# Structural characterisation of ligand, redox and catalytic states in haem enzymes using *in crystallo* spectroscopies and serial crystallography

Tadeo Moreno-Chicano

A thesis submitted for the degree of

Doctor of Philosophy

School of Biological Sciences

University of Essex

Date of submission 31/08/2018

# Abstract

X-ray crystallography is a powerful tool for the study of biomolecules and has determined for instance the structures of numerous enzymes in order to unveil their conformation, interactions with substrates and ligands, and better understand their function. However catalytic processes carried out by enzymes can be of great complexity, involving several steps and intermediates prior to arriving to the final reaction product. Therefore, they are difficult to structurally characterise, since one of the limitations of X-ray crystallography is that provides static 'snapshots' of the average conformation within the crystal. Another important limitation is established by radiation damage, which results from the interaction of the X-ray beam with the protein crystal, and results in photo-reduced species which are not fully representative of the state of the measured protein. This is particularly relevant for metalloproteins such as haem proteins, since their metal centres are prone to reduction upon irradiation.

In this research project, I aimed to overcome these limitations to study four haem enzymes, DtpA, DtpAa, DHP and NOD; and their catalytic cycles, trying to obtain intact species of relevant functional states. To this aim I have used complementary tools such as single-crystal spectroscopies, and the emerging methodologies of serial crystallography, applied at XFEL and synchrotron sources. Spectroscopic information of the crystals provided by *in crystallo* spectroscopies was used to validate the structures obtained by X-ray

crystallography, identify generated species, and assign them within the enzyme catalytic mechanism (e.g. peroxidase cycle).

Serial crystallography methodologies were used to measure room-temperature damage-free structures at XFEL sources of our protein targets, such as the resting state Fe(III). The application of serial crystallography at synchrotron sources, delivered low-dose structures of the haem enzymes, and allowed to perform dose-series measurements with the intention to follow radiation-induced processes at redox metal sites of haem enzymes.

# Acknowledgements

First of all, I would like to specially thank Dr. Mike Hough for giving me the opportunity to develop this project and guiding me during these three years to overcome the challenges that the project presented, it has been a great time here at Essex where I feel that I have developed as a scientist. I am also thankful to the Leverhulme Trust for the funding that allowed the conduction of this research, together with the postdoc position held by Dr. Demet Kekilli, who was a great help and inspiration during the first half of the PhD, and whom I am deeply thankful and hope to see at some other point of my career. Special thanks also to Dr. Richard Strange who also guided me during this time and for many interesting conversations; and Dr. Jonathan Worrall who was also a crucial guidance and support in the DyPs project and gave me the confidence in my work, for which I will be always thankful, and for many other enjoyable moments. I would also like to thank Prof. Mike Wilson for his contagious passion for science and for many insightful conversations about the project. In the same way, I thank all the staff in the department, which build up an invaluable community atmosphere where one can work comfortably. Of course, I would also like to thank to all my work colleagues in lab 5.08, past and present, specially to Megan, for her friendship and so many late hours in the lab working and talking and making time pass smoothly; Karl, for being a great flat mate and friend and support during these years; Oliver, for all the coffees and conversations with different views ☺; Amanda, for all the help with many things and the English breakfasts!; and also Ali, for sharing with me good and tough

moments, and exciting trips to Japan and Chicago. I would specially like to thank Luisa, for her crucial support to get me through these final months, and to Sophie, for all those amazing years and for always believing in me.

A big thank you also to the team at Diamond for the unbelievable collaboration with the chips and a great working atmosphere, which allowed the project to develop enormously. I am also grateful to the teams at APS, Hamburg, SLS, ESRF, and the Japanese team at SACLA, that made possible our work and gave it another dimension with exciting results.

I would like to specially thank my parents, Victoria y Antonio, who have always supported and believed in me and without them I could have never reached this goal. Of course, also to my family for all their loving support. And I would like to thank all my friends back home - els xais! - whom I have missed a lot (and I still do), and gave me the real measure of what is worth in life. Finally, I would like to dedicate this work to Milhouse, for his unconditional friendship and who will always be remembered.

# Author's declaration

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

# Abreviations

NOD – Nitric oxide dioxygenase

DHP – Dehaloperoxidase

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

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

SCS – Single crystal spectroscopy

SX – Serial crystallography

SFX – Serial Femtosecond crystallography

SSX – Synchrotron Serial crystallography

SLX – Serial Laue crystallography

PDB – Protein Data Bank

Cyt c' – cytochrome c prime

Leu – leucine

Arg – arginine

His – histidine

Cys - cysteine

NO – nitric oxide

CO – carbon monoxide

PCR – polymerase chain reaction

SDS-PAGE – Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis.

RR – resonance Raman

SCRR – Single-crystal resonance Raman

UV-vis – Ultraviolet-Visible

IR – Infrared

FT – Fourier transform

MX – Macromolecular Crystallography

SLS – Swiss Light Source

DLS – Diamond Light Source

ESRF – European Synchrotron Radiation Facility

XFEL – X-ray Free-Electron Laser

Kan – kanamycin

Amp – ampicillin

MES – 2-(N-morpholino)ethanesulfonic acid

NaCl – sodium chloride

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

5-dALA – aminolevulinic acid

PAA – peracetic acid

IPTG – isopropyl β-D-1-thiogalactopyranoside

LB – Luria-Bertani broth

SDS – sodium dodecyl sulphate

$MgCl_2$ – magnesium chloride

PEG – Polyethelene Glycol

DMSO – Dimethyl Sulfoxide

TRIS - Thrisaminomethane

DEAE – diethylaminoethyl

pI – isoelectric point

rpm – revolutions per minute

kDa – kilodalton

V - Voltage

mW – milliWatt

RMSD – Root Mean Square Deviation

# Chapter 1 Introduction

## 1.1 Introduction to Haem proteins

Haem proteins are a vast and varied family of proteins, which perform a myriad of functions, all with the common characteristic of containing at least one haem cofactor. The presence of the haem group, allows the protein to carry out different roles vital to many cellular processes, and hence the widespread of haemproteins among organisms with diverse phylogenetic origin. They are probably the most studied family of proteins (Reedy et al. 2008), and the first ones to be structurally characterized by X-ray crystallography, i.e. myoglobin and haemoglobin structures, with the known functions of oxygen storage and transport in mammals. The haem chromophore gives to haem proteins their characteristic reddish or brown colour, depending on the oxidation state and binding environment of the cofactor.

### *1.1.1 The haem cofactor*

The haem prosthetic group is an aromatic polycyclic molecule, known as protoporphyrin IX, whose structure backbone is formed by four pyrrole molecules (5-carbon small aromatic molecule) fused together by methylene bridges around a central iron, which they coordinate through each of their $N_1$'s, acting as a tetradentate ligand. When a tetrapyrrole binds a metal is called a porphyrin, and thus haem is an iron-bound porphyrin, whose chemical structure is shown in Figure 1.1.

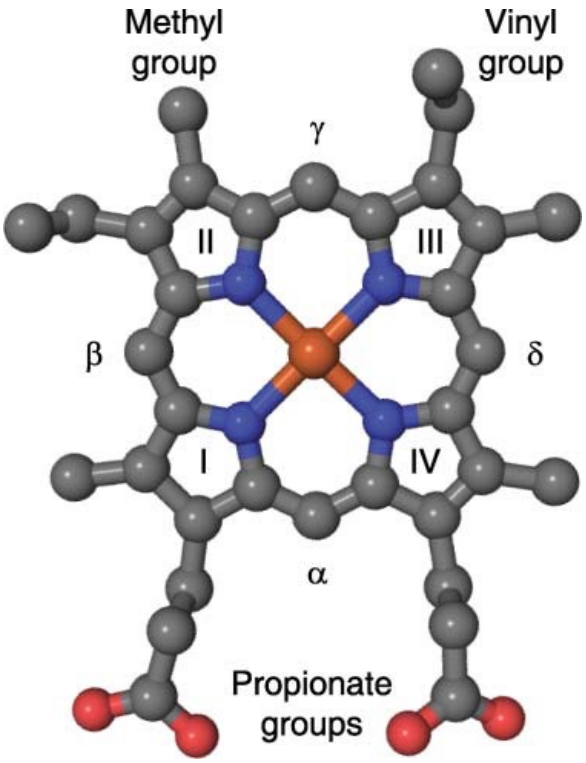


**Figure 1.1:** Haem cofactor (B form), where its backbone formed by four pyrroles (I, II, III, IV) can be appreciated coordinating a central iron ion (in brown) in the centre. Haem ring substituents such as methyl, vinyl and propionate groups can be observed, as well as the haem edges α, β, γ, and δ. (Schneider et al., 2007).

There are different types of haem prosthetic groups, such as a, b, c, d, f, and o types. The first four types a-d are illustrated in Figure 1.2. Of all of them, haem b and haem c represent the most common forms found associated with proteins in living organisms (Reedy et al. 2008). The different haem types differ in the substituents that are found in the outer part of the ring, giving each of them its own functional characteristics. The diversity in functions carried out by haems is accomplished not only by the nature of the cofactor itself, but also by the protein environment surrounding it, the axial ligands coordinating the iron, and the relative accessibility of solvent to the haem site (Schneider et al. 2007; Bowman & Bren 2008). The catalytic function of haem is further extended

by the synergy with other cofactors, such as Flavin, NAD, NADP, and other metal ions (Chapman, Daff & Munro, 1997).

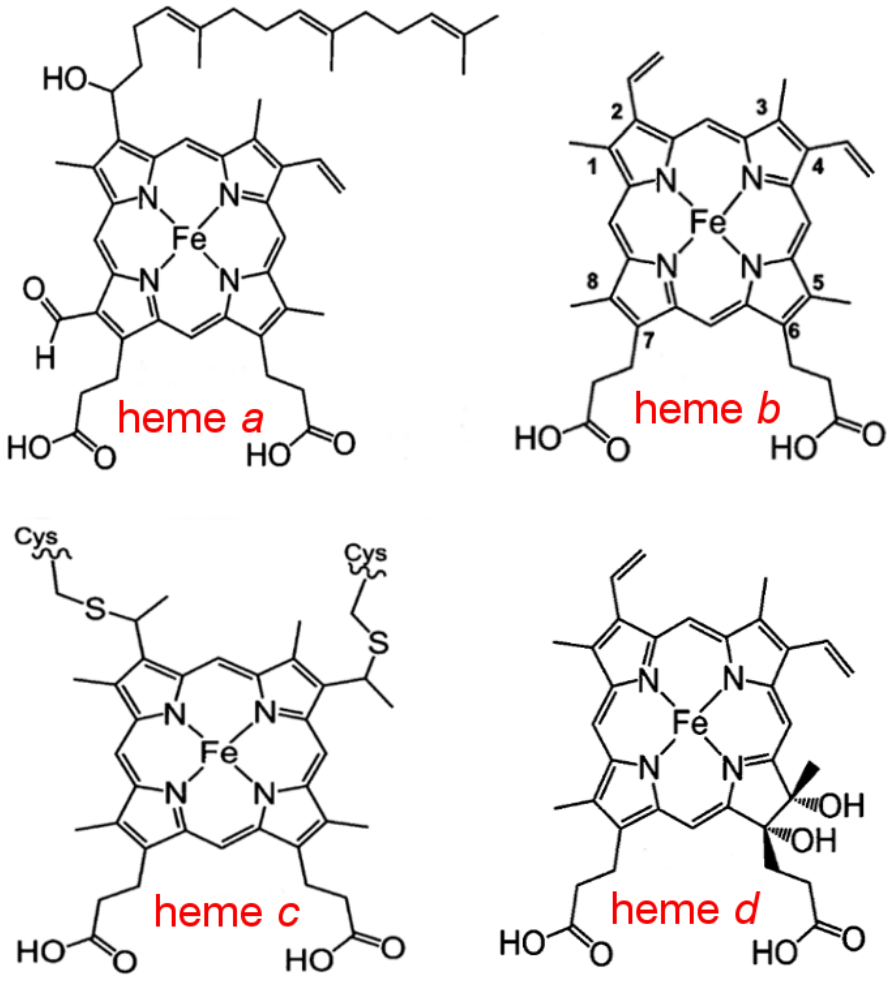


**Figure 1.2:** Haem cofactors a, b, c, and d type, found in haem proteins. (adapted from Liu et al., 2014)

Haem *b* is the precursor of haem a and c. It has four methyl substituents at positions 1, 3, 5, and 8; two vinyl groups at positions 2 and 4, and two propionate groups at positions 6 and 7. Haem b is non-covalently attached to the proteins where is found, which include globins, catalases, peroxidases and cytochrome $b_5$ (Paoli et al. 2002). Heam *a*, has the vinyl group at position 2 replaced by a hydrophobic hydroxyethylfarnesyl side chain, and the methyl

group at position 8 is oxidised to a formyl group, which makes the haem, in overall, more hydrophobic (Liu et al., 2014). Haem *c* is the only kind which is attached to the protein polypeptide though covalent bonds, normally established between the vinyl groups at positions 2 and 4, and Cys residues from the protein, resulting in the formation of thioether bonds. In proteins containing a haem c cofactor (e.g. cytochrome Cs and C's), the Cys residues are found in the highly-conserved motif CXXCH, in which the His residue corresponds to the axial ligand coordinating the haem iron (Paoli et al. 2002).

### *1.1.2 Functional roles of haem proteins.*

Haem proteins are metalloproteins which carry out many different functions and are involved in key cellular processes. As discussed, the functional diversity of haem proteins is determined by the variability of the haem group, how it is bound to the polypeptide chain, and the haem environment and protein fold. Roles carried out by haem proteins include:

- Transport and storage of oxygen ($O_2$) – performed by e.g. haemoglobin, myoglobin, neuroglobin.
- Electron transfer functions – performed by cytochromes a, b and c, found in the electron transport chain.
- Gas-sensor system – some haemproteins are involved in the sensory system of $O_2$, CO and NO.
- Catalysis as enzymes – e.g. peroxidases, catalases, cytochrome P450s, cytochrome c oxidase, ligninases.

When functioning as enzymes, some haem proteins make use of the redox properties of the haem iron to activate oxygen, e.g. $O_2$ or $H_2O_2$, to oxidise a variety of substrates. Despite their high variability, all haem enzymes performing this reaction do it following a similar catalytic mechanism (Moody & Raven 2018), with the generation of ferryl intermediates (iron(IV)-oxo), named Compound I and II, which will be presented in the next section.

### *1.1.3 Haem peroxidases.*

Peroxidases are a large family of enzymes widespread among all living organisms and involved in many biological processes, which use mainly hydrogen peroxide ($H_2O_2$) as electron acceptor to catalyse the oxidation of a broad range of substrates. They can be classified in two main families, haem and non-haem peroxidases (Figure 1.3), depending on whether they harbour this iron cofactor in order to perform their oxidative reactions. Non-haem peroxidases typically use a redox-active cysteine or seleno-cysteine residue for this matter.

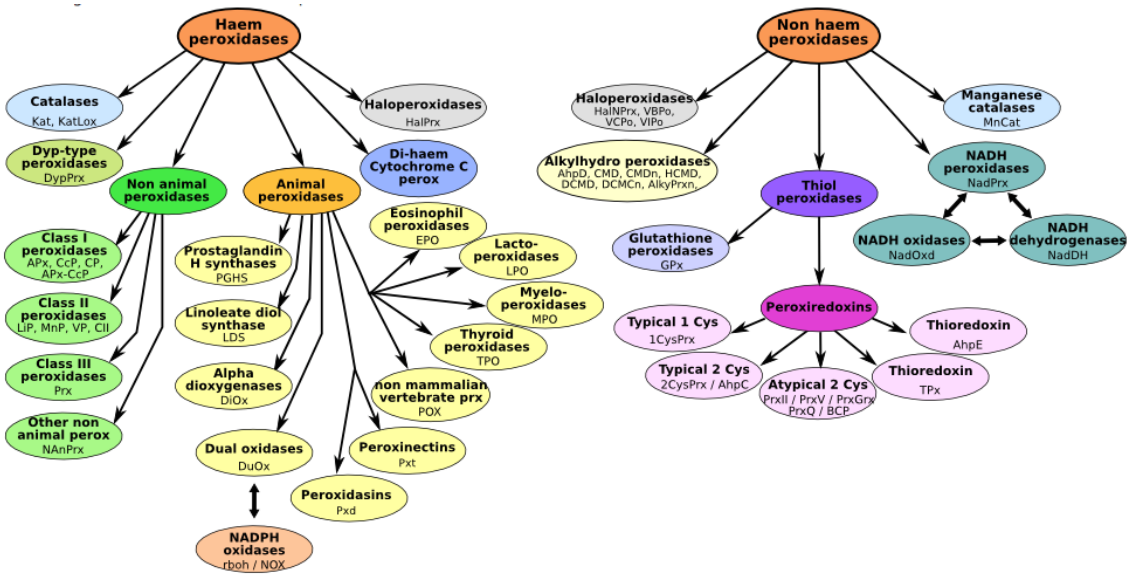


**Figure 1.3:** Classification of families of peroxidases, where haem peroxidases constitute one of the main groups (from the PeroxiBase; http://peroxibase.toulouse.inra.fr).

Haem peroxidases contain at least one haem cofactor where catalysis normally takes place, and they are grouped into plant and animal peroxidase superfamilies (Wellinder, 1992). The plant peroxidases family includes peroxidases from plants, fungi and bacteria, and can be further sub-divided depending if catalysis takes place intracellularly or they are secreted to carry out their function extracellularly. They are well-studied enzymes due to their inherent spectroscopic properties and their ability to degrade a wide range of substrates, and its subsequent biotechnological applications (Christian et al., 2003; Ollikka et al., 1993; Robinson et al., 2001). One of the most known members of this family, Horseradish peroxidase (HRP), has been studied for more than 200 years (Planche, 1810). Structurally, they all present a fold consisting of all $\alpha$-helices where the haem group is embedded (e.g. globin fold). Nevertheless, an exception was found in the latest discovered family, the so-called Dye-decolourising peroxidases, which were found to adopt a ferredoxin-

like fold, with α-helix and β-sheet motifs, and present differences in homology and reactivity with classic peroxidases (Sugano 2009).

### *1.1.3.1 The peroxidase cycle*

The catalytic mechanism carried out at the active site of haem peroxidases is known as the peroxidase cycle, schematized in Figure 1.4. It starts with the ferric resting state Fe(III), which is oxidised by hydrogen peroxide, by two electron equivalents, resulting in the reactive intermediate Compound I (Cmd-I), an iron(IV)-oxo, or ferryl, haem radical cation species. The formed radical can be located in the porphyrin ring or in a neighbouring protein residue (e.g. Tyr, Trp). In classic peroxidases, the formation of Cmd-I is catalysed by a conserved distal His, through the Poulos-Kraut mechanism (Figure 1.5; Poulos & Kraut, 1980), upon binding of hydrogen peroxide to the haem iron forming the species Compound 0 (Cmd-0). Compound I is then sequentially reduced in two steps, firstly accepting an electron from a substrate to form Compound II, an ferryl neutral haem species. Compound II accepts a further electron from the substrate to return to the initial resting state, closing the cycle. The nature of the substrate can be very diverse and depends on the specificity of each peroxidase.

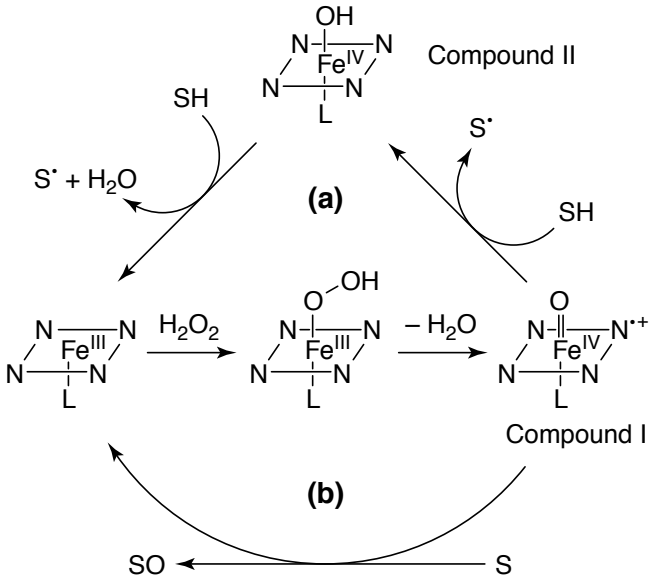


**Figure 1.4:** Catalytic cycle of heme peroxidases: a) one-electron oxidation (classical peroxidase reaction); and b) two-electron oxidation (oxygen transfer). Abbreviations: S, SH, substrate; SO, oxidized substrate (Velde et al. 2001).

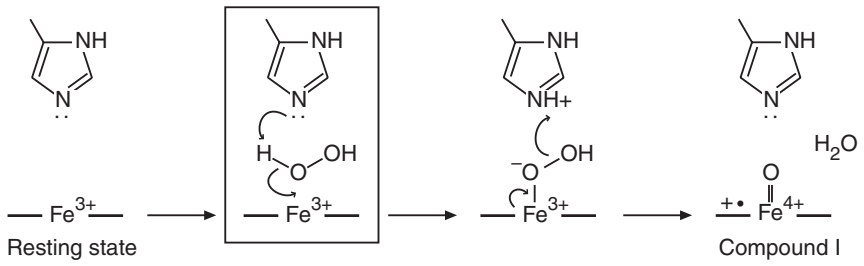


**Figure 1.5:** The Poulos-Kraut mechanism of formation of compound I in haem peroxidases, catalysed by a distal histidine from the protein. (Yoshida et al. 2011)

The nature of the two ferryl intermediates found in haem peroxidases – and in haem enzymes in general (e.g. cytochromes P450s, catalases) –, has been studied for decades since they were discovered. Compound I was firstly observed in Horseradish peroxidase (HRP) by Theorell (1941), and it normally has a green colour. Compound II instead, has a brown appearance and was discovered by Keillin and Mann (1937). One of the main aspects of debate, was

to know whether the ferryl species was protonated, and thus was better described as a double (Fe(IV)=O) or a single bond (Fe(IV)-OH). The use of X-ray crystallography (at 100 K) coupled with spectroscopy used by Berglund et al., (2002), provided a first validated observation of the length of the bond in both intermediates for HRP, which were both described as a double bond at 1.7 Å. Later studies from Gumiero and co-workers (Gumiero et al., 2011) questioned this, since they showed crystal structures (also at 100 K) for both intermediates in ascorbate peroxidase (APX) and Cytochrome c peroxidase (CcP), where the bond length increased in compound II compared to compound I, suggesting its protonation. Recent studies by Kwon et al. (2016), confirmed the protonation of compound II in APX by using neutron diffraction crystallography with crystals in this state. Also by neutron diffraction, Compound I had been shown not to be protonated by Casadei et al (2014), and thus described as a double bond Fe(IV)=O. Recently, the use of X-ray free electron laser (XFEL) sources has opened the possibility of obtaining damage-free intact structures, very relevant in this case since ferryl intermediates are especially prone to photo-reduction. The structure for the Compound I state of CcP was determined (Chreifi et al. 2016), confirming also the double bond of the first ferryl intermediate in the peroxidase cycle.

## 1.2 Radiation damage in protein crystals.

Radiation damage is a known (Nave 1995; Garman 2010; Holton 2009) and unavoidable consequence of using ionizing X-ray radiation to obtain

diffraction data from crystals in order to solve their structure. It is the consequence of the radiation chemistry within the crystal upon interaction with the X-ray beam.

### *1.2.1 Radiation chemistry*

X-ray photons can interact with protein crystals in three ways: i) total absorption of the photon by the photoelectric effect, with ejection of a high-energy photoelectron from the inner shell of the absorbing atom; ii) inelastic Compton scattering of the incident photon, with energy loss to an atomic electron, that can also be ejected; and iii) elastic Thomson (Rayleigh) scattering, with no loss of energy by the incident photon. The latter is the cause of the observed X-ray diffraction pattern. The first two effects are the cause of radiation damage observed in protein crystals. Each ejected photoelectron has enough energy to subsequently induce up to 500 further ionization events, which results in the formation of radical species and the production of solvated electrons within the crystal (Garman, 2010). Since protein crystals contain a high amount of solvent (typically ~50%), the radiolysis of water and other components forming the solvent highly contribute to the formation of these species, in this case hydroxyl radicals and solvated electrons. A scheme showing the reactions describing the photolysis of water (Ward 1988) is shown in Figure 1.6.

The solvated photo-electrons initiate photo-chemical processes that lead to chemical modifications and the formation of free radicals within the crystal (Carugo & Carugo 2005). Free radicals can initiate photochemical processes

and chemically alter the molecules in the solvent or in the protein, increasing the disorder and damaging the crystal. When cryo-temperatures are used to mitigate radiation damage, the diffusion rate of free radicals slows down and this translates into an increase of the lifespan of the crystal (Garman, 2003).

$$\mathrm{H_2O} \xrightarrow{\text{ionizing radiation}} \mathrm{H_2O^{+\bullet}} + \mathrm{e^-}\ \text{(ionization)}$$
$$\mathrm{H_2O} \xrightarrow{\text{ionizing radiation}} \mathrm{H_2O^*}\ \text{(electronic ionization)}$$
$$\mathrm{H_2O^{+\bullet}} + \mathrm{H_2O} \longrightarrow \mathrm{H_3O^+} + {}^{\bullet}\mathrm{OH}$$
$$\mathrm{e^-} + n\mathrm{H_2O} \longrightarrow \mathrm{e^-_{aq}}$$
$$\mathrm{H_2O^*} \longrightarrow \mathrm{H^{\bullet}} + {}^{\bullet}\mathrm{OH}$$
$$\mathrm{e^-_{aq}} + \mathrm{H^+} \longrightarrow \mathrm{H^{\bullet}}.$$

**Figure 1.6:** Cascade of reactions resulting from the interaction of water from the solvent within the crystal with ionizing X-ray radiation (Ward, 1988).

### *1.2.2 Classification and effects of radiation damage*

Radiation damage manifests in different ways and degrees, the most obvious one known as ‘global damage’, which consists in the loss in intensity and resolution of the diffracting properties of the crystal (and subsequent loss of information). This effect was firstly reported in 1962 by Blake and Phillips, with myoglobin crystals measured at room temperature. Another common symptom is the increase of the atomic B-factors, indicating the higher disorder of the structure. But the observation of global damage in a sample is the last stage, when the crystal carries already a high degree of damage. Before this happens, there are also other radiation damage fingerprints that appear in a systematic ordered way as X-ray dose increases (Weik et al., 2000;

Burmeister, 2000; Ravelli and McSweeney, 2000). This other kind of damage is known in the literature as ‘specific radiation damage’, and consists of:

- Increased disorder and breakage of disulphide bonds.
- Decarboxylation of acidic side chains (glutamate and ascorbate), as well as the c-terminal carboxylic group.
- Reduction of metal centres.

Examples of site-specific radiation damage effects are shown in Figure 1.7. The reduction of metal centres is a subtle manifestation of radiation damage especially relevant to us, since we work with metalloproteins (haem proteins) carrying an iron centre (Fe). Active sites and redox centres have been noticed as especially sensitive and prone to photo-reduction by the X-ray beam (Weik et al., 2000; Adam et al., 2004; Yano et al., 2005), due to their electrophilic nature. Thus, the reduction of metal centres is the first kind of damage known to appear in crystals (Pearson et al., 2007), long before disulphide bridges breakage or of course the classic loss of resolution.

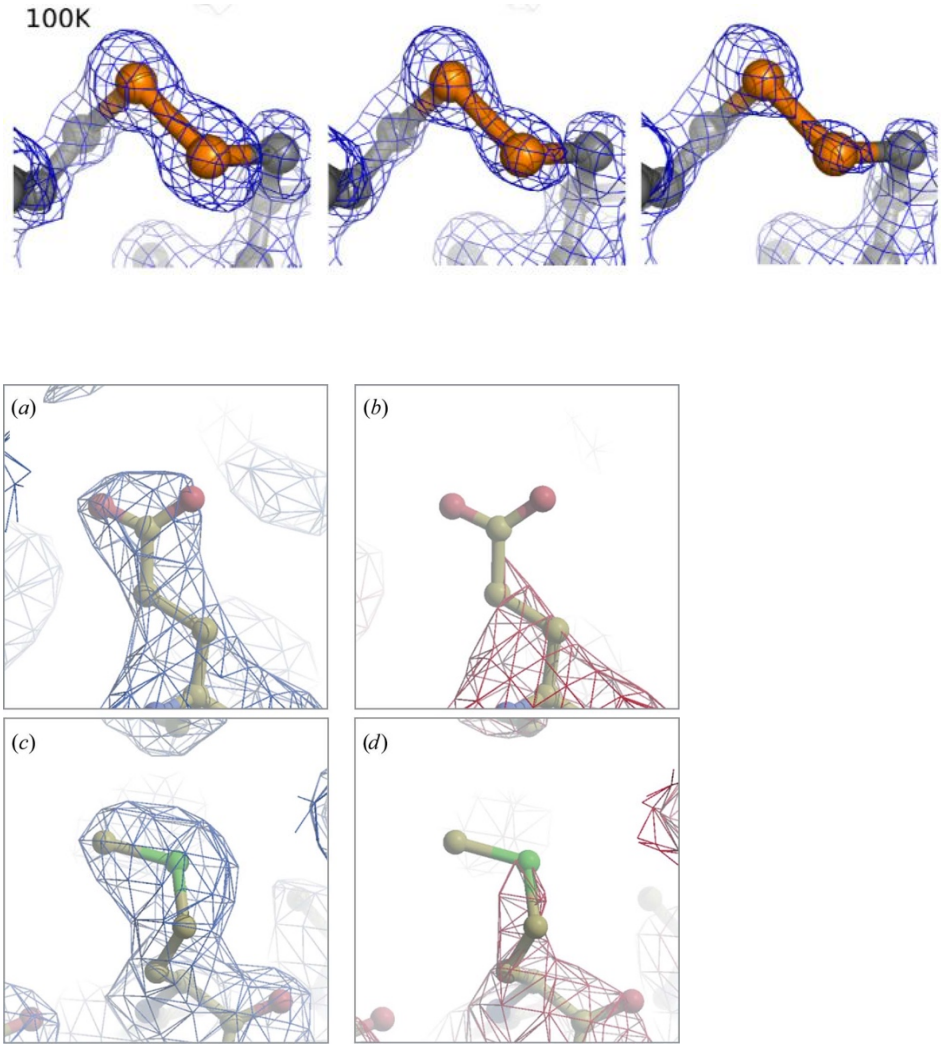


**Figure 1.7:** Site-specific radiation damage effects. In the top panel, breakage of a disulphide bridge at 100 K as absorbed dose increases. At the bottom panel, the decarboxylation of a glutamate residue (a and b) from an apoferritin cryocooled crystal, after the absorption of 2.5 MGy of dose. A loss of electron-density can be appreciated also for a methionine residue (c and d), after absorption of a much higher dose of 50 MGy. (Garman, 2010)

It is important to be aware of the radiation-induced reduction processes, as well as the possible production of radiation artefacts, since it can mislead to wrong interpretations of catalytic mechanisms and biological function. For this reason, tools were developed to assess the extension of specific radiation damage suffered by crystals while exposed to the beam. Single crystal spectroscopies (e.g. UV-vis, Raman, X-ray absorption) represent a great

resource, since they provide us complementary information that allow us to monitor radiation damage *in situ* at the beamline in proteins which contain a chromophore (Berglund et al., 2002; Hough et al., 2008).

### *1.2.3 Mitigating radiation damage*

An initial solution to the effects of radiation damage was to measure cryo-cooled samples at 100 K (Henderson 1990), by installing a nitrogen cryostream pointing onto the sample, which is the common practice nowadays to measure diffraction data. At these cryo-temperatures, crystals lifespan was observed to increase around 70 times (Nave and Garman, 2005), generally long enough for the collection of a complete dataset. With the arrival of third generation synchrotron sources with undulator beamlines in the late 1990s (which are brighter than bending magnet and wigglers ones), and the improvement in the focusing of the beam, its brilliance has improved substantially with a subsequent improvement in data quality and resolution. Furthermore, it is now possible to measure data from microcrystals with micro-focused beams. As a consequence of the higher brightness of the X-ray source, the radiation damage inflicted to crystals has increased notably - even at cryotemperatures - which has become more observed and widespread throughout the structures of proteins deposited in the PDB, and thus more relevant to the structural biology community. As it became more common, there was a growing interest in understanding the chemistry and physics behind this phenomenon (Ravelli and Garman, 2006).

Apart from cryocooling of crystals, solutions have appeared to avoid or minimize radiation damage, where two main approaches have been used. The first one consists in the use of chemical additives, such as sodium ascorbate, that act as radical scavengers and mitigate radiation damage in crystals (Murray and Garman, 2002). The second approach does not imply the modification of any conditions of the crystals, and consists in methodologies to mitigate radiation damage by building composite datasets combining datasets from several measured crystals (Figure 1.8). This approach has allowed to minimize the photo-reduction of active sites and obtain structures for oxidised states of metalloproteins (Kekilli et al., 2014). Another way of applying this same concept is measuring helical scans along the length of a big crystal (Polsinelli et al., 2017), so the beam keeps interacting with fresh material and the X-ray dose is 'spread' among all the total exposed volume. The recent development of serial crystallography methodologies (SX), has opened the possibility of easily and effectively generate composite datasets, but in this case from thousands of microcrystals, delivering low-dose structures (Owen et al., 2017). SX methodologies will be described in Chapter 2, and its application to our protein systems shown in Chapters 4 and 6.

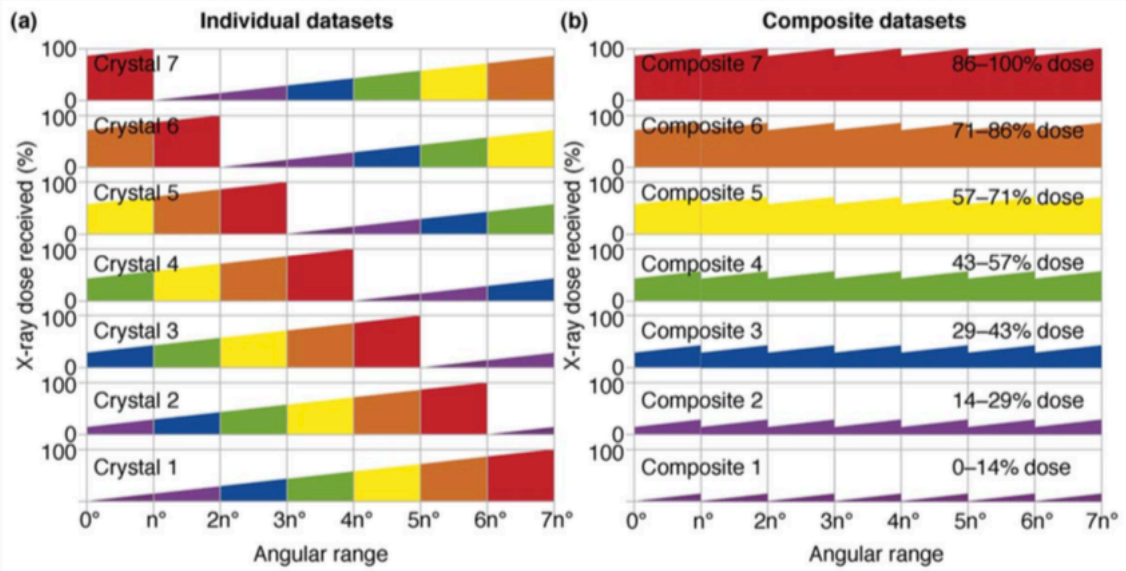


**Figure 1.8:** Composite datasets methodology. Composite datasets (b) formed from different measured crystals (a) of the same protein in its oxidized form. With this approach, it is possible to obtain complete datasets with different percentage of absorbed doses, which is helpful for the determination of the fully oxidized structure as well as reaction intermediates that may appear as the dose increases. (Berglund et al. 2002)

### *1.2.4 Estimation of absorbed dose*

The metric used to measure radiation damage inflicted to a sample is the absorbed dose, defined as the amount of energy deposited per unit mass (Zeldin et al., 2013). The unit of the absorbed dose is the Gray (Gy; 1 Gy = 1 Joule/kg). Radiation damage is proportional to dose, and amount of dose (at 100 K) for protein crystals beyond which the structural information obtained is severely compromised has been estimated to 30 MGy (Owen et al., 2006). This limit is known as the Garman limit, and at this point the crystal generally loses 30% of its diffracting power. At third generation synchrotrons, this dose limit can be easily reached within minutes of total exposure (Macedo et al., 2009).

The absorbed dose by a given crystal in X-ray diffraction experiment can be estimated using the program *RADDOSE-3D* (Zeldin et al., 2013), which can

be run online (www.raddo.se), or downloaded and run with Java in the command line. The output normally used is the diffraction weighted dose (DWD), which provides an estimation taking into account the variability in crystal dimensions. The parameters introduced to run a job in RADDOSE-3D can be divided in sample and experimental parameters. Experimental parameters include irradiation time, beam size and profile, energy and flux of photons, which are normally provided in the headers of generated diffraction images. The two kinds of beam profiles typically used at beamlines are top-hat and Gaussian beams, shown in Figure 1.9. Sample parameters for each crystal include, shape, size, crystal composition (heavy atoms), solvent percentage and buffer content.

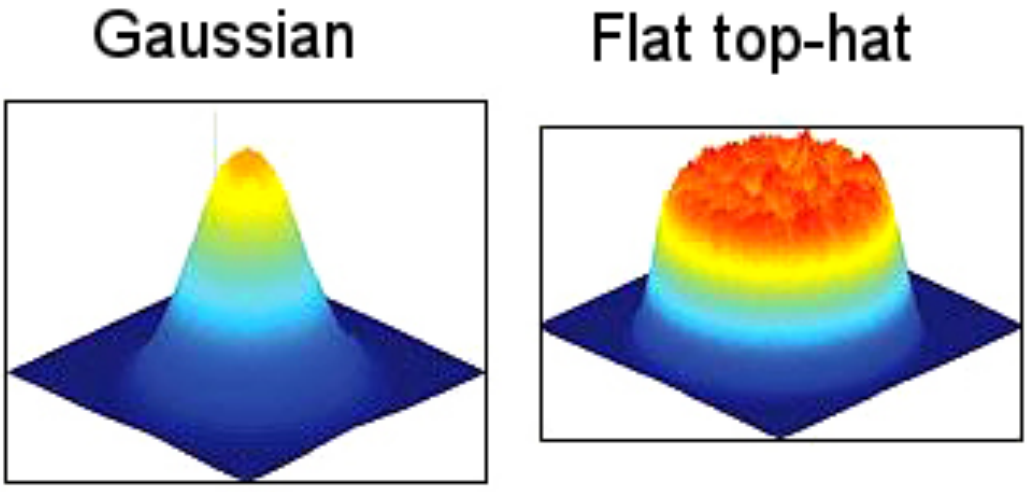


**Figure 1.9:** Three-dimensional surface plot of the Gaussian and flat top- hat beam profiles. (Dr. Demet Kekilli´s thesis)

## 1.3 Enzymatic reactions driven by the X-ray beam within crystals – structural movies.

Radiation damage can suppose a problem when trying to obtain reliable structural data, but it can also be exploited. For instance, for the determination of fully reduced forms of proteins and the elucidation of enzymatic mechanisms (Kekilli et al. 2017; Schlichting et al. 2000). The solvated electrons formed by the irradiation of the crystals with X-rays, are responsible for the reduction of metal centres in metalloproteins, as discussed, - and they can also feed the active site and be used to perform redox reactions. In this way, one can take advantage of the photo-reduction phenomenon to run reactions within crystals and study catalytic mechanisms.

This approach has been successfully applied in the past, where the combination of X-ray crystallography with single crystal spectroscopies was needed for the correct determination of reaction intermediate. A common practice has been to generate catalytic intermediates in the crystal and trap them by flash-cooling in liquid nitrogen for its structure determination afterwards (Gumiero et al., 2011; Pearson et al., 2007; Schlichting et al., 2000). Here, two foundational works on enzymatic crystallography (Berglund et al., 2002; Schlichting et al., 2000) will be briefly described, where the catalytic cycles – including reaction intermediates - for horseradish peroxidase and cytochrome P450cam (both haem proteins) were structurally characterized. A much more recent work, by Horrell and co-workers (2016 and 2018) will be also presented where successive datasets were measured building a structural

movie for a copper nitrite reductase (AcNiR), using an approach named Multiple Structures from One crystal (MSOX). In this work, we also aim to exploit radiation damage to run reactions in haem proteins, making use of all these reported methodologies.

### *1.3.1 Horseradish peroxidase (HRP) structural movie.*

In the case of HRP, a peroxidase found in horseradish root, the structures for the ferric and ferrous forms – as well as for the intermediates Compound I, II and III from HRP catalytic cycles (Figure 1.10) – were structurally characterized, and spectroscopically validated using UV-vis absorption (Figure 1.11). This was accomplished by applying the composite datasets approach mentioned before, in order to prevent the quick reduction of the active centre and consequent loss of the species of interest. In this approach, several crystals were collected and composites of datasets with different percentage of X-ray dose would then be formed from different crystal data wedges (Figure 1.12a).

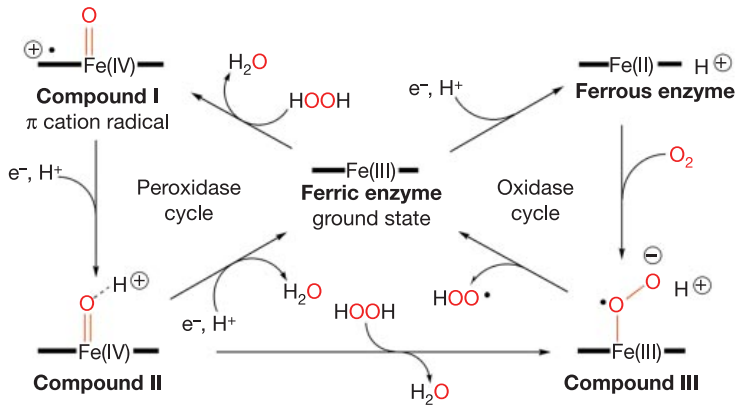


**Figure 1.10:** The five oxidation states of horseradish peroxidase, showing their involvement in the peroxidase and oxidase cycles performed by this protein. (Berglund et al., 2002).

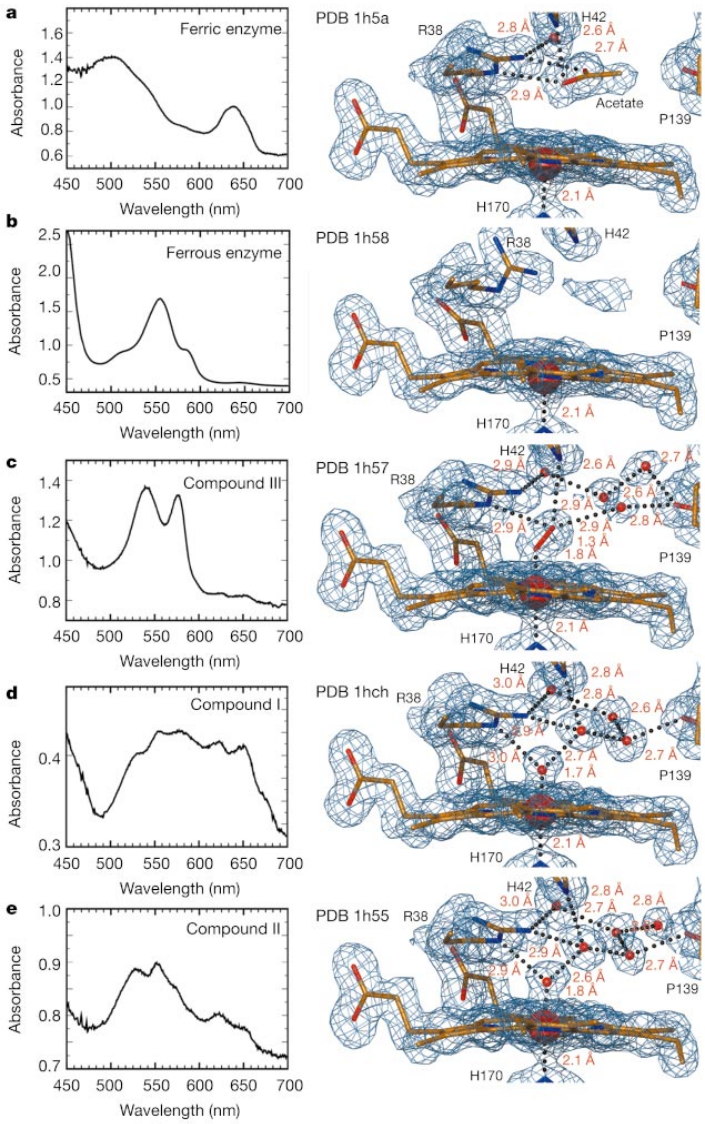


**Figure 1.11:** UV-vis spectra (left column) and refined high-resolution structures (right column) for the five oxidation states of horseradish peroxidase: a, Ferric enzyme; b, Ferrous enzyme; c, Compound III; d, Compound I; e, Compound II. (Berglund et al., 2002)

Applying also the composite methodology, they built up a structural movie, following the photo-reduction of Compound III (one of the intermediates in the oxidase cycle), and its gradual conversion to the ferrous state with subsequent formation of 2 molecules of water, as absorbed dose increased. The five frames of this structural movie can be observed at atomic detail in (Figure 1.12b). Each of them correspond to a certain X-ray dose percentage

range; and as the radiation increases more electrons are introduced into the system, that becomes more photo-reduced.

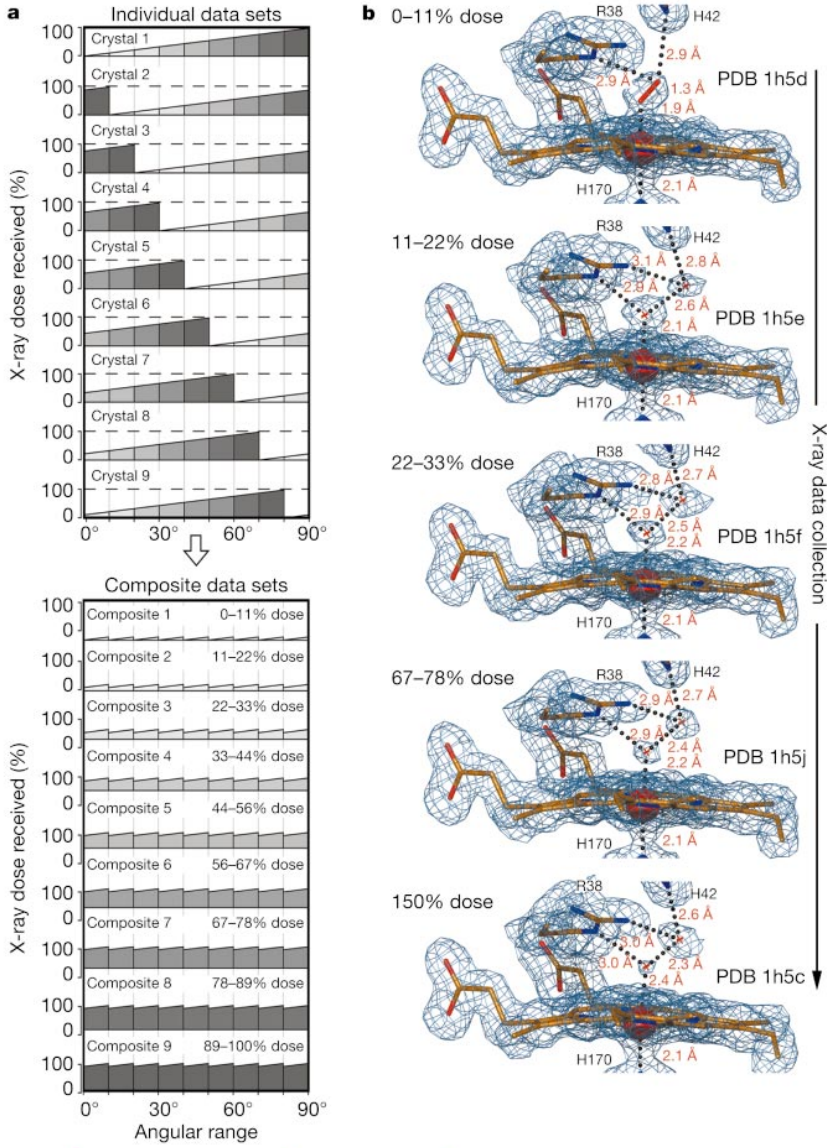


**Figure 1.12:** X-ray-driven catalytic conversion of a dioxygen species in horseradish peroxidase. a) the multicrystal data collection strategy, showing the distribution of the X-ray dose as a function of the rotation angle on individual crystals of HRP. The construction of composite data sets from small chunks of the individual data sets is shown at the bottom. Composite data sets represent structures that received different X-ray doses. This method permits experiments similar to redox titrations. b) electron density maps of the different composite structures with increasing X-ray dose, showing reduction of Compound III. For the last structure, the crystal was pre-exposed to X-rays for 90º before another full X-ray data set was collected on it. (Berglund et al., 2002)

### *1.3.2 Cytochrome P450cam*

Cytochromes P450 are haem-containing enzymes, specifically monooxygenases, named by their absorption maxima of light at 450nm when bound to CO (Ortiz de Montellano, 1995). They stereospecifically catalyse the hydroxylation of hydrocarbons, a reaction that otherwise would need extremely high temperatures for it to happen (even non-specifically). In the case of cytochrome P450cam, it catalyses the stereospecific hydroxylation of camphor to 5-exo-hydroxy-camphor following a complex catalytic cycle shown in Figure 1.13.

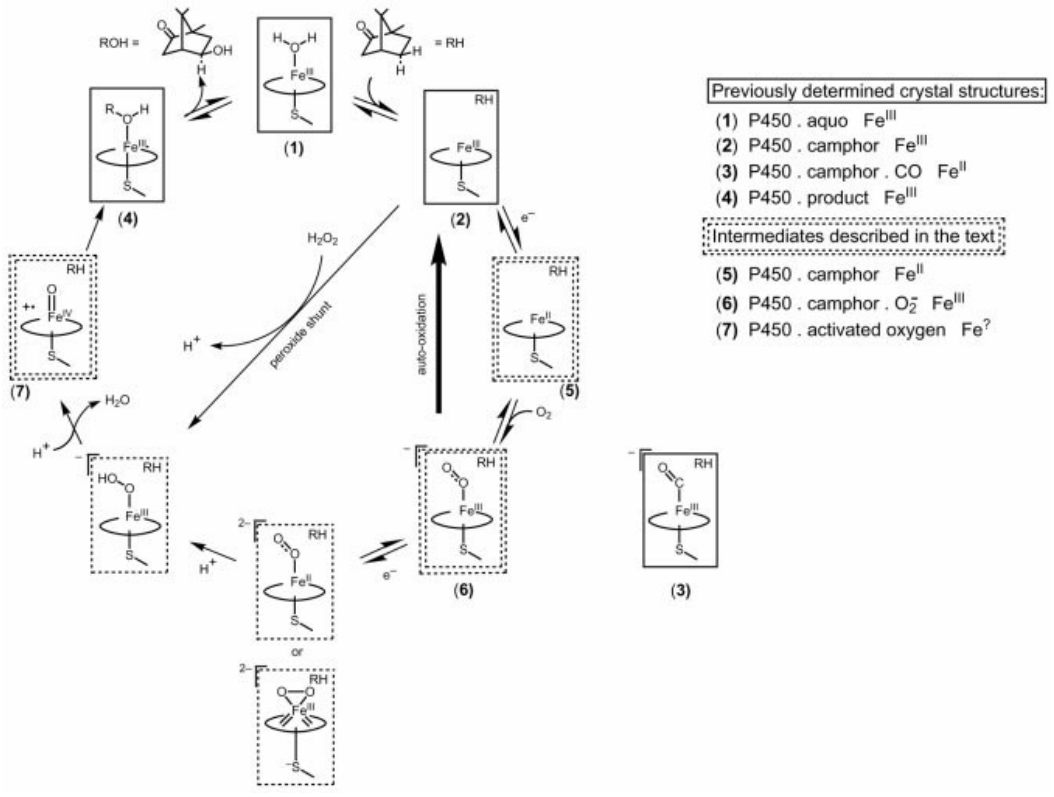


**Figure 1.13:** Schema showing the catalytic cycle of *Pseudomonas putida* P450cam for the hydroxylation of camphor. (Schlichting et al., 2000)

Schlichting and co-workers (2000) were able to obtain structures of some of the intermediates of this catalytic pathway by using a novel approach – building up intermediates in the crystals and freeze-trap them in liquid nitrogen for diffraction experiments later on at the beamline. Specifically, they obtained: the structure for the photo-reduced ferrous complex (Figure 1.14A; species 5 at the cycle in Figure 1.13); the structure for the dioxygen complex (Figure 1.14B and C; species 6 in the cycle in Figure 1.13); and the structure for the oxyferryl intermediate (Figure 1.15; species 7 in Figure 1.13).

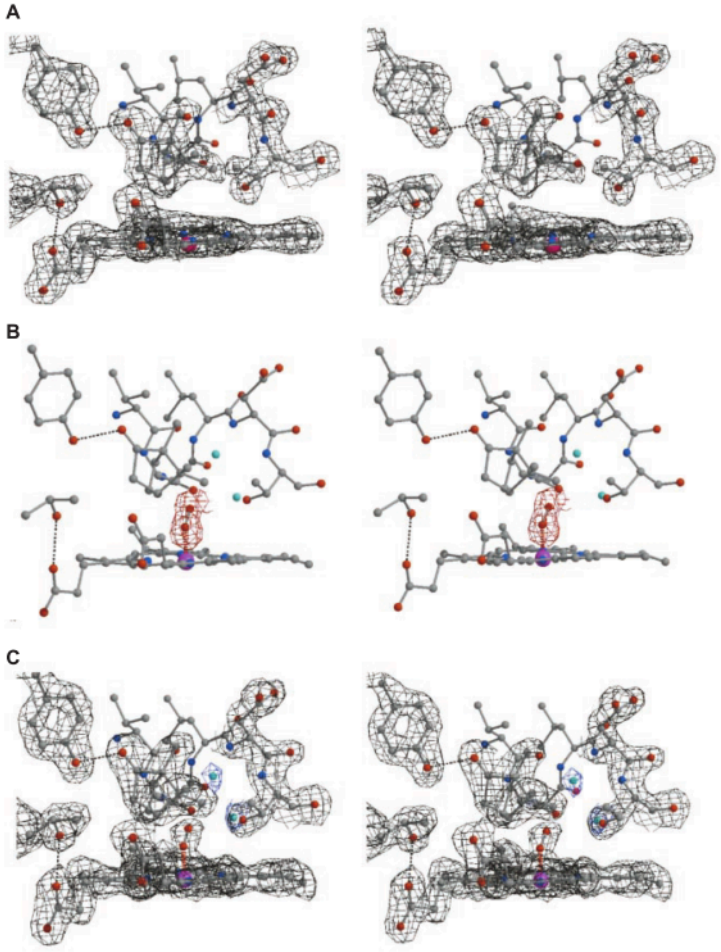


**Figure 1.14:** Stereoview of electron densities and molecular models for different P450 complexes. Map coefficients and contour level are given in parentheses. A) Reduced ferrous form (electron density map $2F_o - F_c$ at $1.3\sigma$). B) Dioxygen complex (simulated annealing omitting the O2 coordinates, $F_o - F_c$ ,$3\sigma$). C) Final model of P450cam dioxygen complex (electron density map $2F_o - F_c$ at $1.3\sigma$). (Schlichting et al., 2000)

In their novel methodological approach, they moved through the cycle from one intermediate to the another by providing to the system the needed conditions for the changes to happen. For example, introducing electrons to the system in different ways: with chemical reducing agents (dithionite - obtention of dioxygen complex) or with electrons induced by X-ray radiation (radiolysis of water molecules in the crystal - obtention of oxyferryl complex). For obtaining the hydroxylated product structure, they just warmed up the crystal (letting it thaw) at the oxyferryl intermediate state and flash-freezed it again in liquid nitrogen to obtain a further dataset. They also used UV-vis single-crystal spectroscopy to monitor and validate the formation of the reduced dioxygen intermediate from ferric microcrystals (Figure 1.16).

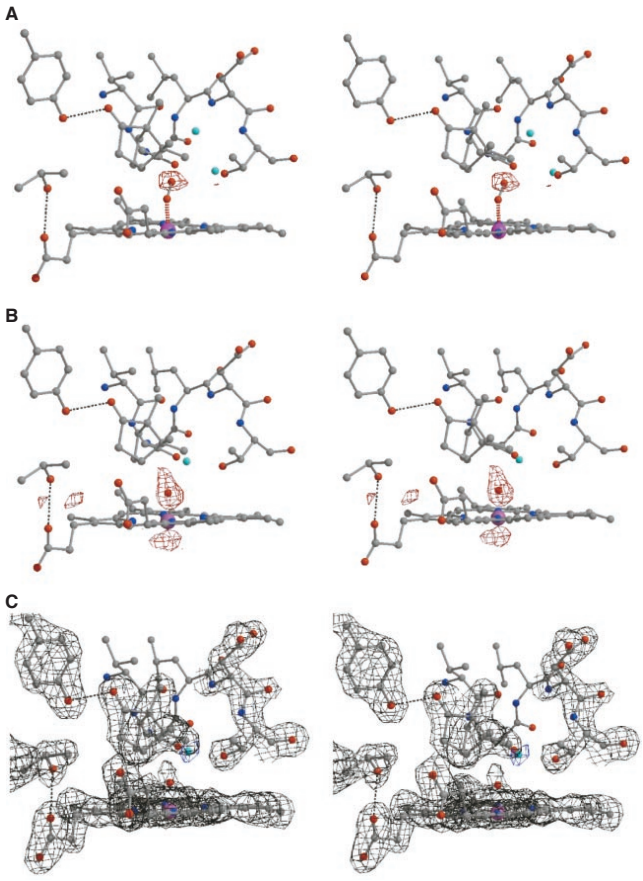


**Figure 1.15:** Stereoviews of electron densities and molecular models for different P450cam complexes. A) Difference density between dataset of dioxygen and oxyferryl intermediates, showing the oxygen atom that is reduced and released as water (Fo-Fo map contoured at

3.6σ). B) Data set of oxyferryl complex (simulated annealing map of the final model omitting the oxygen coordinate, Fo-Fc at 3σ). C) Final model for the oxyferryl intermediate (2Fo-Fc at 1.3σ). (Schlichting et al., 2000)

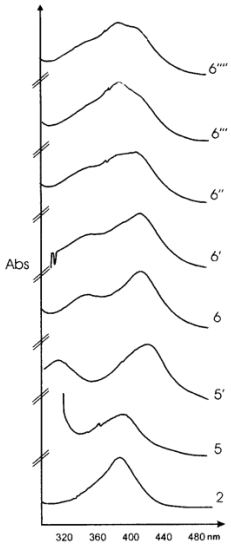


**Figure 1.16:** Spectroscopic analysis of the formation of the reduced dioxygen intermediate of P450cam with microcrystalline slurries mounted in a flow cell and kept at 4ºC. The numbers refer to the species in figure 1.12. **2**, ferric P450.camphor complex; **5**, reduced complex with dithionite present; **5'**, while washing with nitrogenated mother liquor; and **6**, addition of oxygenated mother liquor induces formation of the dioxy complex, which slowly decays back to the ferric complex. **6** to **6''''** show the spectra obtained 220, 360, 440, 600, and 960 s after oxygen addition, respectively. (Schlichting et al., 2000)

### *1.3.3 MSOX measurement with AcNiR*

Recently, Horrell and co-workers (2016) made use of the solvated electrons generated by the X-ray beam to run redox reactions in the copper nitrite reductase AcNiR, from the denitrifying bacteria *Achromobacter cycloclaste*, applying a novel approach. It consisted in measuring Multiple Structures from One crystal (MSOX), and they generated a structural movie build-up by 45 successive datasets measured at 100 K from the same crystal volume, where catalysis turnover, i.e. the reduction of nitrite ($NO_2^-$) to nitric

oxide (NO·), and induced conformational changes could be monitored at the copper active site (example in Figure 1.17, RT-movie). Initially, nitrite could be observed bound at the type2-copper (T2Cu) active site in two conformations (named ‘top-hat’ and ‘side-on’). As absorbed dose by the crystal increased, turnover could be observed by dataset 11, with presence still of partially bound nitrite. By dataset 17, just the product nitric oxide was observed bound at the T2Cu, which appeared substituted by a water molecule in dataset 40. Following this work, they also determined the same structural movie at 190 K (close to glass transition temperature) and room-temperature (Horrell et al., 2018) to investigate conformational and dynamic suppressions that might have occurred at cryo temperatures, and observed that the reaction would take place much more rapidly and at lower doses at higher temperatures (for RT movie, turnover could be appreciated by dataset 5 and 0.15 MGy dose). The structural movie at room-temperature is shown in Figure 1.17.

Therefore, a methodology was established that could be applied to any other redox system if suitable crystals were obtained, and the redox chemistry within them takes place at shorter dose-scales than radiation damage processes.

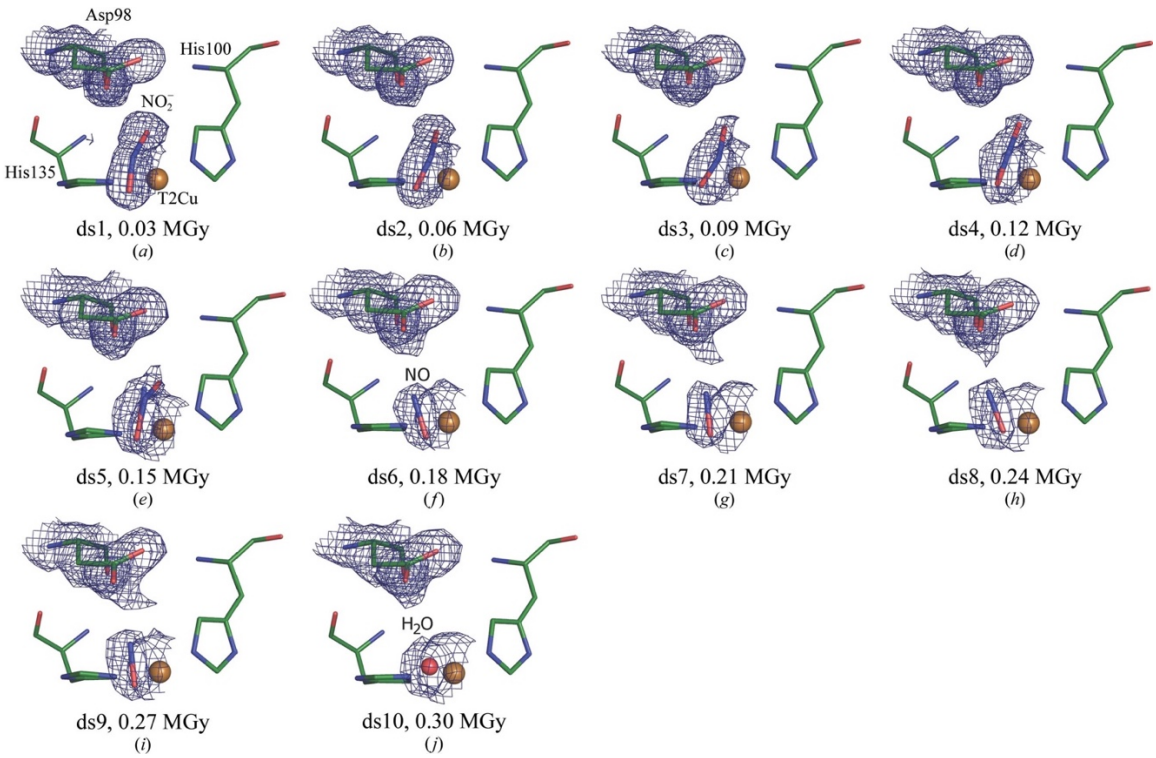


**Figure 1.17:** MSOX movie of catalysis in AcNiR at room temperature. A ten-frame X-ray-induced 'movie' recorded from consecutive serial data sets showing conversion at the type2-Cu site of bound nitrite to NO and then to the $H_2O$ resting state. Initially, in dataset1 (ds1), the nitrite is bound to T2Cu in the 'top-hat' orientation (a), but this rapidly changes in ds2 to a progressively more 'side-on' orientation (b, c and d). The 'side-on' nitrite is converted to NO by ds5(e) and the NO is replaced by water at ds10 (j). Only the proximal position of Asp98 is observed in this sequence of electron-density maps. Asp98, His100, His135, nitrite and NO are represented as sticks, and water molecules and Cu atoms as spheres. Other T2Cu-site residues are omitted for simplicity. The $2F_o - F_c$ density in $e^-/Å^3$ is contoured at 0.45 (ds1), 0.43 (ds2), 0.39 (ds3–ds7) and 0.38 (ds8–ds10). The cumulative X-ray dose absorbed by the crystal is indicated for each snapshot of the series.

## 1.4 Protein systems.

In our methodological approach, we aimed to study enzymatic systems that provide a suitable platform to carry out our research, and they were at the same time novel targets with high interest in their structure and catalytic mechanisms. Protein candidates had to meet two important requirements:

i) They should be proteins which yield high-resolution crystals (both as macro and microcrystals, ideally) needed to observe in detail the active site where the reaction takes place.

ii) Spectroscopic information should be easily measurable from the crystals to characterize different ligand/redox states – i.e. at least one chromophore had to be present in the protein.

All the enzymes studied were haem proteins that belonged to different organisms (bacterium and marine annelid), and catalyse reactions whose products should be apparent at the active site if enough resolution is reached by the crystals (e.g. 2 - 1.5 Å or higher). Therefore, they are all coloured proteins as they all have a haem cofactor bound at one of the protein domains. The presence of this chromophore allows the application of spectroscopies (e.g., UV-vis, fluorescence and Resonance Raman spectroscopies) to monitor its state, providing us a powerful tool to better characterize each protein system. In total, four haem proteins were studied (DHP, DtpA, DtpAa, and NOD), which are described in turn below.

### *1.4.1 DyP-type peroxidase A (DtpA) and Aa (DtpAa) from Streptomyces lividans.*

Dye-decolorizing peroxidases (DyPs) are a relatively novel group of peroxidases identified in 1999 by Kim and Shoda. They are found in bacteria and fungi, but mainly in bacteria. For this reason, it was proposed to name them “bacterial peroxidases” (Fraaije and van Bloois, 2012). They present a lack of homology with classic plant and animal families of peroxidases, and different reactivity. They were named in this curious way after the observation that these enzymes could degrade a broad range of xenobiotic and recalcitrant industrial dyes. The first DyP protein was isolated in 1999 from the fungus *Geotrichum candidum* (Kim & Shoda, 1999). Since then, over 5000 other DyP-type proteins, both from prokaryotic (mainly) and eukaryotic origins have been identified (Colpa et al. 2014). They are scarcely present in archaea. They are sub-classified into the phylogenetically distinct classes A, B, C, and D. Some examples for each subclass are shown in Table 1.1. Many of the DyP-type bacterial enzymes belong to classes B and C, and they perform their activity in the cytoplasm, suggesting their involvement in intracellular metabolic pathways. Enzymes belonging to class A contain a Tat-dependent signal sequence, and therefore perform their catalytic activity outside of the cytoplasm or extracellularly (Van Bloois et al., 2010). Class D contains mainly fungal variants, such as the one discussed above from *G. candidum*.

**Table 1.1:** DyP-type peroxidases subclasses. DyPs can be subdivided into classes A-D. Individual protein name, the organism they originate from and the references associated are listed.

| Subclass | Protein | Organism | Reference |
|---|---|---|---|
| **A** | *Tc*DyP | *Thermomonospora curvata* | Chen et al., 2015 |
| | *Tfu*DYP | *Thermobifida fusca* | van Bloois et al., 2010 |
| | DyPA | *Rhodococcus jostii* RHAI | Ahmad et al., 2011 |
| | EfeB | *Escherichia coli* | Sturm et al., 2006 |
| **B** | DyPB | *Rhodococcus jostii* RHAI | Ahmad et al., 2011 |
| | *Mt*DyP | *Mycobacterium tuberculosis* | Contreras et al., 2014 |
| | *Pp*Dyp | *Pseudomonas putida* | Santos et al., 2013 |
| **C** | DyP2 | *Amycolatopsis sp.* | Brown et al., 2012 |
| | AnaPX | *Anabaena sp.* | Ogola et al., 2009 |
| **D** | *Bad*DyP | *Bjerkandera adusta* | Kim and Shoda, 1999 |
| | *Aau*DyP | *Auricularia auricular-judae* | Strittmatter et al., 2013 |
| | *Il*DyP | *Irpex lacteus* | Salvachúa et al., 2013 |

In this work, we studied two DyP-type proteins found in the bacterium *Streptomyces lividans*: DyP-type peroxidase A (DtpA) and Aa (DtpAa). They both belong to the subclass A within the family of DyP peroxidases. Therefore, both are secreted and function extracellularly, although DtpA could be associated with the external membrane (Petrus et al. 2016). Functional differences for these two enzymes are still unclear, although DtpA is proposed to be involved in the maturation morphological development induced by copper in *S. lividans* (Petrus et al., 2016).

DyPs have several characteristics that differentiate them from classic haem peroxidases: they have a lack of similarity in sequence; they show a wide substrate specificity; and they have the ability to perform at acidic pHs (as low as pH=3.0-3.2; Sugano et al., 2007), which is not usual for peroxidases. In terms of specificity, DyPs show a broad range of substrate reactivity, degrading common substrates to other peroxidases (e.g. azo dyes and lignin), but also anthraquinone-based dyes widely used in industry. In this regard, DyPs represent a promising candidate to develop a methodology to treat wastewaters containing this kind of synthetic dyes, which generate an environmental issue due to their recalcitrant nature. Nevertheless, the natural substrate and role of this family of proteins is still unclear, although bacterial DyPs have been shown to degrade lignin (Ahmad et al. 2011; Van Bloois et al. 2010). For this reason, this lignolytic role has suggested DyPs to be regarded as the bacterial counterpart of the lignin peroxidases found in fungi.

From the structural point of view, DyPs also diverge from classical plant and animal peroxidases, which only feature alpha-helices in the secondary structure (Figure 1.18). Crystal structures of DyP proteins have shown two domains, both adopting a ferredoxin-like fold, where alpha-helical and beta-sheets conformations are observed (Figure 1.18 and Figure 1.19). The haem cofactor is found in just one of these domains, and is bound to the protein through the conserved proximal histidine, present in all peroxidases. At the distal haem site, a new configuration of residues can be found in DyPs, with the absence of the common distal histidine as the most important feature (Figure 1.19). In classic peroxidases, this histidine acts as a proton

donor/acceptor and catalyses the formation of the intermediate Compound I (Ferrari et al. 1999), through the known as Poulos-Kraut mechanism (Poulos and Kraut, 1980). In DyPs, two conserved residues are present at the distal site: an arginine, and an aspartate (Figure 1.19). The presence of the Asp, with a 'free' side chain $pK_a$=3.9, is consistent with the more acidic conditions at which this family of peroxidases can perform, in contrast with the histidine from classic peroxidases, with a $pK_a$=6.0. The Asp residue is also thought to play the main catalytic role in the peroxidase cycle, acting as alternative acid-base catalyst for the formation of Compound I after the binding of $H_2O_2$. A mechanism involving the swing of the Asp into the distal site for catalysis was proposed by Yoshida et al. (2011). However, the presence of the Asp is not so important in subclass-B DyPs, where the Arg is proposed to act as the catalytic residue, for instance in protein DyPB from *R. jostii* (Ahmad et al., 2011). The substitution of the distal Arg by a Lys in the subclass-D *Bad*DyP, resulted in a partial preserved peroxidase activity (Sugano et al., 2007). Therefore, the roles of these residues are different in distinct DyP subclasses.

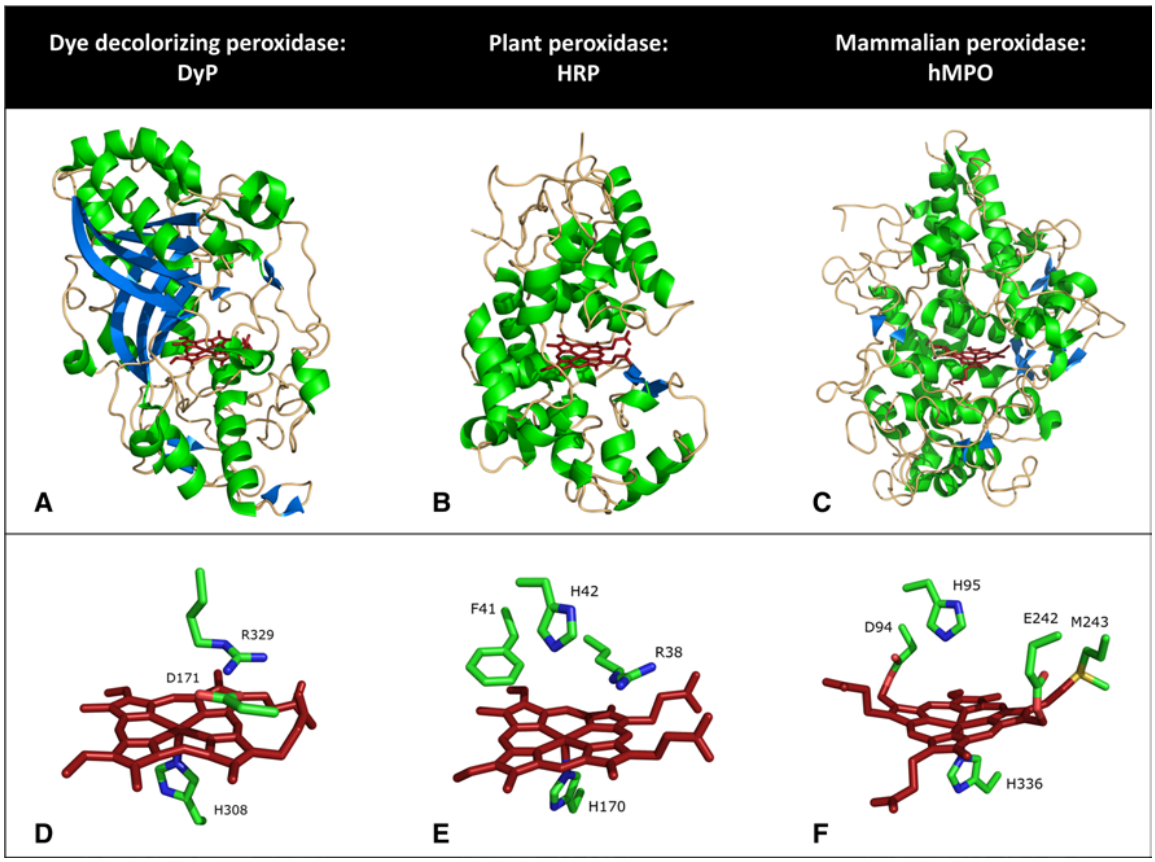


**Figure 1.18:** Structural comparison between DyP from *Bjerkandera adusta* Dec1 (A), HRP from *Armoracia rusticana* (B), and human myeloperoxidase (hMPO) from *Homo sapiens* (C). α-helices are shown in green, β-sheets are in blue, and the heme cofactor is in red. Close view of important amino acids surrounding the heme group of DyP (D), HRP (E), and human myeloperoxidase (F). The proximal histidine of DyP (His308) and both the distal and proximal histidines of HRP (His42 and His170) and human myeloperoxidase (His95 and His336) are indicated, as well as catalytically important residues of DyP (Asp171 and Arg329) and HRP (Arg38 and Phe41). Heme is covalently bound by Asp94, Glu242, and Met243 in human myeloperoxidase. (Colpa et al., 2014)

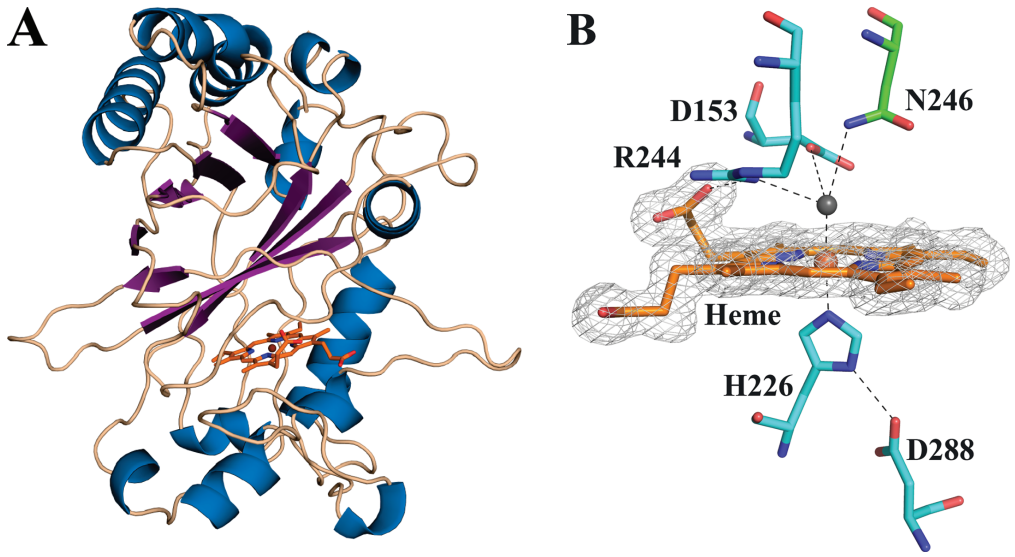


**Figure 1.19:** Structure of the DyPB protein RHA1 from *Rodococcus jostii* (Roberts et al., 2011). A) Overall view of the protein, where the two ferredoxin-like domains can be appreciated, with alpha-helices in blue and beta-sheets in purple. One of them contains the haem prosthetic group (in orange). B) Structural detail of the active site with the most relevant residues shown. Note the aspartic acid (D153) and the arginine (R244) residues at the distal site, conserved among all DyPs. Asparagine residue (N246) in green is only conserved within sub-class B. A distal water is shown above the haem iron as a grey ball.

### *1.4.2 Dehaloperoxidase (DHP), from Amphitrite ornata.*

DHP is a multifunctional haem-containing enzyme found in the marine annelid *Amphitrite ornata* (Lebioda et al. 1999; Franzen et al. 2012). Two main functions were firstly identified for DHP (de Serrano et al., 2007): a haemoglobin function, common in globin proteins, and a peroxidase function. In its haemoglobin function, DHP binds molecular oxygen in its ferrous state and carries out transport and storage functions. Like many other globins, it can also bind other diatomic gases, such as NO and CO (e.g. cytochrome *c*'s). When DHP functions as a peroxidase, catalyses the oxidation of trihalophenols

into the corresponding dihaloquinones, in the presence of hydrogen peroxide (Figure 1.20). DHP can be found in two isoforms, A and B, which differ just in 5 amino acids located around the active site. Both isoforms have shown to catalyse the main known reaction with trihalophenols as a substrate, although differences have been identified (D'Antonio et al., 2010), such as the observation that DHP-B is two to four-fold more active than DHP-A for these substrates.

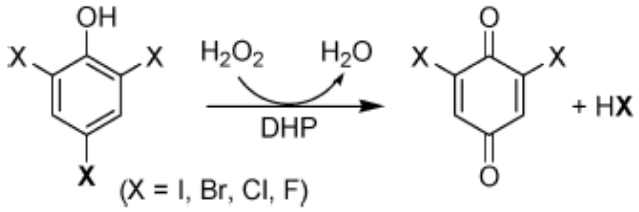


**Figure 1.20:** Reaction catalysed by DHP: trihalophenols are oxidized by hydrogen peroxide, producing dihaloquinones (de Serrano et al., 2007).

Since they belong to the same protein superfamily, DHP shares structural and enzymatic characteristics with other peroxidases such as horseradish peroxidase (HRP). Both HRP and DHP present an all $\alpha$-helices fold, and can accept a wide range of halogenated phenols as substrates (Ferrari et al., 1999). It has been proposed by Lebioda and co-workers (1999) that the peroxidase function in DHP arose from the evolutionary pressure to degrade brominated aromatic compounds generated as repellents by other marine worms (Woodin et al., 1993) who share the benthic ecosystem with *A. ornata*. In this way, the later marine worm would metabolize these harmful substances into more harmless ones. Due to the relatively big size of these aromatic compounds compared to the usual diatomic haem-binding ligands, a large distal pocket evolved at the DHP haem site, where more families of

aromatic compounds can bind, such as indoles (Barrios et al. 2014). Thus, DHP has biotechnological potential to degrade recalcitrant aromatic pollutants used in agriculture and industry (McCombs et al., 2017).

Notably, it was observed that peroxidase activity in DHP can be initiated from both the ferric and ferrous (or oxy-ferrous) oxidation states (D'Antonio and Ghiladi, 2011). When starting from the ferric state, DHP appears to follow the Poulos-Kraut classic mechanism (Poulos and Kraut, 1980), in which a distal histidine catalyses the formation of Compound I after the binding of $H_2O_2$ to the haem iron. Further studies on DHP carried out by Barrios and co-workers (2014), found additional functions performed by the protein: peroxygenase, oxidase, and oxygenase functions. Each of these activities would potentially protect *A. ornata* from the brominated and other halogenated aromatic compounds present in its surrounding environment. This multifunctionality shown by DHP has no precedent in the haem-protein family – a single protein performs five different functions while in other cases it would be needed a different metalloenzyme for each distinct function. This observation raises new questions about structure-function paradigms within the globin family. An unknown and interesting aspect is how DHP switches between functional states, and which conformational changes are involved (Carey et al. 2018).

The first DHP crystal structure (isoform A) showed a homodimer (Figure 1.21) in the asymmetric unit, formed by 15.5 kDa subunits, each of which contains a haem-B cofactor and eight $\alpha$-helices adopting a globin fold (LaCount et al. 2000). An interesting feature was discovered in this work: a binding pocket

at the haem distal site where aromatic substrate analogues would bind (4-iodophenol) up in the pocket, not interacting directly with the heme iron. Such a structural detail had no precedent in haem proteins and Belyea et al. (2005) proposed that the occupation of the pocket by the substrate could trigger the switch between the haemoglobin and the peroxidase functions.

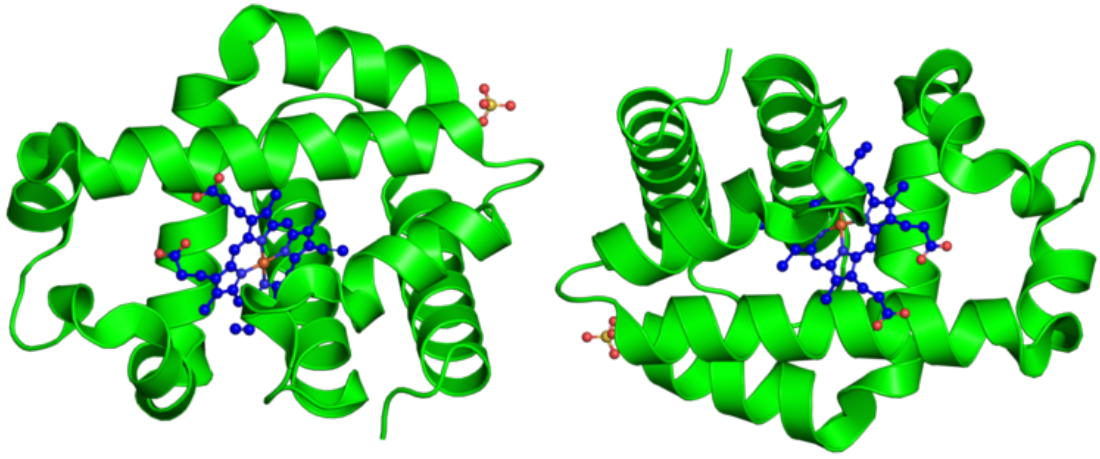

**Figure 1.21:** DHP-A X-ray crystallography structure. General view of the DHP-A dimer found in the asymmetric unit of the first X-ray diffraction structure (PDB code: 1ew6) for this haem enzyme, which is adopting a globin fold.

DHP was extensively characterized structurally and biochemically in the following years (Barrios et al., 2014; Chen et al., 2009; De Serrano et al., 2007; De Serrano et al., 2010) with structures including: wild type and mutant structures for DHP-A, deoxy-ferrous structure of DHP-A, complex of DHP-A with a substrate, structure of DHP-B, structure of the AB heterodimer. A relation of DHP structures is detailed in Table 1.2. At the active site of DHP (Figure 1.22) we can find a proximal histidine (His89), benchmark of the globin family, as well as a distal histidine (His55), conserved in haem peroxidases. The latter is supposed to have an important role in the regulation of the different functions

carried out by DHP, since it has been noted its conformational flexibility swinging in an out of the distal pocket. It is thought to catalyse the peroxidase reaction acting as proton acceptor/donor (Poulos-Kraut mechanism) when hydrogen peroxide is bound to the iron, and mediate the formation of compound I.

**Table 1.2:** Selection of DHP structures. Relation of selected DHP structures found in the literature, specifying the initial state of the crystals (before X-ray diffraction), resolution reached, and publication reference. The substrate TBP stands for 2,4,6-tribromophenol.

| **Sample** | **Initial state** | **Resolution (Å)** | **PDB entry** | **Reference** |
|---|---|---|---|---|
| DHP-A | Ferric | 1.78 | 1ew6 | LaCount et al., 2000 |
| DHP-A | Ferric | 1.62 | 2qfk | De Serrano et al., 2007 |
| DHP-A | Deoxy-ferrous | 1.26 | 3dr9 | Chen et al., 2009 |
| DHP-B | Ferric | 1.58 | 3ixf | De Serrano et al., 2010 |
| DHP-AB dimer | Ferric | 1.52 | Not deposited | De Serrano et al., 2010 |
| DHP-A | Complex with substrate TBP | 1.44 | 4fh6 | Zhao et al., 2013 |

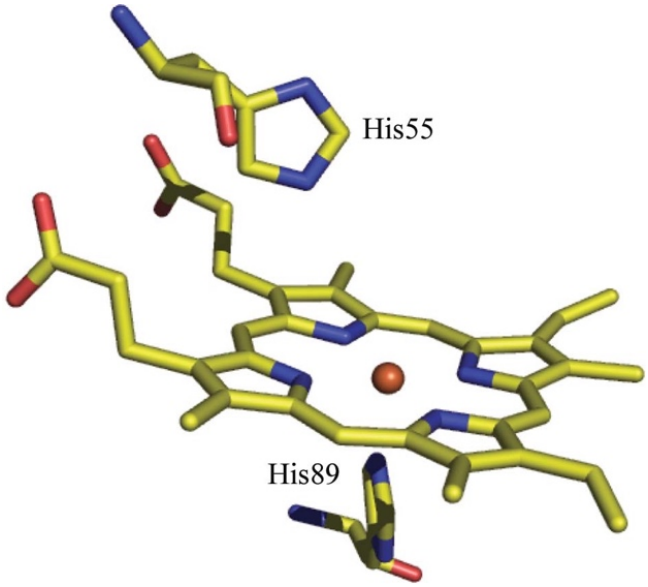


**Figure 1.22:** Representation of the heme prosthetic group of DHP (PDB code 1ew6) relative to the position of the proximal histidine (His89) and the distal histidine (His55) located below and above the heme plane, respectively. In this case, the distal histidine is in the closed conformation swinging into the pocket. (de Serrano et al., 2007)

### *1.4.3 Nitric oxide dioxygenase (NOD) from Escherichia coli*

NOD belongs to the flavohemoglobin (flavoHb) family, which is a group of haem proteins found in bacteria (Mukai et al., 2001). NOD is the flavohemoglobin found in *Escherichia coli*, and it can be also found in the literature named as *hmp* (which stands for haemoprotein; Zhu & Riggs, 1992). Like other flavoHb presents a common protein structure of three differentiated domains with characteristic functions, fused together to yield a protein with novel functional proteins. A globin domain harbours the haem group where the active site is located, and two oxidoreductase domains (NAD-binding and FAD-binding domains), contain the corresponding cofactors and are responsible for shuttling electrons to the haem site to feed the redox reaction catalysed by the enzyme. This general structure was first observed by Ermler and co-workers in (Ermler et al., 1995), when they solved the crystal structure of *Alcaligenes eutrophus* flavoHb (FHP). NODs are enzymes that effectively catalyse the oxidation of nitric oxide (NO·) into nitrate ($NO_3^-$) (Vinogradov & Moens 2008), following the general reaction:

$NO\cdot + O_2 + e^- \rightarrow NO^-_3$

NOD was isolated from NO· resistant *E. coli* colonies, that were able to metabolize this radical (Gardner et al., 1998). When characterized, the protein was found to have been previously described as *E. coli* flavohemoglobin *hmp* (Vasudevan et al., 1991). *Hmp* was in fact the first flavoHb to be described, and the physiological role and structural properties of this group of proteins were unknown. By obtaining this link between NOD and *hmp* a relevant function for

flavohemoglobins was discovered (Gardner et al., 1998). It was proposed that NODs catalyse NO· oxidation as a defence mechanism carried out by pathogenic bacteria, when the host secretes harmful levels of this radical to induce what is known as nitrosative stress. Nitric oxide is a free radical with multiple and diverse biological functions (Kelm and Schrader, 1990; Snyder and Bredt, 1992; Fang, 1997). It is produced in organisms as a result of enzymatic or non-enzymatic oxidations and reductions of nitrogen compounds (e.g. in the nitrogen cycle). In bacteria for example, it is formed by the metabolic reduction of nitrate and nitrite (Goretski et al., 1990). NO· is a known signalling molecule that diffuses through the cellular membrane and is involved in a variety of biological processes (e.g. in blood pressure control in animals). However, when it is present at high concentrations within the cell it is a harmful substance due to its highly reactive radical nature. This property is exploited by the cell as a defence mechanism (Fang, 1997), and toxic amounts are produced by an NO-synthase in response to infections, foreign bodies or tissue injury. Here is where NODs play an important role for the pathogenic bacteria, metabolising and neutralising the high NO· levels, and converting it into the harmless $NO_3^-$.

Since most haemoglobins, if not all, are able to perform the NOD function, it has been proposed that it may be the ancestral original function of haemoglobins rather than the well-known oxygen transport and storage function (Gardner 2012; Vinogradov & Moens 2008). This hypothesis is supported by amino acid sequence similarity of the respective haemoglobin and Flavin domains in flavohemoglobins when compared with haemoglobins

from distantly related animals, plants, bacteria and protists (Hardison, 1996; Zhu and Riggs, 1992; Wittenberg and Wittenberg, 1990; Membrillo-Hernandez and Poole, 1997).

The proposed enzymatic mechanism for NOD (Gardner et al., 1998) consists in several subsequent steps summarised in the following scheme:

(1) $NADH + FAD + H^+ \rightarrow NAD^+ + FADH_2$

(2) $2Fe^{3+} + FADH_2 \rightarrow 2Fe^{2+} + FAD + 2H^+$

(3) $2Fe^{2+} + 2O_2 \rightarrow 2Fe^{3+}\text{-}O_2^-$

(4) $2Fe_3^+\text{-}O_2^- + 2NO\cdot \rightarrow 2Fe_3^+ + 2NO_3^-$

---

(5) $2NO\cdot + 2O_2 + NADH \rightarrow 2NO_3^- + NAD^+ + H^+$

This mechanism is consistent with experimental observations where NOD activity is dependent on NAD, FAD and $O_2$, and sensitive to cyanide ($CN^-$ interferes in reaction 2, binding to $Fe^{3+}$ and preventing the reduction by $FADH_2$; Garner et al., 1998). The flavin group acts as electron carrier from the NAD to the haem cofactor. Oxygen binds with high affinity to the reduced iron (as observed in general for haem proteins), and the conversion of NO· into nitrate occurs even faster, leaving the reduction of the iron as the limiting step. Although well characterised, this mechanism is not yet fully understood at the structural level.

The structure for recombinant NOD (expressed in *E. coli*) was firstly solved by Ilari et al. (2002) at 2.2 Å resolution, from crystals in the ferric state. During purification, a lipid bound to the protein was removed to facilitate

crystallization. The structure overview is shown in Figure 1.23, where the characteristic flavohemoglobin fold with three domains is observed: a C-terminal NAD-binding domain, a FAD-binding domain, and an N-terminal catalytic globin domain where the haem cofactor is located.

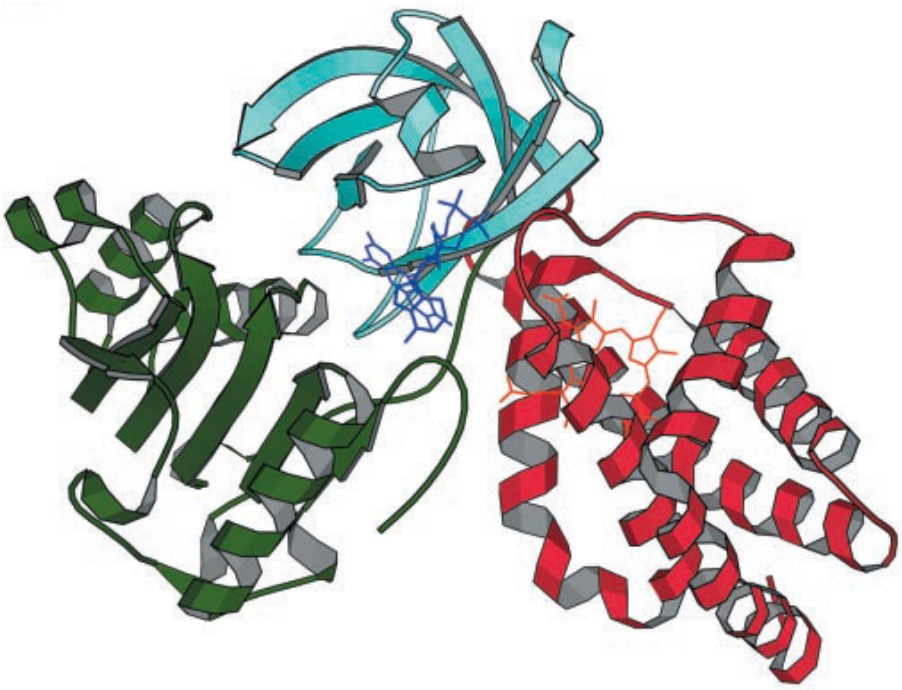

**Figure 1.23:** Overall three-dimensional structure of NOD. The flavin-binding domain is located at the upper part (cyan), the globin domain on the lower right side (red), and the NAD-binding domain on the lower left side (dark green). (Bonamore and Boffi, 2008).

In this study, they also observed an unusual distal configuration at the active site of the protein, if compared with other Hbs and Mbs. The distal site of the haem was occupied by an isopropyl group from Leu57 (Figure 1.24). This apolar residue was located just above the axial position of the iron heme, shielding it from possible interactions in what it was speculated as a ligand selection mechanism. The residues Tyr29 and Gln53, which were proposed to play an important role in catalysis (from conservation, and biochemical

studies), were found in a second coordination shell. These two residues were the major candidates for ligand stabilization at the distal site (Ermler et al., 1995). In this way, Leu57, would be stabilizing a penta-coordinate ferric species, although the solved structure in likely to be in a photo-reduced state after the diffraction experiment (Carugo & Carugo 2005). At the proximal site, the canonical histidine (His85) and phenylalanine (Phe88) are found – conserved among all globin proteins.

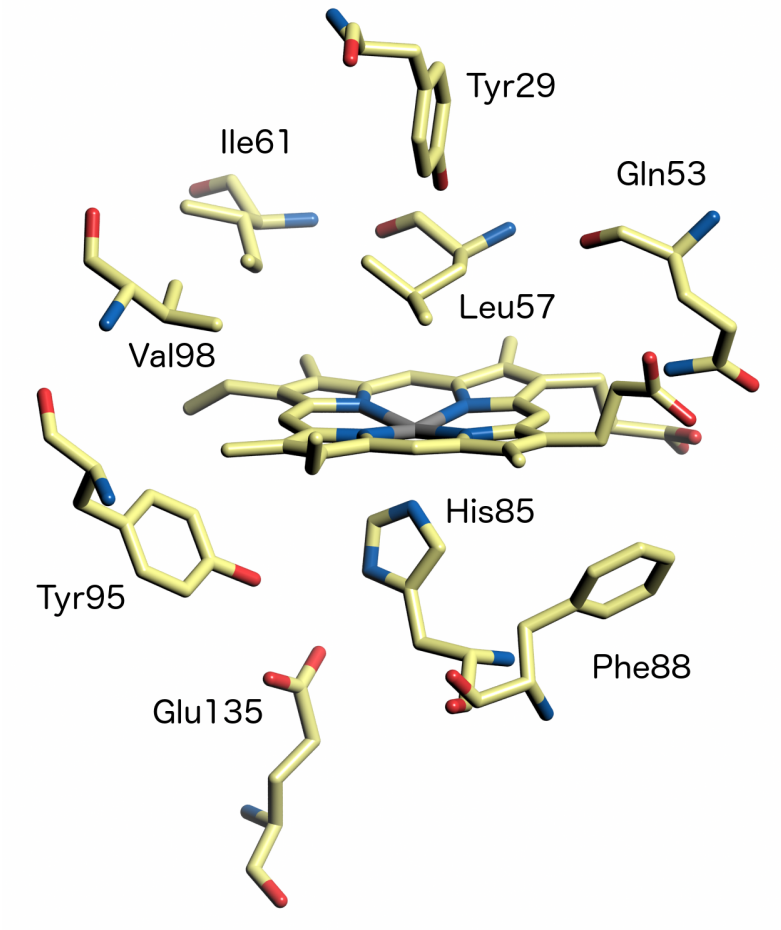


**Figure 1.24:** Structural details of the heme-binding site of the lipid-free five-coordinate structure of NOD, determined from ferric crystals. The heme group is shown together with a selection of amino acid residues located within 5 Å distance from the macrocycle atoms. (PDB entry 1gvh)

Flavohemoglobins usually bind phospholipids, with different affinities depending on the case. For FHP, a tightly bound phospolipid was found in the solved structure occupying the distal pocket at the active site (Ermler et al., 1995; Bonamore et al., 2003). Others, such as Vitroscilla flavoHb (VHP), bind phospholipids weakly and these are lost during purification, and therefore not found in the X-ray structure (Rinaldi et al. 2006). NOD has been shown to bind lipids (Bonamore et al., 2003), but they were removed by Ilari and co-workers (2002) as discussed to help in the crystallization process. These observations suggest that flavoHbs bind lipids when potentially associating with the inner membrane of the cell, and in this way neutralising NO· molecules diffusing through into the cell cytosol. This is supported by the alternative role that has been reported for NOD, which would be involved in the reduction of lipid chains, therefore carrying out a reparation function upon oxidative stress (Bonamore et al., 2003).

# Chapter 2 *In crystallo* spectroscopies and serial crystallography (SX) methodologies.

In this research project, we had the opportunity to apply state-of-the-art techniques and methodologies, such as *in crystallo* spectroscopies and serial crystallography, which have been crucial for:

- Unambiguous determination of the state of the protein to validate a certain structure – *in crystallo* or single crystal spectroscopies (SCSs).
- Measurement of damage-free structures of metalloproteins – serial crystallography applied at XFEL sources, i.e. serial femtosecond crystallography (SFX).
- Follow radiation-induced processes at atomic detail – synchrotron serial crystallography (SSX) applying a dose-series approach.

In the present chapter, some background is given for each of the techniques to put them in context, and the experimental capabilities of the beamlines used in this work at the Swiss Light Source (SLS), the European Synchrotron Radiation Facility (ESRF), Advanced Photon Source (APS), and Diamond Light Source (DLS), are presented and detailed.

## 2.1 Single crystal spectroscopies (SCS).

Single crystal spectroscopies (SCSs) are useful complementary techniques for X-ray crystallography (Pearson & Owen, 2009; Ronda et al., 2015), that can be applied *in situ* at synchrotron sources. With them,

spectroscopic information from protein crystals can be obtained, coupled with the diffraction experiments. This kind of information is of great value since it allows us to assign the redox, ligand, or catalytic state in which the protein is (comparing with protein solution data), which can be difficult (or sometimes not possible) to interpret from the electron-density maps. They have been named elsewhere as *in crystallo* Optical Spectroscopy (von Stetten et al., 2015), and they are found and accessible to users (upon request) at many synchrotron sources (Table 2.1).

The use of *in crystallo* spectroscopies is crucial to, for instance, monitor changes in the crystal induced by the exposure to X-rays (radiation damage). Another remarkable use of these tools is to track redox reactions (in the case of enzymes) and detect the build-up of intermediates with a spectroscopic fingerprint that can then be freeze-trapped for structural determination (or measured at a certain point using time-resolved crystallography). This results in very valuable information, since most of the catalytic intermediates can be hard to detect in solution due to the fast rate of the reactions where they are involved. However, the combination of extremely low temperatures used for routine X-ray crystallography diffraction (100 K), and the restricted environment within the crystal lattice, normally slow down the catalytic reactions to allow us in some cases to unveil its intermediates and mechanism steps (Makinen and Fink, 1977; Schlichting et al., 2000). Therefore, single crystal spectroscopies play a crucial role to locate unambiguously a certain structure within an enzymatic mechanism (Gumiero et al. 2011) and avoid wrong mechanistic

conclusions, which may lead to poor biological interpretations and inefficient biotechnological applications.

**Table 2.1:** Single-crystal spectroscopy instruments at synchrotron sources (updated from Ronda et al., 2015).

| Source location | Available equipment |
|---|---|
| Swiss Light Source (SLS), Villigen, Switzerland - Beamline X10SA (PXII) | UV-vis absorption, Resonance Raman and fluorescence - on-axis geometry. |
| BioCARS, Chicago, IL, USA- Beamline 14-BM-C | On-line 4DX micro-spectrophotometer. |
| ESRF, Grenoble, France (MX beamlines) | On-line micro-spectrophotometer for UV-vis absorption, fluorescence and Raman measurements. CryoBench, for off-line measurements also available. |
| National Synchrotron Light Source (NSLS), Upton, NY, USA – Beamline X26C | On-line 4DX micro-spectrophotometer for UV-vis absorption and Raman measurements. |
| Diamond Light Source, Oxfordshire, UK – MX beamlines | On-line and off-line UV-vis absorption measurements. Available for users upon request. |
| Spring-8, Hyôgo Prefecture, Japan – BL38B1 beamline | UV-vis absorption measurements. |

Since this set of different tools for the crystallographer (e.g. UV-vis, fluorescence and Raman spectroscopies) have become much more available at synchrotrons (and also some lab X-ray sources), there is an ongoing debate within the structural biology community of whether if spectroscopic validation could become a useful complement (or even a requirement in the future) for

the deposition of protein structures (Garman & Weik 2017; Stoner-Ma et al. 2011), adding important information such as whether if the protein has been photo-reduced by the X-ray beam.

Single-crystal spectroscopic techniques applied to crystals were developed in the 1990s by Hadfield and Hadju (1993), and they continued to be developed to this point, when they are present at most synchrotrons. The interest towards these tools and its development has been powered by two main factors:

i) The observation that many enzymes are still able to catalyse reactions in the crystalline state, even if the conditions needed for crystallization are often not ideal for catalysis (e.g. pH, high ionic strength and viscosity; Mozzarelli & Rossi, 1996).

ii) The observation that crystals are affected by the irradiation with the X-ray beam, leading to changes in the redox state and structure of the protein under study during data collection, even at low doses (Pearson et al., 2004; Carugo & Carugo, 2005).

### *2.1.1 Range of spectroscopic techniques*

Each of the different spectroscopic techniques that can be applied to crystals has its strengths, as well as its limitations. They offer complementary information to the structure of a given protein system, which can be combined to answer our research questions. Micro-spectrophotometers (or commonly microspecs) are the instruments used at synchrotron sources to perform single

crystal spectroscopies (SCSs). When they are mounted directly on the beamline to do measurements *in situ* is known as ‘online mode’, with the aim being to probe the region of the crystal exposed to X-rays. Online spectroscopic measurements are of great value to assess the state of a protein within a reaction mechanism, and also to monitor the extent of photo-reduction of a crystal by the X-ray beam. Measurements can be done before or after diffraction experiments in a different location within the synchrotron, in what is known as ‘offline mode’. The offline mode is very useful for the time-consuming tasks of characterization, sample preparation and optimization of experimental procedures. Depending on how microspectrophotometers are installed at the beamline, their geometry can be ‘on-axis’ if the light path coincides with the X-ray beam, or ‘off-axis’, if the light path is situated on a different plane from the beam. The on-axis geometry ensures that the same region of the crystal exposed to the X-rays is probed. In the case of the off-axis set up, it is installed and removed upon need, but it needs to be carefully aligned to probe the same volume of crystal exposed by the beam. With large crystals, this task is particularly difficult, resulting sometimes in the unavoidable contribution to spectra of unexposed regions.

In Table 2.2, the main single-crystal spectroscopic techniques used at synchrotron sources are summarised. As it can be noted, a variety of methods exist, and the presence of vibrational methods entails that a chromophore group in not always needed to be able to obtain spectroscopic information from a certain protein. However, one of the most used vibrational methods is single-

crystal resonance Raman (SCRR), which is based on the laser excitation of a chromophore within the protein to amplify its signal.

**Table 2.2:** Overview of the main spectroscopies currently used in combination with X-ray crystallography, the information that can be obtained from them and their limitations. (Pearson & Owen 2009)

| Spectroscopy | Information obtained | Limitations |
|---|---|---|
| **UV-visible** | Oxidation and ligand states. | Requires a chromophore; optically dense crystals problematic. |
| **Fluorescence** | Local environment, conformational changes. | Requires a chromophore. |
| **Raman** | Vibrational motions of atoms, molecular conformations, distributions of electrons in bonds. | Long acquisition times due to weak signal. Assignment of peaks very complex. Can be dominated by signal from cryoprotectants. |
| **Resonance Raman (RR)** | Vibrational states of aromatic centres, metal centres or coloured chromophores. | Requires a chromophore and laser with emission wavelength matching absorption maximum at sample. |
| **X-ray absorption (XAS)** | XANES (X-ray absorption near-edge structure): oxidation state, covalency, site symmetry.<br><br>EXAFS (Extended X-ray absorption fine structure): radical distances, co-ordination numbers, types of ligand. | Concentrations > 0.1 mM required as well as homogeneous metal site structure. |
| **Fourier-transform Infra-red (FTIR)** | Vibrational motions of atoms, molecular conformations, distributions of electrons in bonds. | Bands of interest can be obscured by intense water bands. |

### *2.1.2 SCS instrumentation at synchrotron beamlines.*

As shown in Table 2.1, there are currently several synchrotron sources with micro-spectrophotometers installed to perform *in crystallo* spectroscopies upon request. These include: the Swiss Light Source (SLS; Owen et al., 2009; Pompidor et al., 2013; Dworkowski et al., 2015), the European Synchrotron Radiation Facility (ESRF; Carpentier et al., 2007; McGeehan et al., 2009; von Stetten et al., 2015), the National Synchrotron Light Source (NSLS, New York; Stoner-Ma et al., 2010; Orville et al., 2011), Spring-8 (Japan; Shimizu et al., 2013), BioCARS (Chicago; Bourgeois et al., 2002; Pearson et al., 2007), and Diamond Light Source (DLS; Owen et al., 2011).

In the present work, UV-vis absorption and Resonance Raman were applied to haem protein crystals, and other spectroscopies were also tested, such as fluorescence, and Raman. Therefore, a brief description will be given for the instruments used at SLS and ESRF, where our methodologies involving X-ray crystallography and *in crystallo* spectroscopies were applied.

#### *2.1.2.1 SCS at MX Beamlines at ESRF*

At the European Synchrotron Radiation Facility (ESRF) in Grenoble single-crystal spectroscopic experiments can be conducted both online and offline. The available micro-spectrophotometer at the facility (McGeehan et al. 2009), can be installed at any MX beamline at ESRF upon request. In our case, online experiments were carried out either at beamline BM30 or at beamline

ID30-A3. Once mounted, the instrument presents an off-axis geometry (Figure 2.1), with the light path on a different plane with respect to the X-ray beam, and at 45º of the goniometer axis γ, to allow sample exchange and access for the cryostream. Thus, due to its geometry, further alignment tasks have to be performed to ensure that same region of the crystal exposed by the beam is probed by the optical path.

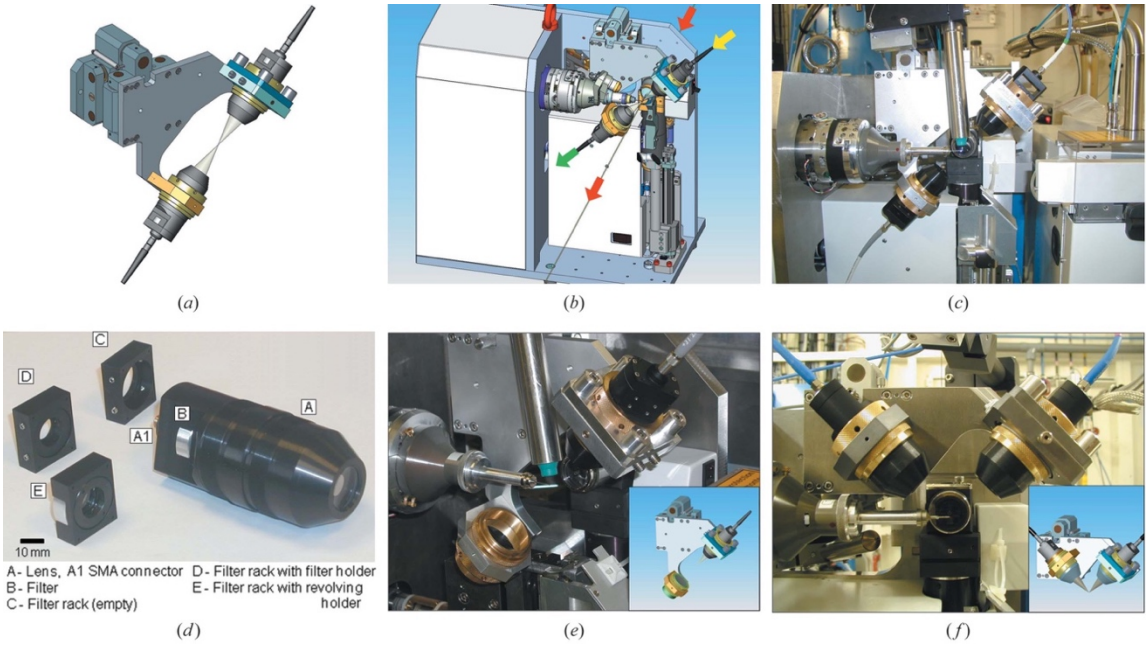


**Figure 2.1:** Portable online microspectrophotometer available at ESRF. (a) Design of the lenses with focal cones depicted in light grey. A CAD drawing of the instrument mounted on a MD2M diffractometer is shown in (b). The path of the X-ray beam is highlighted as red arrows and can be seen to intersect with the light path of the microspectrophotometer (yellow and green arrows). A typical installation is shown on ID14-2 (c). The objectives are mounted at 45º with respect to the γ axis to allow access for the cryostream, the X-ray fluorescence detector and the sample changer. A photograph of the device objective lens and accessories is shown in (d). (e) Single objective design is shown, where the second lower-objective has been replaced by a concave mirror (f) Set-up for fluorescence measurements. (McGeehan et al., 2009)

Offline experiments, can be performed at ESRF at a spectroscopic facility, called the ‘Cryobench’, which has been in constant development since 1999 (Royant et al., 2007; McGeehan et al., 2009; von Stetten et al., 2015), and is located within beamline ID29 (Figure 2.2). The upgraded microspectrophotometer can carry out *in crystallo* measurements at RT or 100 K using a range of spectroscopic techniques: UV-vis absorption (McGeehan et al., 2009), fluorescence (Bourgeois et al., 2002), Raman (Carpentier et al., 2007) and resonance Raman spectroscopies (Figure 2.2). All the different micro-spectrophotometers at the cryobench can be mounted online when needed at any of the MX beamlines at ESRF, to monitor changes in crystals induced by X-ray radiation. The use of this facility is especially suited for the time-consuming tasks of initial spectroscopic characterization and optimization of experimental conditions, making the subsequent time at the beamline more efficient and productive.

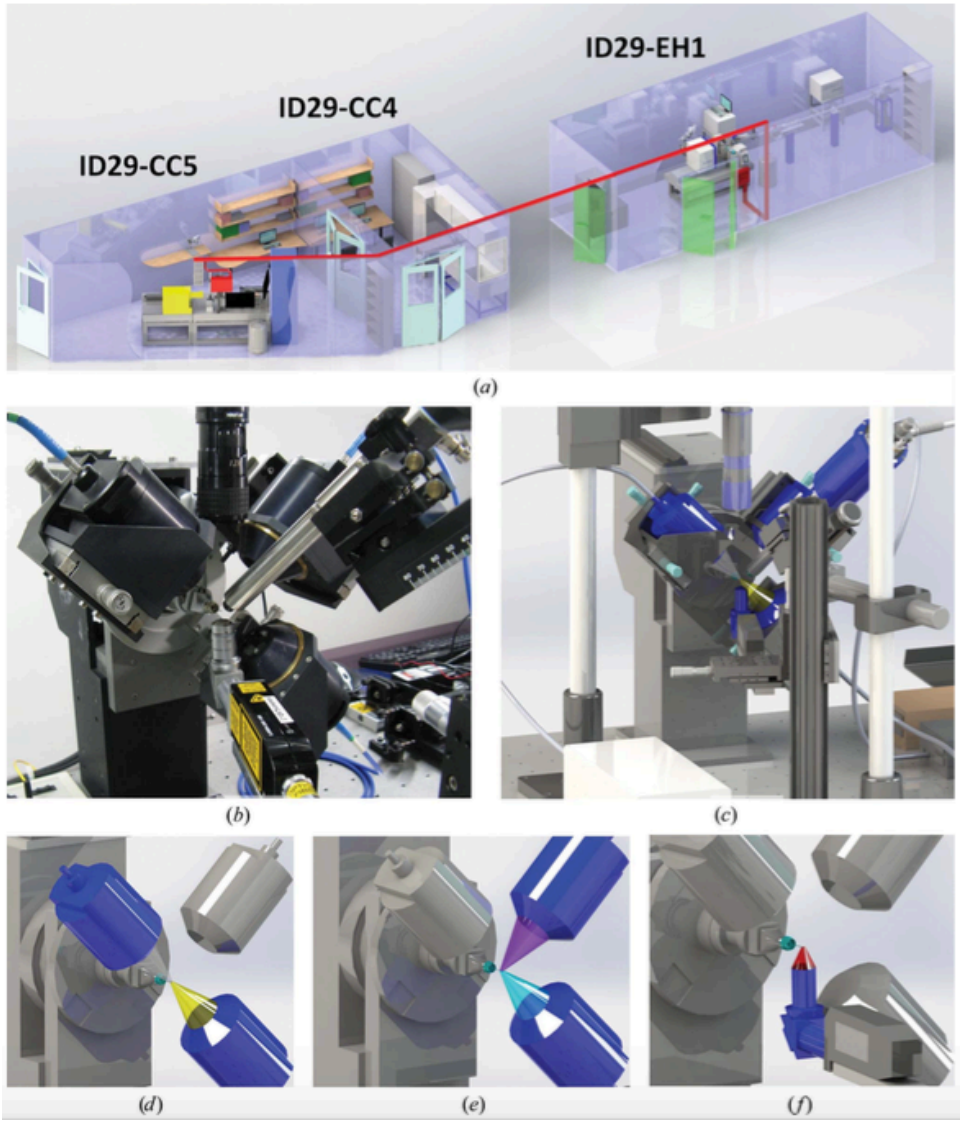


**Figure 2.2:** Cryobench facility at ESRF. a) general layout and situation of the cryobench within beamline ID29, with its control room CC4 and experimental hutch CC5. A Raman fibber is derived from experimental hutch EH1. Photograph in b) or virtual representation in c) of the general disposition of the different microspectrophotometers situated around the sample environment. d) Objectives for UV-vis absorption (transmission mode, 0º disposition). e) Objectives for fluorescence spectroscopy (reflection mode, 90º disposition). f) Objective for Raman spectroscopy (backscattering mode, 180º disposition). All images from von Stetten et al., 2015.

### *2.1.2.2 Beamline X10SA at SLS*

At the Swiss Light Source (SLS), the instrument MS3 is permanently installed online at beamline X10SA (Figure 2.3), able to perform UV-vis

absorption, fluorescence, Raman and Resonance Raman spectroscopic measurements. It features a unique on-axis geometry, where the optical path and the X-ray beam are coaxial, ensuring that the same crystal region that exposed to radiation is probed in the spectroscopic measurements. However, the on-axis geometry has a drawback, which is that no absorption spectra can be measured during data collection since the illumination lamp, lying downstream of the sample, needs to be retracted before diffraction experiments.

The current version of the micro-spectrophotometer is the result of upgrades to improve its capabilities (Owen et al., 2009; Pompidor et al., 2013; Fuchs et al., 2014), and is used as well for sample visualization. It consists of a multi-mode instrument, which can switch between functions without the need of unmounting the sample from the goniometer (Fuchs et al. 2014). The ability to rapidly exchange modes is very powerful as the sample is not altered in any way, i.e. avoiding changes in temperature during manipulation, and precise probing of the same region of the crystal by MX and SCS with reduced alignment time. Since 2010, a Pilatus detector 6M (DECTRIS, Ltd.) can collect datasets in sub-minute timescales. This is especially interesting for observing intermediate species within the crystal whose lifetime are within this time range. The general layout of the experimental set-up at beamline X10SA with the MS3 instrument is shown in Figure 2.3. An off-line facility called the *SLSpectroLab* (www.psi.ch/sls/pxii/spectrolab), near to beamline X10SA, is also available for spectroscopic measurement without the use of X-rays (Pompidor et al., 2013).

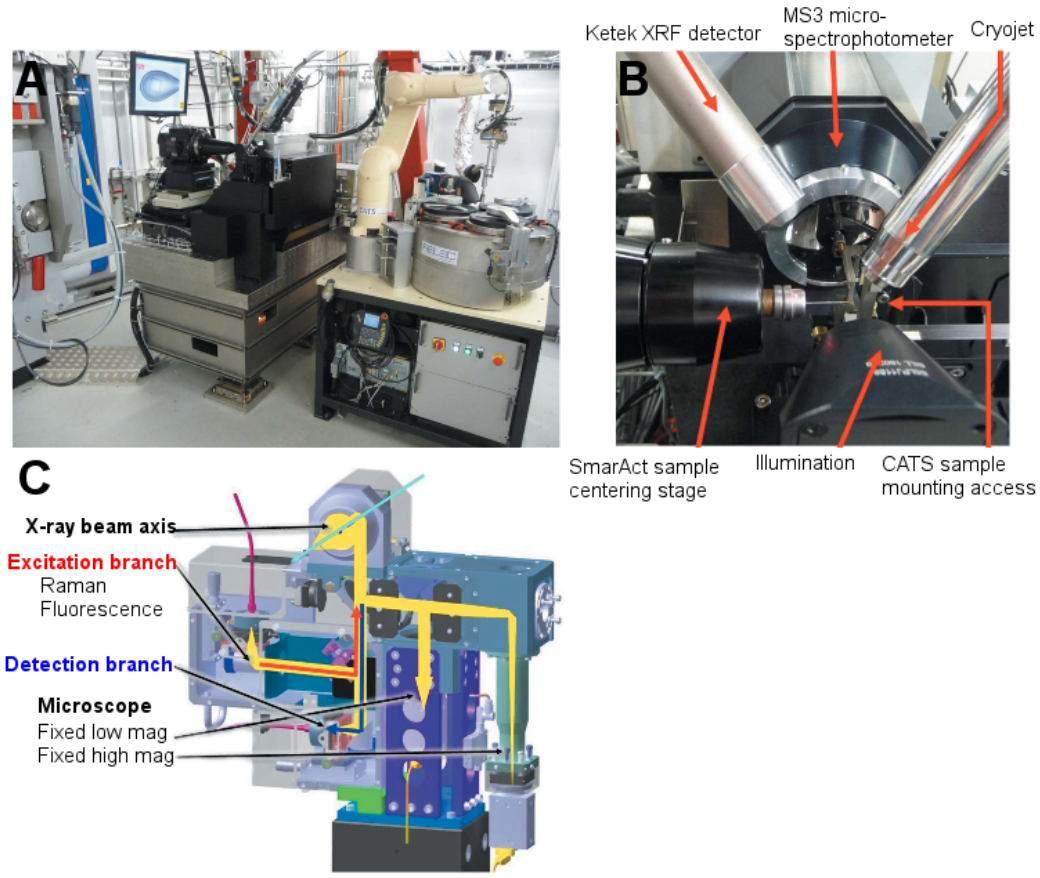


**Figure 2.3:** Beamline X10SA multi-mode MS3 micro-spectrophotometer configuration. (A) Layout of the endstation at beamline X10SA including (from left to right); the Pilatus 6M area detector, D3 diffractometer and the CATS robotic sample changer system. (B) The sample environment of the D3 diffractometer and the closest devices. (C) The MS3 micro-spectrophotometer and its beam pathways (adapted from Fuchs et al., 2014).

### *2.1.3 UV-vis absorption spectroscopy*

UV-vis spectroscopy is the most widespread and widely used among all single-crystal spectroscopies, since it is relatively easy to apply and provides valuable information about the state of the protein (Pearson et al. 2007). Indeed, it was also the first spectroscopic approach used in protein crystals. The most important limitation of this technique is the need of a chromophore

within the protein. This limitation is also common for fluorescence and Resonance Raman. A search in the Protein Data Base done by Stoner-Ma et al. (2011) found that of all the deposited structures until 2010, 11,000 contained a coloured co-factor (approximately 15%). This fact supports the broad applicability of UV-vis absorption to protein crystals.

The first UV-vis online micro-spectrophotometers were developed in the early 1990's to obtain additional information from protein crystals . This design eventually led to the commercial instrument 4DX-ray (SystemsAB, Uppsala, Sweden). This version was implemented in many synchrotrons, for instance at the ESRF cryobench when it was firstly set up in 1999. After that, further developments were done in different facilities such as the on-axis spectrometer at SLS, the different updates and evolutions at the cryobench, or the recently commissioned UV-vis at Diamond light source.

Measuring UV-vis absorption is of great help for determining the redox, ligand or catalytic state of the chromophores within proteins. An example of the applicability of this technique to follow reactions happening *in crystallo* is shown in Figure 2.4. Single-crystal UV-vis spectroscopy can also be used to assess radiation damaged in crystals (e.g. photo-reduction, build-up of radicals and solvated electrons). Another interesting use, owing that many proteins are active in the crystalline state, is for detecting unstable reaction intermediates that have a singular spectroscopic fingerprint (Figure 2.4). In this way, biochemical mechanisms can be followed at real-time within the crystal. For

this reason, UV-vis spectroscopy is an important tool used in experiments aiming at elucidating reaction mechanisms (Berglund et al., 2002).

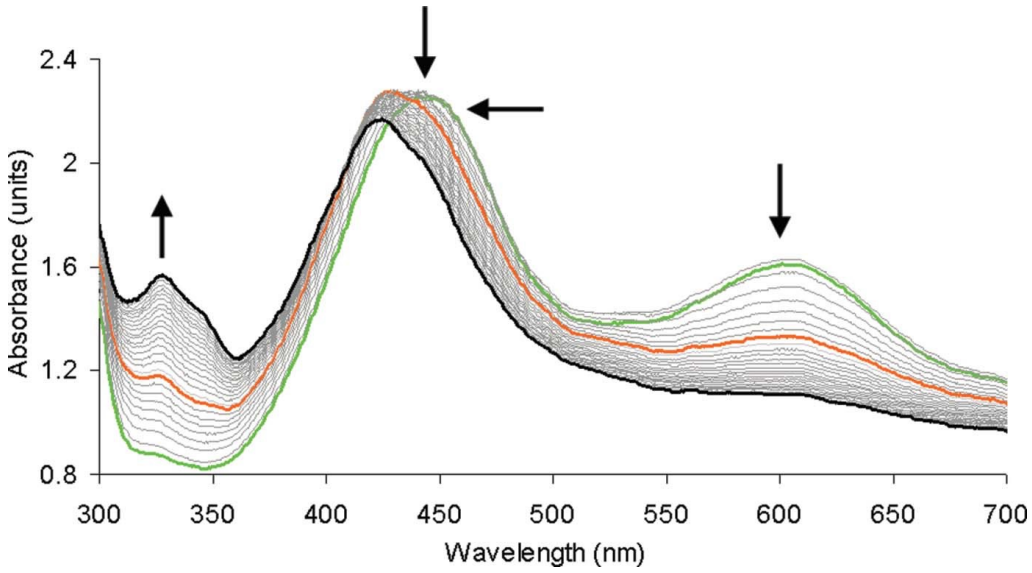


**Figure 2.4:** Reduction of oxidized O-quinone MADH/amicyanin protein crystals (TTQOX/ Cu2+) by X-ray illumination. Arrows indicate the direction of change for the main features. The starting and ending spectra are shown in green and black, respectively; the orange spectrum shows maximum content of the intermediate species TTQOSQ (Pearson et al., 2007).

### *2.1.4 Raman spectroscopy.*

Raman spectroscopy gives us information about vibrational states of the sample under study. It is useful for detecting ligands and chemical species in the protein by assigning characteristic bands when comparing with crystals where the substance is not present (Katona et al. 2007). There are two main approaches: non-resonance and Resonance Raman.

Non-resonance Raman does not need the presence of a chromophore in the system, thus makes spectroscopic measurements possible for all kind of proteins. When applying this technique, the whole protein (or biomolecule probed) contributes to the signal, which results in complex spectra with multiple

peaks to analyse, representative of the different vibrational modes of the bonds between the atoms forming the sample (Figure 2.5). In addition, spectra can be easily dominated by bands from substances used for cryoprotection. Therefore, to be able to detect protein changes with non-resonance Raman, we need a high signal-to-noise ratio, and this signal must not be masked by stronger non-relevant bands.

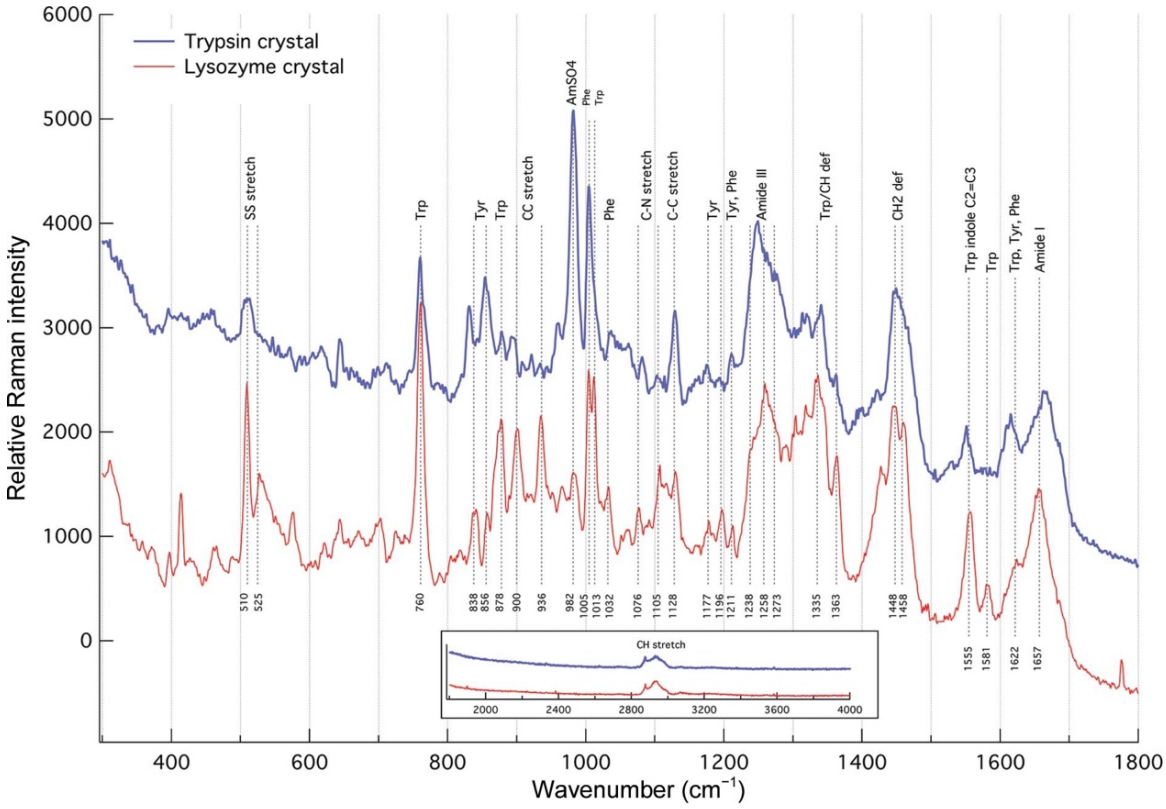


**Figure 2.5:** Non-resonance Raman spectra for trypsin and lysozyme crystals. Note the high abundance of peaks and the presence of a strong ammonium sulphate signal coming from the crystallization condition in the case of trypsin. (von Stetten et al., 2015)

In Resonance Raman, the signal from a chromophore in the protein is enhanced by laser excitation at absorbing wavelengths. In this way, the observed signals are related to the chromophore group and its direct surroundings (Siebert & Hildebrandt, 2008). In previous studies carried out by

Kekilli et al. (2014), it was shown how using Resonance Raman it is possible to unambiguously characterize redox and ligand states of the same protein, where only use of X-ray crystallography would be not sufficient to tell between states. The study was done with a haem-containing cytochrome c' from the denitrifying bacterium *Alcaligenes xylosoxidans* (AxCYTcp), a gas binding protein that discriminates between nitric oxide (NO) and carbon monoxide (CO) by binding them at different faces of the haem group. Using Resonance Raman, Kekilli and co-workers fingerprinted the ligand bound states as well as the redox states adopted by the protein crystals in the different experiments performed (Figure 2.6, Figure 2.7).

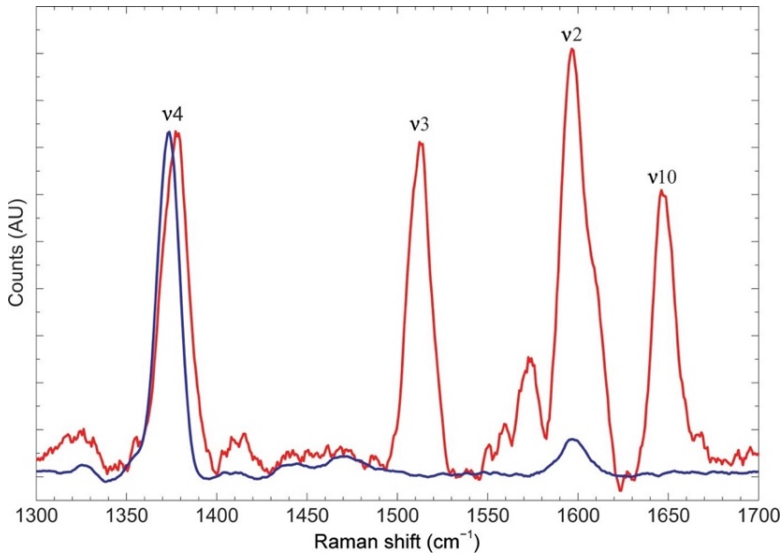


**Figure 2.6:** Single-crystal online resonance Raman spectra of AxCYTcp crystals with CO (in blue) and NO (in red) bound at the active site. (Kekilli et al., 2014)

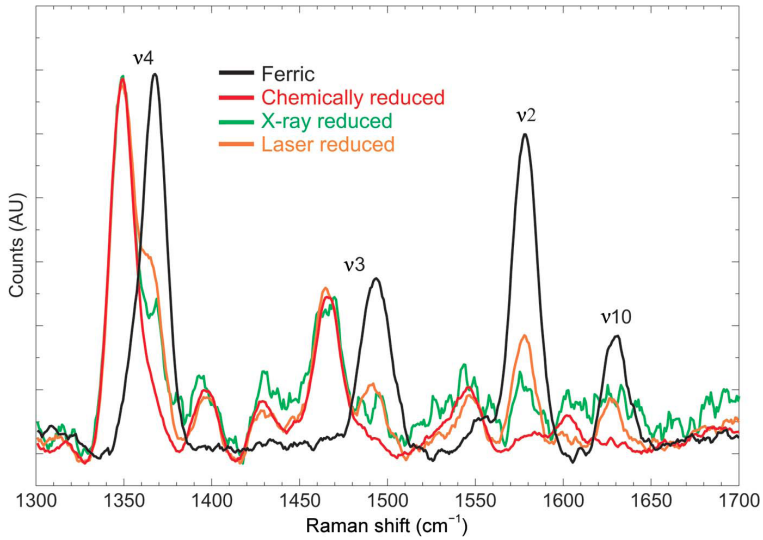


**Figure 2.7:** Single-crystal online resonance Raman spectra of AxCYTcp of crystals in the ferric state and in differently induced reduced states. (Kekilli et al., 2014)

## *2.1.5 X-ray absorption spectroscopy*

X-ray absorption spectroscopy (XAS) can also be a useful complementary technique in order to assign the oxidation state of metal centres by looking at its absorption profile. As an example, Hough and coworkers (2008) monitored the reduction of a Copper Nitrite Reductase by a combination of spectroscopies including UV-vis spectroscopy and X-ray absorption near edge (XANES). The former was used to assess the reduction of one of the copper centres, and the latter was used to monitor the reduction of the second copper centre (Figure 2.8), which had no UV-vis fingerprint. When the protein was reduced, the spectrum would show a shoulder at around 8985 eV not present in the ferric form.

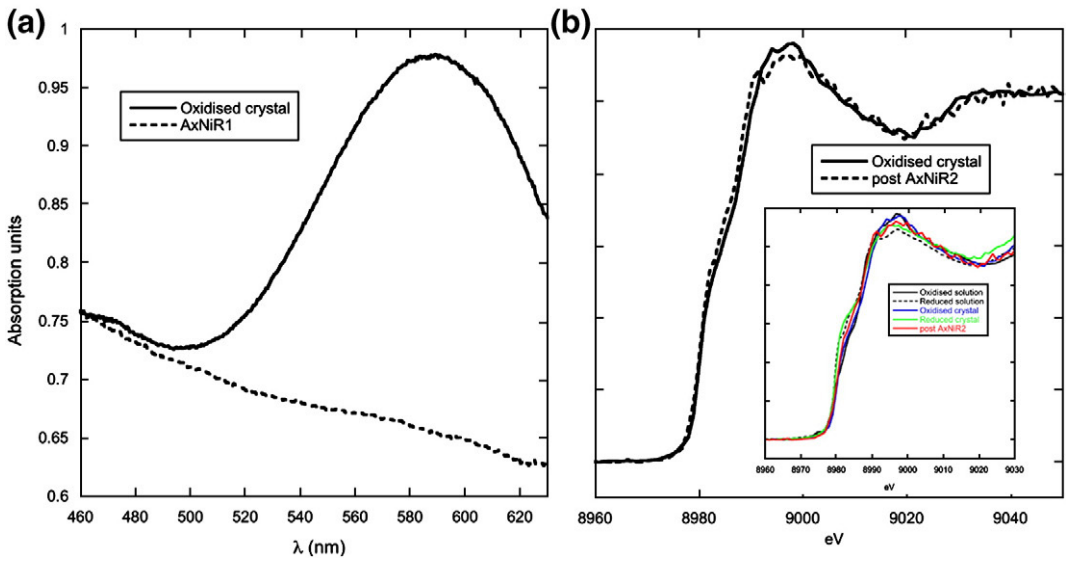


**Figure 2.8:** Combined spectroscopic monitoring of *Alcaligenes xylosoxidans* nitrite reductase (AxNiR) crystals. (a) UV-vis spectra showing the reduction of copper centre 1 after the exposure to X-rays (AxNiR1, in dashed lines). (b) X-ray absorption near edge spectra (XANES) of a crystal before and after exposure (AxNiR2). In the little square within the graph, spectra for oxidised and reduced solution and crystal forms is shown. Note the characteristic shoulder at around 8985 eV when the protein is reduced. (Hough et al., 2008)

## 2.2 Serial crystallography methodologies

During the PhD project the opportunity arose to apply the emerging method of serial crystallography (SX, Chavas, Gumprecht, & Chapman, 2015) to our protein targets. SX was initially developed to be applied to satisfy the sample delivery requirements at X-ray free electron lasers (XFELs). Due to the extreme intensity of the pulses at these sources, protein crystals are destroyed upon interaction with the X-ray beam. For this reason, a sample delivery technique was needed able to present fresh crystals at each XFEL pulse. Therefore, the way to grow crystals had to be changed as well, and optimized to deliver large amounts of microcrystals of suitable sizes for the SX experiments.

Importantly, serial methodologies also allow the determination of room-temperature structures, with the subsequent advantages that this carries (Fraser et al. 2011; Russi et al. 2017), such as the increase in protein dynamics and the conformational diversity, observing protein conformations of more biologically relevance, which may provide new insights into protein function and catalysis.

### *2.2.1 X-ray Free Electron Lasers (XFELs)*

XFELs are the latest generation of X-ray source, produced by a linear accelerator installation, such as the one at SACLA (Figure 2.9), with unique features:

- High brilliance of the beam – at peak brilliance, nine orders of magnitude higher than at conventional synchrotron sources.
- Extremely short pulses – in the range of 5-50 femtoseconds.
- Highly coherent radiation.

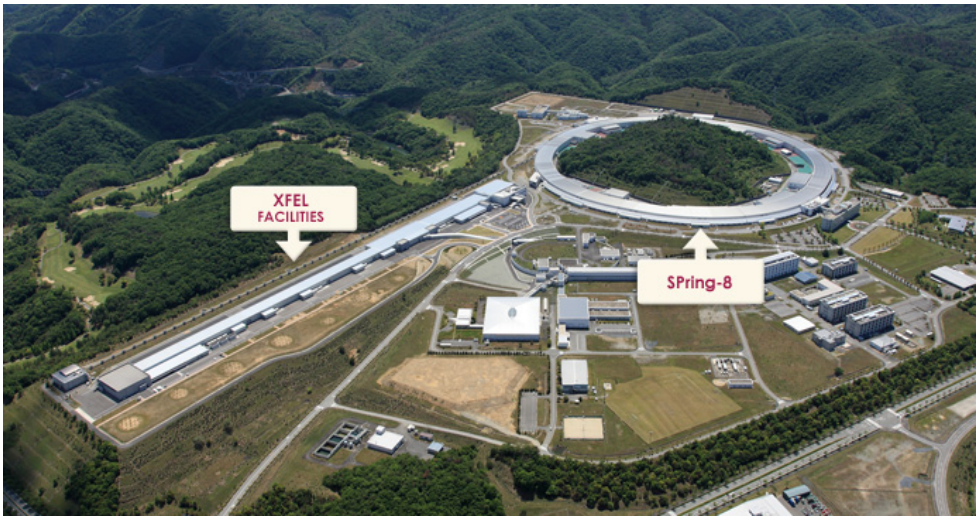


**Figure 2.9:** XFEL facilities at SACLA, Japan. The linear accelerator at the XFEL facility SACLA can be appreciated, within Spring-8 synchrotron complex, in Hyogo prefecture, Japan. The experimental hutch is located in the section closest to the synchrotron ring, whereas the X-rays are generated at the other end.

One of the main attractions of using an XFEL source is the obtention of damage-free structures of protein and other biomolecules (Neutze et al., 2000), after merging the diffraction from several thousands of crystals (Chapman et al. 2011). This is accomplished due the extremely short femtosecond X-ray pulses generated at these facilities, which outrun the radiation damage processes – the so-called diffraction-before-destruction approach. Due to these unique features, this methodology used at XFELs is known as Serial Femtosecond Crystallography (SFX).

The other application of XFELs, which has opened new roadways in the study of the structure of biomolecules, is the possibility of performing time-resolved crystallography experiments with femtosecond resolution (Barends et al. 2015). This has allowed to see novel protein motions and intermediates upon excitation for example with a laser (pump), followed by measurement at

different time-frames with the XFEL pulses (probe), in the so-called pump-probe methodology (Figure 2.10). In the literature example cited, Barends and co-workers could follow ultrafast changes occurred at myoglobin upon photolysis of its CO-bound complex. Therefore, application of time-resolved crystallography at XFEL sources is allowing the scientific community to ´see' biochemical processes in a way that was not possible before, building-up structural movies of protein dynamics and transforming X-ray crystallography into not just the provider of static averaged structures of biomolecules.

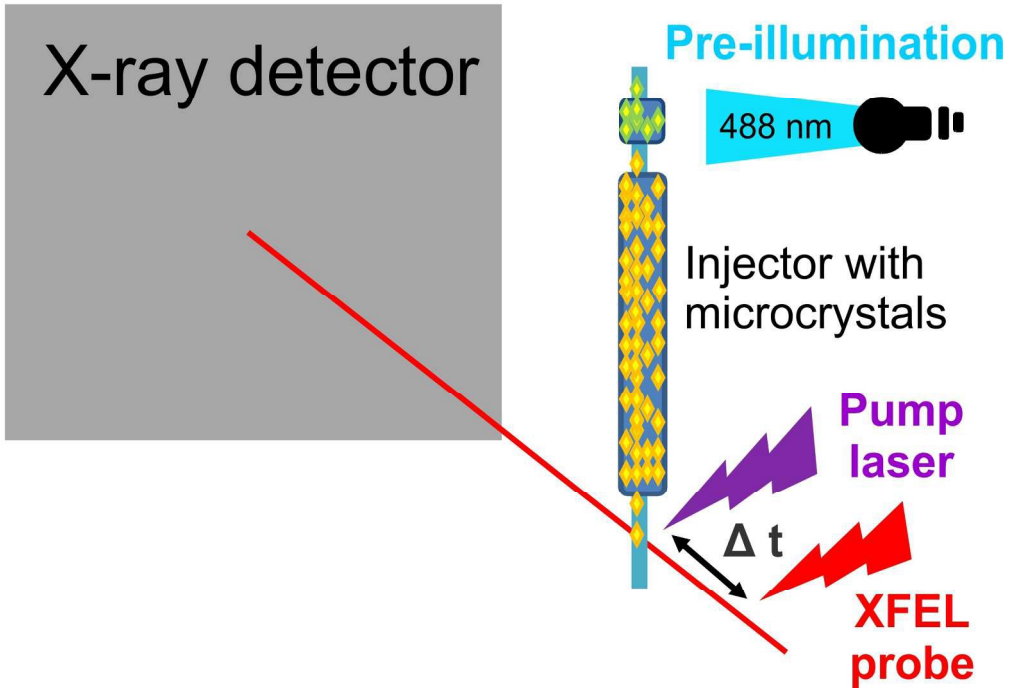


**Figure 2.10:** Scheme of the pump-probe methodology, where the microcrystals are excited with a laser (pump) a controlled amount before (Δt) interaction with the XFEL probe. (Domratcheva & Schlichting 2018)

### *2.2.2 SFX methodologies*

As a new emerging technique, the application of SFX has been approached by different methodologies differing in how the sample is delivered to the beam, each of them with its pros and cons. Generally, they can be classified in two main groups:

- Liquid or viscous **jets** – where the sample is in movement within a transporting medium which passes across the beam.
- **Fixed targets** – where the sample is applied, or grown, onto a support which is screened by the beam in a serial manner.

Gas focused liquid jets, where the sample delivery method used for the determination of the first SFX protein structures (Boutet et al. 2012; Chapman et al. 2011), performed at the Linac Coherent Light Source (LCLS; California, USA). A jet of several microns wide would be directed to pass through the XFEL beam path, carrying microcrystals, one at a time, that would be diffracted at random orientations. A schematic representation of the experimental set up is shown in Figure 2.11. Of course, the reliable production of microcrystals small enough to be jetted (smaller than 10-15 μm) is needed, and the sample consumption for the determination of a single structure is high, since the hit rate is low for this technique. This is one of the big disadvantages of this sample delivery method. The great advantage though, is the possibility to perform time-resolved crystallography experiments, controlling the time since excitation (pump) until interaction with the beam (probe). This same methodology has been applied to membrane proteins, embedded in a highly viscous lipidic

medium, and delivered to the XFEL to beam using an extruder with a much slower sample flow (Botha et al. 2015). Membrane proteins are one of the main targets for SFX structure determination due to its difficulty to grow as large diffracting crystals.

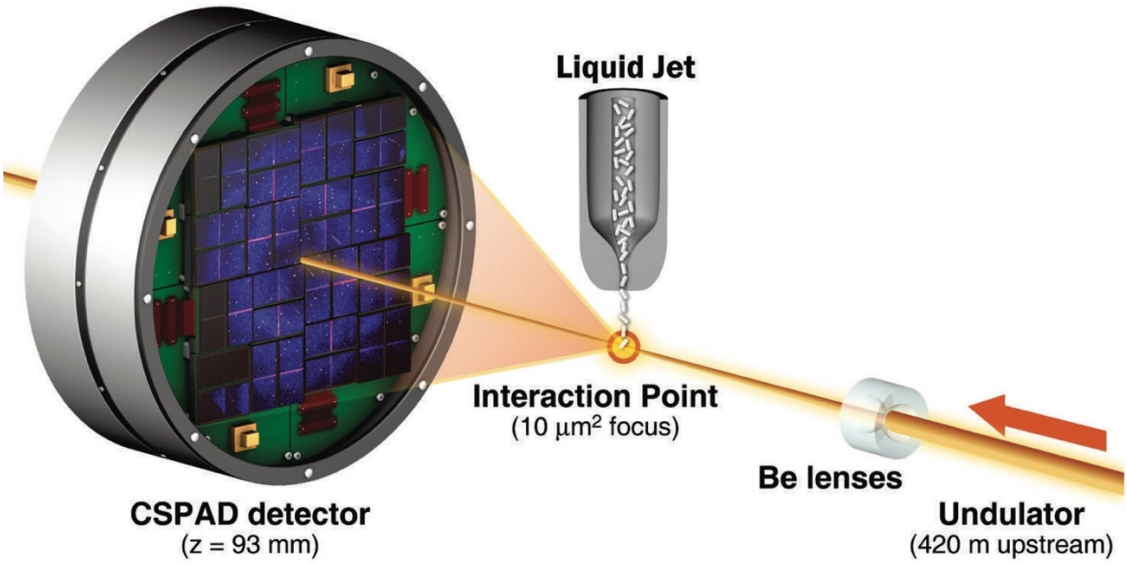


**Figure 2.11:** Experimental geometry used for the serial femtosecond crystallography experiment performed by Boutet et al. (2012) to measure the high resolution XFEL structure of Lysozyme at 1.9 Å. Single pulse diffraction patterns from single crystals flowing in a liquid jet were recorded on a CSPAD detector at a 120 Hz repetition rate at LCLS. Each pulse was focused at the interaction point using 9.4 keV X-rays. The sample-to-detector distance (z) was 93 mm.

The methodology of fixed targets or 'chips' consists in delivering the crystals on a solid support, often made of silicon nitride, where predetermined positions are screened by the XFEL beam in a serial fashion (Frank et al., 2014; Owen et al., 2017). By using this system, there are no limitations in the crystallization conditions used for crystal growing, since high salt or PEG concentrations present difficulties to be jetted. If the chips are measured at RT, a means of protection to prevent the crystals from drying is needed. Different

materials have been used, being the latest one the polymer mylar, forming the windows of a chip-containing frame (Mueller et al., 2015; Ebrahim et al., 2019), and providing also low scattering background. An example of fixed target system is shown in Figure 2.12, where the loading and measurement procedures for a given sample can be appreciated.

The important advantages of using fixed targets are the high hit rates accomplished (depending on the sample even close to 100%), and the consequent low sample consumption. Chips are therefore a highly efficient choice to determine the damage-free structures of metalloproteins yielding intact species, especially if the protein is difficult to express in large quantities, and/or produce a large number of microcrystals. With fixed targets, the size of microcrystals is not a limitation, since they don't have to be jetted, but ideally, they should be in the order of thousands per chip, to have enough data to solve a structure with a low number of chips screened. The downsides of this methodology are: the difficulties to obtain structures of catalytic intermediates, which must remain stable during the time needed for screening (e.g. 7-15 min); and the challenges to perform time-resolved crystallography experiments, since the excitation pump step should be done individually at every measuring point in the chip.

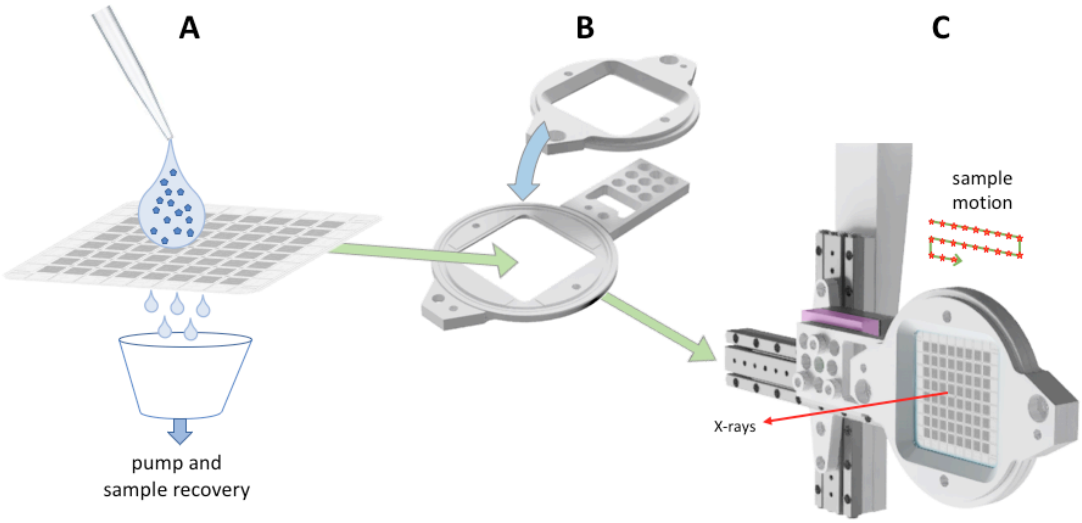


**Figure 2.12:** Schematic of chip loading procedure. (A): microcrystal suspension is pipetted onto the surface of the glow-discharged chip with excess liquid being removed by application of suction to the opposite surface; (B) placing chips into the holder: a thin film of mylar held in place by O-rings seals the chip and prevents drying out; (C) loading of chip and holder assembly onto the beamline sample stage during sample exchange: a kinematic mount (magenta) holds the chip in place in a precise and reproducible position with subsequent alignment carried out using fiducial markings on each chip. The direction of the X-ray beam is indicated as a red arrow while a schematic of chip movement is shown in green. (Ebrahim et al., 2019)

### *2.2.3 Synchrotron serial crystallography (SSX) at I24 (Diamond).*

Serial crystallography has been recently also adapted to a synchrotron beamline environment (Owen et al. 2017) to exploit its characteristics in a more accessible source than the highly competitive XFELs, and at the same time take advantage of a more stable and better characterised X-ray beam. In this case, the technique is referred as Serial Synchrotron Crystallography (SSX), or has been also named elsewhere as Serial Millisecond Crystallography

(SMX; Nogly et al., 2015). As this last name indicates, the length of pulses is in the range of milliseconds (i.e. 12 orders of magnitude longer!). Obviously, radiation damage cannot be outrun with such exposures. However, by using a serial crystallography approach, radiation damage can be greatly minimised since it is spread among all the crystals in the chip. In this way, a low-dose dataset can be obtained (10-50 kGy, Owen et al., 2017, Ebrahim et al., 2019), and in some cases the absorbed dose by the crystal is low enough to observe structural features before any photo-reduction has taken place.

The development of SSX has proven to be very useful to prepare and optimize experiments to be performed at an XFEL source, rendering beamtimes much more efficient and productive. Nevertheless, with SSX one also can perform experiments that cannot be carried out yet at XFELs, such as dose-series measurements (Ebrahim et al., 2019). The fixed target developed by Owen and co-workers installed at beamline I24 (Diamond, Oxford) for SSX experiments is shown in Figure 2.13. It is designed in a compact and modular manner to be easily assembled and used at XFEL sources for SFX experiments as well.

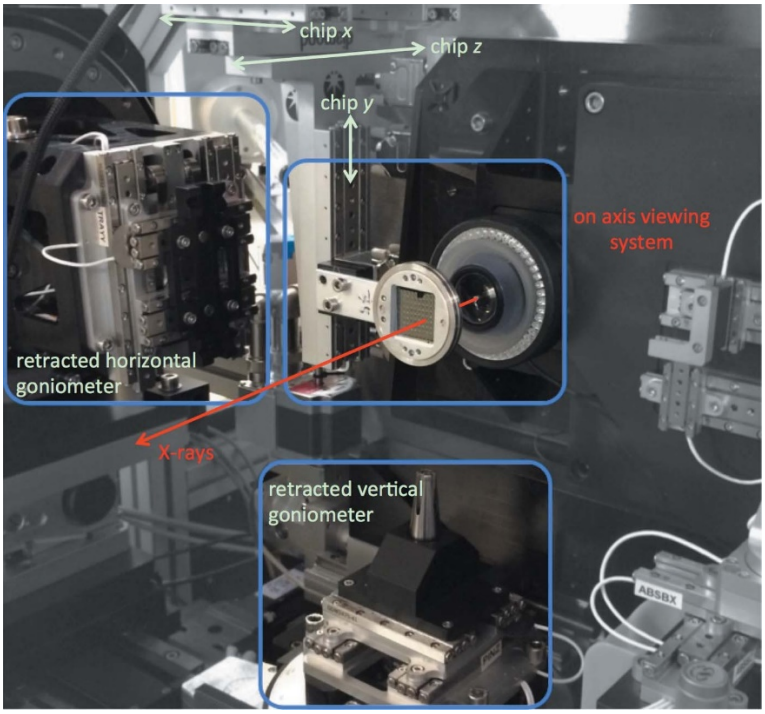


**Figure 2.13:** Instrumentation for fixed-target serial crystallography in situ on beamline I24 at Diamond Light Source. Retraction of both 'standard' goniometers leaves an empty sample environment, allowing the straightforward mounting of hardware for serial crystallography.

### *2.2.4 Pink beam serial crystallography*

Recently, SX has been also applied at a Laue beamline (Meents et al. 2017), BioCARS ID-14B at Advanced Photon Source (APS, Chicago). This X-ray source, also referred to as 'pink beam', is polychromatic with a certain bandpass, producing characteristic Laue diffraction patterns. The experimental set up at APS is shown in Figure 2.14. A fixed target system was used as well in this case, contained in a Helium atmosphere to minimize air scattering background, humidified to prevent the crystals from drying out. Chips were screened through the X-ray beam with the help of a goniometer equipped at the beamline, designed for time-resolved crystallography experiments. Room-

temperature SX datasets were measured from microcrystals from two different proteins (Proteinase K and Phycocyanin), using single-pulse exposures of 100 picoseconds (ps). Therefore, another time frame can be accessed with this technique, in comparison with the ms exposures at synchrotron sources, and time-resolved crystallography experiments can be conducted (Knap et al., 2006; Srajer & Royer, 2008). For a complete dataset, substantially less crystals were needed (e.g. 50-200 crystals), due to the wider bandpass of the X-ray beam, in contrast with the several thousands of crystals needed in a typical SFX experiment. Thus, SX applied at a Laue beamline can be particularly well suited for cases where a limited number of microcrystals are obtained.

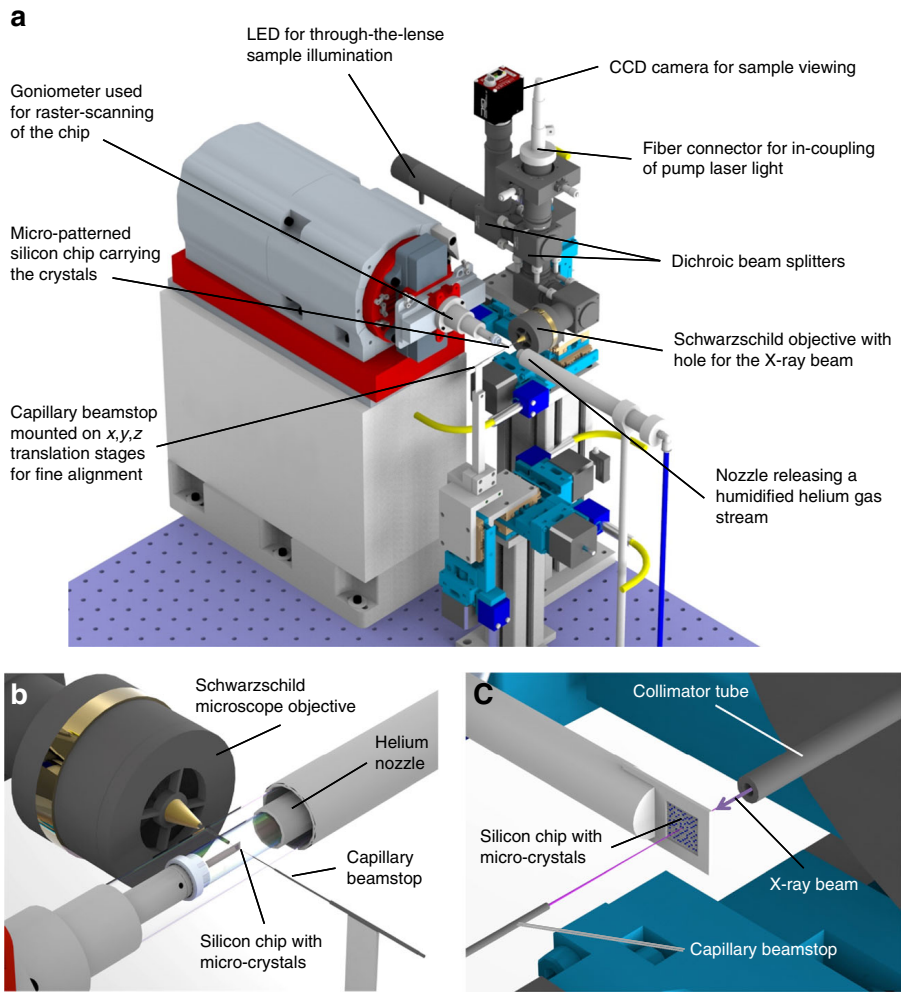


**Figure 2.14:** a) Experimental setup installed at the BioCARS instrument at the APS for fixed target serial crystallography using the pink beam. The micro-patterned silicon chip is raster-scanned through the X-ray beam using the goniometer installed at the beamline. An inline microscope using Schwarzschild optics is used for sample visualization. b) Close-up view showing the chip mounted on the beamline goniometer and being exposed to a stream of humidified helium gas preventing the crystals from drying out. c) Schematic view of phycocyanin crystals mounted onto a micro-patterned silicon chip. The short path of the direct beam between the collimator tube and the capillary beamstop behind the sample is highlighted in pink. By flushing the remaining free beam path with helium, extremely low background scattering levels are achieved. (from Meents et al., 2017)

## 2.3 Thesis aims and objectives.

X-ray crystallography has been for more than 50 years a powerful tool for the study of biomolecules and biochemical processes at atomic resolution. Numerous enzymes have been studied with this technique in order to unveil their structure, interactions with substrates and ligands, and better understand their function. However catalytic processes carried out by enzymes can be of great complexity, involving several steps and intermediates prior to arriving to the final reaction product (Gardner et al., 1998; Schlichting et al., 2000). Therefore, they are difficult to structurally characterize. Especially because one of the limitations of X-ray crystallography is that provides static 'snapshots' of the average conformation within the crystal – making it difficult to study dynamic processes such as enzymatic reactions. Another important limitation is established by radiation damage, which results from the interaction of the X-ray beam with the protein crystal, and results in photo-reduced species which are not fully representative of the state of the measured protein (Garman 2010). This is particularly relevant for metalloproteins such as haem proteins, since their metal centres are prone to reduction upon irradiation (Yano et al., 2005).

In this research project, I aimed to overcome these limitations to study four haem enzymes (DtpA, DtpAa, DHP and NOD), and their catalytic cycles, and try to obtain intact species of relevant functional states. To this aim I used complementary tools, such as those offered by single-crystal spectroscopies, and state-of the-art methodologies such as serial crystallography, applied at XFEL and synchrotron sources. Spectroscopic information of the crystals provided by *in crystallo* spectroscopies was used to validate the structures

obtained by conventional X-ray crystallography, identify generated species, and assign them within the enzyme catalytic mechanism (e.g. peroxidase cycle). My final goal was to structurally characterize the different intermediates involved in the reaction mechanism (e.g. Compound I and II), as ‘time-lapse’ frames to build up a structural movie of enzyme catalysis.

Serial crystallography methodologies were used to measure at XFEL sources (SFX; Schlichting, 2015) room-temperature damage-free structures of the studied enzymes, such as the resting state Fe(III) in haem enzymes, which would be challenging to measure at conventional synchrotron sources. The application of serial crystallography at synchrotron sources (SSX; Owen et al., 2017) delivered low-dose structures of the haem enzymes, and also allowed to apply the MSOX approach to a microcrystal population to carry out dose-series measurements with the intention of generating a structural movie to follow radiation-induced processes at redox metal sites of haem enzymes.

# Chapter 3 Materials and methods.

## 3.1 Molecular cloning.

### *3.1.1 Expression vectors*

Plasmids for recombinant expression of DHP-B (pET16b, Figure 3.1), and for DtpA and DtpAa (pET28a, Figure 3.2) were kindly provided by Dr. Guiladi and Dr. Worrall groups respectively. In the case of NOD, its gene was sub-cloned from a pMMB503EH into a pET28a expression plasmid.

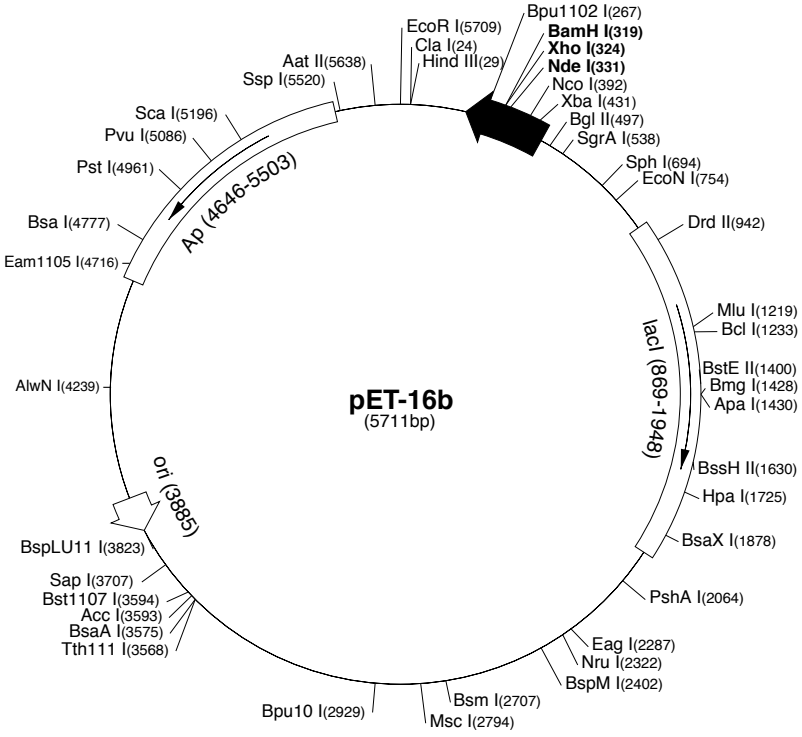


**Figure 3.1:** Plasmid map for the expression vector pET-16b, with Amp resistance.

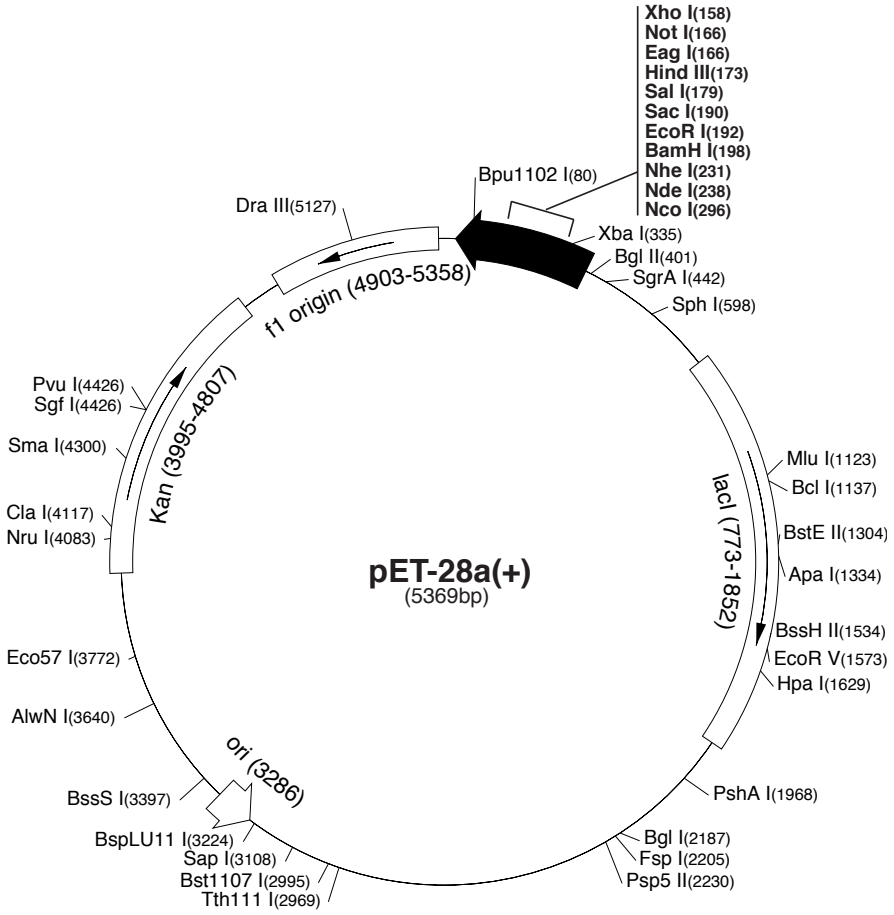


**Figure 3.2:** Plasmid map for the expression vector pET-28a, with Kan resistance.

### *3.1.2 Sub-cloning*

The NOD gene, synthesized by Epoch Bioscience Inc., was amplified by the PCR reaction using designed primers (Eurofins) to introduce NdeI and HindIII restriction sites. The amplicon, as well as the plasmid pET28a (EMD Biosciences), were digested with those restriction enzymes, and the NOD insert was sub-cloned into the vector after ligation, introducing a His-tag at the N-terminal.

### *3.1.3 Ligation*

Ligation reaction was carried out following a general procedure protocol and reactants from Thermo-Fisher Scientific, in order to introduce the NOD gene into the expression vector pET28a (Figure 3.2). A 1:3 stoichiometric ratio of plasmid to insert was used to set up the reaction, which is detailed in Table 3.1. The reaction carried out overnight at 14ºC. A negative control with no NOD insert was also set up in parallel.

**Table 3.1:** Ligation reaction mix for the insertion of the NOD gene into the expression vector pET28a.

| Reagent | Manufacturer | Volume (μl) | Concentration/quantity |
|---|---|---|---|
| NOD insert | Amplified in the lab | 1 | 50 mg (0.02 pmols) |
| pET28a backbone | EMD Biosciences | 4.6 | 37.5 (0.06 pmols) |
| T4 ligase | Thermo-Fisher Scientific | 1 | 5 U/μl |
| T4 ligase buffer | Thermo Fisher Scientific | 2 | 10x |
| Deionized water | - | 11.4 | - |
| | **Total volume** | 20 | |

### *3.1.4 Polymerase chain reaction (PCR).*

The polymerase chain reaction was used to amplify the NOD gene, and to produce different mutants in the case of DtpA (not presented in this work), by site-directed mutagenesis. In both cases, the general set-up reaction was prepared as detailed in Table 3.2. The Pfu polymerase was added last just before starting the PCR reaction, which consisted in a repetition of 18 thermal cycles detailed in Table 3.3.

**Table 3.2:** Reaction mixture for the polymerase chain reaction used in gene amplification and site-directed mutagenesis.

| **Component** | **Manufacturer** | **Concentration** | **Volume (μl)** |
|---|---|---|---|
| DNA template | - | 15 ng/μl | 1 |
| Forward primer | Eurofins | 75 ng/μl | 1 |
| Reverse primer | Eurofins | 75 ng/μl | 1 |
| dNTPs mix | Thermo Scientific | 10x | 3 |
| PCR buffer | Thermo Scientific | 10x | 3 |
| Pfu-turbo polymerase | Thermo Scientific | 2.5 U/μl | 1 |
| Deionised water | - | - | 18.5 |
| DMSO | Fisher | 100% | 1.5 |
| | | **Total volume** | 30 |

**Table 3.3:** Thermal cycling parameters performed during the polymerase chain reaction (PCR). Step 2 to 4 were repeated 18 times.

| **Step** | **Temperature (ºC)** | **Duration** |
|---|---|---|
| 1 - Initial denaturation | 95 | 2 min |
| 2 - Further denaturation | 95 | 50 s |
| 3 - Annealing | $T_m$ of oligonucleotide | 1 min |
| 4 - Elongation | 68 | 13 min |
| 5 - Hold | 4 | - |

### *3.1.5 Heat-shock bacterial transformation and plasmid DNA extraction.*

This technique was used in numerous occasions to introduce the desired plasmid into chemically competent *Escherichia coli* cells. For cloning purposes, XL1-blue strain was used, whereas BL21-DE3 strain was used for over-expression of recombinant protein. The different strains were kept at -80ºC in 100 μl aliquots until use. A standard heat-shock transformation protocol was carried out, detailed in Table 3.4. Once the protocol was finished, cells were spread in LB agar plates supplemented with antibiotic (50 μg/ml Kan for pET28a, 100 μg/ml Amp for pET16b). Plates were incubated overnight at 37ºC and colonies would normally appear the next day. Single colonies were picked up and grown in 10 ml of LB with antibiotic at 220 rpm for a minimum of 18 hours. The pasmid GeneJet Miniprep Kit (Thermo Scientific) was used to extract plasmid DNA following the protocol described by the manufacturer. The purified plasmids were normally stored at -20ºC until needed in subsequent experiments.

**Table 3.4:** Outline of the heat-shock transformation protocol.

| **Transformation stages** | **Condition** |
|---|---|
| First incubation on ice | 30 min |
| Heat-shock at 42ºC | 45 s |
| Second incubation on ice | 2 min |
| Addition of LB media | 750 μl |
| Incubation | 1 h at 37ºC |

## 3.2 Recombinant protein expression

All proteins were expressed in BL21-DE3 cells with pET16b or pET28a vectors with Amp or Kan resistance respectively. A fresh transformation with the plasmid containing the protein gene would be performed before each expression batch. The transformation product would be spread onto LB agar plates with the correspondent antibiotic, and a single colony picked the next day and grown in 10 ml of LB with antibiotic overnight. Cultured cells would be scaled up to 500 ml flasks containing 150 ml of LB media with antibiotic grown at 220 rpm and 37ºC. These cultured cells were used to inoculate the large expression 2 litre flasks, with 1 litre of media each. Expressions were carried out overnight in all cases in Innova-43 or Innova-4300 incubator shakers, at 25-30ºC and 100-200 rpm depending on the protein. Refer to specific protein chapters for detailed information of each expression. In each flask, the appropriate antibiotic would be added before inoculation (50 μg/ml Kan or 100 μg/ml Amp); a heme precursor, which was hemin chloride (0.1 mM) or 5-dALA (0.25mM); an iron source, which was normally iron chloride or citrate (0.1mM); and IPTG (0.5mM), which was added when the cultures have reached an $OD_{600}$ of 0.6-0.8. This was assssed with a Cary-60 (Agilent) UV-vis spectrophotometer. Protein expression was carried out for a minimum of 18 hours, after which cells were pelleted at 5000 rpm for 20 min using a Sorvall Evolution RC centrifuge. At this point, pellets were resuspended in lysis buffer to start the protein purification process, or frozen at -80ºC for future use.

## 3.3 Protein purification

The first step to purify any of the proteins was the lysis of the cells, which was performed with an EmulsiFlex-C5 cell disrupter (Avestin), passing through three times the cells suspended in the lysis buffer to ensure complete breakage of the membrane and release of the protein to the solution. The composition of the lysis buffer was dependent on the purification protocol to be performed. In Table 3.5, composition is shown for a general lysis buffer used for DHP, and the lysis buffer referred as “Buffer A”, used for His-tagged proteins (DtpA, DtpAa, NOD), where a Nickel-affinity column (GE-healthcare) purification step followed the lysis. EDTA was used for the general lysis buffer (when no Ni-NTA column would be used afterwards) to inhibit possible activity of metalloproteases.

**Table 3.5:** Composition of the two kinds of lysis (or resuspension) buffer used for protein purification protocols.

| Component | General lysis buffer | Buffer A |
|---|---|---|
| Buffer | 50 mM Tris pH 8.0 | 50mM Tris pH 7.5 |
| Salt | 100 mM NaCl | 500 mM NaCl |
| Additional additives | 0.1 mM EDTA | 20mM imidazole |
| Reductant | - | - |

For the over-expression of DHP, which was untagged, a protocol from De Serrano et al. (2007) was followed, involving an ammonium sulphate (Thermo-Fisher) salt cut, an ion exchange chromatography (SP-sepharose column, GE-healthcare), and size exclusion chromatography (G75 column, GE-healthcare), using an AKTA purification platform. The protocol for the purification of the rest of the proteins (i.e. DtpA, DtpAa, and NOD), which had a His-tag, was the same, consisting in a His-Trap Nickel-column (GE-

healthcare) purification step, followed by a size exclusion step (G75 or S200, GE-healthcare), again using an AKTA purification platform. The elution buffer normally used for gel filtration purification was: 25 mM HEPES pH 7.5, 100 mM NaCl.

## 3.4 Protein gel electrophoresis and protein concentration

SDS-PAGE electrophoresis was performed after each step, with 15% acrylamide gels, to assess protein purity and select fractions to be pooled for the next step, or for concentration. Gels were run with a Bio-Rad electrophoretic system, typically at 120 V for 90 min. Gels were stained for 20-30 with blue dye, and distained with a solution consisting in: 30% methanol, 10% glacial acetic acid, and 60% of RO water. Distaining solution was changed every ~30 min until bands were visible and well defined. SDS-PAGE gels were inspected with the help of a lightbox.

When protein sample was pure, it was oxidised by use of an excess of potassium ferricyanide (Acros organics), since it was important for our purposes to have a homogeneous oxidation ground state. For removal of the oxidant, as well as for carrying out buffer exchange of the protein when needed, PD-10 columns were used (GE-healthcare). As a final step, the pooled purified protein was concentrated using Vivaspin-10 devices (Sartorius) until reaching the desired concentration, which was usually around 15 mg/ml. The concentration in each case was determined by UV-vis spectrophotometry, as detailed in the next section.

## 3.5 UV-vis spectroscopy analysis

For all the proteins, a Cary-60 UV-vis spectrophotometer (Agilent) was used to determine concentration of protein samples or to establish their oxidation/ligand or catalytic state. Experiments were carried out at RT, acquiring spectra in the wavelength range of 200-800 nm, and using a quartz cuvette (Hellma-Analytics) with 1cm light-path.

Protein concentration was determined by applying the Lambert-Beer equation (Eq.1), where $A$ is the absorbance of the sample, $\varepsilon_\lambda$ the extinction coefficient at a certain wavelength, $l$ is the light-path, fixed to 1cm; and $c$ is the concentration of the protein sample.

$$A = \varepsilon_\lambda \cdot l \cdot c \quad \text{(Eq.1)}$$

For DtpA and DtpAa, which concentration was measured at 280 nm, the extinction coefficient $\varepsilon_{280}$was estimated using the Swiss bioinformatics server ExPASy (Gasteiger et al. 2005), introducing as input the protein amino acid sequence. For NOD and DHP, extinction coefficients at the ferric Soret band, $\varepsilon_{403}$ and $\varepsilon_{406}$ respectively, were obtained from the literature (de Serrano et al, 2007; Helmick et al., 2005). The purity of the protein was also assessed by means of the UV-vis spectroscopy, obtaining a value for each sample of the $R_z$ = $A_{406}/A_{280}$ ratio, and comparing it with the known value for a pure sample. The $R_z$ ratio would allow to know as well wheter if the protein was fully loaded with heme cofactor, or there was a fraction of apo protein.

Chemical reduction of the proteins was carried out by adding a few granules of solid sodium ascorbate or sodium dithionite (Fisher), and the spectral changes upon reduction of the ferric ($Fe^{3+}$) to the ferrous ($Fe^{2+}$) form were monitored. Known heme binding ligands, such as imidazole, cyanide, or nitric oxide, were added in the protein solution in the cuvette to assess binding and obtain the spectral features of the complexes. They would be added to the ferric or ferrous state of the protein depending on which state favoured binding of the ligand. In the case of nitric oxide, 10-20 μl of proliNONOate (Cayman chemicals) solution would be added with a Hamilton syringe (Hampton Research). This compound that would release NO• upon pH drop when in contact with the protein solution.

To initiate the peroxidase cycle in the peroxidases (DHP, DtpA, and DtpAa), a stoichiometric amount of hydrogen peroxide ($H_2O_2$) would be added in several subsequent titrations, to observe and characterize the peroxidase intermediates Compound I and II, and its stability through time. Alternatively, the peroxidase cycle would be activated by adding peracetic acid (PAA), an organic peroxide that was used in the soaking experiments to generate the intermediates within the crystals.

## 3.6 Crystallization

All proteins were crystallized using the hanging or sitting drop vapour diffusion technique, with protein concentrations ranging between 10-25 mg/ml

to obtain the desired shape and size of crystals. Protein buffer used for each target was different, but would include a buffer at a pH in which the protein was stable (e.g. MES, HEPES, sodium acetate) and 50-100 mM of NaCl. If no crystallization conditions were known for a particular protein, commercial screens were performed using an Art Robbins Gryphon crystallization robot, able to set-up drops as small as 0.2 μl in MRC screen plates and minimize protein usage. Crystal hits found in the screens (or already known from literature) were manually optimised in 24-well crystallization plates, setting up hanging drops with volumes between 2-6 μl, were a range of concentrations for the different components in the condition would be tested, as well as different protein:reservoir ratios. The crystallization conditions used for each of the proteins are detailed in Table 3.6. Different crystallization temperatures were also tested for each protein, having the options of: the crystallization room at 18ºC, the robot room at RT, or the crystallization fridge at 4ºC. Crystals would normally appear after 1-4 days after setting up the trays, depending on the protein sample and condition. Crystals were cryo-protected before flash-cooling in liquid nitrogen by soaking for ~10-20 seconds in reservoir solution supplemented with a cryo-protectant tested and optimized beforehand, which typically was glycerol, sucrose, or PEG200. Crystals were harvested with the help of micro-loops (MiTeGen), and mounted in pucks or canes, which in its turn were introduced in a previously cooled dewar which ensured a stable temperature of the samples for the duration of the trip to the synchrotron.

**Table 3.6:** Crystallization conditions used for each of the proteins under study, specifying the precipitant, salt and buffer conforming the solution.

| Protein | Precipitant | Salt | Buffer |
|---|---|---|---|
| DHP | 12-20% PEG 4000 | 200 mM $AmSO_4$ | - |
| DtpA | 18-25% PEG 3000 | - | 50 mM Sodium citrate pH 5.5 |
| DtpAa | 20% PEG 6000 | - | 100 mM HEPES pH 7.0 |
| NOD | 18-22 PEG 3350 | 200 mM $MgCl_2$ | 100 mM Bis Tris pH 5.5 |

## 3.7 Crystal soaking

Protein crystals were soaked in different substrates or ligands whether to obtain complexes of interest, initiate catalytic activity, or chemically reduce the protein. When the substance to be soaked was soluble in water, it would be supplemented at 10-50mM concentration to the relevant crystallization condition. If the compound had limited or no solubility in water, it would be dissolved in DMSO (Fisher) and added to the crystallization solution without exceeding a 5% of DMSO in the final volume. Crystals were soaked normally for 5-15 min, and cryo-protected before flash-cooling in liquid nitrogen. For reduction of the crystals, these were soaked in crystallization solution with added 100 mM of sodium ascorbate or 50 mM of sodium dithionite. They were left for 30-60 min to ensure complete reduction of the haem metal centre, without damaging the crystals. Sometimes, upon binding or reduction, a

change of colour would be apparent from brown to reddish. Crystals were transferred from one solution to another, and finally cryo-cooled in liquid nitrogen using cryo-loops (MiTeGen) for manipulation.

In the case of initiation of the peroxidase cycle, crystals were soaked for 2-3 min in reservoir solution with excess of hydrogen peroxide (50 mM $H_2O_2$) to generate the intermediate species Compound III. For generation of Compound I and II, crystals were soaked in solutions with added 1-5 mM peracetic acid (PAA). If soaked for 1-2 min, crystals would be in the Compound I state, and if left longer (10-15 min) they would develop to the Compound II state. The state of protein forming the crystal was always checked with an on-line single-crystal spectroscopy at the beamline as a form of validation.

**Table 3.7:** Soaking experimental conditions for the different soaks carried out with DHP and DtpA.

| Crystals | Compound | Concentration | DMSO ? | Soaking time |
|---|---|---|---|---|
| DHP | Imidazole | 5 mM | No | 10 min |
| | Benzimidazole | 10 mM | 5 % | 10 min |
| | Benzotriazole | 10 mM | 5 % | 10 min |
| | Indazole | 10 mM | 5 % | 10 min |
| | 5-Bromoindole | 50 mM | 5 % | 5 min |
| | 2,4-dichlorophenol | 5 mM | 5 % | 5 min |
| DtpA | $H_2O_2$ | 50 mM | No | 2-3 min |
| | Peracetic acid | 1-5 mM | No | 2-3 min |

## 3.8 Single-crystal spectroscopic measurements

Single-crystal spectroscopic measurements were performed using the general procedure: **Spectroscopy → diffraction dataset → spectroscopy**. In this way, every crystal would be probed with spectroscopy (e.g. resonance Raman or UV-vis) before any exposure to the X-ray beam to inspect its original state. After collection of a complete diffraction dataset, a further spectrum would be acquired at the same position as the initial one. Photo-reduction events and/or other changes experienced by the protein within the crystal could be tracked in that way. If a series of successive datasets were taken (MSOX methodology, Horrell et al., 2016), then there would be an initial spectroscopic characterization, followed by a spectrum measured after each dataset, which would help following reactions or radiation-induced processes in the crystals.

### *3.8.1 Single crystal resonance Raman (SCRR) measurements.*

SCRR measurements were performed at beamline X10SA at Swiss Light Source (SLS), using the MS3 on-axis multimode spectrophotometer (Dworkowski et al. 2015; Pompidor et al. 2013) presented in Chapter 2. Protein crystals were mounted on the goniometer and maintained at 100 K with a nitrogen cryo-stream. Resonance Raman spectra were acquired before and after X-ray exposure of the sample. The excitation laser (Omicron Laserage, LDM405, 400, CWA) emitted at a wavelength of 405.45 nm and the focused laser spot size at the crystal was of 25 μm diameter. The spectrometer entrance

slit width was 100 mm and 2400 lines $mm^{-1}$ grating blazed at 300 nm was used for dispersion. Spectra were measured using a Newton EMCCD DU-970-UVB (Andor Technologies) detector with 1600x200 pixels and cooled at 193 K. Normally, each spectrum would be acquired with 20 accumulation of 20 seconds each (20x20s).

Raman shifts were calibrated against cyclohexane or 4-acetamidophenol standard samples, as previously described in Dworkowski et al. (2015). The resulting reference spectrum was used to calculate the laser excitation wavelength, which was converted to Raman shifts using the Raman Calibration Tool (RaCaTo). Laser powers at the sample position were monitored using a Thorlabs PMD 100D hand-held power meter. The sample laser power showed a linear dependence upon the input laser power in the range 50-200 mW. SCRR data were measured in the wavenumber range 575-2000 $cm^{-1}$ with particular interest in the porphyrin redox-state and core-size marker regions (1300-1700 $cm^{-1}$). All spectra were processed using the SLS-APE analysis package (Dworkowski et al., 2015), which involves background substraction due to high fluorescence background.

### *3.8.2 Single-crystal UV-vis measurements.*

Single-crystal UV-visible absorption spectroscopy measurements, were carried out both at ESRF and SLS. The micro-spectrophotometer at ESRF (von Stetten et al., 2015) was installed online at any beamline in use before the experiment. Beamlines BM30 (FIP) and ID30A-3 (massif-3) were used in this

work. The instrument was set up in a plane almost perpendicular to the X-ray beam (McGeehan et al., 2009). The objectives were positioned at 45º with respect to the goniometer axis to allow for access of the cryostream and for sample changing. A combined deuterium-halogen lamp (DHL2000, Ocean Optics) was used as the light source. The micro-spectrophotometer was focused into a 25 μm spot and it was aligned to probe the same region as the X-ray beam. Spectroscopic data over the absorbance range 200-1000 nm were measured using the *SpeCuBE* software (von Stetten et al., 2015) and were initially processed using the *SpectraSuite* software (Ocean Optics). Each optical absorption spectrum was the result of 10 accumulations of 200 ms exposures.

For on-line UV-vis measurements at SLS, the MS3 micro-spectrophotometer was changed to the optical absorption mode. The optical spot for the experiments was also focused to a 25 μm light spot, and no alignment was necessary to probe the same region as the X-ray beam due to the on-axis geometry. The same SLS-APE analysis software as for SCRR was used to process the UV-vis spectra.

## 3.9 Serial crystallography methods

### *3.9.1 Batch crystallization*

As discussed in Chapter 2, to satisfy the sample delivery requirements for fixed-target serial crystallography, where thousands of positions are screened in a single run, it was needed to develop the production of a large amounts of microcrystals, and this is best achieved with batch crystallization. Using this

technique, crystals are grown directly in a solution formed by mixing protein solution and crystallization condition. Batches were set up in Eppendorf tubes of 0.5 or 1.5 ml. For each protein, the optimal conditions had to be optimised to reach a point where the desired amount and size of microcrystals were obtained (normally between 10-50 μm), in a monodisperse distribution. Variables considered were:

- Concentration of protein, precipitant, salts and buffer.
- Volume ratio protein:condition, which is important and different for each protein.
- Temperature.
- Seeding.
- Manual mixing or vortexing the batches once set up.
- Rocking the batches while growing.

Specific final batch conditions conditions and optimization process for each sample are detailed in the corresponding protein chapter (Chapter 4 for DyPs, Chapter 6 for DHP and Chapter 7 for NOD).

### *3.9.2 Fixed targets used*

Two different fixed targets or chips were used in this work (Figure 3.3), both developed in recent years (Roedig et al. 2016; Mueller et al. 2015) for serial crystallography, and designed to provide a solid support for the microcrystals adding minimal background in diffraction experiments. For experiments carried out at Diamond (SSX) and SACLA (SFX), the system developed at Diamond was used (Owen et al., 2017), which we have named

as the ‘Diamond’ chip. At the Laue beamline experiments at APS, another kind of fixed targets were used, developed in Heidelberg, and so they were named as ‘Heidelberg’ chips (Figure 3.3).

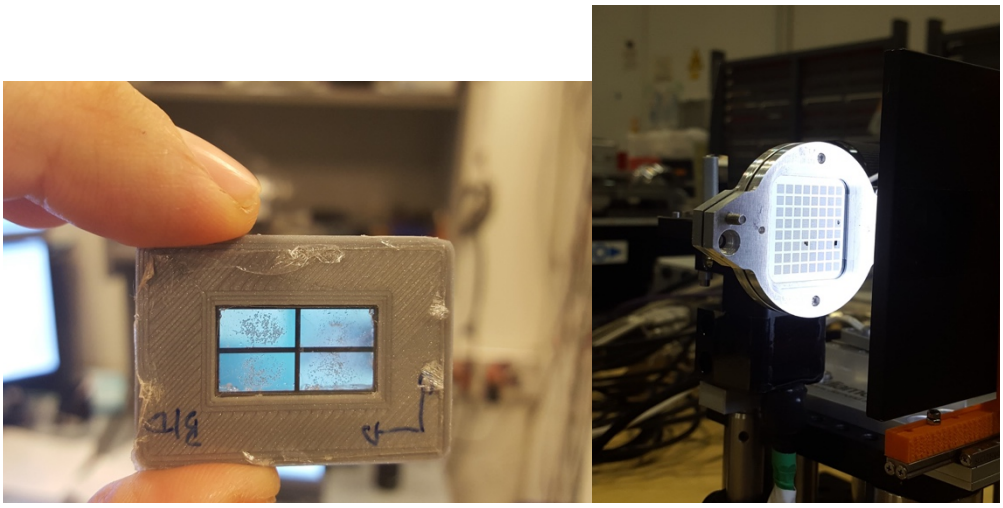

**Figure 3.3:** Chips used for serial crystallography experiments. On the left panel, the ‘Heidelberg’ chip used at APS, and on the top panel the ‘Diamond’ chip used at I24 (Diamond Light Source) and SACLA.

Both fixed target systems are formed by silicon nitride, and introduced in a frame holder, which was sealed with mylar. This polymer would prevent the desiccation of the crystals during the experiments, and had been selected due to its low scattering properties resulting in a low background. Heidelberg chips were divided in four quadrants, each of them formed by a thin layer of silicon nitride presenting a regular array on which the microcrystal solution would be deposited by pipetting.

The Diamond chips contained 25,600 funnel-like apertures which would hold the crystals in position. The upper part wide part was of 120 μm, and the narrow end side would be of variable sizes (7, 12, 37 μm), which would be selected depending on the sample to match the microcrystal size. Their

dimensions are approximately 30x30 $mm^2$ and there are sub-divided in 8x8 'city block', each containing a regular array of 20x20 apertures (Figure 3.4).

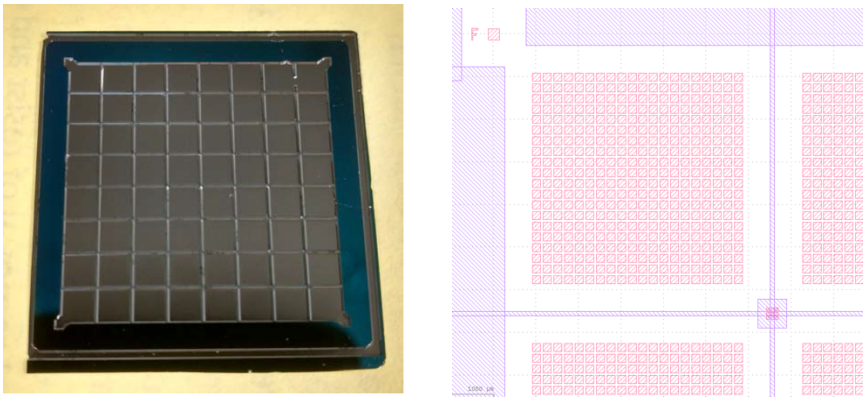

**Figure 3.4:** General view of a Diamond chip where the regular array of 8x8 city blocks can be appreciated (left panel). A graphic scheme showing 20x20 apertures found in each city block is shown on the right panel. (figure kindly provided by Dr. Robin Owen at DLS)

### *3.9.3 Loading of the chips*

Heidelberg chips would be loaded by adding a certain quantity of microcrystals solution in each of the quadrants, then blotted manually from the opposite face of the chip to absorb the excess of liquid, and introduced in a 3D-printed chip case. This container had two windows which were sealed with mylar, and an absorbent frame of material around the chip was soaked with 500-800 μl of crystallization condition to keep the crystals moist.

The loading process for the Diamond chips was similar, but it would be performed inside a portable plastic hood (Figure 3.5), where the moisture was controlled to be always above 70-80 %. Chips would have to be previously glow-discharged (Pelco-easyGlow instrument) to facilitate the spread of the microcrystal solution, which typically would be of 150-200 μl per loading. The excess of liquid was sucked out with the help of a pump, and the chip was closed in-between two frame holders which were also sealed with mylard.

**Figure 3.5:** Chip being loaded in the portable plastic moisture-controlled plastic hood used at SACLA.

### *3.9.4 Chip screening routines.*

In both the Diamond and the Heidelberg chips, the same screening principle was applied, which consists in moving through the selected points of the chip to hit fresh microcrystals each time, although there were some operational differences. In the case of 'Heidelberg' chips, once crystals were loaded onto the chip, an area with the highest crystal density would be manually selected, as well as the step to move along the chip between every shot. These parameters would be introduced in the software used at the BioCARS-14B beamline at APS. Depending on the chosen area and step, the inspection time could change substantially, but it would typically take 30-45 minutes for a run of 5000-7000 positions. Each of the quadrants of the Heidelberg chip had to be screened separately, with manual alignment and selection of experimental parameters in between.

For Diamond chips the screening routine was automated to be carried out at pre-stablished measuring positions (Figure 3.6), i.e. each one of the apertures where crystals were held. Normally, the full chip would be measured at its 25,600 positions with a frequency of 30 Hz, which would take approximately 14 min of measuring time. Each new mounted fixed target at the hutch would be aligned with help of three ‘fiducials’ situated in the corners of the chip. If not fully loaded, the area of the chip to be screen could be modified by selecting/deselecting in program the corresponding city blocks. A common situation was when just the central part of the chip was loaded (due to lack of solution, or just for a quick test), then just the inner 6x6 region would be screened in approximately half the time. The time to run through a whole chip was of around 14 minutes. A scheme showing how the screening routine was carried out is shown in Figure 3.6.

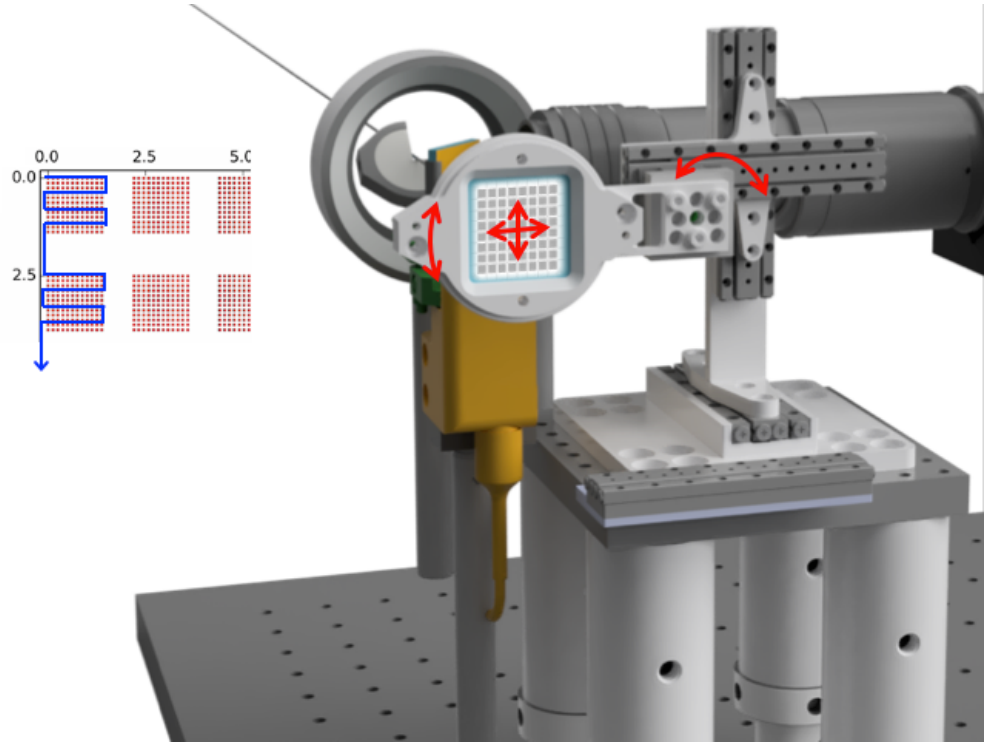


**Figure 3.6:** Representation of the screening routine for a chip mounted in the hutch. On the left can be appreciated how the fixed target moves through the X-ray beam from one city block to another. Slight differences in mounting of each chip (red arrows), are corrected by alignment with the help of three fiducials at the corner of the chip which are visualized with an on-site camera. (figure kindly provided by Owen group at I24, Diamond)

### *3.9.5 SFX data processing with CrystFEL*

During the XFEL beamtime at SACLA, preliminar indexing and on-the-fly estimations of hit rates were carried out with the program *Cheetah* (Barty et al. 2014). *Cheetah* would also select the frames with protein diffraction, in this way making the volume of data to be copied to a hard-drive considerably smaller.

To process the SFX data obtained in the form of *h5* files, *CrystFEL* was used, a piece of software developed by Thomas White and co-workers (2012). CrystFEL consists in a package containing different programs that are all run

from the command line. The normal workflow in crystallography is followed of indexing, scaling and merging, with different options of algorithms to be used in each step.

Within each protein chapter, there will be a detailed report of the process carried out in each case. In this section, a general overview of the workflow and the different steps taken to process the SFX data and generate a final *mtz* file will be explained.

### *3.9.5.1 Indexing with indexamajig*

Firstly, the indexing task would be perfomed, carried out by the program *indexamajig*. The standard parameters used for an initial run would be had been tested initially with the benchmark protein AcNiR, with cell parameters known and consistent diffraction quality. In this test run, initial refinement of the detector geometry would be carried out, and an optimized geometry input file created. The command to run *indexamajig* was:

```
$ indexamajig -i files.lst --peaks=zaef -threshold=300 --
min-gradient=90000 --min-snr=5 --int-radius=3,4,5 --
indexing=asdf -g detector.geom -p parameters.cell -o
testrun.out -j 20
```

Where the different parameters specified are:

- -i files.lst → path to the input diffraction data is provided.
- --peaks=zaef → invokes the peak-finding algorithm *zaef* using the gradient search method by Zaefferer (2000).

- -threshold=300 → a threshold of photon counts for detection of diffraction spots is specified.
- --min-gradient=9000 → minimum square gradient for peak detection.
- --int-radius=3,4,5 → sets the inner, middle and outer radii for three-ring integration during the peak search.
- --indexing=asdf → indexing performed using the asdf CrystFEL algorithm.
- --min-snr=5 → a minimum value for signal to noise ratio is specified.
- -g detector.geom → file containing information about detector geometry, provided by staff at SACLA, following a CrystFEL template.
- -p parameters.cell → symmetry information for the microcrystals is provided, following a CrystFEL template.
- -o testrun.out → generates a stream output file.
- -j=20 → number of processors, with a maximum value of 20.

Measured data from each chip would be processed initially in this way, generating in each case a file 'stream.out' with the indexed reflections. The cell parameters distribution for this first output were examined using the program *cell_explorer*, which generated distribution plots for each parameter. It was run introducing the command:

```
$ cell_explorer firstRun.out
```

Cell parameters were therefore updated for the experimental characteristics of the microcrystals in each case. A second run of *indexamajig* was carried out with these optimized cell parameters, generating a second

refined output file ‘secondRun.out’. The detector geometry was also optimized using this second output file together with the initially used geometry file, using the program *detector-shift*. The command used in this case would be:

```
$ detector-shift .out secondRun.out detector.geom
```

Apart from refining the detector geometry, *detector-shift* also plotted the disagreement (if any), expressed as detector shift in x and y, between the location of experimental and predicted spots in every indexed frame. A refined geometry file was generated ‘refined.geom’, which would be used to perform a last run of *indexamajig*. The final stream file ‘finalRun.out’ for different chip with the same sample could be joined to form a unique file to continue with the scaling and merging steps. The command used to join the different output files was:

```
$ cat chip1.out chip2.out chip3.out > allchips.out
```

#### *3.9.5.2 Merging with process_hkl*

The merging process was carried out with the simplest approach, called *Monte Carlo* method, used by the program *process_hkl* in CrystFEL (White et al. 2012). Merging with the program *partialator* within CrystFEL was initially tested, but *process_hkl* was used for all SFX datasets after advice from Dr. Takanori Nakane (from the Japanese team collaborating with ours at SACLA). Process_hkl was included in the script called *merge.sh*, which would be run from the command line giving as input the final output file with all the reflections from *indexamajig*. The pointgroup in which the protein was crystallized was also needed as an input. The pointgroup with an added centrosymmetric

element would be given, to merge the Friedel pairs, just needed for anomalous data. This was recommended by CrystFEL devolopers (online tutorial; http://www.desy.de/~twhite/crystfel) and resulted in a better merging.

#### *3.9.5.3 Data quality statistics CC and Rsplit*

Data quality statistics were generated running the crystFEL script stat.sh, which would need as input the three output files from merge.sh, i.e. hkl, hkl1 and hkl2 files. The hkl file had all the data, which was splitted in two (hkl1 and hkl2) to make comparisons and generate statistics. Simmetry information for the crystal, would have also to be provided as input to stat.sh. Several figures of merit were generated, being the correlation coefficient (CC) and $R_{split}$ the most relevant for us.

The CC statistic is the equivalent of the $CC_{1/2}$ figure of merit used in conventional X-ray crystallography datasets (Diederichs & Karplus 2013). It has a value between 1 and 0, and compares the data split it two halves and assess how equal similar they are, being 1 when they are identical and 0 when they have no similarity at all. This parameter gives an idea if the data added is meaningful, with a threshold of CC>0.5 commonly used to establish the resolution limit.

The figure of merit $R_{split}$, is a data quality statistic used by the XFEL and serial crystallography community (White et al. 2012), to estimate the quality of the diffraction data. To generate it, data are split in two halves, merged independently (grouped in even and odd frames), and then compared. It is

expressed normally in %, and the lower the value, the better the agreement between the two halves. As a general rule, an $R_{split}$ value of 35% is reasonable, below 20% is good, and below 10% is excellent, although these are rough guidelines.

#### *3.9.5.4 Generation of mtz file*

Finally, once a resolution cut-off has been decided in agreement with CC and $R_{split}$ statistics, an *mtz* file was generated using the *CrystFEL* script 'create-mtz´, giving as input the *hkl* file from the merging process. It was also necessary to specity the symmetry information and cell parameters for the crystal. The obtained *mtz* was used as input experimental data for the programs used within the CCP4i2 suite (Potterton et al. 2018).

### *3.9.6 SSX data processing with DIALS*

The experimental set-up for the fixed target system developed by Diamond (Owen et al. 2017) is shown in Chapter 2. Chips of different apertures (e.g. 7 μm, 12 μm, 37 μm) were available and chosen depending on the size of microcrystals for each sample. Data collection parameters, such as detector distance, beam dimensions and exposure time, were also tailored to the sample, using normally an energy of 12800 eV (0.97 Å).

Serial crystallography data measured at beamline I24 (Diamond Light Source) was stored in *cbf* files, which were processed using *dials.stills_process* within the program *DIALS* (Winter et al. 2018). This task carried out the indexing and integration of the diffraction spots, and was specially developed to process serial crystallography data, formed by single still diffraction patters

from thousands of crystals, and therefore no rotation of crystals is involved (as in conventional synchrotron data collection). Integrated data were scaled and merged using *PRIME* (Uervirojnangkoorn et al. 2015). This would generate an *mtz* file to proceed with molecular replacement and refinement steps, as well as data statistics to assess the quality of the data such as $R_{split}$, $CC_{1/2}$, and completeness.

## 3.10 X-ray data collection and processing

Conventional synchrotron X-ray diffraction data (i.e. not serial crystallography data) were collected at three different sources: Diamond light source (DLS), ESRF, and Swiss light source. X-ray wavelength used normally ranged between 0.8-1 Å. Data collection was carried out a 100 K temperature, unless stated otherwise. Diffraction experiment parameters (e.g. exposure time, detector distance, % transmission and rotation per frame) were decided in each case and controlled using the in-house software present at each synchrotron beamline. Diffraction spots were normally indexed and integrated using the program XDS (Kabsch 2010), and otherwise this job was automatically carried out by the processing pipelines available for example at DLS and ESRF. For the usage of XDS, modification of the input file XDS.INP for each crystal was needed. In this file, important experimental and sample information was specified: detector parameters, cell parameters, space group, detector distance, resolution range, number of images, rotation, wavelength. The output file from XDS was XDS_ASCII.HKL, and would be used as an input for the program *Aimless* (Evans & Murshudov 2013), which would assess the

space group probability, and scale and merge the diffraction data. Once run, Aimless would produce and mtz file, and provide a table, such as the one shown in Figure 3.7, where the most important data collection parameters and data quality statistics were detailed (e.g. resolution, $CC_{1/2}$, I/σ, $R_{merge}$). Data resolution cut-off would be performed following the criteria $CC_{1/2}$ > 0.5 (Diederichs & Karplus 2013). The merged mtz, containing the experimental data, was used for the next steps of structure determination: molecular replacement and refinement. To determine the Matthew´s coefficient, the homonym program (Matthews, 1968; Kantardjieff & Rupp, 2003) in the Ccp4 suite was run to confirm the number of molecules in the asymmetric unit and the solvent content in the crystal.

Summary data for Project: DyP_H2O2 Crystal: ESRF_2oct15_DyP_H2O2 Dataset: DyP_0_5__H2O2

| | Overall | InnerShell | OuterShell |
|---|---|---|---|
| Low resolution limit | 47.98 | 47.98 | 1.63 |
| High resolution limit | 1.60 | 8.76 | 1.60 |
| Rmerge (within I+/I-) | 0.087 | 0.025 | 0.904 |
| Rmerge (all I+ and I-) | 0.100 | 0.029 | 1.003 |
| Rmeas (within I+/I-) | 0.118 | 0.036 | 1.212 |
| Rmeas (all I+ & I-) | 0.116 | 0.035 | 1.175 |
| Rpim (within I+/I-) | 0.079 | 0.025 | 0.801 |
| Rpim (all I+ & I-) | 0.060 | 0.019 | 0.606 |
| Rmerge in top intensity bin | 0.026 | - | - |
| Total number of observations | 302888 | 1795 | 14772 |
| Total number unique | 81138 | 528 | 3992 |
| Mean((I)/sd(I)) | 10.3 | 32.7 | 1.5 |
| Mn(I) half-set correlation CC(1/2) | 0.997 | 0.996 | 0.504 |
| Completeness | 100.0 | 99.0 | 99.9 |
| Multiplicity | 3.7 | 3.4 | 3.7 |

**Figure 3.7:** Example of a data quality statistics table generated by Aimless (Evans & Murshudov 2013).

## 3.11 Molecular replacement

Since, all the proteins where previously determined in the literature, all the structures of DHP, DtpA and DtpAa were solved by the molecular replacement method either with *Phaser* (McCoy et al. 2007) or *MolRep* (Vagin

& Teplyakov 1997), both within the Ccp4 suite. As inputs, the mtz containing the experimental data in each case, as well as the relevant pdb file for each case, obtained from the Protein Data Bank (PDB; wwpdb.org), were provided. Alternatively, the experimental data would be refined against an existing model from another structure that would had been previously refined. Success of the molecular replacement job was assessed by the R values, and by looking at the solution in Coot, to see if there was agreement between the model and the experimental data. In the case of NOD, since the conformation of the protein was different from the one deposited in the PDB, just a part of the molecule was well positioned after the molecular replacement job. To complete structure solution, an auto-modelling job was carried out the program *Bucaneer* (Cowtan 2006), giving as inputs the experimental data, the phases from the MR job, and the amino acid sequence of the protein. In this way, most of the protein backbone and side chains were well positioned and refined.

## 3.12 Refinement and model building

Several rounds of refinement and model building were routinely carried out for each of the structures included in this work. Refinement was performed using the program *Refmac5* (Murshudov et al., 2011) within the ccp4i or i2 suites. This would generate a refined model, an electron density map ($2F_c - F_o$), and a difference map ($F_c - F_o$), which could be all visualized using the model building program *Coot* (Emsley et al., 2010). Improvements of the model would be carried out by manual re-building, relying on the information from the

electron density maps, which would indicate for example wrong conformations of side chains or double conformation of a certain part of the protein. The density map was usually visualized at 1$\sigma$ contour level, and the difference map at 3$\sigma$. Where positive electron density was observed, molecules consistent with this density would be modelled, such as molecules of water, double conformation of side chains, ligands, or components from the crystallization solution such as ions (e.g. $SO_4^-$). As a general procedure, a general check-up of the protein backbone and the amino acid sidechains would be carried out after the first refinement round, which would be isotropic (i.e. assuming B-factor does not vary depending on direction within the crystal). After amending the model, and adjusting the occupancies of double conformations of side chains, a further refinement would be performed, which in this case would be anisotropic if the resolution was > 1.5 Å. Otherwise, isotropic refinement would be kept during the rest of the process. At this point water molecules, ligands and ions would be added, followed by rounds of refinement until the model could not be further improved. Restraint refinement was carried out by adding riding hydrogen atoms to the model and adjusting the weighting factor to obtain acceptable RMSD bond angles and lengths. Special restraints were applied to the haem cofactor, using a modified “MEH” dictionary file, kindly provided by Dr. A. Lebedev. During the whole refinement process, attention would be taken to the $R_{work}$ and $R_{free}$ values, keeping the gap between them as short as possible, and ensuring that both were diminishing after each round of refinement, meaning that the improvement in the model was not biased.

## 3.13 Validation of crystal structures

The validation of each of the produced protein structures was carried out with online servers: *MolProbity* (Chen et al. 2010) would check the protein geometry, and therefore would just need a refined pdb file as input; whereas with *JCSG QC-Check* (Elsliger et al. 2010), an assessment considering also the electron density was performed, with pdb and mtz files as inputs.

Normally, several runs of MolProbity would be carried out, which would pick up geometric issues such as steric clashes, poor rotamers, Ramachandran outliers, and bad bonds and/or angles. These were amended by re-modelling the structure in Coot (Emsley et al., 2010), followed by a round of refinement with Refmac5 (Murshudov et al., 2010), and a further run of MolProbity to assess the improvement of the model. The same workflow was applied with JCSG QC-check, but in this case also considering: B-factors, waters or residues not supported by enough electron density, anisotropic electron density, floating waters, etc. Several cycles were repeated until the model was of high quality and no serious issues were present, i.e. the parameters above mentioned (e.g. Ramachandran outliers, rotamers, bonds) were within its acceptable limits. When the protein structure was considered finished, it was deposited at the Protein Data Base server (wwpdb.org), where a further validation check-up would be performed.

# Chapter 4 Structural characterization of reaction intermediates and resting states in the peroxidase cycle of Dye-decolourising Peroxidases (DyPs).

## 4.1 Introduction

As discussed in Chapter 1, DtpA and DtpAa are two peroxidases from the DyP family found in the bacteria *Streptomyces lividans*. These enzymes were initially studied by the Worrall group at Essex (Petrus et al. 2016) as being one of the proteins involved in the maturation process triggered by copper of the bacterium. The peroxidase cycle of the protein was investigated in solution by means of UV-vis spectroscopy, and the intermediates Compound I and II were identified upon activation with hydrogen peroxide. Since this protein was a heme enzyme that met the requirements for our methodologic approach, it was selected as a platform to study and characterize the peroxidase cycle for DyPs, as had previously been carried out for classic peroxidases (Berglund et al. 2002). We intended to do this by validating the intermediates with *in crystallo* spectroscopies, and obtain its intact structures by measuring low-dose datasets and possibly apply the composite dataset approach (Berglund et al. 2002). Later, the opportunity arose to apply serial crystallography to these targets at synchrotrons (SSX) and at XFEL (SFX), allowing us to obtain damage-free structures of their resting states, and also follow radiation-induced

changes at the active site by applying the Multiple Structures from One Xtal approach (MSOX; Horrell et al., 2016, 2018) combined with SSX.

## 4.2 Protein expression and purification

Initially, DtpA crystals were kindly provided by Dr. Amanda Chaplin from the Worrall group, and were used for characterization experiments at ESRF with single crystal spectroscopies. After these, protein was expressed and purified following the protocol developed by Dr. Chaplin (Petrus et al. 2016). A pET28a construct encoding the protein from *S. lividans* with an N-terminal added His-tag, was used for over-expression in *E.coli*. Overnight pre-cultures were used to inoculate 2 L flasks with 1.4 L of high salt LB medium (10g tryptone, 10 g NaCl, 5 g yeast extract per litre), also containing 50 μg/ml Kan, which were grown at 37ºC and 220 rpm. At an $OD_{600}$ ~1.2, 0.25 mM of 5-aminolevulinic acid (dALA) and 0.1 mM of iron citrate were added to the solution as a haem-precursor and iron supplement. Cultures were then induced by adding isopropyl β-D- thiogalactopyranoside (IPTG; Melford) to a final concentration of 0.5 mM and pure carbon monoxide (CO) gas bubbled through the culture for ~30 s. Flasks were sealed with parafilm and incubated for 18 h at 30ºC and 100 rpm. Cells were harvested via centrifugation (5,000 rpm for 20 min), and the resulting pink pellets resuspended in 50 mM Tris/HCl, 500 mM NaCl (Fisher) and 20 mM imidazole (Sigma) at pH 8 (buffer A). Cells were lysed using an EmulsiFlex-C5 cell disrupter (Avestin), followed by centrifugation (18,000 rpm, 40 min, 4ºC). The supernatant was loaded onto a 5-ml nickel-sepharose column (GE Healthcare) equilibrated with buffer A, and eluted by a

linear imidazole gradient using buffer B (buffer A with 500 mM imidazole). The protein peak eluted at ~40% buffer B, and fractions were run in an SDS-PAGE gel to check protein purity and molecular weight (Figure 4.1, left panel). The protein ran between the 40-50 kDa markers, consistent with the molecular weight of DtpA (40.8 kDa). Fractions were pooled and loaded onto a S200 Sephadex size exclusion column (GE Healthcare) equilibrated with 10 mM Na Acetate pH 5.0, 100 mM NaCl, for the polishing step. Eluted fractions were assessed for purity using a 15% SDS-PAGE gel (Figure 4.1, right panel). The protein was pooled and concentrated using a 10-VivaSpin concentrator (Sartorius) with a 10 kDa cut-off. Concentration was determined by means of UV-vis spectroscopy, using the extinction at 280 nm ($\varepsilon_{280}$) of 37,470 $M^{-1}cm^{-1}$ for DtpA calculated using ProtParam ExPASy (Gasteiger et al. 2005). When a concentration of ~15 mg/ml was reached, protein was aliquoted and stored at -80ºC.

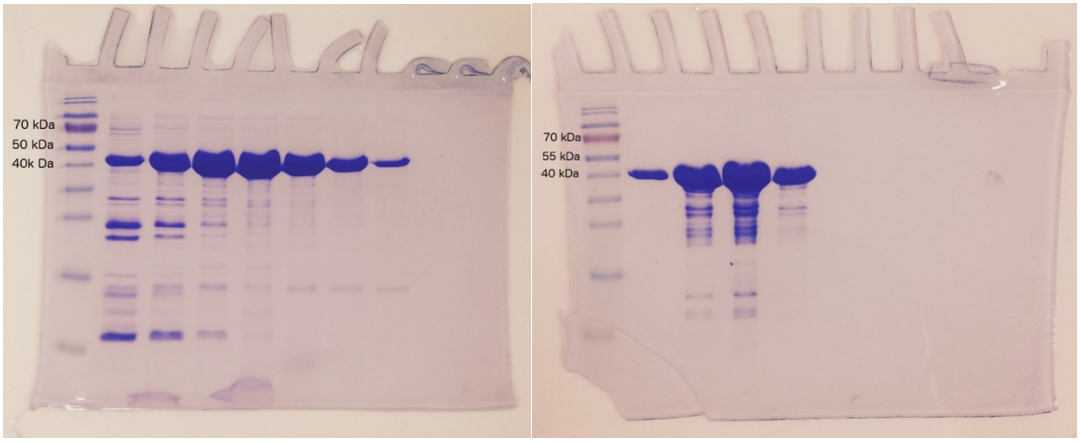


**Figure 4.1:** SDS-PAGE gels of DtpA fractions eluted from a Nickel-sepharose (left panel), and an S200 size exclusion column (right panel).

## 4.3 Initial online-spectroscopy characterization.

As performed for each of the haem enzymes studied in this work, an initial assessment was carried out with the different *in crystallo* spectroscopic techniques available at SLS and ESRF, to decide which of them would be more suitable for each protein system in particular. In the case of DtpA, we started by measuring ferric crystals with the UV-vis micro-spectrophotometer installed at beamline BM30 or at the cryobench (Von Stetten et al. 2015), at ESRF. Refer to Chapter 3 for general experimental parameters and procedures used for this instrument. The resulting UV-vis spectra were of good quality for DtpA crystals, with sharp spectral features observed and no saturation within the range of interest (350-700 nm). Photo-reduction of the protein by the X-ray beam (0.99 MGy of absorbed dose per dataset) could be easily monitored with this technique, as shown in (Figure 4.2). Upon photo-reduction of the haem, several spectral changes could be appreciated: shift of the Soret band to ~420 nm, flattening of the CT band at ~625 nm, and sharpening of the $\beta/\alpha$ bands suggesting a hexa-coordinated heme species.

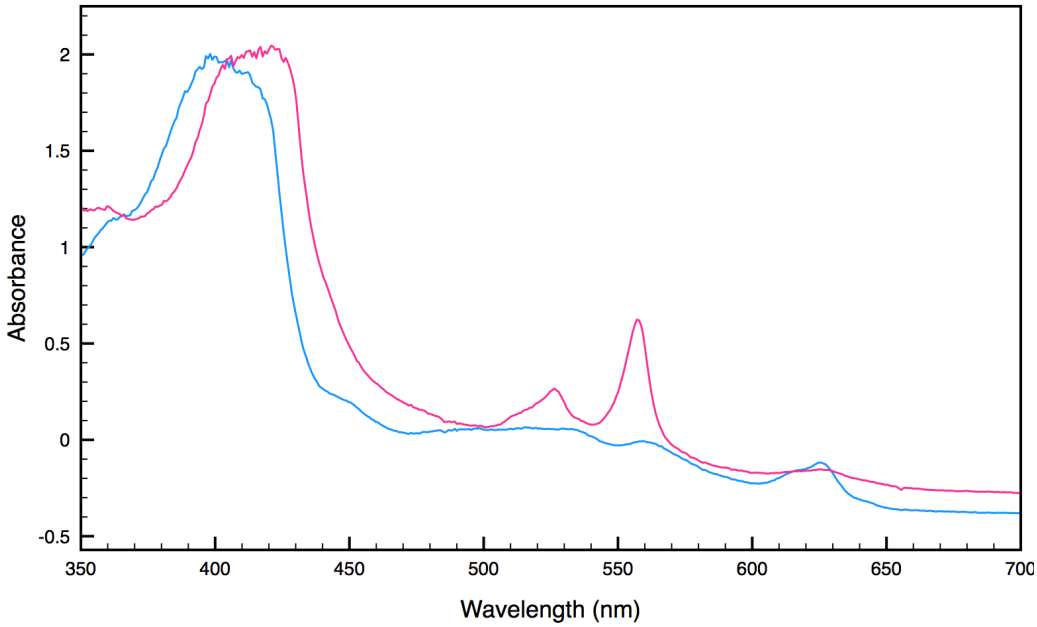


**Figure 4.2:** Online UV-vis spectra measured at 100 K from a DtpA ferric crystal before interaction with X-rays (in blue), and after collection of a complete diffraction dataset (in magenta). Spectra were acquired with the micro-spectrophotometer (von Stetten et al., 2015) installed at beamline BM30, at ESRF (Grenoble, France). Note the several spectral changes appreciated upon photo-reduction of the protein by the X-ray beam.

At beamline X10SA at Swiss Light Source (SLS), we assessed the suitability of applying Resonance Raman (RR) spectroscopy to DtpA, using the MS3 micro-spectrophotometer (Pompidor et al., 2013) equipped at the beamline. General parameters used to acquire SCRR at SLS are detailed in Chapter 3. Spectra were measured from ferric DtpA crystals, in which photo-reduction of the protein could be monitored by observation of the redox marker peak ν4 (Figure 4.3). However, spectra were of lower quality in terms of sharpness and signal of the core size marker peaks (between 1450-1700 $cm^{-1}$). Therefore, it was decided that UV-vis would be the spectroscopy used for this protein system, unless Resonance Raman (or any other spectroscopy) could provide us with some key information at some point of the project (e.g.

checking a specific vibrational mode of a specific bond to validate binding of a ligand).

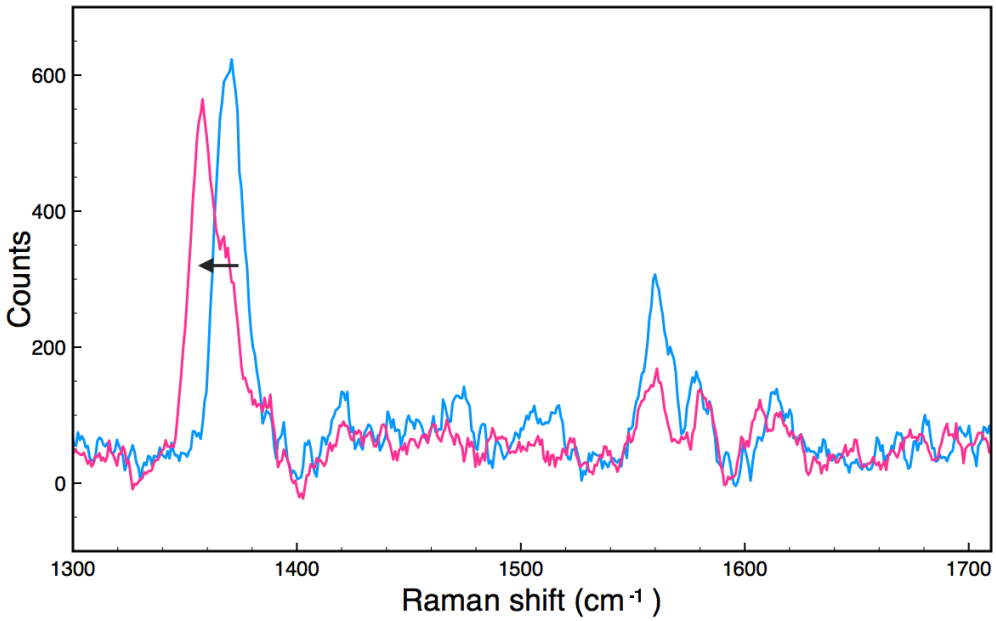


**Figure 4.3:** Resonance Raman online spectra of a ferric DtpA crystal measured at 100 K at beamline X10SA (SLS) before any exposure to the beam (in blue), and after collection of a dataset (in magenta). Shift of the redox marker indication photo-reduction of the protein by the beam is indicated by an arrow. Note that the rest of core size markers in the region 1400-1650 $cm^{-1}$ are not very distinguishable from the background. Spectra were acquired with the on-axis MS3 instrument (Pompidor et al., 2013), with 20 accumulation of 20 s each, and with previous laser excitation at 405.45 nm wavelength. The input laser power was of 50 mW focused at the sample in a 25 μm optical spot.

## 4.4 Photo-reduction of ferric DtpA – validated oxy-ferrous species.

To determine the unknown structure of DtpA by means of X-ray crystallography, ferric DtpA crystals were cryo-cooled in liquid nitrogen and taken to the ESRF, where they were measured. A UV-vis spectrum of a crystal was acquired with the online micro-spectrophotometer installed at the beamline BM30-B (Von Stetten et al. 2015) before exposure to X-rays, showing features

which confirmed the ferric state of the protein (in blue in Figure 4.4), comparing with DtpA solution data (Petrus et al., 2017). A first X-ray diffraction dataset was measured (Dataset-1), after which another UV-vis spectrum was acquired at the same initial point, observing clear spectral changes indicating photo-reduction of the protein by the beam (in orange in Figure 4.4). Specifically, the Soret band shifted to a higher wavelength (~420 nm), and sharp β/α bands appeared. These features were assigned to the oxy-ferrous species of DtpA, in agreement with solution data (Petrus et al., 2016). Therefore, Dataset-1 consisted in a mixture of states, starting with a ferric protein which would become photo-reduced after a certain exposure to the beam to yield the oxy-ferrous species. Hence, a second diffraction dataset (Dataset-2) was collected in order to obtain a 'pure' oxy-ferrous species, followed by a last UV-vis spectrum (in red in Figure 4.4) for validation.

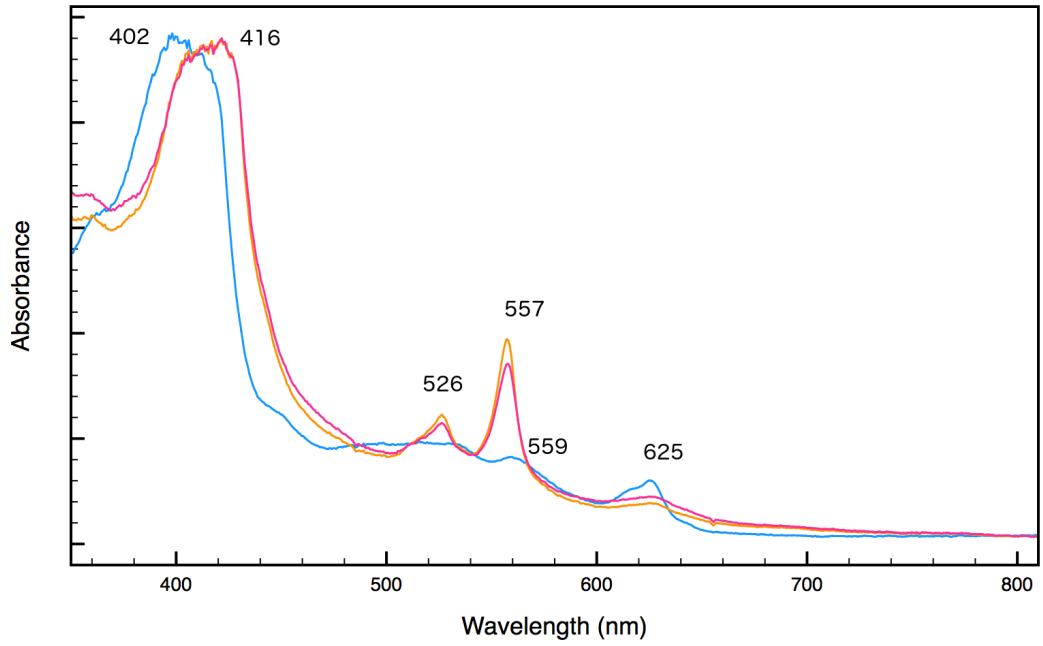


**Figure 4.4:** Online UV-vis spectra of a ferric DtpA crystal: before X-ray exposure, in blue; after collection of a first X-ray diffraction dataset, in orange; and after the collection of a second dataset, in red. The experiment was conducted at beamline BM30 at ESRF (Grenoble, France). Spectra were acquired with the micro-spectrophotometer (von Stetten et al., 2015) installed at

the beamline, with 10 accumulations of 200 ms each. Crystal dimensions were approximately 100x100x30 μm.

The structure from Dataset-1 was initially solved by molecular replacement with *Molrep* (Vagin and Teplyakov, 2010), which was carried out using as a search model the homolog DyP protein from *Streptomyces coelicolor* (PDB entry 4grc; T. Lukk, A. M. A. Hetta, A. Jones, J. Solbiati, S. Majumdar, J. E. Cronan, J. A. Gerlt & S. K. Nair, unpublished work), with a sequence identity of 95%. A solution was found and the structure was determined with model building carried out with *Coot* (Emsley et al., 2010) and rounds of refinement with *Refmac5* (Murshudov et al., 2011). DtpA was found as a dimer in the asymmetric unit (Figure 4.5), with an interface area of interaction of 2667 $Å^2$ between both monomers, calculated with program PISA (Krissinel, E. and Hendrick 2007) in ccp4i2, suggesting that the protein is likely to be found as a dimer *in vivo* as well. DtpA adopts a ferredoxin fold, common among the DyP family (Y. Sugano, 2009), with a mixture of α-helices and β-sheets.

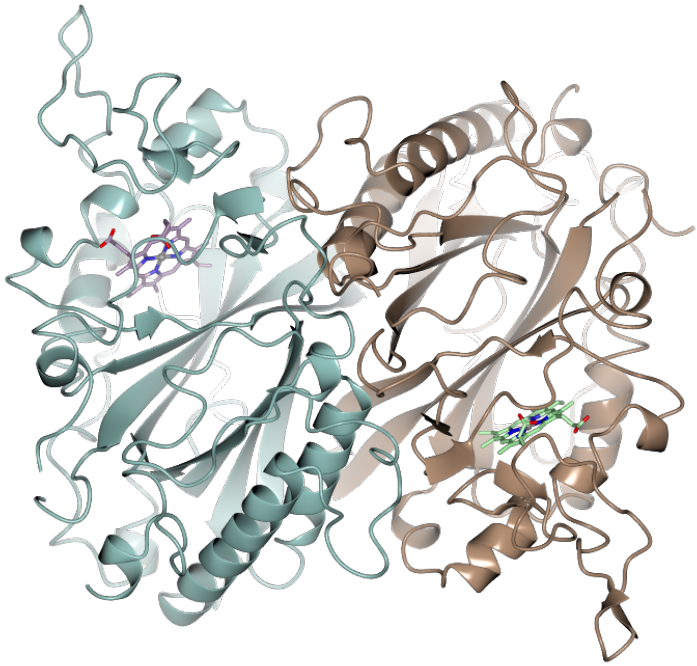

**Figure 4.5:** Overall view of the dimer found in the asymmetric unit of the structure measured from a DtpA ferric crystal at beamline BM30, ESRF. Monomers are coloured differently and haem cofactors are represented with cylinders.

At the active site (monomer B), the haem prosthetic group was found coordinated by the canonical proximal histidine broadly found in haem peroxidases and globins in general. At the distal site, catalytic residues Asp251 and Arg369 were found, conserved among DyPs (Sugano 2009) and giving them its special reactivity. For instance, the presence of the Asp251, a weaker base than the catalytic His found in classic peroxidases, allows the protein to perform at acidic pH's (Sugano, Muramatsu, Ichiyanagi, Sato, & Shoda, 2007). At the distal site in Dataset-1, at the axial position above the haem iron, we could model well a molecular oxygen ($O_2$), and there was also some extra electron density that was modelled as a water (Figure 4.6, left panel), which establishes a hydrogen bond (2.62 Å) with the catalytic Asp251. Presumably, the photo-reduction of the protein happened at low doses, and the dominant

structure in the hybrid structure is the photo-reduced state. However, we aimed to obtain an intact oxy-ferrous structure of DtpA and, to this end, a further X-ray diffraction dataset was collected (Dataset-2). At the active site (heme B) from Dataset-2 (Figure 4.6, right panel), electron density was found just supporting a the molecular oxygen coordinating the photo-reduced heme iron, indicating a 'pure' DtpA oxy-ferrous species which was validated by means of UV-vis *in crystallo* spectroscopy. Data collection and refinement statistics for both datasets are detailed in Table 41.

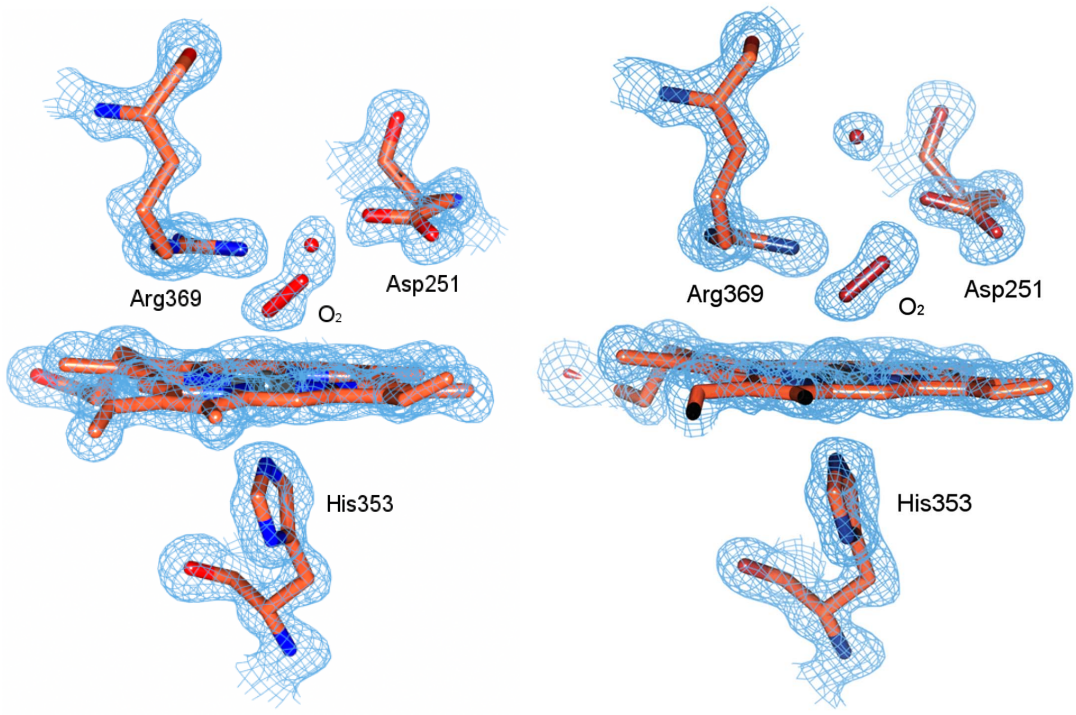


**Figure 4.6:** Representation of the heme site at monomer B from the structures measured from a DtpA ferric crystal, yielding a photo-reduced oxy-ferrous species. On the left, structure from to the first dataset measured (Dataset-1), where a molecular oxygen ($O_2$) and a water molecule (red ball) were modelled at the distal site. Dataset-1 corresponds to a mixed species between the ferric and the oxy-ferrous states. On the right, structure from to a second collected dataset (Dataset-2) is shown, where a validated (by UV-vis *in crystallo* spectroscopy) ´pure' oxy-ferrous species can be observed, with the molecular oxygen coordinating the photo-reduced iron. In both cases, the proximal His353, and the conserved distal residues Asp251 and Arg369 are shown. Electron densities are contoured at 1$\sigma$ (0.532 $e^-/Å^3$).

**Table 4.1:** Data collection and refinement statistics for the two datasets measured from a ferric DtpA crystal at beamline BM30, ESRF (Grenoble, France). Dataset 2 would correspond to the validated oxy-ferrous species, whereas Dataset 1 would be a hybrid species composed by the ferric and photo-reduced states of DtpA.

| | **Dataset 1** | **Dataset 2** |
|---|---|---|
| Space group | $P2_1$ | $P2_1$ |
| Unit cell (Å, °) | a=59.78, b=70.63, c=77.64; β=93.0 | a=59.79, b=70.63, c=77.66; β=93.0 |
| Resolution (Å) | 77.53 – 1.45 (1.47 – 1.45) | 46.16 – 1.49 (1.52 – 1.49) |
| $R_{merge}$ | 0.107 (0.749) | 0.096 (0.776) |
| Unique reflections | 112,162 (5,645) | 103,431 (5,165) |
| I/σ(I) | 8.0 (1.6) | 8.0 (1.3) |
| $CC_{1/2}$ | 0.994 (0.468) | 0.993 (0.509) |
| Completeness (%) | 98.3 (99.6) | 98.3 (99.7) |
| Multiplicity | 2.9 (3.0) | 2.9 (3.0) |
| $R_{factor}$ (%) | 22.1 | 21.9 |
| $R_{free}$ (%) | 25.4 | 24.9 |
| RMSD bond lengths (Å) | 0.014 | 0.015 |
| RMSD bond angles (º) | 1.68 | 1.74 |
| Ramachandran favoured (%) | 97.0 | 97.0 |
| Absorbed dose (MGy) | 0.99 | 1.98 |
| PDB accession code | 5mjh | 5map |

### *4.4.1 Characterising X-ray photo-reduction by the beam.*

With the intention to quantify the dose needed to photo-reduce a given DtpA ferric crystal, a kinetic online UV-vis experiment was designed to monitor the progressive photo-reduction of the protein with the micro-spectrophotometer at ESRF. This simply consisted in continuously acquiring online UV-vis spectra from a mounted crystal (at a previously optimized position) every two seconds after opening the shutter, until the features for the oxy-ferrous species were obvious and no further changes were appreciable in

the spectrum. The beam used had a top-hap shape and 100x100 μm to ensure an even exposure in all the crystal used (approximately 75x75x30 μm), and it was inspected with a light spot of 25 μm of diameter. This kinetic measurement was also carried out at beamline BM30 at ESRF, and its outcome can be seen in Figure 4.7, were an initially ferric crystal (spectrum in blue) develops to a photo-reduced form (spectrum in magenta) upon beam irradiation (1.0 Å wavelength). With this experimental information, and focusing on the spectral changes at wavelength 555 nm, where a sharp α-band develops during the photo-reduction process, the percentage of reduced protein was plotted as a function of the absorbed dose by the crystal (Figure 4.7, bottom panel). Doses were calculated with the online software RADDOSE-3D (Zeldin et al., 2013). As it can be noted in the figure, most the protein becomes reduced after just 10 kGy of X-ray dose (with half-dose of ~3-4 kGy), which represents just around 1% of the dose needed for a measurement of a complete dataset at that beamline (0.99 MGy estimated for Dataset-1). Thus, photo-reduction occurs very rapidly for DtpA, after the first few measured diffraction frames (~4.5 kGy absorbed per pattern).

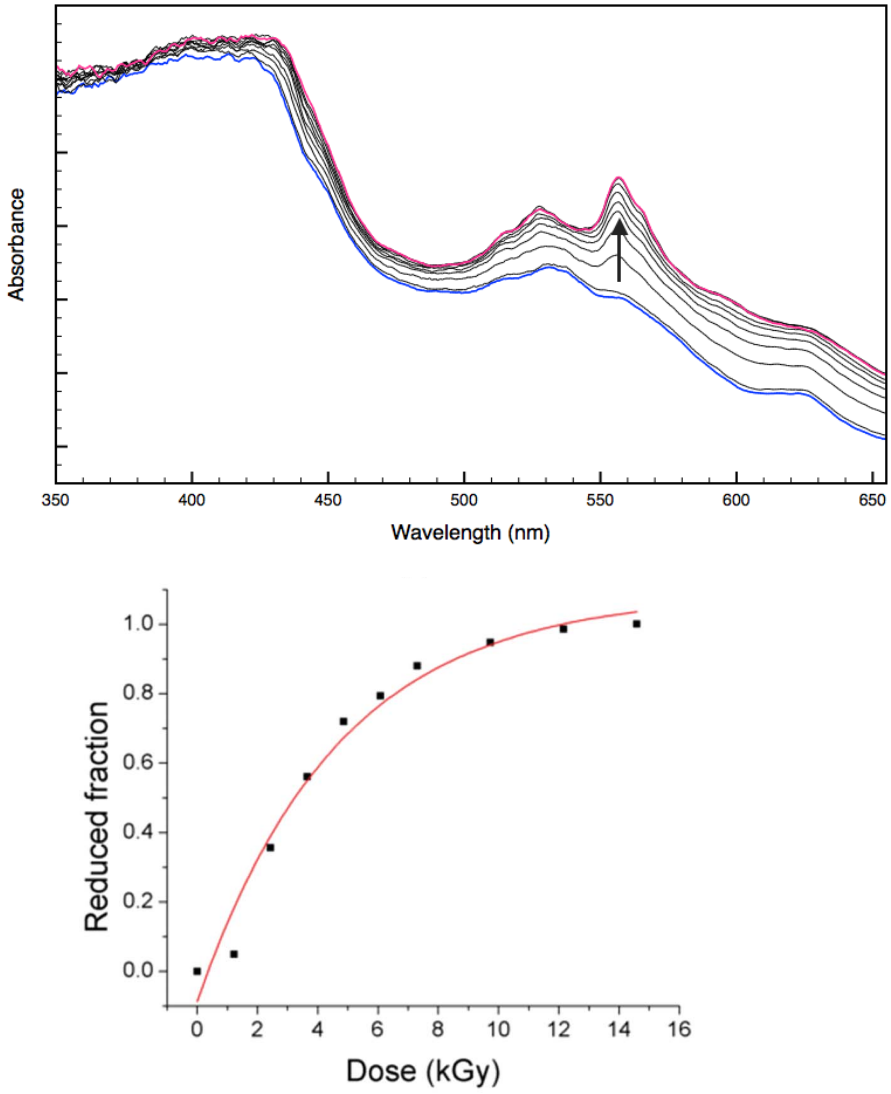


**Figure 4.7:** Photo-reduction of DtpA by the X-ray beam. Top panel - kinetic single crystal UV-vis spectra of an initially ferric DtpA crystal (in blue), with measurements taken every two seconds after the aperture of the beam shutter, and reaching a final photo-reduced oxy-ferrous state (in magenta). Bottom panel- Representation of the reduced fraction of DtpA (calculated for wavelength 555 nm) exposed to the beam as a function of the absorbed dose by the illuminated area of the crystal. Doses in each point were estimated using RADDOSE-3D (Zeldin et al., 2013). Experimental data were fitted to a hyperbolic saturation function using *Origin8* (OriginLab corporation).

## 4.5 DtpA ferric structure

One of our aims in determining the different catalytic states for DtpA, was to obtain the structure of the intact ferric species, since it corresponds to the ground resting state within the peroxidase cycle. However, as shown in the previous section, this was challenging due to the great sensitivity of the metal centre to X-rays and its rapid photo-reduction by the beam at low doses. To overcome this limitation, a specially designed data collection experiment was conducted at ESRF (beamline ID29-3), to mitigate radiation damage and obtain the resting state structure of DtpA. It consisted of performing a helical collection along the length of a large DtpA ferric crystal (~350 μm, Figure 4.8). In this way, the beam would continuously interact with fresh material as the crystal was translated through the helical path, and the total dose used to obtain a complete dataset was spread among a much bigger crystal volume. This collection strategy was coupled with the use of a state-of-the-art EIGER-4M detector (Casanas et al., 2016; Dectris, Ltd.), which could handle very short exposures per frame (maximum frequency of 750 Hz), and a dead time between frames as low as of 3.8 μs, performing ultra-fast data collections.

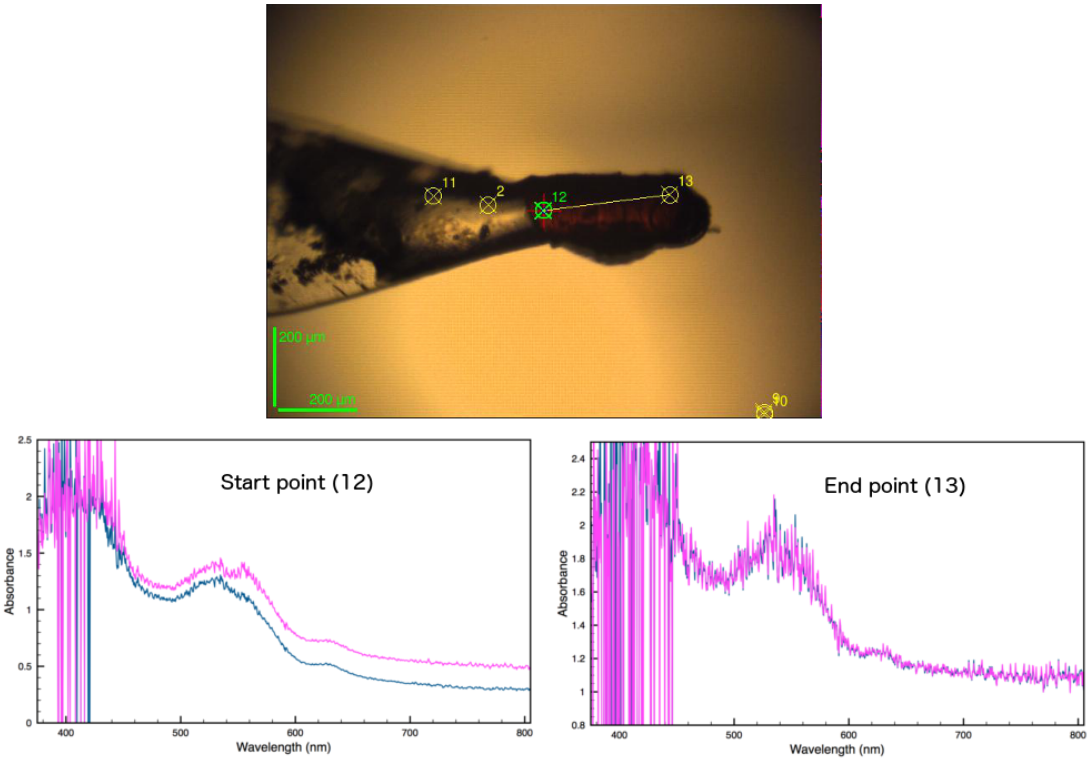


**Figure 4.8:** Helical data collection from a ferric DtpA large crystal to obtain a low-dose structure. Top panel – DtpA ferric crystal mounted at the beamline ID29-A3 at ESRF (Grenoble, France), just before performing the helical X-ray data collection (between points 12 and 13 in the picture) using an EIGER-4M detector (Dectris Ltd.). Bottom panels – spectroscopic validation with the online UV-vis micro-spectrophotometer, with measurements before (in blue) and after (in magenta) data collection, at the start (left) and end (right) of the helical path.

To validate the oxidation state of the structure, the online UV-vis spectrophotometer at ESRF (Von Stetten et al. 2015) was used to carry out measurements in different points along the X-ray measured path (Figure 4.8, bottom panel). Since the geometry of the instrument was off-axis, alignment of the light path was carried out at the beginning of the beamtime to probe the same region as the X-ray beam. However, the crystal needed for this kind of experiment was large, and a certain fraction of unexposed crystal would be unavoidably acquired with each spectrum. Another disadvantage of the size of

the crystal was the substantial increase in optical density, which resulted in most of the points selected yielding poor quality spectra with saturated features. Nevertheless, ferric features in the β/α and CT bands were observed, before and after X-ray exposure (with no evidence of sharp β/α features seen), suggesting that the oxidation state had not changed. The weighted absorbed dose estimated by RADDOSE-3D (Zeldin et al. 2013) for this crystal was of 12 kGy.

The resulting DtpA structure presented no signs of photo-reduction at the active site (Figure 4.9). At the distal axial position, instead of a molecular oxygen ($O_2$), generated during photo-reduction, we found electron density corresponding to a tightly-bound water molecule that coordinates the heme Fe(III), at 1.98 Å. This distal water (W1) forming the hexa-coordinate species, is forming a hydrogen bond (2.50 Å) with another near water present in the pocket (W2), which in its turn is establishing another hydrogen bond interaction (2.68 Å) with the catalytic Asp251. These two waters, belong to an arrange of well-defined waters found at the distal site, establishing a proton-transfer network which is putatively involved in catalysis during the peroxidase cycle. Well-ordered waters like these, with spherical and strong electron density are another sign of a single state of the enzyme in the structure, in contrast with the streaky features observed in photo-reduced structures. Another feature of the structure, which could be also observed in the previous photo-reduced DtpA, is the salt bridge (2.72 Å) formed by the conserved distal residue Arg369 and one of the heme propionate groups. Arg369 is also stabilising the bound

distal water by a hydrogen bond interaction (2.91 Å). Data collection and refinement statistics for this structure are detailed in Table 71.

A superposition of the active sites was carried out between the ferric DtpA species and the structures measured during the determination of the oxy-ferrous species (Dataset 1 and 2), which is shown in Figure 4.10. Minor rearrangements can be noted for some of the distal residues, such as the Asp251 or the Phe390, which are slightly displaced in the photo-reduced structures to accommodate the molecular oxygen.

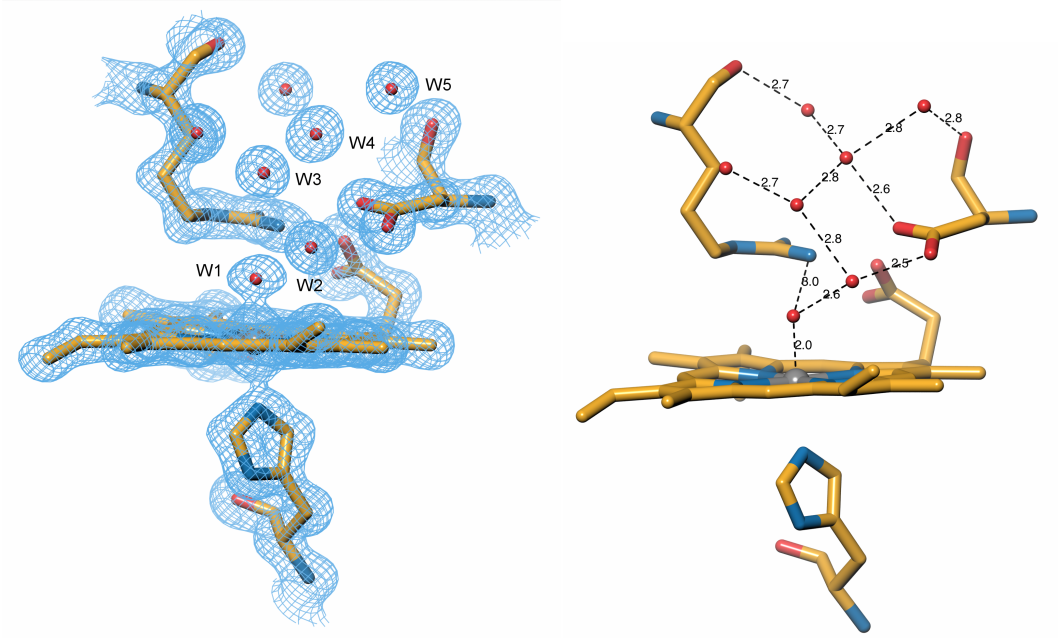


**Figure 4.9:** Heme site of the DtpA ferric species. Left panel - General view of the active site of the ferric DtpA species, with the most important residues shown (His353, Asp251, and Arg369). A tightly bound (1.98 Å) distal water is found coordinating the heme iron, and some of the well-ordered waters forming the proton-transfer network at the distal pocket are shown. Right panel – hydrogen bonds network (dashed lines) of the water molecules at the distal pocket.

**Table :** Data collection and refinement statistics for the ferric DtpA helical dataset, measured at beamline ID30A-3 at ESRF (Grenoble, France), using a Gaussian 15x15μm (FWHM) beam of 0.9677 Å wavelength, and an EIGER-4M detector (Dectris).

| | **Resting state DtpA** |
|---|---|
| **Space group** | $P2_1$ |
| **Unit cell (Å, °)** | a=60.12, b=71.03, c=78.09; β=92.83 |
| **Resolution (Å)** | 40.20 – 1.41 (1.46 – 1.41) |
| **$R_{merge}$** | 0.145 (1.292) |
| **Unique reflections** | 118339 (9905) |
| **I/σ(I)** | 7.8 (1.0) |
| **$CC_{1/2}$** | 0.997 (0.529) |
| **Completeness (%)** | 93.9 (80.4) |
| **Multiplicity** | 5.2 (5.1) |
| **$R_{work}$ (%)** | 18.9 |
| **$R_{free}$ (%)** | 21.6 |
| **RMSD bond lengths (Å)** | 0.011 |
| **RMSD bond angles (º)** | 1.5 |
| **Ramachandran favoured (%)** | 97.0 |
| **Absorbed dose (kGy)** | 12 |
| **PDB accession code** | 6gzw |

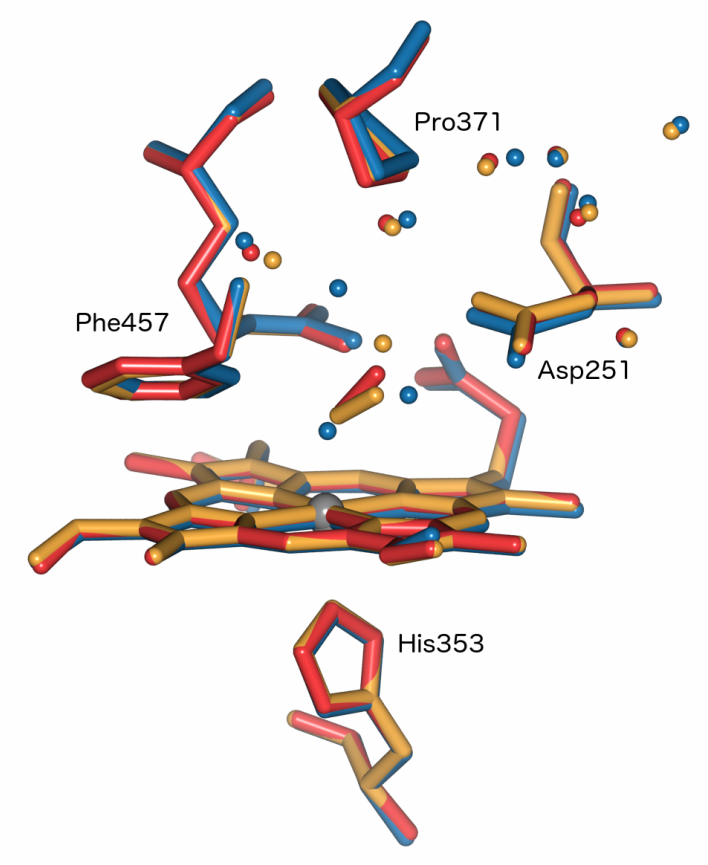


**Figure 4.10:** Superposition of the DtpA heme sites for the ferric species (in blue), oxy-ferrous species (in red), and the partly photo-reduced transition state of Dataset 1 (in yellow). Some conformational changes can be noted, such as the movement of the catalytic Asp251 and the Phe457, to accommodate the molecular oxygen present in the photo-reduced structures.

## 4.6 Activation of the peroxidase cycle of DtpA

An important step in the development of our methodology was to initiate the peroxidase cycle within the DtpA crystals, in order to generate catalytic intermediates and try to obtain its intact structures. For this aim, UV-vis spectroscopy was a crucial tool to assess the state of the protein within the crystal at each stage. We started by using the natural substrate for peroxidases, hydrogen peroxide ($H_2O_2$), which has been used numerous times

in the literature to generate peroxidase intermediates (Berglund et al., 2002; Gumiero et al., 2011; Kuhnel et al., 2007).

### *4.6.1 Activation with Hydrogen peroxide*

In our initial tests and optimization sessions at the cryobench (von Stetten et al., 2015) at ESRF, we characterized the ferric spectrum of DtpA crystals (blue in Figure 4.4) as shown earlier in the chapter. Soaking experiments with hydrogen peroxide into DtpA ferric crystals were also performed, with clear spectral changes occurring when crystals where soaked for 1-2 min in a 0.5% (v/v) $H_2O_2$-supplemented crystallization condition (in magenta in Figure 4.11). Detailed soaking procedures are described in Chapter 3. Upon soaking, DtpA crystals would change from brown to a dark-red colour and produce some bubbles, and eventually would disintegrate if left too long. The features observed in the spectrum for the $H_2O_2$-treated crystals were consistent with the Compound III species for DtpA when comparing them with solution data (Figure 4.12, from Dr. Chaplin´s doctoral thesis; Kekilli et al., 2017). In Table 43, a relation of UV-vis spectral peaks observed for different DtpA species is shown. As it can be appreciated, Compound III could be easily confused with the oxy-ferrous species due to its spectral resemblances. Nevertheless, some examples were found in the literature (Berglund et al., 2002; Kuhnel et al., 2007) where this intermediate was indeed formed in other peroxidases when applying an excess of $H_2O_2$ to the crystals. Therefore, with spectroscopic validation of CmdIII species generated within the crystals, those were cryo-cooled in liquid nitrogen and taken to beamline BM30 (ESRF) for X-ray diffraction experiments. A complete dataset was collected, followed by a

further UV-vis spectrum acquired with the online spectrophotometer (in orange in Figure 4.11). As it can be noted, the state of the protein had clearly changed after exposure to the X-ray beam. However, we were not sure about its identity. After analysis of DtpA solution data (Figure 4.12, Table 43), the radiation-induced state was identified as the deoxy-ferrous species. Unfortunately, due to time restrictions we were not able to measure a further dataset at that beamtime, that would have been that of the validated state. Thus, the measured dataset, corresponded to a mixed species, since initially the protein was in CmdIII state, and upon photo-reduction would develop to a deoxy-ferrous state, with potentially some intermediate states in-between (e.g. Compound-0, as in Kuhnel et al., 2007). A view of the active site of this hybrid species is shown in Figure 4.13. It can be noted how the dioxygen species of Compound III has been photo-reduced, and two water molecules can be observed at the distal site forming a hydrogen bond interaction at 2.55 Å.

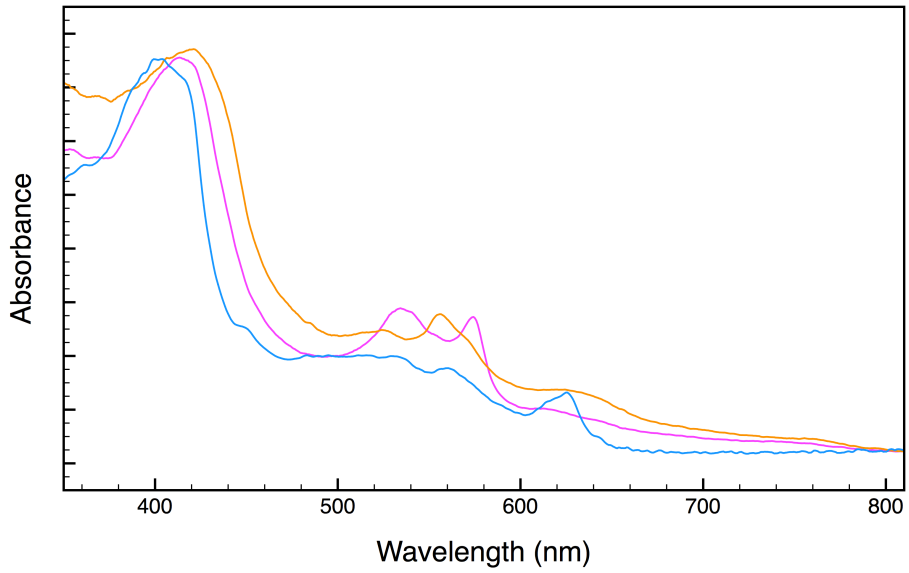


**Figure 4.11:** Single crystal UV-vis spectra of DtpA species: a ferric crystal, in blue; a ferric crystal soaked in 0.5% (v/v) $H_2O_2$ yielding Compound-III species, in magenta; and this same Compound-III species after a collection of a diffraction dataset, in orange. All spectra were

measured using the micro-spectrophotometer available (von Stetten et al., 2015) at ESRF, installed at beamline BM30. Each spectrum is the result of 10 accumulation of 200 ms each.

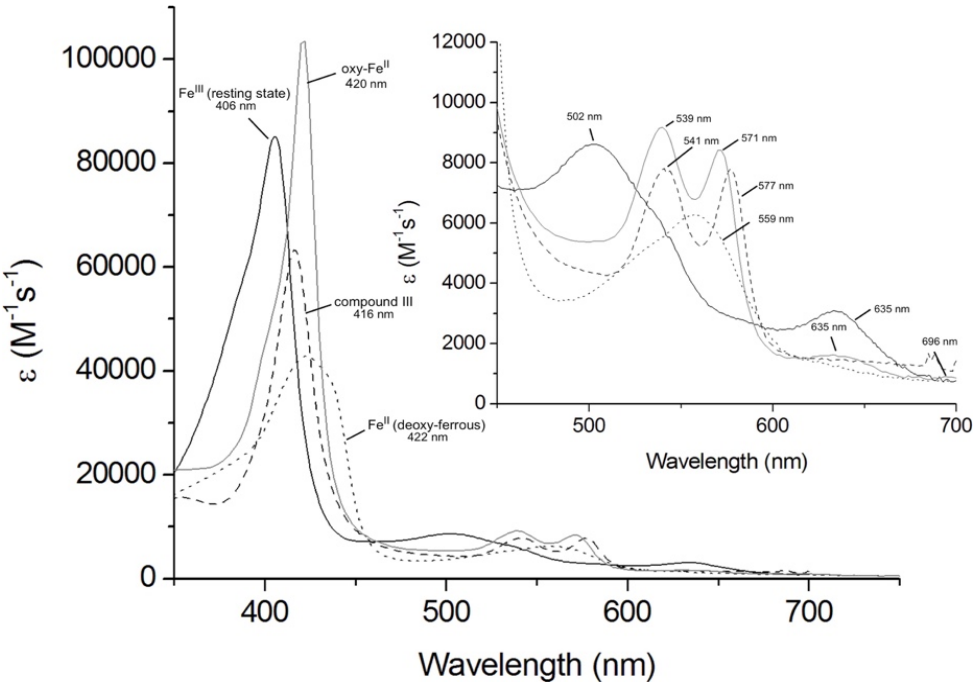


**Figure 4.12:** Solution UV-vis spectra from DtpA showing spectral signatures for resting state, oxy-ferrous and deoxy-ferrous species, and the peroxidase cycle intermediate compound III. Inset on the right shows the region of β/α and CT bands, with maximum peak wavelengths specified. (Dr. Amanda Chaplin thesis)

**Table 4.3:** Absorption maxima of bands observed in solution and *in crystallo* UV-vis spectra measured from different DtpA species, including intermediates from the peroxidase cycle. (sh) stands for shoulder. Data from Dr. Amanda Chaplin's doctoral thesis and this work.

| DtpA species | Soret Band (nm) | β/α bands (nm) | CT band (nm) |
|---|---|---|---|
| Ferric | 406 | 502, 540(sh) | 635 |
| Oxy-ferrous | 420 | 535, 571 | 635 |
| Deoxy-ferrous | 422 | 559 | - |
| Deoxy-ferrous (generated in crystal) | 420 | 525, 557 | - |
| Compound I | 399 | 530, 557 | 614, 644 |

| | | | |
|---|---|---|---|
| Compound II | 419 | 528, 557 | 621 |
| Compound III | 416 | 541, 577 | - |
| Compound III (generated in crystal) | 414 | 535, 574 | - |

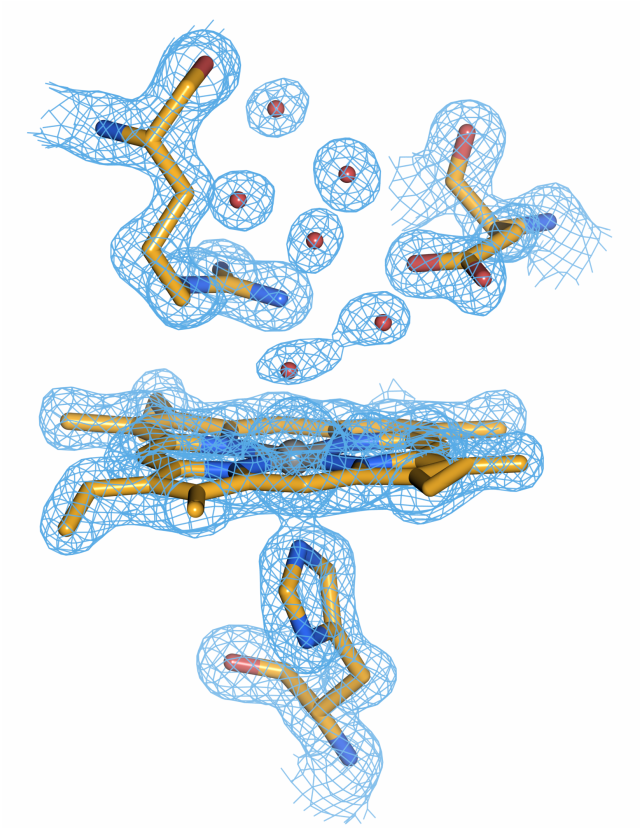

**Figure 4.13:** Active site view of DtpA mixed species corresponding to an initial Compound-III state after photo-reduction by the X-ray beam for collection of a diffraction dataset. Note how the initial dioxygen bond has developed into two molecules of water at the distal site.

In the next visit to ESRF, this time at beamline ID29-A3 (massif-3), DtpA crystals in the Cmd-III state were brought, prepared again by soaking with an excess of 50 mM $H_2O_2$. Firstly, a single crystal UV-vis spectrum was measured (in blue in Figure 4.14) with the online micro-spectrophotometer, to confirm the state in which the protein was at. Compound III features were observed, with

β/α bands at ~530 and ~575 nm, and no CT band. A first X-ray diffraction dataset was measured to photo-reduce the protein and take it to the deoxy-ferrous state (as previously observed), which was confirmed with a further UV-vis measurement (in orange in Figure 4.14). This time a second diffraction dataset was collected to obtain the intact validated species. A final UV-vis spectrum (in red in Figure 4.14) was measured to confirm that the crystal was still at the same state.

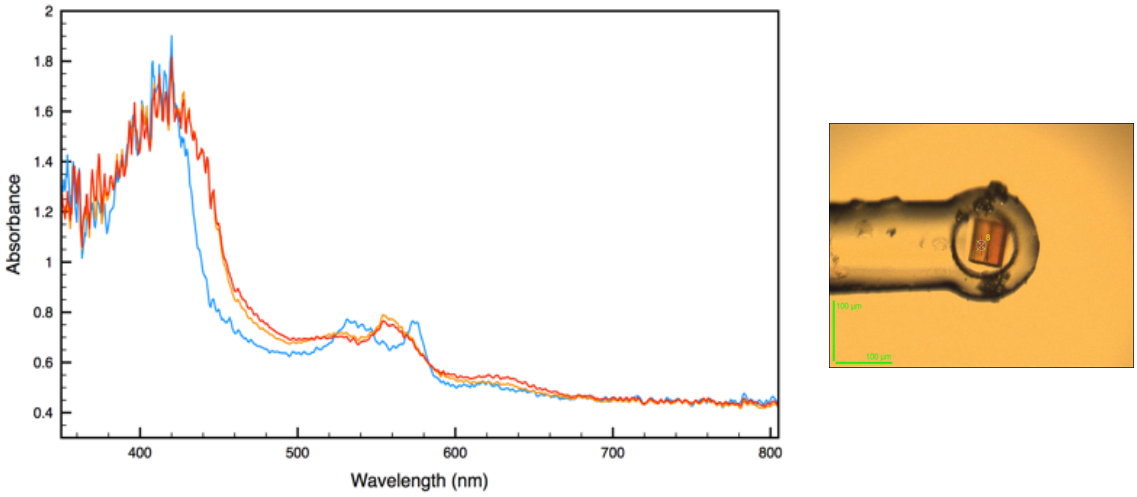


**Figure 4.14:** Single-crystal UV-vis online spectra of a DtpA crystal (right panel): in a starting Compound III state (in blue); after a first X-ray diffraction dataset (in orange), where the protein presents a deoxy-ferrous state; and after a second X-ray diffraction dataset (in red), where DtpA still shows a deoxy-ferrous signature. Spectra were measured with the online micro-spectrophotometer (von Stetten et al., 2015) installed at beamline ID29-A3 at ESRF.

The resultant structure is that of the deoxy-ferrous species, where the dioxygen bound species found in Compound III, would be sequentially photo-reduced by the beam to yield two molecules of water, which would exit the distal site leaving a five-coordinate heme species, as proposed by Berglund et al. (2002). In Figure 4.15, a general view of the active site environment of the deoxy-ferrous species is shown. The empty distal site can be appreciated,

together with one of the putative leaving waters (at 3.43 Å from the heme iron), which is interacting with the catalytic Asp251 through a hydrogen bond (2.78 Å). In Table 44, data collection and refinement statistics are detailed. The mechanism proposed by Berglund et al. (2002) for the sequential photo-reduction of Horseradish peroxidase (HRP) Compound-III species into the deoxy-ferrous species is shown in Figure 4.16.

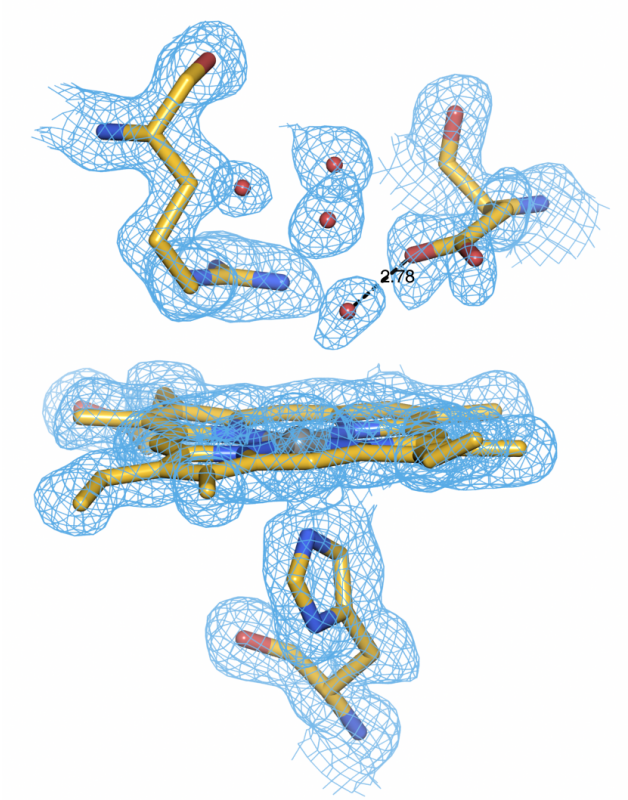


**Figure 4.15:** Representation of DtpA deoxy-ferrous active site. The putative leaving distal water, at 3.53 Å from iron haem, is set to 0.5 occupancy, and is interacting through a hydrogen bond (2.78 Å, dashed line) with the catalytic Asp251. Electron density is shown as a blue mesh contoured to 1 σ (0.426 $e^-/Å^3$).

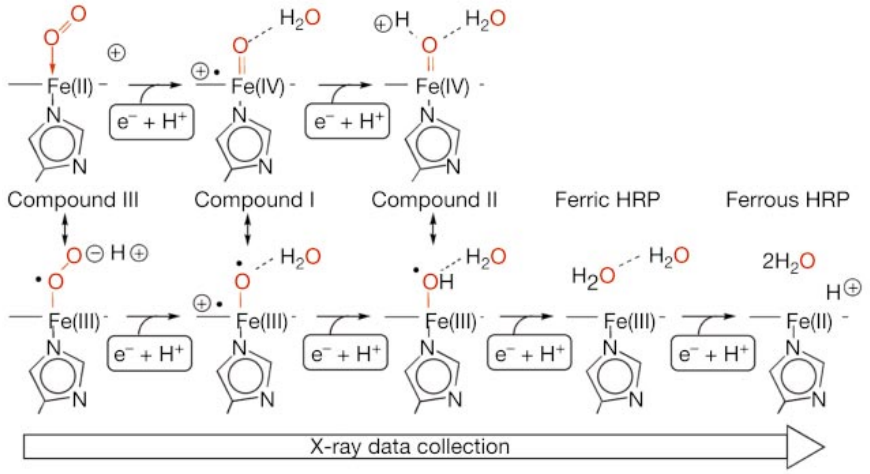


**Figure 4.16:** A proposed mechanism (Berglund et al., 2002) for the reduction of the bound dioxygen species of compound-III in Horseradish peroxidase (HRP) to two molecules of water, yielding the deoxy-ferrous species. Structures linked by double arrows are isoelectronic with each other.

**Table 4.4:** Crystallographic data processing and refinement statistics for deoxy-ferrous DtpA. The dataset was measured at beamline ID30A-3 at ESRF (Grenoble, France), using a wavelength of 0.9677 Å and an EIGER-4M detector (Dectris).

| | **Deoxy-ferrous DtpA** |
|---|---|
| **Space group** | $P2_1$ |
| **Unit cell (Å, °)** | a=59.99, b=70.69, c=77.99; β=92.02 |
| **Resolution (Å)** | 48.36 – 1.61 (1.64 – 1.61) |
| **$R_{symm}$** | 0.075 (0.736) |
| **Unique reflections** | 83002 (3891) |
| **I/σ(I)** | 8.61 (1.31) |
| **Completeness (%)** | 98.1 (92.8) |
| **Multiplicity** | 3.8 (3.9) |
| **$R_{factor}$ (%)** | 21.9 |
| **$R_{free}$ (%)** | 26.4 |
| **RMSD bond lengths (Å)** | 0.0178 |
| **RMSD bond angles (°)** | 1.9 |
| **Ramachandran favoured (%)** | 95.8 |
| **Absorbed dose (MGy)** | 3.52 |

### *4.6.2 Activation with peracetic acid (PAA)*

Several trials were carried out by soaking DtpA ferric crystals in $H_2O_2$-containing solutions spanning the concentration range 0.1-50 mM, but no signs of the intermediates Compound I and II were observed in the UV-vis spectra. Either a ferric signature would be observed, or there would be too much hydrogen peroxide and the enzyme would go straight to Compound III. For this reason, an alternative activation substance of the peroxidase cycle was tested. In the literature, it was noted the widespread usage of the organic peroxide peracetic acid (PAA; Berglund et al., 2002; Kuhnel et al., 2007) to generate the intermediates compound I and II. The chemical structure of PAA is shown in Figure 4.17.

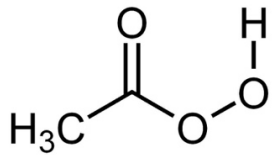


**Figure 4.17:** Chemical structure of peracetic acid.

In the next trip to SLS (Villigen, Switzerland), DtpA ferric crystals were soaked in 2-5mM PAA and a clear colour change from brown to olive green could be observed, indicating the formation of the intermediate Compound I. Crystals were cryo-cooled in liquid nitrogen and mounted on the beamtime for its spectroscopic inspection with the on-axis micro-spectrophotometer (Pompidor et al. 2013) equipped at beamline X10SA. The characteristic signature for DtpA compound I species was observed, with a succession of 4 peaks in the region 500-650 nm (in turquoise in Figure 4.18). An unexpected change was observed after ~2 min of the crystal being mounted at the

goniometer, as the Cmd-I features were fading out (in green in Figure 4.18) to stabilise at compound II state after ~4-5min in total (in purple in Figure 4.18). Therefore, it appears that Compound I is not stable at the beamline environment, even if its cooled at 100 K by the cryo-stream, and no X-rays have interacted yet with the crystal.

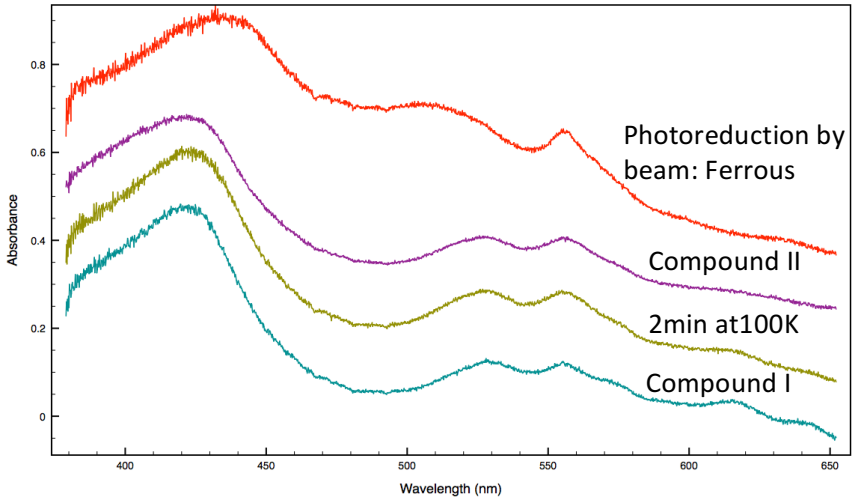


**Figure 4.18:** Single crystal UV-vis spectra of a DtpA peroxidase intermediates. Initially, the crystal was in the Compound I intermediate state (turquoise), and after of a couple of minutes of being mounted at the beamline at 100 K, the spectrum experience changes (green) with some features fading out, until stabilising after ~5min in total at the Compound II state (purple). Following collection of an X-ray diffraction dataset, a further spectrum was acquired showing new features corresponding to a photo-reduced ferrous state of DtpA. All spectra were measure at beamline X10SA, Swiss Light Source, using the micro-spectrophotometer MS3 (Pompidor et al., 2013).

A dataset was collected at this point, with Cmp-II as the starting state of the DtpA crystal. As expected, this high-oxidation state intermediate is prone to reduction and became photo-reduced yielding a DtpA species which was also inspected by UV-vis (in red, Figure 4.18). This state was consistent with a photo-reduced ferrous form of DtpA, and its active site is shown in Figure 4.19,

where a distal water was found coordinating the iron at 2.1 Å. Data collection and refinement statistics for the structure are detailed in Table 45. This hydrated ferrous form has been also described in the literature by Gumiero and co-workers (2011), after photo-reduction of CmpI states found in Cytochrome c peroxidase (CcP) and ascorbate peroxidase (APX), monitoring changes in the crystals also using single-crystal UV-vis spectroscopy.

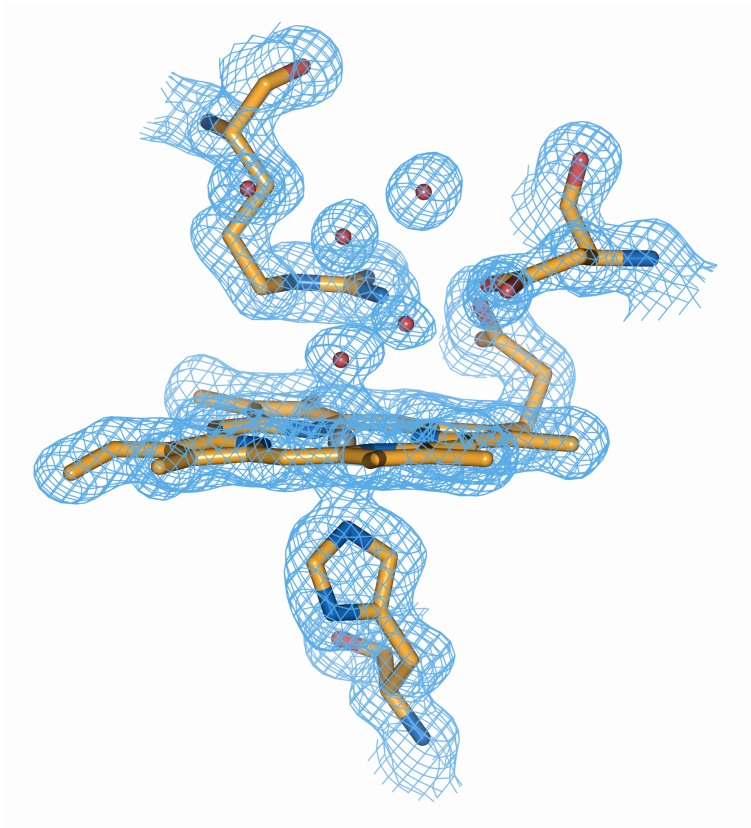

**Figure 4.19:** Active site heme environment for the ferrous hydrated-DtpA species at 1.71 Å resolution, resulting from the photo-reduction of Cmd-II species by the X-ray beam. A distal water at 2.1 Å can be observed coordinating the iron. Electron density map is shown as a blue mesh and contoured at 1σ. Diffraction data was measured at beamline X10SA at Swiss Light Source.

**Table 4.5:** Data collection and refinement statistics for the ferrous structure of DtpA obtained by photo-reduction by the X-ray beam of a crystal in the Compound II state. The dataset was collected at beamline X10SA at Swiss Light Source, at 0.9 Å wavelength.

| | **ferrous DtpA** |
|---|---|
| **Space group** | $P2_1$ |
| **Unit cell (Å, °)** | a=59.75, b=70.31, c=77.85; β=93.13 |
| **Resolution (Å)** | 48.63 – 1.71 (1.74 – 1.71) |
| **$R_{merge}$** | 0.042 (0.247) |
| **Unique reflections** | 69,383 (3,593) |
| **I/σ(I)** | 11.5 (2.2) |
| **$CC_{1/2}$** | 0.998 (0.758) |
| **Completeness (%)** | 99.9 (99.0) |
| **Multiplicity** | 1.9 (1.9) |
| **$R_{factor}$ (%)** | 16.4 |
| **$R_{free}$ (%)** | 20.4 |
| **RMSD bond lengths (Å)** | 0.019 |
| **RMSD bond angles (º)** | 1.88 |
| **Ramachandran favoured (%)** | 97.1 |
| **Absorbed dose (MGy)** | 0.50 |

## 4.7 Serial crystallography with DyPs

The opportunity arose to apply DyP proteins to serial crystallography methodologies at room-temperature, and we performed serial femtosecond crystallography experiments (SFX) at SACLA, and synchrotron serial crystallography (SSX) at Diamond, in order to obtain damage-free structures and low dose/dose-series structures, respectively. For this purpose, it was needed to optimize the crystallization conditions to produce large quantities of microcrystals. Two protein targets were developed, one of them DtpA, with

which we had worked extensively, and also DtpAa, another DyP (belonging as well to subclass A) from *Streptomyces lividans*, which was starting to be investigated at that time.

### *4.7.1 Production of microcrystals.*

The production of microcrystals was carried out in batch crystallization in 1.5ml eppendorfs in relatively large volumes (100-500ul) to obtain a large amount of microcrystals. Potential candidate conditions were searched among commercial screens routinely used. Selected conditions were transferred to batch crystallization set-up, initially screening with small volume batches (20-50ul) to minimize the usage of concentrated protein. Various variables would be screened, such as concentration of precipitant, salts, protein, and also the volume ratio protein to condition used in the batches. When a suitable condition was found to deliver good enough microcrystals, it was scaled up to bigger volumes to obtain large amounts.

#### *4.7.1.1 DtpA microcrystals:*

For DtpA we explored two conditions delivering microcrystals, one using the His-tagged protein (JCSG+_A12) and the other one with the cleaved protein ("Lithium" condition).

- *Protein stock* at 11-15mg/ml, in buffer 10mM NaAcetate pH 5.0, 100mM NaCl.
- *Original hit conditions*:

- **JCSG_A12**: 20% PEG 3350, 200mM $KNO_3$.
- **'Lithium' condition**: 20% PEG 4000, 200mM lithium sulphate, 100mM HEPES pH=7.5

For the 'Lithium' condition, DtpA protein with the cleaved His-tag was used since the uncleaved protein would not crystallize. Batches were set up initially testing volume ratios 1:1, 1:2 and 1:3 protein to condition, being the first the one which delivered best results. If the ratio was higher than 1:1, protein to condition, crystallization would happen too fast and crystals of poor quality would be obtained. The protein fraction used was at 12-14 mg/ml concentration, and the crystallization condition 35% PEG 4000, 200 mM $LiSO_4$, 100 mM HEPES pH 7.5. Batches with 1:1 ratio scaled up to ~500 μl were set up. And initial formation of precipitate would be observed after mixing, and microcrystals would grow overnight (Figure 4.20, left panel) at 18ºC or RT with similar results (although RT appeared to deliver sharper crystals). The mixing process was done always by pipetting up and down several times just after setting up the batch. Vortexing was tested for some batches at the beginning, but turned out to prevent crystallization of the protein, which would aggregate on the walls of the Eppendorf. Final concentrations for one of the large batches with the 'Lithium' condition were:

- 6 mg/ml of cleaved DtpA
- 17.5% PEG 4000
- 100mM LiSO4
- 50mM HEPES pH 7.5.

The JCSG+_A12 condition was shown to produce lots of microcrystals in the commercial screen, and for this reason was selected to take it onto batch trials. DtpA used for this condition had the His-tag on, as obtained from the purification process. Protein concentration used was between 11-15 mg/ml. The crystallization condition fraction used was composed of 20% PEG 3350 and 200 mM $KNO_3$. Several volume ratios protein to condition were tested initially - 1:1, 1:2, 1:3, 1:4, 1:6, 1:8, 1:10. The optimal one for our purposes was the 1:6 ratio, which delivered a large number of microcrystals with sharp edges (Figure 4.20, right panel), and scaled-up well to final volumes of ~500 μl. Batches were mixed by pipetting up and down several times after set up, and crystals would grow overnight at either 18ºC or RT. Final concentrations for one of the JCSG+_A12 batches were:

- 1.7mg/ml DtpA
- 17.1% PEG 3350
- 171 mM $KNO_3$.

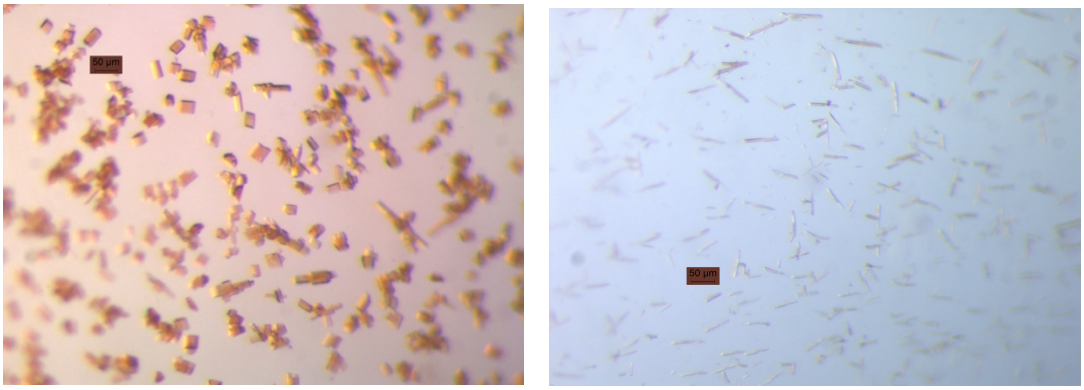

**Figure 4.20:** DtpA microcrystals grown in batch using JCSG+_A12 condition (right panel), and 'lithium' condition (left panel).

#### *4.7.1.2 DtpAa microcrystals:*

In the case of DtpAa, also two conditions were developed providing microcrystals: pHClear_D4, and 'MIB' condition.

- *Protein stock* of 13-15 mg/ml in buffer 10mM Na Acetate pH= 5.0, 200mM NaCl; His-tagged protein used in both cases.
- *Original hit conditions*:
    - **pHClear_D4:** 20% PEG 6000, 100mM HEPES pH=7.0
    - **"MIB":** 25% PEG 1500, 100mM MIB buffer (which stands for Malonate, Imidazole, and Boric acid).

Initially, the 'MIB' condition was developed, which was observed to deliver a large number of crystals in commercial screen drops. Test small batches were set up with volume ratios protein to condition - 1:1, 1:2, 1:3, 1:4, and 1:6. The protein fraction had a concentration of 13-15 mg/ml, and the condition fraction was composed of 25% PEG 1500 and 100 mM MIB buffer. Batches were always mixed by pipetting several times after setting up. DtpAa microcrystals, shown in Figure 4.21, grew after 1-2 days in the 1:2 batch, and lots of them were observed at the 1:6 batch after several days. Therefore, bigger batches were set up at these two ratios. At some point of the optimization process, it was noted that an imidazole complex was likely to be forming when setting up the batch, since there was a change of colour of the solution from brown to reddish. This compound was present in the buffer and is a common known ligand for haemproteins. For this reason, it was decided to develop another condition, since we were interested in the native form of DtpAa in the first place, and the active site should be in a state suitable for soaking,

e.g. for future experiments involving the activation of the protein with hydrogen peroxide (or peracetic acid). Final concentrations of the components in one of the large 'MIB' batches were:

- 6.75 mg/ml of DtpAa
- 16.7% PEG 1500
- 66.7 mM MIB buffer

The pHClear_D4 condition was selected, and initial tests performed at ratios 1:1, 1:2 and 1:3. The protein fraction was at 13-15 mg/ml concentration, and the condition fraction was composed of 20% PEG 6000 and 100 mM HEPES pH 7.0. Batches were always mixed by pipetting several times after set up. Ratio 1:1 resulted in the best outcome, delivering microcrystals overnight in large numbers. It was noted with further optimization testing, that decreasing the concentration of the protein fraction to 6-7 mg/ml would deliver more suitable DtpAa microcrystals, smaller (~15 μm) and sharp. Batches were scaled up to large volumes (~500 μl), delivering crystals in a similar fashion as the small ones. Final concentrations of the components in one of the large pHClear_D4 batches were:

- 3.4 mg/ml DtpAa
- 10% PEG 6000
- 50 mM HEPES pH 7.0

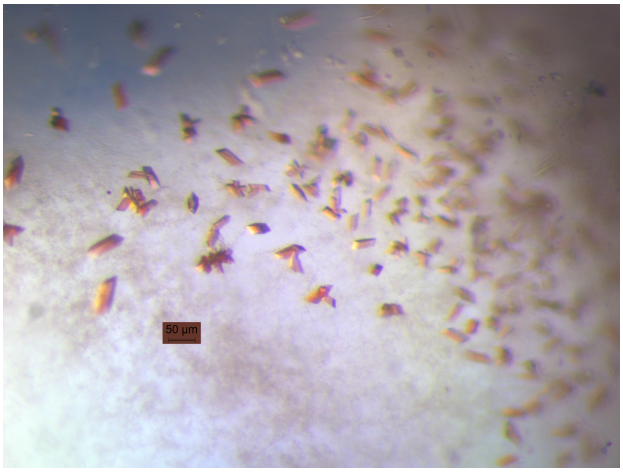


**Figure 4.21:** DtpA microcrystals grown in one of the batches at the 'MIB' condition.

### *4.7.2 Serial femtosecond crystallography (SFX) experiments.*

All produced batches with DtpA and DtpAa microcrystals were taken to the XFEL source SACLA (Hyogo, Japan), were data were collected from fixed targets using 10 fs X-ray pulses. Thus, we could obtain intact RT damage-free species for each of the three structures measured: resting states of DtpA and DtpAa, and the DtpAa – imidazole complex. Experimental procedures of loading and measurement of the of the chips, as well as experimental set up at the hutch at SACLA are detailed in Chapter 3. In Table 46, the different parameters for each of the structures measured are detailed.

**Table 4.6:** Data collection parameters and statistics for the different DyP SFX structures measured at SACLA in October 2017. The X-ray beam used was Gaussian with size 1.25x1.34 microns (FWHM), and the XFEL pulses were 10 fs long, with an energy of 10 keV (1.24 Å). Preliminary estimation of hit rates was performed on-the-fly with the program Cheetah (Barty et al. 2014).

| Sample | Chips measured | Attenuation | Processed patterns | Indexed patterns | Hit rate (%) |
|---|---|---|---|---|---|
| DtpA WT | 5 | 0.3 | 117,260 | 28,941 | 24.68 |
| DtpAa WT | 11 | 0.3 | 237,335 | 64,092 | 27.00 |
| DtpAa + imidazole | 4 | 0.3 | 102,800 | 29,070 | 28.28 |

### *4.7.3 DtpAa SFX structure at SACLA*

Around 20-30 batches (100-200 μl each) were prepared as just detailed in the previous section to obtain ferric DtpAa microcrystals, which were transported to SACLA at room temperature. They were loaded onto silicon nitride chips and measured at the XFEL hutch following procedures detailed in Chapter 3. A UV-vis spectrum was performed with one of the samples in-house at Essex, to confirm the oxidation state of the crystals (Figure 4.22). This spectrum of microcrystals in suspension, matches with the ferric features observed for DtpAa in solution, with the Soret band at a wavelength of ~405 nm. We assume that the same oxidation state remains stable during transportation and loading of the fixed targets. Samples were treated with care and were kept at roughly the same temperature conditions as in the crystallization lab.

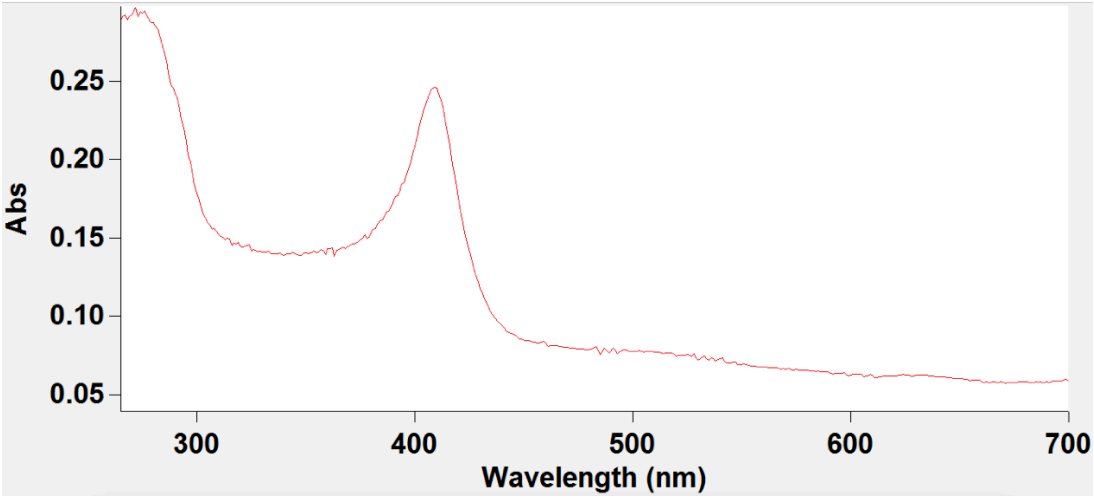


**Figure 4.22:** UV-vis spectrum of suspended DtpAa microcrystals in their own batch crystallization condition. Measurement was carried out with a Cary-60 spectrophotometer (Agilent).

In total, eleven chips were measured for ferric DtpAa microcrystals, obtaining more than 70,000 diffraction patterns to solve the SFX structure. The measurement of diffraction data with an XFEL pulse of 10 femtoseconds ensured the determination of a radiation damage-free structure of ferric DtpAa.

### *4.7.3.1 Data processing of the SFX data*

The software *CrystFEL* (White et al., 2012) was used to process the data measured at SACLA and obtain the damage-free XFEL structure of DtpAa. A comprehensive general explanation of the processing workflow with CrystFEL is detailed in Chapter 3. Since this is the first SFX structure presented in the text, a detailed description of its data analysis will be delivered. For further structures, the processing will be summarised and just the most relevant information will be reported.

CrystFEL has been specifically developed to treat XFEL data. It consists in a package containing different programs that are run from the command line. It follows the normal crystallography workflow of indexing, scaling and merging, with different options of algorithms and parameters to be used in each step to tune the performance. Indexing of the DtpAa diffraction patterns was carried out with the program *indexamajig*, using optimised parameters that were established while indexing an initial AcNiR dataset, a well characterized system that was used as a benchmark protein to refine detector geometry. The following command was used to run *indexamajig*:

```
$ indexamajig -i files.lst --peaks=zaef -threshold=300 --
min-gradient=100000 --min-snr=5 --int-radius=3,4,5 --
indexing=asdf -g 23368-1.geom -p DtpAa.cell -o
testrun.out -j 20
```

The different parameters introduced for the indexing job are explained in Chapter 3. The DtpAa symmetry and cell parameters were provided in the file ‘DtpAa.cell’ (Figure 4.23), generated using a CrystFEL template. Room-temperature cell parameters measured from DtpAa large crystals measured at beamline I24 (Diamond) during a previous beamtime were used as a start point. Each of the DtpAa measured chips was treated in this way, generating in each case a file ‘stream.out’ with the indexed reflections.

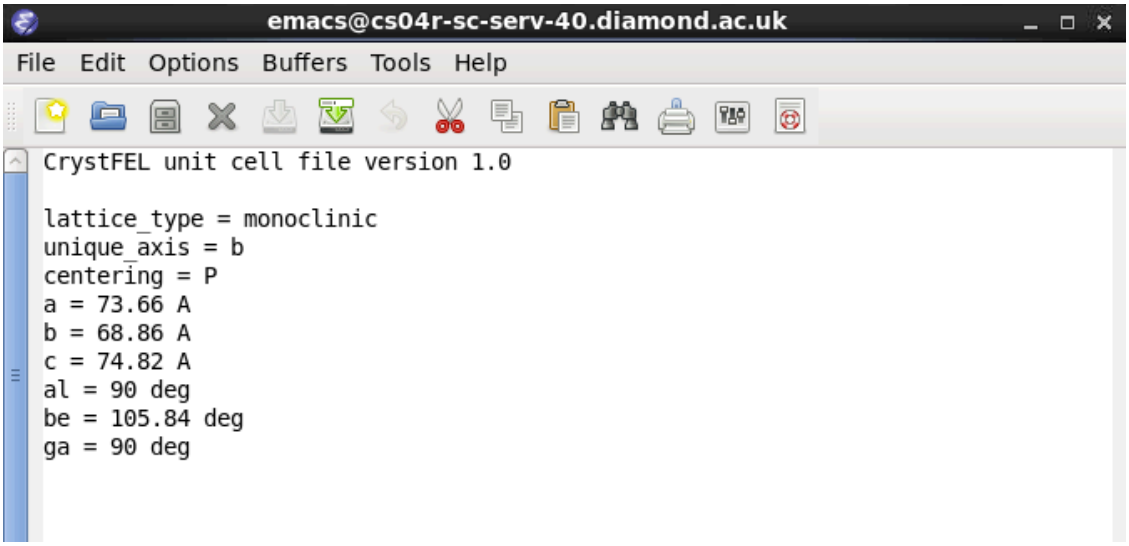


**Figure 4.23:** CrystFEL file 'DtpAa.cell', where cell parameters and symmetry information of DtpAa microcrystals are contained to be inputted to the program *indexamajig*.

An initial run was carried out to refine the cell parameters for the microcrystals. These were optimized by looking at their distribution in each chip using the program *cell_explorer*, introducing the command:

```
$ cell_explorer firstRun_lisbon.out
```

In this example, the distribution of the chip Lisbon (run number 23382) was inspected. Every chip was named with a 6-letter city, to facilitate identification during processing. The outcome from *cell_explorer* is shown in Figure 4.24, were one can see the distribution of cell parameters and how they can be fitted to a Gaussian function to find their mean values and standard deviations.

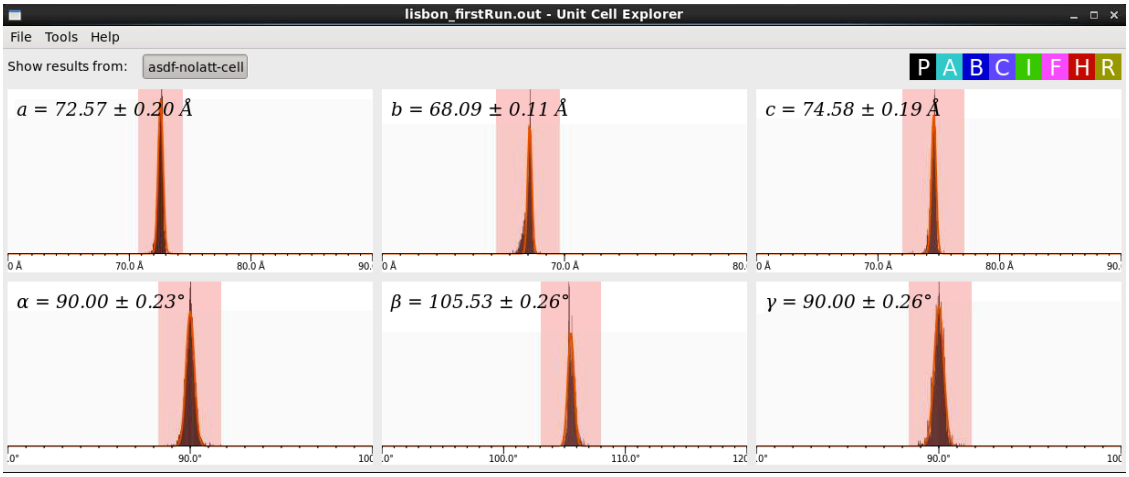


**Figure 4.24:** Distribution plots of cell parameters for the DtpAa chip ´Lisbon', generated by the program *cell_explorer* within *CrystFEL*.

Cell parameters were therefore updated for the experimental distribution of DtpAa microcrystals for each particular chip, and a new file, named 'DtpAa_refined.cell', was generated. A second run was carried out with these improved cell parameters, indexing a higher number of patterns, but not substantially, indicating that the initial given parameters were already close to the real values. At this point, a further run was carried out with *CrystFEL* script *detector_shift* to refine the detector geometry. As inputs, the most recent output with the indexed reflections and the initial detector geometry file were provided. The command used to run *detector_shift* was:

```
$ ./detector-shift run2_lisbon.out detector_lisbon.geom
```

A plot with the distribution of detector shifts for each of the measured crystals was generated by using this script, shown in Figure 4.25. As it can be noted in the figure, just one population of cell dimensions of microcrystals was observed for this sample. *Detector-shift* also generated a refined geometry file,

which in this case was named “run_number-predrefine.geom”. This optimized geometry file would be used for a final run with *indexamajig*, obtaining normally a higher number of indexed patterns. The final stream file, in this case for the Lisbon chip, was named ‘lisbon_refGeom.out’. Data from different chips were merged together using the command:

```
$ cat chip1.out chip2.out chip3.out > chip1-2-3.out
```

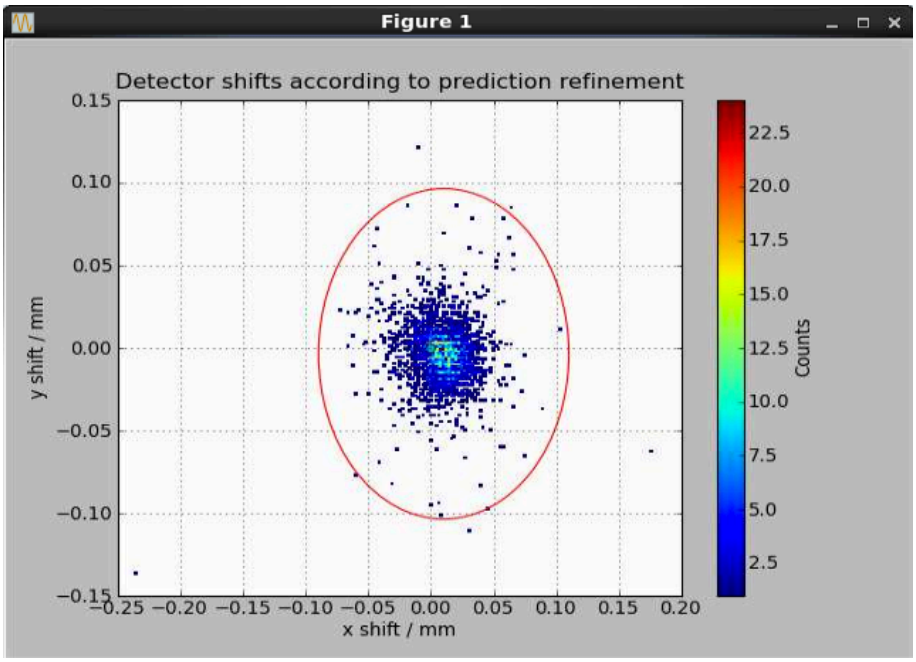


**Figure 4.25:** Plot generated by the program *detector-shift* showing the distribution of this measure for every measured DtpAa microcrystal.

The merging process was carried out with the simplest approach, which is known as Monte Carlo merging (White et al. 2012), using a CrystFEL script called *merge.sh* which in turn would run a program called *process_hkl*. An example of the *merge.sh* script is shown in Figure 4.26. It only needs all the indexed reflections contained in the ‘allchips.out’ file, and the specific pointgroup in which the protein crystallises. In this case for DtpAa, would be 2, since the spacegroup is $P2_1$, but the use of the equivalent pointgroup with a

centrosymmetric element, in this case 2/m_uab (uab specifies the unique axis, which in this case is b), merges the Friedel pairs (just needed for anomalous data), and results in a better merging, as suggested by CrystFEL developers (online tutorial).

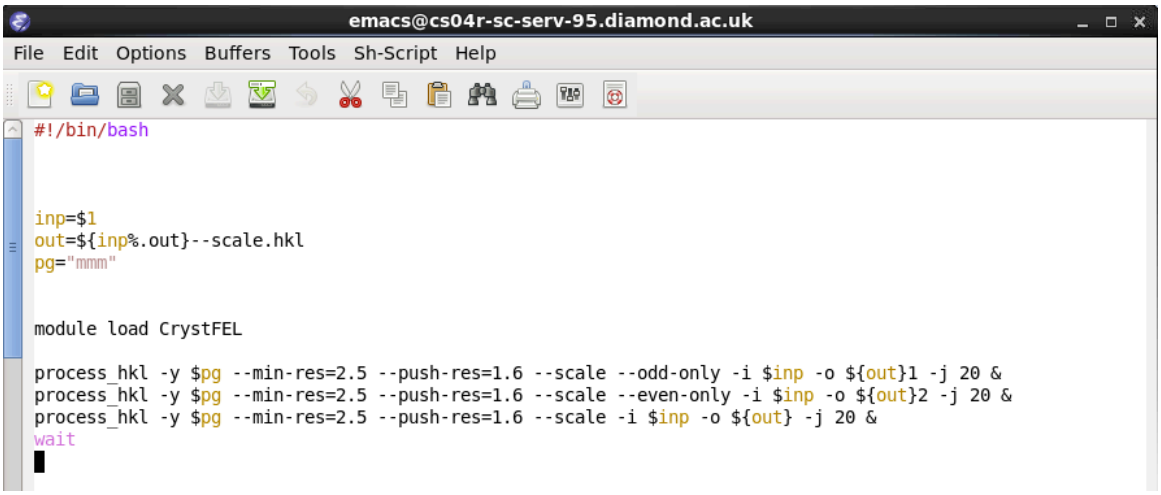


**Figure 4.26:** CrystFEL script merge.sh used for merging DtpAa data, where the input and pointgroup are specified.

Merging using merge.sh, provided three output files, a *hkl* file with all the merged reflections, and files *hkl1* and *hkl2*, where the frames had been split in half for generating data quality statistics afterwards. This was also carried out by using a *CrystFEL* script named *stat.sh*, shown in Figure 4.27. As input, the cell file for DtpAa had to be provided, as well as the different *hkl* files, and the pointgroup specified. Several statistics were generated, being the correlation coefficient (CC) and $R_{split}$ the most relevant to us.

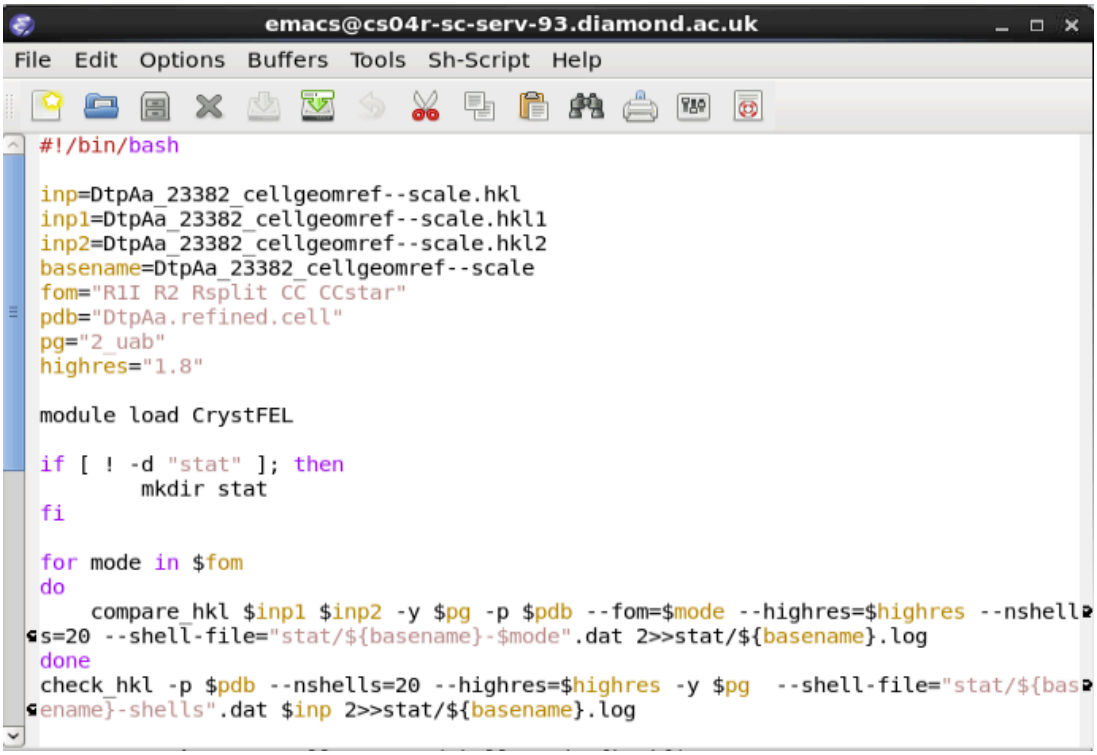


**Figure 4.27:** CrystFEL script stat.sh used for the processing of DtpAa XFEL data. Note that input files are specified, as well as the statistics to be generated and the resolution cut-off.

Each of output files generated by *stat.sh* were *dat* files that could be inspected with the command 'cat'. The file 'shells.dat' (Figure 4.28) contained information about completeness of the data in the different resolution shells.

```
[zlp66871@cs04r-sc-serv-93 stat]$ cat dtpaa-allchips-FINAL--scale-shells.dat
Center 1/nm  # refs Possible  Compl       Meas    Red    SNR    Std dev       Mean     d(A)    Min 1/nm   Max 1/nm
     1.108     3116     3117  99.97    3025695 971.0 16.61    1091.15    1110.61     9.02       0.224      1.992
     2.251     3030     3030 100.00    1922394 634.5 18.09    1446.66    1438.43     4.44       1.992      2.510
     2.691     3042     3042 100.00    1647584 541.6 18.27    1168.97    1123.26     3.72       2.510      2.872
     3.017     3024     3024 100.00    1392056 460.3 16.53     735.91     670.96     3.31       2.872      3.161
     3.283     3015     3015 100.00    1179735 391.3 14.50     416.18     401.78     3.05       3.161      3.405
     3.512     3001     3001 100.00    1153718 384.4 13.71     297.71     287.71     2.85       3.405      3.619
     3.714     2997     2997 100.00    1112464 371.2 12.47     216.73     209.27     2.69       3.619      3.810
     3.896     3015     3015 100.00    1081080 358.6 11.20     161.88     156.63     2.57       3.810      3.983
     4.063     2984     2984 100.00    1031214 345.6  9.66     133.98     123.49     2.46       3.983      4.142
     4.216     3026     3026 100.00    1009861 333.7  8.94     110.98     104.54     2.37       4.142      4.290
     4.360     2960     2960 100.00     950943 321.3  8.07      91.26      86.96     2.29       4.290      4.429
     4.494     2996     2996 100.00     920668 307.3  7.08      74.69      70.90     2.23       4.429      4.559
     4.621     3007     3007 100.00     882640 293.5  6.29      62.15      58.23     2.16       4.559      4.682
     4.741     3011     3011 100.00     843990 280.3  5.48      51.16      47.01     2.11       4.682      4.800
     4.855     2998     2998 100.00     800070 266.9  4.61      38.21      35.49     2.06       4.800      4.911
     4.965     2995     2995 100.00     766633 256.0  3.83      30.98      27.21     2.01       4.911      5.018
     5.069     2967     2967 100.00     727151 245.1  3.23      22.27      20.44     1.97       5.018      5.120
     5.170     2997     2997 100.00     700631 233.8  2.40      17.29      14.00     1.93       5.120      5.219
     5.266     2999     2999 100.00     644024 214.7  1.77      12.02       9.43     1.90       5.219      5.314
     5.360     2982     2982 100.00     559792 187.7  1.22       8.98       6.19     1.87       5.314      5.405
```

**Figure 4.28:** Table contained in the shells.dat file for DtpAa data, where completeness and other statistics are given for the different resolution shells in the dataset.

We looked at the mentioned CC and $R_{split}$ statistics (defined in Chapter 3), to have an idea of the quality of our data and to estimate the resolution at which it should be cut. Examining the values of the CC statistic, shown in Figure 4.29, it can be observed that data quality is good (CC above 0.95 until ~2Å), and that dataset could be cut at 1.85 Å if we take as standard having a CC around 0.5. It is worth noting that due to detector distance and X-ray energy limitations, diffraction data rapidly becomes much worse beyond 1.85 Å resolution.

```
[zlp66871@cs04r-sc-serv-93 stat]$ cat dtpaa-allchips-FINAL--scale-CC.dat
  1/d centre       CC       nref      d / A   Min 1/nm    Max 1/nm
     1.005  0.9867794       2096       9.95      0.268       1.742
     1.967  0.9886549       2036       5.08      1.742       2.193
     2.351  0.9874666       2013       4.25      2.193       2.510
     2.636  0.9904363       2022       3.79      2.510       2.762
     2.869  0.9903825       2034       3.49      2.762       2.975
     3.068  0.9887719       2011       3.26      2.975       3.162
     3.245  0.9858106       1998       3.08      3.162       3.328
     3.404  0.9887945       2009       2.94      3.328       3.480
     3.549  0.9879494       2010       2.82      3.480       3.619
     3.684  0.9864909       1998       2.71      3.619       3.748
     3.809  0.9857632       2011       2.63      3.748       3.869
     3.926  0.9854767       2002       2.55      3.869       3.983
     4.037  0.9840283       2023       2.48      3.983       4.091
     4.142  0.9803238       1992       2.41      4.091       4.193
     4.242  0.9795884       1995       2.36      4.193       4.290
     4.337  0.9747437       1971       2.31      4.290       4.384
     4.428  0.9732755       2042       2.26      4.384       4.473
     4.516  0.9675438       1944       2.21      4.473       4.559
     4.601  0.9616298       2039       2.17      4.559       4.642
     4.682  0.9576744       1971       2.14      4.642       4.722
     4.761  0.9518140       2007       2.10      4.722       4.800
     4.837  0.9385215       2006       2.07      4.800       4.875
     4.911  0.9117695       1967       2.04      4.875       4.947
     4.983  0.8980065       2020       2.01      4.947       5.018
     5.052  0.8827839       1983       1.98      5.018       5.087
     5.120  0.8423071       1990       1.95      5.087       5.154
     5.186  0.8010009       1990       1.93      5.154       5.219
     5.251  0.6606533       1988       1.90      5.219       5.283
     5.314  0.6831171       1969       1.88      5.283       5.345
     5.375  0.4753503       2024       1.86      5.345       5.405
```

**Figure 4.29:** Values of CC data quality statistic in each resolution shell for the resting state DtpAa SFX dataset, to a resolution of 1.85 Å.

The other data quality statistic that we would look for every dataset was $R_{split}$, which is the one used among the serial crystallography community (White et al. 2012). For DtpAa, the values of $R_{split}$ for each resolution shell are shown

in Figure 4.30. As rough guidelines indicated by White and co-workers, an $R_{split}$ below 35% is reasonable, below 20% is good, and below 10% is excellent. In this case, data looks good up till ~2 Å resolution, since is within 20% $R_{split}$, and resolution could be extended until 1.85 Å, matching the CC based cut-off.

```
[zlp66871@cs04r-sc-serv-93 stat]$ cat dtpaa-allchips-FINAL--scale-Rsplit.dat
  1/d centre Rsplit/%       nref      d / A   Min 1/nm    Max 1/nm
       1.005       6.52       2096       9.95      0.268       1.742
       1.967       5.97       2036       5.08      1.742       2.193
       2.351       6.03       2013       4.25      2.193       2.510
       2.636       5.85       2022       3.79      2.510       2.762
       2.869       5.89       2034       3.49      2.762       2.975
       3.068       6.41       2011       3.26      2.975       3.162
       3.245       6.86       1998       3.08      3.162       3.328
       3.404       6.81       2009       2.94      3.328       3.480
       3.549       7.07       2010       2.82      3.480       3.619
       3.684       7.64       1998       2.71      3.619       3.748
       3.809       8.15       2011       2.63      3.748       3.869
       3.926       8.63       2002       2.55      3.869       3.983
       4.037       9.48       2023       2.48      3.983       4.091
       4.142      10.34       1992       2.41      4.091       4.193
       4.242      10.81       1995       2.36      4.193       4.290
       4.337      11.60       1971       2.31      4.290       4.384
       4.428      12.56       2042       2.26      4.384       4.473
       4.516      13.29       1944       2.21      4.473       4.559
       4.601      15.29       2039       2.17      4.559       4.642
       4.682      16.45       1971       2.14      4.642       4.722
       4.761      17.85       2007       2.10      4.722       4.800
       4.837      20.08       2006       2.07      4.800       4.875
       4.911      24.65       1967       2.04      4.875       4.947
       4.983      26.80       2020       2.01      4.947       5.018
       5.052      29.02       1983       1.98      5.018       5.087
       5.120      36.51       1990       1.95      5.087       5.154
       5.186      43.74       1990       1.93      5.154       5.219
       5.251      54.63       1988       1.90      5.219       5.283
       5.314      65.68       1969       1.88      5.283       5.345
       5.375      94.54       2024       1.86      5.345       5.405
```

**Figure 4.30:** Values of $R_{split}$ data quality statistic for the different resolution shells in the SFX DtpAa dataset, until 1.85 Å resolution.

Finally, an *mtz* file was generated using the CrystFEL script *create-mtz*, which is shown in Figure 4.31. The resulting *mtz* file had space group and refined experimental cell parameters specified for the DtpAa microcrystals. From here on, the *mtz* was used as any other dataset from conventional crystallography in the *CCP4i2* software suite (Potterton et al. 2018).

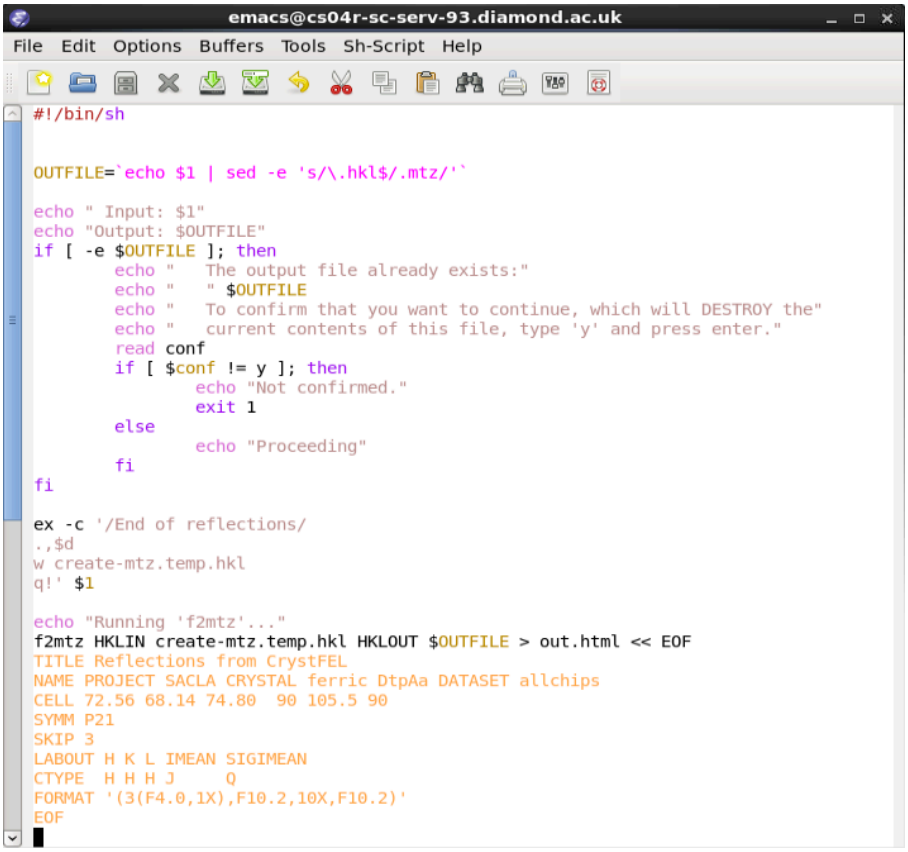


**Figure 4.31:** CrystFEL script create_mtz used in the processing of DtpAa SFX diffraction data to create the final mtz containing all the merged reflections.

### *4.7.3.2 Structure solving and description.*

For the structure solution of DtpAa, merged data were refined against a previously solved at ~1.5 Å resolution composite structure, measured at I24 (Diamond) from DtpAa crystals from different conditions from a same commercial screen plate (data not shown). Hence, this is the first reported room-temperature structure of DtpAa from *S. lividans* coming from a single crystallization condition, and represents more accurately the conformation of the protein in its biological context than a conventional synchrotron structure measured at 100 K. In addition, this structure represents the intact resting state

of DtpAa, with no photo-reduction by the X-ray beam. Final data collection and refinement statistics for the dataset can be found in Table 47.

**Table 4.7:** Data quality and refinement statistics for the SFX structure of resting state DtpAa measured at SACLA (Japan).

| | **Ferric DtpAa** |
|---|---|
| Indexed crystals | 73,281 |
| Average hit rate (%) | 26.5 |
| Space group | $P2_1$ |
| Unit cell (Å) | a=72.57, b=68.14, c=74.80; β=105.50 |
| Resolution (Å) | 37.31 – 1.86 (1.88 – 1.86) |
| Unique reflections | 60162 (3686) |
| $R_{split}$ | 7.30 (94.5) |
| I/σ(I) | 8.8 (1.3) |
| $CC_{1/2}$ | 0.99 (0.48) |
| Completeness (%) | 100 (100) |
| $R_{work}$ (%) | 15.6 |
| $R_{free}$ (%) | 19.2 |
| Rmsd bond lengths (Å) | 0.0129 |
| Rmsd bond angles (º) | 1.864 |
| Ramachandran favoured (%) | 97.3 |

The protein was found adopting a ferredoxin fold, as seen for DtpA previously in the chapter. This characteristic fold for DyPs, with a mix of $\alpha$-helices and $\beta$-sheets secondary structure motifs, can be appreciated in Figure 4.32, where a general overview of the dimer conforming the asymmetric unit is shown. At the active site in monomer A (Figure 4.33), the heme group is found with a proximal His326 coordinating the iron, and at the distal face, residues Asp239 and Arg342 are found, conserved among the DyP family (Sugano 2009). As observed as well in DtpA, the distal Arg342 establishes a salt bridge interaction with one of the haem propionate groups at 2.96 Å, which is a widespread feature within the DyP family. At the axial position at the distal site, a water molecule is found coordinating the octahedral iron at 2.38 Å. This distal

water was also found in the ferric structure of DtpA solved at the ESRF (although with a shorter bond length of 1.98 Å), and indicates that the structure represents the resting state of DtpAa, since there are no signs of photo-reduction. The water is stabilised by the catalytic residue Asp239, which interacts through a hydrogen bond at 2.88 Å. There are many other ordered waters at the distal site, as shown in Figure 4.34, which are forming a proton-exchange network.

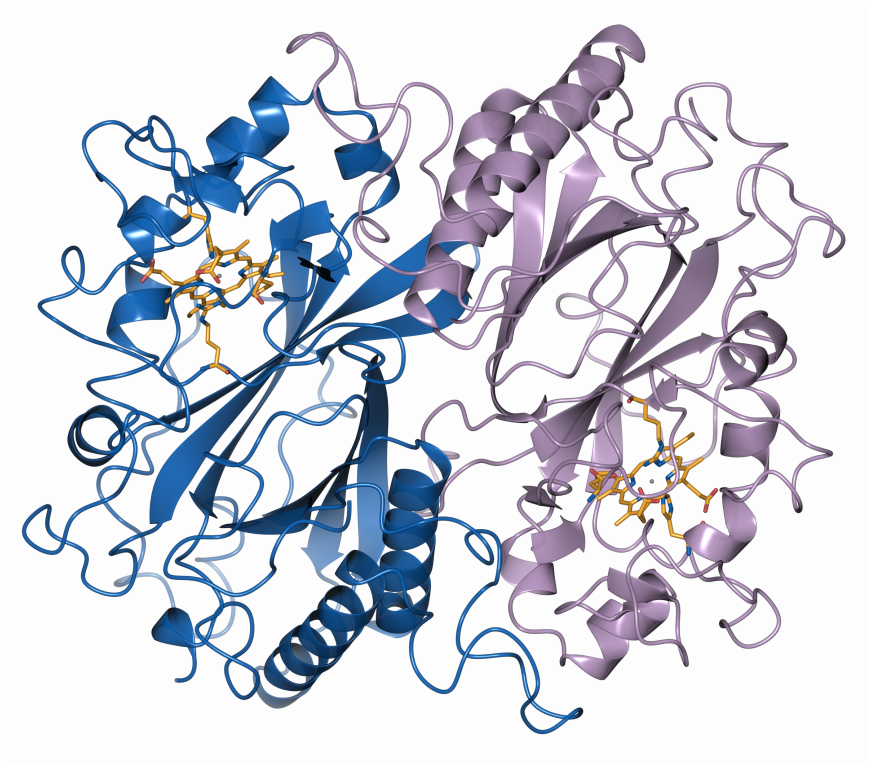

**Figure 4.32:** DtpAa dimer found in the asymmetric unit of the SFX structure measured at SACLA. The two monomers (blue and purple) are represented as ribbons, and haem cofactors as cylinders (yellow).

In monomer B instead, there is a different arrangement at the active site, where an extra density is found at the distal site above the heme, which has been modelled as two molecules of water. Higher above the heme plane, there are more molecules of water modelled, and a fair amount of green extra

electron density is still present. This density might belong to a substance found in the crystallization condition or protein buffer, which has access the pocket through an aperture in the vicinity, although it is not clear. Therefore, the distal water at the axial position could not be observed at heme B, although we assume a Fe(III) state. This asymmetry between the haem sites is also observed in 100 K structure of DtpAa from a single crystal, and hence it is not caused by the crystal size or the measurement at room-temperature.

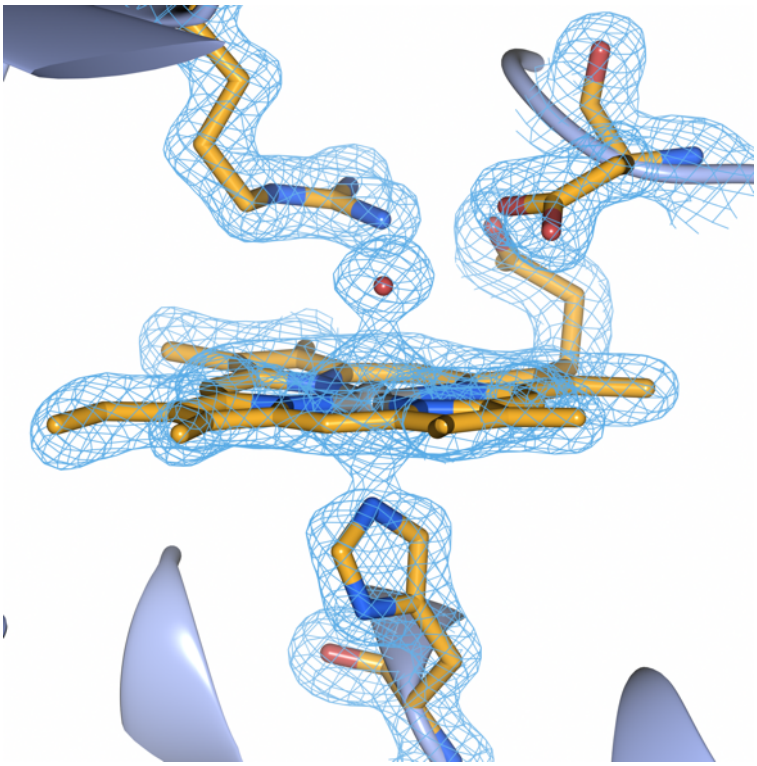

**Figure 4.33:** Representation of the active site at monomer A of the resting state DtpAa SFX structure, measured at SACLA (Japan).

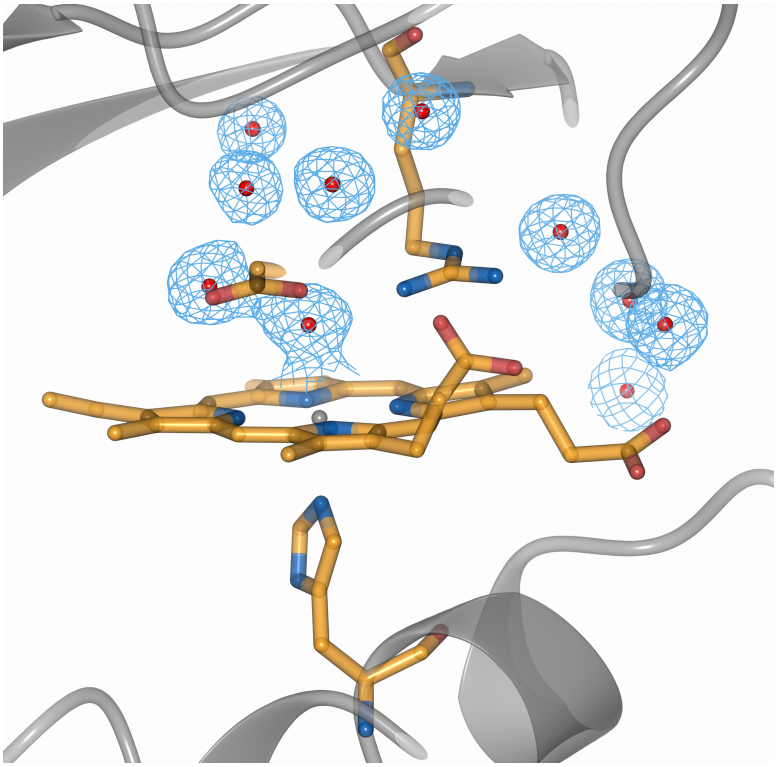

**Figure 4.34:** Chain of ordered waters found at the distal site of heme A in the ferric DtpAa SFX structure, establishing a proton-transfer network.

## *4.7.4 DtpA SFX structure.*

### *4.7.4.1 Data processing*

The case of DtpA resulted quite challenging, since we could not process the data initially. We collected enough data for solving the structure, i.e. more than 20,000 diffraction patterns from microcrystals from a total of five chips (Table 48). However, during data processing we experienced problems with the indexing as well as with the merging steps. When running *indexamajig*, hit rates would be ~10 times less than the ones determined by Cheetah. And when having a look at the cell parameters distribution for the indexed crystals, issues were observed with the determination of cell parameter 'c', and to a minor extent also 'β' (Figure 4.35). The indexing of the data with the algorithm

*TakeTwo* (Ginn et al., 2016), identified a larger number of patterns, however, after the merging process no meaningful solution could be obtained.

**Table 4.8:** Collection parameters and details for chips measured for DtpA at SACLA (October 2017). Hit rates were estimated using the software Cheetah (Barty et al., 2014).

| Chip name | Condition | Blocks measured | Exposure events | Attenuation | Hit rate (%) |
|---|---|---|---|---|---|
| **Himeji** | DtpA ferric | Full chip | 25700 | 0.3 | 23.7 |
| **Jaipur** | DtpA ferric | Full chip | 25700 | 0.3 | 7.7 |
| **Kalyan** | DtpA ferric | Central 6x6 | 14460 | 0.3 | 12.7 |
| **Xalapa** | DtpA ferric | Full chip | 25700 | 0.3 | 39.8 |
| **Yamato** | DtpA ferric | Full chip | 25700 | 0.3 | 34.2 |

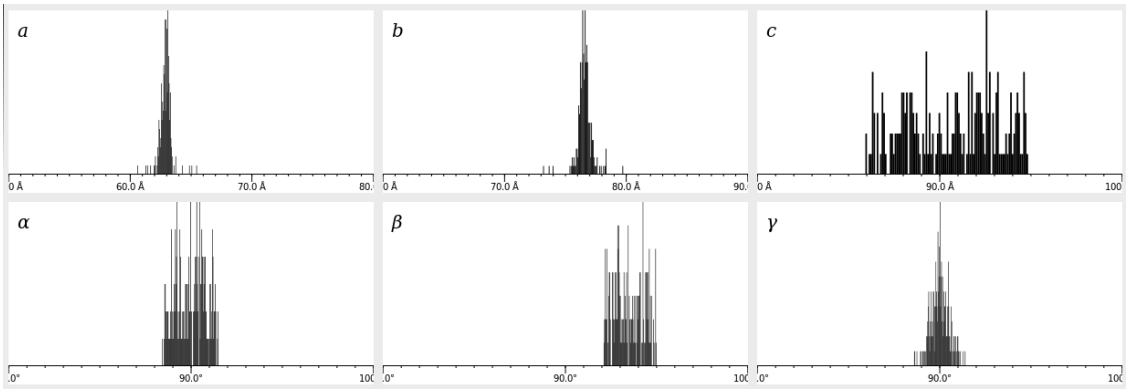


**Figure 4.35:** Cell parameters distribution plots from *cell_explorer* using the initial cell parameters from DtpA structures previously determined. Note the distribution of the cell parameter 'c', which is clearly not well determined.

Eventually, after unsuccessful attempts, the dataset was handed to the developer of CrystFEL, Thomas White, who kindly had a look at it and found that the 'c' axis of the unit cell appear to have double the value compared to the usual cell. This pointed towards the identification of translational non-crystallographic symmetry (tNCS) in this sample, which was affecting the unit

cell in which the protein was crystallising – hence the issues with indexing and merging the data.

Non-crystallographic symmetry (NCS) operations appear when more than one identical molecules are present in the asymmetric unit. It is a very common feature, since most of the structures solved and deposited in the PDB are dimers or even present a higher oligomeric state. The two objects are not completely symmetrical, and can be related be a pseudo-symmetry axis or any other combination of translation and/or rotation of the molecule. In the case of the DtpA SFX dataset, the translation of the DtpA motif along the axis 'c' was not exact, and this resulted in this axis being doubled in length since both monomers were not equivalent anymore. After the modification of the c axis, the indexing step worked well as it can be appreciated in the distribution of the cell parameters shown in Figure 4.36.


**Figure 4.36:** Distribution plots from *cell_explorer* for the SFX DtpA dataset, where the value of axis 'c' has been doubled compared to the previously determined cell. Note the mean values and standard deviations provided for each of the cell parameters.

Thus, data processing could be resumed with the merging process, which was carried out with the program *process_hkl*, within the script *merge.sh*,

as shown for DtpAa. In the same way, the data quality statistics were generated running the script *stat.sh*. The generated tables with resolution shells and the data quality parameters CC and $R_{split}$ for the DtpA SFX dataset are shown in Figure 4.37, Figure 4.38 and Figure 4.39, respectively. The resolution cut-off for the dataset was applied at 2.10 Å with the same criteria of CC ~0.5 in the last resolution shell. An *mtz* file with all the merged reflections was created by running the script *create-mtz*.

```
[zlp66871@cs04r-sc-serv-41 stat]$ cat dtpa_new_test--scale-shells.dat
```

| Center 1/nm | # refs | Possible | Compl | Meas | Red | SNR | Std dev | Mean | d(A) | Min 1/nm | Max 1/nm |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.994 | 4440 | 4440 | 100.00 | 2156116 | 485.6 | 8.92 | 702.50 | 419.94 | 10.06 | 0.232 | 1.756 |
| 1.983 | 4336 | 4336 | 100.00 | 1385803 | 319.6 | 9.37 | 599.38 | 410.17 | 5.04 | 1.756 | 2.211 |
| 2.371 | 4313 | 4313 | 100.00 | 1174754 | 272.4 | 9.49 | 509.36 | 359.98 | 4.22 | 2.211 | 2.531 |
| 2.658 | 4335 | 4335 | 100.00 | 1062284 | 245.0 | 8.89 | 390.66 | 255.78 | 3.76 | 2.531 | 2.785 |
| 2.893 | 4292 | 4292 | 100.00 | 963824 | 224.6 | 8.09 | 265.86 | 176.36 | 3.46 | 2.785 | 3.000 |
| 3.094 | 4295 | 4295 | 100.00 | 840909 | 195.8 | 6.69 | 158.79 | 108.48 | 3.23 | 3.000 | 3.188 |
| 3.272 | 4294 | 4294 | 100.00 | 752384 | 175.2 | 5.57 | 112.59 | 73.79 | 3.06 | 3.188 | 3.356 |
| 3.432 | 4276 | 4276 | 100.00 | 743011 | 173.8 | 4.88 | 81.24 | 51.49 | 2.91 | 3.356 | 3.509 |
| 3.579 | 4265 | 4265 | 100.00 | 734880 | 172.3 | 4.19 | 55.57 | 37.87 | 2.79 | 3.509 | 3.649 |
| 3.714 | 4304 | 4304 | 100.00 | 719915 | 167.3 | 3.47 | 40.57 | 26.98 | 2.69 | 3.649 | 3.780 |
| 3.841 | 4307 | 4307 | 100.00 | 705129 | 163.7 | 2.92 | 34.72 | 21.59 | 2.60 | 3.780 | 3.902 |
| 3.959 | 4263 | 4263 | 100.00 | 680799 | 159.7 | 2.34 | 25.86 | 15.47 | 2.53 | 3.902 | 4.016 |
| 4.071 | 4258 | 4258 | 100.00 | 664921 | 156.2 | 2.11 | 19.74 | 12.97 | 2.46 | 4.016 | 4.125 |
| 4.177 | 4287 | 4287 | 100.00 | 650792 | 151.8 | 1.93 | 19.48 | 11.77 | 2.39 | 4.125 | 4.228 |
| 4.277 | 4252 | 4252 | 100.00 | 624853 | 147.0 | 1.65 | 14.68 | 9.35 | 2.34 | 4.228 | 4.327 |
| 4.374 | 4260 | 4260 | 100.00 | 600532 | 141.0 | 1.36 | 13.32 | 7.37 | 2.29 | 4.327 | 4.421 |
| 4.466 | 4317 | 4317 | 100.00 | 583081 | 135.1 | 1.26 | 10.89 | 6.56 | 2.24 | 4.421 | 4.511 |
| 4.554 | 4231 | 4231 | 100.00 | 543750 | 128.5 | 1.10 | 9.49 | 5.39 | 2.20 | 4.511 | 4.598 |
| 4.639 | 4287 | 4287 | 100.00 | 517486 | 120.7 | 0.94 | 9.40 | 4.57 | 2.16 | 4.598 | 4.681 |
| 4.722 | 4246 | 4246 | 100.00 | 479590 | 113.0 | 0.84 | 7.54 | 3.86 | 2.12 | 4.681 | 4.762 |

**Figure 4.37:** Resolution shells for the DtpA SFX dataset to a resolution of 2.10 Å, providing completeness and other data quality statistics in each case.

```
[zlp66871@cs04r-sc-serv-41 stat]$ cat dtpa_new_test--scale-CC.dat
  1/d centre       CC       nref      d / A   Min 1/nm    Max 1/nm
     0.994  0.9790348       4440      10.06      0.233       1.756
     1.983  0.9684555       4336       5.04      1.756       2.211
     2.371  0.9788943       4313       4.22      2.211       2.531
     2.658  0.9788096       4335       3.76      2.531       2.785
     2.893  0.9802799       4292       3.46      2.785       3.000
     3.094  0.9760384       4296       3.23      3.000       3.188
     3.272  0.9711352       4293       3.06      3.188       3.356
     3.432  0.9642012       4276       2.91      3.356       3.509
     3.579  0.9527067       4265       2.79      3.509       3.649
     3.714  0.9396367       4304       2.69      3.649       3.780
     3.841  0.9316423       4307       2.60      3.780       3.902
     3.959  0.8981198       4263       2.53      3.902       4.016
     4.071  0.8434953       4258       2.46      4.016       4.125
     4.177  0.8498341       4287       2.39      4.125       4.228
     4.277  0.7589848       4252       2.34      4.228       4.327
     4.374  0.7340161       4260       2.29      4.327       4.421
     4.466  0.6443945       4317       2.24      4.421       4.511
     4.554  0.5943237       4231       2.20      4.511       4.598
     4.639  0.6000691       4287       2.16      4.598       4.681
     4.722  0.4505190       4246       2.12      4.681       4.762
```

**Figure 4.38:** Values for the CC data quality statistic for each of the resolution shells in the DtpA SFX dataset, with a resolution cut-off applied to 2.10 Å.

```
[zlp66871@cs04r-sc-serv-41 stat]$ cat dtpa_new_test--scale-Rsplit.dat
  1/d centre Rsplit/%       nref      d / A   Min 1/nm    Max 1/nm
     0.994      10.23       4440      10.06      0.233       1.756
     1.983      10.59       4336       5.04      1.756       2.211
     2.371      10.31       4313       4.22      2.211       2.531
     2.658      10.56       4335       3.76      2.531       2.785
     2.893      11.05       4292       3.46      2.785       3.000
     3.094      13.39       4296       3.23      3.000       3.188
     3.272      15.63       4293       3.06      3.188       3.356
     3.432      18.62       4276       2.91      3.356       3.509
     3.579      21.49       4265       2.79      3.509       3.649
     3.714      27.21       4304       2.69      3.649       3.780
     3.841      31.34       4307       2.60      3.780       3.902
     3.959      40.98       4263       2.53      3.902       4.016
     4.071      47.94       4258       2.46      4.016       4.125
     4.177      51.70       4287       2.39      4.125       4.228
     4.277      63.49       4252       2.34      4.228       4.327
     4.374      78.73       4260       2.29      4.327       4.421
     4.466      85.25       4317       2.24      4.421       4.511
     4.554      98.38       4231       2.20      4.511       4.598
     4.639     113.45       4287       2.16      4.598       4.681
     4.722     132.68       4246       2.12      4.681       4.762
```

**Figure 4.39:** Values of $R_{split}$ data quality statistic for each of the resolution shells in the DtpA SFX dataset, with a resolution cut-off applied to 2.10 Å.

### *4.7.4.2 Structure solving and description.*

A Molecular replacement job was performed with *Molrep* (Vagin and Teplyakov, 1997) using the oxy-ferrous DtpA structure as a search model (PDB accession code: 5map). Two dimers were found in the asymmetric unit (Figure

4.40), as a result from the double 'c' axis length due to tNCS. After a first round of refinement with *Refmac5* (Murshudov et al., 2011), $R_{work}$ and $R_{free}$ went down to acceptable values of 22.2% and 27.7%, respectively. Model building after rounds of refinement was carried out in *Coot* (Emsley et al., 2010). Data collection and refinement statistics for this structure are shown in Table 78. It is worth to highlight that this is the first RT structure reported for DtpA, and therefore it is a more representative conformation of the biological conditions where it functions.

**Table 4.9:** Data quality and refinement statistics for the SFX structure of resting state DtpA, measured from microcrystals at SACLA (Japan), using 10 fs XFEL pulses of 1.24 Å wavelength.

| | **SFX ferric DtpA** |
|---|---|
| Indexed crystals | 23,774 |
| Average hit rate (%) | 20.27 |
| Space group | $P2_1$ |
| Unit cell (Å) | a=62.88, b=76.44, c=155.64; β=97.10 |
| Resolution (Å) | 42.86 – 2.10 (2.14 – 2.10) |
| Unique reflections | 85,666 (4,508) |
| $R_{split}$ | 13.8 (140.8) |
| I/σ(I) | 4.3 (0.8) |
| $CC_{1/2}$ | 0.98 (0.445) |
| Completeness (%) | 100 (100) |
| $R_{work}$ (%) | 20.8 |
| $R_{free}$ (%) | 26.9 |
| Rmsd bond lengths (Å) | 0.0139 |
| Rmsd bond angles (º) | 1.712 |
| Ramachandran favoured (%) | 95.12 |

The haem sites of the different monomers (four in total), present a conformation like the one shown in Figure 4.41 (haem B), with the proximal His353 coordinating the iron, and the conserved residues Asp251 and Arg369 at the distal site, as expected. A water molecule is found in the pocket stabilised by Asp251 through a hydrogen bond (2.96 Å), but too far from the iron (3.55 Å)

to coordinate it and form as hexacoordinate species like the one found in the synchrotron ferric DtpA structure shown previously in the chapter. In one out of the four monomers (heme A), there is no presence of any water at the distal site, which is empty. A comparison has been drawn (Figure 4.41, right panel) between the heme environments of the present structure and the ferric structure measured at 100 K with a helical collection at ESRF. There are just slight conformational differences, for example of the catalytic Asp251, and the water that is coordinating, present in both structures. The major difference between the structures is the bound distal water hexacoordinating the heme iron at 1.98 Å, which is present in the ferric ESRF but not in the SFX SACLA structure. However, as it can be appreciated in Figure 4.41 (left panel), there is a bit of positive density (in green) at the distal position, which suggests the low partial occupancy of a water molecule. It is important to note that both structures come from crystals grown in different crystallization conditions. This is because we were unable to produce microcrystals from the original condition used to obtain large DtpA crystals. The differences for instance in pH (pH 5.5 for large crystals, and pH 7.5 for microcrystals), might explain the differences observed in heme coordination.

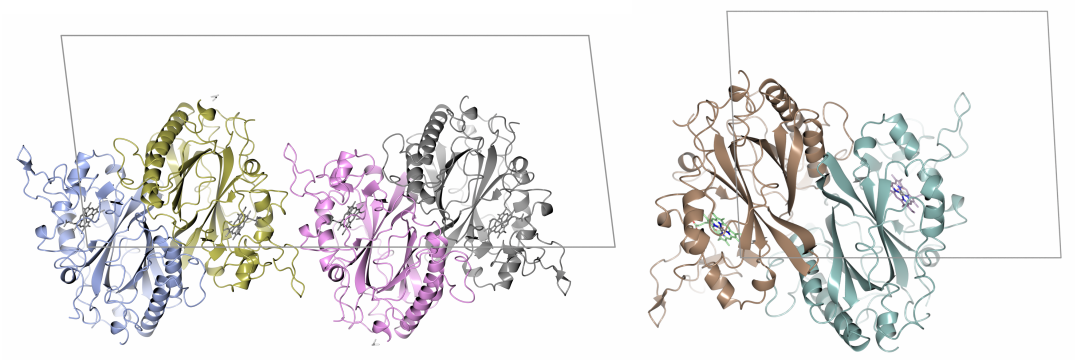

**Figure 4.40:** Asymmetric unit (AUs) of the ferric DtpA SFX structure (on the left) constituted by two dimers, in contrast with the AU found in conventional DtpA structures (on the right), showing just a single dimer. In both cases, the view of the AU is aligned on the 'ac' face of the unit cell.

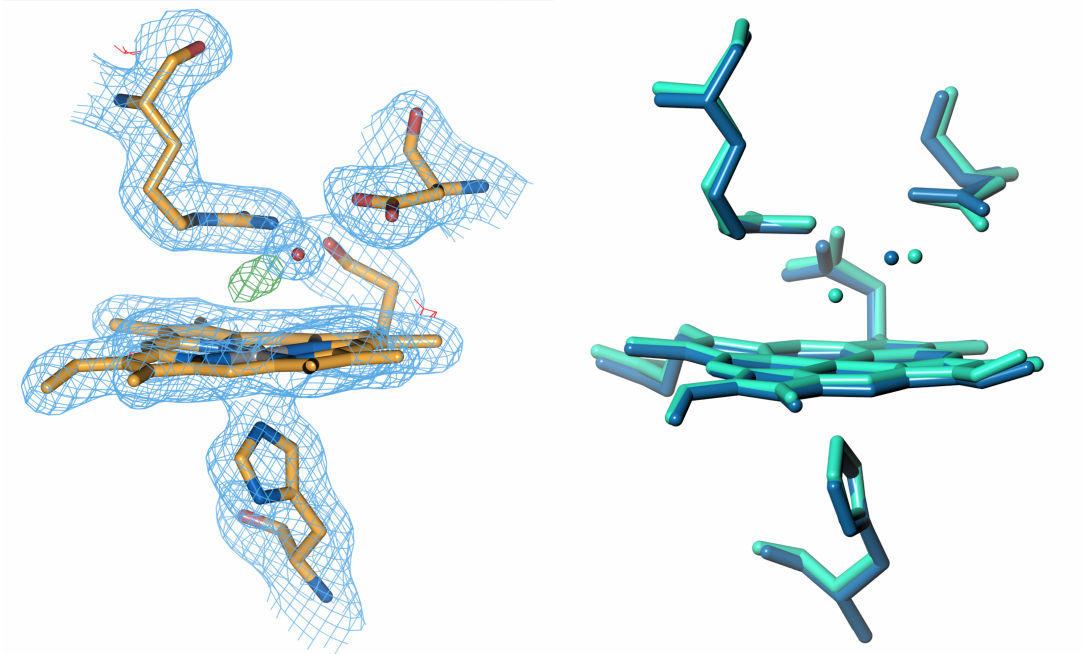

**Figure 4.41:** Heme site of the resting state DtpA SFX room temperature structure measured at SACLA (Japan). Heme environment from monomer B is represented on the left panel with the electron density of the $2F_o - F_c$ map shown in blue and contoured at $1\sigma$. The positive density of the difference $F_o - F_c$ map is shown in green and contoured to $3\sigma$. On the right panel a superposition is shown between the DtpA RT structure from SACLA (in blue), and the synchrotron ferric DtpA cryo-structure measured at ESRF with a helical collection (in turquoise).

### *4.7.5 Imidazole bound DtpAa SFX structure.*

DtpAa microcrystals were produced in several batches with the crystallization condition containing imidazole (MIB condition detailed previously). Batches were loaded onto fixed targets and measured with the usual screening routine with 10 fs XFEL pulses at the SACLA hutch. A total of 4 chips were measured, which are detailed in Table 410.

**Table 4.10:** Collection parameters and details for chips measured with DtpAa – imidazole microcrystals at SACLA (October 2017). Hit rates were estimated using the software Cheetah (Barty et al., 2014).

| Chip name | Condition | Blocks measured | Exposure events | Attenuation | Hit rate (%) |
|---|---|---|---|---|---|
| **Europe** | DtpAa - imidazole | Full chip | 25700 | 0.3 | 12.3 |
| **Fushun** | DtpAa - imidazole | Full chip | 25700 | 0.3 | 25.1 |
| **Getafe** | DtpAa - imidazole | Full chip | 25700 | 0.3 | 27.5 |
| **Huesca** | DtpAa - imidazole | Full chip | 25700 | 0.3 | 48.3 |

#### *4.7.5.1 Data processing*

The standard workflow with *CrystFEL* (as detailed in Chapter 3 and earlier in this chapter) was carried out starting with an initial indexing job with *indexamajig*, using as input cell parameters the ones for ferric DtpAa RT crystals. Unit cell parameters were optimized, using the program *cell_explorer*, and the detector geometry refined using the script *detector-shift*, as previously

described. The final distribution of cell parameters from data measured from all four chips is shown in Figure 4.42. The plot generated by *detector-shift*, showing a single population of microcrystals is shown in Figure 4.43.

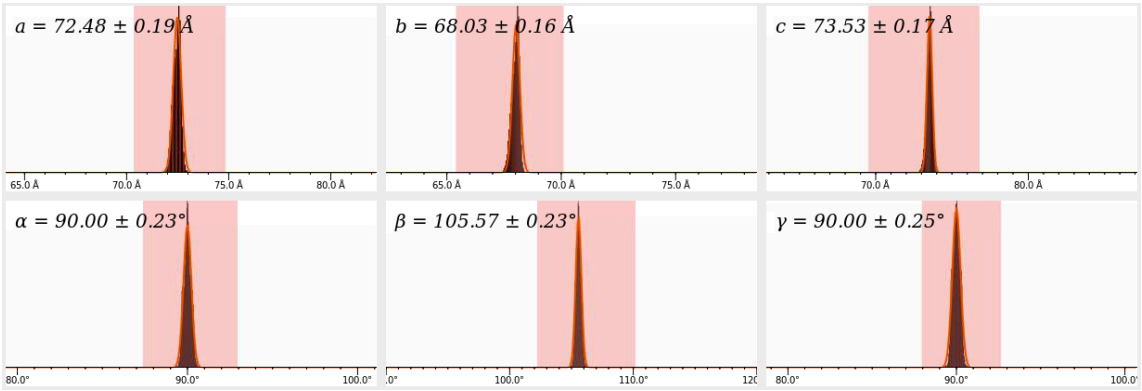


**Figure 4.42:** Cell parameter distribution plots for DtpAa – imidazole microcrystals (data from all four chips), generated by the *CrystFEL* program *cell_explorer*.

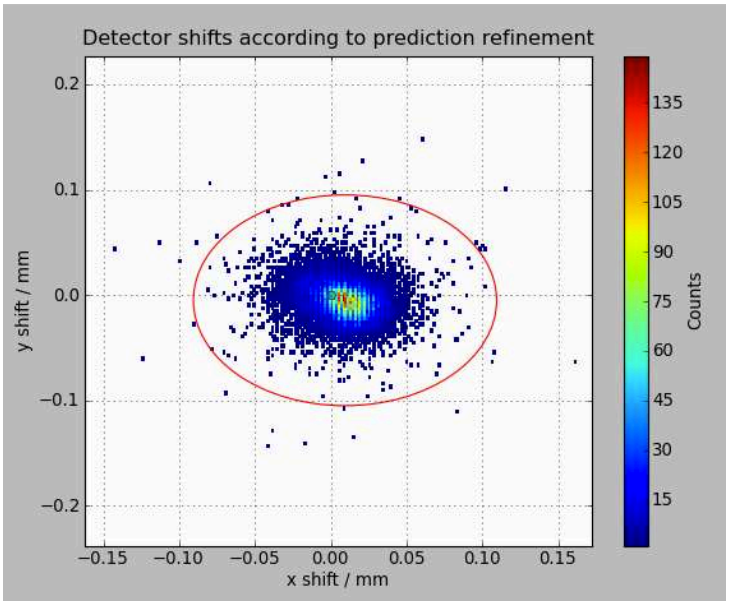


**Figure 4.43:** Plot generated by the CrystFEL script detector-shift, where 'x' and 'y' shifts of the detector are specified for each of the crystals measured. A single population of DtpAa – imidazole microcrystals can be appreciated.

The final stream file from *indexamajig* with all the indexed reflections was merged using the program *process_hkl*, by running the script *merge.sh*,

specifying pointgroup '2/m_uab' for this sample. Data quality statistics were generated with the script *stat.sh*, to a resolution of 1.85 Å. Resolution shells information for this dataset and the different values of CC and $R_{split}$ in each of them are specified in Figure 4.44, Figure 4.45 and Figure 4.46, respectively. Values of the correlation coefficient (CC) are acceptable until a resolution of 1.87 Å, following the standard cut-off criteria of CC~0.5 in the last resolution shell. In the case of $R_{split}$, values indicate good data quality (below 20%) until ~2.2 Å, and resolution can be extended to 1.85 Å, which was the selected final resolution cut-off for the dataset. Finally, an *mtz* file was generated as always with the script *create_mtz*.

```
[zlp66871@cs04r-sc-serv-95 stat]$ cat dtpaa-imid-scale-shells.dat
```

| Center 1/nm | # refs | Possible | Compl | Meas | Red | SNR | Std dev | Mean | d(A) | Min 1/nm | Max 1/nm |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.131 | 3056 | 3056 | 100.00 | 685045 | 224.2 | 6.12 | 995.21 | 1092.33 | 8.84 | 0.269 | 1.993 |
| 2.251 | 2975 | 2975 | 100.00 | 434562 | 146.1 | 6.94 | 1605.67 | 1655.75 | 4.44 | 1.993 | 2.510 |
| 2.691 | 2985 | 2985 | 100.00 | 371442 | 124.4 | 7.32 | 1450.47 | 1437.49 | 3.72 | 2.510 | 2.873 |
| 3.017 | 2955 | 2955 | 100.00 | 314354 | 106.4 | 7.18 | 986.29 | 929.02 | 3.31 | 2.873 | 3.162 |
| 3.284 | 2945 | 2945 | 100.00 | 269257 | 91.4 | 6.88 | 655.41 | 622.78 | 3.05 | 3.162 | 3.406 |
| 3.512 | 2954 | 2954 | 100.00 | 264151 | 89.4 | 6.88 | 463.94 | 450.39 | 2.85 | 3.406 | 3.619 |
| 3.714 | 2946 | 2946 | 100.00 | 255638 | 86.8 | 6.70 | 355.67 | 344.91 | 2.69 | 3.619 | 3.810 |
| 3.896 | 2966 | 2966 | 100.00 | 247965 | 83.6 | 6.46 | 280.18 | 273.32 | 2.57 | 3.810 | 3.983 |
| 4.063 | 2937 | 2937 | 100.00 | 237012 | 80.7 | 6.09 | 251.17 | 229.46 | 2.46 | 3.983 | 4.142 |
| 4.216 | 2920 | 2920 | 100.00 | 229823 | 78.7 | 5.80 | 207.21 | 196.50 | 2.37 | 4.142 | 4.290 |
| 4.360 | 2940 | 2940 | 100.00 | 224841 | 76.5 | 5.43 | 174.95 | 169.68 | 2.29 | 4.290 | 4.429 |
| 4.494 | 2936 | 2936 | 100.00 | 218977 | 74.6 | 4.96 | 151.58 | 142.74 | 2.23 | 4.429 | 4.559 |
| 4.621 | 2941 | 2941 | 100.00 | 213151 | 72.5 | 4.51 | 130.53 | 123.07 | 2.16 | 4.559 | 4.682 |
| 4.741 | 2931 | 2931 | 100.00 | 208789 | 71.2 | 4.04 | 107.68 | 100.05 | 2.11 | 4.682 | 4.800 |
| 4.855 | 2955 | 2955 | 100.00 | 205523 | 69.6 | 3.59 | 82.36 | 76.63 | 2.06 | 4.800 | 4.911 |
| 4.965 | 2907 | 2907 | 100.00 | 197789 | 68.0 | 3.14 | 67.09 | 60.05 | 2.01 | 4.911 | 5.018 |
| 5.069 | 2915 | 2915 | 100.00 | 195178 | 67.0 | 2.61 | 52.39 | 45.42 | 1.97 | 5.018 | 5.120 |
| 5.170 | 2967 | 2967 | 100.00 | 193780 | 65.3 | 2.18 | 39.64 | 33.43 | 1.93 | 5.120 | 5.219 |
| 5.266 | 2918 | 2918 | 100.00 | 181997 | 62.4 | 1.62 | 27.90 | 21.82 | 1.90 | 5.219 | 5.314 |
| 5.360 | 2919 | 2919 | 100.00 | 163780 | 56.1 | 1.08 | 20.67 | 13.87 | 1.87 | 5.314 | 5.405 |

**Figure 4.44:** Resolution shells for the SFX dataset DtpAa – imidazole, providing completeness and other data quality statistics in each case.

```
[zlp66871@cs04r-sc-serv-95 stat]$ cat dtpaa-imid-scale-CC.dat
  1/d centre       CC       nref      d / A   Min 1/nm    Max 1/nm
     1.131  0.8878136       3056       8.84      0.270       1.993
     2.251  0.9238036       2975       4.44      1.993       2.510
     2.691  0.9245559       2986       3.72      2.510       2.873
     3.017  0.9266961       2955       3.31      2.873       3.162
     3.284  0.9330881       2944       3.05      3.162       3.406
     3.512  0.9384775       2954       2.85      3.406       3.619
     3.714  0.9349027       2946       2.69      3.619       3.810
     3.896  0.9314487       2966       2.57      3.810       3.983
     4.063  0.9321189       2937       2.46      3.983       4.142
     4.216  0.9263675       2920       2.37      4.142       4.290
     4.360  0.9291410       2940       2.29      4.290       4.429
     4.494  0.9106784       2936       2.23      4.429       4.559
     4.621  0.8942204       2941       2.16      4.559       4.682
     4.741  0.8778841       2931       2.11      4.682       4.800
     4.855  0.8385154       2955       2.06      4.800       4.911
     4.965  0.7485427       2907       2.01      4.911       5.018
     5.069  0.7696611       2915       1.97      5.018       5.120
     5.170  0.7026605       2967       1.93      5.120       5.219
     5.266  0.6153660       2918       1.90      5.219       5.314
     5.360  0.4606615       2919       1.87      5.314       5.405
```

**Figure 4.45:** Values of CC data quality statistic for each of the resolution shell of the SFX dataset DtpAa – imidazole, to a resolution limit of 1.85 Å.

```
[zlp66871@cs04r-sc-serv-95 stat]$ cat dtpaa-imid-scale-Rsplit.dat
  1/d centre Rsplit/%       nref      d / A   Min 1/nm    Max 1/nm
     1.131      18.79       3056       8.84      0.270       1.993
     2.251      16.31       2975       4.44      1.993       2.510
     2.691      16.00       2986       3.72      2.510       2.873
     3.017      15.84       2955       3.31      2.873       3.162
     3.284      15.73       2944       3.05      3.162       3.406
     3.512      15.36       2954       2.85      3.406       3.619
     3.714      15.43       2946       2.69      3.619       3.810
     3.896      15.79       2966       2.57      3.810       3.983
     4.063      16.60       2937       2.46      3.983       4.142
     4.216      17.65       2920       2.37      4.142       4.290
     4.360      17.68       2940       2.29      4.290       4.429
     4.494      19.77       2936       2.23      4.429       4.559
     4.621      22.20       2941       2.16      4.559       4.682
     4.741      24.48       2931       2.11      4.682       4.800
     4.855      27.96       2955       2.06      4.800       4.911
     4.965      34.00       2907       2.01      4.911       5.018
     5.069      37.00       2915       1.97      5.018       5.120
     5.170      47.50       2967       1.93      5.120       5.219
     5.266      60.32       2918       1.90      5.219       5.314
     5.360      89.53       2919       1.87      5.314       5.405
```

**Figure 4.46:** Values of $R_{split}$ data quality statistic for each of the resolution shell of the SFX dataset DtpAa – imidazole, to a resolution limit of 1.85 Å.

### *4.7.5.2 Structure solution*

Data for the DtpAa – imidazole complex were refined with *Refmac5* (Murshudov et al., 2011) against the solved SFX structure of ferric DtpAa.

Model building for example of the imidazole molecule and crystallographic waters was carried out with *Coot* (Emsley et al., 2010), and refinement cycles with *Refmac5*. Data quality and refinement statistics for the room temperature SFX structure are detailed in Table 411.

**Table 4.11:** Data collection and refinement statistics for the SFX DtpAa – imidazole dataset, measured at room temperature at SACLA (Japan), with XFEL pulses of 10 fs at 1.24 Å wavelength.

| | **SFX DtpAa – imidazole complex** |
|---|---|
| Indexed crystals | 20,321 |
| Average hit rate (%) | 19.8 |
| Space group | $P2_1$ |
| Unit cell (Å) | a=72.50, b=68.04, c=73.54; β=105.58 |
| Resolution (Å) | 35.69 – 1.85 (1.89 – 1.85) |
| Unique reflections | 58,972 (3,637) |
| $R_{split}$ | 17.44 (89.53) |
| I/σ(I) | 4.6 (1.2) |
| $CC_{1/2}$ | 0.95 (0.46) |
| Completeness (%) | 100 (100) |
| $R_{work}$ (%) | 18.8 |
| $R_{free}$ (%) | 22.7 |
| Rmsd bond lengths (Å) | 0.0149 |
| Rmsd bond angles (º) | 1.916 |
| Ramachandran favoured (%) | 97.34 |

DtpAa in the presence of imidazole crystallized in the space group $P2_1$, and was found as expected forming a dimer in the asymmetric unit. In the heme site of one of the monomers (heme B), a molecule of imidazole was found and modelled with full occupancy (Figure 4.47). Since the crystals were grown in an imidazole-containing solution, DtpAa was effectively co-crystallized with the ligand, forming a hexacoordinate heme complex. Imidazole coordinates the heme iron at the axial position through its N3 at 2.19 Å distance, and is stabilised by the catalytic Asp239 through a hydrogen bond interaction with N1

at 2.76 Å. Regarding the other monomer, in heme A we can observe a distal water coordinating the iron at 2.39 Å, as in the ferric DtpA species. However, in this case some additional electron density is present at the distal site (Figure 4.47) in comparison with the resting state (Figure 4.33), suggesting a low partial binding of imidazole at heme A, even if this site represents a predominant ferric species. The presence of another molecule of imidazole it was noted in the SFX structure, located at the interface between the two monomers (Figure 4.48).

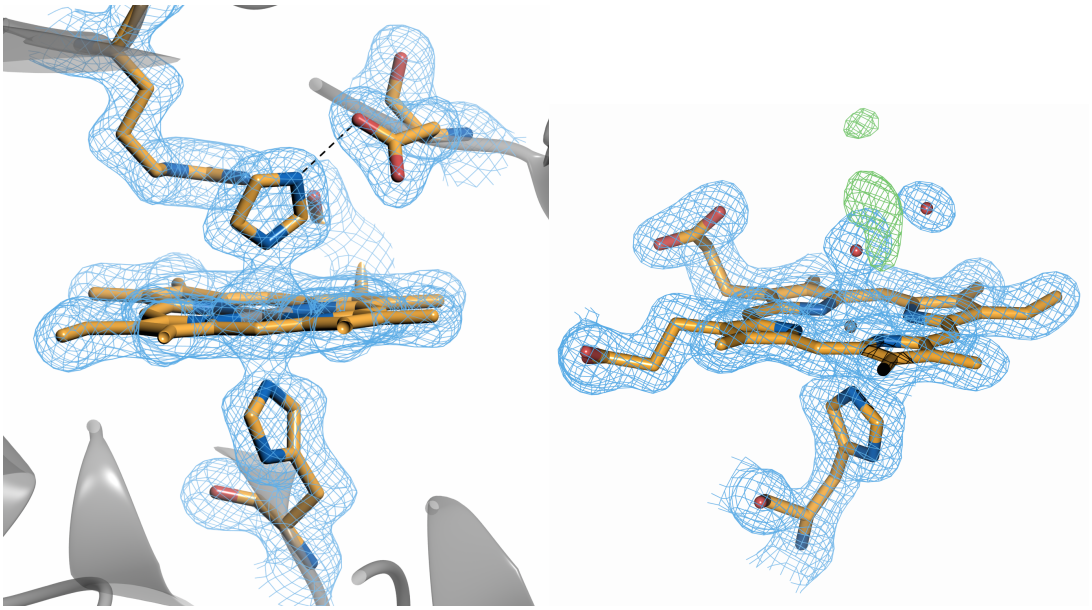

**Figure 4.47:** Active sites from the SFX DtpAa – imidazole complex structure. On the left, the heme site at monomer B is shown, where the aromatic ligand can be observed coordinating the heme iron. On the right, heme A is shown, where a distal water is present coordinating the iron. Electron density from the $2F_o - F_c$ map is shown as a blue mesh, and is contoured at 1 σ. The positive difference electron density map $F_o - F_c$ is shown as a green mesh contoured at 3σ. Note the additional density at heme A, suggesting a slight occupancy of imidazole at this site.

Examining the superposition between the DtpAa ferric and the imidazole-bound structures, just minimal conformational changes could be

observed, such as the movement of the distal residues Asp239 and Phe363 in order to make space to accommodate the aromatic ligand. The RMSD between the two monomer B (where the ligand binds) motifs is 0.11 $Å^2$, and between the two dimers is 0.16 $Å^2$.

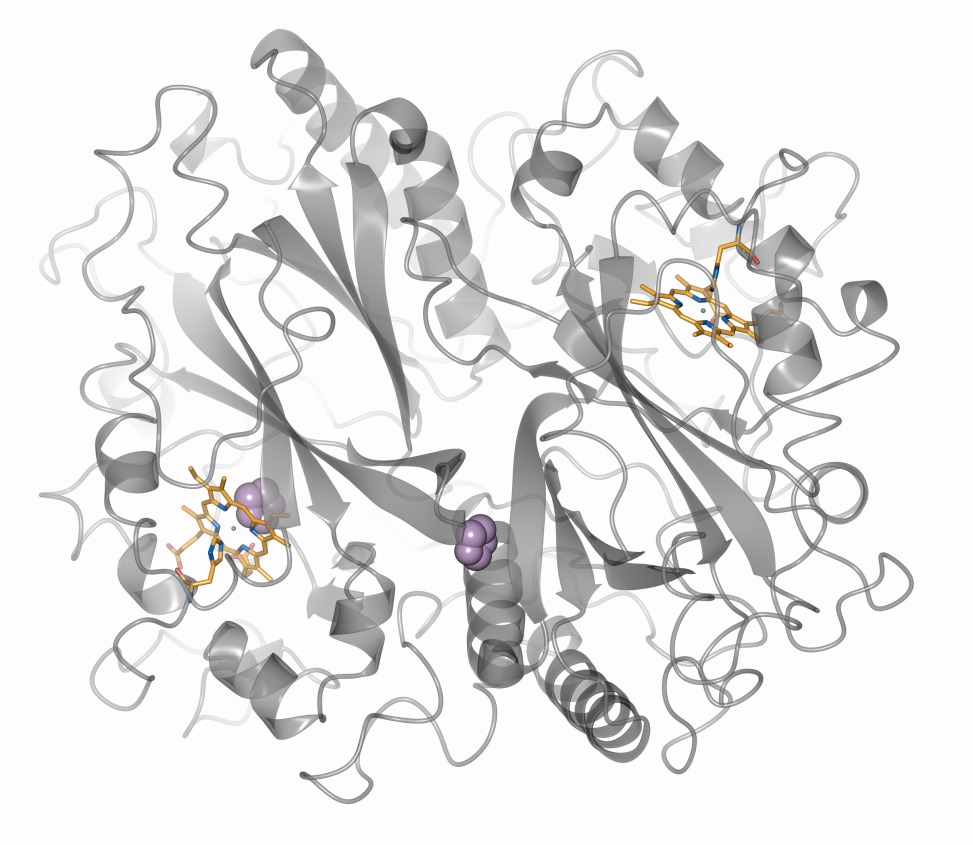

**Figure 4.48:** Overall view of the dimer forming the asymmetric unit of the DtpAa – imidazole SFX structure. Heme cofactors are represented in yellow, and the two ligand molecules found in the structure are represented as purple spheres. One of them is bound to the distal site of heme B (left side of the dimer), and the second one can be found at the interface between the two monomers.

### *4.7.6 Dose-series SSX at Diamond with DtpAa.*

As seen in Chapter 2, the development of serial crystallography methodologies to satisfy XFEL requirements, has opened the possibility to apply the variety of these systems in synchrotron sources, where beamtime is much more accessible and the beam properties are highly stable. Its implementation at synchrotrons has been very useful, if not necessary, to test and optimize the experiments to be performed at the XFEL beamtime, and making it much more efficient and productive. The fixed target system that we used at SACLA (described in Chapter 3) had been also tailored to be used at beamline I24, at Diamond. In fact, its development and optimization by the Diamond team was carried out at I24 (Owen et al. 2017), before moving onto XFEL experiments.

One of the advantages of using fixed targets at a synchrotron, is that since the intensity of the beam is not as extreme as in an XFEL (9 orders of magnitude less intense at peak values), several diffraction patterns can be collected from a same microcrystal without it being destroyed, building up in this way a succession of structures with increasing accumulated doses. These are effectively different frames forming a structural movie. We have name this kind of measurement a ‘dose-series’ experiment, and it is an application of the Multiple Structures from One crystal (MSOX) methodology (Horrell et al. 2016; Horrell et al. 2018) to serial crystallography (i.e. to a microcrystal population). With this approach, one would be able to follow conformational changes induced by photo-reduction or redox catalytic processes fed by the beam, at atomic detail. A schematic illustration for an example of a dose-series (Ebrahim

et al., in press) with 10 dose points of 10 ms each can be appreciated in Figure 4.49.

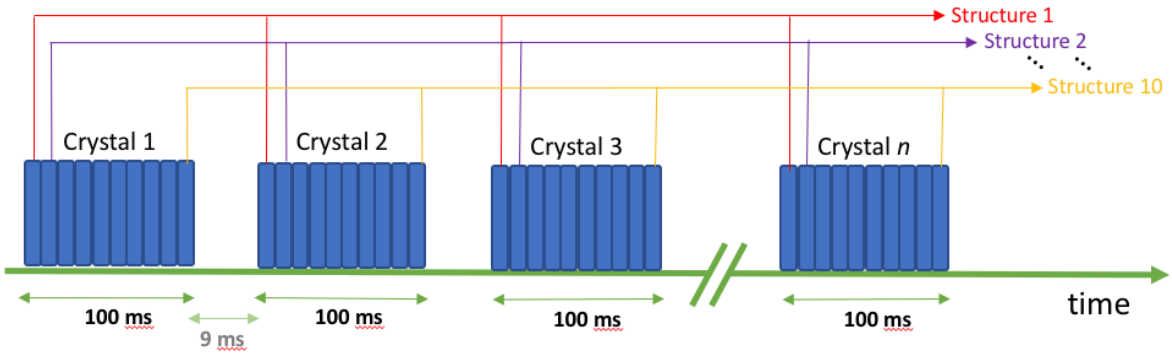


**Figure 4.49:** Schematic representation of a dose-series experiment performed on fixed target, where each microcrystal is measured 10 time successively, with X-ray exposures of 10ms each, before moving onto the next point of measurement. (Ebrahim et al., in press)

We performed such an experiment with the Diamond fixed targets at beamline I24, using protein DtpAa, which delivered the most suitable microcrystals for this kind of experiment, and where we had an XFEL structure to provide a zero point of radiation-damage. A DtpAa dose-series of 5 consecutive dose points or structures was measured, with experimental details shown in Table 412.

**Table 4.12:** Experimental parameters for the dose-series experiment performed using fixed target serial crystallography with DtpAa microcrystals at beamline I24, Diamond Light Source (Oxfordshire, UK).

| | |
|---|---|
| **Sample** | Ferric DtpAa, WT |
| **Chip name** | Fugees |
| **Detector distance (mm)** | 340 |
| **Beam size (μm)** | 9x6 |
| **Flux ($e^-$/s)** | $3.0x10^{12}$ |
| **Transmission (%)** | 100 |
| **Energy (eV)** | 12800 |

| | |
|---|---|
| **Inspected positions** | 25,600 |
| **Shots per position** | 5 |
| **Exposure time per shot (ms)** | 10 |
| **Dose per shot (kGy)** | 46 |
| **Average hit rate (%)** | 38.0 |

Diffraction data was stored in sequentially numbered images in *cbf* format. Data processing until obtention of final mtz files for all the DtpAa dose-series was carried out by Mr. Martin Appleby, at Diamond Light Source. Diffraction patterns were indexed using 'dials.still_process' in *DIALS* version 1.8.5 (Winter et al., 2018), with subsequent scaling and merging performed using prime (Uervirojnangkoorn et al., 2015). Five *mtz* files were generated as a final output, one per each dose point. Data collection and refinement statistics for the whole dose-series are shown in Table 413. Resolution cut-off was assessed using $CC_{1/2}$ and $R_{split}$ parameters. Each of the structures was refined against the SFX solved structure of DtpAa, using *Refmac5* (Murshudov et al., 2011) in the CCP4i suite. Model building was performed in *Coot* (Emsley et al., 2010) in between rounds of refinement, and validation was performed using tools within *Coot* and *Molprobity* (Chen et al., 2010).

**Table 4.13:** Data collection, processing and refinement statistics for the five structures forming the DtpAa dose-series. Data were collected from a single chip (25,600 positions) with 5 sequential images each of 10 ms per position. The beam size was 9x6 microns, with an incident flux of $3x10^{11}$ ph/s at a wavelength of 0.9686 Å. The diffraction weighted dose per dataset was 46 kGy.

| | **Dose point 1** | **Dose point 2** | **Dose point 3** | **Dose point 4** | **Dose point 5** |
|---|---|---|---|---|---|
| Cumulative dose (kGy) | 46 | 92 | 138 | 184 | 330 |
| Number of integrated frames | 9,777 | 9,839 | 9,949 | 9,800 | 9,237 |
| Spacegroup | $P2_1$ | $P2_1$ | $P2_1$ | $P2_1$ | $P2_1$ |
| Cell dimensions (Å, º) | 73.03, 68.30, 74,76; 105.65 | 73.06, 68.30, 74,74; 105.69 | 73.09, 68.30, 74,76; 105.74 | 73.14, 68.29, 74,78; 105.78 | 73.18, 68.28, 74,82; 105.81 |
| Resolution (Å) | 29.54 – 1.80 (1.84 - 1.80) | 29.55 - 1.80 (1.84 - 1.80) | 29.55-1.80 (1.84-1.80) | 29.55-1.80 (1.84-1.80) | 29.56 – 1.80 (1.86 – 1.80) |
| $CC_{1/2}$ | 0.92 (0.32) | 0.92 (0.31) | 0.92 (0.32) | 0.92 (0.33) | 0.91 (0.30) |
| Multiplicity | 13.3 (10.8) | 13.3 (10.8) | 13.3 (10.9) | 13.3 (10.8) | 13.3 (10.6) |
| Completeness (%) | 100 (100) | 100 (100) | 100 (100) | 100 (100) | 100 (100) |
| Number of reflections | 67,734 (3,891) | 65,740 (3,883) | 65,759 (3,888) | 65,820 (3,906) | 65,867 (3,885) |
| R/Rfree (%) | 18.5 / 21.3 | 18.4 / 22.0 | 19.7 / 22.9 | 21.1 / 24.7 | 22.4/ 26.1 |
| RMSD bond lengths (Å) | 0.014 | 0.014 | 0.013 | 0.011 | 0.012 |
| RMSD bond angles (deg) | 1.91 | 1.93 | 1.84 | 1.77 | 1.813 |
| Ramachandran plot | | | | | |
| Most favoured (%) | 97.3 | 96.9 | 97.5 | 97.3 | 97.0 |
| Allowed (%) | 2.7 | 3.1 | 2.5 | 2.7 | 3.0 |

### *4.7.6.1 Dose-series structural movie of ferric DtpAa*

The structural analysis of the data was started by looking at the first dose point, which yielded a low-dose structure of 46 kGy (Figure 4.50). This structure presents almost an identical conformation of the protein as that seen in the damage-free resting state DtpAa SFX structure measured at SACLA. A nuance indicating a slight photo-reduction would be the longer bond Fe – O between the heme iron and the distal bound water, which is 2.42 Å, while in the XFEL

structure is 2.39 Å. A superposition of the heme environment for both structures is shown in Figure 4.50 (right panel), where the high degree of similarity can be noted, with an RMSD between them of 0.11 Å (for monomer A; 0.13 Å for the whole structure).

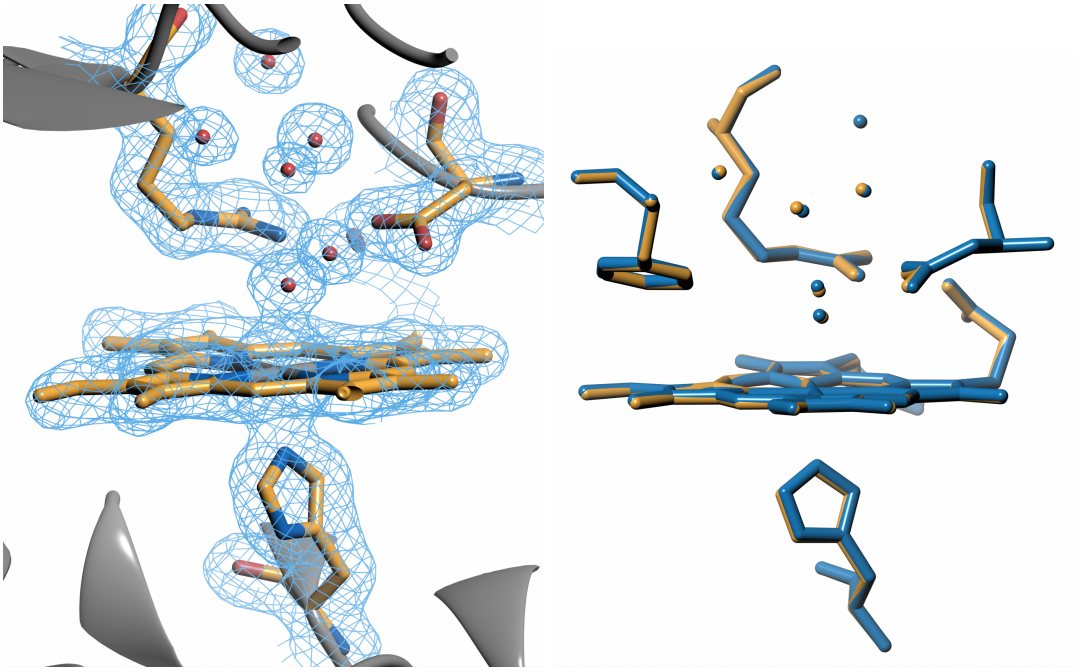

**Figure 4.50:** DtpAa SSX structure corresponding to the first dose point of the dose-series performed at I24, Diamond. On the left panel, representation of heme site A with the distal residues Asp239 and Arg342 and the network of distal waters, with electron density contoured at 1σ. On the right panel, superposition of this same low-dose structure with the damage-free SFX structure of resting state DtpAa, showing an almost identical conformation.

Initial dose-induced changes in a metalloproteins would be expected at the active site, in this case the heme cofactor, since it is well reported how metals firstly affected by radiation damage (Garman & Weik, 2015; Yano et al., 2005). When examining the heme site of DtpAa along the dose-series, we could observe subtle changes occurring at the distal site, specifically at the water at the axial position above the haem iron, and also in the rearrangement

of the network of waters in the distal pocket. In Figure 4.51 the evolution of the active site through the 5 dose-points is shown. As discussed, in the first dose point, just after 10ms of exposure to X-rays and 46 kGy dose absorbed (dose rate of 4.6 MGy/s), the resulting DtpAa structure is close to the ferric damage-free species from SACLA. As dose absorbed by the protein increases, the iron centre becomes more photo-reduced by the beam, and the bond between the iron atom and the distal water increases steadily until the water can be appreciated completely detached from the metal and with low-occupancy, representing a disordered distal water. In fact, the whole network of waters found in the distal pocket, which are well defined and spherical at the beginning of the series, become more disordered as dose increases, showing diminished and more irregular electron density. For instance, distal water W2 is barely supported by electron density in final two dose-points, and waters W3 and W5 get closer and its electron densities merge due to disorder. Therefore, with the dose-series approach we could structurally characterize radiation-induced changes at the active site of resting state DtpAa.

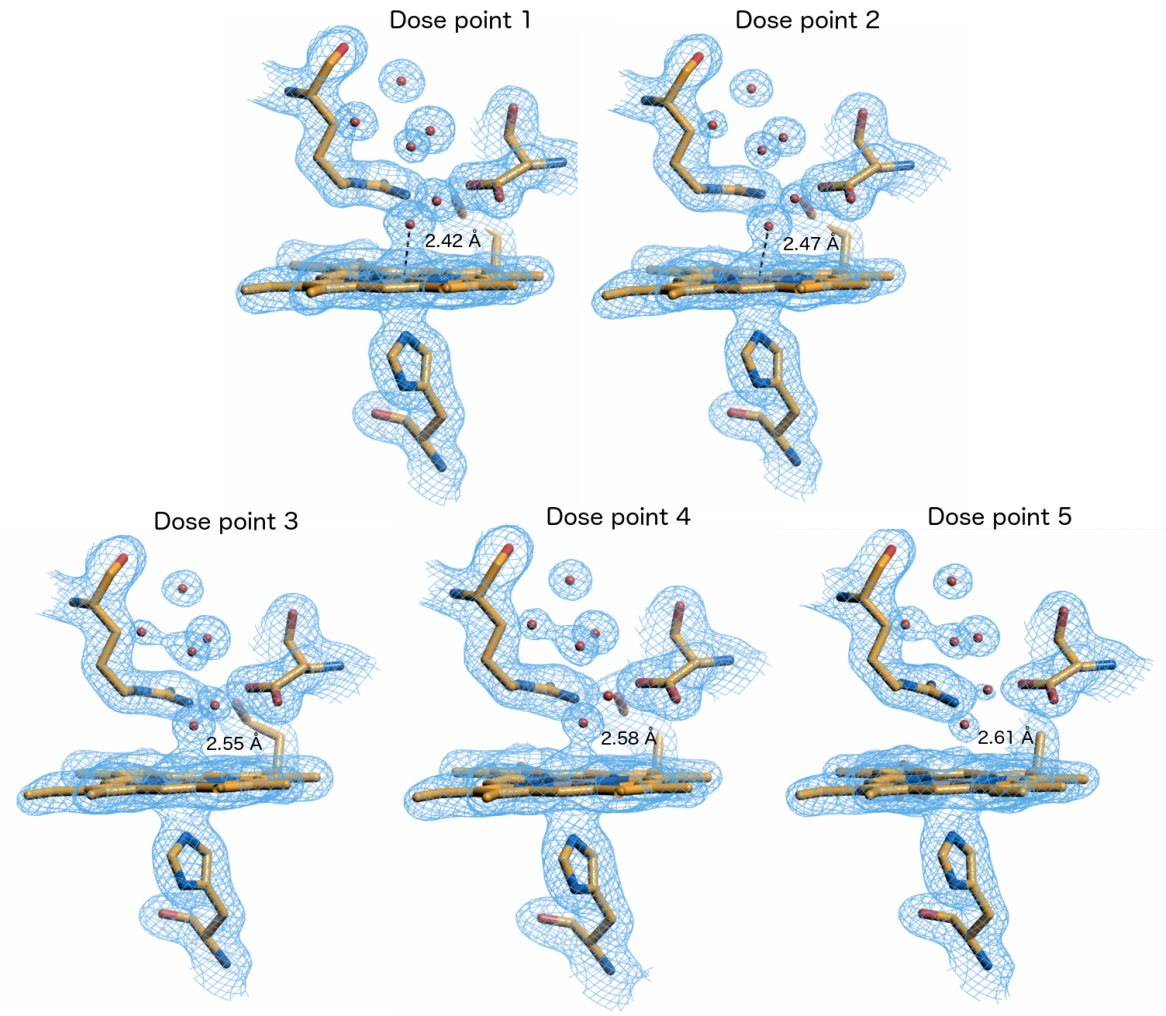


**Figure 4.51:** Active site environments for each of the structures or dose-points forming the dose-series experiment from ferric DtpAa microcrystals, carried out at beamline I24 (Diamond) using fixed target Synchrotron Serial crystallography (SSX).

## 4.8 Discussion

DyP enzymes DtpA and DtpAa from *Streptomyces lividans*, were extensively studied during this work with X-ray crystallography combined with different methodologies, such as single crystal spectroscopies and serial crystallography. It has been shown how the synergy of techniques and methodologies is a powerful tool to better characterize its structures and that of their intermediates and complexes, and thus better understanding how they

function. In this chapter, we report the structure for different species of DtpA and DtpAa, for instance those involved in the catalytic cycle of the former one. For the structures measured from large single crystals at conventional synchrotron sources, they were validated using *in crystallo* spectroscopies, which allowed us to assign the correct state of the protein. The structures measured with an XFEL source as SACLA, using the SFX technique, provided the damage-free resting state for both enzymes. Finally, using serial crystallography at synchrotrons (SSX), the novel dose-series approach is presented and applied to DtpAa, in which we report radiation-induced changes at the active site of the ferric state at atomic detail.

With the use of *in crystallo* spectroscopies, the photo-reduction of DtpA could be monitored, both with UV-vis and with single crystal resonance Raman (SCRR), obtaining the validated oxy-ferrous species. Each of the spectroscopies were performed at different sources, with different micro-spectrophotometers each of which featuring a different geometry. In the case of SCRR, it was carried out at beamline X10SA at SLS, where the MS3 micro-spectrophotometer presents an on-axis geometry (Pompidor et al., 2013), which allows to measure spectra from the same crystal region which has been probed by the X-ray beam. For the measurement of UV-vis instead, it was performed using the off-axis micro-spectrophotometer available at ESRF (von Stetten et al., 2015), which had to be aligned to probe the same region of the crystal as the beam. This did not suppose an issue with small crystals, such as the ones used for the photo-reduction experiments, which were around 50 μm flat plates, and the beam had a top hat profile 100x100 μm, affecting equally

the whole crystal. However, for the UV-vis validation of the ferric structure with this same instrument at ESRF, it presented more of a challenge, since the dimensions of the crystal needed were much larger (~350x200x75 μm) and the beam size was much smaller (15x15 μm FWHM). The light spot of the micro-spectrophotometer was focused to 25 μm, and the light path was perpendicular to that of the of the X-ray beam – hence, there were unexposed parts of the crystal contributing to the measurement of the spectrum. This did not present an issue in this case, since an intact species could be observed at the active site, but could well be a limitation for the characterization of other species, such as peroxidase intermediates, when using large size crystals. Furthermore, it could lead to incorrect determination of species and consequent misinterpretation of reaction mechanisms. Using the UV-vis instrument at ESRF we also performed a progressive reduction experiment, which permitted us to quantify the dose needed to photo-reduce DtpA, and showing another utility of single crystal spectroscopies. The estimated dose to fully photo-reduced DtpA was ~12-14 kGy, a very low value compared to the dose needed to measure a complete dataset (0.99 MGy for DtpA at that beamline). However, the weighted dose calculated for the ferric dataset measured at ESRF with a helical collection was of 12 kGy, which contrasts with the previous calculation. This could be due to the different modes of data collection, as it was static for the progressive reduction, and with rotation along a helical path in the other case. Beamlines were also of very different characteristics; the progressive reduction was performed with a top hat big beam (100x100 μm); whereas for the determination of the low-dose ferric structure a focused beam (15x15 μm) was used. The crystal sizes and shapes were also radically different, and could

also have something to do, although the calculation of the weighted dose should have into account the crystal dimensions. But there could have been a misestimation of crystal sizes in one or both cases. Ultimately, the collection of low-doses datasets of 10-20 kGy is considered to be a ‘safe’ dose to obtain intact species of metalloproteins, and our ferric DtpA structure is in agreement with this observation.

The peroxidase cycle was characterized with the crucial help of *in crystallo* UV-vis spectroscopy, to assign the correct species generated within the DtpA crystals, and to validate the obtention of photo-reduced states involved in the catalytic mechanism. We could determine in this way, the structures of the oxy-ferrous, deoxy-ferrous and ferrous forms of DtpA, all of which were generated in situ at the beamline exploiting radiation damage, as seen also in the literature (Schlichting et a., 2000). Peroxidase intermediates compound I, II, and III were observed in the crystal, but their structures could not be determined due to its rapid photo-reduction when measured. The performance of composite datasets (Berglund et al., 2002) is planned for following scheduled beamtime at SLS, to allow us to build up a low-dose dataset and try to obtain these intermediates, which would be important as it would complete the structural characterization of the DyPs peroxidase cycle. However, this might be particularly challenging for the determination of Compound I, if it results not stable at the beamline and behaves in the same way as in the previous *in crystallo* measurements reported in this chapter. Mounted at the beamline (X10SA at SLS) at 100 K, it would develop to Compound II within 4-5 min. We do not have yet an explanation for this

observation, since it means a 1e$^-$ reduction redox process to yield Cmp-II, and there is no X-ray beam to deliver electrons. However, it is known that Cmd-I spontaneously develops to Cmd-II (Moody & Raven 2018), an a protein residue is thought to provide the needed electron, and thus generating a radical within the protein. Different soaks of peracetic acid will be tested to try to find the conditions that stabilize Cmd-I. Another interesting experiment to perform would be to test if Cmd-0 can be generated after slight irradiation by the beam, as shown by Kuehnel et al. (2007) with chloroperoxidase.

The optimization of batch conditions delivering microcrystals of DtpA and DtpAa, allowed the application of the DyP targets onto fixed targets to perform serial femtosecond crystallography at SACLA. This approach was successful and delivered the damage-free resting state structures for both enzymes, which also represent the first room temperature (RT) structures of the DyPs in *Streptomyces lividans.* Thus, their conformation is more representative of that of the enzymes in its biological conditions. Unfortunately, we could not compare rigorously the RT SFX structure of DtpA with the the cryo-structure ferric species measured at ESRF, since crystals were grown at different conditions with differences of pH (pH 5.5 vs 7.5). Important differences were noted between the structures, such as the non presence of the bound distal water at the distal site of the SFX structure; although some positive electron density could be appreciated, suggesting a partial occupancy of a distal water. We think that these differences could be due to the pH in each condition, which affects the protonation states of the distal residues (e.g. Asp251), and ultimately the coordination of the heme. Regarding the oxidation

states of the DyPs studied, a suspension of microcrystals from a batch in its own crystallization condition was measured with a UV-vis spectrophotometer in the lab, to certify its ferric features, before the trip to SACLA. Batches were treated carefully and maintained at all times in a container protecting them from light and heat sources. Microcrystals were assumed to have kept the oxidation state, since DtpA and DtpAa are stable in its ferric form at room temperature (from experience in the lab), and DtpA crystals had been transported in other occasions at RT to SLS maintaining ferric features when inspected. We do not know if the loading process onto the chips could potentially change the oxidation state of the crystals, but it was performed in a moisture-controlled hood and the silicon nitride forming the fixed targets is an inert material. Most importantly, the measured structures presented the resting state distal water that was seen in the ferric DtpA cryo-structure, and if the proteins would have been reduced at any point before measurement at SACLA, most likely the oxy-ferrous complex would have been formed, as seen how oxygen easily promptly binds to DtpA (even at 100 K) in the photo-reduced structure reported at the beginning of the chapter. Apart from the two resting state structures of DyPs, the SFX structure of the DtpAa – imidazole complex was also determined, showing at the same time how with SFX small ligands can be effectively identified. Thus, the XFEL beamtime at SACLA was especially productive, and this was in great part because of the valuable time spent at beamline I24 (Diamond) with SSX experiments to test diffraction and optimize hit rates.

The implementation of serial crystallography at synchrotron sources (SSX), in our case using fixed targets, has then turned out to be crucial for the

preparation of the scarce and competitive XFEL beamtimes. Furthermore, SSX has opened new possibilities for the measurement of metalloproteins such as the one shown with the dose-series experiment with ferric DtpA microcrystals. In it several successive datasets are collected from a same microcrystal, building up structural movie where one can follow radiation-induced or catalytic processes of the protein studied. In our case five dose points were measured, using exposures of 10 ms, and obtaining a lag phase effect where no resolution decrease due to radiation damage was observed. This appear to be another benefit from SSX allowing the measurement of datasets with very short exposures and hence accumulated dose. In fact, the first measured dataset from the series is almost identical to the damage-free SFX structure from SACLA, opening the debate if whether it is necessary to measure intact states of proteins at XFEL sources, or focus their use in time-resolved crystallography experiments taking advantage of its unique capabilities. Along the five structural frames of the radiation-driven process, we could monitor the photo-reduction of the heme group, observing subtle changes in the network of distal waters, including the bound axial water, which became more disordered as the absorbed dose increased. Finally, it was not known if the photo-reduction process of DtpA had finished and stabilised with those five dose points. For this reason, another dose-series experiment was performed, this time with 20 dose-points, to investigate if the process further develops to, for instance, the formation of an oxy-ferrous species. These data are currently under analysis and will be reported elsewhere.

# Chapter 5 Characterization of DHP redox and ligand states - azole complexes.

## 5.1 Introduction

As discussed in the first chapter, DHP is a haem multifunctional enzyme, which is involved in the oxidation of trihalophenols into dihaloquinones, as a main known task functioning as a peroxidase. After extensive research (De Serrano et al. 2007; Chen et al. 2009; De Serrano et al. 2010; Barrios et al. 2014), DHP has been noted to carry out additional functions, such as oxygenase, oxidase and peroxygenase, catalysing many other reactions involving mainly organic halogenated compounds. In this work, we aimed to characterize some of DHP's redox/ligand states to identify structural changes that determined the differences in functionality of the protein. The isoform B of DHP was the only studied, and we will refer to it throughout the text as DHP-B or simply as DHP.

## 5.2 Protein expression and purification

An expression vector pET16b encoding tag-less DHP-B was kindly provided by Dr. Guiladi group, and was transformed into *E. coli* expression cells BL21-DE3. Cells were pre-cultured overnight in 10 ml containers supplemented with ampicillin (100 μg/ml). Protein expression was carried out in 2 litre flasks inoculated with the pre-cultures, and supplemented with ampicillin (100 μg/ml),

0.1 mM hemin as heme source, and 0.1mM iron citrate as $Fe^{3+}$ source. Cells were grown at 37°C and 220 rpm and induced with 0.5mM IPTG when $OD_{600}$ ≈ 0.8, at which point conditions were changed to 25°C and 200 rpm, and the expression was left overnight for a minimum of 18 hours. Cells were spun at 5,000 rpm for 20 min, and the pellets resuspended in 100 mM KPi buffer pH 8.0, and lysed with an Emulsiflex-C5. The lysate was centrifuged at 18,000 rpm for 40 min to get rid of cellular debris. The purification of the tag-less protein involved several steps, which were followed as specified in the protocol developed by De Serrano and colleagues (2007). It started with a salt fractionation step, adding ammonium sulphate up to 55% saturation (point at which DHP starts to precipitate). After removing the precipitate via centrifugation (18,000 rpm) for 20 min, the sample was dialysed with an 8 kDa cut-off membrane against 5 L of 20 mM KPi pH 6.4 buffer at 4°C. Four changes of buffer were performed, each for a minimum of 6 hours, and the last one being 20 mM KPi pH 5.4. At this point, protein was loaded into a cation-exchange SP-sepharose column for further purification, performing a salt gradient from 0 to 500 mM NaCl on an AKTA-purifier. The fractions from the eluted peak were pooled and run through a G-75 size exclusion column for final polishing of the protein sample. Protein purity was assessed by SDS-PAGE (Figure 5.1), and all pure fractions were pooled and oxidised adding $K_3(FeCN_6)$ in excess to oxidise the protein and obtain a homogeneous ferric sample. A PD-10 column (GE-healthcare) was used to get rid of this chemical and buffer exchange the protein into the final buffer: 20mM MES pH 6.5. The pure protein sample was concentrated using a Vivaspin-10 concentrator (Sartorius) with 10kDa cut-off, until a concentration of ~15 mg/ml. Concentration was determined by UV-vis

spectroscopy, using the extinction coefficient at the haem absorption maximum for the ferric protein ($\varepsilon_{406nm}$ = 116.4 $mM^{-1}cm^{-1}$; D'Antonio et al., 2010).

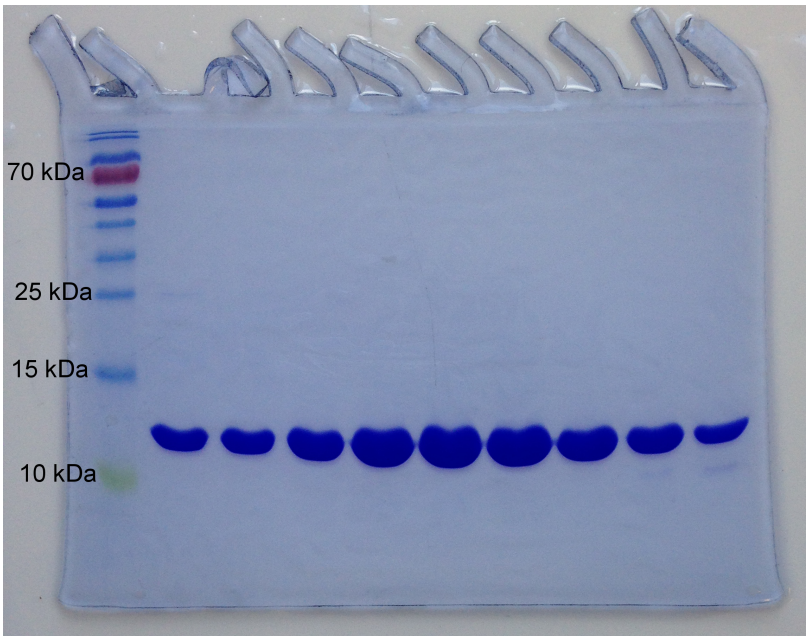


**Figure 5.1:** SDS-PAGE of DHP fractions from a G75 size exclusion polishing run. All fractions were pooled and concentratedProtein marker ladder in was loaded in the first well on the left, with molecular weight for relevant protein marker bands indicated.

## 5.3 Crystallisation

DHP-b crystals were grown using the hanging drop vapour diffusion method at 4ºC (Figure 5.2), at the crystallization condition described in the literature (De Serrano et al. 2007). The protein concentration used was 12 mg/ml and drops of 4-6 μl (1:1 or 1:2 volume ratio protein to reservoir) were equilibrated in 24-well crystallization plates against a reservoir containing 12-20 % PEG 4000 and 150-250 mM ammonium sulphate. Generally, crystals would appear after 2-3 days. It was noted that the freshness of the PEG used was important to significantly improve crystallization – DHP crystals would grow

in larger and sharper crystals which would eventually produce better diffraction quality.

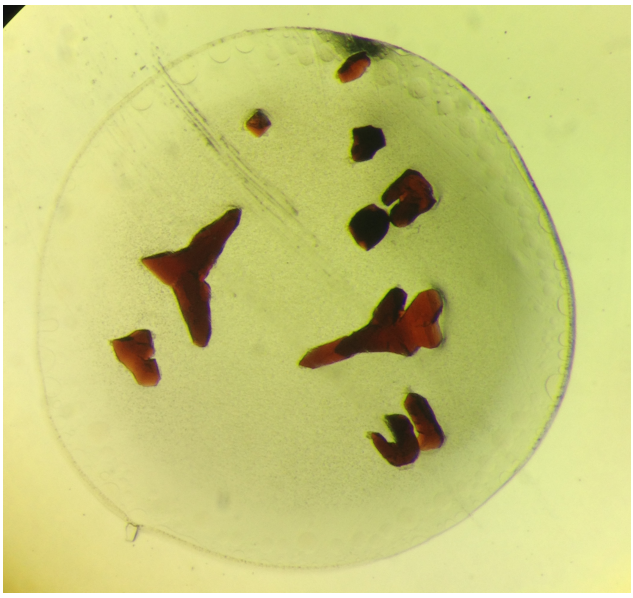

**Figure 5.2 :** Ferric DHP-B crystals grown in a crystallization drop grown at 4ºC. Note its high optical density. The dimensions of the DHP crystals ranged from 50-400 μm, and diffracted to atomic resolution.

## 5.4 Soaking experiments

DHP-B ferric crystals were soaked in prepared crystallization solutions (20 % PEG-4000 and 200 mM ammonium sulphate) with added 5-10 mM concentrations of a different ligand in each case. Ligands soaked in the present work were: azide, imidazole, cyanide, benzimidazole, indazole, and benzotriazole. Some of them, such as azide, cyanide and imidazole are small ligands known to bind haem proteins (Scheidt & Chipman 1986; Perutz 1990). The rest are double-ring aromatic azole ligands were selected as they bind to DHP-B (tested by Guiladi lab) and represent persistent organic pollutants (POPs) with industrial and agricultural applications (Gaikwad et al., 2015;

Giger, Schaffner, & Kohler, 2006; Rodríguez, J. J. S., Padrón, M. E. T., Aufartová, J., and Ferrera & S., 2010). They are able to interact with DHP since, acting as a peroxidase it performs the oxidation of trihalophenols to dihaloquinones, and it holds a pocket able to accommodate big aromatic ligands, a feature which is unusual in haemproteins, which normally bind small ligands (e.g. diatomic gases, cyanide, nitric oxide). For imidazole, cyanide and azide, stock solutions were prepared in water, and the azole ligands were dissolved in DMSO due to their limited solubility in water. Soaking solutions where DMSO was used were prepared to a final concentration of 2 % DMSO, to minimize the damage to the crystals by this chemical. As a general soaking procedure, crystals were left in the soaking solution for 5-10 minutes (sometimes a colour change would be appreciable after a few seconds), and then transferred to a cryo-condition containing 20% glycerol added to the reservoir, prior to flash-cooling in liquid nitrogen.

## 5.5 Data collection and spectroscopic measurements at synchrotrons.

Crystals were measured at Swiss light source (SLS; Villigen, Switzerland); Diamond light source (DLS; Oxford, UK); and the European Synchrotron radiation facility (ESRF; Grenoble, France). At SLS we carried out both diffraction and single crystal resonance Raman (SCRR) experiments on the same crystal, using beamline X10SA, which is equipped with a Pilatus 6M-F area detector (Dectris Ltd.), and an on-axis micro-spectrophotometer (Pompidor et al., 2013). The procedure followed was to measure SCRR for

each crystal at the start, before any exposure to X-rays, and then perform an X-ray diffraction data collection, followed by a further SCRR spectrum. Thus, with the second spectrum we could appreciate any changes experienced by the protein upon beam irradiation. The X-ray selected wavelength for the experiments was 1.0 Å, unless stated otherwise, normally with 0.1º rotation per frame, exposure times of 0.1 s and flux around $2 \cdot 10^{11}$ photons/s. The laser excitation wavelength for the SCRR measurements was 405 nm, with an optical spot size of 25 μm. The power applied at the sample for each measurement was 4.95 mW, and each spectrum was acquired with 20 accumulations of 20 seconds each (20x20s).

At the ESRF, we coupled X-ray diffraction with on-line UV-vis spectroscopy measurements, applying the same approach: spectrum → dataset → spectrum. These experiments used the on-line UV-vis spectrophotometer available at ESRF (McGeehan et al. 2009), as described in Chapter 2. Spectroscopic off-line measurements of the crystals were performed at the cryobench facility (Von Stetten et al. 2015), which was of great help to test and optimize conditions for the online experiments. Spectra acquired at the cryobench were measured for 0.6 s x 10 accumulations.

## 5.6 Initial spectroscopic characterization of DHP crystals.

As a general procedure to assess the possibilities of the protein system, initial characterization experiments were carried out at SLS and ESRF, where tests with the different spectroscopies available were applied to DHP crystals on-line and off-line (at the cryobench). These included UV-vis, Resonance

Raman, Raman, and fluorescence spectroscopies. It was observed that we could easily monitor photo-reduction of the protein in the crystal with both UV-vis and SCRR spectroscopies.

### *5.6.1 Monitoring DHP photo-reduction with SCRR at SLS*

Ferric DHP crystals diffracted at the SLS to atomic resolution (below 1.2 Å) and were spectroscopically inspected by SCRR, obtaining a good signal. As usual, spectroscopic measurements were taken before and after X-ray diffraction, and clear spectral changes were observed (Figure 5.3), indicating that by using SCRR we could obtain valuable information about the state of the protein at its active site. Several peaks are defined within our area of interest in the RR spectrum, between 1300 and 1700 $cm^{-1}$, the strongest one being the haem redox marker, which has a value of 1370 $cm^{-1}$ when DHP is in its ferric state (black in Figure 5.3). The remaining peaks (above 1400 $cm^{-1}$) are known in the literature as core size markers (Mark et al., 2008), and are not as well defined, although changes can be appreciated upon beam irradiation. In this case, the core size markers values observed for DHP before irradiation (i.e. 1560 and 1613 $cm^{-1}$) are in agreement with RR solution data from ferric DHP-B (D'Antonio et al. 2010). After exposure to X-rays, the redox marker is observed to populate two peaks (purple in Figure 5.3): a predominating one at 1352 $cm^{-1}$, corresponding to the reduced form of the haem group; and the other one at 1370 $cm^{-1}$, corresponding to the ferric form. Therefore, using SCRR we could monitor photo-reduction of the DHP and validate its redox state. By integrating the areas of the ferric and ferrous peaks one could also quantify the

proportion of photo-reduced protein in the exposed region by the excitation laser.

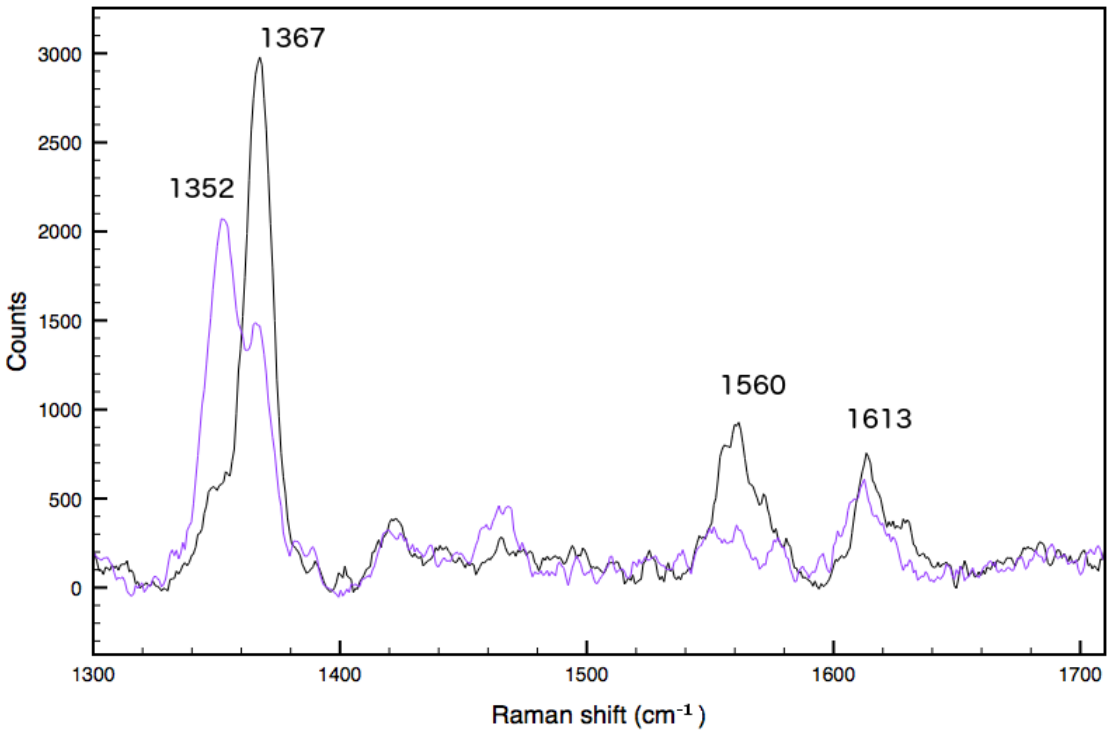


**Figure 5.3:** Single Crystal Resonance Raman (SCRR) spectra of a ferric DHP-B crystal (in black), and the same crystal after exposure to X-rays (in purple) for collection of a diffraction dataset. Several peaks can be observed, the strongest one being the characteristic haem redox marker at around 1370 $cm^{-1}$. Note that this peak becomes split in two after exposure to X-rays due to photo-reduction of the active site, with the reduced redox marker peak at around 1350 $cm^{-1}$. Data were measured using the on-axis MS3 micro-spectrophotometer at beamline X10SA, Swiss Light Source (SLS).

### *5.6.2 Monitoring DHP photo-reduction with UV-vis spectroscopy at ESRF.*

At the ESRF, both on-line and off-line (at the cryobench) UV-vis spectra were measured for DHP crystals using the experimental set ups described in Chapter 2. Ferric DHP and soaks with different ligands were tested, demonstrating that this spectroscopic technique could be also useful to monitor

redox and ligand state changes for this protein. In Figure 5.4, an example of on-line spectra measured from ferric DHP-b crystals at beamline BM30 before and after diffraction are shown. Due to its high optical density, DHP crystals absorbed light strongly and the spectrum would be usually saturated below 500 nm, making it difficult to obtain a complete and well defined spectrum. However, when using small crystal (around 50-100 microns), some features could be observed such as the $\alpha$, $\beta$ and CT bands, which showed important changes upon photo-reduction of the crystal by the X-ray beam (Figure 5.4).

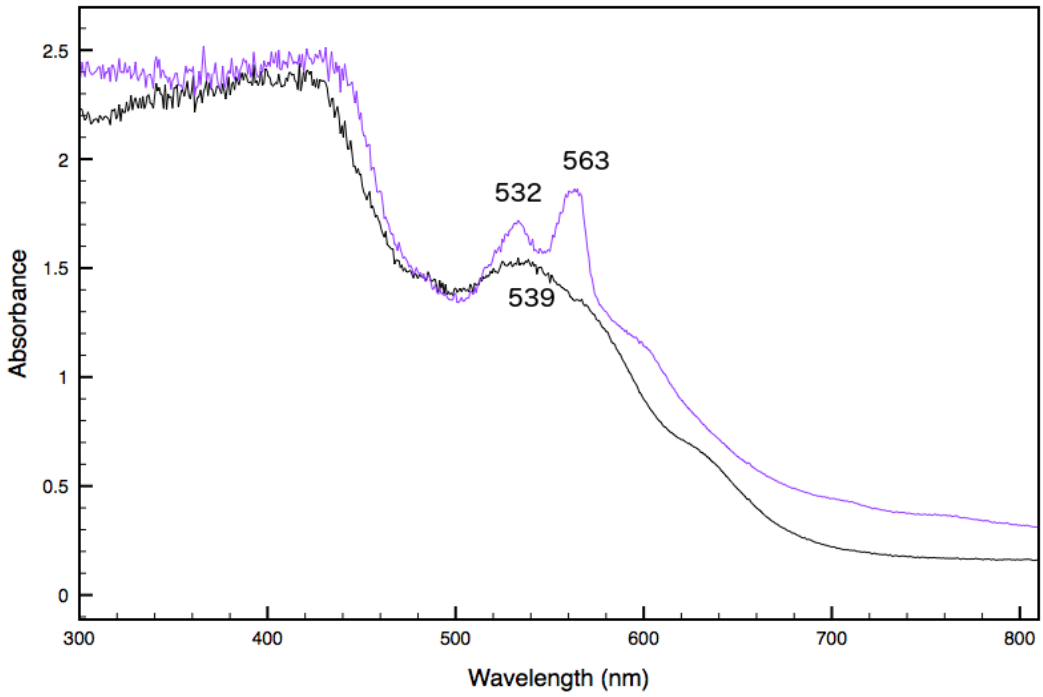


**Figure 5.4:** On-line UV-vis spectra of a DHP-b ferric crystal, acquired before any exposure to X-rays (in black) using the on-line micro-spectrophotometer at beamline BM30 at ESRF. A further spectrum of the same crystal after collection of a diffraction dataset is shown in purple. Note the saturation observed at lower wavelengths (below 450nm). Some spectral features can be appreciated such as the $\beta$ and $\alpha$ bands (above 500 nm), and the CT band (around or above 600 nm).

### *5.6.3 Selection of most suitable single-crystal spectroscopy for DHP.*

Although valuable complementary information about the state of the protein could be obtained by means of UV-vis spectroscopy, as we have just shown, the high optical density of the DHP crystals made single-crystal resonance Raman the most suitable technique for this protein system. As a general rule, optically dense crystals (i.e. with a higher concentration of chromophore in them) are optimal for resonance Raman, since the signal is more amplified due to the higher concentration of the chromophore within the depth to which the excitation laser penetrates the crystal. In contrast, crystals which are less optically dense (i.e. paler) present a weaker RR signal, but are more suitable samples to be inspected with UV-vis spectroscopy, normally providing a well-defined spectrum with ideally no saturated regions. Therefore, in the case of DHP, SCRR was the selected technique to study this system, which would be carried out with the MS3 micro-spectrophotometer at SLS. This selection was not dogmatic though, and UV-vis would still be tested for its usefulness to identify certain states of the protein and follow reactions.

## 5.7 Partially photo-reduced DHP-B structure at 0.99 Å

As shown by both spectroscopic techniques used, most of the DHP protein illuminated by the beam was photo-reduced after X-ray data collection. Therefore, the initially measured structures of DHP at SLS and ESRF represent a mixture of redox states: ferric and ferrous. Here we include the highest resolution dataset that we obtained, which is one measured at SLS reaching

0.99 Å resolution, where more structural details could be appreciated. Data collection and refinement statistics for this partly photo-reduced DHP-B structure are detailed in Table 51. The SCRR data corresponding to this structure is then that shown in Figure 5.3. The protein was found forming a dimer in the asymmetric unit of the crystal (Figure 5.5), and presented a globin fold with 7 helices bundled around a heme-B cofactor, in agreement with literature structures (PDB code 3ixf; De Serrano et al., 2010).

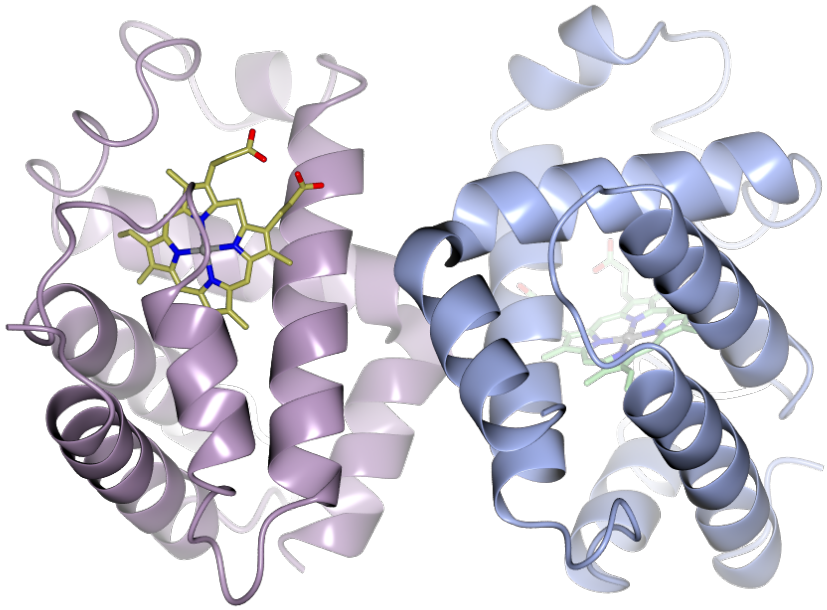

**Figure 5.5:** Overall view of the asymmetric unit for the partially photo-reduced structure of DHP-b, where two protein monomers can be found (blue and purple). The protein adopts a classic globin fold formed by seven $\alpha$-helices per monomer, shown as ribbons. Heme cofactors are represented as cylinders in yellow or green, depending on the monomer. Figure generated using CCP4MG (McNicholas et al., 2011).

**Table 5.1:** Data collection and refinement statistics for the partially photo-reduced structure of DHP-b from an X-ray diffraction dataset measured at Swiss Light Source (Villigen, Switzerland). Values in parenthesis refer to the outermost resolution shell.

| | **Partly photo-reduced DHP-b** |
|---|---|
| Space group | $P2_12_12_1$ |
| Unit cell (Å) | a = 60.94, b = 67.33, c = 67.49 |
| Resolution (Å) | 47.67 – 0.99 |
| $R_{merge}$ | 0.072 (1.216) |
| Total reflections | 965,898 (41,651) |
| Unique reflections | 154,588 (7,579) |
| $I/\sigma(I)$ | 11.8 (1.2) |
| $CC_{1/2}$ | 0.999 (0.480) |
| Completeness (%) | 100 (100) |
| Multiplicity | 6.2 (5.5) |
| $R_{work}$ (%) | 15.7 |
| $R_{free}$ (%) | 17.8 |
| Rmsd bond lengths (Å) | 0.012 |
| Rmsd bond angles ($^o$) | 1.821 |
| Ramachandran favoured (%) | 98.4 |
| Accumulated dose (MGy) | 0.35 |

The atomic resolution reached in this case by DHP crystals resulted in a substantial improvement from other DHP structures in the literature (1.58 Å and 1.62 Å, from De Serrano et al. 2007; De Serrano et al. 2010), which allowed us to observe more details in the structure. Moreover, this established a promising platform for the observation of potential conformational changes occurred upon binding or when trying to drive reactions *in situ*. A representation of both active sites found in the DHP-b dimer is shown in Figure 5.6, where the quality of the atomic resolution data can be appreciated. Indeed, a surprising feature was observed at the active site of monomer B: a double conformation of the heme group, indicating the formation of a hexacoordinate species with

His55 coordinating the iron at the distal site (Figure 5.6, right panel). This interesting feature, is known in the literature as a hemichrome species (Riccio et al. 2002). Not only the heme cofactor presents a double conformation, but also several stretches of main chain surrounding heme B. At heme A, a single conformation of the cofactor was observed, and the catalytic His55 was found swung out of the pocket in an ‘open conformation (Figure 5.6, left panel), and interacting through a hydrogen bond with one of the propionate groups of the heme. As it can be observed, Phe21 at the distal pocket presents a double conformation, and so does His55 with a slight twist.

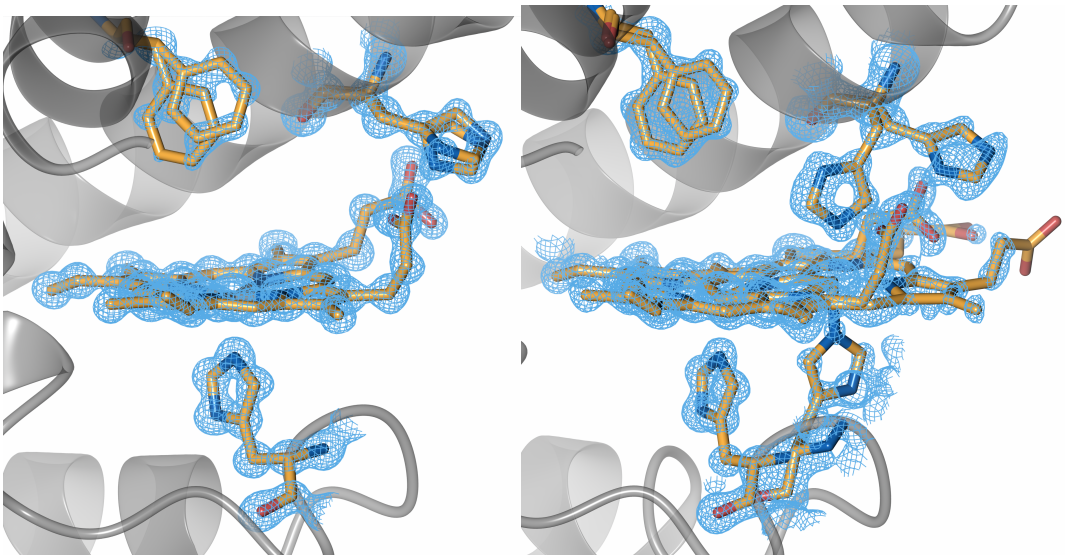

**Figure 5.6:** Heme sites in monomer A (left) and B (right) for the DHP cryo-structure at 0.99 Å obtained from ferric crystals. Electron density map is shown as a blue mesh and is contoured at 1σ. Note the double conformations of His55 and Phe21 in heme A. At heme B, the heme cofactor adopts a double conformation, and the hemichrome species can be appreciated, hexacoordinated by His55 and His89, which also presents a double conformation.

### *5.7.1 Hemichrome species.*

Hemichromes are low-spin six-coordinate species found in ferric Fe(III) haemoglobins, which are mainly produced by partial denaturation of the protein, and are usually associated with blood disorders (Riccio et al. 2002). Hemichromes can be reversible or irreversible, and are reversibly formed when haemoglobins are exposed to certain conditions (Dr. T. Molchanova doctoral thesis, 1981), such as the presence of:

- Fatty acids.
- Aliphatic alcohols (e.g. 1-butanol).
- High concentrations of glycerol.
- Polyethylenglycol (PEG).
- Dehydration.

Since DHP-b crystals were grown in ~15-20% polyethyleneglycol 4000 and were cryo-protected with 20% glycerol, one these substances would be the putative cause for the presence of the hemichrome species in the structure. However, if present in solution, resonance Raman could help to indentify the presence of a low-spin hexacoordinate species, which would be consistent with the hemichrome, potentially giving it a more relevant role than a crystallographic artefact.

## 5.8 Initial ligand bound structures: Imidazole and azide complexes.

As a test for ligand soaking to ferric DHP crystals and to see the viability spectroscopically characterize the formed complexes, soaking experiments were performed with known ligands for DHP: imidazole and azide. To this aim, ferric DHP crystals were soaked in crystallization condition supplemented with 5 mM of either of the ligands for 5-10 min, and after that cryo-protected with 20% glycerol and flash-cooled in liquid nitrogen. Crystal structures of complexes were measured at SLS and SCRR spectra were measured before and after data collection. Diffraction data were processed as usual (detailed in Chapter 3), and both ligands, imidazole and azide, were found bound at the distal site of the haem with full occupancy, directly interacting with the iron, as shown in Figure 5.7. Data collection and refinement statistics are shown in Table 52.

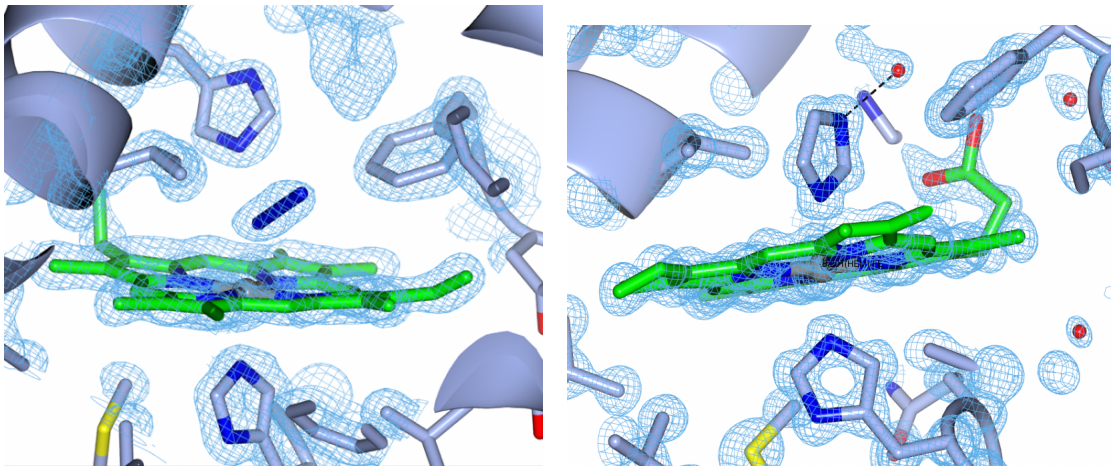

**Figure 5.7:** View of the heme site from the structure of DHP crystals complexed with azide (left panel) and imidazole (right panel). Electron density is shown as a blue mesh and is contoured at 1$\sigma$ (0.4733 $e^-/Å^3$ for the azide complex, and 0.4725 $e^-/Å^3$). Datasets for both crystals were measured at beamline X10SA at Swiss Light Source (SLS; Villigen, Switzerland).

**Table 5.2:** Data collection and refinement statistics for imidazole and azide DHP-b crystal complexes, measured at beamline X10SA, Swiss Light Source (Villigen, Switzerland). Values in parentheses correspond to those of the highest shell of resolution.

| | **Imidazole** | **Azide** |
|---|---|---|
| Space group | $P2_12_12_1$ | $P2_12_12_1$ |
| Unit cell | a=61.36 Å, b=67.36 Å, c=67.52 Å<br>$\alpha=\beta=\gamma=90°$ | a=60.90 Å, b=67.42 Å, c=67.55 Å<br>$\alpha=\beta=\gamma=90°$ |
| Resolution | 47.73 – 1.06 Å | 47.72 – 1.41 Å |
| $R_{merge}$ | 0.016 (0.099) | 0.052 (0.385) |
| Total reflections | 418946 (6143) | 218504 (5150) |
| Unique reflections | 117842 (3254) | 52072 (1604) |
| $I / \sigma(I)$ | 32.4 (6.4) | 14.3 (2.8) |
| $CC_{1/2}$ | 1.0 (0.976) | 0.998 (0.885) |
| Completeness | 93.8% (53.3) | 96% (56.1) |
| Multiplicity | 3.6 (1.9) | 4.2 (3.2) |
| Mosaicity | 0.06 | 0.31 |
| Rwork | 13.3 % | 14.5 % |
| Rfree | 14.3 % | 19.0 % |
| Rmsd bond lengths | 0.009 Å | 0.015 Å |
| Rmsd bond angles | 1.362 º | 1.767 º |
| Rmsd chiral volume | 0.128 $Å^3$ | 0.158 $Å^3$ |
| Ramachandran favoured | 97.9 % | 98.1 % |

Initial single-crystal Resonance Raman spectra were measured before any X-ray exposure to prevent photo-reduction of the crystals and obtain intact spectral signatures of the complexes. As shown in Figure 5.8, both spectra from ligand complexes show differences with the one belonging to the ferric form. The Raman shifts for the main peaks observed in the spectra are detailed in Table 53, where subtle differences can also be appreciated between the two ligand-bound states.

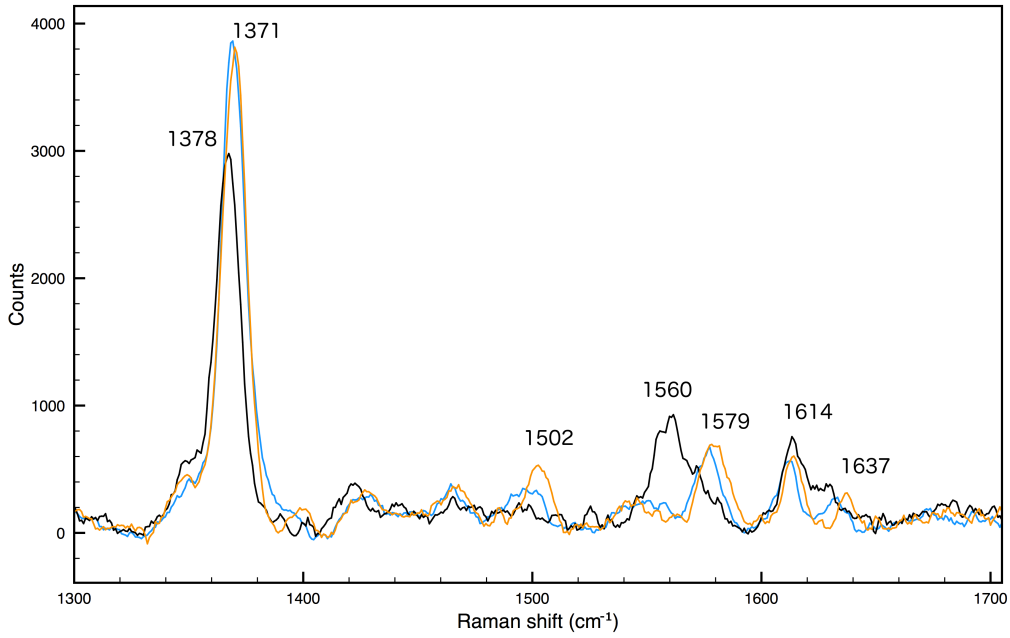


**Figure 5.8:** Single-crystal Resonance Raman (SCRR) spectra of DHP-b ferric crystals (in black), and the same crystals soaked with imidazole (in orange), and with azide (in blue). Spectra were measured with the on-axis MS3 microspectrophotometer at beamline X10SA, at Swiss Light Source (Villigen, Switzerland).

**Table 5.3:** Single-crystal Resonance Raman peaks observed in crystal and in solution for the ferric form of DHP-b, and its complexes with imidazole and azide. Solution data from Barrios et al. (2014).

| **Sample** | **Core size marker peaks ($cm^{-1}$)** | | | | |
|---|---|---|---|---|---|
| | Redox marker $\nu_4$ | $\nu_3$ | $\nu_2$ | $\nu_{c=c}$ | $\nu_{10}$ |
| Ferric solution | 1372 | 1477 | 1563/1584 | - | 1622 |
| Ferric crystal | 1368 | - | 1560 | 1614 | - |
| Imidazole-bound solution | 1374 | 1504 | 1582 | 1622 | 1641 |
| Imidazole-bound crystal | 1371 | 1502 | 1579 | 1614 | 1637 |
| Azide-bound crystal | 1369 | 1464/1497 | 1577 | 1612 | 1633 |

## 5.9 Azole-ligand DHP complexes studies.

The material presented in this section of the chapter has been published in the journal *Biochemistry* (McCombs et al. 2017). Structurally related to imidazole, azoles are a family of nitrogenated aromatic compounds that have a wide range of applications in both industry and agriculture (Giger et al., 2006; Rodríguez et al., 2010; Gaikwad et al., 2015), but little has been researched about their potential environmental impacts where they are found as contaminants (e.g. groundwater and marine ecosystems). DHP has a large distal cavity evolved to bind halogenated phenols and other large organic compounds such as indoles (Lebioda et al. 1999); and it also binds imidazole as many haem proteins do. Thus, solution experiments were performed with three other azoles apart from imidazole (Figure 5.9), to see if they would bind to DHP. An exhaustive spectroscopic characterization using UV-visible and Resonance Raman for different azole complexes in solution was carried out at Guiladi lab (North Caroline State University) by Dr. Nikolette McCombs (Dr. McCombs doctoral thesis). Spectral changes indicating binding of the compound were found for imidazole, benzimidazole and benzotriazole, as in can be observed in Figure 5.10.

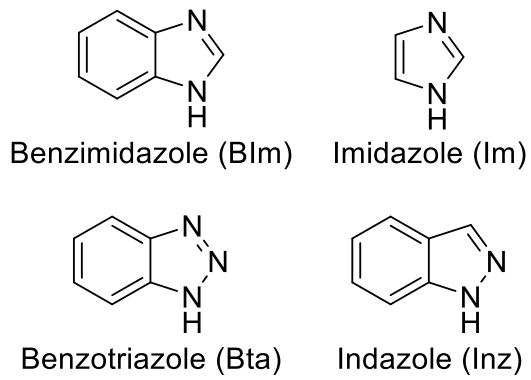


**Figure 5.9:** Azole compounds used as ligands to obtain complexes with DHP-b.

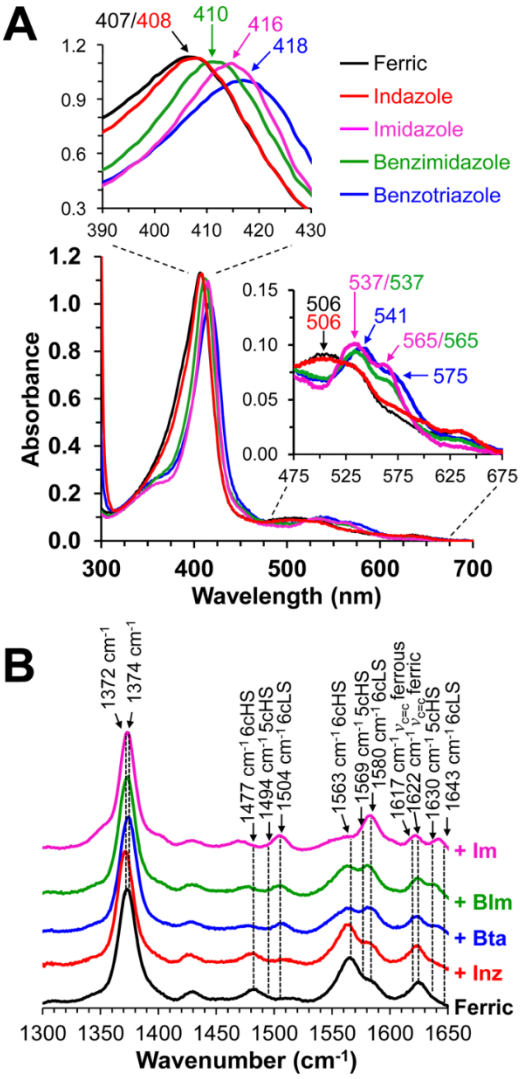


**Figure 5.10:** A) UV-vis spectra of WT ferric DHP (10 μM, black) in the presence of 100 equiv of imidazole (pink), benzimidazole (green), benzotriazole (blue), and indazole (red) in a 10% MeOH/100 mM KPi mixture (v/v) at pH 7. B) Solution resonance Raman spectra of DHP-azole complexes a in panel A at 75 μM enzyme with 50 equiv of azole. (McCombs et al., 2017)

To study the binding modes of each of the azole compounds, the X-ray crystallography structures were needed, and soaking experiments were performed with DHP crystals, as detailed in Chapter 3. The same compounds as in the solution studies were used: indazole, benzimidazole, imidazole and benzotriazole (Figure 5.9). Crystals of the different azole complexes were measured at Diamond Light Source (Oxford, UK) and SLS (Villigen, Switzerland), obtaining atomic resolution structures in each case. Data collection and refinement statistics for the complexes can be found in Table 54. All the azole ligands are found in the distal pocket of the haem (Figure 5.11), and interact directly with the haem iron with the exception of indazole. A superposition of the different heme environments for each of the ligands is shown in Figure 5.12, where the most relevant residues are included in each case. In Table 55, some important binding distances and hydrogen bonds are highlighted.

**Table 5.4:** Data collection and refinement statistics for the DHP – azole complexes measured at Swiss Light Source (SLS) and Diamond Light Source (DLS). Values in parentheses correspond to those of the last shell of diffraction.

| | **Imidazole (SLS)** | **Indazole (Diamond)** | **Benzimidazole (Diamond)** | **Benzotriazole (Diamond)** |
|---|---|---|---|---|
| Space group | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ |
| Unit cell (Å) | a=61.36, b=67.36, c=67.52; $\alpha=\beta=\gamma=90°$ | a=61.25, b=68.41, c=68.43; $\alpha=\beta=\gamma=90°$ | a=60.79, b=68.30, c=68.53; $\alpha=\beta=\gamma=90°$ | a=60.68, b=67.07, c=69.01; $\alpha=\beta=\gamma=90°$ |
| Resolution | 47.69 – 1.06 Å | 48.38 – 1.12 Å | 45.41 – 1.08 Å | 45.57 – 1.14 Å |
| $R_{meas}$ | 0.021 (0.137) | 0.072 (1.085) | 0.053 (1.591) | 0.043 (0.952) |
| Total reflections | 418946 (6143) | 610796 (27505) | 848212 (49268) | 400342 (22871) |
| Unique reflections | 117842 (3254) | 108968 (7633) | 122447 (8989) | 100913 (7178) |
| $I / \sigma(I)$ | 32.4 (6.4) | 13.3 (1.3) | 16.7 (1.2) | 17.6 (1.4) |
| $CC_{1/2}$ | 1.0 (0.976) | 1.0 (0.6) | 1.0 (0.6) | 1.0 (0.6) |
| Completeness | 93.8 (53.3) | 98.4 (94.6) | 99.8 (99.8) | 98.0 (95.3) |
| Multiplicity | 3.6 (1.9) | 5.6 (3.6) | 6.9 (5.5) | 4.0 (3.2) |
| Rwork | 13.27% | 14.67% | 16.20% | 15.17% |
| Rfree | 14.27% | 17.12% | 18.96% | 17.24% |

| | | | | |
|---|---|---|---|---|
| Rmsd bond lengths | 0.0085 Å | 0.0152 Å | 0.0140 Å | 0.0146 Å |
| Rmsd bond angles | 1.3622 ° | 1.6317 ° | 1.6426 ° | 1.5572 ° |
| Rmsd chiral volume | 0.1277 $Å^3$ | 0.1711 $Å^3$ | 0.1286 $Å^3$ | 0.1310 $Å^3$ |
| Ramachandran favoured | 97.94% | 98.33% | 97.66% | 97.47% |
| Absorbed doses (MGy) | 0.32 | 0.63 | 0.48 | 0.15 |
| PDB accession code | 5lkv | 5lk9 | 5k1l | 5llz |

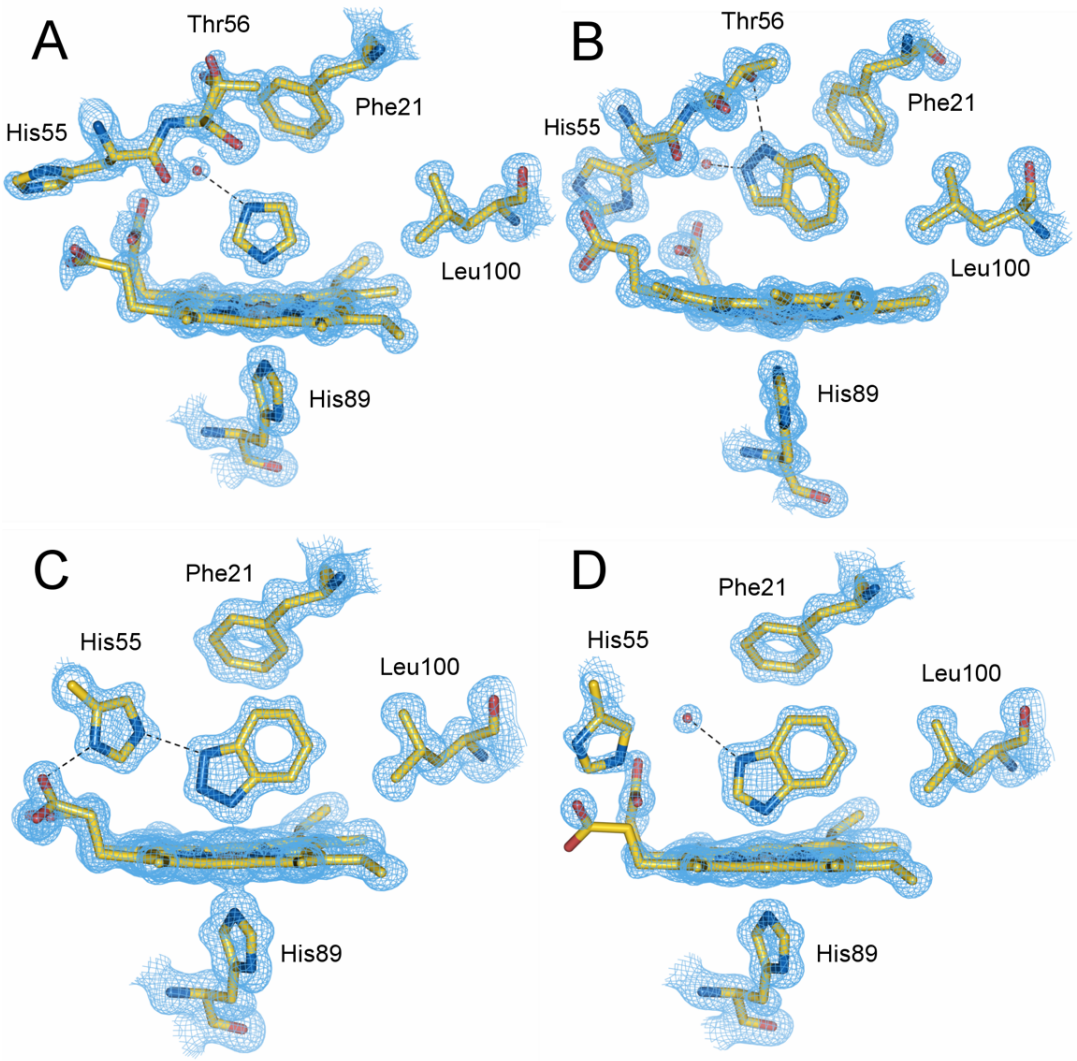


**Figure 5.11:** Binding modes and distal pocket environments, showing the most relevant surrounding residues, for the different DHP – azole complexes studied: A) Imidazole, B) Indazole, C) Benzotriazole, and D) Benzimidazole. All electron densities are contoured at 1$\sigma$. Hydrogen bonds represented as dashed lines. (McCombs et al. 2017)

In the DHP-Imidazole complex (Figure 5.11A), the ligand is found in both active sites of the dimer in the asymmetric unit. It forms a hydrogen bond with

a water molecule present in the pocket and coordinates the iron. In its turn, His55 swings out of the distal pocket when imidazole binds. For the DHP-Bta complex (Figure 5.11C), the benzotriazole molecule interacts directly with the iron at the haem distal pocket, and its stabilized by forming a hydrogen bond with His55, which swings in into the pocket in this case. Bta is just found at one of the active sites (heme B), while at the other one we observe a hemichrome species with the His55 interacting with the iron and forming a hexacoordinated complex. In the DHP-BIm complex (Figure 5.11D), benzimidazole also complexes the heme iron in the axial position resulting in a hexacoordinate species. It also forms a hydrogen bond with a water present in the pocket. Benzimidazole is found in both active sites with good occupancy. Finally, in the DHP-Inz complex (Figure 5.11B), the indazole ligand is found in the distal pocket but it does not complex the heme iron. Instead, it is bound higher in the pocket, more distant from the heme plane than the other ligands and presents hydrogen bond interactions with Thr56 and a neighbouring water molecule. It also presents hydrophobic interactions with the aromatic ring of Phe21. Indazole is present in both active sites of the DHP dimer but with lower occupancy in one of them (heme A). For all the double ring aromatic ligands, residues Phe21, Phe35, Val59, and Leu100, provide a hydrophobic environment in the pocket to favour their binding. In Figure 5.12, a superposition of the different binding modes for the azole-ligands can be observed. His55 presents a notable conformational flexibility depending on which azole is bound at the pocket.

**Table 5.5:** Binding distances and hydrogen bonds for the different azole – DHP-b complexes.

| Complex | Fe-ligand distance (Å) | Hydrogen bonds (Å) | | | |
|---|---|---|---|---|---|
| | | $H_2O$-ligand | $H_2O$-propionate | $H_2O$-Tyr38 | Ligand-His55 |
| Imidazole | 2.05 | 2.64 | 2.89 | - | - |
| Indazole | 3.95 | 2.63 | 2.64 | 2.47 | - |
| Benzotriazole | 2.02 | - | - | - | 2.59 |
| Benzimidazole | 2.04 | 2.79 | 2.91 | - | - |

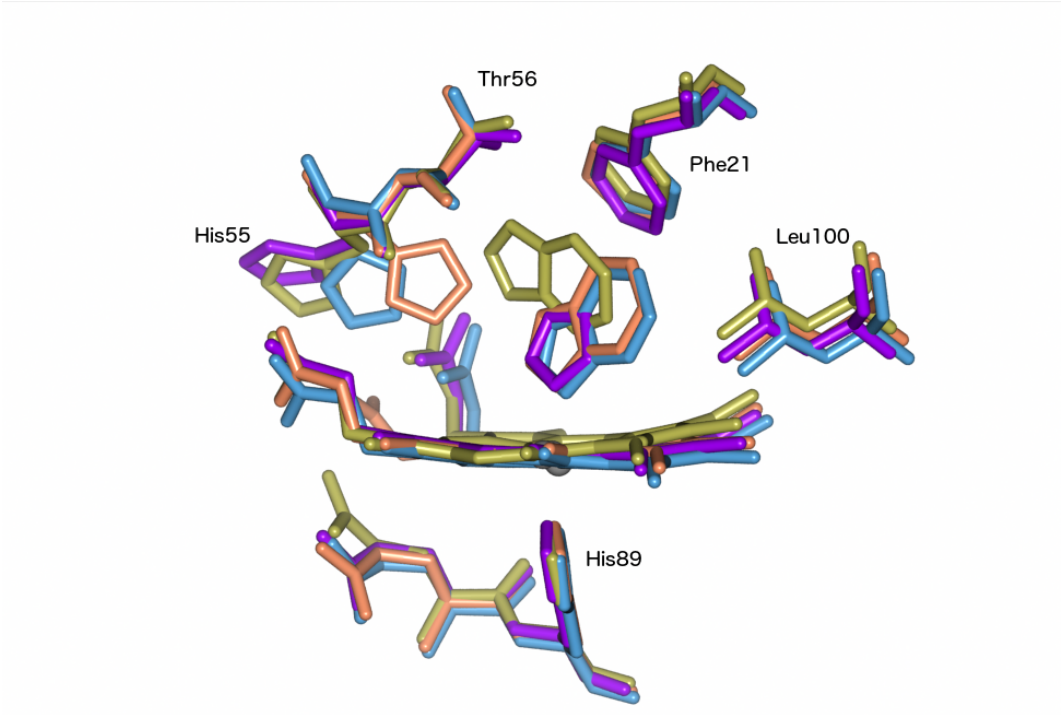


**Figure 5.12:** Superposition of the different binding environments around the heme active site (monomer B) for the different DHP-B azole complexes. The most relevant residues are included. Note the conformational changes of some of the residues, especially the catalytic conserved residue His55, depending on which ligand is bound at the distal pocket. Imidazole complex shown in purple, indazole in gold, benzotriazole in orange, and benzimidazole in blue. (McCombs et al. 2017)

## 5.10 Single crystal Resonance Raman (SCRR) measurements of the azole complexes.

The DHP-b azole-complexes were measured firstly at Diamond in conventional remote beamtime to assess if the soaking experiments had been successful. Once this was verified, crystal complexes soaked in the same conditions were prepared to be measured at Swiss Light Source (SLS), were they would be inspected with SCRR before and after X-ray diffraction using the on-axis micro-spectrophotometer MS3 equipped at beamline X10SA. The usual procedure of spectroscopy → dataset → spectroscopy was followed, as detailed in Chapter 3. This approach would allow us to fingerprint each one of the complexes, and at the same time assess the oxidation state of the protein, and whether it had changed after the diffraction experiment. For comparison, we had the RR solution data for each complex (Figure 5.10B; McCombs et al., 2017), measured at Dr. Guiladi lab in North Carolina. The acquired Single Crystal Resonance Raman (SCRR) spectra for each of the azole complexes, before (black) and after (orange) X-ray diffraction, are shown in Figure 5.13. The strongest peak in each case, the redox marker $\nu_4$, is initially in Raman shift values corresponding to the ferric form (around 1370 $cm^{-1}$), and shifts to lower values (around 1350 $cm^{1}$) upon photo-reduction of the heme during X-ray diffraction. In the case of the indazole complex, it is especially prone to reduction and the protein was becoming reduced from the excitation laser while measuring the SCRR spectrum. This fact can be observed in the resulting spectrum before X-ray diffraction (in black), were the redox marker is split in

two, since is populated by ferric and ferrous signal, and becomes further reduced upon X-ray data collection.

For all the complexes, core size markers can be also observed, and in some cases subtle changes are apparent upon reduction. In Table ***56***, a detailed comparison between solution and crystal Resonance Raman data is shown. Differences in core size marker peaks could be explained in cases were just one of the active sites is occupied by the ligand with full occupancy (benzotriazole and indazole), since there is a contribution to the spectrum of non-ligated species, such as the hemichrome form found in one of the active sites in the benzotriazole complex.

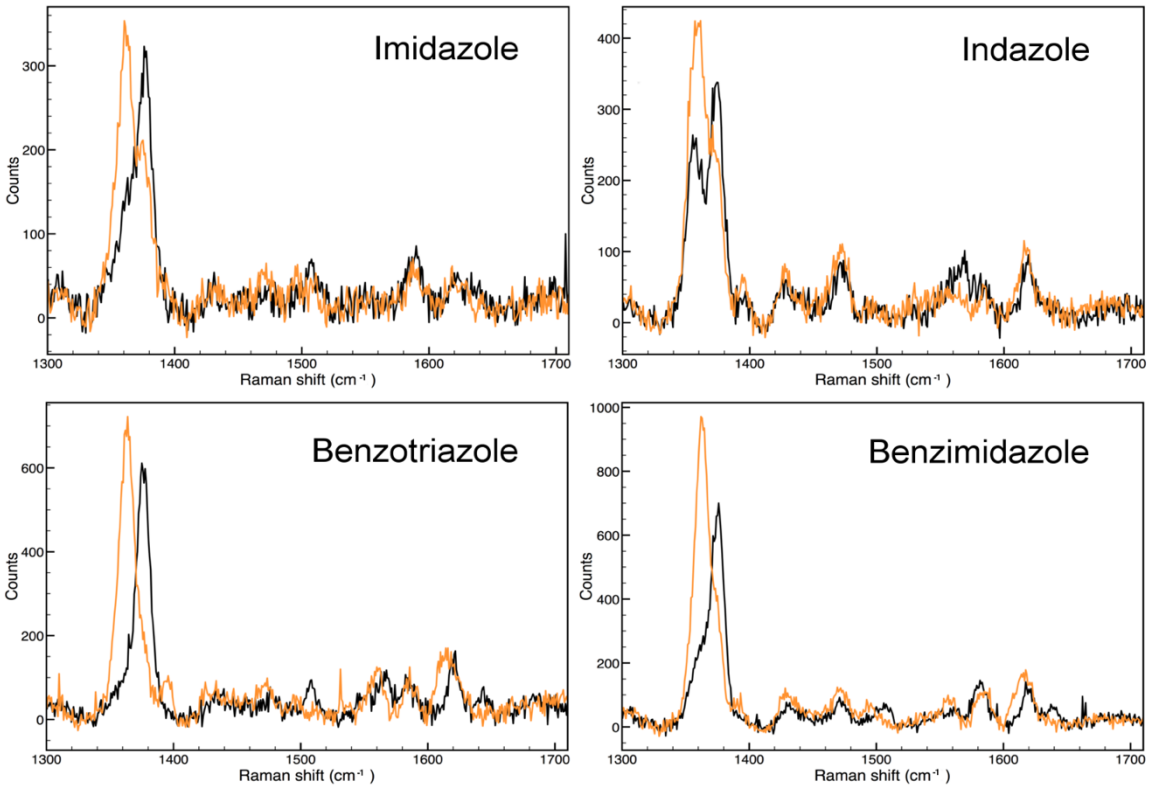

**Figure 5.13:** Single crystal Resonance Raman (SCRR) for each of the DHP-b azole complexes, measured with the on-axis microspectrophotometer at beamline X10SA, SLS (Villigen, Switzerland). In each of the panels, a first measurement before any exposure to X-rays is shown in black, and a measurement after the collection of a dataset is shown in orange. (McCombs et al., 2017)

**Table 5.6:** Observed Resonance Raman peaks for the azole – DHP-b complexes in crystal and in solution.

| | **Redox marker $\nu_4$ ($cm^{-1}$)** | **Core size marker peaks ($cm^{-1}$)** | | | |
|---|---|---|---|---|---|
| | | $\nu_3$ | $\nu_2$ | $\nu_{c=c}$ | $\nu_{10}$ |
| Ferric (solution) | 1374 | 1477 | 1563/1580 | 1622 | - |
| Ferric (crystal) | 1367 | - | 1560 | 1613 | - |
| Imidazole (solution) | 1374 | 1504 | 1580 | 1620 | 1637 |
| Imidazole (crystal) | 1371 | 1504 | 1577 | 1615 | 1636 |
| Benzimid. (solution) | 1374 | 1504 | 1560/1575 | 1620 | 1635 |
| Benzimid. (crystal) | 1358/1366 | 1466 | 1576 | 1612 | 1630 |
| Benzotriazole(solution) | 1374 | 1504 | 1563/1580 | 1620 | - |
| Benzotriazole (crystal) | 1351/1366 | 1466 | 1560 | 1613 | - |
| Indazole (solution) | 1372 | 1477 | 1560/1580 | 1620 | - |
| Indazole (crystal) | 1354 | 1465 | - | 1610 | - |

## 5.11 MSOX with DHP ferric crystal

Following the Multiple Structures from One Crystal (MSOX) approach, described by Horrell et al. (2016), a series of eight consecutive structures from one DHP ferric crystal were acquired at SLS. In our case, MSOX was coupled with spectroscopic inspection of the crystal with SCRR to obtain information about the state of the protein within the crystal. The spectra were measured all at the same point of the crystal, around which the rotation range for X-ray data collection was set up, to ensure that the part of the crystal exposed to the beam was also probed by spectroscopy. The consecutive SCRR spectra measured

are shown in Figure 5.14. The first of them (in black) was measured from a ferric DHP crystal before exposure to X-rays, and the rest were measured in succession after every collected dataset. After the first dataset, RR shows the splitting of the $\nu_4$ redox marker, with the ferrous peak at 1350 $cm^{-1}$ being predominant. It stays like this for 4 more datasets, and from that point the signal at 1370 $cm^{-1}$ becomes predominant. Core size markers experienced a substantial change after photo-reduction from collection of the first dataset, and remained like that for the rest of the MSOX series. Dataset 1, 4 and 8 were inspected to have a structure from the initial, middle and final stage of the series. Data collection and refinement statistics for them are detailed in Table 57. Haem environments for both monomers of the dimer in the asymmetric unit from each structure are represented in Figure 5.15. The hemichrome is present at haem B from the beginning of the series, although it appears to slightly increase through the process. At haem A, the development of electron density can be monitored at the distal site, indicating the partial occupancy of an oxy-ferrous species or a distal ligand bound to the photo-reduced iron.

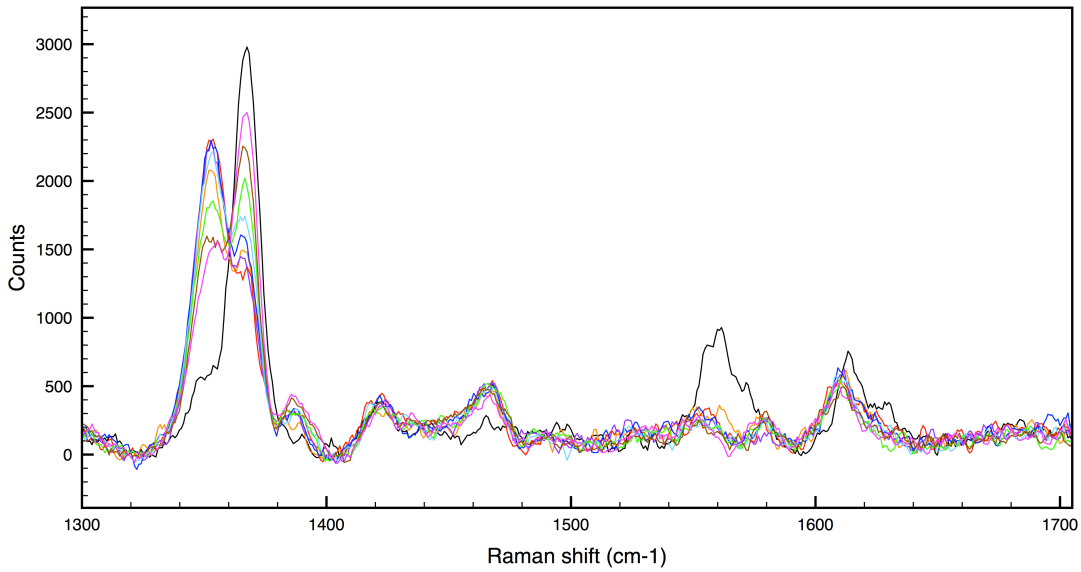


**Figure 5.14:** Single Crystal Resonance Raman (SCRR) spectra collected in succession from a DHP crystal. Initially the crystal was in its ferric state and a spectrum was acquired before any exposure to X-rays (in black). The following spectra were measured each of them after the collection of a consecutive series of eight X-ray diffraction datasets. The succession in which spectra were measured is: black, orange, red, purple, blue, light blue, green, brown, magenta. All measurements were carried out with the MS3 microspectrophotometer at beamline X10SA, SLS (Villigen, Switzerland).

**Table 5.7:** Data quality and refinement statistics for datasets 1, 4 and 8 for the MSOX series (total of eight successive datasets) measured from a ferric DHP crystal at beamline X10SA (SLS), using an X-ray wavelength of 0.8 Å.

| | **Dataset 1** | **Dataset 4** | **Dataset 8** |
|---|---|---|---|
| Space group | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ |
| Unit cell (Å) | a=60.98, b=67.42, c=67.59 | a=61.02, b=67.45, c=67.63 | a=61.14, b=67.46, c=67.70 |
| Resolution (Å) | 47.73 – 1.07 (1.09 – 1.07) | 47.76 – 1.07 (1.08 – 1.07) | 47.79 – 1.06 (1.08 – 1.06) |
| $R_{merge}$ | 0.050 (0.094) | 0.049 (0.106) | 0.050 (0.427) |
| Unique reflections | 118,921 (4,350) | 119,410 (3,939) | 119,842 (3,249) |
| I / σ(I) | 14.5 (5.3) | 16.3 (4.8) | 13.2 (1.9) |
| $CC_{1/2}$ | 0.995 (0.964) | 0.995 (0.955) | 0.997 (0.712) |
| Completeness (%) | 96.2 (72.4) | 95.8 (64.9) | 95.2 (53.3) |
| Multiplicity | 3.8 (1.8) | 3.8 (1.8) | 3.8 (1.7) |
| Rwork (%) | 12.8 | 12.9 | 13.9 |
| Rfree (%) | 14.1 | 14.2 | 15.6 |
| Rmsd bond lengths (Å) | 0.018 | 0.016 | 0.015 |
| Rmsd bond angles (°) | 1.95 | 1.89 | 1.85 |
| Ramachandran favour (%) | 98.4 | 98.4 | 98.4 |
| Accumulated dose (MGy) | 0.45 | 1.8 | 3.6 |

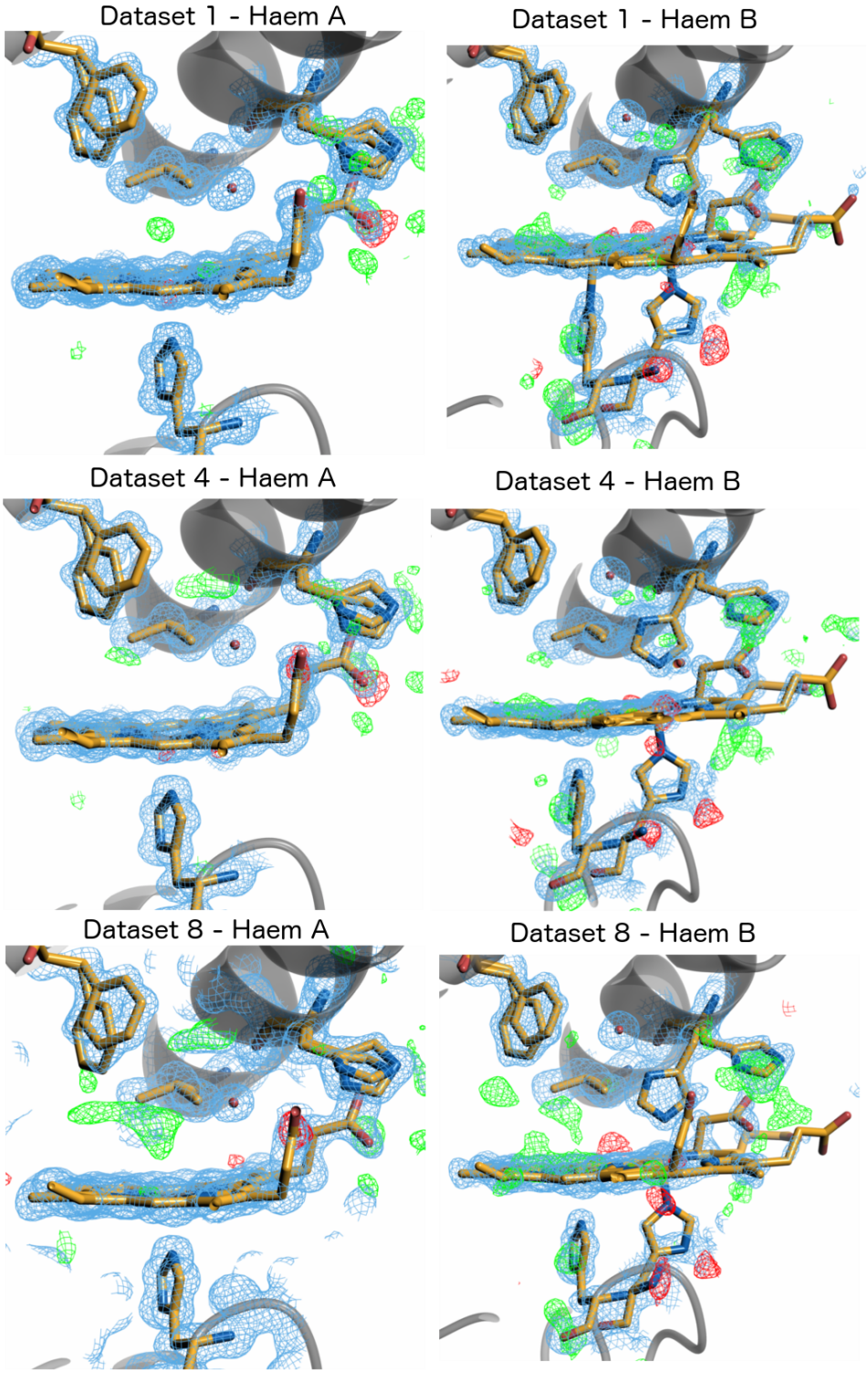


**Figure 5.15:** Haem sites of monomers A and B for the datasets 1, 4 and 8 of the MSOX series (total of eight datasets) measured from a ferric DHP crystal at beamline X10SA (SLS). Electron density from the $2F_o - F_c$ map contoured at $1\sigma$, and the difference $F_o - F_c$ map is contoured at $3\sigma$.

## 5.12 Discussion

In this chapter, we have carried out a structural and spectroscopic characterization of DHP and its complexes with different ligands, learning more about this protein system and its functioning. The spectroscopies used *in crystallo* have been UV-vis and resonance Raman (RR), the latter being the most suitable for the optically dense DHP crystals. Both techniques were found useful to monitor the photo-reduction of the protein by the X-ray beam, showing clear spectral changes: shift of the $\nu_4$ redox marker to ferrous values in RR, and shift of the Soret band to a higher wavelength in UV-vis. Single crystal resonance Raman (SCRR) consistently delivered good quality spectra for this system and, apart from the oxidation state of the protein, allowed to fingerprint different complexes with ligands (e.g. azide, azoles) monitoring differences in the core size markers of the RR spectra (region 1400-1700 $cm^{-1}$). Even if not extensively explored in this work, the use of UV-vis spectroscopy with DHP crystals could be useful, for instance, to identify and validate for instance intermediate states in the peroxidase cycle, provided that the crystals are small enough to at least not saturate the spectral region (500-700 nm), where the $\alpha/\beta$ and CT bands appear. This was successfully tested with small DHP ferric crystals, where clear spectral changes were observed upon activation of the peroxidase cycle by soaking the crystals in 50 mM $H_2O_2$ (shown in Figure 5.16).

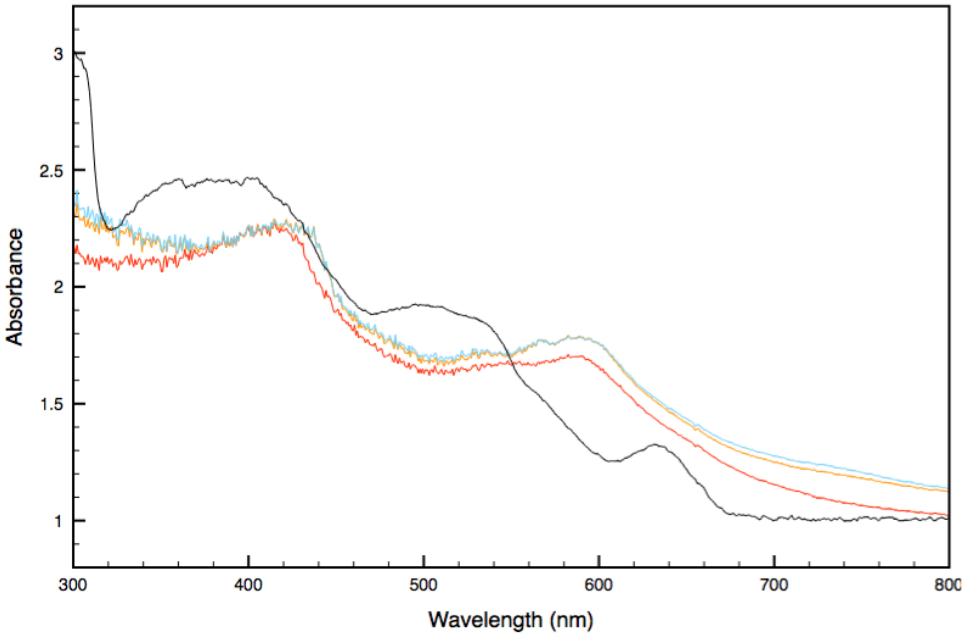


**Figure 5.16:** Single-crystal UV-vis spectra measured with the online micro-spectrophotometer installed at beamline BM30 (ESRF, Grenoble). In black is shown the spectrum of a ferric DHP crystal, which presents a substantially different profile from a $H_2O_2$-treated ferric DHP crystal (in red). Additional spectra of this same $H_2O_2$-treated crystal were measured after acquisition of one diffraction dataset (in orange), and a second one (in light blue). Note the subtle differences of the UV-vis profile after X-ray exposure, such as a shift in the Soret band (around 420 nm).

The structure of DHP-B was determined at an atomic resolution of 0.99 Å, resulting in a substantial improvement compared to previously published structure, which presented a resolution of 1.58 Å (De Serrano et al. 2010). This permitted the appreciation of the structure at a greater detail and a new feature was observed: a double conformation of the heme cofactor at one of the monomers, where a hemichrome species could be observed. In this species, the heme iron is hexacoordinated by His89 at the proximal and His55 at the distal site. It is thought to be the result of a partial denaturation of the protein surrounding the heme due to the presence of glycerol (cryo-protection) and polyethylene glycol (crystallization condition) in the crystal, as investigated by

Molchanova (1981). Another possibility could be that it is formed as a result of photo-reduction of the protein. To confirm this hypothesis, an intact structure of the DHP-b resting state would be needed, which could be obtained for instance by means of a low-dose composite dataset approach, or measuring an SFX damage-free structure at an XFEL. Therefore, the presence and possible role of the hemichrome found in DHP-b structure are not completely clear, although it is likely to be an artefact due to the crystallization conditions. However, if its presence could be confirmed in solution (e.g. spectroscopic means), it could have a functional role within a biological context, which would make it of much more interest to further study it.

As a model to study the interaction of Permanent Organic Pollutants (POPs) with haemproteins, complexes between DHP and azole compounds were studied. The X-ray crystallography structures at atomic resolution coupled with in situ resonance raman characterization were performed for: imidazole, indazole, benzimidazole, and benzotriazole. Two binding modes were observed for the selected azoles, where Im, BIm and Bta would bind directly to the heme iron, and indazole would bind higher up in the distal pocket, establishing a hydrogen bond with Thr56 and a hydrophobic interaction with Phe21. The importance of the conserved His55 in the catalytic activity of DHP could be appreciated in the conformational flexibility that this residue presents depending on the ligand that is bound in the pocket. ‘Open’ and ‘closed’ conformations were observed as well as intermediate states. This residue is likely to have an important role in the process of switching functions at the DHP active site. Since one of these functions is that of oxygen carrier, the presence

of bound compounds such as azoles, could potentially disrupt the haemoglobin function, ultimately affecting the normal functioning of, in this case, the worm *A.ornata*. All the azole complexes were spectroscopically characterized by single crystal resonance Raman (SCRR), each one presenting an spectral fingerprint in the core size markers region (1400-1700 $cm^{-1}$). This technique has been successfully applied by Kekilli et al. (2014) to fingerprint ligand and oxidation states in cytochrome c', and it has been shown for DHP how could be applied for the identification for example of catalytic intermediates. The protein was found to be in the ferric state for all the azole complexes as determined from the position of the redox marker $\nu_4$ at ~1370 $cm^{-1}$. In all cases, DHP would become photo-reduced after data collection, with $\nu_4$ shifting to ~1350 $cm^{-1}$. The indazole complex presented a special behaviour; the protein would become reduced by the activation laser while measuring the RR spectrum. Thus, in this case the heme had a lower threshold energy to become photo-reduced. Studies of laser photo-reduction of haemoproteins was performed by Kekilli and co-workers (2017), who proposed the existence of this energy threshold below which the protein remains unaffected. Since indazole is the only compound which does not bind directly to the iron, but binds above the heme plane at the distal pocket, it could be that leaving a more exposed iron results in it being more prone to photo-reduction. Nevertheless, no laser photo-reduction is observed when performing SCRR in ferric crystals, where no ligand is bound. The selected input power was of 50 mW for all the SCRR measurements, was measured as beain ~1.0 mW at sample (hand-held power meter), which has been assessed as a 'safe' power for other haemproteins (Kekilli et al. 2017). Regarding the role of the azole compounds, for the ones

binding directly to the haem iron (i.e. imidazole, benzotriazole and benzimidazole), they are not expected to be substrates, which are normally found in binding conformations higher in the distal pocket above the haem (Zhao et al., 2013). The haem-bound azoles would prevent binding of $H_2O_2$ and/or $O_2$ and potentially have an effect in the peroxidase and haemoglobin functions. Indole instead, could be metabolized by DHP as many other aromatic compounds such as indoles (Barrios et al. 2014), since there would be room for catalysis at the distal haem site.

Finally, a series of eight consecutive structures were measured from a DHP ferric crystal applying the MSOX approach (Horrell et al., 2016), in this case coupled with SCRR measurements in-between data collections to assess the state of the protein. Clear signs of specific radiation damage can be observed in the last structures in the series, for examples with the decarboxylation of Glu and Asp residues. Along the series, an oxy-ferrous species appears to be developing at one of the DHP heme sites due to photo-reduction. This species is not formed with the dose applied for measurement of a complete dataset, in contrast with DtpA (refer to Chapter 4), where the oxy-ferrous complex is formed rapidly just after 10-20 kGy of dose. The hemichrome species is present at heme B since the first collected dataset, and does not seem to change substantially along the series. If its formation is due to photo-reduction it would have happened during the measurement of the first structure.

# Chapter 6 DHP serial crystallography.

Serial crystallography (SX) using DHP microcrystals in silicon nitride 'chips' (Mueller et al. 2015) has been performed at X-ray sources with different characteristics: synchrotron, Laue (X-ray polychromatic beam), and XFEL beamlines. As discussed in Chapter 2, to apply this methodology it was needed to develop batch crystallization conditions which delivered thousands of microcrystals. A considerable effort was put into finding and optimizing the right conditions. Once they were found, this protein system proved to be consistent and delivered SX results at every X-ray source where it was applied. Ligand soaking using microcrystals was also demonstrated, obtaining damage-free SFX structures of the complexes in a sample and time efficient manner.

## 6.1 Production of DHP microcrystals.

### *6.1.1 Initial trials and optimization*

The production of DHP microcrystals in batch conditions was initially developed in preparation for the Laue beamline at APS, which was the first experiment involving SX. The initial condition to be optimized was that used for crystallization of large DHP crystals in drops: 200mM ammonium sulphate, 20% PEG 4000, grown at 4ºC. No crystals appeared at the beginning, using similar conditions to those in the crystallization trays. What it was found for this protein is that it would only crystallize at high PEG concentration, and thus it was not until the ratio of protein to condition (refer to Chapter 3) was increased (while at the same time increasing DHP concentration of the protein fraction)

that we started to observe crystallization in the batches. At this point, by keeping the working protein to condition ratios, and increasing the protein concentration of the batch, smaller crystals were obtained. In Table 61, some examples of initially set up batches are shown, specifying the outcome observed after several days (3-10 days). Seeding was also tried to obtain microcrystals with DHP, by vortexing one of the batches with big crystals with a seed bead (Molecular Dimensions) to obtain a seed stock. After set up, DHP batches would be seeded with up to 5% of volume of the seed stock, and after 4-6 days some microcrystals would appear in some cases, in conditions where bigger crystals had appeared previously. Due to the long time (1 to 2 weeks) needed for the microcrystals to grow and the unreliability of the method, seeding was not further optimized for this protein system.

**Table 6.1:** Selection of initial batch trials set up with DHP to obtain the conditions to produce microcrystals.

| Condition | Batch final concentration | | Batch set-up | | | Seeded? | Outcome |
|---|---|---|---|---|---|---|---|
| | % PEG | $AmSO_4$ (mM) | Protein (mg/ml) | Ratio prot:cond | Total volume (µl) | | |
| **DHP-1** | 15 | 100 | 6.65 | 1 to 1 | 50 | No | No crystals |
| **DHP-2** | 15 | 100 | 13.5 | 1 to 1 | 50 | No | No crystals |
| **DHP-4** | 15 | 200 | 6.65 | 1 to 1 | 50 | No | No crystals |
| **DHP-5** | 22.5 | 100 | 6.65 | 1 to 1 | 50 | After some days | Large crystals after seeding. |
| **DHP-11** | 30 | 100 | 6.75 | 1 to 3 | 50 | No | Lots of large/medium sized crystals |
| **DHP-13** | 35 | 200 | 5.4 | 1 to 4 | 50 | No | Lots of medium sized crystals. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **DHP-15** | 37.8 | 200 | 3.86 | 1 to 6 | 70 | No | Lots of ~50μm crystals. |
| **DHP-26** | 36.4 | 200 | 5.0 | 1 to 10 | 110 | Yes | Small and micro-crystals (<50μm) |
| **DHP-27** | 32.5 | 200 | 3.75 | 1 to 3 | 50 | Yes | Small and micro-crystals (<50μm) |

### *6.1.2 Final optimized batch condition.*

The best batch condition delivered thousands of DHP microcrystals of 15-35 μm in size in a single batch of 400-500 μl. They were grown at 4ºC within 5-8 days. Although the first crystals would appear after 1 day, the equilibration of the batch solution would take several days, while the protein would keep crystallizing and producing more microcrystals. The details of this batch condition were:

- Protein stock at ~30 mg/ml, in buffer 20mM MES pH 6.5
- Ratios protein to condition: 1:4 or 1:3
- PEG-4000 % in the precipitant fraction as high as possible.

In Table 62, two examples are presented of optimized final batches for DHP. During set up, batches were kept always on ice, and they were vortexed at the end for ~30 seconds to homogenize the solution, which in this case was highly viscous due to the high percentage of PEG 4000 in it. Through the experience of preparing batches with different DHP and crystallization stocks, it was noted that the freshness of the PEG 4000 was a crucial factor to obtain microcrystals. Even if the PEG solution was kept in the fridge, just after some weeks would

not deliver microcrystals in the same way. What would happen is that a notable precipitation could be observed immediately after set up, followed by formation of 100-200 μm crystals after several days at 4ºC. Thus, fresh PEG was prepared to crystallize each protein preparation of DHP. In Figure 6.1, two examples of DHP protein batches containing microcrystals are shown.

**Table 6.2:** Optimized final DHP batches which delivered microcrystals taken to serial crystallography experiments.

| | **DHP-I** | **DHP-II** |
|---|---|---|
| Ratio protein to condition | 1 to 3 | 1 to 4 |
| Total volume of batch (μl) | 400 | 500 |
| Protein conc. in batch (mg/ml) | 7.5 | 6.0 |
| PEG-4000 % in batch | 35.0 | 36.7 |
| $AmSO_4$ conc. in batch (mM) | 200 | 200 |
| Mixing method | Vortexing | Vortexing |
| Crystallization temp. (ºC) | 4 | 4 |

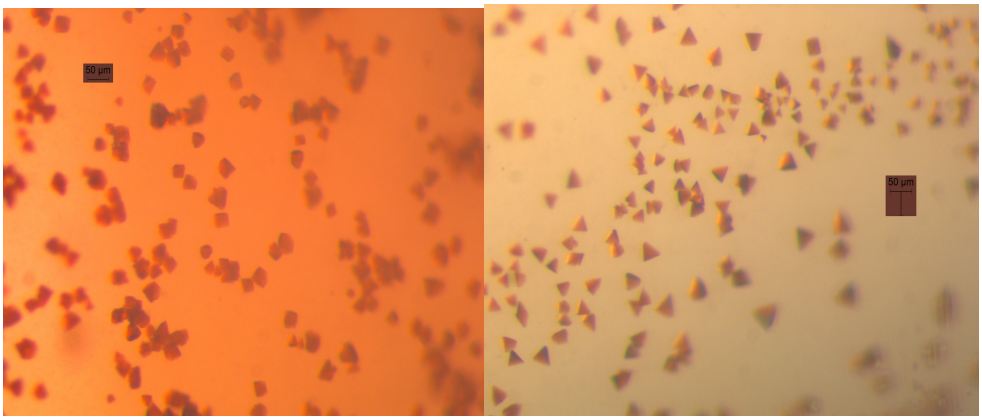


**Figure 6.1:** Images of DHP batches containing microcrystals of sizes ranging 20-40 μm, prepared using a ratio 1:4 (protein to condition). Final concentration in both cases were approximately: 6 mg/ml, 35 % PEG 4000 and 200 mM Ammonium sulphate. Note that the condition at the top is still crystallising and has a considerable amount of DHP in solution.

## 6.2 Synchrotron serial crystallography (SSX) at Diamond.

DHP microcrystals were taken to beamline I24 at Diamond (Oxford, UK), to be applied in the serial crystallography system, using silicon nitride chips, developed by Owen and co-workers (2017), capable of obtaining low-dose structures of 10-50 kGy using diffraction data from several thousands of crystals. At the initial SSX beamtime, diffraction of the crystals was assessed, as well as the hit rates, and loading of the chips was optimized. Of course, ultimately the aim was to obtain an SSX DHP structure, if the crystals were diffracting well and there were enough of them for a complete dataset. The required number of crystals was estimated to be at least several thousands. All the information obtained in these SSX experiments would be valuable and feed our XFEL beamtime at SACLA, since the experimental system (using chips and end station) was going to be the same. One of the aims, indeed, was to compare SSX with SFX structures and see if any differences could be observed due to radiation damage, even at low doses.

DHP microcrystals of sizes ranging 30-50 microns were loaded in a humidity controlled hood (refer to Chapter 3), and inspected at its 25,600 positions following the screening procedure detailed in Chapter 3. In total, diffraction still frames from more than 20,000 crystals were obtained. Five chips were measured for ferric DHP, and detailed parameters of the measurement for each of them are shown in Table 63. In the first visit to I24 (first three chips), hit rates were low, and the conditions for microcrystal production and loading

of the chips were optimized to substantially increase the hit rates in the second visit (last two chips: iborra and lukaku).

**Table 6.3:** Experimental parameters and statistics for chips loaded with ferric DHP microcrystals measured at beamline I24, Diamond (Oxford, UK), at 12800 eV energy.

| Name of chip | Beam size ($\mu m^2$) | Detector distance (mm) | Exp. time (ms) | Trans. (%) | City blocks | Measured frames | Indexed by Dials | Hit rate (%) | Res. Crystals (Å) |
|---|---|---|---|---|---|---|---|---|---|
| Geraldo | 9x6 | 500 | 25 | 50 | Full chip | 25600 | 563 | 2.2 | 1.80 |
| Hillary | 9x6 | 350 | 25 | 50 | Full chip | 25600 | 1694 | 6.6 | 1.35 |
| Joffrey | 9x6 | 350 | 25 | 50 | Column A, B, C, D | 12800 | 966 | 7.5 | 1.35 |
| Joffrey2 | 9x6 | 350 | 25 | 50 | Columns E, F, G, H | 12800 | 921 | 7.2 | 1.35 |
| Iborra | 8x8 | 310 | 20 | 100 | Full chip | 12800 | 7984 | 31.2 | 1.70 |
| Lukaku | 8x8 | 310 | 20 | 100 | Full chip | 12800 | 8131 | 31.8 | 1.50 |

Fixed targets with the smaller aperture, i.e. 7μm, were used for all measurements and the beam was made as small as possible in size to tailor it to the apertures. From these initial experiments, it was noted that hit rates were not very high (around 5-7%), and therefore ideally it was needed to further optimize the production of DHP microcrystals, to obtain them in smaller sizes, which would result in a larger number of crystals per batch and in this way hit rates would increase and we would minimize the number of chips needed to obtain a complete dataset. Detector distance was initially set up at 500mm for lower resolution, and after observing that the microcrystals were diffracting at high resolution, the detector was moved closer, to 350 mm (around 1.5 Å). The measurement of the last prepared chip (Joffrey) was carried out in two different

runs, each measuring a half of the chip. The first half (columns A, B, C, D) was diffracted at the room temperature at the beamline at diamond. The second half (columns E, F, G, H), was diffracted after incubating the chip at 30ºC for 30 min, to simulate the temperature conditions at the hutch at SACLA (27ºC), and in this way see if the DHP microcrystals might be affected by this conditions. A similar number of indexed patterns were found in each of the halves, which were processed separately, and the resolution reached by the crystals did not decrease after being exposed to 30°C, which suggested the viability of performing this same experiment with success at SACLA.

### *6.2.1 Processing of SSX chip data for DHP*

The generated data from DHP crystals diffracted at I24 using SSX were processed using the software *DIALS* (Winter et al. 2018) for indexing and the program *PRIME* (Uervirojnangkoorn et al. 2015) for scaling and merging. Once an *mtz* file was generated with all the merged data, programs within the ccp4i2 suite were used as usual: *Phaser* for molecular replacement, *Refmac5* for refinement, *Coot* for model building.

A subprogram within *DIALS* was used, called 'dials.still_process', especially developed to treat still diffraction patterns from many crystals to then merge them in a single dataset. It was run from the command line and each job was submitted to the cluster at Diamond. To run *DIALS*, two files had to be generated for the processing of each chip: 'dhp.phil' and 'chip_name_process.sh'. The 'dhp.phil' file contained information about the samples such as space group and cell parameters, geometry information of the detector, and indexing and refinement parameters to carry out the spot finding

job. In Figure 6.2, the initial part of the file dhp_process.phil, used for indexing DHP chips (Table 63) is shown.

dhp_process.phil

```
verbosity=10

#spotfinder.lookup.mask=/dls/i24/data/2017/nt14493-58/processing/stills_process/
aldous/mask.pickle
#spotfinder.filter.min_spot_size=2#integration.lookup.mask=/dls/i24/data/2017/
nt14493-63/processing/stills_process/aldous/mask.pickle

#geometry {
#  detector {
#    panel {
#      fast_axis = 1,-1.73842e-05,-0.000834956
#      slow_axis = -1.91427e-05,-0.999998,-0.00210614
#      origin = -217.245,226.774,-349.8
#      }
#    }
#  }

indexing {
  known_symmetry {
    space_group = P212121
    unit_cell = 61.34 67.39 69.17 90 90 90
  }


  refinement_protocol.n_macro_cycles = 1
  refinement_protocol.d_min_start=2.5
  basis_vector_combinations.max_refine=3
  stills.indexer=stills
  stills.method_list=fft1d real_space_grid_search
  multiple_lattice_search.max_lattices=3
}

verbosity=10
integration {
  integrator=stills
```

**Figure 6.2:** Initial part of file 'dhp_process.phil', used for indexing with *DIALS* the dhp chips. Geometry parameters of the detector, symmetry information for the DHP crystals, and indexing and refinement parameters for the job are specified.

The other file needed was the 'chip_name_process.sh', shown in Figure 6.3, which was an executable script to run the *DIALS* job in the Diamond computing cluster, specifying the path to find the diffraction images (stored as *cbf* files). The script also directs to the dhp.phil to obtain the needed information for the indexing process. The number of processors was also specified (being 20 the maximum), and an output file 'geraldo_process.out' was generated.

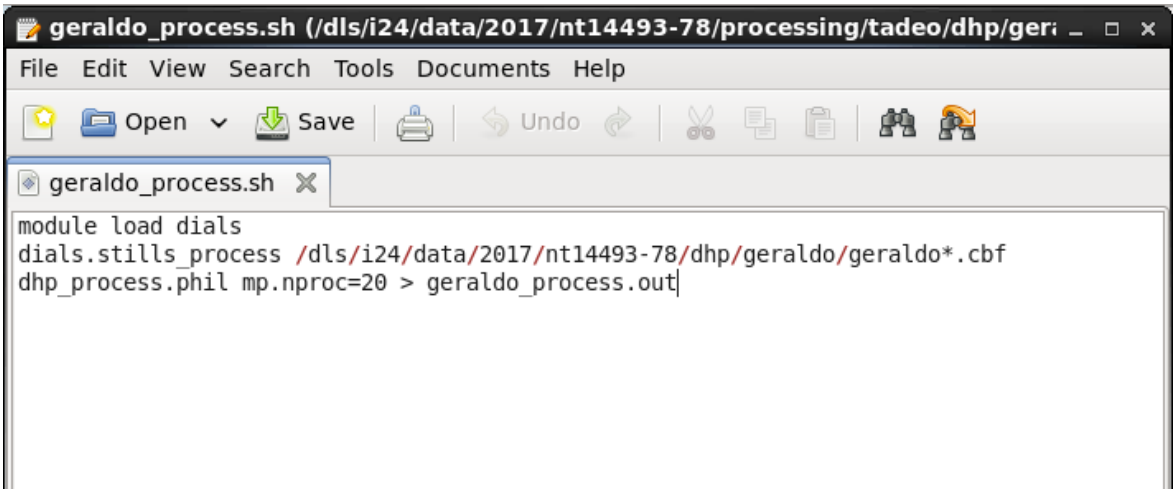


**Figure 6.3:** Script 'geraldo_process.sh' for the indexing with DIALS of the DHP-b loaded chip 'geraldo'.

To send the executable .sh file to the cluster, the following command was introduced:

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

In it, submission to the low priority queue was specified, as well as the number of processors (20 in this case), and the file to be executed in the cluster (geraldo_process.sh). Indexed and integrated files would continually appear in the working folder until the run was finished. The following and last step of the processing of the SSX data was to carry out the merging of the generated integrated files with *Prime*. For this purpose a file named 'prime.phil' had to be tailored to each specific sample. Merging parameters would be specified in this file, including the resolution range at which data should be cut, and symmetry information of the crystals. An example of 'prime.phil' file, used for merging DHP data, is shown in Figure 6.4.

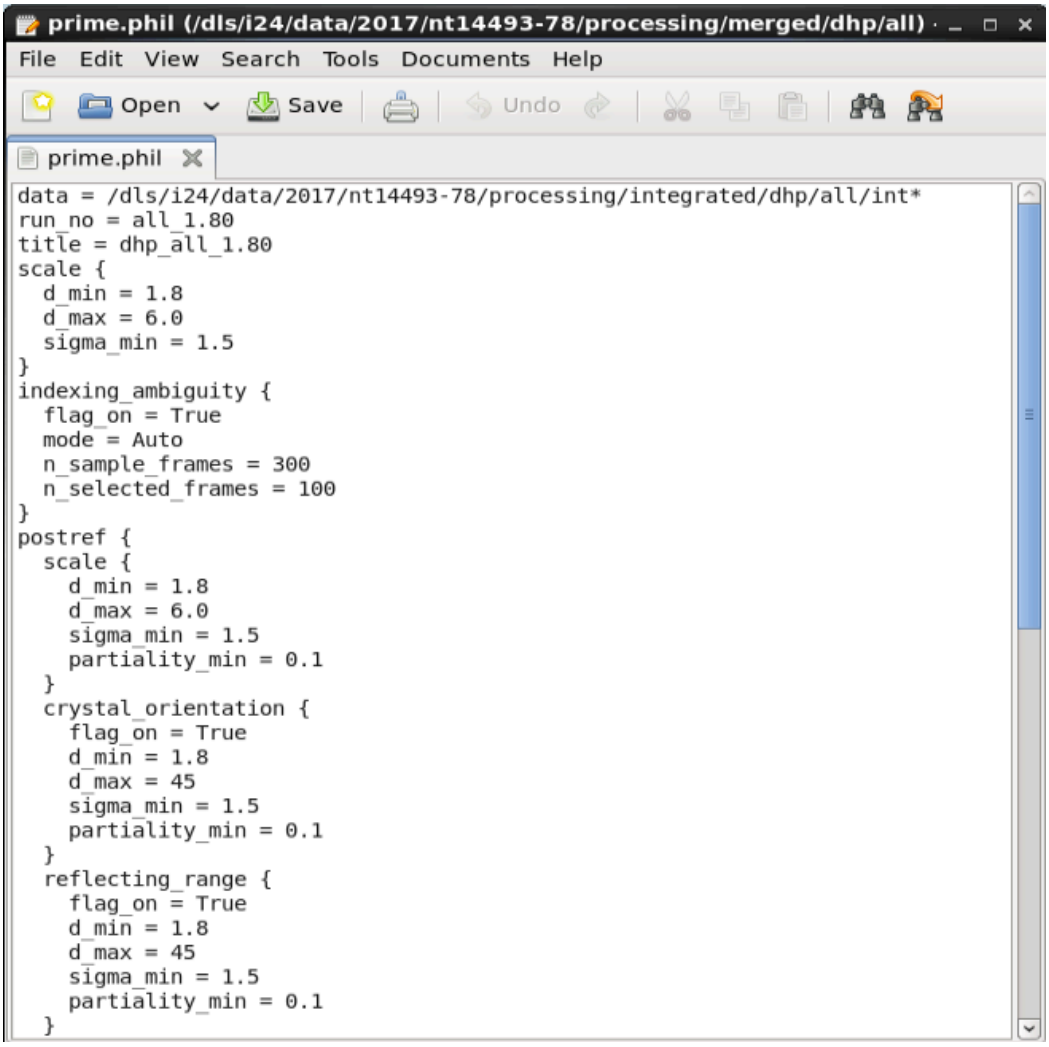


**Figure 6.4:** First half of the script 'prime.phil' used for merging DHP SSX data from the different chips.

*Prime* would be run by firstly loading DIALS in the command line, and then introducing the command and specifying the file with the merging job parameters:

```
$ module load dials
$ prime.run prime.phil
```

The merging job would generate an *mtz* file to continue with the structure solving process, as well as different statistics to assess data quality: $R_{split}$, $CC_{1/2}$, resolution, refined cell parameters, completeness. After a preliminary *Prime* run, a resolution cut-off would be decided, following the criteria of $CC_{1/2}$ > 0.5 in the last resolution shell. Cell parameters would be also adjusted if refined differently from the introduced values. In Figure 6.5, an example is shown of a post-refinement table generated while merging one of the DHP chips, where the different quality statistics for each resolution shell are specified.

Summary for lukaku_1.5/postref_cycle_3_merge.mtz

| Bin | Resolution Range | Completeness | <N_obs> | \|Rmerge | Rsplit | CC1/2 | N_ind | \|CCiso | N_ind | \|CCanoma | N_ind\| | <I/sigI> |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 01 | 37.80 - 4.07 | 100.0 | 2482 / 2483 | 230.69 | 85.87 | 8.75 | 98.01 | 2482 | 0.00 | 0 | 0.00 | 0 | 11.43 |
| 02 | 4.07 - 3.23 | 100.0 | 2370 / 2371 | 165.01 | 82.06 | 10.01 | 97.25 | 2370 | 0.00 | 0 | 0.00 | 0 | 8.93 |
| 03 | 3.23 - 2.82 | 100.0 | 2345 / 2345 | 144.20 | 83.06 | 12.19 | 95.27 | 2345 | 0.00 | 0 | 0.00 | 0 | 4.94 |
| 04 | 2.82 - 2.56 | 100.0 | 2320 / 2320 | 129.15 | 83.84 | 12.01 | 96.30 | 2320 | 0.00 | 0 | 0.00 | 0 | 3.89 |
| 05 | 2.56 - 2.38 | 100.0 | 2316 / 2316 | 118.71 | 84.78 | 13.53 | 93.57 | 2316 | 0.00 | 0 | 0.00 | 0 | 3.38 |
| 06 | 2.38 - 2.24 | 100.0 | 2305 / 2305 | 106.74 | 85.57 | 14.08 | 94.45 | 2305 | 0.00 | 0 | 0.00 | 0 | 2.80 |
| 07 | 2.24 - 2.13 | 100.0 | 2307 / 2307 | 103.93 | 86.32 | 15.50 | 93.17 | 2307 | 0.00 | 0 | 0.00 | 0 | 2.32 |
| 08 | 2.13 - 2.04 | 100.0 | 2273 / 2273 | 94.88 | 86.78 | 15.96 | 92.93 | 2273 | 0.00 | 0 | 0.00 | 0 | 1.95 |
| 09 | 2.04 - 1.96 | 100.0 | 2300 / 2300 | 88.59 | 87.87 | 16.84 | 92.15 | 2300 | 0.00 | 0 | 0.00 | 0 | 1.69 |
| 10 | 1.96 - 1.89 | 100.0 | 2300 / 2300 | 80.29 | 89.11 | 19.32 | 90.99 | 2300 | 0.00 | 0 | 0.00 | 0 | 1.38 |
| 11 | 1.89 - 1.83 | 100.0 | 2286 / 2286 | 74.10 | 89.04 | 20.81 | 89.20 | 2286 | 0.00 | 0 | 0.00 | 0 | 1.20 |
| 12 | 1.83 - 1.78 | 100.0 | 2266 / 2266 | 67.95 | 89.32 | 21.83 | 87.37 | 2266 | 0.00 | 0 | 0.00 | 0 | 1.09 |
| 13 | 1.78 - 1.73 | 100.0 | 2292 / 2292 | 58.67 | 90.86 | 24.74 | 85.61 | 2292 | 0.00 | 0 | 0.00 | 0 | 0.96 |
| 14 | 1.73 - 1.69 | 100.0 | 2280 / 2280 | 51.55 | 91.07 | 26.97 | 83.09 | 2280 | 0.00 | 0 | 0.00 | 0 | 0.92 |
| 15 | 1.69 - 1.65 | 100.0 | 2248 / 2248 | 41.87 | 92.42 | 32.68 | 79.47 | 2248 | 0.00 | 0 | 0.00 | 0 | 0.79 |
| 16 | 1.65 - 1.62 | 100.0 | 2287 / 2287 | 31.44 | 98.82 | 43.41 | 38.06 | 2287 | 0.00 | 0 | 0.00 | 0 | 0.96 |
| 17 | 1.62 - 1.58 | 100.0 | 2271 / 2271 | 24.86 | 92.40 | 42.07 | 71.34 | 2271 | 0.00 | 0 | 0.00 | 0 | 0.77 |
| 18 | 1.58 - 1.55 | 100.0 | 2285 / 2286 | 20.58 | 90.93 | 45.20 | 63.55 | 2285 | 0.00 | 0 | 0.00 | 0 | 0.78 |
| 19 | 1.55 - 1.53 | 100.0 | 2252 / 2252 | 14.13 | 89.12 | 52.29 | 57.09 | 2252 | 0.00 | 0 | 0.00 | 0 | 0.83 |
| 20 | 1.53 - 1.50 | 99.9 | 2262 / 2265 | 9.87 | 88.85 | 63.11 | 51.62 | 2262 | 0.00 | 0 | 0.00 | 0 | 0.88 |
| | TOTAL | 100.0 | 46047 / 46053 | 83.94 | 84.71 | 12.35 | 98.37 | 46047 | 0.00 | 0 | 0.00 | 0 | 2.65 |

**Figure 6.5:** Merging statistics generated after a run of the program *prime*, with different data quality statistics detailed for each resolution shell, in this case for the DHP chip ´lukaku'.

A second *Prime* run would be carried out with the optimized cell parameters and resolution cut-off. The resulting *mtz* file would be taken forward for routine structure solution using the ccp4i2 package, starting with a molecular replacement job with either *Phaser* or *Molrep*, using as a model one

of the previously solved DHP structures (McCombs et al., 2017). The data initially refined to R values of 0.20 and 0.23 for $R_{work}$ and $R_{free}$ respectively, and manual model building was carried out with *Coot* (Emsley et al., 2010) in-between rounds of refinement with *Refmac5* (Murshudov et al., 2011).

### *6.2.2 SSX structure of DHP-b from ferric crystals.*

The structure here described comes from the best dataset measured for DHP, which was measured from the chip 'lukaku' that diffracted to 1.50 Å resolution. Before the beamtime, a UV-vis absorption spectrum was measured from the microcrystals (in suspension in its own batch condition) to assess their oxidation state, and is shown in Figure 6.6. As it can be noted, DHP microcrystals showed a ferric spectral signature, with a Soret band ~400 nm.

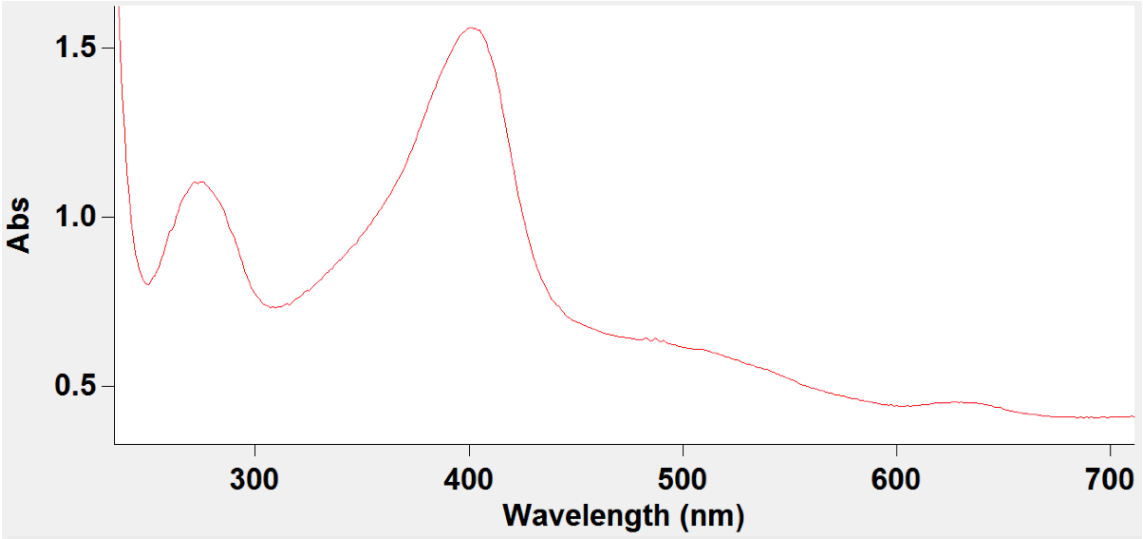


**Figure 6.6:** UV-vis absorption spectrum of a suspension of DHP microcrystals suspended in the batch solution where they were grown. Measurement was carried at room-temperature using a Cary-60 spectrophotometer (Agilent).

In total, diffraction from ~9000 DHP microcrystals was measured from the chip lukaku, which was enough for a complete dataset. Data were not merged with other chips, since data quality when this was done was worst and, in this way, we ensured that all the data were coming from the same batch of microcrystals. Data quality and refinement statistics are shown in Table 64.

The room-temperature SSX structure of DHP presented a homodimer in the asymmetric unit, and the protein adopted the characteristic globin fold, as seen previously (De Serrano et al. 2010). However, this time a new feature could be observed due to the low-dose dataset (82 kGy) achieved with SSX combined with the high resolution obtained for this sample. At haem B, a well-defined distal water was observed coordinating the iron at 2.54 Å (Figure 6.7, right panel). This feature had not been seen before in previous atomic resolution structures due to photo-reduction of the haem site and the subsequent rearrangement of the coordination sphere. Once the molecule of water was modelled in, there was no extra electron-density observed in the $F_o$ - $F_c$ difference map. Thus, the modelling of an $O_2$ molecule was not considered, since the oxy-ferrous species had not been seen by photo-reduction of DHP at much higher doses (refer to Chapter 5). In the SSX structure, the catalytic His55 is stabilising the bound water through a weak hydrogen bond at 3.30 Å. At the active site of monomer A, as observed in other DHP structures in this work, the haem group presents a double conformation, one of them being a hemichrome species (Figure 6.7, left panel), where His55 is hexacoordinating the haem iron at 2.59 Å at the axial position. The proportion of hemichrome in the double conformation for this SSX structure appears to be higher than for

others measured for DHP, and for instance the alternative conformation of the proximal His89 could be modelled.

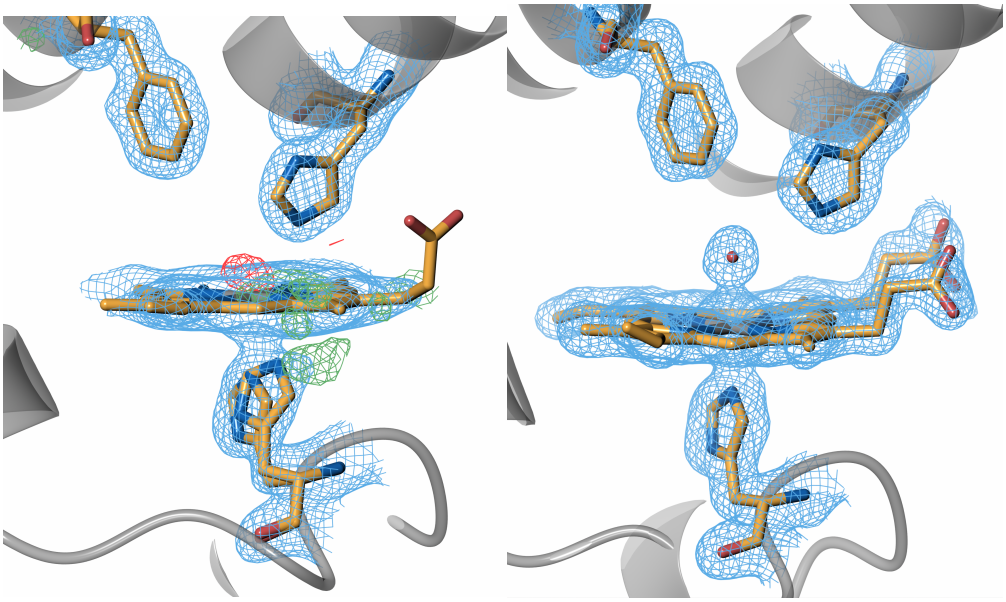

**Figure 6.7:** Haem sites for the synchrotron serial crystallography (SSX) structure from DHP ferric crystals measured at beamline I24 at Diamond (Oxford, UK). Monomer A is shown on the left and monomer B on the right. Note the hemichrome species in monomer A, and the distal water hexacoordinating the heme iron at the axial position in monomer B. Electron density is shown as a blue mesh and contoured at 1σ (0.345 $e^-/Å^3$).

**Table 6.4:** Data collection and refinement statistics for the synchrotron serial crystallography (SSX) structure measured from DHP ferric microcrystals at beamline I24, Diamond (Oxford, UK). Values between parentheses are those corresponding to the last resolution shell.

| | **DHP SSX structure** |
|---|---|
| Space group | $P2_12_12_1$ |
| Unit cell (Å) | a=61.18, b=67.00, c=68.93 |
| Resolution (Å) | 37.80 – 1.50 (1.53 – 1.50) |
| Unique reflections | 46053 (2264) |
| Isotropic B-factor ($Å^2$) | 17.35 |
| $R_{split}$ (%) | 11.36 (52.18) |
| I / σ(I) | 3.74 (0.77) |
| $CC_{1/2}$ | 0.98 (0.6) |
| Completeness (%) | 100 (100) |
| $R_{work}$ (%) | 19.3 |

| | |
|---|---|
| $R_{free}$ (%) | 21.5 |
| Rmsd bond lengths (Å) | 0.0159 |
| Rmsd bond angles (º) | 1.794 |
| Ramachandran favoured (%) | 98.3 |
| Absorbed dose (kGy) | 82 |

## 6.3 Pink beam SX at APS.

As discussed in Chapter 2, Laue radiation is also commonly known within the field as 'pink beam', which consists of a polychromatic radiation having a bandpass of wavelengths. Therefore, more than one spot can be obtained for a given family of crystallographic planes with Miller indices (*h, k, l*), producing characteristic diffraction patterns. The main attractive for using a pink beam was the inspection of different time-frames compared to conventional synchrotron sources – Laue beamlines produce pulses of 100 ps. This would eventually open the possibility to perform time-resolved crystallography experiments, once the method was developed enough. Through a collaboration with Dr. Diana Monteiro and Prof. Martin Trebbin, from the Center for Ultrafast Imaging at the University of Hamburg, we had the opportunity to apply our methodologies at the Laue beamline BioCARS 14-IDB at Advanced Photon Source (APS, Chicago, USA). We visited APS twice, spending a total of 8 days, and therefore we had the chance to test different methodologies while collecting useful data. Since the first APS trip took place before any XFEL beamtime, these constituted our first trials to develop the production of DHP microcrystals (and from other protein targets). The optimization was at its first stages and the DHP crystals grown in batches were rather large: ranging 50-200 μm. This supposed a limitation in terms of

obtaining a big number of patterns, not even mentioning the possibility of jetting the crystals for time-resolved experiments. However, these bigger crystals would maximize the resolution of the measured dataset, bearing also in mind that for a Laue dataset fewer diffraction patterns are commonly needed for a complete dataset (Meents et al. 2017).

All the data collected at APS was obtained using the 'Heidelberg' chips described in Chapter 3. After loading the crystals, the chip was fitted in a holder and mounted in the goniometer (Figure 6.8). The set up at the beamline allowed us a frequency of 2 Hz for chip inspection, and thus 30-45 min would be needed to screen each quadrant of the chip. To ensure collection of complete datasets, 3-4 chips were measured for DHP (same for the other targets).

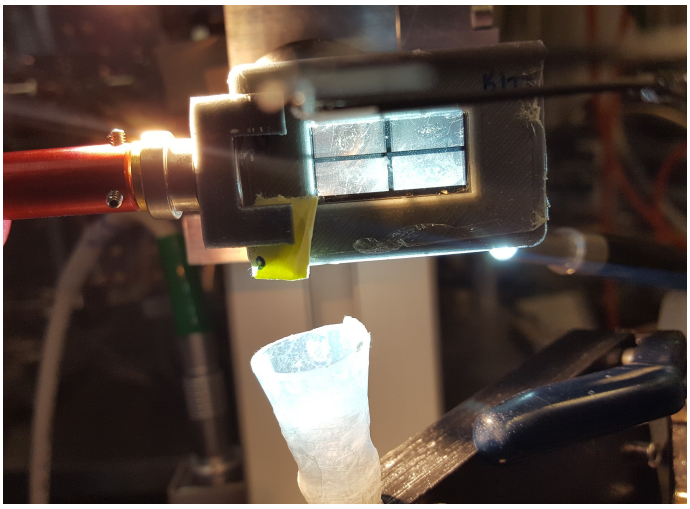

**Figure 6.8:** Heidelberg chip mounted at the BioCARS ID-14B beamline onto the goniometer with help of an adaptor holder. Note that beneath the chip, the drain from the jet experiments can be appreciated.

The screening routine procedure is detailed in Chapter 3. An area highly populated with crystals (assessed visually) would be selected and parameters

introduced using the software at the beamline. Different ‘bunch-modes’ were possible at this Laue beamline, consisting in trains of pulses 100ps long each. Data were measured with a ‘super-bunch’ mode to maximise resolution, which had a total train length of 500 ns and a total current of 102 mA. A single bunch containing 16 mA was isolated from the remaining bunches by symmetrical 1.6 μs gaps. The remaining current was distributed in 8 groups of 7 consecutive bunches with a maximum of 11 mA per group. In Figure 6.9, a Laue diffraction pattern for DHP is shown, diffracted with the pink beam at APS.

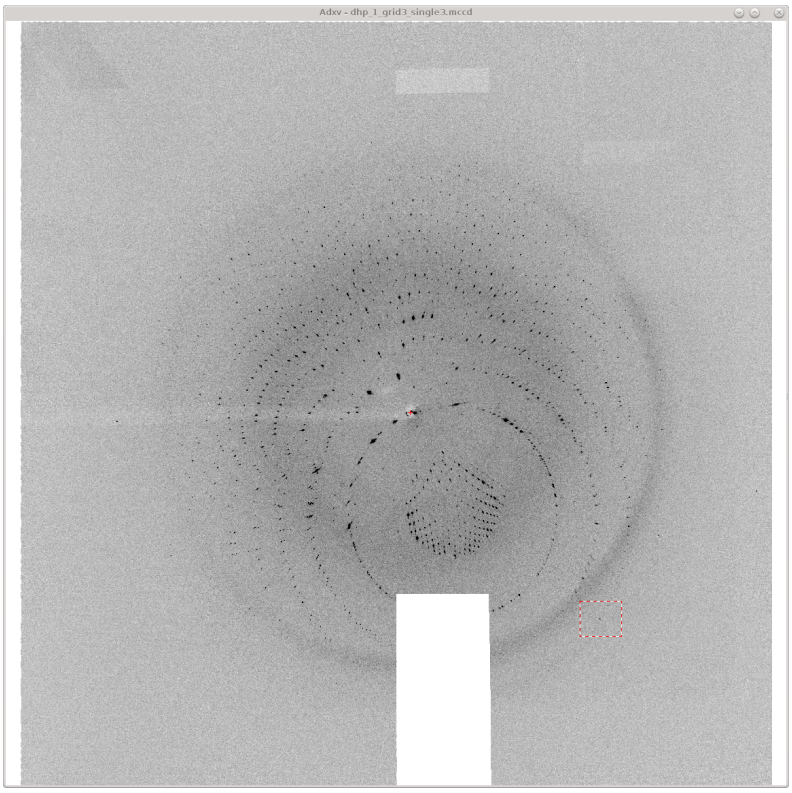

**Figure 6.9:** Laue pattern obtained from DHP-B crystals grown in batch. Characteristic polychromatic diffraction can be appreciated.

Laue data was processed by APS scientist Dr. Vukica Srajer, using the proprietory software Laueview and Precognition. A *hkl* file was generated as an outcome, which was converted to an *mtz* file using the program *F2mtz* within

the CCP4 suite, and used for molecular replacement and refinement as usual using the CCP4i2 user interface. The resolution of the merged data was 2.0 Å, with a 3σ cut-off criteria applied. At this lower resolution, the DHP resulting structure was, as could be expected, less detailed than the SSX structure. A dimer was found in the asymmetric unit and the heme environments for each of the monomers are shown in Figure 6.10. In contrast with the structure measured using SSX, we can observe a double conformation of His55 as a main feature in both monomers. The hemichrome in monomer A is not observable, although there is a slight positive density at the alternative position of the iron and one of the His55 conformation would be in place for hexacoordinating the iron. Data collection and refinement statistics for this serial Laue crystallography (SLX) structure are detailed in Table 65.

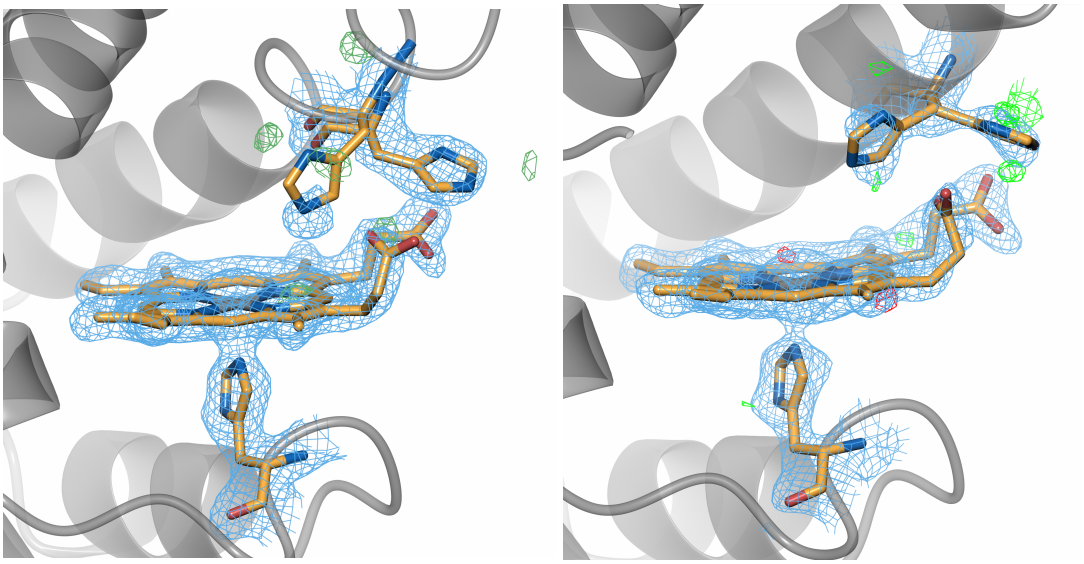

**Figure 6.10:** Heme environments for monomer A (left) and B (right) of the DHP serial Laue crystallography structure (SLX) measured at APS (Chicago, USA). Note the double conformation of His55 in both cases.

**Table 6.5:** Data collection and refinement statistics for the serial Laue crystallography (SLX) dataset from DHP ferric crystals, measured at beamline BioCARS 14-IDB in Advanced Photon Source (APS, Chicago)

| | **DHP Laue structure** |
|---|---|
| Space group | $P2_12_12_1$ |
| Unit cell (Å) | a=60.87, b=67.30, c=67.73 |
| Resolution (Å) | 67.72 – 2.01 (2.06 - 2.01) |
| Total reflections | 14969 (735) |
| I / σ(I) | 20.3 (6.2) |
| Completeness (%) | 79.2 (31.7) |
| $R_{work}$ (%) | 19.3 |
| $R_{free}$ (%) | 25.0 |
| Rmsd bond lengths (Å) | 0.0148 |
| Rmsd bond angles (º) | 1.736 |
| Ramachandran favoured (%) | 97.93 |

## 6.4 Serial Femtosecond Crystallography (SFX) at SACLA

Prepared batches with DHP microcrystals were taken to SACLA (Japan) for measurement, contained in a thermos flask to keep them at 4ºC during the transport to the XFEL beamline. In total, there were 20-30 batches, each of them with 300-400 μl of solution with microcrystals. Once there, they were loaded onto silicon nitride chips (Mueller et al., 2015), which were mounted in the XFEL hutch for diffraction.

### *6.4.1 Experimental set-up at the hutch at SACLA*

The fixed target SX system has been described in Chapter 2 and the chips and screening routines in Chapter 3. However, for the experimental set up at SACLA there were special features different from the synchrotron environment. The most relevant one was the inclusion of the whole system within a box (Figure 6.11) to perform the experiment under a helium atmosphere. Even if it

resulted in a cramped sample environment, the modular design of the fixed target system ensured its straightforward installation. The use of a helium atmosphere minimized the scattering background (Meents et al. 2017) compared to air and, in this way, diffraction data quality was improved. A circular aperture on one side (Figure 6.11, top panel) ensured access to the chamber to enable rapid sample exchange. The movement of the chip and stages between inspecting positions had to be synchronized with the pulse frequency at SACLA, which was 30 Hz.

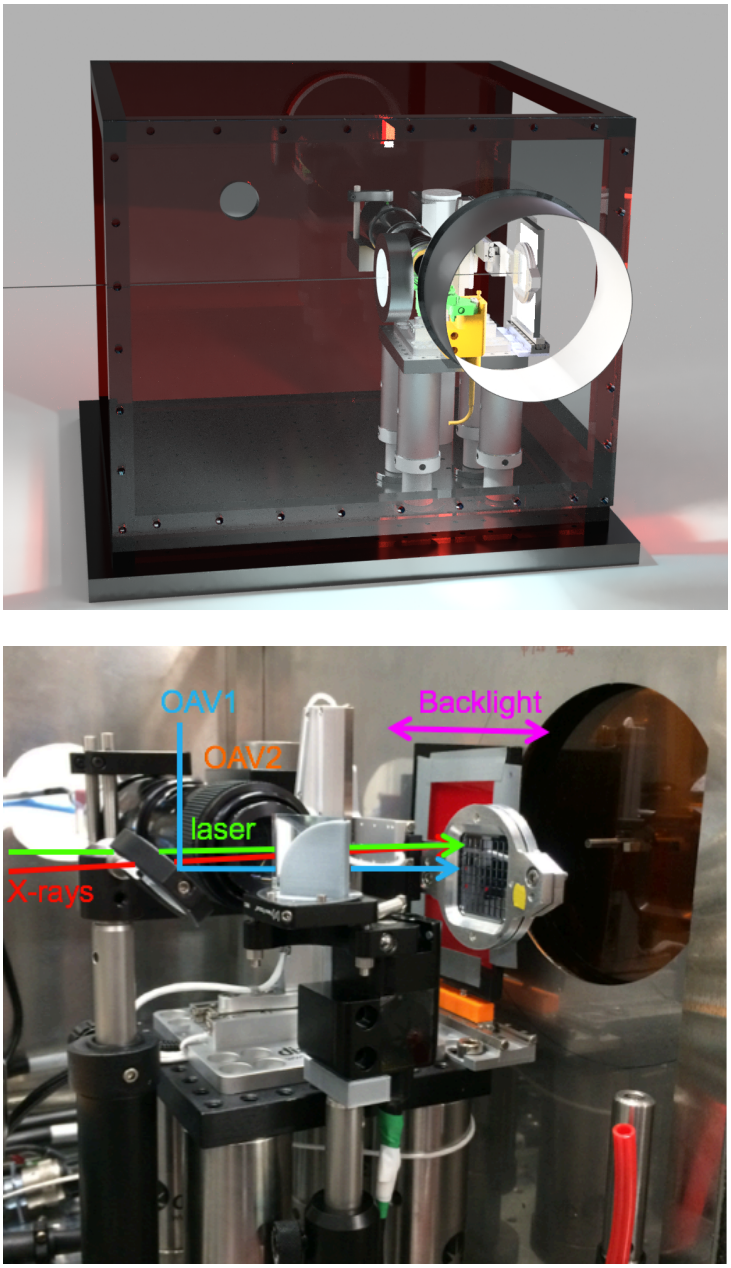


**Figure 6.11:** Experimental set up at SACLA. In the upper panel, a graphic representation of the helium chamber in which the instrument was contained. A circular aperture that was used for sample exchange can be noted. In the lower panel a close-up look at the modular serial crystallography system as mounted inside the chamber. The direction of X-rays, excitation laser, and sample visualization system (OAV1) are indicated.

### *6.4.2 Measured SFX data for DHP.*

A total of three structures were determined: the resting state of DHP, and two ligand complexed states, with substrates 5-bromoindole (5-BrInd) and 2,4-

dichlorophenol (DCP). The known DHP substrates (Carey et al. 2018) were selected to test the ability of perform ligand soaking efficiently with SFX. For the DHP ferric crystals they were applied as they were in the batches. Some pipetting was needed to detach some of the crystals from the Eppendorf walls and resuspend all of those that were accumulated at the bottom of the tube. In the case of the ligand soaks, the compounds were dissolved in DMSO, due to its limited solubility in water, and a 500 mM stock was prepared in each case. The needed amount of stock was added to the batches to obtain a final concentration of 50mM and 5mM for 5-bromoindole (5-BrInd) and 2,4-dichlorophenol (DCP), respectively. The concentration used for DCP was lowered since it was found to damage the crystals and affect diffraction quality at 50 mM. Once the substrate was added, the batches were mixed with the pipette several times and left 5 minutes soaking prior to loading onto the chip. In Table 6.2, collection details for each of the chips are shown. A minimum of four chips were measured for the determination of each of the structures. For easier identification when working with them and during data processing, each chip was named with a six-letter city name. All SFX data for DHP were measured in the October-2017 SACLA beamtime, and are shown in Table 66, and an example of an XFEL diffraction pattern for DHP is shown in Figure 6.12. For all the experiments the following parameters were constant:

- Energy = 10.00 keV
- FEL pulse length = 10 fs
- Beam size = 1.25x1.34 μm
- Repetition rate = 30 Hz

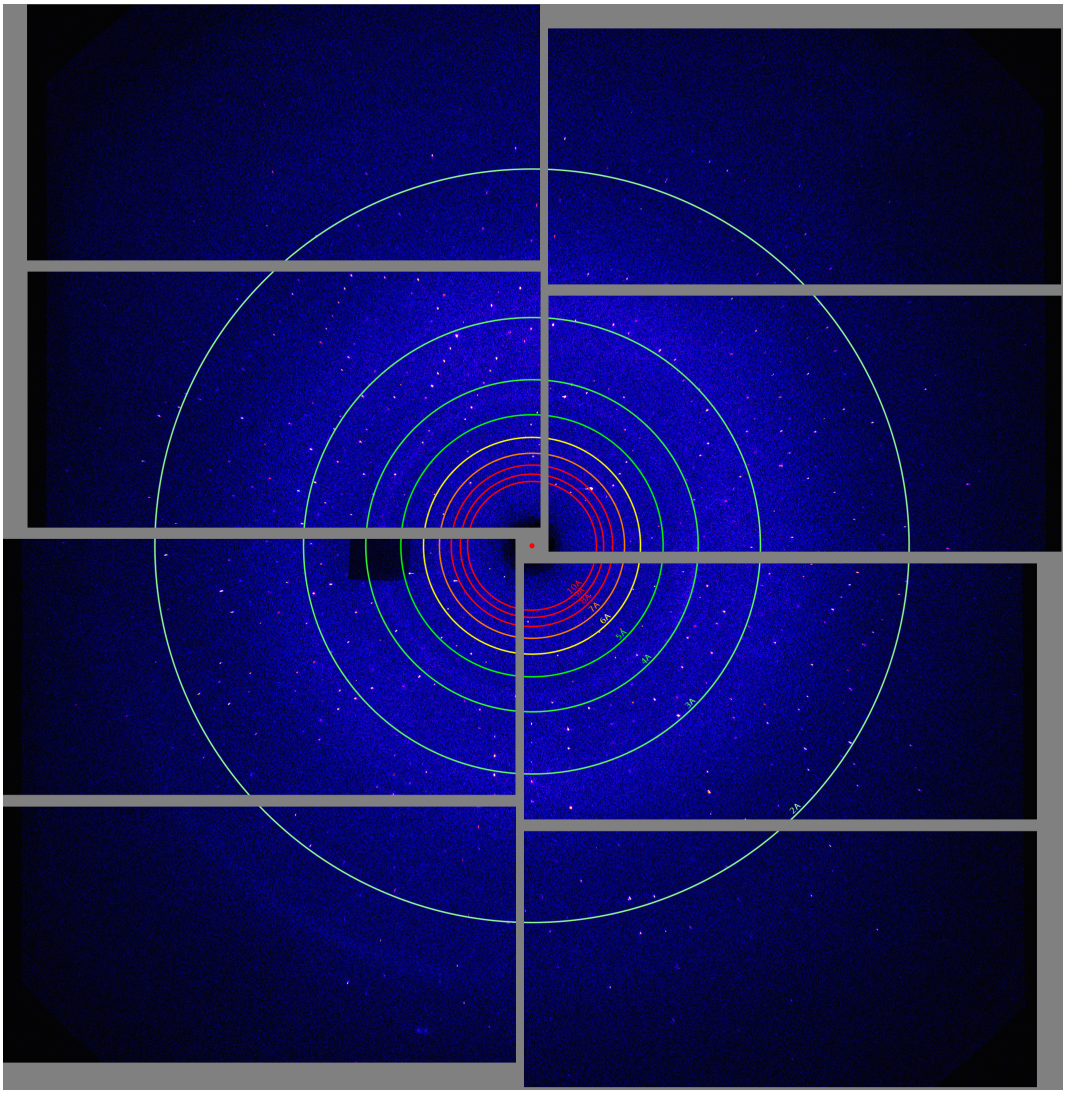

**Figure 6.12:** XFEL diffraction pattern of a ferric DHP microcrystal. Resolution circles range from 2 Å in the outer to 10 Å in the inner circle.

**Table 6.6:** Collection parameters and details for the chips measured for DHP at SACLA (October 2017). The full chip was inspected in each case, with 25,700 exposure events per run. Beam attenuation was set a 0.3. Hit rates were estimated using the software Cheetah (Barty et al., 2014).

| Chip name | Condition | Hit rate (%) |
|---|---|---|
| Ebetsu | DHP ferric | 12.3 |
| Fuqing | DHP ferric | 11.9 |
| Geneva | DHP ferric | 16.6 |
| Teruel | DHP ferric | 41.6 |

| | | |
|---|---|---|
| Verona | DHP ferric | 37.6 |
| Latina | DHP + 5-BrInd | 62.1 |
| Moscow | DHP + 5-BrInd | 41.3 |
| Oxnard | DHP + 5-BrInd | 67.8 |
| Sakura | DHP + 5-BrInd | 13.6 |
| Qidong | DHP + DCP | 27.0 |
| Toledo | DHP + DCP | 68.1 |
| Umlazi | DHP + DCP | 37.2 |
| Valera | DHP + DCP | 20.2 |

### *6.4.3 Processing of the SFX data*

The preliminary hit finding and estimation of hit rates was carried using the program Cheetah (Barty et al. 2014), and this task was performed by Miss Helen Duyvesteyn in the team. The software CrystFEL (White et al. 2012) was used to process all data measured at SACLA and obtain the SFX damage-free structures. This software was specifically developed to treat serial crystallography XFEL data, although in new updates is also able to deal with synchrotron data. As discussed in Chapter 3, it consists of a package containing different programs that are run from the command line. It follows the normal crystallography workflow of indexing, scaling and merging, with different options of algorithms to be used in each step. Indexing of the diffraction patterns from DHP fixed targets was carried out with the CrystFEL program *indexamajig*, using the following command:

```
$ indexamajig -i files.lst --peaks=zaef -threshold=300 --
min-gradient=90000   --min-snr=5   --int-radius=3,4,5   --
indexing=asdf -g 23368-1.geom -p DHP.cell -o testrun.out -
j 20
```

The different parameters are defined in Chapter 3. Each of the chips was processed in this way, generating in each case a file stream.out with the indexed reflections. As an example of the workflow followed for each of the determined DHP SFX structures, the processing of the data for **resting state DHP** will be described from here on. The *DHP.cell* file (Figure 6.13) containing the cell parameters and symmetry for DHP, was generated using a CrystFEL template, introducing the RT parameters obtained at an SSX beamtime at I24 (Diamond, Oxford).

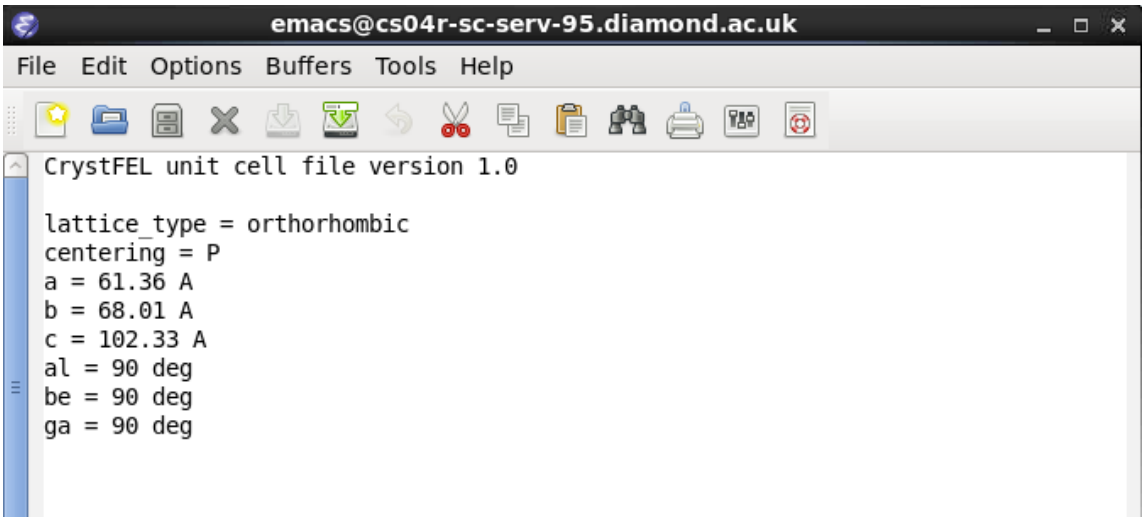


**Figure 6.13:** Capture of CrysFEL file DHP.cell, with symmetry and cell parameters specified for this sample.

An initial indexing run for each chip was carried out to refine the cell parameters for the microcrystals. After the first run, the cell parameters distribution was examined using the CrystFEL tool *cell_explorer*, which

generated distribution plots for every cell parameter (Figure 6.14), calculating the mean value and standard deviation. The command to run this program was (in this example for the chip ‘geneva’):

```
$ cell_explorer firstRun_geneva.out
```

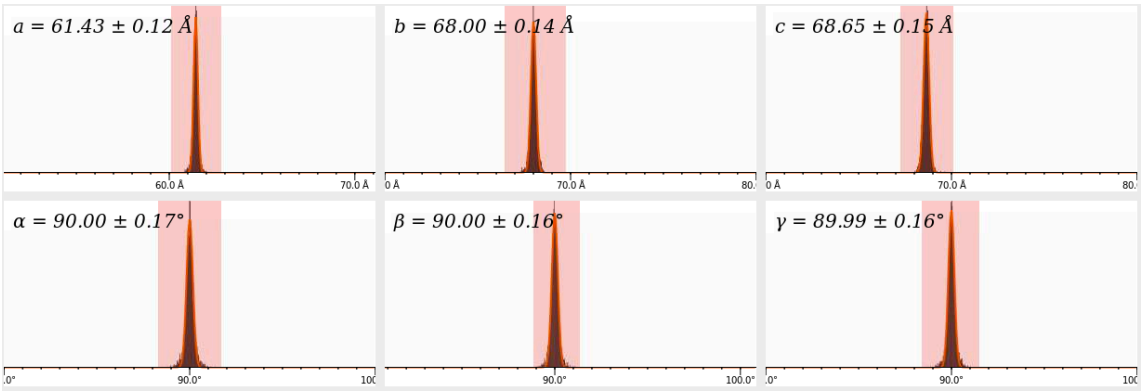


**Figure 6.14:** Distribution of cell parameters among all the measured crystals for the chip ‘geneva’, loaded with ferric DHP micro-crystals. Mean values and standard deviation were calculated for the data included within a selected range (shaded in red).

Cell parameters were therefore updated for the experimental distribution of DHP microcrystals, and a new file, called *DHP_refined.cell*, was generated. A second run of *indexamajig* was carried out with these optimized cell parameters, indexing a higher number of patterns, but not substantially, indicating that the initial given parameters were very close (since they have been obtained from a very similar sample). With the output of the second run and the used geometry file for the detector, a script called *detector-shift* was run to check and optimise the detector geometry (which had been refined initially, refer to Chapter 3). It also plots the disagreement (if any), expressed as detector shift in x and y, between the location of experimental and calculated

spots in every indexed frame. The plot corresponding to the chip ‘geneva’ is shown in Figure 6.15. The command introduced to run *detector-shift* was:

```
$ ./detector-shift run2_geneva.out 23372-1.geom
```

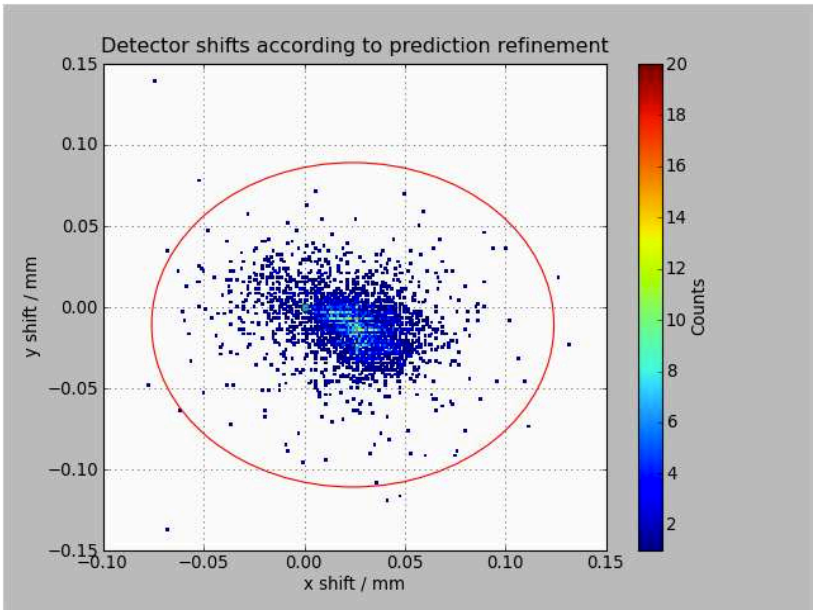


**Figure 6.15:** Plot generated by running the CrystFEL script *detector_shift* for the ferric DHP chip ‘geneva’.

*Detector-shift* would generate as output a refined geometry file, in this case called ‘23372-1-predrefine.geom’. With it, a final run with *indexamajig* was performed, obtaining a higher number of indexed patterns. The final stream file for the ‘geneva’ chip was called ´geneva_refGeom.out’. Data from the different chips was joined together to form a file ready for scaling and merging. The command used to join the different output files was:

```
$ cat chip1.out chip2.out chip3.out > allchips.out
```

The merging process was carried out with the simplest approach, a Monte Carlo method, used by the program *process_hkl* (White et al., 2012). This program was included in the script called *merge.sh*, which we would run with the joined output file. An example of this script is shown in Figure 6.16. The indexed and integrated reflections, *allchips.out*, would be given as input, and we would also need to provide the point group in which the protein crystallized. For DHP, it would be 222, since the spacegroup is $P2_12_12_1$. As recommended by CrystFEL developers (tutorial), the equivalent pointgroup with a centrosymmetric element was used, which in this case would be mmm, to merge the Friedel pairs (just needed for anomalous data), and result in a better merging.

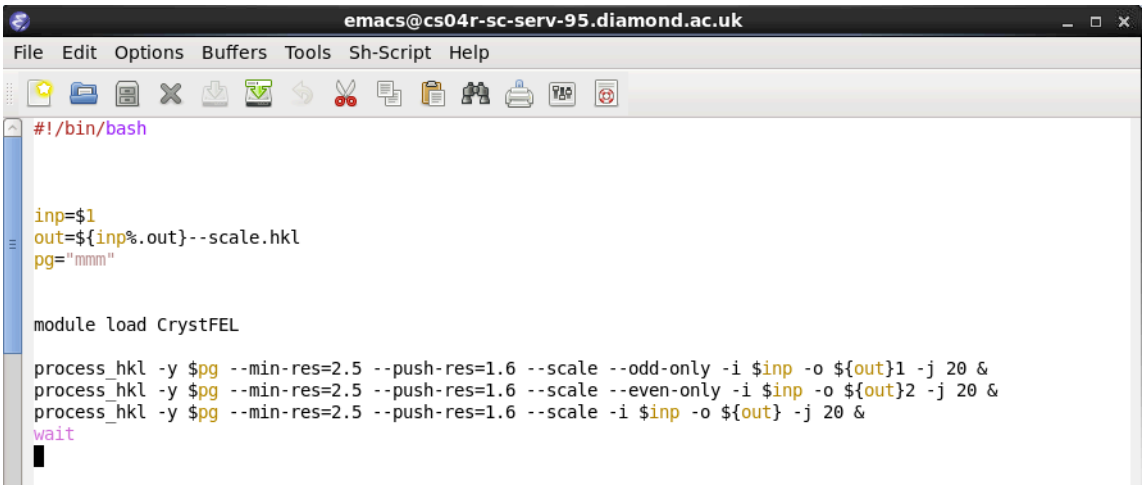


**Figure 6.16:** CrystFEL script merge.sh, which runs the program process_hkl, performing a Monte Carlo merging.

Merging using merge.sh, provided three output files, a .hkl file with all the merged reflections, and files .hkl1 and .hkl2, where the data had been split

in two halves for generating data quality statistics afterwards. This was also carried out by a CrystFEL script called ‘stat.sh’, shown in Figure 6.17. As input, the symmetry information for DHP had to be provided, as well as the .hkl files, and the pointgroup specified. Several statistics were generated, being the correlation coefficient (CC) and $R_{split}$ the most relevant for us (defined in Chapter 3).

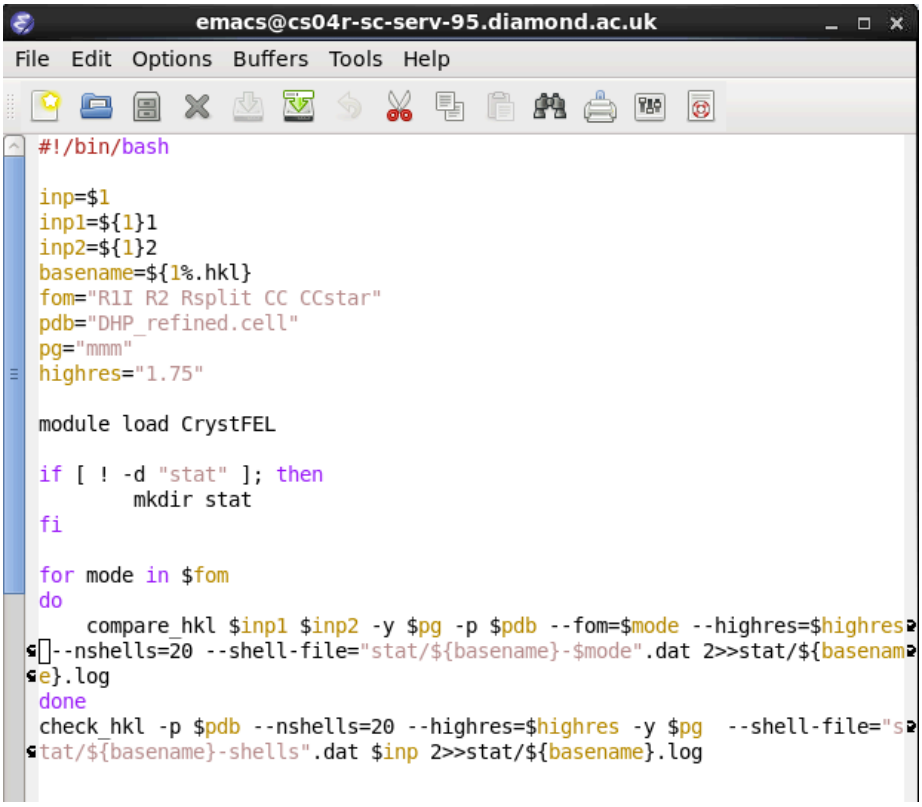


**Figure 6.17:** Caption of the script stat.sh for DHP, where the different figures of merit (fom) to be generated were specified, providing crystal symmetry information, the pointgroup and a resolution cut-off.

Each of the files generated by ‘stat.sh’ were *dat* files containing the corresponding information. ‘Shells.dat’ contained information about completeness and other data quality statistics in the different resolution shells. DHP data were complete all along the selected resolution limit of 1.75 Å (Figure

6.18), this means every possible reflection has been measured at least once, but does not tell us anything about multiplicity or if the data will be meaningful.

```
[zlp66871@cs04r-sc-serv-95 stat]$ cat dhp_e-f-g-t-v--scale-shells.dat
```

| Center 1/nm | # refs | Possible | Compl | Meas | Red | SNR | Std dev | Mean | d(A) | Min 1/nm | Max 1/nm |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.185 | 1630 | 1630 | 100.00 | 1781288 | 1092.8 | 15.66 | 2402.42 | 2478.53 | 8.44 | 0.263 | 2.106 |
| 2.380 | 1537 | 1537 | 100.00 | 1133753 | 737.6 | 16.88 | 2615.81 | 2813.90 | 4.20 | 2.106 | 2.653 |
| 2.845 | 1504 | 1504 | 100.00 | 941159 | 625.8 | 16.65 | 1613.40 | 1742.94 | 3.52 | 2.653 | 3.037 |
| 3.189 | 1487 | 1487 | 100.00 | 762028 | 512.5 | 15.48 | 873.80 | 871.70 | 3.14 | 3.037 | 3.342 |
| 3.471 | 1484 | 1484 | 100.00 | 715829 | 482.4 | 15.11 | 583.74 | 593.36 | 2.88 | 3.342 | 3.600 |
| 3.713 | 1483 | 1483 | 100.00 | 695972 | 469.3 | 14.39 | 367.87 | 364.73 | 2.69 | 3.600 | 3.826 |
| 3.926 | 1474 | 1474 | 100.00 | 662646 | 449.6 | 13.93 | 300.63 | 300.94 | 2.55 | 3.826 | 4.027 |
| 4.119 | 1485 | 1485 | 100.00 | 643548 | 433.4 | 12.94 | 225.01 | 224.33 | 2.43 | 4.027 | 4.211 |
| 4.295 | 1449 | 1449 | 100.00 | 607116 | 419.0 | 11.62 | 155.86 | 153.43 | 2.33 | 4.211 | 4.379 |
| 4.457 | 1479 | 1479 | 100.00 | 600197 | 405.8 | 10.61 | 120.31 | 120.80 | 2.24 | 4.379 | 4.536 |
| 4.609 | 1464 | 1464 | 100.00 | 576679 | 393.9 | 9.23 | 85.68 | 84.73 | 2.17 | 4.536 | 4.682 |
| 4.751 | 1474 | 1474 | 100.00 | 565081 | 383.4 | 7.66 | 75.02 | 58.26 | 2.10 | 4.682 | 4.820 |
| 4.885 | 1462 | 1462 | 100.00 | 548516 | 375.2 | 6.50 | 42.12 | 40.43 | 2.05 | 4.820 | 4.950 |
| 5.012 | 1442 | 1442 | 100.00 | 528717 | 366.7 | 5.08 | 28.28 | 25.93 | 2.00 | 4.950 | 5.074 |
| 5.133 | 1450 | 1450 | 100.00 | 520390 | 358.9 | 3.93 | 19.20 | 17.30 | 1.95 | 5.074 | 5.192 |
| 5.248 | 1462 | 1462 | 100.00 | 499883 | 341.9 | 2.73 | 12.66 | 10.45 | 1.91 | 5.192 | 5.305 |
| 5.359 | 1454 | 1454 | 100.00 | 444094 | 305.4 | 1.82 | 8.14 | 6.39 | 1.87 | 5.305 | 5.413 |
| 5.465 | 1458 | 1458 | 100.00 | 364947 | 250.3 | 1.08 | 5.19 | 3.62 | 1.83 | 5.413 | 5.517 |
| 5.567 | 1458 | 1458 | 100.00 | 286325 | 196.4 | 0.64 | 4.63 | 2.29 | 1.80 | 5.517 | 5.617 |
| 5.666 | 1439 | 1439 | 100.00 | 225306 | 156.6 | 0.34 | 4.17 | 1.27 | 1.76 | 5.617 | 5.714 |

**Figure 6.18:** Shell.dat file for DHP resting state, where different data quality information such as completeness can be obtained for the different resolution shells. The cut-off applied was 1.75 Å and hence the last resolution shell includes diffraction spots within the 1.80-1.76 Å range.

To have a real estimation of the quality of our data and know where to apply the resolution cut-off, the statistics CC and Rsplit had to be inspected. Having a look at CC for ferric DHP, shown in Figure 6.19, it can be observed that data quality is good (CC above 0.95 until ~2Å), and that the resolution cut-off could be applied at 1.85 Å, if the criteria of CC > 0.5 is applied.

```
[zlp66871@cs04r-sc-serv-95 stat]$ cat dhp_e-f-g-t-v--scale-CC.dat
  1/d centre       CC       nref      d / A   Min 1/nm    Max 1/nm
     1.185  0.9856749       1630       8.44      0.264       2.106
     2.380  0.9863494       1537       4.20      2.106       2.653
     2.845  0.9863993       1504       3.52      2.653       3.037
     3.189  0.9844816       1487       3.14      3.037       3.342
     3.471  0.9843328       1484       2.88      3.342       3.600
     3.713  0.9858696       1483       2.69      3.600       3.826
     3.926  0.9843973       1474       2.55      3.826       4.027
     4.119  0.9838754       1485       2.43      4.027       4.211
     4.295  0.9823541       1449       2.33      4.211       4.379
     4.457  0.9761027       1479       2.24      4.379       4.536
     4.609  0.9646750       1464       2.17      4.536       4.682
     4.751  0.9655787       1474       2.10      4.682       4.820
     4.885  0.9533704       1462       2.05      4.820       4.950
     5.012  0.9353684       1442       2.00      4.950       5.074
     5.133  0.8967105       1450       1.95      5.074       5.192
     5.248  0.7968781       1462       1.91      5.192       5.305
     5.359  0.6745461       1454       1.87      5.305       5.413
     5.465  0.3415315       1458       1.83      5.413       5.517
     5.567  0.1883131       1458       1.80      5.517       5.617
     5.666  0.0560186       1439       1.76      5.617       5.714
```

**Figure 6.19:** Table generated by the script stat.sh, where correlation coefficient (CC) statistic for ferric DHP data is shown for each resolution shell until a 1.75 Å cut-off.

Similarly, $R_{split}$ was inspected to assess resolution cut-off according to this statistic, shown in Figure 6.20. Data looks very good up till 2 Å resolution (below 20% $R_{split}$), and it can be extended until 1.85 Å, matching with the CC based cut-off.

```
[zlp66871@cs04r-sc-serv-95 stat]$ cat dhp_e-f-g-t-v--scale-Rsplit.dat
  1/d centre Rsplit/%       nref       d / A   Min 1/nm    Max 1/nm
       1.185       6.95       1630        8.44       0.264       2.106
       2.380       6.47       1537        4.20       2.106       2.653
       2.845       6.59       1504        3.52       2.653       3.037
       3.189       7.13       1487        3.14       3.037       3.342
       3.471       7.02       1484        2.88       3.342       3.600
       3.713       7.11       1483        2.69       3.600       3.826
       3.926       7.43       1474        2.55       3.826       4.027
       4.119       7.53       1485        2.43       4.027       4.211
       4.295       8.41       1449        2.33       4.211       4.379
       4.457       9.75       1479        2.24       4.379       4.536
       4.609      10.94       1464        2.17       4.536       4.682
       4.751      13.27       1474        2.10       4.682       4.820
       4.885      15.32       1462        2.05       4.820       4.950
       5.012      19.35       1442        2.00       4.950       5.074
       5.133      26.05       1450        1.95       5.074       5.192
       5.248      38.36       1462        1.91       5.192       5.305
       5.359      59.74       1454        1.87       5.305       5.413
       5.465     107.54       1458        1.83       5.413       5.517
       5.567     181.16       1458        1.80       5.517       5.617
       5.666     329.01       1439        1.76       5.617       5.714
```

**Figure 6.20:** Outcome table from stat.sh, showing the values of $R_{split}$ for the different resolution shells until 1.75 Å for the ferric DHP dataset.

Finally, an mtz file was generated using the script ‘create-mtz´ (Figure 6.21), which run the CCP4 program *f2mtz*. The hkl file from the merging was given as input. The resulting mtz file had specified the space group and experimental cell parameters from the resting state DHP microcrystals. From here on, the mtz was used as any other dataset from synchrotron sources in the CCP4i2 suite. Firstly, a molecular replacement job was carried out with Molrep (Vagin & Teplyakov 1997) using as a model the previous cryo-structure solved at 0.99 Å resolution. In Table 67, data quality statistics after the processing of the three different DHP structures are shown. Note that the soaking procedure did not affect the diffraction quality of the crystals even if in some cases a 10% of DMSO was present in solution.

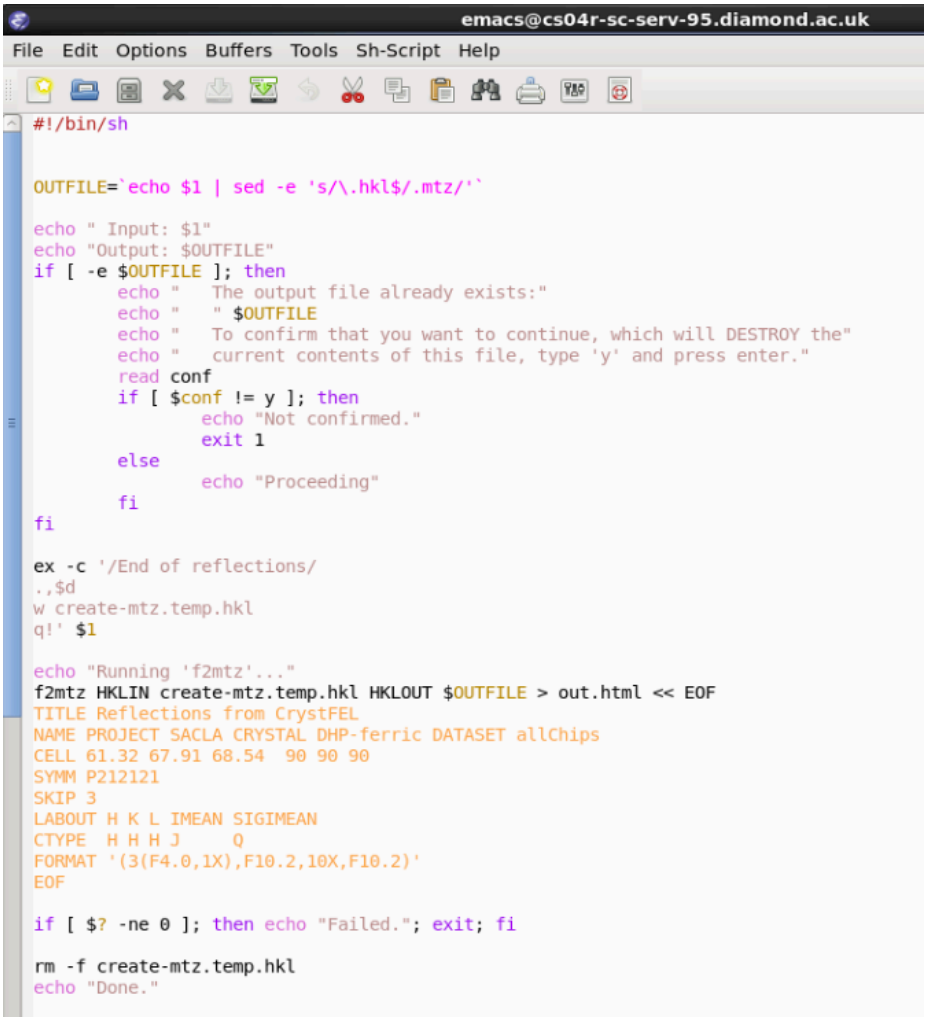


**Figure 6.21:** CrystFEL spript 'create-mtz', where parameters for each particular dataset were introduced in the orange colored region at the bottom. In this case, parameters are specified for ferric DHP data.

**Table 6.7:** Data quality statistics for the different DHP SFX structures measured at SACLA (Japan).

| Condition | DHP ferric | DHP + 5-BrInd | DHP + DCP |
|---|---|---|---|
| Total measured frames | 128,500 | 114,053 | 77,100 |
| Indexed frames | 33184 | 53,687 | 40,396 |
| Average hit rate (%) | 25.82 | 47.07 | 52.39 |
| Resolution (Å) | 37.91 - 1.85 | 37.77 - 1.85 | 37.73 - 1.85 |
| Unique reflections | 27,222 | 26,780 | 26,843 |

| | | | |
|---|---|---|---|
| Completeness (%) | 100 (100) | 100 (100) | 100 (100) |
| CC | 0.99 (0.57) | 0.99 (0.59) | 0.99 (0.53) |
| $R_{split}$ | 7.13 (75.97) | 6.24 (74.71) | 6.84 (77.38) |

## 6.5 SFX structure of the DHP resting state

The measurement of diffraction data with an XFEL pulse of 10 femtoseconds ensured the determination of a radiation damage-free of ferric DHP. Although the DHP microcrystals were not probed with any spectroscopy (e.g. UV-vis) just before loading onto the chips to check the redox state, this was for a well-established condition that we have used before and which delivers ferric crystals, as validated for example with Resonance Raman at SLS (for single crystals at 100 K, Figure 5.3). Furthermore, a UV-vis spectrum was acquired (Figure 6.6) from DHP microcrystals in suspension from another protein batch, grown in the same condition. The position of the Soret band at ~400 nm, was consistent with a ferric state of the protein. Thus, we assume that the microcrystals were in the ferric state, since they were kept at 4 degrees during the trip to Japan, and also at the beamline until they were loaded onto the chips.

At the room-temperature SFX structure of resting state DHP, the protein crystallized as a dimer in the asymmetric unit (Figure 6.22), with each monomer adopting a classic globin fold as expected. Distinct haem sites for each monomer could be appreciated, as in the cryo-structure at 0.99 Å. In heme A (Figure 6.23, left panel), His55 was in a conformation swung into the distal

pocket, whilst in other cryo-structures had been seen swung out of the pocket or a double conformation of the residue could be observed when measuring ferric crystals. On top of the haem iron, an extra positive electron density (from difference map $F_o$ - $F_c$ contoured at 3$\sigma$) can be appreciated, which suggests the partial occupancy of a distal water coordinating the iron. When it was modelled in though, it was not supported by electron-density and therefore it was removed from the model. A coordinating water at the axial haem position can be present in peroxidases in their ground state (e.g. DtpA and DtpAa in Chapter 4). This observation is supported by RR solution data from Dr. Guiladi lab, indicating the existence of a predominant six-coordinate high-spin species for ferric DHP-B in solution. Another important feature of heme site A, is the presence of a double conformation for the heme cofactor, one of them being the hemichrome species, which was also observed in the cryo-structure at 0.99 Å previously described in the chapter. Of course, due to the lower resolution, the species can not be appreciated as clearly as in the 100K. However, a clear negative red electron density (Figure 6.23, left panel) can be seen in the main heme iron and an additional positive difference electron density observed in the alternative iron site, which is coordinated by His55 closer to the heme plane than in heme site B. A positive electron density in the $F_o$ – $F_c$ difference map is also observed in the proximal site just beneath the putative location of the secondary iron, suggesting an alternative conformation of the main proximal histidine to form the hexacoordinated species characteristic of the hemichrome.

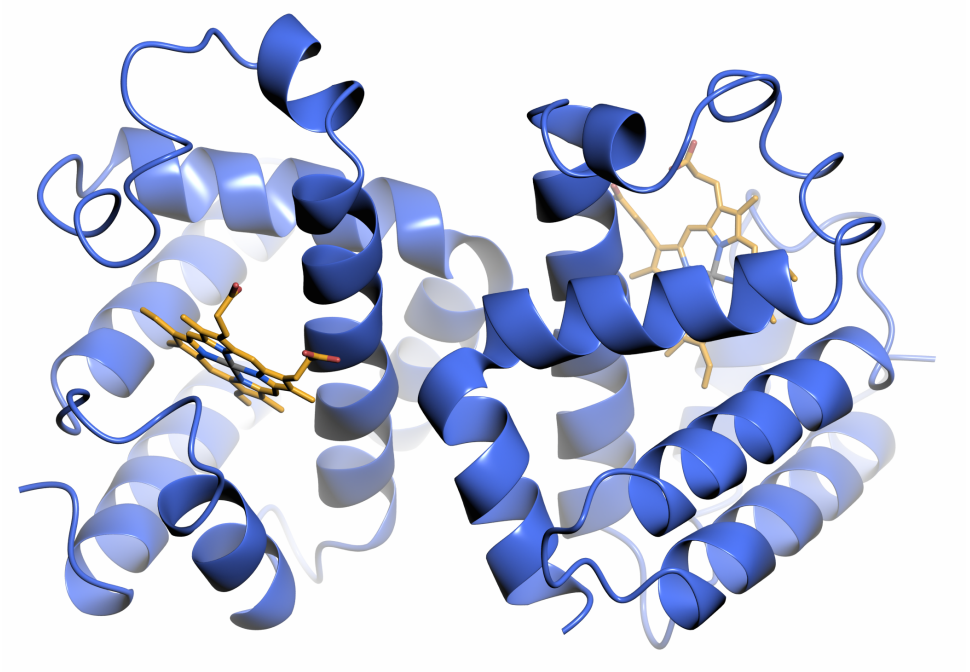

**Figure 6.22:** Overall representation of the ferric DHP dimer found in the asymmetric unit of the SFX structure measured at SACLA. Haem groups shown as cylinders in yellow.

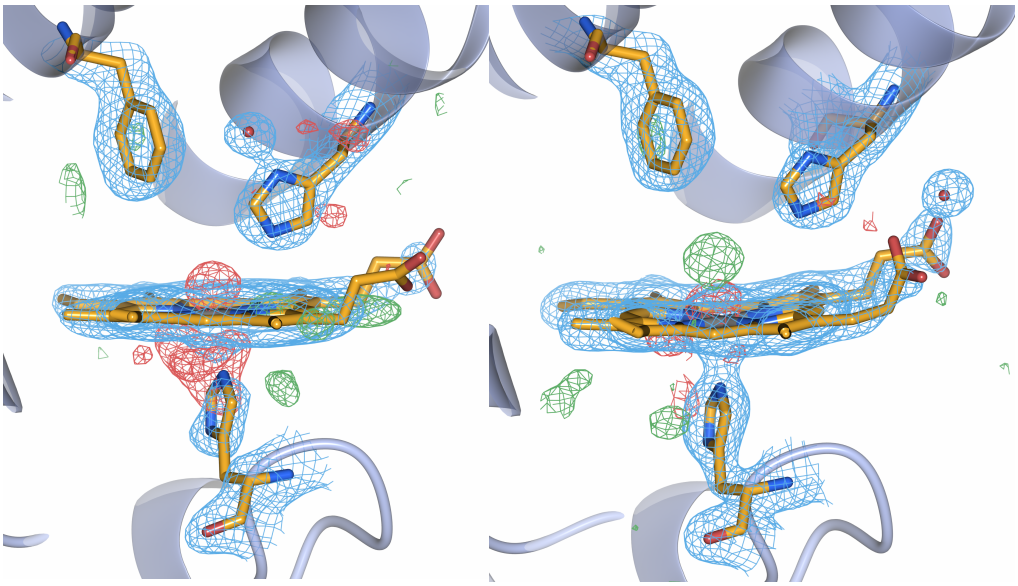

**Figure 6.23:** Damage-free XFEL structure of DHP-B. Heme site of monomer A is shown on the left, where a partial hemichrome species can be appreciated. Active site of monomer B is shown on the right panel, where an extra electron density is observed (green cloud upon the iron), suggesting the presence of a partially occupated distal water.

For a clearer visualization of the double conformation of the heme, the electron density map $2F_o - F_c$ contoured at 4 $\sigma$ is shown in Figure 6.24, where two positions of the iron can be appreciated. The observation of the hemichrome in the XFEL structure, rules out radiation damage and cryoprotection with glycerol as possible causes for its formation, since measurements were performed at room temperature in the presence of just the crystallization condition. It is left to determine whether the hemichrome is a crystal artefact formed by exposure to the crystallization conditions, or if it could have a biological relevance.

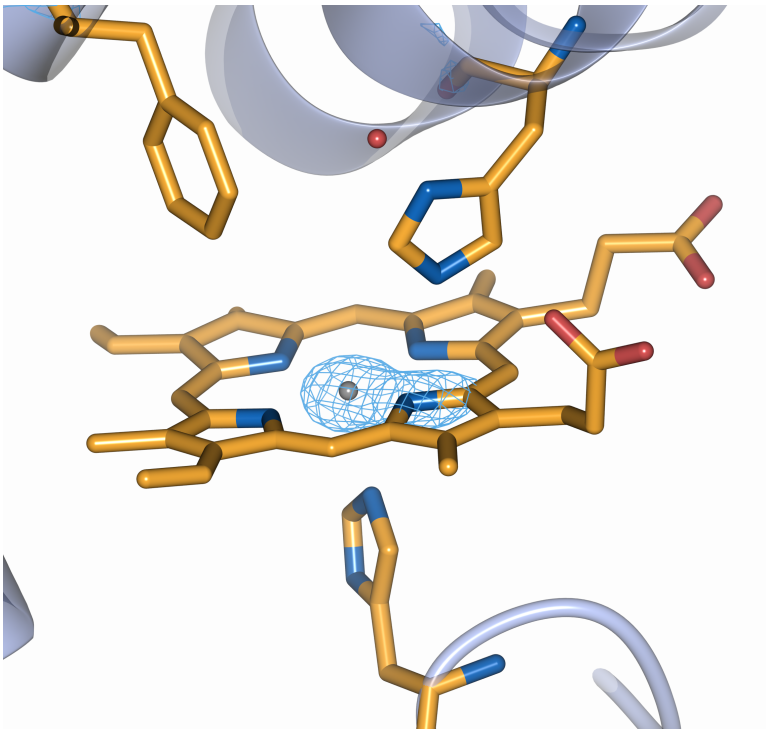

**Figure 6.24:** Representation of heme site A, where the hemichrome species is found. The electron density map $2F_o - F_c$ has been contoured at 4$\sigma$ to highlight the two positions of the iron, indicating a double conformation of the heme cofactor.

## 6.6 SFX structure of 5-bromoindole – DHP complex.

The compound 5-bromoindole (5-BrInd) belongs to the family of haloindoles, which are substrates for DHP (Barrios et al. 2014). These compounds have been isolated from several species of the *Hemichordata* phylum, that cohabitate with *A. ornata* in costal estuaries (Gribble 2000), and they are thought to act as repellents which have been evolutionary targeted by *A. ornata* and DHP. DHP possesses a distal binding pocket for small aromatic molecules (LaCount et al. 2000) that sets it apart from traditionally studied globins (e.g. haemoglobin, myoglobin), and other haem enzymes. The protein oxidises bromoindoles acting as a peroxigenase (Barrios et al. 2014), transferring an oxygen atom (from the $H_2O_2$) to the aromatic compounds.

As discussed previously, soakings were set up by adding 5-BrInd dissolved in DMSO to a final concentration in the batch of 50 mM. The soaking of this compound into large DHP crystals had not been reported. Soaking times before loading the microcrystals onto the chips was 4-5 minutes. The processing of the data for the complex followed the same workflow with *CrystFEL* as the one just shown for the DHP resting state dataset. The four measured chips for this complex were indexed using the same parameters in *indexamajig*. Diffraction data were joined in a single output file containing diffraction from 53,687 microcrystals, with cell parameter distributions shown in Figure 6.25, as calculated by the program *cell_explorer*.

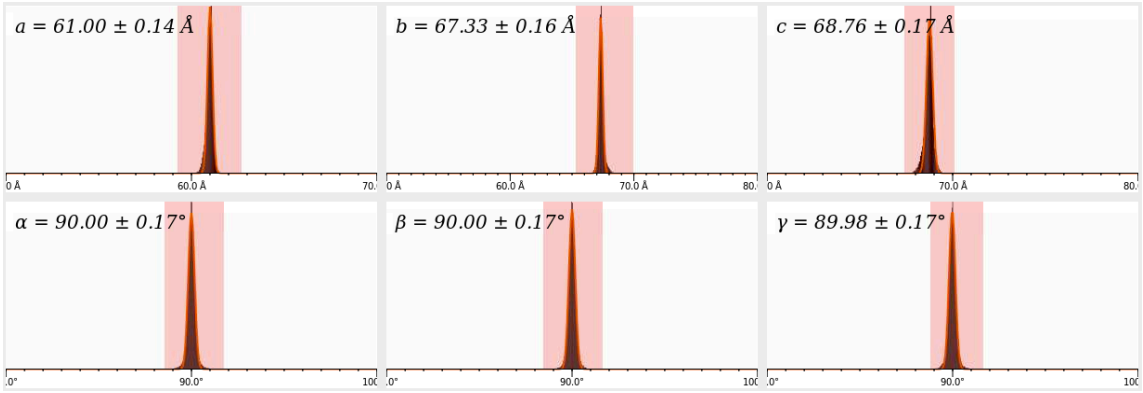


**Figure 6.25:** Distribution plots of cell parameters generated by *cell_explorer* for all the microcrystals measured for the DHP-5-bromoindole complex. Mean values and standard deviations are detailed for each parameter.

The merging process was performed with the program *process_hkl* within the script *merge.sh*, and the data quality statistics were generated with *stat.sh*. Resolution shells for the complex are shown in Figure 6.26, and data quality statistics CC and $R_{split}$ are shown in Figure 6.27 and Figure 6.28, respectively. The resolution cut-off applied for the dataset was at 1.85 Å, in agreement with both statistics. The final *mtz* file was generated with the script *create-mtz*.

```
[zlp66871@cs04r-sc-serv-41 stat]$ cat dhp-5bromo-all-scale-shells.dat
Center 1/nm  # refs Possible  Compl       Meas   Red   SNR    Std dev       Mean     d(A)    Min 1/nm   Max 1/nm
     1.133     1489     1490  99.93    2970987 1995.3 19.10     252.28     343.72     8.83       0.218      2.047
     2.313     1391     1391 100.00    1887874 1357.2 20.48     356.10     491.23     4.32       2.047      2.579
     2.766     1372     1372 100.00    1642914 1197.5 20.71     303.26     386.11     3.62       2.579      2.952
     3.101     1368     1368 100.00    1399478 1023.0 19.03     218.23     232.77     3.23       2.952      3.249
     3.375     1355     1355 100.00    1235540 911.8 17.42     156.93     161.17     2.96       3.249      3.500
     3.610     1352     1352 100.00    1225987 906.8 16.35     113.16     108.28     2.77       3.500      3.719
     3.817     1333     1333 100.00    1164942 873.9 14.97      83.35      78.75     2.62       3.719      3.915
     4.004     1354     1354 100.00    1135631 838.7 14.10      67.29      64.33     2.50       3.915      4.093
     4.175     1325     1325 100.00    1073650 810.3 12.44      48.38      45.48     2.39       4.093      4.257
     4.333     1338     1338 100.00    1042735 779.3 11.06      38.26      35.55     2.31       4.257      4.410
     4.481     1343     1343 100.00    1010742 752.6 10.08      30.49      28.03     2.23       4.410      4.552
     4.619     1336     1336 100.00     968691 725.1  8.57      22.54      20.15     2.17       4.552      4.686
     4.749     1329     1329 100.00     933425 702.4  7.33      18.96      15.48     2.11       4.686      4.812
     4.873     1317     1317 100.00     899678 683.1  6.09      11.95      10.67     2.05       4.812      4.933
     4.990     1328     1328 100.00     879936 662.6  5.05       8.30       7.60     2.00       4.933      5.048
     5.102     1315     1315 100.00     840775 639.4  3.91       6.30       5.28     1.96       5.048      5.157
     5.210     1335     1335 100.00     822607 616.2  2.89       3.91       3.36     1.92       5.157      5.263
     5.313     1316     1316 100.00     736942 560.0  1.95       2.44       2.01     1.88       5.263      5.364
     5.413     1334     1334 100.00     632731 474.3  1.34       1.88       1.37     1.85       5.364      5.461
     5.508     1309     1309 100.00     496076 379.0  0.86       1.53       0.91     1.82       5.461      5.556
```

**Figure 6.26:** Resolution shells for the merged dataset of the DHP-5-bromoindole complex.

```
[zlp66871@cs04r-sc-serv-41 stat]$ cat dhp-5bromo-all-scale-CC.dat
  1/d centre       CC       nref      d / A   Min 1/nm    Max 1/nm
     1.156  0.9857832       1489       8.65      0.265       2.048
     2.314  0.9850576       1392       4.32      2.048       2.579
     2.766  0.9890257       1372       3.62      2.579       2.952
     3.101  0.9889204       1366       3.22      2.952       3.249
     3.375  0.9885572       1355       2.96      3.249       3.500
     3.610  0.9901094       1352       2.77      3.500       3.719
     3.817  0.9894333       1335       2.62      3.719       3.915
     4.005  0.9876059       1353       2.50      3.915       4.094
     4.176  0.9858673       1324       2.39      4.094       4.257
     4.334  0.9859281       1339       2.31      4.257       4.410
     4.481  0.9785103       1342       2.23      4.410       4.552
     4.619  0.9814591       1336       2.17      4.552       4.686
     4.749  0.9672270       1330       2.11      4.686       4.813
     4.873  0.9655465       1316       2.05      4.813       4.933
     4.990  0.9435967       1328       2.00      4.933       5.048
     5.102  0.9172247       1315       1.96      5.048       5.157
     5.210  0.8604259       1335       1.92      5.157       5.263
     5.313  0.7308405       1316       1.88      5.263       5.364
     5.413  0.5940541       1334       1.85      5.364       5.461
     5.508  0.3816324       1309       1.82      5.461       5.556
```

**Figure 6.27:** Values of data quality statistic CC in the different resolution shells for the DHP – 5-bromoindole complex dataset.

```
[zlp66871@cs04r-sc-serv-41 stat]$ cat dhp-5bromo-all-scale-Rsplit.dat
  1/d centre Rsplit/%       nref      d / A   Min 1/nm    Max 1/nm
     1.156       5.72       1489       8.65      0.265       2.048
     2.314       5.65       1392       4.32      2.048       2.579
     2.766       5.10       1372       3.62      2.579       2.952
     3.101       5.56       1366       3.22      2.952       3.249
     3.375       5.75       1355       2.96      3.249       3.500
     3.610       6.19       1352       2.77      3.500       3.719
     3.817       6.59       1335       2.62      3.719       3.915
     4.005       7.09       1353       2.50      3.915       4.094
     4.176       7.69       1324       2.39      4.094       4.257
     4.334       8.30       1339       2.31      4.257       4.410
     4.481       9.97       1342       2.23      4.410       4.552
     4.619      10.53       1336       2.17      4.552       4.686
     4.749      13.01       1330       2.11      4.686       4.813
     4.873      15.42       1316       2.05      4.813       4.933
     4.990      19.30       1328       2.00      4.933       5.048
     5.102      25.47       1315       1.96      5.048       5.157
     5.210      34.60       1335       1.92      5.157       5.263
     5.313      53.01       1316       1.88      5.263       5.364
     5.413      74.71       1334       1.85      5.364       5.461
     5.508     122.62       1309       1.82      5.461       5.556
```

**Figure 6.28:** Values of data quality statistic Rsplit in the different resolution shells for the DHP – 5-bromoindole complex dataset.

The complex was solved by molecular replacement, and after an initial round of refinement, a large positive electron-density in the $F_o$ - $F_c$ difference map was observed in the distal pocket (Figure 6.29), into which the substrate

molecule was modelled. As can be observed in Figure 6.30, 5-BrInd establishes a hydrogen bond (3.01 Å) through the nitrogen $N_1$ of the indole with the $N_{\varepsilon 2}$ of His55, which is swung into the distal pocket to stabilise the binding of the ligand. In its turn, His55 is interacting through another hydrogen bond (2.89 Å) with one of the propionate groups. The substrate also presents a stacking interaction with Phe21 (Figure 6.31), which together with Phe60, are the only residues that change their conformation upon binding of 5-BrInd in comparison with the native structure. As it can be appreciated in Figure 6.31, both Phe sidechains flip parallel to the indole plane for stacking and to accommodate the aromatic compound in the pocket.

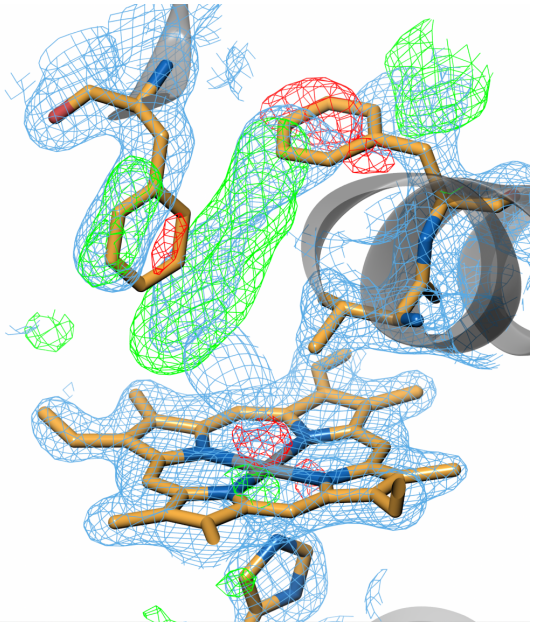

**Figure 6.29:** Detail of the distal pocket of the DHP – 5-BrInd complex structure after the first round of refinement with Refmac5 using the previously solved SFX ferric DHP as a model. A large positive feature in the $F_o - F_c$ difference map (contoured at 3σ) corresponding to the bound substrate can be observed. Note also the conformational change of Phe21 and Phe60 indicated as well by the difference map.

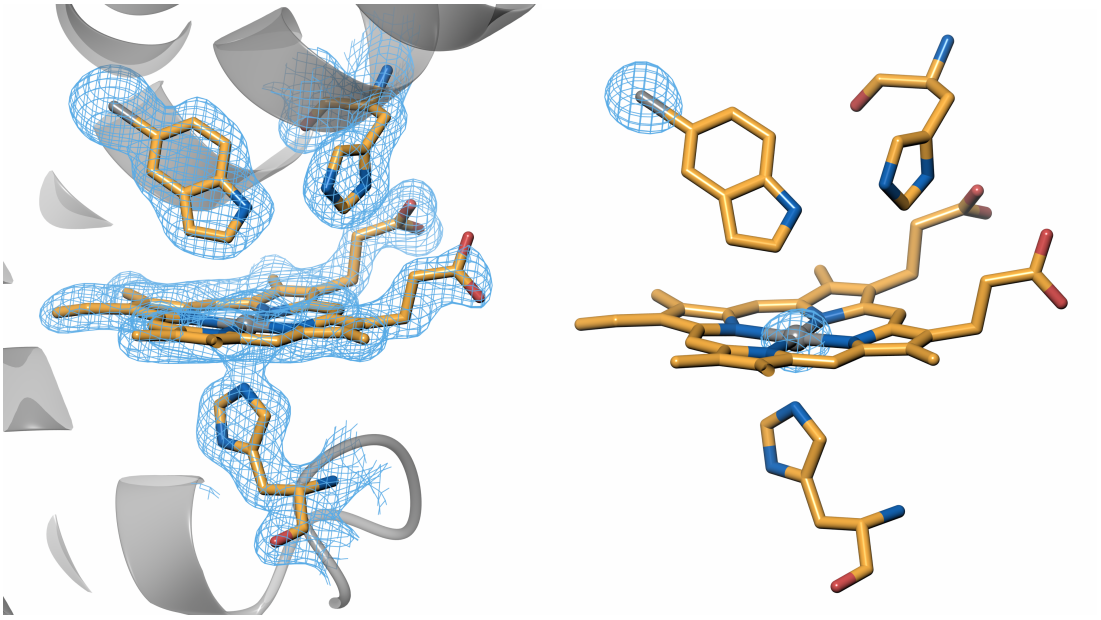

**Figure 6.30:** Substrate 5-bromoindole bound at the DHP-b heme site. The substrate molecule establishes a hydrogen bond interaction with His55, which in turn is also interacting with one of the propionate group. Electron density is contoured at 1$\sigma$ (0.276 $e^-/Å^3$) in the left panel, where the density and the heme environment can be appreciated. In the right panel, the electron density has been contoured to 5$\sigma$ to appreciate the density for the bromine atom of the substrate molecule, and also the iron from the heme group is observed.

The ligand is surrounded by a hydrophobic distal environment with Phe21, Leu100, Phe35, Phe24, Val59, Ile20, Phe 60, Leu104, Ala17, and Met63 in close proximity (Figure 6.32), providing the conditions to favour the binding of the aromatic compound. 5-BrInd is found in both pockets of the monomers forming the dimer in the asymmetric unit of the crystal, in hemeB with full occupancy, and in hemeA with partial occupancy (0.7). HemeA appears to present the hemichrome species partially, noticed by the presence of a lower occupancy alternate conformation This suggests that the substrate is bound to heme A in those molecules where the hemichrome is not formed, and there is enough space at the distal site for 5-BrInd to bind.

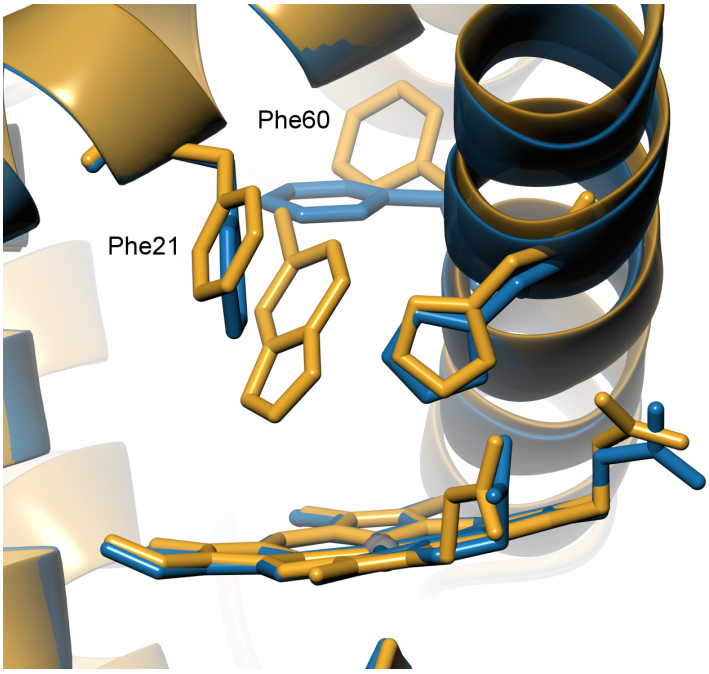


**Figure 6.31:** Superposition of DHP-b ferric structure (in blue) and its complex with 5-bromoindole (in orange). A conformational change of Phe21 and Phe60 can be appreciated to accommodate the substrate in the distal pocket.

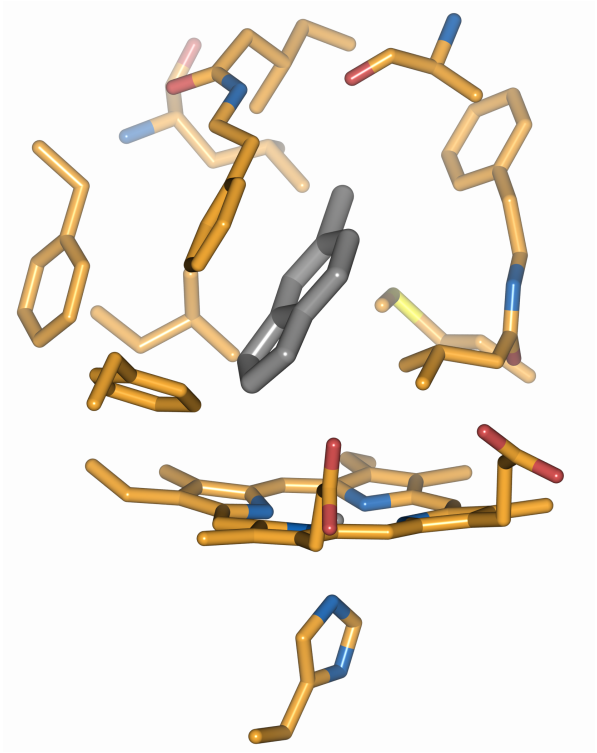

**Figure 6.32:** Hydrophobic environment surrounding the substrate 5-bromoindole bound at the distal pocket. In the left panel, the hydrophobic residues surrounding the ligand in the pocket are represented as cylinders.

## 6.7 SFX structure of 2,4-dichlorophenol (DCP) – DHP complex.

The organic compound DCP belongs to the large family of halogenated phenols that are also substrates for DHP (Barrios et al. 2014), which oxidises them carrying out its peroxidase function. This kind of compounds are secreted as repellents by other marine organisms sharing the benthic ecosystem with *Amphitrite ornata* (Lebioda et al. 1999), which evolved a distal binding pocket in DHP to be able to metabolise them.

Soaking of DCP into DHP microcrystals was performed in the same way as for 5-BrInd, to a final concentration of 50mM in the batch and soaked for 3-4 min. However, in this case, the diffraction patterns observed while running the first chip indicated that the crystals were being damaged in the soaking process. Thus, soaking concentration was lowered down to 5mM, at which point diffraction looked as good as with unsoaked crystals.

### *6.7.1 Data processing*

Processing of the SFX data for this complex was carried in *CrystFEL* just as detailed previously for resting state DHP. The three measured chips for the DCP complex were indexed using the same parameters in *indexamajig*. Diffraction data was joined in a single output file containing diffraction from 40,396 microcrystals, with cell parameter distributions shown Figure 6.33, as

calculated by the program *cell_explorer*. In this case, two populations of crystals can be appreciated, one of them being predominant.

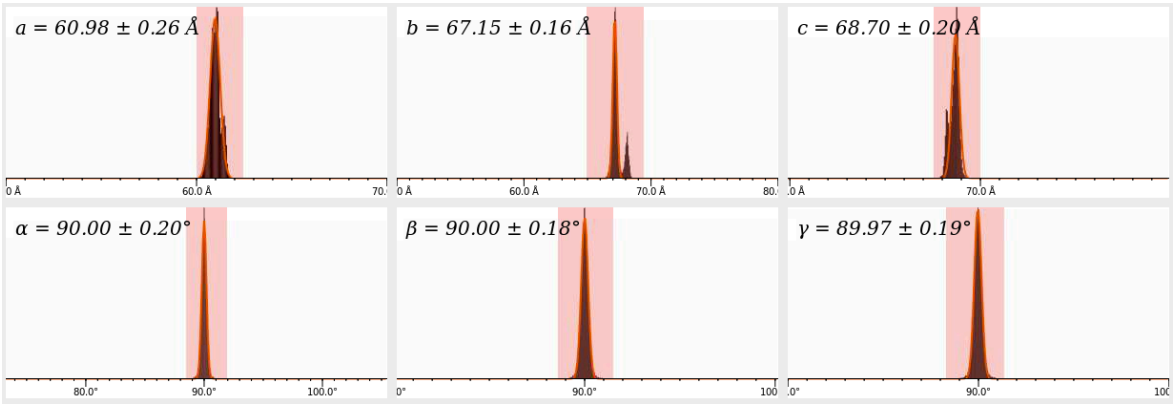


**Figure 6.33:** Distribution plots of cell parameters for the DHP – DCP dataset generated by the program cell_explorer within CrystFEL.

Merging of the data was performed using the program *process_hkl*, which was run executing the script merge.sh. Data quality statistics were generated by running the script stat.sh. Resolution shells for the complex are shown in Figure 6.34, and data quality statistics CC and $R_{split}$ are shown in Figure 6.35 and Figure 6.36, respectively. The resolution cut-off was decided to be applied at 1.85 Å, in agreement with both statistics. The final *mtz* file was generated with the script *create-mtz*.

```
[zlp66871@cs04r-sc-serv-41 stat]$ cat DHP-DCPtest--scale-shells.dat
Center 1/nm  # refs Possible  Compl       Meas   Red   SNR    Std dev       Mean     d(A)    Min 1/nm   Max 1/nm
     1.156     1484     1484 100.00    2122775 1430.4 16.53    1877.77    1928.84     8.65       0.264      2.048
     2.314     1390     1390 100.00    1344879 967.5 17.89    2132.71    2357.75     4.32       2.048      2.579
     2.766     1365     1365 100.00    1134019 830.8 17.69    1492.37    1601.40     3.62       2.579      2.952
     3.101     1356     1356 100.00     944543 696.6 16.35     875.94     851.49     3.22       2.952      3.249
     3.375     1355     1355 100.00     833314 615.0 15.68     534.06     556.69     2.96       3.249      3.500
     3.610     1340     1340 100.00     820189 612.1 15.23     361.14     368.08     2.77       3.500      3.719
     3.817     1344     1344 100.00     793329 590.3 14.38     263.19     266.70     2.62       3.719      3.915
     4.005     1338     1338 100.00     759145 567.4 13.61     216.76     220.03     2.50       3.915      4.094
     4.176     1318     1318 100.00     724042 549.3 12.46     152.38     156.34     2.39       4.094      4.257
     4.334     1346     1346 100.00     715749 531.8 11.24     114.27     117.21     2.31       4.257      4.410
     4.481     1317     1317 100.00     681562 517.5 10.08      95.55      94.59     2.23       4.410      4.552
     4.619     1337     1337 100.00     674213 504.3  8.73      69.22      68.76     2.17       4.552      4.686
     4.749     1323     1323 100.00     650308 491.5  7.47      59.95      51.10     2.11       4.686      4.813
     4.873     1307     1307 100.00     631037 482.8  6.34      37.66      36.33     2.05       4.813      4.933
     4.990     1337     1337 100.00     630567 471.6  5.22      26.09      25.33     2.00       4.933      5.048
     5.102     1317     1317 100.00     606636 460.6  4.08      18.80      17.61     1.96       5.048      5.157
     5.210     1325     1325 100.00     594163 448.4  2.85      14.13      11.54     1.92       5.157      5.263
     5.313     1306     1306 100.00     543573 416.2  2.06       8.72       7.07     1.88       5.263      5.364
     5.413     1323     1323 100.00     470717 355.8  1.38       5.84       4.55     1.85       5.364      5.461
     5.508     1315     1315 100.00     383588 291.7  0.87       4.75       3.01     1.82       5.461      5.556
```

**Figure 6.34:** Resolution shells for the DHP – DCP SFX dataset, generated by the CrystFEL script stat.sh.

```
[zlp66871@cs04r-sc-serv-41 stat]$ cat DHP-DCPtest--scale-CC.dat
  1/d centre       CC       nref      d / A   Min 1/nm    Max 1/nm
     1.157  0.9880483       1484       8.65      0.265       2.048
     2.314  0.9881828       1390       4.32      2.048       2.580
     2.766  0.9824039       1365       3.62      2.580       2.952
     3.101  0.9870668       1356       3.22      2.952       3.249
     3.375  0.9852114       1355       2.96      3.249       3.500
     3.610  0.9863679       1340       2.77      3.500       3.719
     3.817  0.9832961       1344       2.62      3.719       3.915
     4.005  0.9835393       1338       2.50      3.915       4.094
     4.176  0.9797637       1318       2.39      4.094       4.257
     4.334  0.9788825       1346       2.31      4.257       4.410
     4.481  0.9789659       1317       2.23      4.410       4.552
     4.619  0.9736526       1337       2.17      4.552       4.686
     4.749  0.9673772       1323       2.11      4.686       4.813
     4.873  0.9596850       1307       2.05      4.813       4.933
     4.990  0.9356424       1337       2.00      4.933       5.048
     5.102  0.8852492       1317       1.96      5.048       5.157
     5.210  0.6702980       1325       1.92      5.157       5.263
     5.313  0.5831725       1306       1.88      5.263       5.364
     5.413  0.5275908       1323       1.85      5.364       5.461
     5.508  0.3020130       1315       1.82      5.461       5.556
```

**Figure 6.35:** Values for data quality statistic CC in the different resolution shells for the DHP – DCP complex dataset.

```
[zlp66871@cs04r-sc-serv-41 stat]$ cat DHP-DCPtest--scale-Rsplit.dat
  1/d centre Rsplit/%       nref      d / A   Min 1/nm    Max 1/nm
     1.157       6.36       1484       8.65      0.265       2.048
     2.314       6.05       1390       4.32      2.048       2.580
     2.766       6.31       1365       3.62      2.580       2.952
     3.101       6.78       1356       3.22      2.952       3.249
     3.375       6.89       1355       2.96      3.249       3.500
     3.610       6.86       1340       2.77      3.500       3.719
     3.817       7.33       1344       2.62      3.719       3.915
     4.005       7.50       1338       2.50      3.915       4.094
     4.176       8.15       1318       2.39      4.094       4.257
     4.334       8.73       1346       2.31      4.257       4.410
     4.481       9.70       1317       2.23      4.410       4.552
     4.619      11.01       1337       2.17      4.552       4.686
     4.749      13.09       1323       2.11      4.686       4.813
     4.873      15.03       1307       2.05      4.813       4.933
     4.990      18.61       1337       2.00      4.933       5.048
     5.102      25.18       1317       1.96      5.048       5.157
     5.210      40.23       1325       1.92      5.157       5.263
     5.313      51.64       1306       1.88      5.263       5.364
     5.413      77.38       1323       1.85      5.364       5.461
     5.508     128.33       1315       1.82      5.461       5.556
```

**Figure 6.36:** Values for data quality statistic Rsplit in the different resolution shells for the DHP – DCP complex dataset.

## *6.7.2 Structure solution and description*

In the same was as with the 5-BrInd complex, after processing the data, an unambiguous electron density for a bound ligand was found in a similar position in the distal pocket (Figure 6.37). Furthermore, a conformational change of Phe21 and Phe60 was also obvious from them electron density maps. The ligand molecule was generated with the help of the ligand builder within CCP4i2 (Lidia), where a sketch could be draw for DCP (Figure 6.38) and restrains generated in an output *cif* file. The ligand was modelled into the extra electron density and, after a round of refinement with Refmac5, R factors were $R_{work}$=0.185 and $R_{free}$=0.223, and the molecule fitted well the electron density (Figure 6.39).

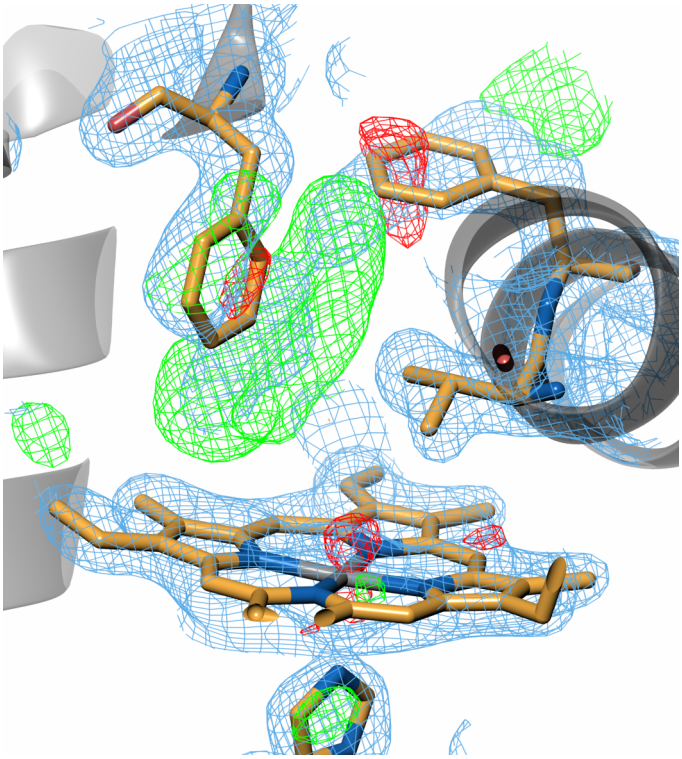

**Figure 6.37:** Detail at the distal sited of the DHP-DCP complex after an initial round of refinemen. A clear large positive electron density in the $F_o$ – $F_c$ difference map (contoured at 3σ) can be observed above the heme plane, corresponding to the bound ligand. Note also the conformational change of Phe21 and Phe60 also suggested by the difference map.

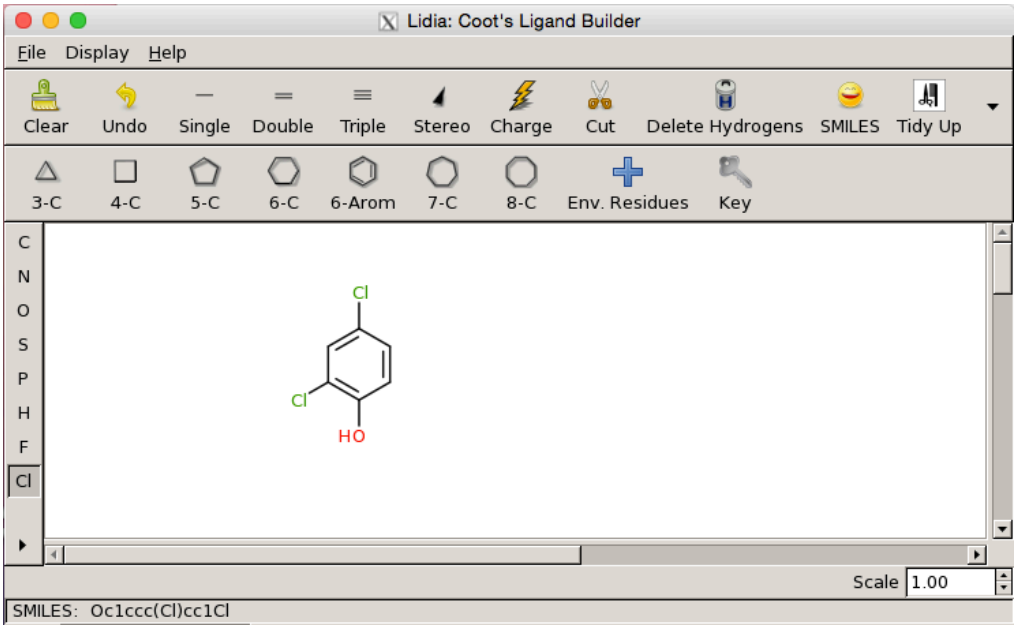


**Figure 6.38:** Sketch generated with the ligand builder within ccp4i2 for the ligand 2,4-dichlorophenol (DCP). Once finalized and saved, restrains would be also generated and stored in an output *cif* file.

We could be certain the bound compound was DCP, since if the contour level was increased, density for the two chlorides in the positions 2 and 4 was clear (Figure 6.39). As discussed, Phe21 and Phe60 flipped to end up in a position parallel to the aromatic ring of DCP, making room for it to bind and providing stacking forces to stabilize it in the pocket. The conserved distal residue His55 is again important to stabilize the substrate molecule, establishing a stronger hydrogen bond interaction (2.70 Å) than for 5-BrInd, this time through the alcohol group in DCP (Figure 6.39). In its turn His55 establishes another hydrogen bond interaction at 2.98 Å with one of the propionates, a feature as well seen in the other complex.

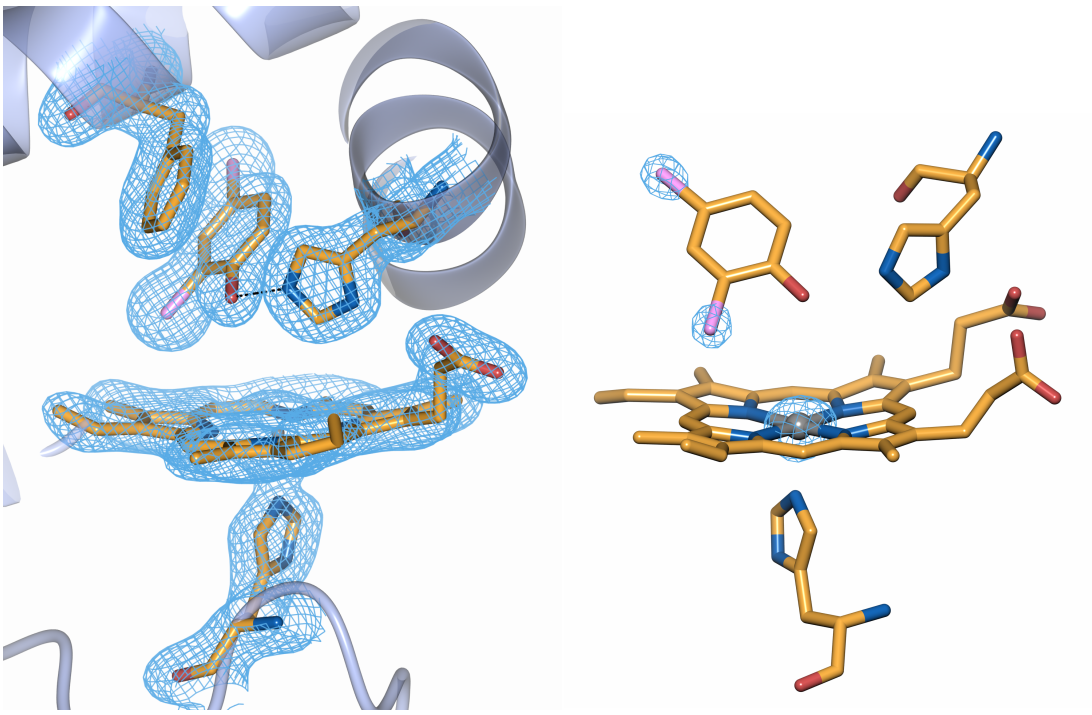

**Figure 6.39:** SFX structure of the DHP-b complex with 2,4-dichlorophenol (DCP) at 1.85 Å. In the left panel, monomer B of the dimer is shown, where the ligand bound at the pocket can be observed, stabilized by stacking interactions with Phe21, and a hydrogen bond interaction with His55. Chloride atoms are shown in pink, and the electron density map is contoured at 1σ

(0.267 $e^-/Å^3$). In the right panel, the $2F_o - F_c$ map is contoured at $5\sigma$ to highlight the chloride atoms, with the heme iron as the strongest density.

A superposition between the DCP complex and the SFX resting state of DHP was carried out (Figure 6.40), featuring the flip of the phenylalanines from a ‘closed gate’ to an ‘open gate’ conformation, enabling the substrate to bind. A partial hemichrome can again be observed in heme A, where the binding of 2,4-dichlorophenol is also partial (around 60% occupancy), suggesting that the compound just binds to resting state ferric species.

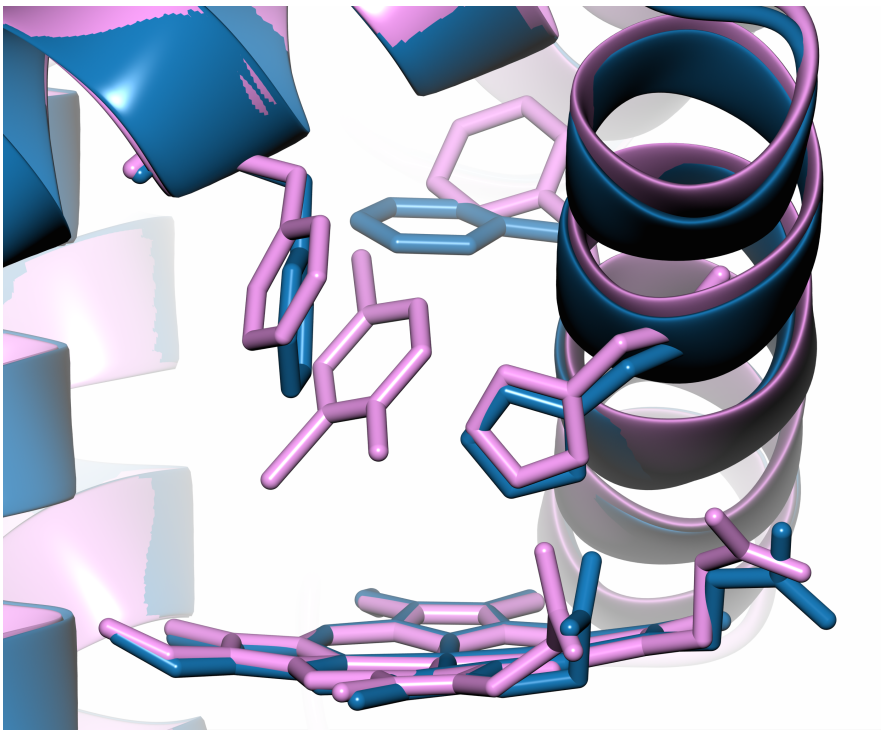

**Figure 6.40:** Superposition of DHP-b distal pocket surrounding a heme B for the resting state (in blue), and the DCP complex (in pink). Conformational changes of Phe21 and Phe60 can be appreciated, allowing the access and binding of aromatic substrates at the distal site.

## 6.8 Discussion

### *6.8.1 Serial crystallography applied to DHP*

Serial crystallography has been applied for DHP at X-ray sources with different characteristics (synchrotron, Laue, and XFEL), and proved to be an efficient technique to obtain room temperature structures of this haem enzyme. This approach is time and sample efficient compared to other SX methods (e.g. liquid jets), just needing around 10,000 microcrystals, measured among 2-4 chips for DHP samples (depending on how well loaded they are), each one of them run in 15 min. So, in a time of around one hour a complete dataset for a SFX or SSX structure could be obtained.

The three structures obtained from DHP ferric crystals by serial crystallography, i.e. SSX, SLX, and SFX, represent the first reported room-temperature structures for DHP. A damage-free resting structure in the case of SFX, and low-dose structures from SSX and SLX. The advantages of measuring room-temperatures structures are the observation or alternative conformations of side and main chains (Fraser et al. 2011; Russi et al. 2017), since the protein dynamics are in great measure suppressed by the use of cryotemperatures in conventional crystallography. In this way, the DHP RT structures provide a more biologically relevant model of the protein, which is closer to the conditions where the enzyme normally functions.

Synchrotron serial crystallography (SSX) provided a DHP structure very similar to the SFX damage-free structure, and even with more detail due to the highest resolution achieved. This could be due to the higher energy used in the

SSX experiments at Diamond (12,800 eV) compared to SFX at SACLA (10,000 eV), and by the fact that in SSX the detector can be situated closer to the sample than in SFX experiments. Apart from giving a similar outcome as an XFEL, the SSX methodology also permits to perform a dose-series approach (as seen for DtpAa in Chapter 4) and follow photo-reduction or catalytic processes.

In the intact SFX structure at 1.85 Å of resting state DHP, at haem B a positive electron density in the difference ($F_o$ – $F_c$) map could be observed, suggesting the partial occupancy of a distal water bound to the ferric haem iron. In contrast, in the SSX structure with a resolution of 1.50 Å, a clear full occupancy distal water was found at this same position. The reason for this observation is not completely clear, since in the SFX structure full occupancy waters are well defined in other parts of the structure, and it could be even down to the batch sample used. The batch of DHP used for the SSX structure, had been set up for a much longer time (~3 months), and that could explain for instance the higher occupancy of the hemichrome in the structure, since the protein had been substantially more time in contact with the PEG, with subsequent higher partial denaturation of the protein. There is also the possibility that the electron density found at the distal haem in the SSX structure corresponds to a molecular oxygen bound of an oxy-ferrous species formed by photo-reduction (Kekilli et al. 2017). However, this is unlikely since such a species has been never observed when measuring large crystals with much higher absorbed doses (refer to Chapter 5). The resonance Raman spectroscopic solution data from ferric DHP (Figure 5.10; D'Antonio et al.,

2010) indicates the presence of a high-spin six-coordinate species, which would agree with a distal water coordinating the haem.

### *6.8.2 Comparison between RT DHP structures from SFX, SSX and SLX, and the cryostructure of DHP.*

The different room-temperature SX structures for DHP were compared with the cryostructure of DHP at 0.99 Å presented in Chapter 5, to observe the possible conformational changes between them. At the same time, a comparison could be drawn as well among the RT structures from the different SX methods used (i.e. SFX, SSX and SLX). The superposition of all four DHP structures is shown in Figure 6.41, where they can be observed adopting very similar conformations, with the main differences observed in the loops surrounding the haem cofactors (represented as cylinders), especially the loops around haem A on the right side of the figure. Therefore, monomer B presents in general a better superposition for all structures. The overall and per monomer rmsd values between all the structures with respect to the damage-free SFX structure are detailed in Table 68. These conformational changes in the loops around the haem groups are likely to be due to the presence of the hemichrome at different states/distribution (and the protein chain around it) in each structure. Remarkably, the hemichrome is observed in monomer A for all the SX structures measured from microcrystals, whereas is present in monomer B for the cryostructure from a large DHP crystal. Thus, it suggests that the method of crystallization used (e.g. batch or hanging drop vapour diffusion) could affect its formation.

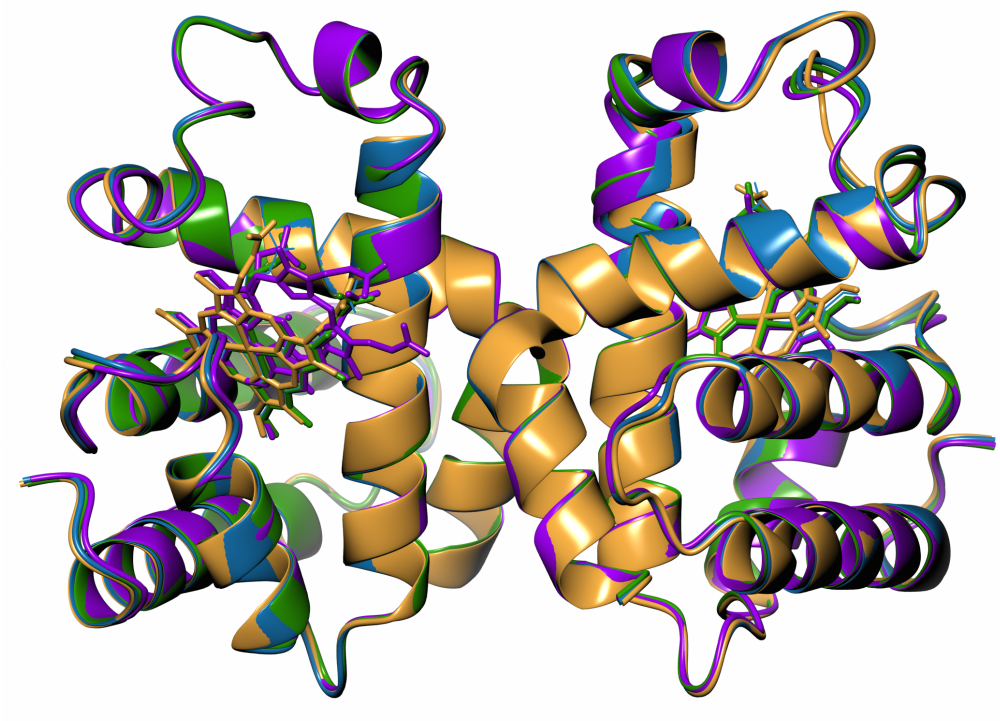

**Figure 6.41:** Superposition between the different RT SFX structures and the cryo structure for DHP. The SFX structure from SACLA is shown in blue, the SSX structure from Diamond in orange, the SLX structure from APS in green, and the 100 K structure measured at SLS in purple. All dimers are represented as ribbons, with monomer B on the left and monomer A on the right. Haem groups are represented as cylinders. Rmsd values between the structures are detailed in Table 68 .

**Table 6.8:** Values of rmsd differences from the superposition between the RT SFX DHP structure (used as reference) and the DHP cryostructure and other SX structures (i.e. SSX and SLX). The superpositions were performed using the SSM method with CCP4-MG (McNicholas et al. 2011).

| Structure | Rmsd overall (Å) | Rmsd monomer A (Å) | Rmsd monomer B (Å) |
|---|---|---|---|
| **100 K DHP** | 0.40 | 0.36 | 0.27 |
| **RT SSX DHP** | 0.30 | 0.34 | 0.17 |
| **RT SLX DHP** | 0.29 | 0.27 | 0.21 |

In addition, a ‘worm’ representation of the all four structures coloured by temperature factor (Wilson B-factor) was also performed and is shown in Figure 6.42, with notable differences observed. The cryostructure of DHP presents the lowest B-factor values, whereas the RT SFX structure (measured at the highest temperature of 28ºC), in contrast, presents the highest. For all structures, the highest B-factor values, i.e. more disorder, are present again in the loops surrounding the haem groups. This observation confirms the higher protein dynamics and conformational variety observed in RT structures (Fraser et al. 2011), which are suppressed at cryotemperatures.

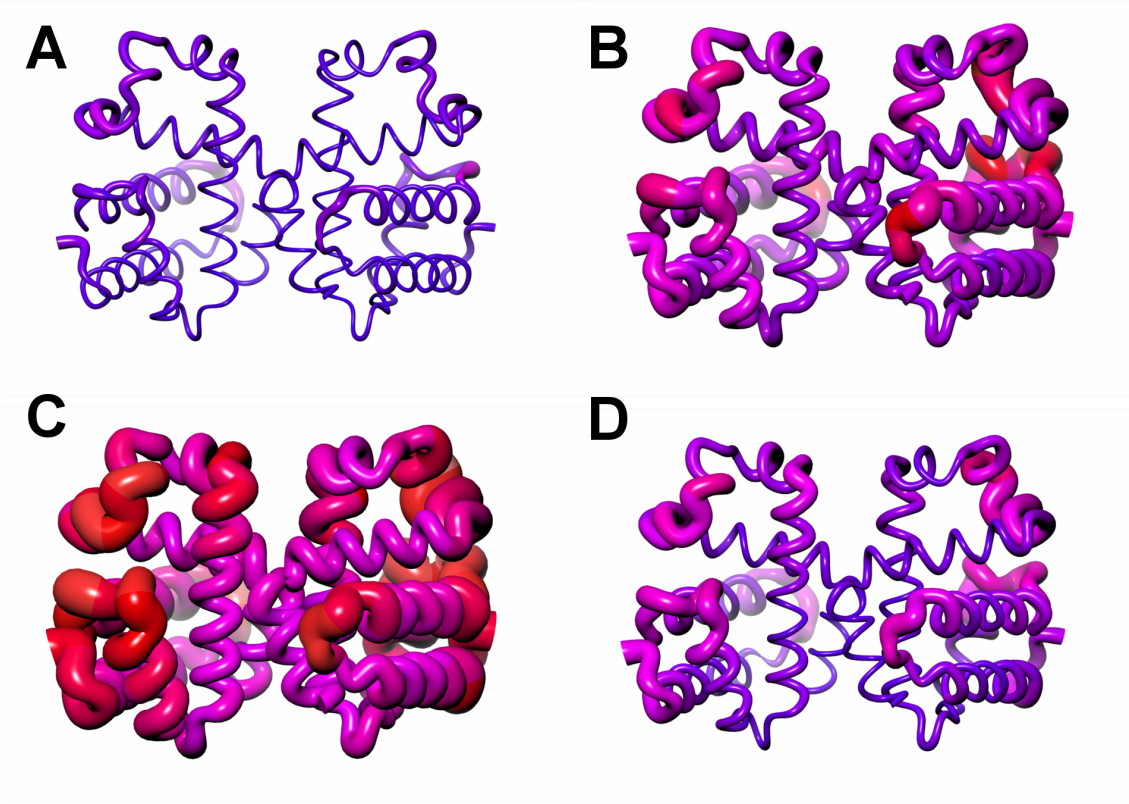


**Figure 6.42:** Representation as ‘worm’ of the main chain for DHP structures, coloured by temperature B-factor. A) DHP structure at 100 K from large crystal; B) room-temperature SSX structure; C) room-temperatrue SFX structure; and D) room-temperature SLX structure.

### *6.8.3 The hemichrome species*

Another of the features observed in the all the SX structures of DHP was the double conformation of haem B, with presence of a hemichrome species, which has been previously observed in haemproteins (Riccio et al. 2002). This hexacoordinated state was also observed in the atomic resolution cryo-structure (at a different monomer) described in Chapter 5, and the reasons for its formation were supposed to be the utilization of glycerol as a cryoprotectant or the result of photoreduction of the heme. With its observation also in SX approaches, these two factors can be ruled out, since all the experiments were carried out a RT, and the SFX structure represents the damage-free state of DHP. However, the hemichrome species could still be an artefact observed in crystal structures by the effect for example of the PEG (Molchanova, 1981) used as precipitant, but it could also have a biological relevance. In fact, signal for a low-spin six-coordinate species can be observed in resonance Raman spectra of DHP-B in solution (Figure 5.10B), that would be consistent with the presence of hemichrome in solution. In any case, further investigation upon this species presence in solution (e.g. by spectroscopic means) would be needed to confirm whether the hemichrome could play a biological role or not.

As it has been discussed, in the superposition of the different structures (RT SX, and cryo) from ferric DHP crystals, it has been observed how the bigger structural differences were observed in the loops surrounding the haem sites (Figure 6.41). This is probably due to different stages of the double conformation, including the hemichrome, observed at this site depending on the sample and methodology used.

### *6.8.4 Soaking experiments with DHP microcrystals*

Ligand soaks were efficiently achieved using DHP microcrystals. Two complexes with known substrates (i.e. DCP and 5-BrInd) were obtained in a highly efficient manner, in this way highlighting the possibility of performing high-throughput ligand screening with a serial crystallography approach. As discussed, this SX methodology is sample and time efficient, since, for instance, it just needs 2-3 DHP big batches (300-400 μl each) measured in 3-4 chips over a single hour of beamtime to deliver a novel structure, when the conditions are optimized. Future plans include to estimate the minim number of patterns from microcrystals needed to observe the presence of a ligand bound in the electron-density $F_o$ - $F_c$ difference map (publication in progress). Furthermore, the structure for the complex of DHP with 5-BrInd, could not be obtained by soaking in large DHP crystals (Carey et al. 2018), which indicates that the use of microcrystals can be useful in cases with solubility and/or diffusion issues.

A common feature observed in both complex structures with DHP has been the conformational change of the phenalanines Phe21 and Phe60 to allow access of the soaked compound to the distal pocket, and providing stacking forces to stabilize it. We have named this observation as the ‘Phenalanine gates’ mechanism, since there is a transition from a closed to an open conformation upon binding of the ligand. Recently, it has been also observed for a third time in a pyrrole complex measured at 100 K form a DHP single crystal (my unpublished data). This suggests that this is the conformational mechanism that DHP has develop in order to accept a wide

range of aromatic substrates in the pocket, for subsequent oxidation with $H_2O_2$. In the mentioned pyrrole structure, a surprising presence of a glycerol molecule was observed (from cryo-protection) bound at the iron heme. This supports the assumption that the *Phe-gates* are closed when not exposed to small aromatic ligands, since no glycerol has been observed to be bound before in other DHP cryo-structures. The fact that pyrrole is a notably small compound, left room in the open pocket for binding of another ligand present in the crystallization condition.

### *6.8.4.1 Comparison with previous DHP – DCP complexes*

DHP crystals were previously soaked with the compound DCP in Dr. Reza Guiladi´s lab (North Caroline State University, USA) by Dr. Leiah Carey, obtaining two structures with different binding modes of the substrate (PDB codes 5wmz and 5wn6). The two complexes were named as DCPα and DCPβ depending on the proximity of the compound to the α and β edges of the heme cofactor. A superposition of these two complexes with our SFX DHP-DCP structure is shown in Figure 6.43. The superposition was carried out with monomer B, since is the one which presents full occupancy of the ligand, and the rms differences with DCPα and DCPβ are 0.41 and 0.38 Å respectively. It can be appreciated how the binding mode of DCPα is the one adopted by the compound in our structure, just with minor differences. A hydrogen bond at 2.6 Å (2.7 Å for SFX structure) is established between the alcohol group of DCP and His55, which is swung in into the pocket. For complex DCPβ, the binding orientation of the compound is substantially different, this time interacting

through a hydrogen bond with Thr56 at 2.7Å, while His55 swings out of the pocket to an open conformation and allow space for the ligand.

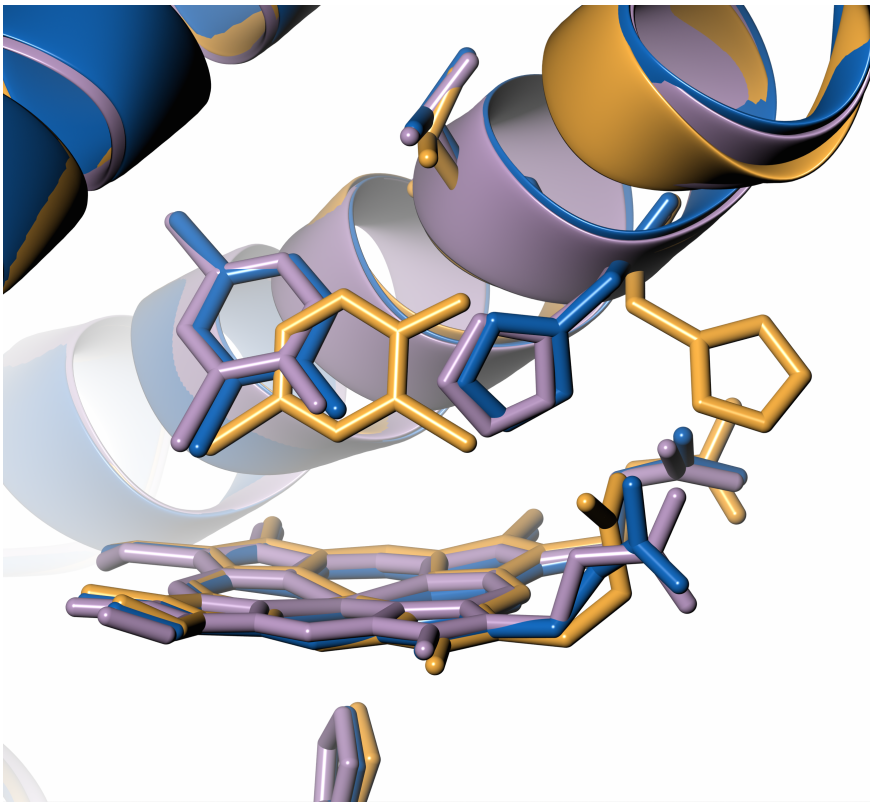

**Figure 6.43:** Superposition of the DHP - DCP complexes. In blue, the SFX structure presented in this chapter; in purple DHP-DCPα (PDB code 5wmz), and in orange DHP-DCPβ (PDB code 5wn6), both determined by Dr. Guiladi´s group. The binding orientation for the DCPα is the same as the one found in our SFX structure. DCPβ, in contrast, presents a different orientation, with conformational changes of His55 and Thr56.

### *6.8.5 Comparison/synergy of different serial crystallography methods for DHP*

In this chapter, different sources have been used with the same methodologic approach for DHP, and thus a comparison can be established between the structures obtained (as shown in Figure 6.41). Having structural information measured in different manners can provide a wider view and

validate the structure of a protein. Especially in the case of metalloproteins, such as DHP, particularly prone to photo-reduction, and able to initiate redox reaction feeding by the beam. In this way, we can explore the synergies of using different sources and methodologies, which can contribute with different nuances to create a better idea of how the protein works. Data quality in the case of DHP was best for the SSX method, which delivered a model very similar to the damage-free structure from an XFEL, and provided more defined features such as the observation do the haem-bound water. Regarding SLX, it provided a structure of limited resolution (~2 Å), but could be obtained from a significantly smaller number of crystals (50-100) with fast exposures (ps).

Experiments should be conducted to exploit the advantages and characteristics of each particular X-ray source. For instance, with an XFEL pulse and a liquid jet delivery system, time-resolved pump-probe experiments in the femtosecond scale can be conducted (Barends et al. 2015), investigating the protein motion following activation of the enzyme. However, for this aim the production of 10-20 μm microcrystals in large quantities suitable for jetting is needed.

In contrast, SSX and SFX using fixed targets, has been proven very efficient (both time and sample wise) for the determination of low-dose or damage-free structures. Other advantages of using chips is their tolerance of heterogeneous crystal sizes and easy loading process. Furthermore, SSX permits the performance of dose-series to characterize radiation damage processes or catalytic reaction in the ms time-scale. In the case of Serial Laue crystallography (SLX), it allows to perform time-resolved serial crystallography

experiments in the ps time-scale range, a possibility that could be explored in future experiments with DHP crystals.

Regarding the loading process, the two chip systems used in this work (i.e. ‘Diamond’ and ‘Heidelberg’ chips) showed similar efficiency. The measuring time was much faster for the Diamond chips (15 min) due to the higher pulse frequency of 30 Hz at SACLA and Diamond, comparted with 2 Hz at APS. Thus, the sample stayed much longer in the ‘Heidelberg’ chips (1-2 hours) at RT, which could have affected DHP crystals (grown at 4ºC), although they have shown to be stable at RT in the different experiments conducted.

As an example of complementarity of methods seen in this chapter, for the resting state of DHP, the damage-free SFX structure from SACLA suggested the presence of a distal axial water hexacoordinating the heme iron; and with the measurement of SSX data reaching higher resolution from the crystals, this distal water could be clearly appreciated and its presence confirmed. In the case of the Serial Laue dataset, its lower resolution did not allow much detail. The most interesting feature for the SLX structure is the double conformation of His55, which might the result of photo-reduction by the beam, or could be due to a different sample preparation.

In conclusion, in this chapter we have shown how RT structures of a given protein, in our case DHP, can be efficiently measured with a fixed-target serial crystallography approach, that can be implemented at sources of different characteristics. This methodology is especially suited to the measurement of intact species of metalloproteins with SFX or low-dose

structures with SSX; and it has been also proved its efficiency to determine ligand-bound complex structures upon soaking of the microcrystals with relevant compounds.

# Chapter 7 Structural characterization of nitric oxide dioxygenase (NOD) in its lipid-bound form

## 7.1 Introduction

As discussed in Chapter 1, the main reaction catalysed by NOD is the oxidation of NO· into $NO_3$ (Gardner et al., 1998), thought to function as a defence mechanism for the bacteria to neutralise harmful concentrations of nitric oxide secreted by the host organism, with general reaction:

$$2NO\cdot + 2O_2 + NADPH \rightarrow 2NO_3^- + NADP^+ + H^+$$

This protein target was one of the initial candidates to develop structural movies of enzyme catalysis, since it is a haem redox enzyme, which could use the solvated electrons generated by the X-ray beam to carry out its catalytic mechanism. At the same time, we could monitor the state of the active site by obtaining spectroscopic information from the haem chromophore group. We aimed thus to run the dioxygenase reaction *in crystallo* driven by the X-ray beam, and ideally measure different structures or 'frames' of a structural movie where changes at the haem site could be observed.

However, difficulties in its expression and crystallization as well as the complex nature of the catalytic mechanism that we aimed to characterise, resulted in initially slow progress with this protein system. Nevertheless, some of the aspects that we have learned about NOD are very interesting and will be shown and detailed throughout the chapter.

## 7.2 Cloning of His-tagged NOD

Difficulties were experienced in the purification and crystallization of NOD following and adapted version of the protocol in the literature (Ilari et al., 2002), which consisted of a three-step chromatographic purification: a DEAE column, followed by an hydroxylapatite column (Bio-Rad), and a final S200 size-exclusion column. The purity of the isolated protein using this protocol was low and crystallization attempts at the literature crystallization condition were unsuccessful. Therefore, it was decided to sub-clone the tag-less NOD gene (originally synthesized by Epoch Bioscience Inc.) encoded in a pMMB503EH plasmid, into a pET28a plasmid (maps for both plasmids included in Chapter 3) to introduce a His-tag at the N-terminal side of the protein, and make its purification easier.

Primers were designed and ordered (Eurofins) to introduce the NOD gene cutting the pET28a plasmid at the restriction enzymes sites NdeI and HindIII. NOD gene was amplified with the designed primers performing a PCR reaction, in this way generating the NdeI and HidIII restriction sites, one at each end of

the gene. The PCR amplification reaction was set up with the following components:

- 1 μl of NOD template gene (15 ng/μl)
- 1 μl of Forward primer (100 ng/μl)
- 1 μl of Reverse primer (100 ng/μl)
- 1 μl of dNTP mix (10 mM)
- 3 μl of PCR buffer (x10)
- 1 μl of DMSO
- 21.5 μl of deionised water
- 0.5 μl of polymerase Pfu.

The polymerase was added at the end, before setting up the thermal cycles (described in Chapter 3) and perform the PCR reaction. The production of the NOD amplicon was assessed running 5 μl of the final reaction mixture in a 1% agarose gel for 60 min at 50 V (Figure 7.1). A clear band can be observed for the amplified NOD gene at ~ 1.2 kb, including the extended ends with restriction sites.

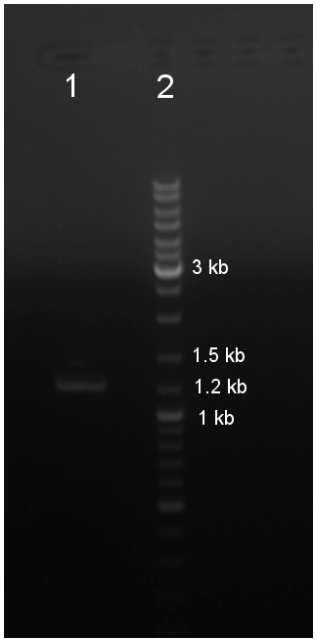


**Figure 7.1:** Agarose gel (1 %) of the NOD amplicon (first lane) obtained by PCR reaction, showing an approximate length of ~1.2 kb, as assessed with the Gene Ruler DNA marker (second lane), where the length in kb for some relevant bands is specified.

Both the amplicon and plasmid pET28a were digested with restriction enzymes NdeI and HindIII for 2.5 hours at 37ºC, with the corresponding buffer (buffer Red, Thermo-Fisher). The digested samples were run in a 1% agarose gel, to separate them form enzymes and reaction components. The bands were cut and the dna fragments extracted using the QIAGEN gel extraction kit. The NOD insert was ligated into the pET28a digested backbone setting up a ligation reaction using a 1 to 3 ratio (stoichiometric), backbone to insert. The reaction mix was composed of:

- 4.6 μl of pET28a backbone (50 ng, 0.02 pmols)
- 1 μl of NOD insert (37.5 ng, 0.06 pmols)
- 2 μl of ligase buffer (10x, Thermo-Fisher)
- 11.4 μl of nuclease-free water (Thermo-Fisher)
- 1 μl of T4 DNA ligase (Thermo-Fisher)

The ligation reaction was carried out at RT overnight. In the next morning 5 μl of the product were transformed into competent cells XL1-blue (protocol detailed in Chapter 3), and plated in LB-agar plates with kanamycin resistance (50 μg/ml; resistance encoded in pET28a) and incubated at 37ºC overnight. Single colonies grew, and one of them was picked and grown in LB with Kan (50 μg /ml) at 220 rpm and 37ºC for 18 hours, to then perform a miniprep (QIAGEN kit) and extract the putative plasmid containing the insert. A 5 μl aliquot of the purified plasmid was sent for sequencing (Source Biosciences), and the raw data generated analysed with the software FourPeaks (nucleobytes.com) to confirm the insertion of the NOD gene (Figure 7.2).

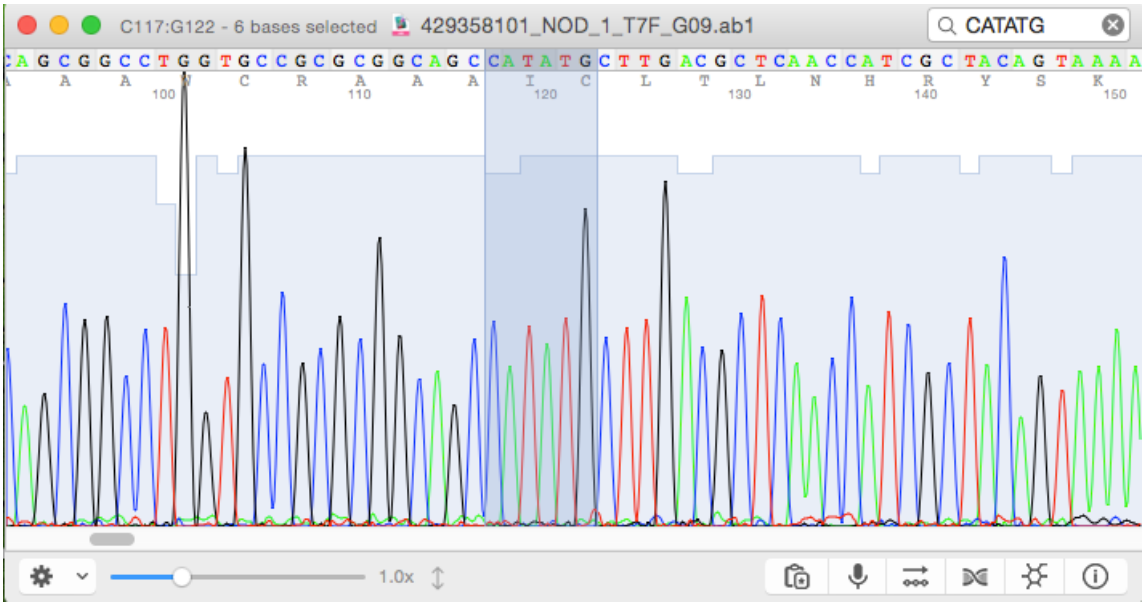


**Figure 7.2:** Sequencing data for the NOD clone shown by the program FourPeaks. The ligated pET28a plasmid with sequence using primers for T7 promoter within the vector. In the image, data sequence using the T7 forward (F) primes is shown, and the highlighted area corresponds to the restriction enzyme NdeI, after which the protein encoding sequence starts.

## 7.3 Protein expression and purification

The pET28a plasmid containing the His-tagged NOD was transformed into expression *E. coli* cells BL21-DE3, which were plated in LB-agarose dishes supplemented with Kan (50 μg/ml), and incubated at 37ºC overnight. Single colonies were picked the following day and grown overnight in LB media with Kan (50 μg/ml) at 37ºC and 220 rpm. These cultures were used to inoculate three 2 litre flasks (10 ml of culture per flask) with 1 litre of media (supplemented with Kan) each for a large expression. The expression of recombinant NOD was induced by addition of IPTG to the cultures. The flasks were grown at 37ºC and 200 rpm until an $OD_{600}$ ~ 0.8 was reached, when 0.5 mM of IPTG (all stated values corresponding to final concentrations) were added to the solution. At this point 0.1 mM hemin (Acros organics) and 50 μM iron citrate (Sigma) were also added as haem precursor and iron source. Expression was carried out overnight (~18-20h) at a temperature of 25°C and 200 rpm. Next morning cells were pelleted after centrifugation at 5,000 rpm for 20 min. Pellets were resuspended in lysis buffer consisting in: 20 mM imidazole, 50mM Tris pH 8.0 and 500 mM NaCl (Buffer A). Cell lysis was performed using an Emulsiflex C-5 (Avestin), running through the cell suspension three times at 12,000 psi to ensure complete lysis. The lysate was centrifuged at 18,000 rpm during 40 min to get rid of cellular debris. The purification protocol consisted in two steps: capture of the protein with a Nickel-sepharose His-trap column (GE-healthcare), and polishing of the eluted protein with size exclusion chromatography. Lysate in buffer A, was run through the nickel column, and the retained protein band was eluted with buffer B (500mM imidazole, 50 mM Tris pH 8.0, 100mM NaCl) applying a gradient with the AKTA purifying platform.

Fractions containing the NOD protein, which presented a red colour, were pooled and concentrated until ~2ml, for injection into a G75 size exclusion column (GE-healthcare). The elution buffer, in which the column was previously equilibrated, was: 20 mM HEPES pH 7.5, 100 mM NaCl. A single peak was observed in the chromatogram, and protein fractions were assessed for purity by a 15% SDS-PAGE gel (Figure 7.3). Pure fractions were pooled and NOD was oxidised with an excess of potassium ferricyanide, followed by buffer exchange in the same buffer with a PD-10 column to get rid of the oxidising agent. The protein solution was concentrated until ~25 mg/ml, using the extinction coefficient at 403 nm of 88.6 $mM^{-1}cm^{-1}$ for ferric NOD (Helmick et al. 2005). Finally, the protein was aliquoted and stored at -80ºC.

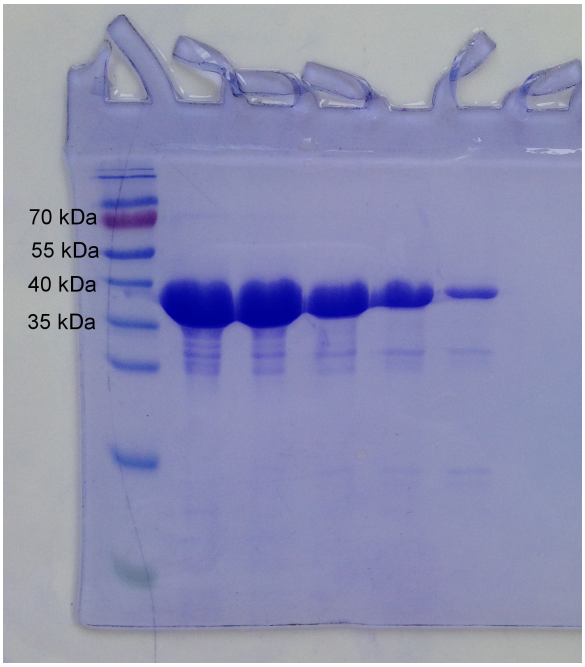


**Figure 7.3:** SDS-PAGE gel of fractions containing NOD-His, eluted from a G75 size exclusion column (GE-healthcare). The protein runs to the same distance in the gel as the 40 kDa marker, consistent with the molecular weight of NOD (43.8 kDa).

## 7.4 Crystallisation of lipid-bound protein

As discussed in the introduction of NOD in Chapter 1, this protein had been previously crystallized and its structure determined by Ilari et al (2002), who reported a lipid-free NOD. Although flavohemoglobins are known to bind lipids with variable affinities depending on the case (Ollesch et al., 1999; Rinaldi et al., 2006), in this case the lipid was removed by Ilari and co-workers due to difficulties to crystallise the lipid-bound protein. Also, they spectroscopically characterized using UV-visible absorption the lipid-free and lipid-bound forms by titration experiments (Figure 7.4; Bonamore, Farina, et al., 2003).

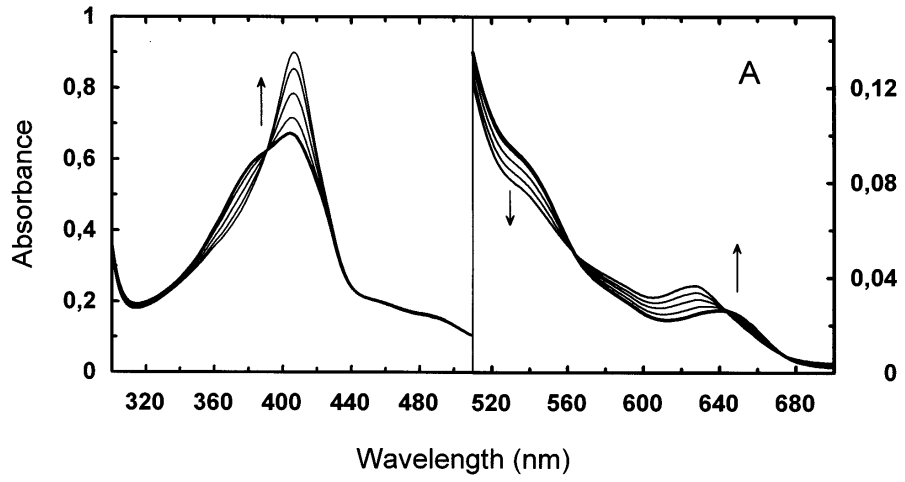


**Figure 7.4:** Titration experiments with NOD and oleic acid. The protein is initially in its lipid-free form (thick line), and oleic acid is titrated observing clear spectral changes until stabilizing at the fully lipid-bound form of NOD. Note the characteristic shoulder observed in Soret band at ~375 nm, and the shift of the CT band from 640 to 625 nm upon binding of the lipid. (Bonamore, Farina, et al. 2003)

Our expressed recombinant NOD stock was inspected by UV-vis spectroscopy, and found to have the spectral features corresponding to the lipid-bound form, as it can be observed in Figure 7.5. This was expected as the

lipid is naturally tightly bound to NOD, and can only be removed by means of an extra hydrophobic chromatography step (Ilari et al. 2002).

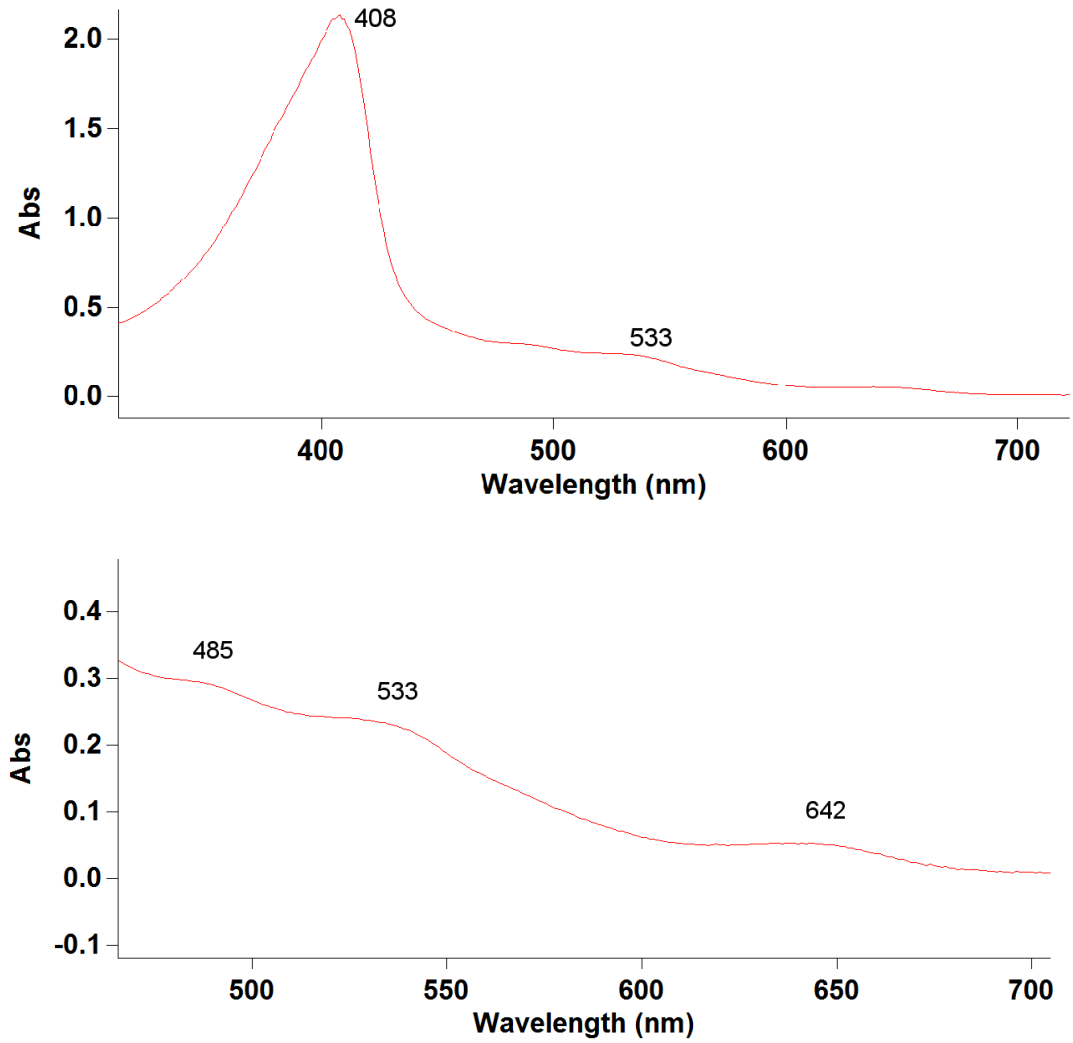


**Figure 7.5:** UV-vis spectrum measured from ferric NOD in solution (upper panel), showing features corresponding to a lipid-bound state. In the lower panel an inset is shown of the 500-700 nm spectral region for a clearer visualization or the β/α and CT bands. The spectrum was acquired using a Cary-60 spectrophotometer (Agilent).

Commercial crystallisation screens with both the his tagged and the cleaved protein were performed using a Gryphon robot (Art Robbins). After 4-5 days, no hits were observed for the cleaved protein, but for the His-tagged NOD there were a notable amount of hits among the crystallization screen

plates. All the hit conditions showed a similar crystal morphology, which consisted of thin layer plates stacked in bunches (Figure 7.6).

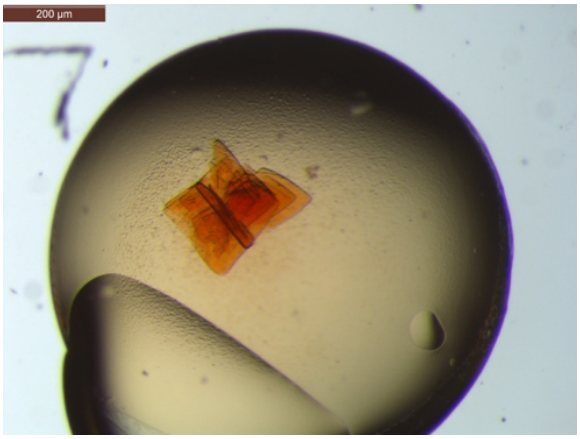


**Figure 7.6:** View of one of the optimized drops for NOD for the condition containing 20% PEG 3350, 200 mM $MgCl_2$ and 100 mM BisTris pH 5.5. Thin plate-shaped crystals can be appreciated, which grew in stacked bunches after 4-5 days at 18ºC.

## 7.5 Single-crystal online spectroscopic characterisation

These plate-shaped crystals were cryo-protected in 25% PEG200-supplemented crystallization condition, and flash-cooled in liquid nitrogen to take them to SLS for measurement. The standard methodology of spectroscopy → dataset → spectroscopy was carried out. In this case, as it could be expected by the thin and pale nature of the crystals, UV-vis spectra acquired with the on-axis MS3 spectrophotometer at SLS (Pompidor et al. 2013) were of good quality and spectral features could be observed in the range 450-650 nm. The area around 400 nm, where the Soret band usually appears, was close to saturation and nuances could not be appreciated, that could tell us about the binding and/or oxidation state of the protein (e.g. maximum and shape of Soret band). In Figure 7.7, single-crystal online spectra are shown for ferric NOD before (in blue), and after (in orange) X-ray data

collection. In the oxidised spectrum, we can observe a wide band at around 480 nm, which corresponds to the oxidised Flavin (Bonamore et al. 2001),therefore we can be certain about the oxidation state of both the haem and FAD cofactors. The β/α bands for ferric NOD are not distinguishable and merged in a wide band around 540nm. The CT band could be observed at ~620 nm wavelength, which was consistent with a lipid-bound form of NOD (Bonamore, Farina, et al. 2003), and confirmed the state of the protein within the crystal. After collection of a complete X-ray diffraction dataset, a further online UV-vis spectrum of the crystal was measured, where clear signs of photo-reduction could be appreciated (Figure 7.7); for instance, the shift of the Soret band, and appearance of sharp β/α bands, consistent with a hexa-coordinated ferrous haem. Thus, a partially photo-reduced species could be expected when solving the structure, since the initial and final states of the protein were different (Kekilli et al, 2017).

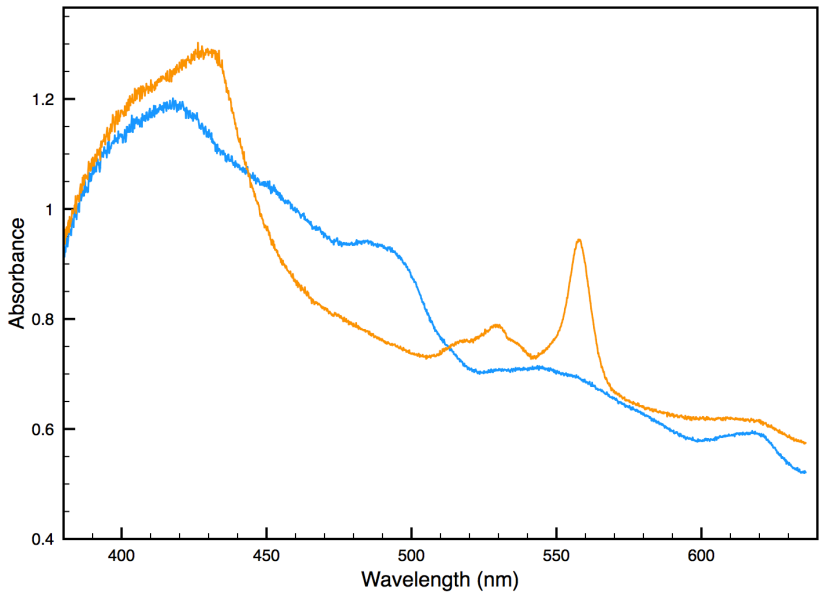


**Figure 7.7:** Single-crystal online UV-vis spectra of a ferric NOD crystal before (in blue) and after (in orange) exposure to X-rays for collection of a diffraction dataset. Measurements were carried out using the MS3 on-axis micro-spectrophotometer (Pompidor et al., 2013), equipped

at beamline X10SA at SLS. Note the clear spectral changes upon photo-reduction of the protein.

## 7.6 Data collection and determination of NOD lipid-bound structure.

Data collection from an NOD ferric crystal was performed at beamline X10SA at SLS at 0.8 Å wavelength and using a Pilatus-6M detector (Dectris), reaching a resolution limit of 1.93 Å (Table 71). As shown in the previous section, single-crystal online UV-vis spectra were acquired before and after X-ray diffraction at a pre-established point (Figure 7.7). Diffraction patterns were indexed and integrated with *XDS* (Kabsch 2010), and scaled and merged using *Aimless* (Evans & Murshudov 2013) within the CCP4 suite. Structure solution was carried out by molecular replacement with *Phaser* (McCoy et al. 2007), using as a search model the lipid-free NOD structure from the literature (PDB code 1gvh). A solution was found with two NOD motifs, however, a considerable part of the protein model would not agree with the experimental electron density; just in some regions, such as around the haem active site, the calculated model would fit the electron density. A new conformation of the protein could be then presumed. Therefore, an automatic model building job was run using the program *Buccaneer* (Cowtan 2006), giving as inputs the experimental reflections, the phases from the MR solution, and the NOD sequence. *Buccaneer* traced several main-chain polypeptide stretches by finding the position of the C$\alpha$ of each residue, and side chain conformations were also determined in some cases. Model building was carried out using

*Coot* (Emsley et al., 2010) to extend and join the different stretches and form the protein chains, as well as model side chains of numerous protein residues. The linker and tag region were found to be mostly ordered, and several residues were modelled manually. Refinement of the structure was carried out with *Refmac5* (Murshudov et al. 2011), with model building in between cycles with *Coot* to model, for instance, waters, ligands, and correct protein conformers. Data collection and refinement statistics for the structure are shown in Table 71.

**Table 7.1:** Data collection and refinement statistics for the dataset measured from a ferric lipid-bound NOD crystal, at beamline X10SA at SLS (Villigen, Switzerland), using a wavelength of 0.8 Å, and a Pilatus-6M detector (Dectris).

| | **Lipid-bound NOD** |
|---|---|
| **Space group** | C2 |
| **Unit cell (Å, °)** | a=90.72, b=90.10, c=115.22; β=95.67 |
| **Resolution (Å)** | 45.14 – 1.93 (1.97 – 1.93) |
| **$R_{merge}$** | 0.099 (1.233) |
| **Unique reflections** | 69,047 (4,412) |
| **I/σ(I)** | 6.8 (0.9) |
| **$CC_{1/2}$** | 0.996 (0.329) |
| **Completeness (%)** | 99.6 (99.9) |
| **Multiplicity** | 3.4 (3.3) |
| **$R_{factor}$ (%)** | 20.6 |
| **$R_{free}$ (%)** | 25.3 |
| **RMSD bond lengths (Å)** | 0.017 |
| **RMSD bond angles (º)** | 1.97 |
| **Ramachandran favoured (%)** | 96.1 |
| **Absorbed dose (MGy)** | 0.55 |

Once the structure was solved, NOD was found forming a dimer in the asymmetric unit as shown in Figure 7.8, in contrast with the published structure

(Ilari et al. 2002), where NOD existed as a monomer in the asymmetric unit. It was noted that the residues from the extended region were establishing contacts with monomers from neighbouring dimers, in this way helping to stabilise the crystallisation of the protein (Figure 7.9). In each of the monomers is clearly distinguishable the separation of the protein in two domains: the globin domain, all alpha-helical and containing the haem cofactor; and the oxidoreductase domain, mainly beta-sheets and harbouring the Flavin cofactor. At the haem active site of monomer A (Figure 7.10), the canonical His85 is found at the proximal site coordinating the iron, and the distal site is dominated by Tyr29, which is stabilising through a hydrogen bond (2.80 Å) an axial ligand at the haem distal site, which has been modelled as molecular oxygen ($O_2$). Thus, this species would be the oxy-ferrous NOD, induced by photo-reduction by the X-ray beam, consistent with online UV-vis spectroscopy (Figure 7.7). At the distal pocket Gln53, Leu57 and Val98 are also found, with Gln53 establishing a weak hydrogen bond with Tyr29 at 3.04 Å, and helping to stabilise its conformation. The described haem environment differs remarkably from the one found in the lipid-free NOD form (Figure 7.10), where the apolar Leu57 was found dominating the empty distal pocket above the haem iron. Furthermore, Tyr29 and Gln53, which were supposed to have an active role in catalysis (from mutants and biochemical experiments), were found in a second coordination sphere relatively far from the distal site. Therefore, we propose that the here reported structure is that of the lipid-bound form of NOD, supported by spectroscopic evidence and a different conformation of the protein, including a different distal environment with Tyr29 and Gln53 playing a

main role stabilising haem ligands. However, the lipid molecule could not be found in the electron density map, and would be presumably disordered.

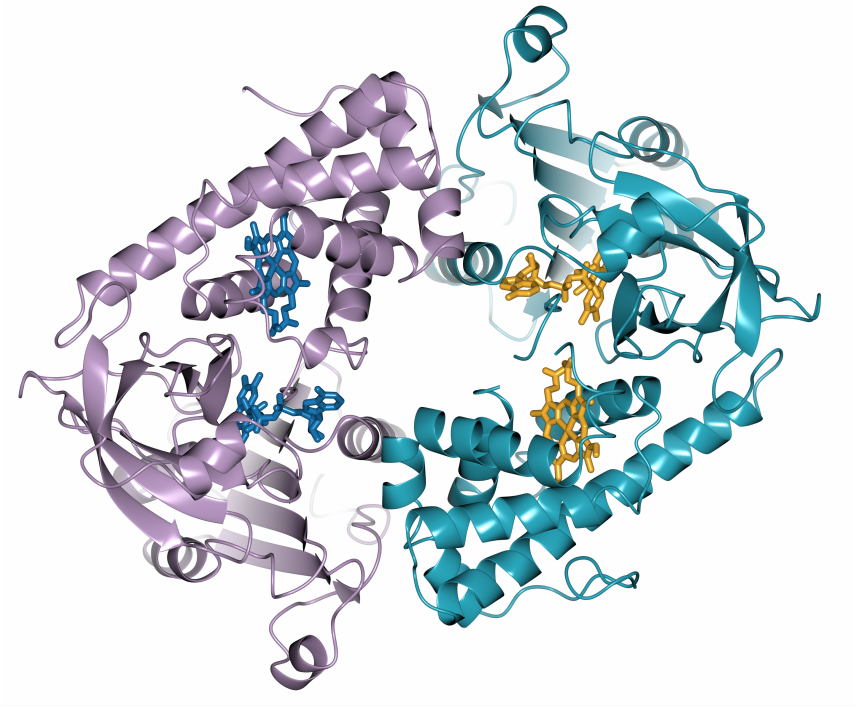

**Figure 7.8:** Overall view of the NOD dimer found in the asymmetric unit of the lipid-bound structure. Haem and FAD cofactors are shown as cylinders (blue and orange) embedded in the protein structure.

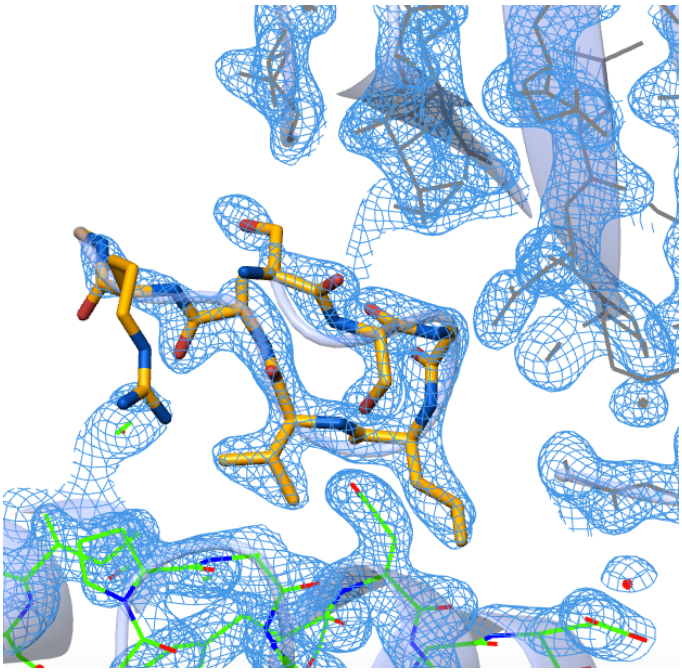

**Figure 7.9:** Representation of the ordered residues of the tag and linker (in yellow) added at the N-terminal side of NOD, where it can be appreciated how they establish contacts between

the monomer they belong to (in green) and the same monomer of the neighbor dimer (in grey). Secondary structural motifs are also shown for both monomers.

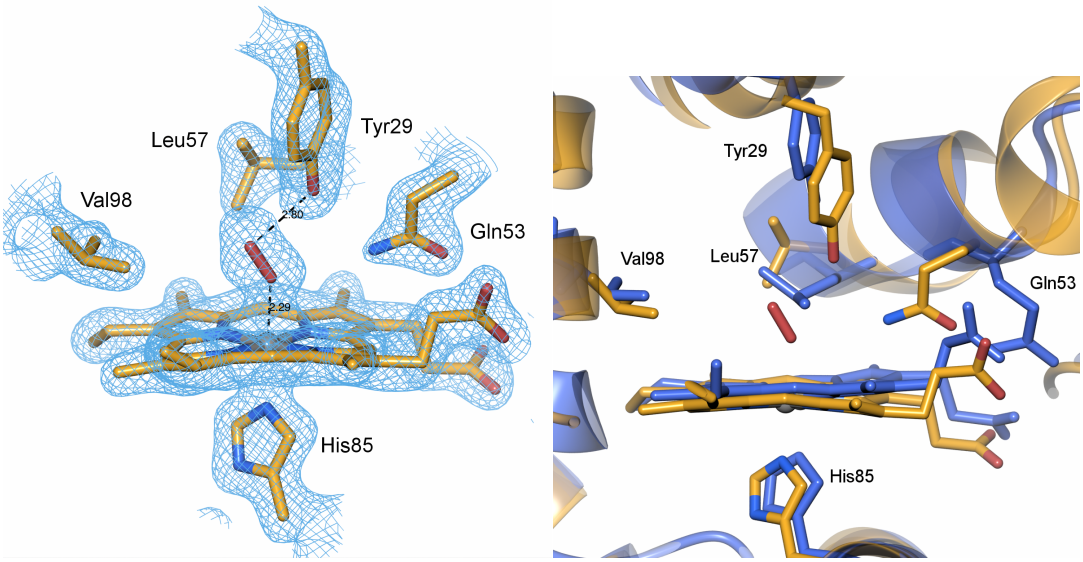


**Figure 7.10:** Haem environment of one of the NOD monomers. On the left panel, active site at monomer A of the lipid-bound NOD structure presented in this work. Distal Tyr29 can be observed stabilising a molecular oxygen ligand ($O_2$) at 2.80 Å, which in its turn is coordinating the iron at 2.29 Å. The observed oxy-ferrous species is the result is the result of photo-reduction of the protein by the X-ray beam, as monitored by online UV-vis spectroscopy with the MS3 microspectrophotometer at beamline X10SA, SLS. Conformation of distal residues is shown, as well as the proximal His85. Electron density is shown as a blue mesh and contoured at 1σ (0.2562 $e^-/Å^3$). On the right panel, the haem environment for the lipid-free NOD structure (PDB code 1gvh) is shown. Note the different conformations adopted by distal residues in both cases.

### *7.6.1 In crystallo catalysis in NOD crystals?*

At monomer B of the 1.93 Å NOD structure, a slightly different conformation of the haem environment was found, where the Gln53 (and the loop containing it) was not as well ordered as in monomer A (Figure 7.11). An interesting feature could be appreciated at the distal site – positive electron density in the difference map was observed after modelling in an oxygen molecule, suggesting the presence of a bigger ligand (Figure 7.11). This was also supported by the shift of Tyr29, which was situated this time at 3.3 Å of the modelled $O_2$, allowing more space at the distal site. Interestingly, the

density had a triangular shape, such as the product from the dioxygenase reaction: $NO_3^-$. Therefore, a nitrate molecule was modelled at the distal site instead of the molecular oxygen, and after a cycle of refinement with *Refmac5* fitted convincingly the electron density in the $2F_o - F_c$ map (Figure 7.12), showing no extra density in the $F_o - F_c$ difference map. This observation of turnover in the NOD crystal, would imply the existence of catalytic activity *in situ* driven by the X-ray beam, as seen in other redox enzymes (Horrell et al. 2016; Horrell et al. 2018; Schlichting et al. 2000). Similar shaped ligands such as acetate were ruled out, as not present in the crystallization solution (20% PEG3350, 200 mM $MgCl_2$, 100 mM BisTris pH 5.5), protein and purification buffers (20 mM HEPES pH 7.5, 50 mM NaCl) or cryoprotectant (25% PEG200). The B-factor values for the iron and ligand are shown in Table 72. However, no nitric oxide was supplemented to the crystal, so the issue remains of where does it come from and how does it reach the active site and produce nitrate. This matter will be discussed in the last section of the chapter.

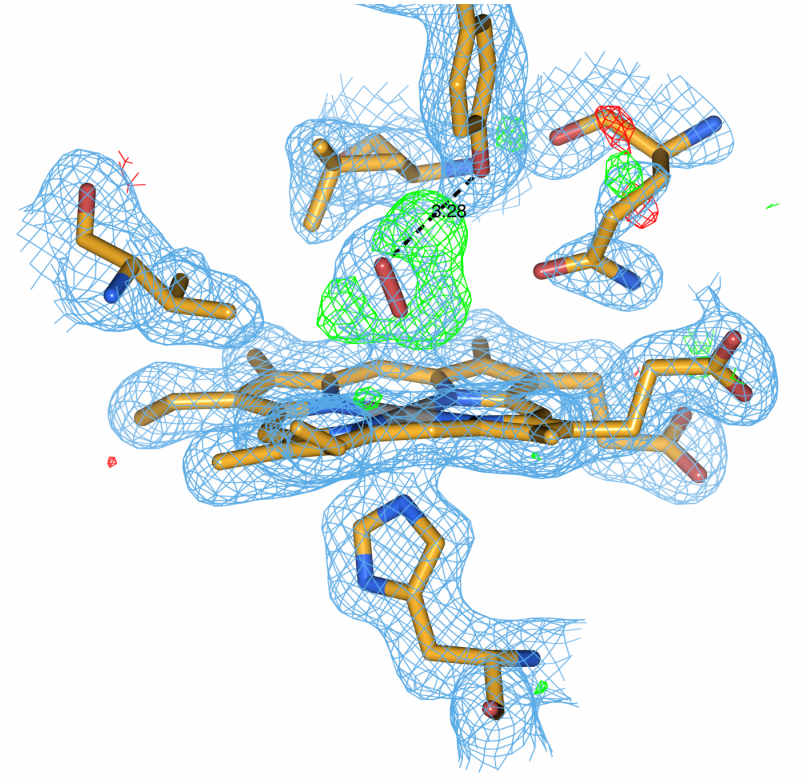


**Figure 7.11:** Representation of the preliminary haem site B in the NOD structure from ferric crystals. After a round of refinement (Refmac5, Murshudov et al., 2011) with a full-occupancy oxygen molecule modelled as haem axial ligand, a large positive electron density in the $F_o$ – $F_c$ difference map with triangular shape can be observed at this position (green mesh, contoured at 3σ), suggesting the present of a larger ligand. Note the disorder shown by Gln53 which presents a partial occupancy.

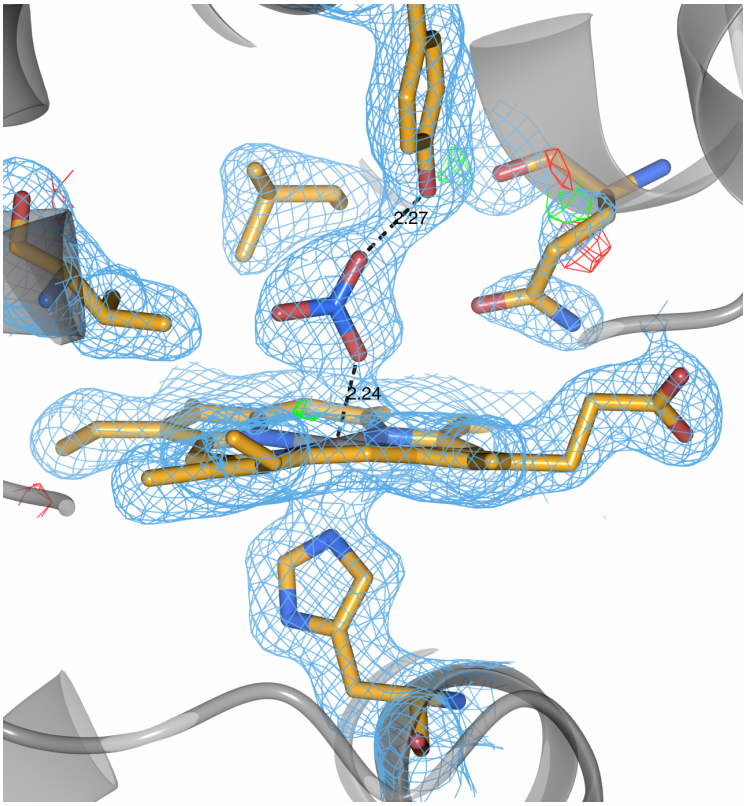


**Figure 7.12:** Nitrate molecule modelled at the distal site of monomer B of the NOD structure at 1.93 Å resolution. Interactions established by the ligand with the haem iron and Tyr29 are shown as dashed lines and lengths indicated. Electron density is contoured at 1σ (0.2562 $e^-$ /$Å^3$).

**Table 7.2:** B-factor values for the each of the atoms forming the nitrate ligand, bound at haem B of the NOD structure (1.93 Å); and the atoms which is interacting with: Fe from the haem group and the oxygen from the tyrosine residue.

| **Molecule/residue** | **Atom** | **B-factor ($Å^2$)** |
|---|---|---|
| Nitrate ($NO_3$) | N | 43.86 |
| | $O_1$ | 45.74 |
| | $O_2$ | 39.93 |
| | $O_3$ | 42.67 |
| Haem | Fe | 25.10 |
| Tyr29 | O (hydroxyl) | 33.26 |

## 7.7 NOD complex with nitric oxide (NO·)

To investigate if NOD would adopt a similar distal conformation for binding a different ligand, we used nitric oxide, which a known ligand for haemproteins and besides is involved in the dioxygenase reaction catalysed by NOD. For this purpose, we soaked ferric NOD crystals for 1 hour in crystallization solution supplemented with ~8 mM of ProliNONOate (Cayman chemicals). This compound was prepared in 25 mM NaOH and would liberate NO· upon a drop in pH, i.e. when in contact with the soak solution at pH ~5.5. Crystals were cryo-cooled in liquid nitrogen and taken to SLS for spectroscopic and X-ray diffraction measurements. Following the standard procedure in this work, spectroscopic data of the crystals were acquired before exposure to the X-ray beam. In this case, an online UV-vis spectrum was measured with the MS3 instrument (Pompidor et al. 2013) at beamline X10SA (SLS), which was used to assess and confirm the formation of the complex (in blue in Figure 7.13). In the spectrum, two distinct $\beta$ and $\alpha$ bands can be appreciated, and the Soret band is shifted to a wavelength of 420 nm upon binding; the same features are observed as well at similar wavelengths when the complex is formed in solution (Figure 7.14). An X-ray diffraction dataset was collected, followed by an additional UV-vis spectrum at the initial point (in orange in Figure 7.13), were some changes could be appreciated, such as the shifts of the Soret band (indicating photo-reduction) and the $\beta/\alpha$ bands to lower wavelengths and presenting different shapes, suggesting changes at the active site. After the collection of a second diffraction dataset, a final UV-vis spectrum was acquired (in red in Figure 7.13), where no further changes were observed. The photo-reduced form of the NOD-NO crystal presents resemblances with the photo-

reduced spectrum of a ferric NOD crystal after collection of a diffraction dataset (Figure 7.13, bottom panel), although different nuances can be appreciated for both spectra.

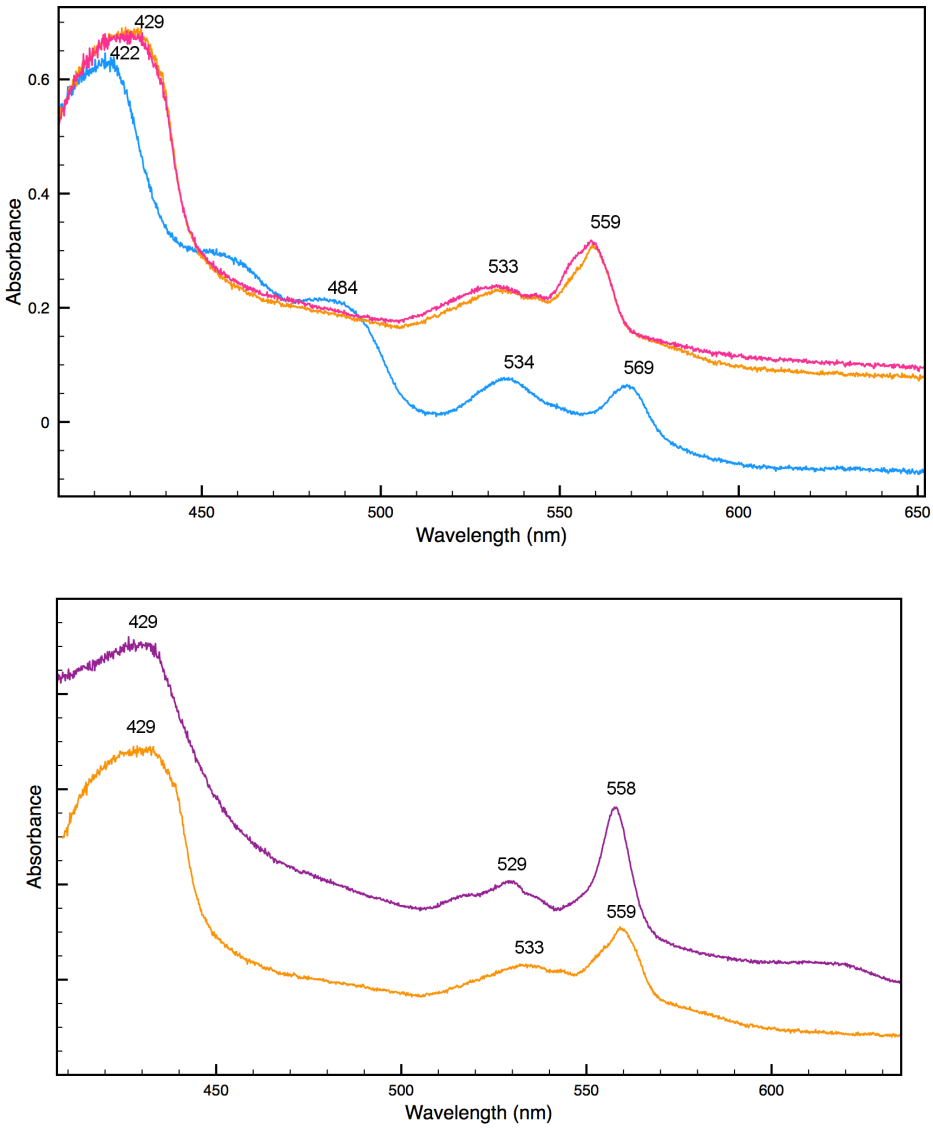


**Figure 7.13:** Online UV-vis spectra of the NOD-NO complex. In the top panel, spectrum of the NOD-NO crystal: before any exposure to the X-ray beam, in blue; after collection of a first X-ray diffraction dataset, in orange; and after collection of a second dataset, in red. In the bottom panel, a comparison is shown between the spectrum of the NO complex after collection of a dataset (in orange), and the spectrum from a ferric NOD crystal after collection of a dataset (in purple), both crystals expected to be photo-reduced. Both spectra are off-set from each other for clarity. Similarities can be drawn, although different nuances are observed. Measurements were carried out at beamline X10SA at SLS, equipped with the on-axis MS3 microspectrophotometer (Pompidor et al., 2013).

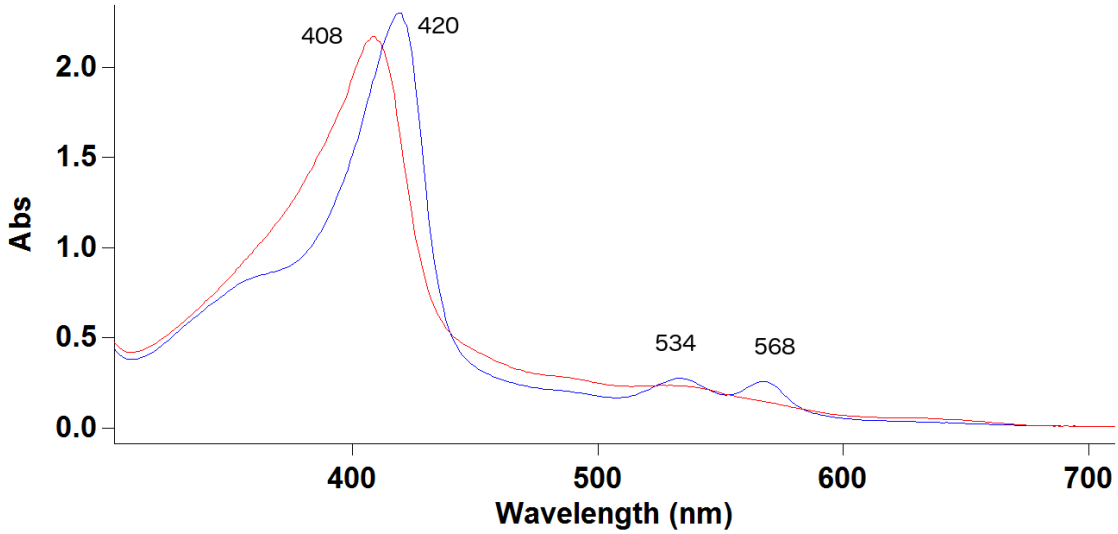


**Figure 7.14:** UV-vis protein solution spectra for the NOD-NO complex (in blue), formed upon addition of proliNONOate to NOD in its ferric state (in red). Spectra were acquired using a Cary-60 UV-vis spectrophotometer (Agilent).

Data collection was carried out at beamline X10SA (SLS) at 0.8 Å wavelength using a Pilatus-6M detector (Dectris), to a resolution of 2.18 Å (Table 73). Structure solution for the first collected dataset was performed by molecular replacement with *Molrep* (Vagin & Teplyakov 1997), using as a search model the previously solved structure from ferric crystals. As usual, refinement was carried out with *Refmac5* (Murshudov et al. 2011) and manual model building with *Coot* (Emsley et al. 2010). Data quality and refinement statistics are detailed in Table 73.

The structure of the NOD-NO complex presented a conformation very similar to that determined from NOD ferric crystals, and the active site environment found is shown in Figure 7.16. At the distal haem face there was electron density for a diatomic ligand, which was modelled as a NO·. However,

we are not completely sure if this is the case since the spectrum suffered changes after irradiation with the beam; therefore, it is most likely a mixed species structure where the initial structure was that of the NOD-NO complex. For comparison, a solution spectrum of the NOD-NO complex formed after the reduction of the protein with sodium dithionite was measured (Figure 7.15), although spectral features were poor. Due to the resemblance of the active site environment with the oxy-ferrous species (obtained by photo-reduction of ferric NOD) a superposition between both structures is shown in Figure 7.16, and a comparison has been drawn in Table 74, including bond lengths and angles.

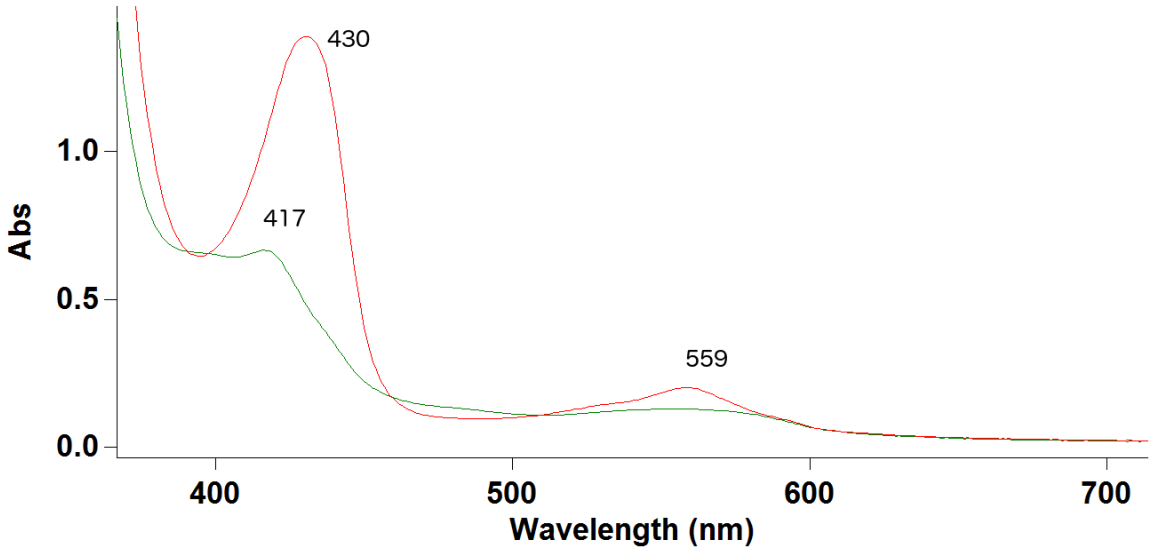


**Figure 7.15:** UV-vis protein solution spectra of NOD chemically reduced upon addition of a few grains of solid sodium dithionite (in red), and after addition of proliNONOate to that same reduced sample (in green). The formation of the NOD-NO complex can be noted although the β/α bands are not distinguishable enough to know their wavelength.

**Table 7.3:** Data collection and refinement statistics for the dataset measured from a complex NOD-NO crystal, at beamline X10SA at SLS (Villigen, Switzerland), using a wavelength of 0.8 Å, and a Pilatus-6M detector (Dectris). Values in parenthesis correspond to those of the last resolution shell.

| | **NOD-NO complex** |
|---|---|
| Space group | C2 |
| Unit cell (Å, °) | a=91.11, b=90.10, c=115.06; β=95.79 |
| Resolution (Å) | 45.32 – 2.18 (2.24 – 2.18) |
| $R_{merge}$ | 0.056 (0.747) |
| Total observations | 82,135 (5,500) |
| Unique reflections | 46,057 (3,189) |
| I/σ(I) | 8.7 (1.7) |
| $CC_{1/2}$ | 0.96 (0.72) |
| Completeness (%) | 94.9 (76.3) |
| Multiplicity | 1.8 (1.7) |
| $R_{factor}$ (%) | 23.1 |
| $R_{free}$ (%) | 27.3 |
| RMSD bond lengths (Å) | 0.015 |
| RMSD bond angles (º) | 1.92 |
| Ramachandran favoured (%) | 95.9 |
| Absorbed dose (MGy) | 0.49 |

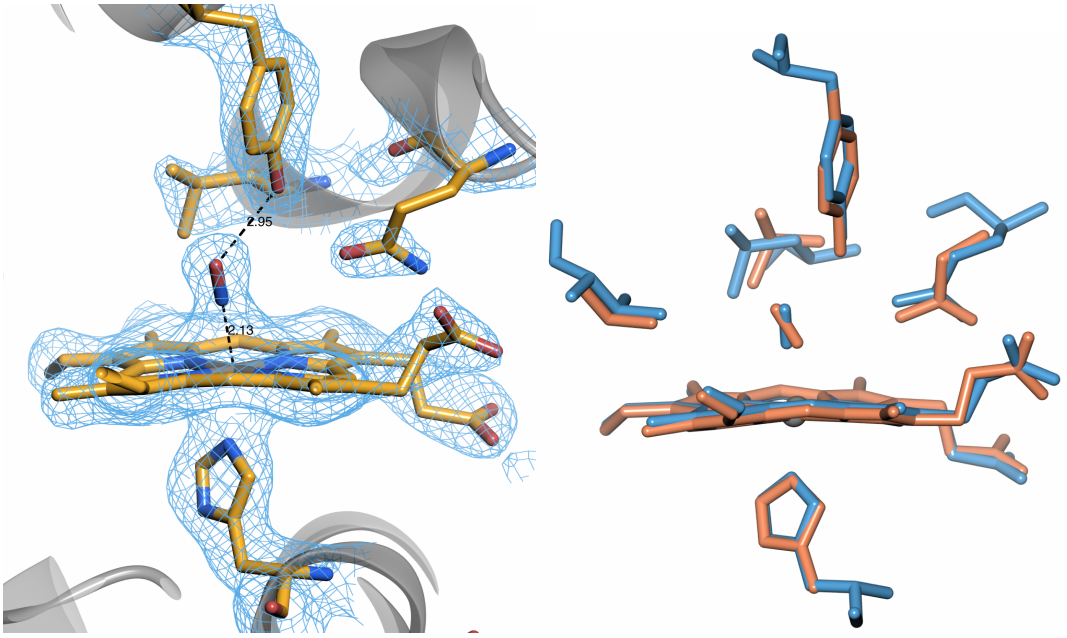


**Figure 7.16:** Active site environment (monomer B) for the structure determined from NOD-NO complex crystal (left panel). Tyrosine-29 can be seen stabilising the distal ligand a through hydrogen bond interaction at 2.95 Å. Distance between interaction of the ligand with Tyr29 and the haem iron shown as dashed lines. In the right panel, a superposition between the NO complex (in light blue) and the oxy-ferrous NOD species (in coral) is shown, where their resemblance can be noted.

**Table 7.4:** Geometry values for selected bonds and angles at the haem active site of the oxy-ferrous NOD species and the putative NOD-NO complex structures.

| **Geometry value (X: O or N)** | **NOD oxy-ferrous species** | **Putative NOD-NO complex** |
|---|---|---|
| Fe-X-O angle (º) | 144 | 153 |
| Fe-X distance (Å) | 2.29 | 2.13 |
| O-Tyr distance (Å) | 2.80 | 2.95 |
| Ligand inner bond (Å) | 1.27 | 1.23 |

## 7.8 NOD chemically reduced structure.

NOD crystals were chemically reduced before cryo-cooling them, to investigate the effect that this process would have in the conformation of the protein. Two substances were used for this purpose, which are the broadly used and well established reductants for protein samples sodium ascorbate and sodium dithionite. With ascorbate, which is a milder reductant, several soaks were tried with 100-200 mM concentration for 30-60 min. Trials were not successful, since the crystals would become visually damaged, and no diffraction was obtained from them. In the case of sodium dithionite, instead, crystals would survive after 30 min soak in a 50mM supplemented solution, even if this reductant is harsher than ascorbate. Reduced crystals were cryo-cooled and taken to Diamond to be examined.

Diffraction data for the reduced NOD crystals were measured at beamline I04-1 (DLS) at a wavelength of 0.9282 Å, to 2.05 Å resolution (Table 75). Unfortunately, this beamline at Diamond had not an available spectroscopy to probe the crystals *in situ*, so no validation of its oxidation state could be performed. Nevertheless, while solving the structure, a significant difference with the structure coming from ferric crystals was found: a loop in the vicinity of the distal site, which was disordered for every crystal to date, appeared ordered in this case, presenting a secondary structure of $\alpha$-helix in a stretch. In Figure 7.17, a comparison between the disordered and the ordered loop from the reduced structure is shown. We can appreciate how in the structure measured from ferric crystals (and became photo-reduced by the X-ray beam), just some residues of the loop can be modelled, and a stretch of 10 residues (45-54) are

disordered, and just some electron density can be observed for the main chain. In the chemically reduced structure all the loop is ordered and just some side chains are disordered (such as Arg49). This conformational change of the protein could have implication in its reactivity, and could give insights into protein function. This will be further analysed in the discussion section of this chapter.

**Table 7.5:** Data collection and refinement statistics for the dataset measured from a chemically reduced NOD crystal, at beamline I04-1 at Diamond Light Source (Didcot, UK), using a wavelength of 0.9282 Å, and a Pilatus 6M-F detector (Dectris). Values in parenthesis correspond to those of the highest resolution shell.

| | **Chemically-reduced NOD** |
|---|---|
| Space group | C2 |
| Unit cell (Å, °) | a=91.77, b=91.40, c=116.30; β=95.50 |
| Resolution (Å) | 29.06 – 2.07 (2.13 – 2.07) |
| $R_{merge}$ | 0.053 (0.633) |
| Total observations | 107,948 (8,294) |
| Unique reflections | 57,703 (4,414) |
| I/σ(I) | 9.1 (1.3) |
| $CC_{1/2}$ | 0.99 (0.44) |
| Completeness (%) | 99.0 (97.9) |
| Multiplicity | 1.9 (1.9) |
| $R_{factor}$ (%) | 20.0 |
| $R_{free}$ (%) | 24.5 |
| RMSD bond lengths (Å) | 0.017 |
| RMSD bond angles (º) | 2.03 |
| Ramachandran favoured (%) | 96.1 |
| Absorbed dose (MGy) | 0.69 |

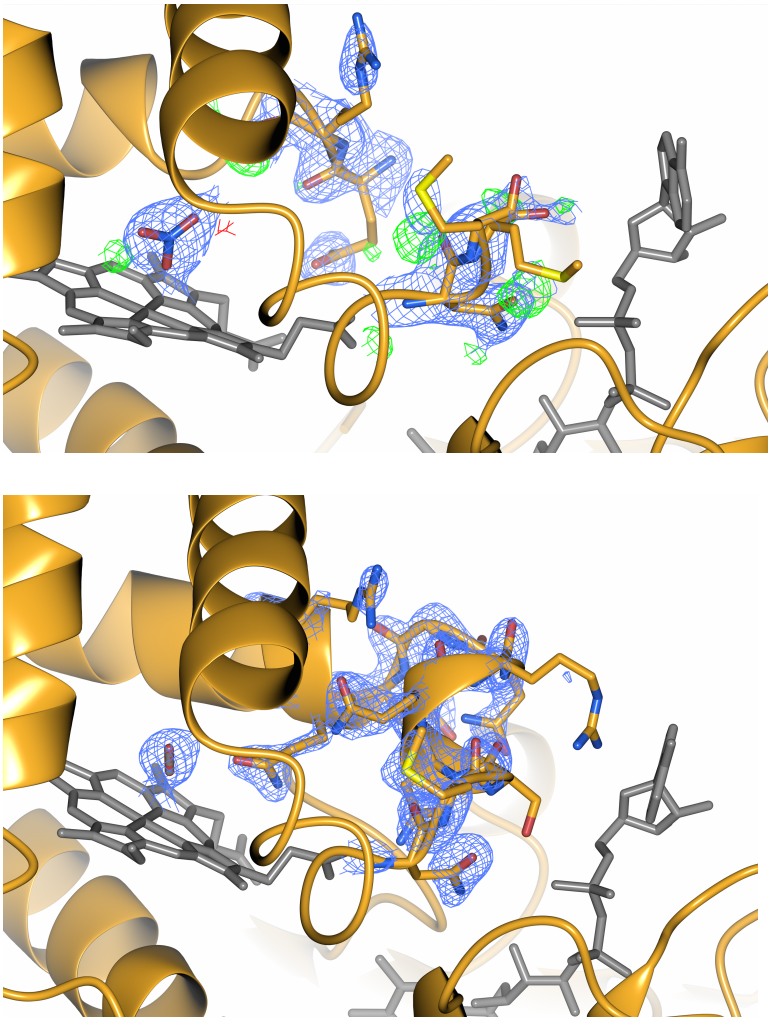

**Figure 7.17:** Structure of chemically-reduced NOD. View of the loop 36-54 for the structure obtained from ferric NOD crystals (upper panel), and the one obtained from a chemically reduced NOD crystal (bottom panel). It can be noted how the loop is well ordered in the reduced state of NOD, presenting some secondary structure features, while in the other case just some residues are defined. Electron density contoured at 1σ for both structures (0.2562 $e^-/Å^3$ upper panel, and 0.2348 $e^-/Å^3$ bottom panel)

An interesting observation is that for the structure resulting from a ferric NOD crystal, electron density at the distal site is large enough to model a nitrate ion (as shown previously), meaning a supposed catalytic activity of the protein in the crystal. In contrast, in the chemically reduced structure, the electron density at the distal site was substantially smaller (Figure 7.17), and could just support a diatomic ligand, which was modelled as a molecular oxygen,

although its identity is not clear and will be discussed later. Therefore, it could be that the oxidation state and/or conformational change of NOD regulates its reactivity. This matter will also be discussed at the end of the chapter.

## 7.9 MSOX from ferric NOD

A Multiple Structures from One crystal (MSOX; Horrell et al., 2016) experiment was applied to an NOD ferric crystal, with the intention to check whether catalysis existed in the crystals driven by the X-ray beam at the haem site. By measuring sequential low dose datasets we hypothesised that we could observe a growing density at the distal site as the reaction progressed. A series of six datasets were measured at SLS with the parameters specified in Table 76.

**Table 7.6:** Experimental parameters for the MSOX experiment with a ferric NOD crystal, carried out at beamline X10SA at SLS, using a Pilatus 6M detector (Dectris).

| Parameter | Value |
|---|---|
| Wavelength (Å) | 0.8 |
| Flux ($e^-$/s) | $1.59x10^{11}$ |
| Transmission (%) | 10 |
| Exposure time (s) | 0.1 |
| Oscillation (º/frame) | 0.1 |
| Transmission (%) | 10 |
| Dose per dataset (MGy) | 0.47 |

Online spectroscopic measurements were also performed before and after data collection for the first three datasets (half of the series), and are shown in Figure 7.18. An initially ferric NOD crystal (in blue), becomes photo-reduced after the measurement of the first dataset (in orange). The spectrum remains almost identical after further collection of the second and third datasets (in red and purple).

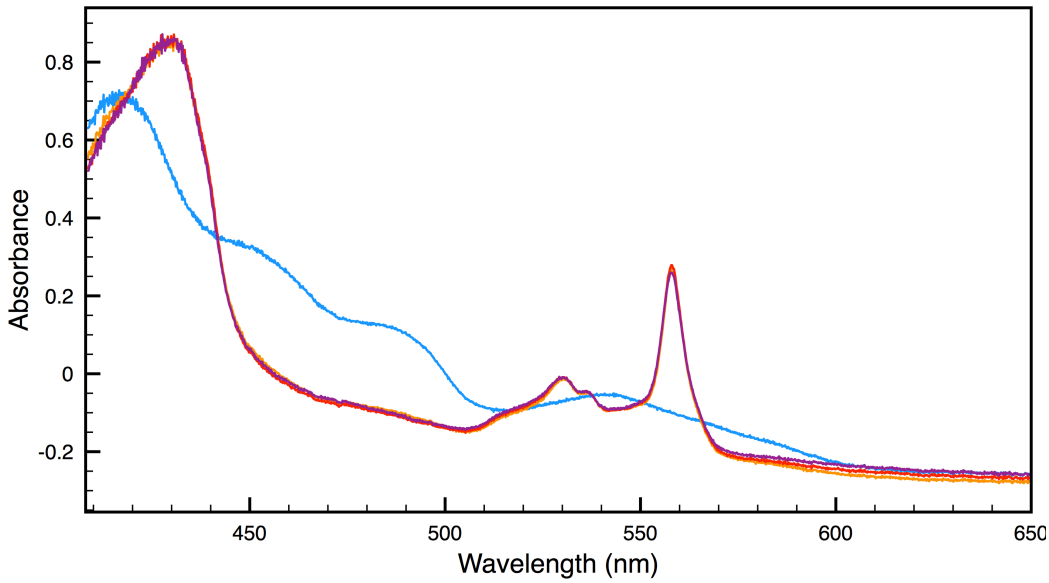


**Figure 7.18:** Single-crystal online UV-vis spectra measured in the MSOX experiment from an NOD ferric crystal. Measurement of the ferric NOD crystal before exposure to X-rays, in blue; after the collection of a first diffraction dataset, in orange; after the collection of a second dataset, in red; and after the collection of a third dataset, in purple. Spectra were acquired with the on-axis MS3 microspectrophotometer (Pompidor et al., 2013) equipped at beamline X10SA, SLS (Villigen, Switzerland).

Each of the structures were solved separately, by a job of molecular replacement with *Molrep* (Vagin & Teplyakov 1997), followed by refinement rounds of *Refmac5* (Murshudov et al. 2011) with manual model building in between using *Coot* (Emsley et al. 2010). The active site (monomer B) in each case was analysed and compared by examining the difference maps $F_o - F_c$ at the distal site. In Figure 7.19, the electron density at the active site for each

dataset is shown. In dataset 1, already a substantial extra electron density is observed after modelling a molecular oxygen at the axial position and performing a round of refinement. In the successive datasets, this extra electron density increases, stabilising around dataset 3, and its triangular shape becomes more apparent. However, this increase in the electron density throughout the experiment is not substantial, probably due to the reaction being almost complete after the measurement of the first dataset with absorption of 0.47 MGy of dose. This is also in agreement with the spectroscopic UV-vis data, where no further changes or evolving of the spectrum is observed after the first dataset. Thus, the species present would be that of the product $NO_3$ complex. Therefore, a future MSOX experiment applying less dose per dataset would be desirable to be performed in future visits to SLS (or ESRF) to validate *in crystallo* turnover at the NOD haem site.

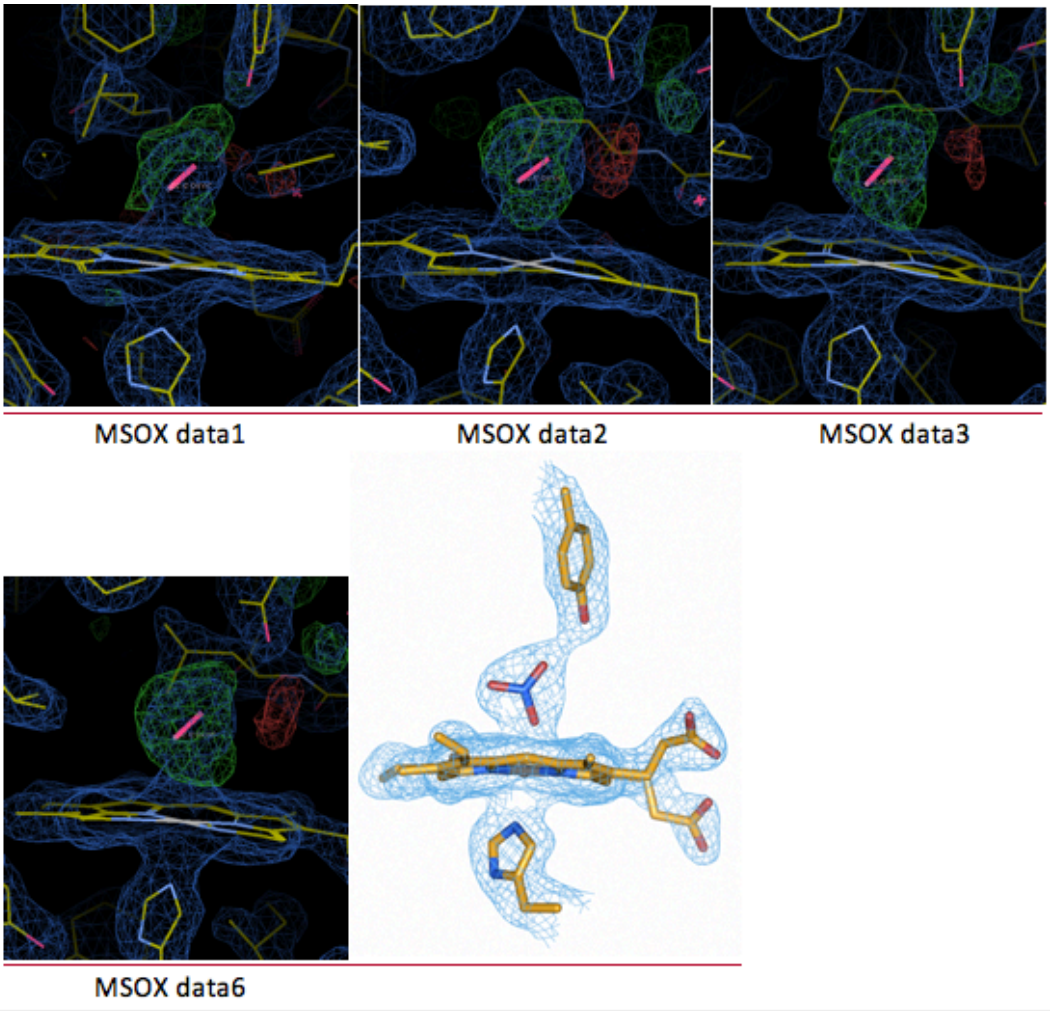


**Figure 7.19:** NOD active site for different datasets of the MSOX experiment. View of the active site for selected datasets is shown, with possible *in crystallo* turnover at the haem site appreciated. Positive extra electron density in green from the $F_o$ – $F_c$ difference map, shown contoured at 3σ. Note the slight increase of green density which stabilises at dataset 3. A final view of the active site where a nitrate molecule has been modelled at the active site is shown in the bottom right panel. Electron density from the $2F_o$ – $F_c$ map is shown as a blue mesh contoured at 1σ. Diffraction data were measured at beamline X10SA at SLS (Villigen, Switzerland), using a Pilatus 6M detector (Dectris) and 0.8 Å wavelength.

## 7.10 NOD SFX structure

Another information that was needed for NOD in order to interpret what could have been happening in the crystal upon photo-reduction, was to

determine the damage-free structure. In addition, an SFX structure would represent the first room-temperature structure for NOD, and would provide a general more relevant conformation to its biological conditions when acting as an enzyme. The ferric resting state of the metalloprotein would provide an active site view before any electrons were introduced to the system. If no catalysis was happening in the crystal while data collection, the density at the distal site should be considerably smaller and no nitrate molecule could possibly be modelled. Apart from this, the obtention of the ferric state of the protein had value in itself and could also give insights into conformational changes in the protein or other photo-induced changes.

### *7.10.1 Production of NOD crystals for serial crystallography*

Despite extensive efforts, this protein could not be crystallized in batches to deliver microcrystals. In fact, no crystals were observed in batches in the range of conditions that were tested; which included different ratios of protein to condition (e.g. 1:1, 1:2, 1:3, 1:4), and different concentrations of precipitant (e.g. 10, 15, 20, 25 % of PEG 3350) and salt (e.g. 100, 200 and 300 mM of $MgCl_2$).

As an alternative, we selected an optimal range of conditions within the optimization hanging-drop plates for large crystals, prioritising the formation of large amounts of small crystals (since no microcrystals could be generated), which consisted of:

- 21-23% PEG 3350, 200mM $MgCl_2$, and 100 mM BisTris pH 5.0-5.5

The volume ratio of protein to condition was also optimised (ratios 1:1, 2:1, 1:2, and 1:3 tested) following the same criteria, and the ratio 2:1 was found to

be the one delivering a larger number of crystals. Protein concentration of the stock used was 26.7 mg/ml. Once these parameters were optimized, ~150 drops were set up. One day before travelling to SACLA, crystals were taken from the drops by applying reservoir solution (20-30 μl) to each drop and resuspending the crystals by pipetting up & down several times as they would tend to be attached to a PEG skin formed at the surface of the drop. By this mechanical procedure large crystals would break up and produce a larger number of crystals. All the harvested crystals, which varied notably in dimensions ranging from 20-100 μm, were pooled in several Eppendorfs which were transported to Japan in a box at RT. Once at the beamline, they were loaded onto a chip (refer to Chapter 3 for loading procedure) before measurement with the femtosecond XFEL pulses. A total of 6 chips were measured in total in two trips to SACLA (Oct2017 and May2018), which are shown in Table 77. In the first trip, just ~3500 crystals were indexed between the two measured chips, which was not enough data to produce a dataset of sufficient quality. Further chips were measured with NOD crystals grown at the same condition, obtaining ~2000 more crystal hits to merge with the previous data. In total, from the six chips ~5500 crystals were indexed. The X-ray energy of the pulses was 10 keV in the first trip and 11 keV in the second.

**Table 7.7:** Fixed targets measured at SACLA for NOD crystals in visits in October 2017 (first two chips) and 2018 (last four chips). The full chip with 25,700 positions was measured in each case, using XFEL pulses of 10 fs duration. Preliminary hit rate estimation was carried out with software Cheetah (Barty et al., 2014).

| Chip name | Energy (keV) | Hit rate (%) | Comments |
|---|---|---|---|
| Zurich | 10.0 | 10.1 | |

| | | | |
|---|---|---|---|
| Bogota | 10.0 | 5.5 | |
| Salami | 11.0 | - | Data-collection software crash - Data lost. |
| Bamboo | 11.0 | 2.7 | |
| Xiguas | 11.0 | 2.1 | |
| Dorito | 11.0 | 2.5 | |

XFEL diffraction data were processed in CrystFEL (White et al. 2012), with the same workflow as shown in chapter 3, and also used for the other protein systems. Indexing was carried out with program *indexamajig*, optimizing some of the search parameters to maximize the number of crystals indexed:

```
$ indexamajig -i files.lst --peaks=zaef -threshold=100 --min-gradient=10000 --min-snr=2 --int-radius=2,3,5 --indexing=asdf -g 23368-1.geom -p NOD.cell -o nod_firstRun.out -j 20
```

For instance, the threshold and the signal to noise ratio (snr) were lowered to detect a higher number of spots, which for this protein were less intense than for DHP or DtpA.

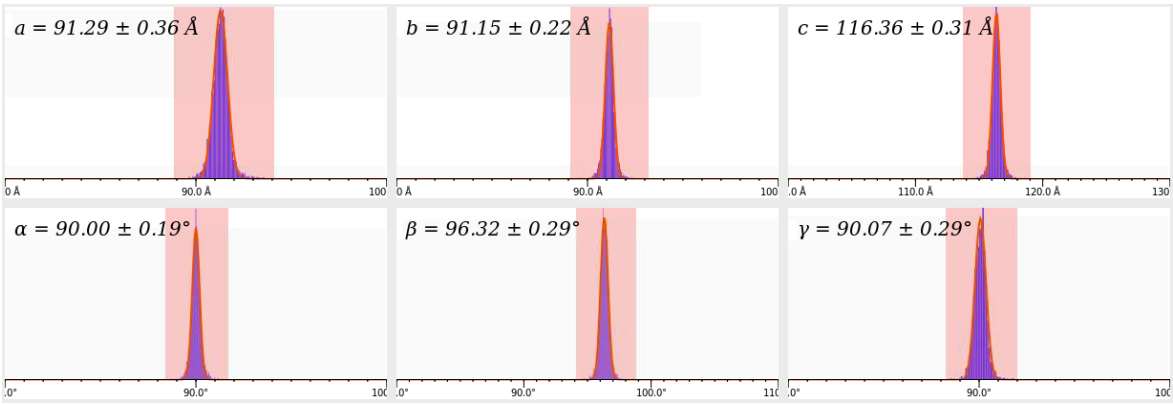


**Figure 7.20:** Cell parameters distribution for the NOD ferric crystals measured at SACLA, generated with the program *cell_explorer* within *CrystFEL* (White et al., 2012).

Diffraction data from chips from both beamtimes were joined together in a final output file. The distribution of cell parameters for the total population of indexed crystals was calculated using the program *cell_explorer* within *CrystFEL*, and the plots generated are shown in Figure 7.20. Merging of the data was carried with the program *process_hkl*, included in the script 'merge.sh', specifying pointgroup 2/m_uab. Data quality statistics were generated by using the script 'stat.sh'. The shells information for the SFX NOD dataset, as well as the CC and $R_{split}$ statistics are shown in Figure 7.21, Figure 7.22 and Figure 7.23, respectively. The resolution cut-off was applied at 2.90 Å, applying the criteria of CC > 0.5 in the highest resolution shell.

| Center 1/nm | # refs | Possible | Compl | Meas | Red | SNR | Std dev | Mean | d(A) | Min 1/nm | Max 1/nm |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.757 | 1163 | 1163 | 100.00 | 145780 | 125.3 | 5.13 | 236.17 | 207.24 | 13.22 | 0.218 | 1.295 |
| 1.462 | 1134 | 1134 | 100.00 | 96434 | 85.0 | 4.24 | 98.35 | 88.36 | 6.84 | 1.295 | 1.630 |
| 1.747 | 1090 | 1090 | 100.00 | 79215 | 72.7 | 4.06 | 92.59 | 77.20 | 5.72 | 1.630 | 1.865 |
| 1.959 | 1151 | 1151 | 100.00 | 76716 | 66.7 | 4.16 | 117.42 | 100.70 | 5.10 | 1.865 | 2.053 |
| 2.132 | 1114 | 1114 | 100.00 | 68701 | 61.7 | 4.34 | 128.83 | 115.95 | 4.69 | 2.053 | 2.211 |
| 2.280 | 1118 | 1118 | 100.00 | 64905 | 58.1 | 4.19 | 125.52 | 109.05 | 4.39 | 2.211 | 2.349 |
| 2.411 | 1104 | 1105 | 99.91 | 60578 | 54.9 | 3.99 | 99.79 | 89.94 | 4.15 | 2.349 | 2.473 |
| 2.529 | 1117 | 1117 | 100.00 | 58376 | 52.3 | 3.66 | 82.08 | 69.00 | 3.95 | 2.473 | 2.586 |
| 2.637 | 1123 | 1123 | 100.00 | 56351 | 50.2 | 3.42 | 71.12 | 57.66 | 3.79 | 2.586 | 2.689 |
| 2.737 | 1082 | 1082 | 100.00 | 52393 | 48.4 | 3.14 | 53.76 | 46.44 | 3.65 | 2.689 | 2.785 |
| 2.830 | 1137 | 1137 | 100.00 | 52727 | 46.4 | 2.92 | 47.97 | 40.49 | 3.53 | 2.785 | 2.875 |
| 2.917 | 1071 | 1071 | 100.00 | 48970 | 45.7 | 2.57 | 41.05 | 31.17 | 3.43 | 2.875 | 2.960 |
| 3.000 | 1125 | 1125 | 100.00 | 47700 | 42.4 | 2.39 | 34.21 | 27.31 | 3.33 | 2.960 | 3.040 |
| 3.078 | 1102 | 1102 | 100.00 | 44507 | 40.4 | 2.24 | 28.72 | 23.35 | 3.25 | 3.040 | 3.116 |
| 3.152 | 1102 | 1102 | 100.00 | 43824 | 39.8 | 1.94 | 22.34 | 18.25 | 3.17 | 3.116 | 3.188 |
| 3.223 | 1099 | 1099 | 100.00 | 41166 | 37.5 | 1.68 | 20.35 | 14.85 | 3.10 | 3.188 | 3.257 |
| 3.291 | 1153 | 1154 | 99.91 | 40432 | 35.1 | 1.52 | 17.42 | 12.92 | 3.04 | 3.257 | 3.324 |
| 3.356 | 1068 | 1068 | 100.00 | 37342 | 35.0 | 1.37 | 15.16 | 11.30 | 2.98 | 3.324 | 3.388 |
| 3.419 | 1132 | 1132 | 100.00 | 39266 | 34.7 | 1.21 | 12.93 | 8.93 | 2.93 | 3.388 | 3.449 |
| 3.479 | 1093 | 1093 | 100.00 | 38151 | 34.9 | 1.09 | 10.66 | 7.45 | 2.87 | 3.449 | 3.509 |

**Figure 7.21:** Resolution shells for the SFX resting NOD dataset, extended to a resolution of 2.85 Å. Statistics generated with the script stat.sh within the program CrystFEL.

| 1/d centre | CC | nref | d / A | Min 1/nm | Max 1/nm |
|---|---|---|---|---|---|
| 0.757 | 0.8535770 | 1160 | 13.21 | 0.219 | 1.295 |
| 1.462 | 0.8660724 | 1134 | 6.84 | 1.295 | 1.630 |
| 1.747 | 0.8543285 | 1089 | 5.72 | 1.630 | 1.865 |
| 1.959 | 0.8752582 | 1149 | 5.10 | 1.865 | 2.053 |
| 2.132 | 0.8524870 | 1111 | 4.69 | 2.053 | 2.211 |
| 2.280 | 0.7858234 | 1118 | 4.39 | 2.211 | 2.349 |
| 2.411 | 0.8164739 | 1099 | 4.15 | 2.349 | 2.473 |
| 2.529 | 0.8547527 | 1116 | 3.95 | 2.473 | 2.586 |
| 2.637 | 0.8783713 | 1118 | 3.79 | 2.586 | 2.689 |
| 2.737 | 0.8383719 | 1078 | 3.65 | 2.689 | 2.785 |
| 2.830 | 0.8333889 | 1135 | 3.53 | 2.785 | 2.875 |
| 2.917 | 0.8462977 | 1067 | 3.43 | 2.875 | 2.960 |
| 3.000 | 0.8084636 | 1123 | 3.33 | 2.960 | 3.040 |
| 3.078 | 0.7878979 | 1093 | 3.25 | 3.040 | 3.116 |
| 3.152 | 0.6596215 | 1098 | 3.17 | 3.116 | 3.188 |
| 3.223 | 0.6311762 | 1097 | 3.10 | 3.188 | 3.257 |
| 3.291 | 0.6365017 | 1142 | 3.04 | 3.257 | 3.324 |
| 3.356 | 0.5491208 | 1066 | 2.98 | 3.324 | 3.388 |
| 3.419 | 0.5484172 | 1127 | 2.93 | 3.388 | 3.449 |
| 3.479 | 0.4051842 | 1090 | 2.87 | 3.449 | 3.509 |

**Figure 7.22:** Values of the CC data quality statistic for each of the resolution shells for the SFX resting NOD dataset, extended to a resolution of 2.85 Å. Statistics generated with the script stat.sh within the program CrystFEL.

| 1/d centre | Rsplit/% | nref | d / A | Min 1/nm | Max 1/nm |
|---|---|---|---|---|---|
| 0.757 | 22.92 | 1160 | 13.21 | 0.219 | 1.295 |
| 1.462 | 24.29 | 1134 | 6.84 | 1.295 | 1.630 |
| 1.747 | 26.07 | 1089 | 5.72 | 1.630 | 1.865 |
| 1.959 | 25.02 | 1149 | 5.10 | 1.865 | 2.053 |
| 2.132 | 24.90 | 1111 | 4.69 | 2.053 | 2.211 |
| 2.280 | 26.47 | 1118 | 4.39 | 2.211 | 2.349 |
| 2.411 | 25.41 | 1099 | 4.15 | 2.349 | 2.473 |
| 2.529 | 27.75 | 1116 | 3.95 | 2.473 | 2.586 |
| 2.637 | 28.45 | 1118 | 3.79 | 2.586 | 2.689 |
| 2.737 | 30.39 | 1078 | 3.65 | 2.689 | 2.785 |
| 2.830 | 33.38 | 1135 | 3.53 | 2.785 | 2.875 |
| 2.917 | 36.16 | 1067 | 3.43 | 2.875 | 2.960 |
| 3.000 | 39.54 | 1123 | 3.33 | 2.960 | 3.040 |
| 3.078 | 42.42 | 1093 | 3.25 | 3.040 | 3.116 |
| 3.152 | 54.26 | 1098 | 3.17 | 3.116 | 3.188 |
| 3.223 | 63.59 | 1097 | 3.10 | 3.188 | 3.257 |
| 3.291 | 66.58 | 1142 | 3.04 | 3.257 | 3.324 |
| 3.356 | 76.70 | 1066 | 2.98 | 3.324 | 3.388 |
| 3.419 | 83.95 | 1127 | 2.93 | 3.388 | 3.449 |
| 3.479 | 99.33 | 1090 | 2.87 | 3.449 | 3.509 |

**Figure 7.23:** Values of the Rsplit data quality statistic for each of the resolution shells for the SFX resting NOD dataset, extended to a resolution of 2.85 Å. Statistics generated with the script stat.sh within the program CrystFEL.

Finally, an *mtz* file was generated using the script ‘create-mtz’, and was exported to be used in the ccp4i2 environment. The data were refined against the previous cryo-structure of NOD from ferric crystals at 1.93 Å, described earlier(Emsley et al. 2010) in the chapter. It refined well after an initial round of Refmac5 (Murshudov et al., 2011), obtaining initial R values of 0.20 / 0.27. Improvements of the model were carried out by further rounds of manual modelling with Coot (Emsley et al. 2010) followed by refinement cycles. In Table 78, data collection and refinement statistics for the NOD SFX structure are summarised. Although electron density is of limited quality at this resolution, it is enough to observe some features. The active site was inspected and some extra density was observed at both distal sites for the two monomers forming the dimer, into which a molecule of water was initially modelled (Figure 7.24). Some positive electron density could be observed after modelling the water, and thus we tried to model a molecular oxygen (Figure 7.24), although

the crystals were not in the oxy-ferrous state (from previous spectroscopic data). The modelling of a $NO_3$ molecule was not supported by enough electron density. The diatomic ligand appeared to be better supported by the density than the water molecule. However, due to the low amount of measured crystals for an SFX structure and the low resolution achieved, model bias can be a risk. The electron density feature at the distal haem site without modelling any ligand into it is shown in Figure 7.25. The spherical shape of the density would agree with diffraction from a single atom, the O of the water molecule. The presence of a distal water in the ferric NOD species is likely since a bound water ligand coordinating the haem iron has been observed in other haem enzymes, such as cytochrome c peroxidase and ascorbate peroxidase (Sharp et al. 2003; Bonagura et al. 2003). Indeed, an example can be observed in this same work, for the resting state of DtpA and DtpAa in Chapter 4.

**Table 7.8:** Data quality and refinement statistics for the SFX structure of resting state NOD, measured at SACLA (Japan), using 10 fs XFEL pulses of 1.24 and 1.13 Å wavelength.

| | **SFX ferric NOD** |
|---|---|
| Indexed crystals | 5,478 |
| Average hit rate (%) | 4.3 |
| Space group | C2 |
| Unit cell (Å) | a=91.29, b=91.15, c=116.36; β=96.32 |
| Resolution (Å) | 45.58 – 2.90 (2.96 – 2.90) |
| Unique reflections | 21,165 (1,105) |
| $R_{split}$ | 28.97 (84.44) |
| $CC_{1/2}$ | 0.88 (0.52) |
| Completeness (%) | 100 (100) |
| $R_{work}$ (%) | 17.5 |
| $R_{free}$ (%) | 24.7 |
| Rmsd bond lengths (Å) | 0.0121 |
| Rmsd bond angles (º) | 1.811 |
| Ramachandran favoured (%) | 95.4 |

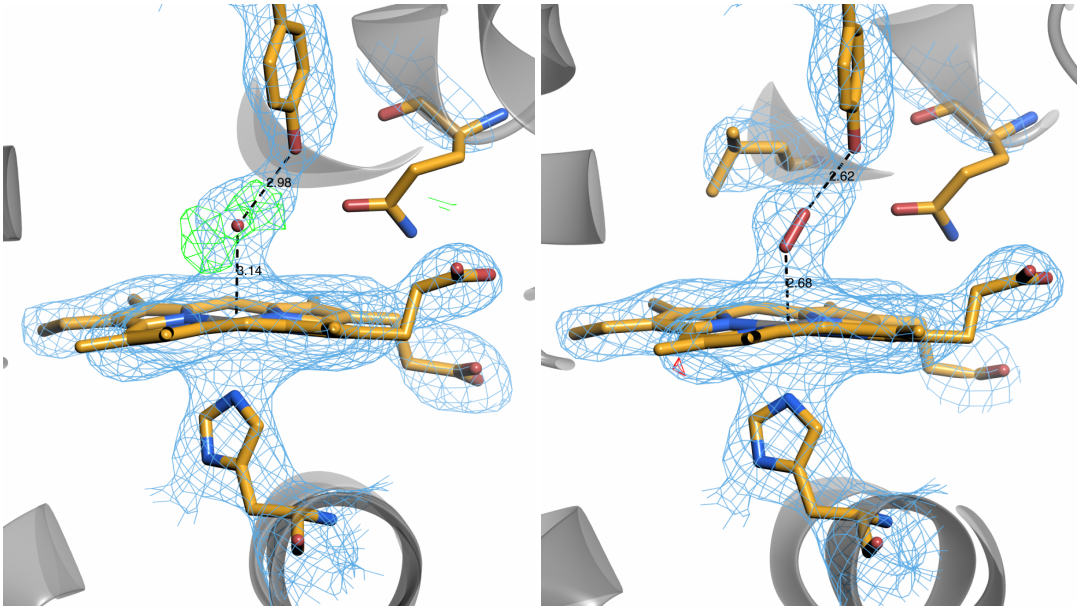


**Figure 7.24:** Active site in monomer B for the NOD SFX structure. In the left panel, after modelling a water molecule at the active site, where some positive electron density difference map can be appreciated (contoured at 3σ); and in the right panel, when a molecular oxygen is modelled at the distal haem site. Bond distances with the ligands are indicated by dashed lines. Note how all the sidechain of distal residue Gln53 is disordered in this monomer. Electron density maps in both cases contoured at 1σ.

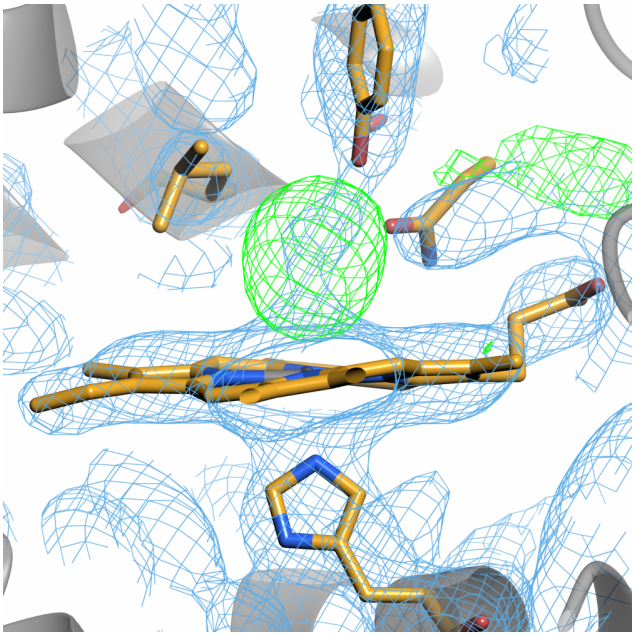

**Figure 7.25:** Active site in monomer B for the NOD SFX structure, without modelling any ligand into the positive electron density of the difference map (contoured at 3σ). Electron density of map $2F_o - F_c$ is represented as a blue mesh and is contoured at 1σ.

For comparision between RT and cryo structure of NOD, a superposition was carried out between the SFX NOD structure and the 100 K structure at 1.93 Å resolution, and is shown in Figure 7.26. As it can be observed, the overall conformation of the protein in both structures is very similar, with an rmsd of 0.42 Å, and just minor differences in some loops can be noticed.

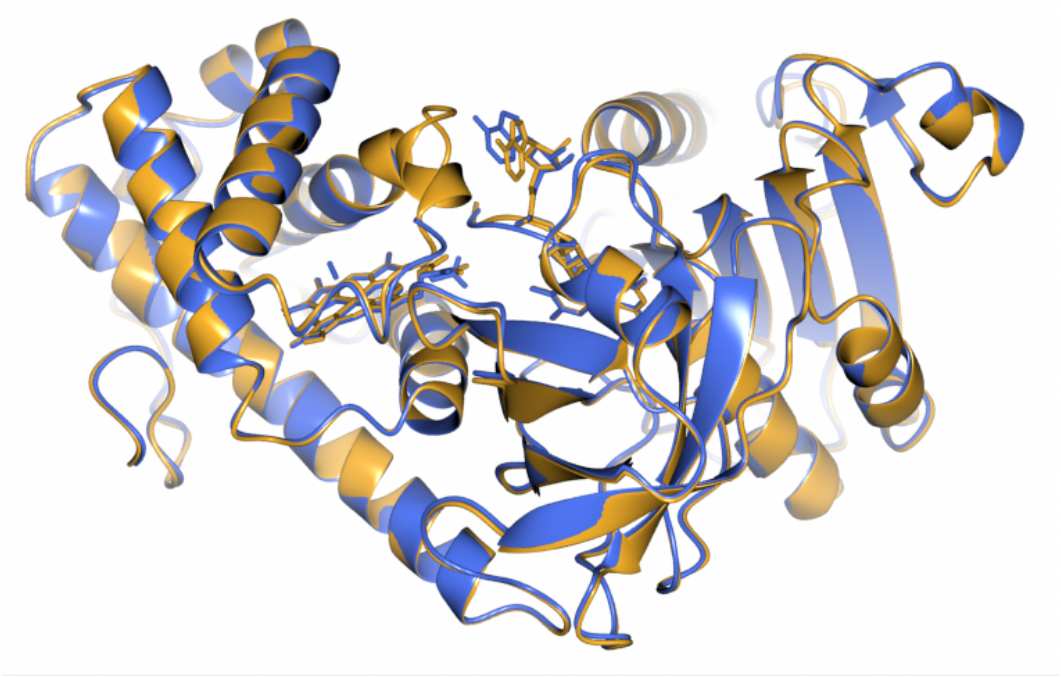

**Figure 7.26:** Superposition of the room-temperature SFX NOD structure (in blue) and the cryostructure of NOD (in orange). Note the very similar conformation of the protein in both structures. Superposition carried out using the SSM method with CCP4MG (McNicholas et al. 2011).

## 7.11 Discussion

In this chapter, we have performed structural studies of the haem enzyme Nitric Oxide Dioxygenase (NOD), coupled with the measurement of single-crystal spectroscopies to assign the correct oxidation and ligand state of the crystals measured, and identity the photo-reduced species produced upon interaction with the X-ray beam. A damage-free SFX structure at low resolution was also determined using a fixed target approach with thousands of NOD crystals.

### *7.11.1 Lipid-bound NOD structure*

Purified NOD ferric protein was found to be in its lipid bound state by assessment with the spectroscopic characterization carried out by Bonamore and co-workers (Bonamore, Farina, et al. 2003). This is expected, since the lipid is tightly bound to NOD and can only be removed by means of hydrophobic chromatography (Bonamore, Gentili, et al. 2003).

The protein was crystallized and the crystals obtained maintained spectral features of the lipid-bound protein, as inspected with *in crystallo* UV-vis just before diffraction experiments (Figure 7.7). In particular, the CT band at ~645 nm, since the Soret band is largely saturated. Upon collection of a complete diffraction dataset, the protein presents a clearly photo-reduced spectrum, as observed for other haemproteins (Kekilli et al. 2017). The structure obtained from the NOD ferric crystal, is thus a mixed species. A different conformation of the protein could be observed from the one previously reported for lipid-free NOD (PDB code 1gvh, Ilari et al., 2002). As shown earlier in the chapter, the haem environment at the active site was remarkably different from the lipid-free active site, with Tyr29 and Gln53 assuming a main at the distal site to stabilise the ligands bound at the axial position of haem, in contrast with Leu57 dominating the distal site above the iron from the reported structure. Furthermore, the conformation of the Flavin cofactor is also different from the lipid-free NOD structure. Therefore, we believe that the NOD structure reported here is that of the lipid-bound form, supported by spectroscopic evidence, and by the remarkable different conformation that is presented by the enzyme. However, we cannot be fully confident in this statement, since no lipid was found in electron-density maps, which is likely due to disorder and the limited

resolution of the crystal. Some features of extra electron-density are present in the structure, although not large enough to support a molecule such a phospholipid, whose identity is besides unknown. The main phospholipids found in the bacterial membrane are phosphatidylethanolamines (PE) and phosphatidylglycerols (PG), (Gidden, Denson, Liyanage, Ivey, & Lay, 2010)

In our NOD structure, the Tyr29 is found assuming a main role in the stabilisation of distal ligands by hydrogen bond interactions, with the help of Gln53, which its conformation by also establishing a weak hydrogen interaction with Tyr29. From conservation and biochemical work with mutants for both residues (Tyr29 and Gln53), they showed a main role in catalytic reactions carried out by the protein (Gardner, Martin, Gardner, Dou, & Olson, 2000), and thus the surprising finding of the haem active site of the lipid-free protein, where they were found in a second coordination shell, with main presence of Leu57 just on top of the iron (Ilari et al., 2002). In addition, is also surprising that the lipid-free NOD structure does not present any signs of photo-reduction by the beam (looking at the electron density map deposited), when is well documented how easily prone to photo-reduction metal centres are (Yano et al. 2005; Kekilli et al. 2014; Kekilli et al. 2017). A dose estimation should be performed reproducing the parameters of the data collection carried out by Ilari et al. (2002), but most likely was enough to photo-reduce NOD. A plausible explanation could be that, in the active site configuration induced by the removal of the lipid, Tyr29 is out of reach and cannot stabilise the distal ligands of the species generated by photo-reduction. It could be that Tyr29 is crucial to ‘control’ and stabilise ligands at the active site, and therefore mediate in catalysis, as the biochemical work suggests (Gardner et al., 2000). That could

imply that the lipid-free form of NOD is not able to catalyse the oxidation of NO to $NO_3$ if the conserved residue cannot stabilize the distal substrates and products. Further enzyme catalysis experiments (Gardner et al., 1998), should be carried out to verify this hypothesis. On the same line of argument, it would be interesting to research if the lipid-free form of NOD has any functionality. In which case NOD could be a multifunctional enzyme which carries out a role when it is associated with the cellular membrane (e.g. dioxygenase function), and a different one when it is in the cytoplasm.

#### *1.1.1.1 Possible in crystallo turnover*

At the active site of the structure obtained from NOD ferric crystals, differences could be noted at each monomer of the dimer found in the asymmetric unit. In monomer A, a molecular oxygen was modelled, thus resulting in the oxy-ferrous state of NOD. This species is often generated at the active site of haem enzymes upon photo-reduction, as for instance seen for DtpA in Chapter 4 (Schlichting et al., 2000; Kekilli et al., 2017). At monomer B, a triangular electron-density feature was identified and modelled as a product nitrate ($NO_3$) molecule. As discussed earlier in this chapter, ligands of similar shape (e.g. acetate) were ruled out since not present in the crystallization condition, buffer or cryoprotectant. The protein could have picked up NO or $NO_3$ from *E. coli*, where it was expressed, although we would have noticed it spectroscopically, and thus can also be ruled out. After purification, the spectrum of NOD was that of its ferric form. The presence of $NO_3$ would mean turnover at one of the haem sites of the crystal, and thus the *in crystallo* catalytic activity of the enzyme driven by the X-ray beam. It is well reported in

the literature how the solvated electrons generated by the beam upon interaction with the crystal can be exploited to run catalytic reaction in redox enzymes (Horrell et al. 2016; Schlichting et al. 2000).

The spectrum of the final photo-reduced state presents some complexity, since multiple bands seem to be observed in the β band region (Figure 7.7), whereas the α band is observed as a single peak. Normally, when an oxy-ferrous species is formed at both monomers, the observed β/α bands appear as single peaks (Kekilli et al., 2017), such as the NOD-NO complex formed in solution shown in Figure 7.14. Thus, the spectroscopic information from the photo-reduced NOD suggests the presence of different species at the haem sites, which agrees with the electron-density features observed.

An unexplained factor in the possibility of turnover at the NOD haem site upon photo-reduction is that no nitric oxide (or any precursor) was added to the crystal, and its origin would be unknown to us. However, we propose that NO could be generated by the interaction of the X-ray beam with some component in the crystal containing nitrogen, such as a protein residue (e.g. Arg, Gln), buffer molecule (BisTris was used), or even $N_2$ present in the solvent; although such chemistry has not been reported before. If the formation of nitrate could be confirmed, this might give new insights aspects of radiation chemistry still not fully understood and to be further investigated.

### *7.11.2 Conformational change in chemically reduced NOD*

A conformational change in the loop 46-54 near the haem site could be appreciated from the structure of chemically reduced (with 50mM sodium dithionite) crystals in comparison with the structure measured from NOD ferric crystals (shown in Figure 7.17), where that stretch is normally disordered. It could be that this conformational change has implications in the reactivity of the protein, and maybe some biological relevance. It is worth noting that even if the initial structure with the disordered loop was measured from NOD ferric crystals, the protein was photo-reduced during data collection and is likely to be more representative of a ferrous haem state (Kekilli et al., 2017). However, it has been reported with other haem proteins how redox changes in the protein within the crystal do not necessarily produce conformational changes at 100 K (Kekilli et al., 2014). Another possibility would be that the conformational change takes place because of the change in pH conditions upon adding the reductant.

Apart from the discussed conformational changes in the loop near the haem site of NOD, there were also substantial differences in the electron-density from the ligand bound at the haem site. In the structure coming from the ferric NOD, a nitrate molecule could be modelled in, suggesting potential turnover as discussed; in the chemically reduced structure instead, just a diatomic ligand could be supported by the electron-density, which was modelled as a molecular oxygen. Therefore, it could be that the oxidation state and/or conformational change of NOD regulates its reactivity. However, the addition of sodium dithionite to a solution results in the elimination of all the

oxygen in the samples, as this compound scavenges it. This fact would explain the lack of turnover at the active site for the reduced crystals, but would not explain why still the oxy-ferrous species is formed, since there shouldn't be oxygen available in the crystal. It could well be that the complex observed after measurement of the reduced crystals it's not an oxy-ferrous species, but maybe other, such as an NO complex – in which case we could not explain its origin, as in the case of the putative turnover commented before. An experiment which could be performed in the future is try to identify spectroscopically the bound ligand; for instance, using SCRR and assess if a peak for the N-O bond or the O-O bond can be observed, in this way validating the species formed after collection of a diffraction dataset. If the formation of an NO-complex could be verified, that would imply that the NO radical is likely to be generated by the interaction of the X-rays and some component forming the crystal.

Although the usage of dithionite would explain well the lack of catalytic activity at the active site, the argument of the conformational change in the protein could be also a plausible explanation. In the structures measured from ferric crystals, we can see an example supporting that argument. In one of the monomers we see a disordered loop, as discussed previously. But the other monomer conforming the dimer in the asymmetric unit, monomer A, has an ordered loop just as in the reduced structure, and no extra density is observed when modelling the molecular oxygen. Therefore, it could be that the ordered conformation of the loop prevents the catalysis to happen by means of accessibility to the pocket or another mechanism.

### 7.11.3 MSOX experiment to capture catalysis

Successive datasets were collected in an MSOX (Horrell et al. 2016; Horrell et al. 2018) experiment, in order to try and capture enzyme catalysis and verify the catalytic activity of NOD crystals upon irradiation. However, the dose absorbed by the crystal per dataset was rather high (0.47 MGy), even if data collection was designed to minimize it. As a result, there was not a substantial increase of electron density along the series of datasets, and turnover (if any) would have been complete after the first dataset. To be able to assess if it exists, we should be able to perform a lower dose MSOX experiment without compromising resolution, which in this case is already at 2.1Å. Another solution would be to apply MSOX in with serial crystallography methodology (Ebrahim et al., in press). In this way, we would spread the dose among thousands of crystals forming the dataset, and we would have a considerably lower dose per dataset (10-50 kGy, e.g. Owen et al. 2017). This work is in progress, and it is mainly challenging due to the difficulty of obtaining NOD microcrystals in batches.

### 7.11.4 SFX structure of NOD

The damage-free structure of the ground state of NOD was characterized by SFX to obtain information about the configuration of the haem sites before photo-reduction has occurred, and confirm the turnover at haem B. Although the resolution is low (2.9 Å), some features are still appreciable. At both distal sites, a bound water was modelled coordinating the iron, feature that has been observed in the ferric state of other haem enzymes (Sharp et al. 2003;

Bonagura et al. 2003). A nitrate molecule could not be supported by the electron-density at haem B, although at this resolution model bias is a concern. A diatomic ligand, could be supported for example at both haem sites, although this is not as plausible as the distal water scenario if the structure is damage-free, since the crystals were in its ferric state before the SFX experiment. This condition has been tested many times at synchrotrons with single-crystal spectroscopies (e.g. SLS), transporting the crystals at RT or cryocooled in a dewar; and the presence of oxygen bound at the distal site (i.e. oxy-ferrous species), would be appreciable in spectral features such as sharp $\beta/\alpha$ bands (Kekilli et al. 2017). Therefore, the damage-free structure of NOD would suggest the presence of a distal water at the distal site, and a nitrate molecule cannot be supported by electron density. However, at this low resolution and with a limited number of patterns from ~5,500 crystals, conclusions cannot be drawn with full certainty.

# Chapter 8 General discussion and concluding remarks

In this work, we have studied four different haem enzymes and their catalytic cycles. The use of complementary techniques such as single-crystal spectroscopies (SCSs) has been shown crucial to the correct assignation of the redox, ligand, or catalytic stated of the protein, allowing us to identify the species within the crystal and located in the enzymatic mechanism. The use of SCSs is well reported in the literature and has been a powerful tool to the structural determination of complete catalytic cycles (Berglund et al. 2002; Schlichting et al. 2000; Gumiero et al. 2011). In our case, *in crystallo* spectroscopies allowed us to validate 'pure' states in DtpA (Chapter 4), such as the oxy-ferrous, the ferric, or the deoxy-ferrous species. With their help, we could also identify in the DtpA crystals the formation of the peroxidase cycle intermediates compound I, II and III, although we were not able to structurally determine them since they were rapidly photo-reduced by the X-ray beam. Therefore, methodology is needed for obtaining the validated structures of these intermediates, which could be obtained by means of SFX, SSX, composite datasets or helical collections. Their determination would be desirable, since it would suppose the structural determination of the main intermediates in the catalytic cycle of the DyP family of peroxidases (Sugano 2009).

For DHP (Chapter 5) resonance Raman was the most suitable technique, and we identified the ferric signature and its photo-reduction by the beam. Different DHP complexes with azole compounds were fingerprinted by SCRR and its photo-reduction also confirmed, as well as the oxidation state of the iron Fe(III) in the bound complex. Therefore, even if the spectroscopic was fully needed to confirm the formation of the complex (i.e. the ligands were clearly observed in the electron density), it was still delivering important redox information of the haem site which might be crucial to reach binding mechanistic conclusions. The use of SCRR, and also *in crystallo* UV-vis spectroscopy, should be applied now to characterize different states in the peroxidase cycle carried out by DHP.

In the case of NOD (Chapter 7), we could also monitor by using single-crystal UV-vis, the photo-reduction of the protein, yielding a complex spectrum with putatively a mixture of distinct hexa-coordinate species in each monomer, as observed in the electron density. In monomer A, a oxy-ferrous species is observed, with molecular oxygen bound to the ferrous haem iron. In monomer B instead, a large electron density feature was modelled as a nitrate $NO_3^-$ molecule, which would imply *in crystallo* catalysis in NOD driven by the beam, as seen in other metalloproteins (Horrell et al., 2016). In future work, would be interesting to try to spectroscopically characterise the bound ligand at haem B, for examble by SCRR finding the vibrational modes that the bound product would have, in order to validate its formation.

## 8.1 Serial crystallography methods

The application of our protein systems to the emerging serial crystallography methodologies (e.g. SFX, SSX) gave us the opportunity to measure room-temperature intact (or low-dose) structures of resting states of the enzymes, as well as ligand complexes (Chapters 4, 6 and 7). To be able to use SX methodologies, is crucial to find the right batch crystallization conditions for the generation of large numbers of microcrystals. This proved to be a demanding and time-consuming optimization task with many variables to consider. Nevertheless, when an optimal condition was found, it would ensure the efficient obtention of SFX or SSX structures with a low number of measured chips needed (typically 2-4).

The measurement of room-temperature structures of the different haem enzymes in this work (and for any protein or biomolecule) is of great value, since they represent better the protein dynamics and conformational variability (Fraser et al. 2011; Russi et al. 2017) adopted by the proteins in their biological context, and therefore are more relevant.

The performance of experiments at XFEL sources is highly competitive and with higher costs associated. However, using synchrotron serial approaches (SSX) delivers similar results in terms of intact structures, with the obtention of low-dose datasets, and sometimes even better resolution of the structures as seen for instance in this work (Chapter 4 and 6). If the availability of XFELs does not improve in the coming years, this observation would open

the debate if whether XFEL beamtime should be focused on the kind of experiments that can only be performed with the remarkable properties of these facilities – femtosecond time-resolved crystallography. Currently, this is performed using a jet sample delivery system, although efforts are being made towards the development of these methodologies in fixed targets as well; and others, such as the tape sample delivery approach (Fuller et al., 2017).

Fixed targets have proven very efficient for the determination of resting states of metalloproteins in this work (in our case of DHP, DtpA, DtpAa and NOD), or any stable protein state where radiation damage needs to be avoided. They are easy to use, can be mounted at any source, due to its modular design (Ebrahim et al, in press), and are time and sample efficient, just needing a few mg of protein crystallized in the form of microcrystals, and 15 min for the inspection of more than 25,000 positions.

One of the challenges with fixed targets are the measurement of intermediate species, such as compound I and II in the peroxidase cycle, since they have to be stable enough to last for the time needed for the measurement (~15 min) of the chip. Therefore, if the intermediate is evolving or decaying, (in our case the transition of Cmd I to Cmp II, or from Cmd II back to resting state), one would obtain data for a mixed species and makes it difficult to obtain discrete intermediate states. Furthermore, since the SFX experiment is carried out at RT, the catalytic processes take places more rapidly and subsequently intermediates are short-lived. Also, it is not possible yet to perform *in situ* spectroscopies to validate the intermediate states, and thus the duration of the

intermediate in the microcrystal sample has to be estimated before-hand in the lab to provide a time-frame of measurement possibilities. In our last trip to SACLA, we tried to obtain the peroxidase cycle intermediate Cmd I for the DtpAa and also for DHP, activating both enzymes by addition of the organic peroxide peracetic acid (data not shown and under analysis). Results do not seem conclusive, although the bond between the Fe and the oxygen atom at the distal site shrinks for DtpAa, indicating some activity at the active site. However, we are not sure if this structure corresponds to an intact intermediates, or is it a mixed species. For this reason, we aim to measure this same intermediate in a synchrotron source with a composite dataset approach to compare with the SFX structure. If we could confirm the identity of the intermediate would represent the prove that catalytic intermediates can be obtained at room-temperature using a fixed target SFX approach.

The synergy/combination between different serial methodologies has been proved to be a powerful tool to characterise protein systems. For instance, XFEL beamtime at SACLA was much more productive because of the time we had to test and optimize experiments at the SSX beamtime at I24. Also, in the case of DHP structure (Chapter 6), the higher resolution achieved with SSX allowed us to observe with more detail the feature of a distal water bound at the axial position of the haem iron. Furthermore, the utilization of SSX permits to carry out experiments such as the dose-series (Chapter 4), which are not possible currently at XFELs, and where one can follow photo-reduction and/or catalytic processes at atomic detail in a ‘structural movie’.

Therefore, the utilisation of serial crystallography methodologies are pushing the boundaries of X-ray crystallography allowing us to study protein dynamics and catalysis with femtosecond resolution (time-resolved femtosecond crystallography), and not just ‘static’ averaged structures of biomolecules forming a crystal.

# References


Ahmad, M. et al., 2011. Identification of DypB from rhodococcus jostii RHA1 as a lignin peroxidase. *Biochemistry*.

Barends, T.R.M. et al., 2015. Direct observation of ultrafast collective motions in CO myoglobin upon ligand dissociation. *Science*, 350(6259), pp.445–450.

Barrios, D.A. et al., 2014. Peroxygenase and oxidase activities of dehaloperoxidase-hemoglobin from Amphitrite ornata. *Journal of the American Chemical Society*, 136(22), pp.7914–7925.

Barty, A. et al., 2014. *Cheetah* : software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *Journal of Applied Crystallography*, 47(3), pp.1118–1131. Available at: http://scripts.iucr.org/cgi-bin/paper?S1600576714007626.

Berglund, G.I., Carlsson, G.H. & Al, E., 2002. Supl. 1 The catalytic pathway ov horseradish peroxidase at high resolution. *Nature*, 417(May), pp.10–11.

Van Bloois, E. et al., 2010. A robust and extracellular heme-containing peroxidase from Thermobifida fusca as prototype of a bacterial peroxidase superfamily. *Applied Microbiology and Biotechnology*, 86(5), pp.1419–1430.

Bonagura, C.A. et al., 2003. High-resolution crystal structures and spectroscopy of native and Compound I cytochrome c peroxidase. *Biochemistry*, 42(19), pp.5600–5608.

Bonamore, A., Gentili, P., et al., 2003. Escherichia coli flavohemoglobin is an efficient alkylhydroperoxide reductase. *Journal of Biological Chemistry*, 278(25), pp.22272–22277.

Bonamore, A., Farina, A., et al., 2003. Interaction with membrane lipids and heme ligand binding properties of Escherichia coli flavohemoglobin. *Biochemistry*, 42(19), pp.5792–5801.

Bonamore, A., Chiancone, E. & Boffi, A., 2001. The distal heme pocket of Escherichia coli flavohemoglobin probed by infrared spectroscopy. *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, 1549(2), pp.174–178.

Botha, S. et al., 2015. Room-temperature serial crystallography at synchrotron X-ray sources using slowly flowing free-standing high-viscosity microstreams. *Acta Crystallographica Section D: Biological Crystallography*, 71, pp.387–397.

Boutet, S. et al., 2012. High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography. *Science*, 337(6092), pp.362–364.

Bowman, S.E.J. & Bren, K.L., 2008. The Chemistry and Biochemistry of Heme c: Functional Bases for Covalent Attachment. *Natural Product Reports*, 25(6), pp.1118–1130.

Brown, M.E., Barros, T. & Chang, M.C.Y., 2012. Identi fi cation and Characterization of a Multifunctional Dye Peroxidase from a Lignin-Reactive Bacterium.

Carey, L.M. et al., 2018. Selective tuning of activity in a multifunctional enzyme as revealed in the F21W mutant of dehaloperoxidase B from Amphitrite ornata. *Journal of Biological Inorganic Chemistry*, 23(2), pp.209–219. Available at: https://doi.org/10.1007/s00775-017-1520-x.

Carugo, O. & Carugo, K.D., 2005. When X-rays modify the protein structure:

Radiation damage at work. *Trends in Biochemical Sciences*, 30(4), pp.213–219.
Casadei, C.M. et al., 2014. Neutron cryo-crystallography captures the protonation state of ferryl heme in a peroxidase. *Science*, 345(6193), pp.193–197.
Casanas, A. et al., 2016. EIGER detector: Application in macromolecular crystallography: *Acta Crystallographica Section D: Structural Biology*, 72(9), pp.1036–1048.
Chapman, H.N. et al., 2011. Femtosecond X-ray protein nanocrystallography. *Nature*, 470(7332), pp.73–78.
Chavas, L.M.G., Gumprecht, L. & Chapman, H.N., 2015. Possibilities for serial femtosecond crystallography sample delivery at future light sources. *Structural Dynamics*, 2(4).
Chen, C. et al., 2015. Characterization of Dye-decolorizing Peroxidase ( DyP ) from Thermomonospora curvata Reveals Unique Catalytic Properties of A-type DyPs *. , 290(38), pp.23447–23463.
Chen, V.B. et al., 2010. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 66(1), pp.12–21.
Chen, Z. et al., 2009. Distal histidine conformational flexibility in dehaloperoxidase from Amphitrite ornata. *Acta Crystallographica Section D: Biological Crystallography*, 65(1), pp.34–40.
Chreifi, G. et al., 2016. Crystal structure of the pristine peroxidase ferryl center and its relevance to proton-coupled electron transfer. *Proceedings of the National Academy of Sciences*, 113(5), pp.1226–1231. Available at: http://www.pnas.org/lookup/doi/10.1073/pnas.1521664113.
Colpa, D.I., Fraaije, M.W. & Van Bloois, E., 2014. DyP-type peroxidases: A promising and versatile class of enzymes. *Journal of Industrial Microbiology and Biotechnology*, 41(1), pp.1–7.
Contreras, H. et al., 2014. Characterization of a Mycobacterium tuberculosis Nanocompartment and Its Potential Cargo Proteins *. , 289(26), pp.18279–18289.
Cowtan, K., 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D: Biological Crystallography*, 62(9), pp.1002–1011.
D'Antonio, J. et al., 2010. Spectroscopic and mechanistic investigations of dehaloperoxidase B from amphitrite ornata. *Biochemistry*, 49(31), pp.6600–6616.
Diederichs, K. & Karplus, P.A., 2013. Better models by discarding data? *Acta Crystallographica Section D: Biological Crystallography*, 69(7), pp.1215–1222.
Domratcheva, T. & Schlichting, I., 2018. Spiers Memorial Lecture Introductory lecture: The impact of structure on photoinduced processes in nucleic acids and proteins. *Faraday Discussions*, 207, pp.9–26. Available at: http://dx.doi.org/10.1039/C8FD00058A.
Dworkowski, F.S.N. et al., 2015. Challenges and solutions for the analysis of in situ, in crystallo micro-spectrophotometric data. *Acta Crystallographica Section D: Biological Crystallography*, 71, pp.27–35.
Eaton, W.A. & Hochstrasser, R.M., 1967. Electronic Spectrum of Single Crystals of Ferricytochrome-c. *The Journal of Chemical Physics*, 46(7), pp.2533–2539.
Ebrahim, A. et al., 2019. Resolving polymorphs and radiation-driven effects in microcrystals using fixed-target serial synchrotron crystallography research papers. *Acta Crystallographica Section D: Biological Crystallography*, 75.

Elsliger, M.A. et al., 2010. The JCSG high-throughput structural biology pipeline. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 66(10), pp.1137–1142.

Emsley, P. et al., 2010. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 66(4), pp.486–501.

Ermler, U. et al., 1995. Crystal structure of the flavohemoglobin from Alcaligenes eutrophus at 1 . 75 A resolution. *The EMBO Journal*, 14(24), pp.6067–6077.

Evans, P.R. & Murshudov, G.N., 2013. How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography*, 69(7), pp.1204–1214.

Ferrari, R.P., Laurenti, E. & Trotta, F., 1999. Oxidative 4-dechlorination of 2,4,6-trichlorophenol catalyzed by horseradish peroxidase. *Journal of Biological Inorganic Chemistry*, (4), pp.232–237.

Franzen, S., Thompson, M.K. & Ghiladi, R.A., 2012. The dehaloperoxidase paradox. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1824(4), pp.578–588. Available at: http://dx.doi.org/10.1016/j.bbapap.2011.12.008.

Fraser, J.S. et al., 2011. Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proceedings of the National Academy of Sciences*, 108(39), pp.16247–16252. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.1111325108.

Fuchs, M.R. et al., 2014. D3, the new diffractometer for the macromolecular crystallography beamlines of the Swiss Light Source. *Journal of Synchrotron Radiation*, 21(2), pp.340–351.

Fuller, F.D. et al., 2017. Drop-on-demand sample delivery for studying biocatalysts in action at XFELs. *Nature Methods*, 14(4), pp.87–92.

Gaikwad, D.D. et al., 2015. Synthesis of indazole motifs and their medicinal importance: An overview. *European Journal of Medicinal Chemistry*, 90, pp.707–731. Available at: http://dx.doi.org/10.1016/j.ejmech.2014.11.029.

Gardner, a M. et al., 2000. Steady-state and transient kinetics of {E}scherichia coli nitr ic-oxide dioxygenase (flavohemoglobin). *J. Biol. Chem.*, 275(17), pp.12581–12589.

Gardner, P.R., 2012. Hemoglobin : A Nitric-Oxide Dioxygenase. , 2012. Available at: file:///Users/Victoria/Library/Application Support/Mendeley Desktop/Downloaded/Gardner - 2012 - Hemoglobin A Nitric-Oxide Dioxygenase.pdf.

Gardner, P.R. et al., 1998. Nitric oxide dioxygenase: an enzymic function for flavohemoglobin. *Proceedings of the National Academy of Sciences of the United States of America*, 95(September), pp.10378–10383.

Garman, E.F., 2010. Radiation damage in macromolecular crystallography: What is it and why do we care? *NATO Science for Peace and Security Series A: Chemistry and Biology*, pp.69–77.

Garman, E.F. & Weik, M., 2015. Radiation damage to macromolecules: Kill or cure? *Journal of Synchrotron Radiation*, 22(2), pp.195–200.

Garman, E.F. & Weik, M., 2017. X-ray radiation damage to biological macromolecules: Further insights. *Journal of Synchrotron Radiation*, 24(1), pp.1–6.

Gasteiger, E. et al., 2005. Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook*, pp.571–607. Available at: http://dx.doi.org/10.1385/1-59259-890-0:571.

Gidden, J. et al., 2010. NIH Public Access. , 283, pp.178–184.

Giger, W., Schaffner, C. & Kohler, H.P.E., 2006. Benzotriazole and tolyltriazole as aquatic contaminants. 1. Input and occurrence in rivers and lakes. *Environmental Science and Technology*, 40(23), pp.7186–7192.

Gribble, G.W., 2000. The natural production of organobromide compounds. *Environmental Science & Pollution Research*, 7(1), pp.37–49.

Gumiero, A. et al., 2011. Nature of the ferryl heme in compounds I and II. *Journal of Biological Chemistry*, 286(2), pp.1260–1268.

Hadfield, A. & Hadju, J., 1993. A fast and portable microspectrophotometer for protein crystallography. *Journal of Applied Crystallography*, 26, pp.839–842.

Helmick, R.A. et al., 2005. Imidazole antibiotics inhibit the nitric oxide dioxygenase function of microbial flavohemoglobin. *Antimicrobial Agents and Chemotherapy*, 49(5), pp.1837–1843.

Henderson, R., 1990. Cryo-protection of protein crystals against radiation damage in electron and X-ray diffraction. *Proceedings of the Royal Society B: Biological Sciences*, 241(1300), pp.6–8.

Holton, J.M., 2009. A beginner's guide to radiation damage. *Journal of Synchrotron Radiation*, 16(2), pp.133–142.

Horrell, S. et al., 2018. Enzyme catalysis captured using multiple structures from one crystal at varying temperatures. *IUCrJ*, 5, pp.283–292.

Horrell, S. et al., 2016. Serial crystallography captures enzyme catalysis in copper nitrite reductase at atomic resolution from one crystal. *IUCrJ*, 3, pp.271–281.

Hough, M.A. et al., 2008. Crystallography with Online Optical and X-ray Absorption Spectroscopies Demonstrates an Ordered Mechanism in Copper Nitrite Reductase. *Journal of Molecular Biology*, 378(2), pp.353–361.

Ilari, A. et al., 2002. The X-ray structure of ferric Escherichia coli flavohemoglobin reveals an unexpected geometry of the distal heme pocket. *Journal of Biological Chemistry*, 277(26), pp.23725–23732.

Kabsch, W., 2010. Xds. *Acta Crystallographica Section D: Biological Crystallography*, 66(2), pp.125–132.

Katona, G. et al., 2007. Raman-Assisted Crystallography. *Science*, 316(April), pp.449–453.

Keilin, D. & Mann, T., 1937. On the haematin compound of peroxidase. *Proceedings of the Royal Society B: Biological Sciences*, I22B, pp.119–133.

Kekilli, D. et al., 2014. Fingerprinting redox and ligand states in haemprotein crystal structures using resonance Raman spectroscopy. *Acta Crystallographica Section D: Biological Crystallography*, 70(5), pp.1289–1296.

Kekilli, D. et al., 2017. Photoreduction and validation of haem-ligand intermediate states in protein crystals by in situ single-crystal spectroscopy and diffraction. *IUCrJ*, 4, pp.263–270.

Kim, S.J.U.N. & Shoda, M., 1999. Purification and Characterization of a Novel Peroxidase from Geotrichum candidum Dec 1 Involved in Decolorization of Dyes These include : Purification and Characterization of a Novel Peroxidase from Geotrichum candidum Dec 1 Involved in Decolorization of . , 65(3), pp.1029–1035.

Knapp, J.E. et al., 2006. Allosteric action in real time: Time-resolved crystallographic studies of a cooperative dimeric hemoglobin. *Proceedings of the National Academy of Sciences*, 103(20), pp.7649–7654. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.0509411103.

Krissinel, E. and Hendrick, K., 2007. Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology*, 372, pp.774–797.

Kuhnel, K. et al., 2007. Structure and quantum chemical characterization of chloroperoxidase compound 0, a common reaction intermediate of diverse heme enzymes. *Proceedings of the National Academy of Sciences*, 104(1), pp.99–104. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.0606285103.

Kwon, H. et al., 2016. Direct visualization of a Fe(IV)-OH intermediate in a heme enzyme. *Nature Communications*, 7(Iv), pp.1–6. Available at: http://dx.doi.org/10.1038/ncomms13445.

LaCount, M.W. et al., 2000. The crystal structure and amino acid sequence of dehaloperoxidase from Amphitrite ornata indicate common ancestry with globins. *Journal of Biological Chemistry*, 275(25), pp.18712–18716.

Lebioda, L. et al., 1999. An enzymatic globin from a marine worm. *Nature*, 401(6752), pp.445–446.

Liu, J. et al., 2014. Metalloproteins containing cytochrome, iron-sulfur, or copper redox centers. *Chemical Reviews*, 114(8), pp.4366–4369.

McCombs, N.L. et al., 2017. Interaction of Azole-Based Environmental Pollutants with the Coelomic Hemoglobin from Amphitrite ornata: A Molecular Basis for Toxicity. *Biochemistry*, 56(17), pp.2294–2303.

McCoy, A.J. et al., 2007. Phaser crystallographic software. *Journal of Applied Crystallography*, 40(4), pp.658–674.

McGeehan, J. et al., 2009. Colouring cryo-cooled crystals: Online microspectrophotometry. *Journal of Synchrotron Radiation*, 16(2), pp.163–172.

McNicholas, S. et al., 2011. Presenting your structures: The CCP4mg molecular-graphics software. *Acta Crystallographica Section D: Biological Crystallography*, 67(4), pp.386–394.

Meents, A. et al., 2017. Pink-beam serial crystallography. *Nature Communications*, 8(1). Available at: http://dx.doi.org/10.1038/s41467-017-01417-3.

Moody, P.C.E. & Raven, E.L., 2018. The Nature and Reactivity of Ferryl Heme in Compounds i and II. *Accounts of Chemical Research*, 51(2), pp.427–435.

Mueller, C. et al., 2015. Fixed target matrix for femtosecond time-resolved and in situ serial micro-crystallography. *Structural Dynamics*, 2(5), pp.1–16. Available at: http://dx.doi.org/10.1063/1.4928706.

Mukai, M. et al., 2001. Flavohemoglobin, a Globin with a Peroxidase-like Catalytic Site. *Journal of Biological Chemistry*, 276(10), pp.7272–7277.

Murshudov, G.N. et al., 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, 67(4), pp.355–367.

Nave, C., 1995. Radiation damage in protein crystallography. *Radiation Physics and Chemistry*, 45(3), pp.483–490.

Neutze, R. et al., 2000. Potential for biomolecular imaging with femtoscond X-ray pulses. *Nature*, 406(August), pp.752–757.

Nogly, P. et al., 2015. Lipidic cubic phase serial millisecond crystallography using synchrotron radiation. *IUCrJ*, 2, pp.168–176.

Ogola, H.J.O. et al., 2009. Molecular characterization of a novel peroxidase from the cyanobacterium Anabaena sp. strain PCC 7120. *Applied and Environmental Microbiology*, 75(23), pp.7509–7518.

Ollesch, G. et al., 1999. Phospholipid bound to the flavohemoprotein from Alcaligenes eutrophus. *European Journal of Biochemistry*, 262(2), pp.396–405.

Owen, R.L. et al., 2009. A new on-axis multimode spectrometer for the macromolecular crystallography beamlines of the Swiss Light Source. *Journal of Synchrotron Radiation*, 16(2), pp.173–182.

Owen, R.L. et al., 2017. Low-dose fixed-target serial synchrotron crystallography. *Acta Crystallographica Section D: Structural Biology*, 73(4), pp.373–378.

Paoli, M., Marles-Wright, J. & Smith, A., 2002. Structure-function relationships in heme-proteins. *DNA and cell biology*, 21(4), pp.271–280. Available at: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L8222796%5Cnhttp://rug.on.worldcat.org/atoztitles/link/?sid=EMBASE&issn=00778923&id=doi:&atitle=Structure+function+relationships+in+lipotropins&stitle=ANN.+NEW+YORK+ACAD.+SCI.&title=Ann.

Pearson, A.R. et al., 2007. Tracking X-ray-derived redox changes in crystals of a methylamine dehydrogenase/amicyanin complex using single-crystal UV/Vis microspectrophotometry. *Journal of Synchrotron Radiation*, 14(1), pp.92–98.

Pearson, A.R. & Owen, R.L., 2009. Combining X-ray crystallography and single-crystal spectroscopy to probe enzyme mechanisms. *Biochemical Society Transactions*, 37(2), pp.378–381. Available at: http://biochemsoctrans.org/lookup/doi/10.1042/BST0370378.

Perutz, M.F., 1990. Mechanisms regulating the reactions of human hemoglobin with oxygen and carbon monoxide. *Annu. Rev. Physiol.*, 52, pp.1–25.

Petrus, M.L.C. et al., 2016. The DyP-type peroxidase DtpA is a Tat-substrate required for GlxA maturation and morphogenesis in Streptomyces. *Open Biology*, 6(1).

Pompidor, G. et al., 2013. A new on-axis micro-spectrophotometer for combining Raman, fluorescence and UV/Vis absorption spectroscopy with macromolecular crystallography at the Swiss Light Source. *Journal of Synchrotron Radiation*, 20(5), pp.765–776.

Potterton, L. et al., 2018. CCP 4 i 2: The new graphical user interface to the CCP 4 program suite. *Acta Crystallographica Section D: Structural Biology*, 74(August 2017), pp.68–84.

Poulos, T.L. & Kraut, J., 1980. The stereochemistry of peroxidase catalysis. *Journal of Biological Chemistry*, 255(17), pp.8199–8205.

Reedy, C.J., Elvekrog, M.M. & Gibney, B.R., 2008. Development of a heme protein structure - Electrochemical function database. *Nucleic Acids Research*, 36(SUPPL. 1), pp.307–313.

Ren, Z. et al., 1999. Laue crystallography : Coming of age Laue crystallography : coming of age. , (July).

Riccio, A. et al., 2002. The crystal structure of a tetrameric hemoglobin in a partial hemichrome state. *Proceedings of the National Academy of Sciences of the United States of America*, 99(15), pp.9801–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=125021&tool=pmcentrez&rendertype=abstract.

Rinaldi, A.C. et al., 2006. Interaction of Vitreoscilla hemoglobin with membrane lipids. *Biochemistry*, 45(13), pp.4069–4076.

Rodríguez, J. J. S., Padrón, M. E. T., Aufartová, J., and Ferrera, Z. & S., 2010. Benzimidazole Fungicides in Environmental Samples: Extraction and Determination Procedures. In O. Carisse, ed. *Fungicides*. Rijeka, Croatia: InTech.

Roedig, P. et al., 2016. Room-temperature macromolecular crystallography using a micro-patterned silicon chip with minimal background scattering. *Journal of Applied Crystallography*, 49, pp.968–975.

Ronda, L. et al., 2015. From protein structure to function via single crystal optical spectroscopy. *Frontiers in Molecular Biosciences*, 2(April), pp.1–14. Available at: http://journal.frontiersin.org/article/10.3389/fmolb.2015.00012/abstract.

Rossi, G.L. & Bernhard, S.A., 1970. Are the Structure and Function of an Enzyme the Same in Aqueous Solution and in the Wet Crystal ? *Journal of Molecular Biology*, 49, pp.85–91.

Russi, S. et al., 2017. Conformational variation of proteins at room temperature is not dominated by radiation damage. *Journal of Synchrotron Radiation*, 24(1), pp.73–82.

Salvachúa, D. et al., 2013. Enzyme from Irpex lacteus and Its Application in Enzymatic Hydrolysis of Wheat Straw. , 79(14), pp.4316–4324.

Santos, A., Mendes, S. & Brissos, V., 2013. New dye-decolorizing peroxidases from Bacillus subtilis and Pseudomonas putida MET94 : towards biotechnological applications.

Scheidt, W.R. & Chipman, D.M., 1986. Preferred Orientation of Imidazole Ligands in Metalloporphyrins. *Journal of the American Chemical Society*, 108(6), pp.1163–1167.

Schlichting, I., 2015. Serial femtosecond crystallography: The first five years. *IUCrJ*, 2(2013), pp.246–255.

Schlichting, I. et al., 2000. The Catalytic Pathway of Cytochrome P405cam at Atomic Resolution. *Science*, 287(2000), pp.1615–1622.

Schneider, S. et al., 2007. Diversity and conservation of interactions for binding heme in b-type heme proteins. *Natural Product Reports*, 24(3), pp.621–630.

De Serrano, V. et al., 2010. Structure of dehaloperoxidase B at 1.58 Å resolution and structural characterization of the AB dimer from Amphitrite ornata. *Acta Crystallographica Section D: Biological Crystallography*, 66(5), pp.529–538.

De Serrano, V. et al., 2007. X-ray crystal structural analysis of the binding site in the ferric and oxyferrous forms of the recombinant heme dehaloperoxidase cloned from Amphitrite ornata. *Acta Crystallographica Section D: Biological Crystallography*, 63(10), pp.1094–1101.

Sharp, K.H. et al., 2003. Crystal structure of the ascorbate peroxidase-ascorbate complex. *Nature Structural Biology*, 10(4), pp.303–307.

Šrajer, V. et al., 2000. Extraction of accurate structure-factor amplitudes from Laue data: wavelength normalization with wiggler and undulator X-ray sources. *Journal of Synchrotron Radiation*, 7(4), pp.236–244. Available at: https://doi.org/10.1107/S0909049500004672.

Srajer, V. & Royer, W., 2008. Time-resolved x-ray crystallography of heme proteins. *Methods in Enzymology*, 437(07), pp.379–395.

Von Stetten, D. et al., 2015. In crystallo optical spectroscopy (icOS) as a complementary tool on the macromolecular crystallography beamlines of the ESRF. *Acta Crystallographica Section D: Biological Crystallography*, 71, pp.15–26.

Stoner-Ma, D. et al., 2011. Single-crystal Raman spectroscopy and X-ray crystallography at beamline X26-C of the NSLS. *Journal of Synchrotron Radiation*, 18(1), pp.37–40.

Strittmatter, E. et al., 2013. First crystal structure of a fungal high-redox potential dye-decolorizing peroxidase substrate interaction sites and long-range electron transfer. *Journal of Biological Chemistry*, 288(6), pp.4095–4102.

Sturm, A. et al., 2006. YcdB from Escherichia coli Reveals a Novel Class of. , 281(20), pp.13972–13978.

Sugano, Y., 2009. DyP-type peroxidases comprise a novel heme peroxidase family. *Cellular and Molecular Life Sciences*, 66(8), pp.1387–1403.

Sugano, Y. et al., 2007. DyP, a unique dye-decolorizing peroxidase, represents a

novel heme peroxidase family: ASP171 replaces the distal histidine of classical peroxidases. *Journal of Biological Chemistry*, 282(50), pp.36652–36658.

Uervirojnangkoorn, M. et al., 2015. Enabling X-ray free electron laser crystallography for challenging biological systems from a limited number of crystals. *eLife*, 2015(4), pp.1–29.

Vagin, A. & Teplyakov, A., 1997. MOLREP: An Automated Program for Molecular Replacement. *Journal of Applied Crystallography*, 30(6), pp.1022–1025.

Velde, F. Van De, Rantwijk, F. Van & Sheldon, R.A., 2001. Improving the catalytic performance of peroxidases in organic synthesis. *Trends in Biotechnology*, 19(2), pp.73–80.

Vinogradov, S.N. & Moens, L., 2008. Diversity of globin function: Enzymatic, transport, storage, and sensing. *Journal of Biological Chemistry*, 283(14), pp.8773–8777.

Ward, J.F., 1988. DNA Damage Produced by Ionizing Radiation in Mammalian Cells: Identities, Mechanisms of Formation, and Reparability. *Progress in Nucleic Acid Research and Molecular Biology*, 35(C), pp.95–125.

White, T.A. et al., 2012. CrystFEL: A software suite for snapshot serial crystallography. *Journal of Applied Crystallography*, 45(2), pp.335–341.

Winter, G. et al., 2018. DIALS: Implementation and evaluation of a new integration package. *Acta Crystallographica Section D: Structural Biology*, 74, pp.85–97.

Yano, J. et al., 2005. X-ray damage to the Mn4Ca complex in single crystals of photosystem II: A case study for metalloprotein crystallography. *Proceedings of the National Academy of Sciences*, 102(34), pp.12047–12052. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.0505207102.

Yoshida, T. et al., 2011. The catalytic mechanism of dye-decolorizing peroxidase DyP may require the swinging movement of an aspartic acid residue. *FEBS Journal*, 278(13), pp.2387–2394.

Zaefferer, S., 2000. New developments of computer-aided crystallographic analysis in transmission electron microscopy research papers. *Journal of Applied Crystallography*, 33(v), pp.10–25.

Zeldin, O.B. et al., 2013. Predicting the X-ray lifetime of protein crystals. *Proceedings of the National Academy of Sciences*, 110(51), pp.20551–20556. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.1315879110.

Zhao, J. et al., 2013. Structural and kinetic study of an internal substrate binding site in dehaloperoxidase-hemoglobin A from Amphitrite ornata. *Biochemistry*, 52(14), pp.2427–2439. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23480178.

Zhu, H. & Riggs, A.F., 1992. Yeast flavohemoglobin is an ancient protein related to globins and a reductase family. *Proceedings of the National Academy of Sciences of the United States of America*, 89(11), pp.5015–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=49219&tool=pmcentrez&rendertype=abstract.