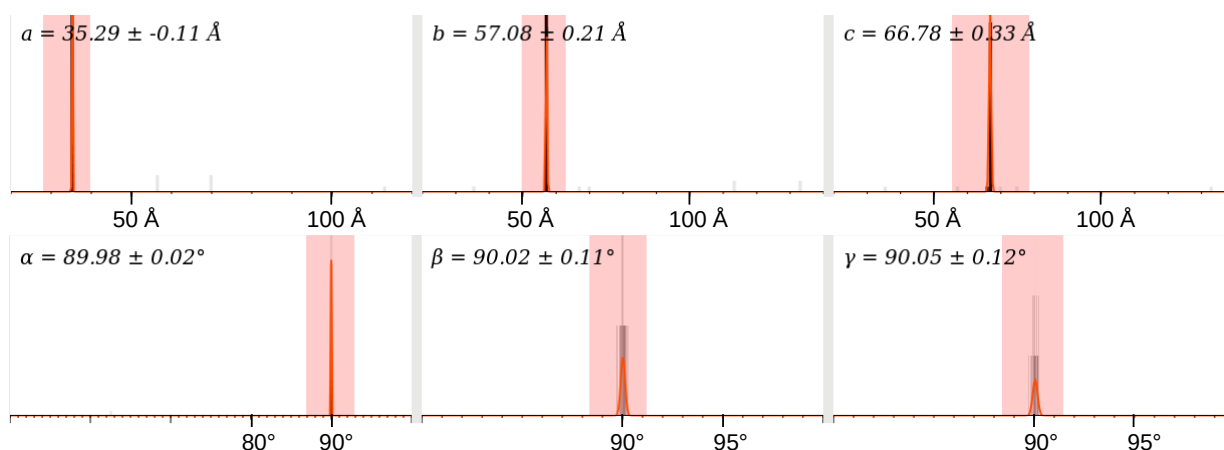


FIGURE 4.9: CrystFEL indexing solution obtained from the simulated diffraction data generated from the published MyD88 structure (PDB code 4EO7) showing the unit cell distribution from the simulated single-crystal diffraction.

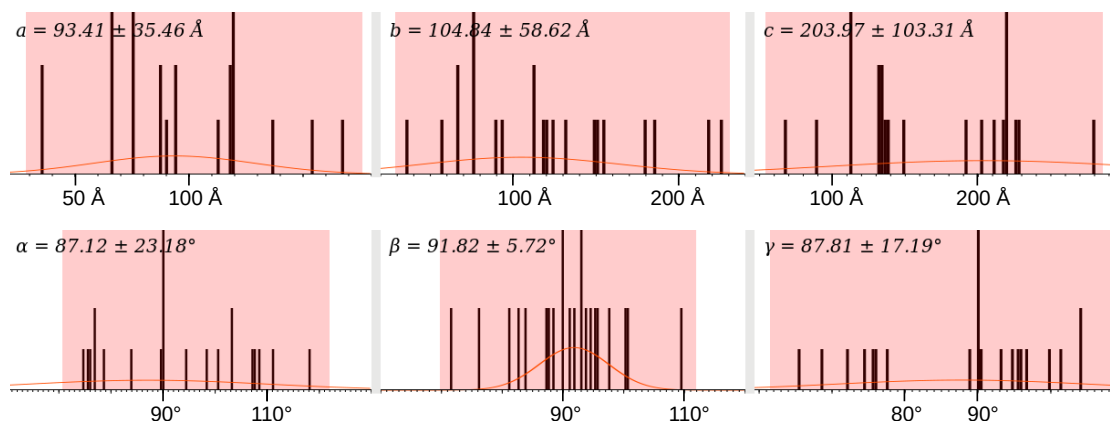


FIGURE 4.10: CrystFEL indexing solution obtained from the simulated diffraction data generated from the published MyD88 structure (PDB code 4EO7) showing the unit cell distribution from the simulated multi-crystal diffraction.

Knowing this, a sub-set of the experimental data was indexed with the unit cell from the known structure (PDB code 4EO7), in order to check if a correct indexing solution could be obtained. A successful indexing solution with this sub-set was not identified. Further in the future, after having successfully identified the correct unit cell from analysing the data obtained at the LCLS using the GDVN delivery system, indexing of the chip data was again attempted with this correct unit cell. A successful indexing solution again proved elusive, which suggested that there were simply too few Bragg peaks per crystal lattice to obtain an indexing solution. Data collection and analysis statistics can be found in Table 4.1.

4.4.2 LCP Injector

Sample Preparation

The LCP injector was tested at the LCLS during a protein crystal screening (PCS) parasitic beamtime. The MyD88 crystals were mixed with LCP and then loaded for use with the HVI delivery system, producing a stable HV stream (Fig. 4.11) suitable for data collection.

Data Collection & Analysis



FIGURE 4.11: An image showing the HVI nozzle (top cross) with a stream of MyD88 crystals in LCP. The lower cross on the image indicates the location at which the sample and X-ray beam interact. The ripples of the sample (seen below the lower cross) demonstrate the effect of the X-ray beam on the sample after exposure.

Analysis of the diffraction data obtained using the HVI delivery system, was performed using the Psocake GUI on the SLAC national accelerator servers. Bragg diffraction from the MyD88 crystals was seen, along with region of strong signal caused by the LCP (inner circle), and solvent signal (outer ring) (Fig. 4.12a).

Two mask versions were tested. The first mask included bad pixels, detector edges, an inner circle (LCP mask), and outer ring (solvent mask)(Fig. 4.12b). After inspecting the detector images after hit finding, images were observed that had multiple distinct solvent rings, with peaks present between these solvent rings that could be identified during hit finding when the solvent region was excluded from masking (Fig. 4.13a). In an attempt to minimise the solvent region excluded from analysis due to the initial masking, two thinner annulus shaped masks were tested (Fig. 4.13b). Further testing showed negligible differences in hit finding and successful indexing when the solvent region was left completely unmasked, so therefore final hit finding included the inner region, the bad pixels, and the detector edges.

Final hit finding parameters in Psocake used an adaptive algorithm, and included a minimum of 2 pixels/peak, a maximum of 30 pixels/peak, a minimum pixel intensity of 150, a minimum summed threshold (intensity of all pixels within a peak summed) of 300, an SNR of 7, a rank of 3, a radius of 5, and a dr of 2. From 846,827 diffraction patterns, 514,167 contained more than 10 peaks, and were classified as hits (60.7% hit rate). All Bragg peaks identified via the hit finding stage were summed, to generate a 2D powder pattern (Fig. 4.14). From this powder pattern, diffraction was observed at up to 3.5 Å resolution.

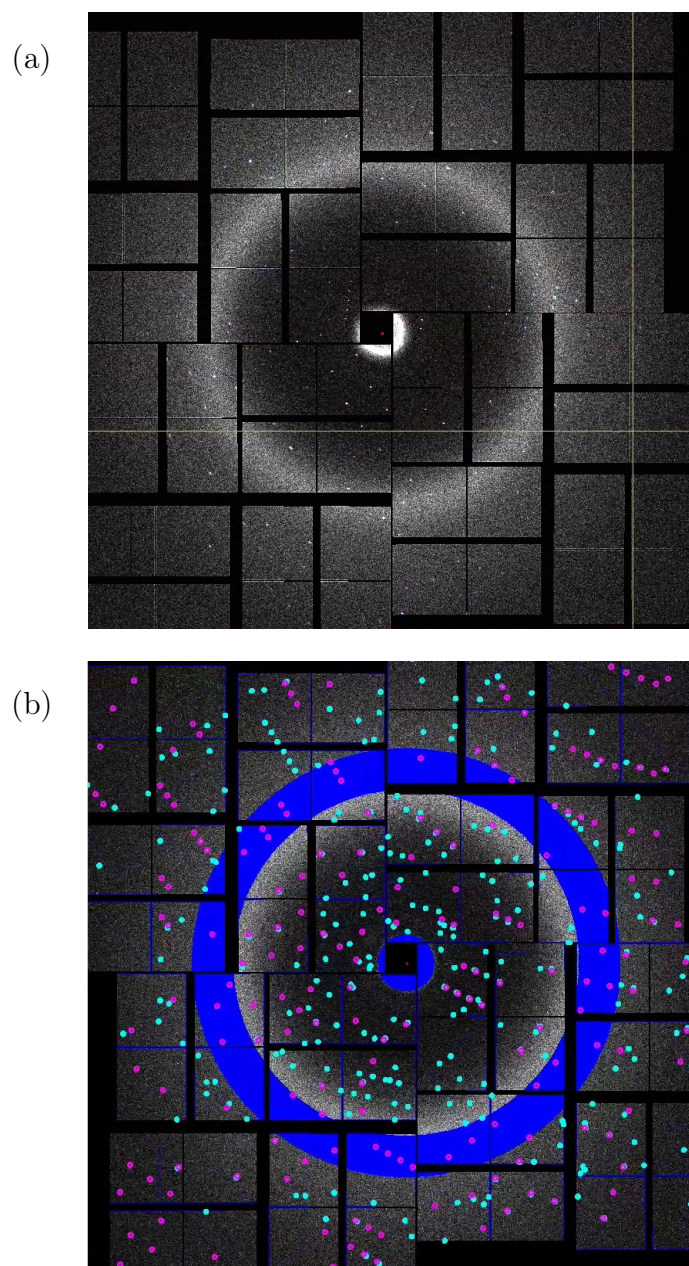


FIGURE 4.12: Detector images collected from MyD88 crystals in LCP, on the CXI beamline at the LCLS. (a) An example of a detector image obtained. The peaks can be clearly identified, but there are also two regions where crystal diffraction is hard to distinguish from background signal; a centre region of high intensity due to the LCP, and a less intense ring region, caused by the solvent. (b) The same image as shown in (a) after a detector mask has been applied (blue) to exclude the LCP, solvent, bad pixels and detector edges, and after hit finding and indexing. Locations of peaks identified from hit finding correspond to the cyan rings and predicted peak locations from an indexing solution correspond to the pink rings.

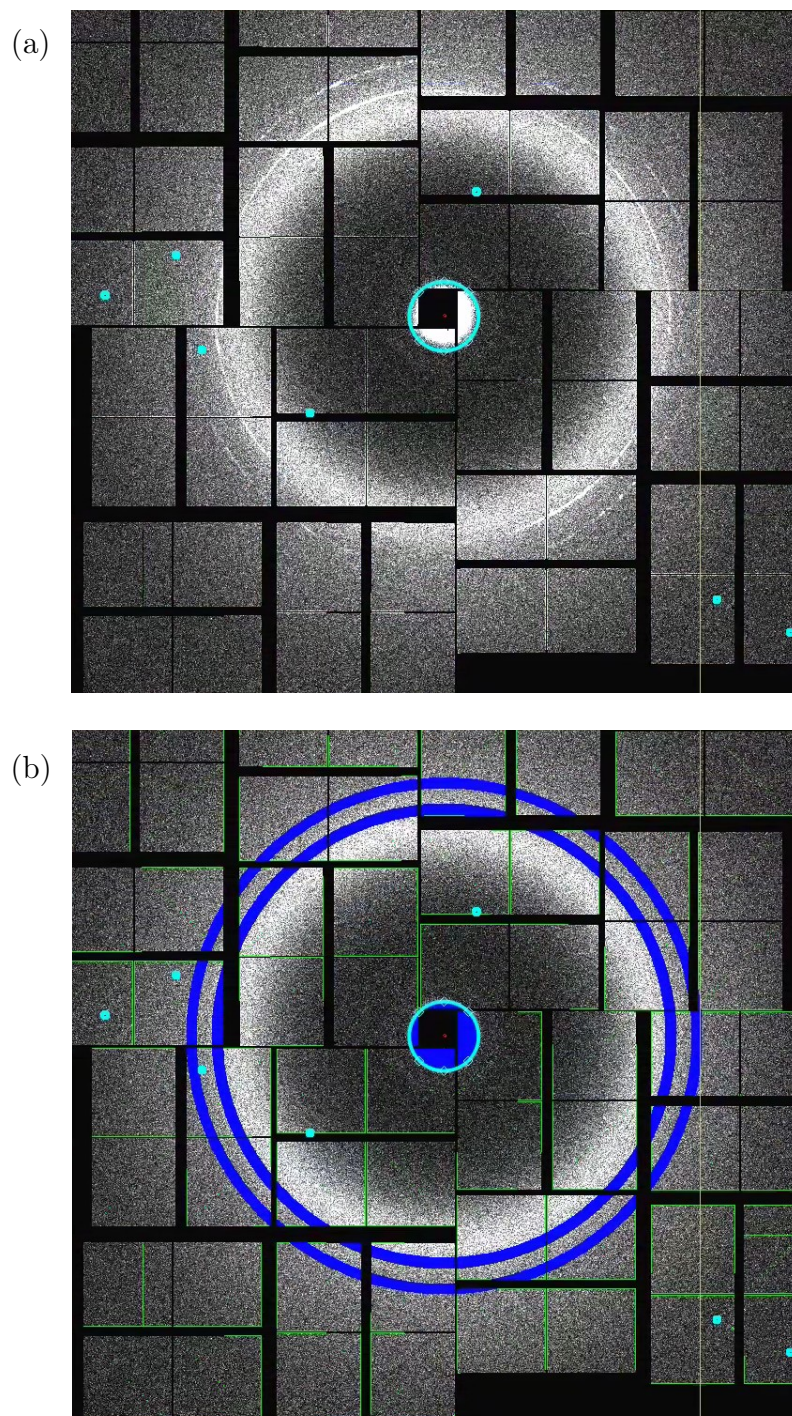


FIGURE 4.13: Another example of a detector image collected from MyD88 crystals in LCP, on the CXI beamline at the LCLS, with peaks identified from hit finding circled by cyan rings. (a) A detector image after hit finding, where multiple distinct solvent rings can be seen in the image. In an attempt to minimise the region excluded from analysis due to masking, two thinner annulus shaped masks were tested (b).

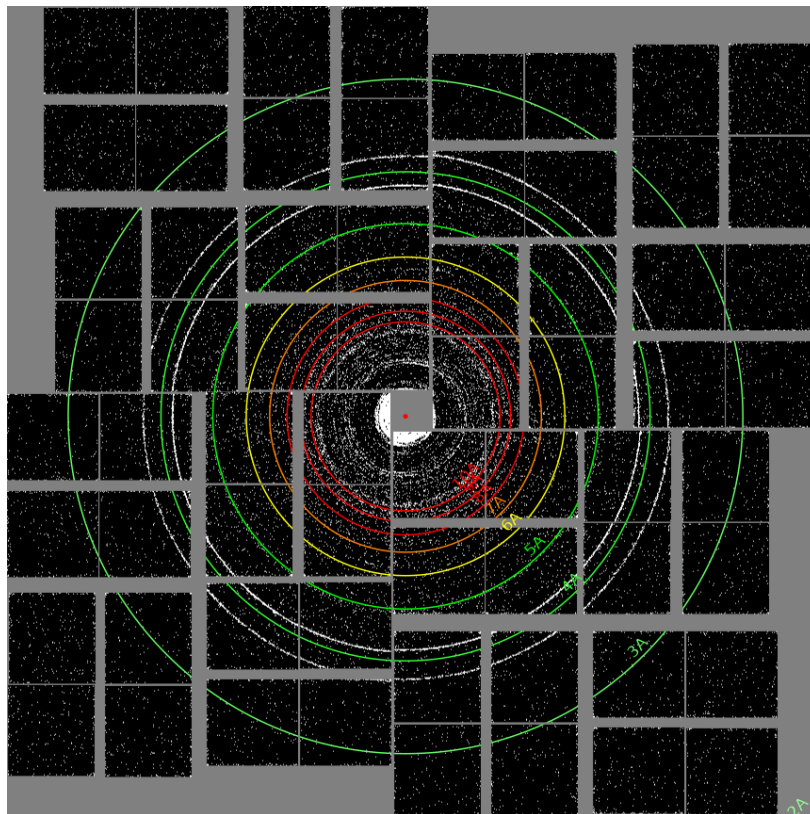


FIGURE 4.14: A 2D powder pattern generated using the data collected from MyD88 crystals in LCP at the CXI beamline at the LCLS. The powder pattern was generated by summing all the peaks identified during hit finding together into one image in order to visually estimate the resolution of our data.

Following hit finding, indexing was performed using the DirAx, MOSFLM and XGANDALF algorithms, using the Psocake GUI provided on the SLAC servers. Parameters included integration radii of 4, 5, 7, with a tolerance range of 5\AA for unit cell lengths, and 1.5° for angles. Indexing was run with no prior cell information, and with retry, and multi-indexing parameters enabled. From the 514,167 images counted as hits, 155,781 unit cells were obtained (30.3% indexing rate). Figure 4.15 shows the indexing results for the diffraction data obtained from MyD88 crystals in LCP at the CXI beamline at the LCLS. The indexing results show broad, undefined unit cell distributions, similar to the indexing results observed from the Petra III chip data, suggesting issues with crystal clumping leading to multi-crystal diffraction. Data collection and analysis statistics can be found in Table 4.1.

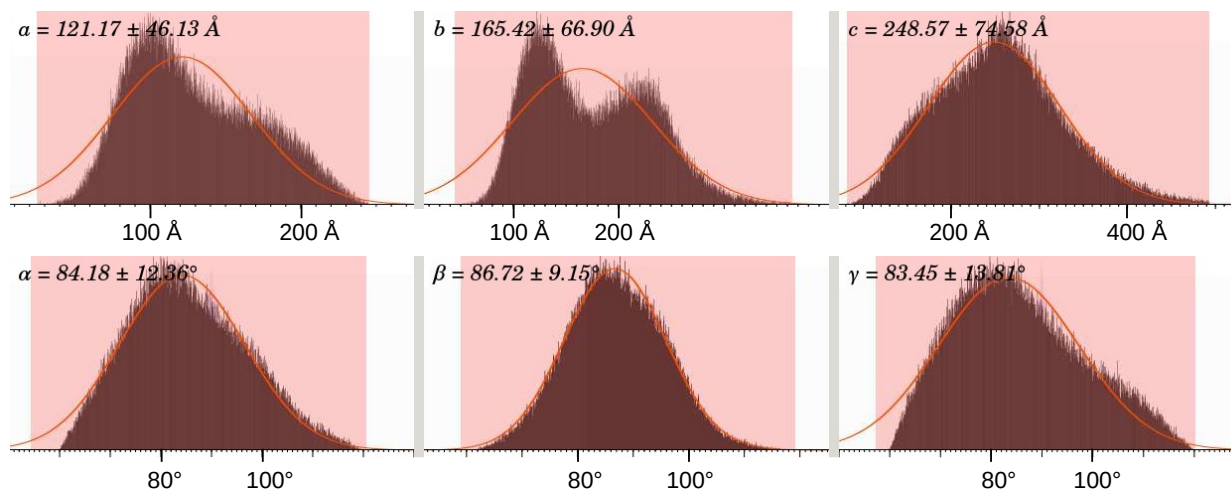


FIGURE 4.15: Indexing results for the MyD88 crystals in LCP. DirAx, MOSFLM and XGANDALF algorithms were used for indexing within the Psocake GUI. Unit cell distributions show broad, undefined unit cell parameters.

4.4.3 GDVN Injector

Sample Preparation

The original goal of this beamtime was for the optimisation of crystal concentration and GDVN flow parameters. The optimal crystal concentration was 7.5×10^8 crystals/ml (a 2x dilution from the neat crystal solution), as this resulted in the highest observed hit-rate. An optimal sample flow rate was obtained at 20 $\mu\text{L}/\text{min}$, producing a stable sample stream (Fig. 4.16). However, sample delivery protocols were further optimised, as the crystals had a tendency to stick to the in-line filters causing a decreased hit-rate. Several different filters were tested (10 μm & 20 μm stainless steel filters, and 10 μm & 20 μm Peek filters) with the 20 μm stainless steel filter working the best and providing the highest hit-rate. Flushing the filter with water at regular intervals also helped to prevent clogging, allowing us to maintain a stable hit-rate. During this experiment, regular sample mixing in order to mitigate crystal settling in the reservoir, was not automated, therefore the sample was manually vortexed before sample loading, and during data collection, in order to prevent this issue.

Data Collection & Analysis



FIGURE 4.16: The tip of the GDVN nozzle (top) with the jet stream shown in this figure flowing from the nozzle vertically down.

Initial data analysis was performed on a subset of the data collected at 186 mm from the detector. The Psocake GUI was used for the initial analysis, with the adaptive hit finding algorithm selected. Parameters were optimised, with a peak being identified if it consisted of a minimum of 3 pixels, and a maximum of 30 pixels. Thresholds were optimised to a single pixel threshold of 150, and a minimum peak threshold of 350. The optimal SNR was 6.4, with a rank of 3, radius of 4, and a dr of 2.

This initial analysis used a 5 peaks/pattern threshold, for a pattern to be classified as a 'hit'. Initial indexing with this subset of GDVN data resulted in a monoclinic C-centred unit cell indexing solution, with approximate lengths and angles of 100.5 Å, 31.5 Å, 54.5 Å, 90°, 107.4°, and 90°.

After the initial analysis of this data subset was performed, which included initial hit finding and indexing, further hit finding and indexing was then performed on the full data set (at the three different detector distances) utilising the hit finding algorithm "Peakfinder8", in Cheetah, which used the hit finding and indexing parameters that were previously optimised in Psocake, as the initial parameters. The initial hit finding performed in Cheetah included the use of a mask during hit finding in order to exclude bad pixels, detector edges, the jet-streak region, and, in the case of the 111 mm detector distance, the corner regions of the detector that were affected by detector shadowing (Fig. 4.17).

Each detector distance was treated as a separate data set for optimisation of hit finding parameters. The 186 mm and 286 mm detector distances both showed optimal hit finding with the following parameters; an SNR of 6, a minimum of 2 pixels/peak, and an ADC threshold of 150. While hit finding was optimal with these hit finding parameters, the data obtained with a 286 mm detector distance showed extremely low hit rates (0.025%),

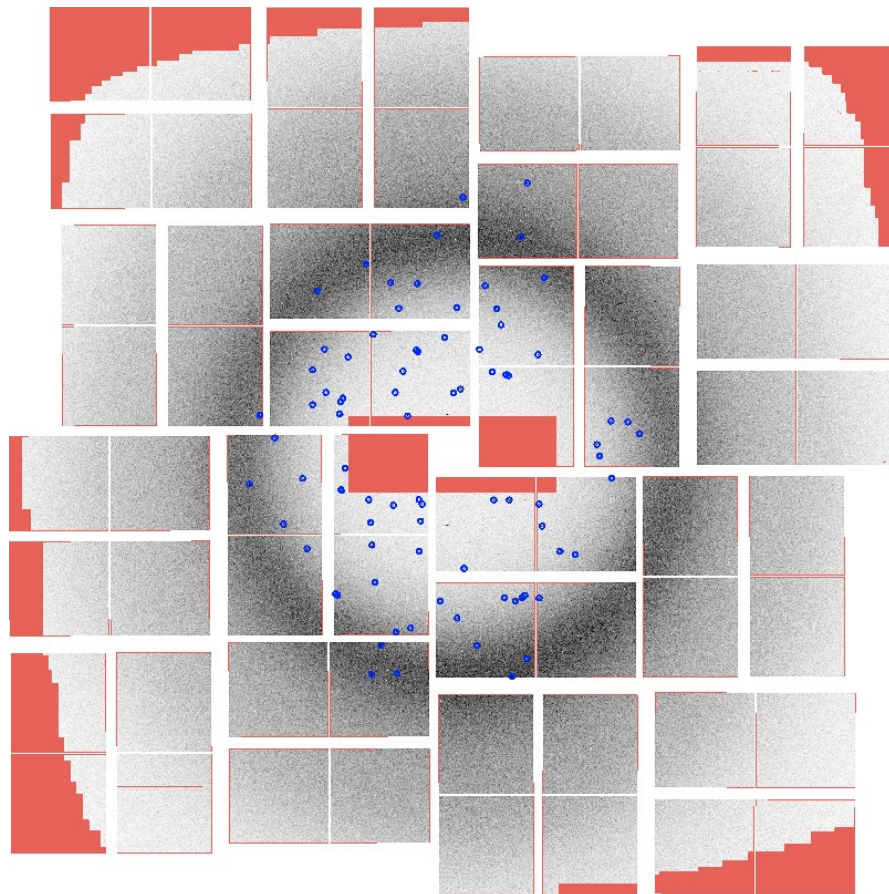


FIGURE 4.17: A GDVN MyD88 crystal detector image obtained from the CXI beamline at the LCLS, at a 111 mm detector distance. Initial hit finding included a mask (red) that was applied to all detector images to exclude bad pixels, detector edges, the jet streak region, and the detector shadowing region. Detector shadowing was only observed in data obtained at a 111 mm detector distance, and therefore the detector shadowing region was only masked for that data subset. Blue circles indicate locations where hit finding has identified Bragg peaks.

and therefore further analysis excluded this data subset. Data collected at the 111 mm detector distance was also optimised, with optimal parameters that included an SNR of 6.5, a minimum of 1 pixel/peak, and an ADC threshold of 400. A minimum of 15 peaks/image was required for all detector distances, for a detector image to be counted as a hit. After this initial hit finding, further observations of the detector images showed that many of them had little to no jet streaking (Fig. 4.18), and therefore hit finding was redone with the same parameters, but with the jet streak region unmasked, to ensure no low resolution peaks were inadvertently masked out.

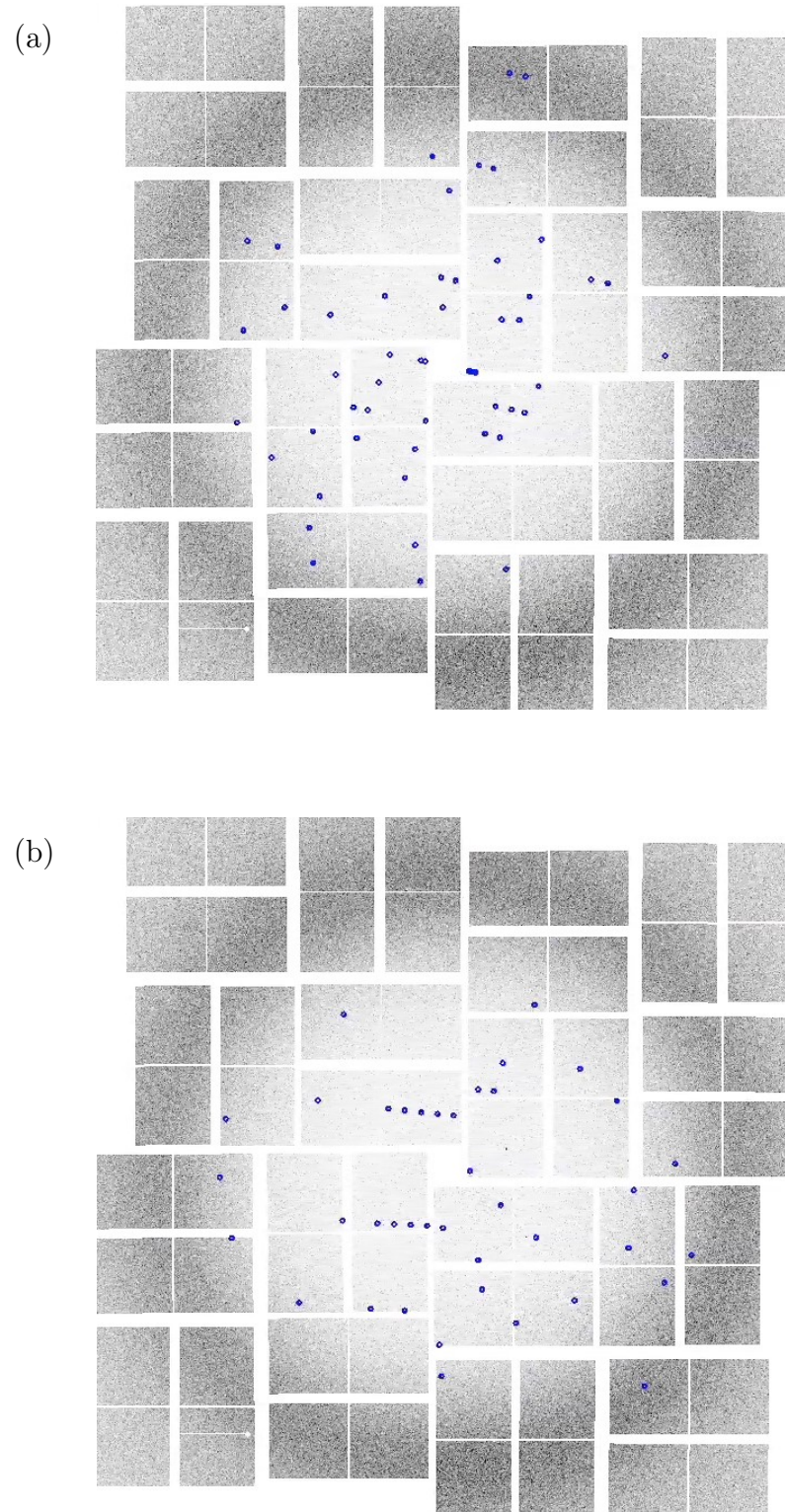


FIGURE 4.18: GDVN MyD88 crystal diffraction patterns obtained from the CXI beamline at the LCLS after hit finding, showing narrow, minimal jet streaking (a) and no jet streaking (b).

A 2D powder plot (Fig. 4.19), was also generated for the GDVN data, that summed the peaks identified during hit finding into a single image, in order to visualise the quality, completeness and resolution of the data.

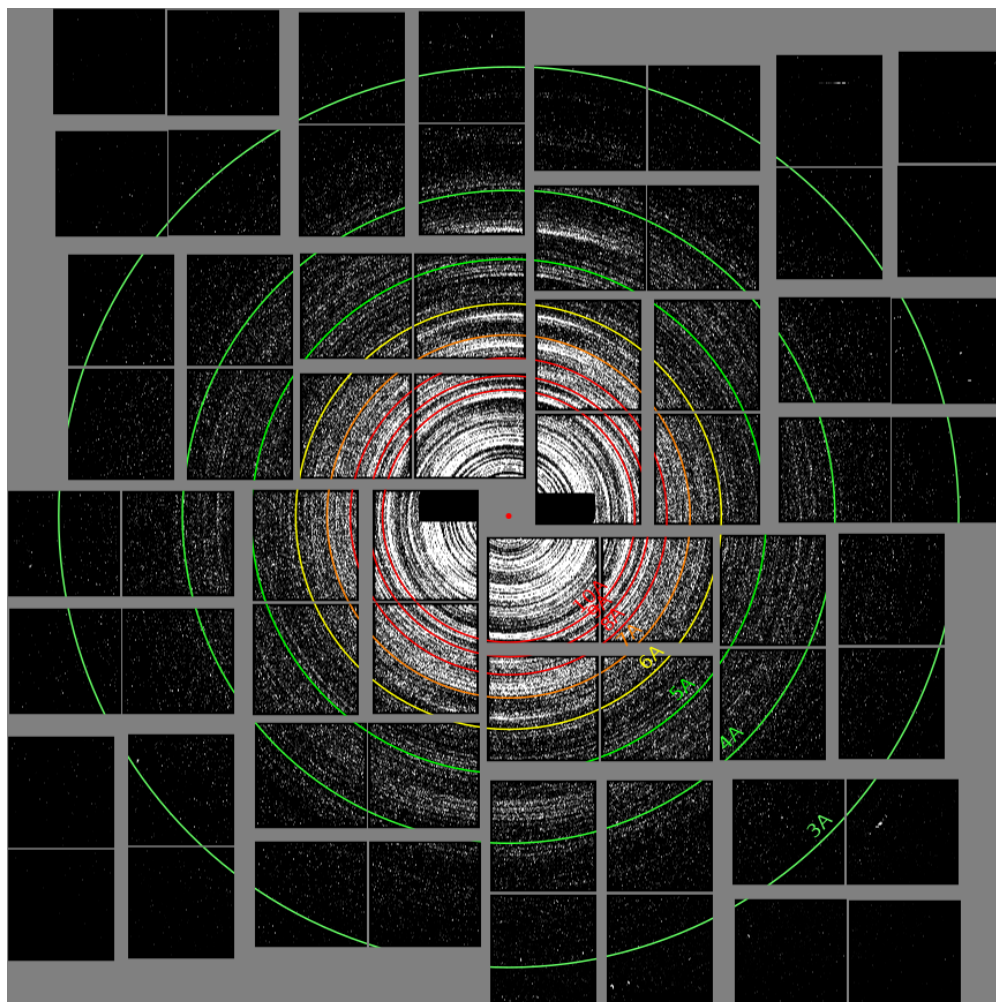


FIGURE 4.19: A 2D powder pattern generated from data obtained from GDVN delivered MyD88 crystals, using the CXI beamline at the LCLS. The powder pattern was generated by summing all the peaks identified during hit finding together into one image in order to visually estimate the resolution of our data.

Indexing was performed using Indexamajig, using the prior unit cell information obtained from the initial analysis. The output from this indexing stage was then used to optimise the detector position, rotation and distance using the Geoptimiser program in the CrystFEL software suite. The 186 mm detector distance was corrected to 181 mm while the 111 mm detector distance was corrected to 106 mm. After correcting the detector positioning using

Geoptimiser, indexing was performed, resulting in two slightly different unit cells from the two detector distances.

The indexing solution from the 181 mm data subset resulted in a unit cell that had a narrow, and symmetrical distribution. The indexing solution from the 106 mm data subset resulted in a slightly smaller unit cell, as well as a slight broadening, and visible skew to the unit cell distributions, that indicated that the detector distance and positioning needed further corrections. Therefore a second round of optimising using Geoptimiser was performed on the 106 mm data subset, with a final corrected detector distance of 106.1 mm. This correction resulted in agreement between the indexing solutions for the 181 mm and 106.1 mm data subsets, with both data subsets now providing the same unit cell solution.

Indexing was run using the MOSFLM, XGANDALF, and DirAx algorithms, with prior unit cell information being provided to the MOSFLM and XGANDALF algorithms during the final round of indexing. Bragg peaks were integrated using integration radii of 3, 4, and 5 [176, 189, 190]. Data subsets from the two detector distances were then merged. Figure 4.20 shows the indexing solutions for the full data set, with narrow and defined distributions observed.

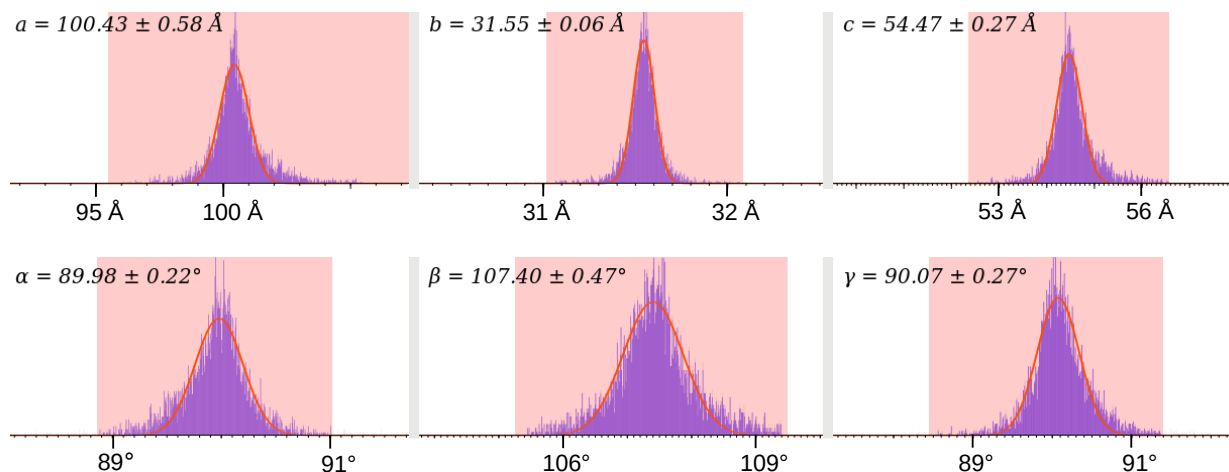


FIGURE 4.20: Indexing solutions obtained from data collected at 181 mm and 106.1 mm detector distances using GDVN delivery at the CXI beamline at the LCLS. The unit cell distributions have narrow and well defined peaks.

The final analysis resulted in a hit-rate of 1.31%, and an indexing rate of 34.93%. A unit cell of $a, b, c = 100.4, 31.5, 54.5$ and $\alpha, \beta, \gamma = 90.0^\circ, 107.4^\circ, 90.0^\circ$ was obtained. Table 4.1

provides an overview of the data collection statistics for the three delivery methods. Further analysis was performed on this data set, and the structure of MyD88 co-crystallised with MAL was solved to a resolution of 2.3 Å which is discussed in detail in Chapter 5.

TABLE 4.1: Crystallography data collection and analysis statistics. MyD88 crystal diffraction was obtained at the P11 beamline at the Petra III synchrotron using a microfluidic silicon chip system. MyD88 crystal diffraction was also obtained at the CXI beamline at the LCLS XFEL using the GDVN and HVI sample delivery systems. From these experiments, data analysis was performed, and the results of hit finding and indexing are presented in this table.

	Petra III Chip	LCLS LCP	LCLS GDVN
Total number of images	37,746	846,827	1,029,868
Total number of hits	10,754	514,167	13,528
Percentage hit rate (%)	28.48	60.72	1.31
Total number indexed	2,025	155,781	4,725
Indexing rate (%)	18.83	30.30	34.93
Resolution range (Å)	undetermined	undetermined	30.93-2.30 (2.38-2.30)
Data completeness (%)	undetermined	undetermined	91.4 (60.2)
Space group	undetermined	undetermined	C2
Unit cell			
a, b, c (Å)	undetermined	undetermined	100.40, 31.50, 54.50
α , β , γ (°)	undetermined	undetermined	90.00, 107.40, 90.00

4.4.4 Comparison

In order to attempt a comparison between the diffraction patterns obtained from the three delivery methods (chip, HVI, and chip), 1D plots were created from the 2D summed powder plots. The 1D plot generated from the LCP showed two intense peaks due to the solvent rings, and when overlayed with the 1D plots from the GDVN and chip data, there was no peak that could clearly be identified as being common to all three plots. However, we were able to identify a common peak between the chip and GDVN data. Therefore a comparison between the GDVN and chip radially averaged 1D plots was performed, with the data sets normalised to a common peak present in both (Fig. 4.21).

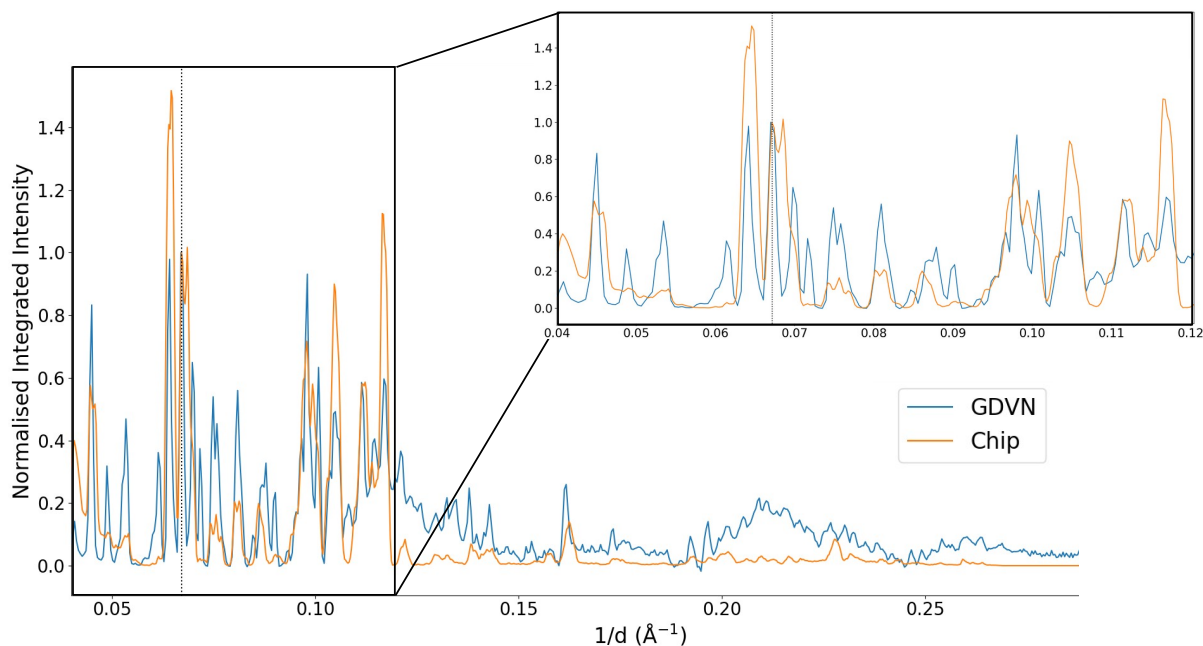


FIGURE 4.21: A comparison showing the normalised integrated intensity 1D plots for the GDVN and microfluidic silicon chip data, with an insert showing a clearer view of the area within the rectangle. The dotted line shown, is the common peak that was used to normalise both data sets.

4.5 Discussion

Single crystals that are well ordered and are greater than 10 μm in size, are ideal for conventional crystallography, however we are now in the era of nanocrystallography, where advances

in synchrotron science [25, 27, 196], and the introduction of the XFEL [41–44] have made it possible to collect data on crystals that are much smaller than this. Sample preparation and delivery however, are still a significant challenge for serial crystallography. When dealing with the combination of small crystals and crystal aggregation, the challenge of obtaining high quality serial crystallography data is further exacerbated.

In this chapter, we have demonstrated that choosing the best delivery system matched to a specific target is a fundamental requirement for achieving quality results. We have assessed three different delivery systems specifically developed for serial crystallography, with the goal of collecting high quality data in order to solve the structure of MyD88 co-crystallised with MAL. The MyD88 crystal system produces needle-like crystals that are very thin (100 nm) in one direction and are therefore too small for conventional crystallography. The MyD88 crystals also have a tendency to aggregate, which could cause issues during delivery and make it difficult to obtain single crystal diffraction. The chosen delivery systems tested include the GDVN system, HVI system, and a microfluidic silicon chip system. Each system has its pros and cons which are discussed in further detail in the following chapter.

The first sample delivery system investigated was the microfluidic silicon chip. Fixed-target chip-based delivery systems were initially designed with the goal of reducing sample consumption. Out of the three delivery systems tested, this holds true, with the microfluidic silicon chip system exhibiting the best performance with respect to sample consumption. With this delivery system, it is theoretically possible to collect an entire data set with as little as 10 μ L of the crystal solution, making it an appealing choice for when the protein system under investigation is difficult to express or grow in large volumes, such as is the case for the majority of membrane-bound proteins.

However, while this system outperformed the GDVN and LCP systems, with respect to sample consumption, crystal aggregation on the chip was observed, and this issue did not have a solution that could be added to the protocol on-the-fly. Specifically, it was during the stage where the excess liquid was drawn off via the underside of the chip, that crystal aggregation occurred, resulting in multi-crystal diffraction, which led to data analysis complications. This issue is a known problem with microfluidic chip systems, where Soares et al.,

2014 studied this same aggregation phenomenon, reporting that aggregation and preferential orientation occurred with various fixed-target delivery systems including micro-meshes, in-situ plates, and conveyor belts, and also that the aggregation was more pronounced with smaller crystals [197].

It is therefore necessary in the future to determine the best and most reproducible way in which to draw the excess liquid off the chip so that a more even distribution of crystals can occur, while being able to maintain a high enough concentration of crystals to maintain a high hit-rate. A number of factors were optimised with the goal of trying to minimise this aggregation issue, including the amount of mother liquor left on the chip, the initial concentration of crystals in the solution, and the humidity surrounding the sample setup when drawing off the mother liquor, and the crystal shape. Despite our best efforts at optimising these various parameters, crystal aggregation could not be avoided. While we were unsuccessful at minimising the aggregation problems for our needle-like MyD88 crystals, it is hypothesised that for crystals that grow to similar lengths in all 3 dimensions, it may still be possible to obtain an even distribution on the chip. This is because the chip's chosen pore size would be similar in size to the crystal's size, independent of which orientation the crystal was in when initially deposited on the chip's surface.

Once the sample was prepared and frozen in liquid nitrogen on the chip, the chip was then mounted onto a standard crystallography goniometer for data collection. This step was challenging as the current setup required manual mounting. Due to the strong magnetic connection between the chip (glued to a brass pin) and the goniometer, the force exerted would sometimes result in the chip snapping off the pin causing a loss of sample. This issue increased the time required for sample preparation, which meant less data was collected. Data collection was performed under a steady stream of liquid nitrogen, and over time layers of ice gradually built up on the surface of the chip, leading to the presence of increasingly intense ice rings in the detector images. More chips, and shorter collection times were therefore necessary to counteract this ice issue. The chip method was tested at room temperature in an attempt to circumvent ice build-up and involved the addition of a humidifier system. At the time of this experiment, the system proved unstable as the crystal solution was continually evaporating during data collection irrespective of the presence of the humidifier system,

and crystal diffraction was not detected.

While there were many challenges involved in the setup of this delivery system, once a chip had been successfully prepared and mounted, it was possible to collect a complete diffraction data set from a single chip. This method also provided very low background signal compared to the other methods tested. The low background signal was primarily due to: 1) the excess solution being drawn off the chip prior to the sample-beam interaction resulting in minimal solvent scatter, and 2) the capillary beam-stop being positioned as close as possible to the sample, minimising stray light entering the detector. While the low background signal allowed us to observe Bragg peaks from single crystal diffraction, due to a combination of the beam diameter, intensity, and crystal size, these Bragg peaks were sparse and seen mostly at low resolution. Successful indexing of diffraction patterns containing a single crystal was therefore difficult, as there were not enough peaks present for reliable analysis. Indexing solutions were still difficult to achieve for diffraction patterns that did have the necessary number of peaks, as these patterns tended to consist of very closely spaced peaks, indicating that they were most likely the product of multi-crystal diffraction.

At the time that this experiment was performed, it was not possible to test this fixed-target delivery system at an XFEL. The use of an XFEL rather than a synchrotron, could potentially have solved the issues relating to beam size and intensity, allowing the collection of higher resolution diffraction, and an increase to the number of Bragg peaks detected per pattern. Crystal aggregation issues would still be present, but with newer indexing algorithms, some that have been developed specifically to handle multi-crystal diffraction, it is possible that crystal aggregation would not be a significant hurdle. A micro-patterned silicon chip in combination with a high-speed goniometer called the "Roadrunner", has now been successfully tested at the LCLS, therefore making it possible to obtain XFEL data from MyD88 using a microfluidic silicon chip system [79]. Working in the same way as a synchrotron setup, but using a much faster goniometer system, so that the high pulse repetition rate of an XFEL is not underutilised due to goniometer speed limitations. The study by Roedig et. al. (2017), successfully determined the crystal structures of two viruses (picornavirus bovine enterovirus 2 (BEV2) and cytoplasmic polyhedrosis virus type 18 polyhedrin) requiring micrograms of sample, and less than 14 and 10 minutes of total data collection

time respectively [79].

Other fixed-target delivery systems available, included the nylon loop and conveyor belt methods [53]. The nylon loop method typically requires dozens of large crystals, of at least 50 μm in size. The crystals are rotated through the beam while also being moved vertically and horizontally in a helical manner, so that radiation induced damage is not observed in the diffraction patterns [53, 198]. It is therefore not a viable option for MyD88, as the crystals are too small. The conveyor belt delivery system can be successfully used with micron-sized crystals, but this method was designed for the purpose of time-resolved studies and requires two delivery steps; 1) delivery of the crystal sample to the conveyor belt, and 2) delivery of the sample via the conveyor belt to the beam [73, 77]. It is therefore unnecessarily complicated for a static structural determination of MyD88.

The second delivery system investigated in this study was the HVI injector using LCP as the delivery medium. The HVI injector was tested to determine if this approach could reduce crystal aggregation, therefore minimising the occurrence of multi-crystal diffraction. LCP was originally used as a medium for membrane-based protein crystal growth and delivery, mimicking cell membrane conditions, as it contains both hydrophobic and hydrophilic domains. It is now also used as a viscous delivery medium for crystals grown in solution. It was hypothesised that the combination of the hydrophobic and hydrophilic domains would reduce the amount of crystal aggregation that was occurring in the sample by maintaining an even concentration of crystals throughout the LCP. This injector has been successfully used for the collection of data with other crystal systems grown in solution [54, 57, 59, 66, 67], and therefore we considered it for our system. From the data collected using the HVI delivery system, the SNR was significantly higher (7.0 SNR) when compared to data collected at Petra III, using the chip delivery system (2.25 SNR). A higher SNR from the data obtained with the HVI delivery system at the LCLS is not surprising. While the background signal was extremely low for data collection with the chip system (most background counts ranged between 5-15 ADU), the sample signal was also extremely low due to the comparatively large beam diameter, and low intensity. Therefore, to fairly compare the SNR, the chip system should also be tested at an XFEL, so that the difference in signal intensity from the crystal compared to the background, is more apparent.

The LCP HVI experiment used parasitic beamtime, which meant we could not optimise the beam parameters for our sample. The use of parasitic beamtime is commonly used at XFELs as it allows two groups to collect data simultaneously; the primary users have control of the beam parameters and collect data with the primary beam, and the secondary (parasitic) users who have access to the refocused X-ray beam for data collection. In this case, the primary users selected a beam diameter of 3.5 μm with 10% X-ray transmission, and these conditions could therefore not be adjusted. The indexing solutions for the HVI data displayed the same broad and undefined unit cell distributions that were present in data obtained using the microfluidic silicon chip.

While difficulty in obtaining high resolution data is in part due to the large beam diameter and low transmission, the broad and undefined unit cell is indicative of multi-crystal diffraction. It was hoped that the hydrophilic and hydrophobic components of the LCP would encourage a more even distribution of the crystals, but from the presence of multi-crystal diffraction, it appears that the crystals may have become concentrated within the water channel domains of the LCP, although further testing is required to confirm this. While the HVI system may not prove suitable as a delivery medium for the needle-like MyD88 crystals that are first grown in solution, it cannot be completely discounted. MyD88 crystals grown in solution and then mixed with LCP seemed to clump together within water channels, but this would not be the case for a protein that is initially crystallised in LCP.

Furthermore, for crystals grown in solution, different high viscous media such cellulose or silicon grease which have less hydrophobic or hydrophilic components, may be a better choice for our crystal system. For the MyD88 crystals in particular, exchanging the LCP for a different HV media may allow for a more even crystal distribution. Depending on the stability, size, and polarity of crystals grown in solution, different HV media need to be tried and tested to determine whether the crystals remain evenly distributed throughout, or whether they tend to clump together.

The GDVN was the third delivery system that was investigated. The experimental conditions available when the GDVN was tested offered a number of advantages over the other experiments. During this experiment, we were the primary users, and therefore we could tune and optimise the beam size and intensity to the best conditions for our sample,

that is in contrast to when the HVI system was tested. The GDVN and LCP beamtimes were also advantageous compared to when the chip was used, as they were both performed at an XFEL. This meant a much higher intensity beam, suitable for collecting data from nano sized crystals could be used, while data collection using a microfluidic silicon chip was performed at a synchrotron, with a much lower flux.

The design of a GDVN system, allows the sample stream to be focused down to a diameter ranging from 0.1 - 2 μm , whereas the HVI injector is only capable of producing streams with diameters of 50 - 70 μm , dependent on nozzle size. Due to the much smaller diameter capability of the GDVN system, it is much more likely that a single crystal will be present within the X-ray beam interaction region, and therefore the GDVN is more likely to produce single crystal diffraction. The SNR observed from the GDVN data was 6.0, comparable to the SNR from data obtained using the HVI system.

As both experiments involved crystals in a solution or viscous medium, and both were performed at an XFEL, it is not unexpected that similar SNRs would be obtained. While some of the diffraction patterns obtained using the GDVN delivery system did have Bragg peaks positioned closely together, indicative of multi-crystal diffraction, this was seen far less frequently than what was observed when using the chip and HVI delivery systems. The indexing result obtained from the GDVN diffraction data showed narrow unit cell distributions which allowed us to correctly identify the unit cell for our crystal system.

One issue seen with the GDVN system, was clogging of the nozzle. When this issue occurs it results in a loss of beamtime while the nozzle is replaced. While the composition of the crystal solution needs to be optimised for crystal stability and background signal, it is also necessary to avoid specific solvents or concentrations of solvent that would result in clogging, allowing the GDVN system to perform optimally. It is therefore necessary to avoid high salt concentrations and specific buffers that are known to cause icing or salt crystals when exposed to vacuum conditions.

Another issue that occurred during experimental time was that crystals were sticking to the inline filters and causing blockages. In order to prevent crystal clogging and sticking during this experiment, the sample reservoir, where the crystals were sitting, was vortexed every 10 minutes. This seemed to prevent the crystals from settling in the sample reservoir.

It is now a routine part of data collection that the reservoir is continuously rotated in the hutch to eliminate issues involving crystal settling and aggregating, and therefore alleviating a major cause of multi-crystal diffraction. Flushing the sample lines intermittently with water was also performed to ameliorate the issue of inline filter clogging.

Newer variations of gas-focused nozzles have been developed, and are now being used, including the double flow focusing nozzle (DFFN) (described in detail in section 2.1.1). A DFFN works similarly to the GDVN, but the crystal solution is first surrounded by a sheath liquid such as ethanol, before both of these liquids are then surrounded by a high pressure gas stream for focusing. This method is therefore better for sample consumption than the GDVN. The addition of a sheath liquid to this type of gas-focused system, can also help prevent clogging of crystals at the nozzle exit, as the sheath liquid can act as a barrier for the sample, preventing evaporation or salting out. It was not possible to choose the DFFN delivery system for comparison for this chapter, as data collection using the GDVN was performed prior to its development.

There are several advantages to using injector type sample delivery for small crystals over static systems. The experimental setup of injection-based delivery systems involve less steps than static-based delivery methods and are therefore less time consuming. This is because sample preparation is minimal as the crystals need only be filtered prior to experimental use or in the case of the LCP injector, mixed with MO or a high viscosity media [54, 57, 59, 66, 67]. Both injector-based delivery systems are performed at room temperature eliminating the need for freezing the samples and avoiding the need for a cryo-protectant within the crystal solution, therefore allowing the sample to be delivered to the beam at a more biologically relevant temperature.

With respect to sample consumption, the microfluidic silicon chip system proved to be the most efficient, while the GDVN was the most inefficient. Using our chip, with a pore area of 1.5 mm x 1.5 mm, approximately 44,400 detector images could be obtained from a crystal volume of 2.5 μL (with 0.1 degree oscillation at each position). If we compare this with the consumption rates of the HVI and GDVN systems (where 2.5 μL of crystal solution is equivalent to 6.25 μL of LCP solution), using the HVI system, we obtain 15,000 detector images (at a 120 Hz repetition rate XFEL), while using the GDVN system, we would obtain

only 900 detector images (at a 120 Hz repetition rate XFEL) (see Table 4.2).

While the chip system is by far the best choice for the most efficient use of sample, when using the chip system at the Petra III synchrotron, it has the slowest data collection rate. To collect data from the MyD88 crystals at Petra III using the chip, an exposure time of 2 seconds was required. This meant that approximately 12 hours is needed to fully utilise a single chip and obtain 44,400 detector images. To obtain that same number of detector images at a 120 Hz XFEL using either the HVI or GDVN system, less than 10 minutes of data collection time would be required (Table 4.2). In future, with the use of the microfluidic silicon chip system at an XFEL, this data collection rate may increase to collection rates comparable to the injection-based delivery methods but will still require additional time to re-position for the next row of pores on the chip to be scanned.

TABLE 4.2: Sample delivery efficiency comparison. Based on a sample volume of 2.5 μL .

Method	Images obtained from 2.5 μL	Time to obtain 44,400 images
Chip system at Petra III	44,400	740 minutes
HVI system at LCLS	15,000	6.2 minutes
GDVN system at LCLS	900	6.2 minutes

Faster repetition rate XFELs are also available, with pulse train repetition rates of 10 Hz and intra-train repetition rates of up to 4.5 MHz. These high repetition rates translate to up to 27,000 pulses/second and therefore 44,400 images could theoretically be obtained in as little as 1.64 seconds. While these Megahertz collection rates are possible with GDVN-based delivery, high-speed-high-precision goniometer hardware and software has at this time only been developed with the capability of matching pulse rates of up to 120 Hz. HVI systems, initially designed to minimise sample wastage when performing serial crystallography at synchrotrons, are not capable of jetting at speeds fast enough to clear the radiation damaged sample from the interaction region and replenish with new sample before a second pulse occurs. It is therefore the GDVN method at Megahertz repetition rate facilities, that currently allows for the highest data collection rates to be obtained.

In reality, the optimal crystal concentrations for each delivery method may not result in the same hit rates. For low viscosity solutions that are delivered to the beam via GDVN, increasing the crystal concentration may not be feasible due to low crystal yield, while on the other hand increasing the crystal concentration too much could lead to clogging in the delivery lines and/or a predominance of multi-crystal diffraction.

The same issues are present for crystals grown in solution and then mixed with LCP for use in an HVI system. To maintain lipidic cubic phase, a maximum of 50% crystal solution is mixed with Monoolein. Furthermore, with too much water present, crystals will stay within the solution instead of embedding within the LCP with this leading to highly concentrated water channels throughout the LCP, or to the excess water being pushed out at the start of jetting, along with the majority of the crystals. Fixed-target delivery systems, such as microfluidic silicon chips have the potential to achieve hit rates of 100%, with studies showing that when a fixed-target system is used alongside in situ spectroscopy, a hit-rate approaching 100% is experimentally achievable [199].

When comparing the overall hit rate and indexing rate for serial crystallography the silicon chip and LCP setups provided much higher hit rates compared to the GDVN setup. These high hit rates indicate a high concentration of crystals in both instances. However, the silicon chip pore and crystal size combination lead to crystals pooling together in clumps on top of the chip rather than being evenly aligned within the pores, while it was suspected that the LCP caused the crystals to accumulate within its water channels, leading to an increase in crystal concentration. Therefore for the MyD88 crystal system the GDVN was the optimal sample delivery system.

4.6 Conclusion

Two key findings can be drawn from this serial crystallography sample delivery comparison chapter. Firstly, which of the three methods trialled was best for the MyD88 crystal system under investigation, and secondly, in what scenario would each of the three delivery methods deliver the best outcomes.

For the MyD88 crystal system, crystal aggregation was the limiting factor. The needle-like crystals were very prone to crystal aggregation, with a large proportion of the data from both the microfluidic silicon chip and HVI system producing multi-crystal diffraction that could not be correctly indexed. The GDVN therefore proved to be the most effective for collecting data from our MyD88 crystal system. In particular the ability to obtain a thin stream (1 μm), as well as a smaller beam size (2 μm) was critical for obtaining single crystal diffraction patterns. The GDVN delivery system allowed us to obtain high quality single crystal diffraction that was correctly indexed, and a unit cell determined. This provided the initial steps of our data analysis journey, ending in the structural determination of MyD88 which can be found in Chapter 5.

The HVI delivery method, if further optimised with a more appropriate HV media for our crystals could still be a possible choice for a delivery system in future. In this instance, because of the small size of our crystals, and our inability to clearly check if there is an even crystal distribution in the syringe after the crystals are mixed with the MO, a significant amount of time would be required for optimisation. Also, with so many developments and improvements to GDVN and chip-based delivery systems, the need to test and optimise the HV delivery media means that for use at XFELs, this system may remain a system of necessity for those membrane-bound proteins that require an LCP environment for successful growth. Furthermore, whereas the GDVN can be used for both static and time-resolved crystallography studies, the nature of the HVI system means that for light activated time-resolved crystallography studies, issues involving pre-illumination can occur, and for mix-and-diffuse time-resolved crystallography, the longer diffusion times of HV media may limit the scope of its usefulness [58].

As mentioned, it is hypothesised that the microfluidic silicon chip system could potentially optimise the pore sizing to better match the crystal size so that the crystals might rest above the pores as the mother liquor is drawn off, thereby minimising aggregation and allowing them to be positioned in a evenly distributed array, as seen in Figure 4.3. While this step could possibly minimise crystal aggregation, drawing off the excess liquid will always have an intrinsic risk of causing crystal aggregation and preferential orientation. A better solution to this issue can be seen with the development of on-chip crystallisation by Lieske et al.

2019 [200]. This method aimed to eliminate crystal aggregation issues, while also producing higher quality crystals due to the minimised handling steps involved when crystals are grown on-chip. This method was successful in demonstrating that on-chip grown crystals remain in place upon the removal of the mother liquor, with no aggregation observed [200].

New chip designs, such as the "Roadrunner II chip", have also improved what framing and mounting materials are used, so that the chip is easily mounted in a defined and reproducible position [200]. Lieske et al. [200], has also developed strategies for maintaining the correct humidity throughout the setup and experimental phases. After the excess mother liquor is removed from the chip, the "humidor", a chip cover can be slid over the chip in order to help maintain a humid environment, before the chip is mounted onto the goniometer. For experimentation, a constant stream of humidified helium is used to prevent the crystals from drying out [200], therefore eliminating the need for cryogenic data collection conditions.

The GDVN has been the workhorse of delivery systems for serial crystallography experiments. While the GDVN consumes much larger sample volumes in comparison to other systems, the development of the DFFN has meant that this disadvantage can be mitigated. However, the high rate of sample consumption necessary for both the GDVN and DFFN delivery systems, means that crystals that are limited by their low yield production may be more suited to static, low sample consumption methods. This is not the case for MyD88, and with the GDVN delivery system at the LCLS, we have collected and analysed data to determine the structure of MyD88 (Chapter 5).

Chapter 5

The Structural Determination of MyD88

5.1 Introduction

Solving the structure of proteins through serial crystallography using XFELs, provides the opportunity to collect data from nanocrystals at room temperature, under near-physiological conditions. The retrieval of molecular structures under these conditions can help with understanding how proteins interact in their native state, inside an organism. Knowing how these structures fold, and the details of protein interactions, allows scientists to design drugs based on experimental data rather than by trial and error. As mentioned in Chapter 4, the Toll/interleukin-1 receptor (TIR) domain of MyD88, is involved in a Toll-like receptor (TLR) mediated inflammatory response in the human body. It has been associated with a number of pathological states including infectious, autoimmune, inflammatory, cardiovascular and cancer-related disorders. Cytoplasmic signalling within a cell by TLRs starts with their domains interacting with TIR-containing adaptor proteins that include both MyD88 and MyD88-Adaptor-Like protein (MAL). Recruitment of these adaptors to the TLRs via TIR:TIR interactions is what initialises downstream signalling pathways, leading to the induction of pro-inflammatory genes, which means increased expression. Although, TLR pathways have been well characterised [177, 179, 181, 182, 201–205], molecular information for signalling proteins is still quite limited, and this impedes the development of therapeutic strategies, and the understanding of the effects of polymorphic variants on human disease.

The structure of MyD88 has previously been determined via X-ray crystallography [183], however its structure has not been solved in the case where MyD88 is co-crystallised with its native binding partner MAL. The structure of MAL is known and will provide biologically relevant insights into how the MyD88 and MAL interact. Therefore, the goal of this chapter is to solve and present the structure of MyD88, co-crystallised with MAL, and to expand our knowledge of how MyD88 and MAL interactions occur within our cells.

The team at La Trobe University collaborated with Dr Thomas Ve, at Griffith University, and Professor Bostjan Kobe, at the University of Queensland, who provided biological and biochemical expertise and co-crystallised MyD88 and MAL proteins. A number of synchrotron and XFEL experiments were performed on the MyD88 crystals, as discussed in Chapter 4. These experimental facilities included Petra III, LCLS, and NSLS II, although diffraction was not observed at the NSLS II as the experimental parameters could not be freely optimised. The data obtained at the LCLS using a GDVN delivery system yielded high quality diffraction data which was used to solve the MyD88 structure [110], presented in this chapter. I am joint first author on this publication; my contribution was to solve the SFX structure and compare the SFX and microED results. A comparison of the SFX and microED structure was performed in order to determine whether the different techniques would result in structural differences. It should be noted though that the microED experiment was performed under cryogenic conditions, while the SFX experiment was performed at room temperature, which needed to be taken into account when comparing the resulting structures. The publication describes the successful structural determination of MyD88, and also emphasises its relevance to biological function. Furthermore, the publication compares our MyD88 structure, co-crystallised with MAL, with other MyD88 structures that are already published in order to gain insight into how co-crystallisation affects the structure, and the availability of specific regions for protein interactions. The result was the development of a model for MyD88 oligomerisation, induced by the initial binding of MAL that had previously been unknown.

5.2 Publication

TABLE 5.1: Candidate and co-author contributions

Manuscript Title	“MyD88 TIR domain higher-order assembly interactions revealed by microcrystal electron diffraction and serial femtosecond crystallography”
Journal	Nature Communications doi: 10.1038/s41467-021-22590-6
Publication Status	Published
Authors	Max T.B. Clabbers [†] , Susannah Holmes[†] , Timothy W. Muusse, Parimala Vajjhala, Sara J. Thygesen, Alpeshkumar K. Malde, Dominic J.B. Hunter, Tristan I. Croll, Leonie Flueckiger, Jeffrey D. Nanson, Md. Habibur Rahaman, Andrew Aquila, Mark S. Hunter, Mengning Liang, Chun Hong Yoon, Jingjing Zhao, Nadia A. Zatsepin, Brian Abbey, Emma Siernecki, Yann Gambin, Katelyn J. Stacey, Connie Darmanin, Bostjan Kobe, Hongyi Xu & Thomas Ve
Nature and extent of candidates contribution	Thirty-five percent including data collection, analysis, SFX structural determination, comparative analysis between SFX and microED structures, and interpretation of the results.
Nature and extent of co-authors contributions	Sixty-five percent including MyD88 mutation studies, crystallisation, data collection, analysis, microED structural determination, and interpretation of results.

[†]These authors contributed equally to this work

ARTICLE


<https://doi.org/10.1038/s41467-021-22590-6>

OPEN

MyD88 TIR domain higher-order assembly interactions revealed by microcrystal electron diffraction and serial femtosecond crystallography

Max T. B. Clabbers^{1,10,11}, Susannah Holmes^{2,11}, Timothy W. Muusse³, Parimala R. Vajjhala³, Sara J. Thygesen³, Alpeshkumar K. Malde⁴, Dominic J. B. Hunter^{3,5,6}, Tristan I. Croll⁷, Leonie Flueckiger², Jeffrey D. Nanson³, Md. Habibur Rahaman³, Andrew Aquila⁸, Mark S. Hunter⁸, Mengning Liang⁸, Chun Hong Yoon⁸, Jingjing Zhao¹, Nadia A. Zatsepin², Brian Abbey², Emma Sieracki⁵, Yann Gambin⁵, Katryn J. Stacey^{3,6,9}, Connie Darmanin², Bostjan Kobe^{3,6,9}, Hongyi Xu¹ & Thomas Ve⁴

MyD88 and MAL are Toll-like receptor (TLR) adaptors that signal to induce pro-inflammatory cytokine production. We previously observed that the TIR domain of MAL (MAL^{TIR}) forms filaments in vitro and induces formation of crystalline higher-order assemblies of the MyD88 TIR domain (MyD88^{TIR}). These crystals are too small for conventional X-ray crystallography, but are ideally suited to structure determination by microcrystal electron diffraction (MicroED) and serial femtosecond crystallography (SFX). Here, we present MicroED and SFX structures of the MyD88^{TIR} assembly, which reveal a two-stranded higher-order assembly arrangement of TIR domains analogous to that seen previously for MAL^{TIR}. We demonstrate via mutagenesis that the MyD88^{TIR} assembly interfaces are critical for TLR4 signaling in vivo, and we show that MAL promotes unidirectional assembly of MyD88^{TIR}. Collectively, our studies provide structural and mechanistic insight into TLR signal transduction and allow a direct comparison of the MicroED and SFX techniques.

¹Department of Materials and Environmental Chemistry, Stockholm University, Stockholm, Sweden. ²Australian Research Council Centre of Excellence in Advanced Molecular Imaging, Department of Chemistry and Physics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Victoria, Australia. ³School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland, Australia. ⁴Institute for Glycomics, Griffith University, Southport, Queensland, Australia. ⁵EMBL Australia Node in Single Molecule Science, University of New South Wales, Kensington, New South Wales, Australia. ⁶Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia. ⁷Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK. ⁸Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, California, USA. ⁹Australian Infectious Diseases Research Centre, The University of Queensland, Brisbane, Queensland, Australia. ¹⁰Present address: Department of Biological Chemistry, University of California Los Angeles, Los Angeles, California, USA. ¹¹These authors contributed equally: Max T. B. Clabbers, Susannah Holmes. ✉email: c.darmanin@latrobe.edu.au; b.kobe@uq.edu.au; hongyi.xu@mmk.su.se; t.ve@griffith.edu.au

Toll-like receptors (TLRs) detect pathogens and endogenous danger-associated molecules, initiating innate immune responses that lead to the production of pro-inflammatory cytokines. Signaling by TLRs is initiated by dimerization of their cytoplasmic TIR (Toll/interleukin-1 receptor [IL-1R]) domains, followed by recruitment of the TIR-containing adaptor proteins, including MyD88 (myeloid differentiation primary response gene 88) and MAL (MyD88 adaptor-like/TIRAP) (Fig. 1)¹. Combinatorial recruitment of these adaptors via TIR:TIR interactions orchestrates downstream signaling, leading to induction of the pro-inflammatory genes. In previous work, we showed that MAL TIR domains (MAL^{TIR}) spontaneously and reversibly form filaments in vitro. They also formed co-filaments with TLR4 TIR domains (TLR4^{TIR}) and nucleated the assembly of MyD88^{TIR} into crystalline arrays². These results suggested signaling by cooperative assembly formation (SCAF), a mechanism prevalent in innate-immunity and cell-death pathways^{3,4}, and we proposed a model for signal amplification, in which the TLR4, MAL and MyD88 TIR domains sequentially and cooperatively assemble into a higher-order TIR domain complex. This assembly then induces the formation of the Myddosome, involving the death domains of MyD88 and the protein kinases, IRAK2 and IRAK4, leading to proximity-based activation of these kinases (Fig. 1)^{5,6}. The 7 Å cryogenic electron microscopy (cryo-EM) structure of the MAL^{TIR} filament revealed a hollow tube composed of 12 two-stranded protofilaments of TIR domains and mutational analyses revealed that protein interactions within these protofilaments are likely to represent higher-order TIR-domain interaction interfaces during in vivo signaling, although the structures formed within cells may be more limited in size⁷. However, the structural basis of how MyD88^{TIR} and TLR4^{TIR} domains self-assemble and interact with MAL^{TIR} remained uncharacterized.

Here we set out to structurally characterize the MyD88^{TIR} crystalline assemblies observed in our previous work². As the

crystals were too small for conventional X-ray crystallography, we employed the complementary techniques of microcrystal electron diffraction (MicroED) and serial femtosecond crystallography (SFX). MicroED^{8,9} enables structure determination of submicrometre-sized crystals. In MicroED data collection, the crystal is continuously rotated in a transmission electron microscope (TEM)^{10–12}, analogous to the rotation method used in X-ray crystallography¹³, and to related three-dimensional electron diffraction methods in TEM¹⁴. MicroED can complement existing methods in structural biology such as conventional X-ray crystallography, where growing crystals of sufficient size and crystallinity is often the major barrier to structure determination^{15–18}. Indeed, many failed crystallization trials have been shown to contain microcrystals^{19–21}. Furthermore, small macromolecular crystals potentially have reduced defects^{22–24}, and controlled perturbations to the sample, such as soaking and vitrification, may be applied rapidly and more uniformly^{24–26}. MicroED has already enabled protein structure determination from microcrystals^{9,10,23,27–30}, structure solution of a previously uncharacterized metalloenzyme³¹, structure determination of membrane proteins from microcrystals embedded in lipidic cubic phase^{32–35} and the visualization of ligand-binding interactions^{25,36}. Furthermore, MicroED enables the study of biomolecules that naturally aggregate or assemble into microcrystals, facilitating structure determination of several short peptide fragments from thin prion protofibrils^{23,37–39}. Such naturally occurring crystalline assemblies are of special interest, as they can reveal the interactions occurring in assemblies within cells, illustrating the underlying mechanisms guiding the assembly formation and providing relevant structural insights.

More or less in parallel to the development of MicroED, SFX has emerged as a powerful technique for structure determination and the study of protein dynamics of microcrystalline samples^{24,40–43}. SFX exploits the femtosecond-scale duration of extremely brilliant X-ray free-electron laser (XFEL) pulses for the collection of high-quality diffraction data at room temperature, which occurs before the onset of structure-altering radiation damage^{44–47}. In SFX, diffraction data are collected as single snapshots from randomly oriented microcrystals^{45,47}. With the crystals delivered to the beam at room temperature and minimal sample handling, challenges associated with cryo-cooling and potential protein conformation restrictions are avoided⁴⁸. SFX has facilitated structure determination from submicrometre crystals of radiation-sensitive proteins^{49–51} and membrane proteins such as G protein-coupled receptors^{52–55}. SFX has also enabled time-resolved studies of light-sensitive proteins with unprecedented temporal resolution^{55–57}, enabling the study of reactions initiated by ligand binding and exploiting the sub-micrometre crystal size for rapid reaction initiation^{49–51,54,56–58}. In particular, SFX has advanced fibril studies, e.g., amyloids or microtubules, where the fibrous biomolecule assemblies may have partial or no crystallinity, approaching the regime of single-molecule imaging^{59,60}.

Here we present MicroED and SFX structures of the MAL-induced MyD88^{TIR} microcrystals at 3.0 Å and 2.3 Å resolution, respectively. Importantly, both structures show several distinct remodelled loop regions that adopt conformations that are different from previously determined monomeric X-ray and nuclear magnetic resonance (NMR) structures^{61,62}. Crystal packing analysis revealed that the MAL-induced MyD88^{TIR} crystals have a two-stranded higher-order assembly arrangement of TIR domains identical to that observed previously within spontaneously formed MAL^{TIR} filaments², and mutagenesis studies demonstrated that the interfaces within these higher-order MyD88^{TIR} assemblies are important for signaling. This identical architecture suggested a unidirectional templating mechanism for

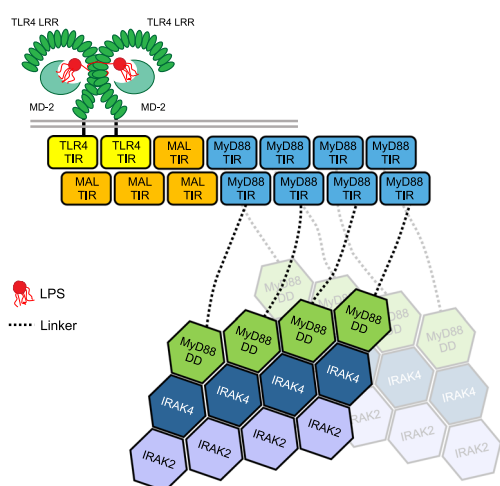


Fig. 1 Schematic diagram of the SCAF model for TLR signaling. Pathogen-associated molecular patterns (e.g., LPS) binding to the extracellular LRR domain of a TLR (e.g., TLR4) induces dimerization of its TIR domains, which leads to the recruitment of an adaptor TIR domain (e.g., MAL^{TIR}) to the extended surface created by the TLR4^{TIR} dimer. Elongation of this trimer through recruiting additional adaptor's TIR domains (e.g., MAL^{TIR} or MyD88^{TIR}) into a higher-order complex leads to clustering of MyD88 DDs and subsequent recruitment of IRAKs through DD interactions. The initial TIR dimerization and trimerization steps are likely to be unfavourable and rate limiting, whereas subsequent monomer additions are more favourable, rapid and cooperative. LRR, leucine-rich repeat domain; LPS, lipopolysaccharide; TLR, Toll-like receptor; TIR, Toll/interleukin-1 receptor domain; DD, death domain.

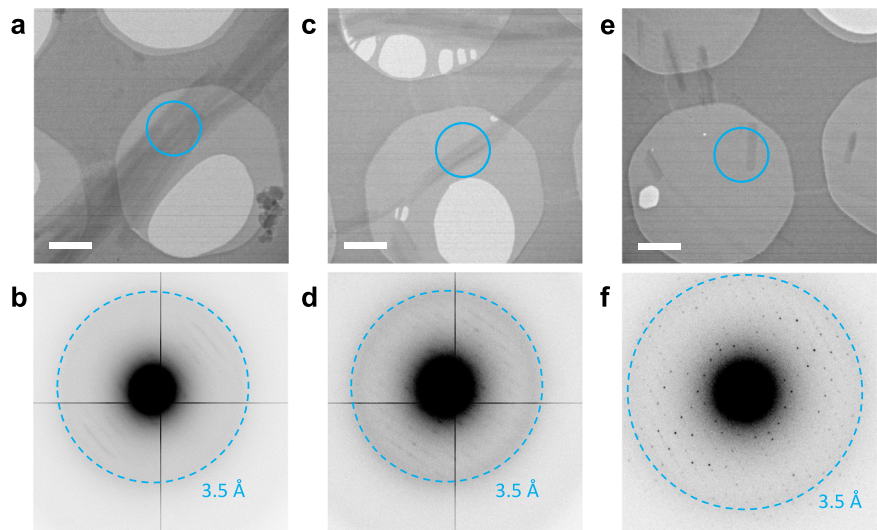


Fig. 2 MicroED data collection from MyD88^{TIR} microcrystals. **a, b** Electron micrograph of aggregated microcrystals, only showing poor-quality diffraction data. Scale bar, 1 μm. **c, d** Multiple microcrystals are overlapping, showing multiple lattices in their corresponding diffraction patterns, complicating data indexing. Scale bar, 1 μm. **e, f** Single hydrated microcrystal, showing high-quality diffraction data up to 3.0 Å resolution. Scale bar, 1 μm. The cyan rings on the micrographs indicate the 1.5 μm diameter parallel beam, defined by the selected area aperture, used for MicroED data collection. Electron diffraction patterns were collected with an angular increment of 0.68° per frame, at a dose rate of 0.12 e⁻/Å² per frame. The data in **a–f** are representative of three EM grids prepared using 3 μl of a 1:50 MAL^{TIR}:MyD88^{TIR} crystal solution.

Table 1 Data collection statistics.		
Data collection	MicroED ^{a,b}	SFX ^c
Temperature (K)	77	300
Space group	C2	C2
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	99.06, 31.01, 54.30	100.40, 31.50, 54.50
<i>α</i> , <i>β</i> , <i>γ</i> (°)	90.00, 107.70, 90.00	90.00, 107.40, 90.00
Resolution (Å)	30.54–3.00 (3.11–3.00)	30.93–2.30 (2.38–2.30)
<i>R</i> _{merge}	0.46 (0.95)	-
<i>R</i> _{split}	-	0.34 (1.3)
Mean <i>I</i> /σ(<i>I</i>)	4.8 (1.8)	2.6 (1.4)
CC _{1/2}	0.95 (0.43)	0.90 (0.36)
CC*	0.99 (0.77)	0.97 (0.73)
Completeness (%)	73.7 (57.3)	91.4 (60.2)
Multiplicity	12.2 (6.0)	24.2 (3.4)

Values in parentheses are for the highest-resolution shell. Intensity statistics were generated from phenix.table_one⁹⁸ for the MicroED data and from CrystFEL^{107,128} for the SFX data.

^aMerged data from 18 crystals.

^bMicroED data were truncated at mean *I*/σ(*I*) ≥ 1.5 and CC_{1/2} ≥ 0.4⁹⁴.

^cMerged data from 4725 indexed snapshots out of 13,528 hits.

nucleation and assembly of the higher-order MyD88^{TIR} oligomers, which we confirmed using crystal growth assays. Moreover, structural comparison of the MyD88^{TIR} higher-order assembly and monomeric MyD88^{TIR} enabled us to understand the conformational changes that MyD88^{TIR} monomers undergo upon joining the higher-order assembly. Collectively, our studies shed light on the hierarchical nature of the SCAF mechanism operating in TLR and IL-1R pathways.

Results

Data acquisition. The MAL-induced MyD88^{TIR} microcrystals were typically 100–200 nm in diameter, making them ideally suited to both MicroED (Fig. 2) and SFX.

The microcrystals were deposited on Quantifoil EM grids and vitrified for screening and MicroED data acquisition

(Supplementary Fig. 1). The microcrystals had a tendency to aggregate, forming large bundles that diffracted poorly (Fig. 2a, b). Furthermore, the bent and overlapping crystals complicated the data interpretation (Fig. 2c, d). Using a small parallel electron beam of 1.5 μm diameter, defined by the selected area aperture, only single thin hydrated microcrystals were selected for MicroED data collection (Fig. 2e, f). The MyD88^{TIR} microcrystals diffracted to 3.0 Å resolution and provided high-quality electron diffraction data (Fig. 2f). Data from 18 crystals were integrated, scaled and merged (Table 1). The overall completeness is limited, owing to a preferred orientation of the MyD88^{TIR} microcrystals on the grid and because of the limited tilt range of the goniometer.

The MyD88^{TIR} microcrystals were studied in parallel using SFX. Initially, serial crystallography was attempted on a fixed target at the PETRAIII PII beamline, with a beam size of 2 × 2 μm. However, in this setup, data collection and analysis were complicated by the frequent bundling of microcrystals into larger aggregates (Supplementary Fig. 2). To reach higher resolution and overcome microcrystal aggregation, the sample was delivered as a stream of solvated microcrystals with a gas-dynamic virtual nozzle (GDVN) injector^{63,64} to a pulsed XFEL beam at the Linac Coherent Light Source (LCLS), SLAC National Acceleratory Laboratory⁶⁵. By using a micro-focused beam (nominally 1 × 1 μm full width at half maximum (FWHM)), and optimizing crystal concentration (7.5 × 10⁸ crystals/ml), high-quality diffraction patterns from individual crystals were collected. Overall, the SFX dataset comprised 4725 indexed patterns from 13,528 hits (35% indexing rate) out of 1,029,868 detector frames (average hit rate of 1.3%). The lattice parameters derived from the SFX data were found to be slightly larger than in the MicroED data collected under cryo-conditions (Table 1).

Structure solution, model building and refinement. The structure of MyD88^{TIR} was initially solved using the MicroED data by molecular replacement, finding a well-contrasting unique solution in space group C2. The solution was found with a search model derived from a distantly related Toll-related receptor 2 (TRR2) TIR domain, sharing only 30% sequence identity with

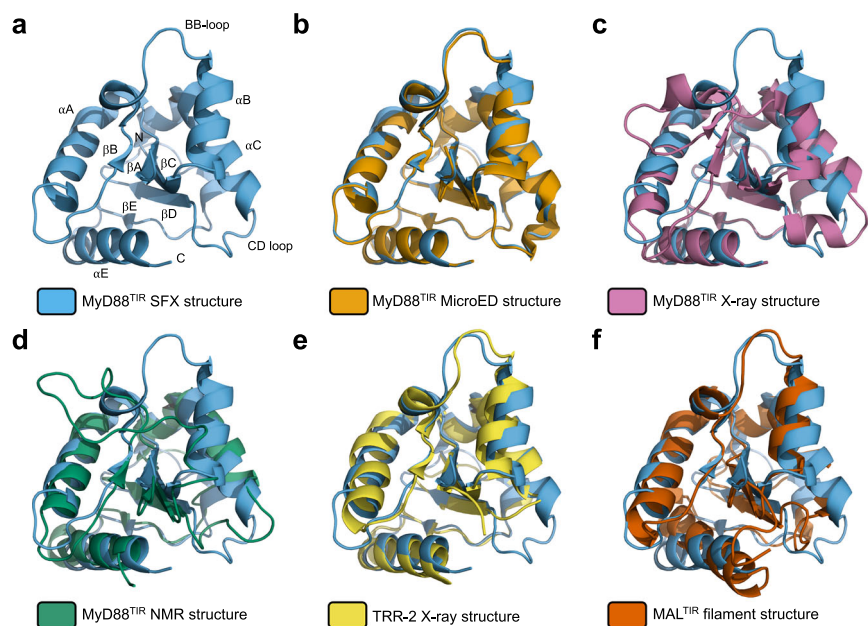


Fig. 3 MyD88^{TIR} structure comparison. **a** Ribbon diagram (blue) of a monomer from the MyD88^{TIR} higher-order assembly structure. Structural elements are labelled sequentially in TIR domains, with the BB-loop connecting strand βB with helix αB, according to the established nomenclature¹²⁹. **b–f** Superposition of the MyD88^{TIR} SFX structure (blue), with **b** the MyD88^{TIR} MicroED structure (orange); **c** the monomeric MyD88^{TIR} X-ray crystal structure (PDB ID 4EO7; magenta); **d** the monomeric MyD88^{TIR} NMR solution structure (PDB ID 2Z5V; green); **e** the crystal structure of the TIR domain of the Toll-related receptor TRR2 from the lower metazoan *Hydra vulgaris* (PDB ID 4W8G; yellow); and **f** the MAL^{TIR} higher-order assembly cryo-EM structure (PDB ID 5UZF; red).

Table 2 Refinement statistics.			
Refinement	MicroED	SFX ^a	SFX ^b
Refinement program	phenix.refine	REFMAC5	phenix.refine
Resolution (Å)	30.54–3.00	30.94–2.30	30.93–2.30
No. reflections	2436	6352	6687
<i>R</i> _{work} / <i>R</i> _{free}	0.223/0.280	0.220/0.270	0.239/0.281
Mean <i>B</i> -factor (Å ²)	52.01	40.00	45.60
R.M.S. deviations			
Bond lengths (Å)	0.005	0.002	0.001
Bond angles (°)	0.524	1.191	0.370
Ramachandran			
Favoured (%)	97.79	98.53	99.26
Allowed (%)	2.21	1.47	0.74
Outliers (%)	0.00	0.00	0.00
Clashscore	4.38	1.70	3.94
Rotamer outliers (%)	0.00	0.76	0.00

^aSFX structure refinement using the REFMAC5 refinement programme.
^bSFX structure refinement using the MicroED structure refinement protocol.

MyD88^{TIR} (Fig. 3). The structure of MyD88^{TIR} was iteratively built and refined using the MicroED data (Table 2) and, despite moderate completeness and resolution, the electrostatic potential map showed well-resolved features and enabled remodelling of the loop regions that differed from the previously determined monomeric crystal and solution structures^{61,62} (Figs. 3 and 4a). The higher-resolution SFX structure (2.3 Å) was first solved using the MicroED MyD88^{TIR} model as a template for molecular replacement followed by iterative rebuilding and refinement using a different protocol compared to the MicroED structure (Table 2, SFX^a). To enable a direct comparison between the MicroED and SFX models, we also solved, rebuilt and refined the SFX

MyD88^{TIR} structure using an identical protocol as described for the MicroED data (Table 2, SFX^b). The SFX^a map (Fig. 4b) showed well-resolved features, including water molecules that were not modelled in the MicroED structure. To check whether the MicroED and SFX maps were biased by the search model, simulated annealing (SA) composite omit maps were calculated, confirming the interpretation of our structural models (Supplementary Fig. 3). As the microcrystals contain a small proportion of MAL^{TIR} molecules, there may be a contribution of this heterogeneity to the diffraction, but this is likely to have a negligible effect. Accordingly, there is no evidence of the presence of MAL^{TIR} molecules in the electron density and electrostatic potential maps of the MAL-induced MyD88^{TIR} crystals.

Structural comparison of MyD88^{TIR} structures. The MicroED and SFX^b MyD88^{TIR} structures, which were built and refined using the same protocol, are almost identical, with a root mean square deviation (RMSD) of 0.4 Å for 138 Ca atoms. Minor differences in some side-chain conformations can be observed, which is most likely due to the flexibility of certain regions resulting in poorly defined electron density or as a result of the difference in the data collection temperature (Supplementary Fig. 4). The MyD88^{TIR} SFX^a structure was used for the comparison with other TIR domain structures and for the analyses of interaction interfaces within the crystal. The structure of MyD88^{TIR} within the MAL-induced higher-order assembly exhibited conformational differences from the known NMR (RMSD of 2.4 Å for 107 Ca atoms)⁶¹ and X-ray (RMSD of 2.0 Å for 118 Ca atoms)⁶² structures of monomeric MyD88^{TIR}. This is especially apparent in the region encompassing the BB loop and αB helix, and in the CD loop (Fig. 3). The conformational differences are likely due to participation of these regions in TIR:TIR interactions within the MAL-induced higher-order assemblies. Among the known TIR domain structures, the MAL^{TIR} filament structure (Fig. 3) and the TLR1, TLR2, TLR6 and IL-RACP crystal structures possess similar BB-loop and αB-helix conformations².

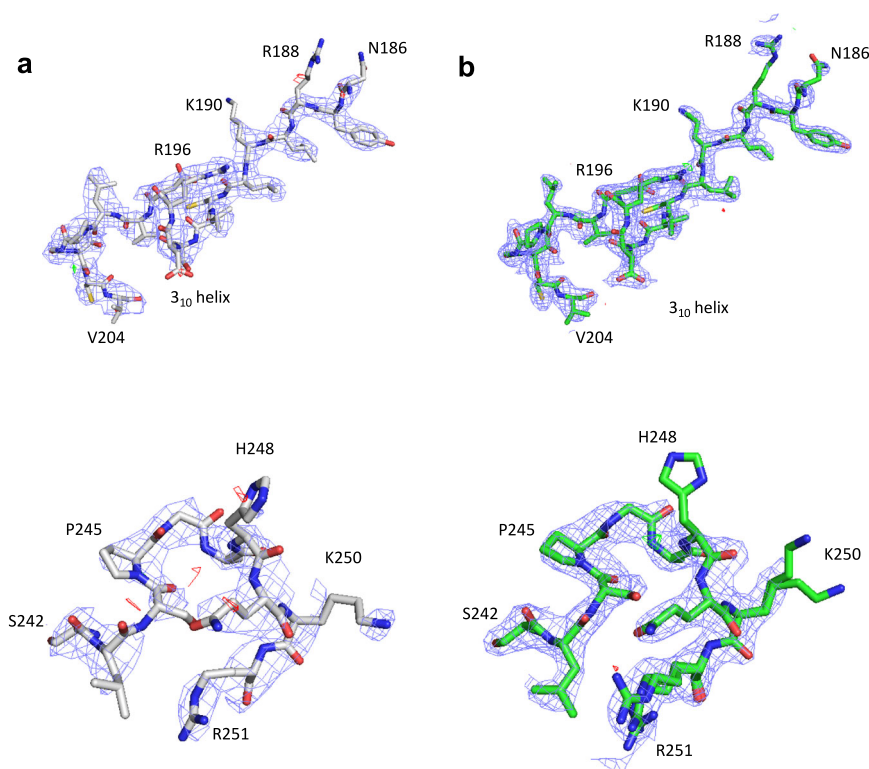


Fig. 4 Structure determination and model building of the MyD88^{TIR} higher-order assembly by MicroED and SFX. Models and maps are presented of the remodelled BB loop (residues 186–204; top) and CD loop (residues 242–251; bottom) for the **a** MicroED and **b** SFX^a structures. The carbon atoms in the MicroED and SFX^a structures are shown in grey and green, respectively. Nitrogen, oxygen and sulfur atoms are shown in blue, red and yellow, respectively. The electrostatic scattering potential (MicroED) and electron density (SFX) 2mFo – DFC maps (blue isomesh) are contoured at 1.2 σ , and the difference mFo – DFC maps (green and red isomesh for positive and negative density, respectively) are contoured at 2.8 σ . No missing reflections were restored using weighted Fc values for map calculations.

MyD88^{TIR} interaction interfaces in the microcrystal. Analysis of the crystal packing reveals MyD88^{TIR} higher-order assemblies, each consisting of two offset parallel strands of TIR domains, with subunits in a head-to-tail arrangement forming each strand (Fig. 5a–c and Supplementary Tables 1–3). Formation of the MyD88^{TIR} assemblies is mediated by two major types of asymmetric TIR domain interactions: one within each of the two strands (intrastrand interface) and one between the two strands (interstrand interface).

Based on the SFX structure, the intrastrand interface involves opposite sides of the MyD88^{TIR} domain, which together buries ~18.0–18.6% (1500 Å²) of the total surface area per subunit in the structure. It is composed of interactions between residues located in the BB loop of one subunit (BB surface) and the β D and β E strands and the α E helix on the next subunit (EE surface) (Fig. 5b–d and Supplementary Table 1). The highly conserved proline residue (P200 in MyD88) in the BB loop is buried in a shallow pocket between the β E strand and the α E helix consisting of residues I253, C274, L290 and A292. Hydrogen bonds (Supplementary Table 1) and a hydrophobic stacking interaction between the side chains of W284 and R196 stabilize the interface. The conformation of the BB loop is also stabilized by an internal salt-bridge between E183 and R196 (Fig. 5d).

The interstrand interface buries ~12.0–12.2% of the total surface area per subunit (991 Å²) and is composed of interactions between residues located on the α B and α C helices of one molecule (BC surface) and the CD loop and the α D helical region of the partner molecule (CD surface) (Fig. 5b, c, e and Supplementary Table 2). Several residues (W205, F235, K238, F239, L241, P245, I267 and F270) contribute hydrophobic interactions to this interface (Fig. 5e).

The interactions between the MyD88^{TIR} two-stranded assemblies, which form a continuous sheet in the microcrystals, involve residues predominantly located in the α A helix and the CD and EE loops (Supplementary Table 3). The interface buries ~7–8% (570 Å²) of the total surface area per subunit and is less extensive than the intrastrand and interstrand interactions (Fig. 5f and Supplementary Table 3). These inter-assembly interactions are most likely analogous to non-biological crystal contacts in macromolecular crystals^{2,66}.

Mutation of MyD88^{TIR} intrastrand and interstrand residues perturbs assembly formation and signaling. We previously showed that alanine mutations of R196, D197, P200, W284 and R288 in the intrastrand interface, and K238, L241, S266 and R269 in the interstrand interface disrupted MAL-induced MyD88^{TIR} microcrystal formation in solution². To demonstrate the biological importance of the interaction interfaces, we tested the effect of interface residue mutations in a HEK293 TLR4 reporter cell line with an nuclear factor- κ B (NF- κ B)-driven mScarlet-I reporter and with endogenous MYD88 knocked out (Fig. 6a, b and Supplementary Fig. 5). Intrastrand mutations R196A, W284A, I253D and R288A abolished NF- κ B activation by the TLR4 ligand lipopolysaccharide (LPS), whereas P200A in the BB loop substantially reduced activation (Fig. 6a and Supplementary Fig. 5d). In the interstrand interface, mutants K238A, L241A, F270A and F270E had little or no LPS response. An alanine mutation of F239 in this interface, which predominantly is involved in hydrophobic interactions with α B helix residues within the same subunit, only led to ~20% loss of activity. Mutants localized at the periphery of the interstrand interface had either intact signaling (P245H and R269A)

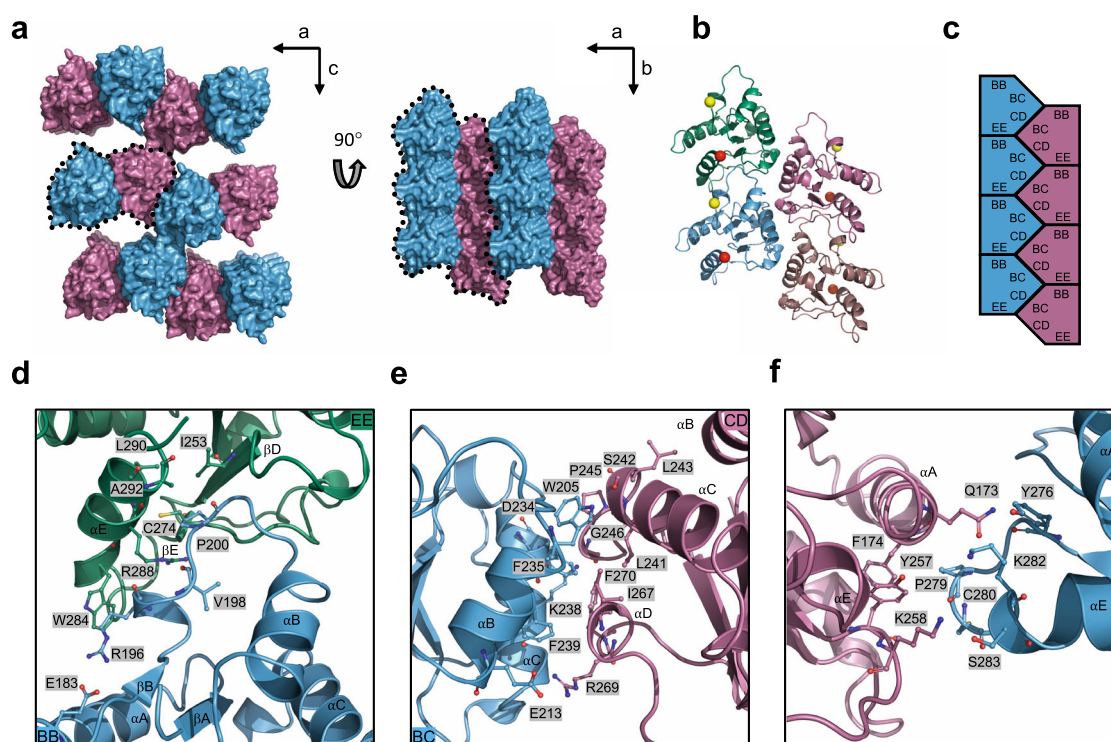


Fig. 5 Structure of the MyD88^{TIR} higher-order assembly microcrystal. **a** Surface representation of the MyD88^{TIR} microcrystal, consisting of two-stranded higher-order assemblies (black dotted lines). The two strands are shown in blue and magenta, respectively. **b** Ribbon diagram of the MyD88^{TIR} higher-order assembly. A yellow sphere indicates the N terminus of each TIR monomer and a red sphere indicates the C terminus of each TIR monomer. The two strands are shown in blue and green, and magenta and dark salmon, respectively. **c** Schematic diagram of the MyD88^{TIR} microcrystals and the two types of asymmetric interactions within the higher-order assembly. BB surface consist of residues in BB loop; EE surface consist of residues in β D and β E loops, and the α E helix; BC surface consist of residues in α B and α C helices; whereas CD surface consist of residues in CD loop and the α D helical region. **d, e** Detailed interactions within the higher-order assembly **d** intrastrand interface and **e** interstrand interface. **f** Detailed interactions between the two-stranded higher-order assemblies, forming the sheet structure.

or ~20% loss in activity (D234A). Mutant K282A, located at the interface forming the sheet structure that is considered not biologically important (Fig. 5f), also had intact signalling. These signaling results agree very closely with analyses of the LPS-induced clustering of expressed MyD88 in cells (Fig. 6b and Supplementary Fig. 5e, f). The results are also consistent with our previous study on spontaneous and MAL-induced MyD88 clustering², except that here, using a cell line deficient in endogenous MyD88, an effect of interstrand mutations can be clearly seen.

Disease-related mutations and post-translational modification sites modulate assembly formation. Several MyD88 TIR domain missense mutations (V204F, S206C, I207T, S209R, S230N, M219T, L252P and T281P) sustain lymphoma cell survival due to constitutive NF- κ B signaling^{67–69}. Mapping of these residues onto the MyD88^{TIR} assembly revealed that the S209R mutation is likely to directly impact interstrand interactions, whereas the T281P mutation may impact intrastrand interactions (Supplementary Fig. 6). L252 is buried and not directly involved in higher-order assembly interactions, but molecular dynamics simulations suggest that this mutation is likely to modulate the conformation of the CD loop⁷⁰, which is critical for interstrand interactions in the MyD88 higher-order assembly. To directly test the hypothesis that these disease-related mutations increase MyD88 higher-order assembly formation, we analysed their effects on clustering in both cell-based and cell-free systems (Fig. 6a–c). Consistent with previous reports, expression of the S209R, L252P and T281P mutants in our reporter cell line showed increased basal NF- κ B activation (Fig. 6a). L252P showed

no further inducibility by LPS, whereas S209R and T281P were LPS responsive. All three mutants had increased basal clustering compared to wild-type (WT) MyD88, which was further increased by LPS for S209R and T281P (Fig. 6b). The aggregation propensity of these mutants was also evaluated by single-molecule spectroscopy, by measuring the brightness of the fluorescence time traces of cell-free expressed green fluorescent protein (GFP)-tagged proteins⁷¹ (Fig. 6c). The S209R and T281P mutants had increased aggregation propensity, forming larger particles than WT MyD88 (Supplementary Fig. 7). By contrast, the L252P mutant formed smaller particles than WT MyD88 (Supplementary Fig. 7), but the complexes were found in higher numbers and formed at lower protein concentrations, as previously reported⁷².

The MyD88 TIR domain has been reported to be phosphorylated on S242 (α C helix) and S244 (CD loop), with phosphomimetic mutations of these residues leading to opposite effects on NF- κ B activation: the S244D mutation becomes hyperactive, whereas the S242D mutation has an inhibitory effect^{70,73}. S242 forms a hydrogen bond with W205 in the MyD88 higher-order assembly and mutation of this residue to an aspartate is thus likely to destabilize the interstrand interface (Supplementary Fig. 6). S244 is not directly involved in higher-order assembly interactions, but similar to L252P, molecular dynamics simulations suggest that the S244D mutation causes a change in the CD loop conformation⁷⁰. When the ability of MyD88 to cluster in HEK293 cells was tested (Fig. 6b), the S244D phosphomimetic mutation increased MyD88^{FL} clustering, whereas S242D inhibited clustering, which is in perfect agreement with NF- κ B activation by these mutants (Fig. 6a). Similar data were

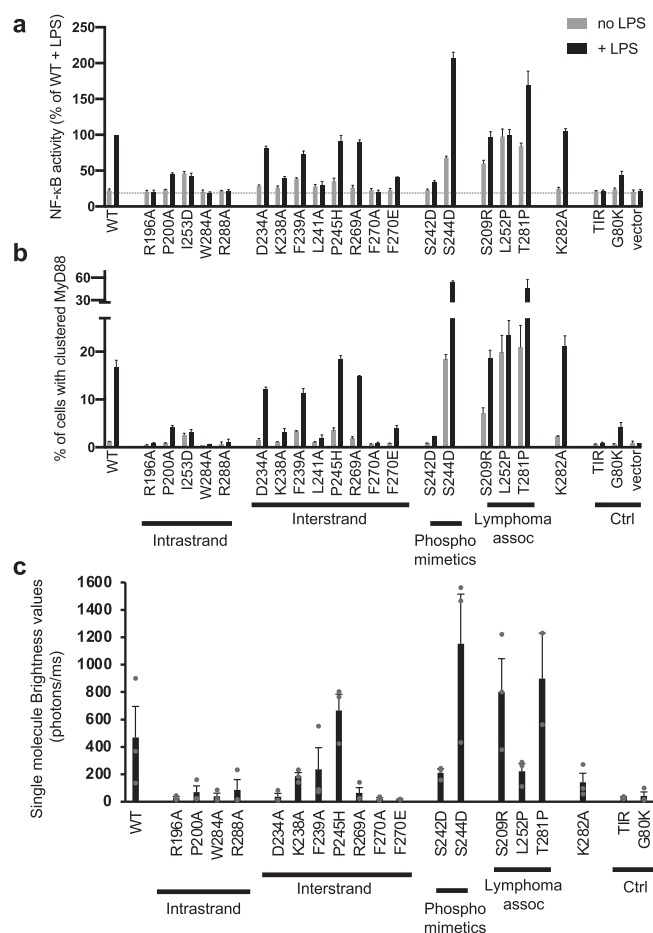


Fig. 6 Interface, disease-associated and phosphomimetic mutations modulate MyD88 signaling and assembly. **a**, **b** Effects of MyD88 mutations on LPS-induced signaling and MyD88 clustering were tested in HEK293 cells expressing TLR4, MD2 and CD14, with MYD88 knocked out and stably transfected with an NF-κB-driven mScarlet-I fluorescent reporter. The cells were transfected with plasmids expressing wild-type or mutant V5-tagged MyD88, or empty vector, and then treated with (black bars) or without (grey bars) LPS (100 ng/mL) overnight, immunostained to detect MyD88-V5 and analysed by flow cytometry. Cells with very low expression of MyD88 were used for analysis to avoid spontaneous signaling (Supplementary Fig. 5b, c). The mean ± range from $n=2$ independent experiments is shown. The death-domain mutation G80K, which has previously been shown to prevent MyD88 clustering¹³⁰, and a TIR domain alone construct provided negative controls. **a** NF-κB activation measured by the geometric mean fluorescence intensity of the mScarlet-positive population relative to LPS-treated cells expressing wild-type MyD88. The dotted line indicates level of activation in cells with empty vector. **b** The percentage of cells with clustered MyD88 was determined based on the elevated height-to-area ratio of the MyD88 signal, which is observed when MyD88 clusters² (Supplementary Fig. 5e). **c** Wild-type MyD88 and mutants were expressed in a cell-free system with an N-terminal GFP tag and the fluorescent samples were analysed by single-molecule spectroscopy on a home-made confocal microscope. To characterize the propensity of wild-type MyD88 and mutants to form higher-order assemblies, the average brightness values (equation (1)) of the proteins were calculated⁷². The results show that S209R, S244D, P245H and T281P mutants have higher propensity than wild-type MyD88 to oligomerize. The mean ± SEM of $n=3$ or $n=2$ (F270E and T281P) experiments using different lysate batches with two technical repeats per experiment is shown. The G80K mutant and the TIR domain were used as negative controls.

observed in the single-molecule assay, using cell-free expressed proteins (Fig. 6c). Overall, our new data strongly suggest that MAL-induced MyD88 TIR-domain clustering directly correlates with the level of NF-κB activation and therefore support the relevance of our structure as a model of MyD88 TIR domain association in vivo.

Comparison of MAL^{TIR} and MyD88^{TIR} assemblies. To gain deeper insights into TIR-domain assembly formation, we compared the MyD88^{TIR} microcrystal structure (Fig. 5) with our previously published cryo-EM structure of the MAL^{TIR} filament². Both assemblies share a common overall architecture with head-to-tail intrastrand interactions mediated by the BB and EE surfaces, and interstrand interactions mediated by the BC and CD surfaces (Supplementary Fig. 8a). The conformations of the αE helix and the EE and CD loops are different in MyD88 compared to MAL (Fig. 3 and Supplementary Fig. 8a), resulting in an increase in the buried surface of both the intrastrand and interstrand MyD88^{TIR} interactions (Supplementary Table 4).

The conformational differences in the αE helix and EE loop also lead to differences in the interface between the two-stranded higher-order assemblies (Supplementary Fig. 8b). In the MyD88^{TIR} microcrystal, these interactions involve the αA helices and the CD and EE loops, whereas in the MAL^{TIR} cryo-EM structure the αA, αC and αD helices and the AA and EE loops contribute to these interactions. The differences in these interactions result in distinct packing of the two-stranded higher-order assemblies, MAL^{TIR} forming a tube consisting of 12 protofilaments, whereas MyD88^{TIR} forms a continuous sheet (Supplementary Fig. 8c).

MAL^{TIR} nucleates MyD88^{TIR} assembly formation unidirectionally. MAL^{TIR} nucleates the assembly of the MyD88^{TIR} microcrystals². The similar architecture observed in the MAL^{TIR} and MyD88^{TIR} higher-order assemblies suggests a molecular-templating mechanism for nucleation and assembly, in which MAL^{TIR} serves as a platform to promote unidirectional assembly of MyD88^{TIR} through intra- and interstrand interactions. To test this hypothesis, we captured MyD88^{TIR} microcrystal growth using differential interference contrast (DIC) and fluorescence microscopy. Either MAL^{TIR} or GFP-MAL^{TIR} fusion proteins acted as nucleators of assembly formation and GFP-MAL^{TIR} nucleates the same type of MyD88^{TIR} microcrystals as MAL^{TIR} (Supplementary Fig. 9). Short MAL^{TIR}-MyD88^{TIR} crystal seeds were washed to remove MAL and then mixed with MyD88^{TIR}. The results revealed that MyD88^{TIR} assembly formation was unidirectional, with a substantial number of seeds observed with growth from one end only (Fig. 7a and Supplementary Movie 1). However, the tendency of MyD88 microcrystals to aggregate also presented a problem here, as the assemblies could also be seen growing in multiple directions from seed aggregates (Supplementary Fig. 10). GFP fluorescence is observed throughout the GFP-MAL^{TIR}-MyD88^{TIR} crystal seeds, suggesting MAL^{TIR} can also incorporate within the MyD88^{TIR} higher-order assembly, which is consistent with our previous report showing that MAL^{TIR} and MyD88^{TIR} can form smaller heterogeneous complex structures when mixed at a 1:1 ratio². As the concentration of GFP-MAL^{TIR} used for preparing the seeds (0.25–2 μM) is significantly lower than the critical concentration for MAL^{TIR} filament formation (30 μM)², and the initial concentration of MyD88^{TIR} is ~50–400× higher than GFP-MAL^{TIR} or MAL^{TIR}, the seeds must predominantly consist of MyD88^{TIR} molecules, with a small fraction of MAL^{TIR} molecules localized at one end and also scattered throughout the seed. Furthermore, the MyD88^{TIR} assemblies continue to grow after removal of GFP-

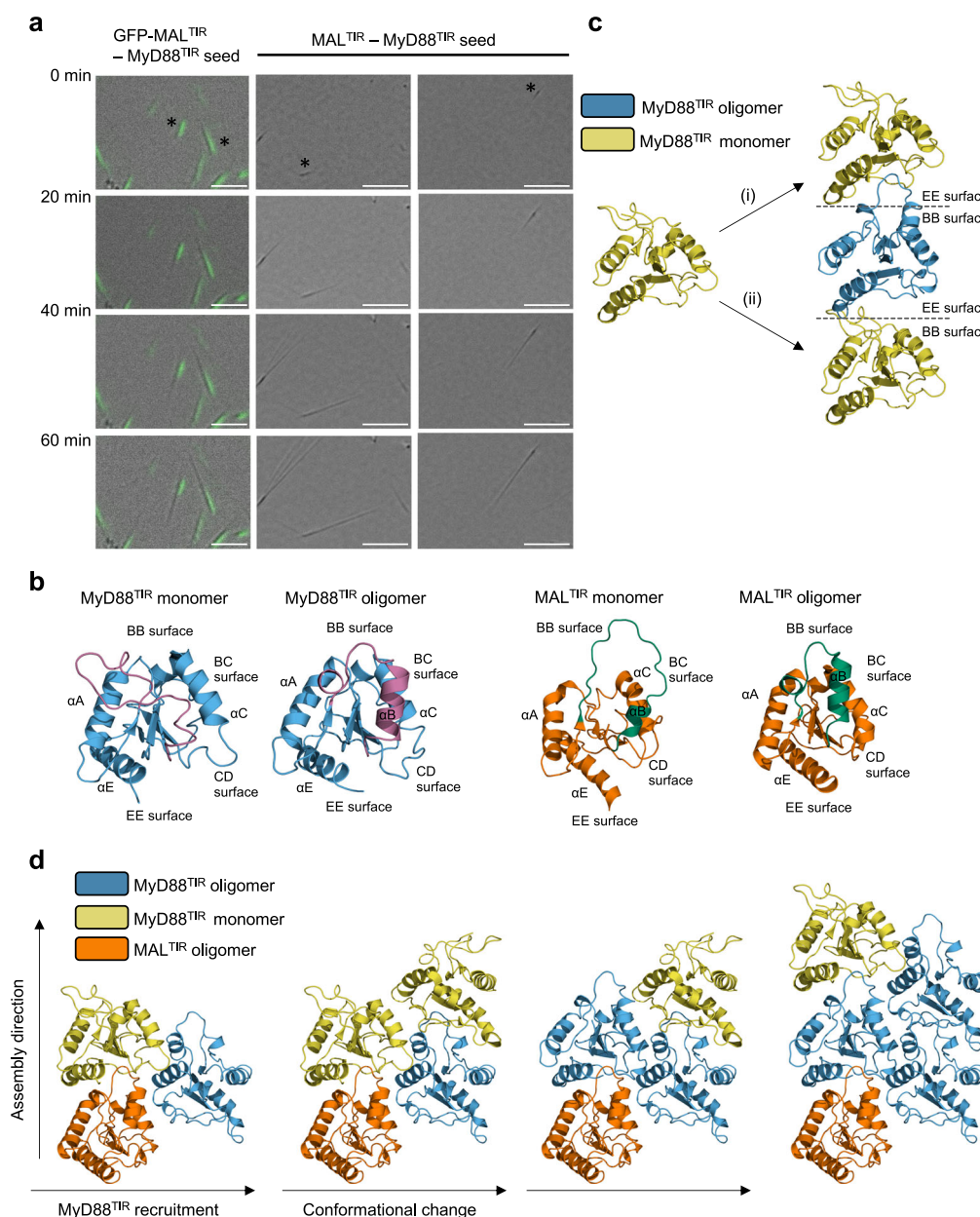


Fig. 7 MAL^{TIR} nucleates MyD88^{TIR} assembly formation unidirectionally. **a** Time-lapse imaging of MyD88^{TIR} microcrystal formation. Representative images of microcrystals growing from single GFP-MAL^{TIR}-MyD88^{TIR} and MAL^{TIR}-MyD88^{TIR} seeds are shown. The seeds were washed to remove MAL and then mixed with MyD88^{TIR}. Data are representative of five independent experiments. Asterisks denote seeds with unidirectional growth. Scale bars: left panel 5 μ m; middle and right panels 10 μ m. **b** Ribbon diagrams of MyD88^{TIR} (NMR solution structure of monomeric MyD88^{TIR} (PDB ID 2Z5V) and higher-order assembly structure) and MAL^{TIR} (NMR solution structure of monomeric MAL^{TIR} (PDB ID 2NDH) and higher-order assembly cryo-EM structure (PDB 5UZZ)), highlighting the rearrangement of the BB loop and α B helix (magenta in MyD88^{TIR} and green in MAL^{TIR}) during the monomer-to-oligomer transition. **c** Two models of interstrand interactions, transitioning between MyD88^{TIR} monomer (yellow) and MyD88^{TIR} higher-order assembly (blue): (i) EE surface of MyD88^{TIR} monomer docks onto BB surface of MyD88^{TIR} higher-order assembly. This interaction does not require any conformational changes in the BB loop and α B helix to occur prior to binding. (ii) BB surface of MyD88^{TIR} monomer docks onto EE surface of MyD88^{TIR} higher-order assembly. This interaction requires significant conformational changes in the BB loop and α B helix to occur prior to binding and is therefore less favoured. **d** Model of MyD88^{TIR} unidirectional assembly formation. The conformational changes in BB loop and α B helix required for the recruitment of new TIR domain subunits are induced by interstrand interactions. The higher-order assembly conformations of MAL^{TIR} and MyD88^{TIR}, and the monomeric conformation of MyD88^{TIR} are shown in orange, blue and yellow, respectively.

MAL^{TIR}, demonstrating that MAL^{TIR} is only required for MyD88^{TIR} assembly nucleation and not elongation.

To predict whether any of the inter- and intrastrand interface surfaces in MAL^{TIR} are preferred for the interaction with MyD88^{TIR}, we calculated the predicted buried surface areas of possible MAL^{TIR} and MyD88^{TIR} interactions. The calculations

showed that the MAL^{TIR} BB surface–MyD88^{TIR} EE surface interaction has the largest buried surface area (Supplementary Table 4). We also mapped the electrostatic potential on the surface of MAL^{TIR} and MyD88^{TIR}, and found that the MAL^{TIR} BB surface and MyD88^{TIR} EE surface are the only interaction interfaces that are highly charge complementary (Supplementary

Fig. 11). Furthermore, molecular dynamics simulations on MAL^{TIR}:MyD88^{TIR} complexes revealed that complexes involving the MAL BB and MyD88 EE surfaces are more stable than complexes involving the MAL EE and MyD88 BB surfaces (Supplementary Fig. 12). Consistent with these analyses, we have previously demonstrated that mutations in the MAL^{TIR} BB surface prevented full-length MAL-induced MyD88 clustering both in vitro and in cells (R121A, P125A and P125H)².

We also compared the structures of MAL^{TIR} and MyD88^{TIR} monomers with their respective structures within higher-order assemblies. This comparison revealed large conformational differences in the BB and BC surface regions (BB loop and α B helix), whereas the EE and CD surface regions adopt similar conformations (Fig. 7b). Models of the recruitment of monomeric MyD88^{TIR} to a growing strand demonstrate that recruitment of new subunits to the assembly via their EE surfaces requires only minimal conformational changes prior to binding, whereas recruitment of new subunits to the assembly via their BB surfaces requires large rearrangements of both the BB loop and α B helix prior to binding, and would therefore be predicted to be less favourable (Fig. 7c). Overall, our structural analyses suggest that in the nucleation and elongation steps of MyD88^{TIR} assembly formation, the EE surface of incoming MyD88^{TIR} molecules dock onto the BB surface of MAL^{TIR} or MyD88^{TIR} subunits. Interstrand interactions via BC and CD surfaces then trigger a rearrangement of the α B helix and BB loop in these newly incorporated TIR domain molecules, enabling them to interact with the EE surface of new incoming MyD88^{TIR} subunits (Fig. 7d).

Discussion

Over the last decade, crystallography has expanded in several different directions, both in terms of electron crystallography, through developments in MicroED^{8,9}, and in terms of X-ray crystallography, through SFX^{45–47}. Here we used MicroED and SFX to determine the structure of the MyD88 TIR domain from hydrated microcrystalline arrays at 3.0 Å and 2.3 Å resolution, respectively. Both of these techniques have their advantages and disadvantages. SFX utilizes high-intensity X-rays to generate high-resolution structures at room temperature and is able to use injector sample delivery systems to overcome crystal aggregation issues at the expense of high sample consumption (typically 0.3–12 mg of protein)^{74–76}. By contrast, MicroED is able to minimize sample consumption (<1 µg protein) allowing for near-complete sampling of reciprocal space using the rotation method of vitrified microcrystals at cryogenic temperature using only a few or even just a single crystal. However, this can come at the expense of often having worse crystallographic quality metrics than is typically achieved in X-ray crystallography. Future advancements in this method, such as serial electron diffraction^{77,78}, improved electron diffraction detectors, and accurate modelling of the electrostatic potential, taking into account the charged state of atoms and the potential distribution, are likely to improve map quality and provide information about charge interactions^{30,79}. For SFX, developments in mix-and-inject experiments at XFELs using nano-focused X-ray beams^{80,81} alongside advancements in data analysis⁸² will provide future opportunities to conduct time-resolved studies of protein assembly formation. The eventual goal of structural biology at XFELs is to try and push the limits of signal-to-noise, to the point where it is possible to image single molecules in solution⁸³.

In our investigation, only subtle differences were observed between the MicroED and SFX structures, which may be explained by the differences in the data resolution and completeness, flexibility of certain regions, and difference between cryogenic and room temperature data collection. To our

knowledge, only one other group has reported a comparison of these two techniques on the same protein crystal system²⁴. Their work showed a slight expansion of the unit cell in the SFX case, which was linked to differences in the data collection temperature. Our room-temperature SFX data also showed a slight increase in lattice parameters along the *a*-axis, when compared to the cryogenic MicroED data (Table 1), indicating the lattice change is related to the temperature difference between the two data sets.

SCAF, which involves assembly of higher-order oligomers for transmission of receptor activation information to cellular responses, is an emerging theme in signal transduction⁴ and operates in several innate-immunity and cell-death pathways including inflammasome signaling⁸⁴, RIG-I-like receptor⁸⁵ and TLR pathways^{2,5}. In this study, we found that the MAL-induced MyD88^{TIR} crystalline assemblies contain a two-stranded head-to-tail arrangement of TIR domains, as previously observed for the TIR domains of the adaptor protein MAL². Analysis of single amino-acid MyD88 mutations for their effect on cellular signaling support the biological relevance of the defined interfaces. Previous functional analyses have measured spontaneous signaling by MyD88 overexpressed in HEK293 cells⁷⁰. Our analysis here has several advantages. First, we used cells with endogenous MYD88 knocked out, which gives a more stringent determination of the function of mutants. This improvement allowed us to demonstrate the importance of residues in the interstrand interactions, which were not apparent in our earlier study². Second, through the use of flow cytometry, we can analyse single-cell responses and select only cells with MyD88 expressed at very low levels to avoid spontaneous signaling. This gives us the ability to observe the response of the mutants to LPS treatment in an intact signaling pathway, avoiding artefacts of overexpression. With this technique, we demonstrated that the R196A mutant is completely inactive, whereas prior work showed it promoted 56% of WT NF- κ B activity in the presence of endogenous MyD88, despite having defective TIR domain interactions⁷⁰. Consequently, we are confident in the biological relevance of the signaling assay reported here, which confirmed the importance of several critical residues in both the intra- and interstrand interfaces of the MyD88^{TIR} assembly.

We provide evidence demonstrating that MAL^{TIR} serves as a platform to promote unidirectional assembly of MyD88^{TIR} oligomers. One feature of unidirectional elongation is establishment of hierarchy in the higher-order oligomers, in which upstream molecules can nucleate the assembly formation of downstream molecules, but not vice versa, and appears to be a common feature in many innate-immunity pathways. For example, elongation of the BCL10 adaptor in the CARMA1–BCL10–MALT1 assembly is unidirectional, with growth at one end only as revealed by confocal imaging⁸⁶, and structures of the RIG-I:MAVS CARD, the FADD:caspase-8 DED and the MyD88:IRAK4:IRAK2 DD assemblies revealed that the RIG-I, FADD and MyD88 oligomers recruit their downstream partners via only one CARD, DD and DED surface, respectively^{5,85,87}.

Our data add support to a sequential and cooperative mechanism for TLR signal transduction, in which receptor and adaptor TIR domains assemble via the inter- and intrastrand interactions observed in the MyD88^{TIR} and MAL^{TIR} higher-order assemblies, leading to formation of a TIR-domain signalosome. This would then promote clustering of MyD88 DDs to form the Myddosome, with recruitment and activation of IRAKs⁵. The Myddosome defined in vitro is a helical array of DD of MyD88–IRAK4–IRAK2 in a 6:4:4 arrangement⁵. In contrast to this mechanism suggesting stepwise recruitment of MyD88 proteins, it has recently been proposed that some MyD88 pre-exists in unstimulated cells in a free oligomeric complex via DD interactions, but cannot recruit IRAK4 due to the TIR domain blocking

access to the IRAK4 binding surface⁶. Upon receptor activation, it is proposed that MyD88 TIR domains are recruited into the TLR4^{TIR}-MAL^{TIR} signaling complex, releasing the autoinhibition and enabling recruitment of IRAKs to the pre-formed MyD88 oligomer. Further data are needed to validate either of these models, but there are a number of caveats regarding the possibility of pre-formed autoinhibited complexes. First, MyD88 DD surfaces involved in IRAK4 interactions are also required for the assembly of MyD88 DDs into a hexamer and binding of MyD88 TIR domains to these surfaces is likely to prevent DD oligomer formation altogether. Second, there is a sharp concentration dependence for oligomerization of both full-length MyD88 and MyD88 DD in vitro⁷² and the threshold for MyD88 clustering in cells is readily exceeded by overexpression. The spontaneous signaling seen with overexpression⁸⁸ argues against MyD88 clusters being intrinsically inhibited for IRAK recruitment. At normal cellular concentrations, autoinhibition is likely to play a role in limiting self-association of MyD88⁷². Stepwise TIR domain-mediated recruitment into a TLR signalosome would then increase the local concentration of DD, leading to Myddosome assembly.

In conclusion, our study provides new insights into the architecture and assembly mechanism of TIR-domain signalosomes in TLR pathways, and at the same time allows for a comparison of the complementary techniques of MicroED and SFX. The detailed TIR:TIR interactions reported in this study may also provide templates for designing small-molecule mimics of the important interfaces to inhibit MyD88 higher-order assembly formation for potential therapeutic applications.

Methods

Protein production. Overlapping PCR was used to generate a construct encoding a GFP-MAL^{TIR} fusion protein (EGFP residues 3–239; MAL residues 79–221) with a GSGGS linker, which was cloned into the pMCSG7 expression vector by ligation-independent cloning⁸⁹. For additional information regarding the primers used, please see Supplementary Table 5. MyD88^{TIR} (residues 155–296 in pET28b, C-terminal His₆-tag)², MAL^{TIR} (residues 79–221 in pMCSG7, N-terminal His₆-tag and c-Myc-tag)² and GFP-MAL^{TIR} were produced in *Escherichia coli* BL21 (DE3) cells, using auto-induction media⁹⁰. Cells were grown at 30–37 °C until the mid-exponential phase (OD_{600nm} of 0.6–0.8) was reached. The temperature was then reduced to 15–20 °C and the cultures were grown for ~16 h before harvesting. The cells were lysed in 50 mM HEPES (pH 7–8), 500 mM NaCl and 1 mM dithiothreitol, using sonication. The resulting supernatant was applied onto a 5 ml HisTrap FF column (GE Healthcare). The bound protein was eluted using a linear gradient of imidazole from 30 to 250 mM and the fractions containing the protein of interest were pooled, concentrated and applied onto a Superdex 75 HiLoad 26/60 gel-filtration column (GE Healthcare) pre-equilibrated with 10 mM HEPES pH 7.5 and 150 mM NaCl. The peak fractions were pooled, concentrated to a final concentration of 1–10 mg/ml and stored in aliquots at –80 °C.

MyD88^{TIR} crystallization. MAL-induced MyD88^{TIR} crystals were produced by incubating MAL^{TIR} (0.5–3 μM) with MyD88^{TIR} domain (60–100 μM) in 10 mM HEPES pH 7.5–8, 150 mM NaCl at 25–37 °C for 60–120 min. GFP-MAL-induced MyD88^{TIR} crystals were produced by incubating GFP-MAL^{TIR} (0.5–3 μM) with MyD88^{TIR} (95 μM) in 10 mM HEPES pH 7.5, 150 mM NaCl at 30 °C for 20–120 min. To produce seeds for crystal growth analysis, the incubation (30 °C) of GFP-MAL^{TIR} (0.5–3 μM) with MyD88^{TIR} (95 μM) in a total volume of 50 μl was stopped after 20 min.

MyD88^{TIR} crystal growth assays. GFP-MAL^{TIR}-MyD88^{TIR} and MAL^{TIR}-MyD88^{TIR} seeds were centrifuged at 2000 × g for 5 min and washed three times with 250 μl 10 mM HEPES pH 7.5, 150 mM NaCl. The seeds were resuspended in 100 μl 10 mM HEPES pH 7.5, 150 mM NaCl and diluted 1:3200 in the same buffer. Five microlitres of diluted seed was added to the well of an imaging plate (μ-Plate 96 well ibiTreat sterile, Ibidi) with 45 μl MyD88^{TIR} (95 μM). The plate was centrifuged at 1500 × g for 5 min and immediately transferred to microscope for imaging. During imaging, the plate was incubated at 30 °C on the Nikon Eclipse Ti2 inverted microscope. DIC and GFP fluorescence images were taken using the ×40 objective lens with ×1.5 magnification.

MicroED sample preparation and data acquisition. The MyD88^{TIR} crystal samples were prepared by depositing 3 μl of 1:50 MAL:MyD88 microcrystal solution on a Quantifoil 3.5/1.0 (300 mesh) Cu holey carbon EM grid. Excess liquid

was blotted away and the sample was vitrified by flash-cooling in liquid ethane, using a FEI Vitrobot Mark IV (blot force 0, blotting time 6 s). The sample was transferred to a Gatan 914 high-tilt cryo-transfer holder. MicroED data were collected on a JEOL JEM-2100 (LaB6 filament) TEM operated at 200 kV equipped with a Tamepix hybrid pixel detector (Amsterdam Scientific Instruments). Screening and MicroED data collection, using the rotation method, were performed via the Instamatic software interface⁹¹. Diffraction data were collected under parallel beam conditions from an area of ~1.5 μm diameter, defined by a selected area aperture. The sample-to-detector distance was 1830 mm. Data were collected with an exposure time of 1.5–2.0 s and an angular increment of 0.68–0.92° per frame. The electron dose rate applied during data collection was ~0.08 e[−]/Å²/s. The average tilt range covered per individual crystal was about 30°, corresponding to a total exposure dose of ~5.5 e[−]/Å².

MicroED data processing and structure determination. Data of 18 crystals were integrated, scaled and merged using XDS⁹² and AIMLESS⁹³. Data were truncated approximately at the average $I/\sigma(I) \geq 1.5$ and $CC_{1/2} \geq 0.494$ (Table 1). A distantly related TIR domain homologue, TRR2 from *Hydra vulgaris* (PDB ID 4W8G), was identified as a suitable search model using the automated molecular replacement pipeline MrBUMP⁹⁵. An optimized search model was generated using Sculptor⁹⁶. The structure of MyD88^{TIR} was subsequently solved using molecular replacement in Phaser⁹⁷ in the PHENIX software suite⁹⁸. The model was iteratively built and refined using Coot⁹⁹, phenix.refine¹⁰⁰ and interactive structure optimization using molecular dynamics in ISOLDE¹⁰¹. The model was refined using a 5% test set for R_{free} , individual isotropic B -factors, electron scattering factors and automated optimization of the data vs. stereochemistry and data vs. ADP (atomic displacement parameters) weighting. The geometry of the structural models was validated using MolProbity¹⁰². A SA composite omit map was calculated over the entire contents of the unit cell using phenix.composite_omit_map¹⁰⁰, sequentially omitting 5% fractions of the structure. No missing reflections were filled in for map calculations.

Serial crystallography, PETRAIII synchrotron. Based on the Roedig et al.¹⁰³ design, 2.5 μl of microcrystals in crystallization buffer containing 16% glycerol (3.3×10^8 – 1×10^9 crystals/ml) were deposited on a chip with a pore size of 1 μm (manufactured by Sauna P/L) under a humidified environment. The excess mother liquor was filtered from the crystal by drawing off the solution on the underside of the chip, leaving behind a thin layer of crystals. The chip was then immediately flash frozen in liquid nitrogen and mounted onto the standard goniometer system, under a cryo stream, on the P11 beamline at the PETRAIII synchrotron. The beamline was set up at 12 keV with a beam size of 2 μm and flux measured at 6.4×10^{11} p/s. A 100 μm pinhole was used and a capillary beam stop designed by the beamline scientists was incorporated into the beamline. The detector distance was set to 588.3 mm and data were collected using the fly scan mode (exposure of 2 s, step size of 10 μm, oscillation of 0.1° and 2 frames per crystal) on a Pilatus 6 M detector.

SFX sample preparation and data acquisition. The crystal concentration tested ranged from 2.5×10^8 to 2×10^9 crystals/ml for SFX measurements with the optimal crystal concentration of 7.5×10^8 crystals/ml. The crystals were filtered through a 20 μm stainless steel filter prior to loading into the sample reservoir for sample injection. SFX data were collected at the coherent X-ray Imaging (CXI) endstation at the LCLS, SLAC National Accelerator Laboratory⁶⁵. A GDVN was used, with an optimized flow rate of 20 μl/min. The XFEL operated at a rate of 120 Hz, delivering 9.6 keV (1.3 Å), X-ray pulses of ~40 fs duration with an estimated 8.3×10^{11} photons/pulse at the interaction position, assuming ~50% intensity loss along the beamline. The beam was focused to a diameter of 1 μm FWHM. The data were collected on a Cornell-SLAC Pixel Array Detector^{104,105} at a distance of 0.181 and 0.1061 m, for ~111 and 32 min (796,710 and 233,158 detector frames), respectively. The GDVN overcame most of the multi-crystal issues but clogging in the injector lines and in-line filters was an issue. The in-line Peek filters were replaced with 20 μm stainless steel filters and the sample reservoirs were vortexed every 15 min to prevent the crystals from settling. The highest hit rates (~4.5%) were achieved by cycling between delivering sample and washing the sample delivery lines with water every 10 min during data collection. Data were collected from a total of 3.2 ml of crystal solution (~4.3 mg of MyD88^{TIR} mixed with 0.17 mg MAL^{TIR}) at room temperature.

SFX data processing and structure determination. Hit finding and detector calibration was performed using Cheetah¹⁰⁶ with hit finder 8 and a minimum of 15 peaks per image, and with minimal jet masking used. The CrystFEL software suite¹⁰⁷ was then used for indexing, utilizing MOSFLM¹⁰⁸, XGANDALF¹⁰⁹ and DirAx¹¹⁰ as indexing algorithms and merged with a partialator (using scaling without partiality modelling), followed by data reduction using AIMLESS⁹³ in the CCP4 software suite¹¹¹.

The SFX MyD88^{TIR} structure was solved, rebuilt and refined using two different protocols, SFX^a and SFX^b. SFX^a: MyD88^{TIR} structure was solved by molecular replacement using Phaser⁹⁸ and a polyalanine model of the MicroED structure as a template. The structure was iteratively rebuilt and refined using Coot⁹⁹ and REFMAC5¹¹² within the CCP4 suite¹¹¹. A 10% R_{free} test set was used for

refinement. The model was first refined using individual isotropic *B*-factors; however, in the final steps of refinement, hydrogens were added to the model and TLS parameters were included to model anisotropic displacements. The geometry of the structural model was validated using MolProbity¹⁰². SFX^b. The structure was solved, rebuilt, refined and validated using an identical protocol as described for the MicroED data. A SA composite omit map was calculated using the same protocol as described for the MicroED data.

Structural analyses. The Dali¹¹³, PISA¹¹⁴ and PIC¹¹⁵ servers and PyMOL (version 2.2.3 Schrödinger, LLC) were used to analyse the structures. Electrostatic potentials were calculated using APBS¹¹⁶. Figures were prepared using PyMOL.

Plasmids and site-directed mutagenesis. The cDNA encoding luciferase in the NF- κ B-driven reporter plasmid (pNF κ B-Luc, Stratagene) was replaced with that of the fluorescent protein mScarlet-1 (Supplementary Table 5). The resulting plasmid (pNF κ B-mScarlet) drives expression of the fluorescent protein, mScarlet-I, upon NF- κ B activation. Single point mutations of MyD88 were produced by Genscript in a pEF6-MyD88-V5-His₆ plasmid encoding residues 1–296 of human MyD88².

Cell lines and cell culture. HEK-Blue human TLR4 cells (InvivoGen) were stably transfected with the reporter plasmid pNF κ B-mScarlet and a single-cell clone was obtained (HEK-Blue-TLR4-NF κ B-mScarlet cell line). The MYD88 gene was knocked out in this cell line using the CRISPR-Cas9 system and a single-cell clone that did not show any detectable MyD88 expression or LPS response was obtained (HEK-Blue-TLR4-NF κ B-mScarlet-MyD88 knockout (KO) cell line). All cells were maintained in Dulbecco's modified Eagle medium with 4.5 g/l glucose, 110 mg/l sodium pyruvate supplemented with Glutamax-1, 10% heat-inactivated fetal bovine serum, 50 U/ml penicillin and 50 μ g/ml streptomycin (reagents from Life Technologies). All cells were tested and shown to be mycoplasma-free.

Evaluation of the effects of MyD88 mutations on higher-order assembly and TLR4 signaling in HEK-Blue-TLR4-NF- κ B-mScarlet-MyD88 KO cells by flow cytometry. To assess the ability of mutant MyD88 to form a higher-order assembly and to restore TLR4 signaling in the HEK-Blue-TLR4-NF- κ B-mScarlet-MyD88 KO cell line, 400,000 cells were plated in antibiotic-free media in a 12-well plate and transfected 3–4 h later with 200 ng plasmids expressing WT or mutant MyD88 or empty vector alone, using Lipofectamine 2000 (Thermo Fisher Scientific) according to manufacturer's instructions². After ~16 h, transfection media were replaced with medium with 5% serum and 6–8 h later the cells were treated with or without ultrapure *E. coli* LPS (100 ng/mL; Invivogen) for ~16 h. Cells were collected and fixed for 30 min with 4% paraformaldehyde and immunostained overnight with anti-V5 rabbit monoclonal (D3H8Q) antibody (Cell Signaling Technology) at a 1 : 2000 dilution, followed by goat anti-rabbit-Alexa Fluor-488 (Life Technologies) at a 1 : 10,000 dilution for 1 h². The stained cells were run on a BD Cytotex S flow cytometer and the data were analysed using the FlowJo software. Cells were first gated to exclude debris on a side scatter vs. forward scatter (FSC) plot and then gated to select single cells on a FSC-width vs. FSC-area plot (Supplementary Fig. 5a). A plot of MyD88-V5 signal vs. mScarlet-I reporter signal showed that the reporter was activated upon TLR4 stimulation with LPS, in cells with MyD88 levels that were below the detection threshold of MyD88-V5 (Supplementary Fig. 5b), and higher levels of expression led to progressively more spontaneous signaling. Thus, cells below the detection threshold of MyD88-V5 were assessed for reporter activation and for the ability of MyD88 to cluster into a higher-order assembly. Reporter activation is expressed as the mScarlet-I mean fluorescence intensity in the mScarlet-I-positive cells or as the percentage of mScarlet-I-positive cells. MyD88 clustering was determined from a plot of MyD88-V5 signal peak height vs. area² (Supplementary Fig. 5e). The clustering assay is based on the fact that the signal from detection of a clustered protein, with a fluorescent antibody, results in a fluorescent pulse with increased peak height and decreased width compared to the signal from cells expressing diffuse protein².

Preparation of cell-free extracts. The *Leishmania tarentolae* Parrot strain was obtained as LEXSY host P10 from Jena Bioscience GmbH, Jena, Germany, and cultured in TBGG media (12 g/L tryptone, 24 g/L yeast extract, 0.8% glycerol, 5.55 mM glucose, 17 mM KH₂PO₄, 72 mM K₂HPO₄) containing 0.2% v/v penicillin/streptomycin (Life Technologies) and 0.05% w/v hemin (MP Biomedicals)¹¹⁷. Cells were collected by centrifugation at 2500 \times g, washed twice by resuspension in 45 mM HEPES pH 7.6, containing 250 mM sucrose, 100 mM potassium acetate and 3 mM magnesium acetate, and resuspended to 0.25 g cells/g suspension. Cells were placed in a cell disruption vessel (Parr Instruments, USA) and incubated under 7000 kPa nitrogen for 45 min, then lysed by a rapid release of pressure. The lysate was clarified by sequential centrifugation at 10,000 \times g and 30,000 \times g and anti-splice leader oligonucleotide was added to 10 mM. The lysate was then desalted into 45 mM HEPES pH 7.6, containing 100 mM potassium acetate and 3 mM magnesium acetate, and snap-frozen until required.

Protein expression using cell-free extracts. The MyD88 mutants produced by Genscript in the pEF6-MyD88-V5-His₆ vector were Gateway™ cloned into the

pCellFree G03 vector to produce N-terminally GFP-tagged proteins (Supplementary Table 5)¹¹⁸. Cell-free lysates from three different preparations were supplemented with a feeding solution containing nucleotides, amino acids, T7 polymerase, HEPES buffer and a creatine/creatine kinase ATP regeneration system at a ratio of lysate to feed solution of 0.21 and a final Mg²⁺ concentration of 6 mM. Purified plasmid DNA, at a concentration between 100 and 400 ng/mL, was added to the expression reaction at a ratio of 1 : 9 (v/v) and the reaction allowed to proceed for 3 h at 27 °C. Fluorescently tagged expressed proteins were detected by SDS-polyacrylamide gel electrophoresis using a Chemidoc MP imaging system (Bio-Rad, Laboratories Pty Ltd, Gladesville, NSW, Australia). Gels were imaged without further processing, using the inbuilt Alexa 488 (GFP), Alexa 546 (mCherry) and Cy5 (prestained markers) settings to verify expression¹¹⁷.

Single-molecule spectroscopy and brightness analysis. The expressing lysates were diluted 1 in 10 in a buffer containing 50 mM HEPES pH 7.5 and 150 mM NaCl directly in a custom-made 192-well silicone plate. Samples were analysed at room temperature on a Zeiss Axio Observer microscope equipped with a \times 40/1.2 NA water-immersion objective (Zeiss C-Apochromat), used to focus a 488 nm laser and collect fluorescence. The fluorescence signal was collected in 1 ms time bins and filtered by a 565 nm dichroic mirror and a 525/50 nm band pass filter optimized for GFP detection¹¹⁹. For brightness analysis, the average intensity (μ) and SD (σ) of the signal were calculated for each 100 s time trace and brightness (B) was calculated as $B = \sigma^2/\mu$ ¹²⁰ (1). For determining the number of large polymers, all expressing lysates were diluted to the same average fluorescence (1000 photons/ms) and raw fluorescence traces were collected as described above, then analysed for the frequency of events of given size. A threshold of 4000 photons/ms was used to discriminate large and small MyD88 assemblies.

Molecular dynamics simulations. All molecular dynamics simulations were performed with the MicroED MyD88^{TIR} structure and using the GPU version Gromacs 2019.3¹²¹ on the Gadi cluster at the National Computing Infrastructure, Australia. The Gromos 54A7^{122,123} force field was used to model the proteins. Each complex was placed in a truncated octahedron periodic box with a 1.4 nm distance between the protein surface and the edge of the box wall. The protonation state of titratable groups was chosen appropriate to pH 7.0. Each system was simulated under periodic boundary conditions in a rectangular box. The pressure was maintained at 1 bar, by weakly coupling the system to a semi-isotropic pressure bath, using an isothermal compressibility of 4.6×10^{-5} bar⁻¹ and a coupling constant $\tau_P = 1$ ps¹²⁴. The temperature of the system was maintained at 298 K by independently coupling the protein–ligand complex, lipids and water to an external temperature bath with a coupling constant $\tau_T = 0.1$ ps, using a Berendsen thermostat¹²⁴. All bond lengths were constrained using the LINCS algorithm¹²⁵. The Simple-Point Charge¹²⁶ water model was used and constrained using the SETTLE algorithm¹²⁷. Each system was energy-minimized for 1000 steps, using the steepest descent method, followed by a position-restrained MD simulation, where all heavy atoms of protein were restrained to their original position using 1000 kJ/mol/nm², allowing water molecules to equilibrate. The restraints were removed and the whole system was allowed to equilibrate for 5 ns. The MD simulations were performed for 100 ns in duplicate, starting with different initial velocity distribution for each system. All coordinates, velocities, forces and energies were saved every 10,000 steps for analysis. The stability of the protein and protein–ligand complexes was evaluated, by measuring the RMSD of protein backbone atoms by fitting the backbone atoms of protein and comparing the initial and final structures.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data supporting the findings of this manuscript are available from the corresponding authors upon reasonable request. A reporting summary for this article is available as a Supplementary Information file. The atomic coordinates and structure factors of the MicroED and high-resolution SFX models (SFX^a and SFX^b) have been deposited in the Protein Data Bank under accession codes 7BEQ, 7L6W and 7BER, respectively. Raw MicroED data are available from the SGrid Data Bank (doi:10.15785/SBGRID/814). SFX data are available at CXIB.org (<https://doi.org/10.11577/1767965>). Source data are provided with this paper.

Received: 24 December 2020; Accepted: 18 March 2021;

Published online: 10 May 2021

References

1. Ve, T., Gay, N. J., Mansell, A., Kobe, B. & Kellie, S. Adaptors in toll-like receptor signaling and their potential as therapeutic targets. *Curr. Drug Targets* **13**, 1360–1374 (2012).

2. Ve, T. et al. Structural basis of TIR-domain-assembly formation in MAL- and MyD88-dependent TLR4 signaling. *Nat. Struct. Mol. Biol.* **24**, 743–751 (2017).
3. Nanson, J. D., Kobe, B. & Ve, T. Death, TIR, and RHIM: self-assembling domains involved in innate immunity and cell-death signaling. *J. Leukoc. Biol.* **105**, 363–375 (2019).
4. Vajjhala, P. R., Ve, T., Bentham, A., Stacey, K. J. & Kobe, B. The molecular mechanisms of signaling by cooperative assembly formation in innate immunity pathways. *Mol. Immunol.* **86**, 23–37 (2017).
5. Lin, S. C., Lo, Y. C. & Wu, H. Helical assembly in the MyD88-IRAK4-IRAK2 complex in TLR/IL-1R signalling. *Nature* **465**, 885–890 (2010).
6. Moncrieffe, M. C. et al. MyD88 death-domain oligomerization determines Myddosome architecture: implications for Toll-like receptor signaling. *Structure* **28**, 281–289.e283 (2020).
7. Latty, S. et al. Activation of Toll-like receptors nucleates assembly of the MyDDosome signaling hub. *eLife* **7**, 1–15 (2018).
8. Nannenga, B. L. & Gonen, T. The cryo-EM method microcrystal electron diffraction (MicroED). *Nat. Methods* **16**, 369–379 (2019).
9. Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. Three-dimensional electron crystallography of protein microcrystals. *eLife* **2**, e01345 (2013).
10. Nannenga, B. L., Shi, D., Leslie, A. G. W. & Gonen, T. High-resolution structure determination by continuous-rotation data collection in MicroED. *Nat. Methods* **11**, 927–930 (2014).
11. Nederlof, I., van Genderen, E., Li, Y. W. & Abrahams, J. P. A Medipix quantum area detector allows rotation electron diffraction data collection from submicrometre three-dimensional protein crystals. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1223–1230 (2013).
12. Shi, D. et al. The collection of MicroED data for macromolecular crystallography. *Nat. Protoc.* **11**, 895–904 (2016).
13. Arndt, U. W. & Wonacott, A. J. *The Rotation Method in Crystallography* (Noth-Holland, 1977).
14. Gemmi, M. et al. 3D electron diffraction: the nanocrystallography revolution. *ACS Cent. Sci.* **5**, 1315–1329 (2019).
15. Holton, J. M. & Frankel, K. A. The minimum crystal size needed for a complete diffraction data set. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 393–408 (2010).
16. Luft, J. R., Wolfley, J. R. & Snell, E. H. What's in a drop? Correlating observations and outcomes to guide macromolecular crystallization experiments. *Cryst. Growth Des.* **11**, 651–663 (2011).
17. Nave, C. & Hill, M. A. Will reduced radiation damage occur with very small crystals? *J. Synchrotron Rad.* **12**, 299–303 (2005).
18. Sanishvili, R. et al. Radiation damage in protein crystals is reduced with a micron-sized X-ray beam. *Proc. Natl. Acad. Sci. USA* **108**, 6127–6132 (2011).
19. Calero, G., Cohen, A. E., Luft, J. R., Newman, J. & Snell, E. H. Identifying, studying and making good use of macromolecular crystals. *Acta Crystallogr. F Struct. Biol. Commun.* **70**, 993–1008 (2014).
20. Stevenson, H. P. et al. Transmission electron microscopy for the evaluation and optimization of crystal growth. *Acta Crystallogr. D Biol. Crystallogr.* **72**, 603–615 (2016).
21. Stevenson, H. P. et al. Use of transmission electron microscopy to identify nanocrystals of challenging protein targets. *Proc. Natl. Acad. Sci. USA* **111**, 8470–8475 (2014).
22. Cusack, S. et al. Small is beautiful: protein micro-crystallography. *Nat. Struct. Mol. Biol.* **5**, 634–637 (1998).
23. de la Cruz, M. J. et al. Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED. *Nat. Methods* **14**, 399–402 (2017).
24. Wolff, A. M. et al. Comparing serial X-ray crystallography and microcrystal electron diffraction (MicroED) as methods for routine structure determination from small macromolecular crystals. *IUCr* **7**, 306–323 (2020).
25. Clabbers, M. T. B., Fisher, S. Z., Coinçon, M., Zou, X. & Xu, H. Visualizing drug binding interactions using microcrystal electron diffraction. *Commun. Biol.* **3**, 1–8 (2020).
26. Martynowycz, M. W. & Gonen, T. Ligand incorporation into protein microcrystals for MicroED by on-grid soaking. *Structure* **29**, 88–85 (2020).
27. Clabbers, M. T. B. et al. Protein structure determination by electron diffraction using a single three-dimensional nanocrystal. *Acta Crystallogr. D Biol. Crystallogr.* **73**, 738–748 (2017).
28. Liu, S. & Gonen, T. MicroED structure of the NaK ion channel reveals a Na⁺-partition process into the selectivity filter. *Commun. Biol.* **1**, 38 (2018).
29. Nannenga, B. L., Shi, D., Hattne, J., Reyes, F. E. & Gonen, T. Structure of catalase determined by MicroED. *eLife* **3**, e03600 (2014).
30. Yonekura, K., Kato, K., Ogasawara, M., Tomita, M. & Toyoshima, C. Electron crystallography of ultrathin 3D protein crystals: atomic model with charges. *Proc. Natl. Acad. Sci. USA* **112**, 3368–3373 (2015).
31. Xu, H. et al. Solving a new R2lox protein structure by microcrystal electron diffraction. *Sci. Adv.* **5**, eaax4621 (2019).
32. Martynowycz, M. W., Khan, F., Hattne, J., Abramson, J. & Gonen, T. MicroED structure of lipid-embedded mammalian mitochondrial voltage-dependent anion channel. *Proc. Natl. Acad. Sci. USA* **117**, 32380 (2020).
33. Martynowycz, M. W. et al. MicroED structure of the human adenosine receptor determined from a single nanocrystal in LCP. Preprint at <https://doi.org/10.1101/2020.1109.1127.316109> (2020).
34. Polovinkin, V. et al. Demonstration of electron diffraction from membrane protein crystals grown in a lipidic mesophase after lamella preparation by focused ion beam milling at cryogenic temperatures. *J. Appl. Crystallogr.* **53**, 1416–1424 (2020).
35. Zhu, L. et al. Structure determination from lipidic cubic phase embedded microcrystals by MicroED. *Structure* **28**, 1149–1159 (2020).
36. Purdy, M. D. et al. MicroED structures of HIV-1 Gag CTD-SP1 reveal binding interactions with the maturation inhibitor bevirimat. *Proc. Natl. Acad. Sci. USA* **115**, 13258–13263 (2018).
37. Gallagher-Jones, M. et al. Sub-ångström cryo-EM structure of a prion protofibril reveals a polar clasp. *Nat. Struct. Mol. Biol.* **25**, 131–134 (2018).
38. Rodriguez, J. A. et al. Structure of the toxic core of α -synuclein from invisible crystals. *Nature* **525**, 486–490 (2015).
39. Sawaya, M. R. et al. Ab initio structure determination from prion nanocrystals at atomic resolution by MicroED. *Proc. Natl. Acad. Sci. USA* **133**, 11232–11236 (2016).
40. James, D. et al. Improving high viscosity extrusion of microcrystals for time-resolved serial femtosecond crystallography at X-ray lasers. *J. Vis. Exp.* 10.3791/59087 (2019).
41. Pandey, S. et al. Time-resolved serial femtosecond crystallography at the European XFEL. *Nat. Methods* **17**, 73–78 (2020).
42. Pandey, S., Poudyal, I. & Malla, T. N. Pump-probe time-resolved serial femtosecond crystallography at X-ray free electron lasers. *Crystals* **10**, 628 (2020).
43. Zatsel, N. A., Li, C., Colasur, P. & Nannenga, B. L. The complementarity of serial femtosecond crystallography and MicroED for structure determination from microcrystals. *Curr. Opin. Struct. Biol.* **58**, 286–293 (2019).
44. Boutet, S. et al. High-resolution protein structure determination by serial femtosecond crystallography. *Science* **337**, 362–364 (2012).
45. Chapman, H. N. et al. Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–77 (2011).
46. Schlichting, I. Serial femtosecond crystallography: the first five years. *IUCr* **2**, 246–255 (2015).
47. Spence, J. C. H. XFELs for structure and dynamics in biology. *IUCr* **4**, 322–339 (2017).
48. Fraser, J. S. et al. Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. USA* **108**, 16247–16252 (2011).
49. Ishigami, I. et al. Snapshot of an oxygen intermediate in the catalytic reaction of cytochrome c oxidase. *Proc. Natl. Acad. Sci. USA* **116**, 3572–3577 (2019).
50. Kern, J. et al. Structures of the intermediates of Kok's photosynthetic water oxidation clock. *Nature* **563**, 421–425 (2018).
51. Suga, M. et al. An oxyl/oxo mechanism for oxygen-oxygen coupling in PSII revealed by an x-ray free-electron laser. *Science* **366**, 334–338 (2019).
52. Ishchenko, A., Cherezov, V. & Liu, W. Preparation and delivery of protein microcrystals in lipidic cubic phase for serial femtosecond crystallography. *J. Vis. Exp.*, e54463 (2016).
53. Liu, W., Ishchenko, A. & Cherezov, V. Preparation of microcrystals in lipidic cubic phase for serial femtosecond crystallography. *Nat. Protoc.* **9**, 2123–2134 (2014).
54. Nango, E. et al. A three-dimensional movie of structural changes in bacteriorhodopsin. *Science* **354**, 1552–1557 (2016).
55. Nogly, P. et al. Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser. *Science* **361**, eaat0094 (2018).
56. Pande, K. et al. Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein. *Science* **352**, 725–729 (2016).
57. Tenboer, J. et al. Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science* **346**, 1242–1246 (2014).
58. Batyuk, A. et al. Native phasing of X-ray free-electron laser data for a G protein-coupled receptor. *Sci. Adv.* **2**, e1600292 (2016).
59. Branden, G. et al. Coherent diffractive imaging of microtubules using an X-ray laser. *Nat. Commun.* **10**, 2589 (2019).
60. Popp, D. et al. Flow-aligned, single-shot fiber diffraction using a femtosecond X-ray free-electron laser. *Cytoskeleton (Hoboken)* **74**, 472–481 (2017).
61. Ohnishi, H. et al. Structural basis for the multiple interactions of the MyD88 TIR domain in TLR4 signaling. *Proc. Natl. Acad. Sci. USA* **106**, 10260–10265 (2009).
62. Snyder, G. A. et al. Molecular mechanisms for the subversion of MyD88 signaling by TcpC from virulent uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **110**, 6985–6990 (2013).
63. DePonte, D. P. et al. Gas dynamic virtual nozzle for generation of microscopic droplet streams. *J. Phys. D Appl. Phys.* **41**, 195505 (2008).
64. Weierstall, U., Spence, J. C. H. & Doak, R. B. Injector for scattering measurements on fully solvated biospecies. *Rev. Sci. Instrum.* **83**, 35108 (2012).

65. Liang, M. et al. The coherent X-ray imaging instrument at the Linac Coherent Light Source. *J. Synchrotron Radiat.* **22**, 514–519 (2015).
66. Kobe, B. et al. Crystallography and protein-protein interactions: biological interfaces and crystal contacts. *Biochem. Soc. Trans.* **36**, 1438–1441 (2008).
67. Avbelj, M. et al. Activation of lymphoma-associated MyD88 mutations via allosterically-induced TIR-domain oligomerization. *Blood* **124**, 3896–3904 (2014).
68. Ngo, V. N. et al. Oncogenically active MYD88 mutations in human lymphoma. *Nature* **470**, 115–121 (2011).
69. Yu, X. et al. MYD88 L265P mutation in lymphoid malignancies. *Cancer Res.* **78**, 2457 LP–2452462 (2018).
70. Vyncke, L. et al. Reconstructing the TIR side of the Myddosome: a paradigm for TIR-TIR interactions. *Structure* **24**, 437–447 (2016).
71. Sieracki, E. et al. Nanomolar oligomerization and selective co-aggregation of alpha-synuclein pathogenic mutants revealed by single-molecule fluorescence. *Sci. Rep.* **6**, 37630 (2016).
72. O'Carroll, A. et al. Pathological mutations differentially affect the self-assembly and polymerisation of the innate immune system signalling adaptor molecule MyD88. *BMC Biol.* **16**, 1–16 (2018).
73. Xie, L. et al. Protein phosphatase 2A catalytic subunit α plays a MyD88-dependent, central role in the gene-specific regulation of endotoxin tolerance. *Cell Rep.* **3**, 678–688 (2013).
74. Mills, G., Bean, R. & Mancuso, A. P. First experiments in structural biology at the european X-ray free-electron laser. *Appl. Sci.* **10**, 1–19 (2020).
75. Wiedorn, M. O. et al. Megahertz serial crystallography. *Nat. Commun.* **9**, 4025 (2018).
76. Zhao, F. Z. et al. A guide to sample delivery systems for serial crystallography. *FEBS J.* **286**, 4402–4417 (2019).
77. Bücker, R. et al. Serial protein crystallography in an electron microscope. *Nat. Commun.* **11**, 996 (2020).
78. Smeets, S., Zou, X. & Wan, W. Serial electron crystallography for structure determination and phase analysis of nanocrystalline materials. *J. Appl. Crystallogr.* **51**, 1262–1273 (2018).
79. Yonekura, K. & Maki-Yonekura, S. Refinement of cryo-EM structures using scattering factors of charged atoms. *J. Appl. Crystallogr.* **49**, 1517–1523 (2016).
80. Hejazian, M., Darmanin, C., Balaur, E. & Abbey, B. Mixing and jetting analysis using continuous flow microfluidic sample delivery devices. *RSC Adv.* **10**, 15694–15701 (2020).
81. Trebbin, M. et al. Microfluidic liquid jet system with compatibility for atmospheric and high-vacuum conditions. *Lab Chip* **14**, 1733–1745 (2014).
82. Martin, A. V. et al. Fluctuation X-ray diffraction reveals three-dimensional nanostructure and disorder in self-assembled lipid phases. *Comms. Mater.* **1**, 40 (2020).
83. Bielecki, J. et al. Electrospray sample injection for single-particle imaging with x-ray lasers. *Sci. Adv.* **5**, eaav8801 (2019).
84. Lu, A. et al. Unified polymerization mechanism for the assembly of ASC-dependent inflammasomes. *Cell* **156**, 1193–1206 (2014).
85. Wu, B. et al. Molecular imprinting as a signal-activation mechanism of the viral RNA sensor RIG-I. *Mol. Cell* **55**, 511–523 (2014).
86. David, L. et al. Assembly mechanism of the CARMA1–BCL10–MALT1–TRAF6 signalosome. *Proc. Natl. Acad. Sci. USA* **115**, 1499–1504 (2018).
87. Fu, T.-M. et al. Cryo-EM structure of Caspase-8 tandem DED filament reveals assembly and regulation mechanisms of the death-inducing signaling complex. *Mol. Cell* **64**, 236–250 (2016).
88. Burns, K. et al. MyD88, an adapter protein involved in interleukin-1 signaling. *J. Biol. Chem.* **273**, 12203–12209 (1998).
89. Stols, L. et al. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expr. Purif.* **25**, 8–15 (2002).
90. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
91. Cichocka, M. O., Ångström, J., Wang, B., Zou, X. & Smeets, S. High-throughput continuous rotation electron diffraction data acquisition via software automation. *J. Appl. Cryst.* **51**, 1652–1661 (2018).
92. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
93. Evans, P. R. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).
94. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
95. Keegan, R. M. & Winn, M. D. Automated search-model discovery and preparation for structure solution by molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 447–457 (2007).
96. Bunkóczi, B. & Read, R. J. Improvement of molecular-replacement models with Sculptor. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 303–312 (2010).
97. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
98. Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1948–1954 (2002).
99. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. D. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
100. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
101. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D Biol. Crystallogr.* **74**, 519–530 (2018).
102. Chen, V. B. et al. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
103. Roedig, P. et al. Room-temperature macromolecular crystallography using a micro-patterned silicon chip with minimal background scattering. *J. Appl. Crystallogr.* **49**, 968–975 (2016).
104. Herrmann, S. et al. CSPAD-140k: a versatile detector for LCLS experiments. *Nucl. Instrum. Methods Phys. Res. A* **718**, 550–553 (2013).
105. Philipp, H. T., Hromalik, M., Tate, M., Koerner, L. & Gruner, S. M. Pixel array detector for X-ray free electron laser experiments. *Nucl. Instrum. Methods Phys. Res. A* **649**, 67–69 (2011).
106. Barty, A. et al. Cheetah: Software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *J. Appl. Crystallogr.* **47**, 1118–1131 (2014).
107. White, T. A. et al. CrystFEL: a software suite for snapshot serial crystallography. *J. Appl. Crystallogr.* **45**, 335–341 (2012).
108. Leslie, A. G. W. Integration of macromolecular diffraction data. *Acta Crystallogr. D Biol. Crystallogr.* **D55**, 1696–1702 (1999).
109. Gevorkov, Y. et al. XGANDALF - extended gradient descent algorithm for lattice finding. *Acta Crystallogr. A* **75**, 694–704 (2019).
110. Duisenberg, A. J. M. Indexing in single-crystal diffraction with an obstinate list of reflections. *J. Appl. Crystallogr.* **25**, 92–96 (1992).
111. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
112. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
113. Holm, L. Using Dali for protein structure comparison. *Methods Mol. Biol.* **2112**, 29–42 (2020).
114. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
115. Tina, K. G., Bhadra, R. & Srinivasan, N. PIC: protein interactions calculator. *Nucleic Acids Res.* **35**, 473–476 (2007).
116. Unni, S. et al. Web servers and services for electrostatics calculations with APBS and PDB2PQR. *J. Comput. Chem.* **32**, 1488–1491 (2011).
117. Hunter, D. J. B., Bhumkar, A., Giles, N., Sieracki, E. & Gambin, Y. Unexpected instabilities explain batch-to-batch variability in cell-free protein expression systems. *Biotechnol. Bioeng.* **115**, 1904–1914 (2018).
118. Gagoski, D. et al. Gateway-compatible vectors for high-throughput protein expression in pro- and eukaryotic cell-free systems. *J. Biotechnol.* **195**, 1–7 (2015).
119. Gambin, Y. et al. Single-molecule analysis reveals self assembly and nanoscale segregation of two distinct cavin subcomplexes on caveolae. *eLife* **3**, e01434 (2014).
120. Gambin, Y. et al. Confocal spectroscopy to study dimerization, oligomerization and aggregation of proteins: a practical guide. *Int. J. Mol. Sci.* **17**, 655 (2016).
121. Kutzner, C. et al. More bang for your buck: improved use of GPU nodes for GROMACS 2018. *J. Comput. Chem.* **40**, 2418–2431 (2019).
122. Oostenbrink, C., Villa, A., Mark, A. E. & Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25**, 1656–1676 (2004).
123. Schmid, N. et al. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **40**, 843 (2011).
124. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
125. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
126. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. Interaction models for water in relation to protein hydration BT- intermolecular forces. In *Proc. 14th Jerusalem Symposium on Quantum Chemistry and Biochemistry*, 13–16 (Springer, 1981).
127. Miyamoto, S. & Kollman, P. A. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).

128. White, T. Processing serial crystallography data with CrystFEL: a step-by-step guide. *Acta Crystallogr. D Biol. Crystallogr.* **75**, 219–233 (2019).
129. Xu, Y. et al. Structural basis for signal transduction by the toll/interleukin-1 receptor domains. *Nature* **408**, 111–115 (2000).
130. Ferrao, R. et al. IRAK4 dimerization and trans-autophosphorylation are induced by Myddosome assembly. *Mol. Cell* **55**, 891–903 (2014).

Acknowledgements

We thank Ronan Keegan and Tim Gruene for insightful discussions on molecular replacement and model building. We thank Xiaodong Zou (X.Z.) for insightful discussions and critical manuscript reading. We acknowledge the use of the Centre for Microscopy and Microanalysis at the University of Queensland. The work was supported by the Swedish Research Council (2017–05333 to H.X., 2019–00815 to X.Z.), the Knut and Alice Wallenberg Foundation (2018.0237 to X.Z.), the SciLifeLab technology development project (MicroED@SciLifeLab to H.X.), the Wellcome Trust (209407/Z/17/Z supporting T.I.C.), the National Health and Medical Research Council (NHMRC grants 1107804 and 1160570 to B.K. and T.V., 1071659 to B.K. and 1108859 to T.V.) and the Australian Research Council (ARC) Laureate Fellowship (FL180100109 to B.K.). T.V. received ARC DECRA (DE170100783) funding. C.D., B.A., N.A.Z. and S.H. acknowledge the support of the Australian Research Council through the Centre of Excellence in Advanced Molecular Imaging (CE140100011). This work was supported by computational resources provided by the Australian Government through NCI-Gadi under the National Computational Merit Allocation Scheme (Project cj47) and Queensland Cyber Infrastructure Foundation (QCIF, Project f49). Portions of this research were carried out at the Linac Coherent Light Source (LCLS), a National User Facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The CXI instrument was funded by the LCLS Ultrafast Science Instruments (LUSI) project funded by the U.S. Department of Energy, Office of Basic Energy Sciences. Use of the LCLS, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract number DE-AC02-76SF00515. Portions of this research were carried out at the light source PETRAIII, P11 beamline at DESY, a member of the Helmholtz Association (HGF). We thank Dr. Anja Burkhardt, Dr. Alke Meents and Dr. Olga Lorbeer for assistance with using the beamline P11.

Author contributions

T.V., B.K., H.X. and C.D. conceived the project and provided project leadership. T.V. designed the crystallization experiments. T.V., J.D.N., C.D. and S.J.T. produced crystals. S.J.T. and K.J.S. designed and performed crystal growth experiments and analysed the data. T.V., J.D.N. and M.H.R. expressed and purified proteins. H.X. designed MicroED experiments. H.X., J.Z. and M.T.B.C. were involved in MicroED sample preparation, screening and data collection. M.T.B.C. and H.X. performed analysis and structure

determination using the MicroED data. T.I.C. performed model building. C.D. designed and led SFX experiments. C.D., L.F., A.A., M.S.H. and M.L. collected SFX data. S.H., N.A.Z., C.H.Y., B.A. and C.D. analysed SFX data. T.V., M.T.B.C. and S.H. analysed the MicroED and SFX structures. A.K.M. designed and performed the molecular dynamics experiments and analysed the data. T.V. designed site-directed mutants. K.J.S., P.R.V. and T.W.M. designed and performed cell biology experiments, and analysed the data. E.S., Y.G. and D.J.B.H. designed and performed single-molecule spectroscopy experiments, and analysed the data. T.V., B.K., C.D., H.X., S.H. and M.T.B.C. wrote the initial draft. All authors contributed to editing and writing the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-22590-6>.

Correspondence and requests for materials should be addressed to C.D., B.K., H.X. or T.V.

Peer review information *Nature Communications* thanks Nicholas J. Gay, Tsan Xiao and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

5.3 Summary

In this paper (for supplementary information refer to Appendix B) we have successfully solved a new, biologically relevant structure of MyD88, co-crystallised with MAL using both SFX and microED techniques. We have directly compared these cutting edge techniques, and no major conformational differences were observed.

We also determined that our MyD88 structures are different to the structures of MyD88 previously published, with conformational differences observed in key regions that relate to MyD88 oligomerisation. Using biological assays and modelling studies (carried out by our collaborators) we were able to demonstrate that the oligomerisation seen in our crystals, is the biologically relevant oligomerisation that occurs naturally within our cells in response to the recruitment of MyD88 by MAL. This new and biologically relevant structure of MyD88 provides new insights into the architecture and assembly mechanisms of TIR-domain signalosomes in TLR pathways. The TIR:TIR interactions detailed in this paper may also provide clues for designing small molecular mimics of the interfaces important for function, so that they can be used to inhibit MyD88 higher-order assembly formation for the purpose of potential therapeutic applications.

Chapter 6

Conclusions and Future Work

6.1 Summary of thesis

X-ray crystallography is a powerful tool for determining the atomic structure of proteins, but even with the availability of highly coherent and brilliant XFEL sources, X-ray crystallography is still reliant on the availability of high quality crystals. However, in terms of the overall crystal size, the availability of intense, highly focused X-ray sources have meant that micro and even nanocrystals can be used to generate high resolution diffraction data. The first hard X-ray free electron laser facility, the LCLS, led the charge in terms of developing SFX and determining protein structures nanocrystals. SFX has now become a routine tool for protein structural determination and is the technique that enabled us to solve the structure of MyD88 (Chapter 5), since these crystals were too small for conventional crystallography, and even for SMX (Chapter 4).

Over the past ten years a number of different delivery systems have been developed for serial crystallography, both for use at synchrotrons, and at XFELs. Depending on specific crystal characteristics, such as size, shape and sample availability, the system best suited for a particular target can vary. Three different delivery systems were tested and compared with the MyD88 crystals in order to evaluate the disadvantages and advantages of each system (Chapter 4). The systems included the GDVN at the LCLS, the HVI delivery system at the LCLS, and a microfluidic silicon chip delivery system at Petra III. The comparison of

these delivery systems thus involved considering the fundamental differences in the source characteristics (XFEL vs synchrotron). A major issue that was observed with MyD88 was crystal aggregation leading to multi-crystal diffraction. This caused problems for analysis using data obtained with both the chip, and LCP delivery systems, but was largely avoided with the use of the GDVN system. Using the GDVN resulted in high resolution single crystal diffraction (Chapter 4 & 5). With further developments and improvements to fixed-target chip systems, and additional preliminary testing of the media used for the HVI system these systems still merit further investigation as they are far better at minimising sample consumption, than the GDVN. In addition, the chip system is also capable of yielding a much higher hit-rate than either the HVI or GDVN system. However, the benefits of each delivery system need to be weighed against the sample requirements. In our case, the GDVN delivery system was the best choice for the MyD88 crystals, as it minimised the issue of crystal aggregation, and a sufficient amount of sample could be produced that sample consumption was not a major issue.

With the diffraction data obtained from the GDVN experiment at the CXI beam-line at the LCLS (Chapter 4), the biologically relevant structure of MyD88, co-crystallised with MAL was solved to a resolution of 2.30Å (Chapter 5). This new structure provides us with new insights into the assembly mechanisms of TIR-domain containing signalosomes such as MyD88 and MAL in the context of TLR pathways. Furthermore, details of the TIR:TIR interactions described may provide important clues regarding the design and development of small molecule mimics, that mimic the interfaces that are important for function. These small molecule mimics could be used to target and inhibit MyD88 higher-order assembly formation, which has potential therapeutic applications. In this study, the two complementary techniques of SFX and microED were also compared, with a detailed characterisation of both the SFX structure and the microED structure to determine whether any differences were present. The SFX data, collected at room temperature showed a slight increase in lattice parameters along the a-axis compared to the microED data, which was attributed to the cryogenic data collection conditions [206] for microED compared to room temperature data collection for the SFX data. The only other relevant difference was that the SFX data could be collected to a much higher resolution of 2.30Å vs 3.0Å with microED. Therefore

both techniques were complementary to each other and provided similar results with respect to the structure.

With the megahertz repetition rates now available at XFELs, as well as improvements and new developments to beam-line components, sample delivery systems, detector capabilities, and data analysis pipelines, the opportunities for widening the applicability of serial crystallography has been steadily increasing. We have developed a technique at the EuXFEL which takes advantage of its unique pulse-train structure we refer to as "multi-hit" SFX. Understanding the unique pulse-train format at the EuXFEL, and yielding multiple hits on the same crystal, this multi-hit SFX was successfully demonstrated with lysozyme crystals (Chapter 3). From data collected at the EuXFEL SPB/SFX beamline, crystals that were hit twice by consecutive X-ray pulses were successfully identified.

The process of identifying multi-hit crystals was carried out by firstly filtering for diffraction patterns obtained from consecutive pulses within a pulse train. The patterns are then further filtered, with only consecutive patterns that have less than 10° difference in their assigned orientations. These patterns are then classified as multi-hit, with a single crystal responsible for multiple diffraction patterns (Chapter 3). Having confirmed that the selection criteria for classifying multi-hits is accurate, the presence of these multi-hit crystals meant that questions arose regarding how this is possible. These questions included understanding how crystals could be hit more than once within the beam interaction region, and how parameters such as sample velocity, beam shape and diameter influence the occurrence of multi-hits. Furthermore, the question arose as to how many diffraction patterns could be produced by a single crystal via this multi-hit phenomenon (Chapter 3). Radiation dose was also a key consideration in order to determine whether the diffraction quality of crystals that were hit multiple times would be impacted. This was critical in deciding whether the structure obtained from the second hit data set would be affected by radiation-induced damage to the crystal.

To assess the possible effects of radiation damage in multi-hit SFX, the diffraction data was split into three isolated data subsets that included: i) crystal diffraction patterns from crystals that were only hit once, ii) from crystals hit twice, the diffraction patterns obtained from the first pulse, and iii) from crystals hit twice, the diffraction patterns from the second

pulse. Once the data was split, a structural solution was determined for each subset, solved independently so a comparison could then be performed. The comparison of these three structures was particularly focused on any change to the di-sulphide bonds, but no significant differences were observed. Therefore for the lysozyme crystals investigated, all diffraction patterns, regardless of whether they were a multi-hit pattern, were all equally useful for structural determination. This opens up the possibility of using this multi-hit phenomenon to develop new approaches to techniques such as tr-SFX. In conclusion, the work presented here demonstrates how serial crystallography is evolving, enabling new types of experiments to be performed as well as supporting other emerging techniques such as micro-ED.

6.2 Future Work

While single crystal diffraction patterns from MyD88 crystals were successfully obtained using the combination of a GDVN delivery system and XFEL sources, single crystal diffraction patterns were difficult to obtain with the other setups tested. The microfluidic silicon chip system in combination with a synchrotron source had two major issues: first, the crystals tended to aggregate, causing multi-crystal diffraction, and second, the small crystal size combined with the lower flux due to micro-focusing, limited the number of Bragg peaks that were detected. From experiments published by Soares et al. [197] the aggregation issues associated with the microfluidic chip system during the excess liquid removal step, are common. Lieske et al. [200] have developed an on-chip crystallisation method, allowing the crystals to be grown directly on the chip minimising handling of the crystals. This on-chip crystallisation method also provided a solution to the issue of crystal aggregation, with no aggregation observed using this on-chip crystallisation method [200].

With this new methodology in mind, a future experiment with MyD88 using on-chip crystallisation would offer an alternate method for addressing the aggregation issue that may mean that the silicon chip could be a viable alternative to the GDVN for MyD88. However, due to the small size of the MyD88 crystals, high resolution data obtained at a synchrotron source could still prove challenging to obtain. Therefore, utilising this on-chip crystallisation and delivery system at an XFEL, rather than a synchrotron would theoretically offer the

best solution to all of the issues observed with the MyD88 crystals using the microfluidic silicon chip. With the new goniometer systems that have been developed [79, 200], high quality diffraction data can be collected using fixed-target systems at 120 Hz XFELs, and furthermore, using the on-chip crystallisation and delivery system at an XFEL, a direct comparison between the fixed-target chip system and the moving target GDVN system, can now be performed with identical beam parameters, allowing for a far more quantitative approach.

From crystal diffraction data collected at the LCLS using the GDVN system, a 2.30Å high resolution biologically relevant structure for MyD88 was obtained (Chap. 5). Within this MyD88 structure were small regions of high flexibility, indicative of regions where multiple conformations are present. Within these regions there was an insufficient amount of electron density for the conformation to be fully solved. Therefore, with additional high quality diffraction data, and at a higher resolution, these areas within the MyD88 structure could have their multiple conformations fully solved. A second SFX experiment could also be conducted under cryogenic conditions using the on-chip crystallisation method suggested previously, in order eliminate any changes in the structure due to temperature differences, enabling the data sets from both microED and SFX to be as similar as possible. Also, having obtained a SFX data set under cryogenic conditions, a secondary comparison between room-temperature SFX and cryogenic SFX could also be performed, allowing us to further examine the effect of temperature on the MyD88 structure in as much detail as possible.

To investigate MyD88 further, with the aim of understanding the mechanisms involved in the recruitment of MyD88 by MAL towards immune activation of the TLR4 pathway, time-resolved serial crystallography studies should now be performed. These time-resolved studies should aim to capture the process of MAL initiating the higher-order assembly formation of MyD88, creating a "molecular movie" that would allow us to visualise this assembly formation process. This type of time-resolved experiment requires a mix-and-diffuse method to be used allowing the MyD88 and MAL to be mixed in a controlled way, allowing us to obtain diffraction data at the time point of interest. Developments in mixing devices make this possible, with new microfluidic mixing devices, as well as drop-on-drop delivery system, that utilise acoustic droplet injectors for droplet delivery [75, 77, 157, 207, 208].

Data collected at the SPB/SFX beamline at the EuXFEL provided an opportunity to investigate multi-hit SFX. Multi-hit SFX involves a single crystal that is hit more than once, as it is delivered to the XFEL beam via a moving target delivery system, such as the GDVN delivery system (Chapter 3). While further testing and optimisation is needed in order to fully determine whether it is possible to obtain a near 100% fraction of multi-hits during an experiment, using multi-hit serial crystallography as a time-resolved technique opens up the possibility of obtaining not just one intermediate structure, but multiple intermediate structures during a single experiment. For example, with conventional light-activated time-resolved serial crystallography, crystals are delivered to the X-ray beam using a GDVN delivery system, but before reaching the X-ray interaction region, a structural change is initiated via an optical laser at a specified time. This allows us to collect data from an intermediate structure as it passes through the beam interaction region. Taking advantage of our multi-hit technique, a second snapshot could be collected, allowing us to see a second intermediate structure with a time interval as small as 222 ns (using the 4.5 MHz capabilities of the EuXFEL) between the two intermediate snapshots. Observing multiple intermediate structures from a single crystal will allow us to truly create a "molecular movie", which could potentially form another valuable approach to SFX for larger, micron-sized crystals.

The field of serial crystallography is growing and diverging, with rapid development in beamline instrumentation, sample delivery, experimental techniques, and novel data analysis approaches. Key scientific questions in structural biology, whose answers were previously unattainable, now have the potential to be addressed. With the demonstration of multiple viable structures using multi-hit SFX, there is now a new avenue to explore in time-resolved crystallography. My work on the biologically important MyD88 structure has also laid the foundations for a series of follow up studies investigating its conformational dynamics. With these results in mind, I anticipate that the work I have presented here will continue to be developed and will have a positive impact on the field of serial crystallography.

Appendix A

Chapter 3 Supplementary Information

Supplementary Information

Supplementary Methods

Injector speed calculations

Using equation 1, the jet speed of the 13 $\mu\text{L}.\text{min}^{-1}$ / 85 $\text{mg}.\text{min}^{-1}$ flow rate combination could be theoretically calculated from the 78 m/s experimentally determined jet speed due to identical liquid flow rates (13 $\mu\text{L}.\text{min}^{-1}$). Equation (1), gives v as jet velocity, dP as gas pressure drop over the discharge orifice (energy source) and ρ as the liquid density. Because the gas flow rate is proportional to the pressure drop (dP in equation (1)), the jet speed (v) changes by this factor when comparing to the experimentally found jet speeds, where g is gas flow rate, equation (2).

$$\frac{v_1}{v_2} = \frac{\sqrt{2dP_1/\rho_1}}{\sqrt{2dP_2/\rho_2}} \quad (1)$$

if g is dependent on dP and $\rho_1 = \rho_2$ then

$$\frac{v_1}{v_2} = \sqrt{g_1/g_2} \quad (2)$$

Supplementary Figures

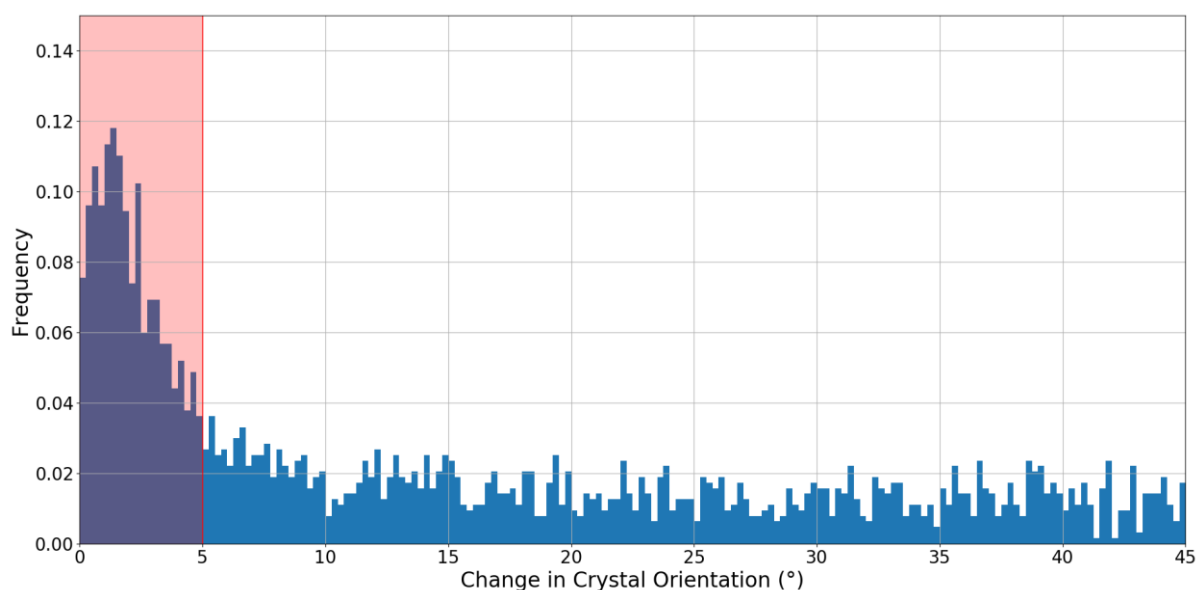


Figure S1 Frequency of images as a function of the relative change in crystal rotation for **b axis.** It is characterized by reciprocal space vector \vec{b} , between consecutive diffraction measurements (separated by 886 ns) within the X-ray pulse train. The flow rate was 42 m/s. An increase in the frequency of images with a change in orientation of less than 5 degrees can be seen due to the occurrence of double hits.

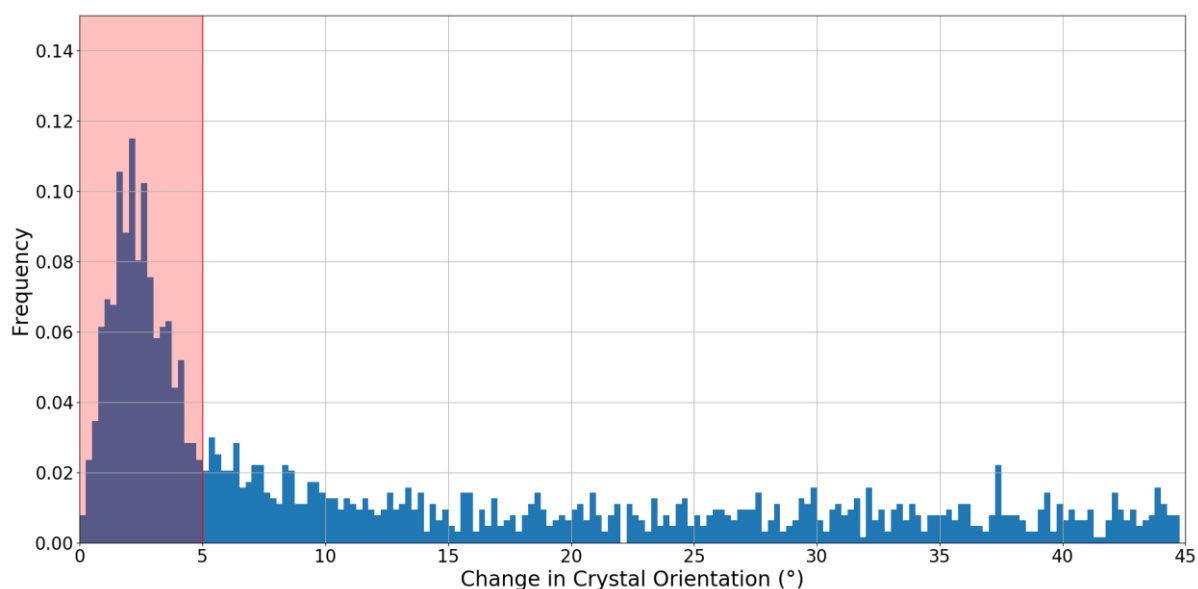


Figure S2 Frequency of images as a function of the relative change in crystal rotation for **c axis.** It is characterized by reciprocal space vector \vec{c} , between consecutive diffraction measurements (separated by 886 ns) within the X-ray pulse

train. The flow rate was 42 m/s. An increase in the frequency of images with a change in orientation of less than 5 degrees can be seen due to the occurrence of double hits.

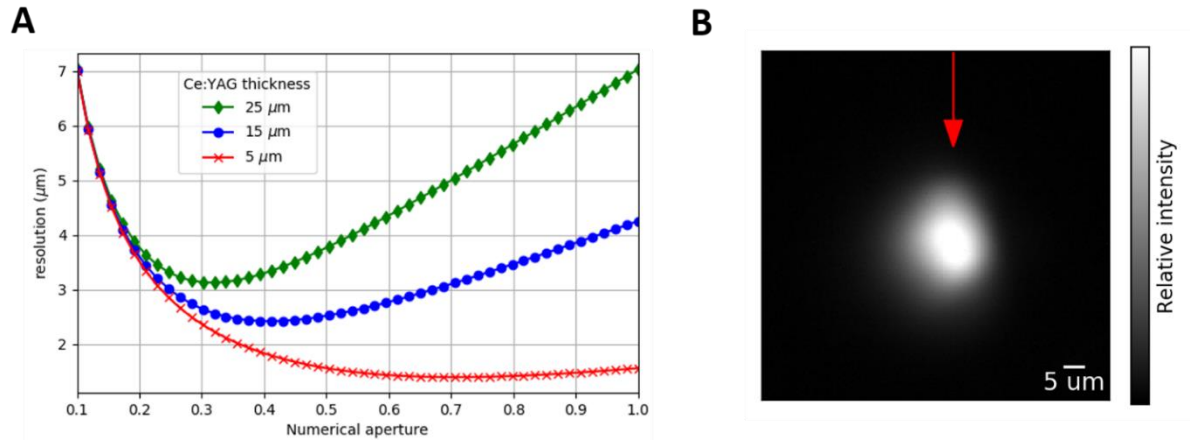


Figure S3 YAG information A. Resolution of a CE:YAG with varying thickness. The plot was generated based on Koch *et. al.*⁶¹, for three different YAG thickness and various numerical aperture values. The numerical aperture for the optical imaging system for the SPB beamline at the EuXFEL is between 0.3 to 0.4, providing a resolution estimate between 2-3 μm for the 15 μm YAG thickness. **B.** Average YAG image determined during the experiment with the red arrow indicating the jet directions.

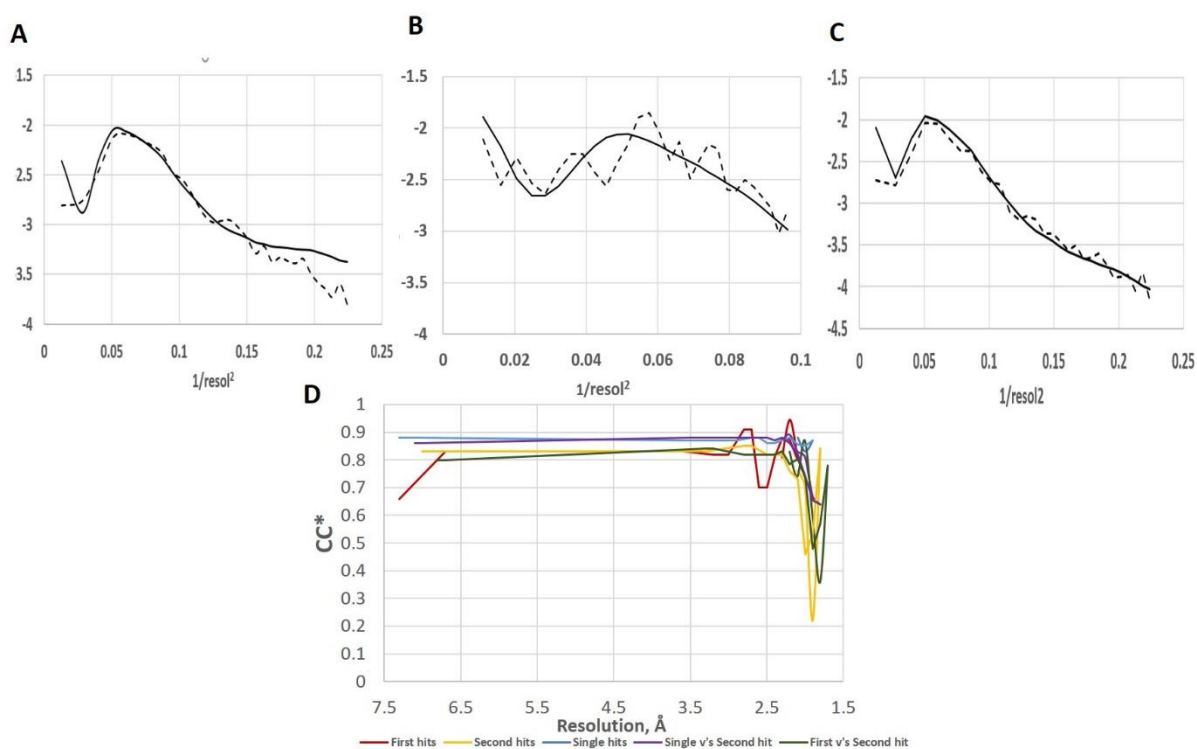


Figure S4. Statistical plots for the lysozyme structure. Wilson plot of merged data using Phenix for **A.** single hit crystals and **B.** crystals hit for the first time in the multi-hit scenario **C.** crystals hit for the second time in the multi-hit scenario. The dotted line indicates $\ln(I/I_{th})$ and the solid line represents (Reference_protein). **D.** CC* for data separated into single hits (blue), first hits (red) and second hits (yellow) as well as correlation of merged data of the second hit relative to the first hit (green) and single hit (purple) data sets, indicating they are very similar. The first hit structure is red and the second hit structure is green.

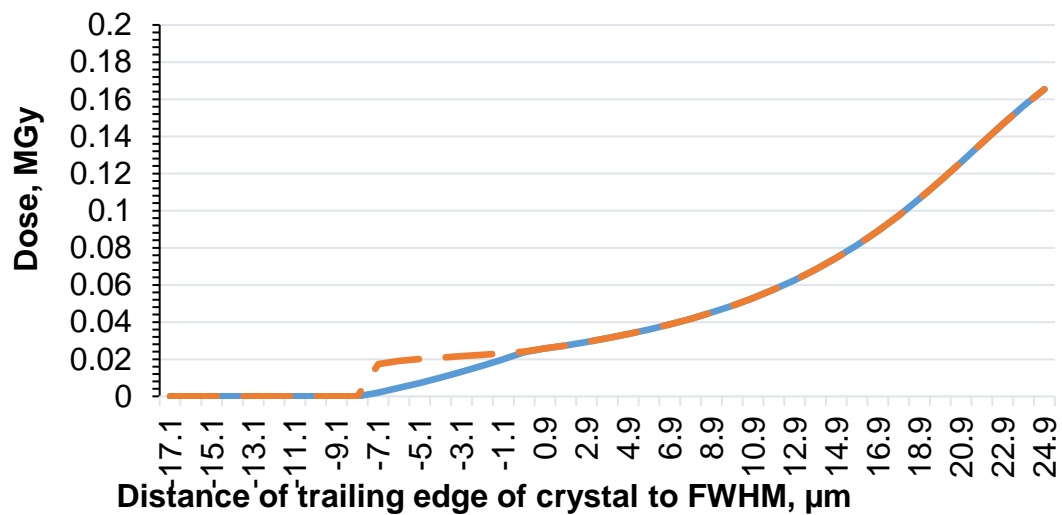


Figure S5 Dose plot showing the dose a crystal would receive as it enters the first half of the X-ray beam, left of the full width half maxima (FWHM). The dose is calculated based on 8 x 8 μm crystal size using Marman *et. al.*⁴² algorithm. Two plots are provided; dose estimated on the whole crystal (dotted orange line) and the dose calculated only on the exposed part of the crystal (solid green line).

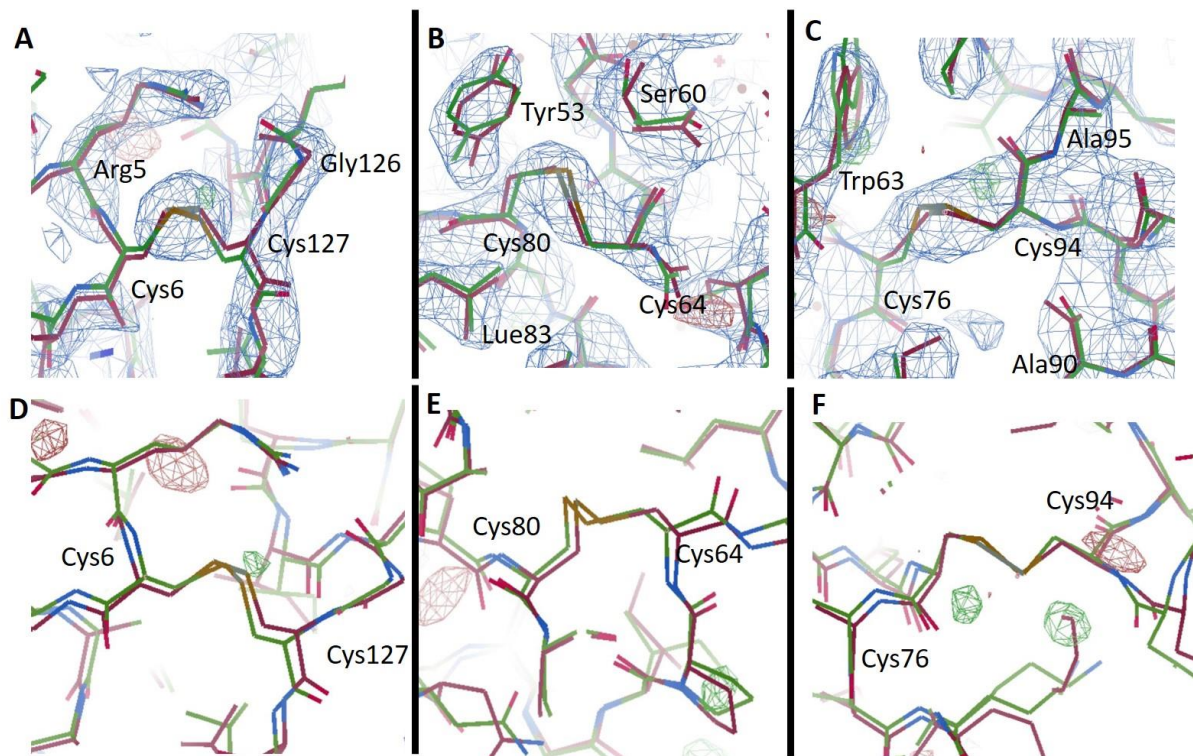


Figure S6 . Density maps surrounding the other three di-sulfide bonds in the lysozyme structure **A.** Cys6 and Cys127, **B.** Cys80 and Cys64 and **C.** Cys76 and

Cys94 and their corresponding difference electron density (DED) maps, **D.**, **E.** and **F.** The 2Fo-Fc map at 1 sigma is shown in blue and Fourier difference (Fo-Fc) maps at 3 sigma with green indicating positive density and red negative density.

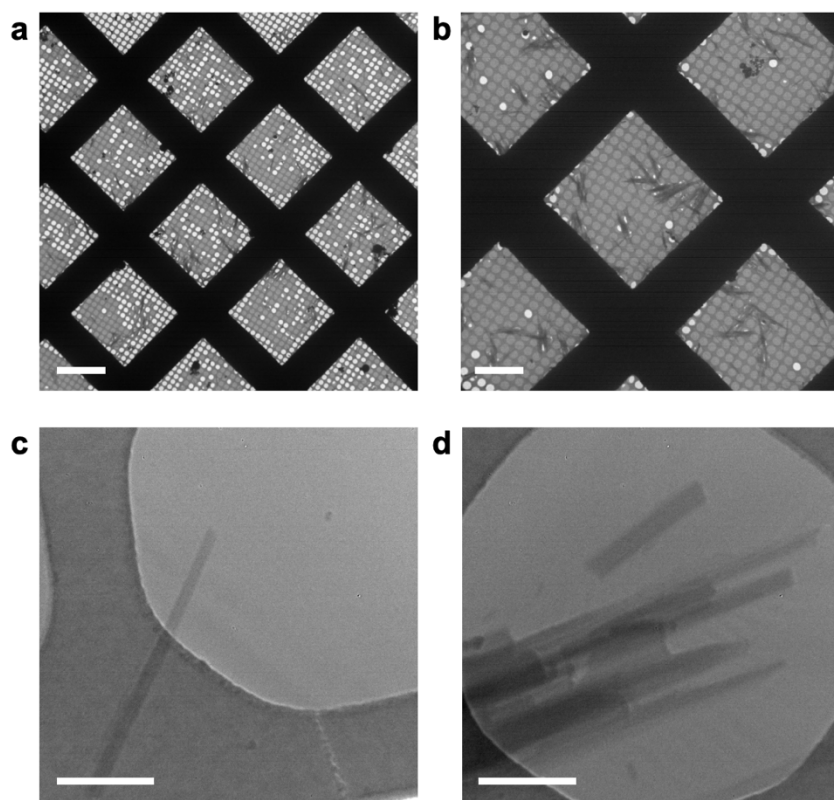
Appendix B

Chapter 5 Supplementary Information

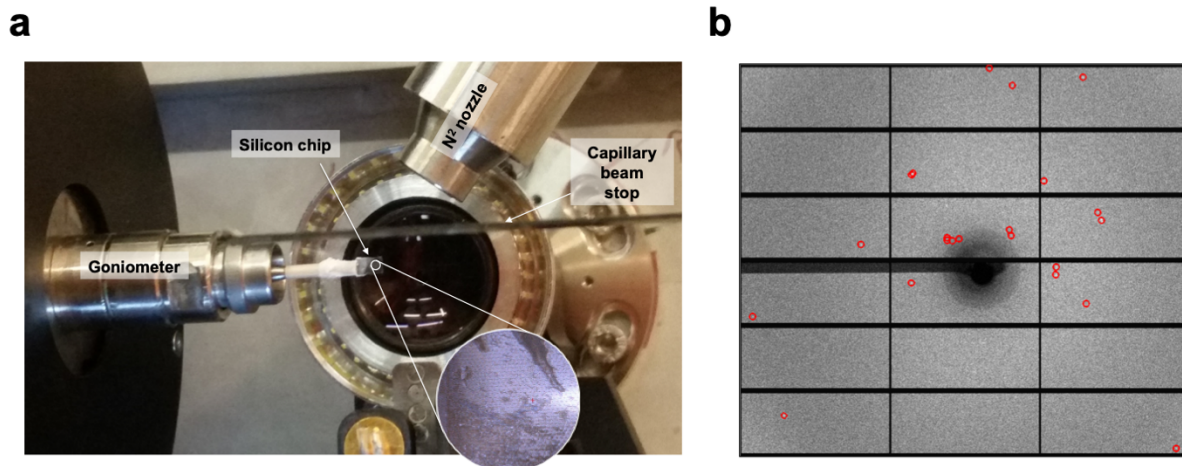
MyD88 TIR domain higher-order assembly interactions revealed by microcrystal electron diffraction and serial femtosecond crystallography

Clabbers *et al.*

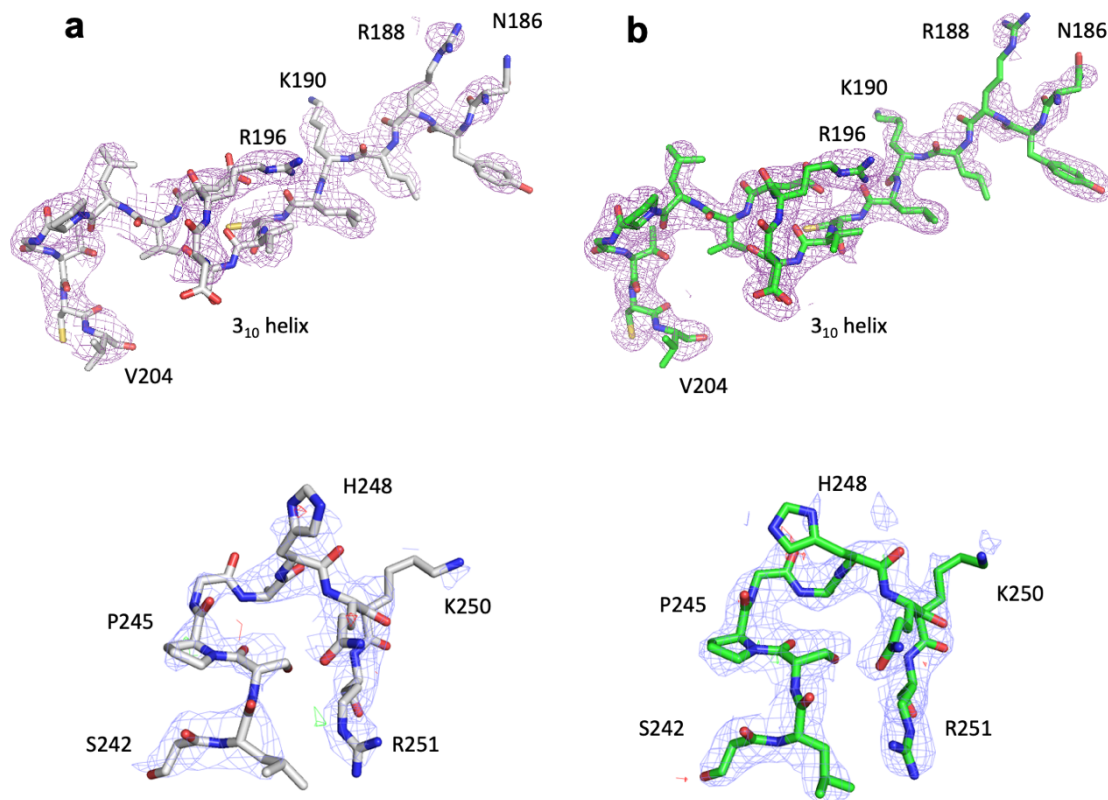
Supplementary Figures



Supplementary Figure 1. Micrographs of MyD88^{TIR} higher-order assembly microcrystals at different magnifications. **a.** Grid overview showing even distribution of microcrystals. Scale bar, 30 μm . **b.** Grid squares show substantial aggregation of microcrystals into bundles. Scale bar, 15 μm . **c.** Single hydrated microcrystal over a hole in the carbon support layer. Scale bar, 1 μm . **d.** Bundle of hydrated MyD88^{TIR} microcrystals. Scale bar, 1 μm .



Supplementary Figure 2. Crystallography set-up at PETRA III using the static chip system. **a.** MyD88^{TIR} sample set-up at the P11 beam-line at the Petra III synchrotron. The sample was placed on a silicon chip which was subsequently frozen in liquid nitrogen and mounted on the goniometer under the cryogenic nitrogen stream. The insert shows a close-up image of microcrystal aggregation on the silicon chip with a crystal concentration of 3.3×10^8 crystals/ml. This led to multiple crystal diffraction patterns in individual images, which could not be reliably indexed. **b.** A diffraction image obtained via the chip method at the P11 beam-line. Red circles highlight Bragg peaks identified via CrystFEL.



Supplementary Figure 3. Model building and refinement of MyD88^{TIR}. Models and simulated annealing (SA) composite omit maps (contoured at 1.0σ) are presented for the **a.** MicroED and **b.** SFX^b structures. Both structures were solved, built and refined using an identical protocol to enable a direct comparison. The BB loop (magenta isomesh, residues 186-204; top) and CD loop (blue isomesh, residues 242-251; bottom) regions are shown. The carbon atoms in the MicroED and SFX structures are shown in grey and green, respectively. Nitrogen, oxygen and sulfur atoms are shown in blue, red and yellow, respectively. The SA composite omit maps were calculated over the entire unit cell.

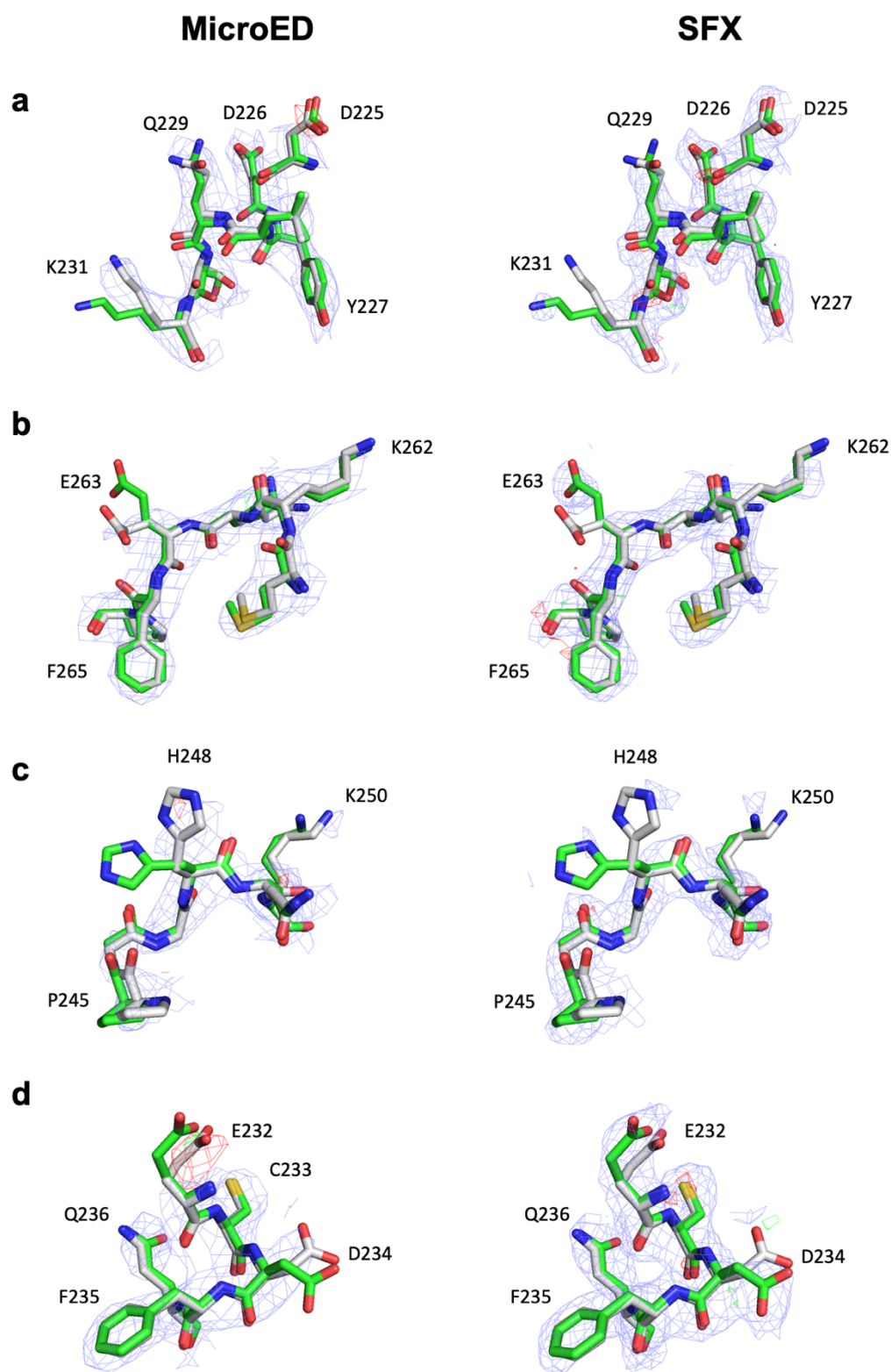
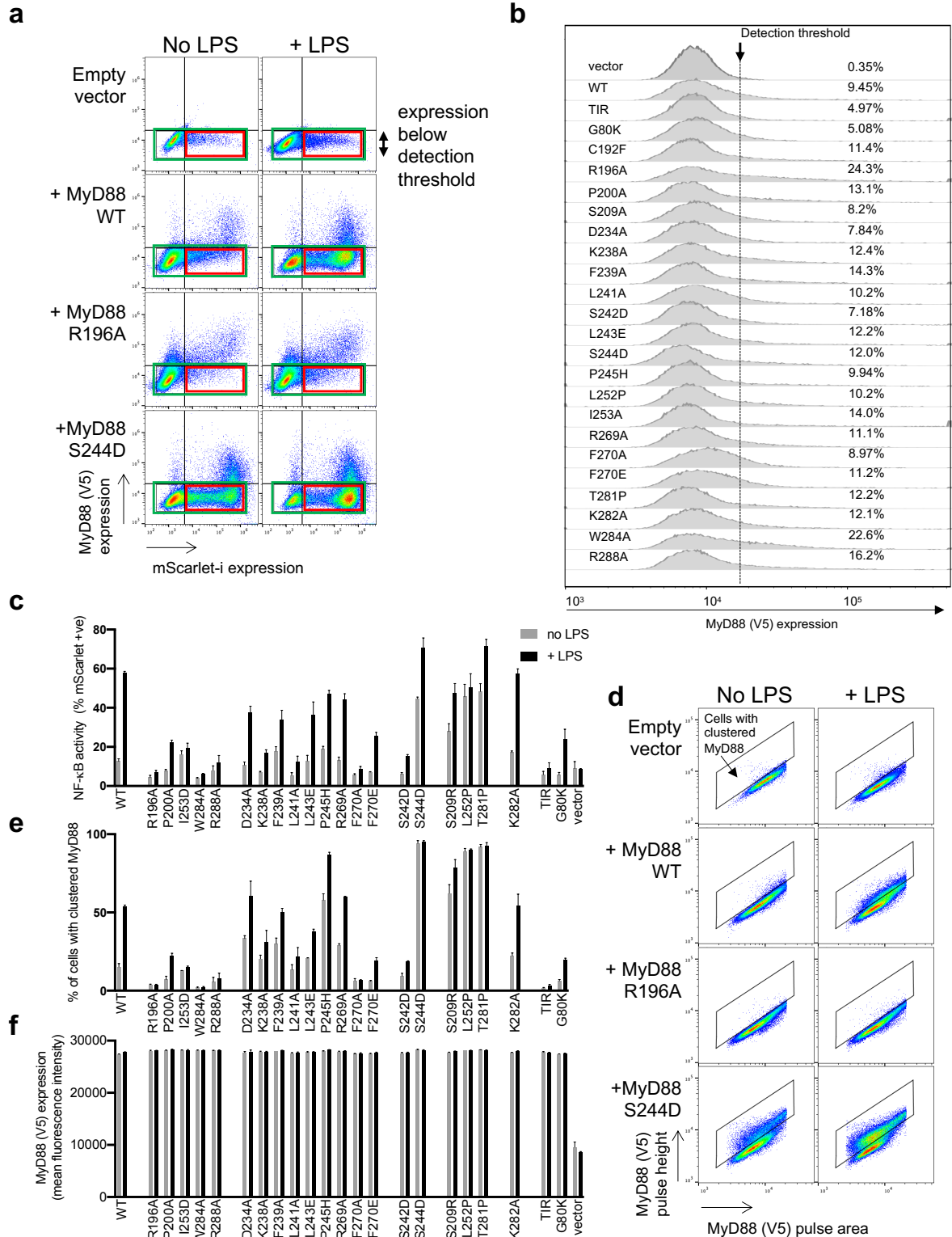


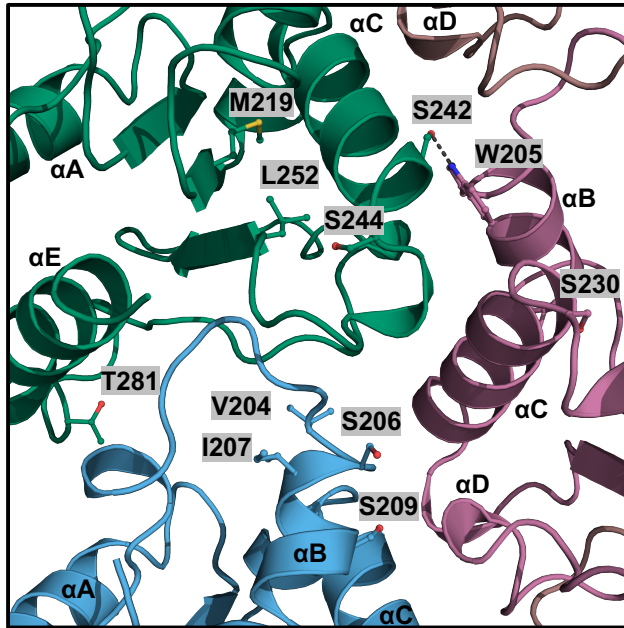
Figure S4. Structural comparison between the MicroED and SFX derived structures. The aligned MicroED (grey) and SFX^b (green) models are shown with electrostatic potential (MicroED; left panel) and electron density (SFX^b; right panel) maps. The models and maps are presented at four different sites; **a** D225-K231, **b** K261-P265, **c** P246-K250 and **d** E232-Q236. The structures were solved, built and refined using an identical protocol to enable a direct

comparison between the two methods. The carbon atoms in the MicroED and SFX structures are shown in grey and green, respectively. Nitrogen, oxygen and sulfur atoms are shown in blue, red and yellow, respectively. Electrostatic scattering and electron density $2mFo-DFc$ maps (blue isomesh) are contoured at 1.0σ , and the difference $mFo-DFc$ maps (green and red isomesh, for positive and negative density, respectively) are contoured at 3.0σ . No missing reflections were restored using weighted Fc values for map calculations.

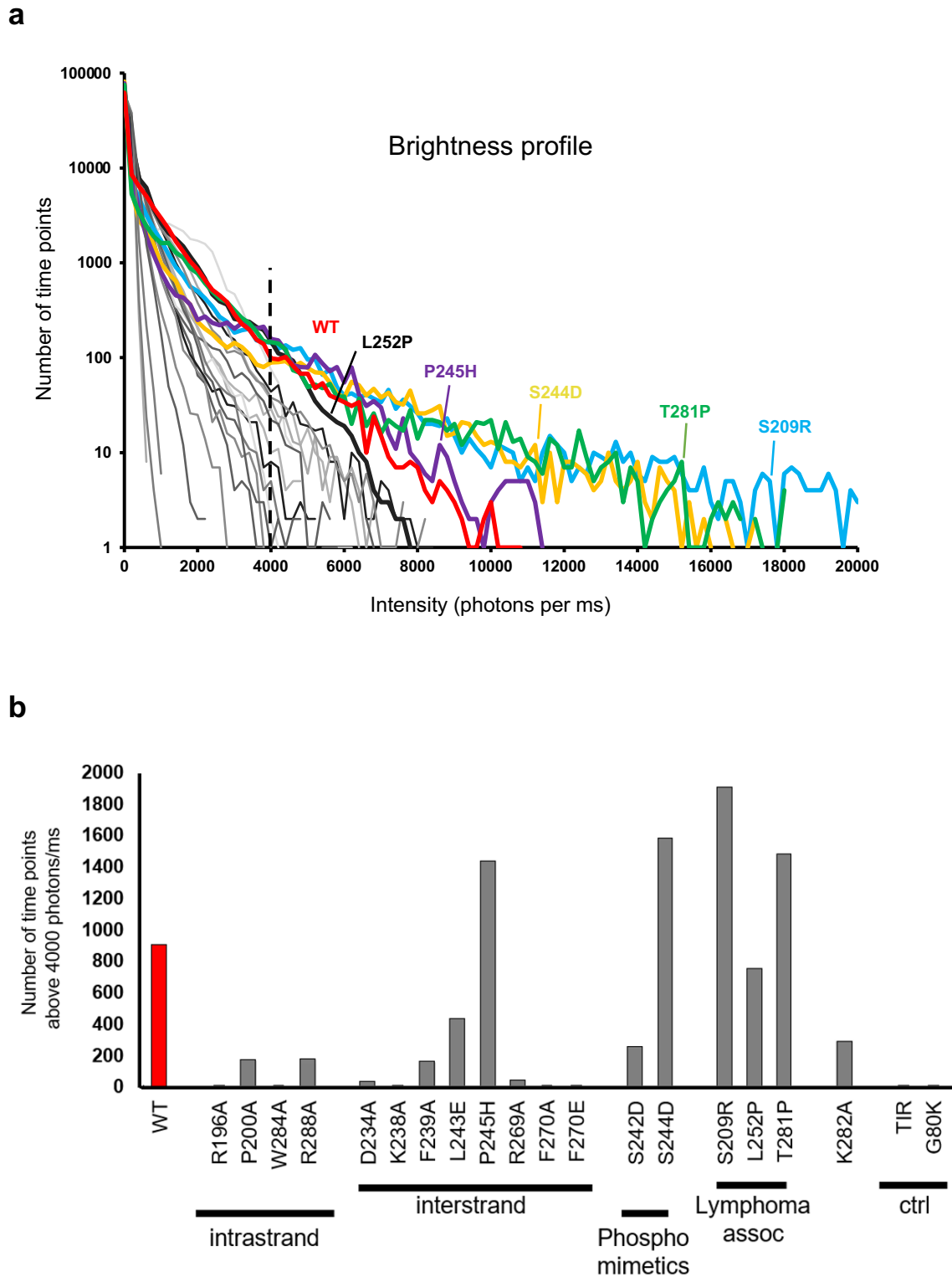


Supplementary Figure 5. Details of the flow cytometry-based assay for evaluation of MyD88 mutants. A HEK293 reporter cell line was prepared that stably expresses human TLR4, MD-2 and CD14, with an integrated NF- κ B-driven mScarlet-I reporter plasmid, and endogenous MyD88 knocked out. The resulting cell line (HEK-Blue-TLR4-NF- κ B-mScarlet-MyD88 KO) was transfected with plasmids expressing V5-tagged wild-type or mutant MyD88 and then treated with or without 100 ng/ml LPS. The amount of plasmid used was titrated down to give minimal spontaneous signaling. **a**. Plot of MyD88-V5 expression vs mScarlet

expression for selected samples with or without LPS treatment. Some basal mScarlet expression is seen in cells transfected with empty vector, and is at least partly related to transfection stress. Cells with WT MyD88 levels that were not above the background fluorescence level seen with empty vector transfection, still show induction of mScarlet-I expression upon stimulation with LPS. Because the cells are strongly responding to LPS stimulation we surmise that levels of MyD88 below the detection threshold are sufficient for function. This very low level of expression was selected for analysis as higher levels show spontaneous signaling. Defective mutants, such as R196A, do not respond to LPS at this MyD88 expression level. By contrast, constitutively active mutants, such as S244D, show elevated mScarlet expression in the absence of LPS. The geometric mean fluorescence signal within the red window was used for analysis of signaling response in Fig. 6a, to avoid including potentially variable numbers of untransfected cells. **b.** Histograms of cells transfected with the different MyD88 mutant plasmids show that all constructs have some expression at a level above the signal for the empty vector, confirming that all mutants can be expressed. **c.** NF- κ B activation expressed as the percentage of mScarlet-positive cells within the green boxed window in panel a shows similar effects of mutations to Fig. 6a. **d.** MyD88 clustering in Fig. 6b was determined from fluorescence peak height-vs-area plots for MyD88-V5 signal² from cells with low MyD88 within the green box gate in panel a. In cells with clustered MyD88, the peak height-to-area ratio of the MyD88-V5 signal is elevated, compared to that in cells with diffuse MyD88 localisation. Data for Fig. 6b were obtained from the percentage of cells with clustered MyD88 i.e. those falling within the marked gate. **e.** and **f.** MyD88 clustering was also determined at concentrations above the MyD88 detection threshold. At the expression level shown in panel f, there was a similar trend to that obtained at levels below the detection threshold (Fig. 6b), but greater saturation of the response.

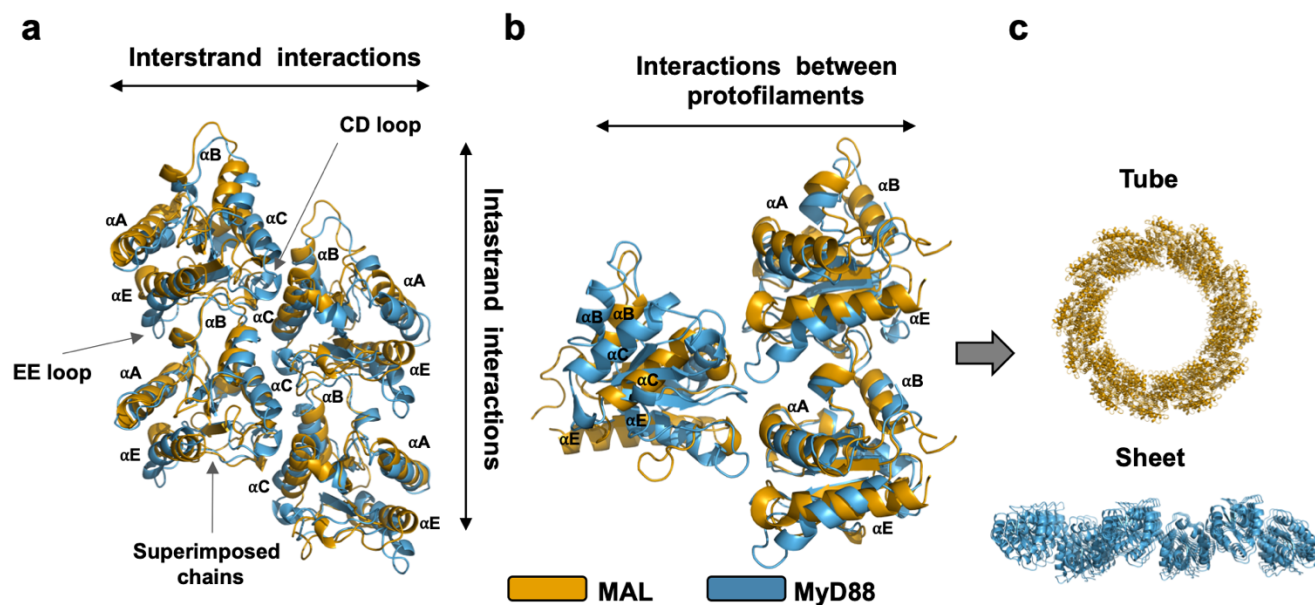


Supplementary Figure 6. Disease-associated and phosphomimetic mutations modulate MyD88 signaling and assembly. Ribbon diagram of the MyD88^{TIR} higher-order assembly with disease-associated and phosphorylated mutations highlighted in stick representation.

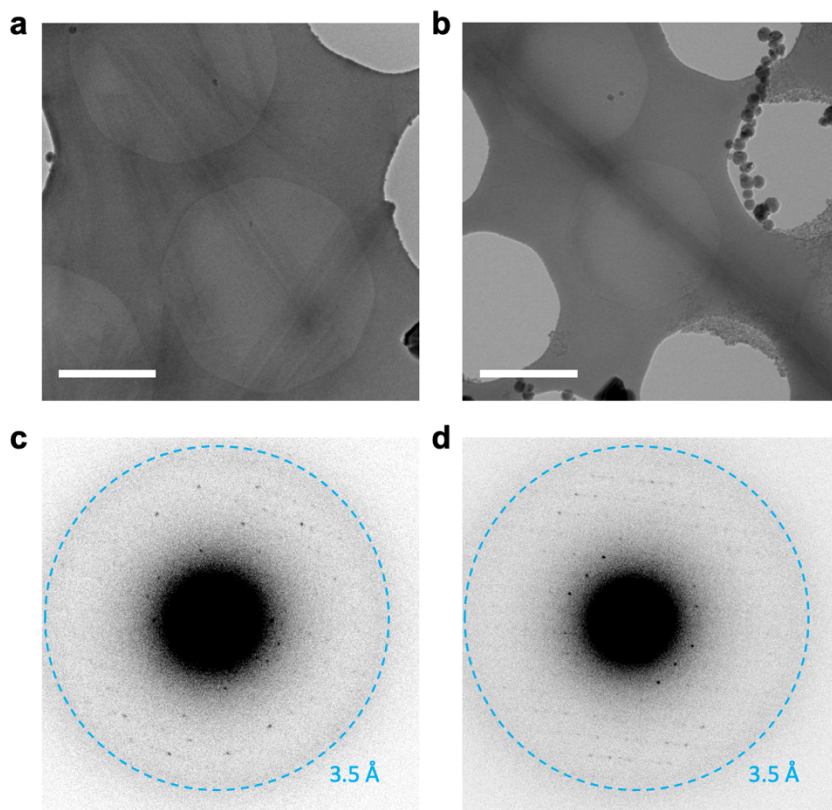


Supplementary Figure 7. Single-molecule analyses of MyD88^{TIR} assembly formation. a. Brightness profiles of MyD88 mutants. The brightness value in Fig. 6c represents the averaged value of many different particles. In inhomogeneous mixtures the brightness value alone cannot distinguish between rare, large particles or more frequent, but smaller particles. The maximum sizes of the assemblies were therefore analysed in more detail using the photon counting histogram (PCH) method⁷² where the frequency of observed events are plotted as a function of their amplitude. Smaller particles yield many peaks at lower intensities while for larger

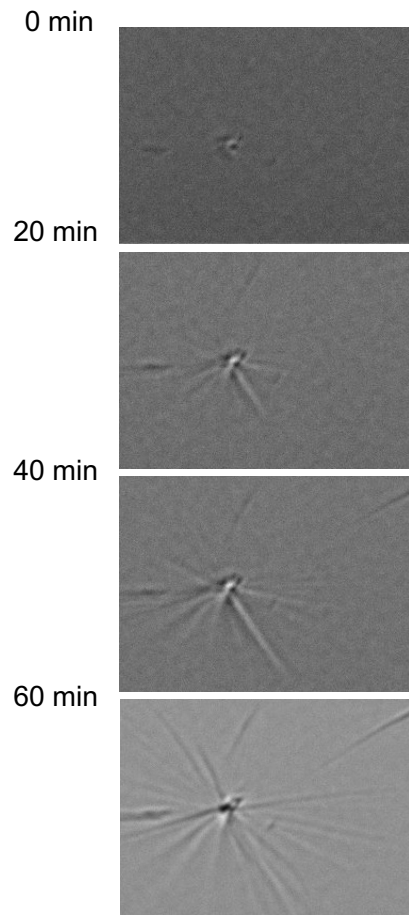
particles the distribution is skewed towards higher intensities. In the case of MyD88, the PCH analysis show that only the S209R, S244D, P245H, and T281P form larger particles than wild-type MyD88. **b.** The ability of wild-type MyD88 and mutants to form large particles was also assessed by counting the number of events with high intensities. When the number of events above an arbitrary selection threshold of 4,000 photons/ms were plotted, the results clearly showed that the S209R, S244D, P245H and T281P mutants form larger particles than wild-type MyD88.



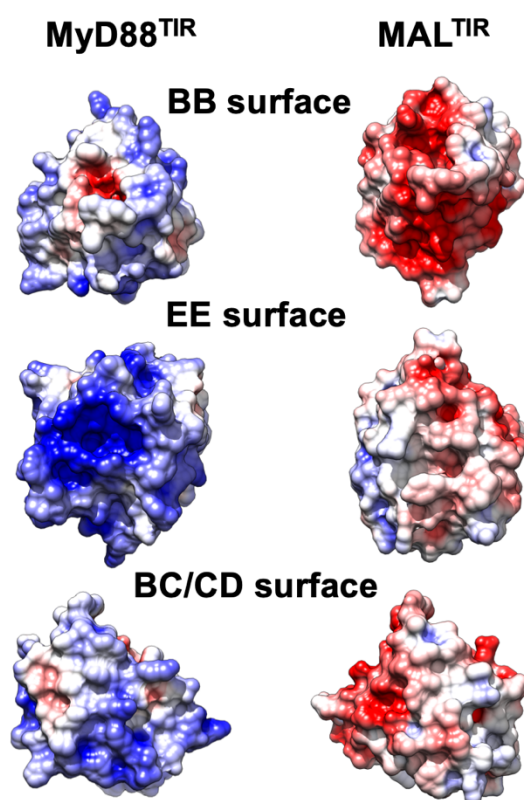
Supplementary Figure 8. Structural comparison of the MyD88^{TIR} microcrystal and MAL^{TIR} filament. **a.** Ribbon diagram comparing the MyD88^{TIR} and MAL^{TIR} higher-order assemblies. **b.** Ribbon diagram comparing the interactions between two-stranded higher-order assemblies in the MyD88^{TIR} microcrystal and MAL^{TIR} filament. **c.** Arrangement of protofilaments in the MAL^{TIR} filaments and two-stranded assemblies in MyD88^{TIR} microcrystals.



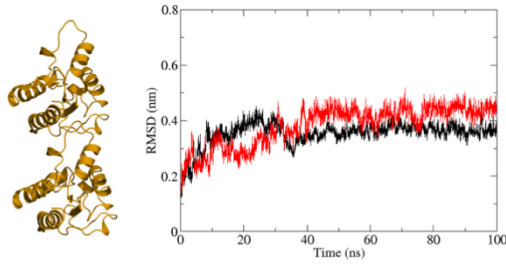
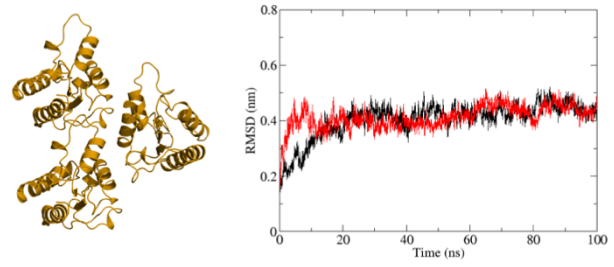
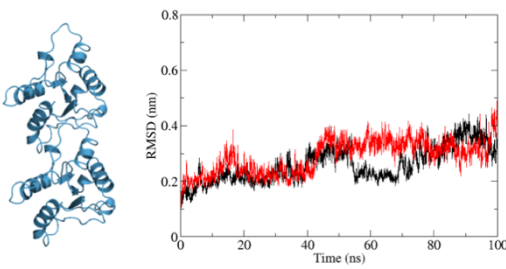
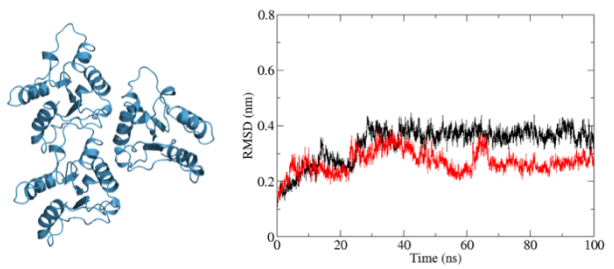
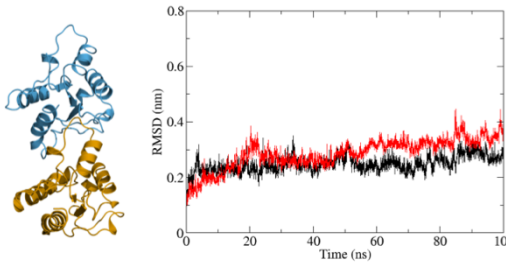
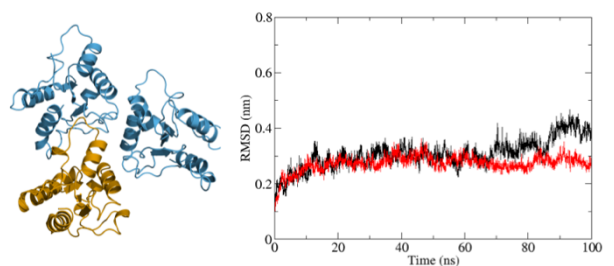
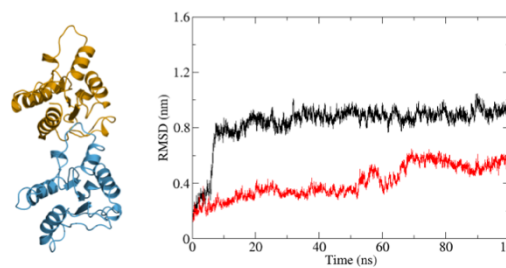
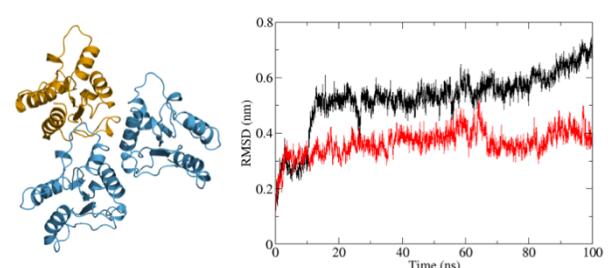
Supplementary Figure 9. Characterisation of GFP-MAL^{TIR} induced MyD88^{TIR} microcrystals by MicroED. **a,b.** Micrographs showing frozen hydrated GFP-MAL^{TIR}-seeded microcrystals on an EM grid. Scale bars: **a** 2 μm ; **b** 1.5 μm . **c, d.** Electron-diffraction patterns of GFP-MAL^{TIR}-seeded MyD88^{TIR} microcrystals, showing Bragg spots to 3.0 \AA resolution. Diffraction patterns were collected with an angular increment of 0.68° , and a dose rate of $0.12 \text{ e}/\text{\AA}^2$ per frame. The diffraction patterns were indexed in space group *C2* with unit cell parameters $a = 101.077$, $b = 30,635$, $c = 57.713$, and $\beta = 110.326^\circ$.



Supplementary Figure 10. MAL^{TIR} nucleates MyD88^{TIR} assembly formation unidirectionally. Time-lapse imaging of MyD88^{TIR} microcrystal formation. Representative images of multiple microcrystals, growing from a cluster of MAL^{TIR}-MyD88^{TIR} seeds, are shown. Data are representative of five independent experiments.



Supplementary Figure 11. Surface representations of MyD88^{TIR} and MAL^{TIR} with electrostatic potential mapped to the surface. Electrostatic potentials were calculated using APBS¹²⁸. Coloring is continuous going from blue (potential +10 kT/e) through white to red (potential -10 kT/e).

(i) MAL^{TIR} dimer**(ii) MAL^{TIR} trimer****(iii) MyD88^{TIR} dimer****(iv) MyD88^{TIR} trimer****(v) MAL^{TIR} BB:MyD88^{TIR} EE dimer****(vi) MAL^{TIR} BB:MyD88^{TIR} EE trimer****(vii) MAL^{TIR} EE:MyD88^{TIR} BB dimer****(viii) MAL^{TIR} EE:MyD88^{TIR} BB trimer**

Supplementary Figure 12. Molecular-dynamics simulations MyD88^{TIR} and MAL^{TIR} complexes. Starting structures and atom-positional RMSDs of the backbone atoms (C_α, N, C) from duplicate 100 ns molecular dynamics simulations (black and red) of dimeric and trimeric MAL^{TIR} (i—ii) and MyD88^{TIR} (iii–iv) complexes, and predicted MAL^{TIR} and MyD88^{TIR} complexes (v–viii) using a box with a 1.4 nm distance between the protein surface and the edge of the box wall. The complexes analysed are highlighted in ribbon representation (MAL^{TIR} in orange; MyD88^{TIR} in blue). BB = BB surface of intrastrand interaction; EE= EE surface of intrastrand interaction. The RMSD analysis revealed that the complexes involving the MAL BB surface and MyD88 EE surface (v–vi) are more stable than complexes involving the MAL EE surface and MyD88 BB surface (vii–viii).

Supplementary Video 1. MAL^{TIR} nucleates MyD88^{TIR} assembly formation unidirectionally. Movie of MyD88^{TIR} microcrystal formation from GFP-MAL^{TIR}-MyD88^{TIR} seeds.

Table S1: MyD88^{TIR} higher-order assembly interactions. Interfacing residues of the intrastrand interface, interstrand interface and the interfaces between the two-stranded higher-order assemblies based on the high-resolution SFX structure as defined by PISA¹¹³.

Interstrand interface		Intrastrand interface		Interface between two stranded assemblies	
Residues	HS ¹	Residues	HS ¹	Residues	HS ¹
P169		C203		S170	H (K282)
I172		W205	H (S242)	I172	
Q173		S206		Q173	H (Y276)
Q176		S209		F174	
I179		E213	S (R269)	Q176	
V193		K231		E177	
S194		E232		R180	
D195		D234	S (248)	Y257	
R196		F235		K258	
V198	H (R288)	K238		A259	
L199		F239		Y276	H (Q173)
P200		L241		T277	
G201		S242	H (W205)	N278	
T202	H (T272)	L243		P279	
C203		S244		C280	
V204		P245		K282	H (S170)
W205		G246		S283	
S206		H248	S (D234)		
I207		S266			
G246		I267			
A247		R269	S (E213)		
H248		F270			
K250					
R251					
L252					
I253					
P254					
R269					
F270					
I271					
T272	H (T202)				
C274					
N278					
C280					
T281					
W284					
R288	H (V198)				
L289					
K291					
A292					
L295					

¹ Residues making hydrogen bond (H) or salt bridge (S).

Table S2. Comparison of the buried surface areas of different interfaces.

Interface	Buried surface area (\AA^2)
MAL BB surface: MAL EE surface	1116.2
MyD88 BB surface: MyD88 EE surface ¹	1415.6
MAL BB surface: MyD88 EE surface (predicted) ¹	1576.4
MyD88 BB surface: MAL EE surface (predicted) ¹	1233.8
MAL BC surface: MAL CD surface	798.4
MyD88 BC surface: MyD88 CD surface ¹	965.2
MAL BC surface: MyD88 CD surface	764.4
MyD88 BC surface: MAL CD surface	934

¹ The calculated buried surface area is based on the MicroED MyD88^{TIR} structure

References

- [1] John Meurig Thomas. “The birth of X-ray crystallography”. *Nature* 491 (2012). DOI: 10.1038/491186a.
- [2] William Lawrence Bragg and William Henry Bragg. “The structure of some crystals as indicated by their diffraction of X-rays”. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 89 (1913). DOI: 10.1098/rspa.1913.0083.
- [3] William Henry Bragg and William Lawrence Bragg. “The structure of the diamond”. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 89 (1913). DOI: 10.1098/rspa.1913.0084.
- [4] J. C. Kendrew et al. “A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis”. *Nature* 181 (1958). DOI: 10.1038/181662a0.
- [5] Carlos Mendez et al. “Making use of Symmetries in the Three-Dimensional Elastic Inverse Homogenization Problem”. *International Journal for Multiscale Computational Engineering* (2019). DOI: 10.1615/IntJMultCompEng.2019029111.
- [6] Brian W Matthews. “Racemic crystallography—Easy crystals and easy structures: What’s not to like?” *Protein Science : A Publication of the Protein Society* 18 (2009). DOI: 10.1002/pro.125.
- [7] M. I. Aroyo, ed. *International Tables for Crystallography: Space-group symmetry*. 2nd ed. Vol. A. Chester, England: International Union of Crystallography, 2016. DOI: 10.1107/97809553602060000114.

- [8] Kati Valtonen. “CHARACTERIZATION OF WATER-BASED BARRIER COATINGS”. PhD thesis. 2002. DOI: 10.13140/RG.2.2.12963.91681.
- [9] P. P. Ewald. “Die Berechnung optischer und elektrostatischer Gitterpotentiale”. *Annalen der Physik* 369 (1921). DOI: <https://doi.org/10.1002/andp.19213690304>.
- [10] P. P. Ewald. “Introduction to the dynamical theory of X-ray diffraction”. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* 25 (1969). DOI: 10.1107/S0567739469000155.
- [11] Thomas A. White. “Post-refinement method for snapshot serial crystallography”. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (2014). DOI: 10.1098/rstb.2013.0330.
- [12] Mark Ladd and Rex Palmer. “I X-rays, X-ray Diffraction, and Structure Factors”. *Structure Determination by X-ray Crystallography*. Ed. by Mark Ladd and Rex Palmer. Boston, MA: Springer US, 2003. DOI: 10.1007/978-1-4615-0101-5_3.
- [13] G. Taylor. “The phase problem”. *Acta Crystallographica Section D: Biological Crystallography* 59 (2003). DOI: 10.1107/S0907444903017815.
- [14] William Lawrence Bragg and Max Ferdinand Perutz. “The structure of haemoglobin”. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 213 (1952). DOI: 10.1098/rspa.1952.0136.
- [15] Herbert Hauptman. “The Direct Methods of X-ray Crystallography”. *Science* 233 (1986).
- [16] D. M. Blow and F. H. C. Crick. “The treatment of errors in the isomorphous replacement method”. *Acta Crystallographica* 12 (1959). DOI: 10.1107/S0365110X59002274.
- [17] R. J. Read. “New ways of looking at experimental phasing”. *Acta Crystallographica Section D: Biological Crystallography* 59 (2003). DOI: 10.1107/S0907444903017918.
- [18] Z Dauter, M Dauter, and E.j Dodson. “Jolly SAD”. *Acta Crystallographica Section D* 58 (2002). DOI: 10.1107/S090744490200118X.
- [19] E. Dodson. “Is it jolly SAD?” *Acta Crystallographica Section D: Biological Crystallography* 59 (2003). DOI: 10.1107/S0907444903020936.

- [20] J. C. Phillips and K. O. Hodgson. “The use of anomalous scattering effects to phase diffraction patterns from macromolecules”. *Acta Crystallographica Section A* 36 (1980). DOI: <https://doi.org/10.1107/S0567739480001817>.
- [21] Bi-Cheng Wang. “Resolution of phase ambiguity in macromolecular crystallography”. *Methods in Enzymology*. Vol. 115. Diffraction Methods for Biological Macromolecules Part B. Academic Press, 1985. DOI: 10.1016/0076-6879(85)15009-3.
- [22] A. T. Brünger et al. “Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination”. *Acta Crystallographica Section D: Biological Crystallography* 54 (1998). DOI: 10.1107/S0907444998003254.
- [23] Kevin D. Cowtan and Kam Y. J. Zhang. “Density modification for macromolecular phase improvement”. *Progress in Biophysics and Molecular Biology* 72 (1999). DOI: 10.1016/S0079-6107(99)00008-5.
- [24] T. C. Terwilliger. “Automated structure solution, density modification and model building”. *Acta Crystallographica Section D: Biological Crystallography* 58 (2002). DOI: 10.1107/S0907444902016438.
- [25] D. Flot et al. “The ID23-2 structural biology microfocus beamline at the ESRF”. *Journal of Synchrotron Radiation* 17 (2010). DOI: 10.1107/S0909049509041168.
- [26] Michelle S. Miller et al. “Getting the Most Out of Your Crystals: Data Collection at the New High-Flux, Microfocus MX Beamlines at NSLS-II”. *Molecules* 24 (2019). DOI: 10.3390/molecules24030496.
- [27] Anja Burkhardt et al. “Status of the crystallography beamlines at PETRA III”. *The European Physical Journal Plus* 131 (2016). DOI: 10.1140/epjp/i2016-16056-0.
- [28] Egor Sobolev et al. “Megahertz single-particle imaging at the European XFEL”. *Communications Physics* 3 (2020). DOI: 10.1038/s42005-020-0362-y.
- [29] A.p Mancuso et al. “The Single Particles, Clusters and Biomolecules and Serial Femtosecond Crystallography instrument of the European XFEL: initial installation”. *Journal of Synchrotron Radiation* 26 (2019). DOI: 10.1107/S1600577519003308.

- [30] Naomi E Chayen. “Turning protein crystallisation from an art into a science”. *Current Opinion in Structural Biology* 14 (2004). DOI: 10.1016/j.sbi.2004.08.002.
- [31] Neer Asherie. “Protein crystallization and phase diagrams”. *Methods. Macromolecular Crystallization* 34 (2004). DOI: 10.1016/j.ymeth.2004.03.028.
- [32] Irene Russo Krauss et al. “An Overview of Biological Macromolecule Crystallization”. *International Journal of Molecular Sciences* 14 (2013). DOI: 10.3390/ijms140611643.
- [33] Fabrice Gorrec and Jan Löwe. “Automated Protocols for Macromolecular Crystallization at the MRC Laboratory of Molecular Biology”. *Journal of Visualized Experiments : JoVE* (2018). DOI: 10.3791/55790.
- [34] K. Zhukovsky. “Undulators for Short Pulse X-Ray Self-Amplified Spontaneous Emission-Free Electron Lasers”. *High Energy and Short Pulse Lasers*. IntechOpen, 2016. DOI: 10.5772/64439.
- [35] Brian McNeil. “First light from hard X-ray laser”. *Nature Photonics* 3 (2009). DOI: 10.1038/nphoton.2009.110.
- [36] P. Emma et al. “First lasing and operation of an ångstrom-wavelength free-electron laser”. *Nature Photonics* 4 (2010). DOI: 10.1038/nphoton.2010.176.
- [37] Christopher J. Milne et al. “SwissFEL: The Swiss X-ray Free Electron Laser”. *Applied Sciences* 7 (2017). DOI: 10.3390/app7070720.
- [38] David Pile. “First light from SACLA”. *Nature Photonics* 5 (2011). DOI: 10.1038/nphoton.2011.178.
- [39] In Soo Ko et al. “Construction and Commissioning of PAL-XFEL Facility”. *Applied Sciences* 7 (2017). DOI: 10.3390/app7050479.
- [40] Hans Weise and Winfried Decking. “Commissioning and First Lasing of the European XFEL”. *Proceedings of the 38th Int. Free Electron Laser Conf. FEL2017* (2018). In collab. with Bishofberger Kip (Ed.), Carlsten Bruce (Ed.), and Schaa RW (Ed.) Volker. DOI: 10.18429/JACOW-FEL2017-MOC03.
- [41] Iulia Georgescu. “The first decade of XFELs”. *Nature Reviews Physics* 2 (2020). DOI: 10.1038/s42254-020-0204-6.

-
- [42] Claudio Pellegrini. “The development of XFELs”. *Nature Reviews Physics* 2 (2020). DOI: 10.1038/s42254-020-0197-1.
- [43] “The next decade of XFELs”. *Nature Reviews Physics* 2 (2020). DOI: 10.1038/s42254-020-0206-4.
- [44] Uwe Bergmann, Vittal Yachandra, and Junko Yano. *X-Ray Free Electron Lasers : Applications in Materials, Chemistry and Biology*. Energy and Environment Series. [Cambridge]: Royal Society of Chemistry, 2017.
- [45] Marie Luise Grünbein et al. “Megahertz data collection from protein microcrystals at an X-ray free-electron laser”. *Nature Communications* 9 (2018). DOI: 10.1038/s41467-018-05953-4.
- [46] Max O. Wiedorn et al. “Megahertz serial crystallography”. *Nature Communications* 9 (2018). DOI: 10.1038/s41467-018-06156-7.
- [47] W. Decking et al. “A MHz-repetition-rate hard X-ray free-electron laser driven by a superconducting linear accelerator”. *Nature Photonics* 14 (2020). DOI: 10.1038/s41566-020-0607-z.
- [48] J. M. Holton and K. A. Frankel. “The minimum crystal size needed for a complete diffraction data set”. *Acta Crystallographica Section D: Biological Crystallography* 66 (2010). DOI: 10.1107/S0907444910007262.
- [49] Johndale C. Solem. “Imaging biological specimens with high-intensity soft x rays”. *JOSA B* 3 (1986). DOI: 10.1364/JOSAB.3.001551.
- [50] Richard Neutze et al. “Potential for biomolecular imaging with femtosecond X-ray pulses”. *Nature* 406 (2000). DOI: 10.1038/35021099.
- [51] Henry N. Chapman et al. “Femtosecond Diffractive Imaging with a Soft-X-ray Free-Electron Laser”. *Nature Physics* 2 (2006). DOI: 10.1038/nphys461. arXiv: physics/0610044.
- [52] Sébastien Boutet et al. “High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography”. *Science* 337 (2012). DOI: 10.1126/science.1217737.

- [53] Feng-Zhu Zhao et al. “A guide to sample delivery systems for serial crystallography”. *The FEBS Journal* 286 (2019). DOI: 10.1111/febs.15099.
- [54] M. L. Grünbein and G. Nass Kovacs. “Sample delivery for serial crystallography at free-electron lasers and synchrotrons”. *Acta Crystallographica Section D: Structural Biology* 75 (2019). DOI: 10.1107/S205979831801567X.
- [55] I. Martiel, H. M. Müller-Werkmeister, and A. E. Cohen. “Strategies for sample delivery for femtosecond crystallography”. *Acta Crystallographica Section D: Structural Biology* 75 (2019). DOI: 10.1107/S2059798318017953.
- [56] D. P. DePonte et al. “Gas dynamic virtual nozzle for generation of microscopic droplet streams”. *Journal of Physics D: Applied Physics* 41 (2008). DOI: 10.1088/0022-3727/41/19/195505.
- [57] Ki Hyun Nam. “Sample Delivery Media for Serial Crystallography”. *International Journal of Molecular Sciences* 20 (2019). DOI: 10.3390/ijms20051094.
- [58] Dominik Oberthuer et al. “Double-flow focused liquid injector for efficient serial femtosecond crystallography”. *Scientific Reports* 7 (2017). DOI: 10.1038/srep44628.
- [59] Dianfan Li and Martin Caffrey. “Structure and Functional Characterization of Membrane Integral Proteins in the Lipid Cubic Phase”. *Journal of Molecular Biology. Molecular Mechanisms in Integral Membrane Enzymology* 432 (2020). DOI: 10.1016/j.jmb.2020.02.024.
- [60] S. Botha et al. “Room-temperature serial crystallography at synchrotron X-ray sources using slowly flowing free-standing high-viscosity microstreams”. *Acta Crystallographica Section D: Biological Crystallography* 71 (2015). DOI: 10.1107/S1399004714026327.
- [61] Przemyslaw Nogly et al. “Lipidic cubic phase serial millisecond crystallography using synchrotron radiation”. *IUCrJ* 2 (Pt 2 2015). DOI: 10.1107/S2052252514026487.
- [62] J. M. Martin-Garcia et al. “Serial millisecond crystallography of membrane and soluble protein microcrystals using synchrotron radiation”. *IUCrJ* 4 (2017). DOI: 10.1107/S205225251700570X.

- [63] Tobias Weinert et al. “Serial millisecond crystallography for routine room-temperature structure determination at synchrotrons”. *Nature Communications* 8 (2017). DOI: 10.1038/s41467-017-00630-4.
- [64] P Berntsen et al. “The serial millisecond crystallography instrument at the Australian Synchrotron incorporating the “Lipidico” injector”. *Rev. Sci. Instrum.* (2019).
- [65] Marjan Hadian-Jazi et al. “Analysis of Multi-Hit Crystals in Serial Synchrotron Crystallography Experiments Using High-Viscosity Injectors”. *Crystals* 11 (2021). DOI: 10.3390/cryst11010049.
- [66] Uwe Weierstall et al. “Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography”. *Nature Communications* 5 (2014). DOI: 10.1038/ncomms4309.
- [67] Wei Liu, Andrii Ishchenko, and Vadim Cherezov. “Preparation of microcrystals in lipidic cubic phase for serial femtosecond crystallography”. *Nature Protocols* 9 (2014). DOI: 10.1038/nprot.2014.141.
- [68] R. Fromme et al. “Serial femtosecond crystallography of soluble proteins in lipidic cubic phase”. *IUCrJ* 2 (2015). DOI: 10.1107/S2052252515013160.
- [69] Robert KY Cheng. “Towards an Optimal Sample Delivery Method for Serial Crystallography at XFEL”. *Crystals* 10 (2020). DOI: 10.3390/cryst10030215.
- [70] Richard Ellson et al. “Transfer of Low Nanoliter Volumes between Microplates Using Focused Acoustics—Automation Considerations”. *JALA: Journal of the Association for Laboratory Automation* 8 (2003). DOI: 10.1016/S1535-5535-03-00011-X.
- [71] F. Mafuné et al. “Microcrystal delivery by pulsed liquid droplet for serial femtosecond crystallography”. *Acta Crystallographica Section D: Structural Biology* 72 (2016). DOI: 10.1107/S2059798316001480.
- [72] Christian G. Roessler et al. “Acoustic Injectors for Drop-On-Demand Serial Femtosecond Crystallography”. *Structure* 24 (2016). DOI: 10.1016/j.str.2016.02.007.

-
- [73] Franklin D. Fuller et al. “Drop-on-demand sample delivery for studying biocatalysts in action at X-ray free-electron lasers”. *Nature Methods* 14 (2017). DOI: 10.1038/nmeth.4195.
- [74] Z. Su et al. “Serial crystallography using automated drop dispensing”. *Journal of Synchrotron Radiation* 28 (2021). DOI: 10.1107/S1600577521006160.
- [75] Agata Butryn et al. “An on-demand, drop-on-drop method for studying enzyme catalysis by serial crystallography”. *Nature Communications* 12 (2021). DOI: 10.1038/s41467-021-24757-7.
- [76] Majid Hejazian, Eugeniu Balaur, and Brian Abbey. “Recent Advances and Future Perspectives on Microfluidic Mix-and-Jet Sample Delivery Devices”. *Micromachines* 12 (2021). DOI: 10.3390/mi12050531.
- [77] Kenneth R. Beyerlein et al. “Mix-and-diffuse serial synchrotron crystallography”. *IUCrJ* 4 (2017). DOI: 10.1107/S2052252517013124.
- [78] Jan Kern et al. “Structures of the intermediates of Kok’s photosynthetic water oxidation clock”. *Nature* 563 (2018). DOI: 10.1038/s41586-018-0681-2.
- [79] Philip Roedig et al. “High-speed fixed-target serial virus crystallography”. *Nature Methods* 14 (2017). DOI: 10.1038/nmeth.4335.
- [80] Philip Hart et al. “The CSPAD megapixel x-ray camera at LCLS”. *X-Ray Free-Electron Lasers: Beam Diagnostics, Beamline Instrumentation, and Applications*. X-Ray Free-Electron Lasers: Beam Diagnostics, Beamline Instrumentation, and Applications. Vol. 8504. International Society for Optics and Photonics, 2012. DOI: 10.1117/12.930924.
- [81] A. Allahgholi et al. “AGIPD, a high dynamic range fast detector for the European XFEL”. *Journal of Instrumentation* 10 (2015). DOI: 10.1088/1748-0221/10/01/C01023.
- [82] S. Kim. “Auto-indexing oscillation photographs”. *Journal of Applied Crystallography* 22 (1989). DOI: 10.1107/S0021889888010878.

-
- [83] Stefan Zaefferer. “New developments of computer-aided crystallographic analysis in transmission electron microscopy”. *Journal of Applied Crystallography* 33 (2000). DOI: 10.1107/S0021889899010894.
- [84] Anton Barty et al. “Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data”. *Journal of Applied Crystallography* 47 (Pt 3 2014). DOI: 10.1107/S1600576714007626.
- [85] Yaroslav Gevorkov. “Real-time image analysis and data compression in high throughput X-ray diffraction experiments”. Masters thesis. Technische Universität Hamburg-Harburg, 2016.
- [86] M. Hadian-Jazi et al. “A peak-finding algorithm based on robust statistical analysis in serial crystallography”. *Journal of Applied Crystallography* 50 (2017). DOI: 10.1107/S1600576717014340.
- [87] H. R. Powell. “The Rossmann Fourier autoindexing algorithm in MOSFLM”. *Acta Crystallographica Section D: Biological Crystallography* 55 (1999). DOI: 10.1107/S0907444499009506.
- [88] A. J. M. Duisenberg. “Indexing in single-crystal diffractometry with an obstinate list of reflections”. *Journal of Applied Crystallography* 25 (1992). DOI: 10.1107/S0021889891010634.
- [89] W. Kabsch. “Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector”. *Journal of Applied Crystallography* 21 (1988). DOI: 10.1107/S0021889888007903.
- [90] W. Kabsch. “Automatic indexing of rotation diffraction patterns”. *Journal of Applied Crystallography* 21 (1988). DOI: 10.1107/S0021889887009737.
- [91] Kenneth R. Beyerlein et al. “FELIX: an algorithm for indexing multiple crystallites in X-ray free-electron laser snapshot diffraction images”. *Journal of Applied Crystallography* 50 (Pt 4 2017). DOI: 10.1107/S1600576717007506.

- [92] H. M. Ginn et al. “TakeTwo: an indexing algorithm suited to still images with known crystal parameters”. *Acta Crystallographica Section D: Structural Biology* 72 (2016). DOI: 10.1107/S2059798316010706.
- [93] Y. Gevorkov et al. “XGANDALF – extended gradient descent algorithm for lattice finding”. *Acta Crystallographica Section A: Foundations and Advances* 75 (2019). DOI: 10.1107/S2053273319010593.
- [94] C. Li et al. “SPIND: a reference-based auto-indexing algorithm for sparse serial crystallography data”. *IUCrJ* 6 (2019). DOI: 10.1107/S2052252518014951.
- [95] A. G. W. Leslie. “The integration of macromolecular diffraction data”. *Acta Crystallographica Section D: Biological Crystallography* 62 (2006). DOI: 10.1107/S0907444905039107.
- [96] Chia-Ying Huang et al. “In situ serial crystallography for rapid de novo membrane protein structure determination”. *Communications Biology* 1 (2018). DOI: 10.1038/s42003-018-0123-6.
- [97] S. Botha et al. “De novo protein structure determination by heavy-atom soaking in lipidic cubic phase and SIRAS phasing using serial synchrotron crystallography”. *IUCrJ* 5 (Pt 5 2018). DOI: 10.1107/S2052252518009223.
- [98] Ashley C. W. Pike et al. “An overview of heavy-atom derivatization of protein crystals”. *Acta Crystallographica. Section D, Structural Biology* 72 (Pt 3 2016). DOI: 10.1107/S2059798316000401.
- [99] M. F. Perutz. “Isomorphous replacement and phase determination in non-centrosymmetric space groups”. *Acta Crystallographica* 9 (1956). DOI: 10.1107/S0365110X56002485.
- [100] Keitaro Yamashita et al. “An isomorphous replacement method for efficient de novo phasing for serial femtosecond crystallography”. *Scientific Reports* 5 (2015). DOI: 10.1038/srep14017.
- [101] Takanori Nakane et al. “Membrane protein structure determination by SAD, SIR, or SIRAS phasing in serial femtosecond crystallography using an iododetergent”. *Proceedings of the National Academy of Sciences* 113 (2016). DOI: 10.1073/pnas.1602531113.

- [102] I. Schlichting. “Experimental phasing of serial femtosecond crystallography data”. *IUCrJ* 4 (2017). DOI: 10.1107/S2052252517012167.
- [103] T. Zhang et al. “Serial crystallographic analysis of protein isomorphous replacement data from a mixture of native and derivative microcrystals”. *Acta Crystallographica Section D: Biological Crystallography* 71 (2015). DOI: 10.1107/S139900471501603X.
- [104] D. Liebschner. “Towards the automation of in situ experimental phasing”. *Acta Crystallographica Section D: Structural Biology* 76 (2020). DOI: 10.1107/S205979832001178X.
- [105] K. Nass et al. “Protein structure determination by single-wavelength anomalous diffraction phasing of X-ray free-electron laser data”. *IUCrJ* 3 (2016). DOI: 10.1107/S2052252516002980.
- [106] Alexander Gorel et al. “Two-colour serial femtosecond crystallography dataset from gadoteridol-derivatized lysozyme for MAD phasing”. *Scientific Data* 4 (2017). DOI: 10.1038/sdata.2017.188.
- [107] Steven E Ealick. “Advances in multiple wavelength anomalous diffraction crystallography”. *Current Opinion in Chemical Biology* 4 (2000). DOI: 10.1016/S1367-5931(00)00122-8.
- [108] T. Nakane et al. “Native sulfur/chlorine SAD phasing for serial femtosecond crystallography”. *Acta Crystallographica Section D: Biological Crystallography* 71 (2015). DOI: 10.1107/S139900471501857X.
- [109] Garry L. Taylor. “Introduction to phasing”. *Acta Crystallographica Section D: Biological Crystallography* 66 (Pt 4 2010). DOI: 10.1107/S0907444910006694.
- [110] Max T. B. Clabbers et al. “MyD88 TIR domain higher-order assembly interactions revealed by microcrystal electron diffraction and serial femtosecond crystallography”. *Nature Communications* 12 (2021). DOI: 10.1038/s41467-021-22590-6.
- [111] J. P. Abrahams and A. G. W. Leslie. “Methods used in the structure determination of bovine mitochondrial F1 ATPase”. *Acta Crystallographica Section D* 52 (1996). DOI: 10.1107/S0907444995008754.

- [112] T.c Terwilliger. “Statistical density modification with non-crystallographic symmetry”. *Acta Crystallographica Section D* 58 (2002). DOI: 10.1107/S0907444902016360.
- [113] E. Alharbi et al. “Comparison of automated crystallographic model-building pipelines”. *Acta Crystallographica Section D: Structural Biology* 75 (2019). DOI: 10.1107/S2059798319014918.
- [114] A. Perrakis, R. Morris, and V. S. Lamzin. “Automated protein model building combined with iterative structure refinement”. *Nature Structural Biology* 6 (1999). DOI: 10.1038/8263.
- [115] K Cowtan. “The Buccaneer software for automated model building. 1. Tracing protein chains”. *Acta Crystallographica Section D* 62 (2006). DOI: 10.1107/S0907444906022116.
- [116] T.c Terwilliger et al. “Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard”. *Acta Crystallographica Section D* 64 (2008). DOI: 10.1107/S090744490705024X.
- [117] G.m Sheldrick. “A short history of SHELX”. *Acta Crystallographica Section A* 64 (2008). DOI: 10.1107/S0108767307043930.
- [118] G.m Sheldrick. “Experimental phasing with SHELXC/D/E: combining chain tracing with density modification”. *Acta Crystallographica Section D* 66 (2010). DOI: 10.1107/S0907444909038360.
- [119] Gerrit Langer et al. “Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7”. *Nature Protocols* 3 (2008). DOI: 10.1038/nprot.2008.91.
- [120] Xavier Martinez et al. “Molecular Graphics: Bridging Structural Biologists and Computer Scientists”. *Structure* 27 (2019). DOI: 10.1016/j.str.2019.09.001.
- [121] P. Emsley and K. Cowtan. “Coot: model-building tools for molecular graphics”. *Acta Crystallographica Section D: Biological Crystallography* 60 (2004). DOI: 10.1107/S0907444904019158.
- [122] Schrödinger, LLC. “The PyMOL Molecular Graphics System, Version 1.8”. 2015.

- [123] Eric F. Pettersen et al. “UCSF Chimera—A visualization system for exploratory research and analysis”. *Journal of Computational Chemistry* 25 (2004). DOI: 10.1002/jcc.20084.
- [124] William Humphrey, Andrew Dalke, and Klaus Schulten. “VMD: Visual molecular dynamics”. *Journal of Molecular Graphics* 14 (1996). DOI: 10.1016/0263-7855(96)00018-5.
- [125] P. V. Afonine et al. “Towards automated crystallographic structure refinement with phenix.refine”. *Acta Crystallographica Section D: Biological Crystallography* 68 (2012). DOI: 10.1107/S0907444912001308.
- [126] G. N. Murshudov et al. “REFMAC5 for the refinement of macromolecular crystal structures”. *Acta Crystallographica Section D: Biological Crystallography* 67 (2011). DOI: 10.1107/S0907444911001314.
- [127] Axel T. Brünger. “Free R value: a novel statistical quantity for assessing the accuracy of crystal structures”. *Nature* 355 (1992). DOI: 10.1038/355472a0.
- [128] P Andrew Karplus and Kay Diederichs. “Assessing and maximizing data quality in macromolecular crystallography”. *Current opinion in structural biology* 34 (2015). DOI: 10.1016/j.sbi.2015.07.003.
- [129] P. Andrew Karplus and Kay Diederichs. “Linking Crystallographic Model and Data Quality”. *Science* 336 (2012). DOI: 10.1126/science.1218231.
- [130] Greta Assmann, Wolfgang Brehm, and Kay Diederichs. “Identification of rogue datasets in serial crystallography”. *Journal of Applied Crystallography* 49 (2016). DOI: 10.1107/S1600576716005471.
- [131] Rob W.W. Hooft, Chris Sander, and Gerrit Vriend. “Objectively judging the quality of a protein structure from a Ramachandran plot”. *Bioinformatics* 13 (1997). DOI: 10.1093/bioinformatics/13.4.425.
- [132] A. Gorel, I. Schlichting, and T. R. M. Barends. “Discerning best practices in XFEL-based biological crystallography – standards for nonstandard experiments”. *IUCrJ* 8 (2021). DOI: 10.1107/S205225252100467X.

-
- [133] A Shilova et al. “Current status and future opportunities for serial crystallography at MAX IV Laboratory”. *Journal of Synchrotron Radiation* 27 (2020). DOI: 10.1107/S1600577520008735.
- [134] Jose M. Martin-Garcia. “Protein Dynamics and Time Resolved Protein Crystallography at Synchrotron Radiation Sources: Past, Present and Future”. *Crystals* 11 (2021). DOI: 10.3390/cryst11050521.
- [135] D. K. Schneider et al. “FMX – the Frontier Microfocusing Macromolecular Crystallography Beamline at the National Synchrotron Light Source II”. *Journal of Synchrotron Radiation* 28 (2021). DOI: 10.1107/S1600577520016173.
- [136] A. Meents et al. “Pink-beam serial crystallography”. *Nature Communications* 8 (2017). DOI: 10.1038/s41467-017-01417-3.
- [137] Y. Gevorkov et al. “pinkIndexer – a universal indexer for pink-beam X-ray and electron diffraction snapshots”. *Acta Crystallographica Section A: Foundations and Advances* 76 (2020). DOI: 10.1107/S2053273319015559.
- [138] Kyuseok Yun et al. “Coherence and pulse duration characterization of the PAL-XFEL in the hard X-ray regime”. *Scientific Reports* 9 (2019). DOI: 10.1038/s41598-019-39765-3.
- [139] Wojciech Roseker et al. “Double-pulse speckle contrast correlations with near Fourier transform limited free-electron laser light using hard X-ray split-and-delay”. *Scientific Reports* 10 (2020). DOI: 10.1038/s41598-020-61926-y.
- [140] Karol Nass et al. “Structural dynamics in proteins induced by and probed with X-ray free-electron laser pulses”. *Nature Communications* 11 (2020). DOI: 10.1038/s41467-020-15610-4.
- [141] Felix Lehmkuhler et al. “Emergence of anomalous dynamics in soft matter probed at the European XFEL”. *Proceedings of the National Academy of Sciences* 117 (2020). DOI: 10.1073/pnas.2003337117.

-
- [142] Alexander Kozlov et al. “Hybrid Plasma/Molecular-Dynamics Approach for Efficient XFEL Radiation Damage Simulations”. *Crystals* 10 (2020). DOI: 10.3390/cryst10060478.
- [143] Claudiu A. Stan et al. “Liquid explosions induced by X-ray laser pulses”. *Nature Physics* 12 (2016). DOI: 10.1038/nphys3779.
- [144] Gabriel Blaj et al. “Generation of high-intensity ultrasound through shock propagation in liquid jets”. *Physical Review Fluids* 4 (2019). DOI: 10.1103/PhysRevFluids.4.043401.
- [145] Daniel Ursescu et al. “Generation of shock trains in free liquid jets with a nanosecond green laser”. *Physical Review Fluids* 5 (2020). DOI: 10.1103/PhysRevFluids.5.123402.
- [146] Marie L. Grünbein et al. “Observation of shock-induced protein crystal damage during megahertz serial femtosecond crystallography”. *Physical Review Research* 3 (2021). DOI: 10.1103/PhysRevResearch.3.013046.
- [147] Claudiu Andrei Stan et al. “The Magnitude and Waveform of Shock Waves Induced by X-ray Lasers in Water”. *Applied Sciences* 10 (2020). DOI: 10.3390/app10041497.
- [148] Alexander Gorel et al. “Shock Damage Analysis in Serial Femtosecond Crystallography Data Collected at MHz X-ray Free-Electron Lasers”. *Crystals* 10 (2020). DOI: 10.3390/cryst10121145.
- [149] Tokushi Sato et al. “Femtosecond timing synchronization at megahertz repetition rates for an x-ray free-electron laser”. *Optica* 7 (2020). DOI: 10.1364/OPTICA.396728.
- [150] M. L. Shelby et al. “A fixed-target platform for serial femtosecond crystallography in a hydrated environment”. *IUCrJ* 7 (2020). DOI: 10.1107/S2052252519014003.
- [151] Mohammad Vakili et al. “Microfluidic polyimide gas dynamic virtual nozzles for serial crystallography”. *Review of Scientific Instruments* 91 (2020). DOI: 10.1063/5.0012806.

- [152] Austin Echelmeier et al. “Segmented flow generator for serial crystallography at the European X-ray free electron laser”. *Nature Communications* 11 (2020). DOI: 10.1038/s41467-020-18156-7.
- [153] Henry N. Chapman et al. “Femtosecond X-ray protein nanocrystallography”. *Nature* 470 (2011). DOI: 10.1038/nature09750.
- [154] Jason Tenboer et al. “Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein”. *Science* 346 (2014). DOI: 10.1126/science.1259357.
- [155] Allen M Orville. “Recent results in time resolved serial femtosecond crystallography at XFELs”. *Current Opinion in Structural Biology* 65 (2020). DOI: 10.1016/j.sbi.2020.08.011.
- [156] Haiwang Yong et al. “Observation of the molecular response to light upon photoexcitation”. *Nature Communications* 11 (2020). DOI: 10.1038/s41467-020-15680-4.
- [157] Juraj Knoška et al. “Ultracompact 3D microfluidics for time-resolved structural biology”. *Nature Communications* 11 (2020). DOI: 10.1038/s41467-020-14434-6.
- [158] Robert Dods et al. “Ultrafast structural changes within a photosynthetic reaction centre”. *Nature* 589 (2021). DOI: 10.1038/s41586-020-3000-7.
- [159] Samuel L. Rose et al. “An unprecedented insight into the catalytic mechanism of copper nitrite reductase from atomic-resolution and damage-free structures”. *Science Advances* 7 (2021). DOI: 10.1126/sciadv.abd8523.
- [160] Eriko Nango et al. “A three-dimensional movie of structural changes in bacteriorhodopsin”. *Science* 354 (2016). DOI: 10.1126/science.aah3497.
- [161] Przemyslaw Nogly et al. “Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser”. *Science* 361 (2018). DOI: 10.1126/science.aat0094.
- [162] Petr Skopintsev et al. “Femtosecond-to-millisecond structural changes in a light-driven sodium pump”. *Nature* 583 (2020). DOI: 10.1038/s41586-020-2307-8.

- [163] Ji-Hye Yun et al. “Early-stage dynamics of chloride ion-pumping rhodopsin revealed by a femtosecond X-ray laser”. *Proceedings of the National Academy of Sciences* 118 (2021). DOI: 10.1073/pnas.2020486118.
- [164] Mohamed Ibrahim et al. “Untangling the sequence of events during the S2 \rightarrow S3 transition in photosystem II and implications for the water oxidation mechanism”. *Proceedings of the National Academy of Sciences* 117 (2020). DOI: 10.1073/pnas.2000529117.
- [165] Elin Claesson et al. “The primary structural photoresponse of phytochrome proteins captured by a femtosecond X-ray laser”. *eLife* 9 (2020). Ed. by Werner Kühlbrandt and Cynthia Wolberger. DOI: 10.7554/eLife.53514.
- [166] Melissa Carrillo et al. “High-resolution crystal structures of transient intermediates in the phytochrome photocycle”. *Structure* 29 (2021). DOI: 10.1016/j.str.2021.03.004.
- [167] Johan Bielecki, Filipe R. N. C. Maia, and Adrian P. Mancuso. “Perspectives on single particle imaging with x rays at the advent of high repetition rate x-ray free electron laser sources”. *Structural Dynamics* 7 (2020). DOI: 10.1063/4.0000024.
- [168] D. Assalauova et al. “An advanced workflow for single-particle imaging with the limited data at an X-ray free-electron laser”. *IUCrJ* 7 (2020). DOI: 10.1107/S2052252520012798.
- [169] Kartik Ayyer et al. “3D diffractive imaging of nanoparticle ensembles using an x-ray laser”. *Optica* 8 (2021). DOI: 10.1364/OPTICA.410851.
- [170] Tomas Ekeberg et al. “Three-Dimensional Reconstruction of the Giant Mimivirus Particle with an X-Ray Free-Electron Laser”. *Physical Review Letters* 114 (2015). DOI: 10.1103/PhysRevLett.114.098102.
- [171] I. V. Lundholm et al. “Considerations for three-dimensional image reconstruction from experimental data in coherent diffractive imaging”. *IUCrJ* 5 (2018). DOI: 10.1107/S2052252518010047.

- [172] Ruslan P. Kurta et al. “Correlations in Scattered X-Ray Laser Pulses Reveal Nanoscale Structural Features of Viruses”. *Physical Review Letters* 119 (2017). DOI: 10.1103/PhysRevLett.119.158102.
- [173] Max Rose et al. “Single-particle imaging without symmetry constraints at an X-ray free-electron laser”. *IUCrJ* 5 (2018). DOI: 10.1107/S205225251801120X.
- [174] Akihiro Suzuki et al. “Design of a liquid cell toward three-dimensional imaging of unidirectionally-aligned particles in solution using X-ray free-electron lasers”. *Physical Chemistry Chemical Physics* 22 (2020). DOI: 10.1039/C9CP03658J.
- [175] Oleksandr Yefanov et al. “Evaluation of serial crystallographic structure determination within megahertz pulse trains”. *Structural Dynamics* 6 (2019). DOI: 10.1063/1.5124387.
- [176] T. A. White et al. “CrystFEL: a software suite for snapshot serial crystallography”. *Journal of Applied Crystallography* 45 (2012). DOI: 10.1107/S0021889812002312.
- [177] Luke A. J. O’Neill and Andrew G. Bowie. “The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling”. *Nature Reviews Immunology* 7 (2007). DOI: 10.1038/nri2079.
- [178] Luke A. J. O’Neill. “Targeting signal transduction as a strategy to treat inflammatory diseases”. *Nature Reviews Drug Discovery* 5 (2006). DOI: 10.1038/nrd2070.
- [179] Luke AJ O’Neill. “Therapeutic targeting of Toll-like receptors for inflammatory and infectious diseases”. *Current Opinion in Pharmacology* 3 (2003). DOI: 10.1016/S1471-4892(03)00080-8.
- [180] Jeffrey D. Nanson, Bostjan Kobe, and Thomas Ve. “Death, TIR, and RHIM: Self-assembling domains involved in innate immunity and cell-death signaling”. *Journal of Leukocyte Biology* 105 (2019). DOI: 10.1002/JLB.MR0318-123R.
- [181] Thomas Ve et al. “Structural basis of TIR-domain-assembly formation in MAL- and MyD88-dependent TLR4 signaling”. *Nature Structural & Molecular Biology* 24 (2017). DOI: 10.1038/nsmb.3444.

- [182] Hidenori Ohnishi et al. “Structural basis for the multiple interactions of the MyD88 TIR domain in TLR4 signaling”. *Proceedings of the National Academy of Sciences* 106 (2009).
- [183] Greg A. Snyder et al. “Molecular mechanisms for the subversion of MyD88 signaling by TcpC from virulent uropathogenic *Escherichia coli*”. *Proceedings of the National Academy of Sciences* 110 (2013).
- [184] Lan Zhu et al. “Serial Crystallography for Structure-Based Drug Discovery”. *Trends in Pharmacological Sciences* 41 (2020). DOI: 10.1016/j.tips.2020.08.009.
- [185] P. Roedig et al. “A micro-patterned silicon chip as sample holder for macromolecular crystallography experiments with minimal background scattering”. *Scientific Reports* 5 (2015). DOI: 10.1038/srep10451.
- [186] Philip Roedig et al. “Room-temperature macromolecular crystallography using a micro-patterned silicon chip with minimal background scattering”. *Journal of Applied Crystallography* 49 (2016). DOI: 10.1107/S1600576716006348.
- [187] Connie Darmanin et al. “Protein crystal screening and characterization for serial femtosecond nanocrystallography”. *Scientific Reports* 6 (2016). DOI: 10.1038/srep25345.
- [188] V Mariani et al. “OnDA: online data analysis and feedback for serial X-ray imaging”. *Journal of Applied Crystallography* 49 (2016). DOI: 10.1107/S1600576716007469.
- [189] Thomas A. White et al. “Recent developments in CrystFEL”. *Journal of Applied Crystallography* 49 (2016). DOI: 10.1107/S1600576716004751.
- [190] Thomas A. White. “Processing serial crystallography data with it CrystFEL: a step-by-step guide”. *Acta Crystallographica Section D* 75 (2019). DOI: 10.1107/S205979831801238X.
- [191] Martin Caffrey and Vadim Cherezov. “Crystallizing membrane proteins using lipidic mesophases”. *Nature Protocols* 4 (2009). DOI: 10.1038/nprot.2009.31.
- [192] M. Caffrey. “A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes”. *Acta Crystallographica Section F: Structural Biology Communications* 71 (2015). DOI: 10.1107/S2053230X14026843.

- [193] Sébastien Boutet and Garth J. Williams. “The Coherent X-ray Imaging (CXI) instrument at the Linac Coherent Light Source (LCLS)”. *New Journal of Physics* 12 (2010). DOI: 10.1088/1367-2630/12/3/035024.
- [194] D. Damiani et al. “Linac Coherent Light Source data analysis using psana”. *Journal of Applied Crystallography* 49 (2016). DOI: 10.1107/S1600576716004349.
- [195] J. Thayer et al. “Data systems for the Linac coherent light source”. *Advanced Structural and Chemical Imaging* 3 (2017). DOI: 10.1186/s40679-016-0037-7.
- [196] Daniele de Sanctis et al. “ID29: a high-intensity highly automated ESRF beamline for macromolecular crystallography experiments exploiting anomalous scattering”. *Journal of Synchrotron Radiation* 19 (2012). DOI: 10.1107/S0909049512009715.
- [197] A. S. Soares et al. “Solvent minimization induces preferential orientation and crystal clustering in serial micro-crystallography on micro-meshes, in situ plates and on a movable crystal conveyor belt”. *Journal of Synchrotron Radiation* 21 (2014). DOI: 10.1107/S1600577514017731.
- [198] Aina E. Cohen et al. “Goniometer-based femtosecond crystallography with X-ray free electron lasers”. *Proceedings of the National Academy of Sciences of the United States of America* 111 (2014). DOI: 10.1073/pnas.1418733111.
- [199] S. Oghbaey et al. “Fixed target combined with spectral mapping: approaching 100% hit rates for serial crystallography”. *Acta Crystallographica Section D: Structural Biology* 72 (2016). DOI: 10.1107/S2059798316010834.
- [200] J. Lieske et al. “On-chip crystallization for serial crystallography experiments and on-chip ligand-binding studies”. *IUCrJ* 6 (2019). DOI: 10.1107/S2052252519007395.
- [201] L. a. J. O’Neill. “The role of MyD88-like adapters in Toll-like receptor signal transduction”. *Biochemical Society Transactions* 31 (2003). DOI: 10.1042/bst0310643.
- [202] Aisling Dunne et al. “Structural Complementarity of Toll/Interleukin-1 Receptor Domains in Toll-like Receptors and the Adaptors Mal and MyD88”. *Journal of Biological Chemistry* 278 (2003). DOI: 10.1074/jbc.M301742200.

-
- [203] Katherine A. Fitzgerald et al. “Mal (MyD88-adaptor-like) is required for Toll-like receptor-4 signal transduction”. *Nature; London* 413 (2001). DOI: 10.1038/35092578.
- [204] Ailís O’Carroll et al. “Pathological mutations differentially affect the self-assembly and polymerisation of the innate immune system signalling adaptor molecule MyD88”. *BMC Biology* 16 (2018). DOI: 10.1186/s12915-018-0611-7.
- [205] Masahiro Yamamoto et al. “Role of Adaptor TRIF in the MyD88-Independent Toll-Like Receptor Signaling Pathway”. *Science* 301 (2003). DOI: 10.1126/science.1087262.
- [206] A. M. Wolff et al. “Comparing serial X-ray crystallography and microcrystal electron diffraction (MicroED) as methods for routine structure determination from small macromolecular crystals”. *IUCrJ* 7 (2020). DOI: 10.1107/S205225252000072X.
- [207] Richard Neutze and Keith Moffat. “Time-resolved structural studies at synchrotrons and X-ray free electron lasers: opportunities and challenges”. *Current Opinion in Structural Biology. Carbohydrates and glycoconjugates ? Biophysical methods* 22 (2012). DOI: 10.1016/j.sbi.2012.08.006.
- [208] George D. Calvey et al. “Mixing injector enables time-resolved crystallography with high hit rate at X-ray free electron lasers”. *Structural Dynamics* 3 (2016). DOI: 10.1063/1.4961971.