# NEUTRON CRYSTALLOGRAPHY OF A HEME PEROXIDASE

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Cecilia Maria Casadei

Department of Biochemistry

University of Leicester

**2015**

# Neutron crystallography of a heme peroxidase

**Cecilia Maria Casadei**

**Abstract**

Heme peroxidases are an ubiquitous family of catalytic iron-containing proteins. These enzymes are responsible for the catalysis of hydrogen peroxide removal from the cell through the formation of high valent transient states of the heme cofactor. The purpose of the present research project is to investigate the reaction pathway of cytochrome *c* peroxidase. The study of hydrogen related chemical features in the active site of this enzyme in the resting state and in the transient species is of paramount importance for the clarification of the reaction mechanism. In these circumstances neutron crystallography is the technique of choice. Neutron crystallography allows the direct localization of deuterium substituted hydrogen atoms in macromolecules in absence of radiation-induced damage.

The structures of the resting state and of cryo-trapped compound I of cytochrome *c* peroxidase were determined by neutron crystallography. The nature of the catalytic center was investigated and in particular the protonation state of the heme iron axial ligand and of the key catalytic residues was established. These findings contributed to the understanding of the reaction pathway from the resting state to the intermediate species compound I. Complementary spectroscopic techniques were employed to assess compound I formation in single crystals and its stability in the conditions of the neutron crystallography experiment.

The temperature dependence of the key catalytic features of cytochrome *c* peroxidase in the resting state was investigated by neutron crystallography. Significant alterations of protonation states were found in the resting state at cryogenic temperature. In addition the nature of the heme iron distal ligand was found to be affected by temperature in the resting state and it was possible to relate this phenomenon to the temperature induced spin state change observed by spectrophotometry in the visible region.

As part of the present project, contribution was given to the development of the cryogenic temperature sample environment at the neutron macromolecular diffractometer LADI-III of the Institut Laue-Langevin broadening the experimental capabilities of the instrument. A wide range of experiments are made possible by the new set-up including the cryo-trapping of catalytic intermediates, the study of the temperature dependence of structural features and the investigation of species and complexes that are not stable at ambient temperature.

# Acknowledgments

I would like to thank my supervisors Prof. P. Moody, Dr. M. Blakeley and Prof. E. Raven for their support and their precious advice, the collaborators who took part in the project: A. Ostermann and T. Schrader from FRM II; J. Devos, S. Martinez, R. Cubitt and S. Teixeira at the ILL; D. Von Stetten at the ESRF; M. G. Concilio and A. Fielding at the University of Manchester; H. Kwon, S. Kapetanaki, A. Gumiero and J. Basran from the University of Leicester.

On the personal side, a special thank to my parents Rosa and Folco, and my boyfriend Stefano. They have always supported and encouraged me. I would also like to thank my friends from the Physics department Daniele, Matteo and Laura, that despite the distance have been supportive throughout my PhD.

# Contents

# List of Tables

# List of Figures

# List of abbreviations

**A.U.** absorbance units.

**CcP** cytochrome *c* peroxidase.

**CCP4** Collaborative Computational Project 4.

**EPR** Electronic Paramagnetic Resonance.

**FPLC** fast protein liquid chromatography.

**FRM II** Forschungs-Neutronenquelle Heinz Maier-Leibniz.

**GUI** graphical user interface.

**LB** lysogeny broth.

**MLZ** Heinz Maier-Leibnitz Zentrum 2.

**MPD** (±)-2-methyl-2,4-pentanediol.

**NTA** nitrilotriacetic acid.

**O.D.** Optical Density.

**PDB** Protein Data Bank.

**PMSF** phenylmethane-sulfonyl fluoride.

**SDS** sodium dodecyl sulfate.

**SDS-PAGE** sodium dodecyl sulfate polyacrylamide gel electrophoresis.

**SLS** Swiss Light Source.

**UV-VIS** Ultraviolet-visible.

# Part I

# Introductory aspects

# Chapter 1

# Neutron crystallography of heme peroxidases

This chapter presents an introduction to the characteristics of the heme peroxidase family of enzymes with particular focus on cytochrome *c* peroxidase, the protein investigated in the present work. The general processes involved in the catalytic mechanism are well established and are introduced in this chapter together with an overview of the studies carried out on cytochrome *c* peroxidase prior to the present work. The open questions concerning the structural details of the catalytic intermediates and the relevance of those in view of the study of the reaction mechanism are introduced. The choice of neutron crystallography as a primary tool for investigating such intermediates is discussed together with the role of complementary spectroscopic techniques. To conclude, the aim of the present work is stated.

## 1.1 Heme peroxidases

Heme peroxidases are a family of catalytic iron-containing proteins that are found in nearly all living organisms. These enzymes catalyse hydrogen peroxide removal from the cell *via* oxidation of a substrate. Hydrogen peroxide is a potentially hazardous metabolic by-product whose reduction to water is achieved by means of the formation of high valent

intermediates of the enzyme active site.

The heme peroxidase active site is characterized by the presence of the heme cofactor. The heme is the result of the complexation of iron into a porphyrin macrocycle as shown in Figure 1.1. The iron exhibits different oxidation states depending on the catalytic species. The iron shows distorted octahedral coordination where the equatorial ligands are the nitrogen atoms of the pyrrolic groups of the heme cofactor. The axial ligands are known as the proximal and distal ligands. The proximal ligand is endogenous and is provided by the imidazole ring of a histidine side chain. The distal ligand is exogenous namely it doesn't belong to the polypeptide chain. Its characterization together with the study of the nature of the coordinative bond formed with the heme iron is one of the aspects investigated in this work.

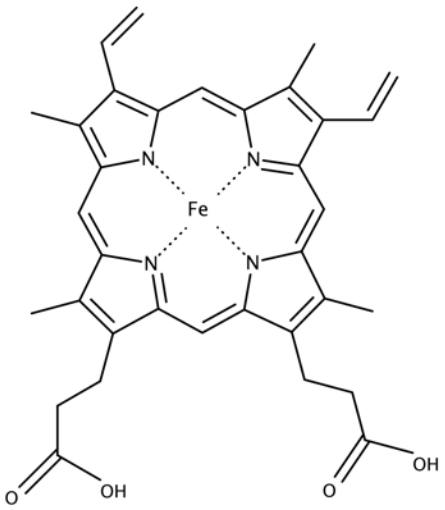


Figure 1.1: The heme cofactor.

Heme peroxidases share a common catalytic cycle involving two intermediate species known as compound I (Cpd I) and compound II (Cpd II) which represent a distinctive feature of the heme peroxidase family. The general processes involved in the catalytic pathway are well established and can be formulated as follows [1]:

$$\mathrm{P} + \mathrm{H_2O_2} \rightarrow \mathrm{Cpd\ I} + \mathrm{H_2O} \tag{1.1}$$

$$\text{Cpd I} + \text{HS} \rightarrow \text{Cpd II} + \text{S}\cdot \tag{1.2}$$

$$\text{Cpd II} + \text{HS} \rightarrow \text{P} + \text{H}_2\text{O} + \text{S}\cdot \tag{1.3}$$

where P represents the peroxidase in the resting state while HS and S· are the substrate in the reduced and oxidised form respectively. The resting state of the enzyme presents a ferric heme *i.e.* a heme iron in the oxidation state $Fe^{3+}$. In the first step of the catalytic cycle the enzyme in the resting state reacts with hydrogen peroxide to form the transient compound I, which is oxidised by two electron equivalents above the resting state. Compound I is reduced back to the peroxidase ferric state through two sequential single electron transfer steps. In the first step compound I is reduced by oxidation of the substrate to a second transient species known as compound II which is one oxidation equivalent above the resting state. The oxidized intermediates compound I and II are known as ferryl species characterized by $Fe^{4+}$ in the active site. Compound II is finally reduced back to the ferric enzyme *via* oxidation of a substrate molecule.

## 1.2 Cytochrome *c* peroxidase

The present work focuses on the heme peroxidase enzyme cytochrome *c* peroxidase (C*c*P) from *Saccharomyces cerevisiae* present in the intermembrane space of *S. cerevisiae* mitochondria. Cytochrome *c* peroxidase is a 34 kDa monomeric protein with 294 residues and one non-covalent bound heme cofactor per monomer.

The physiological substrate of cytochrome *c* peroxidase is the heme protein cytochrome *c* (cyt *c*). C*c*P catalyzes hydrogen peroxide reduction *via* oxidation of ferrocytochrome *c* presenting an $Fe^{2+}$ metal center to ferricytochrome *c* characterized by $Fe^{3+}$. Equations 1.1 - 1.3 take the form:

$$\text{C}c\text{P} + \text{H}_2\text{O}_2 \rightarrow \text{Cpd I} + \text{H}_2\text{O} \tag{1.4}$$

$$\text{Cpd I} + \text{cyt } c\ [\text{Fe}^{2+}] \rightarrow \text{Cpd II} + \text{cyt } c\ [\text{Fe}^{3+}] \tag{1.5}$$

$$\text{Cpd II} + \text{cyt } c\ [\text{Fe}^{2+}] + 2\text{H}^+ \rightarrow \text{C}c\text{P} + \text{H}_2\text{O} + \text{cyt } c\ [\text{Fe}^{3+}] \tag{1.6}$$

where compound I is characterized by a ferryl ion $Fe^{4+}$ and a protein-based cationic radical Trp-191$^{\cdot+}$ [2] while compound II presents a ferryl heme only.

Recombinant cytochrome *c* peroxidase is expressed with high yield in *E. coli*. Its purification and crystallization do not present particular challenges. C*c*P crystallizes in space group $P2_12_12_1$ of the orthorombic system and it was the first of all peroxidases to have its structure solved [3], [4]. The first cytochrome *c* peroxidase X-ray crystal structure was determined in 1980 [4]. This very first structure was solved by experimental phasing *via* multiple isomorphous replacement by using mercury derivatives accompanied by density modification procedures. The first Protein Data Bank (PDB) entry for cytochrome *c* peroxidase is from 1984 with accession code 2CYP [5]. After 2CYP, over one hundred X-ray structures of *S. cerevisiae* cytochrome *c* peroxidase were published in the PDB including those of mutants and those of complexes with a wide variety of ligands. Figure 1.2 shows the structure of C*c*P (PDB entry 4CVI).

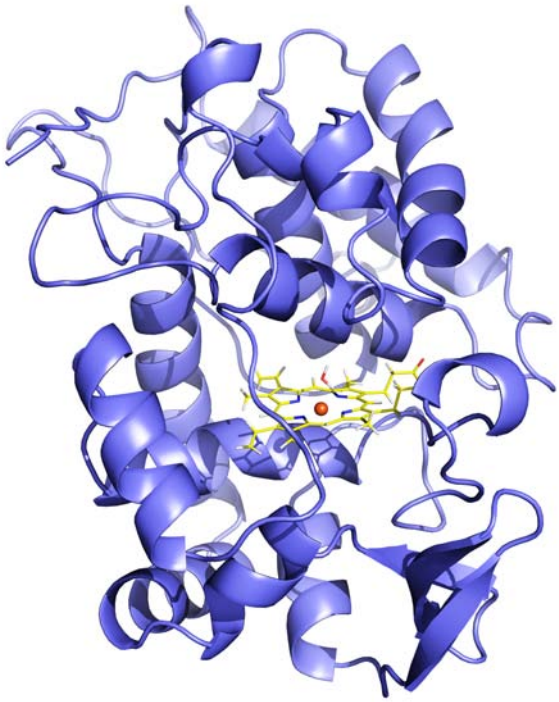

Figure 1.2: The structure of C*c*P. PDB entry 4CVI published as part of the present work.

Cytochrome *c* peroxidase became the paradigm for heme peroxidases. Nevertheless C*c*P presents at least two distinctive features compared to the other peroxidases. Its substrate is a macromolecule while other heme peroxidases catalyze the oxidation of small molecules. In addition compound I radical is localized on the side chain of a residue in the polypeptide while other peroxidases exhibit a porphyrin-based radical.

The structure of the 1:1 complex between C*c*P and its physiological substrate cytochrome *c* from *S. cerevisiae* was also solved, providing some insight into the position of the substrate binding site and the nature of the inter-protein interaction within the complex [6]. The characteristics of the enzyme-substrate interaction in C*c*P have been the object of numerous studies and it is today established that while the enzyme presents two substrate binding sites, only the site determined by crystallography in [6] is responsible for electron transfer between redox partners [7].

## 1.3 Cytochrome *c* peroxidase catalytic intermediates

Although general understanding of the reaction cycle of cytochrome *c* peroxidase was achieved, the complete rationalization of the mechanism leading to the formation of the transient species compound I and compound II as well as the structural features of the active site of these intermediates was far from being well established at the time when this research project was proposed.

Extensive work was carried out with the purpose of unraveling the structure and properties of the high valent ferryl intermediates compound I and compound II of the catalytic cycle and the precise nature of these transient species has been a matter of debate for decades. A multiplicity of biochemical and biophysical techniques were used to characterize cytochrome *c* peroxidase and its transient intermediates involving different laboratories and a variety of approaches. A review of these long-standing efforts is presented in [8], [9], [10].

Once the general reaction pathway is established (Equations 1.4-1.6), the understanding of the chemical processes taking place during the reaction cycle requires knowledge of

the three dimensional configuration of the active site at atomic level in the resting state and in the intermediate species. This includes the localization of hydrogen atoms *i.e.* the determination of protonation states of the catalytic residues in the active site, that of the heme iron distal ligand and the study of the hydrogen bond network in the heme pocket. Although the understanding of hydrogen related chemistry is essential for the elucidation of reaction pathways in redox enymes since electron transfer is often accompanied by proton transfer, the determination of hydrogen related features is a challenge in most circumstances.

While the heme iron distal ligand was characterized as a water molecule in physiological conditions, the nature of the ligand in the transient species is matter of debate and that, together with the state of protonation of catalytic residues in the heme pocket constitutes fundamental information for the formulation of the reaction mechanism. In ferryl compound I and II the question is whether the the active site exhibits a ferryl-oxo or a ferryl-hydroxide group as shown in Figure 1.3.


Figure 1.3: Possible formulations of the ferryl heme in compound I and compound II of C*c*P . Left: iron-oxo; right: iron-hydroxide.

The problem has been addressed by a number of authors making use of a variety of techniques. The two ferryl species can be discerned based on the nature of the iron-oxygen bond rather than by direct detection of the hydrogen atom. Resonance Raman methods were employed allowing the characterization of single and double bonds based on the study of the iron-oxygen vibrational frequency. However laser excitation in this type of spectroscopy alters the state of photolabile intermediates [11]. Early approaches to the problem included extended X-ray absorption fine structure spectroscopy where the structure of the iron X-ray absorption edge is studied to extract information on the

chemical environment. X-ray induced damage gives rise to ambiguous and unreliable results. A review of these early methods is given in [8]. Photodamage induced by X-ray exposure is discussed thoroughly in the following.

X-ray crystallography methods were used by different authors to gain information about the nature of the ferryl heme. X-ray diffraction is a direct probe of protonation states only when resolution beyond 1.2 Å is achieved. Atomic X-ray form factors, which constitute a measure of the scattering power of an element, are directly proportional (at transferred wavevector equal to zero) to the number of electrons of the atomic species. The atomic X-ray form factor of hydrogen is much smaller than that of the other elements in a protein structure. Hence high resolution beyond 1.2 Å is required to localize hydrogen atoms in the electron density map of a protein structure and most protein crystals do not diffract to such extent. Cytochrome *c* peroxidase does not constitute an exception. X-ray diffraction can then be used only as an indirect probe of the ferryl oxygen protonation state based on the iron-oxygen bond length which is expected to be relatively short and around 1.6 - 1.7 Å in the iron-oxo case and longer in the iron-hydroxide case. This type of study provides indirect information and greatly depends on the degree of accuracy and precision with which atomic positions and interatomic distances are determined.

Both the direct visualization of hydrogens in electron density maps and the indirect inference of the ferryl oxygen protonation state based on the iron-oxygen distance are hampered by the occurrence of radiation damage effects following X-ray exposure, which are particularly severe in metallo-proteins [12], [13]. Photodamage during X-ray exposure is the main factor hampering the reliability of X-ray structural information in heme proteins. X-rays interact with electrons and at wavelengths of the order of $10^{-10}$ m typical of diffraction experiments X-ray photons carry an energy of the order of $10^4$ eV. According to [14] the main source of radiation damage in protein crystals is due to the production of free radicals in water molecules. Others [12] propose that following X-ray exposure, Auger electrons are liberated mainly by light elements. Thermalization of Auger electrons within the sample produces a large number of cascade electrons, of the order of $10^3$ per absorbed X-ray photon. In either case radiation induced damage and in particular alteration of the

oxidation state of a redox center are the unavoidable consequence of X-ray exposure.

X-ray exposure at cryogenic temperature limits the extent of radiation induced damage. However in the case of metallo-proteins these effects are in general particularly severe and low temperature data collection does not suffice in the prevention of photodamage. In cytochrome *c* peroxidase it was well established that radiation damage produces photoreduction effects in the heme cofactor [15] and in particular in C*c*P compound I photodamage becomes apparent at X-ray doses exceeding 0.02 MGy, well below the dose received by the crystalline sample during a typical diffraction data collection.

An inconsistent picture emerged from early studies of compound I ferryl ligand protonation state. A review of the issue is given in [16] where the incompatibility of the results obtained by different techniques and research groups is investigated. It is now recognized that these early approaches *i.e.* X-ray crystallography, X-ray absorption and resonance Raman spectroscopy were affected by radiation-induced damage hampering the reliability of the findings.

The problem of photo-induced damage in X-ray crystallography was circumvented by developing multiple crystal data collection protocols at cryogenic temperature. Several isomorphous crystals are used, none of which receives a dose larger than the threshold value of 0.02 MGy and complete sampling of reciprocal space is achieved by merging the datasets [15], [17]. Even though the multiple crystal approach at cryogenic temperature is a valuable method for minimizing radiation damage, this problem which is inherent to X-ray beam exposure cannot be fully eliminated.

The work presented in this thesis is aimed at addressing the question concerning hydrogen positions in the catalytic site of cytochrome *c* peroxidase reaction intermediates by using a different approach. Neutron crystallography and neutron cryo-crystallography are the techniques of choice when the biologically relevant question concerns hydrogen related structural features and chemical processes. The next section provides an overview of neutron macromolecular diffraction describing the unique features of this technique and the factors that limit its applicability.

## 1.4 Neutron macromolecular crystallography

Neutron macromolecular crystallography can provide complementary and supplementary information to that gained from X-ray studies. Nevertheless the study of biological macromolecules by means of neutron diffraction techniques has been rather limited and of over 100,000 protein structures deposited in the Protein Data Bank only 83 have been determined by means of neutron crystallography.

The use of neutron macromolecular diffraction has been historically limited by the inherent low flux of neutron sources with $10^6$ to $10^8$ neutrons cm$^{-2}$ s$^{-1}$ at the sample position. In comparison, X-ray beams with flux of $10^{16}$ photons cm$^{-2}$ s$^{-1}$ are available at synchrotron sources [18]. The inherent low flux of neutron beams results in long data collection times and the requirement for large volume crystals. In the early days neutron protein crystallography experiments would require weeks or months to achieve complete data collection and crystal volumes of several cubic millimeters. The recent developments of neutron beamline technology and methods, the advances in sample preparation together with the progress in refinement methods and computational tools have dramatically improved the speed of data collection and the quality with which neutron protein structures can now be determined [19]. In addition, smaller sample volumes can now be used and larger unit cells can be studied.

The properties of neutron-matter interaction provide a two-fold advantage in protein crystallography studies.

- The lack of neutron induced radiation damage ensures that the structure being investigated is not altered by the experimental probe. While X-rays interact with electrons, neutrons interact with atomic nuclei and within the energy range used in crystallography the neutron-sample interaction does not produce observable radiation damage [14]. Wavelengths of the order of 1 Å correspond to neutron energy of the order of $10^{-2}$ eV while at the same wavelength photon energy is of the order of $10^4$ eV. This is due to different energy-wavelength relations for massive and massless particles: $E = p^2/2m = h^2/2m\lambda^2$ in the case of (non-relativistic) neutrons and

$E = hc/\lambda$ in the case of photons where $E$ is the neutron or photon energy, $m$ is the neutron mass, $\lambda$ the neutron or photon wavelength, $h$ represents Planck's constant and $c$ the speed of light.

- In addition neutron crystallography is capable of localizing deuterium substituted hydrogen atoms, providing essential information concerning hydrogen-related chemistry. This aspect will be further developed in the following sections and the theoretical background is derived in Chapter 2 and discussed in detail in Section 2.4.4 of the present manuscript.

### 1.4.1 Deuteration and perdeuteration

This section reports an introduction to the peculiarities of neutron scattering that lead to the possibility of studying hydrogen related structural features in macromolecules via replacement of hydrogen for deuterium. These aspects are considered in depth in Chapter 2 where the basics of the theory of low energy neutron scattering are presented.

The X-ray atomic form factor of hydrogen and its isotope deuterium equals $0.28 \cdot 10^{-12}$ cm. Other elements consituting proteins show larger values of the form factor, for example in the case of carbon the form factor amounts to $1.69 \cdot 10^{-12}$ cm (Table 1.1). Hydrogen atoms can be localized only in X-ray crystal structures determined to resolution of 1.2 Å and beyond [20]. As stated in Section 1.3 heme enzymes are very sensitive to photo-induced damage which can lead to the alteration of the oxidation state of the redox center, preventing the use of high X-ray doses required in ultra-high resolution diffraction data collection. In addition ionized hydrogen atoms *i.e.* protons do not contribute to X-ray scattering due to the lack of electrons. Hence isolated protons cannot be detected by means of X-ray crystallography.

Neutron crystallography of deuterated or perdeuterated samples provides an efficient tool for determining the position of deuterium substituted hydrogen atoms in absence of radiation damage. The neutron scattering length of protons equals $\langle b \rangle = -0.37 \cdot 10^{-12}$ cm where the average $\langle\ \rangle$ is performed over the spin states of the neutron-nucleus two particles system. Due to its negative scattering length protons appear as negative

density in neutron scattering density maps, in contrast to most isotopes in protein crystals that appear as positive density peaks (Table 1.1). This generally leads to cancellation of neutron scattering density of -$CH_2$ and -$CH_3$ groups at resolution greater than 1.5 Å hampering the accurate interpretation of the neutron map [18].

In addition protons have a large incoherent scattering cross-section giving rise to significant background in the crystallography experiment [18]. The values of the coherent and incoherent scattering cross-sections of the proton are reported hereafter:

$$\sigma_{\text{coherent}}(^1\text{H}) = 1.76\,\text{barn} \tag{1.7}$$

and

$$\sigma_{\text{incoherent}}(^1\text{H}) = 80.27\,\text{barn}. \tag{1.8}$$

The origin of incoherent scattering resides in the spin dependence of the neutron scattering length, as shown by the following equations:

$$\sigma_{\text{coherent}} = 4\pi|\langle b\rangle|^2 \tag{1.9}$$

and

$$\sigma_{\text{incoherent}} = 4\pi\langle|b - \langle b\rangle|^2\rangle \tag{1.10}$$

where $b$ is spin dependent and $\langle\ \rangle$ denotes the average over singlet and triplet spin states for the proton - neutron system [14].

By contrast deuterium has a positive neutron scattering length of $0.66 \cdot 10^{-12}$ cm that is similar to the scattering lengths of the other atoms found in proteins, namely carbon, oxygen, nitrogen and sulphur (Table 1.1). In addition deuterium incoherent scattering cross section is about 40 times smaller than that of hydrogen.

The signal to noise ratio in neutron diffraction data from a fully hydrogenated system is dominated by the large incoherent background. Replacing hydrogen by deuterium increases the coherent scattering signal and decreases the incoherent background so that the signal-to-noise ratio typically improves by nearly one order of magnitude [18]. Deuteration can be achieved either partially, by soaking the crystals in a deuterated mother liquor,

or fully (perdeuteration) by expressing the protein under deuterated conditions *in vivo*. In the first case hydration water will be replaced by deuterated water and exchangeable hydrogens (*i.e.* hydrogen atoms that are bound to oxygen or nitrogen) will be exchanged for deuterium. In the latter case hydrogen atoms that are bound to carbon will also be replaced. Due to the significant background reduction, perdeuteration enables radically smaller crystals to be used in neutron data collection [21], [22] and [23].

In the present thesis the terms 'deuterium-exchanged', '$D_2O$-soaked' and 'partially deuterated' are used to refer to crystals where only hydrogens bound to oxygen and nitrogen have been exchanged for deuterium. On the other hand the terms 'perdeuterated' and 'fully deuterated' are used as synonyms to denote protein where the totality of hydrogen atoms have been replaced by deuterium.

Table 1.1: X-ray atomic form factor at scattering angle equal to zero, neutron coherent scattering length, neutron coherent and incoherent scattering cross sections for the chemical species ${}^{\mathrm{A}}_{\mathrm{Z}}\mathrm{X}$ where A denotes the mass number and Z the atomic number. Where the mass number is omitted, data refer to the natural mixture of isotopes.

| ${}^{\mathrm{A}}_{\mathrm{Z}}\mathrm{X}$ | $f_{\mathrm{X-ray}}(10^{-12}\mathrm{cm})$ | $\langle b \rangle (10^{-12}\mathrm{cm})$ | $\sigma_{\mathrm{coherent}}$ (barn) | $\sigma_{\mathrm{incoherent}}$ (barn) |
|---|---|---|---|---|
| ${}^{1}_{1}\mathrm{H}$ | 0.28 | $-0.37$ | 1.76 | 80.27 |
| ${}^{2}_{1}\mathrm{H}$ | 0.28 | $+0.67$ | 5.59 | 2.05 |
| ${}_{6}\mathrm{C}$ | 1.69 | $+0.66$ | 5.55 | 0.00 |
| ${}_{7}\mathrm{N}$ | 1.97 | $+0.94$ | 11.01 | 0.50 |
| ${}_{8}\mathrm{O}$ | 2.25 | $+0.58$ | 4.23 | 0.00 |
| ${}_{15}\mathrm{P}$ | 4.20 | $+0.51$ | 3.31 | 0.01 |
| ${}_{16}\mathrm{S}$ | 4.48 | $+0.28$ | 1.02 | 0.01 |

### 1.4.2 Neutron cryo-crystallography

Cold neutrons (6.7 meV, 3.5 Å) do not cause radiation damage effects in protein crystals. Therefore neutron structures are routinely determined at ambient temperatures. However cryo-crystallography techniques and flash-cooling strategies can be adopted for trapping transient, labile or intermediate species in protein crystals [24].

Protein crystals typically contain 35 - 70% aqueous solvent and fast cooling rates are required to avoid crystalline ice formation that would otherwise damage the crystal lattice. Protein crystals are therefore typically flash-cooled by rapid immersion in fluids at temperatures lower than the 140 K glass transition temperature of water with the purpose of vitrifying the solvent in and around the crystal.

In neutron crystallography, where the available neutron fluxes at reactor and spallation sources are relatively low and large crystals are required for data collection, the flash-cooling of protein crystals is particularly challenging due to the finite rate of heat transfer from the interior of the crystal to its surface and from its surface to the exterior cooling agent. Temperature gradients across the sample's volume can lead to increased mosaic spread hampering the crystal order. Extensive experimental work was done by various research groups in order to optimize cooling protocols by using different cryoprotectants and cooling agents in order to enable the freeze-trapped capture of kinetic intermediates in proteins.

As part of the present project, a cryogenic temperature set-up has been adapted and installed on the Laue diffractometer LADI-III at the Institut Laue-Langevin with accessible temperature range between 80 and 400 K substituting the previous liquid helium-based low temperature sample environment [24], [25]. The description of the cryogenic temperature set-up at LADI-III is presented in Chapter 3 of the present work.

It is worthwhile observing that flash-cooling rates are slower for larger crystals increasing the crystal disorder (mosaic spread) and hampering the resolution of the experiment. In addition the formation of cryo-artefacts in the structure is possible and this must be taken into account while interpreting the experimental results. On the other hand the reduced thermal motion at cryogenic temperatures generally results in lower Debye-Waller

factors enhancing the visualization of the more mobile parts of the system.

## 1.5 Complementary spectroscopies

In the present research project the structural characterization by crystallography of the catalytic intermediates of cytochrome *c* peroxidase reaction pathway was accompanied by complementary studies aimed at assessing the formation of the species of interest. The diagnostics of the enzyme intermediates was carried out by ultraviolet-visible spectrophotometry and electronic paramagnetic resonance.

The heme group constituting the core of the catalytic activity of C*c*P is also a chromophore since electronic transitions involving the cofactor give rise to absorbance peaks in the visible region of the electromagnetic spectrum. The spectral features of ultraviolet-visible light absorption are therefore related to the structure of the active site. Although it is not possible to quantitatively correlate absorbance peaks to electronic transitions in the chromophore *i.e.* predict the spectral features based on the study of molecular orbitals of the metal-ligands system, ultraviolet-visible spectrophotometry still provides valuable information. It is in fact possible to empirically correlate spectral features in the visible region of the electromagnetic spectrum to the various species of interest, and it was observed that minute structural variations in the chromophore lead to significant absorbance changes. Ultraviolet-visible spectra can be collected in solution and in crystalline samples. Like crystallography, temperatures of 100 K and 298 K are accessible.

Electronic paramagnetic resonance yields information on unpaired electrons and spin state of the metal center and highlights the presence of radical species. Single crystal and solution samples can be studied, however operation temperatures are generally below 10 K, different from typical crystallography studies. A review of electronic paramagnetic resonance techniques in heme proteins characterization can be found in [26].

## 1.6 The purpose of this work

The purpose of the present work is the full structural characterization of the active site of cytochrome *c* peroxidase in the resting state and in its catalytic intermediates. In particular, hydrogen-related chemical information turns out to be essential for the clarification of reaction mechanism in this enzyme.

The main focus of this work is the determination of the nature of the heme iron distal ligand and in particular its protonation state as well as the protonation state of key catalytic residues in the active site of cytochrome *c* peroxidase in the resting state and in the reaction intermediates.

Neutron crystallography allows the localization of deuterium substituted hydrogen atoms in protein structures therefore it is the main technique employed in the present work. Complementary techniques are also used. X-ray crystallography is essential to gain supplementary structural information while spectroscopic techniques are adopted in the characterization of catalytic intermediates and the assessment of reaction protocols.

# Chapter 2

# Low energy neutron scattering

This chapter presents a general elementary introduction to the theory describing the interaction between low energy neutrons and matter. Most information presented here is based on reference [27]. An overview of the theory of low energy neutron scattering is presented in the book by G. L. Squires [28]. The purpose of this chapter is to describe in a non-exhaustive fashion the theoretical framework that puts into context the neutron diffraction experiments carried out in the present work.

Elastic scattering is considered first and the expression of the scattering amplitude within the Born approximation is derived (Sections 2.1 and 2.2). This is an equation of paramount importance in almost any neutron scattering experiment. Extension of the formalism to the general case of inelastic scattering leads to the derivation of the expression of Fermi's Golden Rule, the master equation of scattering and the dynamic structure factor (Section 2.3).

In Section 2.4 the general expression of the dynamic structure factor is reformulated as the sum of two terms corresponding to the phenomena of coherent and incoherent scattering. This aspect is particularly relevant in the experimental work carried out within the present research project since it provides the theoretical background for understanding the requirement for deuteration and perdeuteration in neutron macromolecular crystallography (Section 1.4).

The physical interpretation of coherent and incoherent scattering is presented in Sec-

tion 2.5 by deriving the relationship between scattering cross sections and correlation functions. Finally, M. von Laue diffraction condition is derived in Section 2.6 by considering the coherent and elastic component within the general expression of the dynamic structure factor.

## 2.1 The scattering amplitude

Non-relativistic scattering of one particle of mass $m$ by a static (*i.e.* time-independent) potential:

$$V = V(\vec{r}) \tag{2.1}$$

is considered. In these circumstances the scattering event is necessarily elastic[1]. Time symmetry and the conservation of energy are in fact related by Noether's theorem [29]. Schrödinger equation for the particle wavefunction $\psi(\vec{r}, t)$ reads:

$$i\hbar \frac{\partial}{\partial t}\psi(\vec{r}, t) = \mathcal{H}\psi(\vec{r}, t) \tag{2.2}$$

where $\hbar$ is Planck's constant $h$ divided by $2\pi$ and $\mathcal{H}$ is the system's Hamiltonian operator. In the assumption of time-independent potential, time and space variables in Schrödinger equation can be separated yielding:

$$\psi(\vec{r}, t) = u(\vec{r})e^{-i\frac{E}{\hbar}t} \tag{2.3}$$

where $E$ is the energy of the system. In Equation 2.3 $u(\vec{r})$ is the eigenfunction of the Hamiltonian $\mathcal{H}$ with eigenvalue $E$, namely it is a solution of the stationary Schrodinger equation:

$$\mathcal{H}u(\vec{r}) = Eu(\vec{r}). \tag{2.4}$$

By introducing the explicit expression of the Hamiltonian the following equation is obtained:

$$[-\frac{\hbar^2}{2m}\Delta + V(\vec{r})]u(\vec{r}) = Eu(\vec{r}). \tag{2.5}$$

[1] In addition, assumption is made that the mass of the scatterer is much larger than the mass $m$ of the incident particle so that target recoil can be neglected.

where $\Delta$ represents the Laplace operator.

In the asymptotic regime, where $r \to \infty$ and $V \to 0$, the general eigenfunction can be written as:

$$u_{\vec{k}}(\vec{r}) \to \frac{1}{\sqrt{V}}(e^{i\vec{k}\vec{r}} + f_{\vec{k}}(\theta, \phi)\frac{e^{ikr}}{r}). \tag{2.6}$$

Equation 2.6 is a solution of 2.5 in the asymptotic regime with eigenvalue:

$$E = \frac{\hbar^2 k^2}{2m}. \tag{2.7}$$

The particle wavefunction $\psi(\vec{r}, t)$ is related to the particle probability density $\rho(\vec{r}, t)$ by :

$$\rho(\vec{r}, t) = |\psi(\vec{r}, t)|^2 \tag{2.8}$$

and the probability current $\vec{J}(\vec{r}, t)$ is defined as:

$$\vec{J}(\vec{r}, t) = -i\frac{\hbar}{2m}(\psi^*\vec{\nabla}\psi - \psi\vec{\nabla}\psi^*). \tag{2.9}$$

In the case of a plane wave:

$$Ce^{i\vec{k}\vec{r}} \tag{2.10}$$

the probability current equals:

$$\vec{J}(\vec{r}, t) = \rho(\vec{r}, t)\frac{\hbar\vec{k}}{m} = \rho(\vec{r}, t)\vec{v}_g \tag{2.11}$$

where $\vec{v}_g$ is the group velocity.

Expression 2.6 is the sum of two terms. The first:

$$e^{i\vec{k}\vec{r}} \tag{2.12}$$

represents a plane wave with momentum $\vec{p} = \hbar\vec{k}$, this can be considered as the unperturbed wavefunction in absence of any scattering potential. The second term:

$$f_{\vec{k}}(\theta, \phi)\frac{e^{ikr}}{r} \tag{2.13}$$

describes a spherical wave with wavevector of absolute value $k = 2\pi/\lambda$. This term represents the scattered wave. The modulus of the wavevector is not altered in elastic scattering.

The scattering amplitude $f_{\vec{k}}$ modulates the spherical wave amplitude as a function of the angular coordinates $\theta$ and $\phi$.

The incident flux $J_i$ and the scattered flux $J_s$ are related by:

$$J_s = J_i \frac{1}{r^2} \left| f_{\vec{k}}(\theta, \phi) \right|^2 . \tag{2.14}$$

The ratio between the scattered flux within the angular element $d\Omega = \sin\theta d\theta d\phi$ and the incident flux is:

$$d\sigma := r^2 d\Omega \frac{J_s(\theta, \phi)}{J_i}. \tag{2.15}$$

The differential scattering cross section therefore is:

$$\frac{d\sigma}{d\Omega} = \left| f_{\vec{k}}(\theta, \phi) \right|^2 \tag{2.16}$$

and it has the dimensions of an area.

## 2.2 The integral equation of scattering and the Born approximation

Examination of expression 2.6 reveals that the scattering event encodes all information about the sample in the scattering amplitude function $f_{\vec{k}}(\theta, \phi)$. This section presents the derivation of the relationship between the scattering amplitude and the interaction potential $V = V(\vec{r})$ within the range of validity of the Born approximation.

The interaction potential is assumed to be static. The equation to be considered is therefore the stationary Schrödinger equation 2.5, here rewritten for convenience:

$$[-\frac{\hbar^2}{2m}\Delta + V(\vec{r})]u(\vec{r}) = Eu(\vec{r}) \tag{2.17}$$

where the scattering potential acts as a perturbative term. Since the scattering is elastic, $E$ is given by the unperturbed eigenvalue:

$$E = \frac{\hbar^2 k^2}{2m}. \tag{2.18}$$

By replacing this explicit expression of the eigenvalue into the stationary form of Schrödinger equation and by isolating the terms including the interaction potential the following expression is derived:

$$-\frac{\hbar^2}{2m}(\Delta + k^2)u_{\vec{k}}(\vec{r}) = -V(\vec{r})u_{\vec{k}}(\vec{r}). \tag{2.19}$$

The purpose is that of obtaining an expression for the Hamiltonian's eigenfunction $u_{\vec{k}}(\vec{r})$. Before proceeding with the derivation a digression about the classical problem of Poisson's equation in electrostatics is proposed with the aim of providing some insight into the formalism of the mathematical treatment.

### 2.2.1 Digression: Poisson's equation in electrostatics

It it worthwhile interrupting shortly the description of the scattering problem to present a formally equivalent problem in the domain of electrostatics. The electric field $\vec{E}$ generated by a stationary distribution of charges $\rho(\vec{r})$ is given by the following relationship:

$$\vec{E} = -\vec{\nabla}\Phi \tag{2.20}$$

where $\Phi$ is a scalar potential. This can be combined with the differential form of Gauss law that provides an expression for the divergence of the electric field:

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_0} \tag{2.21}$$

where $\epsilon_0$ is the vacuum permittivity, yielding Poisson's equation:

$$\Delta\Phi = -\frac{\rho}{\epsilon_0}. \tag{2.22}$$

In general Poisson's equation can be written as:

$$\Delta u = v \tag{2.23}$$

where $v$ represents the source term. It is worthwhile observing that Poisson's equation is linear *i.e.* the linear combination of solutions is itself a solution. The Green's function $G(\vec{r}, \vec{r}\,')$ is defined as the solution of Poisson's equation in case of a point source placed at position $\vec{r}\,'$ according to:

$$\Delta G(\vec{r}, \vec{r}\,') = \delta(\vec{r} - \vec{r}\,'). \tag{2.24}$$

Any general extended source can be expressed as a weighted sum of point sources:

$$v(\vec{r}) = \int d\vec{r}\,' v(\vec{r}\,')\delta(\vec{r} - \vec{r}\,'). \tag{2.25}$$

By exploiting the linearity of equation 2.23 the general solution is expressed by:

$$u(\vec{r}) = \int d\vec{r}\,' v(\vec{r}\,')G(\vec{r}, \vec{r}\,'). \tag{2.26}$$

Expression 2.26 solves equation 2.23. This can be easily verified by introducing 2.26 into 2.23 and inverting the order of derivation and integration.

In the case of equation 2.22 one particular form of the Green's function is:

$$G(\vec{r}, \vec{r}\,') = -\frac{1}{4\pi}\frac{1}{|\vec{r} - \vec{r}\,'|} \tag{2.27}$$

leading to the well-known form of the electrostatic potential:

$$\Phi(\vec{r}) = \frac{1}{4\pi\epsilon_0}\int d\vec{r}\,' \frac{\rho(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|}. \tag{2.28}$$

### 2.2.2 Derivation of the integral form of Schrödinger equation

It is worthwhile observing that equation 2.19, here repeated for convenience:

$$-\frac{\hbar^2}{2m}(\Delta + k^2)u_{\vec{k}}(\vec{r}) = -V(\vec{r})u_{\vec{k}}(\vec{r}) \tag{2.29}$$

is a linear equation *i.e.* any linear combination of solutions is itsef a solution. In analogy to the procedure presented in Section 2.2.1 a source term $j_{\vec{k}}(\vec{r})$ is defined:

$$j_{\vec{k}}(\vec{r}) = V(\vec{r})u_{\vec{k}}(\vec{r}). \tag{2.30}$$

Equation 2.29 can be written in the form of Poisson's equation:

$$\mathcal{D}_{\vec{r}}u_{\vec{k}}(\vec{r}) = j_{\vec{k}}(\vec{r}) \tag{2.31}$$

by defining the operator:

$$\mathcal{D}_{\vec{r}} := \frac{\hbar^2}{2m}(\Delta + k^2). \tag{2.32}$$

The operator $\mathcal{D}_{\vec{r}}$ is:

$$\mathcal{D}_{\vec{r}} = -(\mathcal{H}_0 - E) \tag{2.33}$$

where $\mathcal{H}_0$ is the free-particle Hamiltonian:

$$\mathcal{H}_0 := -\frac{\hbar^2}{2m}\Delta. \tag{2.34}$$

The Green's function $G(\vec{r}, \vec{r}\,')$ is defined as the solution of equation 2.31 in presence of a point-like source placed at position $\vec{r}\,'$:

$$\mathcal{D}_{\vec{r}} G(\vec{r}, \vec{r}\,') = \delta(\vec{r} - \vec{r}\,'). \tag{2.35}$$

Any extended source $j_{\vec{k}}(\vec{r})$ can be expressed as a superposition of point-like sources:

$$j_{\vec{k}}(\vec{r}) = \int d\vec{r}\,' j_{\vec{k}}(\vec{r}\,')\delta(\vec{r} - \vec{r}\,'). \tag{2.36}$$

The linearity of equation 2.31 can be exploited to obtain the following expression for the solution $u_{\vec{k}}(\vec{r})$:

$$u_{\vec{k}}(\vec{r}) = C(\vec{r}) + \int d\vec{r}\,' G(\vec{r}, \vec{r}\,') j_{\vec{k}}(\vec{r}\,'). \tag{2.37}$$

By introducing the wavefunction given in 2.37 into equation 2.31, by inverting the order of integration and derivation and by using expressions 2.35 and 2.36 the following condition is obtained:

$$\mathcal{D}_{\vec{r}} C(\vec{r}) = 0. \tag{2.38}$$

The wavefunction 2.37 is a solution of equation 2.31 given that condition 2.38 is fulfilled, that is:

$$\frac{\hbar^2}{2m}(\Delta + k^2)C(\vec{r}) = 0. \tag{2.39}$$

Therefore $C$ represents the wavefunction of the unperturbed particle $u^0_{\vec{k}}(\vec{r})$ i.e. the plane wave:

$$e^{i\vec{k}\vec{r}}. \tag{2.40}$$

Equation 2.37 becomes:

$$u_{\vec{k}}(\vec{r}) = u^0_{\vec{k}}(\vec{r}) + \int d\vec{r}\,' G(\vec{r}, \vec{r}\,') V(\vec{r}\,') u_{\vec{k}}(\vec{r}\,'). \tag{2.41}$$

Equation 2.41 is known as the *integral form of the Schrödinger equation*. It is worthwhile stressing that this equation does not provide an explicit expression for the solution of the scattering problem $u_{\vec{k}}(\vec{r})$ since the latter enters both sides of the equation.

### 2.2.3 The expression of the Green's function

The Green's function introduced in equation 2.35 can be expressed as:

$$G(\vec{r},\vec{r}\,') = -\frac{1}{4\pi}\frac{2m}{\hbar^2}\frac{e^{ik|\vec{r}-\vec{r}\,'|}}{|\vec{r}-\vec{r}\,'|}. \tag{2.42}$$

The derivation of this result is reported in [27]. Equation 2.42 represents a spherical wave propagating from $\vec{r}\,'$ to $\vec{r}$. By using this expression the solution of the scattering problem 2.41 can be reformulated as:

$$u_{\vec{k}}(\vec{r}) = u^0_{\vec{k}}(\vec{r}) - \frac{1}{4\pi}\frac{2m}{\hbar^2}\int d\vec{r}\,'\frac{e^{ik|\vec{r}-\vec{r}\,'|}}{|\vec{r}-\vec{r}\,'|}V(\vec{r}\,')u_{\vec{k}}(\vec{r}\,'). \tag{2.43}$$

The physical interpretation of equation 2.43 is straightforward. The first term of the wavefunction $u^0_{\vec{k}}(\vec{r})$ represents the unperturbed solution *i.e.* a plane wave. Viceversa the second term represents the contribution generated by the interaction with the perturbing potential $V(\vec{r})$. This term is expressed by means of the space integral of the Green's function multiplied by the source term. Every point $\vec{r}\,'$ acts as a source of a spherical wave that propagates from $\vec{r}\,'$ to the point $\vec{r}$ where $u$ is computed. The spherical wave is expressed by the Green's function that plays the role of a *propagator*. The amplitude of the spherical wave is proportional to the value of the source term in $\vec{r}\,'$ that is:

$$j_{\vec{k}}(\vec{r}\,') = V(\vec{r}\,')u_{\vec{k}}(\vec{r}\,'). \tag{2.44}$$

In other words a point $\vec{r}\,'$ acts a source of spherical waves if the interaction potential $V(\vec{r}\,')$ and the wavefunction $u_{\vec{k}}(\vec{r}\,')$ simultaneously assume non zero values. This is consistent with the interpretation of the wavefunction as a probability density expressed in equation 2.8.

As mentioned above, the presence of the wavefunction $u_{\vec{k}}(\vec{r}\,')$ within the source term of the integral in equation 2.43 makes the problem recursive. In other words the solution $u_{\vec{k}}(\vec{r})$ enters the source term of the perturbaton (equation 2.30) and is at the same time the result of the pertubation.

### 2.2.4 The Born series and the Born approximation

Equation 2.43 can be expressed in a compact form by defining the Green's operator $\mathcal{G}$ and by expressing the wavefunctions by means of the Dirac notation:

$$\mathcal{G}\left|u_{\vec{k}}\right\rangle = \int d\vec{r}' G(\vec{r},\vec{r}')u_{\vec{k}}(\vec{r}'). \tag{2.45}$$

Equation 2.43 can therefore be rewritten as:

$$\left|u_{\vec{k}}\right\rangle = \left|u^0_{\vec{k}}\right\rangle + \mathcal{G}V\left|u_{\vec{k}}\right\rangle. \tag{2.46}$$

By replacing the expression of $\left|u_{\vec{k}}\right\rangle$ on the right hand side of the equation in an iterative fashion, the following expression of the *Born series* is obtained:

$$\left|u_{\vec{k}}\right\rangle = \sum_{n=0}^{\infty}(\mathcal{G}V)^n\left|u^0_{\vec{k}}\right\rangle. \tag{2.47}$$

This equation expresses the solution of the scattering problem as the sum of infinite terms.

The term of order $n = 0$ represents the unperturbed wavefunction:

$$\left|u^0_{\vec{k}}\right\rangle = e^{i\vec{k}\vec{r}} \tag{2.48}$$

where the normalization factor has been omitted. The first order term $n = 1$ represents the scattered wave resulting from the interaction between the unperturbed wave function and the scattering potential, that is:

$$(\mathcal{G}V)\left|u^0_{\vec{k}}\right\rangle = -\frac{1}{4\pi}\frac{2m}{\hbar^2}\int d\vec{r}'\frac{e^{ik|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|}V(\vec{r}')u^0_{\vec{k}}(\vec{r}'). \tag{2.49}$$

The successive terms represent multiple scattering phenomena. For example the term with order $n = 2$ accounts for the interaction between the scattered wave 2.49 and the scattering potential.

The Born approximation consists in truncating the Born series at the order $n = 1$ *i.e.*:

$$u_{\vec{k}}(\vec{r}) = u^0_{\vec{k}}(\vec{r}) - \frac{1}{4\pi}\frac{2m}{\hbar^2}\int d\vec{r}'\frac{e^{ik|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|}V(\vec{r}')u^0_{\vec{k}}(\vec{r}'). \tag{2.50}$$

In the asymptotic regime $r \to \infty$ the factor $\frac{1}{|\vec{r}-\vec{r}'|}$ approaches $\frac{1}{r}$ and the following expression holds:

$$u_{\vec{k}}(\vec{r}) = u^0_{\vec{k}}(\vec{r}) - \frac{1}{4\pi}\frac{2m}{\hbar^2}\frac{e^{ikr}}{r}\int d\vec{r}' e^{-ik\hat{r}\cdot\vec{r}'}V(\vec{r}')u^0_{\vec{k}}(\vec{r}'). \tag{2.51}$$

By replacing:

$$|u_{\vec{k}}^{0}\rangle = \frac{1}{(2\pi)^{3/2}} e^{i\vec{k}\vec{r}} \tag{2.52}$$

and with definitions:

$$\vec{k} := \vec{k}_i \qquad k\hat{r} := \vec{k}_f \qquad \vec{k}_i - \vec{k}_f := \vec{Q} \tag{2.53}$$

the equation of the Born approximated wavefunction in the asymptotic regime reads:

$$u_{\vec{k}}(\vec{r}) = u_{\vec{k}}^{0}(\vec{r}) - \frac{1}{4\pi}\frac{2m}{\hbar^2}\frac{e^{ikr}}{r}\frac{1}{(2\pi)^{3/2}}\int d\vec{r}\,' V(\vec{r}\,') e^{i\vec{Q}\vec{r}\,'}. \tag{2.54}$$

By comparison with expression 2.6, also valid in the asymptotic regime, and here rewritten for convenience:

$$u_{\vec{k}}(\vec{r}) \to \frac{1}{\sqrt{V}}(e^{i\vec{k}\vec{r}} + f_{\vec{k}}(\theta,\phi)\frac{e^{ikr}}{r}) \tag{2.55}$$

an expression of the scattering amplitude is derived:

$$f_{\vec{Q}}(\Omega) = -\frac{1}{2\pi}\frac{m}{\hbar^2}\int d\vec{r}\,' V(\vec{r}\,') e^{i\vec{Q}\vec{r}\,'}. \tag{2.56}$$

Within the conditions of validity of the Born approximation the scattering amplitude is proportional to the Fourier transform of the scattering potential.

In the Born approximation multiple scattering phenomena are not accounted for. A condition for the validity of this approximation is derived by imposing the rapid convergence of the Born series 2.47. In this case the modulus of first order term is much smaller than the modulus of the term of order zero. By carrying out the calculations [27] for the condition of validity of the Born approximation it is observed that this approach is not suitable in the case of the scattering of low energy neutrons on the nuclear potential of a single nucleus. Such phenomenon is treated in a more appropriate fashion by the partial waves description developed in [27]. Nevertheless the vast majority of neutron scattering experiments are interpreted in the Born approximation. This is possible since condensed matter experiments deal with ensembles of nuclei.

It is worthwhile observing that in the case of scattering of X-rays the problem must be approached in the second quantization framework as described in reference [30]. In this picture the Hamiltonian describing the interaction between the photon and the electrons employed in the calculation of the system transition rate is expressed in the second

quantization formalism *i.e.* by means of creation and annihilation operators for both electrons and photons. The scattering amplitude is expressed by an equation analogous to expression 2.56 as the Fourier transform of the electron density of the scatterer.

## 2.3 The general case of inelastic scattering

The scattering problem can be formulated in a general way as the interaction between probe particles and a target possessing internal degrees of freedom. Consideration of the Schrödinger equation of the combined system including both probe and target states leads to the quantum mechanical description of inelastic scattering. The formalism that is developed in this case is analogous to that of Section 2.2, the main difference being that in the general scenario of inelastic scattering the wavefunctions describing target states must be considered and this makes the derivation more intricate.

In this framework Fermi's Golden Rule is derived: this is the probability density of a transition of the coupled probe-particle system. The derivation is carried out within the stationary Schrödinger equation formalism. An alternative derivation could be given by means of the time dependent perturbation theory [31]. Nevertheless the stationary formalism can be adopted in the assumption that the potential is switched on slowly with respect to the relevant time scales and that the system can be considered quasi-stationary at any time. This corresponds to considering a transition related to the time evolution of the system from $t \to -\infty$ to $t \to +\infty$.

Central results of scattering theory such as the master equation of scattering and the expression of the dynamic structure factor are derived in the final part of the present section.

### 2.3.1 The combined probe-target system

In the general scenario assumption is made that a probe particle of mass $m$ is scattered by a target $T$ whose mass is largely superior to that of the probe particle. Assumption is made that the target is a complex system with internal degrees of freedom and that it

can undergo transitions between quantum states.

Be $\{ \xi \}$ the ensemble of coordinates characterizing the target system and $\vec{r}$ the probe particles coordinates as in Section 2.2. The stationary Schrödinger equation of the combined system is considered:

$$[\mathcal{H}_T(\xi) - \frac{\hbar^2}{2m}\Delta_{\vec{r}} + V(\vec{r},\xi)]\psi(\vec{r},\xi) = E_{\text{total}}\psi(\vec{r},\xi). \tag{2.57}$$

In this equation $\mathcal{H}_T(\xi)$ is the target Hamiltonian operator, $V(\vec{r},\xi)$ describes the probe-target interaction potential, $\psi(\vec{r},\xi)$ is the wavefunction of the combined system and $E_{\text{total}}$ denotes the total energy of the combined system. In analogy to the derivation in Section 2.2 the term depending on the interaction potential can be isolated yielding:

$$[E_{\text{total}} - \mathcal{H}_T(\xi) + \frac{\hbar^2}{2m}\Delta_{\vec{r}}]\psi(\vec{r},\xi) = V(\vec{r},\xi)\psi(\vec{r},\xi). \tag{2.58}$$

Assumption is made that the solutions of the stationary Schrödinger equation for the isolated target system are the eigenfunction $\phi_\nu(\xi)$ with eigenvalues $E_\nu$ where $\nu$ denotes the quantum numbers characterizing the target state:

$$\mathcal{H}_T(\xi)\phi_\nu(\xi) = E_\nu\phi_\nu(\xi). \tag{2.59}$$

The direct product states:

$$|\Theta_{\nu,\vec{k}}\rangle = |\phi_\nu\rangle \otimes |\vec{k}\rangle \tag{2.60}$$

where $|\vec{k}\rangle$ denotes the plane wave with wavevector $\vec{k}$, form a basis of the Hilbert space of the combined system. In real space representation [2] these basis states are expressed as:

$$\langle\vec{r},\xi|\Theta_{\nu,\vec{k}}\rangle = \Theta_{\nu,\vec{k}}(\vec{r},\xi) = \frac{1}{(2\pi)^{3/2}}e^{i\vec{k}\vec{r}}\phi_\nu(\xi). \tag{2.61}$$

They are eigenfunctions of the non-interacting probe-target system:

$$[\mathcal{H}_T(\xi) - \frac{\hbar^2}{2m}\Delta_{\vec{r}}]\Theta_{\nu,\vec{k}}(\vec{r},\xi) = \mathcal{H}_0\Theta_{\nu,\vec{k}}(\vec{r},\xi) = E_0\Theta_{\nu,\vec{k}}(\vec{r},\xi) \tag{2.62}$$

[2] In real space representation the eigenstates $|x\rangle$ of the position operator are assumed as a basis of the Hilbert space. They fulfill the following relations:

$\hat{x}\,|x\rangle = x\,|x\rangle\,,$

$\langle x|\psi\rangle = \int dx'\psi(x')\delta(x-x') = \psi(x)$

$\langle x|x'\rangle = \int dx''\delta(x-x'')\delta(x'-x'') = \delta(x-x')$

where the total energy of the system in the non-interacting case is:

$$E_0 = E_\nu + \frac{\hbar^2 k^2}{2m}. \tag{2.63}$$

The wavefunctions $|\Theta_{\nu,\vec{k}}\rangle$ are normalized and they fulfill orthogonality and closure relations.

In presence of a scattering potential with limited space extension the total energy of the combined system $E_{\text{total}}$ equals the energy of the uncoupled combined system $E_0$ before the scattering event.

### 2.3.2 The Green's function in the combined system framework

In analogy with the derivation presented in Section 2.2 relative to elastic scattering, equation 2.58 can be written by defining a source term for the combined system:

$$j(\vec{r},\xi) := V(\vec{r},\xi)\psi(\vec{r},\xi) \tag{2.64}$$

yielding:

$$\mathcal{D}\psi(\vec{r},\xi) = j(\vec{r},\xi). \tag{2.65}$$

The operator $\mathcal{D}$ is defined as:

$$\mathcal{D} := [E_{\text{total}} - \mathcal{H}_T(\xi) + \frac{\hbar^2}{2m}\Delta_{\vec{r}}] \tag{2.66}$$

that is:

$$\mathcal{D} := -(\mathcal{H}_0 - E_{\text{total}}) \tag{2.67}$$

where $\mathcal{H}_0$ is the unperturbed Hamiltonian $\mathcal{H}_T(\xi) - \frac{\hbar^2}{2m}\Delta_{\vec{r}}$. The total energy $E_{\text{total}}$ equals the energy of the non interacting system before the scattering event:

$$E_{\text{total}} = E_{\nu,i} + \frac{\hbar^2 k_i^2}{2m} \tag{2.68}$$

where $E_{\nu,i}$ is the initial energy of the target and $\frac{\hbar^2 k_i^2}{2m}$ is the initial energy of the probe particle.

The Green's function is now defined as the solution of equation 2.65 in presence of a point-like source term:

$$\mathcal{D}G(\vec{r},\vec{r}';\xi,\xi') = \delta(\vec{r}-\vec{r}')\cdot\delta(\xi-\xi'). \tag{2.69}$$

The Green operator $\mathcal{G}$ can formally be expressed as the inverse of $(E_{\text{total}} - \mathcal{H}_0)$:

$$\mathcal{G} := \frac{1}{E_{\text{total}} - \mathcal{H}_0}. \tag{2.70}$$

The Green function can be calculated as:

$$G(\vec{r},\vec{r}';\xi,\xi') = \langle \vec{r},\xi|\mathcal{G}|\vec{r}',\xi'\rangle\,. \tag{2.71}$$

The full derivation is presented in reference [27]. The final result is:

$$G(\vec{r},\vec{r}';\xi,\xi') = -\frac{m}{2\pi\hbar^2}\sum_{\nu}\phi_\nu(\xi)\phi_\nu^*(\xi')\frac{e^{ik_f|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} \tag{2.72}$$

where the sum runs over the target states $\nu$ and $k_f$ depends on the isolated target energy $E_\nu$ according to:

$$E_{\nu,i} + \frac{\hbar^2 k_i^2}{2m} = E_\nu + \frac{\hbar^2 k_f^2}{2m}. \tag{2.73}$$

Analogous to the elastic case 2.42 the Green's function 2.72 is expressed as a linear combination of spherical waves. Each sperical wave is characterized by a wavevector $k_f$ given by 2.73 and represents the scattered probe particle corresponding to the final target state $\phi_\nu(\xi)$. The relative weight of each spherical wave in the linear combination is given by $\phi_\nu(\xi)\phi_\nu^*(\xi')$.

In the basis of the unperturbed system wavefunctions $|\Theta_{\mu,\vec{k}}\rangle$ the integral form of the Schrödinger equation is written as:

$$\psi_{\mu,\vec{k}}(\vec{r},\xi) = \Theta_{\mu,\vec{k}}(\vec{r},\xi) + \int d\vec{r}' \int d\xi' G(\vec{r},\vec{r}';\xi,\xi')V(\vec{r}',\xi')\psi_{\mu,\vec{k}}(\vec{r}',\xi') \tag{2.74}$$

where $\psi_{\mu,\vec{k}}(\vec{r},\xi)$ is the solution of 2.65 evolving from the unperturbed system solution $\Theta_{\mu,\vec{k}}(\vec{r},\xi)$ as the interaction $V(\vec{r},\xi)$ is introduced. Analogous to equation 2.41 in Section 2.2, this is a recursive relation since the solution $\psi_{\mu,\vec{k}}(\vec{r},\xi)$ appears on both sides of the equation.

### 2.3.3 The Born approximation

In the Born approximation multiple scattering events are not accounted for and assumption is made that the interaction potential $V(\vec{r},\xi)$ acts directly on the unperturbed wavefunction $\Theta_{\mu,\vec{k}}(\vec{r},\xi)$. The exact solution $\psi_{\mu,\vec{k}}(\vec{r},\xi)$ is replaced by the unperturbed function $\Theta_{\mu,\vec{k}}(\vec{r},\xi)$ in the right hand side of equation 2.74:

$$\psi_{\mu,\vec{k}}(\vec{r},\xi) = \Theta_{\mu,\vec{k}}(\vec{r},\xi) + \int d\vec{r}' \int d\xi' G(\vec{r},\vec{r}';\xi,\xi') V(\vec{r}',\xi') \Theta_{\mu,\vec{k}}(\vec{r}',\xi') \tag{2.75}$$

that is:

$$\begin{aligned} &\psi_{\mu,\vec{k}}(\vec{r},\xi) = \\ &\Theta_{\mu,\vec{k}}(\vec{r},\xi) - \frac{m}{2\pi\hbar^2} \sum_{\nu} \phi_\nu(\xi) \int d\vec{r}' \int d\xi' \phi_\nu^*(\xi') \frac{e^{ik_f|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} V(\vec{r}',\xi') \Theta_{\mu,\vec{k}}(\vec{r}',\xi'). \end{aligned} \tag{2.76}$$

By writing the explicit function describing the unperturbed probe particles and by considering the asymptotic limit $r \to \infty$ where the factor $\frac{1}{|\vec{r}-\vec{r}'|}$ approaches $\frac{1}{r}$ the following expression is obtained:

$$\begin{aligned} &\psi_{\mu,\vec{k}}(\vec{r},\xi) = \Theta_{\mu,\vec{k}}(\vec{r},\xi) + \\ &- \frac{m}{(2\pi)^{5/2}\hbar^2} \sum_{\nu} \phi_\nu(\xi) \frac{e^{ik_f r}}{r} \int d\vec{r}' \int d\xi' \phi_\nu^*(\xi') e^{-ik_f|\hat{r}\cdot\vec{r}'|} V(\vec{r}',\xi') e^{i\vec{k}\vec{r}'} \phi_\mu(\xi'). \end{aligned} \tag{2.77}$$

By defining the transferred wavevector:

$$\vec{Q} := \vec{k} - \vec{k}_f \tag{2.78}$$

the following equation is obtained:

$$\begin{aligned} &\psi_{\mu,\vec{k}}(\vec{r},\xi) = \Theta_{\mu,\vec{k}}(\vec{r},\xi) + \\ &- \frac{m}{2\pi\hbar^2} \frac{1}{(2\pi)^{3/2}} \sum_{\nu} \phi_\nu(\xi) \frac{e^{ik_f r}}{r} \int d\vec{r}' \int d\xi' \phi_\nu^*(\xi') e^{i\vec{Q}\vec{r}'} V(\vec{r}',\xi') \phi_\mu(\xi') \end{aligned} \tag{2.79}$$

where the integral can be formulated in Dirac notation:

$$\langle \phi_\nu, \vec{k}_f | V | \phi_\mu, \vec{k} \rangle \propto \int d\vec{r}' \int d\xi' \phi_\nu^*(\xi') e^{i\vec{Q}\vec{r}'} V(\vec{r}',\xi') \phi_\mu(\xi'). \tag{2.80}$$

Equation 2.79 shows that the scattering event can be interpreted in terms of a wavefunction composed of individual scattering channels that do not interfere with each other. Given an initial probe state $|\vec{k}\rangle$ and an initial target state $|\mu\rangle$ the scattered function is given by the sum over the scattering channels $\nu$ of spherical waves $\frac{e^{ik_f r}}{r}$. Each scattering channel corresponds to a final target state $\phi_\nu(\xi)$ and a final probe state characterized by a wavevector $\vec{k}_f$ that complies with the conservation of energy. In the sum appearing in 2.79 each final state $\phi_\nu(\xi)\frac{e^{ik_f r}}{r}$ is weighted by a factor proportional to the integral $\langle\phi_\nu, \vec{k}_f|V|\phi_\mu, \vec{k}\rangle$.

### 2.3.4 Fermi's Golden rule

The probability density of the transition related to an individual scattering channel is an expression of wide use and it is known as *Fermi's Golden Rule*. Fermi's Golden Rule is derived by calculating the probe particle flux associated to an individual scattering channel in 2.79. The probe particle flux of a single scattering channel is related to the differential cross section of the channel which in turn can be used to derive the expression of the scattering amplitude and the probability density of the transition related the single channel. The derivation of Fermi's Golden Rule is reported hereafter.

It is possible to isolate the individual scattering channels by using operators of the type:

$$P_{\phi_{\mu_f}} = |\phi_{\mu_f}\rangle \langle\phi_{\mu_f}| \otimes \mathcal{I}_{\vec{k}} \tag{2.81}$$

where $P_{\phi_{\mu_f}}$ is the projection operator over one particular final target state $|\phi_{\mu_f}\rangle$ and $\mathcal{I}_{\vec{k}}$ is the unit operator in the subspace of particle states and it is given by:

$$\mathcal{I}_{\vec{k}} = \sum_{\vec{k}} |\vec{k}\rangle \langle\vec{k}| . \tag{2.82}$$

The projectors can be used to derive the projection of $\psi_{\mu,\vec{k}}$ over each final target state $\phi_{\mu_f}$ by calculating:

$$\psi_{\mu\to\mu_f,\vec{k}\to\vec{k}_f} = \langle\vec{r}, \xi|P_{\phi_{\mu_f}}|\psi_{\mu,\vec{k}}\rangle \tag{2.83}$$

where the expression of $\psi_{\mu,\vec{k}}$ obtained within the Born approximation is used. The probe

particle flux related with each scattering channel $\mu_f$ is given by:

$$\vec{J}_{\mu\to\mu_f,\vec{k}\to\vec{k}_f} = \\ -i\frac{\hbar}{2m}\int d\xi(\psi^*_{\mu\to\mu_f,\vec{k}\to\vec{k}_f}\vec{\nabla}_{\vec{r}}\,\psi_{\mu\to\mu_f,\vec{k}\to\vec{k}_f} - \psi_{\mu\to\mu_f,\vec{k}\to\vec{k}_f}\vec{\nabla}_{\vec{r}}\,\psi^*_{\mu\to\mu_f,\vec{k}\to\vec{k}_f}). \tag{2.84}$$

It is worthwhile observing that in this equation what is calculated is the flux of the probe particles. Therefore the gradient $\vec{\nabla}_{\vec{r}}$ is calculated with respect to the probe particle coordinates $\vec{r}$ and integration over the target coordinates $\xi$ is performed.

By carrying out the calculations the differential cross section:

$$\frac{d\sigma}{d\Omega} := r^2\frac{J_{\text{scattered}}(r,\Omega)}{J_{\text{incident}}} \tag{2.85}$$

can be calculated yielding for each scattering channel [3]:

$$\left(\frac{d\sigma}{d\Omega}\right)_{\mu\to\mu_f,\vec{k}\to\vec{k}_f} = m^2\left(\frac{2\pi}{\hbar}\right)^4\frac{k_f}{k}\left|\langle\phi_{\mu_f},\vec{k}_f|V|\phi_\mu,\vec{k}\rangle\right|^2. \tag{2.86}$$

In the elastic case $\mu_f = \mu$ and equation 2.86 becomes:

$$\left(\frac{d\sigma}{d\Omega}\right)_{\mu\to\mu,\vec{k}\to\vec{k}_f} = m^2\left(\frac{2\pi}{\hbar}\right)^4\left|\langle\phi_{\mu},\vec{k}_f|V|\phi_\mu,\vec{k}\rangle\right|^2. \tag{2.87}$$

The scattering amplitude is (Equation 2.16):

$$f_{\mu\to\mu_f,\vec{k}\to\vec{k}_f}(\Omega) = -\frac{4\pi^2 m}{\hbar^2}\langle\phi_{\mu_f},\vec{k}_f|V|\phi_\mu,\vec{k}\rangle. \tag{2.88}$$

The differential cross section can be interpreted in terms of the transition probability density $w_{\mu\to\mu_f,\vec{k}\to\vec{k}_f}$:

$$d\sigma_{\mu\to\mu_f,\vec{k}\to\vec{k}_f} = \frac{w_{\mu\to\mu_f,\vec{k}\to\vec{k}_f}}{J_{\text{incident}}}d\Omega. \tag{2.89}$$

Equation 2.86 yields the following expression for the probability density of a transition which is known as Fermi's Golden Rule:

$$w_{\mu\to\mu_f,\vec{k}\to\vec{k}_f} = \frac{2\pi}{\hbar}\left|\langle\phi_{\mu_f},\vec{k}_f|V|\phi_\mu,\vec{k}\rangle\right|^2\rho(k_f) \tag{2.90}$$

where the density of final states is:

$$\rho(k_f) = \frac{mk_f}{\hbar^2}. \tag{2.91}$$

[3] The incident flux is $J_{\text{incident}} = \frac{1}{(2\pi)^3}\frac{\hbar k}{m}$.

Fermi's Golden Rule expresses the fact that given a transition of a probe-target system $\mu \to \mu_f, \vec{k} \to \vec{k}_f$ induced by the interaction potential $V$, the probability of such transition is directly proportional to the square of the matrix element:

$$\left|\langle \phi_{\mu_f}, \vec{k}_f | V | \phi_\mu, \vec{k} \rangle\right|^2 = \left| \int d\vec{r}\, e^{i\vec{Q}\vec{r}} \int d\xi\ \phi^*_{\mu_f}(\xi) V(\vec{r}, \xi) \phi_\mu(\xi) \right|^2 . \tag{2.92}$$

An experimental measurement corresponds to an integration over a set of final states where the integration space is given by solid angle and energy of the scattered particle.

### 2.3.5 The master equation of scattering

The *master equation of scattering* is an expression of the double differential cross section related to the transition that leads the probe particle from the state $|\vec{k}_i\rangle$ to the state $|\vec{k}_f\rangle$. This expression is of wide use in scattering experiments since all possible target states - which are generally not individually accessible in an experiment - are considered simultaneously.

The initial and final states of the probe-target system can be expressed by the direct products:

$$|i\rangle = |\vec{k}_i\rangle \otimes |\lambda_i\rangle = |\vec{k}_i, \lambda_i\rangle \tag{2.93}$$

$$|f\rangle = |\vec{k}_f\rangle \otimes |\lambda_f\rangle = |\vec{k}_f, \lambda_f\rangle \tag{2.94}$$

where $|\lambda\rangle$ denotes the target state characterized by the set of target quantum numbers $\lambda$. With this notation equation 2.86 can be formulated as follows:

$$\left(\frac{d\sigma}{d\Omega}\right)_{\lambda_i \to \lambda_f, \vec{k}_i \to \vec{k}_f} = m^2 \left(\frac{2\pi}{\hbar}\right)^4 \frac{k_f}{k_i} \left|\langle \vec{k}_f, \lambda_f | V | \vec{k}_i, \lambda_i \rangle\right|^2 . \tag{2.95}$$

The equation expressing the double differential scattering cross section is obtained by introducing in 2.95 the condition for the conservation of energy $E_f + E_{\lambda_f} = E_i + E_{\lambda_i}$ by using the factor $\delta(E_f + E_{\lambda_f} - E_i - E_{\lambda_i})$ where $E_i$ and $E_f$ are the initial and final energy of the probe particle while $E_{\lambda_i}$ and $E_{\lambda_f}$ denote the initial and final energy of the target:

$$\left(\frac{d^2\sigma}{d\Omega dE_f}\right)_{\lambda_i \to \lambda_f, \vec{k}_i \to \vec{k}_f} = \tag{2.96}$$

$$m^2 \left(\frac{2\pi}{\hbar}\right)^4 \frac{k_f}{k_i} \left|\langle \vec{k}_f, \lambda_f | V | \vec{k}_i, \lambda_i \rangle\right|^2 \delta(E_f + E_{\lambda_f} - E_i - E_{\lambda_i}).$$

In an experimental context the probe particle states $|\vec{k}_i\rangle$ and $|\vec{k}_f\rangle$ are generally known. The target is found initially in a superposition of states $|\lambda_i\rangle$, each with probability $p(\lambda_i)$. The target state $|\lambda_f\rangle$ is generally not detected in experiments. It is hence convenient to consider all final target states compatible with the probe particle transition of interest *i.e.* from state $|\vec{k}_i\rangle$ to state $|\vec{k}_f\rangle$. By summing over all target final states and by introducing the statistical distribution of the target initial states $p(\lambda_i)$ the master equation of scattering is obtained:

$$\left(\frac{d^2\sigma}{d\Omega dE_f}\right)_{\vec{k}_i\to\vec{k}_f} = \tag{2.97}$$
$$m^2\left(\frac{2\pi}{\hbar}\right)^4\frac{k_f}{k_i}\sum_{\lambda_i,\lambda_f}p(\lambda_i)\left|\langle\vec{k}_f,\lambda_f|V|\vec{k}_i,\lambda_i\rangle\right|^2\delta(E_f+E_{\lambda_f}-E_i-E_{\lambda_i}).$$

It is worthwhile recalling that the formalism developed so far is based on the Born approximation.

### 2.3.6 The interaction potential

The interaction potential can be formulated as the sum over $N$ individual scatterers $j$ of two-particle interactions:

$$V(\vec{r}) = \sum_{j=1}^{N} V_j(\vec{r}-\vec{R}_j) \tag{2.98}$$

where $\vec{R}_j$ denotes the positions of the individual scatterers. By replacing this expression in the matrix element:

$$\langle\vec{k}_f,\lambda_f|V|\vec{k}_i,\lambda_i\rangle \tag{2.99}$$

and by carrying out the calculations the following expression is obtained:

$$\langle\vec{k}_f,\lambda_f|V|\vec{k}_i,\lambda_i\rangle = \frac{1}{(2\pi)^3}\sum_{j=1}^{N}V_j(\vec{Q})\,\langle\lambda_f|e^{i\vec{Q}\cdot\vec{R}_j}|\lambda_i\rangle \tag{2.100}$$

where $V_j(\vec{Q})$ is the Fourier transform of the interaction potential of the $j^{\text{th}}$ scatterer according to:

$$V_j(\vec{Q}) = \int d\vec{r}\, e^{i\vec{Q}\cdot\vec{r}}V_j(\vec{r}). \tag{2.101}$$

$V_j(\vec{Q})$ is known as the *form factor* of the interaction potential $V_j(\vec{r})$.

In neutron scattering a point-like form of the interaction potential is adopted and this is known as *Fermi pseudo-potential*:

$$V_j(\vec{r}) = \frac{2\pi\hbar^2}{m} b_j \delta(\vec{r}) \tag{2.102}$$

yielding:

$$V_j(\vec{Q}) = \frac{2\pi\hbar^2}{m} b_j. \tag{2.103}$$

The double differential scattering cross section is reformulated as:

$$\left(\frac{d^2\sigma}{d\Omega dE_f}\right)_{\lambda_i\to\lambda_f,\vec{k}_i\to\vec{k}_f} = \frac{k_f}{k_i}\left|\sum_{j=1}^{N} b_j \langle\lambda_f|e^{i\vec{Q}\cdot\vec{R}_j}|\lambda_i\rangle\right|^2 \delta(E_{\lambda_i} - E_{\lambda_f} + \hbar\omega). \tag{2.104}$$

where:

$$\hbar\omega := E_i - E_f \tag{2.105}$$

is the energy transferred from the probe to the target.

### 2.3.7 The dynamic structure factor

The master equation of scattering is obtained using as a starting point Equation 2.104, summing over initial states considering their statistical distribution $p(\lambda_i)$ and by summing over final states compatible with the conservation of energy.

$$\left(\frac{d^2\sigma}{d\Omega dE_f}\right)_{\vec{k}_i\to\vec{k}_f} = \frac{k_f}{k_i}\sum_{\lambda_i,\lambda_f} p(\lambda_i)\left|\sum_{j=1}^{N} b_j \langle\lambda_f|e^{i\vec{Q}\cdot\vec{R}_j}|\lambda_i\rangle\right|^2 \delta(E_{\lambda_i} - E_{\lambda_f} + \hbar\omega). \tag{2.106}$$

A $\delta$-function in energy can be expressed as an integral over time:

$$\delta(E_{\lambda_i} - E_{\lambda_f} + \hbar\omega) = \frac{1}{2\pi\hbar}\int_{-\infty}^{+\infty} dt\; e^{-\frac{i}{\hbar}(E_{\lambda_i} - E_{\lambda_f} + \hbar\omega)t}. \tag{2.107}$$

Be $\mathcal{H}_T$ the target Hamiltonian:

$$\mathcal{H}_T|\lambda\rangle = E_\lambda|\lambda\rangle \tag{2.108}$$

the following expression holds:

$$e^{\frac{i}{\hbar}\mathcal{H}_T t}|\lambda\rangle = e^{\frac{i}{\hbar}E_\lambda t}|\lambda\rangle. \tag{2.109}$$

The sum over final states appearing in the master equation can be formulated by writing the explicit form of the modulus squared and by introducing the integral form of the Dirac delta:

$$\sum_{\lambda_f} \left| \sum_{j=1}^{N} b_j \langle \lambda_f | e^{i\vec{Q}\cdot\vec{R}_j} | \lambda_i \rangle \right|^2 \delta(E_{\lambda_i} - E_{\lambda_f} + \hbar\omega) \tag{2.110}$$

$$= \sum_{\lambda_f} \sum_{j,j'=1}^{N} b_j b_{j'}^* \langle \lambda_f | e^{i\vec{Q}\cdot\vec{R}_j} | \lambda_i \rangle \langle \lambda_i | e^{-i\vec{Q}\cdot\vec{R}_{j'}} | \lambda_f \rangle \delta(E_{\lambda_i} - E_{\lambda_f} + \hbar\omega)$$

$$= \frac{1}{2\pi\hbar} \int dt \sum_{\lambda_f} \sum_{j,j'=1}^{N} b_j b_{j'}^* \langle \lambda_f | e^{\frac{i}{\hbar}E_{\lambda_f}t} e^{i\vec{Q}\cdot\vec{R}_j} e^{-\frac{i}{\hbar}E_{\lambda_i}t} | \lambda_i \rangle \langle \lambda_i | e^{-i\vec{Q}\cdot\vec{R}_{j'}} | \lambda_f \rangle e^{-i\omega t}.$$

The energy can be replaced by the Hamilton operator:

$$\sum_{\lambda_f} \left| \sum_{j=1}^{N} b_j \langle \lambda_f | e^{i\vec{Q}\cdot\vec{R}_j} | \lambda_i \rangle \right|^2 \delta(E_{\lambda_i} - E_{\lambda_f} + \hbar\omega) \tag{2.111}$$

$$= \frac{1}{2\pi\hbar} \int dt \sum_{\lambda_f} \sum_{j,j'=1}^{N} b_j b_{j'}^* \langle \lambda_i | e^{-i\vec{Q}\cdot\vec{R}_{j'}} | \lambda_f \rangle \langle \lambda_f | e^{\frac{i}{\hbar}\mathcal{H}_T t} e^{i\vec{Q}\cdot\vec{R}_j} e^{-\frac{i}{\hbar}\mathcal{H}_T t} | \lambda_i \rangle e^{-i\omega t}$$

$$= \frac{1}{2\pi\hbar} \int dt \sum_{j,j'=1}^{N} b_j b_{j'}^* \langle \lambda_i | e^{-i\vec{Q}\cdot\vec{R}_{j'}} e^{\frac{i}{\hbar}\mathcal{H}_T t} e^{i\vec{Q}\cdot\vec{R}_j} e^{-\frac{i}{\hbar}\mathcal{H}_T t} | \lambda_i \rangle e^{-i\omega t}$$

where the closure relation for the $|\lambda\rangle$ states has been used:

$$\sum_{\lambda_f} |\lambda_f\rangle \langle \lambda_f| = \mathcal{I}. \tag{2.112}$$

By definition of time evolution operator it follows:

$$e^{\frac{i}{\hbar}\mathcal{H}_T t} e^{i\vec{Q}\cdot\vec{R}_j} e^{-\frac{i}{\hbar}\mathcal{H}_T t} = e^{i\vec{Q}\cdot\vec{R}_j(t)} \tag{2.113}$$

and:

$$\sum_{\lambda_f} \left| \sum_{j=1}^{N} b_j \langle \lambda_f | e^{i\vec{Q}\cdot\vec{R}_j} | \lambda_i \rangle \right|^2 \delta(E_{\lambda_i} - E_{\lambda_f} + \hbar\omega) \tag{2.114}$$

$$= \frac{1}{2\pi\hbar} \int dt \sum_{j,j'=1}^{N} b_j b_{j'}^* \langle \lambda_i | e^{-i\vec{Q}\cdot\vec{R}_{j'}} e^{i\vec{Q}\cdot\vec{R}_j(t)} | \lambda_i \rangle e^{-i\omega t}.$$

The double differential cross section is:

$$\left(\frac{d^2\sigma}{d\Omega dE_f}\right)_{\vec{k}_i\to\vec{k}_f} = \tag{2.115}$$
$$\frac{k_f}{k_i}\frac{1}{2\pi\hbar}\sum_{\lambda_i} p(\lambda_i)\sum_{j,j'=1}^{N} b_j b_{j'}^{*}\int dt\,\langle\lambda_i|e^{-i\vec{Q}\cdot\vec{R}_{j'}}e^{i\vec{Q}\cdot\vec{R}_j(t)}|\lambda_i\rangle\, e^{-i\omega t}.$$

The *scattering function* or *dynamic structure factor* is defined as:

$$\left(\frac{d^2\sigma}{d\Omega dE_f}\right)_{\vec{k}_i\to\vec{k}_f} = \frac{k_f}{k_i}S(\vec{Q},\omega) \tag{2.116}$$

and its expression is:

$$S(\vec{Q},\omega) = \frac{1}{2\pi\hbar}\sum_{\lambda_i} p(\lambda_i)\sum_{j,j'=1}^{N} b_j b_{j'}^{*}\int dt\,\langle\lambda_i|e^{-i\vec{Q}\cdot\vec{R}_{j'}}e^{i\vec{Q}\cdot\vec{R}_j(t)}|\lambda_i\rangle\, e^{-i\omega t} \tag{2.117}$$

where:

$$\vec{R}_{j'} = \vec{R}_{j'}(t=0). \tag{2.118}$$

Equation 2.117 does not present a sum over final target states, instead the time dependence is explicit in the integral term.

It is worth recalling that this expression is valid within the Born approximation. In this approximation multiple scattering effects are neglected and assumption is made that every scatterer is exposed to the incident and unattenuated flux. In the Born approximation framework the scattering function 2.117 turns out to be proportional to the number of scatterers. According to expression 2.117, for a given incident flux the scattered flux increases without bounds with increasing number of scatterers. This unphysical result is a shortcoming of the introduced approximation. The lack of proportionality between the scattered flux and the sample size at high values of the scattered flux is known as the extinction effect. This is due to the fact that the assumption underlying the Born approximation namely the homogeneity of the incident beam over the sample volume doesn't hold in these circumstances. The extinction effect is accounted for in the formalism of the dynamical scattering theory [32].

## 2.4 Coherent and incoherent scattering

The present section describes the phenomena of coherent and incoherent scattering. The approach is that of separating coherent and incoherent terms in the expression of the dynamic structure factor and to derive the corresponding double differential scattering cross sections. The role of nuclear spin in incoherent scattering is discussed. Particular emphasis is put on the description of scattering by protons and deuterons and on the experimental requirement for the production of deuterated and perdeuterated samples in macromolecular neutron crystallography.

### 2.4.1 Coherent and incoherent terms within the dynamic structure factor

The scattering function 2.117 can be reformulated as follows:

$$S(\vec{Q},\omega) = \sum_{j,j'=1}^{N} b_j b_{j'}^{*}\, S_{j,j'}(\vec{Q},\omega) \tag{2.119}$$

where:

$$S_{j,j'}(\vec{Q},\omega) = \frac{1}{2\pi\hbar}\sum_{\lambda_i} p(\lambda_i) \int dt\, \langle\lambda_i|e^{-i\vec{Q}\cdot\vec{R}_{j'}}\, e^{i\vec{Q}\cdot\vec{R}_j(t)}|\lambda_i\rangle\, e^{-i\omega t}. \tag{2.120}$$

The neutron scattering length $b$ of a nucleus depends on the isotope and on the spin state of the combined neutron - nucleus system. Isotopes and nuclear spins of a chemical element are in general distributed in a statistical manner over the atomic sites occupied by the element of interest. Consequently expression 2.119 can be formulated as:

$$S(\vec{Q},\omega) = \sum_{j,j'=1}^{N} \overline{b_j b_{j'}^{*}}\, S_{j,j'}(\vec{Q},\omega) \tag{2.121}$$

where $\overline{b_j b_{j'}^{*}}$ refers to the average over isotopes and spin states.

There is no correlation between the scattering length of a nucleus belonging to a given chemical species and the atomic position, since the scattering length is determined by the

isotope and the spin state of the neutron - nucleus system. Hence [4]:

$$\overline{b_j b^*_{j'}} = \overline{b_j}\ \overline{b^*_{j'}} \qquad \text{with} \qquad j \neq j' \tag{2.122}$$
$$\overline{b_j b^*_{j'}} = \overline{b_j^2} \qquad \text{with} \qquad j = j'.$$

Average values are defined as follows:

$$\overline{b_j} = \sum_i p_j^i\ b_j^i \tag{2.123}$$
$$\overline{b_j^2} = \sum_i p_j^i\ (b_j^i)^2$$

where $p_j^i$ represent the probability that the element at position $\vec{R}_j$ has scattering length $b_j^i$ and the equation:

$$\sum_i p_i = 1 \tag{2.124}$$

is fulfilled, with $i$ denoting the possible isotopes and spin states.

Equation 2.121 can be formulated as the sum of two terms:

$$\begin{aligned} & S(\vec{Q},\omega) \qquad (2.125)\\ &= \sum_{j\neq j'} \overline{b_j}\ \overline{b^*_{j'}} S_{j,j'}(\vec{Q},\omega) + \sum_{j=1}^{N} \overline{b_j^2} S_{j,j}(\vec{Q},\omega) \\ &= \sum_{j,j'} \overline{b_j}\ \overline{b^*_{j'}} S_{j,j'}(\vec{Q},\omega) + \sum_{j=1}^{N} (\overline{b_j^2} - \overline{b_j}^2) S_{j,j}(\vec{Q},\omega) \end{aligned}$$

Expression 2.125 allows the identification of two contributions within the equation defining the double differential cross section 2.116 namely:

$$\frac{d^2\sigma}{d\Omega dE_f} = \left(\frac{d^2\sigma}{d\Omega dE_f}\right)_{\text{coherent}} + \left(\frac{d^2\sigma}{d\Omega dE_f}\right)_{\text{incoherent}}. \tag{2.126}$$

The *coherent* contribution to the cross section is obtained by considering for every pair of scatterers $j$ and $j'$ the average values $\overline{b_j}$ and $\overline{b^*_{j'}}$:

$$\left(\frac{d^2\sigma}{d\Omega dE_f}\right)_{\text{coherent}} = \tag{2.127}$$
$$\frac{k_f}{k_i}\frac{1}{2\pi\hbar}\sum_{\lambda_i} p(\lambda_i) \sum_{j,j'=1}^{N} \overline{b_j}\ \overline{b^*_{j'}} \int dt\ \langle\lambda_i| e^{-i\vec{Q}\cdot\vec{R}_{j'}} e^{i\vec{Q}\cdot\vec{R}_j(t)} |\lambda_i\rangle\ e^{-i\omega t}.$$

[4] Two statistical variables $A$ and $B$ are uncorrelated if: $\overline{AB} = \overline{A}\ \overline{B}$

On the other hand the *incoherent* term presents a single sum over individual scatterers $j$. For every scatterer the variance of the neutron scattering length appears, where the scattering length average value is calculated by considering the statistical distribution of isotopes and spin states.

$$\left(\frac{d^2\sigma}{d\Omega dE_f}\right)_{\text{incoherent}} = \tag{2.128}$$
$$\frac{k_f}{k_i}\frac{1}{2\pi\hbar}\sum_{\lambda_i} p(\lambda_i)\sum_{j=1}^{N}(\overline{b_j^2}-\overline{b_j}^2)\int dt\,\langle\lambda_i|e^{-i\vec{Q}\cdot\vec{R}_j}e^{i\vec{Q}\cdot\vec{R}_j(t)}|\lambda_i\rangle\,e^{-i\omega t}.$$

### 2.4.2 Monoatomic systems

A monoatomic system is a system composed of a single chemical element. In these circumstances the total coherent and incoherent scattering cross sections can be defined:

$$\sigma_{\text{coherent}} = 4\pi\overline{b}^2 \tag{2.129}$$
$$\sigma_{\text{incoherent}} = 4\pi(\overline{b^2}-\overline{b}^2) = 4\pi\overline{(b-\overline{b})^2}.$$

Equation 2.127 in the monoatomic case yields:

$$\left(\frac{d^2\sigma}{d\Omega dE_f}\right)_{\text{coherent}} = \tag{2.130}$$
$$\frac{\sigma_{\text{coherent}}}{4\pi}\frac{k_f}{k_i}\frac{1}{2\pi\hbar}\sum_{\lambda_i} p(\lambda_i)\sum_{j,j'=1}^{N}\int dt\,\langle\lambda_i|e^{-i\vec{Q}\cdot\vec{R}_{j'}}e^{i\vec{Q}\cdot\vec{R}_j(t)}|\lambda_i\rangle\,e^{-i\omega t}$$

while equation 2.131 yields:

$$\left(\frac{d^2\sigma}{d\Omega dE_f}\right)_{\text{incoherent}} = \tag{2.131}$$
$$\frac{\sigma_{\text{incoherent}}}{4\pi}\frac{k_f}{k_i}\frac{1}{2\pi\hbar}\sum_{\lambda_i} p(\lambda_i)\sum_{j=1}^{N}\int dt\,\langle\lambda_i|e^{-i\vec{Q}\cdot\vec{R}_j}e^{i\vec{Q}\cdot\vec{R}_j(t)}|\lambda_i\rangle\,e^{-i\omega t}.$$

### 2.4.3 The spin state of the combined neutron - nucleus system

The phenomenon of neutron scattering from the nuclear potential is characterized by rotational symmetry *i.e.* the scattering potential is isotropic $V(\vec{r}) = V(r)$. Noether's

theorem relates rotational symmetry to the conservation of angular momentum [5]. The total angular momentum of the neutron - target nucleus system is a conserved quantity. In the case of $s$-wave scattering [6] no orbital contribution is present and the total angular momentum operator $\hat{\mathbf{J}}$ is the sum of the spin operator of the probe particle $\hat{\mathbf{S}}$ and of the target nucleus $\hat{\mathbf{I}}$:

$$\hat{\mathbf{J}} = \hat{\mathbf{S}} + \hat{\mathbf{I}}. \tag{2.132}$$

The neutron spin is $S = 1/2$ and the total spin quantum numbers are:

$$\begin{aligned} J^+ &= I + 1/2 \\ J^- &= I - 1/2. \end{aligned} \tag{2.133}$$

The neutron scattering length of an isotope depends on the total spin state of the neutron - nucleus system. Therefore the nuclear scattering length of an isotope with nuclear spin $I$ assumes the values $b^+$ and $b^-$ corresponding to the total spin quantum numbers $J^+$ and $J^-$ respectively. Each $\hat{\mathbf{J}}$ eigenstate shows a degeneracy $n_J = 2J + 1$. Hence the probability of each total spin state is:

$$\begin{aligned} p^+ &= \frac{n^+}{n^+ + n^-} \\ p^- &= \frac{n^-}{n^+ + n^-}. \end{aligned} \tag{2.134}$$

The statistical mean values of the scattering length are:

$$\begin{aligned} \overline{b} &= p^+ b^+ + p^- b^- \\ \overline{b^2} &= p^+ (b^+)^2 + p^- (b^-)^2. \end{aligned} \tag{2.135}$$

### 2.4.4 Neutron scattering from protons and deuterons

In this paragraph the general formalism developed in the previous sections is applied to the case of neutron scattering from protons and deuterons. The values of the coherent

[5] In quantum mechanical terms, when the potential is isotropic the Hamiltonian of the system commutes with the angular momentum operators. Hence a common set of energy and angular momentum eigenstates exists.

[6] $s$-wave scattering appears to be the relevant term in partial waves expansion approach [27]

and incoherent scattering cross sections are derived for the system composed of the single isotope ${}^1_1\mathrm{H}$ and for that composed of ${}^2_1\mathrm{H}$.

**Neutron scattering from protons** Both neutrons and protons are characterized by spin quantum number 1/2. A basis of the 4-dimensional Hilbert space of the combined neutron - proton spin system is given by:

$$\begin{aligned} |\uparrow\uparrow\rangle &= |\uparrow\rangle_n \otimes |\uparrow\rangle_p \\ |\downarrow\uparrow\rangle &= |\downarrow\rangle_n \otimes |\uparrow\rangle_p \\ |\uparrow\downarrow\rangle &= |\uparrow\rangle_n \otimes |\downarrow\rangle_p \\ |\downarrow\downarrow\rangle &= |\downarrow\rangle_n \otimes |\downarrow\rangle_p \end{aligned} \tag{2.136}$$

where the indexes $n$ and $p$ refer to the neutron and the proton respectively. The total spin operator $\hat{\mathbf{J}}$ is not diagonal in this basis. The total spin quantum number assumes the following values:

$$\begin{aligned} J^+ &= I + 1/2 = 1 \\ J^- &= I - 1/2 = 0 \end{aligned} \tag{2.137}$$

yielding two eigenvalues of $\hat{\mathbf{J}}^{\mathbf{2}}$:

$$\begin{aligned} &\hbar^2 J^+(J^+ + 1) \\ &\hbar^2 J^-(J^- + 1). \end{aligned} \tag{2.138}$$

Each eigenvalue corresponds to a subspace of dimension $2J+1$ spanned by the eigenvectors $|J, M_J\rangle$:

$$\begin{aligned} J = 1 \qquad & |1,1\rangle = |\uparrow\uparrow\rangle \\ & |1,0\rangle = \frac{1}{\sqrt{2}}(|\downarrow\uparrow\rangle + |\uparrow\downarrow\rangle) \\ & |1,-1\rangle = |\downarrow\downarrow\rangle \end{aligned} \tag{2.139}$$

$$J = 0 \qquad |0,0\rangle = \frac{1}{\sqrt{2}}(|\downarrow\uparrow\rangle - |\uparrow\downarrow\rangle) \tag{2.140}$$

with:

$$\hat{\mathbf{J}}^2 \left|J, M_J\right\rangle = \hbar^2 J(J+1) \left|J, M_J\right\rangle \tag{2.141}$$
$$\hat{J}_z \left|J, M_J\right\rangle = \hbar M_J \left|J, M_J\right\rangle .$$

Neutron scattering lengths depend on $J$ and the experimental values in the case of protons are:

$$\begin{array}{lll} J^+ = 1 & b^+({}^1_1\mathrm{H}) = +10.4 \text{ fm} & p^+ = \dfrac{3}{4} \\ J^- = 0 & b^-({}^1_1\mathrm{H}) = -47.4 \text{ fm} & p^- = \dfrac{1}{4}. \end{array} \tag{2.142}$$

This yields:

$$\overline{b}({}^1_1\mathrm{H}) = -3.8 \text{ fm}. \tag{2.143}$$

The neutron scattering length of the proton has a negative value. The corresponding neutron scattering cross sections are:

$$\sigma_{\text{coherent}}({}^1_1\mathrm{H}) = 4\pi\overline{b}^2 = 1.8 \text{ barn} \tag{2.144}$$
$$\sigma_{\text{incoherent}}({}^1_1\mathrm{H}) = 4\pi(\overline{b^2} - \overline{b}^2) = 79.8 \text{ barn}.$$

The value of the incoherent scattering cross section of protons greatly exceeds that of the coherent cross section.

**Neutron scattering from deuterons** Deuterons have nuclear spin quantum number is $I = 1$ and the total spin quantum number for the combined system takes the values:

$$J^+ = I + 1/2 = 3/2 \tag{2.145}$$
$$J^- = I - 1/2 = 1/2.$$

The experimental values of the neutron scattering length of the deuteron are:

$$\begin{array}{lll} J^+ = 3/2 & b^+({}^2_1\mathrm{H}) = +9.5 \text{ fm} & p^+ = \dfrac{4}{6} \\ J^- = 1/2 & b^-({}^2_1\mathrm{H}) = +1.0 \text{ fm} & p^- = \dfrac{2}{6}. \end{array} \tag{2.146}$$

This yields:

$$\overline{b}(^2_1\mathrm{H}) = +6.7 \text{ fm}. \tag{2.147}$$

The neutron scattering length of the deuteron has a positive value. The corresponding neutron scattering cross sections are:

$$\begin{aligned} \sigma_{\text{coherent}}(^2_1\mathrm{H}) &= 4\pi\overline{b}^2 = 5.6 \text{ barn} \\ \sigma_{\text{incoherent}}(^2_1\mathrm{H}) &= 4\pi(\overline{b^2} - \overline{b}^2) = 2.0 \text{ barn}. \end{aligned} \tag{2.148}$$

The incoherent scattering cross section of deuterons is significantly smaller than the coherent cross section.

The values of scattering lengths and cross sections presented in this paragraph explain why partial or total deuteration is required in neutron macromolecular crystallography. This aspect of neutron diffraction was anticipated in Section 1.4.

Protein crystals are composed of about 50% hydrogen atoms. Equation 2.143 shows that protons have a negative value of the scattering length. For this reason protons appear as negative density in neutron scattering density maps, in contrast to most isotopes in protein crystals that appear as positive density peaks. This leads to neutron density cancellation around neighboring atoms in -$CH_2$ and -$CH_3$ groups at resolutions greater than 2 Å, hampering the accurate interpretation of neutron maps [18].

The diffraction signal is produced by elastic coherent scattering as shown in Section 2.6. Protons exhibit a large incoherent scattering cross section as reported in Equation 2.144 that does not contribute to the diffracted intensity. This gives rise to a large and significant background in the crystallography experiment [18].

By contrast, deuterons have a positive neutron scattering length as shown in Equation 2.147. This value is similar to the scattering lengths of the other atoms found in proteins, namely carbon, oxygen, nitrogen and sulphur (Table 1.1 in Chapter 1). Hence no density cancellation occurs in fully deuterated samples where aliphatic groups are -$CD_2$ and -$CD_3$.

In addition incoherent scattering from deuterons is much smaller than that from protons as results by comparing the values of the incoherent scattering cross section. Partial

or total deuteration of protein crystals in neutron macromolecular crystallography is required to limit the incoherent background signal in diffraction data acquisition.

## 2.5 The relationship between scattering cross sections and correlation functions

The present section describes the relationship between the double differential scattering cross section and the correlation function of the number density operator providing some insight into the physics of coherent and incoherent scattering phenomena.

### 2.5.1 Formulation of the scattering function in terms of quantum statistical average

The expression of the scattering function 2.117 here rewritten for convenience:

$$S(\vec{Q},\omega) = \frac{1}{2\pi\hbar}\sum_{\lambda_i} p(\lambda_i) \sum_{j,j'=1}^{N} b_j b_{j'}^{*} \int dt\, \langle \lambda_i | e^{-i\vec{Q}\cdot\vec{R}_{j'}} e^{i\vec{Q}\cdot\vec{R}_j(t)} | \lambda_i \rangle\, e^{-i\omega t} \tag{2.149}$$

can be reformulated by using the following definition of quantum statistical average:

$$\langle e^{-i\vec{Q}\cdot\vec{R}_{j'}} e^{i\vec{Q}\cdot\vec{R}_j(t)} \rangle = \sum_{\lambda_i} p(\lambda_i)\, \langle \lambda_i | e^{-i\vec{Q}\cdot\vec{R}_{j'}} e^{i\vec{Q}\cdot\vec{R}_j(t)} | \lambda_i \rangle \tag{2.150}$$

where $|\lambda_i\rangle$ represents a complete basis for the target states. Equation 2.149 is written as:

$$S(\vec{Q},\omega) = \sum_{j,j'=1}^{N} b_j b_{j'}^{*}\ \ S_{j,j'}(\vec{Q},\omega) \tag{2.151}$$

with:

$$S_{j,j'}(\vec{Q},\omega) = \frac{1}{2\pi\hbar}\int dt\, \langle e^{-i\vec{Q}\cdot\vec{R}_{j'}} e^{i\vec{Q}\cdot\vec{R}_j(t)} \rangle\, e^{-i\omega t}. \tag{2.152}$$

Equation 2.151 highlights the fact that the scattering function can be expressed as a double sum where the indexes $j$ and $j'$ run over the individual scatterers.

Individual scatterers can be classified according to their type or class $\kappa$. This can for instance denote the chemical element. The sum over individual scatterers $j$ can be

formulated according to:

$$\sum_{j} = \sum_{\kappa} \sum_{j \in \{j_\kappa\}} \tag{2.153}$$

where the sum is performed over the different classes $\kappa$ and for each class $\kappa$ a second sum is carried out over the individual scatterers $j_\kappa$ belonging to that class. By introducing this notation Equation 2.151 yields:

$$S(\vec{Q},\omega) = \sum_{\kappa,\kappa'} b_\kappa b^*_{\kappa'} \;\; S_{\kappa,\kappa'}(\vec{Q},\omega) \tag{2.154}$$

with:

$$S_{\kappa,\kappa'}(\vec{Q},\omega) = \frac{1}{2\pi\hbar} \sum_{j \in \{j_\kappa\}, j' \in \{j_{\kappa'}\}} \int dt \, \langle e^{-i\vec{Q}\cdot\vec{R}_{j'}} e^{i\vec{Q}\cdot\vec{R}_j(t)} \rangle \, e^{-i\omega t}. \tag{2.155}$$

This quantity is known as the *partial scattering function*. In an analogous fashion the *partial self scattering function* is defined according to:

$$S^{\text{self}}_{\kappa}(\vec{Q},\omega) = \frac{1}{2\pi\hbar} \sum_{j \in \{j_\kappa\}} \int dt \, \langle e^{-i\vec{Q}\cdot\vec{R}_{j}} e^{i\vec{Q}\cdot\vec{R}_j(t)} \rangle \, e^{-i\omega t}. \tag{2.156}$$

### 2.5.2 The correlation function of the number density operator

The probability density of a particle in the quantum state $|\lambda\rangle$ of being at position $\vec{R}$ at time $t$ is:

$$\psi^*_\lambda\big(\vec{R}(t)\big)\psi_\lambda\big(\vec{R}(t)\big) = \int d\vec{r} \, \delta\big(\vec{r} - \vec{R}(t)\big)\psi^*_\lambda(\vec{r})\psi_\lambda(\vec{r}) = \langle\lambda|\hat{\rho}(\vec{r},t)|\lambda\rangle \tag{2.157}$$

where the number density operator is defined as:

$$\hat{\rho}(\vec{r},t) = \delta\big(\vec{r} - \vec{R}(t)\big). \tag{2.158}$$

In the case of a system composed of several particles this expression takes the form:

$$\hat{\rho}_\kappa(\vec{r},t) = \sum_{j \in \{j_\kappa\}} \delta\big(\vec{r} - \vec{R}_j(t)\big). \tag{2.159}$$

The eigenvalues of this operator provide the probability density of presence of a scatterer of class $\kappa$ at position $\vec{r}$.

The *correlation function* of the number density operator is defined as:

$$G_{\kappa,\kappa'}(\vec{r},t) = \int d\vec{r}\,' \langle \hat{\rho}_\kappa(\vec{r}\,' - \vec{r})\hat{\rho}_{\kappa'}(\vec{r}\,',t)\rangle \tag{2.160}$$
$$= \sum_{j\in\{j_\kappa\},j'\in\{j_{\kappa'}\}} \int d\vec{r}\,' \left\langle \delta\big(\vec{r}\,' - \vec{r} - \vec{R}_j\big)\delta\big(\vec{r}\,' - \vec{R}_{j'}(t)\big)\right\rangle$$

where $\vec{R}_j = \vec{R}_j(t=0)$. This function relates the probability of finding a scatterer of class $\kappa$ at time $t=0$ and position $\vec{r}\,' - \vec{r}$ to that of finding a scatterer belonging to the class $\kappa'$ at time $t$ and position $\vec{r}\,'$. Similarly the *partial self pair correlation function* can be defined as follows:

$$G^{\text{self}}_{\kappa}(\vec{r},t) = \sum_{j\in\{j_\kappa\}} \int d\vec{r}\,' \left\langle \delta\big(\vec{r}\,' - \vec{r} - \vec{R}_j\big)\delta\big(\vec{r}\,' - \vec{R}_j(t)\big)\right\rangle . \tag{2.161}$$

This function describes the probability of finding a scatterer of class $\kappa$ at time $t=0$ and position $\vec{r}\,' - \vec{r}$ in presence of the same scatterer at time $t$ and position $\vec{r}\,'$. It is worthwhile stressing that the delta functions appearing in equations 2.160 and 2.161 in general do not commute since $\vec{R}_j$ is a quantum mechanical operator.

### 2.5.3 The relationship between scattering functions and correlation functions

The relationship between the partial scattering function $S_{\kappa,\kappa'}(\vec{Q},\omega)$ in Equation 2.155 and the correlation function of the number density operator $G_{\kappa,\kappa'}(\vec{r},t)$ defined in Equation 2.160 is given by:

$$G_{\kappa,\kappa'}(\vec{r},t) \propto \int d\vec{Q} \int d\omega\, e^{-i(\vec{Q}\vec{r}-\omega t)} S_{\kappa,\kappa'}(\vec{Q},\omega). \tag{2.162}$$

The correlation function of the number density operator $G_{\kappa,\kappa'}(\vec{r},t)$ is the double Fourier transform in reciprocal space and energy of the partial scattering function $S_{\kappa,\kappa'}(\vec{Q},\omega)$. Hence:

$$S_{\kappa,\kappa'}(\vec{Q},\omega) \propto \int d\vec{r} \int dt\, e^{i(\vec{Q}\vec{r}-\omega t)} G_{\kappa,\kappa'}(\vec{r},t). \tag{2.163}$$

The partial scattering function $S_{\kappa,\kappa'}(\vec{Q},\omega)$ is the double Fourier transform in space and time of the correlation function of the number density operator $G_{\kappa,\kappa'}(\vec{r},t)$.

Equation 2.163 highlights the fact that the scattering experiment provides the Fourier analysis of space and time correlations of the number density operator.

### 2.5.4 Coherent and incoherent scattering cross sections expressed through correlation functions

By using equations 2.116, 2.154, 2.155 and 2.163 the following equation expressing the relationship between the double differential scattering cross section and the correlation function of the number density operator is obtained:

$$\left(\frac{d^2\sigma}{d\Omega dE_f}\right)_{\vec{k}_i\to\vec{k}_f} \propto \frac{k_f}{k_i}\sum_{\kappa,\kappa'} b_\kappa b^*_{\kappa'} \int d\vec{r} \int dt\; e^{i(\vec{Q}\vec{r}-\omega t)} G_{\kappa,\kappa'}(\vec{r},t). \tag{2.164}$$

In the case of a monoatomic system the coherent and incoherent contributions can be separated:

$$\left(\frac{d^2\sigma}{d\Omega dE_f}\right)^{\text{coherent}}_{\vec{k}_i\to\vec{k}_f} \propto \sigma_{\text{coherent}} \frac{k_f}{k_i} \int d\vec{r} \int dt\; e^{i(\vec{Q}\vec{r}-\omega t)} G(\vec{r},t) \tag{2.165}$$

$$\left(\frac{d^2\sigma}{d\Omega dE_f}\right)^{\text{incoherent}}_{\vec{k}_i\to\vec{k}_f} \propto \sigma_{\text{incoherent}} \frac{k_f}{k_i} \int d\vec{r} \int dt\; e^{i(\vec{Q}\vec{r}-\omega t)} G^{\text{self}}(\vec{r},t). \tag{2.166}$$

These equations shed light on the physical meaning of coherent and incoherent scattering phenomena. They show that while coherent scattering provides information about correlations between scatterers, incoherent scattering yields information about the correlation of a particle with itself.

In neutron crystallography the relevant contribution is the coherent (and elastic) term. This gives rise to the diffraction spots as shown in section 2.6. In this experimental technique incoherent scattering is considered as a nuisance contributing to the background level of the experiment. As pointed out in Section 2.4.4 in neutron macromolecular diffraction experiments it is necessary to keep the incoherent scattering signal to a minimum. This results in the need for sample deuteration and perdeuteration. The production of fully deuterated (perdeuterated) samples requires a significant experimental effort (Appendix B).

## 2.6 Neutron diffraction

In this section the derivation of the mathematical condition for diffraction is presented. This is equivalent to the equations describing diffraction derived by M. von Laue in 1912 [33]. The general expression of the dynamic structure factor is considered. The position of each individual scatterer at any time is expressed in terms of displacement from the equilibrium position in the harmonic oscillator picture. The Debye-Waller factor is defined and by considering the term corresponding to coherent elastic phenomena in the dynamic structure factor formula the diffraction condition is obtained.

### 2.6.1 The Debye-Waller factor

The starting point are equations 2.154 and 2.155 here repeated for convenience:

$$S(\vec{Q},\omega) = \sum_{\kappa,\kappa'} b_\kappa b^*_{\kappa'} \;\; S_{\kappa,\kappa'}(\vec{Q},\omega) \tag{2.167}$$

with:

$$S_{\kappa,\kappa'}(\vec{Q},\omega) = \frac{1}{2\pi\hbar} \sum_{j\in\{j_\kappa\},j'\in\{j_{\kappa'}\}} \int dt \, \langle e^{-i\vec{Q}\cdot\vec{R}_{j'}} e^{i\vec{Q}\cdot\vec{R}_j(t)} \rangle \, e^{-i\omega t} \tag{2.168}$$

where $\vec{R}_{j'}$ stands for $\vec{R}_{j'}(t=0)$. Assumption is made that the system is harmonic and the individual atomic positions are expressed in terms of displacement $\vec{u}_j(t)$ from the equilibrium position $\vec{R}^0_j(t)$ according to:

$$\vec{u}_j(t) = \vec{R}_j(t) - \vec{R}^0_j. \tag{2.169}$$

The quantum statistical average in equation 2.168 can be written:

$$\begin{aligned} \langle e^{-i\vec{Q}\cdot\vec{R}_{j'}} e^{i\vec{Q}\cdot\vec{R}_j(t)} \rangle &= \langle e^{-i\vec{Q}\cdot[\vec{R}^0_{j'}+\vec{u}_{j'}(t=0)]} e^{i\vec{Q}\cdot[\vec{R}^0_j+\vec{u}_j(t)]} \rangle \\ &= e^{-i\vec{Q}\cdot(\vec{R}^0_{j'}-\vec{R}^0_j)} \, \langle e^{-i\vec{Q}\cdot\vec{u}_{j'}(t=0)} e^{i\vec{Q}\cdot\vec{u}_j(t)} \rangle \end{aligned} \tag{2.170}$$

where the equilibrium positions $\vec{R}^0_j$ can be treated as classical vectors while the atomic displacements $\vec{u}_{j'}$ are quantum mechanical operators. In the harmonic approximation it can be shown [27] that the expectation value of the exponential function appearing in

2.170 can be expressed by means of exponential functions of expectation values according to:

$$\langle e^{-i\vec{Q}\cdot\vec{u}_{j'}(t=0)} e^{i\vec{Q}\cdot\vec{u}_j(t)}\rangle = e^{-\frac{1}{2}\langle[\vec{Q}\cdot\vec{u}_{j'}]^2\rangle} e^{-\frac{1}{2}\langle[\vec{Q}\cdot\vec{u}_j(t)]^2\rangle} e^{\langle[\vec{Q}\cdot\vec{u}_{j'}][\vec{Q}\cdot\vec{u}_j(t)]\rangle}. \tag{2.171}$$

The Debye-Waller factor is defined as:

$$e^{-W_j(\vec{Q})} = e^{-\frac{1}{2}\langle[\vec{Q}\cdot\vec{u}_j]^2\rangle}. \tag{2.172}$$

By using this definition the partial scattering function takes the form:

$$\begin{aligned} & S_{\kappa,\kappa'}(\vec{Q},\omega) \qquad (2.173)\\ & = \frac{1}{2\pi\hbar} \sum_{j\in\{j_\kappa\}, j'\in\{j_{\kappa'}\}} \int dt\, \langle e^{-i\vec{Q}\cdot\vec{R}_{j'}} e^{i\vec{Q}\cdot\vec{R}_j(t)}\rangle\, e^{-i\omega t} \\ & = \frac{1}{2\pi\hbar} \sum_{j\in\{j_\kappa\}, j'\in\{j_{\kappa'}\}} \int dt\, e^{-i\vec{Q}\cdot(\vec{R}^0_{j'}-\vec{R}^0_j)} \langle e^{-i\vec{Q}\cdot\vec{u}_{j'}} e^{i\vec{Q}\cdot\vec{u}_j(t)}\rangle\, e^{-i\omega t} \\ & = \frac{1}{2\pi\hbar} \sum_{j\in\{j_\kappa\}, j'\in\{j_{\kappa'}\}} \int dt\, e^{-i\vec{Q}\cdot(\vec{R}^0_{j'}-\vec{R}^0_j)} e^{-W_{j'}(\vec{Q})} e^{-W_j(\vec{Q})} e^{\langle[\vec{Q}\cdot\vec{u}_{j'}][\vec{Q}\cdot\vec{u}_j(t)]\rangle} e^{-i\omega t} \end{aligned}$$

where $\vec{R}_{j'}$ stands for $\vec{R}_{j'}(t=0)$ and $\vec{u}_{j'}$ stands for $\vec{u}_{j'}(t=0)$.

### 2.6.2 Coherent elastic scattering

The exponential of the expectation value $\langle[\vec{Q}\cdot\vec{u}_{j'}][\vec{Q}\cdot\vec{u}_j(t)]\rangle$ appears in the integrand within the expression of $S_{\kappa,\kappa'}(\vec{Q},\omega)$. Provided that the exponent assumes small values the exponential function can be expanded in Taylor series about the origin and by considering only the first term of the expansion the exponential equals 1 and is time independent. This term gives rise to elastic scattering:

$$\begin{aligned} & S^{\text{elastic}}_{\kappa,\kappa'}(\vec{Q},\omega) \qquad (2.174)\\ & = \frac{1}{2\pi\hbar} \sum_{j\in\{j_\kappa\}, j'\in\{j_{\kappa'}\}} \int dt\, e^{-i\vec{Q}\cdot(\vec{R}^0_{j'}-\vec{R}^0_j)} e^{-W_{j'}(\vec{Q})} e^{-W_j(\vec{Q})} e^{-i\omega t} \\ & = \sum_{j\in\{j_\kappa\}, j'\in\{j_{\kappa'}\}} e^{-i\vec{Q}\cdot(\vec{R}^0_{j'}-\vec{R}^0_j)} e^{-W_{j'}(\vec{Q})} e^{-W_j(\vec{Q})} \delta(\hbar\omega) \end{aligned}$$

where the integral representation of the Dirac delta has been used. It is worthwhile observing that due to the presence of the Dirac delta term $\delta(\hbar\omega)$ this expression is non-zero

only when $\omega = 0$ *i.e.* there is no energy transfer between probe and target, corresponding to the case of elastic scattering.

In a crystal the real space lattice can be defined by means of a set basis vectors $\vec{a}_1, \vec{a}_2, \vec{a}_3$ by:

$$\vec{l} = l_1\vec{a}_1 + l_2\vec{a}_2 + l_3\vec{a}_3 \tag{2.175}$$

and the atomic equilibrium positions are:

$$\vec{R}_j^0 = \vec{l} + \vec{d} \tag{2.176}$$

where $\vec{d}$ describes the position within the unit cell.

The coherent elastic differential cross section is:

$$\begin{aligned}
&\left(\frac{d^2\sigma}{d\Omega dE_f}\right)_{\text{coherent}}^{\text{elastic}} \\
&= S_{\text{coherent}}(\vec{Q}, \omega = 0) \\
&= \sum_{\kappa,\kappa'} \overline{b_\kappa}\ \overline{b^*_{\kappa'}}\ \ S_{\kappa,\kappa'}(\vec{Q}, \omega = 0) \\
&= \sum_{l,l'} e^{i\vec{Q}\cdot(\vec{l}-\vec{l}')} \sum_{d,d'} \overline{b_d}\ \overline{b^*_{d'}}\ \ e^{i\vec{Q}\cdot\vec{d}} e^{-i\vec{Q}\cdot\vec{d}'} e^{-W_{d'}(\vec{Q})} e^{-W_d(\vec{Q})} \delta(\hbar\omega) \\
&\propto N_{\text{UC}} \sum_l e^{i\vec{Q}\cdot\vec{l}} \left| \sum_d \overline{b_d}\ \ e^{i\vec{Q}\cdot\vec{d}} e^{-W_d(\vec{Q})} \right|^2 \delta(\hbar\omega)
\end{aligned} \tag{2.177}$$

where $N_{\text{UC}}$ is the number of unit cells in the sample and the relation:

$$\sum_l e^{i\vec{Q}\cdot(\vec{l}-\vec{l}')} = \sum_l e^{i\vec{Q}\cdot\vec{l}} \tag{2.178}$$

holds due to the translational invariance of the system.

The nuclear form factor of the unit cell is defined according to:

$$F(\vec{Q}) = \sum_d \overline{b_d}\ e^{-W_d(\vec{Q})} e^{i\vec{Q}\cdot\vec{d}} \tag{2.179}$$

Any reciprocal lattice vector can be expressed via:

$$\vec{G} = h\vec{b}_1 + k\vec{b}_2 + l\vec{b}_3 \tag{2.180}$$

where the reciprocal lattice basis vectors are:

$$\vec{b}_1 = 2\pi \frac{\vec{a}_2 \times \vec{a}_3}{\vec{a}_1 \cdot (\vec{a}_2 \times \vec{a}_3)}$$
$$\vec{b}_2 = 2\pi \frac{\vec{a}_3 \times \vec{a}_1}{\vec{a}_1 \cdot (\vec{a}_2 \times \vec{a}_3)}$$
$$\vec{b}_3 = 2\pi \frac{\vec{a}_1 \times \vec{a}_2}{\vec{a}_1 \cdot (\vec{a}_2 \times \vec{a}_3)}. \tag{2.181}$$

It turns out that:

$$\vec{l} \cdot \vec{G} = 2\pi n \tag{2.182}$$

with $n$ integer, for any direct lattice vector $\vec{l}$ and any reciprocal lattice vector $\vec{G}$.

The sum expressed in Equation 2.178 is equal to zero except when the reciprocal space vector $\vec{Q}$ corresponds to a reciprocal lattice vector $\vec{G}$, for which 2.182 holds. It follows that:

$$\sum_l e^{i\vec{Q}\cdot\vec{l}} \propto \frac{1}{V_{\mathrm{UC}}} \sum_{\vec{G}} \delta(\vec{Q} - \vec{G}) \tag{2.183}$$

where $V_{\mathrm{UC}}$ is the unit cell volume. Equation 2.177 takes the form:

$$\left(\frac{d^2\sigma}{d\Omega dE_f}\right)^{\mathrm{elastic}}_{\mathrm{coherent}} = \frac{N_{\mathrm{UC}}}{V_{\mathrm{UC}}} \left|F(\vec{Q})\right|^2 \sum_{\vec{G}} \delta(\vec{Q} - \vec{G})\ \delta(\hbar\omega) \tag{2.184}$$

where the term $\delta(\vec{Q} - \vec{G})$ expresses M. von Laue diffraction conditions. This equation is of paramount importance since it highlights the information content of the diffraction experiment. A complementary approach to the diffraction experiment is obtained interpreting the diffracted intensity in terms the squared Fourier transform of the real space scattering density. Such approach is developed in details in reference [34] and [35].

## 2.7 Concluding remarks

The capability of neutron diffraction to localize deuterium substituted hydrogen atoms allows the determination of protonation states, hydrogen bond networks and orientation of water molecules in macromolecular structures. This capability together with the requirement for partial or full deuteration of samples is explained in terms of neutron scattering lengths and cross sections.

Neutron scattering lengths are an intrinsic property of isotopes and enter the scattering equations *via* Fermi's expression of the nuclear potential. Cross sections are obtained for monoatomic systems by expressing the scattering function as the sum of coherent and incoherent contributions. The incoherent term arises as a consequence of the spin dependence of scattering lengths. This contribution is particularly important in the case of protons generating large background in diffraction measurements. The problem can be circumvented by isotope substitution replacing hydrogen for deuterium.

M. von Laue conditions for diffraction are obtained by considering the double differential scattering cross section in the coherent and elastic scenario.

# Bibliography


[1] H. B. Dunford, *Peroxidases and catalases: Biochemistry, Biophysics, Biotehnology and Physiology* Wiley (2010).

[2] M. Sivaraja, D. B. Goodin, M. Smith and B. M. Hoffman *Science* **245**, 738-740 (1989).

[3] T. L. Poulos, S. T. Freer, R. A. Alden, N. H. Xuong, S. L. Edwards, R. C. Hamlin and J. Kraut, *The Journal of Biological Chemistry* **253**, 3730-3735 (1978).

[4] T. L. Poulos, S. T. Freer, R. A. Alden, S. L. Edwards, U. Skogland, K. Takio, B. Eriksson, N. H. Xuong, T. Yonetani and J. Kraut, *The Journal of Biological Chemistry* **255**, 575-580 (1980).

[5] B. C. Finzel, T. L. Poulos and J. Kraut, *The Journal of Biological Chemistry* **259**, 13027-13036 (1984).

[6] H. Pelletier and J. Kraut *Science* **258**, 1748-1755 (1992).

[7] T. L. Poulos in *Heme Peroxidases* (E. L. Raven, H. B. Dunford eds) RSC Publishing (2015).

[8] R. K. Behan and M. T. Green *Journal of Inorganic Biochemistry* **100**, 448-459 (2006).

[9] A. N. Volkov, P. Nicholls and A. R. Worrall *Biochimica and Biophysica Acta* **1807**, 1482-1503 (2011).

[10] T. L. Poulos *Chemical Reviews* **114**, 3919-3962 (2014).

[11] J. Terner, V. Palaniappan, A. Gold, R. Weiss, M. M. Fitzgerald, A. M. Sullivan, C. M. Hosten, *Journal of Inorganic Biochemistry* **100**, 480501 (2006).

[12] C. M. Wilmot, T. Sjoegren, G. H. Carlsson, G. I. Berglund and J. Hajdu *Methods in Enzymology* **353**, 301-318 (2002).

[13] H. P. Hersleth, Y. W. Hsiao, C. H. Goerbitz and K. K. Andersson *Chemistry and Biodiversity* **5**, 2067-2089 (2008).

[14] N. Nimura and A. Podjarny, *Neutron Protein Crystallography.* Oxford Science Publications (2011).

[15] A. Gumiero, C. L. Metcalfe, A. R. Pearson, E. L. Raven and P. C. E. Moody, *The Journal of Biological Chemistry* **286**, No. 2 (2011).

[16] M. T. Green *Journal of the American Chemistry Society* **128**, 1902-1906 (2006).

[17] Y. T. Meharenna, T. Doukov, H. Li, S. M. Soltis and T. L. Poulos *Biochemistry* **49**, 2984-2986 (2010).

[18] F. Meilleur, D. A. A. Myles and M. P. Blakeley, *Eur. Biophys. J.* (2006).

[19] M. P. Blakeley, P. Langan, N. Nimura and A. Podjarny, *Current Opinion in Structural Biology* **18**, (2008).

[20] Howard E., Sanishvili R.,Cachau R. E., Mitschler A., Chevrier B., Barth P., Lamour V., Van Zandt M., Sibley E., Bon C., Moras D., Schneider T. R., Joachimiak A. and Podjarny A.,*Proteins* **55**, 792-804 (2004).

[21] I. T. Weber, M. J. Waltman, M. Mustyakimov, M. P. Blakeley, D. A. Keen, A. K. Ghosh, P. Langan and A. Y. Kovalevsky, *J. Med. Chem.* **56**, 5631-5635 (2013).

[22] E. I. Howard, M. P. Blakeley, M. Haertlein, I. Petit-Haertlein, A. Mitschler, S. J. Fisher, A. Cousido-Siah, A. G. salvay, A. Popov, C. Muller-Dieckmann, T. Petrova and A. Podjarny, *J. Mol. Recognit.* **24**, 724-732 (2011).

[23] M. P. Blakeley, F. Ruiz, R. Cachau, I. Hazemann, F. Meilleur, A. Mitschler, S. Ginell, P. afonine, O. N. Ventura, A. Cousido-Siah, M. Haertlein, A. Joachimiak, D. Myles and A. Podjarny, *Proc. Natl. Acad. Sci. U. S. A.* **105**, 1844-1848 (2008).

[24] D. A. A. Myles, F. Dauvergne, M. P. Blakeley and F. Meilleur, *Journal of Applied Crystallography* **45**, (2012).

[25] M. P. Blakeley, A. J. Kalb, J. R. Helliwell and D. A. Myles, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16405-16410 (2004).

[26] G. Palmer in *The Porphyrins* Vol. IV, Academic Press Inc. (1979).

[27] H. Schober, *Journal of Neutron Research* **17**, 109-357 (2014).

[28] G. L. Sqires, *Introduction to the theory of thermal neutron scattering* Dover Publications (1978).

[29] E. Noether *Nachr. D. Knig. Gesellsch. D. Wiss. Zu Gttingen, Math-phys. Klasse*, 235257 (1918).

[30] R. Santra, *Journal of Physics B* **42**, 023001 (2009).

[31] P.A.M. Dirac, *Proceedings of the Royal Society A* **114**, 243265 (1927).

[32] V. F. Sears, *Neutron Optics* Oxford University Press (1989).

[33] C. Kittel, *Introduction to Solid State Physics* New York: John Wiley & Sons (2005).

[34] D. Sherwood and J. Cooper, *Crystals, X-rays and proteins: comprehensive protein crystallography* Oxford University Press (2010).

[35] D. S. Sivia, *Elementary scattering theory for X-ray and neutron users* Oxford University Press (2011).

# Part II

# Technical aspects in neutron macromolecular crystallography

# Chapter 3

# The LADI-III diffractometer and the low temperature sample environment

This chapter describes the quasi-Laue neutron macromolecular diffractometer LADI-III at the Institut Laue-Langevin (Grenoble). The general characteristics of the instrument are presented in the first part while the second part describes the low temperature sample environment that was developed at LADI-III as part of the present research project.

## 3.1 The quasi-Laue neutron diffractometer LADI-III

The neutron diffractometer LADI-III is an instrument devoted to macromolecular crystallography at the high flux neutron source of the Institut Laue-Langevin (ILL). Figure 3.1 is a photograph of the instrument while a technical drawing of the machine is shown in Figure 3.2. This section presents an introduction to the production of neutron beams at the ILL reactor source and describes the main components of the instrument LADI-III. The final part of this section presents a discussion of the specifics of the quasi-Laue

technique adopted at LADI-III.

Figure 3.1: LADI-III neutron diffractometer for macromolecular crystallography at the neutron source of the Institut Laue-Langevin (Grenoble) showing the cylindrical detector.

**Production of neutrons** The Institut Laue-Langevin reseach reactor produces a steady state thermal neutron flux of $10^{15}$ neutrons $cm^{-2}$ $s^{-1}$ at a distance of 40 cm from the reactor core by $^{235}U$ fission in a highly enriched fuel element. Neutrons are $D_2O$ moderated. Collisions with the moderator give rise to a Maxwell-Boltzmann distribution of the velocity modulus of neutrons with mean velocity depending on the moderator temperature. A cold source constituted of boiling $D_2$ at the temperature of 25 K is installed close to the reactor core. Collisions with the fluid in the cold source produce neutrons with wavelengths above approximately 3 Å.

**Transport of neutrons** Neutron transport is based on the phenomenon of total external reflection. A neutron guide is a hollow tube, commonly made of nickel. As for electromagnetic radiation, the refractive index $n$ of neutrons in a medium is defined as the ratio $k/k_0$ where $k$ and $k_0$ are the propagation vectors in the medium and in vacuum

respectively. Unlike electromagnetic radiation, the refractive index of neutrons in most materials is smaller than 1. This gives rise to the phenomenon of total external reflection at the vacuum/medium interface when the angle between the wavevector of incident neutrons and the interface surface is smaller than a critical angle $\theta_c$. The critical angle depends on the medium and on the neutron wavelength and amounts to $\theta_c/\lambda = 0.1°/\text{Å}$ in the case of nickel. A nickel guide therefore tolerates a beam divergence of $\pm 0.1°/\text{Å}$.

Guides accepting larger divergence can be produced by using alternating layers of nickel and titanium. Nickel and titanium present a large contrast of refractive indexes and nickel/titanium multilayers show an artificial periodicity in the refractive index. Layer thickness is set so to give rise to a Bragg peak at $\theta > \theta_c$ therefore extending the angular range of reflectivity of the device above the natural critical angle of nickel. Further extension of the angular range is achieved by producing nickel/titanium multilayers with layers of variable thickness.

The diffractometer LADI-III is located at the end of the cold neutrons guide H143. This guide is characterized by a critical angle of $0.1°/\text{Å}$, corresponding to a beam divergence of $\pm 0.35°$ for 3.5 Å neutrons. The H143 guide is composed of a straight section close to the reactor core and a curved section close to the instrument. The latter has a curvature radius of 2000 m and is used to reduce the number of $\gamma$ photons and fast neutrons that impinge on the detector coming from the reactor core. Since the critical angle is directly proportional to the neutron wavelength, most fast neutrons are not guided through the curved guide and do not reach the experimental set-up.

**The wavelength broad band filter** The peculiarity of the macromolecular neutron diffractometer LADI-III is that a quasi-Laue configuration is used. Quasi-Laue techniques are diffraction data collection methods that involve the use of a polychromatic beam produced by selecting a wavelength range from the spectral distribution of the source. This method presents advantages with respect to monochromatic data collection and with respect to Laue techniques which make use of the full spectral distribution of the source. The last part of this section presents a discussion of the characteristics of quasi-Laue

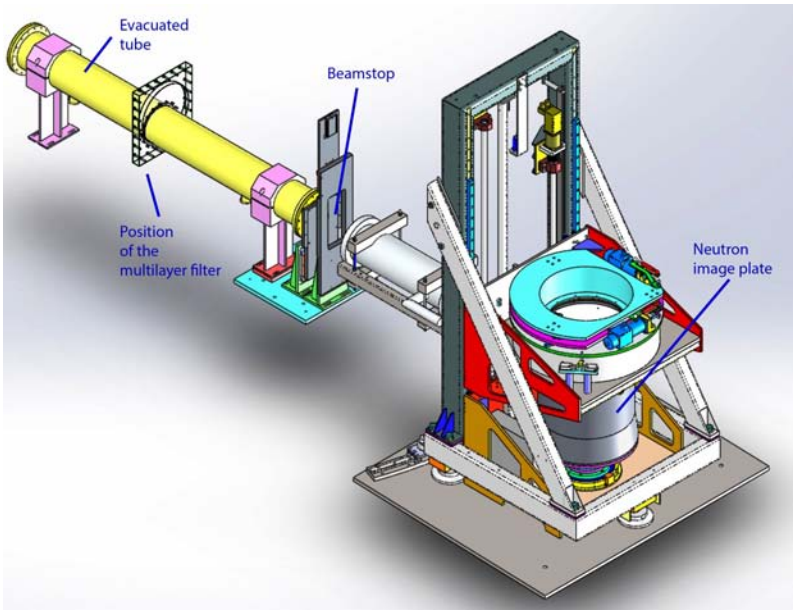


Figure 3.2: Technical drawing of the neutron macromolecular diffractometer LADI-III at the Institut Laue Langevin (Grenoble) including (from left to right) the vacuum tube, the wavelength filter position, the direct beam stop, the collimation elements and the cylindrical detector.

techniques.

At the LADI-III beamline a quasi-Laue neutron beam is produced by means of a nickel/titanium multilayer filter. The multilayer reflector is composed of 40 silicon crystal mirrors, each with 748 sets of alternating nickel and titanium layers with variable thickness between 74 Å and 90 Å [1]. According to [2] the wavelength range selected by such device depends on the beam divergence and layers with variable thickness are used to increase the acceptance angle of the filter.

The neutron beam propagates in evacuated tubes towards the sample position and the beam diameter is progressively reduced by using neutron absorber screens with circular apertures of progressively smaller diameter. Lithium ($^6$Li) fluoride is used as a neutron absorber for this purpose. The collimation screen located closer to the sample position is removable allowing the use of different aperture diameters comprised between 0.5 and 2.9 mm. It is advantageous to match the beam diameter with the sample size. This allows to illuminate the whole sample volume while keeping the background levels on the detector

to a minimum.

The wavelength range at the sample position is characterized by $\Delta\lambda/\lambda \sim 25\%$ and the range can be shifted from central wavelength of 3.5 Å to 5.3 Å by varying the angle between the incoming neutrons and the surface of the multilayer filter, which is typically of the order of $1^\circ - 2^\circ$.

The flexibility of the instrument in terms of wavelength range is a characteristic of paramount importance. As explained at the end of this section in case of very large unit cells the use of longer wavelengths is preferable in order to reduce the number of spatially overlapped reflections on the detector. On the other hand in the case of small unit cells and well diffracting crystals the use of shorter wavelengths is more advisable with the purpose of extending the resolution limit of the data collection.

**The neutron image plate detector** The LADI-III diffractometer presents a cylindrical position sensitive detector with radius of 200 mm and height of 450 mm. A Fujifilm neutron image plate is used for neutron detection and is placed in the interior of the cylindrical drum that sorrounds the sample. This type of neutron detector has been extensively characterized and described [3]. The basic principle of operation is summarized hereafter.

The neutron image plate consists of a phosphor layer with thickness of the order of $10^2\mu$m containing a neutron converter, a storage phosphor and an organic binder, coated on a polymer film. Upon neutron exposure the neutron converter gadolinium in the form of $Gd_2O_3$ releases secondary radiation. Conversion electrons in the energy range between 30 and 200 keV and $\gamma$ photons are released. Although the storage phosphor is sensitive to both electrons and $\gamma$ rays, conversion electrons have the dominant effect in these circumstances. Secondary electrons are adsorbed in the storage phosphor composed of BaFBr doped with europium ions $Eu^{2+}$. Subsequently to secondary radiation absorption, electron-hole pairs are generated. Electrons are trapped in halide vacancies forming $F$-centers or colour centers. Holes are stored in proximity of the dopant ion. $F$-centers accumulate during neutron exposure with an average number of 640 $F$-centers generated per incident neutron.

After neutron exposure the plate is read by using a helium/neon laser with wavelength

of 632.8 nm. The laser system employed at LADI-III operates with power set at 70 mW. The plate is scanned in phonographic mode by translating the read-head vertically while the drum rotates about its axis. The pixel size is user selected at the read-out stage, possible values are $125^2$, $250^2$ or $500^2$ $\mu$m$^2$. During photostimulation an $F$-center is excited from the $1s$ ground state to the $2p$ excited state. Electron-hole pairs recombine and the released energy is transferred to the dopant ion which emits energy through its characteristic luminescence at 390 nm accompanying the $5d$ to $4f$ transition. The photostimulated light is collected at the photocathode of a photomultiplier and converted into photoelectrons. The photoelectrons are accelerated through a system of dynodes with applied voltage of 650 V and the amplified current is detected by the data acquisition system.

The position sensitive neutron image plate is characterized by a good spatial resolution of approximately 100 $\mu$m [3], the dynamic range extends over five orders of magnitude [4] and the detective quantum efficiency defined as the ratio between the square of the signal-to-noise ratio on output and the square of the signal-to-noise ratio on input is 46 % [5].

**Considerations about the quasi-Laue technique** Quasi-Laue data collection presents advantages with respect to monochromatic techniques and with respect to the use of the full spectral distribution of the source that characterizes Laue methods. The use of a polychromatic neutron beam allows a more rapid survey of reciprocal space compared to the monochromatic scenario since a large number of reflections are simultaneously stimulated over all incident wavelengths.

The quasi-Laue configuration represents a two-fold advantage compared to the employment of the full spectral range of the source. First the number of stimulated reflections is reduced leading to a smaller number of spatially overlapped diffraction spots. Moreover the signal to noise ratio is improved. In fact each reflection is stimulated at a particular wavelength (or by a narrow range of wavelengths) and sits upon an incoherent background signal that accumulates over all wavelengths in the incident beam [6]. Hence the selection of a narrower wavelength range reduces the incoherent background signal on the detector.

**Considerations about the use of cold neutrons** The use of long wavelengths *i.e.* cold neutrons instead of thermal ones is advantageous since the observed intensity of Laue reflections is directly proportional to the wavelength squared [7]. In addition the density of reflections on the detector is proportional to the unit cell volume [1] and to the difference between the inverse cubes of the wavelength limits employed [8]. By adopting long wavelengths the number of stimulated reflections decreases together with the number of spatially overlapped spots on the detector.

## 3.2 The low temperature experimental set-up at LADI-III

Cold neutron exposure does not produce observable radiation damage in biomolecular crystals. Hence neutron diffraction data are routinely collected at ambient temperature unlike X-ray crystallography analyses which are generally performed at cryogenic temperatures to limit photodamage by reducing the reactivity of photogenerated free radicals. Ambient temperature structural determinations allow the study of biomolecules in conditions that are close to physiological ones avoiding the insurgence of cryo-induced artifacts. Nevertheless data collection at cryogenic temperature presents a number of advantages and noteworthy applications.

Reduced thermal motion can lead to increased data quality and increased number of ordered solvent molecules in the protein structure [9]. In addition cryogenic temperature experiments allow the study of cryo-trapped reaction intermediates [10] (Chapter 9), the investigation of the temperature dependence of structural features in macromolecules (Chapter 12) and more generally the study of species and complexes that are not stable at ambient temperature. Not least, since most X-ray crystallography studies are performed at 100 K, neutron data collected at the same temperature allow direct comparison of structures and the employment of joint refinement procedures (Chapter 5). The possibility of carrying out diffraction experiments in cryogenic conditions in addition to routine ambient

[1] The moduli of reciprocal lattice basis vectors are inversely proportional to the direct lattice ones. Hence in the case of large unit cell biomolecular crystals reciprocal lattice points are tightly spaced.

temperature studies therefore represents a major improvement of a neutron macromolecular diffractometer.

As part of the present research project a low temperature sample environment was developed at the instrument LADI-III by customization of the commercial Cobra - Non Liquid Nitrogen Open Flow System (Oxford Cryosystems) and in particular of the beamline sample holder with S. Martinez (Institut Laue-Langevin).

In the past low temperature studies were carried out at LADI-I (the previous version of the LADI-III diffractometer) by using a liquid helium closed-cycle cryorefrigerator at the temperature of 15 K [11]. The 15 K neutron structures of lysozyme [12], [13], concanavalin A [9] and rubredoxin [13] were determined. Protein crystals were cryo-cooled by means of a two-step procedure. First the cryo-protected specimen was flash-cooled at 77 K by rapid immersion in liquid nitrogen. In the second step it was transferred to the cold head of the cryorefrigerator with temperature at the sample position of 15 K.

It was extremely challenging to successfully transfer the delicate samples by using the 15 K displex set-up and a wider use of low temperature data collection protocols necessitated the installation of a new low temperature apparatus allowing a safer sample transfer to the designated position. A user-friendly low temperature set up was adapted to the LADI-III beamline, installed, tested and used in data collection as part of the present project.

A Cobra - Non Liquid Nitrogen Open Flow System from Oxford Cryosystems was employed. The instrument is capable of cooling samples to 80 K in an open stream of dry nitrogen and heating them up to 400 K. Figure 3.3 presents a scheme of the instrument. A nitrogen generator purifies nitrogen from air producing a nitrogen stream at ambient temperature with purity above 97%. Cooling is achieved by cyclically compressing and expanding helium in a closed cycle cooler. A water-cooled helium compressor compresses the gas that is then expanded in the refrigerator. The nitrogen gas is cooled by passing it through heat exchangers. Helium expansion subtracts heat from the nitrogen gas. The latter is delivered at the sample position by the instrument coldhead shown in Figure 3.4. Expanded helium is returned to the compressor and recycled.

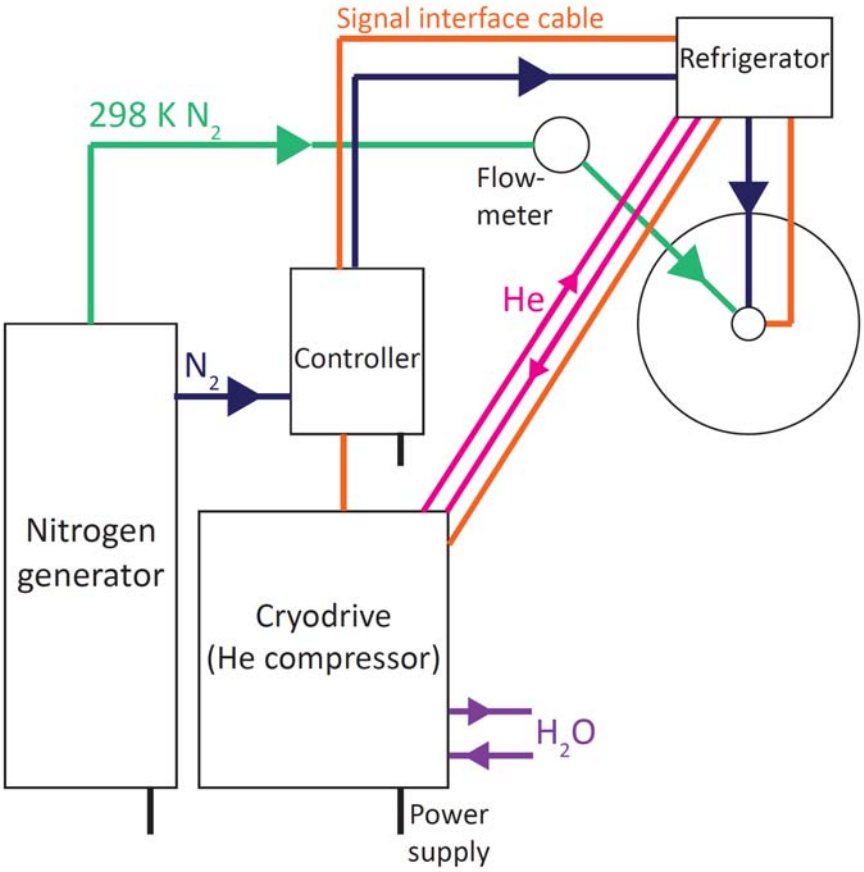


Figure 3.3: Layout of the Cobra Non Liquid Nitrogen Open Flow System (Oxford Cryosystem) installed on the LADI-III instrument. Colour code: blue, nitrogen to be cooled; green, nitrogen at 298 K, magenta, helium; violet, cooling water; orange, signal interace cables; black, power supply.

Installation of the commercial version of the Cobra cryosystem required adaptation of the sample holder of the LADI-III instrument. A new sample holder was produced by S. Martinez as shown in Figure 3.4 and 3.5. A small motor was mounted below the goniometerhead to drive the sample rotation about the spindle axis. A standard goniometerhead (Huber, model 1005) was used, compatible with standard X-ray crystallography sample mounts. It is possible to cryo-cool the sample *in situ* by blocking the nitrogen stream while positioning the crystallography pin on the goniometerhead. It is also possible to transfer pre-cooled crystals to the instrument. Both methodologies were successfully employed

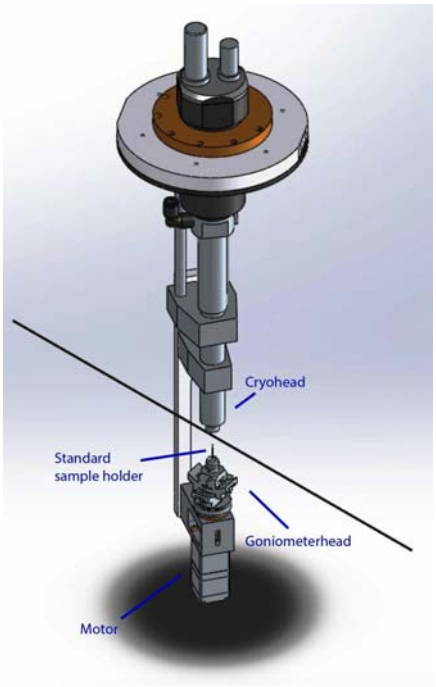


Figure 3.4: The customized sample holder for the insertion of the Cobra set-up cryohead in the LADI-III neutron diffractometer. The figure shows the aluminium support, the goniometerhead and a small motor which drives the sample rotation about the spindle axis positioned below the goniometer.

within the present work. Recovery of the crystal after neutron exposure for subsequent X-ray data collection is also possible.

The sample holder is pre-aligned off-line by means of the same alignment set-up used in ambient temperature experiments [2]. The sample position is optimized by inspecting the alignment *in situ* with a telescope.

The temperature distribution in the region close to the coldhead was measured to determine the best sample position. It was observed that when the refrigerator temperature is set at 100 K, the temperature at the sample position is below 110 K at distances from the coldhead smaller than 20 mm as shown in Figure 3.6 (right). It was also observed

[2]The off-line set-up is prepared based on the known beam position determined by exposing for one second a strong neutron absorber, usually a $B_4C$ sample of small size and observing the image produced on the detector.

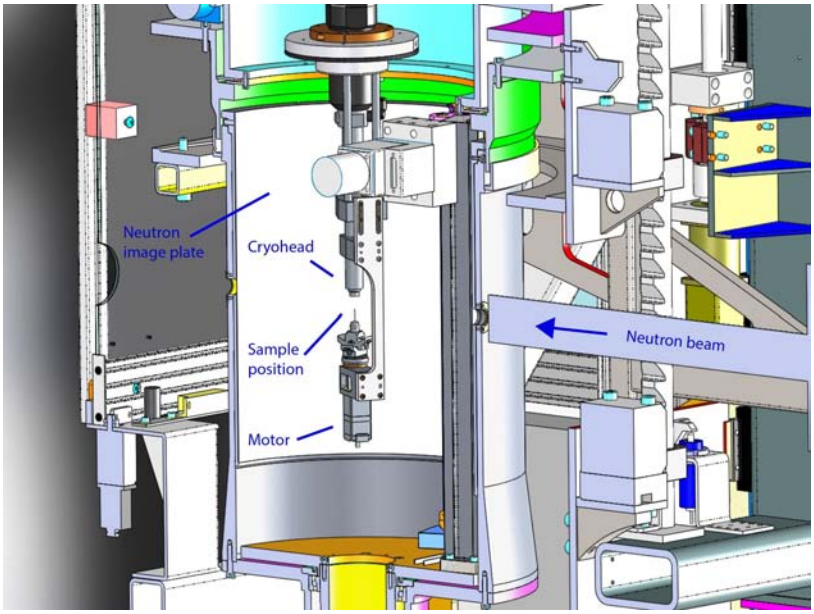


Figure 3.5: Technical drawing showing a section of the LADI-III diffractometer. The sample holder was adapted to allow the insertion of the Cobra set-up coldhead.

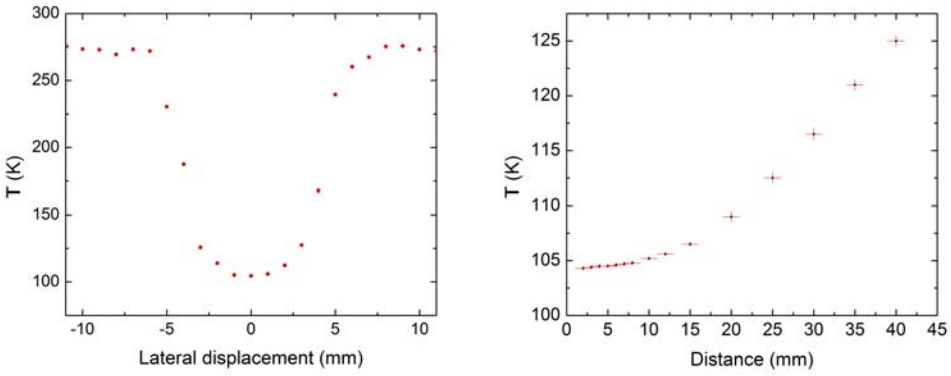


Figure 3.6: Left: temperature as a function of lateral displacement at a sample positioned 5 mm below the Cobra cryohead. Right: temperature as a function of distance from the Cobra cryohead. In both measurements the refrigerator temperature was set to 100 K.

that the temperature varies greatly with lateral displacement. Figure 3.6 (left) shows that when the sample is positioned 5 mm below the coldhead, the maximum lateral displacement that it can undergo without considerable increase of the temperature amounts to a few millimeters.

## 3.3 Concluding remarks

Compared to synchrotron X-ray fluxes, neutron fluxes available at sources worldwide are smaller by many orders of magnitude. This lead to the requirement for protein crystals with large volume of several $mm^3$ in early days neutron crystallography experiments. The growth of large protein crystals has traditionally been a bottleneck in neutron macromolecular crystallography limiting the range of applicability of this technique.

Recent advances in neutron instrumentation and detectors, in sample preparation with the production of fully deuterated macromolecules and in structural refinement software with the development of X-ray and neutron joint refinement procedures, have broadened the range of biological problems that can be tackled by neutron macromolecular diffraction.

The quasi-Laue diffractometer LADI-III at the Institut Laue-Langevin provides a high neutron flux at the sample position of $10^8$ neutrons $cm^{-2}$ $s^{-1}$ and is equipped with a large position sensitive detector allowing the structural characterization of challenging samples with large unit cells up to 150 Å on cell edge or small crystal volumes of 0.2 $mm^3$ in the case of partially deuterated samples and 0.05 $mm^3$ in the case of fully deuterated samples.

The Cobra - Non Liquid Nitrogen Open Flow System low-temperature set-up allows structural analyses at 100 K with a user-friendly sample mounting stage broadening the range of application of the technique and in particular allowing the cryo-trapping of reaction intermediates.

# Chapter 4

# Quasi-Laue neutron data processing

The present chapter provides a description of the various stages involved in polychromatic neutron diffraction data processing. A concise presentation of quasi-Laue neutron data collection at the macromolecular diffractometer LADI-III is reported followed by the description of the processing steps required in the treatment of polychromatic data namely the determination and refinement of the single crystal orientation, the integration of diffraction spots, the scaling and merging of reflections. This non-exhaustive description focuses on data processing aspects and software features that are particularly relevant in neutron crystallography work presented in Part III.

## 4.1 Quasi-Laue neutron data collection

The neutron macromolecular diffractometer LADI-III at the Institut Laue-Langevin high flux reactor is an instrument devoted to quasi-Laue neutron data collection on single crystals of biomolecules. The characteristics of this diffractometer are presented in Chapter 3 of the present work. The peculiarity of LADI-III is that a polychromatic neutron beam is used so that a large number of reflections can be recorded simultaneously.

The single crystal is held stationary during neutron exposure. The crystal is rotated

about the spindle axis and an image is collected at each orientation with rotation steps generally of 7°. Knowledge of the symmetry properties of the reciprocal lattice is used to determine the appropriate rotation range about the spindle axis. After exploring such rotation range the crystal is tilted changing the orientation of the crystallographic axes with respect to the spindle axis and rotation is performed about the spindle axis with the purpose of improving the reciprocal space sampling. The tilting procedure can be repeated more than once to maximize reciprocal space coverage.

## 4.2 Indexing and integration of quasi-Laue neutron diffraction data

Quasi-Laue diffraction images are individually processed by means of the software *LAUEGEN* [14], [15], [16]. *LAUEGEN* is used in the initial steps of polychromatic data processing namely data indexing, refinement of orientation parameters and diffraction spots integration.

**The Laue Data Module parameters file** *LAUEGEN* works with a set of parameters associated with individual diffraction images. Initial values are read in from a Laue Data Module (*LDM*) keyworded parameters file [17] which defines the image file to be processed and the values of the associated parameters. The *LDM* file includes parameters related to the crystal such as the unit cell parameters, the orientation and the resolution limit of diffraction; parameters describing the experimental set-up such as the wavelength range and the crystal to detector distance as well as parameters concerning images and diffraction spots such as pixel size, spot size and distortion parameters.

**Indexing of diffraction images** *LAUEGEN* incorporates algorithms for the indexing of Laue diffraction images, namely the determination of the orientation of the sample's crystallographic axes with respect to the laboratory frame of reference. As derived in Section 2.6.2 diffracted intensity appears as the transferred wavevector associated to the scattered neutron beam equals a reciprocal lattice vector. Therefore each diffraction spot

can be assigned to the set of reciprocal lattice indexes $h\ k\ l$ characterizing the corresponding reciprocal lattice vector, for which purpose knowledge of the crystal orientation is required. Automatic indexing is based on the method presented in [18] and requires knowledge of unit cell parameters, crystal to detector distance and position of the direct neutron beam. Cell dimensions determined by X-ray diffraction from the same crystal or from an isomorphous crystal are used. The *Find Orientation* menu in *LAUEGEN* graphical user interface allows the user to select low index spots directly on the diffraction image to be used by the autoindexing routine. The indexing algorithm implemented in *LAUEGEN* uses detector coordinates of low index spots provided by the user to calculate the corresponding reciprocal lattice vectors and the angles between them. Angles between reciprocal lattice vectors are matched against a table of angles calculated in advance based on lattice parameters. This allows to assign indexes to the spots being considered and to determine the orientation of the lattice with respect to the laboratory frame.

**Laue pattern simulations** *LAUEGEN* calculates the expected spot positions in the polychromatic diffraction pattern based on the values assigned to the relevant *LDM* parameters *i.e.* the cell parameters, the crystal to detector distance, the wavelength range, the diffraction limit and the crystal orientation. The predicted pattern can be visualized by using the *Laue Simulations* menu. This option allows to display different properties of the diffraction spots. For illustration purposes Laue simulations obtained in cytochrome *c* peroxidase neutron data processing are presented hereafter. Figure 4.1 shows the Laue pattern predicted by assuming a wavelength range between 3 and 4 Å and diffraction extent of 2.2 Å. Figure 4.2 shows the same pattern where only spots flagged as spatially overlapped are displayed. It is worthwhile observing that the majority of the spots are considered as spatially overlapped. The accurate integration of such spots is therefore particularly important. Figure 4.3 highlights multiple spots namely those spots which consist of harmonic overlaps. In addition to colour coded Laue simulations, interactive simulations are available in *LAUEGEN* which allow to observe the effect of varying the soft limits namely the wavelength range and the diffraction extent as shown in Figure

4.4. With regard to this figure, it is worthwhile stressing that the number of spatially overlapped spots increases dramatically with the width of the wavelength range employed in the data collection hindering the use of the white neutron beam. Figure 4.5 highlights multiple reflections, analogous to Figure 4.3, focusing for the sake of illustration on the multiplicity information of the nodal 0 0 -1 resulting from the superposition of 12 harmonic reflections.

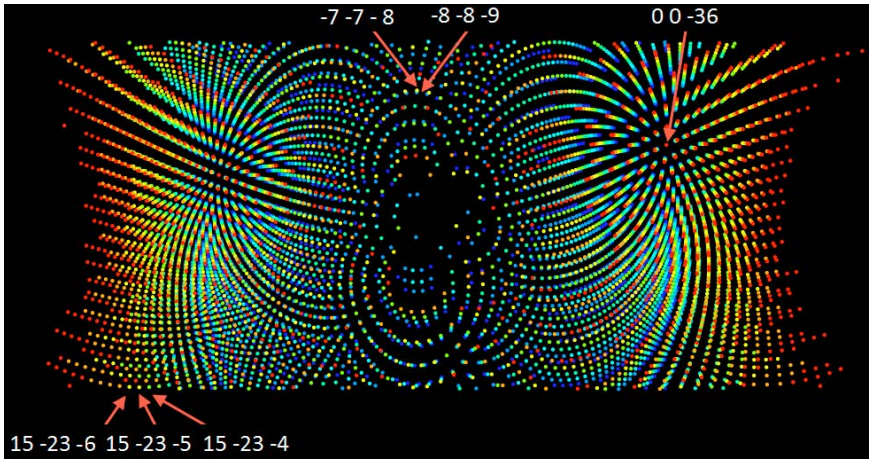


Figure 4.1: Laue diffraction pattern simulation from *LAUEGEN* for cytochrome *c* peroxidase by assuming neutron wavelengths in the range from 3 to 4 Å and diffraction extent of 2.2 Å. Spots are colour coded by wavelength with red corresponding to 4 Å and blue corresponding to 3 Å. Multiple spots (harmonic overlaps) are colour coded according to their highest wavelength component. The reciprocal space indexes of some spots have been reported for the sake of illustration.

**Orientation refinement** Once the image has been successfully indexed, refinement of the crystal orientation prior to spots integration is of paramount importance for improving the accuracy of integration. Crystal orientation refinement is achieved by using the *Match Spots and Refine* or the *Nodals Search and Refine* options and by carrying out refinement cycles iteratively until the Laue prediction matches the observed diffraction pattern. In the first option the user selects directly on the image pairs of predicted and observed spot positions to be used in orientation refinement. In the second option the user assigns to the nodal spots selection index an integer value comprised between 1 and 12. The

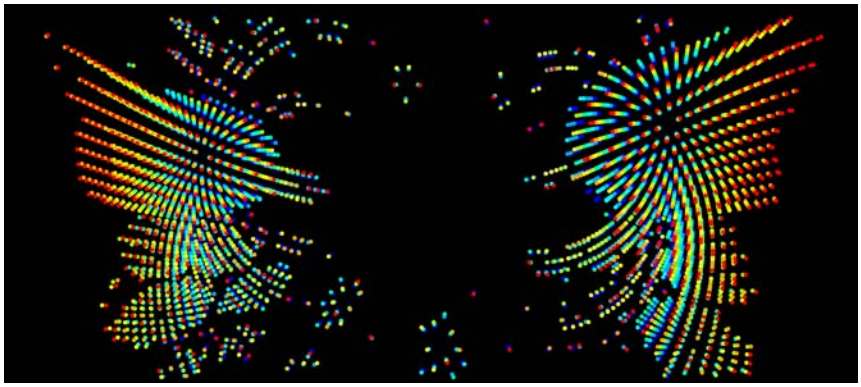

Figure 4.2: Laue diffraction pattern simulation from *LAUEGEN* for cytochrome *c* peroxidase by assuming neutron wavelengths in the range from 3 to 4 Å, diffraction extent of 2.2 Å and spot size of 2 mm by 2 mm. Only spots flagged as spatially overlapped are shown. These spots are integrated provided that they are not closer than a user defined value $\epsilon$. Spots are colour coded by wavelength with red corresponding to 4 Å and blue corresponding to 3 Å. Multiple spots (harmonic overlaps) are colour coded according to their highest wavelength component.

program considers those spots whose absolute values of $h$ $k$ $l$ indexes after division by their highest common factor do not exceed the selected nodal index value. Spatially overlapped spots *i.e.* spots corresponding in general to different indexes and stimulated by different wavelengths which happen to appear close to each other on the detector, are not considered for refinement purposes. The software searches for observed spots in an ellipse surrounding the predicted spot position. Localization of the observed spot position is carried out by considering pixels with an intensity value greater than a set threshold value above the local background. In both orientation refinement options the initial root mean square deviation between predicted and observed spots positions is calculated. Non linear least squares refinement is carried out and the parameters to be refined are user selected. These include the orientation of the crystal, the crystal to detector distance and distortion parameters. Unit cell parameters are normally kept fixed if the values determined by X-ray diffraction from the same crystal are used. The post-refinement value of the root mean square deviation is displayed. Refinement runs can be repeated by increasing the nodal spots selection index so that a larger number of spots is included

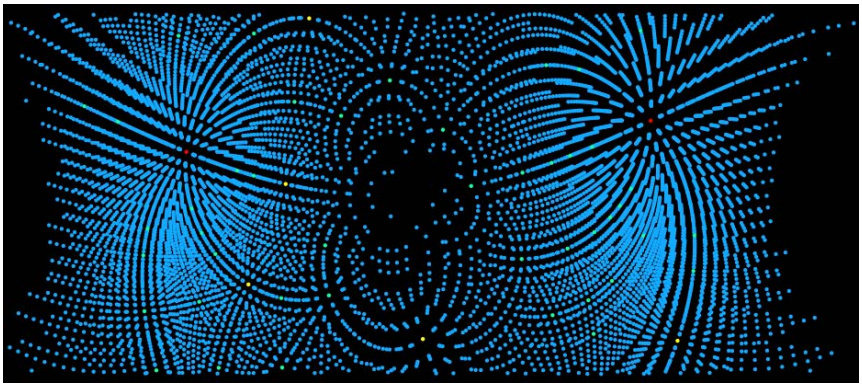

Figure 4.3: Laue diffraction pattern simulation from *LAUEGEN* for cytochrome *c* peroxidase by assuming neutron wavelengths in the range from 3 to 4 Å and diffraction extent of 2.2 Å. Spots are colour coded according to their multiplicity i.e. number of harmonic overlaps. Single spots are shown in blue. Multiplicity is equal to 2 for spots in cyan, yellow spots are triple and red spots show multiplicity higher than 3.

until a good agreement between the predicted and the observed pattern is obtained.

**Diffraction spots integration** Spots integration is carried out by profile fitting. For this purpose the image is divided into a maximum of 17 radial bins. The integration procedure involves background plane constants calculation, determination of an elliptical integration profile per bin, profile fitting to determine the integrated intensity of every diffaction spot and evaluation of the standard deviation associated to each integrated intensity. The details of the procedure are reported in [15] and [19]. For each image a corresponding intensities file in *MTZ* format is produced. This file reports information for each diffraction spots comprising the $h\ k\ l$ indexes, the spot coordinates on the detector, the integrated intensity and its associated error as well as the wavelength at with the reflection was stimulated, the image number, information about the spot's multiplicity and spatial overlap flag.

**Spot overlaps** Laue images present spatially overlapped and harmonic overlapped spots. Spots that are closer than a user determined threshold value $\delta$ are considered as spatially overlapped spots. These spots are integrated provided that they are not closer than the

user defined value $\epsilon$. Harmonic overlaps *i.e.* energy overlapped spots or multiple reflections are those reflections which are stimulated at different wavelengths and correspond to parallel reciprocal lattice vectors, which accumulate on the same point on the detector. Nodal reflections (see definition above) are often multiple reflections and are important in data indexing. Harmonic overlap deconvolution is not performed at the data integration stage. Multiple spots constitute less than 1% of the spots [20], they are integrated without deconvolution and are discarded during the scaling stage.

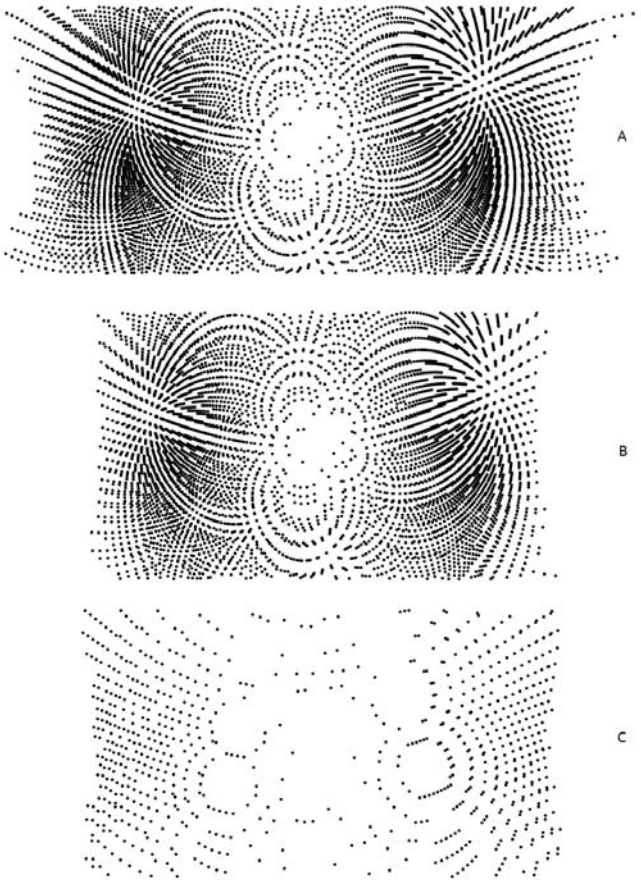


Figure 4.4: Laue diffraction pattern simulation from *LAUEGEN* for cytochrome *c* peroxidase showing the effect of soft limits variations. (A) Wavelength range from 3 to 4 Å and diffraction extent of 2.2 Å. (B) Wavelength range from 3 to 4 Å and diffraction extent of 2.5 Å. (C) Wavelength range from 3.5 to 3.6 Å and diffraction extent of 2.2 Å.

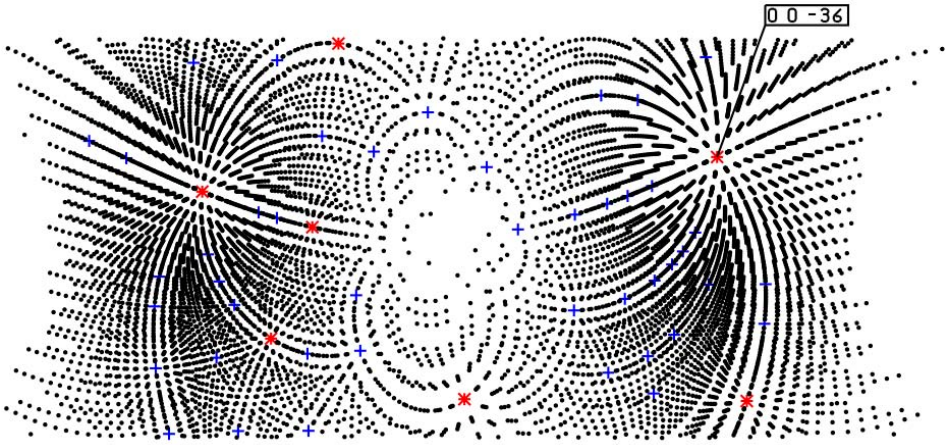


Figure 4.5: Laue diffraction pattern simulation from *LAUEGEN* for cytochrome *c* peroxidase by assuming neutron wavelengths in the range from 3 to 4 Å and diffraction extent of 2.2 Å. Spots are coded according to their multiplicity i.e. number of harmonic overlaps. Doubles are shown in blue and higher multiples are red. For illustration purposes a multiple spot is highlighted. It presents nodal indexes (0 0 -1) and multiplicity of 12 with minimum indexes (0 0 -47) and maximum indexes (0 0 -36).

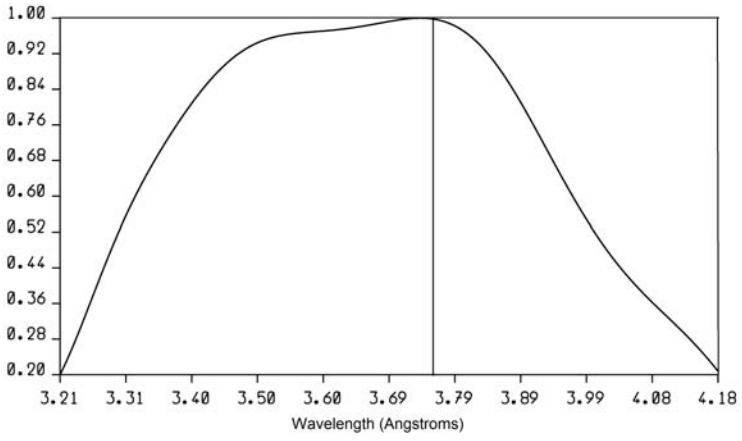


Figure 4.6: Wavelength normalization curve derived by *LSCALE*. Inverse of the wavelength normalization factor as a function of wavelength.

## 4.3 Scaling and merging of quasi-Laue neutron crystallography data

The polychromatic data processing software *LAUEGEN* produces as a result of spots integration an output file containing unscaled and unmerged reflection intensities and the associated error. Since a quasi-Laue neutron beam is used reflection intensities are affected by the wavelength distribution of the incident flux. Hence wavelength normalization is of paramount importance for the accurate determination of intensities. This section describes data scaling and merging steps in quasi-Laue data processing and defines the statistical indicators used in Part III of the present work.

**Wavelength normalization** Scaling of the integrated intensities is performed in order to correct for fluctuations in the integrated intensities due to incident flux variations and in particular to the spectral distribution of the incident neutron beam. Scaling of polychromatic data is achieved by means of the software *LSCALE* [21]. The wavelength normalization curve is derived using the intensities of symmetry equivalent reflections recorded at different wavelengths. Figure 4.6 presents the wavelength normalization curve namely the inverse of the wavelength scaling factor as a function of wavelength. This curve reflects the spectral distribution of the incident neutron beam.

**Reflections merging and statistical indicators** Scaled reflections are merged by using the software *SCALA* [22] which outputs an *MTZ* format file containing scaled and merged reflection intensities and the associated error. *SCALA* also calculates data collection and processing statistical indicators. Some of these indicators are particularly informative on data quality and are reported in data collection tables in Part III of this work. Data quality indicators include the average ratio between the measured spot intensity and its associated error as well as the completeness of the collected data namely the fraction of reciprocal space sampled given the resolution limits used in data processing. The multiplicity of the collected data is also reported: this is the redundancy with which on average every reflection has been sampled. Finally the $R_{merge}$ value is calculated by

using the following expression:

$$R_{merge} = \frac{\sum_h \sum_j |I_{hj} - \langle I_h \rangle|}{\sum_h \sum_j I_{hj}}$$

where $I_{hj}$ is the scaled intensity related to the $j^{th}$ observation of reflection $h$, $\langle I_h \rangle$ is the mean intensity of reflection $h$ and sum is performed over all observations $j$ of reflection $h$ and over all reflections $h$. $R_{merge}$ indicates the extent of the agreement between measurements of equivalent reflections. Finally scaled and merged intensities are converted into structure factor amplitudes and associated error by means of the program *TRUNCATE* [23] which applies Wilson statistics for the handling of negative intensity observations.

## 4.4 Concluding remarks

Quasi-Laue data processing involves the indexing and integration of polychromatic reflections, the scaling and merging of observed intensities and the conversion of the integrated intensities into structure factors amplitudes.

Indexing, integration and scaling of quasi-Laue data present peculiarities that are specific to polychromatic diffraction experiments and are summarized hereafter.

- Unlike the monochromatic case, polychromatic data indexing requires cell parameters to be provided as an input.

- The use of a polychromatic beam allows the detection of a large number of reflections per image compared to the monochromatic pattern. However the need to deal with spatially overlapped diffraction spots *i.e.* spots corresponding in general to different indexes and stimulated by different wavelengths which happen to appear close to each other on the detector represents the main difficulty related to this method and is a source of inaccuracy in diffraction spots integration.

- Scaling of polychromatic data involves the determination of the wavelength normalization curve that accounts for the spectral distribution of the incident neutron beam.

The processed neutron diffraction data are used together with X-ray diffraction data in structural model refinement as described in Chapter 5.

# Chapter 5

# X-ray and neutron structural joint refinement

Chapter 4 describes how diffraction images are processed to obtain the set of experimental observations:

$$h \quad k \quad l \quad F_{\text{obs}} \quad \sigma(F_{\text{obs}})$$

where $F_{\text{obs}}$ denotes the experimental value of the structure factor amplitude of reflection $h\,k\,l$ and $\sigma(F_{\text{obs}})$ is its associated error.

Assumption is made that the model of a structure similar to that being studied is available. In these circumstances observed structure factor amplitudes are phased by molecular replacement methods which consist of positioning the structural model in the unit cell by rotation and translation searches with the purpose of matching model and experimental Patterson functions *i.e.* the Fourier transform in real space of reciprocal space intensity. Structure factor phases calculated from the correctly positioned model are assigned to the experimental amplitudes and enter the Fourier transform algorithms for the computation of scattering density maps in real space. The determination of the protein structure of interest is achieved by refining the existing structural model against the experimental data in an iterative fashion.

Cytochrome *c* peroxidase structural refinement involves the use of one of the vari-

ous models available in the Protein Data Bank (www.rcsb.org) as a starting structural model. Phasing by molecular replacement is carried out through a rigid body refinement run in the macromolecular refinement routine *phenix.refine* of the software *PHENIX* [34]. The model of the structure is determined by X-ray and neutron joint refinement which consists of refining model parameters against neutron and X-ray experimental data sets simultaneously.

This chapter introduces some generalities concerning macromolecular structural refinement focusing on the aspects which are of particular relevance in the crystallographic studies presented in Part III of this work. The details of the protocol for X-ray and neutron joint structural refinement employed within this research project are also described.

## 5.1 Generalities on macromolecular structural refinement

Macromolecular structural refinement is obtained by refining model parameters against experimental data. Model parameters include for each atom in the structure real space coordinates $x\ y\ z$, occupancy $n \in [0, 1]$ and $B$-factor $B = 8\pi^2 \langle u^2 \rangle$ where $\langle u^2 \rangle$ is the mean of the atomic displacement squared and accounts for the effects of temperature and static disorder [1]. An accurate model of the molecular structure is generally obtained by cycling global restrained reciprocal space refinement and local real space model corrections.

Restrained reciprocal space refinement consists of modifying real space parameters of the structural model with the purpose of optimizing the agreement between observed and calculated structure factor amplitudes. The totality of the sampled reciprocal space is considered simultaneously and geometrical restraints are introduced to assure reasonable stereochemistry of the polypeptide. Real space model corrections consist of producing local modifications in the model in real space to match electron and neutron scattering density maps computed by using experimental data, by means of graphics software such as *COOT* [25].

[1] With reference to Chapter 2 the Debye-Waller factor in Equation 2.179 corresponds to $\exp(-W) = \exp\left(-B\frac{\sin^2\theta}{\lambda^2}\right)$.

The following paragraphs present a description of the underlying ideas in maximum likelihood structural refinement, provide definitions of the statistical indicators that are used to monitor refinement progress and report information on the computation of electron and neutron scattering density maps.

**Maximum likelihood structural refinement** Structural refinement of cytochrome *c* peroxidase presented in Part III of this work was carried out by using the program *phenix.refine* of the software suite *PHENIX* [24]. This program makes use of maximum likelihood algorithms to perform global reciprocal space restrained refinement.

Maximum likelihood methods are based on Bayes theorem [26] and their employment in crystallography is treated in [27] and reviewed in [28]. In crystallography Bayes theorem establishes the proportionality between the model likelihood function *i.e.* the posterior probability of the model given the observed set of experimental data, and the product between the prior probability of the model and the data likelihood function. Data likelihood is the conditional probability of observing the set of data in the assumption that a given model is valid. In structural refinement the observed structure factor amplitudes represent the experimental data while the parameterized model of the molecular structure allows computation of model structure factor amplitudes and phases.

The purpose in maximum likelihood based refinement methods is the determination of the set of model parameters that corresponds to the minimum of the function:

$$-\log[p(\{F^{\mathrm{calc}}\};\{F^{\mathrm{obs}}\})]$$

where $\{F^{\mathrm{calc}}\}$ is the set of model structure factor amplitudes and $\{F^{\mathrm{obs}}\}$ is the set of experimental amplitudes. The notation $p(\quad;\quad)$ refers to a conditional probability. By assuming independence of reflection amplitudes and independence of stereochemical restraints this corresponds to minimization of:

$$-\sum_{\vec{h}}\log[p(F^{\mathrm{obs}}_{\vec{h}};F^{\mathrm{calc}}_{\vec{h}})]-\sum_{r}\log[p(b^{\mathrm{model}}_{r};b^{\mathrm{target}}_{r})] \tag{5.1}$$

where the first term is the data likelihood function and the sum runs over sampled reflections $\vec{h}$. The second term is the prior probability of the model. It represents prior

stereochemical knowledge in terms of the sum over geometrical restraints $r$ of the logarithm of the probability of the value $b_r^{\text{model}}$ assumed by the stereochemical parameter $b_r$ in the considered structural model given its ideal value $b_r^{\text{target}}$.

**Data likelihood function** The derivation of an explicit expression for the first term of Equation 5.1 is sketched in the following. Amplitude and phase errors in the contribution of individual atoms to the unit cell structure factor are considered. This results *via* the central limit theorem in a two-dimensional Gaussian distribution of complex $\boldsymbol{F}^{\text{obs}}$ about $D\boldsymbol{F}^{\text{calc}}$:

$$p(\boldsymbol{F}^{\text{obs}}; \boldsymbol{F}^{\text{calc}}) = \frac{1}{\pi\sigma_\Delta^2} \exp\left(-\frac{\left|\boldsymbol{F}^{\text{obs}} - D\boldsymbol{F}^{\text{calc}}\right|^2}{\sigma_\Delta^2}\right) \tag{5.2}$$

where complex terms are in bold. The Luzzati factor $D$ [29] is a measure of coordinate error $\vec{\Delta}x$ in the model according to:

$$D = \langle \cos\left(2\pi\vec{h}\vec{\Delta}x\right)\rangle . \tag{5.3}$$

The distribution 5.2 is characterized by a standard deviation $\sigma_\Delta$ that includes the effects of both model incompleteness and model errors.

Upon marginalization of the phase the conditional distribution of observed amplitudes $p(F^{\text{obs}}; F^{\text{calc}})$ is obtained. This non-Gaussian distribution is known as the Rice distribution. Normalized structure factor amplitudes $E$ are generally used and the following expression for the probabilities entering the first term of Equation 5.1 is obtained [30]:

$$p(E_{\text{obs}}; E_{\text{calc}}) = \frac{2E_{\text{obs}}}{1-\sigma_{\text{A}}^2+\sigma_E^2} \exp\left(-\frac{E_{\text{obs}}^2+\sigma_{\text{A}}^2 E_{\text{calc}}^2}{1-\sigma_{\text{A}}^2+\sigma_E^2}\right) I_0\left(\frac{2E_{\text{obs}}\sigma_{\text{A}}E_{\text{calc}}}{1-\sigma_{\text{A}}^2+\sigma_E^2}\right) \tag{5.4}$$

where the variance has been inflated by considering the error $\sigma_E^2$ associated with the experimental normalized structure factor and $I_0$ is the hyperbolic Bessel function of order 0. The distribution 5.4 is described by a single parameter $\sigma_{\text{A}}^2 = 1 - \sigma_\Delta^2$. The parameter $\sigma_{\text{A}}$ includes the effects of model incompleteness and coordinate errors in the model.

$\sigma_{\text{A}}$ can be evaluated as a function of resolution by the methods described in [31] and can be in turn used to evaluate the Sim weight $m$ [32] and the Luzzati factor $D$. The Sim weight accounts for phase errors $\Delta\phi$ in the model:

$$m = \langle \cos(\Delta\phi)\rangle \tag{5.5}$$

while the Luzzati factor describes model coordinate errors as expressed in Equation 5.3. The parameters $m$ and $D$ are of paramount importance for the generation of density maps that minimize model bias as illustrated in the following.

**Prior stereochemical knowledge** Prior knowledge of the stereochemical properties of structural entities is accounted for by the second term in equation 5.1. This term includes geometrical restraints in the maximum likelihood calculations. In the case the stereochemical parameter $b_r$ is normally distributed the probabilities entering the second term of equation 5.1 take the form:

$$p(b_r^{\text{model}}; b_r^{\text{target}}) = \frac{1}{\sqrt{2\pi}\sigma_{b,r}} \exp\left(-\frac{(b_r^{\text{model}} - b_r^{\text{target}})^2}{2\sigma_{b,r}^2}\right) \tag{5.6}$$

where $\sigma_{b,r}$ describes the distribution of the ideal value about its mean $b_r^{\text{target}}$.

**Electron and neutron scattering density maps** Electron density and nuclear scattering density are reconstructed *via* Fourier transformation of the respective complex structure factors (see Equation 2.179). Molecular replacement density maps suffer from severe model bias since model phases are used. Fourier coefficients that minimize model bias were devised based on the consideration of model incompleteness and parameter errors [31]. $\sigma_{\text{A}}$-weighted density maps are given by:

$$\rho(x,y,z) = \sum_h \sum_k \sum_l (2mF_{hkl}^{\text{obs}} - DF_{hkl}^{\text{calc}}) e^{i\phi_{hkl}^{\text{calc}}} e^{-2\pi i(hx+ky+lz)}. \tag{5.7}$$

Both the Sim weight $m$ and the Luzzati factor $D$ are estimated from the parameter $\sigma_{\text{A}}$ that can be evaluated as a function of resolution by the methods described in [31]. Equation 5.7 defines the $\sigma_{\text{A}}$-weighted $2\text{F}_{\text{o}} - \text{F}_{\text{c}}$ map. The following relationship has been established [31]:

$$F^{\text{obs}} e^{i\phi^{\text{obs}}} \sim (2mF^{\text{obs}} - DF^{\text{calc}}) e^{i\phi^{\text{calc}}} \tag{5.8}$$

showing that the $\sigma_{\text{A}}$-weighted $2\text{F}_{\text{o}} - \text{F}_{\text{c}}$ map presents density peaks corresponding to features in the experimental density. The corresponding difference map is:

$$\rho(x,y,z) = \sum_h \sum_k \sum_l 2(mF_{hkl}^{\text{obs}} - DF_{hkl}^{\text{calc}}) e^{i\phi^{\text{calc}}} e^{-2\pi i(hx+ky+lz)}. \tag{5.9}$$

Equation 5.9 defines the $\sigma_A$-weighted $F_o - F_c$ map or difference map. X-ray $\sigma_A$-weighted $F_o - F_c$ maps present positive peaks in correspondence of missing scatterers in the model and negative density corresponding to scatterers in the model that do not appear in the probed structure. In the case of neutron scattering density maps, positive peaks in the difference map correspond either to a missing positive scatterer or to a negative scatterer in the model that is not present in the corresponding probed structrure. Negative peaks correspond either to a positive scatterer in the model without counterpart in the probed structure or to a missing negative one, where positive and negative refer to the neutron scattering length of the isotope being considered.

**Maps representations** Electron density and nuclear scattering density maps are represented as surfaces of homogeneous density. Electron or nuclear scattering density are computed at grid points in the three dimensional real space and a mesh connecting points with the same density is produced to represent the iso-density surface. In the case of X-ray maps each contour level corresponds to a set value of the electron density which is normally expressed as the number of electrons in 1 $Å^3$. The contouring level is generally expressed with reference to the standard deviation of the density function $\rho(\vec{r})$:

$$\sigma_{\rho(\vec{r})} = \sqrt{\frac{1}{g}\sum_{i=1}^{g}[\rho(\vec{r}_i) - \langle\rho\rangle]^2} \tag{5.10}$$

where the sum is performed over $g$ grid points and $\langle\rho\rangle$ is the average value of the density over the considered volume. A map contoured at $x\sigma$ is a surface of homogeneous electron or nuclear scattering density where the density value is $x$ standard deviations above the mean density level.

**Statistical indicators and cross validation** Refinement progress can be monitored by means of a statistical indicator, the R-value, that measures the discrepancy between the observed and the model structure factor amplitudes. Following the definition of this indicator, an increase in the number of model parameters can lead to an improvement in the R-value even when these parameters are not physically meaningful. To avoid the

risk of model overparameterization a cross validation method was devised which consists of excluding a small percentage of data, the test set, from refinement [33]. Two statistical indicators are computed: the linear residual $R_{work}$ calculated by including reflections employed in refinement *i.e.* the working set, and $R_{free}$ calculated by considering only reflections from the test set. These statistical indicators are defined as follows:

$$R_{work} = \frac{\sum_{\vec{h} \notin \{\text{test set}\}} \left| F_{\vec{h}}^{obs} - F_{\vec{h}}^{calc} \right|}{\sum_{\vec{h} \notin \{\text{test set}\}} F_{\vec{h}}^{obs}} \tag{5.11}$$

and

$$R_{free} = \frac{\sum_{\vec{h} \in \{\text{test set}\}} \left| F_{\vec{h}}^{obs} - F_{\vec{h}}^{calc} \right|}{\sum_{\vec{h} \in \{\text{test set}\}} F_{\vec{h}}^{obs}}. \tag{5.12}$$

Overparameterization results in diverging values of $R_{work}$ and $R_{free}$ with differences between the two indicators that exceed 10%.

## 5.2 Neutron and X-ray joint structural refinement

The purpose of neutron macromolecular crystallography is the localization of deuterium substituted hydrogen atoms in a protein structure. Hydrogens account for approximately 50% of the atoms in a protein structure. Hence inclusion of hydrogen and deuterium atoms in the structural model increases the number of parameters to be refined by a factor of two. Simultaneous consideration of X-ray and neutron diffraction data is required to improve the data to parameters ratio with respect to model refinement against neutron data only [35].

In addition at medium resolution -$CH_2$ and -$CH_3$ groups do not contribute appreciably to scattering factors since carbon and hydrogen scattering lengths have values of opposite sign. Hence X-ray diffraction data are required to correctly position these groups.

For this purpose an X-ray diffraction data set is collected either from the same crystal used in neutron exposure or from a different crystal prepared by adopting the same protocol. Exposure to cold neutrons with energy in the range 8.0 - 4.6 meV and corresponding wavelength between 3.2 and 4.2 Å, does not induce appreciable radiation damage in protein crystals as discussed in Chapter 1, allowing the use of the same sample for subsequent

X-ray exposure. Neutron and X-ray diffraction data are processed to obtain for each data set structure factor amplitudes and associated errors. Quasi-Laue neutron diffraction data processing is outlined in Chapter 4 while concise descriptions of both X-ray and neutron data processing protocols adopted in cytochrome *c* peroxidase structural studies are provided in Part III of the present work.

Structural model refinement against X-ray and neutron diffraction data is carried out by means of the *PHENIX* software suite (version 1.9) [34]. First a structural model is refined by using X-ray data only. Then the refined model enters the joint refinement procedure. Hydrogen and deuterium atoms are included in the model and parameters are refined against X-ray and neutron data. This section describes the protocol employed in the two phases of the structural refinement of cytochrome *c* peroxidase.

### 5.2.1 Preparation of the initial structural model

The starting structural model is obtained as follows. A deposited *PDB* model (http://www.rcsb.org) of the structure of cytochrome *c* peroxidase is used as the initial model for refinement against X-ray diffraction data only. The utility *phenix.refine* of the *PHENIX* suite is employed to refine the model against X-ray data alone. A rigid body refinement run is carried out first. The input files are:

- The model file. This is a *PDB* format file containing atomic coordinates, occupancies and *B*-factors. At this stage the model does not include hydrogen and deuterium atoms as well as the heme iron distal ligand and solvent molecules.

- The X-ray data file. This is an *MTZ* format file containing the processed data. For each reflection $h\ k\ l$ the observed structure factor amplitude and its associated error are reported in columns:

  $$h \quad k \quad l \quad F_{\text{obs}} \quad \sigma(F_{\text{obs}}).$$

  If required this file can be edited by the *sftools* utility of the *CCP4i* software suite [36].

$R_{free}$ flags generation is required at this stage. This is done in the *Input data* tab of *phenix.refine* graphical user interface (GUI) by using the *Options* menu in the *X-ray data* section. The *Refinement settings* tab of the GUI is used to specify the refinement details. In this case 1 cycle of rigid body refinement with *n-gaussian* scattering table is carried out.

The software outputs a refined model *PDB* file and a data file in the *MTZ* format (*filename_data*.mtz) containing data and $R_{free}$ flags:

$$h \quad k \quad l \quad F_{\text{obs}} \quad \sigma(F_{\text{obs}}) \quad R_{\text{free}}\,\text{flag}.$$

This file will be used as input in following refinement cycles. The software also outputs a map coefficients file in the *MTZ* format.

The rigid body refined model can now be refined against X-ray data alone in a few runs of atomic coordinates, occupancies and individual isotropic B-factors refinement. 10 cycles of refinement are generally used in each refinement run. Between runs, model modifications such as the inclusion of solvent molecules in the structure are made by means of the modelling software *COOT* [25] and refinement runs are repeated in an iterative fashion. When a satisfying X-ray refined model is obtained the process can be interrupted and the refined model can be used as the starting point for the joint structural refinement.

### 5.2.2 Structural joint refinement

Structural joint refinement is carried out by using the software *phenix.refine* of the *PHENIX* suite [37]. The input files are:

- The model file. For this purpose, the X-ray refined model file obtained as described above is used.
- The X-ray diffraction data file in *MTZ* format with columns:

$$h \quad k \quad l \quad F_{\text{obs}} \quad \sigma(F_{\text{obs}}) \quad R_{\text{free}}\,\text{flag}.$$

- The neutron diffraction data file in *MTZ* format with columns:

$$h \quad k \quad l \quad F_{\text{obs}} \quad \sigma(F_{\text{obs}}).$$

If neutron diffraction data scaling is performed by *SCALEPACK* [18] the resulting data file in *SCA* format can be converted into *MTZ* format by the *CCP4i* program *scalepack2mtz*.

Before being employed as a joint refinement input in *phenix.refine*, the X-ray refined model is modified as follows:

- Solvent molecules, the heme iron distal ligand and alternative conformations are removed by editing directly the *PDB* file in a text editor.
- Hydrogen and deuterium atoms are included by means of the *ReadySet* routine of the *PHENIX* suite [34]. In the case of a perdeuterated structure the following *ReadySet* options are used:
  - *Add hydrogen to model if absent*
  - *Neutron refinement options: all deuterium*

  while in the case of a $D_2O$-soaked crystal:
  - *Add hydrogen to model if absent*
  - *Neutron refinement options: H/D at exchangeable sites and H elsewhere*

  In this case both hydrogen and deuterium are introduced at exchangeable sites, each with occupancy of 0.5.

The structural joint refinement is carried out in *phenix.refine* by feeding in the three input files listed above. The user declares the *Data type* corresponding to each input file and the *Data labels* for X-ray and neutron diffraction data in the *Input data* tab of *phenix.refine* graphical user interface (GUI). $R_{free}$ flags for neutron data are generated at this stage by using the *Options* menu in the *Neutron data* section. The software produces $R_{free}$ flags so that for each reflection the X-ray data and the neutron data $R_{free}$ flags have the same value. Refinement settings and the scattering table are selected in the *Refinement settings* tab of the GUI. The first refinement run consists of 1 cycle of rigid body refinement. The scattering table is *n-gaussian*.

The software outputs a refined model file in *PDB* format and a data file in *MTZ* format (*filename_data*.mtz) containing X-ray and neutron data as well as X-ray and neutron $R_{free}$ flags:

$$h \quad k \quad l \quad F_{obs,x} \quad \sigma(F_{obs,x}) \quad R_{free}\,flag_x \quad F_{obs,n} \quad \sigma(F_{obs,n}) \quad R_{free}\,flag_n$$

This file is used as input in the following refinement cycles. The software also outputs a map coefficients file in *MTZ* format.

The rigid body refined model is used in refinement against X-ray and neutron data simultaneously through runs of atomic coordinates, occupancies and individual isotropic B-factors refinement. It is worthwhile mentioning that only non-unit occupancy factors are refined unless differently specified. 10 cycles of refinement are generally used in each refinement run. These refinement runs are carried out without generation of new $R_{free}$ flags. The option *Hydrogen refinement methods: individual* is selected in the *All parameters - Advanced - Hydrogens* menu to enable refinement of hydrogen and deuterium atoms parameters.

Between runs model modifications such as the inclusion of solvent molecules in the structure are made by means of the modelling software *COOT* [11] and refinement runs are repeated in an iterative manner. After inclusion of solvent molecules oxygen atoms in the model by means of *COOT*, the software *ReadySet* is used to generate the corresponding deuterium atoms by using the option *Add deuterium to solvent molecules*. Solvent molecules deuterium atoms are included in the *PDB* file with an occupancy of 0.01. It is appropriate to edit the *PDB* file in a text editor and set such occupancies to 1.00.

Model modifications such as the inclusion of water molecules are carried out in *COOT* based on both neutron and X-ray $\sigma_A$-weighted $F_o - F_c$ maps. Neutron $\sigma_A$-weighted $2F_o - F_c$ and $F_o - F_c$ maps are used to determine the deuteration state of acidic sites and hydrogen bond networks.

The study of the nature *i.e.* the deuteration state of the heme iron axial ligand is carried out based on neutron $\sigma_A$-weighted $2F_o - F_c$ and $F_o - F_c$ maps only. The nature of the heme iron distal ligand is determined by testing all possible ligands namely $D_2O$, $OD^-$, O. The trial ligand is introduced in the model and a five cycles refinement run is

performed. Inspection of the neutron $\sigma_A$-weighted $F_o - F_c$ maps in the different scenarios indicates which ligand modeling is the most adequate to account for the observed scattering density.

## 5.3 The diffraction precision index

As discussed in Chapter 1 neutron macromolecular crystallography provides direct evidence of amino acid and ligand protonation states unlike X-ray crystallography whereby the protonation state of the heme iron axial ligand is inferred based on the iron - oxygen distance. By using neutron diffraction the nature of the iron ligand is determined based on the neutron scattering density observed in the region as described in Section 5.2. Nevertheless interatomic distances are occasionally reported within the present manuscript as complementary information.

The error associated to a distance is computed based on the approach presented in [38]. The diffraction precision index is the uncertainty in the position of an atom $x$ having a $B$-factor equal to the average $B$-factor of the structure $B_{\text{avg}}$ and it is computed as follows:

$$\sigma_{x,B_{\text{avg}}} = 0.18\ (1+s)^{1/2}\, V_{\text{M}}^{-1/2}\, C^{-5/6}\, \text{R}_{\text{free}}\, d_{\text{min}}^{5/2} \tag{5.13}$$

where $s$ is the solvent content given by the ratio between the number of refined solvent atoms and the total number of atoms, $V_{\text{M}}$ is Matthew's volume namely the volume per mass unit in Å$^3$/Da, $C$ represents the completeness (Section 4.3), $\text{R}_{\text{free}}$ is the refinement statistical indicator defined in Section 5.1 and $d_{\text{min}}$ is the minimum direct lattice interplanar spacing. The uncertainty on the position of atom $x_i$ showing $B$-factor $B_i$ is:

$$\sigma_{x_i} = \sigma_{x,B_{\text{avg}}} \left( \frac{B_i}{B_{\text{avg}}} \right)^{1/2}. \tag{5.14}$$

The error associated to the distance $d_{ij}$ between atoms $x_i$ and $x_j$ is:

$$\sigma_{d_{ij}} = (\sigma_{x_i}^2 + \sigma_{x_j}^2)^{1/2}. \tag{5.15}$$

## 5.4 Concluding remarks

Modern software for macromolecular structural refinement makes use of maximum likelihood algorithms overcoming the limitations of traditional least squares methods which are based on the assumption that $p(F^{\text{obs}}; F^{\text{calc}})$ is a normal distribution. Maximum likelihood techniques naturally incorporate prior stereochemical information through the prior probability of the model giving rise to geometrical restraints in refinement. The software *phenix.refine* used in structural refinement within the present work makes use of maximum likelihood methods. Density maps coefficients are also calculated within the maximum likelihood framework [31] with the purpose of minimizing model bias.

The purpose of neutron crystallography experiments in cytochrome *c* peroxidase is the localization of deuterium substituted hydrogen atoms and in particular in correspondence of the heme iron distal ligand. X-ray and neutron joint structural refinement presents the advantage of increasing the number of experimental observations. This compensates the increase in the number of model parameters that follows the inclusion of hydrogen and deuterium atoms in the structural model and prevents data overfitting. Although X-ray diffraction data provide an important contribution to model refinement, the study of the solvent structure and of the heme iron distal ligand in cytochrome *c* peroxidase is carried out based on neutron scattering density maps only.

# Bibliography


[1] J. Habash, J. Raftery, S. Weisgerber, A. Cassetta, M. S. Lehmann, P. Hoghoj, C. Wilkinson, J. W. Campbell and J. R. Helliwell *J. Chem. Soc., Faraday Trans.* **93**, (24) 4313-4317 (1997).

[2] B. P. Schoenborn *SPIE Neutron Optical Devices and Applications* **1738**, 192-198 (1992).

[3] M. Thoms, D. Myles and C. Wilkinson *Nuclear Instruments and Methods in Physics Research A* **424**, 26-33 (1999).

[4] N. Niimura, Y. Karasawa, I. Tanaka, J. Miyahara, K. Takahashi, H. Saito, S. Koizumi and M. Hidaka *Nuclear Instruments and Methods in Physics Research A* **349**, 521-525 (1994).

[5] C. Wilkinson, M. S. Lehmann, F. Meilleur, M. P. Blakeley, D. A. A. Myles, S. Vogelmeier, M. Thoms, M. Walsh and G. J. McIntyre *J. Appl. Cryst.* **42**, 749-757 (2009).

[6] D. A. A. Myles, C. Bon, P. Langan, F. Cipriani, J. C. Castagna, M. S. Lehmann and C. Wilkinson *Physica B* 241-243, 1122-1130 (1998).

[7] F. Cipriani, F. Dauevergne, A. Gabriel, C. Wilkinson, M. S. Leheman *Biophysical Chemistry* **53**, 5-14 (1994).

[8] C. Wilkinson and M. S. Lehmann *Nuclear Instruments and Methods in Physics Research A* **310**, 411-415 (1991).

[9] M. P. Blakeley, A. J. Kalb, J. R. Helliwell, D. A. A. Myles *Proc. Natl. Acad. Sci.* **101**, 16405-16410 (2004).

[10] C. M. Casadei; A. Gumiero; C. L. Metcalfe; E. J. Murphy; J. Basran; M. G.Concilio; S. C. M. Teixeira; T. E. Schrader; A. J. Fielding; A. Ostermann; M. P. Blakeley; E. L. Raven; P. C. E. Moody,*Science* **345**, 193-197 (2014).

[11] D. A. A. Myles, F. Dauvergne, M. P. Blakeley and F. Meilleur *Journal of Applied Crystallography* **45** 686-692 (2012).

[12] D. A. A. Myles *Transactions ACA* **38** (2003)

[13] F. Meilleur, M. P. Blakeley and D. A. A. Myles in *Hydrogen and hydration sensitive structural biology* 75-85 KubaPro, Tokyo (2005).

[14] J. W. Campbell,*J. Appl. Cryst.* **28**, 228-236 (1995).

[15] J. W. Campbell, Q. Hao, M. M. Harding, N. D. Nguti, C. Wilkinson, *J. Appl. Cryst.* **31**, 496-502 (1998).

[16] I. J. Clifton, E. M. H. Duke, S. Wakatsuki and Z. Ren, *Methods in Enzymology* **277**, 448-467 (1997).

[17] J. W. Campbell, I. J. Clifton, M. M. Harding and Q. Hao, *J. Appl. Cryst.* **28**, 635-640 (1995).

[18] M. Elder *Information Quarterly for Protein Crystallography* **19** (1986).

[19] M. G. Rossmann, *J. Appl. Cryst.* **12**, 225-238 (1979).

[20] D. W. J. Cruickshank, J. R. Helliwell and K. Moffat *Acta Cryst. A* **43**, 656-674 (1987).

[21] S. Arzt, J. W. Campbell, M. M. Harding, Q. Hao, J. R. Helliwell, *J. Appl. Cryst.* **32**, 554-562 (1999).

[22] M. S. Weiss, *J. Appl. Cryst.* **34**, 130-135 (2001).

[23] S. French and K. Wilson, *Acta Cryst.* A**34**, 517-525 (1978).

[24] P. V. Afonine, M. Mustyakimov, R. W. Grosse-Kunstleve, N. W. Moriarty, P. Langan and P. D. Adams *Acta Crystallogr. D* **66**, 1153-1163 (2010).

[25] P. Emsley and K. Cowtan, *Acta Crystallogr. D* **60**, 21262132 (2004).

[26] T. Bayes and R. Price *Philosophical Transactions of the Royal Society of London* **53**, 370-418 (1763).

[27] G. N. Murshudov, A. A. Vagin and A. J. Dodson *Acta Cryst. D* **53**, 240-255 (1997).

[28] A. J. McCoy *Acta Cryst. D* **60**, 2169-2183 (2004).

[29] V. Luzzati Acta Cryst. **5**, 802-810 (1952).

[30] R. Srinivasan and G. N. Ramachandran *Acta Cryst.* **19**, 1008-1014 (1965).

[31] R. J. Read *Acta Cryst. A* **42**, 140-149 (1986)

[32] G. A. Sim *Acta Cryst.* **12**, 813-815 (1959).

[33] A. T. Brunger *Methods Enzymol.* **277**, 366-396 (1997).

[34] P. D. Adams, P. V. Afonine, G. Bunkczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart, *Acta Crystallogr. D* **66** 213221 (2010).

[35] P. D. Adams, M. Mustyakimov, P. V. Afonine and P. Langan *Acta Cryst. D* **65**, 567-573 (2009).

[36] 'Collaborative Computational Project, Number 4, The CCP4 suite: Programs for protein crystallography.' *Acta Crystallogr. D* **50**, 760763 (1994).

[37] P. V. Afonine, M. Mustyakimov, R. W. Grosse-Kunstleve, N. W. Moriarty, P. Langan, P. D. Adams *Acta Crystallogr. D* **66**, 1153-1163 (2010).

[38] M. Gurusaran, M. Shankar, R. Nagarajan, J. R. Helliwell and K. Sekar *IUCrJ* **1** (2014).

# Part III

# Crystallographic and spectroscopic studies of cytochrome *c* peroxidase

# Chapter 6

# Cytochrome *c* peroxidase in the resting state at 298 K

This chapter presents the structural characterization of cytochrome *c* peroxidase in the resting state at 298 K. The neutron structure of the catalytic site is presented along with the UV-visible spectrophotometry study. The purpose of the work described in this chapter is the determination of structural features such as the protonation state of catalytic residues and the nature of the iron axial ligand in the distal heme pocket in the resting state of the enzyme at 298 K.

The knowledge of the structural details of the active site in the resting state close to physiological temperature is of paramount importance for the clarification of the reaction mechanism of cytochrome *c* peroxidase in living organisms. The lack of radiation damage induced by exposure of the protein crystal to cold neutrons allows neutron diffraction data collection at 298 K. This provides unique information concerning the resting state of C*c*P whose structural features are temperature dependent as described in Chapter 7 and 8.

The experimental details of neutron and X-ray diffraction data collection, joint neutron and X-ray structural refinement as well as concise technical information concerning spectrophotometry data acquisition are reported in Section 6.1 of the present chapter. The results of diffraction and spectroscopic experiments are presented in Section 6.2. The

results described in this chapter were published in [1] and the atomic coordinates were deposited in the Protein Data Bank under accession code 4CVI.

## 6.1 Experimental methods

Neutron diffraction data were collected from a single crystal of $D_2O$-soaked cytochrome *c* peroxidase in the ferric state at 298 K. X-ray diffraction data were collected from the same crystal at 298 K. As discussed in Section 1.4 exposure to cold neutrons does not induce appreciable radiation damage in protein crystals allowing the use of the same sample for subsequent X-ray data collection. The single crystal employed in the neutron and X-ray diffraction experiments is shown in Figure 6.1.

This section reports the experimental details of neutron and X-ray diffraction data collection, the procedure employed in structural refinement as well as the technical details concerning spectrophotometry data collection in ferric C*c*P at 298 K.

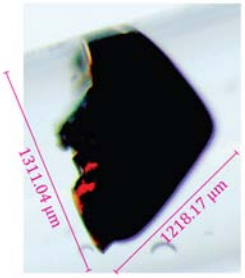


Figure 6.1: $D_2O$-soaked C*c*P crystal in the resting state employed in the neutron and X-ray diffraction experiments at 298 K. The crystal's volume is approximately 1 $mm^3$.

**Neutron data collection** A single crystal of $D_2O$-soaked cytochrome *c* peroxidase in the resting state with volume of approximately 1 $mm^3$ was used. The crystal was mounted in a quartz capillary, surrounded by a small amount of mother liquor and sealed with wax. The amount of mother liquor sorrounding the crystal was kept to a minimum to limit the contribution to scattering and to prevent the crystal from slipping along the capillary wall. A drop of mother liquor was positioned at one end of the capillary tube to maintain

the humidity of the environment. Quasi-Laue neutron diffraction data to 2.4 Å resolution were collected at 298 K on the LADI-III beamline [2] at the Institut Laue-Langevin. As is typical for a Laue experiment, the crystal was held stationary at a different spindle angle setting for each exposure. In total 19 images were collected with an average exposure time of 23.2 hours per image. Figure 6.2 reports one of the images. Three different crystal orientations were adopted. The neutron data were processed using the program *LAUEGEN* [3] modified to account for the cylindrical geometry of the detector [4]. The software *LSCALE* [5] was used to determine the wavelength normalization curve using the intensities of symmetry equivalent reflections measured at different wavelengths. No explicit absorption corrections were applied. Data were merged in *SCALA* [6]. Scaled and merged intensities were converted into structure factor amplitudes and associated error by using the software *TRUNCATE* [7]. The statistics are shown in Table 6.1 and indicators are defined in Section 4.3.

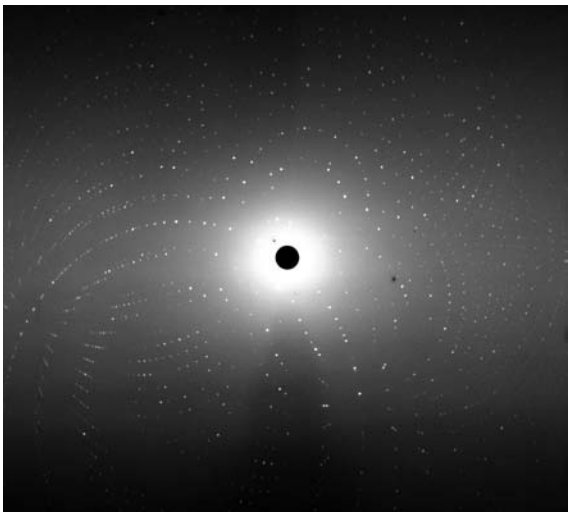

Figure 6.2: Quasi-Laue diffraction pattern of cytochrome *c* peroxidase in the resting state at 298 K recorded at LADI-III (Institut Laue-Langevin).

**X-ray data collection** X-ray diffraction data were collected at 298 K from the same single crystal used for neutron data collection. Copper K$\alpha$ radiation characterized by wavelength $\lambda = 1.5418$ Å from a GeniX Cu HF generator was used. A total of 204 images

of 0.5° oscillation and 60 s exposure per image were recorded on a Mar345 detector. Diffraction intensities were indexed and measured using the *XDS* package [8]. Statistics are shown in Table 6.2 and indicators are defined in Section 4.3.

Table 6.1: Neutron data collection statistics for ferric C*c*P at 298 K. Values in parentheses refer to the highest resolution shell and $d$ is the direct lattice interplanar distance.

| | |
|---|---|
| Space group | $P2_12_12_1$ |
| Cell dimensions | |
| a, b, c (Å) | 51.7 76.8 107.6 |
| **Data collection** | |
| Wavelength (Å) | 3.2 - 4.2 |
| Range of $d$ spacing (Å) | 39.70 - 2.40 (2.53 - 2.40) |
| $R_{merge}$ | 0.151 (0.194) |
| $I/\sigma(I)$ | 7.6 (4.5) |
| Completeness (%) | 75.2 (48.0) |
| Redundancy | 3.8 (2.4) |

**Model refinement** Structural refinement was performed by using as starting model the X-ray structure of ferric C*c*P with Protein Data Bank accession code 2YCG. The starting model was modified to remove multiple conformations and solvent molecules. Structural refinement was carried out using the *PHENIX* software suite [9]. First the model was refined by means of the X-ray terms alone, followed by joint X-ray and neutron refinement [10]. Hydrogen and deuterium atoms were added with the program *ReadySet* [9]. $D_2O$ molecules were introduced according to positive peaks above $3\sigma$ in the $F_o - F_c$ nuclear scattering density map and model modifications were made by means of the modelling program *COOT* [11]. The position of the iron coordinated water molecule was determined using the neutron data only. 198 $D_2O$ molecules were added in total. Final refinement

Table 6.2: X-ray data collection statistics for ferric C*c*P at 298 K. Values in parentheses refer to the highest resolution shell and $d$ is the direct lattice interplanar distance.

| | |
|---|---|
| Space group | $P2_12_12_1$ |
| Cell dimensions | |
| a, b, c (Å) | 51.7 76.8 107.6 |
| **Data collection** | |
| Wavelength (Å) | 1.5418 |
| Range of $d$ spacing (Å) | 14.80 - 2.10 (2.15 - 2.10) |
| $R_{merge}$ | 0.061 (0.243) |
| $I/\sigma(I)$ | 14.6 (4.6) |
| Completeness (%) | 94.7 (97.3) |
| Redundancy | 3.5 (2.4) |

statistics are shown in Table 6.3 and indicators are defined in 5.1.

**UV-visible spectrophotometry** Cytochrome *c* peroxidase in the resting state was characterized by UV-visible spectrophotometry at 298 K both in solution and in a single crystal. Solution spectra were measured by means of a SHIMADZU UV-240IPC recording spectrophotometer. The error associated with the determination of peak positions is of the order of the wavelength step used in the data collection *i.e.* ±1 nm. Single crystal spectra were acquired by using the micro-spectrophotometry set-up described in Appendix D.1. The error associated with peak positions amounts to a few nanometers due to the difficulty in positioning features in a spectrum with significant measurement noise on the vertical axis.

Table 6.3: Neutron and X-ray structural joint refinement statistics for ferric C*c*P at 298 K. The subscripts n and X refer to neutron and X-ray diffraction data respectively.

| **Joint refinement** | |
|---|---|
| $d_{min,n}$ (Å) | 2.4 |
| $d_{min,X}$ (Å) | 2.1 |
| Number of unique reflections $N_n$ | 12695 |
| Number of unique reflections $N_X$ | 24307 |
| $R_{work,n}$ | 0.1759 |
| $R_{work,X}$ | 0.1337 |
| $R_{free,n}$ | 0.2433 |
| $R_{free,X}$ | 0.1766 |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.012 |
| Bond angles (°) | 1.349 |

## 6.2 Interpretation of the results

The neutron crystallography structure of C*c*P in the resting state at 298 K allowed the unambiguous determination of hydrogen related structural features in the macromolecule. Protonation states were determined *via* inspection of the strong neutron scattering density peaks emerging in presence of deuterium substituted hydrogen atoms as discussed in Section 5.2.

The nuclear scattering density map of the active site is shown in Figure 6.3 while Figure 6.4 shows the electron density and the nuclear scattering density of the catalytic residue His-52. At the end of the section, panel 6.8 reports side by side the neutron and X-ray maps of the distal and proximal heme pockets, for direct comparison, together with the model of the structure. The present section reports the description of the structural features of

the active site and the characterization of the species by UV-visible spectrophotometry.

**The structure of the active site** The catalytic site of C*c*P in the resting state at 298 K presents the following structural features. On the distal side of the heme (Figure 6.8, top), His-52 is neutral as $N_\delta$ is deuterated but $N_\epsilon$ is not. The $N_\epsilon$ of Trp-51 is deuterated; Arg-48 is fully deuterated and positively charged. The neutron structure unambiguously identifies the iron distal ligand as a heavy water molecule, rather than an $OD^-$ ion, at a distance of $(2.7 \pm 0.4)$ Å from the iron (the uncertainty is calculated according to Section 5.3). The heme is therefore considered as five-coordinate. The water ligand donates a hydrogen bond to the $\pi$ orbitals of the porphyrin macrocycle and may also interact with a poorly ordered water that is found in the active site (not shown) and that displaces the guanidinium group of Arg-48, found in the 'out' position. On the proximal side (Figure 6.8, bottom), $N_\epsilon$ of His-175 is not deuterated and is $(2.0 \pm 0.3)$ Å from the iron. The $N_\delta$ of His-175 is deuterated and within hydrogen bond distance of the unprotonated $O_{\delta 2}$ of Asp-235; likewise, $O_{\delta 1}$ of Asp-235 is unprotonated and within hydrogen bond distance of the deuterated $N_\epsilon$ of Trp-191. Both heme propionates are unprotonated. Figure 6.5 presents a schematic interpretation of the results.

**UV-visible spectrophotemetry** Cytochrome *c* peroxidase in the resting state was characterized by UV-visible spectrophotometry at 298 K both in solution and in a single crystal. The relevance of this type of characterization is explained in Section 1.5. Figure 6.6 presents the UV-visible spectrum of C*c*P in solution. Absorbance peaks were observed at 408 nm, 505 nm, 541 nm (shoulder) and 640 nm with uncertainty of ±1 nm. Figure 6.7 shows the single crystal UV-visible spectrum. The Soret band (408 nm) saturates in the single crystal due to the elevated optical density. Peaks are observed at wavelengths of 501 nm and 621 nm with errors amounting to a few nanometers.

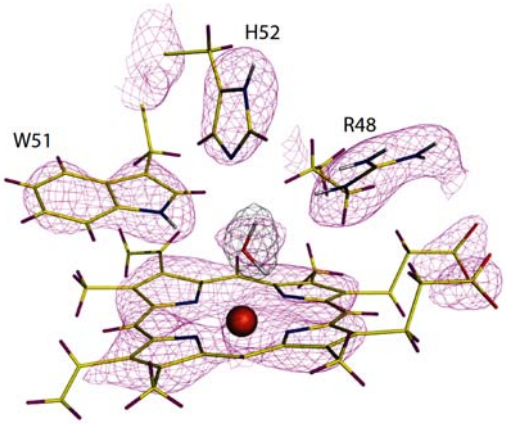


Figure 6.3: The structure of ferric C*c*P in the distal heme pocket at 298 K. The nuclear scattering density [$\sigma_A$-weighted $2F_o - F_c$ map contoured at 2.6 root mean square (rms)] is shown in magenta. The black contour [$\sigma_A$-weighted neutron $F_o - F_c$ map contoured at $5\sigma$] shows the difference density calculated by omitting the water ligand. The following colour scheme applies to the model of the structure: hydrogen atoms, magenta; deuterium, white; oxygen, red; nitrogen, blue; carbon, yellow; iron, orange sphere.

## 6.3 Concluding remarks

Cytochrome *c* peroxidase in the resting state at 298 K was characterized by neutron crystallography and by ultraviolet-visible spectrophotometry. The model of the structure of C*c*P in the resting state at 298 K was refined to a resolution of 2.4 Å allowing the unambiguous determination of the catalytic residues protonation states as well as the nature of the iron distal ligand. Of particular interest are the following observations:

- The neutron structure of cytochrome *c* peroxidase in the resting state at 298 K shows that $N_\epsilon$ of catalytic residue His-52 is not protonated.
- Furthermore the iron distal axial ligand is a $D_2O$ molecule.

The characterization by UV-visible spectophotometry showed that solution and single crystal spectra are similar to each other and in good agreement with what was previously reported [12].

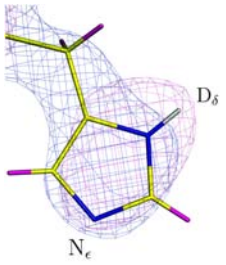


Figure 6.4: The side chain of the catalytic residue His-52 in ferric C*c*P at 298 K. The nuclear scattering density [$\sigma_A$-weighted $2F_o - F_c$ map contoured at 2.6 root mean square (rms)] is shown in magenta. The electron density [$\sigma_A$-weighted $2F_o - F_c$ map contoured at 2.4 rms] is shown in blue. Colour scheme as in Figure 6.3.

Both the crystallographic structure of C*c*P in the resting state at 298 K and its absorbance spectrum in the visible region were used in the present research project as a benchmark for drawing comparisons between different reaction intermediates and different temperatures. The results presented in this Chapter were employed in the study of the reaction pathway from the resting state of cytochrome *c* peroxidase to its transient compound I discussed in Chapter 11 and in the investigation of the temperature dependence of the characteristics of the active site presented in Chapter 12.

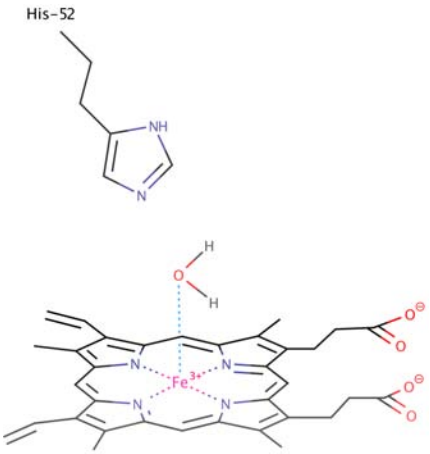


Figure 6.5: The protonation state of the heme iron distal ligand and catalytic residue His-52 in C*c*P resting state at 298 K.

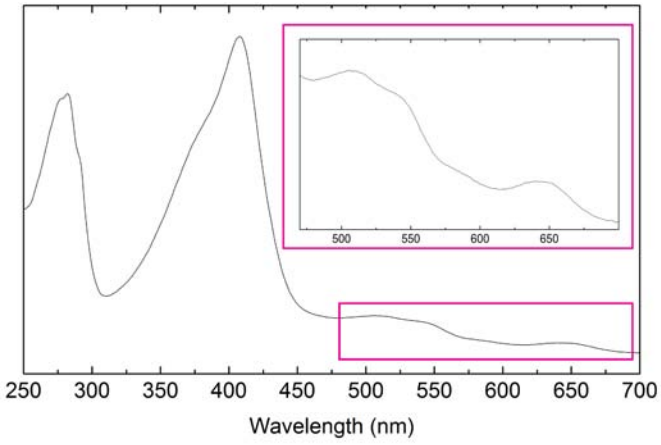


Figure 6.6: UV-VIS absorption spectrum of C*c*P in solution at 298 K. The vertical axis reports the optical density. The insert shows the spectral region between 470 and 700 nm. Buffer: 10 mM $KH_2PO_4$, 150 mM KCl, pH 6.5.

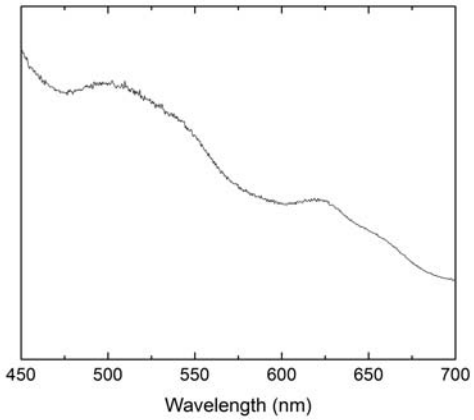


Figure 6.7: UV-VIS absorption spectrum of a single crystal of C*c*P in the resting state at 298 K. The vertical axis reports the optical density.

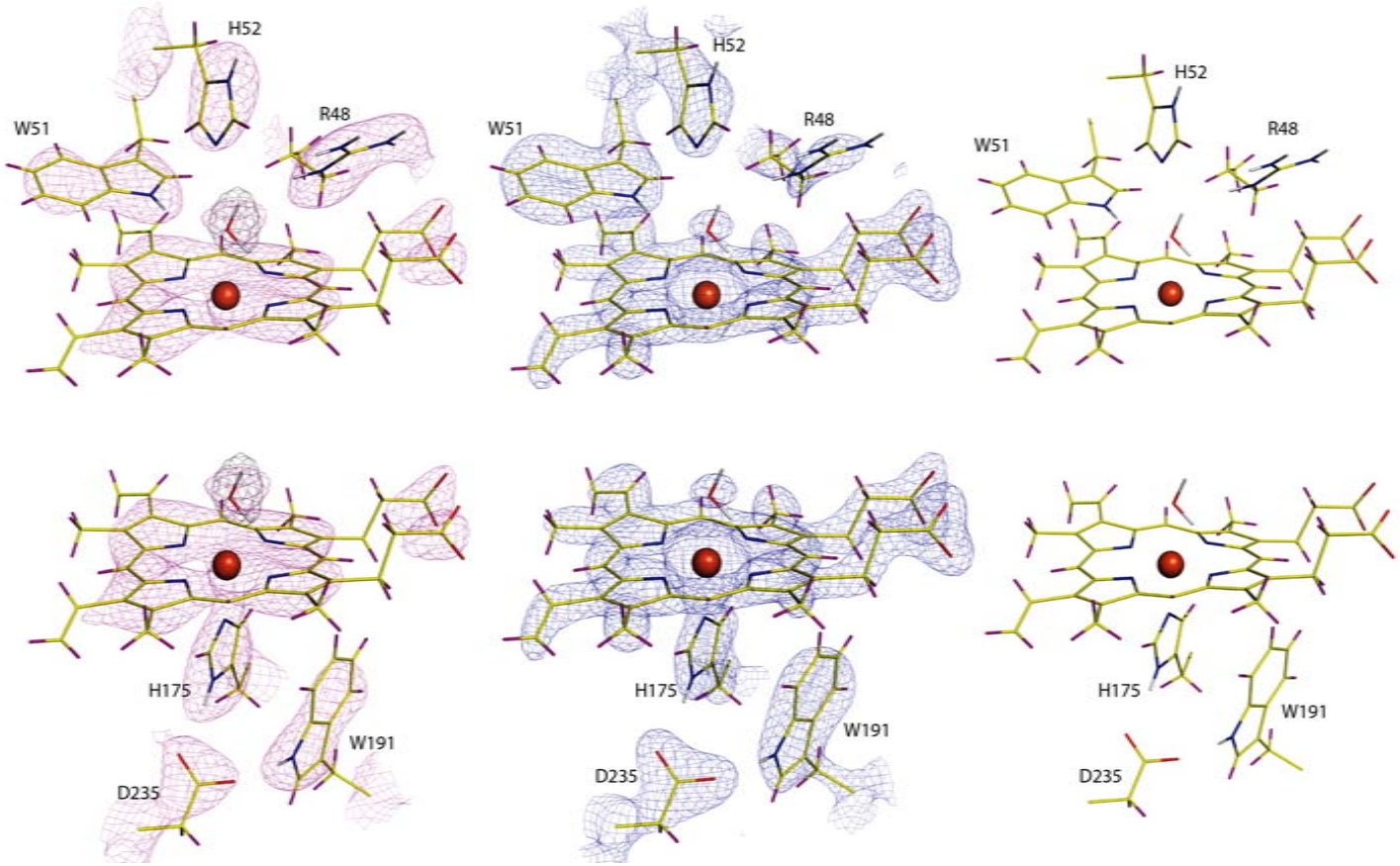


Figure 6.8: **The structure of ferric C*c*P in the region of the heme at 298 K.** Left: nuclear scattering density [$\sigma_A$-weighted $2F_o - F_c$ map contoured at 2.6 root mean square (rms)] in the distal and proximal heme pocket is shown in magenta. The black contour [$\sigma_A$-weighted neutron $F_o - F_c$ map contoured at $5\sigma$] shows the difference density calculated by omitting the water ligand. Centre: electron density [$\sigma_A$-weighted $2F_o - F_c$ map contoured at 2.4 rms] in the distal and proximal pocket is shown in blue. Right: the model of the structure of the distal and proximal heme pocket. Colour scheme as in Figure 6.3.

# Chapter 7

# Cytochrome *c* peroxidase in the resting state at 100 K

This chapter presents the study of cytochrome *c* peroxidase in the resting state at 100 K. Neutron crystallography studies were complemented by spectroscopic investigations. In particular the dependence on temperature of the resting state absorbance peaks in the visible region was determined.

This work is aimed at the investigation of the temperature dependence of the structural features in the active site of ferric C*c*P. The structure of C*c*P at cryogenic temperature was compared to that at 298 K presented in Chapter 6 and the temperature dependence of protonation states in the catalytic site was studied.

The experimental details of neutron and X-ray diffraction data acquisition, structural refinement and spectrophotometry data collection are reported in Section 7.1 of the present chapter. Section 7.2 presents the results of crystallographic and spectroscopic experiments. The analysis of the temperature dependence of structural features in the active site of C*c*P resting state is presented in Chapter 12.

## 7.1 Experimental methods

Neutron diffraction data were collected from a single crystal of deuterium exchanged cytochrome *c* peroxidase in the resting state at 100 K. The single crystal employed in the neutron crystallography experiment is shown in Figure 7.1. X-ray diffraction data were collected from a different but isomorphous crystal at 100 K.

This section reports the experimental details of diffraction data collection, the procedure employed in structural refinement and the technical information concerning spectrophotometry data acquisition in ferric C*c*P at 100 K.

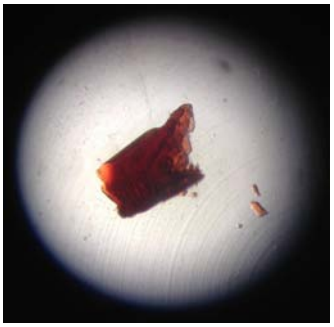

Figure 7.1: $D_2O$-soaked C*c*P crystal in the resting state employed in the neutron diffraction experiment at 100 K. The crystal's volume is approximately 0.1 $mm^3$.

**X-ray data collection** X-ray diffraction data were collected at 100 K from a single crystal of $D_2O$-soaked C*c*P in the resting state. Copper $K_\alpha$ radiation characterized by wavelength $\lambda = 1.5418$ Å from a GeniX Cu HF generator was used. A total of 270 images of 0.5° oscillation and 60 s exposure per image were recorded on a Mar345 detector. Diffraction intensities were indexed and measured using the *XDS* package [8]. Statistics are shown in Table 7.1 and indicators are defined in Section 4.3.

**Neutron data collection** A single crystal of $D_2O$-soaked cytochrome *c* peroxidase in the resting state with volume of approximately 0.1 $mm^3$ was used. The crystal was mounted in a LithoLoop (Molecular Dimensions) on a magnetic cap (Molecular Dimensions). The crystal was cryo-cooled directly in a stream of nitrogen gas at 100 K from

Table 7.1: X-ray data collection statistics for ferric C*c*P at 100 K. Values in parentheses refer to the highest resolution shell and $d$ is the direct lattice interplanar distance.

| | |
|---|---|
| Space group | $P2_12_12_1$ |
| Cell dimensions | |
| a, b, c (Å) | 51.0 75.9 106.9 |
| **Data collection** | |
| Wavelength (Å) | 1.5418 |
| Range of $d$ spacing (Å) | 15.00 - 2.00 (2.05 - 2.00) |
| $R_{merge}$ | 0.048 (0.117) |
| $I/\sigma(I)$ | 26.6 (13.3) |
| Completeness (%) | 99.3 (99.7) |
| Redundancy | 5.3 (5.3) |

the Oxford Cryosytems Cobra - Non Liquid Nitrogen Open Flow System (Section 3.2), of the instrument LADI-III. Quasi-Laue neutron diffraction data to 2.2 Å resolution were collected at 100 K on the LADI-III beamline [2] at the Institut Laue-Langevin. The crystal was held stationary at a different spindle angle setting for each exposure. In total 21 images were collected with exposure time of 24 hours per image. Two different crystal orientations were adopted. The neutron data were processed using the program *LAUEGEN* [3] modified to account for the cylindrical geometry of the detector [4]. The values of the unit cell parameters determined by X-ray crystallography where introduced as an input to the indexing algorithm. The software *LSCALE* [5] was employed to determine the wavelength normalization curve using the intensities of symmetry equivalent reflections measured at different wavelengths. No explicit absorption corrections were applied. Data were merged in *SCALA* [6]. Scaled and merged intensities were converted into structure factor amplitudes and associated error by means of the program *TRUNCATE* [7]. Data

collection statistics are shown in Table 7.2 and indicators are defined in Section 4.3.

Table 7.2: Neutron data collection statistics for ferric C*c*P at 100 K. Values in parentheses refer to the highest resolution shell and $d$ is the direct lattice interplanar distance.

| | |
|---|---|
| Space group | $P2_12_12_1$ |
| Cell dimensions | |
| a, b, c (Å) | 51.0 75.9 106.9 |
| **Data collection** | |
| Wavelength (Å) | 3.0 - 4.0 |
| Range of $d$ spacing (Å) | 39.37 - 2.20 (2.32 - 2.20) |
| $R_{merge}$ | 0.163 (0.187) |
| $I/\sigma(I)$ | 6.9 (4.7) |
| Completeness (%) | 69.0 (47.3) |
| Redundancy | 3.4 (2.2) |

**Model refinement** Structural refinement was performed by using as a starting model the structure with Protein Data Bank accession code 4CVJ. The starting model was modified to remove hydrogen and deuterium atoms, the heme iron distal axial ligand as well as solvent molecules. Structural refinement was carried out using the *PHENIX* software suite [9]. First the model was refined by means of the X-ray terms alone followed by joint X-ray and neutron refinement. Hydrogen and deuterium atoms were added with the program *ReadySet* [9]. $D_2O$ molecules were introduced according to positive peaks above $3\sigma$ in the $F_o - F_c$ nuclear scattering density map and model modifications were made by means of the modelling program *COOT* [11]. 323 $D_2O$ molecules were added in total. Final refinement statistics are shown in Table 7.3 and indicators are defined in Section 5.1.

Table 7.3: Neutron and X-ray structural joint refinement statistics for ferric C*c*P at 100 K. The subscripts n and X refer to neutron and X-ray diffraction data respectively.

| **Joint refinement** | |
|---|---|
| $d_{min,n}$ (Å) | 2.2 |
| $d_{min,X}$ (Å) | 2.0 |
| Number of unique reflections $N_n$ | 14720 |
| Number of unique reflections $N_X$ | 28569 |
| $R_{work,n}$ | 0.1963 |
| $R_{work,X}$ | 0.1435 |
| $R_{free,n}$ | 0.2539 |
| $R_{free,X}$ | 0.1831 |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.010 |
| Bond angles (°) | 1.622 |

**UV-visible spectrophotometry** Single crystal UV-visible spectra were acquired by using the micro-spectrophotometry set-up described in Appendix D.1. The error associated with the determination of peak positions ranges from one to a few nanometers depending on the measurement noise on the vertical axis.

## 7.2 Interpretation of the results

The neutron cryo-crystallography structure of C*c*P in the resting state allowed the determination of the catalytic residues protonation states and provided information concerning the nature of the iron axial ligand in the distal heme pocket. The structure at cryogenic temperature was compared to that at 298 K (Chapter 6) and the temperature dependence of protonation states in the active site was observed. In addition inspection of UV-visible

spectra showed significant modifications of spectral features with varying temperature corroborating what was observed *via* neutron crystallography.

The present section reports the description of the structural features of the active site at cryogenic temperature and the comparison with the structure determined at 298 K. The temperature dependence of single crystal absorption in the visible region of the electromagnetic spectrum is also presented.

**The structure of the active site** The catalytic site of C*c*P in the resting state at 100 K presents the following structural features. Compared to the enzyme at 298 K, most of the protonation states of the catalytic residues are retained in the structure at 100 K. His-52 is an exception showing single deuteration on $N_\delta$ at 298 K and double deuteration on $N_\delta$ and $N_\epsilon$ at 100 K (Figure 7.2). Unlike the structure at 298 K the iron distal ligand is not an ordered $D_2O$ molecule at cryogenic temperature. Nevertheless the identification of the heme iron distal ligand in this structure remained ambiguous motivating further investigation of the low temperature features of the resting state by using a perdeuterated sample (Chapter 8).

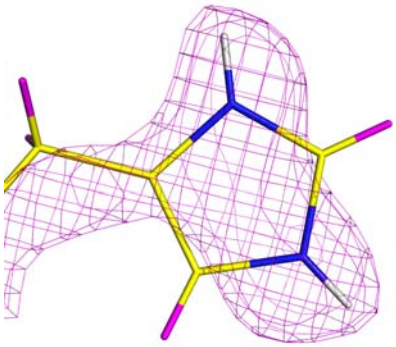

Figure 7.2: The side chain of the catalytic residue HIS-52 in ferric C*c*P at 100 K. The nuclear scattering density [$\sigma_A$-weighted $2F_o - F_c$ map contoured at 2.8 root mean square (rms)] is shown in magenta. The following colour scheme applies to the model of the structure: hydrogen atoms, magenta; deuterium, white; nitrogen, blue; carbon, yellow.

**UV-visible spectrophotemetry** The temperature dependence of the characteristics of the active site in cytochrome *c* peroxidase in the resting state was studied by single crystal UV-visible spectrophotometry (Section 1.5). Figure 7.3 shows the single crystal ultraviolet-visible spectra of ferric C*c*P at different temperatures. The Soret band (close to 408 nm) saturates in the single crystal due to the elevated optical density. At cryogenic temperature peaks are observed at wavelengths of 542 nm, 579 nm and 624 nm (shoulder) with errors of several nanometers on the peak positions. With increasing temperature these peaks disappear and features at 501 nm and 621 nm (with errors amounting to a few nanometers) arise in agreement with [12].

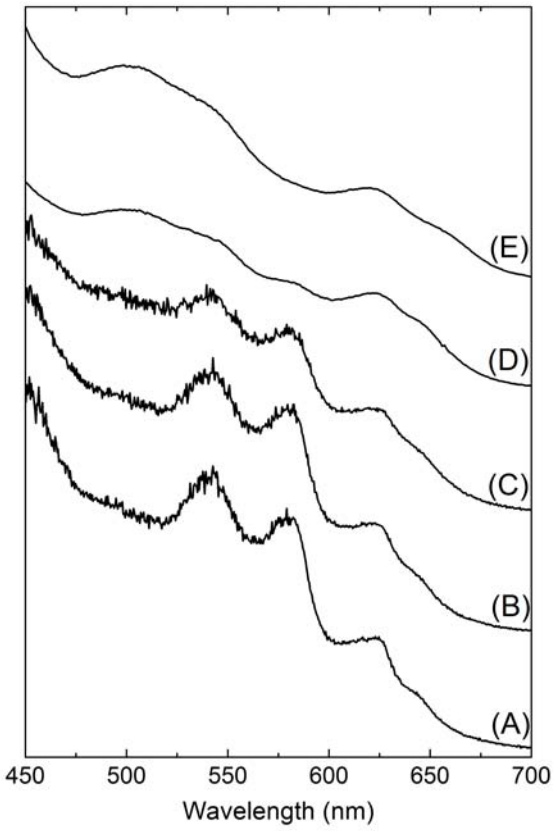


Figure 7.3: Temperature dependence of single crystal ultraviolet-visible absorption spectra of C*c*P in the resting state. (A) 100 K, (B) 150 K, (C) 200 K, (D) 230 K and (E) 298 K. The vertical axis reports the optical density on an arbitrary scale. The spectra have been shifted along the vertical axis for clarity.

## 7.3 Concluding remarks

Cytochrome *c* peroxidase in the resting state at 100 K was characterized by neutron crystallography and by single crystal UV-visible spectrophotometry. The neutron crystallography structure and the absorbance peaks in the visible region at cryogenic temperature were compared to their counterparts at 298 K revealing a significant temperature dependence of the characteristics of the active site.

The neutron structure of C*c*P in the resting state at 100 K was determined to a resolution of 2.2 Å allowing the unambiguous identification of the catalytic residues protonation states and providing information concerning the iron distal ligand. The structure of the active site at cryogenic temperature was compared to that at 298 K and the following features were observed to be temperature dependent:

- $N_\epsilon$ of catalytic residue His-52 is deuterated at cryogenic temperature and it is not at 298 K.
- Furthermore the heme iron distal ligand is not an ordered $D_2O$ molecule at 100 K, different from what was observed at 298 K.

It was not possible to unambiguously identify the heme iron distal ligand and the temperature dependence of the neutron crystallography structure of C*c*P resting state was further investigated by using a perdeuterated sample in Chapter 8.

Ultraviolet-visible absorption spectra were collected as a function of temperature. A smooth transition was observed from low temperature spectral features at 100 K to high temperature features at 298 K corroborating the temperature dependence observed in neutron crystallography structures. This analysis suggests that UV-visible spectrophotometry in C*c*P is sensitive to small structural changes in the chromophore allowing the use of spectra in the visible region for the identification of the species.

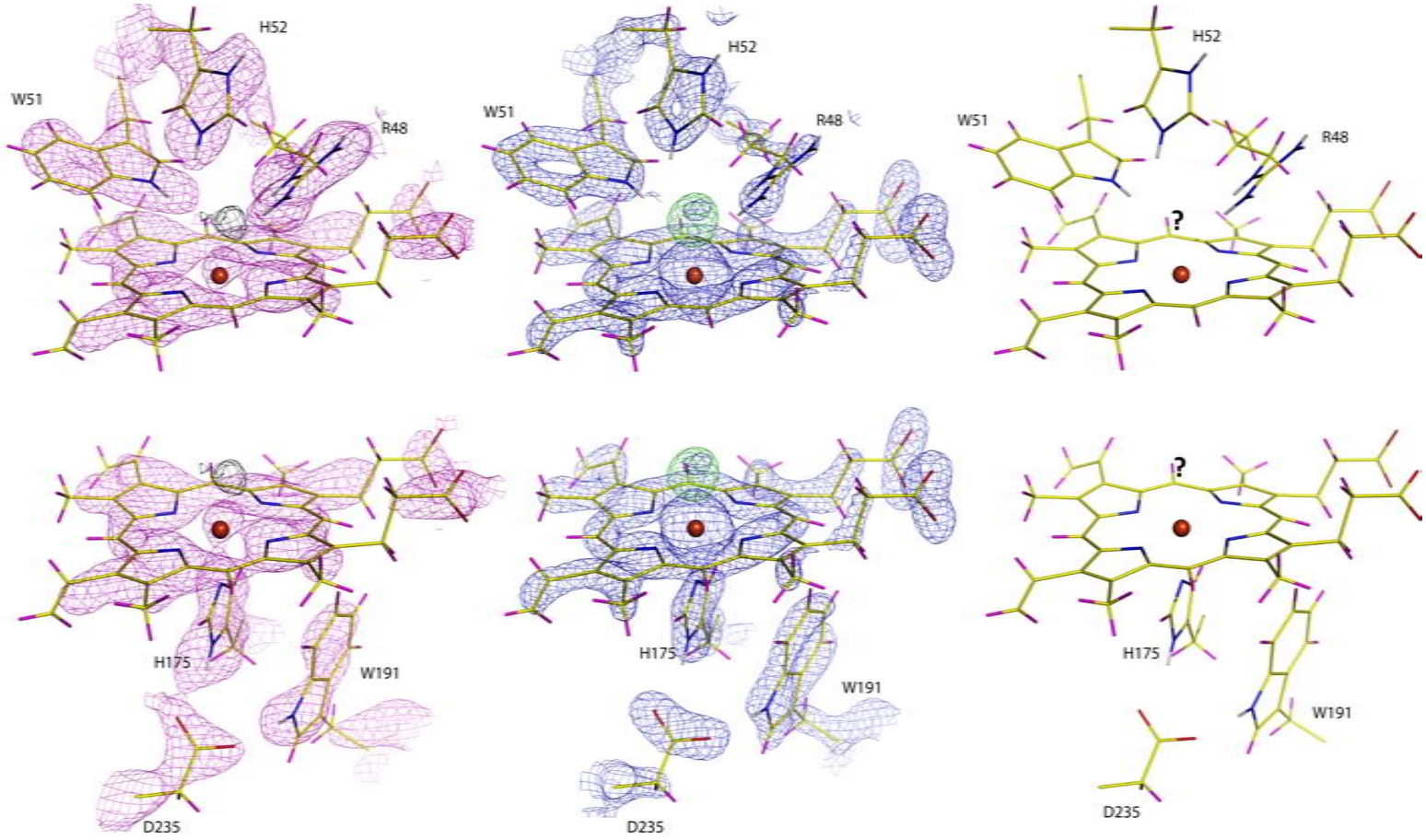


Figure 7.4: **The structure of ferric CcP in the region of the heme at 100 K.** Left: nuclear scattering density [$\sigma_A$-weighted $2F_o - F_c$ map contoured at 2.4 root mean square (rms)] in the distal and proximal heme pocket is shown in magenta. The black contour [$\sigma_A$-weighted neutron $F_o - F_c$ map contoured at $3.5\sigma$] shows the difference density calculated by omitting the axial ligand. Centre: electron density [$\sigma_A$-weighted $2F_o - F_c$ map contoured at 2.8 rms] in the distal and proximal pocket is shown in blue. The green contour [$\sigma_A$-weighted electron density $F_o - F_c$ map contoured at $6\sigma$] shows the difference density calculated by omitting the axial ligand. Right: the model of the structure of the distal and proximal heme pocket. Colour scheme as in Figure 7.2.

# Chapter 8

# Perdeuterated cytochrome *c* peroxidase in the resting state at 100 K

This chapter presents the crystallographic and spectroscopic studies concerning perdeuterated cytochrome *c* peroxidase in the resting state at 100 K. The purpose of this work is the investigation of the temperature dependence of the characteristics of the active site in ferric C*c*P. The neutron crystallography structure of C*c*P at cryogenic temperature was compared to that at 298 K presented in Chapter 6 and the temperature dependence of structural features in the catalytic site was studied.

For the reasons described in Section 2.4.4, the use of fully deuterated protein improved significantly the quality of neutron diffraction data without substantial alterations of the crystal structure. The absence of major deuteration induced artifacts was investigated by X-ray crystallography and ultraviolet-visible spectrophotometry. In particular the dependence of absorbance peaks in the visible region of the electromagnetic spectrum on the extent of deuteration of the protein was studied.

The experimental details of neutron and X-ray diffraction data collection, X-ray structural refinement, joint neutron and X-ray structural refinement as well as spectrophoto-

metry data acquisition methods are reported in Section 8.1 of the present chapter. The results of diffraction and spectroscopic experiments are presented in Section 8.2.

## 8.1 Experimental methods

Fully deuterated (perdeuterated) cytochrome *c* peroxidase was expressed, purified and crystallized as reported in Appendix B. X-ray diffraction data were collected from a single crystal of fully deuterated cytochrome *c* peroxidase in the ferric state at 100 K and the structural model was refined at this stage by using X-ray data alone. Neutron diffraction data were collected from a different but isomorphous crystal at 100 K. The model was refined by joint neutron and X-ray strucural refinent. The crystallographic studies were complemented by UV-visible spectrophotometry measurements.

This section reports the experimental methods used in neutron and X-ray diffraction data collection, the procedures employed in X-ray and joint structural refinement as well as the technical details concerning spectrophotometry data collection in fully deuterated ferric C*c*P at 100 K. C*c*P perduterated crystals employed in the X-ray diffraction and UV-vis spectrophotometry experiments are shown in Figure 8.1 while Figure 8.2 reports a photograph of the single crystal used in the neutron crystallography data collection.

Figure 8.1: Perdeuterated cytochrome *c* peroxidase crystals employed in the X-ray diffraction work and ultraviolet-visible spectrophotometry experiments. The crystals are in a dialysis button with diameter of 7 mm.

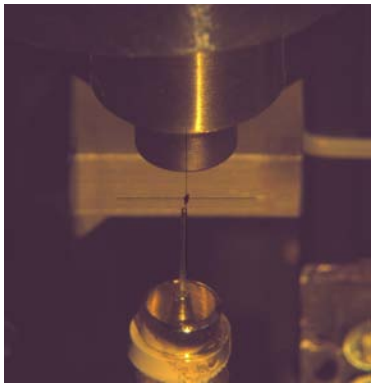

Figure 8.2: Perdeuterated cytochrome *c* peroxidase crystal mounted in the cryostream at the monochromatic neutron diffractometer BIODIFF.

**X-ray diffraction data collection** X-ray crystallography data collection from perdeuterated C*c*P at 100 K was carried out at the Henry Wellcome Laboratories for Structural Biology of the University of Leicester. A single crystal of fully deuterated C*c*P in the ferric state was mounted in a nylon loop and maintained at 100 K using an Oxford Cryosystems cryostream. A total of 270 images of 0.5° oscillation and 5 s exposure per image were collected using copper $K_\alpha$ radiation (1.5418 Å) from a Rigaku 007 HF X-ray generator fitted with Osmic VariMax HF optics and a Rigaku Saturn 944+ CCD detector. Data were processed by means of the *CCP4* suite of programs (Collaborative Computational Project 4 [13]). Recorded intensities were indexed and measured using *iMOSFLM* [14] then scaled and merged using *AIMLESS* [15]. Statistics are shown in Table 8.1 and indicators are defined in Section 4.3.

**Model refinement** Phasing by molecular replacement was done by using the software *Phaser* [16] adopting as a starting model the structure with PDB code 2XIL. Model refinement was done by *REFMAC5* software [17] and all model modifications were done by the modelling program *COOT* [11]. Final refinement statistics are shown in Table 8.1 and indicators are defined in 5.1.

Table 8.1: X-ray data collection and structural refinement statistics for perdeuterated ferric C*c*P at 100 K. Values in parentheses refer to the highest resolution shell and $d$ is the direct lattice interplanar distance.

| | |
|---|---|
| Space group | $P2_12_12_1$ |
| Cell dimensions | |
| a, b, c (Å) | 50.9 75.7 106.9 |
| **Data collection** | |
| Range of $d$ spacing (Å) | 12.39 - 1.81 (1.85 - 1.81) |
| $R_{merge}$ | 0.11 (0.43) |
| $I/\sigma(I)$ | 9.8 (2.3) |
| Completeness (%) | 99.3 (95.5) |
| Redundancy | 5.1 (3.2) |
| **Refinement** | |
| $d_{min}$ (Å) | 1.81 |
| Number of unique reflections | 36298 |
| $R_{work}$ | 0.15 |
| $R_{free}$ | 0.19 |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.02 |
| Bond angles (°) | 2.0 |

**Neutron data collection** A large single crystal (1.8 mm x 0.5 mm x 0.3 mm) of perdeuterated cytochrome *c* peroxidase was mounted in a 1 mm LithoLoop (Molecular Dimensions MD7-147) on a magnetic cap (CrystalCap Magnetic, ALS, Hampton Research HR4-779). The crystal was cryo-cooled directly in a stream of nitrogen gas at 100 K from the OxfordCryosytems series 700+ Cryostream of the instrument BIODIFF. Neu-

tron diffraction data were recorded to 1.83 Å resolution at 100 K using the monochromatic neutron diffractometer BIODIFF at the FRM II research reactor (Garching, Germany). BIODIFF uses a cylindrical neutron-sensitive image plate detector that completely surrounds the crystal (diameter: 400 mm, height: 450 mm). At a wavelength of 3.37 Å (with $\Delta\lambda/\lambda = 3\%$), 115 frames were recorded with a rotation range of 0.5° per frame and an exposure time of 120 minutes per frame. To increase the completeness a second crystal (0.6 mm x 0.8 mm x 1.2 mm), isomorphous to the first, was mounted on the instrument [1] and a second series of 59 frames were collected at a wavelength of 3.37 Å (with $\Delta\lambda/\lambda = 3\%$) with a rotation range of 0.5° per frame and an exposure time of 80 minutes per frame. The diffraction data were indexed and integrated using *DENZO* (v.1.96.2) [18] and scaled up to 1.83 Å resolution using *SCALEPACK* (v.2.3.6) [18]. Statistics are shown in Table 8.2 and indicators are defined in Section 4.3.

Table 8.2: Neutron data collection statistics for perdeuterated ferric C*c*P at 100 K. Values in parentheses refer to the highest resolution shell and $d$ is the direct lattice interplanar distance.

| | |
|---|---|
| Space group | $P2_12_12_1$ |
| Cell dimensions | |
| a, b, c (Å) | 50.9 75.7 106.9 |
| **Neutron data collection** | |
| Range of $d$ spacing (Å) | 107.00 - 1.83 (1.88 - 1.83) |
| $R_{merge}$ | 0.21 (0.47) |
| $R_{pim}$ | 0.17 (0.47) |
| $I/\sigma(I)$ | 2.87 (1.08) |
| Completeness (%) | 89.1 (65.8) |
| Redundancy | 2.0 (1.3) |

[1] The first crystal was lost during the transfer to liqid nitrogen required for changing the goniometer angle.

**Joint neutron and X-ray structural refinement** The crystals used for neutron and X-ray data collection were essentially isomorphous. The structural model determined by the use of X-ray data alone was modified to remove the iron axial ligand and all the solvent molecules. It was then used as the starting point for structural refinement using the *PHENIX* software suite [9]. A joint X-ray and neutron refinement was carried out [10]. A rigid body refinement was performed first, followed by multiple iterations of atomic coordinates, occupancies and individual isotropic B-factors refinement. Deuterium atoms were added with the program *ReadySet* [9]. $D_2O$ molecules were added according to positive peaks above $4\sigma$ in the $F_o - F_c$ difference nuclear scattering density map and all model modifications were made with the modelling program *COOT* [11]. 361 water molecules were added in total. No contamination from hydrogen was accounted for during structural refinement although interestingly hydrogen presence was found localized in backbone amide groups. The occupancy of His-52 $D_\delta$ and $D_\epsilon$ atoms (Figure 8.5) as well as the occupancy of the deuterium atom in the heme iron axial ligand were refined using a starting value of 0.1. The position and identity of the heme iron distal ligand were determined using the neutron data only. The nature of the distal ligand was investigated by using different trial models and by inspection of the $\sigma_A$-weighted $F_o - F_c$ neutron map computed after model refinement in each case. An oxygen atom, a $D_2O$ molecule and an $OD^-$ group were tested as a trial ligand. Final refinement statistics are shown in Table 8.3 and indicators are defined in 5.1.

**UV-visible spectrophotometry** Fully deuterated cytochrome *c* peroxidase in the resting state was characterized by single crystal UV-visible spectrophotometry at 100 K. Spectra were acquired by using the micro-spectrophotometry set-up described in Appendix D.1. The error associated with peak positions amounts to a few nanometers due to the degree of uncertainty in positioning spectral features with significant measurement noise on the vertical axis. The study of the effect of C*c*P deuteration on the absorbance in the visible region was carried out at the SLSpectroLAB facility [19] of the X10SA beamline at the Swiss Light Source (Villigen, Switzerland).

Table 8.3: Neutron and X-ray structural joint refinement statistics for perdeuterated ferric C*c*P at 100 K. The subscripts n and X refer to neutron and X-ray diffraction data respectively.

| **Joint refinement** | |
|---|---|
| $d_{min,n}$ (Å) | 1.83 |
| $d_{min,X}$ (Å) | 1.81 |
| Number of unique reflections $N_n$ | 32816 |
| Number of unique reflections $N_X$ | 38204 |
| $R_{work,n}$ | 0.2072 |
| $R_{work,X}$ | 0.1484 |
| $R_{free,n}$ | 0.2544 |
| $R_{free,X}$ | 0.1846 |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.010 |
| Bond angles (°) | 1.283 |

## 8.2 Interpretation of the results

The X-ray crystallography structure of fully deuterated C*c*P in the resting state at 100 K was used for direct comparison with the structure of its hydrogenated counterpart (Protein Data Bank accession code 2YCG). Photoreduction effects from X-ray exposure were monitored by UV-visible spectrophotometry.

The neutron crystallography structure of perdeuterated C*c*P in the resting state at 100 K allowed the determination of protonation states of catalytic residues in the active site. In addition the nature of the heme iron distal ligand was investigated. The temperature dependence of the structural features in the active site was studied complementing the analysis presented in Chapter 7.

The present section reports the results of the comparison between hydrogenated and

fully deuterated C*c*P X-ray structures in the resting state at 100 K, the description of the structural features of the active site in the neutron structure of perdeuterated C*c*P at 100 K and the study of the dependence of absorption in the visible region on the deuteration level of the protein.

**The X-ray structure of fully deuterated C*c*P at 100 K** The structure of fully deuterated C*c*P in the resting state at 100 K was determined by X-ray crystallography and used for comparison with the hydrogenated counterpart 2YCG. X-ray induced photoreduction effects were monitored by UV-visible spectrophotometry. The reduced form of the ferric enzyme is known as ferrous state and shows characteristic absorption peaks in the UV-visible spectrum. The spectra of ferric samples before and after X-ray exposure are reported in Figure 8.3 showing no wavelength shift of absorption peaks. The UV-visible analysis doesn't show any feature ascribable to radiation-induced reduction of the metal center. The structure of perdeuterated C*c*P in the resting state at 100 K does not show appreciable differences compared to fully hydrogenated 2YCG. The root mean square distance between $C_\alpha$ of the two structures aligned with *Superpose* (*CCP4* software suite [13]) is 0.25 Å. In particular, the X-ray structure of perduterated ferric C*c*P shows that the Fe - O distance in the catalytic site is (1.9 ± 0.1) Å (see Figure 8.4) in agreement with 2YCG. The uncertainty on the distance was computed as reported in Section 5.3.

**The protonation state of catalytic residue His-52** The neutron structure of perduterated C*c*P in the resting state at 100 K shows that $N_\epsilon$ of catalytic residue His-52 is deuterated with partial occupancy of the $D_\epsilon$ atom (Figure 8.5). This observation is in partial agreement with what was determined in the deuterium exchanged counterpart in Chapter 7 where $N_\epsilon$ is fully deuterated. As pointed out in Chapter 7, this is a major difference with respect to the structure of the resting state at 298 K where $N_\epsilon$ of His-52 is not deuterated.

**The nature of the heme axial ligand** The heme iron distal axial ligand in the structure of fully deuterated C*c*P at 100 K was identified as $OD^-$ with partial occupancy of

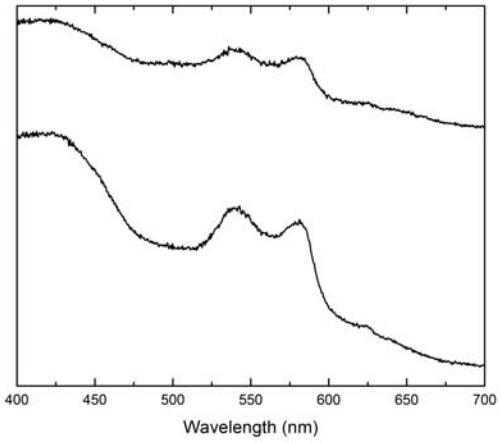


Figure 8.3: UV-vis absorption spectra of crystals of perdeuterated ferric C*c*P at 100 K. The vertical axis reports the dimensionless absorbance. Unexposed (top) and after X-ray data collection (bottom). For clarity the spectra have been shifted along the vertical axis. The spectra have been collected using two different crystals. The different volumes of the samples explain the difference in the observed peak intensities.

the deuterium atom as shown in Figure 8.6. This result was validated by testing different ligands. An oxygen ligand was introduced in a trial model giving rise after structural refinement to a positive peak in the neutron $F_o - F_c$ map appearing at $2.4\sigma$ as shown in Figure 8.7. Subsequently a $D_2O$ molecule was tested as a ligand resulting in a negative peak in the neutron $F_o - F_c$ map appearing at $2.8\sigma$ at the position of one deuterium atom as shown in Figure 8.8. Conversely no peak was observed in the neutron difference map upon introduction of an $OD^-$ ligand in the model. This is different from what was observed in the structure of C*c*P resting state at 298 K where the iron distal ligand is a $D_2O$ molecule. Figure 8.9 presents a schematic interpretation of the results.

**Dependence of UV-visible absorption on the protein's deuteration level** The presence of deuteration induced modifications in the structure of C*c*P active site was investigated by single crystal UV-visible spectrophotometry. Figure 8.10 presents the spectra at 100 K of hydrogenated, $D_2O$-soaked and fully deuterated C*c*P in the resting

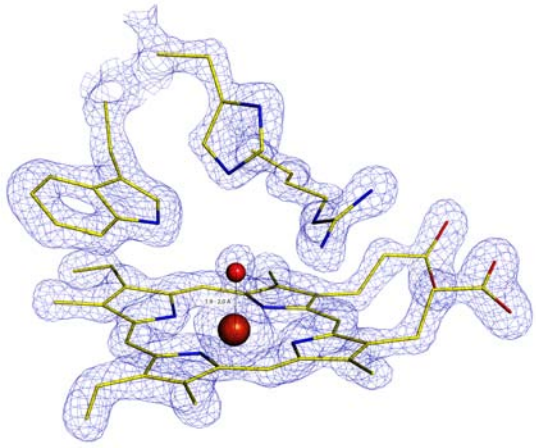

Figure 8.4: X-ray $2F_o - F_c$ map contoured at 2.5 rms for the distal heme pocket of perdeuterated C*c*P in the ferric state at 100 K. The Fe - O distance is (1.9 ± 0.1) Å. Iron is represented as a sphere in orange, oxygen in red, carbon in yellow and nitrogen in blue.

state. The three spectra appear to be similar showing that deuteration does not produce major changes in the structure, in agreement with what was determined by comparing the X-ray structures of hydrogenated and fully deuterated C*c*P . Nevertheless small changes appear and in particular the absorbance shoulder at 625 nm is suppressed in the deuterium exchanged and fully deuterated spectra.

## 8.3 Concluding remarks

Fully deuterated cytochrome *c* peroxidase in the resting state at 100 K was characterized by X-ray and neutron crystallography as well as by UV-visible spectrophotometry. X-ray crystallography and ultraviolet-visible spectrophotometry were employed primarily with the purpose of investigating the presence of deuteration induced structural changes. The X-ray structure didn't reveal any deuteration induced artifact when compared to the X-ray structure of the fully hydrogenated counterpart. UV-visible spectra also didn't show major changes when comparing different deuteration levels, although a modification in the absorption at long wavelengths was observed.

The neutron structure of perdeuterated C*c*P in the resting state at 100 K was deter-

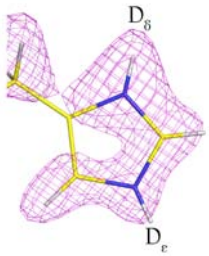


Figure 8.5: The structure of the catalytic residue His-52 in perdeuterated ferric C*c*P at 100 K. The nuclear scattering density [$\sigma_A$-weighted $2F_o - F_c$ contoured at 2.0 root mean square (rms)] is shown in magenta. Colour scheme: deuterium atoms, white; nitrogen, blue; carbon, yellow.

mined to a resolution of 1.83 Å allowing the unambiguous identification of the catalytic residues protonation states. The nature of the iron distal ligand was also investigated. The following structural features are of particular interest:

- The neutron structure of perdeuterated cytochrome *c* peroxidase in the resting state at 100 K shows that $N_\epsilon$ of catalytic residue His-52 is partially deuterated.
- Furthermore the iron distal axial ligand is an $OD^-$ group, with partial occupancy of the deuterium atom.

By comparison with the results presented in Chapter 6 it appears that the structure of the active site is temperature dependent. This result will be discussed further in Chapter 12.

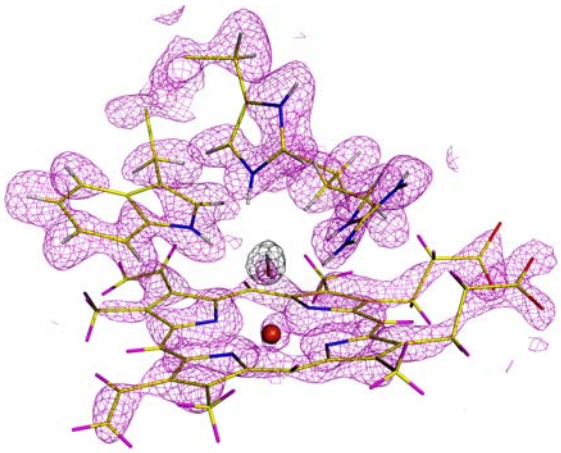

Figure 8.6: The structure of perdeuterated ferric C*c*P in the region of the heme at 100 K. Nuclear scattering density [$\sigma_A$-weighted $2F_o - F_c$ contoured at 2.0 root mean square (rms)] in the distal heme pocket is shown in magenta. The black contour [$\sigma_A$-weighted neutron $F_o - F_c$ map contoured at $4\sigma$] shows the difference density calculated by omitting the $OD^-$ ligand. Colour scheme: hydrogen atoms, magenta; deuterium, white; oxygen, red; nitrogen, blue; carbon, yellow; iron, orange.

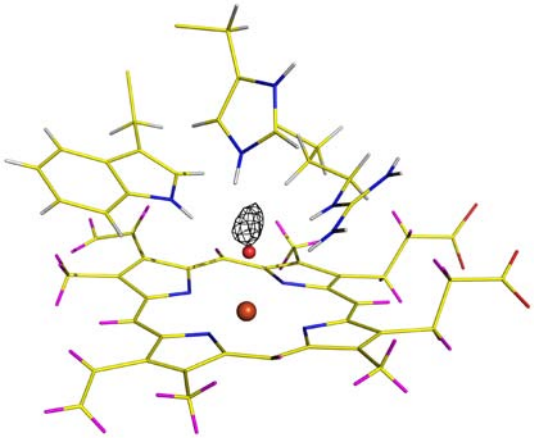

Figure 8.7: Determination of the nature of the heme iron distal ligand in perdeuterated ferric C*c*P at 100 K. Positive nuclear scattering density [$\sigma_A$-weighted $F_o - F_c$ contoured at 2.2 $\sigma$] is represented in black and shows the position of the deuterium atom of the heme iron distal ligand. Colour scheme as in Figure 8.6.

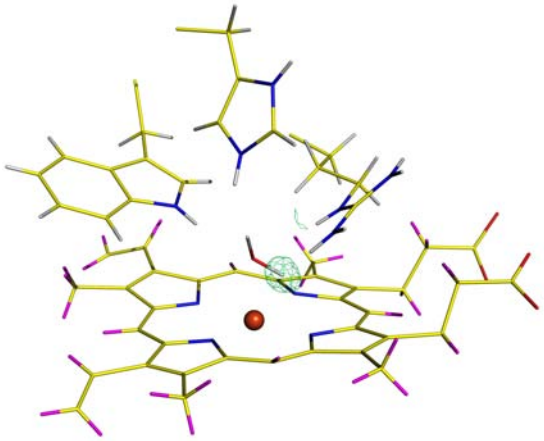

Figure 8.8: Determination of the nature of the heme iron distal ligand in perdeuterated ferric C*c*P at 100 K. Negative nuclear scattering density [$\sigma_A$-weighted $F_o - F_c$ contoured at 2.6 $\sigma$] is shown in green. Colour scheme as in Figure 8.6.

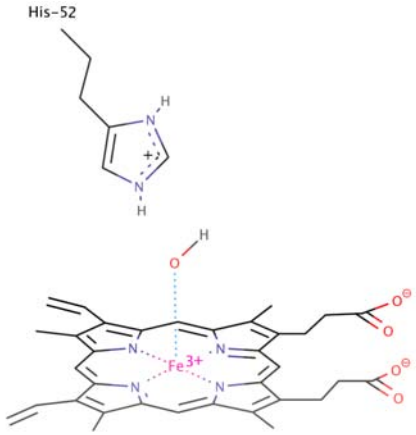


Figure 8.9: The protonation state of the heme iron distal ligand and catalytic residue His-52 in C*c*P resting state at 100 K.

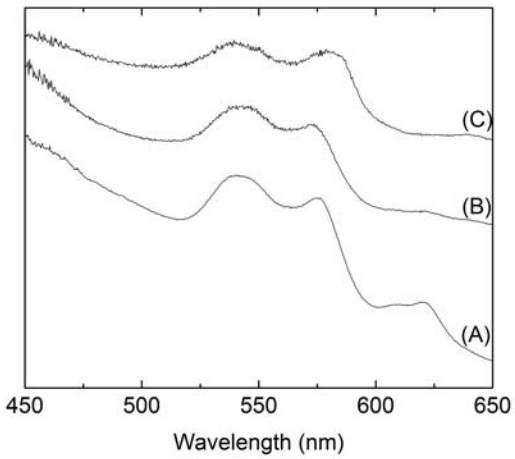


Figure 8.10: Single crystal UV-VIS absorption spectra of C*c*P in the resting state at 100 K at different deuteration levels. The vertical axis reports the dimensionless absorbance. (A) Fully hydrogenated, (B) deuterium exchanged and (C) perdeuterated. The vertical axis reports the optical density on an arbitrary scale. The spectra have been shifted along the vertical axis for clarity.

# Chapter 9

# Cytochrome *c* peroxidase compound I

This chapter presents the structural characterization of the transient intermediate compound I of cytochrome *c* peroxidase. The purpose of this work is the determination of the protonation state of the catalytic residues and of the iron axial ligand in the distal heme pocket in C*c*P compound I. The knowledge of such structural features of the active site is essential for the understanding of the catalytic mechanism in cytochrome *c* peroxidase.

Compound I is a transient species at 298 K. The exposure times employed in neutron crystallography are long compared to the half-life of cytochrome *c* peroxidase compound I at 298 K. For this reason neutron data collection in compound I was carried out at cryogenic temperature.

Complementary techniques were employed to assess compound I formation upon reaction with hydrogen peroxide and its stability at cryogenic temperature. In particular UV-visible spectrophotometry and electronic paramagnetic resonance (EPR) studies were carried out and both techniques were used to characterize C*c*P single crystals and in solution. Importantly, the structure of compound I determined at cryogenic temperature appeared to be relevant for the clarification of reaction mechanism in physiological conditions.

The experimental details of neutron and X-ray diffraction data collection, joint neutron and X-ray structural refinement as well as concise technical information concerning spectrophotometry and EPR data acquisition are reported in Section 9.1 of the present chapter. The results of diffraction and spectroscopic experiments are presented and discussed in Section 9.2. The results were published in [1] and the atomic coordinates were deposited in the Protein Data Bank under accession code 4CVJ. The interpretation of the structure of cytochrome *c* peroxidase compound I in terms of the reaction pathway is presented in Chapter 11.

## 9.1 Experimental methods

Neutron diffraction data were collected from a single crystal of $D_2O$-soaked C*c*P compound I at 100 K. The single crystal employed in neutron data collection is shown in Figure 9.1. X-ray crystallography data were collected at 100 K from an isomorphous crystal prepared by adopting the same procedure.

This section describes the experimental details of neutron and X-ray crystallography data collection, the procedure adopted in structural refinement and the technical details concerning the complementary techniques employed for the characterization of compound I.

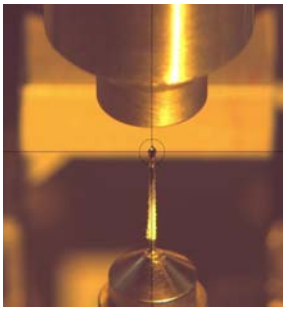

Figure 9.1: $D_2O$-soaked C*c*P compound I crystal employed in the neutron diffraction experiment at 100 K.

**Neutron data collection** A $D_2O$-soaked single crystal of C*c*P with volume of 1.2 mm x 0.8 mm x 0.7 mm was employed for neutron data collection in compound I. This catalytic intermediate was produced by reaction of C*c*P in the resting state with hydrogen peroxide [21]. Compound I is a transient species at 298 K. Nevertheless it can be considered stable at cryogenic temperature as established by the studies presented in section 9.2. The following experimental protocol was therefore adopted. C*c*P was reacted by soaking the single crystal in $H_2O_2$-containing mother liquor for 60 minutes at 277 K with peroxide concentration of 30 mM. The crystal was mounted on a LithoLoop on a magnetic cap (Molecular Dimensions) and it was cryo-cooled directly in a stream of nitrogen gas at 100 K from an Oxford Cryosystems cryostream. Monochromatic neutron diffraction data were collected at 100 K using the BIODIFF beamline at the FRM II research reactor at the Heinz Maier-Leibnitz Zentrum 2 (MLZ). Diffraction was observed to 2.5 Å resolution. A total of 201 frames of 0.3° oscillation and 120 minutes exposure per image were recorded at a wavelength of 3.39 Å. To increase data completeness the crystal was tilted by 45° from the rotation axis and a second series of 67 frames were collected at a wavelength of 3.98 Å with an oscillation of 0.3° and 120 minutes exposure. The diffraction data were indexed and integrated using *DENZO* (version 1.96.2) and scaled using *SCALEPACK* (version 2.3.6) [18]. Statistics are shown in Table 9.1 and the statistical indicators are defined in Section 4.3.

**X-ray data collection** X-ray diffraction data were collected at the Henry Wellcome Laboratories for Structural Biology of the University of Leicester. A $D_2O$-soaked single crystal of C*c*P was reacted as above and maintained at 100 K using an Oxford Cryosystems cryostream. A total of 270 images of 0.5° oscillation and 1 s exposure were collected using copper $K_\alpha$ radiation with wavelength $\lambda = 1.5418$ Å from a Rigaku 007 HF X-ray generator fitted with Osmic VariMax HF optics and a Rigaku Saturn 944+ CCD detector. Recorded intensities were indexed and measured using *iMOSFLM*, then scaled and merged using *AIMLESS* [13]. Statistics are shown in Table 9.2 and indicators are defined in Section 4.3.

Table 9.1: Neutron data collection statistics for C*c*P compound I at 100 K. Values in parentheses refer to the highest resolution shell and $d$ is the direct lattice interplanar distance.

| | |
|---|---|
| Space group | $P2_12_12_1$ |
| Cell dimensions | |
| a, b, c (Å) | 51.19 75.83 107.59 |
| **Data collection** | |
| Wavelength (Å) | 3.39 and 3.98 |
| Range of $d$ spacing (Å) | 50.0 - 2.5 (2.6 - 2.5) |
| $R_{merge}$ | 0.173 (0.482) |
| $I/\sigma(I)$ | 4.6 (1.5) |
| Completeness (%) | 90.7 (71.8) |
| Redundancy | 2.3 (1.7) |

**Model refinement** Structural refinement was carried out by using as starting model the X-ray structure of C*c*P compound I with Protein Data Bank accession code 2XIL. The starting model was modified to remove multiple conformations, the ferryl oxygen and solvent molecules. Structural refinement was performed using the *PHENIX* software suite [9]. The model was first refined by means of the X-ray terms alone. This was followed by joint X-ray and neutron structural refinement [10]. Hydrogen and deuterium atoms were introduced with the program *ReadySet* [9]. $D_2O$ molecules were added according to positive peaks in the $F_o - F_c$ nuclear scattering density map and model modifications were made by means of the modelling software *COOT* [11]. The position of the ferryl oxygen was determined based on the $F_o - F_c$ nuclear scattering density map only. 315 water molecules were added in total: 298 were modeled as full $D_2O$ molecules while 17 were partially disordered showing spherical density and were modeled as oxygen atoms only. Final refinement statistics are shown in Table 9.3 and indicators are defined in 5.1.

Table 9.2: X-ray data collection statistics for C*c*P compound I at 100 K. Values in parentheses refer to the highest resolution shell and $d$ is the direct lattice interplanar distance.

| | |
|---|---|
| Space group | $P2_12_12_1$ |
| Cell dimensions | |
| a, b, c (Å) | 51.19 75.83 107.59 |
| **Data collection** | |
| Wavelength (Å) | 1.5418 |
| Range of $d$ spacing (Å) | 17.00 - 2.18 (2.25 - 2.18) |
| $R_{merge}$ | 0.074 (0.164) |
| $I/\sigma(I)$ | 16.5 (5.9) |
| Completeness (%) | 99.3 (94.9) |
| Redundancy | 4.7 (2.6) |

**Single crystal UV-visible spectrophotometry** A single crystal of $D_2O$-soaked cytochrome *c* peroxidase in the resting state was cryo-cooled at 100 K and its UV-visible absorbance spectrum was registered by using the microspectrophotometry set-up described in Appendix D.1. The spectrum of the resting state at 100 K was used for comparison with that of C*c*P compound I at the same temperature. Compound I was produced by reaction with hydrogen peroxide by soaking the single crystal in mother liquor containing $H_2O_2$ 30 mM for 10 minutes [1] at 277 K. The crystal was subsequently cryo-cooled and its absorbance was measured as above. In single crystal spectra the error associated with the determination of peak positions amounts to a few nanometers due to the difficulty in localizing peaks in a spectrum with significant measurement noise.

**Solution UV-visible spectrophotometry** Cytochrome *c* peroxidase was characterized by solution UV-visible spectrophotometry at 298 K both in the resting state and in

[1] Soaking times were adapted to the crystal's volume.

Table 9.3: Neutron and X-ray structural joint refinement statistics for C*c*P compound I at 100 K. The subscripts n and X refer to neutron and X-ray diffraction data respectively.

| **Joint refinement** | |
|---|---|
| $d_{min,n}$ (Å) | 2.5 |
| $d_{min,X}$ (Å) | 2.2 |
| Number of unique reflections $N_n$ | 13661 |
| Number of unique reflections $N_X$ | 22053 |
| $R_{work,n}$ | 0.1873 |
| $R_{work,X}$ | 0.1488 |
| $R_{free,n}$ | 0.2720 |
| $R_{free,X}$ | 0.2052 |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.011 |
| Bond angles (°) | 1.261 |

transient compound I. The latter was produced by reacting the resting state with 1 equivalent of hydrogen peroxide. The UV-visible absorption spectra of cytochrome *c* peroxidase in solution were measured by means of a SHIMADZU UV-240IPC UV-visible recording spectrophotometer. The error associated with the wavelength determination is of the order of the wavelength step used in the data collection *i.e.* ±1 nm.

**Electronic paramagnetic resonance** Cytochrome *c* peroxidase compound I was characterized by electronic paramagnetic resonance (EPR) both in solution and in single crystals. EPR measurements were performed at the Photon Science Institute of the University of Manchester [2] by using the set-up described in Appendix D.2. In single crystals compound I was produced by soaking the crystal in mother liquor containing $H_2O_2$ 30 mM

[2]EPR measurements were carried out with the help of Dr Alistair Fielding and Maria Grazia Concilio.

for 60 minutes at 277 K. In solution compound I was formed by reacting the resting state with 5 equivalents of hydrogen peroxide. The mixing time was 15 seconds.

The EPR signal is normally reported as a function of the applied magnetic field or of the corresponding values of the Landé factor (g-values). The following equation defines the relationship between the two quantities:

$$g = \frac{h\nu}{\mu_B B} \tag{9.1}$$

where $h$ is Planck's constant, $\nu$ is the value of the microwave frequency employed in the experiment and equals 9.402 GHz in this case, $\mu_B$ denotes the Bohr magneton and $B$ refers to the applied magnetic field.

## 9.2 Interpretation of the results

This section describes the results concerning the structural and spectroscopic studies of cytochrome *c* peroxidase compound I. The neutron crystallography structural determinations are presented first, followed by the complementary studies by UV-visible spectrophotometry and electronic paramagnetic resonance.

### 9.2.1 Neutron crystallography

The neutron crystallography structure of C*c*P compound I allowed the unambiguous determination of the nature of the iron distal ligand and the protonation state of the catalytic residues including that of His-52 in the distal heme pocket.

The present section reports the description of the structural features of the active site in compound I as determined by neutron cryo-crystallography by adopting the procedure described in Section 5.2. The nuclear scattering density map of the active site is shown in Figure 9.2 while Figure 9.3 shows the electron density and the nuclear scattering density of the catalytic residues. At the end of the chapter, panel 9.13 reports side by side the neutron and X-ray maps of the distal and proximal heme pockets, for direct comparison, together with the model of the structure.

**The structure of the active site** The catalytic site of C*c*P compound I at 100 K presents the characteristics described hereafter. Compared to the enzyme in the resting state, most of the protonation states of the catalytic residues are retained in the structure of compound I, with the exception of His-52. In compound I His-52 shows deuteration at both $N_\delta$ and $N_\epsilon$ (see Figure 9.2 and 9.3 D). $N_\epsilon$ of the Trp-191 radical is deuterated: the side chain therefore constitutes a protonated cation radical [20]. Compared to the ferric enzyme, Arg-48 has moved into the heme pocket. A water molecule in the active site donates a hydrogen bond to the $\pi$ orbitals of the porphyrin ring and accepts a hydrogen bond from His-52 (Figure 9.13). The nuclear scattering density map indicates that the ferryl oxygen is not deuterated in the structure. This is consistent with the observed hydrogen bond network where the ferryl oxygen acts as a hydrogen bond acceptor to the $N_\epsilon$ of Trp-51 and to the $N_\epsilon$ of Arg-48, both of which are deuterated. This geometry can be interpreted in terms of a non-protonated Fe(IV)=O species. The $2F_o - F_c$ neutron map calculated in the absence of the ferryl oxygen shows a peak ($3.8\sigma$) in nuclear scattering density at $(1.6 \pm 0.4)$ Å from the iron atom (Figure 9.13, left) where the uncertainty was determined according to the procedure presented in Section 5.3. This interatomic distance is consistent with the most recent C*c*P compound I structures determined by X-ray crystallography [21], [22]. Nevertheless bond lengths provide only indirect information on the protonation state. Conversely the neutron structure shows unambiguously that the ferryl oxygen is not protonated. Figure 9.4 presents a schematic interpretation of the results.

### 9.2.2 UV-Visible spectrophotometry

This section describes the characterization of cytochrome *c* peroxidase compound I by UV-visible spectrophotometry both in single crystals and in solution. While UV-visible spectroscopy provides valuable information concerning the production of compound I by reaction with peroxide and its stability at cryogenic temperature, the results obtained by this technique are not unambiguous due to the similarity between compound I and compound II spectra in the region between 450 and 700 nm. Electronic paramagnetic

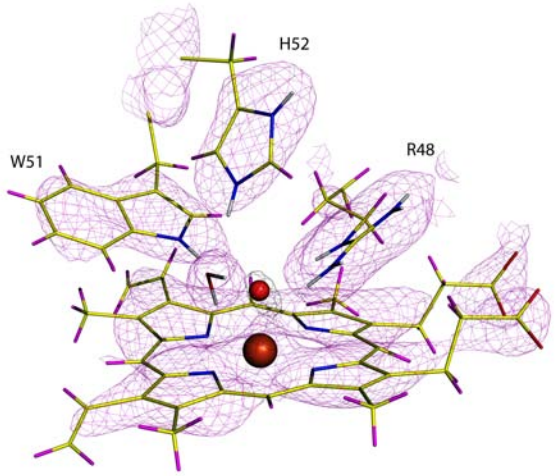


Figure 9.2: The structure of C*c*P compound I in the distal heme pocket at 100 K. The nuclear scattering density [$\sigma_A$-weighted $2F_o - F_c$ map contoured at 2.2 root mean square (rms)] is shown in magenta. The black contour [$\sigma_A$-weighted neutron $F_o - F_c$ map contoured at $3.5\sigma$] shows the difference density calculated by omitting the iron distal ligand. The following color scheme applies to the model of the structure: hydrogen atoms, magenta; deuterium, white; oxygen, red; nitrogen, blue; carbon, yellow; iron, orange sphere.

resonance measurements were therefore required to complement the UV-visible absorbance experiment.

**Assessment of compound I formation upon reaction with peroxide by single crystal UV-visible spectrophotometry** Figure 9.5 (A) reports the UV-visible spectra of C*c*P in the resting state at 100 K. In single crystals the Soret absorbance band (close to 400 nm) saturates due to the elevated optical density and the informative spectral region is therefore comprised between 450 nm and 700 nm. The resting state at 100 K shows absorbance peaks at wavelengths of 542 nm, 579 nm and 624 nm with errors amounting to a few nanometers. It is worth observing that the absorbance of C*c*P in the resting state is temperature dependent as can be seen by comparison with Figure 6.7 in Section 6.2. By reacting the single crystal with peroxide as described in Section 9.1 compound I is formed. The spectrum at 100 K is shown in Figure 9.5 (B). Compound I shows absorbance peaks

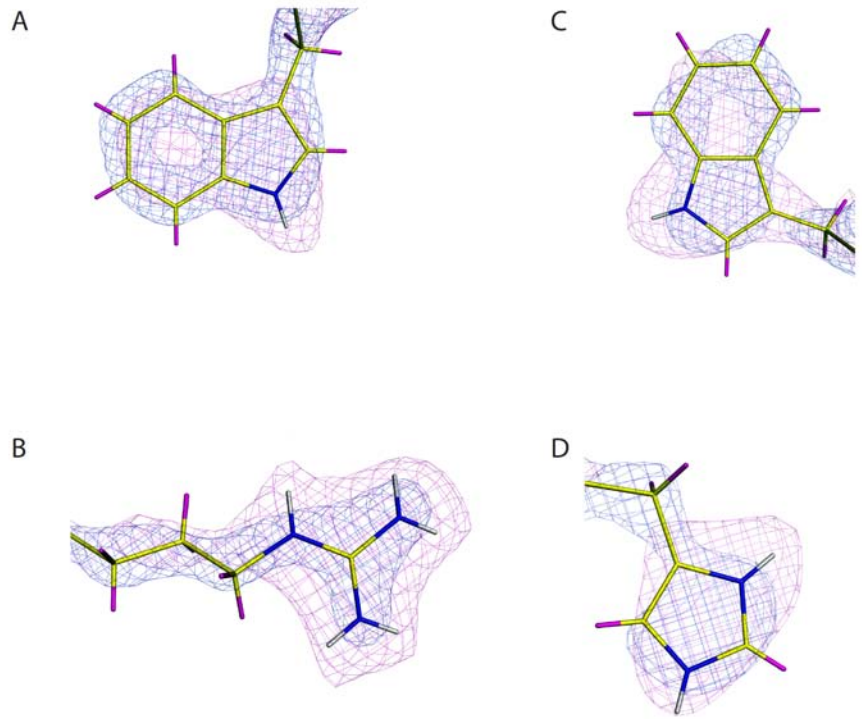


Figure 9.3: The side chains of the catalytic residues in C*c*P compound I at 100 K. The nuclear scattering density [$\sigma_A$-weighted $2F_o - F_c$ map contoured at 2.2 root mean square (rms)] is shown in magenta. The electron density [$\sigma_A$-weighted $2F_o - F_c$ map contoured at 2.6 rms] is shown in blue. Color scheme as in Figure 9.2. (A) Trp-51, showing that $N_\epsilon$ is deuterated. (B) Arg-48, showing the deuteration of all the nitrogen atoms of the guanidinium group. (C) Trp-191, showing that $N_\epsilon$ is deuterated. (D) His-52, showing deuteration at both $N_\delta$ and $N_\epsilon$.

at 529 nm, 560 nm and 630 nm with errors amounting to a few nanometers, in agreement with [21].

**Assessment of compound I stability at cryogenic temperature by single crystal UV-visible spectrophotometry** Figure 9.6 (A) shows the UV-visible absorption spectrum of a single crystal of C*c*P compound I at 100 K. After absorption data collection the crystal was recovered and stored in liquid nitrogen over 20 days. Figure 9.6 (B) shows the spectrum of the crystal after storage at 77 K. Comparison of the two spectra shows

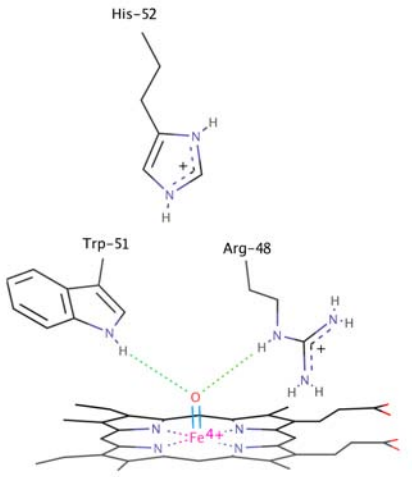


Figure 9.4: The protonation state of the heme iron distal ligand and catalytic residues in C*c*P compound I. The hydrogen bonds between the ferryl oxygen and the distal residures Trp-51 and Arg-48 are shown as dashed lines in green. The details of the heme macrocycle have been omitted for clarity.

that no decay occurs during the time delay employed in the experiment.

**Study of compound I temperature dependence by single crystal UV-visible spectrophotometry** As discussed in the case of cytochrome *c* peroxidase in the resting state, structural features of an enzyme's active site such as the protonation state of catalytic side chains can vary with temperature. Information concerning the reaction mechanism of an enzyme can be inferred by considering structural features determined at temperature close to that present in physiological conditions. In cytochrome *c* peroxidase the UV-visible spectrum is informative of the structure of the active site, in particular a small structural modification in the active site gives rise to a change in the absorbance in the visible region of the electromagnetic spectrum. This was described in Chapter 7 where the temperature dependence of the structure of the active site in the resting state of C*c*P was observed to give rise to temperature dependent spectral features in the visible region. Compound I crystallographic structure can only be accessed at cryogenic temper-

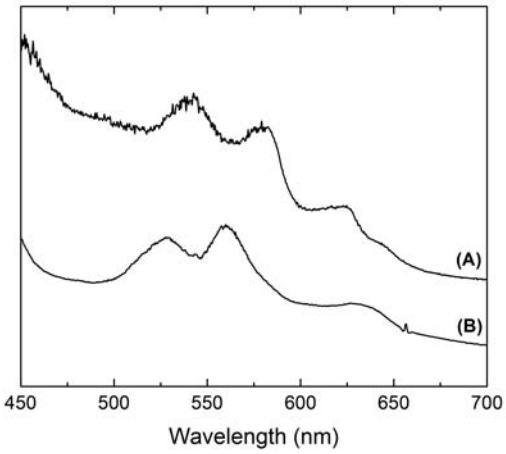


Figure 9.5: UV-VIS absorption spectrum of a single crystal of C*c*P in the resting state at 100 K (A) and of compound I at 100 K (B). The vertical axis reports the optical density.

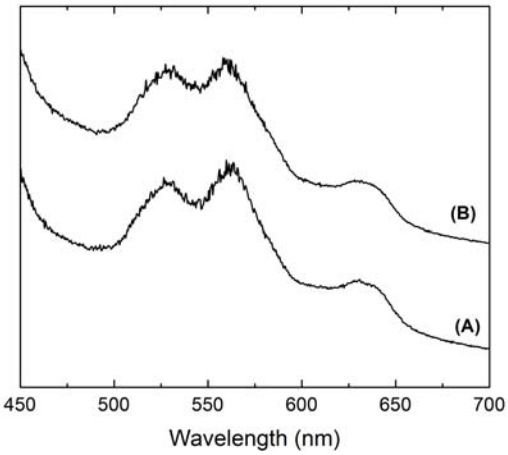


Figure 9.6: UV-VIS absorption spectrum at 100 K of a single crystal of $D_2O$-soaked C*c*P compound I (A). Spectrum of the same crystal after storage at 77 K for 20 days (B). The vertical axis reports the optical density.

ature since the exposure time required in the neutron diffraction experiment is longer by approximately two orders of magnitude than compound I half life at 298 K. Different from the resting state scenario, UV-visible spectrophotometry in compound I revealed that the active site structure is temperature independent in compound I as shown in Figure 9.7. This allowed the use of the neutron crystallography structure of compound I determined

at cryogenic temperature to extract information concerning the reaction mechanism of cytochrome *c* peroxidase in physiological conditions.

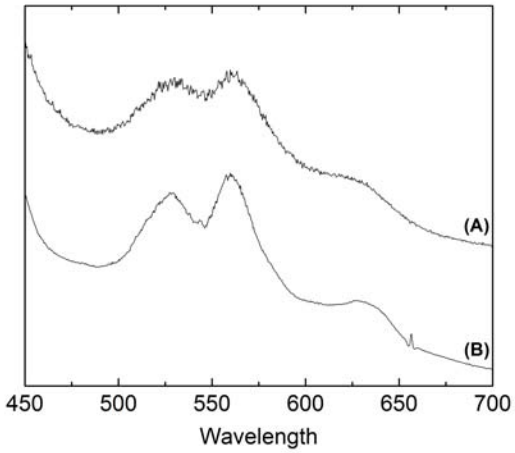


Figure 9.7: Single crystal UV-VIS absorption spectra of C*c*P compound I at 298 K (A) and at 100 K (B). The vertical axis reports the optical density on an arbitrary scale. The spectra have been shifted along the vertical axis for clarity.

**Assessment of compound I formation upon reaction with peroxide by solution UV-visible spectrophotometry** Formation of compound I in solution was assessed *via* UV-visible spectrophotometry by the methodology described in Section 9.1. Figure 9.8 shows the spectrum of C*c*P resting state at 298 K (blue) and of compound I at the same temperature (red). C*c*P resting state UV-visible spectrum at 298 K shows absorbance peaks at 408 nm, 506 nm, 540 nm (shoulder) and 640 nm with error of ±1 nm. Compound I spectrum at 298 K presents peaks at 419 nm, 530 nm, 560 nm and 627 nm (shoulder) with error of ±1 nm.

**Assessment of compound stability at cryogenic temperature by solution UV-visible spectrophotometry** Compound I solution was stored at 193 K for 130 days. The solution was thawed and the absorbance was measured at 298 K (green spectrum in Figure 9.9). No significant changes in the spectral features appeared. Several spectra were

measured at different time delays after thawing as shown in Figure 9.9. A blue shift of the Soret band was observed indicating the decay of compound I to the ferric state at 298 K.

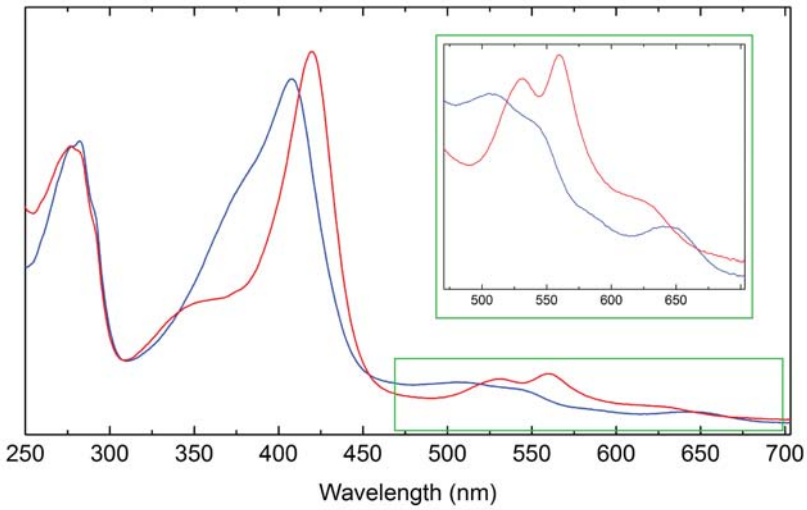


Figure 9.8: UV-VIS absorption spectra of C*c*P in solution at 298 K. The resting state is shown in blue and compound I in red. The vertical axis reports the optical density. Buffer: 10 mM $KH_2PO_4$, 150 mM KCl, pH 6.5.

**Determination of compound I half life in solution at 298 K** The stability of compound I in solution at 298 K was measured by reacting the resting state with hydrogen peroxide and by monitoring the decay of the absorbance at 419 nm over time. Figure 9.10 reports the absorbance at 419 nm as a function of time as determined experimentally [3] (blue dots). Data were fitted by a negative exponential with equation:

$$A_{(409\,\mathrm{nm})} = C_1 + C_2 e^{-kt}. \tag{9.2}$$

The best fit curve is shown in Figure 9.10 (solid line) where the fitting parameter $k$ assumes the value:

$$k = (6.5 \pm 0.1) \cdot 10^{-3}\,\mathrm{min}^{-1}. \tag{9.3}$$

[3]Data were collected by Dr. Jaswir Basran, Department of Biochemistry and Henry Wellcome Laboratories for Structural Biology, University of Leicester.

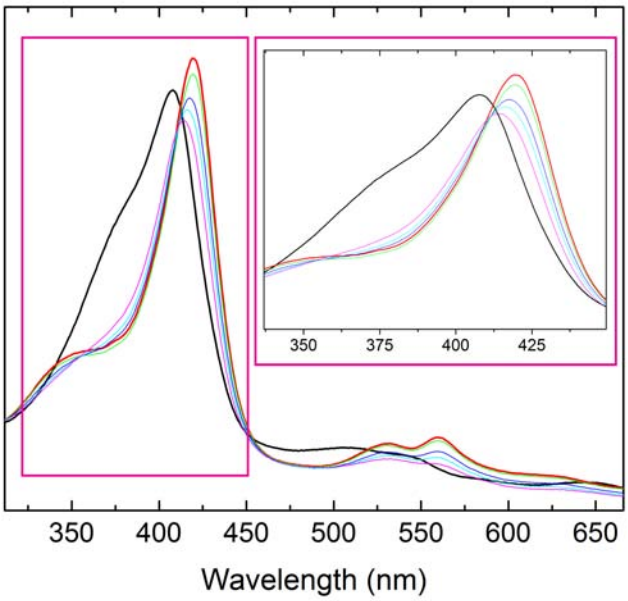


Figure 9.9: UV-VIS absorption spectra of C*c*P in solution at 298 K. The resting state (black spectrum) and compound I (red) spectra are shown. Compound I spectra have been recorded before and after storage at 193 K for 130 days (red and green spectra respectively). After thawing, compound I spectra have been recorded at 298 K at different time delays: 10 minutes, blue spectrum; 25 minutes, cyan and 66 minutes, magenta. The vertical axis reports the optical density. Buffer: 10 mM $KH_2PO_4$, 150 mM KCl, pH 6.5.

The half-life for the decay of compound I to ferric state at 298 K, defined as the value $\tau_{1/2}$ that fulfills:

$$e^{-k\tau_{1/2}} = \frac{1}{2} \tag{9.4}$$

amounts to:

$$\tau_{1/2} = 107 \pm 2\,\text{minutes}. \tag{9.5}$$

This result is in approximate agreement with early data [23].

### 9.2.3 Electronic Paramagnetic Resonance

The investigation of compound I formation by reaction of the resting state with peroxide and the assessment of compound I stability at cryogenic temperature performed by

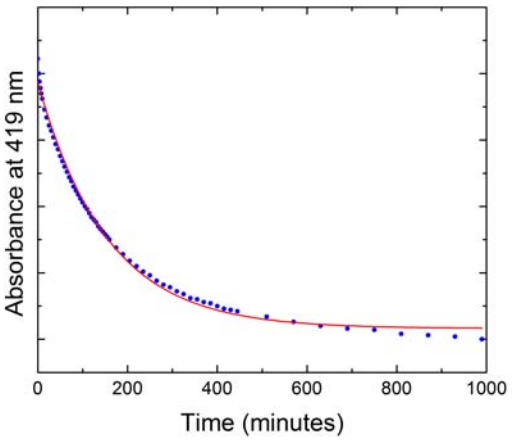


Figure 9.10: Absorbance at $\lambda = 419$ nm of C*c*P compound I in solution at 298 K as a function of time after reacting the resting state with 1.5 equivalents of hydrogen peroxide. Buffer: 10 mM $KH_2PO_4$, 150 mM KCl, pH 6.5.

UV-visible spectroscopy were complemented by electronic paramagnetic resonance measurements. There are two sources of ambiguity in cytochrome *c* peroxidase UV-visible experiments. First the spectrum of the resting state at 100 K is to some extent similar to that of compound I at 100 K in the spectral region accessible in single crystals, that is from 450 nm to 700 nm. In addition compound I and compound II UV-visible spectra are very similar [24].

By contrast EPR spectra permit to distinguish unambiguously the resting state and compound I, since both species exhibit characteristic signals. Nevertheless, it is worth pointing out that EPR-silent compound II cannot be detected in a direct way in EPR spectra.

An overview of electronic paramagnetic resonance in heme proteins is given in [25]. It is worthwhile mentioning that electronic paramagnetic resonance measurements are routinely carried out at temperatures below 10 K, this is required to limit the thermal population of the excited electronic Zeeman level in a magnetic field.

This section reports the study of compound I formation and stability at cryogenic temperature by EPR both in single crystals and in solution.

**Single crystal electronic paramagnetic resonance** Compound I formation by reaction with hydrogen peroxide in single crystals was studied by comparing the EPR signal of the resting state with that of a reacted crystal. Figure 9.11 shows the EPR spectrum of a single crystal in the resting state at 5 K (top) and that of a single crystal of C*c*P compound I at 8 K (bottom). The ferric state shows a structured signal localized near $g = 6$. This is the characteristic signal of high spin ferric [Fe(III)] heme in agreement with [26]. Compound I is characterized by Fe(IV) and a cation radical species localized on Trp-191. Due to the presence of the radical species, compound I exhibits a characteristic EPR signal at Landé factor $g$ close to 2, the free electron Landé factor.

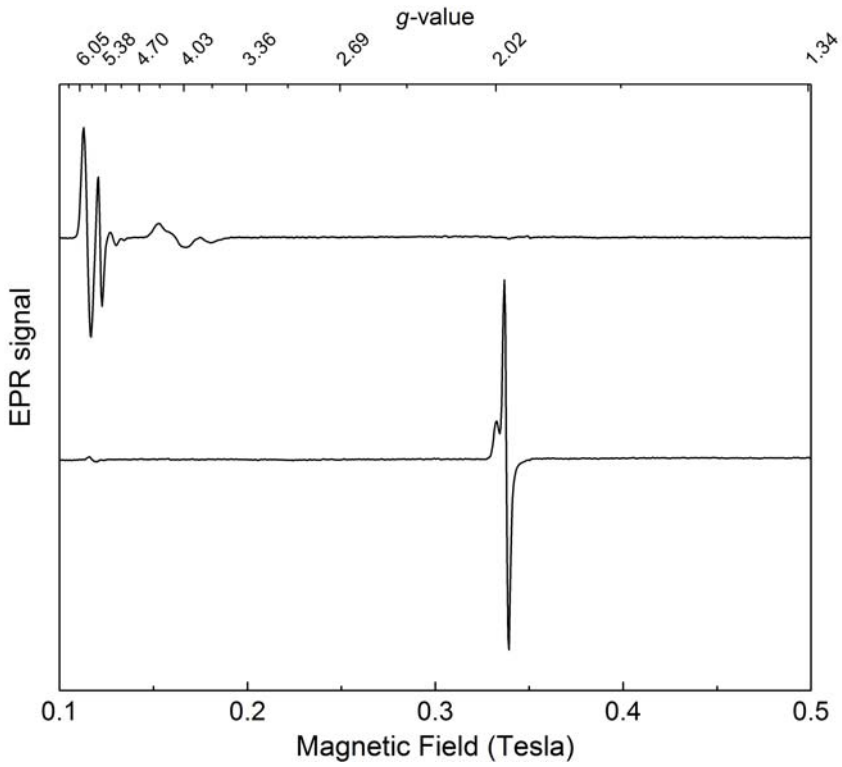


Figure 9.11: EPR spectra of a single crystal of C*c*P in the resting state at 5 K (top) and that of compound I at 8 K (bottom). The horizontal axis reports the magnetic field amplitude and the corresponding Landé factor values. The signal reported on the vertical axis is the derivative of the absorption signal. Both spectra were recorded with the following experimental conditions: 9.402 GHz, 0.5 mTesla modulation amplitude, 1 mW, 12 scans, 1024 points.

**Solution electronic paramagnetic resonance** Figure 9.12 (A) shows the EPR spectrum of a solution of C*c*P in the resting state at 4.5 K. C*c*P was reacted with hydrogen peroxide as described in 9.1 and the EPR spectrum was measured as shown in Figure 9.12 (B). The ferric EPR signal disappeared and compound I radical signal at Landé factor close to 2 was observed. The sample was stored in liquid nitrogen during 20 days. Figure 9.11 (C) shows the spectrum of the sample after storage. No signal decay was observed at $g$ close to 2 and no ferric signal arose at $g$ close to 6. Compound I appeared to be stable at 77 K.

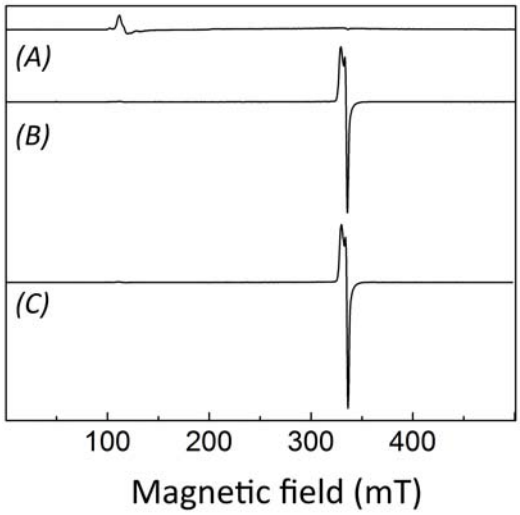


Figure 9.12: EPR spectra of C*c*P in solution at 4.5 K. (A) Ferric C*c*P. (B) C*c*P compound I immediately after mixing with hydrogen peroxide. (C) The same sample as in (B) after storage in liquid nitrogen for 20 days. Spectra were recorded with the following experimental conditions: 9.402 GHz, 0.1 mTesla modulation amplitude, 0.05 mW, 4 scans, 2048 points. Buffer: 10 mM $KH_2PO_4$, 150 mM KCl, pH 6.5.

## 9.3 Concluding remarks

The structure of the intermediate state compound I of cytochrome *c* peroxidase was investigated by neutron cryo-crystallography. The protonation states of the catalytic residues as well as the nature of the axial ligand of the heme iron in the distal pocket were iden-

tified. The following structural features are of particular relevance for the clarification of the catalytic pathway in cytochrome *c* peroxidase.

- The neutron structure of cytochrome *c* peroxidase compound I at 100 K shows that $N_\epsilon$ of catalytic residue His-52 is protonated.
- Furthermore the iron distal axial ligand is an oxygen atom.

These experimental determinations lead to new hypotheses concerning the reaction pathway from the resting state to compound I intermediate, as developed in Chaper 11

Complementary spectroscopic techniques were employed to assess compound I formation by reaction of the resting state with hydrogen peroxide. Compound I stability at cryogenic temperature was also verified. The absence of temperature dependence of compound I structural features as shown by UV-visible spectrophotometry suggests that the structure of compound I at cryogenic temperature is informative of the reaction mechanism in physiological conditions.

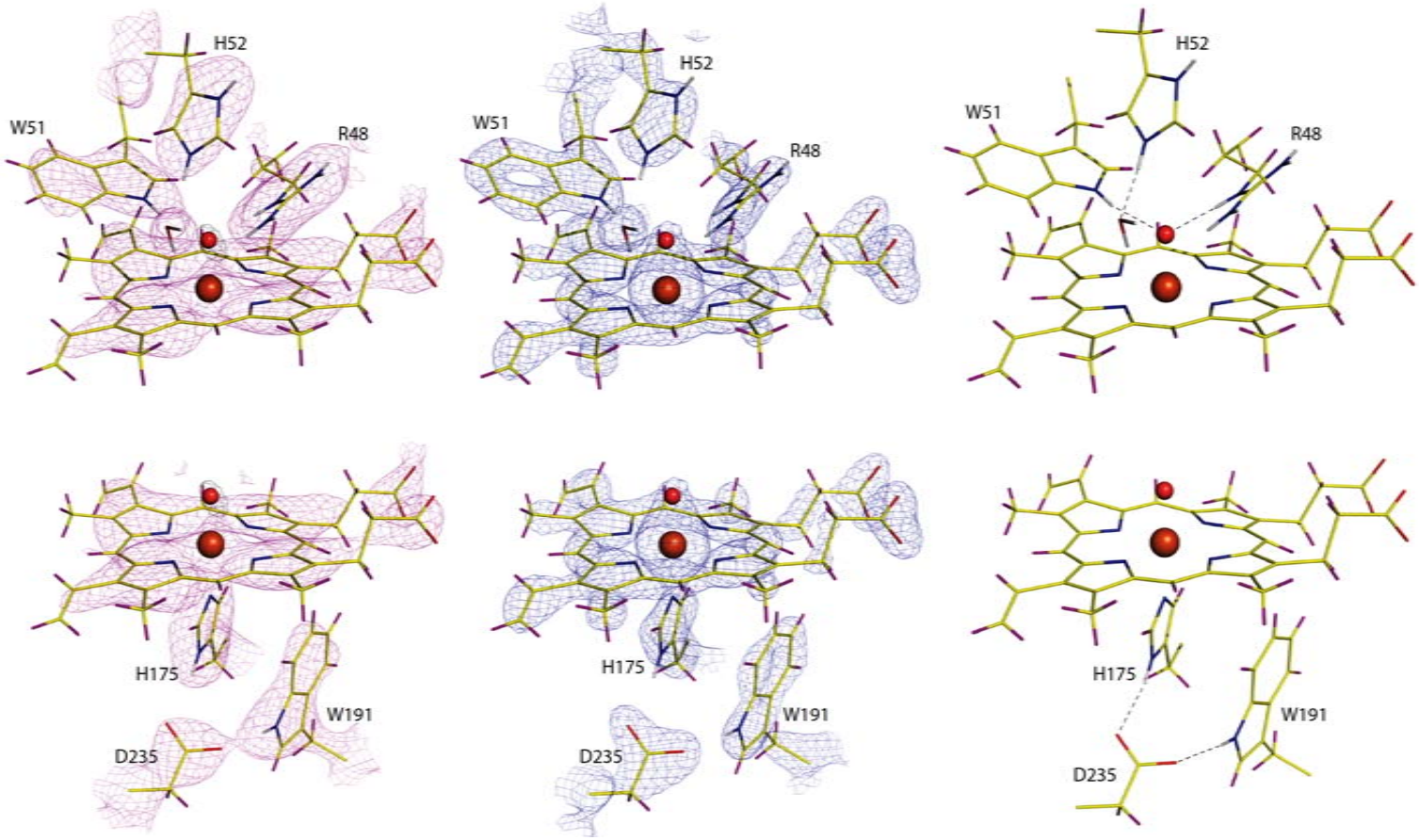


Figure 9.13: **The structure of C*c*P compound I in the region of the heme at 100 K.** Left: nuclear scattering density [$\sigma_A$-weighted $2F_o - F_c$ contoured at 2.2 root mean square (rms)] in the distal and proximal heme pocket is shown in magenta. The black contour [$\sigma_A$-weighted neutron $F_o - F_c$ contoured at $3.5\sigma$] shows the difference density calculated by omitting the ferryl oxygen. Centre: electron density [$\sigma_A$-weighted $2F_o - F_c$ contoured at 2.6 rms] in the distal and proximal pocket is shown in blue. Right: the model of the structure of the distal and proximal heme pocket. Color scheme as in Figure 9.2.

# Chapter 10

# Preliminary studies on C*c*P compound II

This chapter describes preliminary studies aimed at the structural characterization of the transient intermediate compound II of cytochrome *c* peroxidase. Neutron crystallography was employed with the purpose of establishing the protonation state of catalytic residues and of the heme iron distal ligand. The study of hydrogen-related structural features in compound II active site is necessary to complete the understanding of the catalytic mechanism of cytochrome *c* peroxidase (Section 1.2). The structural characterization of compound II is required to clarify the reaction pathway that leads from compound I to compound II and from compound II back to the resting state of the enzyme.

Like compound I, compound II is a transient species at 298 K and for this reason neutron data collection in compound II was carried out at cryogenic temperature. Complementary techniques were employed to assess compound II formation *via* chemical treatment of the resting state both in solution and in single crystals. In particular ultraviolet-visible spectrophotometry and electronic paramagnetic resonance studies were carried out.

Due to the experimental difficulties encountered, the work carried out within the present project consisted only of preliminary experiments, described in the following sections. The experimental details of neutron and X-ray diffraction data collection, X-ray

and neutron structural refinement as well as concise technical information concerning spectrophotometry and electronic paramagnetic resonance data acquisition are reported in Section 10.1 of the present chapter. The preliminary results of the experiments are presented in Section 10.2.

## 10.1 Experimental methods

Fully deuterated (perdeuterated) cytochrome *c* peroxidase was expressed, purified and crystallized as reported in Appendix B. Neutron diffraction data were collected from a single crystal of fully deuterated C*c*P compound II at 100 K. X-ray crystallography data were collected at 100 K from the same crystal subsequent to neutron data collection.

This section describes the experimental details of neutron and X-ray crystallography data collection, the procedure adopted in structural refinement and the technical information concerning the complementary techniques employed for the characterization of compound II.

**Neutron data collection** A fully deuterated single crystal of C*c*P with volume of approximately 1 $mm^3$ was employed in neutron data collection from compound II. This catalytic intermediate was produced in the single crystal by adopting a two-step reaction protocol by Dr H. Kwon (University of Leicester). The crystal in the resting state was first reacted with sodium dithionite for 60 s. This reducing agent produces the ferrous intermediate of the heme iron which corresponds to the oxidation state $Fe^{2+}$. Compound II was formed by reacting the ferrous species with hydrogen peroxide 0.2 mM for 60 s (Figure 10.1). Both steps were carried out in anaerobic conditions and the crystal was cryo-cooled by immersion in liquid nitrogen without exposure to oxygen.

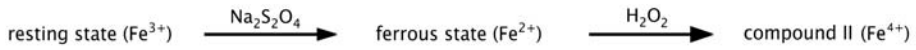


Figure 10.1: Formation of C*c*P compound II by reacting anaerobically the resting state with sodium dithionite and subsequently with hydrogen peroxide.

The crystal was transferred to the sample position of the LADI-III beamline [2] at the Institut Laue-Langevin. The crystal was preserved at cryogenic temperature by a stream of nitrogen gas at 100 K from the Oxford Cryosytems Cobra - Non Liquid Nitrogen Open Flow System (Section 3.2) of the instrument LADI-III. Quasi-Laue neutron diffraction data to 2.2 Å resolution were collected. The crystal was held stationary at a different spindle angle setting for each exposure. In total 30 images were collected with exposure time of 12 hours per image. Three different crystal orientations were adopted. After neutron data collection the crystal was recovered and used for subsequent X-ray data collection.

Neutron data were processed using the program *LAUEGEN* [3] modified to account for the cylindrical geometry of the detector [4]. The values of the unit cell parameters determined by X-ray crystallography where introduced as an input to the indexing algorithm. The software *LSCALE* [5] was employed to determine the wavelength normalization curve using the intensities of symmetry equivalent reflections measured at different wavelengths. No explicit absorption corrections were applied. Data were merged in *SCALA* [6]. Scaled and merged intensities were converted into structure factor amplitudes and associated error by means of the program *TRUNCATE* [7]. Data collection statistics are shown in Table 10.1 and indicators are defined in Section 4.3.

**X-ray data collection** X-ray diffraction data collection was carried out at the Henry Wellcome Laboratories for Structural Biology of the University of Leicester by Dr H. Kwon. Data were collected at 100 K from the same crystal employed in neutron data collection. Two data-sets of 360 and 220 images were collected with oscillation of 0.5° and exposure time of 5 s per image, using copper $K_\alpha$ radiation (wavelength 1.5418 Å) from a Rigaku 007 HF X-ray generator fitted with Osmic VariMax HF optics and a Rigaku Saturn 944+ CCD detector. The two data-sets were processed separately by means of the *CCP4* suite of programs [13]. Recorded intensities were indexed and measured using *iMOSFLM* [14] then scaled and merged using *AIMLESS* [15]. Finally, the two datasets were scaled and merged by using *AIMLESS*. Statistics are shown in Table 10.2 and indicators are defined

Table 10.1: Neutron data collection statistics for perdeuterated C*c*P compound II at 100 K. Values in parentheses refer to the highest resolution shell and $d$ is the direct lattice interplanar distance.

| | |
|---|---|
| Space group | $P2_12_12_1$ |
| Cell dimensions | |
| a, b, c (Å) | 50.9 75.6 106.4 |
| **Data collection** | |
| Wavelength (Å) | 3.04 - 4.05 |
| Range of $d$ spacing (Å) | 39.27 - 2.20 (2.32 - 2.20) |
| $R_{merge}$ | 0.194 (0.252) |
| $I/\sigma(I)$ | 6.2 (4.0) |
| Completeness (%) | 52.9 (33.4) |
| Redundancy | 3.8 (2.1) |

in Section 4.3.

**Model refinement** Structural refinement was performed by using as a starting model the structure with Protein Data Bank accession code 2XJ5. This starting model, which does not contain hydrogen or deuterium atoms in the polypeptide, was modified to remove the heme iron distal ligand and solvent molecules. Structural refinement was carried out using the *PHENIX* software suite [9]. The model was refined by means of the X-ray terms alone. The modelling program *COOT* [11] was used to introduce solvent molecules according to positive peaks in the $F_o - F_c$ electron density map and for all model modifications.

The X-ray refined model was modified to remove solvent molecules and employed as a starting point in refinement against neutron data. Joint X-ray and neutron refinement was also tested with poor results. Model refinement by using neutron diffraction data

Table 10.2: X-ray data collection statistics for perdeuterated C*c*P compound II at 100 K. Values in parentheses refer to the highest resolution shell and $d$ is the direct lattice interplanar distance.

| | |
|---|---|
| Space group | $P2_12_12_1$ |
| Cell dimensions | |
| a, b, c (Å) | 50.9 75.6 106.4 |
| **Data collection** | |
| Wavelength (Å) | 1.5418 |
| Range of $d$ spacing (Å) | 26.61 - 1.80 (1.84 - 1.80) |
| $R_{merge}$ | 0.103 (0.333) |
| $I/\sigma(I)$ | 15.6 (4.2) |
| Completeness (%) | 98.9 (85.6) |
| Redundancy | 10.2 (5.8) |

was carried out in *PHENIX* [9]. Data to a resolution of 2.3 Å were considered with the additional criterion of rejecting reflections with $|F|/\sigma(F) < 3$. Deuterium atoms were added with the program *ReadySet* [9]. $D_2O$ molecules were introduced according to positive peaks above $3\sigma$ in the $F_o - F_c$ nuclear scattering density map and model modifications were made by means of the modelling program *COOT* [11]. 251 $D_2O$ molecules were added in total. Final refinement statistics are shown in Table 10.3 and indicators are defined in Section 5.1.

**Complementary techniques** Complementary experiments were carried out by Dr H. Kwon (University of Leicester). Single crystal ultraviolet-visible spectra were acquired at the Henry Wellcome Laboratories for Structural Biology of the University of Leicester by using the spectrophotometry set-up described in Appendix D.1. Compound II was formed in fully deuterated C*c*P single crystals by reaction with sodium dithionite for 60 s and subsequent treatment with hydrogen peroxide 0.2 mM for 60 s. Both steps were carried out

Table 10.3: Neutron structural refinement statistics for perdeuterated C*c*P compound II at 100 K.

| **Model refinement** | |
|---|---|
| $d_{min}$ (Å) | 2.3 |
| Number of unique reflections | 9983 |
| $R_{work}$ | 0.2568 |
| $R_{free}$ | 0.3273 |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.001 |
| Bond angles (°) | 0.455 |

in anaerobic conditions. The crystal was subsequently cryo-cooled and its absorbance at 100 K was measured. In single crystal spectra the error associated with the determination of peak positions amounts to a few nanometers due to the difficulty in localizing peaks in spectra with significant measurement noise.

Solution electronic paramagnetic resonance measurements were performed at the Photon Science Institute of the University of Manchester with Dr. A. Fielding by using the set-up described in Appendix D.2.

## 10.2 Interpretation of the results

This section presents the results concerning the study of cytochrome *c* peroxidase compound II. Compound II formation was successfully established by ultraviolet-visible spectrophotometry and electronic paramagnetic resonance as described in the following. The final part of this section concerns the neutron crystallography experiment, which did not produce data of sufficient quality to allow an interpretation of protonation states in the active site. The challenges associated with this experiment are discussed.

**Assessment of compound II formation in single crystals by UV-visible spectrophotometry** Compound II formation by reaction with dithionite and subsequent treatment with peroxide was assessed by UV-visible spectrophotometry in single crystals. The spectral features of compound II in the visible region are very similar to those of compound I. For this reason also the intermediate product of the reaction protocol was characterized by spectrophotometry. Figure 10.2 (A) shows the absorbance spectrum in the visible region of a single crystal of C*c*P treated with sodium dithionite. The resting state characterized by ferric heme ($Fe^{3+}$) is reduced to the ferrous state ($Fe^{2+}$) showing absorbance peaks at 517 nm, 558 nm and 585 nm, with errors amounting to a few nanometers. Figure 10.2 (B) shows the visible spectrum of compound II produced in anaerobic conditions by treating a crystal in the resting state with sodium dithionite and subsequently with hydrogen peroxide as described in 10.1. Compound II exhibits absorbance peaks at 530 nm, 560 nm, and 634 nm with errors amounting to a few nanometers, similar to compound I and in agreement with [21].

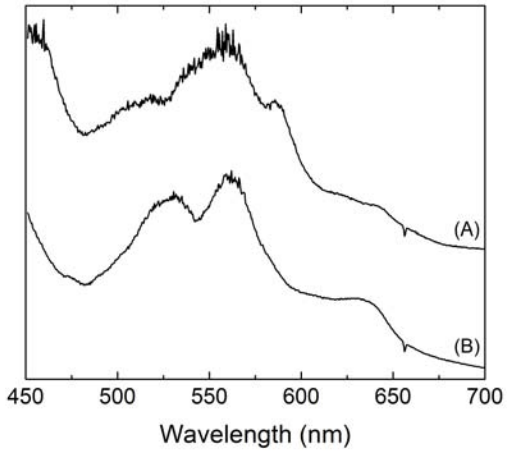


Figure 10.2: UV-VIS absorption spectrum of a single crystal of C*c*P at 100 K. (A) Ferrous state produced by chemical reduction of the resting state. (B) Compound II produced by reacting the ferrous intermediate with hydrogen peroxide. The vertical axis reports the optical density.

**Assessment of compound II formation in solution by electronic paramagnetic resonance** Compound II formation by reaction with dithionite and peroxide in anaerobic

conditions was assessed by solution electronic paramagnetic resonance (EPR) at 5 K. Both compound I and compound II were prepared and the EPR signal was recorded at 5 K. Figure 10.3 (A) presents the EPR spectrum of compound I showing the Trp-191 cation radical signal with Landé factor close to 2. Figure 10.3 (B) reports the spectrum of EPR-silent compound II.

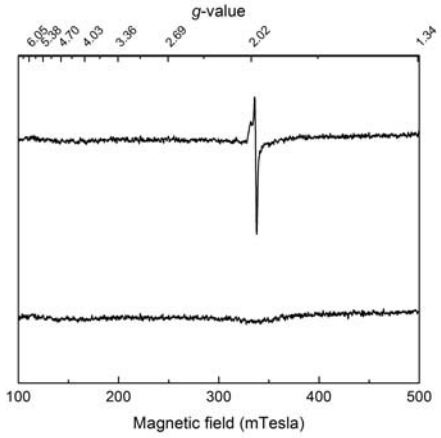


Figure 10.3: Electronic paramagnetic resonance spectra of C*c*P in solution at 5 K. (A) Compound I. (B) Compound II. Both spectra were recorded with the following experimental conditions: 9.402 GHz, 0.1 mTesla modulation amplitude, 100 kHz, 1 mW, 2 scans, 2048 points.

**Neutron crystallography of C*c*P compound II** C*c*P compound II neutron scattering density maps did not allow the structural characterization of deuterium-related features of the active site. Unlike the resting state of the enzyme and the transient species compound I, the formation of compound II in cytochrome *c* peroxidase required chemical treatment of the single crystal with sodium dithionite. This is a strong reducing agent and it was observed that chemical treatment by dithionite perturbs the crystal order. This type of damage affected neutron diffraction to a greater extent compared to X-ray diffraction. In fact in the neutron diffraction experiment the beam diameter was set according to the crystal size with the purpose of illuminating the whole crystal volume (approximately 1 $mm^3$ in compound II). On the other hand, in the X-ray crystallography experiment only a small portion of the crystal was exposed.

Neutron diffraction spots were observed to a resolution of 2.2 Å *i.e.* to reciprocal lattice indexes corresponding to real space minimum interplanar distance of 2.2 Å. This value of resolution is generally considered a very good achievement in a neutron crystallography experiment. However in the case of compound II spots were observed to be streaked hampering the accurate measurement of intensities. In addition the completeness of neutron data collection was poor *i.e.* only a small fraction of reciprocal space was sampled. The number of unique reflections used in model refinement was small, compared to the number of refinable parameters (4 parameters per atoms and over 5000 atoms, including solvent). These factors hampered the quality of compound II neutron scattering density maps leading to the impossibility of unambiguous interpretation of protonation states.

## 10.3 Concluding remarks

The study of the transient intermediate compound II of cytochrome *c* peroxidase was limited to preliminary work within the present project. It was possible to demonstrate compound II formation in C*c*P by reaction of the resting state with sodium dithionite and subsequent oxidation by peroxide in anaerobic conditions. It was also possible to produce diffracting crystals of cryo-cooled, fully deuterated compound II. However the structural characterization of compound II by neutron crystallography was hampered by the difficulty in obtaining good quality diffraction data. In particular, data completeness was very poor, resulting in a small number of unique reflections used in refinement. In addition, the accurate integration of streaked diffraction spots was a challenge, hampering the reliability of results. The formation of neutron diffraction spots with distorted shape was attributed to the use of sodium dithionite in compound II preparation.

As an alternative to the use of sodium dithionite, compound II can be produced in a single crystal of C*c*P by photoreduction of compound I in an X-ray beam [21]. This approach might offer a viable alternative to chemical treatment. In this method the capability of assessing the X-ray dose required to reduce compound I to compound II without producing a significant amount of the fully reduced ferrous species would be critical. In

this case the combined use of ultraviolet visible spectrophotometry and electronic paramagnetic resonance would be required. The two techniques are in fact complementary. The first does not allow to distinguish between compound I and compound II but presents different features in the ferrous state. Conversely, the latter highlights compound I presence, while both compound II and the ferrous species are silent in EPR measurements.

# Bibliography


[1] C. M. Casadei; A. Gumiero; C. L. Metcalfe; E. J. Murphy; J. Basran; M. G.Concilio; S. C. M. Teixeira; T. E. Schrader; A. J. Fielding; A. Ostermann; M. P. Blakeley; E. L. Raven; P. C. E. Moody,*Science* **345**, 193-197 (2014).

[2] M. P. Blakeley, S. C. Teixeira, I. Petit-Haertlein, I. Hazemann, A. Mitschler, M. Haertlein, E. Howard, A. D. Podjarny, *Acta Crystallogr. D* **66**, 1198-1205 (2010).

[3] J. W. Campbell,*J. Appl. Cryst.* **28**, 228-236 (1995).

[4] J. W. Campbell, Q. Hao, M. M. Harding, N. D. Nguti, C. Wilkinson, *J. Appl. Cryst.* **31**, 496-502 (1998).

[5] S. Arzt, J. W. Campbell, M. M. Harding, Q. Hao, J. R. Helliwell, *J. Appl. Cryst.* **32**, 554-562 (1999).

[6] M. S. Weiss, *J. Appl. Cryst.* **34**, 130-135 (2001).

[7] S. French and K. Wilson, *Acta Cryst.* A**34**, 517-525 (1978).

[8] W. Kabsch, *Acta Crystallogr. D* Biol. **66**, 125-132 (2010).

[9] P. D. Adams, P. V. Afonine, G. Bunkczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart, *Acta Crystallogr. D* **66**, 213221 (2010).

[10] P. V. Afonine, M. Mustyakimov, R. W. Grosse-Kunstleve, N. W. Moriarty, P. Langan, P. D. Adams *Acta Crystallogr. D* **66**, 1153-1163 (2010).

[11] P. Emsley and K. Cowtan, *Acta Crystallogr. D* **60**, 21262132 (2004).

[12] A. E. Pond, G. S. Bruce, A. M. English, M. Sono and J. H. Dawson *Inorganica Chimica Acta* **275-276**, 250-255 (198).

[13] 'Collaborative Computational Project, Number 4, The CCP4 suite: Programs for protein crystallography.' *Acta Crystallogr. D* **50**, 760763 (1994).

[14] A. G. W. Leslie and H. R. Powell, *Evolving Methods for Macromolecular Crystallography* **245**, 41-51 (2007).

[15] P. R. Evans, *Acta Cryst. D* **62**, 72-82 (2006).

[16] Airlie J. McCoy, Ralf W. Grosse-Kunstleve, Paul D. Adams, Martyn D. Winn, Laurent C. Storonia and Randy J. Read *Journal of Applied Crystallography* **40**, 658-674 (2007).

[17] G. N. Murshudov, A. A.Vagin and E. J. Dodson, *Acta Cryst. D* **53**, 240-255 (1997).

[18] Z. Otwinowski and W. Minor, *Methods in Enzymology* **276**, 307-326 (1997).

[19] G. Pompidor, F. S. N. Dworkowski, V. Thominet, C. Schulze-Briese and M. R. Fuchs, *J. Synchrotron Rad.* **20**, 765-776 (2013).

[20] J. E. Huyett, P. E. Doan, R. Gurbiel, A. L. P. Houseman, M. Sivaraja, D. B. Goodin and B. M. Hoffman, *J. Am. Chem. Soc.* **117**, 90339041 (1995).

[21] A. Gumiero, C. L. Metcalfe, A. R. Pearson, E. L. Raven and P. C. E. Moody, *J. Biol. Chem.* **286**, 12601268 (2011).

[22] A. J. Sitter, C. M. Reczek and J. Terner, *J. Biol. Chem.* **260**, 75157522 (1985).

[23] J. E. Erman and T. Yonetani, *Biochim. Biophys. Acta* **393**, 350357 (1975).

[24] A. E. Pond, G. S. Bruce, A. M. English, M. Sono and J. H. Dawson, *Inorg. Chim. Acta* **275 276**, 250255 (1998).

[25] G. Palmer, The Porphyrins Volume IV, Chapter 6, Academic Press (1979).

[26] Christopher A. Bonagura, B. Bhaskar, Hideaki Shimizu, Huiying Li, M. Sundaramoorthy, Duncan E. McRee, David B. Goodin and Thomas L. Poulos, *Biochemistry* **42**, 5600-5608 (2003).

# Part IV

# Discussion and conclusions

# Chapter 11

# The revised reaction mechanism

The protonation states of the key catalytic residues and of the heme iron distal ligand were determined by neutron crystallography in the resting state of cytochrome *c* peroxidase and in the intermediate compound I. The present chapter describes the consequences of these findings on the formulation of the reaction mechanism from the resting state of the enzyme to the transient species compound I.

In the first part of this chapter the relevance of crystallographic studies in the clarification of the catalytic pathway in physiological conditions is discussed. In the second part two alternative formulations of the reaction pathway are proposed.

## 11.1 Considerations on cryo-crystallography

This section addresses the question whether the crystallography studies of cytochrome *c* peroxidase presented in this work are relevant for the purpose of clarifying the reaction mechanism. First, the enzyme was verified to be active in the single crystal: upon soaking the crystal with peroxide compound I was formed as verified by ultraviolet-visible spectrophotometry and electronic paramagnetic resonance. In addition, as extensively discussed in Chapter 1, cold neutrons are non-ionizing radiation. Hence the probed structure is not altered by the experimental probe. Last, the use of cryo-crystallography appears to perturb the structure of C*c*P in the resting state. The effect of temperature on the configuration

of the active site in C*c*P is described in Chapter 12. The relevance of the structures determined at cryogenic temperature in the context of the study of C*c*P catalytic mechanism is discussed hereafter.

**The resting state of C*c*P** Cold neutrons do not induce radiation damage in protein crystals and it was possible to determine the neutron crystallography structure of cytochrome *c* peroxidase at 298 K (Chapter 6). The protonation states in the active site of the enzyme were determined at this temperature close to physiological conditions. It is well established that the resting state of C*c*P is temperature dependent [1], [2]. In particular the ultraviolet-visible absorption spectra show a significant change with temperature suggesting a concomitant change in the active site features as shown in Figure 11.1. The structure of the resting state at cryogenic temperature of 100 K was determined (Chapter 7 and 8), showing fundamental differences in protonation states with respect to the structure at 298 K. While the temperature dependence of structural features in the resting state of C*c*P is discussed thoroughly in Chapter 12, it is worthwhile stressing here the implications of this dependence in terms of the study of reaction mechanism. Concerning the resting state of the enzyme, it is necessary to consider the structure at physiological temperature within the context of the investigation of reaction pathway, since this one differs from the structure at cryogenic temperature which is therefore not relevant in the study of the catalytic activity.

**The transient compound I of C*c*P** At ambient temperature the half-life of C*c*P compound I amounts to several hours (Chapter 9). Due to the transient nature of this intermediate with a short half-life compared to the time-scales of typical neutron crystallography experiments, neutron diffraction data collection at ambient temperature is not viable. Differently from the case of the resting state, cryo-crystallography in compound I appears to be justified as a consequence of the lack of temperature dependence of the features of this intermediate. Ultraviolet-visible spectrophotometry studies (presented in Chapter 9, image here reported for convenience in Figure 11.2) show that the spectral features in compound I do not change with temperature suggesting that the crystallographic

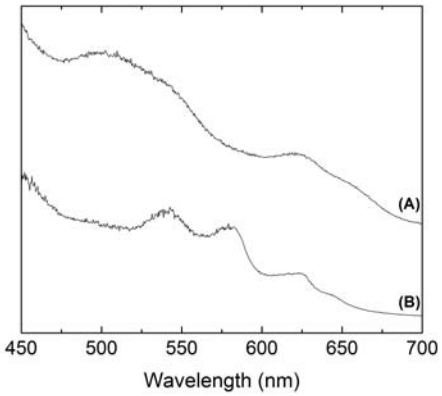


Figure 11.1: Single crystal ultraviolet-visible absorption spectra of C*c*P in the resting state at 298 K (A) and at 100 K (B). The vertical axis reports the optical density on an arbitrary scale. The spectra have been shifted along the vertical axis for clarity.

structure of the cryo-trapped intermediate is informative of the reaction mechanism in physiological conditions.

## 11.2 The catalytic pathway to compound I

The understanding of the mechanism of peroxide oxygen-oxygen bond cleavage by peroxidases leading to the formation of compound I has raised the interest of a number of investigators [3]. This portion of the catalytic cycle of cytochrome *c* peroxidase was invariably formulated by assuming that His-52 in the distal heme pocket is found in the neutral, single-protonated state both in the resting state and in transient compound I [4], [5]. This long-standing description of reaction mechanism assumes that His-52 first acts as a base-catalyst to deprotonate peroxide. Then His-52 acts as an acid-catalyst to protonate peroxide and release water. This mechanism is not compatible with the findings of the present work.

The neutron crystallography structures of cytochrome *c* peroxidase in the resting state and of the transient compound I revealed the protonation states of the residues in the active site and the nature of the heme iron distal ligand. In the resting state His-52 was

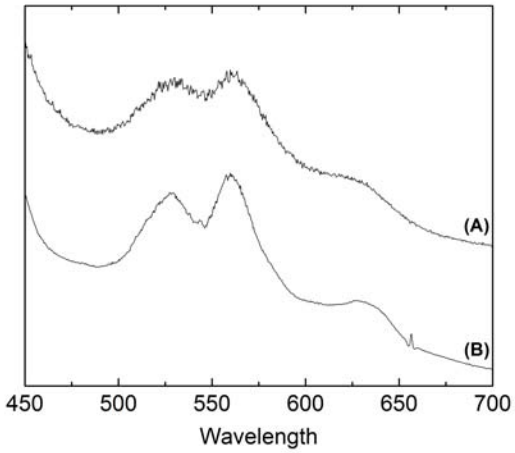


Figure 11.2: Single crystal UV-VIS absorption spectra of C*c*P compound I at 298 K (A) and at 100 K (B). The vertical axis reports the optical density on an arbitrary scale. The spectra have been shifted along the vertical axis for clarity.

observed to be single protonated and hence neutral (Figure 11.3, left) while the heme iron distal ligand was identified as a water molecule (Figure 11.4, left). In transient compound I His-52 was observed as double-protonated (Figure 11.3, right). This was unexpected and revealed that the traditional role of His-52 as a base-catalyst first and an acid-catalyst then, leading to a single protonated His-52 in compound I, needs to be reassessed. The heme iron distal ligand was identified as a non-protonated oxygen atom in compound I (Figure 11.4, right) providing an answer to the long-standing question concerning the nature (ferryl-oxo or ferryl-hydroxide) of this intermediate. These experimental findings lead to the formulation of two alternative reaction pathways from the resting state of C*c*P to the transient compound I.

**The proposed reaction mechanism (I)** The first hypothesis concerning the reaction pathway of cytochrome *c* peroxidase from the resting state to compound I is represented schematically in Figure 11.5. In the resting state of the enzyme $N_\epsilon$ of His-52 is not protonated (A). Due to the geometry of the region, $N_\epsilon$ lone-pair does not interact ideally

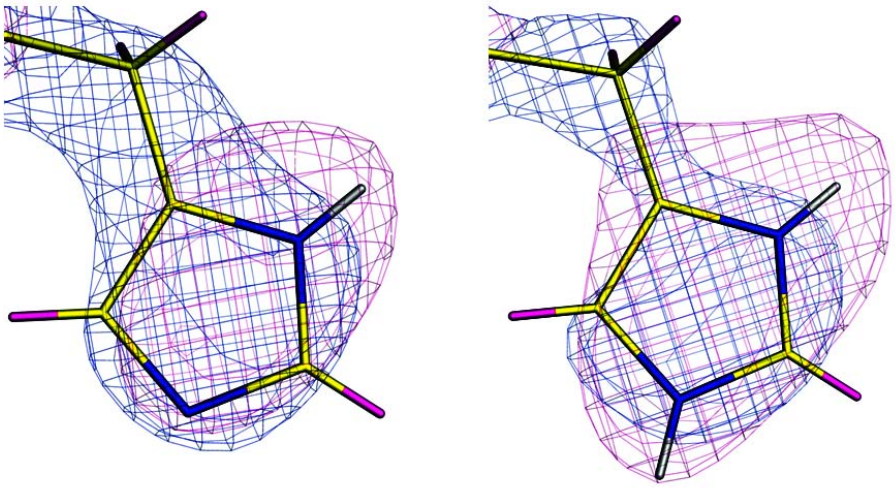

Figure 11.3: The catalytic residue His-52. Left: the resting state at 298 K with nuclear scattering density [$\sigma_A$-weighted $2F_o - F_c$ map contoured at 2.6 root mean square (rms)] in magenta and electron density [$\sigma_A$-weighted $2F_o - F_c$ map contoured at 2.4 rms] in blue. Right: transient compound I at 100 K with nuclear scattering density [$\sigma_A$-weighted $2F_o - F_c$ map contoured at 2.2 root mean square (rms)] in magenta and electron density [$\sigma_A$-weighted $2F_o - F_c$ map contoured at 2.6 rms] in blue. The following color scheme applies to the model of the structure: hydrogen atoms, magenta; deuterium, white; nitrogen, blue; carbon, yellow.

with any hydrogen bond partner. In the first step a hydrogen peroxide molecule replaces the water ligand in the active site forming a peroxide-bound complex (B). $N_\epsilon$ lone-pair interacts favorably with peroxide as shown in (B). His-52 acts as a base-catalyst and extracts a proton from peroxide. A proton is provided by the surroundings (C) producing the cleavage of the oxygen-oxygen bond and the release of one water molecule. Compound I is formed with a positively charged His-52 residue, in agreement with the experimental findings (D). The acid-catalyst role of His-52 occurs after compound I formation, so that a neutral His-52 residue is formed before a second turn-over of the enzyme takes place.

**The proposed reaction mechanism (II)** The second hypothesis is shown in Figure 11.6. Hydrogen peroxide replaces water in the active site and His-52 $N_\epsilon$ lone pair interacts favorably *via* hydrogen bonding with peroxide (B). A proton is supplied by the surround-

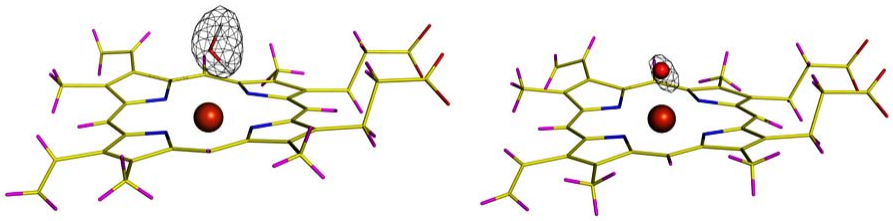

Figure 11.4: The nature of the heme iron distal ligand. Left: the resting state at 298 K with neutron omit map [$\sigma_A$-weighted neutron $F_o - F_c$ map contoured at $5\sigma$] in black. Right: transient Compound I at 100 K with neutron omit map [$\sigma_A$-weighted neutron $F_o - F_c$ map contoured at $3.5\sigma$] in black. The following color scheme applies to the model of the structure: hydrogen atoms, magenta; deuterium, white; oxygen, red; nitrogen, blue; carbon, yellow; iron, orange sphere.

ings and the oxygen-oxygen bond is cleaved. One water molecule is released producing an iron-hydroxide species (C). This species might be stabilized by the interaction with Arg-48 which is positively charged. A direct proton transfer occurs from the hydroxide ligand to the distal histidine $N_\epsilon$ that acts as a base-catalyst in this step producing compound I with a ferryl-oxo species and positively charged His-52 (D). The distal histidine $pK_a$ ($pK_a = -\log K_a$, where $K_a$ is the acid dissociation constant) was reported to have a value between 4 and 5 [4], [6]. The proton transfer from the ferryl hydroxide group to neutral His-52 would be possible if the ligand $pK_a$ was lower than that of His-52. This value is not known reliably for a hydroxide group in the heme environment.

In both scenarios active site water molecules or fully protonated Arg-48 may participate in proton relay networks providing the proton required during compound I formation [Figure 11.5 (C) and 11.6 (B)].

The main difference between the two mechanisms, in addition to the formation of an iron-hydroxide intermediate in pathway II, resides within the order of events. In pathway I His-52 is protonated [step (B) to (C)] before water is lost from the iron(III)-peroxy species. On the other hand in pathway II His-52 protonation follows the collapse of the iron(IV)-hydroxide intermediate [step (C) to (D)].

The experimental determination of the favored pathway is an intricate problem. For ex-

ample, it can be argued that replacement of positively charged Arg-48 for a neutral residue *via* site-directed mutagenesis would result in the suppression of the enzyme turnover rate if pathway II was the preferred route. In fact positively charged Arg-48 is likely to have a role in the stabilization of the iron-hydroxide species in pathway II. However inspection of the proposed mechanisms reveals that this experiment would not help in distinguishing between the two. In fact Arg-48 may be involved in proton relay networks in both scenarios and its removal would not be informative of the preferred route.

Stopped-flow spectroscopic measurements in the visible region might show the different species being formed if the acquisition rate is sufficiently rapid compared to the decay rate of the species of interest. In addition, numerical simulations may point to one of the routes being preferred.

## 11.3 Concluding remarks

The neutron crystallography study of cytochrome *c* peroxidase in the resting state and of the transient compound I lead to the unambiguous determination of protonation states in the active site. In the resting state the temperature dependence of the structural features implies that it is necessary to consider the crystallography structure determined at 298 K when the purpose is that of studying the reaction mechanism. The same does not hold in the case of compound I where ultraviolet-visible measurements do not show temperature-dependent features.

The neutron crystallography structure of compound I revealed that His-52 is double-protonated in this intermediate and this is not compatible with the reaction mechanism proposed before this work. Based on the experimental findings, two alternative reaction pathways were proposed, where His-52 retains a proton upon formation of compound I.

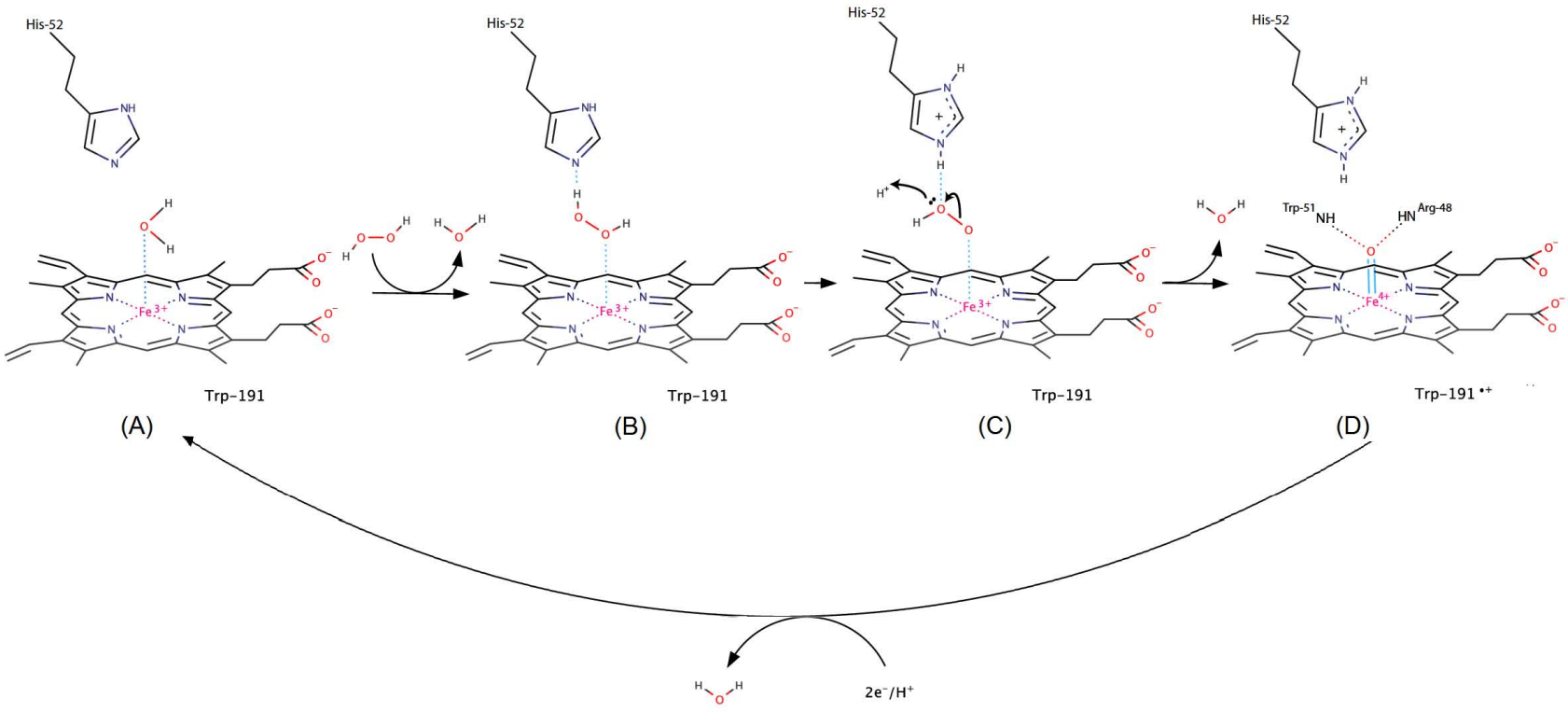


Figure 11.5: The reaction pathway leading to the formation of compound I of cytochrome *c* peroxidase as formulated in hypothesis (I) in the text.

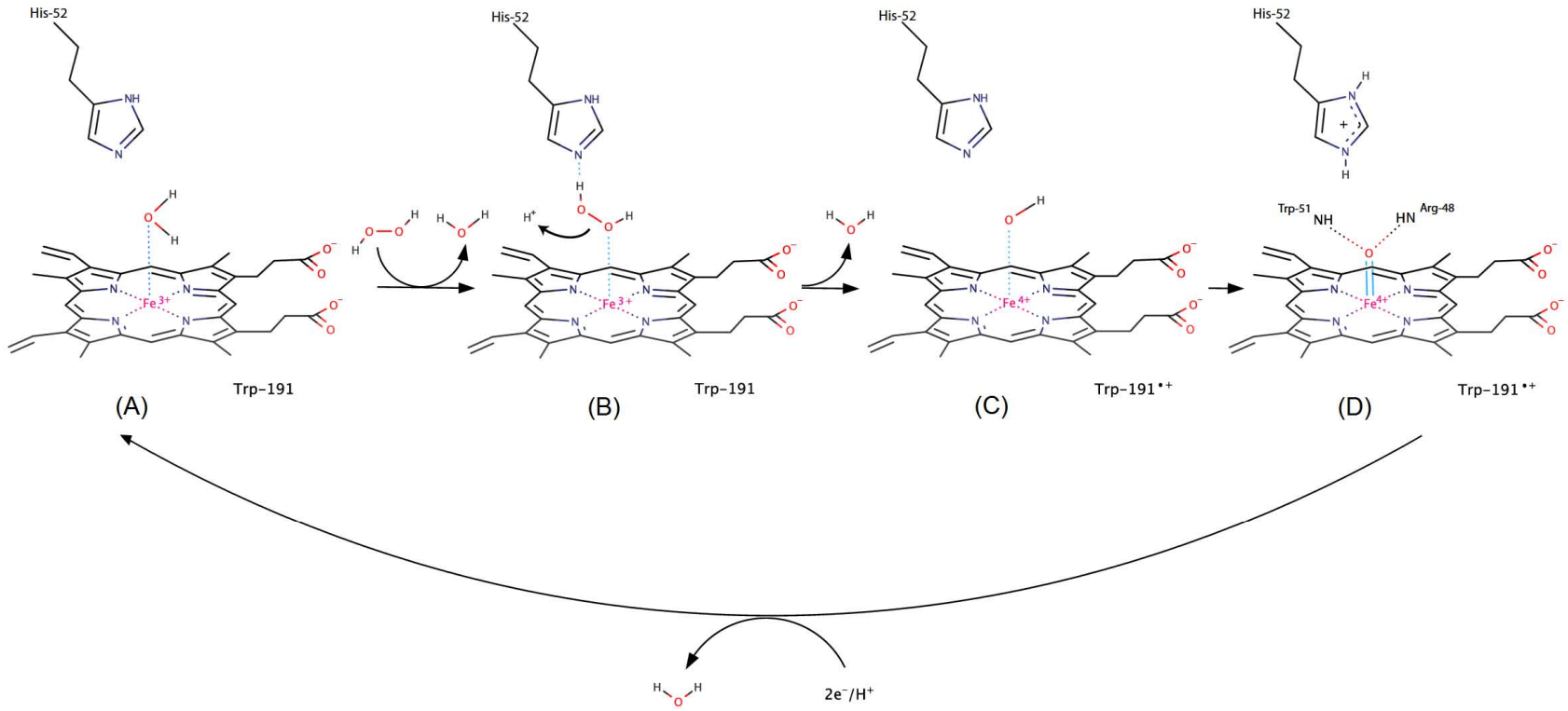


Figure 11.6: The reaction pathway leading to the formation of compound I of cytochrome *c* peroxidase as formulated in hypothesis (II) in the text.

# Chapter 12

# The temperature dependence of key catalytic features

The present chapter describes the effect of temperature on the structural features of the catalytic site of cytochrome *c* peroxidase in the resting state. The correlation between structural features and spin states of the heme iron at different temperatures is discussed.

## 12.1 Neutron crystallography study of temperature dependence

Temperature dependent effects have been observed in a number of biomolecular systems. These effects include alterations in ligand binding sites in protein structures [8], spin state changes in heme proteins [9] and variations in the number and position of water molecules bound to a protein structure [7]. These examples illustrate the diversity of temperature induced alterations in biomolecules.

In cytochrome *c* peroxidase the presence of temperature dependent features was recognized in early investigations [1], [2]. In the present work the temperature dependence of the structural features in the catalytic site of cytochrome *c* peroxidase was investigated primarily by neutron crystallography. This allowed the study of protonation states of the

catalytic residues and of the heme iron axial ligand in the absence of radiation damage both at ambient temperature and at cryogenic temperature of 100 K.

The neutron crystallography structure of the resting state of C*c*P at ambient temperature is presented in Chapter 6 while that of the resting state at cryogenic temperature of 100 K is described in Chapter 7. The structure of fully deuterated CcP in the resting state at 100 K is presented in Chapter 8. Significant differences have been observed by comparing the structure at 298 K to those at 100 K. In particular the protonation state of catalytic His-52 in the distal heme pocket and the nature of the heme iron axial ligand in the distal heme pocket proved to be temperature dependent.

**The protonation state of His-52** The neutron crystallography structure of cytochrome *c* peroxidase in the resting state at 298 K shows that the catalytic residue His-52 in the distal heme pocket is single-protonated (deuterated) at $N_\delta$ (Section 6.2) as shown in Figure 12.1 A. By contrast the structure of the resting state at 100 K reveals that this residue is double-protonated at $N_\delta$ and $N_\epsilon$ (Section 7.2) as reported in Figure 12.1 B. The structure of the fully deuterated version of the resting state shows that His-52 is double-protonated (deuterated) with $D_\epsilon$ site being partially occupied (Section 8.2) as shown in Figure 12.1 C.

Although it is not possible to exclude that variations in the extent of deuteration can lead to minor alterations in the structure, the difference in the occupancy of $D_\epsilon$ observed in the structures at 100 K with different deuteration levels is more likely ascribable to the difference in the quality of the refined structures (structural refinement statistics in Table 7.3 and 8.3). Lower resolution data are less likely to be informative of subtle details such as partial occupancy of atoms.

Since the resting state of C*c*P is temperature dependent, its structure at temperature close to physiological conditions should be considered when the purpose is that of investigating the reaction mechanism (Chapter 11). The protonation state of His-52 in the resting state has fundamental implications on the reaction pathway leading to the formation of compound I. Since it is single-protonated at physiological temperature, His-

52 assumes the role of a base-catalyst during compound I formation. This would not be possible in case of double protonation of this residue in the resting state.

**The nature of the heme iron distal ligand** The neutron crystallography structure of cytochrome *c* peroxidase in the resting state at 298 K allowed the identification of the heme iron distal ligand as a well ordered, fully occupied water molecule (Section 6.2) as reported in Figure 12.2 (left). As anticipated in Chapter 7, the structure of the resting state at 100 K did not lead to the unambiguous identification of the ligand. It was observed that at 100 K the neutron scattering density in the region of the ligand is not compatible with a well-ordered water molecule, but testing of different ligands in the active site did not provide a conclusive answer to the question concerning the protonation state of the iron-bound oxygen atom. The use of a fully deuterated sample allowed to improve data quality and it was possible to study the nature of the distal ligand *via* ligand testing as described in Section 8.2. In the structure determined at 100 K the scattering density in the ligand region was interpreted by modeling the ligand as an hydroxide group (Figure 12.2, right).

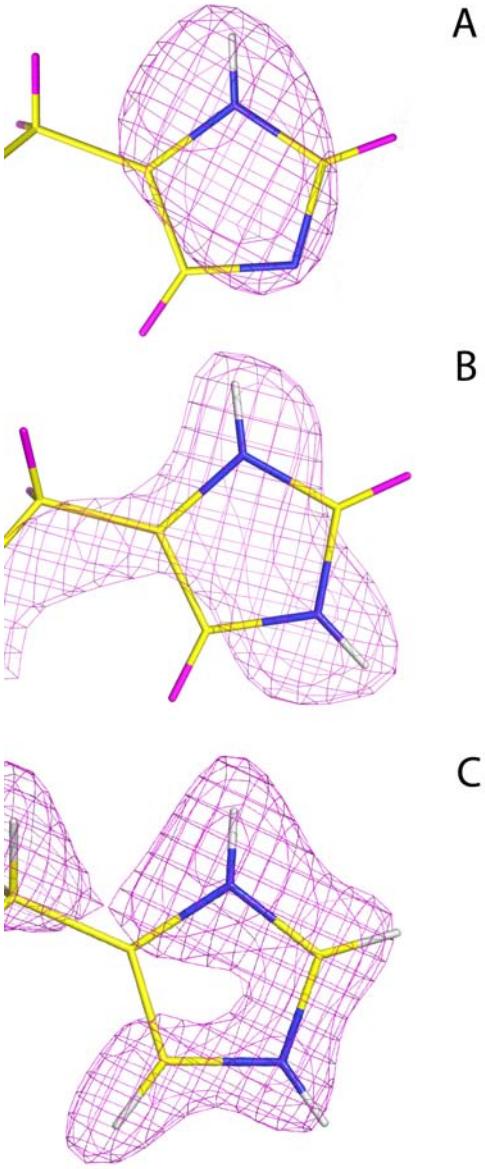


Figure 12.1: The protonation state of His-52 in the resting state of C*c*P . The nuclear scattering density map is shown in magenta. The following colour scheme applies to the model of the structure: hydrogen atoms, magenta; deuterium, white; nitrogen, blue; carbon, yellow. **(A)** Partially deuterated C*c*P at 298 K with $\sigma_A$-weighted $2F_o - F_c$ map contoured at 2.6 root mean square (rms). **(B)** Partially deuterated C*c*P at 100 K with $\sigma_A$-weighted $2F_o - F_c$ map contoured at 2.8 rms. **(C)** Fully deuterated C*c*P at 100 K with $\sigma_A$-weighted $2F_o - F_c$ contoured at 2.0 rms.

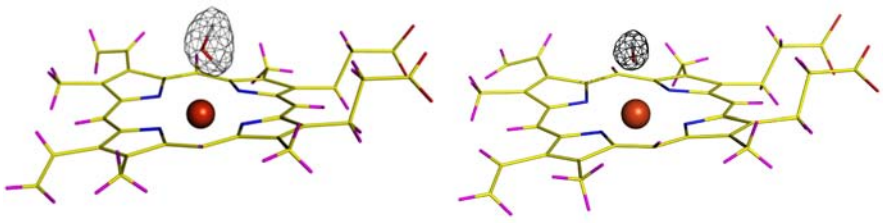

Figure 12.2: The nature of the heme iron distal ligand in the resting state of C*c*P . The following colour scheme applies to the model of the structure: hydrogen atoms, magenta; deuterium, white; nitrogen, blue; carbon, yellow; iron, orange. Left: partially deuterated structure at 298 K with $\sigma_A$-weighted neutron $F_o - F_c$ omit map contoured at $5\sigma$. Right: fully deuterated structure at 100 K with $\sigma_A$-weighted neutron $F_o - F_c$ map contoured at $4\sigma$.

## 12.2 The correlation between structural and spin state variations with temperature

The temperature dependence of absorbance peak positions in the visible spectra of C*c*P in the resting state is well established and its relation with a change in the spin state of the heme iron was discussed in the early days of C*c*P studies [1], [2]. This section illustrates the findings of the present work concerning the relationship between the nature of the heme iron distal ligand as determined by neutron crystallography and the iron spin state as studied by ultraviolet-visible spectrophotometry.

The ferric ion $Fe^{3+}$ is characterized by electronic configuration $[Ar]3d^5$. In heme proteins this ion constitutes the metal center of a distorted octahedral coordination complex and exists in different spin states. In crystal field theory ligands are approximated to point charges and the metal center-ligand interaction to a purely electrostatic coupling. The covalent character of the metal-ligand interaction is not considered and the resulting model is far from being realistic. Nevertheless this simplistic theory accounts for the existence of different spin states of the iron ion as a result of the energy splitting of d-orbitals produced by the octahedral ligand field. The electronic configuration of $Fe^{3+}$ in an octahedral coordination geometry can be either high-spin $t_{2g}^3 e_g^2$ (Figure 12.4 left) or low-spin

$t_{2g}^5$ (Figure 12.4 right).

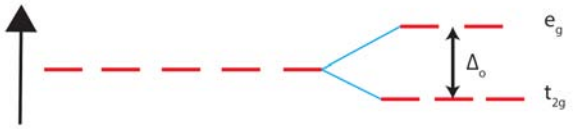


Figure 12.3: Energy splitting of the d orbitals of a transition metal ion in an octahedral ligand field. Left: free ion. Right: octahedral field. The vertical axis represents the direction of increasing energy. The d orbitals are split by the energy $\Delta_o$ in two groups labeled $e_g$ and $t_{2g}$ according to symmetry properties.

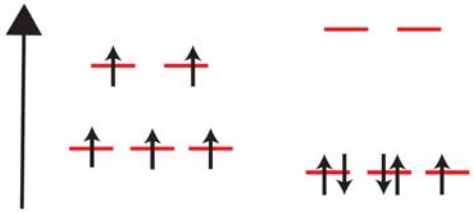

Figure 12.4: Spin configurations for the ion $Fe^{3+}$ in an octahedral ligand field. Left: high-spin configuration. Right: low-spin configuration. The vertical axis represents the direction of increasing energy.

It was established [10] that in cytochrome *c* peroxidase the species $Fe^{3+}$-$OH_2$ is characterized by a high-spin ground state and by a thermally populated low-spin excited state. On the other hand the species $Fe^{3+}$-OH is characterized by a low-spin ground state and by a thermally populated high-spin excited state.

Single crystal absorbance spectra of C*c*P in the resting state are presented as a function of temperature in Chapter 7. The spectra at 100 K and at 298 K are here reported for convenience in Figure 12.5. These measurements are in agreement with what was determined in [1], [2] where the absorbance features in the visible region were assigned to different spin states of the heme iron. In particular, the absorbance features in the spectrum recorded at 298 K were attributed to a prevalence of the high-spin species.

Conversely those in the spectrum at 100 K were assigned to a prevalence of the low-spin species.

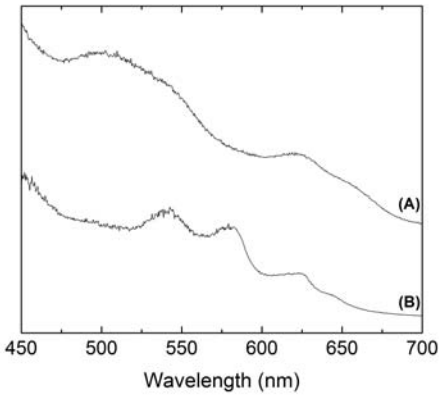


Figure 12.5: Single crystal ultraviolet-visible absorption spectra of C*c*P in the resting state at 298 K (A) and at 100 K (B). The vertical axis reports the optical density on an arbitrary scale. The spectra have been shifted along the vertical axis for clarity.

The structural findings concerning the nature of the heme iron distal ligand presented within this work provide direct evidence that the spin state variation observed by spectrophotometry in the visible region at 298 K and 100 K follows from the presence of different ligands at different temperatures. The neutron crystallography structure of the resting state at 298 K unambiguously identifies the heme iron ligand as a water molecule. As stated above, in these circumstances the high-spin configuration represents the ground state as observed by spectrophotometry in the visible region (Figure 12.5). On the other hand the neutron crystallography structure at 100 K reveals that the iron ligand is an hydroxide group. As anticipated above the low-spin state constitutes the ground state in this scenario, in agreement with the spectrophotometry study (Figure 12.5). These results are summarized schematically in Figure 12.6.

It is worthwhile observing that the electronic paramagnetic resonance spectrum of the resting state at 4.5 K shows a signal close to Landé factor $g = 6$ which is typical of high spin heme (Figure 9.11). The low spin ferric heme signal, which is expected to be the

dominant one due to the larger population of the low spin state at cryogenic temperature, is in fact suppressed by power saturation effects that take place at this temperature (*i.e.* long relaxation times of the low spin species). As a result only the EPR signal of the thermally populated high spin state is observed at 4.5 K, in agreement with [11]. EPR measurements at 100 K show that the low spin EPR signal is the dominant one at this temperature [11], in agreement with the discussion above.

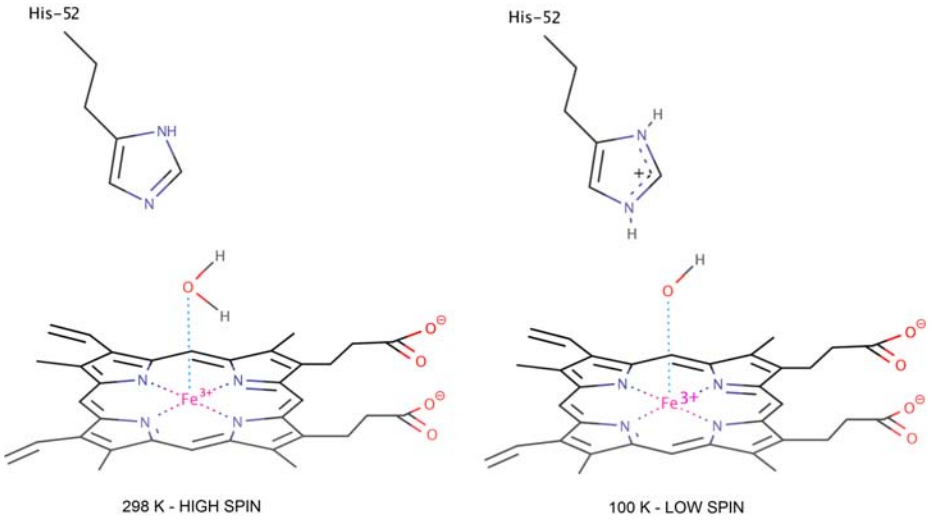


Figure 12.6: Schematic summary of the findings concerning the temperature dependence of structural features and spin states in C*c*P resting state. Left: 298 K. Right: 100 K.

## 12.3 Concluding remarks

The neutron crystallography structures of the resting state of cytochrome *c* peroxidase at 298 K and at 100 K showed remarkable differences. In particular the protonation state of catalytic His-52 and of the heme iron distal ligand were determined to be temperature dependent. The onset of cryo-artifacts in the structure of the active site of C*c*P in the resting state was investigated by neutron crystallography and the relationship with the temperature dependent absorbance features in the visible region was discussed in terms of the spin state of the heme iron.

It is worthwhile stressing as a general point that cryo-induced artifacts may cast doubt on the reliability of structural determinations at cryogenic temperature when the purpose is that of studying the physiological status. In the case of C*c*P resting state the temperature dependence of the structural features of the active site *i.e.* the presence of cryo-artifacts at 100 K, prevented the use of the structure determined at 100 K for the study of reaction mechanism. Instead, the structure determined at ambient temperature was employed for this purpose. The same does not hold in the case of the study of compound I whereby the use of the cryogenic temperature structure for the elucidation of reaction mechanism appeared to be legitimate as discussed in Chapter 11.

# Bibliography


[1] T. Yonetani, D. F. Wilson and B. Seamonds *The Journal of Biological Chemistry* **241** 5347-5352 (1966).

[2] T. Yonetani and A. Ehrenberg Magnetic Resonance in Biological Systems, Procedings of the second international conference held at the Wenner-Gren center, Stockholm, June 1966 155-158 (1967).

[3] T. L. Poulos *Chemical Reviews* **114**, 3919-3962 (2014).

[4] H. B. Dunford *Heme Peroxidases* Wiley, Chichester, UK (1999).

[5] H. B. Dunford *Peroxidases and Catalases: Biochemistry, Biophysics, Biotechnology and Physiology* Wiley, Chichester, UK, ed. 2 (2010).

[6] J. E. Erman, L. B. Vitello, M. A. Miller and J. Kraut *J. Am. Chem. Soc.* **114**, 65926593 (1992).

[7] M. P. Blakeley, A. J. Kalb, J. R. Helliwell, D. A. A. Myles *Proc. Natl. Acad. Sci.* **101**, 16405-16410 (2004).

[8] M. Fischer, B. Shoichet and J. Fraser *ChemBioChem* **16**, 1560-1564 (2015).

[9] A. Feis, M. P. Marzocchi, M. Paoli and G. Smulevich *Biochemistry* **33**, 4577-4583 (1994).

[10] T. Iizuka, M. Kotani and T. Yonetani *Biochimica et Biophysica Acta* **167** 257-267 (1968).

[11] T. Yonetani and H. Anni *The Journal of Biological Chemistry* **262** 9547-9554 (1987).

# Part V

# Appendix

# Appendix A

# Cytochrome *c* peroxidase expression, purification and crystallization

This appendix reports the protocol for the expression, purification and crystallization of cytochrome *c* peroxidase. Deuterium exchanged crystals employed in the neutron crystallography work were obtained either by crystallizing the protein in a $D_2O$ containing medium or by soaking fully hydrogenated crystals in such medium.

## A.1 Introduction

The MKT variant of cytochrome *c* peroxidase (C*c*P) [1] was cloned by the PROTEX service of the University of Leicester into the expression plasmid pLEICS-03. The plasmid carries kanamycin resistance and tobacco etch virus cleavable N-terminus His-tag.

Vector amplification was carried out by A. Gumiero by transformation into *E. coli* XL1 Blue supercompetent cells [2]. Plasmid DNA was isolated and transformed into *E. coli* BL21 DE3 Gold competent cells.

## A.2 Cytochrome *c* peroxidase expression

The expression of cytochrome *c* peroxidase was carried out by means of *E. coli* BL21 DE3 Gold expression system. Vector-carrying cells were selected throughout the expression procedure by exploiting their kanamycin resistance.

An agar-LB plate made up with 15 ml of agar-LB medium (see section C) and containing a final kanamycin concentration of 30 $\mu$g/ml was streaked with the transformed BL21 DE3 Gold *E. coli* cells. The plate was incubated overnight at 37 °C.

A single colony was used to inoculate 5 ml of sterilized LB medium (see section C) containing a final kanamycin concentration of 30 $\mu$g/ml. The culture was incubated for 6 hours at 37 °C and shaken at 225 rpm. The culture was used to inoculate 320 ml of sterilized LB medium (see section C) containing a final kanamycin concentration of 30 $\mu$g/ml. This was incubated overnight at 37 °C and shaken at 225 rpm.

Eight flasks containing 1 l of sterilized LB medium per flask (see section C) and a final kanamycin concentration of 30 $\mu$g/ml were prepared. An aliquot of 40 ml of culture was added to each flask. The flasks were incubated at 37 °C and shaken at 225 rpm until the optical density (O.D.) at wavelength of 600 nm reached the value of 0.8 absorbance units (A.U.). This process took approximatively two hours.

The temperature of the incubator was set at 22 °C. Protein expression was induced by adding a solution of 1 M $\beta$-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. The flasks were incubated at 22 °C and shaken at 225 rpm overnight.

The cells were harvested by centrifugation at 4500 rpm and 4 °C for 15 minutes. The precipitate was frozen and stored at −80 °C.

## A.3 Cytochrome *c* peroxidase purification

A two-step purification procedure was performed with little modification of the method previously reported in [3] and [4] yielding highly pure MKT-C*c*P. The procedure was carried out at 0 °C - 4 °C. Due to the high yield of the expression process the product of 4 l of bacterial culture was split in two or more parts before purification in order to avoid

the overload of chromatographic columns.

The frozen precipitate was suspended in 40 ml of buffer A (see section C). Protease activity was inhibited by adding 1.5 ml of phenylmethane-sulfonyl fluoride (PMSF) solution (see section C). The suspension was sonicated at 0 °C by eight sonication cycles of one minute each, with one minute of pause between cycles to prevent sample overheating. The suspension was spun at 19,000 rpm and 4 °C for 40 minutes. The supernatant was collected and 1.6 ml of hemin solution (see Appendix C) were added in order to induce protein reconstitution. The solution was gently stirred for two hours at 4 °C.

The first purification step was performed by means of nickel-NTA-agarose affinity chromatography (12 ml of resin). At high pH the histidine residues constituting the N-terminus His-tag of C*c*P bind the $Ni^{2+}$ ions. The protein was eluted by setting the pH at a value below the histidine $pK_a$ value.

The nickel-NTA-agarose column was equilibrated with buffer B (Section C) and the protein solution was loaded. The column was washed with 200 ml of buffer B followed by 500 ml of wash buffer (Section C). The protein was eluted by means of elution buffer (Section C) and the first 10 - 20 ml were discarded. The purity of the protein was assessed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) as shown in Figure A.1.

The pH of the protein solution was set to 6 by titration with $KH_2PO_4$ 1M at pH 10 and the solution was concentrated up to approximately 10 mg/ml. Protein concentration was determined *via* UV-visible spectrophotometry by using the known value of the extinction coefficient corresponding to the maximum of the Soret band $\epsilon = 135$ $mM^{-1}$ $cm^{-1}$.

A highly pure product was obtained by gel filtration chromatography using a GE Healthcare 16/60 HiLoad$^{TM}$ Superdex$^{TM}$ 75 chromatographic column. The product of 4 l of culture was split into several batches in order to prevent gel filtration column overloading. The column was previously equilibrated with FPLC buffer and the sample was run at 0.2 ml/min by using the same buffer (Section C).

Figure A.2 shows the chromatogram namely the 280 nm absorption as a function of the elution volume. The gel filtration chromatography fractions corresponding to the

second half of the highest peak of the chromatogram were collected. The size exclusion chromatography yielded a highly pure product as shown in Figure A.1. Figure A.3 reports the UV-visible spectrum of the product. It is worthwhile observing that the spectrum shows absorbance near 640 nm: this is attributed to charge transfer transitions from the porphyrin to the metal cation in high spin ferric systems [5].

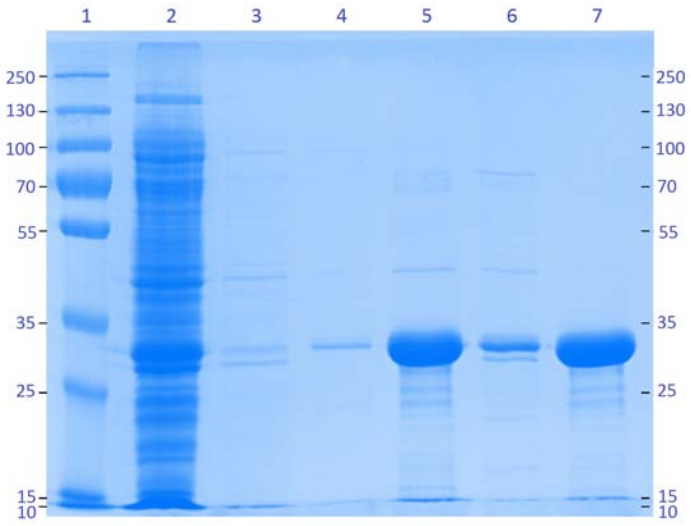


Figure A.1: C*c*P purity assessment by SDS-polyacrylamide gel electrophoresis. Lane 1: reference marker with molecular weights expressed in kDa; lane 2: nickel-NTA-agarose chromatography flowthrough during sample loading; lane 3 and 4: nickel-NTA-agarose chromatography flowthrough while washing; lane 5: C*c*P eluted from nickel-NTA-agarose chromatographic column; lane 6: FPLC peak at large molecular weight and 51 ml in figure A.2; lane 7: C*c*P eluted from gel filtration chromatographic column.

## A.4 Cytochrome *c* peroxidase crystallization

Cytochrome *c* peroxidase was concentrated to 5 - 10 mg/ml and dialysed against double-distilled water at 4 °C for 4 - 6 days in a pre-hydrated dialysis tube of suitable size provided with a 6 - 8 kDa cut-off cellulose membrane (Novagen, D-Tube$^{TM}$ Dialyzers).

After 4 - 6 days of dialysis the protein precipitates forming a dark, amorphous aggregate. Precipitate and supernatant were collected in 1.5 ml tubes and spun in a table-top

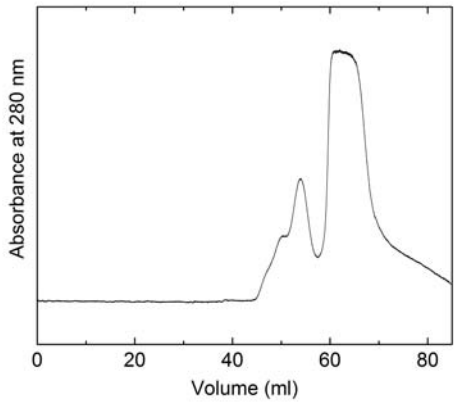


Figure A.2: Gel filtration chromatogram of MKT C*c*P. The fractions corresponding to the second half of the highest peak were collected.

centrifuge at 13,000 rounds per minute (rpm) and 4 °C for 5 minutes. The supernatant was removed and the protein was dissolved in buffer C whose composition is described in section C. The dialysis tube was washed with buffer C. The C*c*P solutions were merged in 1.5 ml tubes and spun in a table-top centrifuge at 13,000 rpm and 4 °C for 5 minutes. A white precipitate was observed. The supernatant was collected and transferred in one or more pre-hydrated dialysis tubes for the next dialysis cycle against double-distilled water at 4 °C.

The dialysis cycle was repeated 5 - 7 times. The first dialysis cycles yielded an amorphous precipitate while the last few cycles yielded needle-like crystals.

Finally the C*c*P solution was transferred to a 100 $\mu$l microdialysis button that was closed by using a pre-hydrated cellulose 12-14 kDa cut-off membrane. The protein was dialysed against C*c*P crystallization buffer whose composition is described in Section C. Crystals of different size were observed after few days. Deuterium exchanged crystals were produced by using the same crystallization buffer in which $D_2O$ was used instead of $H_2O$. A photograph of deuterium exchanged crystals is shown in Figure A.4. It is worthwhile observing that the C*c*P crystallization buffer contains 30% MPD: this diol enhances protein crystallization and acts as cryoprotectant in low temperature experiments preventing the formation of crystalline ice.

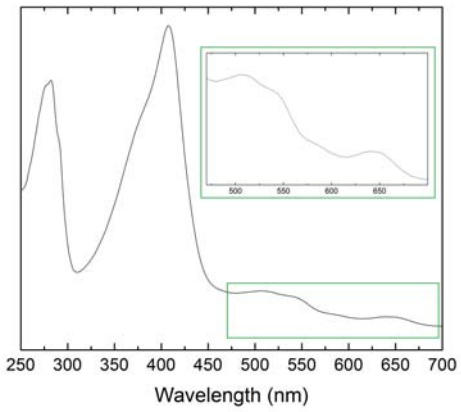


Figure A.3: UV-VIS spectrum of C*c*P collected from the FPLC fractions corresponding to the second half of the highest peak of the chromatogram in Figure A.2. The insert shows the spectral region between 470 and 700 nm.

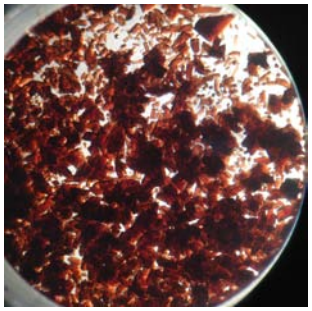

Figure A.4: Deuterium exchanged C*c*P crystals.

# Appendix B

# The production of fully deuterated cytochrome *c* peroxidase

In this appendix the protocol for the expression of the perdeuterated version of cytochrome *c* peroxidase is described, together with the preliminary study of protein expression levels in different media and solubility. The purification of perdeuterated C*c*P was carried out in a fully hydrogenated environment as described in section A.3. Perdeuterated crystals were obtained as reported in section A.4 by carrying out the last crystallization stage in $D_2O$ containing crystallization buffer. The production of perdeuterated cytochrome *c* peroxidase was carried out with Dr Juliette Devos at the D-LAB facility of the Institut Laue-Langevin (Grenoble, France).

The expression of perdeuterated protein is achieved by growing *E. coli* in a fully deuterated medium. The growth of bacteria for the expression of hydrogenated protein is carried out in the nutritionally rich lysogeny broth (LB) as described in Appendix A.1. However a deuterated version of this medium composed of tryptone and yeast extract is not available. A minimal medium containing the minimum amount of nutrients necessary for the growth of bacteria is therefore employed for the growth of *E. coli* in a fully deuterated environment. The minimal medium is composed of $D_2O$, metal salts and a deuterated carbon source, in this case D-glycerol. The growth of bacteria in such medium is slow

compared to the fully hydrogenated and nutritionally rich LB medium. For this reason the cells need to go through a process of adaptation in hydrogenated minimal medium before the growth in fully deuterated minimal medium.

## B.1 Preliminary work

*E. coli* BL21 DE3 Gold competent cells were transformed with the expression plasmid pLEICS-03, as reported in Appendix A.1. The plasmid carries kanamycin resistance allowing the selection of vector-carrying cells throughout the procedure.

This section reports the preliminary study of the level of expression of C*c*P in different media and its solubility. The first part refers to lysogeny broth (LB) medium. The second part to hydrogenated minimal medium and the third part to deuterated minimal medium.

### B.1.1 Expression in LB medium and solubility study

An agar-LB plate made up with 15 ml of agar-LB medium (see section C) and containing a final kanamycin concentration of 30 $\mu$g/ml was streaked with the transformed BL21 DE3 Gold *E. coli* cells. The plate was incubated overnight at 37 °C.

#### Expression study in LB medium

A single colony was used to inoculate 10 ml of sterilized LB medium (see section C) containing a final kanamycin concentration of 30 $\mu$g/ml. The culture was incubated at 37 °C and shaken at 190 rpm. After 24 hours the optical density (O.D.) reached the value 8.5 A.U.

Three aliquots of 250 $\mu$l each were extracted from the culture. Each aliquot was diluted in kanamycin containing LB medium to obtain an O.D. of 0.2 A.U. The three cultures were incubated at 37 °C for several hours. Expression was induced by adding IPTG to a final concentration of 1 mM when the O.D. reached a value between 0.6 and 0.9 A.U. 1 ml of uninduced sample was saved for comparison for each of the three cultures. The first induced culture was incubated for 3 hours at 37 °C, the second was incubated overnight

at 30 °C and the third was incubated overnight at 20 °C. The O.D. of the induced cultures were measured and samples of each culture were extracted.

For each culture uninduced and induced samples were run on a SDS-PAGE to assess the expression level at different temperatures. The sample volumes used in SDS-PAGE were assessed in order to have the same quantity of cells when comparing uninduced and induced samples. SDS-PAGE samples were prepared by centrifugation of the appropriate volume of cell culture in a table top centrifuge at maximum speed for 2 minutes. The supernatant was removed and the precipitate was suspended in 25 $\mu$l of SDS-PAGE sample buffer (see Section C.3) and 25 $\mu$l of urea 8 M. The SDS-PAGE samples were heated at 100 °C for 10 minutes and spun in a table top centrifuge at 8000 rpm for 1 minute.

The SDS-PAGE in Figure B.1 shows uninduced and induced samples at 37 °C in lanes 2 and 3, uninduced and induced samples at 30 °C in lanes 6 and 7, uninduced and induced samples at 20 °C in lanes 10 and 11.

**Solubility study**

Each induced culture was centrifuged at 5,000 rpm for 12 minutes. The supernatant was removed and the precipitate was resuspended in 1 ml of buffer A (see section C). Cell lysis was achieved by bead milling. A volume of 500 $\mu$l of glass beads with a diameter of 0.1 mm (Scientific Industries Inc.) was added to each of the three suspensions. The suspensions were shaken in a cell disruptor (Disruptor Genie, Scientific Industries Inc.) for 6 minutes. Aliquots with a volume of 50 $\mu$l were extracted from each sample and spun in a table top centrifuge at maximum speed fo 2 minutes.

Both precipitate and supernatant were collected for each sample and run on a SDS-PAGE to assess the solubility of the perdeuterated protein. The supernatant was mixed with an equal volume of SDS-PAGE sample buffer (see section C). The precipitate was suspended in buffer A and SDS-PAGE sample buffer 1:1, to a final volume equal to the final volume of the supernatant sample. The SDS-PAGE samples were heated at 100 °C for 10 minutes and spun in a table top centrifuge at 8,000 rpm for 1 minute. The SDS-PAGE in Figure B.1 shows the samples containing centrifugation precipitate and supernatant in

lanes 4 and 5 for the culture induced at 37 °C, in lanes 8 and 9 for the culture induced at 30 °C, and in lanes 12 and 13 for the culture induced at 20 °C.

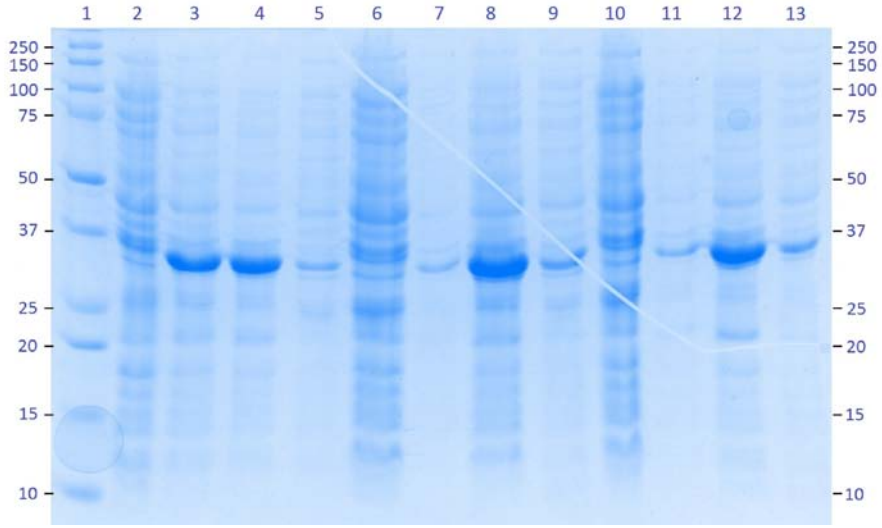


Figure B.1: LB medium expression and solubility study by SDS-PAGE. Lane 1 shows the molecular weight reference marker. Lanes 2 to 5 refer to the case of induction at 37 °C. Lane 2: uninduced sample; lane 3: induced sample; lane 4: precipitate obtained by centrifugation of the cell lysate; lane 5: supernatant. Lanes 6 to 9 refer to the case of induction at 30 °C. Lane 6: uninduced sample; lane 7: induced sample; lane 8: precipitate obtained by centrifugation of the cell lysate; lane 9: supernatant. Lanes 10 to 13 refer to the case of induction at 20 °C. Lane 10: uninduced sample; lane 11: induced sample; lane 12: precipitate obtained by centrifugation of the cell lysate; lane 13: supernatant.

### B.1.2 Expression in hydrogenated minimal medium and solubility study

Expression and solubility studies were carried out in hydrogenated minimal medium (see section C) similar to the studies in LB medium described in section B.1.1.

A single colony was used to inoculate 10 ml of hydrogenated minimal medium (see section C) containing a final kanamycin concentration of 30 $\mu$g/ml. The medium was sterilized by filtration. The culture was incubated at 37 °C and shaken at 190 rpm. After 24 hours the optical density (O.D.) reached the value 1 - 1.4 A.U. showing that the growth

rate in minimal medium was substantially slower than in LB medium. After 34 hours from the start of the incubation at 37 °C 1 ml of culture was extracted and a 10 fold dilution in kanamycin containing hydrogenated minimal medium was done. The diluted culture was incubated at 37 °C for 16 hours and reached an O.D. of around 3.6 A.U. Three aliquots were extracted from the culture. Each aliquot was diluted in kanamycin containing minimal medium to obtain an O.D. of 0.2 A.U. The three cultures were incubated at 37 °C for several hours. Expression was induced by adding IPTG to a final concentration of 1 mM when the O.D. reached a value between 0.6 and 0.9 A.U.

Expression and solubility studies were carried out as described in section B.1.1. The results are shown in figure B.2. Induction at 20 °C yielded the largest amount of protein in the soluble fraction as indicated in figure B.2.

### B.1.3 Expression in deuterated minimal medium and solubility study

Deuterated minimal medium was prepared analogous to hydrogenated minimal medium (see section C) but using $D_2O$ instead of $H_2O$. Perdeuterated glycerol was used as carbon source. In addition, $(NH_4)_2SO_4$, $KH_2PO_4$, $Na_2HPO_4{\cdot}2H_2O$, $(NH_4)_2$-H-citrate were deuterium exchanged. These components were dissolved in $D_2O$ to allow H/D exchange. The solvent was then evaporated. The procedure was repeated multiple times. The antibiotic used throughout the perdeuteration procedure was D-kanamycin namely a $D_2O$ solution of hydrogenated kanamycin. The antibiotic concentration was 30 $\mu$g/ml.

**Step 1** A volume of 1 ml of hydrogenated minimal medium culture was diluted to a final volume of 10 ml in kanamycin containing deuterated minimal medium and incubated at 37 °C overnight to an O.D. of about 4.2 A.U.

**Step 2** A volume of 1 ml of such culture was diluted to a final volume of 12 ml in kanamycin containing deuterated minimal medium and incubated at 37 °C for 24 hours. The O.D. reached a value of about 3.3 A.U.

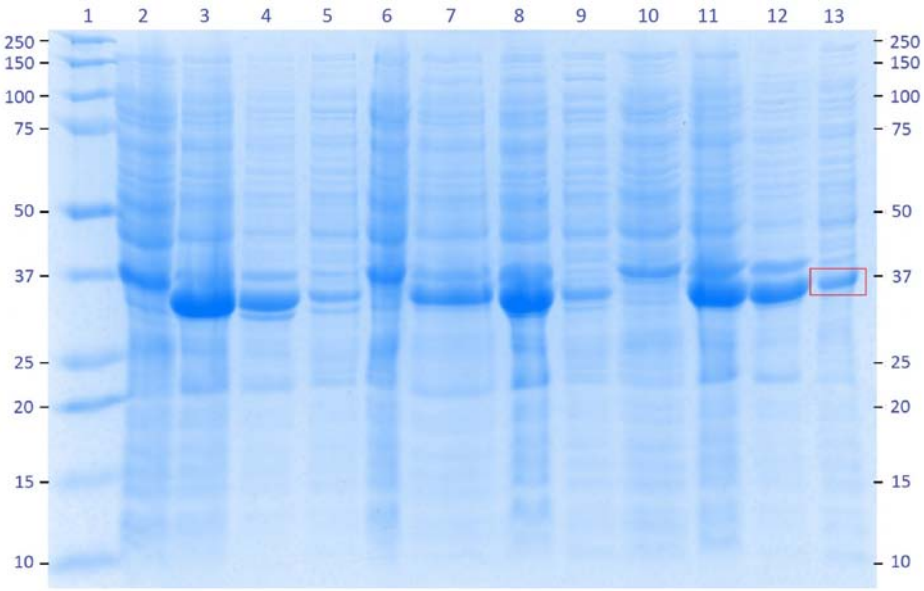


Figure B.2: Hydrogenated minimal medium expression and solubility study by SDS-PAGE. Lane 1 shows the molecular weight reference marker. Lanes 2 to 5 refer to the case of induction at 37 °C. Lane 2: uninduced sample; lane 3: induced sample; lane 4: precipitate obtained by centrifugation of the cell lysate; lane 5: supernatant. Lanes 6 to 9 refer to the case of induction at 30 °C. Lane 6: uninduced sample; lane 7: induced sample; lane 8: precipitate obtained by centrifugation of the cell lysate; lane 9: supernatant. Lanes 10 to 13 refer to the case of induction at 20 °C. Lane 10: uninduced sample; lane 11: induced sample; lane 12: precipitate obtained by centrifugation of the cell lysate; lane 13: supernatant.

Four aliquots of 800 $\mu$l each were diluted in kanamycin containing deuterated minimal medium to final O.D. of about 0.2 A.U. The four samples were incubated at 37 °C until the O.D. reached a value between 0.6 and 0.8 A.U.

Expression was induced by adding IPTG to a final concentration of 1 mM and hemin ($D_2O$ solution with NaOD 0.1 M) to a final concentration of 5 $\mu$g/ml. Expression was carried out at 16 °C and 20 °C. For each expression temperature the effect of adding *E. coli* OD 2 D medium (Silantes, complete medium for *E. coli* with minerals and buffer, $^2$H-labeled) during induction was studied. The effectiveness of different methods of protein extraction from the cells was also studied. In particular sonication and bead milling were

tested.

The SDS-PAGE samples were prepared as described in section B.1.1 and the results are shown in figure B.3 and B.4. Induction at 20 °C yielded the largest amount of protein in the soluble fraction as indicated in figure B.4.

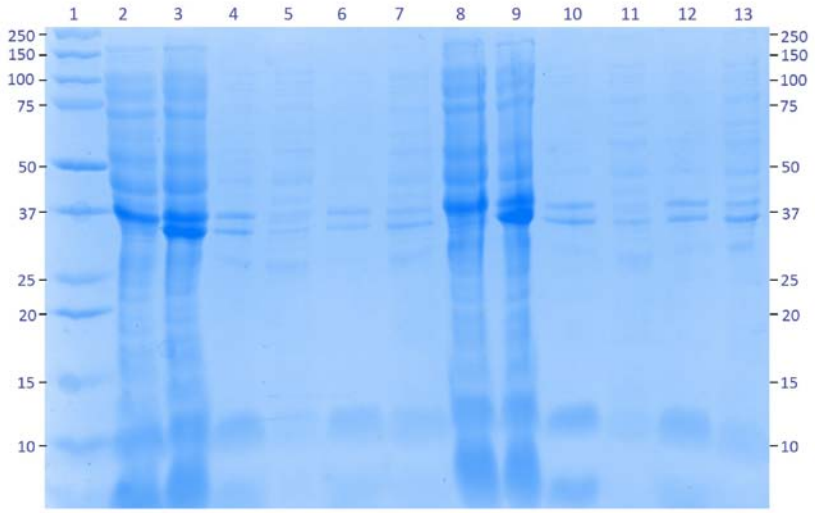


Figure B.3: Deuterated minimal medium expression at 16 °C and solubility study by SDS-PAGE. Lane 1 shows the molecular weight reference marker. Lane 2: uninduced sample; lane 3: induced sample; lane 4: centrifugation precipitate in the case of cell lysis by bead milling; lane 5: centrifugation supernatant in the case of cell lysis by bead milling; lane 6: centrifugation precipitate in the case of cell lysis by sonication; lane 7: centrifugation supernatant in the case of cell lysis by sonication. Lanes 8 to 13 report analogous samples as those in lanes 2 to 7 but with addition of *E. coli* OD 2 D medium (Silantes) during protein expression. Lane 8: uninduced sample; lane 9: induced sample; lane 10: centrifugation precipitate in the case of cell lysis by bead milling; lane 11: centrifugation supernatant in the case of cell lysis by bead milling; lane 12: centrifugation precipitate in the case of cell lysis by sonication; lane 13: centrifugation supernatant in the case of cell lysis by sonication.

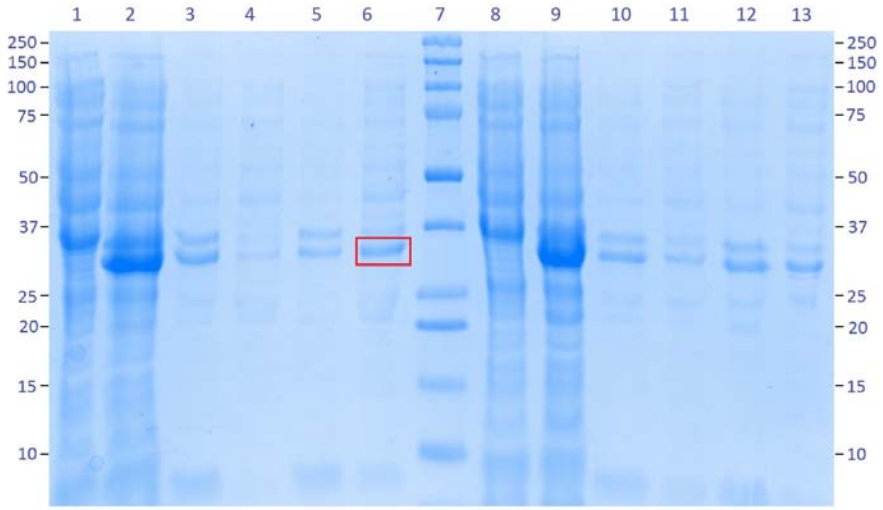


Figure B.4: Deuterated minimal medium expression at 20 °C and solubility study by SDS-PAGE. Lane 1: uninduced sample; lane 2: induced sample; lane 3: centrifugation precipitate in the case of cell lysis by bead milling; lane 4: centrifugation supernatant in the case of cell lysis by bead milling; lane 5: centrifugation precipitate in the case of cell lysis by sonication; lane 6: centrifugation supernatant in the case of cell lysis by sonication. Lane 7 shows the molecular weight reference marker. Lanes 8 to 13 report analogous samples as those in lanes 1 to 6 but with addition of *E. coli* OD 2 D medium (Silantes) during protein expression. Lane 8: uninduced sample; lane 9: induced sample; lane 10: centrifugation precipitate in the case of cell lysis by bead milling; lane 11: centrifugation supernatant in the case of cell lysis by bead milling; lane 12: centrifugation precipitate in the case of cell lysis by sonication; lane 13: centrifugation supernatant in the case of cell lysis by sonication.

## B.2 Protocol for perdeuterated C*c*P expression

The process of adaptation of cell growth in deuterated minimal medium required 5 dilution and growth steps. The culture produced in Step 2 of section B.1.3 was the starting point.

**Step 3** An 800 $\mu$l aliquot from Step 2 was diluted to a final volume of 12 ml in kanamycin containing deuterated minimal medium and incubated at 37 °C for 24 hours.

**Step 4** After 24 hours the O.D. reached a value close to 3.5 A.U. A culture volume of 1.3 ml was extracted and diluted to a final volume of 15 ml in kanamycin containing deuterated minimal medium. The culture was incubated at 37 °C for 24 hours.

**Step 5** After 24 hours the O.D. reached a value close to 3 A.U. An aliquot of 1.3 ml was diluted to a final volume of 15 ml in kanamycin containing deuterated minimal medium and incubated at 37 °C for 24 hours. The O.D. was then 3.1 A.U.

**Step 6** One milliliter of culture was diluted to a final volume of 10 ml in kanamycin containing deuterated minimal medium and incubated at 37 °C to an O.D. of 6 A.U. The cells were stored in a -80 °C freezer as a stock of deuterated minimal medium adapted cells.

An aliquot of 10 ml was extracted from Step 4 culture and diluted to a final volume of 150 ml in kanamycin containing deuterated minimal medium and incubated at 37 °C until the O.D. reached a value of about 1 A.U. The culture was transferred into a Labfors fermenter (Infors) together with 1.2 l of deuterated minimal medium and 1.3 ml of D-kanamycin 30 mg/ml.

The use of a fermenter was motivated by the requirement for high cell density in a fully deuterated environment in absence of contamination from hydrogen. Oxygen was supplied to the culture through a dry air supply line. The $O_2$ pressure in the medium was measured throughout the procedure[1] as reported in figure B.6, blue curve. Temperature and pH were monitored during the process, as shown in figure B.6 in yellow and green respectively. The O.D. was measured manually throughout the procedure, by extracting small samples of culture, and is shown in black in figure B.6.

Fermentation was carried out at 30 °C. The pH was kept constant at the value 6.9. For this purpose a NaOD pump was used supplying 4% NaOD to the culture when the pH value was reported below a threshold value. During the first 14 hours from the start of fermentation the oxygen content decreased due to bacteria growth. After 28 hours the

[1]The $O_2$ sensor was calibrated as follows: $p_{O_2} = 100\,\%$ corresponds to dry air bubbling in distilled water while $p_{O_2} = 0\,\%$ corresponds to nitrogen bubbling in distilled water.

oxygen level increased significantly due to decreasing bacteria growth rate. The culture was then fed through a feeding pump as shown by the red curve in figure B.6 supplying perdeuterated feeding medium whose composition is described in section C.

Expression was induced when the optical density reached the value of 15.4 A.U. The temperature was gradually decreased to 20 °C, hemin was added (see section C) to a final concentration of 5 mg/l and IPTG (see section C) to final concentration 1 mM.

Nineteen hours after induction the cells were harvested by centrifugation at 4500 rpm and 4 °C for 15 minutes. A total of 66.2 g of cell paste were obtained, these were frozen and stored at −80 °C. The results are shown in figure B.5.

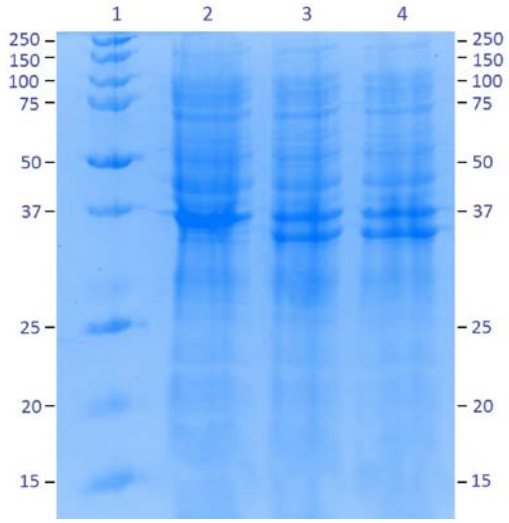


Figure B.5: SDS-PAGE of fermentation samples. Lane 1 shows the molecular weight reference marker. Lanes 2 shows the uninduced sample. Lanes 3 and 4 show the induced sample after 14 and 19 hours respectively from induction.

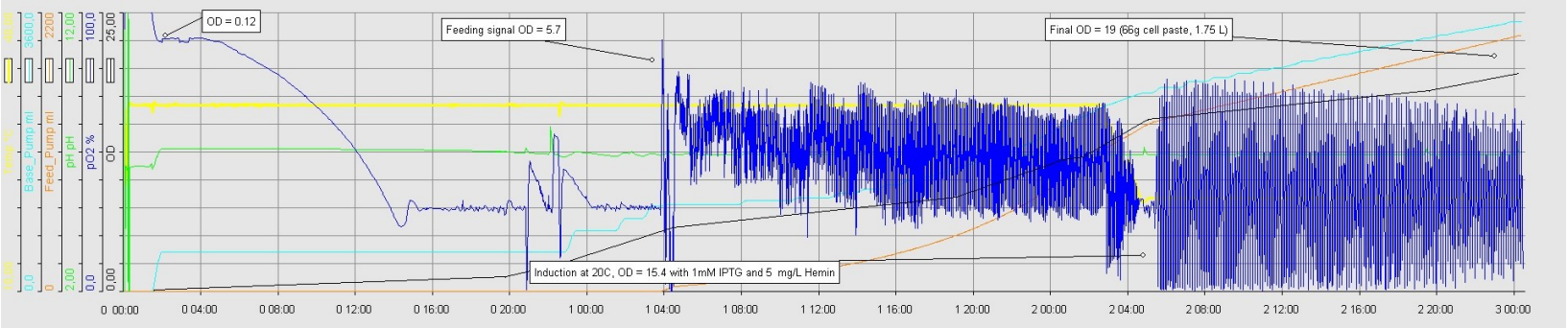


Figure B.6: Behaviour of temperature (yellow), NaOD flow (light blue), feeding medium flow (orange), pH (green), oxygen pressure (blue) and optical density (black) during fermentation.

# Appendix C

# Materials

This appendix reports the composition of media and buffer solutions employed in sections A and B. The materials used for SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) are also described.

## C.1 Materials employed in the expression, purification and crystallization of C*c*P

**LB-Agar medium** 20 g of LB low salt powder (Melford), 15 g of agar and 2 g of glucose made up to 1 l with distilled water. The solution was sterilized.

**Lysogeny broth (LB) medium** 20 g of LB low salt powder (Melford) and 2 g of glucose made up to 1 l with deionized water. The solution was sterilized.

**Phenylmethane-sulfonyl fluoride (PMSF) solution** 34 mg of phenylmethane-sulfonyl fluoride dissolved in 1 ml of propanol.

**Hemin solution** 50 mg of hemin dissolved in 10 ml of NaOH 0.1 M.

**Buffer A** $KH_2PO_4$ 150 mM, pH 6.5.

**Buffer B** $KH_2PO_4$ 100 mM, KCl 300 mM, glycerol 10% w/v, pH 8.

**Wash buffer** $KH_2PO_4$ 100 mM, KCl 300 mM, glycerol 10% w/v, pH 6.

**Elution buffer** $KH_2PO_4$ 100 mM, KCl 300 mM, glycerol 10% w/v, pH 4.2.

**FPLC buffer** $KH_2PO_4$ 10 mM, KCl 150 mM, pH 6.5.

**Buffer C** $KH_2PO_4$ 500 mM, pH 6.5.

**C*c*P crystallization buffer** $KH_2PO_4$ 50 mM, (±)-2-methyl-2,4-pentanediol 30% v/v, pH 6.

## C.2 Materials employed in the perdeuteration of C*c*P

**Minimal medium** In 1 litre: 6.86 g $(NH_4)_2SO_4$, 1.56 g $KH_2PO_4$, 6.48 g $Na_2HPO_4{\cdot}2H_2O$, 0.49 g $(NH_4)_2$-H-citrate, 5 g glycerol, 1 ml $MgSO_4$ 1M, 1 ml metal salts solution.

**Metal salts solution** 0.5 g/l $CaCl_2{\cdot}2H_2O$, 16.7 g/l $FeCl_3{\cdot}6H_2O$, 0.18 g/l $ZnSO_4{\cdot}7H_2O$, 0.16 g/l $CuSO_4{\cdot}5H_2O$, 0.15 g/l $MnSO_4{\cdot}4H_2O$, 0.18 g/l $CoCl_2{\cdot}6H_2O$, 20.1 g/l Na-EDTA.

**NaOD 4%** 90 ml of $D_2O$, 10 ml of NaOD in $D_2O$ 10 M stock, sterile filtering.

**Feeding medium** 30 g of perdeuterated glycerol, 250 ml of perdeuterated minimal medium, sterile filtering.

**Hemin solution** 5 mg/ml of hemin in $D_2O$ with NaOD 0.1 M

## C.3 SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) were run by using the electrophoresis system *Mini Protein Tetra System* (Bio-RAD) at 180 V.

**Running gel - 12% acrylamide - 2 gels** 4 ml of acrylamide : bisacrylamide 37.5 : 1, 3.3 ml of Tris-Cl [tris(hydroxymethyl)aminomethane] 3 M at pH 8.45, 100 $\mu$l of sodium dodecyl sulfate (SDS) 10 %, 1 ml of glycerol 100 %, 1.6 ml of distilled water, 100 $\mu$l of ammonium persulfate 10 %, 10 $\mu$l of tetramethylethylenediamine (TEMED)

**Stacking gel - 4% acrylamide - 2 gels** 0.65 ml of acrylamide : bisacrylamide 37.5 : 1, 1.24 ml of Tris-Cl [tris(hydroxymethyl)aminomethane] 3 M at pH 8.45, 50 $\mu$l of sodium dodecyl sulfate (SDS) 10 %, 3.1 ml of distilled water, 50 $\mu$l of ammonium persulfate 10 %, 10 $\mu$l of tetramethylethylenediamine (TEMED)

**SDS-PAGE sample buffer (5X)** 1.25 ml of Tris 1 M pH 6.8; 1 g of SDS; 3 ml of glycerol 100 %; 0.4 g of DTT, 2.5 mg of Comassie G-250 and distilled $H_2O$ to 10 ml.

**SDS-PAGE running buffer: cathode solution (10X)** 121.1 g of Tris base; 179.2 g of tricine; 10 g of SDS and distilled $H_2O$ to 1 l.

**SDS-PAGE running buffer: anode solution (10X)** 121.1 g of Tris base and distilled $H_2O$ to 1 l; pH 8.9.

# Appendix D

# Complementary techniques

Ultraviolet-visible spectrophotometry and electronic paramagnetic resonance were employed for the characterization of the intermediates in the catalytic cycle of cytochrome *c* peroxidase . The present appendix reports the technical details of the experimental set-ups.

## D.1 Single crystal microspectrophotometry

Absorption spectra of single crystals of C*c*P were collected using an Ocean Optics Maya 2000 PRO spectrometer, with an Ocean Optics DH-2000-BAL UV-VIS-NIR light source and a Hamamatsu S10420 FFT-CCD back-thinned detector with fibre optic coupled to 80 mm diameter 4X reflective lenses (Optic Peter. Lentilly, France) and mounted with a custom mount (BioMedical Mechanical Workshop, University of Leicester) on a Rigaku Raxis IV $\phi$ drive. Crystals were mounted on a nylon loop in a SPINE standard mount. The temperature was maintained at 100 K with an Oxford Cryosystems cryostream. Absorption spectra were acquired by means of the Ocean Optics SpectraSuite software.

## D.2 Electronic paramagnetic resonance

Continuous-wave EPR spectra were recorded at 9.404 GHz on a Bruker EMX spectrometer with a Super-high-Q rectangular cavity and an Oxford ESR- 900 liquid helium cryostat. Quartz sample holders were used both for single crystals and for solution measurements. The operating conditions are stated in the legend of figures.

# Bibliography

[1] D. B. Goodin, M. G. Davidson, J. A. Roe, A. G. Mauk and M. Smith, *Biochemistry* **30**, 4953 - 4962 (1991).

[2] A. Gumiero, C. L. MetCalfe, A. R. Pearson, E. L. Raven and P. C. E. Moody, *The Journal of Biological Chemistry* **286**, No. 2 (2011).

[3] C. Metcalfe, I. K. Macdonald, E. J. Murphy, K. A. Brown, E. L. Raven and P. C. Moody, *J. Biol. Chem.* **283**, 61936200 (2008).

[4] E. J. Murphy, C. L. Metcalfe, J. Basran, P. C. Moody and E. L. Raven, *Biochemistry* **47**, 1393313941 (2008).

[5] G. R. Moore and G. W. Pettigrew, *Cytochromes c: Evolutionary, Structural and Physicochemical aspects.* Springer-Verlag (1990).

# Part VI

# Publications and list of activities

# Neutron cryo-crystallography captures the protonation state of ferryl heme in a peroxidase

Cecilia M. Casadei *et al.*

*Science* **345**, 193 (2014);
DOI: 10.1126/science.1254398



*Science* (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 

be used to track the temperature dependence of the gap function in TRSB superconductors.

## REFERENCES AND NOTES


1. R. Joynt, L. Taillefer, *Rev. Mod. Phys.* **74**, 235–294 (2002).
2. G. R. Stewart, Z. Fisk, J. O. Wills, J. L. Smith, *Phys. Rev. Lett.* **52**, 679–682 (1984).
3. M. R. Norman, *Science* **332**, 196–200 (2011).
4. R. A. Fisher *et al.*, *Phys. Rev. Lett.* **62**, 1411–1414 (1989).
5. K. Hasselbach, L. Taillefer, J. Flouquet, *Phys. Rev. Lett.* **63**, 93–96 (1989).
6. S. Adenwalla *et al.*, *Phys. Rev. Lett.* **65**, 2298–2301 (1990).
7. G. Aeppli, A. Goldman, G. Shirane, E. Bucher, M. Lux-Steiner, *Phys. Rev. Lett.* **58**, 808–811 (1987).
8. S. M. Hayden, L. Taillefer, C. Vettier, J. Flouquet, *Phys. Rev. B* **46**, 8675 (1992).
9. R. N. Kleiman *et al.*, *Phys. Rev. Lett.* **69**, 3120–3123 (1992).
10. U. Yaron *et al.*, *Phys. Rev. Lett.* **78**, 3185–3188 (1997).
11. A. Huxley *et al.*, *Nature* **406**, 160–164 (2000).
12. J. A. Sauls, *Adv. Phys.* **43**, 113–141 (1994).
13. J. D. Strand *et al.*, *Science* **328**, 1368–1369 (2010).
14. J. D. Strand, D. J. Van Harlingen, J. B. Kycia, W. P. Halperin, *Phys. Rev. Lett.* **103**, 197002 (2009).
15. G. M. Luke *et al.*, *Phys. Rev. Lett.* **71**, 1466–1469 (1993).
16. P. D. de Réotier *et al.*, *Phys. Lett. A* **205**, 239–243 (1995).
17. Y. Machida *et al.*, *Phys. Rev. Lett.* **108**, 157002 (2012).
18. P. N. Argyres, *Phys. Rev.* **97**, 334–345 (1955).
19. Other mechanisms, such as chiral crystal structure or gyrotropic order, may also give rise to a finite Kerr effect; however, such sources of Kerr rotation do not exhibit the same magnetic field training behavior observed in TRS-breaking materials (*20*).
20. P. Hosur, A. Kapitulnik, S. A. Kivelson, J. Orenstein, S. Raghu, *Phys. Rev. B* **87**, 115116 (2013).
21. L. D. Landau, L. P. Pitaevskii, E. M. Lifshitz, *Electrodynamics of Continuous Media* (Elsevier, Burlington, MA, ed. 2, 1984).
22. S. Spielman *et al.*, *Phys. Rev. Lett.* **68**, 3472–3475 (1992).
23. J. Xia, Y. Maeno, P. T. Beyersdorf, M. M. Fejer, A. Kapitulnik, *Phys. Rev. Lett.* **97**, 167002 (2006).
24. J. Xia, P. T. Beyersdorf, M. M. Fejer, A. Kapitulnik, *Appl. Phys. Lett.* **89**, 062508 (2006).
25. A. Kapitulnik, J. Xia, E. Schemm, A. Palevski, *New J. Phys.* **11**, 055060 (2009).
26. P. E. Sulewski *et al.*, *Phys. Rev. B Condens. Matter* **38**, 5338–5352 (1988).
27. Materials and methods are available as supplementary materials on *Science* Online.
28. B. Jeanneret, G. L. Gavilano, G. A. Racine, C. Leemann, P. Martinoli, *Appl. Phys. Lett.* **55**, 2336 (1989).
29. J. P. Brison *et al.*, *J. Low Temp. Phys.* **95**, 145 (1994).
30. B. Lussier, B. Ellman, L. Taillefer, *Phys. Rev. B Condens. Matter* **53**, 5145–5148 (1996).
31. V. M. Yakovenko, *Phys. Rev. Lett.* **98**, 087003 (2007).
32. J. Goryo, *Mod. Phys. Lett. B* **24**, 2831–2836 (2010).
33. R. M. Lutchyn, P. Nagornykh, V. M. Yakovenko, *Phys. Rev. B* **80**, 104508 (2009).
34. E. Taylor, C. Kallin, *Phys. Rev. Lett.* **108**, 157001 (2012).
35. P. J. C. Signore *et al.*, *Phys. Rev. B Condens. Matter* **52**, 4446–4461 (1995).
36. S. Fujimoto, *Phys. Rev. B* **72**, 024515 (2005).
37. S. Yip, A. Garg, *Phys. Rev. B Condens. Matter* **48**, 3304–3308 (1993).


## ACKNOWLEDGMENTS

Stimulating discussions with S. Kivelson, A. Huxley, and D. Agterberg; sample characterization by K. Avers (Northwestern); and instrument design assistance from G. Burkhard (Stanford) are greatly appreciated. This work was supported by the U.S. Department of Energy Office of Basic Energy Science, Division of Materials Science and Engineering, at Stanford under contract DE-AC02-76SF00515 and at Northwestern under contract DE- FG02-05ER46248. Construction of the Sagnac apparatus was partially funded by the Stanford Center for Probing the Nanoscale (NSF NSEC 0425897). E.R.S. received additional support from a Gabilan Stanford Graduate Fellowship and the DARE Doctoral Fellowship Program.

## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/345/6193/190/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S4

15 November 2013; accepted 21 May 2014
10.1126/science.1248552

HEME ENZYMES

# Neutron cryo-crystallography captures the protonation state of ferryl heme in a peroxidase

**Cecilia M. Casadei,[1,2] Andrea Gumiero,[3] Clive L. Metcalfe,[3] Emma J. Murphy,[3] Jaswir Basran,[1] Maria Grazia Concilio,[4] Susana C. M. Teixeira,[2,5] Tobias E. Schrader,[6] Alistair J. Fielding,[4] Andreas Ostermann,[7] Matthew P. Blakeley,[2] Emma L. Raven,[3]* Peter C. E. Moody[1]***

**Heme enzymes activate oxygen through formation of transient iron-oxo (ferryl) intermediates of the heme iron. A long-standing question has been the nature of the iron-oxygen bond and, in particular, the protonation state. We present neutron structures of the ferric derivative of cytochrome c peroxidase and its ferryl intermediate; these allow direct visualization of protonation states. We demonstrate that the ferryl heme is an Fe(IV)=O species and is not protonated. Comparison of the structures shows that the distal histidine becomes protonated on formation of the ferryl intermediate, which has implications for the understanding of O–O bond cleavage in heme enzymes. The structures highlight the advantages of neutron cryo-crystallography in probing reaction mechanisms and visualizing protonation states in enzyme intermediates.**

Aerobic organisms have evolved to use the intrinsic oxidizing power of oxygen from the atmosphere. This activation of oxygen is achieved by a catalytic metal center (usually iron or copper) buried within a protein. In the case of iron, high-valent iron-oxo (also known as ferryl) intermediates play a role in a large number of different, and sometimes difficult, biological oxidations catalyzed by various heme and non-heme iron-containing enzymes.

For the heme iron enzymes, the mechanism of oxidation involves initial formation of the iconic compound I intermediate (*1*, *2*). Compound I contains an oxidized ferryl heme, plus either a porphyrin π-cation radical or a protein radical; reduction of compound I by one electron yields the closely related compound II intermediate, which contains only the ferryl heme. These compound I and II intermediates are a defining feature across the heme enzyme family and appear in a diverse group of catalytic heme enzymes that include all the cytochrome P450s, the nitric oxide synthases, and the terminal oxidases, plus the heme dioxygenases and heme peroxidases. Indeed, they are such crucial intermediates in so many processes—including many involved in drug metabolism and other important oxidations—that their structure and reactivity has become a key question for both heme (*3*–*7*) and non heme iron enzymes where similarly transient iron-oxo species are also employed (*8*, *9*).

A long-standing and highly contentious question has been to clarify the bond order and protonation state of the ferryl heme. The question has focused on whether the ferryl is formulated as an Fe(IV)=O [iron(IV)-oxo] or Fe(IV)-OH [iron(IV)-hydroxide] species, but there are a number of reasons why previous methodologies—none of which can measure the protonation state directly—have failed to fully resolve the issue. To begin with, even capturing these transient intermediates in some enzymes [especially in the P450s (*10*, *11*)] has proved very difficult. Early approaches to the problem used both extended x-ray absorption fine structure and resonance Raman methods [reviewed in (*12*)] to examine the iron-oxygen bond as an indirect reporter on the protonation state. These studies indicate a short Fe-O bond length, but the data were not totally consistent, and the photolability of compound I during laser excitation is well documented (*13*) so that interpretation of stretching frequencies from Raman work has not been unambiguous either. More recently, x-ray crystallographic methods have been employed. These methods showed longer iron-oxygen bond lengths, but the x-ray structures of compounds I and II were subsequently shown to have been affected by photoreduction of the iron and are now considered to be unreliable. The more recent use of multiple crystals in x-ray analyses minimizes photoreduction but cannot entirely eliminate the problem. Moreover, hydrogen atoms are difficult to locate in electron density maps due to their weak scattering. Even if very high-resolution data (i.e., better than 1.2 Å) are available, hydrogen atoms can remain obscured due to their mobility. There is also the substantive question of whether such small changes in Fe-O

[1]Department of Biochemistry and Henry Wellcome Laboratories for Structural Biology, University of Leicester, Lancaster Road, Leicester LE1 9HN, UK. [2]Institut Laue-Langevin, 71 Avenue des Martyrs, 38000, Grenoble, France. [3]Department of Chemistry, University of Leicester, University Road, Leicester LE1 7RH, UK. [4]The Photon Science Institute, The University of Manchester, Manchester M13 9PL, UK. [5]EPSAM, Keele University, Keele, Staffordshire ST5 5BG, UK. [6]Jülich Centre for Neutron Science (JCNS), Forschungszentrum Jülich GmbH, Outstation at MLZ, Lichtenbergstraße 1, 85747 Garching, Germany. [7]Heinz Maier-Leibnitz Zentrum (MLZ), Technische Universität München, Lichtenbergstraße 1, D-85748 Garching, Germany.
***Corresponding author. E-mail: emma.raven@le.ac.uk (E.L.R.); peter.moody@le.ac.uk (P.C.E.M.)**

bond length (<0.2 Å) between Fe(IV)=O and Fe (IV)-OH species, even if measured at high resolution and without photoreduction, are determined accurately enough in x-ray experiments to report reliably on the protonation state.

For these reasons, we have adopted a different approach. Neutron diffraction offers a more reliable solution to the problem, because hydrogen and, particularly, its isotope deuterium (D) are visible at much lower resolutions than required with x-rays (*14*, *15*). The approach does present a number of challenges, particularly the need to cool the large crystals required for neutron crystallography down to the cryogenic temperatures required for the study of enzyme intermediates (*16*). But the fact that neutrons are non-ionizing and are scattered by the atomic nuclei (rather than by the electrons, as in x-ray crystallography) means that photoreduction, which has been the source of most of the confusion in previous x-ray work, does not occur at all. In the case of redox enzymes, the elimination of photoreduction is a considerable advantage if the other experimental challenges can be overcome.

In this work, we have used neutron crystallography to determine the structure of ferric cytochrome c peroxidase (CcP) and its transient ferryl intermediate. The structures allow us to visualize the positions of hydrogen and deuterium atoms in the enzyme, as well as to unambiguously identify the protonation state of the ferryl heme.

Cytochrome c peroxidase catalyzes the $H_2O_2$-dependent oxidation of cytochrome c. Compound I of CcP, formed rapidly from the reaction of ferric enzyme with peroxide, contains a ferryl heme plus a stable protein radical on $Trp^{191}$ (*17*). It was the first peroxidase to be crystallized (*18*) and over many years has served as a model enzyme for the understanding of oxygen activation across the family of catalytic heme enzymes. We first solved the neutron structure of ferric CcP to 2.4 Å at room temperature. The structure was determined using x-ray and neutron diffraction data obtained from a $D_2O$-exchanged crystal (see supplementary materials and methods; data and refinement statistics are shown in table S1).

Figure 1 shows nuclear scattering and electron density of ferric CcP in the region of the heme and summarizes the positions of all the hydrogen and deuterium atoms. On the distal side of the heme (Fig. 1, A and C), $His^{52}$ is neutral as $N_\delta$ is deuterated but $N_\varepsilon$ is not. The $N_\varepsilon$ of $Trp^{51}$ is deuterated; $Arg^{48}$ is fully deuterated and positively charged. The neutron structure unambiguously identifies W1 as a heavy water molecule ($D_2O$), rather than an $OD^-$ ion, at a distance of 2.7 Å from the iron. We assign the heme as five-coordinate. On the proximal side (Fig. 1, B and D), $N_\varepsilon$ of $His^{175}$ is not deuterated and is 2.0 Å from the iron. The $N_\delta$ of $His^{175}$ is deuterated and within hydrogen bond distance of the unprotonated $O_\delta 2$ of $Asp^{235}$; likewise, $O_\delta 1$ of $Asp^{235}$ is unprotonated and within hydrogen bond distance of the deuterated $N_\varepsilon$ of $Trp^{191}$. Both heme propionates are unprotonated.

Crystals of $D_2O$-exchanged CcP (see supplementary materials and methods) were reacted

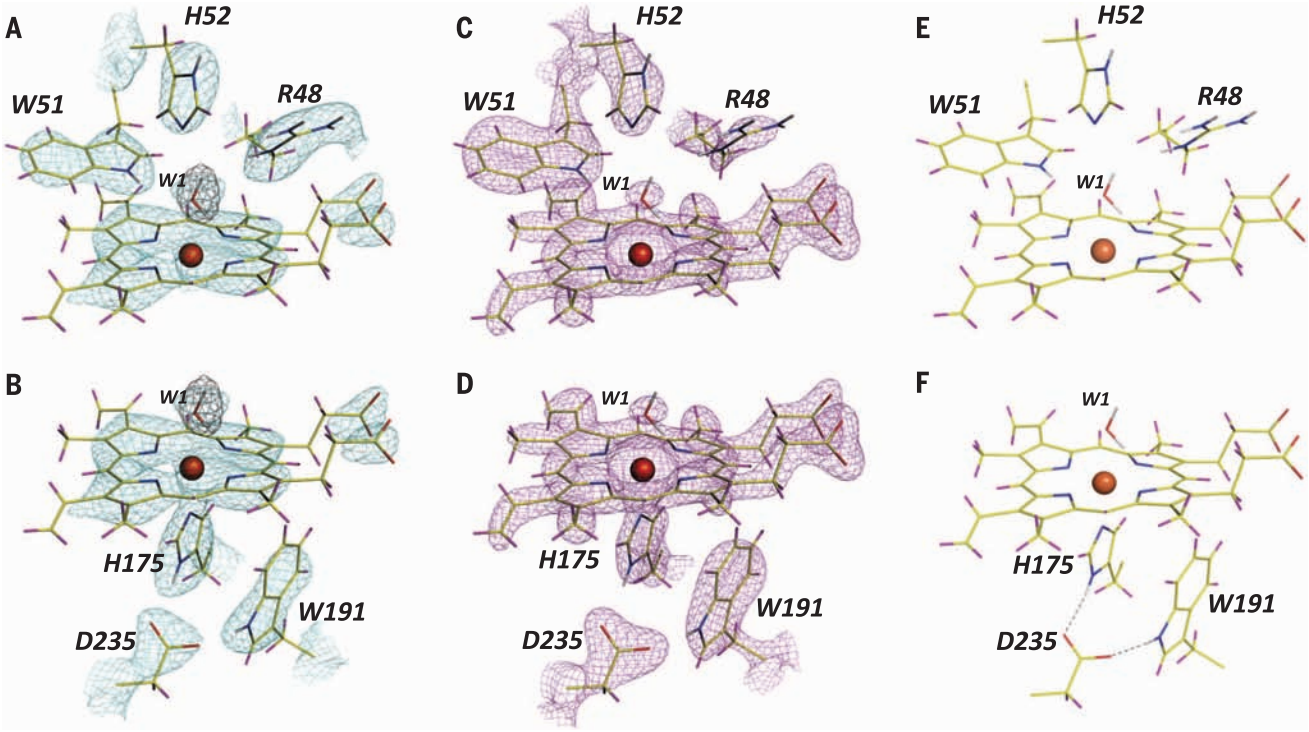


**Fig. 1. The structure of ferric CcP in the region of the heme.** In the neutron experiment, neutrons are diffracted by the atomic nuclei (in contrast to x-rays, which are scattered by electrons). Hydrogen atoms (shown in magenta) have a negative coherent neutron scattering length (−3.74 fm), whereas deuterium atoms (shown in white) have a positive coherent scattering length (6.77 fm). Exchangeable hydrogen atoms, therefore, show up with density similar to carbon (6.65 fm) and oxygen (5.85 fm), whereas nitrogen has a greater scattering length (9.37 fm). The scatter from a methylene ($CH_2$) group sums to nearly zero, so that at moderate resolution (~2 to 2.5 Å) cancellation of nuclear scattering density is observed, and the joint refinement allows use of the x-ray terms to place the nonhydrogen atoms in these regions. Nuclear scattering density [σA-weighted $2F_o$-$F_c$ contoured at 2.6 root mean square (RMS)] in the (**A**) distal and (**B**) proximal heme pocket is shown in cyan. Electron density (σA-weighted $2F_o$-$F_c$ contoured at 2.4 RMS) in the (**C**) distal and (**D**) proximal pocket is shown in magenta. The black contour (σA-weighted neutron $F_o$-$F_c$ contoured at 5.0σ) shows the difference density calculated by omitting the bound water (W1). Protonation states are depicted for (**E**) distal and (**F**) proximal active site residues. The distal pocket shows an ordered water molecule (W1) 2.7 Å away from the five-coordinate heme iron. W1 donates a hydrogen bond to the π orbitals of the porphyrin ring and may be a potential hydrogen bond acceptor to the deuterated $N_\varepsilon$ of $Trp^{51}$. W1 may also interact with a poorly ordered water (not shown) that displaces the guanidinium of $Arg^{48}$, which is seen in the "out" position (*19*). Color scheme: hydrogen atoms, magenta; deuterium, white; oxygen, red; nitrogen, blue; carbon, yellow; iron, orange sphere. Hydrogen bonds are shown as dashed lines. H, His; R, Arg; D, Asp; W (except W1), Trp.

with hydrogen peroxide as previously reported (*19*). Formation of compound I in CcP crystals is established [e.g., (*20*, *21*)] but was verified in several ways. Single-crystal microspectrophotometry at 100 K on similar (smaller) crystals showed the characteristic absorption peaks in the visible region (530, 560, and 632 nm) (fig. S1A), in agreement with the literature (*22*). Spectra of these crystals were unchanged on storage over 20 days (fig. S1B). Single-crystal electron paramagnetic resonance (EPR) on similar crystals prepared and reacted in the same way (see supplementary materials and methods) also showed the appearance of the characteristic $g = 2$ signal from the $Trp^{191}$ radical (*17*) (fig. S2); in solution, compound I of CcP is stable for hours at room temperature (*23*) and indefinitely at 77 K (figs. S3 and S4).

We then used x-ray and neutron diffraction data to determine the 2.5 Å resolution neutron structure of compound I of CcP at 100 K (see supplementary materials and methods). Figure 2 shows nuclear scattering and electron density in the region of the heme and summarizes the positions of all the hydrogen and deuterium atoms. Data and refinement statistics are shown in table S2. Nuclear scattering and electron density maps for individual active site residues are shown in fig. S5. In comparison to the ferric enzyme, most of the protonation states are retained in the compound I structure, with the exception of $His^{52}$, which is now deuterated at both $N_\delta$ and $N_\varepsilon$ (fig. S5D). We observe that the $N_\varepsilon$ of the $Trp^{191}$ radical is deuterated, which identifies the radical species as a (protonated) π-cation radical (*24*). Compared to the ferric enzyme, $Arg^{48}$ has moved into the heme pocket.

Analysis of the nuclear scattering density maps indicates that the ferryl oxygen is not deuterated in the structure. This is confirmed by examination of the hydrogen bond structure: The ferryl oxygen acts as a hydrogen bond acceptor to the $N_\varepsilon$ of $Trp^{51}$ and to the $N_\varepsilon$ of $Arg^{48}$, both of which are accordingly deuterated. We interpret this geometry as consistent with an unprotonated Fe(IV)=O species. An $F_o$–$F_c$ neutron map calculated in the absence of the ferryl oxygen shows a peak (3.8σ) in nuclear scattering density at 1.6 Å from the iron atom (Fig. 2, A and B), which is consistent with the most recent x-ray data [reported as 1.63 Å (*19*) and 1.73 Å (*25*)]. However, bond lengths cannot be determined precisely at this resolution and, as we note above, report only indirectly on the protonation state. The unambiguous observation from the neutron structure is that the oxygen is not protonated, which establishes that the ferryl intermediate is an Fe(IV)=O species at this pH. The neutron data show that there is no hydrogen bond from the distal histidine to the ferryl heme, as has been suggested previously (*26*) in horseradish peroxidase to account for pH-dependent spectroscopic behavior of the ferryl intermediate.

Mechanisms for O–O bond cleavage in peroxidases are invariably drawn (*1*, *27*) showing the distal histidine ($His^{52}$ in the case of CcP) as neutral in both ferric and compound I species, which assumes that both protons of $H_2O_2$ are used in the formation of a water molecule. Our data are not consistent with this long-standing description. $His^{52}$ is indeed neutral in the ferric state (Fig. 1 and fig. S5E) but is deuterated on $N_\varepsilon$ and $N_\delta$ in compound I (Fig. 2 and fig. S5D). We take this to mean that the widely assumed role of the distal histidine in compound I formation—acting first as a base catalyst (to deprotonate peroxide) and then as an acid catalyst (to protonate the peroxide oxygen and release water)—needs to be reassessed.

Possible alternative mechanisms showing movement of protons during O–O bond cleavage are shown in Fig. 3. Ferric CcP is unprotonated on the $N_\varepsilon$ of $His^{52}$, and the $N_\varepsilon$ lone-pair geometry does

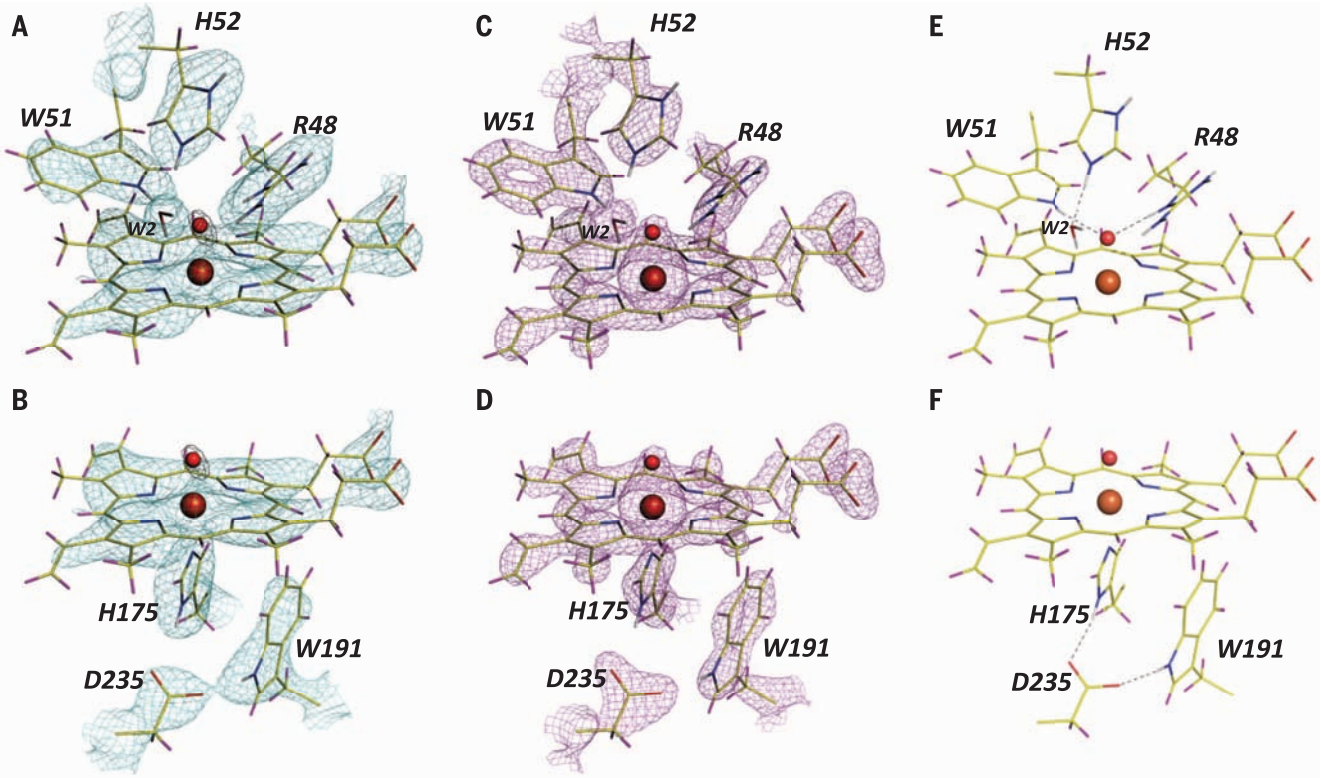


**Fig. 2. The structure of compound I of CcP in the region of the heme.** Nuclear scattering density (σA-weighted $2F_o$-$F_c$ contoured at 2.2 RMS) in the (**A**) distal and (**B**) proximal heme pocket is shown in cyan. Electron density (σA-weighted $2F_o$-$F_c$ contoured at 2.6 RMS) in the (**C**) distal and (**D**) proximal heme pocket is shown in magenta. The black contour (σA-weighted neutron Fo-Fc contoured at 3.5σ) shows the difference density calculated by omitting the ferryl oxygen. Protonation states are depicted for (**E**) distal and (**F**) proximal active site residues. In the distal pocket, the ferryl oxygen acts as an acceptor for hydrogen bonds from the deuterated $N_\varepsilon$ of $Arg^{48}$ and the deuterated $N_\varepsilon$ of $Trp^{51}$. $His^{52}$ is deuterated at both $N_\delta$ and $N_\varepsilon$, the latter acting as a hydrogen bond donor to a bound water molecule (W2). W2 donates a hydrogen bond to the π orbitals of the porphyrin ring. Color scheme is as in Fig. 1, with the iron and ferryl oxygen depicted as orange and red spheres.

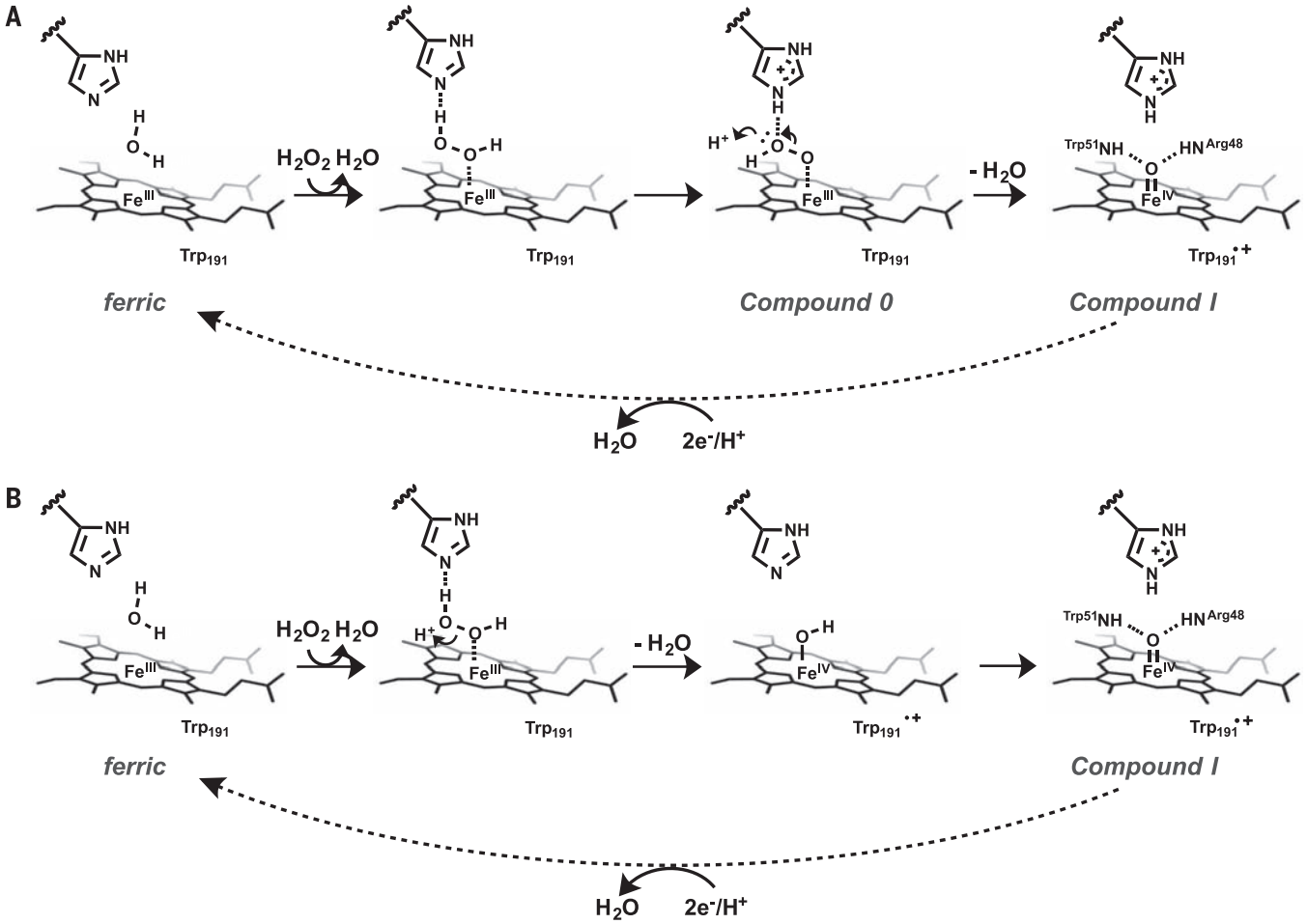


**Fig. 3. Possible movement of protons during O–O bond activation.** (**A**) Formation of a peroxide-bound complex and then compound 0 is followed by O–O bond cleavage. The distal histidine retains a proton upon formation of compound I, which means that an additional proton is required for release of water in both formation and reduction of compound I. (See text for detailed discussion.) The orientation of deuterium atoms on active site water molecules (Figs. 2 and 3) seems to preclude a water-mediated mechanism for O–O bond cleavage as suggested in horseradish peroxidase (*34*) because the deuterium atoms on W2 are oriented away from the ferryl oxygen atom and a hydrogen bond would not be possible (Fig. 2E). (**B**) An alternative mechanism from the same peroxide-bound complex. Cleavage of the peroxide bond may lead initially to formation of a transient iron(IV)-hydroxide species [Fe(IV)-OH], which, via proton transfer, leads to compound I, with Fe(IV)=O and a protonated distal histidine.

not interact ideally with any neighboring proton donors. This lack of ideal hydrogen bonding structure in the ferric form may favor formation of the peroxide complex and the deprotonated compound 0 (*28*), as structures of the closely related compound III (oxy) complexes all suggest that more favorable hydrogen bonding interactions with the bound peroxide are likely (fig. S6). Our data suggest that $His^{52}$ retains a proton upon formation of compound I. In terms of proton count, this presumably means that delivery of an additional proton to compound 0 is needed for compound I formation if water is released as product (Fig. 3A), aligning the mechanism more closely with the related P450s in which delivery of a single proton to compound 0 is also widely assumed (*2*, *29*). $Arg^{48}$, which is protonated (Figs. 1 and 2), could act as the source of the additional proton required for release of water in both formation and reduction of compound I, as it is thought to play a role in proton-relay networks (*30*). Hydroxide (instead of water, Fig. 3A) formation is also possible; if this is the case, a role for $Arg^{48}$ in charge stabilization of $OH^-$ is easily envisaged and would be consistent with early predictions (*31*). Because the proton on $His^{52}$ must be removed before a second turnover of the enzyme can occur, the acid-catalysis role of $His^{52}$ must occur after compound I formation and not [as is often assumed (*1*)] before (Fig. 3A). An alternative and intriguing possibility (Fig. 3B), which could apply generally to any heme protein with a distal histidine in the heme pocket, is that O–O bond cleavage initially forms a transient Fe(IV)-OH species [as in the P450s (*32*)], but the presence of the distal histidine provides an escape route for the proton, thus disfavoring formation of Fe(IV)-OH. Proton transfer in this direction, to form Fe(IV)=O, would be consistent with published $pK_a$s ($K_a$, acid dissociation constant) for the distal histidine across the peroxidase family [estimated $pK_a \approx 4$ to 5 (*1*, *33*)] if the $pK_a$ of the ferryl heme, which is not known reliably for a heme peroxidase, was lower than that.

## REFERENCES AND NOTES


1. H. B. Dunford, *Heme Peroxidases* (Wiley, Chichester, UK, 1999).
2. P. R. Ortiz de Montellano, *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Kluwer Academic/Plenum, New York, ed. 3, 2005).
3. K. D. Karlin, *Nature* **463**, 168–169 (2010).
4. S. G. Sligar, *Science* **330**, 924–925 (2010).
5. J. T. Groves, *Nat. Chem.* **6**, 89–91 (2014).
6. I. Schlichting *et al.*, *Science* **287**, 1615–1622 (2000).
7. G. I. Berglund *et al.*, *Nature* **417**, 463–468 (2002).
8. L. Que Jr., W. B. Tolman, *Nature* **455**, 333–340 (2008).
9. A. R. McDonald, L. Que Jr., *Nat. Chem.* **3**, 761–762 (2011).
10. J. Rittle, M. T. Green, *Science* **330**, 933–937 (2010).
11. D. G. Kellner, S. C. Hung, K. E. Weiss, S. G. Sligar, *J. Biol. Chem.* **277**, 9641–9644 (2002).
12. R. K. Behan, M. T. Green, *J. Inorg. Biochem.* **100**, 448–459 (2006).
13. J. Terner *et al.*, *J. Inorg. Biochem.* **100**, 480–501 (2006).
14. M. P. Blakeley, *Crystallogr. Rev.* **15**, 157–218 (2009).
15. V. L. Davidson, *Nat. Chem.* **3**, 662–663 (2011).
16. M. P. Blakeley, A. J. Kalb, J. R. Helliwell, D. A. Myles, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16405–16410 (2004).
17. M. Sivaraja, D. B. Goodin, M. Smith, B. M. Hoffman, *Science* **245**, 738–740 (1989).
18. T. L. Poulos *et al.*, *J. Biol. Chem.* **255**, 575–580 (1980).
19. A. Gumiero, C. L. Metcalfe, A. R. Pearson, E. L. Raven, P. C. Moody, *J. Biol. Chem.* **286**, 1260–1268 (2011).

20. H. Hori, T. Yonetani, *J. Biol. Chem.* **260**, 349–355 (1985).
21. S. L. Edwards, H. X. Nguyen, R. C. Hamlin, J. Kraut, *Biochemistry* **26**, 1503–1511 (1987).
22. A. E. Pond, G. S. Bruce, A. M. English, M. Sono, J. H. Dawson, *Inorg. Chim. Acta* **275–276**, 250–255 (1998).
23. J. E. Erman, T. Yonetani, *Biochim. Biophys. Acta* **393**, 350–357 (1975).
24. J. E. Huyett *et al.*, *J. Am. Chem. Soc.* **117**, 9033–9041 (1995).
25. Y. T. Meharenna, T. Doukov, H. Li, S. M. Soltis, T. L. Poulos, *Biochemistry* **49**, 2984–2986 (2010).
26. A. J. Sitter, C. M. Reczek, J. Terner, *J. Biol. Chem.* **260**, 7515–7522 (1985).
27. H. B. Dunford, *Peroxidases and Catalases: Biochemistry, Biophysics, Biotechnology, and Physiology* (Wiley, Chichester, UK, ed. 2, 2010).
28. E. Derat, S. Shaik, *J. Phys. Chem. B* **110**, 10526–10533 (2006).
29. S. Shaik, D. Kumar, S. P. de Visser, A. Altun, W. Thiel, *Chem. Rev.* **105**, 2279–2328 (2005).
30. I. Efimov *et al.*, *J. Am. Chem. Soc.* **133**, 15376–15383 (2011).
31. T. L. Poulos, J. Kraut, *J. Biol. Chem.* **255**, 8199–8205 (1980).
32. T. H. Yosca *et al.*, *Science* **342**, 825–829 (2013).
33. J. E. Erman, L. B. Vitello, M. A. Miller, J. Kraut, *J. Am. Chem. Soc.* **114**, 6592–6593 (1992).
34. P. Vidossich *et al.*, *J. Phys. Chem. B* **114**, 5161–5169 (2010).

**ACKNOWLEDGMENTS**

We thank D. Collison for assistance with EPR, P. Ortiz de Montellano for discussions on the mechanisms, S. Fisher for assistance with the joint x-ray/neutron refinement, and the University of Leicester BioMedical Workshop for custom building of microspectrophotometry equipment, and J. Devos and the D-lab for experimental support. Atomic coordinates have been deposited in the Protein Data Bank under accession codes 4CVI for ferric CcP and 4CVJ for compound I. This work was supported by The Leverhulme Trust (grant F/00 212/Q to E.L.R./P.C.E.M.), Biotechnology and Biological Sciences Research Council (grant BB/C001184/1 to E.L.R./P.C.E.M., and a studentship to E.J.M.), The Wellcome Trust (grant WT094104MA to P.C.E.M./E.L.R.), an Institut Laue-Langevin studentship (to C.M.C.) and beam time at LADI-III and BIODIFF (EU FP7 NMI3-II grant 283883), Bruker UK (Sponsorship of A.J.F. and M.G.C.), and beam time at LADI-III and BIODIFF. P.C.E.M., M.P.B., and E.L.R. designed the research; C.M.C., A.G., M.P.B., A.O., S.C.M.T, T.E.S., C.L.M., E.J.M., and P.C.E.M. performed crystallographic experiments; C.M.C., A.G., M.P.B., S.C.M.T., A.O., T.E.S., E.L.R., and P.C.E.M. analyzed crystallographic data; C.M.C., A.J.F., M.G.C., J.B., and P.C.E.M. performed spectroscopic experiments; C.M.C., A.G., A.J.F., M.G.C., E.L.R., and P.C.E.M. analyzed spectra; and E.L.R., M.P.B., and P.C.E.M. wrote the paper, with contributions from all authors.

**SUPPLEMENTARY MATERIALS**

www.sciencemag.org/content/345/6193/193/suppl/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 and S2
References (*35–48*)

4 April 2014; accepted 29 May 2014
10.1126/science.1254398

CATALYSIS

# Assessing the reliability of calculated catalytic ammonia synthesis rates

**Andrew J. Medford,[1,2] Jess Wellendorff,[1,2] Aleksandra Vojvodic,[1] Felix Studt,[1] Frank Abild-Pedersen,[1] Karsten W. Jacobsen,[3] Thomas Bligaard,[1] Jens K. Nørskov[1,2]***

**We introduce a general method for estimating the uncertainty in calculated materials properties based on density functional theory calculations. We illustrate the approach for a calculation of the catalytic rate of ammonia synthesis over a range of transition-metal catalysts. The correlation between errors in density functional theory calculations is shown to play an important role in reducing the predicted error on calculated rates. Uncertainties depend strongly on reaction conditions and catalyst material, and the relative rates between different catalysts are considerably better described than the absolute rates. We introduce an approach for incorporating uncertainty when searching for improved catalysts by evaluating the probability that a given catalyst is better than a known standard.**

With the surge in density functional theory (DFT) calculations of chemical and materials properties, the question of the reliability of calculated results becomes increasingly urgent (*1*), particularly when calculations are used to make predictions of new materials with interesting functionality (*2–5*). Evaluating the reliability of DFT calculations has relied mainly on comparisons to experiments or to data sets of higher-level calculations to provide a measure of the expected accuracy of directly calculated properties such as bond strengths, bond lengths, or activation energies of elementary processes. The question is how such intrinsic uncertainties in calculated microscopic properties transform into error bars on calculated complex properties, defined here as properties that depend on several microscopic properties in a complex way (*6*). Examples of such properties include mechanical strength (*7*), phase stability (*8*), and catalytic reaction rates (*5*).

We estimate the reliability of DFT energies by choosing an ensemble of exchange-correlation functionals to represent the known computational errors for a set of adsorption energies (*9–11*). This ensemble of energies is used to calculate the rates of the ammonia synthesis reaction via microkinetic modeling. We choose this process because it is well described both experimentally and theoretically (*12–16*) and has enough complexity to bring out important aspects of error propagation through multiple layers of simulation. This approach directly captures correlations between systematic errors in the underlying energetics, revealing that uncertainties on the calculated rates exhibit a nontrivial dependence on the reaction conditions as well as the material and that trends in catalytic activity are considerably better described than the absolute rates.

To calculate energies and estimated errors, we apply the Bayesian error estimation functional with van der Waals correlation (BEEF-vdW), a recent exchange-correlation density functional tailored for surface chemistry with built-in error estimation capabilities (*11*). The functional was fitted to describe several different properties, including common adsorbate-surface bond strengths, and an ensemble of density functionals around BEEF-vdW was generated. This Bayesian error estimation (BEE) ensemble was designed to reproduce known energetic errors by mapping them to uncertainties on the exchange-correlation model parameters. Figure 1 illustrates this. Uncertainties on new calculations may then be estimated by mapping back again: Random sampling of a probability distribution for fluctuations of the model parameters leads to a large ensemble of different predictions of the same quantity. The statistical variance of those predictions defines the error estimate on the BEEF-vdW result, $\sigma_{\mathrm{BEE}} = \sqrt{Var(\vec{p})}$, where the ensemble predictions are stacked in vector $\vec{p}$. Further details are provided in (*11*, *17*). This approach to quantitative error estimation in DFT can be viewed as a structured analysis of the sensitivity of DFT results to the choice of exchange-correlation approximation. An appropriately designed ensemble also captures correlated variations between DFT total energies and offers a consistent approach to keeping track of possible sources of error when data from multiple calculations are folded in composite post-DFT frameworks, such as microkinetic models used to analyze catalytic reactions.

For the ammonia synthesis reaction, microkinetic models provide the link between the calculated microscopic properties and the reaction rate or turnover frequency (TOF). Here, we use a relatively simple kinetic model based on $N_2$ dissociation as the rate-limiting step, following the mechanism described by Honkala *et al.* (*16*). This model has previously been shown to capture the experimentally observed trends in catalytic activity for different catalysts (*18*); details of the model can be found in (*17*, *18*). We first consider in some detail the rates of ammonia synthesis over stepped Fe and Ru surfaces, which are the industry-standard catalysts (*14*, *19*, *20*). The calculated ammonia synthesis rate over iron per active (step) site (the TOF) is shown in Fig. 2A as an Arrhenius plot at industrial conditions. The red shaded area indicates that the estimated

[1]SUNCAT Center for Interface Science and Catalysis, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. [2]SUNCAT Center for Interface Science and Catalysis, Department of Chemical Engineering, Stanford University, Stanford, CA 94305, USA. [3]Center for Atomic-scale Materials Design (CAMD), Department of Physics, Technical University of Denmark, DK-2800 Lyngby, Denmark.
***Corresponding author. E-mail: norskov@stanford.edu**

www.sciencemag.org/content/345/6193/193/suppl/DC1

# Supplementary Materials for

## Neutron cryo-crystallography captures the protonation state of ferryl heme in a peroxidase

Cecilia M. Casadei, Andrea Gumiero, Clive L. Metcalfe, Emma J. Murphy, Jaswir Basran, Maria Grazia Concilio, Susana C. M. Teixeira, Tobias E. Schrader, Alistair J. Fielding, Andreas Ostermann, Matthew P. Blakeley, Emma L. Raven,* Peter C. E. Moody*

*Corresponding author. E-mail: emma.raven@le.ac.uk (E.L.R.); peter.moody@le.ac.uk (P.C.E.M.)

Published 11 July 2014, *Science* **345**, 193 (2014)
DOI: 10.1126/science.1254398

**This PDF file includes:**

**Supplementary materials**

# Neutron cryo-crystallography captures the protonation state of ferryl heme in a peroxidase

*Cecilia M. Casadei[1,2], Andrea Gumiero[3], Clive L. Metcalfe[3], Emma J. Murphy[3], Jaswir Basran[1], Maria Grazia Concilio[4], Susana C. M. Teixeira[2,5], Tobias E. Schrader[6], Alistair J. Fielding[4], Andreas Ostermann[7], Matthew P. Blakeley[2], Emma L. Raven[3§] & Peter C. E. Moody[1§]*

## Materials and Methods

**Protein expression, purification, crystallisation and deuteration.** The original MKT variant of C*c*P (*35*) was a gift from Professor Grant Mauk (University of British Columbia) and was cloned into the expression plasmid pLEICS-03 carrying kanamycin resistance and a TEV-cleavable His-Tag sequence and expressed in BL21 DE3 Gold. The protein was purified and crystallised at pH 6.0 according to published methods (*36, 37*). We have used deuterated C*c*P crystals in which labile hydrogen atoms are exchanged for deuterium atoms to enhance their visibility in nuclear scattering density maps. Deuteration of C*c*P was carried out by either transferring the dialysis buttons containing crystals into identical mother liquor made up with $D_2O$ instead of water and leaving the crystals to exchange for at least 24 h, or by crystallisation from mother liquor made up in $D_2O$. Compound I was prepared by reaction with $H_2O_2$ in the manner previously described (*19*).

**Crystallography.** Neutron diffraction data for ferric C*c*P at room temperature were collected using the LADI-III beamline at the Institut Laue-Langevin, Grenoble, France and for Compound I at 100 K using the BIODIFF beamline at the FRM II research reactor at the Heinz Maier-Leibnitz Zentrum (MLZ). X-ray data for ferric C*c*P crystals were collected at the EMBL Grenoble Outstation. Compound I X-ray diffraction data, and all solution and single crystal spectrophotometry were recorded at the University of Leicester. For ferric C*c*P, neutron (2.4 Å) and X-ray (2.1 Å) diffraction data were collected at room temperature and the structure obtained using both sets of data; neutron (2.5 Å) and X-ray (2.18 Å) diffraction data for Compound I were collected at 100K and the structure obtained in the same way. Atomic coordinates have been deposited in the Protein Data Bank under accession codes 4CVI for ferric C*c*P and 4CVJ for Compound I.

*Neutron diffraction data collection of ferric CcP.* A large single crystal (approximately 1 $mm^3$) of deuterated C*c*P was mounted in a quartz capillary, surrounded by a small amount of mother liquor and sealed with wax ready for data collection. Quasi-Laue neutron diffraction data to 2.4 Å resolution were collected at room temperature from the deuterated C*c*P crystal on the LADI-III beamline (*38*) at the Institut Laue-Langevin. As is typical for a Laue experiment, the crystal was held stationary at a different φ setting for each exposure. In total 19 images were collected (with an average exposure time of 23.2 h per image) from 3 different crystal orientations. The neutron data were processed using the program LAUEGEN (*39*) modified to account for the cylindrical geometry of the detector (*40*). The program LSCALE (*41*) was used to determine the wavelength-normalization curve using the intensities of symmetry-equivalent reflections measured at different wavelengths. No explicit absorption corrections were applied. These data were then merged in SCALA (*42*). The statistics are shown in Supplementary Table S1.

*Neutron diffraction data collection of Compound I.* A large single crystal (1.2 x 0.7 x 0.8 mm) of C*c*P was reacted with peroxide (*19*) to form Compound I and mounted on a 1mm LithoLoop on a magnetic cap (Molecular Dimensions). The crystal was cryo-cooled directly in a stream of nitrogen gas at 100 K from an Oxford Cryosystems cryostream. Monochromatic neutron diffraction data were collected to 2.5 Å resolution at 100 K using the BIODIFF beamline at the FRM II research reactor at the Heinz Maier-Leibnitz Zentrum

(MLZ). At a wavelength of 3.39 Å, 201 frames of 0.3° oscillation and 120 minutes exposure were first recorded. To increase the completeness the crystal was tilted by 45° from the rotation axis and a second series of 67 frames were collected at a wavelength of 3.98 Å with an oscillation of 0.3° and 120 minutes exposure. The diffraction data were indexed and integrated using DENZO (v.1.96.2) and scaled using SCALEPACK (v.2.3.6) (*43*). The statistics are shown in Supplementary Table S2.

*X-ray diffraction data collection of ferric CcP.* X-ray diffraction data were collected at room temperature (from the same single crystal used for neutron data collection at LADI-III) using $CuK_\alpha$ radiation ($\lambda$ = 1.5418 Å) from a GeniX Cu HF generator. A total of 204 images of 0.5° oscillation and 60 s exposure were recorded on a Mar345 detector. Diffraction intensities were indexed and measured using the XDS package (*44*). Statistics are shown in Supplementary Table S1.

*X-ray diffraction data collection of Compound I.* A single crystal was reacted as above and maintained at 100K using an Oxford Cryosystems cryostream. 270 images of 0.5° oscillation and 1s exposure were collected using $CuK_\alpha$ radiation ($\lambda$ = 1.5418 Å) from a Rigaku 007 HF X-ray generator fitted with Osmic VariMax HF optics and a Rigaku Saturn 944+ CCD detector. Recorded intensities were indexed and measured using iMOSFLM then scaled and merged using AIMLESS (*45*). Statistics are shown in Supplementary Table S2.

**Structural refinement of ferric C*c*P.** The crystal used for neutron and X-ray data collection was essentially isomorphous with the previously determined X-ray structure of ferric C*c*P (PDB ID 2YCG). 2YCG was modified to remove multiple conformations and all solvent molecules and then used as the starting point for structural refinement using the PHENIX software suite (*46*). Firstly the model was refined with the X-ray terms alone, followed by joint X-ray and neutron refinement (*47*). H- and D-atoms were added with the program *ReadySet (46)*. $D_2O$ molecules were added according to clear positive peaks in the $F_o$ - $F_c$ difference nuclear scattering density maps and all model modifications were made with the modelling program *COOT (48).* The position of the heme coordinated water molecule (W1) was determined using the neutron data only. 198 $D_2O$ water molecules were added in total. Final refinement statistics are shown in Supplementary Table S1.

**Structural refinement of Compound I.** The crystals used for neutron and X-ray data collection were essentially isomorphous with the previously determined X-ray structure of C*c*P Compound I (PDB ID 2XIL). 2XIL was modified to remove multiple conformations, the ferryl oxygen and all solvent molecules and then used as the starting point for structural refinement using the PHENIX software suite (*46*). The model was first refined with the X-ray terms alone, followed by joint X-ray and neutron refinement (*47*). H- and D-atoms were added with the program *ReadySet* (*46*). $D_2O$ molecules were added according to clear positive peaks in the $F_o$ - $F_c$ difference nuclear scattering density maps and all model modifications were made with the modelling program *COOT* (*48*). The position of the ferryl oxygen was determined using the neutron data only. 315 water molecules were added in total; 298 were modelled as full $D_2O$ molecules, while 17 were partially disordered showing spherical density and were modelled as oxygen atoms only. Final refinement statistics are shown in Supplementary Table S2.

**Single crystal microspectrophotometry.** Absorption spectra of single crystals of C*c*P Compound I (obtained by soaking crystals in $H_2O_2$ (30 mM) at 4˚C followed by flash freezing at 77 K, as previously (*19*)) were collected using an Ocean Optics Maya 2000 PRO spectrometer, with an Ocean Optics DH-2000-BAL UV-VIS-NIR light source and a Hamamatsu S10420 FFT-CCD back-thinned detector with fibre optic coupled to 80 mm diameter 4X reflective lenses (Optic Peter, Lentilly, France) and mounted with a custom mount (BioMedical Mechanical Workshop, University of Leicester) on a Rigaku Raxis IV φ drive. Crystals were mounted on a nylon loop in a SPINE standard mount. The temperature was maintained at 100 K with an Oxford Cryosystems cryostream. Absorption spectra were acquired by means of the Ocean Optics SpectraSuite software.

**Single crystal EPR Spectroscopy.** Continuous-wave EPR spectra were recorded at 9.4 GHz on a Bruker EMX spectrometer with a Super-high-Q rectangular cavity and an Oxford ESR-900 liquid helium cryostat. The single crystal was mounted separately on a quartz sample holder. The operating conditions are stated on the Figure legends. Deuterated C*c*P crystals were reacted with peroxide in the same way as for the single crystal microspectrophotometry above.

**Table S1** Data collection and refinement statistics for ferric C*c*P .

| | |
|---|---|
| Space group | $P2_12_12_1$ |
| Cell dimensions | |
| a, b, c (Å) | 51.7 76.8 107.6 |
| **Data collection (Neutron, quasi-Laue λ = 3.2-4.2 Å )** | |
| Resolution (Å) | 39.7 -2.4 (2.53-2.4)* |
| $R_{merge}$ | 0.151 (0.194) |
| I / σI | 7.6 (4.5) |
| Completeness (%) | 75.2 (48.0) |
| Redundancy | 3.8 (2.4) |
| **Data collection (Xray λ = 1.5418 Å)** | |
| Resolution (Å) | 14.8-2.1 (2.15-2.1)* |
| $R_{merge}$ | 0.061 (0.243) |
| I / σI | 14.6 (4.6) |
| Completeness (%) | 94.7 (97.3) |
| Redundancy | 3.5 (2.4) |

**Refinement**

| | |
|---|---|
| $d_{min}$ (Neutron) | 2.4(Å) |
| $d_{min}$ (Xray) | 2.1 (Å) |
| Number of reflections (Neutron) | 12695 |
| Number of reflections (X-ray) | 24307 |
| Rwork/ Rfree (Neutron) | 0.1759/0.2433 |
| Rwork/ Rfree (X-ray) | 0.1337/0.1766 |
| R.m.s deviations | |
| Bond lengths (Å) | 0.012 |
| Bond angles (°) | 1.349 |

*values in parenthesis are for the outer resolution bin

**Table S2** Data collection and refinement statistics for Compound I of C*c*P .

| | |
|---|---|
| Space group | $P2_12_12_1$ |
| Cell dimensions | |
| a, b, c (Å) | 51.19 75.83 107.59 |
| **Data collection (Neutron λ = 3.39 Å, λ = 3.98 Å)** | |
| Resolution (Å) | 50 -2.5 (2.59-2.5)* |
| $R_{merge}$ | 0.173 (0.482) |
| I / σI | 4.6 (1.5) |
| Completeness (%) | 90.7 (71.8) |
| Redundancy | 2.3 (1.7) |
| **Data collection (Xray λ = 1.5418 Å)** | |
| Resolution (Å) | 17-2.18 (2.25-2.18)* |
| $R_{merge}$ | 0.074 (0.164) |
| I / σI | 16.5 (5.9) |
| Completeness (%) | 99.3 (94.9) |
| Redundancy | 4.7 (2.6) |

**Refinement**

| | |
|---|---|
| $d_{min}$ (Neutron) | 2.5(Å) |
| $d_{min}$ (Xray) | 2.18(Å) |
| Number of reflections (Neutron) | 13661 |
| Number of reflections (X-ray) | 22053 |
| Rwork/ Rfree (Neutron) | 0.1873/0.2720 |
| Rwork/ Rfree (X-ray) | 0.1488/0.2052 |
| R.m.s deviations | |
| Bond lengths (Å) | 0.011 |
| Bond angles (°) | 1.261 |

*values in parenthesis are for the outer resolution bin

**Figure S1**
***Casadei et al***

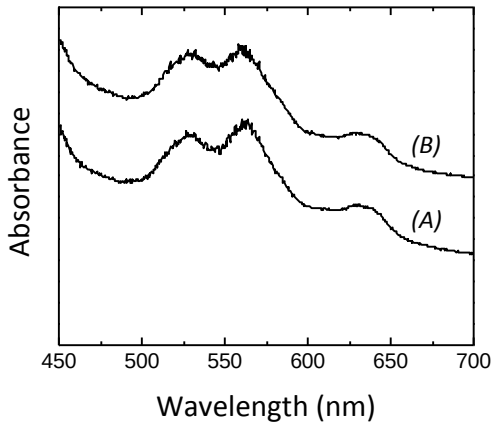


**Figure S1. Single crystal UV-visible spectra of CcP.** (A) UV-visible spectrum (at 100 K) of a deuterium-exchanged single crystal of CcP after soaking with peroxide (see Methods), showing the characteristic Compound I peaks in the visible region. (B) UV-visible spectrum (at 100 K) of the same crystal as in (A), after storage at 77K for 20 days. The spectrum in (B) is offset on the y-axis, for clarity.

**Figure S2**
***Casadei et al***

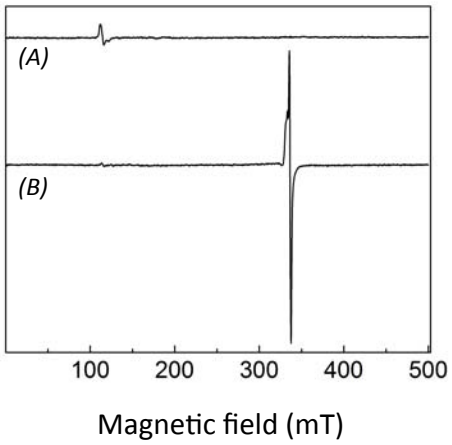


**Figure S2. Single crystal EPR spectra of CcP.** 9 GHz EPR spectra of CcP single crystals. (A) in the ferric state; (B) after soaking in hydrogen peroxide. Spectra recorded at 8 K, 0.1 mT modulation amplitude, 1 mW, 2 scans, 2048 points.

**Figure S3**
***Casadei et al***


Wavelength (nm)

**Figure S3. Solution spectra of CcP.** UV-visible spectra of Compound I of CcP in solution (formed by reaction with 5 equivalents of $H_2O_2$) recorded immediately after addition of peroxide and then again 3 weeks after storage in liquid nitrogen. Solid line is the spectrum collected immediately after mixing with $H_2O_2$ and the dashed line is the spectrum of the same sample collected 3 weeks later. Buffer: 10 mM potassium phosphate, 150 mM KCl, pH 6.5. (We also measured the stability of Compound I in solution, at room temperature, on reaction with 1.5-2 equiv. of peroxide (10 mM potassium phosphate, 150 mM KCl, pH 6.5). The decay of the Compound I absorbance maximum (419 nm) gave a half life for decay of Compound I to ferric as 0.006 ± 0.0001 $min^{-1}$, which is in approximate agreement with early data from Erman *(23)*).

**Figure S4**
***Casadei et al***

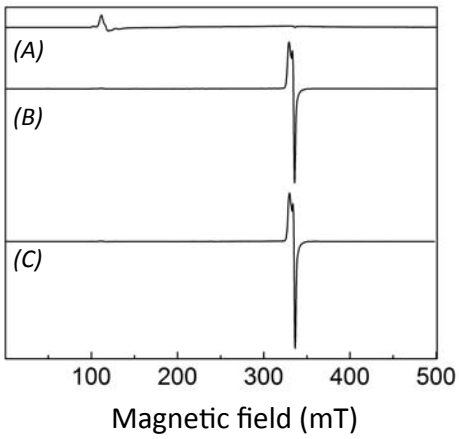


**Figure S4.** 9 GHz EPR spectra of a solution of (A) ferric CcP, (B) Compound I of CcP immediately after mixing with $H_2O_2$, and (C) the same sample as (B) but after storage for 20 days in liquid nitrogen. Spectra were recorded at 4.5 K, 0.1 mT modulation amplitude, 0.04743 mW power, 4 scans, 2048 points. Compound I was prepared by manually mixing 256 µM native enzyme (10 mM potassium phosphate buffer, 150 mM KCl, pH 6.5) with 5-fold/equivolume hydrogen peroxide directly into 4 mm quartz tubes. The mixing time was 15 seconds. A control experiment was performed mixing in equivolume of buffer. All reagents were kept on ice.

**Figure S5**
***Casadei et al***

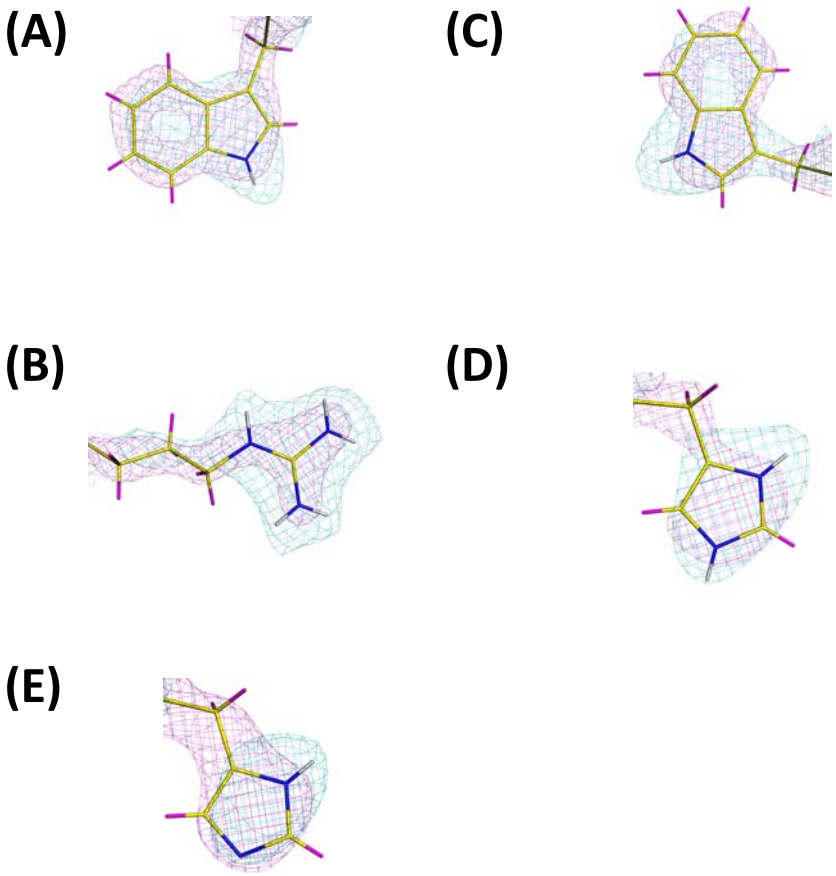


**Figure S5. Electron and nuclear scattering density of the active site side chains observed in Compound I (A-D) and ferric (E) structures.** Electron density from the X-ray structures is shown in magenta, and nuclear scattering density from the neutron structures is overlaid in blue. (A) Trp51, showing that Nε is deuterated. (B) Arg48, showing the deuteration of all the nitrogen atoms of the guanidinium group. (C) Trp191, showing that Nε is deuterated. (D) His52, showing deuteration at both Nδ and Nε. (E) For comparison, His52 in the ferric C*c*P structure is presented, showing His52 as neutral. The contour levels for ferric and Compound I are the same as those in Figures 1 and 2, respectively. The atoms are coloured according to the convention of Figure 1.

**Figure S6**
***Casadei et al***

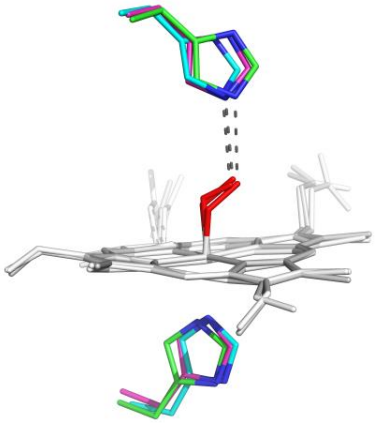

**Figure S6. Structures of peroxidase oxy-complexes.** Overlay of structures of oxy complexes from CcP (in green, PDB 1DCC), ascorbate peroxidase (in magenta, PDB 2XIH) and horseradish peroxidase (in cyan, PDB 1H57). The heme (gray) with the oxy group (red) is shown in all cases.

**References and Notes**


1. H. B. Dunford, *Heme Peroxidases* (Wiley, Chichester, UK, 1999).

2. P. R. Ortiz de Montellano, *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Kluwer Academic/Plenum, New York, ed. 3, 2005).

3. K. D. Karlin, Bioinorganic chemistry: Model offers intermediate insight. *Nature* **463**, 168–169 (2010). Medline doi:10.1038/463168a

4. S. G. Sligar, Glimpsing the critical intermediate in cytochrome P450 oxidations. *Science* **330**, 924–925 (2010). Medline doi:10.1126/science.1197881

5. J. T. Groves, Enzymatic C-H bond activation: Using push to get pull. *Nat. Chem.* **6**, 89–91 (2014). Medline doi:10.1038/nchem.1855

6. I. Schlichting, J. Berendzen, K. Chu, A. M. Stock, S. A. Maves, D. E. Benson, R. M. Sweet, D. Ringe, G. A. Petsko, S. G. Sligar, The catalytic pathway of cytochrome P450cam at atomic resolution. *Science* **287**, 1615–1622 (2000). Medline doi:10.1126/science.287.5458.1615

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

8. L. Que Jr., W. B. Tolman, Biologically inspired oxidation catalysis. *Nature* **455**, 333–340 (2008). Medline doi:10.1038/nature07371

9. A. R. McDonald, L. Que Jr., Iron-oxo complexes: Elusive iron(V) species identified. *Nat. Chem.* **3**, 761–762 (2011). Medline doi:10.1038/nchem.1153

10. J. Rittle, M. T. Green, Cytochrome P450 compound I: Capture, characterization, and C-H bond activation kinetics. *Science* **330**, 933–937 (2010). Medline doi:10.1126/science.1193478

11. D. G. Kellner, S. C. Hung, K. E. Weiss, S. G. Sligar, Kinetic characterization of compound I formation in the thermostable cytochrome P450 CYP119. *J. Biol. Chem.* **277**, 9641–9644 (2002). Medline doi:10.1074/jbc.C100745200

12. R. K. Behan, M. T. Green, On the status of ferryl protonation. *J. Inorg. Biochem.* **100**, 448–459 (2006). Medline doi:10.1016/j.jinorgbio.2005.12.019

13. J. Terner, V. Palaniappan, A. Gold, R. Weiss, M. M. Fitzgerald, A. M. Sullivan, C. M. Hosten, Resonance Raman spectroscopy of oxoiron(IV) porphyrin pi-cation radical and oxoiron(IV) hemes in peroxidase intermediates. *J. Inorg. Biochem.* **100**, 480–501 (2006). Medline doi:10.1016/j.jinorgbio.2006.01.008

14. M. P. Blakeley, Neutron macromolecular crystallography. *Crystallogr. Rev.* **15**, 157–218 (2009). doi:10.1080/08893110902965003

15. V. L. Davidson, Ion-protein coordination: The many faces of a proton. *Nat. Chem.* **3**, 662–663 (2011). Medline doi:10.1038/nchem.1122

16. M. P. Blakeley, A. J. Kalb, J. R. Helliwell, D. A. Myles, The 15-K neutron structure of saccharide-free concanavalin A. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16405–16410 (2004). Medline doi:10.1073/pnas.0405109101

17. M. Sivaraja, D. B. Goodin, M. Smith, B. M. Hoffman, Identification by ENDOR of $Trp^{191}$ as the free-radical site in cytochrome c peroxidase compound ES. *Science* **245**, 738–740 (1989). Medline doi:10.1126/science.2549632

18. T. L. Poulos, S. T. Freer, R. A. Alden, S. L. Edwards, U. Skogland, K. Takio, B. Eriksson, N. Xuong, T. Yonetani, J. Kraut, The crystal structure of cytochrome *c* peroxidase. *J. Biol. Chem.* **255**, 575–580 (1980). Medline

19. A. Gumiero, C. L. Metcalfe, A. R. Pearson, E. L. Raven, P. C. Moody, Nature of the ferryl heme in compounds I and II. *J. Biol. Chem.* **286**, 1260–1268 (2011). Medline doi:10.1074/jbc.M110.183483

20. H. Hori, T. Yonetani, Powder and single-crystal electron paramagnetic resonance studies of yeast cytochrome c peroxidase and its peroxide and its peroxide compound, Compound ES. *J. Biol. Chem.* **260**, 349–355 (1985). Medline

21. S. L. Edwards, H. X. Nguyen, R. C. Hamlin, J. Kraut, Crystal structure of cytochrome *c* peroxidase compound I. *Biochemistry* **26**, 1503–1511 (1987). Medline doi:10.1021/bi00380a002

22. A. E. Pond, G. S. Bruce, A. M. English, M. Sono, J. H. Dawson, Spectroscopic study of the compound ES and the oxoferryl compound II states of cytochrome *c* peroxidase: Comparison with the compound II of horseradish peroxidase. *Inorg. Chim. Acta* **275–276**, 250–255 (1998). doi:10.1016/S0020-1693(97)06106-9

23. J. E. Erman, T. Yonetani, A kinetic study of the endogenous reduction of the oxidized sites in the primary cytochrome c peroxidase-hydrogen peroxide compound. *Biochim. Biophys. Acta* **393**, 350–357 (1975). Medline doi:10.1016/0005-2795(75)90061-6

24. J. E. Huyett, P. E. Doan, R. Gurbiel, A. L. P. Houseman, M. Sivaraja, D. B. Goodin, B. M. Hoffman, Compound ES of cytochrome-c peroxidase contains a Trp pi-cation radical: Characterization by continuous wave and pulsed Q-band external nuclear double resonance spectroscopy. *J. Am. Chem. Soc.* **117**, 9033–9041 (1995). doi:10.1021/ja00140a021

25. Y. T. Meharenna, T. Doukov, H. Li, S. M. Soltis, T. L. Poulos, Crystallographic and single-crystal spectral analysis of the peroxidase ferryl intermediate. *Biochemistry* **49**, 2984–2986 (2010). Medline doi:10.1021/bi100238r

26. A. J. Sitter, C. M. Reczek, J. Terner, Heme-linked ionization of horseradish peroxidase compound II monitored by the resonance Raman Fe(IV)=O stretching vibration. *J. Biol. Chem.* **260**, 7515–7522 (1985). Medline

27. H. B. Dunford, *Peroxidases and Catalases: Biochemistry, Biophysics, Biotechnology and Physiology* (Wiley, Chichester, UK, ed. 2, 2010).

28. E. Derat, S. Shaik, The Poulos-Kraut mechanism of Compound I formation in horseradish peroxidase: A QM/MM study. *J. Phys. Chem. B* **110**, 10526–10533 (2006). Medline doi:10.1021/jp055412e

29. S. Shaik, D. Kumar, S. P. de Visser, A. Altun, W. Thiel, Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. *Chem. Rev.* **105**, 2279–2328 (2005). Medline doi:10.1021/cr030722j

30. I. Efimov, S. K. Badyal, C. L. Metcalfe, I. Macdonald, A. Gumiero, E. L. Raven, P. C. Moody, Proton delivery to ferryl heme in a heme peroxidase: Enzymatic use of the Grotthuss mechanism. *J. Am. Chem. Soc.* **133**, 15376–15383 (2011). Medline doi:10.1021/ja2007017

31. T. L. Poulos, J. Kraut, The stereochemistry of peroxidase catalysis. *J. Biol. Chem.* **255**, 8199–8205 (1980). Medline

32. T. H. Yosca, J. Rittle, C. M. Krest, E. L. Onderko, A. Silakov, J. C. Calixto, R. K. Behan, M. T. Green, Iron(IV)hydroxide p$K_a$ and the role of thiolate ligation in C–H bond

activation by cytochrome P450. *Science* **342**, 825–829 (2013). Medline doi:10.1126/science.1244373

33. J. E. Erman, L. B. Vitello, M. A. Miller, J. Kraut, Active-site mutations in cytochrome *c* peroxidase: A critical role for histidine-52 in the rate of formation of Compound I. *J. Am. Chem. Soc.* **114**, 6592–6593 (1992). doi:10.1021/ja00042a068

34. P. Vidossich, G. Fiorin, M. Alfonso-Prieto, E. Derat, S. Shaik, C. Rovira, On the role of water in peroxidase catalysis: A theoretical investigation of HRP compound I formation. *J. Phys. Chem. B* **114**, 5161–5169 (2010). Medline doi:10.1021/jp911170b

35. D. B. Goodin, M. G. Davidson, J. A. Roe, A. G. Mauk, M. Smith, Amino acid substitutions at tryptophan-51 of cytochrome c peroxidase: Effects on coordination, species preference for cytochrome c, and electron transfer. *Biochemistry* **30**, 4953–4962 (1991). Medline doi:10.1021/bi00234a017

36. C. Metcalfe, I. K. Macdonald, E. J. Murphy, K. A. Brown, E. L. Raven, P. C. Moody, The tuberculosis prodrug isoniazid bound to activating peroxidases. *J. Biol. Chem.* **283**, 6193–6200 (2008). Medline doi:10.1074/jbc.M707412200

37. E. J. Murphy, C. L. Metcalfe, J. Basran, P. C. Moody, E. L. Raven, Engineering the substrate specificity and reactivity of a heme protein: Creation of an ascorbate binding site in cytochrome c peroxidase. *Biochemistry* **47**, 13933–13941 (2008). Medline doi:10.1021/bi801480r

38. M. P. Blakeley, S. C. Teixeira, I. Petit-Haertlein, I. Hazemann, A. Mitschler, M. Haertlein, E. Howard, A. D. Podjarny, Neutron macromolecular crystallography with LADI-III. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 1198–1205 (2010). Medline doi:10.1107/S0907444910019797

39. J. W. Campbell, *LAUEGEN*, an X-windows-based program for the processing of Laue diffraction data. *J. Appl. Cryst.* **28**, 228–236 (1995). doi:10.1107/S002188989400991X

40. J. W. Campbell, Q. Hao, M. M. Harding, N. D. Nguti, C. Wilkinson, *LAUEGEN* version 6.0 and *INTLDM*. *J. Appl. Cryst.* **31**, 496–502 (1998). doi:10.1107/S0021889897016683

41. S. Arzt, J. W. Campbell, M. M. Harding, Q. Hao, J. R. Helliwell, *LSCALE* - the new normalization, scaling and absorption correction program in the Daresbury *Laue* software suite. *J. Appl. Cryst.* **32**, 554–562 (1999). doi:10.1107/S0021889898015350

42. M. S. Weiss, Global indicators of x-ray data quality. *J. Appl. Cryst.* **34**, 130–135 (2001). doi:10.1107/S0021889800018227

43. Z. Otwinoski, W. Minor, *Methods in Enzymology*, vol. 276, *Macromolecular Crystallography*, Part A, C. W. Carter Jr., R. M. Sweet, Eds. (Academic Press, New York, 1997), pp. 307–326.

44. W. Kabsch, *XDS*. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010). Medline doi:10.1107/S0907444909047337

45. Collaborative Computational Project, Number 4, The *CCP4* suite: Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994). Medline doi:10.1107/S0907444994003112

46. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *PHENIX*: A comprehensive Python-based system for

macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010). Medline doi:10.1107/S0907444909052925

47. P. V. Afonine, M. Mustyakimov, R. W. Grosse-Kunstleve, N. W. Moriarty, P. Langan, P. D. Adams, Joint x-ray and neutron refinement with *phenix.refine*. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 1153–1163 (2010). Medline doi:10.1107/S0907444910026582

48. P. Emsley, K. Cowtan, *Coot*: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004). Medline doi:10.1107/S0907444904019158

## Publications related to this PhD work

**Science, 2014**, *Casadei, C. M.; Gumiero, A.; Metcalfe, C. L.; Murphy, E. J.; Basran, J.; Concilio, M. G.; Teixeira, S. C. M.; Schrader, T. E.; Fielding A. J.; Ostermann, A.; Blakeley M. P.; Raven, E. L.; Moody, P. C. E.*, Vol. 345, pages 193-197.
'Neutron cryo-crystallography captures the protonation state of ferryl heme in a peroxidase.'
This paper was selected for F1000Prime, as being of special significance in its field by Faculty Member Victor Davidson.

## Conferences with oral contribution

03.2015 **NMX@ESS - Institut de Biologie Structurale**, Grenoble, France.
Oral contribution (invited speaker): 'Neutron crystallography captures the protonation state of ferryl heme in a Peroxidase.'

12.2014 **Biological Structures Group Winter Meeting 2014 - British Crystallography Association**, Grenoble, France.
Oral contribution (invited speaker): 'Neutron crystallography captures the protonation state of ferryl heme in a Peroxidase.'

10.2014 **2014: Crystal (cl)Year - Molecular Biotechnology Center**, Torino, Italy.
Oral contribution: 'Transient intermediates in heme peroxidases: a multidisciplinary approach.'

06.2014 **1st PSB Blue Symposium - Institut de Biologie Structurale**, Grenoble, France.
Oral contribution: 'Transient intermediates in heme peroxidases: a multidisciplinary approach.'

06.2014 **Current challenges in integrated structural biology - Institut de Génétique et de Biologie Moleculaire et Cellulaire**, Strasbourg, France.
Oral contribution: 'Transient intermediates in heme peroxidases: a multidisciplinary approach.'

## Awards

10.2015 **Shortlisted for Reaxys PhD prize 2015**, Hong Kong.

06.2015 **65th Lindau Nobel Laureate Meeting: interdisciplinary**, Lindau, Germany.
I have been selected for participation to the 65th Lindau Meeting where 650 young scientists met with 65 Nobel Laureates in different disciplines. The meeting is scheduled in June 2015.

10.2014 **2014: Crystal (cl)Year - Molecular Biotechnology Center**, Torino, Italy.
Best oral contribution award.

## Other relevant experience

2015 **'All you need is neutrons' seminar series** , Institut Laue-Langevin, Grenoble, France.
Contribution with lecture 'Scattering theory and the Born approximation'.

09.2014 **1st European Crystallography School**, Pavia, Italy.

2014 **SOPHIE course 2014**, Institut Laue-Langevin, Grenoble, France.
Scientific Oral Presentation Held In English

03.2014 **Hercules European School 2014**, Grenoble, France.
Higher European Research Course for Users of Large Scale Experimental Systems

09.2013 **13th Oxford School on Neutron Scattering**, St Anne's College, University of Oxford, Oxford, UK.

01.2013 **CCP4 Study Weekend**, University of Nottingham, Nottingham, UK.

09.2012 **Training Course in Complementary Optical Spectroscopies in Macromolecular Crystallography**, Grenoble, France.
COSMX 2012 - ESRF, EMBL Grenoble outstation, IBS.