Structural Determinants of Peroxidase Activities

Sarah Pannell

A thesis submitted for the degree of
Doctor of Philosophy (Biochemistry)
at the University of Sussex
April 2011

Declaration

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Signed_____

Sarah Pannell

In memory of my grandmother

Ruth Parker

who would have been surprised how much, and yet how little,

has changed in a biochemistry lab in the past 60 years.

1923 - 2006

Acknowledgements

Having finally completed this thesis it is not surprising that I have an army of people that deserve my profuse thanks.

My supervisors Professor Andy Smith and Dr Darren Thompson have been an amazing source of scientific wisdom (Andy) and coffee (Darren) during the past five years, and I thank them both for their support and the provision of the interesting and stretching project that fills the next 200 pages. It has been a pleasure to work in both their labs.

Dr Wendy Doyle has been a fount of all knowledge to PhD students in the Smith lab for probably more years than she'd like me to note. Without her, very little of the work documented in this thesis would have been possible, as I'd still be figuring out how to work the FPLC. I'm still astonished she manages to put up with us all. Wendy, there probably isn't enough chocolate in the world to thank you!

Both the Lab 13 and CRPC extended lab groups have been great environments to work in, with a ready flow of tea, cake and discussion (scientific or otherwise). I would like to particularly thank, in addition to Wendy, Dr Maria Karatsa-Dodgson, Dr Andrea Fallas, Dr Yalda Javadi, Dr Jonathan Phillips, Dr Karen Marshall, and nearly-Drs Kate Elliott, Caroline Sanvitale, Sunny Singh, Elodie Bentley, Lisa Reeve and Kyle Morris for their help and encouragement. I would also like to thank Dr Khasim Cali, whose previous DPhil work provided the proteins from which Chapters 5 and 6 of this thesis were written and who was kind enough to share his research with me.

Dr Amelia Philpott, who I met (what now seems like centuries ago) as a first year undergraduate, has been a regular sounding board and was always good for a coffee and whinge when my research wasn't going quite the way I planned it to. The Postgraduate Seminar Series that Amelia and I organised and later the Symposium (with Kate Elliott) was a welcome distraction from my own research and helped convince me that I wanted to encourage others to talk about science in my future career. Dr Sarah Robins-Hobden has also been a wonderful reassurance to me (and sometimes provided a kick up the bum) during the writing stage, encouraging me that it is possible to finish a thesis while working, though it was more difficult than either of us really imagined.

Throughout my academic journey my family, particularly my parents, have been a fantastic source of support and encouragement. I specifically thank them for not asking "haven't you finished it yet?" too often (though often enough that I didn't forget it wasn't quite finished yet)!

And last, though by no means least, my biggest thanks go to my ever-tolerant husband Rob, who I blame entirely for getting me into this mess in the first place, and without whose support, patience and love I would never have made it out the other side.

Summary

Horseradish peroxidase (HRP) is a robust enzyme with commercial applications as an immunodiagnostic reporter enzyme and in the catalysis of difficult chemical transformations. The commercial enzyme is still isolated from the roots of the horseradish plant *Armoracia rusticana*, and has been studied as a model haem enzyme system since the early 1940's. Following the development of methods to produce the active recombinant enzyme in *E.coli* (Smith *et al.*, 1990) and completion of the crystallographic structure in 1997 (Gajhede *et al*., 1997) it has been possible to identify the structural requirements for activity and to extend these activities by protein engineering techniques. Three aspects relating to the enhancement of the 'normal' and 'designed' activities of selected variants have been explored at the structure function level in this work.

Earlier work (Gajhede *et al*., 1997; White *et al*., 1997) highlighted residues that interact with aromatic substrates but that also potentially occlude access to the reactive haem edge by larger bulky substrates of potential commercial interest. The X-ray structure of the HRP-C* A140G/F179A variant was solved to 2.0Å. A larger engineered cavity at the haem edge was observed consistent with the ability of the variant to oxidise luminol directly, a property not seen in the wild-type enzyme. The structure factors for other residues in the haem access channel were not significantly affected.

The structure of the S167M HRP variant was also solved, because of interest in its ability to form a novel sulphonium linkage to the haem vinyl group (K. Cali, DPhil thesis, University of Sussex). The sulphur to β vinyl distance was found to be 3.15Å, compared to 1.7Å in the natural sulphonium linkage of myeloperoxidase. This implies that significant thermal motion in the structure is required for the haem-protein crosslink to form, accounting for the relatively slow autocatalytic modification process observed in the presence of hydrogen peroxide.

Previous work (Ngo and Smith, Int. Pat. No. WO/2007/020428) has shown that HRP engineered with a more open distal haem pocket (mimicking that of chloroperoxidase or cytochrome P450s) with a weak surrogate base (provided by a Glu residue as in chloroperoxidase) was capable of both entantioselective sulphoxidation and epoxidation. Building on this work, an alternative variant was designed in which the location of the weak base, provided by a Glu or Asp residue, has been varied within the active site. In particular, the HRP variant R38E:F41A:H42A (EAA) catalyses the production of the 1 and 2 naphthol from naphthalene at a rate of 124±4 min^{-1}, suggesting the generation of an epoxide intermediate in the active site. The wild-type enzyme does not catalyse this reaction or the sulphoxidation reactions described for earlier variants. This is believed to be the first report of aromatic C-H bond activation by an engineered plant peroxidase and is unusual in that C-H bond activation of this type normally requires a P450-type thiolate ligated haem system. Equilibrium binding studies show that naphthalene binds to the engineered haem cavity with an estimated K_d of 30±2 μM. Unfortunately, crystals of HRP variants described by Ngo and Smith (2007) and of the new EAA variant described here could not be obtained, despite many crystallisation attempts under a wide range of conditions.

Commonly used abbreviations

Ascorbate peroxidase	APX
Arthromyces ramosus peroxidase	ARP
2,2'-azino-bis (2-ethylbenzthiazoline-6-sulphonic acid)	ABTS
Chloroperoxidase	CPO
Coprinus cinereus peroxidase	CiP
Deoxyribosenucleic acid	DNA
Dithiothreitol	DTT
Ethylenediaminetetraacetic acid	EDTA
Fast protein liquid chromatography	FPLC
Gas chromatography/mass spectroscopy	GC/MS
Horseradish peroxidase isoenzyme C	HRP-C
Isopropyl β-D-1-thiogalactopyranoside	IPTG
3-morpholinopropanesulphonic acid	MOPS
Phanerochaete Chrysosporium lignin peroxidase	LiP
Recombinant horseradish peroxidase isoenzyme C	HRP-C*
Wild-type recombinant horseradish peroxidase C	WT HRP-C*
Yeast cytochrome *c* peroxidase	C*c*P

Chapter 1

Introduction

1.1 Peroxidases, a general introduction

Peroxidase enzymes, whilst increasingly thought of as being almost completely ubiquitous in the natural world (Koua et al, 2009) have, more recently, been put to use throughout a variety of industries, with diverse roles ranging from bioremediation (Klibanov *et al.*, 1983) to use in chemiluminescent assays (Ewetz and Thore, 1976).

The roles of peroxidase molecules in the natural world are being intensively studied and exploited, with recent interest coming from the potential use of peroxidases in the widening field of biotechnology. For example, peroxidases found in wood rot fungi are being optimised for use in the biochemical industry as methods to utilise the natural product lignin, currently a waste product produced when cellulose is purified from plant matter (Biorenew project, EU Sixth Framework Programme FP6-2004-NMP-NI-4). This optimisation work involves using molecular biological techniques, such as site directed mutagenesis and directed evolution, to improve the stability of peroxidases in order to exploit their properties in an industrial context. The work described in Chapters 3 and 4 of this thesis was carried out as a wider part of the Biorenew project, towards new catalysts for clean chemical biotransformation.

Typically, enzymes classified as peroxidases catalyse the oxidation of substrates with the concurrent break down of hydrogen peroxide (Nagano, *et al.*, 1995). There are some variations to this general reaction, in specific enzyme groupings, which will be remarked upon later in this chapter. The sum of the peroxidative reactions can be expressed as shown in Equation 1.1, the net result of the peroxidative catalytic cycle of reactions that will be expanded upon later.

$$H_2O_2 + 2\ RH \rightarrow 2\ H_2O + 2\ R^{\cdot}$$

Equation 1.1 Net chemical reaction of peroxidases

RH is a substrate oxidised by the enzyme to give a radical product, $R^{\cdot}$.

1.2 Classification of peroxidases

Peroxidase enzymes (EC 1.11.1.X, with "X" designating the specific peroxidase, according to the Nomenclature Committee of the IUBMB) can be divided in to haem and non-haem peroxidases, with a number of subdivisions of each classification. Figure 1.1 shows the hierarchy of peroxidases, with groupings as described in the literature. This work deals exclusively with the haem peroxidases, as detailed below. The non-haem peroxidases contain catalytic centres involving ions such as selenium or vanadium (Smith and Veitch, 1998), rather than the protoporphyrin IX haem group, or its derivatives, discussed in this thesis and shown in Figure 1.2a.

The haem peroxidases are most helpfully divided in to three superfamilies within the class, with groupings based on a combination of similarity of sequence and tertiary structure. All haem peroxidases utilise hydrogen peroxide (H_2O_2) in the peroxidative cycle to produce radical products (Passardi *et al.*, 2007). Each haem peroxidase also contains a variation of the ferriprotoporphyrin IX haem moiety, however the precise haem structure differs between the three superfamilies and will be remarked upon where appropriate (Furtmüller *et al.*, 2005). Haem was originally synthesised by Fischer and the system he used in the labelling of the constituent parts of the haem is retained, despite a later attempt to replace his system (Dunford, 1999).

The three haem peroxidase superfamilies, discussed in detail below, are referred to in this work as the mammalian peroxidases, the plant peroxidases and the indistinct peroxidases. The latter describes a group of haem peroxidases that differ from the mammalian and plant peroxidase superfamilies significantly enough that the identification of a further superfamily is warranted, however no clear name for this superfamily has been decided upon. The nomenclature used when describing these peroxidases has been extensively debated, and some particulars of the debate are described below, however the terms mammalian and plant peroxidase superfamilies are retained in this work as alternative superfamily names have not been firmly established in the recent literature.

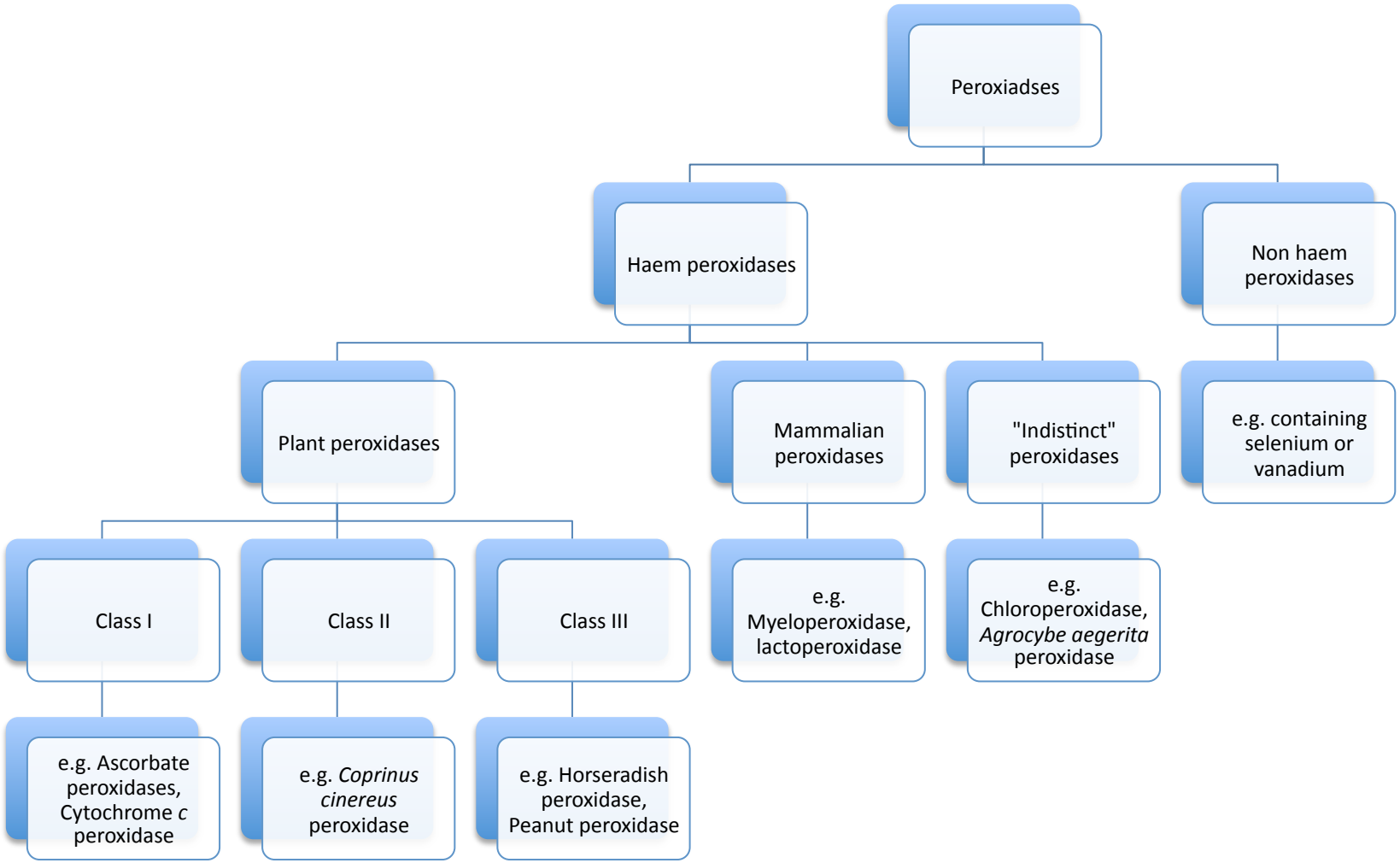
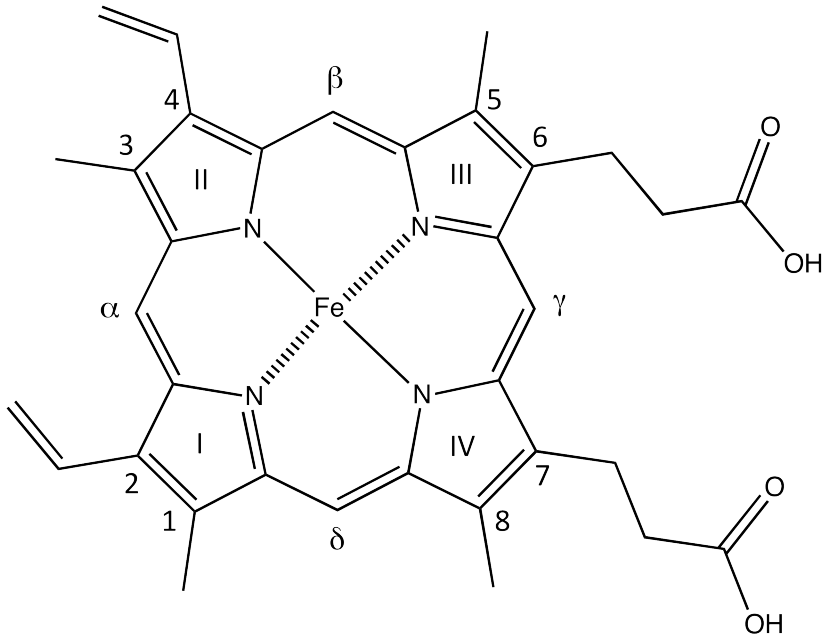
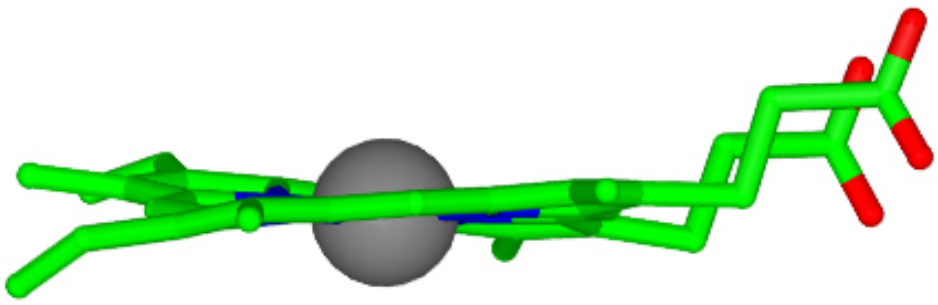


Figure 1.1 Classification of peroxidases
Peroxidases can be grouped in to two large families, with a number of smaller superfamily subdivisions. The plant peroxidase superfamily has been further subdivided in to three subfamilies which were originally determined based on origin, though more recent genome sequence searches have resulted in the discovery of similar sequences in a range of different organisms.

Figure 1.2 Ferriprotoporphyrin IX "haem"
The haem group is the vital centre of peroxidase enzymes. In plant peroxidases this is the near-planar protoporphyrin IX, though this is derivatised in mammalian peroxidases due to covalent linkages between the haem and the protein.



1.2.a Haem labelled with the Fischer system
Four pyrrole rings (I-IV) are interconnected by methene bridges (α-δ). Three types of group are attached to the pyrrole rings – two propionate groups (from carbons labelled 6 and 7), methyl groups (from carbons labelled 1,3,5, and 8) and two vinyl groups (from carbons labelled 2 and 4).



1.2.b Haem viewed from the side
The ferriprotoporphyrin IX haem is approximately planar, with little distortion in the molecule. In plant peroxidases, the haem prosthetic group is attached to the protein through hydrogen bonding of the groups extending from the pyrrole rings, and, additionally, by an effectively covalent link to the iron centre from a proximal histidine residue (Dunford, 1999).

1.2.1 Plant peroxidases

The 'plant peroxidases', some believe, are now a rather inaccurately named superfamily since this class has been found to contain peroxidases of plant and bacterial origin. These new developments in identification of members of this superfamily, brought about by extensive genetic screening, has lead to the recent suggestion that the superfamily should be renamed as the non-animal peroxidases (Koua *et al*, 2009) though this name is not yet widely used.

However it is named, the superfamily can also be subdivided into three classes: Class I, II and III. Whilst the naming of plant peroxidase Classes has previously been based on their origin (Welinder, 1992), this, like the naming of the superfamily, has become less accurate as larger genome sequence searches has revealed sequences of each Class in a wider variety of organisms. The original classification (described in Welinder, 1992) describes the three classes: the 'bacterial peroxidases' (Class I); the 'secretory fungal peroxidases' (Class II); the 'secretory plant peroxidases' (Class III). No new names for the Classes of plant peroxidase have been suggested, rather, they are increasingly referred to simply by their Class number. A broad generalisation suggests that Class I of the plant peroxidases contains bacterial and yeast cytochrome *c* peroxidases, Class II contains peroxidases of fungal origin and Class III are thought of as more typical "plant" peroxidases such as HRP and peanut peroxidase (Welinder, 1992).

Class I APX	Arg38	*Trp41*	His42	Asn71	His163	Asp208
Class I CCP	Arg48	*Trp51*	His52	Asn82	His175	Asp235
Class II CIP	Arg51	Phe54	His55	Asn92	His183	Asp245
Class II LIP	Arg43	Phe46	His47	Asn84	His176	Asp238
Class II MnP	Arg42	Phe45	His46	Asn80	His173	Asp242
Class III HRP-C	Arg38	Phe41	His42	Asn70	His170	Asp247

Table 1.1 Catalytic site residues in members of the plant peroxidase superfamily
Of note, and italicised, are the tryptophan residues found at position 41 or 51 in ascorbate peroxidase and Cytochrome *c* peroxidase respectively. These unique residues will be commented on later in this chapter.

The three Classes of plant peroxidase share less than 20% sequence identity (Smulevich, 1998), however the three Classes share a core "peroxidase" fold. This fold consists of ten α-helices, designated A-J in the literature, with additional helices denoted by their position in relation to the core fold, for example the additional helices F' and F'' in Class III plant peroxidases occur between helices F and G of the core fold. The key feature of the peroxidase fold is that it conserves position, in the active site, of catalytically important residues between the Classes, as displayed in Table 1.1 (Smith and Veitch, 1998) and remarked upon later in this Chapter. Of additional note is the fact that in only one position do these residues differ between plant peroxidase Classes – the phenylalanine found at position 41 in HRP-C is replaced by a tryptophan in Class I peroxidases. These catalytic residues are vital for specific enzymatic activity and a large bulk of work has been dedicated to studying the effect of replacement of the catalytic site residues of a number of different peroxidases, a subject that will be discussed later in this thesis.

Members of the plant peroxidase superfamily have been the subjects of intense study and much of the general understanding of enzymatic mechanisms has come from the study of this enzymatic superfamily (Dunford, 1999). From its initial postulation, the accuracy of the proposed catalytic cycle of plant peroxidases has been further confirmed and continues to be examined in detail, particularly in the light of recent examples of variants where disruption to the catalytic site causes sensitivity and alteration of the cycle.

The plant peroxidase superfamily members have some key identifying spectral features, namely a Soret maximum at approximately 400nm and a further two peaks known as "charge transfer" bands, which, in HRP are found at approximately 500 and 650nm in the visible spectrum, an example of which is shown in Figure 1.3. These characteristic spectral features allows for clear identification of a peroxidase molecule, a characteristic that is highly valuable when preparing the recombinant enzyme from crude cell extract (as described later in Chapter 2 and results reported on in Chapter 3). The exact wavelength of the features, in particular the Soret maximum, is

particular to the specific arrangement of the haem group and the surrounding

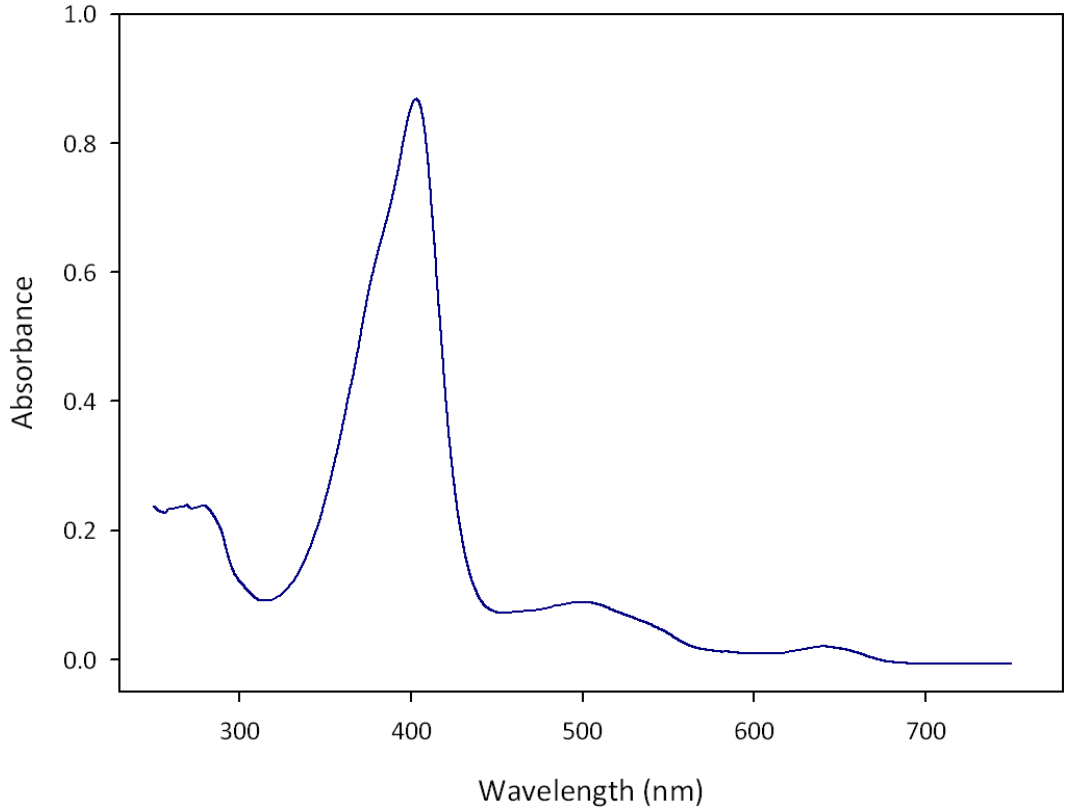


Figure 1.3 UV Visible spectrum of wild type HRP
The UV/Visible spectrum of HRP is characterised by a Soret peak at 403nm, with charge transfer bands found at 500nm and 650nm, indicating the co-ordination state of the haem. A shoulder region seen at approximately 380nm is also characteristic of the wild type.

residues. This has allowed examination of the physical haem environment and broad conclusions to be drawn regarding the nature of the haem without the requirement for high-resolution crystallographic structures.

1.2.1.1 Class I plant peroxidases

Class I peroxidases have been previously suggested to be of prokaryotic origin (Welinder, 1992), however more recently genome sequence searches have revealed Class I peroxidase sequences in plants and fungi in addition to prokaryotes (Zámocký 2004). Class I plant peroxidases contain three distinguishing structural features that separate this class from other plant peroxidase Classes, namely the lack of protein surface glycosylation, the lack of disulphide bonds and finally that structural calcium ions present in other plant peroxidases and shown to be essential in the function of these enzymes, are not seen in Class I plant peroxidases. There is no alternative ion present and the active site pocket has a different structure, remarked upon later in this chapter. Members of the Class I peroxidase family include Cytochrome *c* peroxidases and ascorbate peroxidases as well as the more recently studied catalase-peroxidase, the only member of the plant peroxidase superfamily to possess significant catalase activity in addition to the peroxidase activity (Zámocký, 2004, Smulevich *et al.*, 2006).

1.2.1.2 Class II plant peroxidases

Class II plant peroxidases are still thought to be correct in retaining their original nomenclature as fungal peroxidases. This Class of plant peroxidase is primarily secreted by fungi for the degradation of lignin, a component of plant cell walls (Passardi *et al*, 2007). Four well defined genetic clusters of enzymes have been shown within Class II plant peroxidases, with an additional group of less well clustered gene sequences including the Class II plant peroxidase from *Coprinus cinereus* (Passardi *et al., 2007)*. The four defined gene clusters are: manganese peroxidases, lignin peroxidases, versatile peroxidases (which carry out both catalase and peroxidase functions) and a fourth cluster that have recently been designated as a “new ligninase” group (Passardi *et al.*, 2007).

The Class II peroxidases all contain structural features apparently "missing" from Class I plant peroxidases: the natural wild type forms of Class II peroxidases contain approximately 5% surface glycosylation, structural calcium ions are present in both proximal and distal domains and four conserved disulphide bonds are seen supporting the tertiary structure of the enzyme (Welinder, 1992).

1.2.1.3 Class III plant peroxidases

Class III plant peroxidases have previously been known as the "classical secretory plant peroxidases" (Welinder, 1992) and have retained this designation since extensive examination of genomic sequences has shown that Class III peroxidases are only seen in plant species (Passardi *et al.*, 2007). Class III peroxidases differ from the other two Classes in the superfamily as they have three additional alpha helical regions (Gajhede *et al.*, 1997), though there is significant sequence variation in these regions throughout the class, particularly in the helices designated F' and F'' (Veitch, 2004 a). Two structural calcium ions are seen in Class III, as in Class II, plant peroxidases. It has been shown that for the enzyme horseradish peroxidase, removal of these calcium ions results in a decrease in function of the enzyme of up to 40% (Veitch, 2004 a).

The most widely studied plant peroxidase, and the subject of this thesis, is the Class III plant peroxidase horseradish peroxidase (HRP) (Duroux and Welinder, 2003). HRP has multiple isoenzymes, given alphabetical designation, with numerical classifiers added when further derivations of each isoenzyme was found (Dunford, 1999). Three isoenzymes A have similarly acidic properties, isoenzymes B (three forms) and C (two forms) are neutral and the six isoenzymes E have basic properties, with pI variation between 3 and 12 between the different enzymes (Dunford, 1999). Of the 14 isoforms of HRP, isoenzyme C (HRP-C) is the most abundant and widely studied. Native HRP-C typically contains 18-22% glycosylation (Veitch, 2004 a), with other Class III plant peroxidases approaching 25% surface glycosylation (Veitch, 2004 b).

The enzymatic effect of horseradish peroxidase was originally seen by Planche in the early 1800s (thought to be through the effect of the enzyme on the small amount of 2,5-di-(4-hydroxy-3-methoxyphenyl)-3,4-dimethylfuran contained in his solution)

though the reaction was not attributed to the oxidative power of HRP at the time, rather observed as a feature of addition of the horseradish root to a solution being tested for purity (Veitch, 2004 a). HRP, and many other plant peroxidases, have been studied extensively over the following 200 years, with the complex relationship between protein structure and function the cause of much debate.

1.2.2 Mammalian peroxidase superfamily

The mammalian peroxidase superfamily is a grouping of peroxidases that includes myeloperoxidase and lactoperoxidase and is distinctive from plant and fungal peroxidases, through the presence of an entirely different protein fold. Mammalian peroxidases show no similarity to plant peroxidases at the level of tertiary structure and little similarity at the functional level, with the exception of the catalytic intermediates seen in both superfamilies.

The original naming of this superfamily was established when sequences for its members (myeloperoxidase, lactoperoxidase etc.) were only found in mammals, however since that time there have been "mammalian peroxidase" sequences found in a variety of invertebrates (e.g. Taurog, 1998). This has led to a debate as to the best name for this superfamily. Two recent suggestions for the renaming of the superfamily have resulted from sequence alignments and the discovery of similar sequences in many different organisms.

The recent expansion of the superfamily has been brought about through comprehensive sequence alignment with peroxidase-like domains and, due to the variety of organisms found to contain sequences of high homology, the first suggestion for the renaming is the peroxidase-cyclooxygenase superfamily (Zámocký *et al.*, 2008). This more accurately defines the function of members of the superfamily without defining a specific class of origin. The second suggestion, that this superfamily be renamed the animal peroxidase superfamily, is becoming more common in the literature, though is not yet widely established (Passardi *et al.*, 2007).

For the purpose of this thesis, the previously accepted mammalian peroxidase superfamily nomenclature will be used. Though this may not necessarily accurately reflect the extended superfamily, the peroxidases of interest in this work were part of the early classification system for the superfamily and therefore can sensibly retain this name.

The mammalian peroxidase superfamily has previously been typified by myeloperoxidase, lactoperoxidase and eosinophil peroxidase. These peroxidases, all originally found in mammals, are each involved in different areas of the innate immune response (Carpena *et al.*, 2009).

The superfamily, however it is named, can be further subdivided based on function, though each sequence can be traced back to a common phylogenetic ancestor (Zámocký *et al.*, 2008), an ancestor that is distinct from the plant peroxidase superfamily (Devarajan *et al.*, 2008).

As well as the differences in amino acid sequence between plant and mammalian peroxidases that suggest a different evolutionary route to peroxidative activity, all mammalian peroxidases contain a covalent linkage between the haem moiety and the protein chain (Andersson *et al.*, 1996), a feature that will be discussed at some length later in this work. This linkage is thought to confer stability and defend the enzyme from degradation by the radical reaction products (Huang *et al*, 2006).

A further significant difference between mammalian peroxidases and the other two superfamilies is the ability of mammalian peroxidases to perform halogenation activities, dependent on substrate availability (Furtmüller *et al.*, 2006). This halogenation reaction oxidises halides to hypohalous acids, though myeloperoxidase is uniquely able to oxidise chloride to hypochlorous acid, a feature thought to be due to the presence of a sulphonium linkage between protein and haem (Zederbauer *et al*., 2007). This reaction is not natively seen in plant peroxidases and hence is of interest, particularly when exploring the cause of this reaction.

The key role of mammalian peroxidases is believed to be in the immune response, since their catalytic mechanism allows the production of halides with antimicrobial activity participating in the host's immune reaction to foreign bodies (Zámocký *et al.*, 2008). They have also been thought to contribute to hormone synthesis, particularly thyroid peroxidase, which, as the name suggests, is found in the thyroid gland. The thyroid peroxidase is thought to be responsible for the specific iodination of tyrosine residues for the production of hormones (Zederbauer *et al.*, 2007).

As seen in Class II and III plant peroxidases, mammalian peroxidases also contain calcium ions, however only one calcium is found in each enzyme, with two seen in the dimeric structure of the myeloperoxidase molecule (Zederbauer *et al*, 2007). The structure of the region around this calcium ion is conserved between members of the superfamily and the calcium shown to be vital for structural stability as well as catalytic activity (Furtmüller *et al.*, 2006).

1.2.2.1 Evolutionary history of the superfamily

The mammalian and plant peroxidase superfamilies are believed to have developed similar peroxidative function through convergent evolution (Taurog, 1998) rather than sharing a recent common ancestor (Passardi *et al.*, 2007). They do not share substantial sequence homology and are therefore thought to have developed via evolutionarily distinct routes. Though both superfamilies have the same fundamental intermediate haem species, designated Compound I and Compound II (as mentioned previously), the underlying catalytic activity is not the same.

The theory that the plant and mammalian peroxidase superfamilies arose as a result of independent convergent evolution (Devarajan *et al.*, 2008) is supported by the significant variation in both sequence and tertiary structure between the superfamilies. Further evidence supporting this theory is the formation of significant covalent linkages between the haem and protein in mammalian peroxidases (Zederbauer *et al.*, 2007). A considerably higher degree of sequence homology is exhibited between peroxidases identified as part of the core mammalian peroxidase superfamily than within members of the plant peroxidase superfamily (approximately

20-30%), with myeloperoxidase sharing 70% sequence identity with eosinophil peroxidase (Kooter *et al.*, 1999).

Various studies have suggested that the mammalian peroxidase superfamily originally occurred as a result of gene duplication events, with advantageous mutations arising that have, over evolutionary history, led to the development of distinct members of this superfamily having different functions (Loughran *et al.*, 2008). The different groupings (also known as clades) within the superfamily are suggested to be the myeloperoxidases, eosinophil peroxidases, the lactoperoixdases and the thyroid peroxidases. Of these four groupings, the first three share the same Enzyme Classification number – 1.11.1.7 – with only thyroid peroxidases having the designation 1.11.1.8. Thyroid peroxidase is given this designation as it acts on iodide in the thyroid in the production of iodine coupled tyrosine residues for the production of thyroid hormones, although this unique identification is potentially misleading, since both lactoperoxidase and myeloperoxidase are capable of iodination of tyrosine residues elsewhere in the organism (Ruf and Carrayon, 2005).

1.2.2.2 Haem in mammalian peroxidases

Whilst both plant and mammalian superfamilies contain a haem-based moiety at their catalytic centres, the haem present in the mammalian peroxidase superfamily is a derivation of the standard haem and distinct from the protoporphyrin IX found in members the plant peroxidase superfamily (shown earlier in Figure 1.2). The mammalian peroxidase haem is an iron-porphyrin ring based structure that has two or three covalent linkages between the haem and the protein chain (Zederbauer *et al.*, 2007). These linkages are not found naturally in any plant peroxidase, although work has been done by various groups to engineer covalent linkages in plant peroxidases (for example Raven & co-workers with ascorbate peroxidase in e.g. Metcalfe *et al*, 2004, Piripou *et al.*, 2007). This work will be commented on in more detail in Chapter 5 following work by Cali (2009) to engineer such a linkage in HRP.

Although covalent linkages between the haem moiety and the protein chain are not seen naturally in any plant peroxidase, two distinct types of covalent haem-protein

linkage are seen in the mammalian peroxidase superfamily. These have been shown to form autocatalytically (DePillis *et* al., 1997), although no intermediates have been identified and no detailed mechanism for the creation of these covalent linkages has been proposed. Cytochrome *c* also forms single or multiple covalent thioether linkages between the haem vinyl groups and appropriate cysteine residues in the protein chain (Piripou *et al.*, 2007 a). A single covalent haem-protein link is seen in some members of the cytochrome P450 (CYP) group of enzymes, specifically members of the CYP4A (LeBrun *et al.*, 2002b), CYP4B (Henne *et al.*, 2001) & CYP4F (LeBrun *et al.*, 2002a) subfamilies. A recent view has been proposed, supported through the literature, that location of appropriate residues through proximity to the haem vinyl group allows covalent linkages to form (Piripou *et al.*, 2007 a). In some cases this residue proximity may also rely on the catalytic activity of the enzyme to form, since the cytochrome *c* peroxidase contains a methionine residue at what would appear to be a location predicted to form a covalent haem-protein linkage yet none is seen under physiologically likely conditions (Piripou *et al.*, 2007 b).

In all known mammalian peroxidases, two ester linkages are formed between oxidised methyl groups and suitably located negatively charged residues of the protein chain (aspartate and glutamate in LPO and EPO), a feature that is thought to be conserved throughout mammalian peroxidases (documented in Oxvig *et al.*, 1999, for human eosinophil peroxidase, Andersson *et al.*, 1996, for lactoperoxidase, amongst many others). These are thought to protect the haem from degradation by the reactive products produced in these enzymes as a result of their role in host defence against pathogens (Zederbauer *et al.*, 2007). It has been found more recently that only one of these is essential for catalytic activity in lactoperoxidase (Colas *et al.*, 2002) and to shield the haem from effects of oxygen radical species produced by the catalytic process (Huang *et al.*, 2006).

Myeloperoxidase, however, contains an additional linkage between an unconserved methionine residue at position 243 and the β carbon of the 2-vinyl group of the haem (Fiedler *et al.*, 2000). This sulphonium linkage is thought to be the unique linkage that confers chlorination activity to the myeloperoxidase molecule (Kooter *et al.*, 1999).

Figure 1.4 Haem in myeloperoxidase

The haem in myeloperoxidase, which covalently bonds to the protein chain via two ester linkages and a unique sulphonium linkage, is a derivation of the protoporphyrin IX haem ('haem b') seen in Figure 1.2.

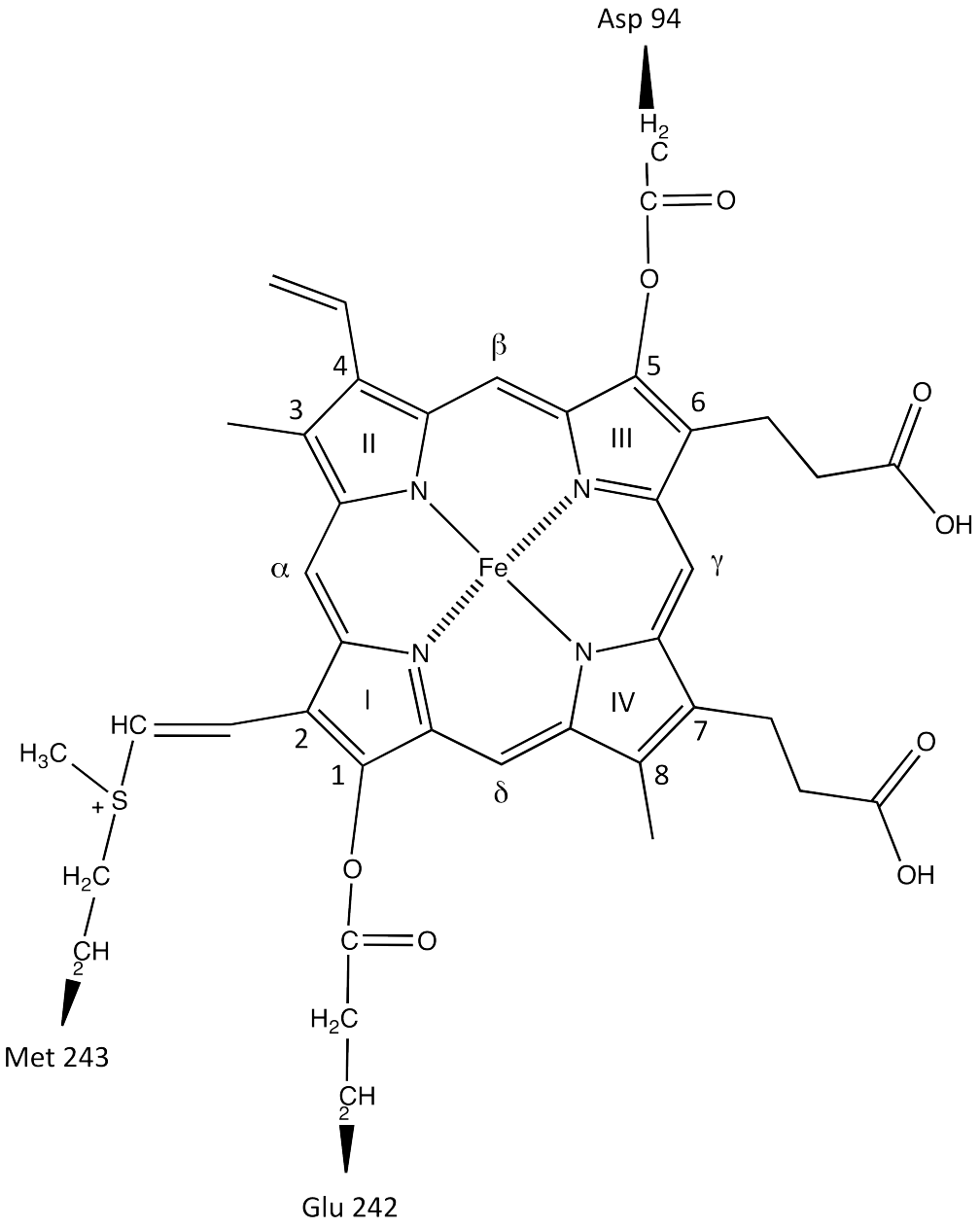


Figure 1.4.a View of the distal face of the haem

Haem-protein linkages are shown, including the unique sulphonium linkage between methionine at position 243 and the haem vinyl group. This model of the linkage was proposed by Fenna, Zeng and Davey (1995).

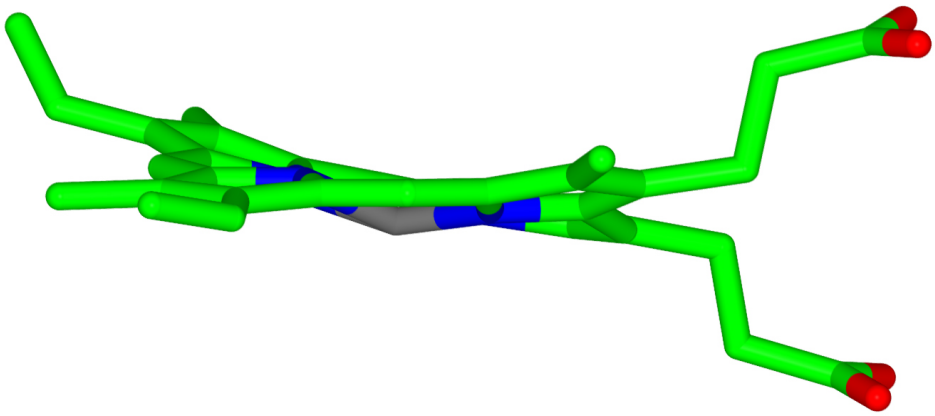


Figure 1.4.b View of the distortion of the haem in MPO

Unlike the planar 'haem b' found in plant peroxidases, the haem seen in myeloperoxidase is severely distorted, with the covalent linkages between haem and protein causing the apparent bending in the porphyrin ring. Haem view taken from MPO structure (Fenna & Zheng, 1995).

Mutating the methionine at position 243 to glutamine and valine resulted in complete loss of chlorination activity, with the threonine substitution of the same residue retaining 15% chlorination activity compared to wild type myeloperoxidase (Kooter *et al.*, 1999). The haem in myeloperoxidase is significantly distorted compared to the planar haem conformation in other peroxidases (Fenna *et al.*, 1994; Devarajan *et al.*, 2008), an example of which is shown in Figure 1.4.a and 1.4.b. The two ester linkages between hydroxylated methyl groups and conserved acidic residues in the protein backbone do not alter the planarity of the haem in other mammalian peroxidases, therefore the addition of the sulphonium linkage appears to considerably alter the haem conformation, which may account for some of the altered activity seen between myeloperoxidases and other mammalian peroxidases (Fenna *et al.*, 1994).

Peroxidase (native)	**Soret maximum (nm)**	**Further maxima in resting state (nm)**	**References**
Myeloperoxidase	430	496, 570, 620, 690	Furtmüller *et al.*, 2006
Recombinant myeloperoxidase variant Met243Thr	413	500, 542, 590, 638	Furtmüller *et al.*, 2006
Lactoperoxidase	413	500, 550, 600, 650	Andersson *et al.*, 1996
Eosinophil peroxidase	413	500, 542, 580, 638	Furtmüller *et al.*, 2006
Horseradish peroxidase C	403	380nm (Soret shoulder), 498, 640	Schokbaum & Lo, 1972
Coprinus cinereus peroxidase	403	505, 534, 649	Smulevich *et al.*, 1994
Cytochrome *c* peroxidase	408	505, 645	Dunford, 1999 (from Yonetani & Anni, 1987)

Table 1.2 Spectroscopic features of mammalian and plant peroxidases
Spectral features of peroxidases, as determined by the nature of the haem centre, allow comparisons between the enzymes. Members of the mammalian peroxidase superfamily, with covalently linked haem groups that are arguably not true haem, have red shifted Soret maxima and complex peaks in the visible region.

1.2.2.3 Characteristics of mammalian peroxidases

The nature of the haem environment in mammalian peroxidases mean that there are great differences seen between the electronic absorption spectra of mammalian and plant peroxidases. Table 1.2 details these spectroscopic differences between myeloperoxidase, lactoperoxidase, eosinophil and plant peroxidases horseradish peroxidase, *Coprinus cinereus* peroxidase and cytochrome *c* peroxidase.

As shown in Table 1.2, myeloperoxidase shows a distinctly different spectrum to other mammalian peroxidases, with a red-shifted Soret maximum, alterations with the other maxima in the visible region of the spectrum. Furtmüller and co-workers produced a series of myeloperoxidase site directed mutants to examine the properties further, showing that it is the unique sulphonium linkage in myeloperoxidase that causes the differences seen. Their variant, which substituted the methionine at position 243 for a threonine and thus preventing the sulphonium linkage from forming, shows an absorption spectrum similar to lactoperoxidase (Furtmüller *et al.*, 2006).

1.2.3 "Indistinct" peroxidases

The final subdivision of the haem peroxidases is a grouping of "indistinct" peroxidases that don't fit well in to either the plant or mammalian peroxidase superfamilies (Smith and Veitch, 1998). This group contains chloroperoxidase, shown in Figure 1.5, a haem peroxidase found in *Caldariomyces fumago* (Shaw and Hager, 1959) that functions in a way more similar to P450s (EC 1.11.1.10) and di-haem cytochrome *c* peroxidases (Smith and Veitch, 1998). The class has also, more recently, been shown to best represent peroxidases such as the peroxygenase from *Agrocybe aegerita* (Hoffrichter and Ullrich, 2006).

This class of peroxidases is biochemically interesting, with chemical functions more similar to those of other haem containing enzymes such as cytochrome P450 monooxygenases (Griffin, B.W. (1991). Chloroperoxidase: a review. In *Peroxidases in Chemistry and Biology.* (Everse, J., Everse, K.E. & Grisham, M.B., eds)), yet utilising peroxidative behaviour to achieve these chemical reactions. Their unique position

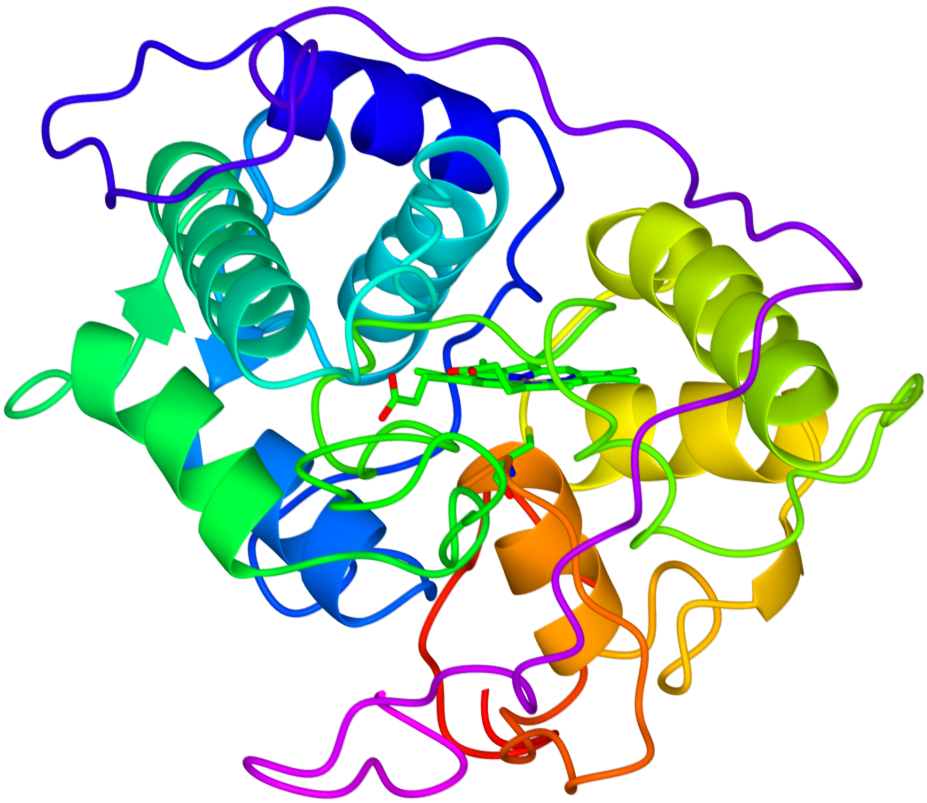


Figure 1.5 Chloroperoxidase molecule

The chloroperoxidase molecule, as with other peroxidase molecules, has a proximal and distal domain, with the haem found at the centre. However the residues found in the area of the haem cavity are distinctly different from other peroxidases. The proximal haem-coordinating residue is a cysteine, seen in P450 enzymes rather than in peroxidases. Also, distal residues, particularly glutamate at position 183, are not seen in plant peroxidases.

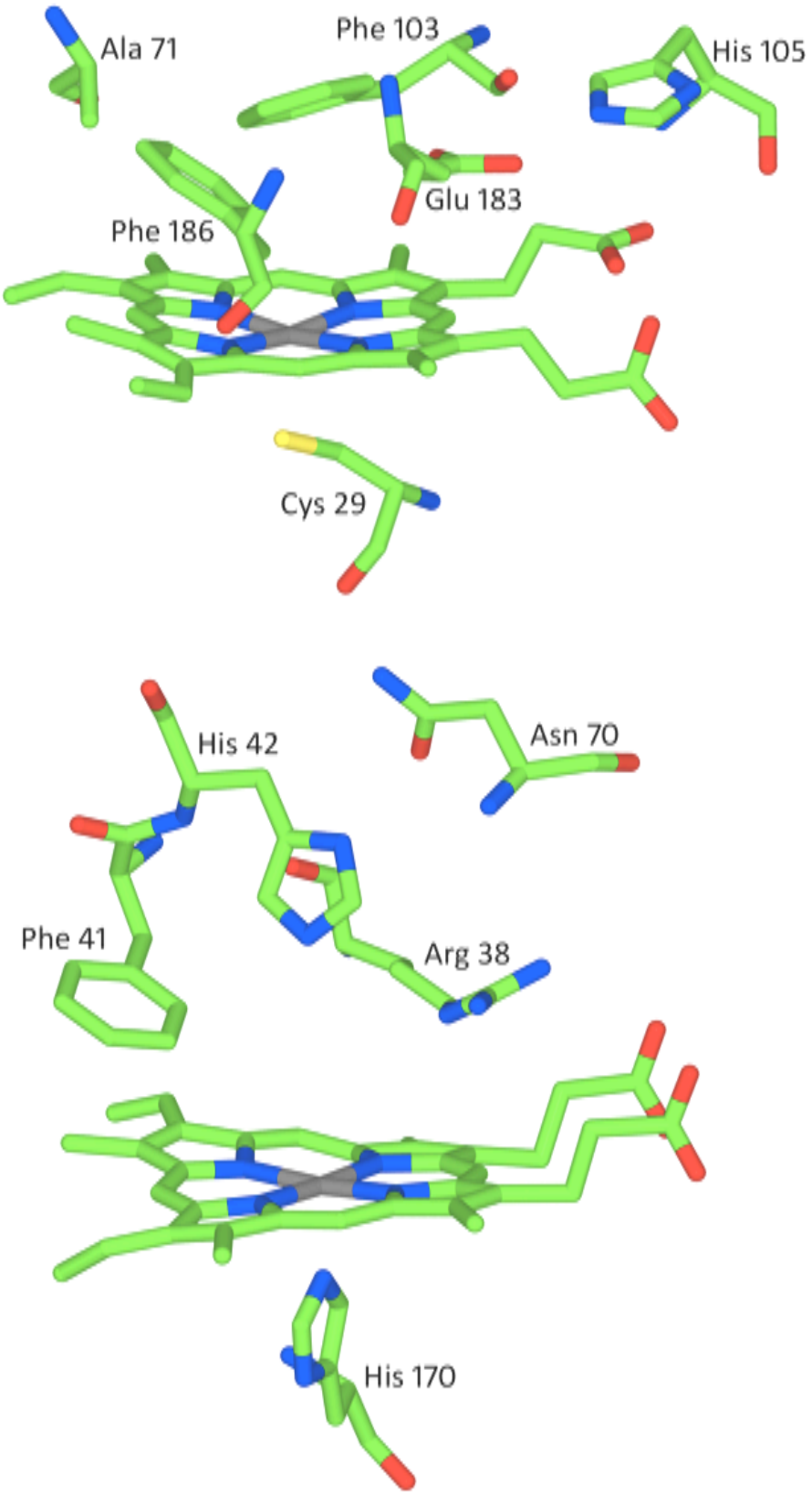


Figure 1.6 Catalytic centres of chloroperoxidase and horseradish peroxidase
The catalytic site of CPO (top) has a more open architecture that allows substrate access to the oxy-ferryl haem directly rather than at the haem edge as seen in HRP (bottom).

between the classes of enzyme suggests that structural characteristics seen in the active site of "indistinct" peroxidases forms the basis for these specific enzymatic activities.

Chloroperoxidase, for example, was recognised to carry out halogenation activities, which are not seen elsewhere in plant peroxidases (Hager *et al.*, 1966). This behaviour has been attributed to the differences seen in the architecture of the active site, which is distinct from other peroxidases. Figure 1.6 shows the catalytic site of chloroperoxidase, with the equivalent residues in a plant peroxidase (HRP) for comparison.

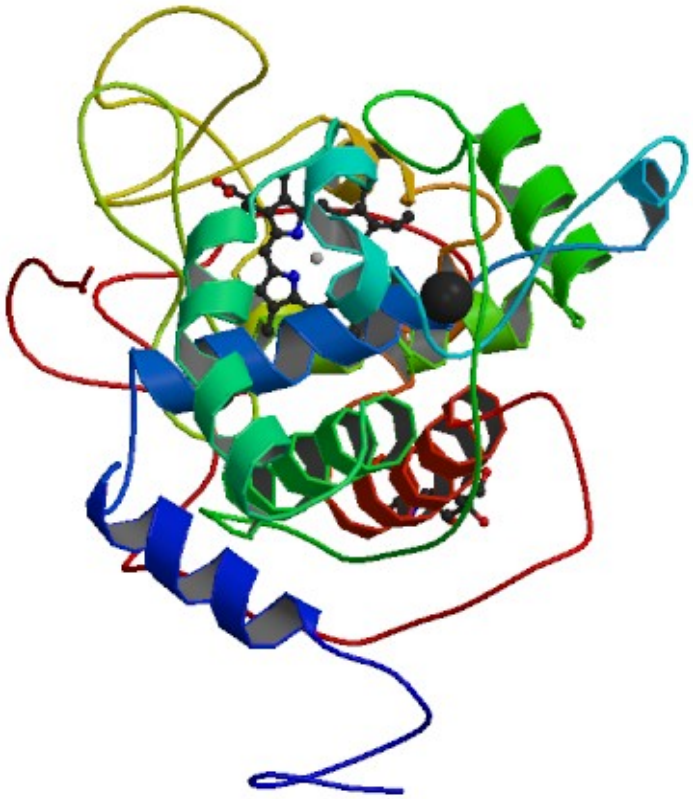
The peroxidase/peroxygenase from *Agrocybe aegerita* has been suggested to fit best within this class of enzyme as it is capable of performing both peroxidase and peroxygenase activities. This enzyme has been described as a chloroperoxidase-cytochrome P450 hybrid (Omura, 2005) and utilises peroxide in the hydroxylation of substrates such as naphthalene or toluene and the halogenation of phenols, particularly having strong bromination activity (Ullrich and Hofrichter, 2005). Spectroscopically, it is distinct from chloroperoxidase, rather sharing the spectrum of the resting state cytochrome P450 (Hofrichter and Ullrich, 2006).

1.2.4 General characteristics common to peroxidases

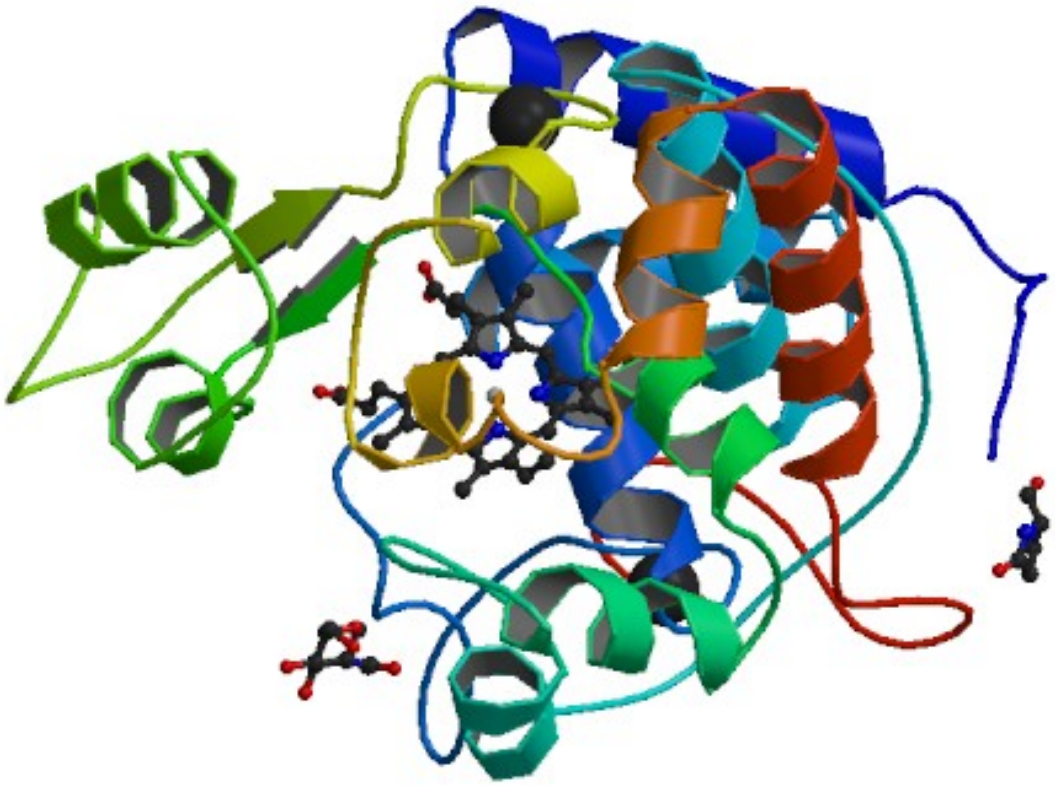
Peroxidases typically have very restricted haem access – reactions are forced, by the structure of the enzyme, to take place at the haem edge (Sundaramoorthy *et al.*, 1995), with direct haem access only produced as a result of structural modification of the enzyme. An exception to this is seen in the enzyme chloroperoxidase, a feature that will be discussed in detail later in this Chapter. The result of the generally restricted access to the haem cavity is that substrate reduction and radical production is facilitated and direct oxygen transfer from the oxy-ferryl haem to the substrate is restricted. In an example of an enzyme overcoming this restriction, the concept of long range electron transfer has been termed for peroxidases, such as versatile

Figure 1.7 Plant peroxidase structures

1.7.a Class I plant peroxidase
Cytochrome C peroxidase structure (Finzel *et al.*, 1984) with PDB access code 2CYP.



1.7.b Class II plant peroxidase
Phanerochaete chrysosporium lignin peroxidase (Piontek *et al.*, 1993) with PDB access code 1LGA.



1.7.c Class III plant peroxidase

Peanut peroxidase (Schuller *et al.,* 1996) with PDB access code 1SCH. The peanut peroxidase structure was solved using a partial molecular replacement sequence from earlier ascorbate peroxidase structure, in combination with an isomorphous derivative data set. This first Class III plant peroxidase structure led to further molecular replacement solutions of other Class III plant peroxidases and, significantly, the first HRP structure by Gajhede and co-workers in 1997.

peroxidase, which are capable of oxidising substrates at the enzyme surface through transfer of electrons from the catalytic haem to reactive residues on the enzyme surface.

1.3 Crystallographic structures of peroxidases

The first crystal structure of a Class III secretory plant peroxidase was of peanut peroxidase and was solved in 1996 by Schuller and co-workers, to a resolution of 2.7Å. The structure of peanut peroxidase is shown in Figure 1.7.c, with PDB access code 1SCH. This was preceded by the publication of structures from Class I and II plant peroxidases, though the structure of the peanut peroxidase was seen as a significant breakthrough in the field. The early Class I peroxidase structures were solved from cytochrome C peroxidase, Figure 1.7.a, (Finzel *et al.*, 1984) and pea cytosolic ascorbate peroxidase (Patterson *et al.*, 1995). Structures of Class II peroxidases lignin peroxidase were solved, initially from *Phanerochaete chrysosporium*, Figure 1.7.b, (Piontek *et al.*, 1993), then later *Arthromyces ramosus* peroxidase (Kunishima *et al.*, 1994), *Coprinus cinereus* peroxidase (Petersen *et al.*, 1994) and manganese peroxidase (Sundaramoorthy *et al.*, 1995).

The structure of the peanut peroxidase showed the classical peroxidase fold seen in Class I and II plant peroxidases, but also the addition in the Class III peroxidase of three helical loop regions. Two of these loops in particular were shown to create the access channel to the haem active site and therefore postulated to play a key role in modulating activity of the enzyme (Veitch, 2004 a).

The solution of the first Class III extracellular plant peroxidase led to the solution, at higher resolution, of horseradish peroxidase using molecular replacement from the peanut peroxidase structure (Gajhede *et al.*, 1997). Subsequent high resolution structures have revealed insights into the active site cleft of HRP-C* that has led to the confirmation of the validity of the catalytic mechanism for HRP-C proposed initially by Poulos and Kraut in 1980. The overall topology of the HRP-C* structure is shown in Figure 1.8.

The first structure of a mammalian peroxidase was of canine myeloperoxidase, solved initially to 3.0Å by Zeng and Fenna (1992), with subsequent solution of other mammalian peroxidases in rapid succession. Despite the advantages of examination of crystallographic structures, further work was required to establish the mechanism by which the mammalian peroxidases covalently attach the haem centre to the protein chains.

As seen in Figure 1.7, the structures of peroxidases in the different superfamilies vary considerably, with the core peroxidase fold incorporated in to tertiary structures that are tailored for each enzyme's catalytic function. Substrates for haem peroxidases also differ greatly between enzymes, as does the way in which substrates access the catalytic centre of the enzyme.

1.4 Horseradish peroxidase

The enzyme horseradish peroxidase (HRP), the central subject of this thesis, has been a well studied enzyme for many years. Its catalytic activity was first observed in the early 19th Century when it was reported that addition of horseradish root to a solution containing guaiacum resulted in a colour change (Veitch, 2004 a). Of course, at this time this reaction was not attributed to the peroxidase contained in the horseradish root. One of the first published uses of the term "peroxidase" appears in 1903 when Bach and Chodat documented their investigations of the chemistry of peroxide in the living cell (Bach and Chodat in Dunford with Everse, Everse and Grisham, (eds) 1991). Since this time, HRP has become central in the understanding of peroxidase chemistry and biochemistry and it continues to be investigated as relationships between peroxidase, oxygenase and globin structure, function and catalytic mechanism are studied in greater depth.

1.4.1 Horseradish peroxidase isoenzymes

There are 14 isoenzymes of HRP (Dunford, 1999), the most common and widely studied being HRP-C. Previous studies had suggested that there were up to 40 isoenzymes seen in preparations of HRP from the horseradish root, though this was

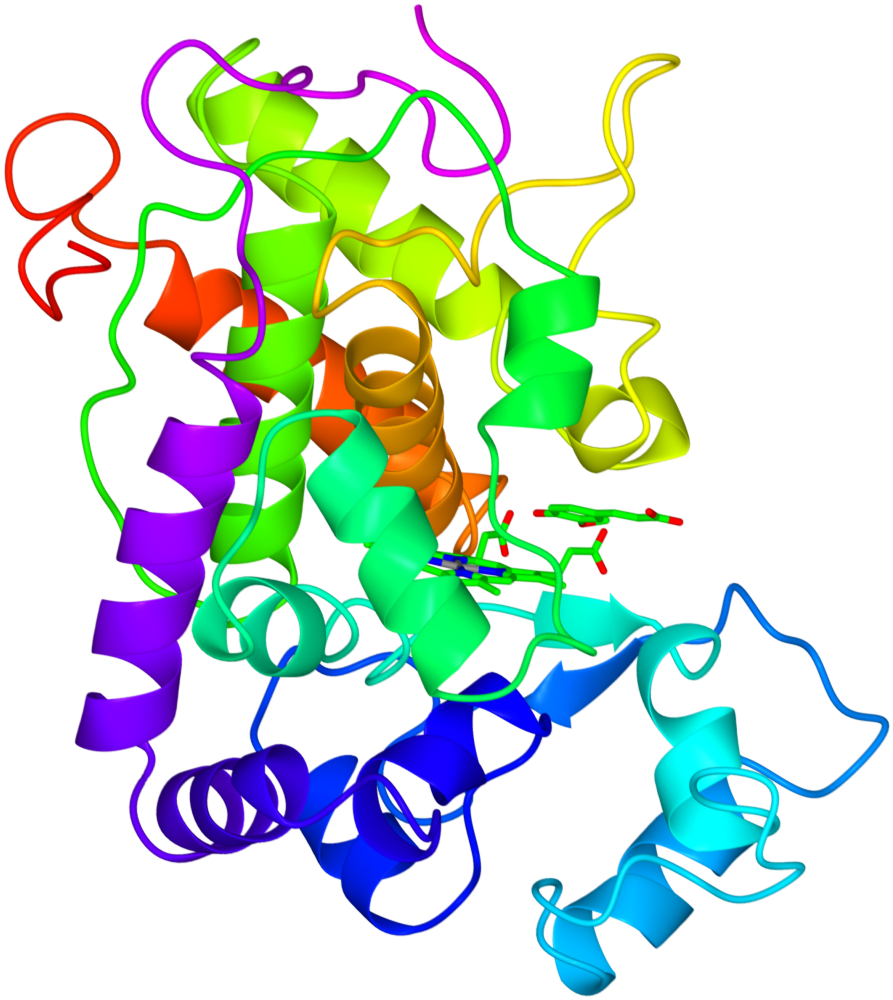


Figure 1.8 Horseradish peroxidase structure
The structure of HRP-C*, shown here from the high resolution structure (PDB code 7ATJ) which was crystallised with ferulic acid in the haem access channel (seen in the above picture in one of three possible identified conformations). Two domains are seen in the HRP molecule, with the haem centre.

later shown to be unlikely, with only minor variations of carbohydrate content distinguishing between different alleged isoenzymes (Dunford, in Everse, Everse and Grisham, 1991).

The amino acid sequence of HRP, the first partial determination of which was for the proximal and distal haem regions of isoenzyme C (Welinder, 1977), is 308 amino acids in length. Minor differences between the isoenzymes occur, as described earlier, and result in the isoenzymes possessing different pIs, with the neutral forms (B & C) readily commercially available (Dunford, 1999). The molecular weight of HRP-C is 33,890 Daltons as calculated from the amino acid sequence, with additional haem, calcium ions and carbohydrate groups taking the final molecular weight to closer to 44,000 Daltons (Welinder, 1979).

1.4.2 Horseradish peroxidase characterisation

In the two decades since the production of the synthetic HRP-C gene (designated HRP-C* in this work) in *E. coli* (Smith *et al.*, 1990), a large body of work has been done to characterise the relationship between structure and function within this enzyme. Site directed mutagenesis of the protein has led to the production of a wide variety of enzyme variants with catalytic activities that diverge from the wild type (e.g. F41L and F41T, Ozaki and Ortiz de Montellano, 1995) , with increased (e.g. Rodriguez-Lopez *et al.*, 1997) or decreased stability (e.g. Colas and Ortiz de Montellano, 2004).

1.5 Horseradish peroxidase catalytic cycle

The catalytic cycle of typical plant peroxidases has now been well established for a number of years, extrapolated across the superfamily from model systems, in particular HRP-C* and the yeast cytochrome *c* peroxidase (Dunford, 1999).

The now accepted model of the peroxidase catalytic cycle was originally proposed in 1980 as a three step reaction mechanism (Poulos and Kraut, 1980) after study of the tertiary structure of cytochrome *c* peroxidase, various active site mutants and a review of previous spectroscopic work. This cycle is shown in Figure 1.9.

The most widely used current model for the study of the plant peroxidase catalytic cycle has been HRP, particularly after cytochrome *c* peroxidase was shown to have an atypical peroxidase catalytic site in which the cation radical is stored on an adjacent tryptophan residue rather than in the porphyrin ring as seen in other plant peroxidases (Veitch, 2004 b). As cytochrome *c* peroxidase is a Class I peroxidase, it does not contain the structural calcium ions seen in other plant peroxidases. Also, the publication of the peanut peroxidase structure by Schuller and co-workers in 1996 led to the revelation that the Class III plant peroxidase haem cavity was more enclosed than previous study and modelling had suggested (Howes *et al.*, 1997).

HRP-C* reaction intermediates have been trapped and well characterised using a range of techniques including the UV/Visible conventional and stopped-flow spectroscopy used in this work. More recently, an elegant use of the dynamic properties of a protein crystal in the X-ray beam has been used to determine the structures of the HRP-C* catalytic cycle intermediates (Berglund *et al.*, 2002) which will be discussed in detail later in this chapter. Mutational studies of key residues in the catalytic site have shown that a conserved hydrogen-bonding network is present with disruptions to this network having a profound effect on catalytic activity.

Although the model system has changed from the original interpretation in cytochrome *c* peroxidase, the catalytic cycle of typical peroxidases can still be broken down into three key steps, shown in Scheme 1.1, in line with the general mechanism for peroxidase catalysis proposed by Poulos and Kraut. Since then, these intermediates have been characterised in a wide range of peroxidases and the general reaction mechanism proposed in 1980 shown to be correct.

The first step (designated i in Scheme 1.1) is the formation of the two-electron intermediate, Compound I, from the resting enzyme, achieved by the oxidation of the enzyme by hydrogen peroxide (H_2O_2). Return back to the resting enzyme is accomplished by two one-electron reductions, first to Compound II (ii) and then from Compound II back to the resting state (iii), both involving small reducing substrates.

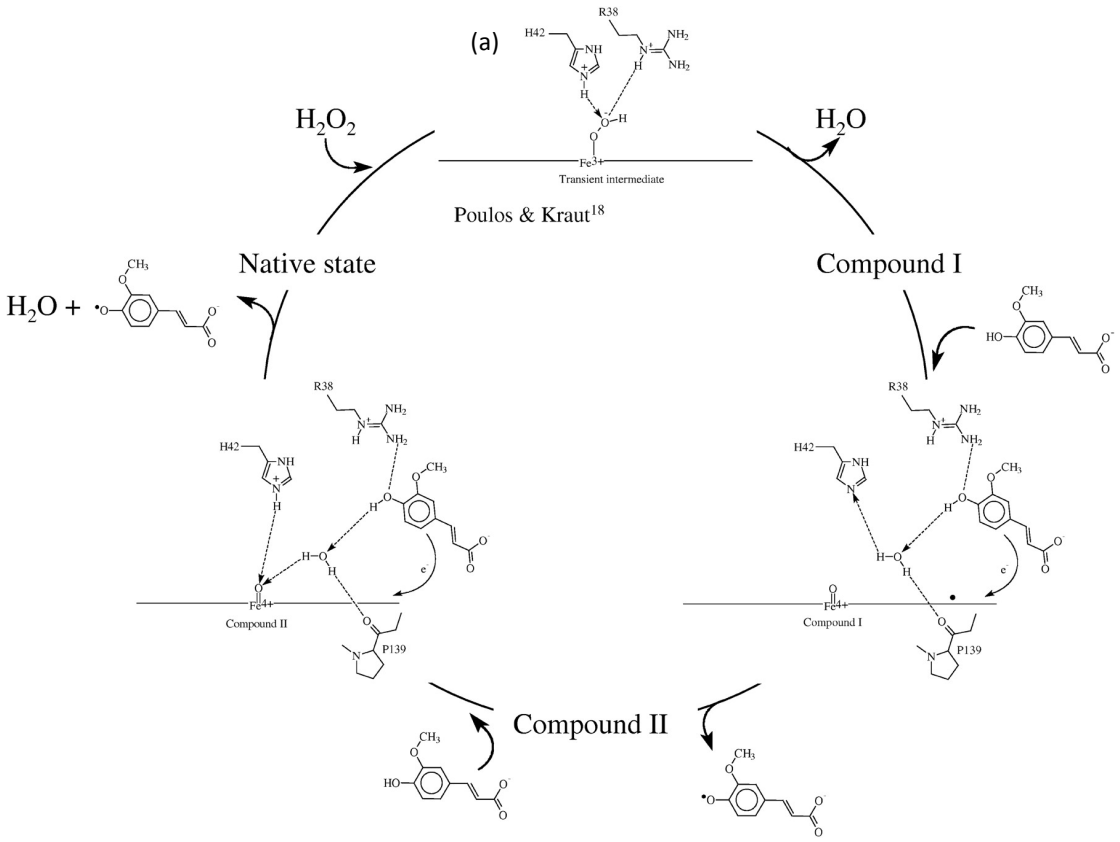


Figure 1.9 Peroxidase catalytic cycle
From Henriksen *et al.*, (1999). Crystal structures of horseradish peroxidise intermediates showed, at high resolution, the stereochemistry of peroxidase catalysis, confirming the previous proposed catalytic cycle (e.g. Poulos and Kraut, 1980).

A proposed transient intermediate is marked (a), was initially proposed in the formation of Compound I. This was later detected in low-temperature experiments and characterised as a Fe(III)-hydroperoxy complex (Veitch, 2004a) and designated as Compound 0.

$$\text{Haem } (Fe^{III+}) + H_2O_2 \rightarrow \text{Haem } (Fe^{IV}=O, \text{ radical charge}) + H_2O$$

i) Resting State → Compound I

$$\text{Haem } (Fe^{IV}=O, \text{ radical charge}) + RH \rightarrow \text{Haem } (Fe^{IV}=O) + H^+ + R^\cdot$$

ii) Compound I → Compound II

$$\text{Haem } (Fe^{IV}=O) + H^+ + RH \rightarrow \text{Haem } (Fe^{III+}) + H_2O + R^\cdot$$

iii) Compound II → Resting State

Scheme 1.1 HRP-C* key catalytic intermediates, from Henriksen, Smith and Gajhede, 1999.
'Haem' designates the protoporphyrin XI haem catalytic centre of the enzyme, showing the oxidation state of the iron centre, RH and $R^\cdot$ represent the small peroxidase substrates and radical substrate reaction products.

These substrates are varied in nature, dependant on the peroxidase in question, but in Class III plant peroxidases such as HRP and *Coprinus cinereus* peroxidase, these tend to be small aromatic amines, indoles, phenols and sulphonates, with equivalent radical products produced (Smith and Veitch, 1998).

1.5.1 Catalytic cycle intermediates

Initial identification of the catalytic intermediates has been possible through UV/Visible spectrophotometry, since the different coordination states of the haem in wild type HRP-C show distinct spectra that can be easily recognised. As the reaction progresses to Compound I with the addition of hydrogen peroxide and the rate determining step of the catalytic cycle is the reduction of Compound II to the resting state, each intermediate can be readily tracked and documented (Poulos, in Bertini *et al.*, 2007). Changes in the UV/Visible spectra of HRP-C* variants are able to give insights into haem functionality without detailed structural information being available.

The peaks seen in the UV/Visible spectra of peroxidases correspond to the electronic state of the haem, with the distinct Soret maxima at approximately 400nm resulting from a π -> π^* electronic transition from the haem iron (Smulevich, 1998). Further weaker peaks in the spectrum between 500nm and 700nm are indicative of the

binding state of the haem, with charge transfer bands seen in the resting state enzyme as a result of charge distribution between the porphyrin ring and the haem iron atom. These are not seen in Compounds I or II, due to the bound oxygen on the iron and, for Compound I, radical charge distribution across the porphyrin ring (Smulevich, 1998; Dunford, 1999).

1.5.1.1 Resting state enzyme

The start and finish point in the catalytic cycle is the Fe(III) resting enzyme. This enzyme has a characteristic UV/Visible absorption spectrum, shown earlier in Figure 1.3. The spectrum shows a clear Soret maximum at 402nm, with charge transfer bands at 530nm and 650nm. These readily defined bands indicate the nature of the haem centre, with the iron centre of the haem moiety in the 3^+ oxidation state. This will be remarked upon later, in Chapters 3-6, since the HRP-C* variants created over the course of this work have considerable variation in the UV/Visible spectrum of the enzyme, which may go some way to explaining the activity of the enzymes.

1.5.1.2 Compound I

Upon binding of hydrogen peroxide, even in the absence of any substrate, the first intermediate of the peroxidase catalytic cycle is formed. A transient intermediate had been hypothesised to be formed prior to Compound I formation, indicated a) in Figure 1.9. This intermediate was later detected upon addition of hydrogen peroxide to wild type HRP-C* at low temperatures. This intermediate, termed Compound 0, was then characterised as an Fe(III)-hydroperoxy complex (Veitch, 2004 a), following observation of this complex in Rodiguez-Lopez *et al.*, (1996) in the HRP-C* variant R38L. This complex shows hydrogen peroxide bound to the haem centre before the dioxygen bond is cleaved forming Compound I.

Compound I, as the intermediate at the first step of the catalytic cycle is known in all peroxidase species, has been extensively characterised and shown to consist of an oxyferryl haem with the π cation radical disseminated through the haem ring and the concomitant production of a molecule of water. A notable exception to this porphyrin

π cation radical is found in lignin peroxidase, where the radical is stored on a redox active tryptophan, residue 171, a unique ability to interact with substrates at a second site away from the catalytic centre (Blodig *et al.*, 1999).

The UV/Visible spectrum of Compound I, seen in Figure 1.10, shows the result of the addition of hydrogen peroxide on the spectrum of wild type HRP-C*. The Soret peak at 403nm has greatly decreased, typically to half the height of the peak of the resting state, and the charge transfer bands appear to level out with a much broader peak from 550 to 700nm.

1.5.1.3 Compound II

Compound II is seen upon addition of one equivalent of a reducing substrate to a solution of Compound I. This is the first of two one-electron reducing steps that complete the catalytic cycle. The formation of Compound II results in the reduction of the π cation radical seen in Compound I, whilst the oxy-ferryl haem remains.

This catalytic intermediate is characterised in the UV/Visible spectrum by the red-shifting of the Soret peak and two peaks close together in the visible region of the spectrum, as shown in Figure 1.10.

1.5.1.3 Compound III

If excess hydrogen peroxide is added to the enzyme in the absence of substrate, a new compound is seen in the UV/Visible spectrum. This compound, identified as Compound III, has dioxygen bound to the haem, with the iron either in the Fe^{2+} state or in the isoelectronic form with charge distributed across the bound oxygen as Fe^{3+}-O_2^- (Berglund *et al.*, 2002) and is not part of the traditional peroxidase catalytic cycle.

Further addition of hydrogen peroxide to the enzyme in the absence of substrate can result in a haem 'bleaching' effect, whereupon superoxide products react with the haem, oxidising pyrrole rings. This results in the release of the iron centre from the haem and corresponding loss of function in the enzyme. This effect has been studied

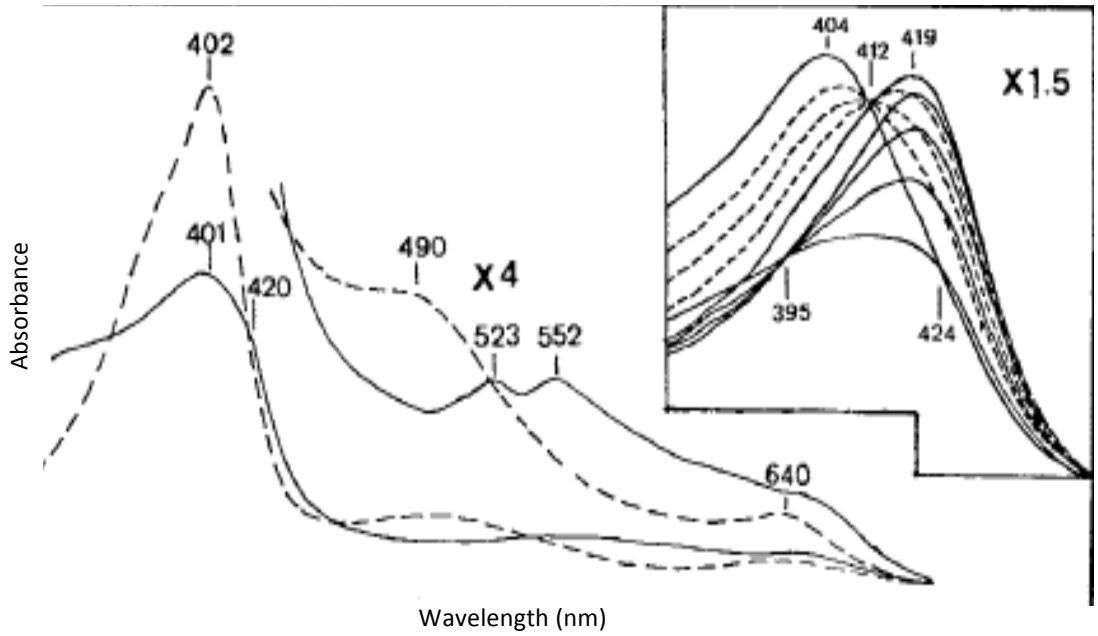


Figure 1.10 –UV visible spectra of catalytic cycle intermediates of HRP (from Smith *et al.,* 1992)
The dashed line in the main figure is representative of a wild type HRP-C* in resting state, with the solid line showing the first intermediate, Compound I. In the 1.5x enlargement of the Soret region, the peak at 404nm shows the resting state and further traces show the changes in absorbance as the enzyme moves between resting state and the steady state intermediate Compound II.

Absorbance intensity at the Soret peak drops by approximately 50% in the formation of Compound I, whilst charge transfer bands level out as the dioxygen bond cleavage distributes a radical charge across the porphyrin ring. The steady state intermediate, Compound II, sees the Soret peak return to near-native intensity, however this peak is shifted to approximately 420nm with oxygen still bound to the iron centre of the haem.

in a wide variety of haem-containing enzymes (e.g. Schaefer *et al.*, 1995, Nagababu and Rifkind, 2000). It has also been noted, for example by Colas and Ortiz de Montellano (2004), that peroxidase mutants can be more sensitive to haem loss through bleaching, attributed to the alterations and destabilisation of the catalytic site.

1.5.2 Catalytic site residues

As seen earlier, in Table 1.1, residues in the catalytic site of plant peroxidases are highly conserved, even between the three distinct Classes. This is echoed in the crystal structure of the HRP-C* active site shown in Figure 1.11, the residues surrounding the haem in peroxidases play a vital role in the stability of the resting enzyme (Henriksen *et al.*, 1998). In HRP-C, the key distal catalytic site residues are arginine (residue 38), phenylalanine (residue 41), histidine (residue 42) and asparagine (residue 170). These, alongside the residues at the entrance to the catalytic site (phenylalanine at position 179, alanine at position 140, and some others) create a small hydrophobic pocket in which HRP-C substrates bind and are oxidised by the reactive haem centre.

On the proximal side of the haem, a histidine (residue 170) provides the 5^{th} co-ordination position to the iron centre of the haem. Histidine 170 is stabilised by hydrogen bonding from a nearby aspartate (residue 247).

These residues are highlighted in Figure 1.11, with a stereo view shown in Figure 1.12, which shows the HRP-C* catalytic site from the high resolution structure solved to 1.45Å by Henriksen and co-workers (1999) (PDB file 7ATJ). The hydrogen bonding network that is crucial in the catalytic activity of the enzyme is also included, extending on the distal side to the calcium ion that is required for enzymatic function.

These structurally conserved residues are clearly vital for peroxidative activity, with substitutions of these residues corresponding to decrease in catalytic activity relative to the severity of the substitution (Veitch and Smith, 2001).

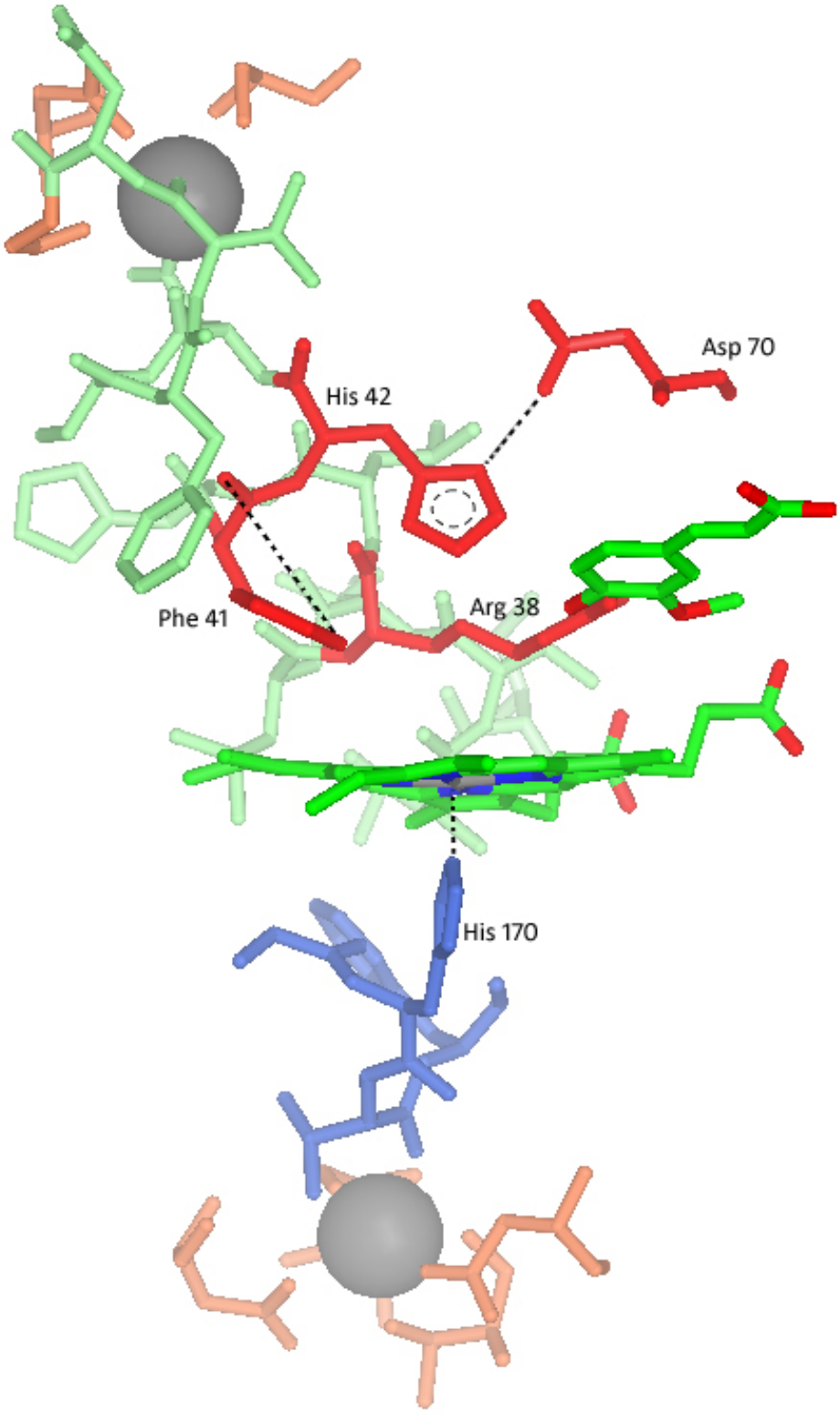


Figure 1.11 Horseradish peroxidase catalytic site
Key catalytic site residues relevant to this work are highlighted, with hydrogen bonding between these residues depicted. These residues are vital for the binding and catalysis of HRP substrates via the haem edge, with Arg38 and His42 particularly involved in dioxygen bond cleavage.
Two calcium atoms are incorporated into the protein structure, each with seven ligands from the protein backbone.

1.5.3 Horseradish peroxidase substrates

As described previously, substrates are only able to access the reactive haem of Class III plant peroxidases at the haem edge. This greatly limits the accessibility of the active centre of the enzyme and restricts the range of catalytic activities to the oxidation of small aromatic substrates using one-electron oxidation rather than the direct transfer of oxygen to the substrate. The residues around the access channel to the haem cavity sterically hinder the substrates and prevent access.

One such substrate, which is known to act as a peroxidase substrate *in vivo*, is (3-(4-hydroxy-3-methoxyphenyl)-2-propenoic acid (better known as ferulic acid), a natural component of plant cell walls (Henriksen *et al.*, 1999, after Fry, 1986). The cross-linking of ferulic acid in the formation of cell walls has been reported to be catalysed by peroxidase enzymes, with the enzyme catalysing the oxidative coupling of the ferulic acid molecules via diferulic acid (Markwalder and Neukom, 1976).

Ferulic acid has been notable in the study of horseradish peroxidase, particularly as HRP-C* has been co-crystallised with ferulic acid in order to generate high quality crystals for X-ray diffraction and subsequent structure solution. The HRP-C*-FA crystals, however, presented a different crystal form to the previous structures (Gajhede *et al.,* 1997, Henriksen *et al.,* 1998). The crystal structures show that the ferulic acid molecule sits in an access cavity towards the haem edge, with additional solvent molecules stabilising the complex. It was not possible to unambiguously model the orientation of the molecule in the site. These structures have lead to the proposal of a model for phenolic substrate oxidation by peroxidases (Henriksen *et al.,* 1999).

Ferulic acid has been regularly used as a cocrystallisation additive in subsequent work, acting to stabilise the crystalline complex.

A further intensively studied HRP substrate is the chemical luminol. The oxidation of luminol is a reaction catalysed by HRP-C and is often utilised in medical diagnostic immunoassays and biosensory applications (Marquette and Blum, 2006), as the

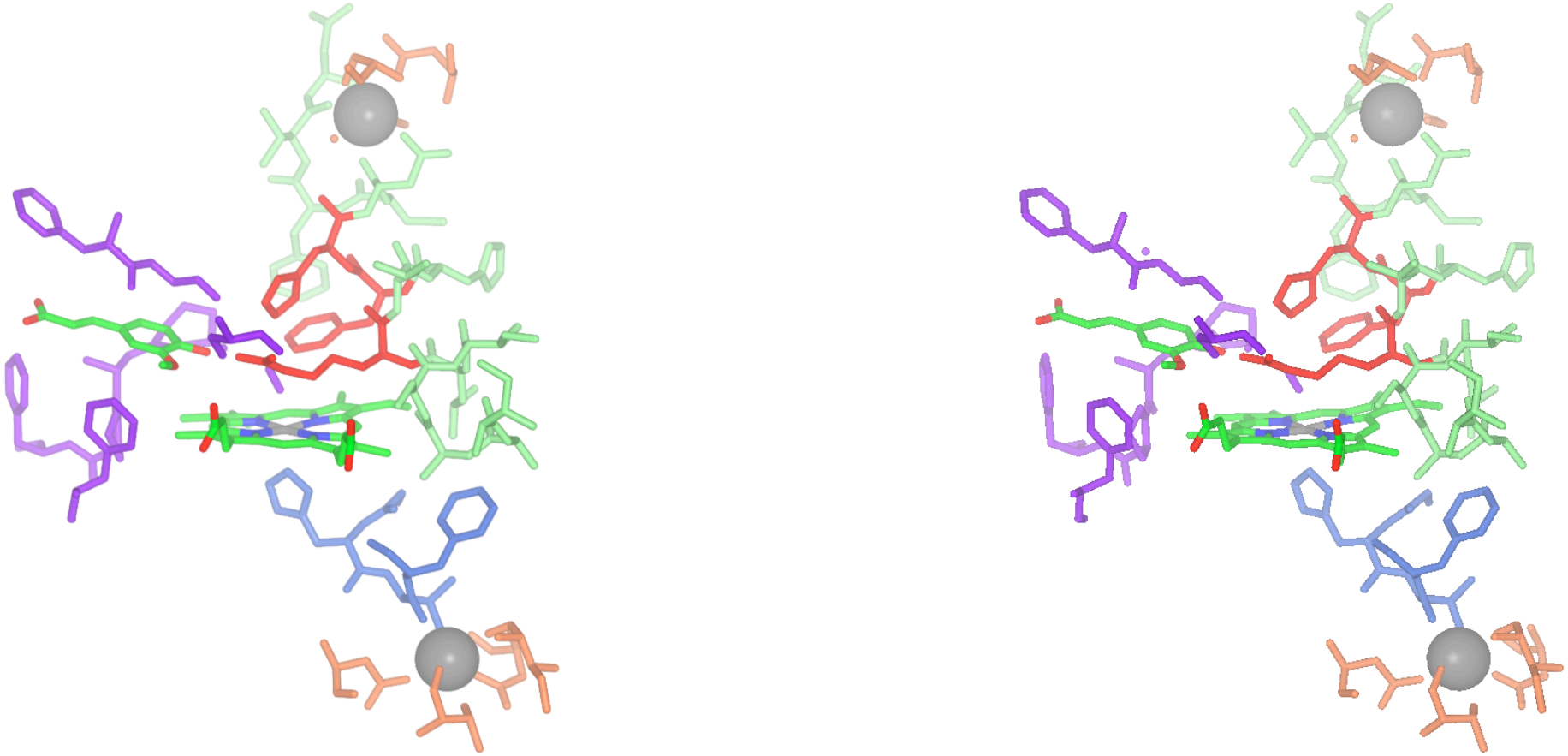


Figure 1.12 Horseradish peroxidase catalytic site (stereo view)
The catalytic activity and thermal stability of HRP is compromised if either of the calcium molecules is missing from the enzyme, with the protein unable to fold correctly without calcium present (Veitch 2004 a).

degradation of the oxidised radical product produces a chemiluminescent signal. The reaction mechanism of luminol oxidation and degradation is shown, in general terms, in Figure 1.13 (Marquette and Blum, 2006).

A more detailed reaction mechanism is only generally hypothesised, with the luminol molecule reacting with both Compound I and II to give the radical product shown in Figure 1.13, which then decays emitting light (Dunford, 1999).

HRP-C is utilised in these assays since the alternatives, using metal cations and complexes, require pH increases, whereas biological catalysts such as HRP-C can be used at near-neutral pH (8-8.5) (Marquette and Blum, 2006). A disadvantage to the use of HRP-C in these reactions is that typically the enzyme requires the addition of enhancer molecules to allow the reaction to progress at a rate where signal can be readily detected (Whitehead *et al.*, 1983), which can lead to misrepresentation of results in the immunoassays to which it is applied.

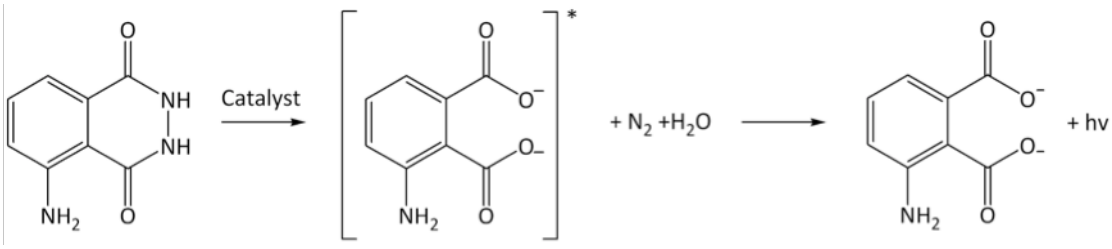


Figure 1.13. The chemiluminescent reaction of luminol, from Marquette and Blum, 2006

The use of such substrates in the study of peroxidases has allowed focus when performing site directed mutagenesis in and around the catalytic centres of the enzyme. In particular, the changes in rate constants in respect to particular substrates can be dramatically altered by apparently subtle changes in the active site and entrance cavity of HRP-C*. Three varieties of variant and their corresponding changes in substrate reactivity are discussed later in this work.

1.6 Non-peroxidase haem enzymes

1.6.1 Cytochromes P450

Cytochromes P450 are haem centred members of the external monooxygenase class of enzymes and are involved in catalysing the direct transfer of a single atom from molecular oxygen to a substrate, producing water as a by product (Bernhardt, 2006), a process which is considered chemically difficult (Ngo and Smith, 2007). Cytochromes P450 are classed in the external monooxygenase family since they make use of an external reductant rather than reducing molecular oxygen to water through the extraction of reducing equivalents from a substrate molecule (internal monooxygenase activity) (Bernhardt, 2006).

The full range of the reactions known to be catalysed by the cytochrome P450 class of enzymes is large but includes reactions such as epoxidation of arenes and alkenes, oxidation of sulphur and nitrogen groups, reduction of epoxides and the oxidative cleavage of carbon-carbon bonds (Sono *et al.*, 1996).

This range of reactions catalysed by members of the cytochrome P450 enzyme family makes cytochrome P450s attractive targets for the chemical and biopharmaceutical industries, with a large amount of work devoted to further understanding the complex relationship between structure and function of these enzymes (Henne *et al.*, 2001).

Although these enzymes, like peroxidases, are considered ubiquitous throughout the animal and plant kingdoms (Henne *et al.*, 2001)., cytochrome P450 enzymes are not considered naturally robust – they are sensitive to changes in conditions that cause inactivation of the enzyme and loss of catalytic activity (Ngo and Smith, 2007). In addition these enzymes require an ancillary reductase, such as NADPH, to function, greatly increasing the cost of using these enzymes as biochemical catalysts in industry.

Work has been done more recently to try to improve the cytochrome P450 for potential industrial uses, using a combination of techniques including site directed mutagenesis and directed evolution (Bernhardt, 2006). The areas targeted for

improvement in this body of work has included attempts to improve the stability of the enzyme, to increase stereo and regio selectivity of hydroxylation and to eliminate the need for the cofactor NADPH, a costly barrier to industrial usage of this group of enzymes (Bernhardt, 2006).

1.7 Engineering oxygenase activity in a peroxidase

The work described in Chapters 3 and 4 of this thesis is a direct descendant of work carried out by Ngo and Smith, that was patented in 2007 (WO/2007/020428, Novel peroxidases and uses). The patent describes the need for a robust peroxygenase to catalyse the selective transfer of oxygen to substrates as seen in the cytochrome P450 class of enzymes described above. This catalyst would be necessarily stable in order to produce yields needed for effective production of pharmaceutical products by this method. The priority in creation of a new oxygen-transfer catalyst is that it is capable of withstanding continuous turnover conditions, potentially with high levels of peroxide that could potentially inactivate a less robust enzyme.

The work described in the patent details how the horseradish peroxidase catalytic centre can be manipulated through site directed mutagenesis to perform peroxygenase activity whilst maintaining the HRP-C substrate hydrogen peroxide, rather than molecular oxygen and reductases to function. This highlights the substitution of particular residues in the active site of HRP-C which, when substituted by residues that reduce the steric hindrance within the active site of the enzyme, allow substrates direct access to the haem centre rather than the restricted haem-edge activity seen in typical plant peroxidases.

Stereoselective oxidation of compounds is particularly important, with future uses of the enzymes in production of pharmaceutical products where consistent yields of one product would greatly reduce costs in downstream processing. The potential for production of styrene epoxides and chiral sulphoxides with high levels of specificity may therefore allow later production of complex chiral ligands. The range of potential substrates that it may be possible to oxidise has not yet been fully explored.

Figure 1.14 shows the catalytic steps in the peroxidase reaction, alongside the cytochrome P450-like reaction of the chloroperoxidase molecule. Following the apparently identical Compound I state, where oxygen from hydrogen peroxide is bound to the haem, the reaction pathways diverge, with chloroperoxidase capable of direct oxidation of substrates. Engineering of the HRP-C* active site has allowed manipulation of the peroxidase pathway, with variant enzymes able to follow the peroxygenase route whilst returning to the resting state upon cycle completion.

This work, including previous results, is discussed more fully in Chapters 3 and 4.

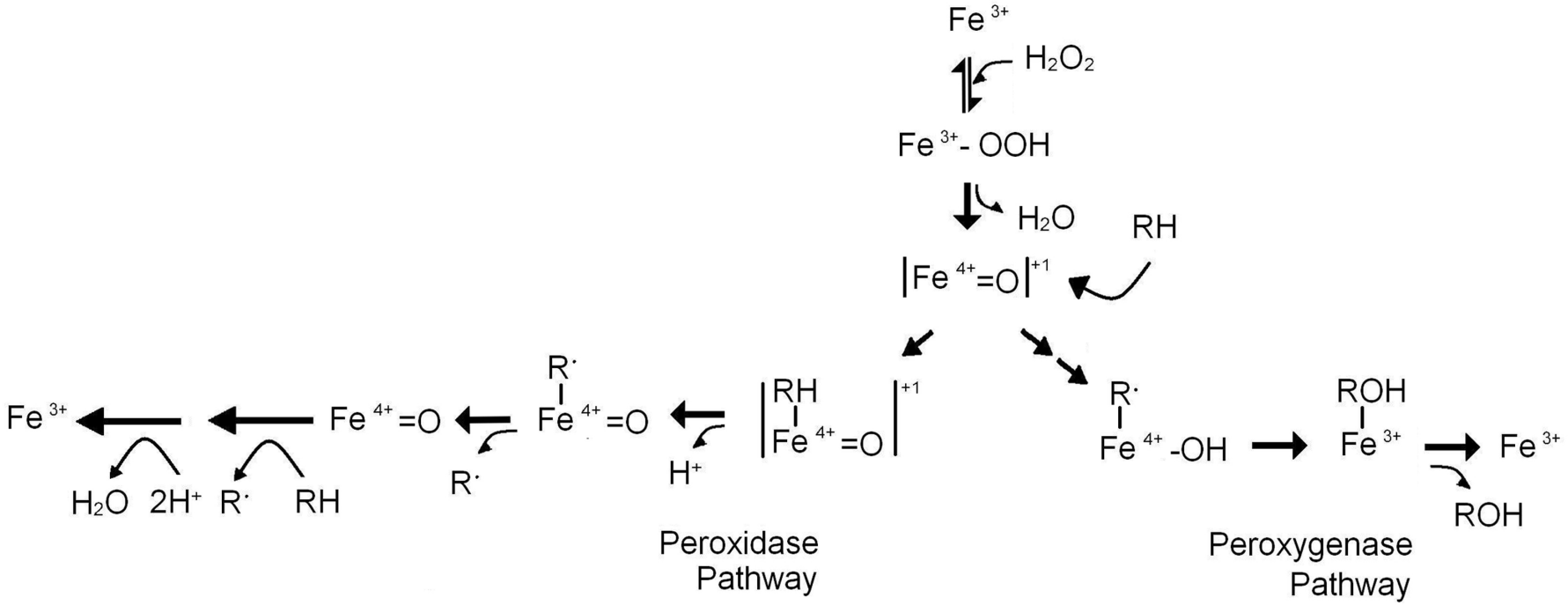


Figure 1.14 Cataltic pathway of peroxidases and peroxygenases

Following Compound I formation, the peroxidase pathway returns the enzyme to resting state via two one-electron reductions, a pathway originally described in 1980 by Poulos and Kraut. Although their original description of the pathway has been modified slightly with subsequent study, the underlying three-step reaction was shown to be correct. More recently, study of peroxygenases has shown a comparable intermediate in Compound I. This common intermediate has suggested that it is the structure of the catalytic site of the relevant enzymes that directs the course of reaction and has led to this work.

Structural techniques

1.8 X-ray crystallography

1.8.1 Historical perspective

Since the publication of the first protein structure solved using X-ray crystallography by John Kendrew and his team in the late 1950s (Kendrew *et al.*, 1958), the field of structural biology has increased almost exponentially. However, regardless of the variety of techniques available to study protein structure, and despite its limitations, X-ray crystallography remains solidly at the heart of the field.

In the decades that have followed the first X-ray crystal structure solution, the technique has developed such that protein crystal structures may be solved in a matter of hours or days, rather than the months and years taken for the first structures.

A number of different techniques have been developed to probe the structures of proteins, particularly important have been the techniques developed to study molecules that are unsuitable for crystallisation and so render X-ray crystallographic analysis of the structure impossible. These techniques have been used to enhance knowledge of enzymatic mechanisms, to suggest enzymatic targets for drug molecules and to explore the properties of proteins in a way that cannot be carried out by other methods. Whilst all techniques for structural analysis of proteins have their advantages and disadvantages, the use of these methods, in combination with other approaches, has enabled the publication of detailed knowledge of 1393 different protein folds, with 36,853 unique protein structures (PDB Statistics, www.rcsb.org/pdb, accessed 20 March 2010).

1.8.2 Other structural techniques

The technique of Nuclear Magnetic Resonance spectroscopy (NMR), previously limited to small molecules, has become an important tool in the study of protein structure. NMR has the advantage over X-ray crystallography of giving an ensemble model of a protein structure in solution, rather than in a fixed crystalline lattice. In the case of

horseradish peroxidases, solution NMR studies of HRP variants were carried out (Veitch *et al.*, 1997) using other peroxidases (for example the peanut peroxidase structure published by Schuller and co-workers in 1996) as structural models before the HRP-C* crystal structure was solved (Gajhede *et al.*, 1997).

Similarly, cryo-electron microscopy has played a role in structural biology at the lower resolution range. Cryo-electron microscopy (cryo-EM), a technique that initially focused on viral capsid proteins due to their high level of symmetry (Whitford, 2005), is able to reach resolutions in which secondary structure elements can be identified. This is achieved by the amalgamation of many images of a protein in different orientations contributing to the overall model. Traditionally, images at 20Å-30Å resolution were produced, however models have been produced that approach 7Å resolution (for example the hep B capsid, Baumeister & Steven, 2000). Cryo-EM is being successfully used to give three-dimensional structural information about macromolecular structures that are beyond of the scope of current X-ray crystallographic techniques. The cryo-EM images are also being combined with portions of high resolution X-ray or NMR data to provide a fuller picture of larger molecules and the disposition of subunit components (Whitford, 2005).

Since each technique has its own clear strengths and weaknesses, combinations of structural techniques have proved successful in establishing structural information about larger and more complex multi-component molecular species.

1.8.3 Protein Data Bank

In the late 1960s it became apparent that the field of structural biology was increasing such that a means of sharing data was becoming necessary. In 1971, a repository was set up by the Brookhaven National Laboratory to collate data emerging from the structural biology community (Announcement in Nature New Biology, 1971) for ease of distribution amongst the scientific community of the time. The original Protein Data Bank (PDB), based at Brookhaven in New York, with copies stored in Cambridge and Tokyo, contained just seven structures at its inception (Berman *et al.*, 2000). By 1977 a total of 47 macromolecules had structural information recorded in the PDB, with a

variety of data available for each molecule based on solution data available from the authors (Bernstein *et al.*, 1977).

The Brookhaven Protein Data Bank continued to grow year-on-year, particularly in the 1980s, when technology reached the pace of advance in crystallographic structure solution (Berman *et al.*, 2004). In 1998, its management passed to the Research Collaboratory for Structural Bioinformatics (RSCB).

The RSCB PDB has continued the legacy left by the Brookhaven Protein Data Bank, in both growth and technological advancement. Original submissions to the Brookhaven Protein Data Bank were entered on punched cards, followed by magnetic tapes (Berman *et al.*, 2004). Now structures are submitted online, having gone through a preparation and checking process by the author, following strict guidelines on format (www.wwpdb.org/docs.html).

In 2009, there were 7,439 structures deposited in the RSCB PDB (total structures deposited between the launch and the end of 2009 is 64342). Of these 7,439 structures, 6,795 have been solved by X-ray crystallography, 574 by NMR and 51 by other structural methods, including electron microscopy (PDB Statistics, www.rscb.org/pdb, accessed 10/12/2009). If the 2009 figures are compared to the equivalent data from 10 years ago, where 1,965 structures were solved by X- ray crystallographic methods, 390 by NMR, and 1 by other methods (PDB Statistics, www.rscb.org/pdb, accessed 23/10/2009), we can see unequivocally how important X-ray crystallography remains in the solution of protein structures, and how much time has been devoted to the understanding of protein structure.

1.8.4 21st Century X-ray crystallography

Current trends in X-ray crystallography have taken two distinct paths to the solution of crystallographic structure. These trends have been largely defined by the organisation in which the structure is to be solved. The small scale crystallography groups, which tends to be located in universities or research institutions, has tended towards the crystallisation of complex protein molecules, often looking at larger complexes,

including protein-nucleotide complexes. The focus in academic institutions tends to be towards exploratory work with questions probed around the structure and function of specific enzymes and their variants, rather than high throughput, whole genome protein crystallography that is becoming more popular with larger more industrial crystallographers.

The other striking feature of the more recent data being deposited in the Protein Data Bank is the increasing size and complexity of the structures being solved (Sanderson and Skelly (eds.) 2007). Large macromolecular complexes are being studied, presenting an even greater challenge to the crystallographer, a challenge that is often favoured by the smaller research group. Examples of the variety of challenges faced when attempting to solve a protein's structure by X-ray crystallography are detailed below.

1.8.5 Crystallographic challenges

1.8.5.1 Challenge 1: Pure, highly concentrated, homogeneous protein

Before protein crystallisation can be even attempted, protein that is pure and homogeneous is required, in quantities that far surpass the amount needed for enzymatic assays. Growth of protein crystals requires a concentrated protein solution in excess of 5mg/ml, and, dependent on the success of the crystallisation attempts and the method used to set up the crystallisation experiments, may require a great deal of protein.

Prior to the development of recombinant DNA technologies, allowing genes to be expressed in host organisms, this need for pure, highly concentrated protein meant that the protein had to be extracted and purified from its original source. A potentially time consuming process, and one that was a barrier to crystallisation of many proteins, since adequate amounts of protein could not be obtained. The advent of recombinant DNA technology during the 1970s (Cohen *et al.*, 1973) has allowed the process of producing pure, homogeneous protein to no longer be such a bottleneck in the crystallisation process. This has been particularly important in the crystallisation of

peroxidases, as often these are heterogeneous *in vivo*, with multiple isoforms present as well as multiple surface glycosylation sites. For example, horseradish peroxidase, the subject of this thesis, has 15 isoforms that are isolatable by protein purification techniques (Veitch, 2004), with some plant species utilising in excess of 100 different peroxidase genes (Welinder, 2000). The most prevalent horseradish peroxidase isoform, HRP-C, has nine potential surface glycosylation sites, giving a predicted, but variable, carbohydrate content of 18%-22% (Veitch, 2004). In 1990, Smith and co-workers circumvented this problem with the production of recombinant HRP-C from *E.coli* (Smith *et al.*, 1990). Detail of producing pure, homogeneous HRP-C* variants for this project is recorded in Chapter 2, sections 2.2 and 2.3, based on the original method.

1.8.5.2 Challenge 2: Crystallisation

Producing protein crystals has replaced the purification of the protein solution as the bottleneck in the process of solving protein structures by X-ray crystallography.

To form crystals suitable for X-ray diffraction, protein must aggregate in a controlled way, such that a three-dimensional crystalline lattice structure forms. The term used for proteins close to precipitation is “super saturation”. A super saturated solution will be on the point of precipitation and must aggregate correctly for crystals appropriate for X-ray diffraction to be formed.

The process of crystallisation from a super saturated protein solution is facilitated by a precipitant solution – a solution that causes a reduction in the protein’s solubility and allows the crystalline lattice to form. The precipitant solution must not disrupt the protein fold and hence cause disordered aggregation. Figure 1.15 shows a few examples of the variety of disordered aggregates and precipitates seen in protein crystallisation attempts by this author, from which suitable protein crystals have not formed.

Much research has been carried out into the precipitant conditions required by proteins to form ordered crystalline aggregates (for example Stura *et al*., 1992) and a

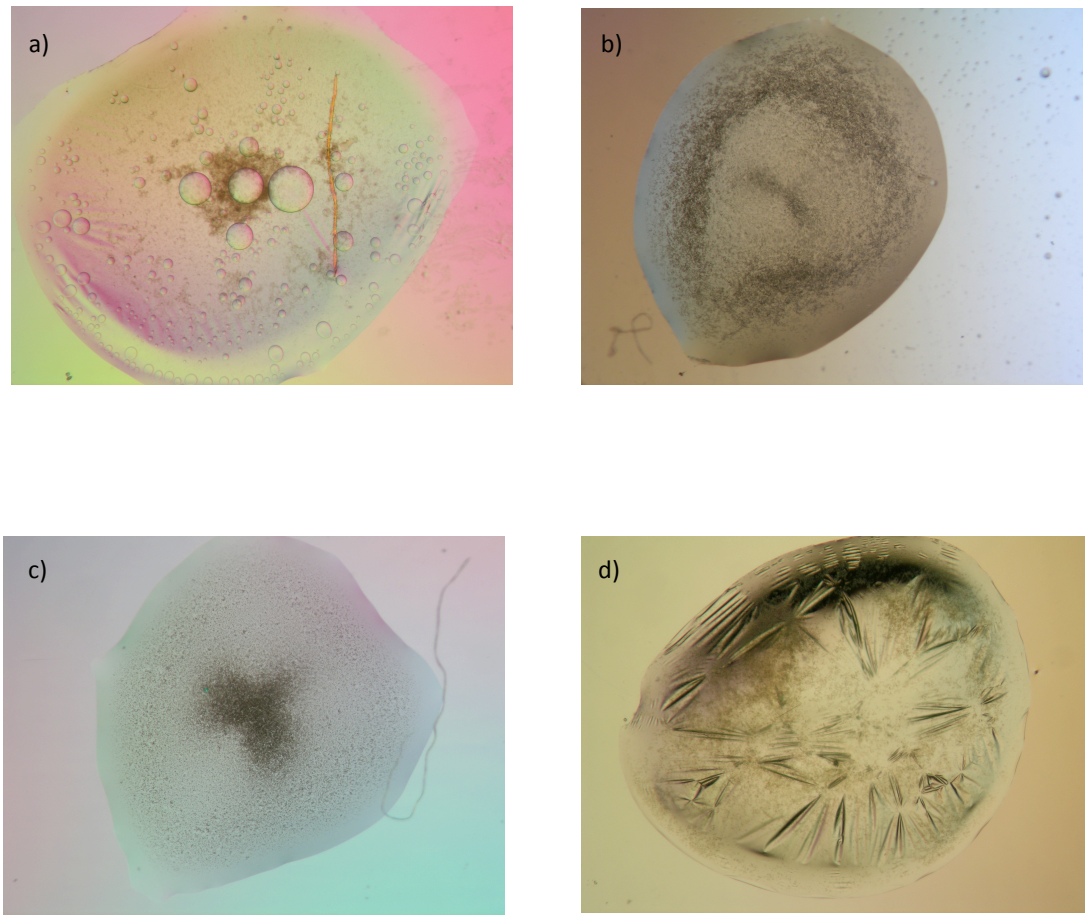


Figure 1.15 Precipitate observed in crystal drops
a) An oil film with protein seen in droplets within the precipitant.
b) Heavy amorphous precipitate.
c) Light amorphous precipitate (edges), heavy amorphous precipitate (centre).
d) A skin has formed on the drop, preventing diffusion between drop and well solution.

number of companies supply optimised kits containing precipitant solutions which have proved widely successful in facilitating the crystallisation process.

Figure 1.16, based on Chayen (2004) and others, shows an example of a protein crystallisation curve and illustrates the central problem in protein crystallography. The axes represent protein and precipitant concentration. There is a small region, marked metastable zone, in which ordered protein crystals can grow. The next region, marked labile (nucleation) zone, shows the region where crystal nuclei can form. This, in itself, can be a problem. If microcrystalline nuclei form, in order to form larger crystals around the nuclei, the solution must be in the crystal growth area, otherwise nuclei will form but crystals won't. This means that conditions that favour nucleation of protein in solution may not favour crystal growth and therefore nucleants will not produce diffraction quality crystals. Above the nucleation region is a region, marked precipitation, in which no crystalline formations will occur. This region, in which protein aggregates in the most disordered way, means that the balance of protein and precipitant concentration is wrong and the protein is highly unlikely to form ordered crystals. Instead the aggregates seen often mean that all the protein in solution has precipitated in a disordered fashion. In the cases where crystal growth is seen in this area, the crystallisation occurs because the protein concentration (y axis on Figure 1.16) drops such that the nucleation region is reached and then further so crystal growth can occur. In these cases enough protein has remained in solution to allow crystal growth.

Crystallisation trial experiments can be set up in a variety of different ways, each with its own merits, however all methods rely on a supersaturated protein solution and a precipitant solution. The different methods each exploit different properties of the crystallographic solubility curve (Figure 1.16) (Chayen, 2004).

Vapour diffusion has been the more common method of protein crystallisation for over 40 years (Chayen, in Sanderson and Skelly, 2007), however this may be gradually superseded by newer methods. Vapour diffusion, however is a process in which a

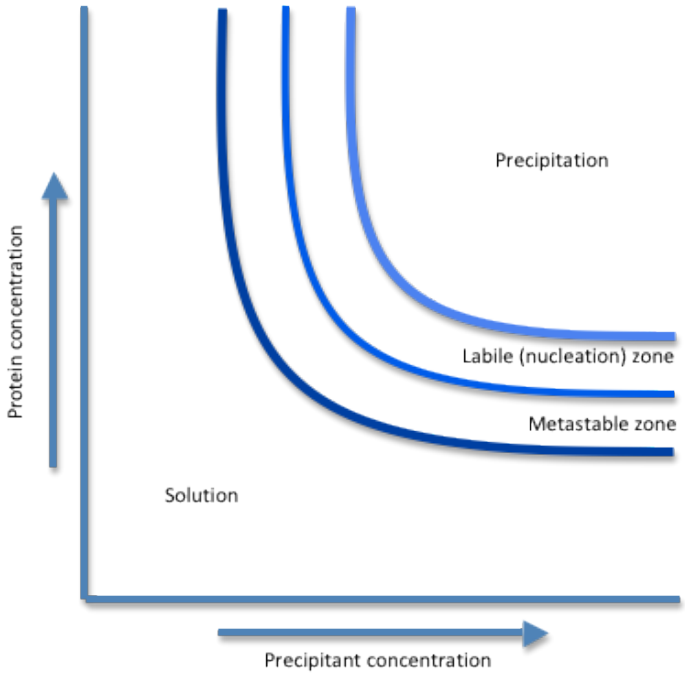


Figure 1.16 Protein solubility curve
With increasing protein and precipitant concentration, different aggregates may be seen in a protein crystallisation drop. These aggregates may be in ordered crystals or, as more common, disordered precipitate, however careful control of the conditions allow diffraction quality crystals to be formed. The solubility curve will be different for different proteins, with some proteins having much narrower labile and metastable zones than others, resulting in challenges in crystallisation conditions.

protein solution is mixed with an equal volume of precipitant and then allowed to equilibrate against a reservoir of precipitant solution. Diffusion slowly occurs between the drop and the precipitant allowing the protein solution to gradually become more supersaturated and, hopefully, crystal nuclei to form. The vapour diffusion method of protein crystallisation is often favoured by academic research laboratories as it is relatively quick to set up by hand, without the need for expensive robotic parts, crystal formation can be easily viewed in the protein-precipitant drop using a light microscope, and crystals grown in these drops can be quickly accessed for mounting and X-ray diffraction.

Two methods of setting up crystallisation experiments exist that facilitate crystal formation by vapour diffusion – the hanging drop and the sitting drop methods. In the hanging drop method, crystallisation plates, typically containing 24 wells that hold 0.5-1ml of precipitant solution, are used. These can either be sealed using a cover slip with a greased ring around the rim of the well allowing a sealed environment or, more recently, with lids fitted with a screw thread that can easily be removed if access to the drop is required. Protein-precipitant drops typically contain up to 5µl of protein and, usually, an equivalent volume of precipitant solution. Multiple protein drops, often at different protein concentrations, can be set on one cover slip, allowing flexibility in the conditions and facilitating maximum crystal growth.

In the sitting drop method, protein and precipitant solutions are placed or “sat” on a bridge with precipitant around or on a shelf at the edge of the well rather than being suspended above the drop. Wells are sealed and vapour diffusion is able to take place in the same way as the hanging drop method. Sitting drop experiments may require smaller volumes of protein and precipitant than hanging drop, favouring this method of crystallisation if crystallisation conditions are not known and therefore a wide range of conditions are to be examined.

The batch method of protein crystallisation, unlike the methods described above, does not require any dynamics between the protein drop and another solution (be it via

vapour diffusion or dialysis). Instead, protein and precipitant solution are mixed and then incubated under paraffin oil. The less dense oil forms a protective coating around the drop, so no evaporation can take place. The majority of precipitants can be used in batch methods, though phenol, dioxane or thymol must be avoided to prevent solution dissolving in the oil (Chayen and Saridakis, 2008).

More recently, the use of gels has been shown to be successful in the growth of high quality macromolecular crystals. This method follows on from over a century of work crystallising small molecules using gel based media (Henisch, 1988) and has been proven to prevent crystal sedimentation as well as showing an overall improvement in the crystallisation process through reduction of convection within the solution (Biertümpfel *et al.*, 2002). It has also been suggested that this method increases the internal order of the crystal, reducing mosaicity and preventing twinning (Biertümpfel *et al.*, 2002).

The equilibrium dialysis method has advantages over alternative methods since it requires small volumes of protein. There have been many different designs of dialysis chamber (for example an early method was designed by Weber and Goodkin (1976), however all rely on the protein solution being separated from the precipitant by a semi-porous membrane, through which gradual diffusion takes place.

Crystal growth has often been described as more of a black art than a science, particularly as there are so many variables that can affect the growth of crystals suitable for X-ray diffraction (David Watkin, British Crystallographic Association annual meeting, Loughborough, 2009). The process of crystal growth is often a frustrating process that may require multiple different strategies and techniques, let alone protein and precipitant concentrations, however the reward of producing X-ray diffraction quality crystals balances the frustration of forming them in the first place.

1.8.5.3 Challenge 3: Data collection

Crystallographic data collection is the most important step in the process after growth of diffraction quality crystals. Crystals are sensitive to the ionising power of the X-ray

radiation, particularly with the short wavelengths of X-rays used at synchrotron sources, and therefore optimisation of data collection is key in obtaining the best data from the crystals available.

Mounting a crystal in an X-ray beam can be achieved in a number of ways, ways that may depend on the crystal and the growth method. Previously, the most common method of mounting crystals in the path of the X-ray beam involved coaxing the crystal into a fine capillary tube with as little liquid as could be achieved. This was then mounted at room temperature and diffraction images taken. This method holds advantages, particularly if there are only a very small number of crystals and potential damage caused by introducing the crystal to stresses that may cause its lattice to be disrupted. The capillary method of mounting crystals, however, has limitations, particularly as the capillary and liquid offer little protection to the crystal from the powerful X-ray beam and crystals are subject to damage by the X-ray beam during data collection.

Much effort has been made in the study of the effects of radiation damage to crystals and how it might be best avoided or exploited. One of the best exploitations of the damage that X-rays cause to crystals is exhibited in the study of the oxidation states of horseradish peroxidase by Berglund and co-workers in 2002. In this work, multiple crystals were used in the collection of data, with single crystal microspectrophotometry employed to analyse the oxidation state of the crystal and separate the data in to the corresponding HRP-C oxidation states at different points in the data collection (Berglund *et al.*, 2002). This enabled the establishment of a full X-ray crystallographic analysis of the five different oxidation states of HRP-C: Compounds I and II of the typical peroxidase catalytic cycle, Compound III, a separate dioxygen bound haem state that is seen if the enzyme is exposed to an excess of hydrogen peroxide (Berglund *et al.*, 2002), and both ferric ($Fe\ III^{-}$) and ferrous ($Fe\ II^{-}$) forms of the haem. Without the understanding and management of X-ray induced reduction in the crystals, the intermediates of the HRP-C catalytic cycle would remain structurally undefined. This work also raised a question about the validity of similar enzymes whose structures have been solved in a particular redox state and deposited in the

PDB, and, in particular, whether the redox state of many published structures is truly known (Berglund *et al.*, 2002). X-ray data collected from crystals subject to prolonged exposure showed reduction of a disulphide bond within the HRP-C* structure (Berglund *et al.*, 2002).

The method now used to protect crystals during data collection is known as cryoprotection, a technique in which the crystal in solution is rapidly cooled, forming a glass effect around the crystal, before X-ray diffraction. Cryoprotection of crystals has different methods available, though all have the same goal: to form vitreous glass, rather than crystalline ice, around the crystal when the crystal is placed at 100K (Abdel-Meguid *et al.*, in Sanderson and Skelly, 2007). The formation of vitreous ice rather than crystalline ice prevents the disruption of the crystalline lattice, of which a significant proportion is solvent (40-90%, Abdel-Meguid *et al.*, in Sanderson and Skelly, 2007). This also prevents the appearance of "ice rings" in the X-ray diffraction pattern, large rings that form as a result of diffraction from the ice crystals and can significantly disrupt the analysis of diffraction data as well as the potential problems caused by distortion of the crystalline lattice.

There are a variety of different techniques and methods used in the cryoprotection of crystals in preparation for data collection. These involve the movement of the crystal into an alternative solution, either for a short or longer period of time, to either coat the crystal in a cryoprotective buffer, or to completely replace the precipitant solution that the crystal was grown in with a cryoprotective buffer. This needs to be done unless the precipitant solution used in crystal growth is able to act as a cryoprotectant, however this is not usually the case.

Three common methods of cryoprotection are described by Abdel-Meguid *et al.*, in Sanderson and Skelly, 2007. These are: the slow equilibration method, the "quick dip" method and the immiscible hydrocarbon method. Though these methods are easily described, the suitability of the method and the specifics of solutions used will be dependent on the crystals grown and therefore the choice of cryoprotectant remains key to high quality data collection.

In both the quick dip and the immiscible hydrocarbon methods, crystals are selected from their growth medium and transferred quickly to a new solution. As the name suggests, the slow equilibrium method gradually moves the crystal from its growth solution to a solution that will perform well under cryo-cooled conditions. This is carried out in stepwise transfer that allows the crystal to gradually soak in the new conditions without disruption of the crystal lattice.

The immiscible hydrocarbon method, a method only really suited to stable crystals that can withstand the hydrocarbon solution they are immersed in, depends on the displacement of the solvent surrounding the crystal in its growth solution by the hydrocarbon during the immersion (Abdel-Meguid *et al.*, in Sanderson and Skelly, 2007). The crystal can then be mounted and X-ray diffraction can take place.

In the case of the quick dip method, by far the most common method of cryoprotection, the solution the crystal is transferred to tends to be a solution similar to the growth solution, but containing one or more additive that has been previously tested to ensure that vitreous ice forms upon cryogenic freezing. Various additives have been suggested as “antifreeze” chemicals that facilitate the formation of vitreous ice in a sample. Since crystals are sensitive to environmental change from the solution they are grown in, cryoprotective solutions are chosen that are not significantly different from the crystal growth solution, or antifreeze additives are mixed with crystal growth solutions to form the optimum balance between ensuring vitreous ice formation when cryo-cooled and reducing the stress on the crystal that may result in the disruption of the crystalline lattice and destruction of the crystal.

A number of different chemicals have been shown to cause vitreous ice formation, relying on the individual crystallographer to experiment with suggested additives and their own crystal growth solution. Although this procedure will largely rely on testing different additives, there are cryoprotectant trial kits available containing the most commonly used solutions.

X-ray source also will influence the quality of data collected. In-house data collection can give good quality data, however the wavelength of the X-rays will always limit the maximum resolution of the data collected. Most in-house data collection is carried out using X-ray kits containing a rotating anode generator, in which electrons are fired at a metal target, typically copper for X-ray diffraction of proteins and protein complexes, producing X-rays (Sanderson, in Sanderson and Skelly, 2007).

The X-rays are produced by the metal target since the high power electrons fired at the target cause electrons in inner orbitals of the metal to be displaced. X-rays are emitted since outer orbital electrons must drop to inner orbitals to replace those dislodged by the bombardment from external electrons (Sanderson, in Sanderson and Skelly, 2007). The copper rotating anode generator produces X-rays of 1.542Å wavelength, restricting the maximum resolution of diffraction collectable from an in-house source.

Shorter wavelength diffraction can be obtained by using a synchrotron source for X-rays, with wavelengths in 1Å range. This also has the advantage of being able to collect higher resolution data sets from poorer quality or smaller crystals, with protein structure determination possible from crystals as small as 30-100µm at maximum length (Wasserman *et al.*, in Sanderson and Skelly, 2007).

1.8.5.4 Challenge 4: Structure solution

The task, once a set of X-ray data has been collected, is to assign phases to the data and set about solving the protein structure using the calculated electron density maps to create a contour pattern in which to fit the amino acid sequence. Unfortunately this is not a simple task in the majority of protein X-ray diffraction patterns as, due to the size of the molecules involved, phases have to be established rather than assigning them in the same way as small molecule chemical crystallography.

The simplest method of assigning phase information to a data set is to effectively borrow the phase information of a previously solved structure, a process known as molecular replacement. This, however, relies on a high-resolution structure with significant expected structural homology to the new structure, combined with manual

refinement of the model structure. Various programmes are available to carry out automated molecular replacement, all relying on algorithms to compare model structure phasing to the information available from the new data. The electron density map produced by the automated systems must then be studied at an individual amino acid level, ensuring that each part of the molecule fits within the parameters of the map. When data collected is at high resolution, this technique is a relatively quick and efficient method of determining protein structure.

Other techniques for solving the crystallographic "phase problem" use methods to interrogate the crystal in different ways, aiming to provoke changes in the diffraction pattern obtained without altering the tertiary structure of the protein. This allows knowledge of where the modifications are expected to occur within the protein structure to be used to extract phase information.

When a suitable phase model is not already available, alternative methods must be used in order to calculate an electron density map with which to work. Various alternatives exist, namely Multiple Isomorphus Replacement (MIR), Multi-wavelength Anomalous Diffraction (MAD) and Single-wavelength Anomalous Dispersion (SAD).

MIR, otherwise known as the heavy-atom method, utilises the fact that each atom in a protein contributes to the diffraction pattern obtained and that it is possible to displace atoms in a protein and substitute these for heavy atoms such as mercury, platinum or gold without disturbing the overall structure of the protein (Rhodes, 2000). This technique enabled the first protein crystallographic structures to be determined, with Kendrew and co-workers attaching different complex metal ligands (such as *p*-chloro-mercuri-benzene sulphonate, potassium mercuri-iodide) to the myoglobin molecule, though at the time they did not know whether the heavy atom binding might have any effect on the molecule (Kendrew *et al.,* 1958). Binding of heavy atoms to the protein, providing they do not affect the unit cell properties, will allow calculation of the location of those heavy atoms in the unit cell. The heavy atoms have an effect on the intensities of reflections in the diffraction image. Using multiple different heavy atom derivatives allows models to be created which locate

the heavy atoms in real space. Structure factors can be assigned to each heavy atom, through which both amplitudes and phases can be calculated. Using these assigned phases, the full structure can be phased (Rhodes, 2000).

MAD, unsurprisingly, relies on the availability of multiple wavelengths of X-ray, combined with the ability of heavy atoms to absorb specific wavelengths when exposed to these X-rays. Oxygen, nitrogen and carbon atoms are unaffected by this technique (Rhodes, 2000). This technique, described by Guss *et al.* in 1988, uses properties of atoms already existing in the protein, such as metal iron centres or sulphur groups (or that can be inserted in the protein expression stage such as selenium through selenomethionine). Diffraction patterns are collected from a crystal using different wavelengths of X-rays (thus is reliant on access to synchrotron sources with variable wavelengths), which produce different intensities of diffraction as the radiation is absorbed by the heavy atoms. From the diffraction intensities, known atoms can be assigned relative location and phase of these atoms can be calculated. This phasing can then be extrapolated across the normal data and electron density maps created (Guss *et al.,* 1988) and refined.

SAD, unlike MAD, does not require the crystal to be subject to extended beam time and hence potentially sustain radiation damage, nor does it require a tuneable X-ray source and hence use of a synchrotron. Particular heavy atoms, for example sulphur atoms, produce anomalous diffraction data at wavelengths achievable on “home” sources (Rose *et al.,* 2001). However, SAD tends to require complete datasets and identification of each anomalous pair (Gonzalez, 2003) in order to successfully complete the phase allocation for the anomalous atoms.

Each structure solution technique has both advantages and disadvantages, therefore each technique is regularly used in solution of crystallographic structures, dependent on the availability of resources, including synchrotron beamtime and existing homologous structures.

The structure solution technique used in this work was molecular replacement, with

the model structure published by Henriksen *et al.*, in 1999 (PDB id: 7ATJ) to be used for phasing information.

1.8.6 Protein structure

The importance of understanding protein structure is clear, giving insight into mechanisms of action of complex proteins and systems and enabling precise targeting of residues for site directed mutagenesis, as well as the enormous implications that understanding of protein structure has on the study of disease. Knowledge of a protein's structure combined with study of detailed function can be developed and capitalised on, with enzyme structures compared and features introduced from other proteins to develop enzymes in the creation of more efficient catalysts.

In the 15 years since the solution of the first Class III plant peroxidase (Schuller *et al.,* 1996) study of the structure-function relationship of peroxidases has allowed detailed descriptions of the catalytic mechanisms of this complex group of enzymes. The work detailed in this thesis adds to this understanding, highlighting how relatively small amino acid substitutions can have a profound effect on enzymatic activity.

1.9 Aims

The aim of this work is to explore the relationship between structure and function in the plant peroxidase HRP-C*, further examining the complex determinants of HRP-C* catalytic activity. This is a question that has continually been asked of peroxidases, although it was particularly important during early work in which peroxidases were being compared to the globins, with structural differences meaning the difference between reversible oxygen binding of globins and the peroxidative activity of peroxidases (Poulos and Kraut, 1980).

Three varieties of HRP-C* variant were studied, each probing a different feature of this enzyme, with the structural determinants of function investigated. Each is an iteration of work previously carried out, expanding the preceding work and probing the structure-function relationship of HRP-C further.

The first portion of the work described in this thesis (Chapters 3 and 4) involved a series of HRP-C* variants termed “peroxygenase” variants. These were initially designed, in previous work, to mimic the active site of the enzyme chloroperoxidase, an enzyme with catalytic properties of both peroxidases and cytochrome P450s. Novel HRP-C* variants were created that altered properties of the previous variants. Previous variants were examined alongside the newly designed variants, expanding the range of substrates these peroxygenase variants were tested with. Variants were subject to examination by UV/Visible spectrophotometry in characterising steady state activity. In addition, attempts were made to grow crystals of these new variants for examination by X-ray crystallography, however, as shown later in this thesis, these attempts did not prove entirely successful.

The second variant species described in this thesis was the S167M HRP-C*, an HRP-C* variant designed and created during the doctoral study of Dr Khasim Cali. This variant introduced a methionine residue that formed covalent linkage between the haem and the protein, modelled on the mammalian peroxidase myeloperoxidase. Crystallographic study of this HRP-C* variant is described in Chapter 5 of this thesis, with new insights in to the formation of the linkage suggested by the crystal structure.

The third HRP-C* variant species is a double mutant HRP-C* variant that reduces sterically hindering residues at the entrance of the haem access channel. Residues 140 and 179 (alanine and phenylalanine respectively) restrict access of substrates to the haem and therefore control rates of catalytic activity for the enzyme. The double mutant A140G/F179A, created by Dr Cali and based on previous work on single mutants by Dr Gareth Jones, was crystallised and an X-ray structure modelled, fully described in Chapter 6 of this thesis. This work reveals the degree to which the reduction in the size of these residues increases the available area at the entrance to the haem access channel, affecting the ability of substrates to access the active site of the HRP-C* molecule. It supports previous results showing an increased rate of activity with the substrate luminol, a finding with potentially significant consequences.

Study of the three different types of HRP-C* variant described above has increased understanding of the relationship between structure and function in this enzyme. In addition, it suggests future direction for further examination of the structure-function relationship of this versatile enzyme.

Chapter 2

Materials and methods

2.1 General remarks

Chemical reagents quoted in this thesis were purchased from Sigma Aldrich unless otherwise stated.

Restriction enzymes were purchased from New England Biolabs. Digests were carried out in the optimum buffer for the enzyme, supplied by New England Biolabs.

Crystallographic reagent kits were purchased from Molecular Dimensions Ltd.

Preparation of protein variants

2.2 Site directed mutagenesis

The polymerase chain reaction (PCR) was used to synthesise the desired protein variants. The commercially available expression vector pFLAG1 (International Biotechnologies Inc) with the synthetic wild type HRP-C (HRP-C*) gene (Smith et al. 1990) inserted into the multiple cloning site was used as a template for PCR. Complementary primers were designed to introduce specific mutations to the HRP-C* gene sequence of pFLAG1-HRP-C* during the PCR reaction, essentially as described by Doyle *et al* (1998). Primers were designed to sit back-to-back on opposite strands of the DNA sequence and contained base substitutions that not only introduced the amino acid point mutations required, but also created an additional restriction enzyme site in order to select for the inclusion of the plasmid containing the mutations in transformants (Doyle et al., 1998). Figure 2.1 shows a diagrammatic representation of this process of site directed mutagenesis using the whole plasmid amplification method.

Protein expression in pFLAG1-HRP-C* is controlled by the tac promoter sequence which is suppressed by the *lac*I^q gene. Protein expression is inducible by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG).

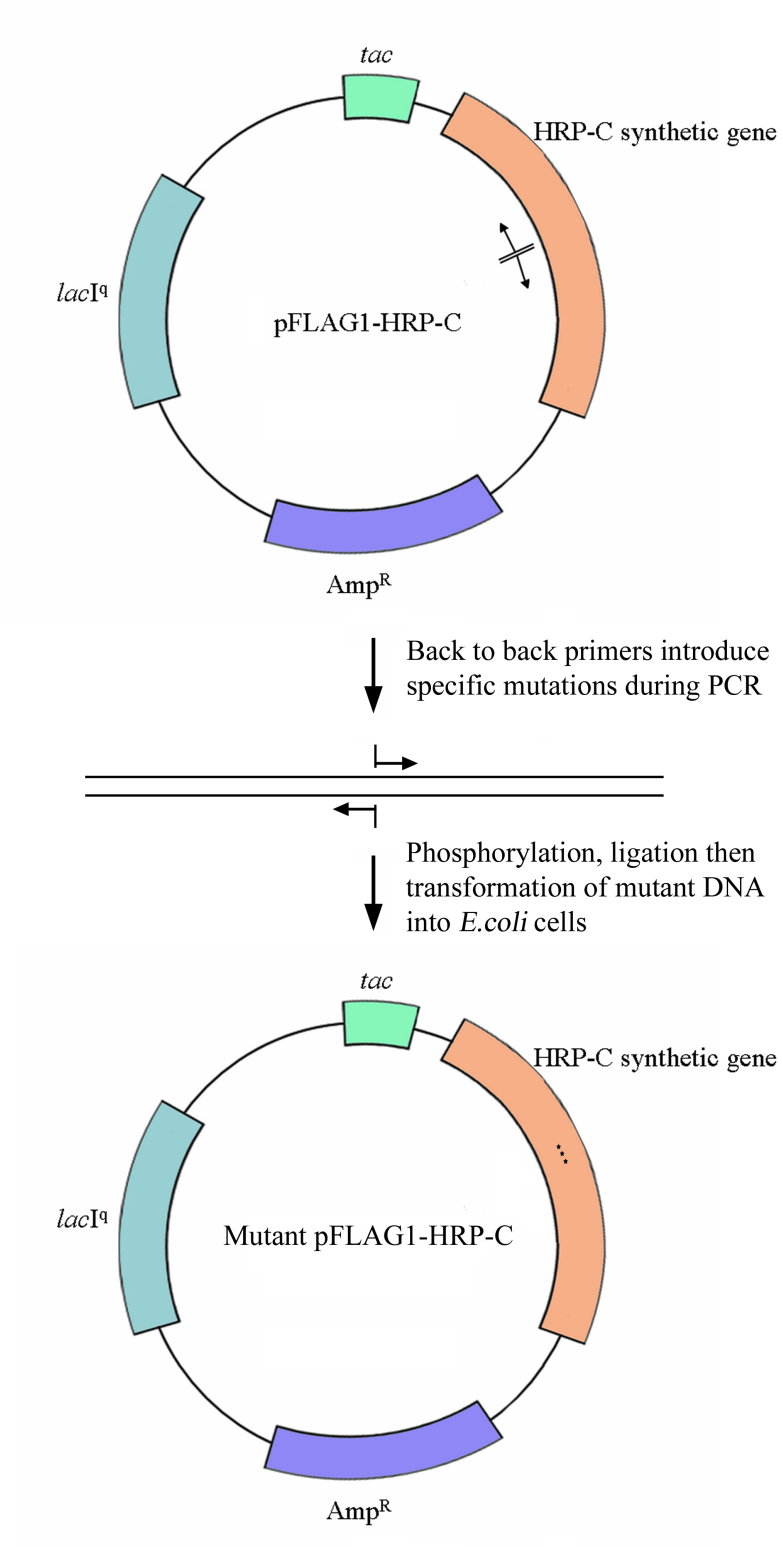


Figure 2.1 Process of site directed mutagenesis of pFLAG1 HRP-C

The whole plasmid amplification method of HRP-C* synthetic gene. Primers were designed that introduced specific mutations in to the gene at the desired locations. Template pFLAG1-HRPC was incubated with variant primers and subject to PCR conditions as described in Chapter 2, with variations to this method described in Chapter 3.

Following successful PCR, as determined by agarose gel electrophoresis, DNA was phosphorylated, ligated and transformed into competent *E.Coli* cells.

A high fidelity polymerase (Pfu polymerase) (New England Biolabs) was used in the PCR reaction to minimise the introduction of errors into the plasmid PCR product.

2.2.1 Primer Design

Oligonucleotide primers for each of the desired mutants were designed, with codon substitutions favourable to *E.coli*, that would allow the desired amino acids to be incorporated into the plasmid DNA (see Table 2.1). As well as including codon substitutions that would replace amino acids, the primers were also designed to incorporate an additional restriction enzyme site to allow selection of correctly mutated sequences. This technique has advantages over other methods of site directed mutagenesis as potential correctly constructed and ligated genes are selectable from transformants at an early stage (Doyle et al., 1998). These initial selections are later confirmed by DNA sequencing, however this method has the ability to greatly reduce the number of incorrect sequences examined.

Variant	Primers (5'-3')	Unique restriction enzyme site	Approx. Tm
R38E/F41A/H42A (EAA)	1. **GCA**GACTGCTTCGTGAATGGTTGC 2. A**GC**GTGCAG**TTC**TAATATTGAAGCAGCGATCC	*PstI* – CTGCAG	64°C 56-76°C
R38A/F41E/H42A (AEA)	3. TCAC**GAAG**CTGACTGCTTCGTGAATGGTTG 4. AGA**GC**TAATATTGAAGCAGCGATCCTG	*SapI* – GCTCTTC	62°C 62°C
R38A/F41A/H42D (AAD)	5. TCAC**GC**C**G**ATGACTGCTTCGTGAATGGTTG 6. AGA**GC**TAATATTGAAGCAGCGATCCTG	*SapI* – GCTCTTC	64°C 62°C

Table 2.1. Designed primers for new HRP-C* variants.
Codons in bold represent codon changes that cause amino acid substitutions. Codons underlined represent silent mutations that allow the introduction of an additional restriction enzyme site.

Potential restriction enzyme digest sites were found using the online program Webcutter (Version 2.0 © Max Heiman, 1997) which identifies restriction enzyme sites, and sites that could be created by silent mutations, within a specified region of input DNA. Each of the desired mutant sequences was input to Webcutter and a variety of restriction enzyme sites were identified. Restriction enzymes were selected

on the basis of proximity to the mutated region – as the site would ideally incorporate the region of the mutation to ensure the codon substitutions had occurred correctly – and by looking at the location and quantity of other recognition sites in the pFLAG1-HRP-C* vector for each potentially suitable enzyme. Ideal restriction sites would be found in the region of the desired mutation and correct mutations would be identifiable, when digested with the enzyme and run on an agarose gel, by creating a DNA fragment of unique length.

Primers were synthesised by MWG Biotech and supplied in a lyophilised form. The primers were solubilised with sterile 10mM Tris pH8.5 to a concentration of 100µM.

2.2.2 Polymerase Chain Reaction

Reagents

Pfu polymerase

Pfu polymerase buffer

dNTPs – dATP, dTTP, dGTP, dCTP

Designed primers

Template DNA

Sterile ddH_2O

Temperature (°C)	Time (min)	Explanation of step
95	10	"Hot start" at the start of the reaction completely denatures the template DNA
50-60	1	Allows annealing of the primers to the template DNA strands
72	20	Extension of the primers by *Pfu* polymerase. Extension rate ~1kb/2min
Repeat steps for 25 cycles		
4	∞	Incubation

Method

Sterile PCR reactions were set up with 5µl Pfu buffer, 5µl 1mM dNTPs, 1µl pFLAG1-HRP-C* template DNA with 5µl of both the forward and reverse primers. The reaction

medium was made up to 49.5µl with ddH_2O. After the five minute "hot start" phase of the programmed PCR reaction, 0.5µl of *Pfu* high fidelity DNA polymerase was added.

25 cycles of the PCR reaction were carried out, after which the tubes were incubated at 4°C.

2.2.3 Confirmation of PCR product

Reagents

PCR reaction sample

Ficoll gel loading buffer

Tris/Borate/EDTA buffer (TBE)

DNA ladder

Ethidium bromide

Sterile ddH_2O

Method

5µl of the PCR reaction medium was removed, to which 3µl of 6x Ficoll gel loading buffer was added. Samples were subject to gel electrophoresis on 1% agarose gels and run with TBE, proportions according to Sambrook *et al*., (1989).

Agarose gels were stained using ethidium bromide, with each gel incubated at room temperature in a dilute solution of ethidium bromide. Stained gels were visualised using a UV transiluminator. PCR reactions that had amplified the plasmid showed a clear band at approximately 6.3kb, measured using a 1kb DNA ladder (New England Biolabs) run alongside the samples on the gel.

2.2.4 Purification of PCR products

Reagents

Dpn1 restriction enzyme

Sterile ddH_2O

Phenol:Chloroform:Isoamylalcohol (25:24:1)

Chloroform

3M sodium acetate solution, pH 5.2

Ethanol (100% at -20C, 70% at room temperature)

2M Tris buffer pH 8.5

Method

Positively identified PCR products were incubated for 1 hour at 37°C with 1µl *Dpn1* to digest any of the methylated template DNA in the sample.

Samples were transferred to sterile microcentrifuge tubes and ddH_2O was added to a final volume of 150µl.

To extract the protein from each sample, an equal volume of Phenol:Chloroform:Isoamylalcohol was added to the sample, the tube was inverted to mix and centrifuged at 13,000 rpm for 2 minutes. The centrifugation causes the separation of an organic layer and an aqueous phase; protein forms at the interface between the two layers. The aqueous phase, containing the DNA, was removed into a new microcentrifuge tube without disturbing the material at the interface or removing any of the organic layer.

An equal volume of chloroform was added to the new aqueous sample and inverted to mix. This was then centrifuged at 13,000 rpm for 1 minute and the aqueous phase extracted as previously.

10% v/v 3M sodium acetate solution, pH 5.2, was added to the sample. 2.5 volumes of 100% ethanol (stored at -20°C) was added to the sample and the solution was kept on ice for 10 minutes to chill. The sample was centrifuged at 13,000 rpm for 5 minutes, the supernatant removed and replaced with 70% ethanol at room temperature. This was again centrifuged at 13,000 rpm for 5 minutes and the supernatant removed. The remaining ethanol was evaporated in a sterile environment and the DNA dried fully. DNA was resuspended (in sterile Tris buffer pH 8.5) to a volume of 8µl.

2.2.5 Ligation of DNA

Reagents

T4 phage polynucleotide kinase (PNK)

10x Ligase buffer

Bovine serum albumin (BSA)

T4 DNA ligase

ddH_2O

Method

Polynucleotide kinase was used to phosphorylate the 5' tail of the DNA prior to ligation.

1µl of T4 PNK and 1µl 10x ligase buffer were added to the DNA obtained in the purification process. 1µl of BSA was also added to the reaction and samples were incubated at 37°C for a minimum of two hours.

Ligation was performed after incubation. 1µl of 10x ligase buffer, 8µl ddH_2O and 1µl T4 DNA ligase was added to each sample. Samples were incubated overnight at 4°C.

2.2.6 Transformation into competent *E.coli* cells

Reagents

Library efficiency chemically competent *E.coli* DH5α cells (Invitrogen)

Ligated DNA

Luria-Bertani (LB) Broth

Luria-Bertani (LB) Agar plates, inoculated with 100µg/ml ampicillin

Method

Chemically competent DH5α *E.coli* cells were thawed on ice and aliquots of 100µl per microcentrifuge tube were used for each transformation. 10µl of variant plasmid DNA was added and the cells mixed gently. Cells were left on ice for 10 minutes before heating for 45 seconds at 42°C in a water bath. Cells were placed on ice for a further

five minutes, after which time 1ml sterile LB Broth was added to each sample. Samples were incubated in a shaking incubator at 37°C and 180rpm for 1 hour, allowing antibiotic resistance genes to be expressed.

Cells were centrifuged at 6000 rpm for 2 minutes, 800µl of supernatant was removed. The pellet was resuspended in the remaining supernatant and 100µl was plated on to sterile LB agar plates, containing ampicillin.

Agar plates were incubated over night at 37°C and plates on which colonies had grown were stored at 4°C for up to one month.

2.2.7 Confirmation of site directed mutagenesis

Reagents

LB Broth
Qiagen miniprep kit (Qiagen)
Restriction enzyme as dictated by designed primers
Appropriate buffer for restriction enzyme
Bovine serum albumin
6x Ficoll gel loading buffer
1% agarose gel
1kb DNA ladder (NEB)
Tris/Borate/EDTA buffer
Ethidium bromide

Method

12 colonies were picked and allowed to grow overnight at 37°C in individual 5ml samples in a shaking incubator.

Overnight growths were prepared as directed by the Qiagen miniprep kit.

Correctly ligated mutant DNA was screened using the unique new restriction enzyme site introduced with the variants. Plasmid DNA was incubated with the appropriate

restriction enzyme, suitable buffer (determined by New England Biolabs, supplied with the enzyme) and BSA for 1 hour at 37°C.

The product of the restriction enzyme digest was stained using 6x Ficoll gel loading buffer and was subject to gel electrophoresis on a 1% agarose gel. A 1kb DNA marker was added to the gel to allow identification of length of bands seen.

Bands on the agarose gel were visualised under UV and colonies identified as positive or negative for the correct insert based on their banding pattern.

2.2.8 Sequencing of positive colonies

Reagents

50ml LB broth

100mg/ml Ampicillin stock

Promega Wizard Midi Prep kit

Tris buffer pH 8.5

Method

Positive colonies were inoculated into flasks containing 50ml of LB broth (a maximum of two positive colonies from each variant were selected at a time). These flasks were incubated in a shaking incubator at 37°C and 180 rpm overnight.

DNA was extracted according to the protocol of the Promega Wizard Midi Prep kit.

DNA concentration was calculated using a Shimadzu UV/Visible spectrophotometer. DNA was diluted 1/50 in sterile Tris buffer pH 8.5 and a spectrum between 320-220nm was taken of the sample, blanked against Tris buffer. The ratio of absorbances at 260nm:280nm gave an indication of the purity of the DNA in the sample. The absorbance at 260nm was used to calculate the concentration of the DNA.

DNA samples to be sent for sequencing were diluted to 100ng/µl. DNA samples, along with standard primers for the pFLAG1 plasmid, were sent to Lark Biotechnologies for

DNA sequencing, results were returned by e-mail. Sequencing confirmed correct point mutations as well as identifying any errors in the PCR or the ligation reaction.

Protein expression and purification

2.3.1 Transformation for protein expression

Reagents

Chemically competent W31110 cells (from a lab-grown stock)

Correctly sequenced variant DNA

LB Broth

LB Agar

50% glycerol

Method

Transformation was carried out as described previously, however, as W31110 cells were used, the heat shock step (incubation at 42°C) was carried out for 2 minutes rather than for 45 seconds. The length of incubation with LB broth allowing antibiotic resistance to be expressed was also reduced from 1 hour to 30 minutes.

Cells were plated on to LB agar plates inoculated with 100µg/ml and incubated at 37°C overnight.

A small number of colonies were selected from the colonies grown on the LB agar plate. These colonies were introduced to 5ml LB and incubated at 37°C overnight. From these cultures, two samples were taken with 1ml 50% glycerol added to each sample and inverted to mix. These samples were frozen in liquid nitrogen and stored at -80°C. The most effective of these, determined by small scale expression testing, were used as starter cultures for future *E.coli* growth.

2.3.2 Small scale protein expression

Reagents

Sterile LB broth

Ampicillin

isopropyl-β-D-thiogalactopyranoside (IPTG)

Method

50µl of overnight or glycerol culture from the cultures grown from correctly sequenced DNA plasmids was inoculated into 5ml of LB broth (fix this sentence!). These cultures were incubated at 37°C and 180rpm, until the optical density (OD) of the cells at 500nm reached 0.6 absorbance units, measured using a Shimadzu UV/Visible spectrophotometer blanked on LB broth.

Protein expression was induced by the addition of 1mM sterile IPTG and cultures were grown for a further three hours. One culture was inoculated but left uninduced as a control culture.

Cells were harvested in microcentrifuge tubes by centrifugation at 7,000 rpm for 2 minutes. The supernatant was discarded and protein extracted from the cell pellets.

2.3.3 Small scale protein preparation

Reagents

Lysis buffer: 20mM Tris, pH 8.0, 1mM EDTA, 1mM DTT, 2M urea

Triton X-100

Protein resuspension solution: 50mM Tris, pH 8.0, 1mM EDTA, 30mM DTT, 10M urea

Method

Cells were resuspended in 500µl lysis buffer in a microcentrifuge tube and sonicated for 30 seconds using a Vibra Cell VCX500 sonicator. Cell suspension was placed on ice to recover. Cells were centrifuged for 15 minutes at 13,000 rpm in a microcentrifuge. The supernatant was removed and retained.

The pellet was resuspended 1ml lysis buffer containing 1% Triton X-100 and incubated at room temperature for 10 minutes. Resuspension was centrifuged for 15 minutes at 13,000 rpm in a microcentrifuge. The supernatant was discarded.

The protein pellet was resuspended in 30µl protein resuspension solution by vortexing. The protein suspension was centrifuged for 10 minutes at 13,000 rpm in a microcentrifuge. The supernatant, containing unfolded HRP-C*, was retained.

2.3.4 SDS polyacrylamide gel electrophoresis

Reagents

12% Resolving gel:

1.25ml 3M Tris, pH8.5

4.3ml 28% Acrylamide

0.05ml 20% Sodium dodecyl sulphate (SDS)

1.9mlddH_2O

2.5ml Ammonium persulphate

5µl N,N,N',N'-Tetramethylethylenediamine (TEMED)

Stacking gel:

1.5ml 1M Tris, pH 7.0 / 5% SDS

1.5ml 28% acrylamide

3.0ml ddH_2O

6.0ml Ammonium persulphate

15µl TEMED

Running buffer:

Tris, pH 8.3

Glycine

20% SDS

ddH_2O

Sample stain:

1g SDS

2ml glycerol

2ml bromophenol blue

1.25 1M Tris, pH 7.0

2ml 2-mercaptoethanol

2.75ml ddH_2O

Gel stain:

0.1% Coomassie Blue

10% Acetic acid

10% Ethanol

Destain:

10% Acetic acid

10% Ethanol

Method

Samples were stained 1:1 with bromophenol blue stain, allowing progress through the gel to be monitored. 15µl of each stained sample was loaded to the gel, along with a wide range molecular mass marker (Sigma Aldrich).

Gels were run at 20mA until the samples had moved into the resolving gel, at which point, current was increased to 40mA. Current was stopped when the dye front had reached the bottom of the gel.

Gels were stained overnight in a Coomassie Blue stain and destained the next day until bands were clearly visible.

2.3.5 Large scale protein expression

Reagents

Sterile Terrific Broth (TB), containing yeast, tryptone and glycerol

Sterile TB Phosphate buffer

50ml overnight *E.coli* culture

Ampicillin

IPTG

Method

450ml baffled flasks of Terrific Broth had 50ml TB phosphate buffer, ampicillin to a final concentration of 100µg/ml and 5ml of overnight culture added under sterile conditions. Flasks were incubated in a shaking incubator at 37°C and 180rpm.

At regular intervals, samples were taken from flasks and the OD of cells at 500nm was recorded using a Shimadzu UV Visible spectrophotometer, against a sample blank, taken prior to culture addition to the flask of terrific broth. When the OD of the flasks approached 1.0 absorbance units, protein expression was induced by the addition of IPTG to a final concentration of 1mM. Protein expression was allowed to continue in the incubator at 37°C and 180rpm for four hours. Cells were harvested using a Beckman J-6B centrifuge, with a JS-4.2 rotor, at 4000rpm for 30 minutes. The supernatant was removed and cell pellets frozen at -20°C.

2.3.6 Folding and purification of HRP-C*

Reagents

Resuspension of harvested cell pellets:

Resuspension solution: 50mM Tris pH 8.0, 10mM EDTA, and 1mM DTT

Lysozyme

Lysis buffer: 2M Urea, 20mM Tris pH 8.0, 1mM EDTA, 1mM DTT, (1% v/v Triton X-100)

Protein suspension solution: 50mM Tris pH 8.0, 6M Urea, 1mM EDTA, 1mM DTT

Bovine serum albumin

Bradford reagent (Biorad)

Folding:

Fold solution: 2.1M Urea, 50mM Tris pH 8.3, 8mM $CaCl_2$, 0.7mM oxidised glutathione

Haemin in 0.1M NaOH

Dialysis buffer: 20mM sodium acetate pH 4.3 containing 1mM $CaCl_2$

FPLC buffers: 20mM sodium acetate pH 4.3 (buffer A) and 20mM sodium acetate pH 4.3 containing 1M NaCl (buffer B)
5mM MOPS buffer pH 7.0

Method

Cell pellets were resuspended in a solution containing 50mM Tris pH 8.0, 10mM EDTA, 1mM DTT. The cell suspension was lysed using a combination of addition of 2mg/ml lysozyme and subsequent sonication using an ultrasonicator (B. Braun Biotech International, Sartorius Group). The cell lysate was centrifuged at 27,000 x g for 30 minutes using a Beckman J2-21 series floor-standing centrifuge with a JA-20 rotor.

Pellets were resuspended in a lysis buffer and allowed to stand in this solution for 15 minutes before centrifugation at 15,000rpm for 30 minutes and the pellets retained. The process was repeated. Pellets were then resuspended in a solution that did not contain the detergent Triton X-100. The protein solution was centrifuged again at 27,000 x g for 30 minutes and the pellets retained.

The resultant pellets were resuspended in a solution of 50mM Tris pH 8.0, 6M Urea, 1mM EDTA, 1mM DTT. This solution was centrifuged at 15,000rpm for 30 minutes and the supernatant retained.

Protein concentration of the supernatant was calculated using a standard curve of known BSA concentrations and 1ml of 1:5 Bradford reagent per cuvette, with absorbance read at 595nm using the spectrophotometer. The protein was diluted to approximately 2mg/ml using the protein suspension solution.

Protein was folded overnight at 4°C in a solution of 2.1M Urea, 50mM Tris pH 8.3, 8mM $CaCl_2$, 0.7mM oxidised glutathione. The following day, haem was added to the fold solution in the form of haemin to a final concentration of 6µM. Haem was left to incorporate into the folding protein for 24 hours at 4°C.

Folded protein was concentrated using a combination of spiral-wound concentrator

(Amicon), first equilibrated with 2M urea, and then further concentrated to a volume of approximately 30ml using an Amicon stirred cell concentrator.

Concentrated enzyme was then dialysed overnight at 4°C against 20mM Sodium Acetate pH 4.3 containing 1mM $CaCl_2$.

Folded protein was purified using a Mono S ion exchange chromatography column pre-equilibrated with a buffer of 20mM Sodium Acetate pH 4.3, 1mM $CaCl_2$. Correctly folded protein was eluted using a salt gradient of 0-1M NaCl in sodium acetate buffer.

For storage, the eluted protein was gel filtered using a PD-10 column into a solution of 5mM MOPS pH 7.0 and beaded in to liquid nitrogen. Protein was stored at -80°C until required.

Characterisation of new protein variants

2.4.1 General species characterisation

Species characterisation of all variants was carried out using either a Shimadzu UV-1601 or UV-2401PC UV/Visible spectrophotometer at room temperature.

2.4.1.1 UV/Visible spectra

UV visible absorption spectra of each enzyme variant were taken between 250nm and 750nm, blanked against 5mM MOPS pH 7.0. These were compared against the wild type HRP-C* spectrum to establish mutational variations in the haem environment.

2.4.1.2 Standard ABTS assay

Reagents

2,2h-azinobis-(3-ethylbenzthiazoline-6-sulphonic acid) (ABTS)

Phosphate citrate buffer, pH 5.0 μ=100mM

H_2O_2

Varying μM enzyme variants

Method

0.3mM ABTS, 2µM H_2O_2 and 5mM phosphate citrate buffer at pH 5.0, was added to a plastic cuvette. The reaction was started by the addition of enzyme variant.

Formation of the green ABTS radical was followed at 414nm until substrates had been depleted or three minutes, whichever was the shorter.

2.4.2 Substrate binding

Spectra between 250nm – 750nm were recorded (except where indicated).

Difference spectra were collected to determine the effect of substrate binding to the enzyme variants over this wavelength range in the absence of peroxide.

All substrates were dissolved in methanol (except naphthalene, which was dissolved in acetonitrile). Difference spectra of enzyme with methanol/acetonitrile present were taken to ensure this would not affect the behaviour of the enzyme at high concentrations.

2.4.3 Substrate oxidation

2.4.3.1 Naphthalene oxidation

This assay was adapted from Kluge *et al*. (2007/2009).

Reagents

1mM H_2O_2

1mM Naphthalene dissolved in acetonitrile

250µl Potassium phosphate buffer, pH 7.0

5µM enzyme variant

ddH_2O to 500µl total volume in cuvette

Method

Naphthalene stock was freshly prepared in acetonitrile, ensuring that the

concentration of stocks were not altered due to the partial volatilisation of naphthalene (Kluge *et al*. (2007)).

The reaction, carried out in a closed quartz cuvette, was started by the addition of 5µM enzyme variant and inversion to mix, before rapid insertion in to the spectrophotometer. The assay was followed at 303nm, tracking the progress of formation of 1-naphthol over the assay period (5-20 minutes). Quoted apparent ε303 nm for 1-naphthol is 2,030$M^{-1}cm^{-1}$ (Kluge *et al.* (2007)).

2.4.3.2 Pre steady state – naphthalene oxidation

Stopped flow experiments were carried out at 25°C using an Applied Photophysics Ltd SX18MV stopped flow spectrophotometer, using the conditions above. Data was collected using Xscan and analysed using Applied Photophysics Ltd's Pro-Kineticist v1.06 (2001).

The enzyme was preincubated with naphthalene to an extent where it was in excess of 80% saturated with naphthalene. This solution was then subject to rapid mixing with hydrogen peroxide and the resultant reaction followed through rapid scans between 300 and 750nm, and using single wavelength scans at 400nm.

2.4.3.3 GC/MS detection of naphthalene oxidation products

GC/MS analysis of products of naphthalene oxidation was carried out based on the method of Kluge *et al* (2007).

Reagents

250µl 1M hydrochloric acid

250µl chloroform

Method

Reactants were acidified with hydrochloric acid, chloroform was added to the reaction and the solution mixed vigorously.

The solution was allowed to settle and the chloroform layer extracted with a micropipette. The chloroform extract was retained for GC/MS analysis, with the remainder discarded.

Samples were injected into the GC/MS with an injection temperature of 250°C, with helium used as the carrier gas.

A splitless programme was chosen to allow greatest sensitivity to be achieved, with a 20 minute timecourse run from a starting temperature of 165°C, raising in increments of 2°C.

Masses were recorded for all compounds in the solution, with specific emphasis placed on M_r 128 and 144, the mass of naphthalene and naphthol respectively.

2.4.3.4 Sulphide oxidation

Reagents

50µM thioanisole, dissolved in methanol

1mM H_2O_2

Phosphate/citrate buffer, pH 7.0

1µM enzyme variant, exchanged into phosphate/citrate buffer, pH 7.0

Method

Oxidation of thioanisole was observed by a change in the uv/visible spectrum of the variant, in particular a drop in absorbance in the region of 254nm.

Reaction was started by addition of the enzyme variant. Repeat spectra were taken every minute between 750nm-220nm, until a haem bleaching effect (observed previously in Ngo, 2007) occurred.

X-ray crystallography

2.5.1 Crystallisation trials

Crystallisation trials were set up using 24 well Linbro plates (Molecular Dimensions Ltd.). Well solution varied but each well contained either 0.5ml (pre-prepared screens) or 1ml of solution. Each well was sealed using silicon high vacuum grease (Dow Corning) and a silicon cover slip (Molecular Dimensions Ltd.). The hanging drop vapour diffusion method was used to precipitate protein, using 1µl of concentrated protein solution (between 5 and 15mg/ml protein) and 1µl of well solution in a single drop on the cover slip.

Pre-prepared crystal screens were purchased from Molecular Dimensions Ltd. and were used as sparse matrix attempts to produce initial crystal "hits" which could be later refined. Each screen was prepared as previously described, using 0.5ml of the pre-prepared solution per well. The screens used in the production of this thesis were: PACT Premier Screens 1 & 2, Structure Screens 1 & 2 and the Clear Strategy Screen. Each boxed screen contained 48 or 50 different solutions that cover a broad range of conditions that have previously been shown to produce a high yield of starter crystals. (PACT screen developed by Newman *et al.* (2005), Structure Screens based on Jancarik & Kim (1999), Clear Strategy Screen developed by Brzozowski and Walton).

Crystallisation trays were incubated at 4°C and regularly checked using a Nikon stereoscope to view drops.

Optimisation screens were set up around conditions that yielded microcrystalline precipitate "hits". These optimisation screens varied precipitant concentration, salt concentration or pH by small amounts in order to attempt to improve on the crystals formed in the initial screens.

2.5.2 X-ray data collection

Crystals were selected for suitability for data collection subjectively, a process which

was dependent on the size, clarity and colour of the crystal. Particularly in drops containing multiple crystals, a variety of crystals were mounted in the X-ray beam to ascertain the best potential for high-resolution data collection.

Prior to crystals being prepared for data collection, an appropriate cryoprotectant solution was determined. This was established so that the crystals could be X-rayed in a solution that would not create disruptive ice crystals when placed in the cryostream at 100K. Appropriateness of cryoprotectant was ascertained by first checking the precipitant solution the crystals were grown in for ice crystal formation. If this solution did not form ice crystals at 100K then the preparation process continued using just the well solution. If ice crystals were formed, well solution was systematically diluted with suitable solutions and tested until a combination was found that did not produce ice crystals.

Data were collected in-house from suitable crystals using a Rigaku RU-H3RHB copper rotating anode generator producing X-rays of 1.542 Å wavelength. The X-rays were focused using Osmine Confocal Max-Flux mirrors through the crystal and diffraction data were collected using an R-axis IV++ imaging plate.

2.5.3 Data processing

Initial data files were processed using MOSFLM (Leslie, 1992) and then processed in batch using programmes in the CCP4 suite (Collaborative Computational Project, Number 4, 1994). These included Scala, RefMac5, and others, detailed in later chapters. Final fitting of models to the electron density maps produced was performed in Coot (Emsley and Cowtan, 2004). Data was subject to checking procedures using SFCheck (Vagin *et al.,* 1999*)* and ProCheck (Laskowski *et al,* 1993) in which phi and psi angles and bond lengths were checked for plausible accuracy.

Finished data were prepared in a format in which structures can be submitted to the RCSB Protein Data Bank (Berman *et al*., 2000).

Chapter 3

HRP peroxygenase variants: design, production and general characterisation

3.1 Introduction

Three new HRP variants were produced during this work, their design based on the previous iteration of HRP "peroxygenase" variants produced by Dr. Emile Ngo in his doctoral work entitled *"Construction and characterisation of horseradish peroxidase mutants that mimic some of the properties of cytochromes P450"* (patented in 2007 with patent number WO/2007/020428). His variants, particularly R38A/F41A/H42E (AAE) and R38A/F41A/H42E/N70H (AAEH), built on earlier work from within the group, particularly unpublished work by Schneider and Smith. The work developed from evidence that showed how active site engineering of HRP could potentially generate new activities, in this case peroxygenase activity, not present in the wild-type enzyme. This Chapter discusses the steps taken in the design and production of the three new variants.

3.1.1 Previous work

A wealth of previous work has examined aspects of the HRP catalytic centre, with work by the groups of Ortiz de Montellano and Tanaka having particular reference to the work described in this thesis. The variants described here and in the previous work were designed to create peroxygenase or oxygen transfer activity in HRP-C*, an activity not seen naturally in the wild type enzyme, but which occurs in choloroperoxidase (CPO), as discussed in Chapter 1, and recently identified in *Agrocybe aegerita* peroxidase.

CPO is, as discussed previously, a peroxidase that fits neither in the plant nor mammalian superfamilies. This "indistinct" peroxidase has unique catalytic properties amongst peroxidases and could be better described as a haem peroxidase-cytochrome P450 hybrid, having properties common to both groups of enzymes (Sundaramoorthy *et al,* 1995).

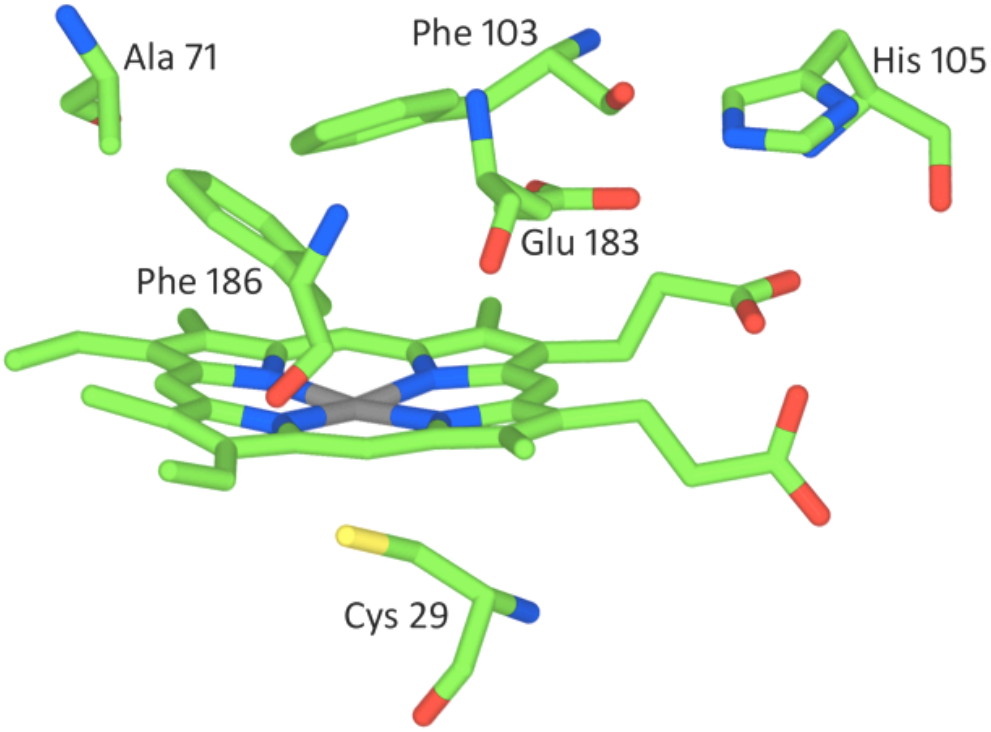


Figure 3.1 Chloroperoxidase active site
Key residues are highlighted in the CPO active site. Particularly of note is the glutamate residue at position 183, not found in the active site of other peroxidases. Also of note is the cysteine residue at position 29, which forms the 5th co-ordination position with the iron centre of the haem, where histidine is found in this position in plant peroxidases.

Crystallographic structure determination of the chloroperoxidase model (Sundaramoorthy *et al.*, 1995) gave indication that a glutamate residue present in the active site may be contributing to the unique activity of this protein molecule. Figure 3.1 shows a representation of the CPO active site, with the glutamate residue highlighted.

Following the publication of the CPO structure in 1995 and as a result of the implication that the glutamate residue was responsible for the unique catalytic activity of the CPO enzyme, a series of HRP-C* variants have been created to probe the role of residues within the HRP-C active site. Initially these examined the function of the glutamate but latterly have attempted to reproduce the function of chloroperoxidase in HRP-C.

3.1.2 HRP variants at position 42

Investigations by Tanaka and co-workers initially examined the glutamate residue described in the chloroperoxidase structure, suggesting that this residue could be supporting oxygen bond cleavage in place of the histidine shown to be responsible for peroxide cleavage in other peroxidases (see Table 1.1 with the conserved histidine, amongst other conserved residues, in plant peroxidase active sites) (Tanaka *et al.*, 1996). This work identified that it was possible to form Compound I at a rate slower than, but comparable to, wild type HRP in the H42E HRP variant. This was a significant step since previous substitution of the histidine at position 42 resulted in a dramatic drop in the enzyme's ability to form Compound I, for example by 10^6 in the case of H42A (Newmyer and Ortiz de Montellano, 1996). The increase in peroxide required to form Compound I in variants in which H42 had been replaced by an hydrophobic amino acid resulted in a haem bleaching effect, however this was not observed in the H42E variant (Tanaka *et al.*, 1996).

3.1.3 HRP-C* variants that enlarge the catalytic site

Variants R38A/F41H/H42E (AHE) and R38H/F41A/H42E (HAE) (Schneider and Smith, unpublished), further examined the roles of residues unique to the active site of

chloroperoxidase and the more “open’ nature of the CPO active site, compared to the relatively closed architecture of the traditional plant peroxidases. As seen in Figure 3.1, the CPO active site is significantly more open than HRP-C, allowing substrates access to the haem centre, unlike in the plant peroxidases where catalytic activity takes place at the haem edge.

This work, in which catalytic site residues were engineered to be more chloroperoxidase-like, produced variants that explored the role of a glutamate residue introduced in the earlier work, and attempted to create a more open active site pocket in which the ferryl oxygen of Compound I is directly accessible to aromatic substrates.

As well as the apparent importance of the introduced glutamate residue in the function of the HRP-C* peroxygenase variants, the significance of a histidine residue within the active site was probed. In earlier work, the most successful chloroperoxidase mimic had a histidine residue introduced at position 38, an alanine at position 41 and a glutamate at position 42 (variant HAE). This suggested that the introduction of a histidine residue and a structurally adjacent glutamate, residues that are seen in the chloroperoxidase active site, combined with a reduction in sterically hindering residues within the HRP active site cavity could dramatically increase the peroxygenase-like activity of HRP-C* (Schneider and Smith, unpublished).

In particular, the HAE variant showed an increase in the specific activity of styrene epoxidation, an activity not seen in typical peroxidases; the activity was in excess of 15 fold higher than the H42E single substitution and greater than 12,000 fold higher than wild type HRP-C*. This feature, which was further developed during the work of Ngo, is explored in more detail in Chapter 4.

HRP-C* variants R38A/F41A/H42E (AAE), R38A/F41A/H42E/N70H (AAEH) and R38A/F41A/H42A/N70H (AAAH) were created by Emile Ngo during his doctoral work. These variants enlarged the catalytic site, further exposing the haem centre and allowing substrates greater access to the ferryl oxygen. These variants had some properties of CPO, particularly with newly engineered ability to perform sulphoxidation

reactions.

An interesting observation was that both AAEH and AAAH appeared to have haem centres that were predominantly 6-coordinated and low spin at pH 7.0, where wild type HRP-C* and other similar variants present substantially 5-coordinated high spin haem between pH 5.0 and 9.0 at room temperature. The significance of this low spin 6-coordination will be discussed later in this chapter and in Chapter 4, since this observation has implications for the mechanisms behind the new activities of the enzyme variants.

3.1.4 Reasoning for variants produced in the current work

The previous work shows the importance of the introduced glutamate residue in facilitating the acid-base mechanism of dioxygen cleavage resulting in Compound I formation, as demonstrated by the differing activities between variants AAEH and AAAH. This evidence shows the necessity of the acidic residue within the active site, facilitating dioxygen bond cleavage, in combination with the enlarging of the haem pocket by removal of large residues.

The variants detailed and characterised in the following two chapters are a logical iteration of structure-based design of HRP-based peroxygenase type variants. Based on the most active of the previous generation of variants, the role of the acidic glutamate in the active site is probed, with new variants designed that alter its position in the catalytic site. Specific residues were also targeted in the HRP cavity in the previous work, retained in this work, to try to introduce the more open active site configuration seen in chloroperoxidase, thus removing the large, sterically hindering residues of HRP that prevent direct access of substrates to the ferryl oxygen during the catalytic process.

The previous work identified the glutamate at position 42 having significant impact on the ability of the engineered enzyme to perform chloroperoxidase-like activity, namely oxidation of the sulphide thioanisole. Variants AAE and AAEH both performed thioanisole sulphoxidation, with AAEH giving 100% stereospecificity of reaction,

producing the (R)-sulphoxide of thioanisole, a crucial feature for future chemical applications and significant since this form and stereoselectivity of thioanisole sulphoxidation is seen in CPO (Wojaczyńska and Wojaczyńska, 2010). Variant AAAH, where an alanine residue rather than a glutamate replaced the histidine residue at position 42, was capable of sulphoxidation of thioanisole, but at a reduced rate and with only 81% stereospecificity. Previous HRP-C* variants gave a mix of R and S forms, with variant HAE giving 48:52% R:S sulphoxide forms. The high level of enantioselectivity implies a more open haem cavity and confirms the effectiveness of producing a chloroperoxidase-like behaviour in HRP-C*.

The work described in Chapters 3 and 4 examines a new generation of variants in which the acidic residue is retained in the active site, but position of the charge in the site is varied, either by reducing the length of the side chain (replacing the glutamate for an aspartate residue whilst retaining the expanded active site architecture as in the previous work) or by moving the glutamate residue to other positions in the haem cavity.

Variant identifier	Location of site directed mutations	Created by
HAE	R38H/F41A/H42E	F. Schneider
AHE	R38A/F41H/H42E	F. Schneider
AAE	R38A/F41A/H42E	E. Ngo
AAEH	R38A/F41A/H42E/N70H	E. Ngo
AAAH	R38A/F41H/H42A/N70H	E. Ngo
EAA	R38E/F41A/H42A	S. Pannell (present work)
AEA	R38A/F41E/H42A	S. Pannell (present work)
AAD	R38A/F41A/H42D	S. Pannell (present work)

Table 3.1 HRP-C* 'peroxygenase' variants as described in this work.
Variant identifiers are used for ease of reading. Where important to the discussion the location of the mutations will be reiterated.

Table 3.1 summarises the variants of particular interest to this work, including both past and present work. For simplicity, these will continue to be referred to by their three letter identifiers rather than reciting the full location of the mutations, though these will be highlighted where crucial to discussion.

3.2 Results and discussion

3.2.1 Production of HRP variants

The three newly designed variants, aimed at further probing the nature and function of residues present in a redesigned HRP catalytic site, were produced using site directed mutagenesis of the pFLAG-HRP gene construct. This process, as described below, allowed production of variant protein through expression in *E.Coli*. Protein was extracted from inclusion bodies, essentially as described by Smith *et al.,* for the first wild type HRP-C* expression and folding. The synthetic HRP-C (HRP-C*) gene was constructed to favour codons commonly found in *E.Coli* systems (Smith *et al.,* 1990). Challenges in the refolding of extracted protein with these variants were encountered, despite following a well-established folding procedure. These issues are described, with notes on alternative methods attempted to increase protein yield.

Fully folded and active protein was produced for two of the three variants designed, with yields being achieved that were comparable to the previous generation of variants, however the folding process is not as successful as wild type HRP, observations on which are detailed below.

3.2.1.1 Polymerase chain reaction

Polymerase chain reactions (PCR) were carried out as described in Chapter 2.2.2, using the designed primers shown in Table 2.1. Primers were synthesised with specific base substitutions that would allow the desired residues to be introduced with minimum mismatching of bases between primers and the pFLAG HRP plasmid template.

3.2.1.1.1 Primer redesign

Primers 3 and 4, designed to produce the variant AEA, shown in Table 2.1 were completely redesigned following repetitive failure to produce any PCR product despite variations of annealing temperatures and changes in conditions that would normally result in primer binding. This included dropping the annealing temperature to 45°C, a level that would, in the majority of cases, cause indiscriminate binding of primer to DNA resulting in incorrect PCR products, however no product was seen. All conditions

and reagents were checked, and where possible replaced, to ensure that solutions used were not contaminated or otherwise compromised. Controls were produced using the same conditions and reagents that subsequently yielded positive results, confirming that there was a problem with the initial primers.

Upon re-examination it was discovered that the original sequence, shown in Table 2.1, was palindromic and potentially was forming a stable hairpin loop, which would explain the difficulties in obtaining any PCR product. The potential loop is shown in Figure 3.2, which illustrates the high quantity of palindromic sequence thought to be the primary cause of PCR failure with these primers. The overlapping regions would mean that, rather than annealing to complementary sequence in the pFLAG HRP-C template, there would be a higher likelihood of the primer binding to itself and hence unusable. Redesigned primers are shown in Table 3.2.

Redesigned primers yielded the expected products (as determined by agarose gel electrophoresis, described previously in Chapter 2 and detailed below) with a high degree of specificity. The products of the PCR reaction with the redesigned primers were taken forward alongside the original products from reactions using primers designed for variants AAD and EAA.

Variant	Primers (5'-3')	Unique restriction enzyme site	Approx. Tm
R38A/F41E/H42A (AEA)	3. T**GAAGCA**CAGTGCTGTAATGGTTGC 4. TCAAG**TGC**TAATATTGAAGCAGCGATCCTGG	*BspH1* - *TCATGA*	64°C 66°C

Table 3.2 – Redesigned primers for new HRP-C* variant AEA.
Codons in bold represent codon changes that cause amino acid substitutions. Codons underlined represent silent mutations that allow the introduction of an additional restriction enzyme site.

3.2.2 Site-directed mutagenesis of HRP

For clarity, only the process of the site directed mutagenesis, expression and folding of one variant, R38A/F41A/H42D – AAD, is followed here. Other variants were produced

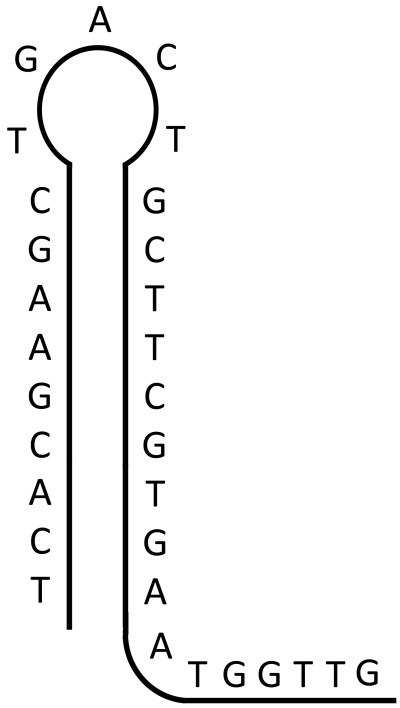


Figure 3.2 – Primer stem loop
Primers originally designed for variant AEA had to be redesigned following repeated failure in PCR reactions, despite changes in conditions expected to give rise to indiscriminate product.
Upon re-examination of the primers it was discovered that the forward primer had a considerable region of palindromic sequence, as demonstrated in this figure, that was not recognised in the original design of the primers.
Subsequent redesign of primers yielded the desired results.

in the same way, with differences only occurring in the annealing temperatures and the restriction enzymes used to identify potential correct variants, particulars of which are recorded in Table 2.1 and 3.1. Any other slight alterations for other variants are remarked upon where relevant.

Initial PCR reactions showed a high level of non-specific binding of the AAD primers when a 5µl sample of the PCR product was subject to agarose gel electrophoresis in order to identify whether the reaction had yielded the specific fragment required. The agarose gel, when stained with ethidium bromide and visualised using a UV light source, showed that the expected 6.3kb PCR product was contaminated with smaller DNA fragments. Annealing temperatures were increased from 54°C to 57.5°C which successfully eliminated the non-specific binding. Figure 3.3 is an example photograph, taken using a camera with a UV filter, of the correct PCR product at approximately 6kb on the right, with a 1kb DNA ladder (New England Biolabs) for scale on the left.

PCR products were purified, ligated and transformed into library efficiency chemically competent DH5α *E.coli* cells as detailed in section 2.2.4-6, this standard procedure was performed without any additional steps. Transformed cells were spread on agar plates inoculated with ampicillin, ensuring that only *E.coli* cells containing the correct pFLAG HRP plasmid were able to grow. In all three cases, large numbers of colonies were produced, indicating that the plasmid had successfully transformed into the DH5α cells. Up to 12 colonies were selected from the overnight-cultured agar plates and each colony incubated separately overnight in 5ml LB broth. DNA was extracted from these cells using the Qiagen miniprep kit using the established procedure detailed in the handbook supplied with the miniprep kit.

3.2.3 Confirmation of mutations

Restriction enzyme digests were performed on the DNA extracted from each culture as the first step in determining whether the plasmid DNA obtained was the correct sequence. Unique restriction enzyme sites were designed in to the new primers as a way of quickly identifying colonies that did not have the correct mutations in DNA, rather than sequencing each potential colony. New restriction enzyme sites resulted

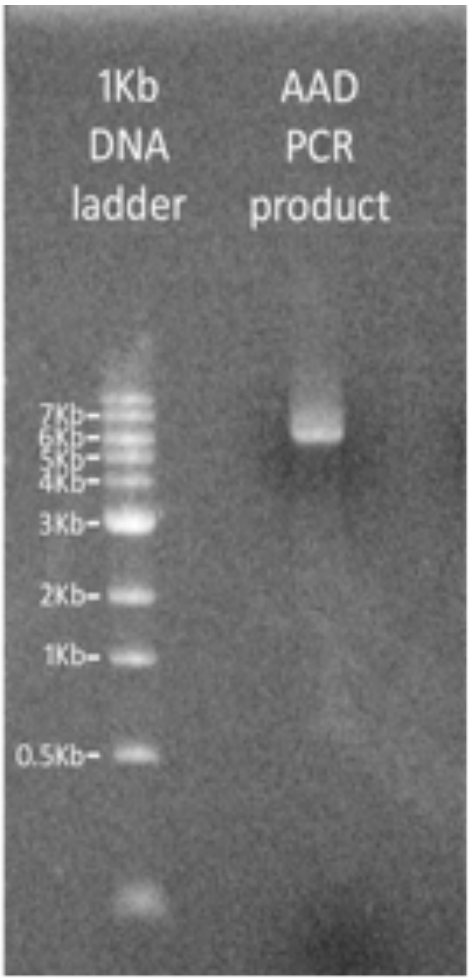


Figure 3.3. PCR product of HRP-C* variant AAD
PCR product made with primers introducing point mutations to DNA. The band shows PCR product at the correct size expected for pFLAG HRP-C* which is assumed to be correct until identified by plasmid sequencing.

in differences seen when digested plasmid DNA was subject to agarose gel electrophoresis, with fragments of DNA of differing lengths. Calculation of restriction enzyme sites and expected fragment lengths prior to primer synthesis enabled restriction enzymes to be selected that gave clear differences between correctly mutated and non-correctly mutated DNA.

Whilst the technique of selectively designing restriction enzyme sites does not guarantee that samples that appear correctly mutated do not have subtle differences in the DNA sequence that are not picked up, it enables DNA that does not contain the new restriction enzyme site, and hence the desired mutations, to be discarded. This means that selection of potentially correctly mutated DNA is able to be made prior to sequencing and hence reduces the need to sequence multiple negative colonies. This was performed both as a cost and time saving exercise.

As described in Table 2.1, the unique restriction enzyme site created by the correct introduction of the AAD point mutations was a *Sap1* site. In the wild type pFLAG HRP-C*, digestion with *Sap1* would show a linear fragment on an agarose gel as there is only one *Sap1* site present in the pFLAG HRP-C* vector, at position 3334. The introduced site, at position 214 in the HRP-C* gene, would give two fragments, of approximately 3.1Kb and 3.0Kb. If the mutation had been incorporated correctly, a large band should be visible at approximately 3Kb, as the resolution of the gel electrophoresis is not such that there would be an obvious visible distinction between 3Kb and 3.1Kb. DNA that had not incorporated the mutations correctly would show one band at approximately 6kb.

Figure 3.4 shows the results of the restriction enzyme digests on the ten clones isolated after mutagenesis with the AAD primers. The visible distinctions between potentially correct colonies and colonies that do not contain the correctly mutated plasmid DNA is evident. Lanes 1, 3 and 5 have a clear band visible at approximately 3Kb, however they also each contain a strong band at approximately 6Kb. This may indicate the correct mutation in the plasmids but that the restriction enzyme has not

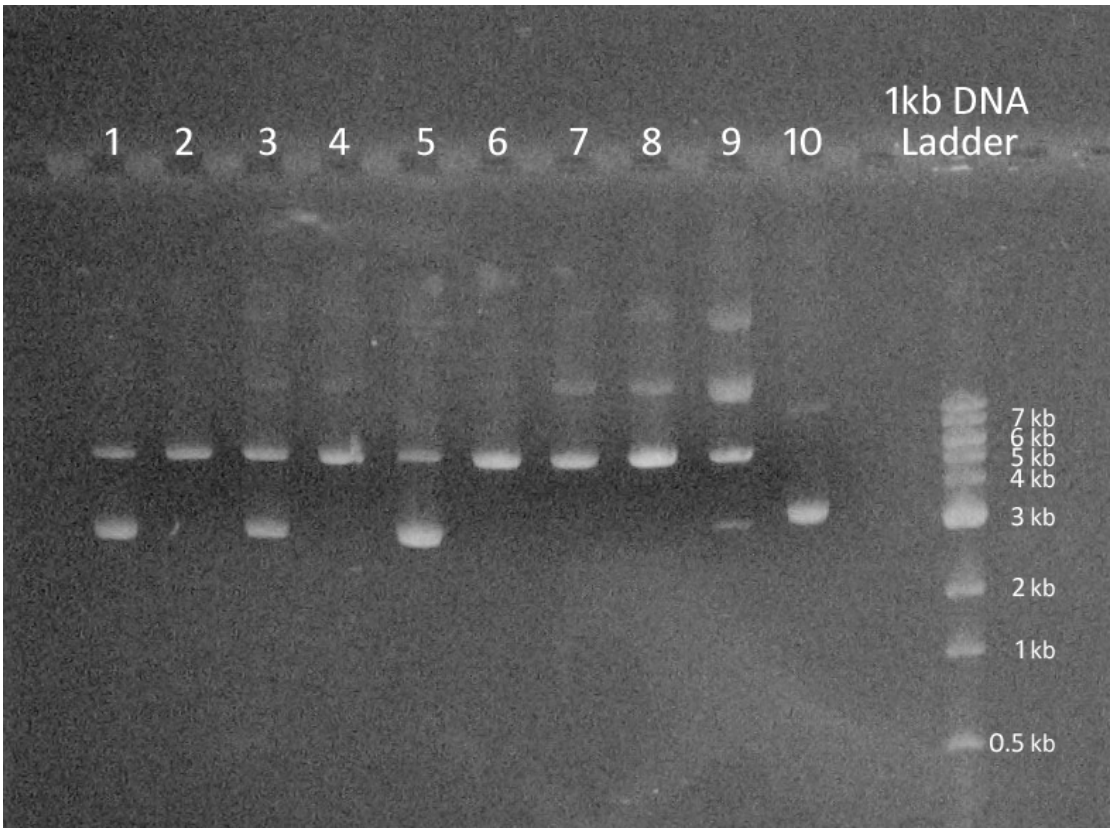


Figure 3.4 Restriction enzyme digest of pFLAG HRP plasmid variant AAD.
DNA, extracted from 5ml overnight expression of colonies in LB media, was incubated with restriction enzyme *Sap1* for one hour. The predicted correct product would produce two DNA fragments at 3Kb and 3.1Kb, which would not be able to be separated on the low resolution 1% agarose gel, however a thick band at approximately 3Kb would give an indication that the plasmid potentially contained the correctly mutated product.
Lanes 1, 3, 5 and 10 show the expected band at approximately 3Kb, however only lane 10 shows an absence of undigested plasmid at 6Kb. Lane 9 shows a faint band at 3kb, possibly demonstrating that the product is correct but the digestion may not have occurred correctly. Lanes 2, 4, 6, 7 and 8 do not show any digested product at 3kb suggesting that the plasmids do not contain the required point mutations.

completely digested. Lane 10 shows a strong band at approximately 3Kb, without a band at 6Kb. Lane 9 shows a faint band at approximately 3Kb, with a much stronger band at approximately 6Kb. This may possibly be a "positive" colony, however since lanes 1, 3, 5 and 10 show strong bands at approximately 3Kb, DNA from the clone in lane 9 was not characterised further. Lanes 2, 4, 6, 7 and 8 do not show any band at 3Kb, suggesting that these plasmids do not contain the desired point mutations that would have introduced the additional restriction enzyme site. Lanes 3-10 also have weak smeared signals above the 6Kb band, with more defined bands in lanes 8-10. This gives a potential mutagenesis efficiency of 50%, though the correct point mutations must be confirmed by DNA sequencing.

The colonies producing DNA shown in lanes 5 and 10 of Figure 3.4, with strong bands in the expected range for a "positive" result, were individually inoculated into 50ml LB broth and grown overnight at 37°C. These were prepared using Promega's Wizard Midi Prep kit and DNA concentration calculated as in section 2.2.7. DNA from colony 10 was diluted to 100ng/µl and posted to Lark Sequencing Technologies, Inc. (now Cogenics, a division of Beckman Coulter) for sequencing.

Sequencing was carried out using pFLAG forward and reverse primers, ensuring that the full sequence of the pFLAG-HRP construct was sequenced. This allowed clear sequencing in the region of the mutation sites, with confirmation from both forward and reverse primers of the changes in the DNA.

Two further colonies of variant AAD were sequenced, as the full plasmid sequencing of the DNA extracted from the growth of Colony 10 showed a non-silent mutation elsewhere in the HRP-C* DNA sequence. DNA from Colony 3 was confirmed by the initial positive gel electrophoresis result and subsequent gene sequencing validation of the initial result, combined with the assurance that the complete HRP-C* gene is correct and no other mutations have been generated during the PCR reaction. Silent mutations that did not alter the final amino acid sequence of the protein would have been acceptable, however non were seen in the AAD, AEA or EAA DNA sequences where the new mutations had been introduced. Full sequence of the AAD HRP-C* is

shown in Appendix 1.

Correctly sequenced DNA was subsequently transformed into W3110 cells for protein expression. Where DH5α cells are optimized for DNA replication, W3110, and its related strain HW1110, are *E.coli* fermentation strains and therefore optimised for high yield protein expression. A comparative small-scale expression test between the W3110 and HW1110 cell lines using the EAA variant plasmid showed little difference in yield between the two strains in expressing the new variant plasmid. This previous reliability of expression with the W3110 cells led to this cell line being chosen for large scale growth and protein expression.

3.2.4 Protein expression of new variants

Small scale protein expression testing also ensured that the new variant plasmids were producing sufficient protein of the approximate expected molecular weight for HRP-C*. Although correct protein folding couldn't be determined at this stage, the small-scale expression step ensured that, prior to larger scale expression, protein of the anticipated approximate correct molecular weight was being produced.

Large-scale protein expression was carried out at 37°C with 500ml of terrific broth culture medium in each of up to 10 2.5l baffled flasks, ensuring adequate aeration of the *E.coli* culture throughout the growth. Growth was started by addition of 5ml *E.Coli* overnight growth culture, following addition of antibiotic and phosphate solution to the terrific broth and U/V irradiation of the flasks to avoid contamination of the growth stock. Induction of protein expression was triggered by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) when the OD_{500nm} of the growth solution, measured against a blank of terrific broth without culture medium, reached 0.8 absorbance units, this absorbance was typically reached between three and four hours after the growth was started. IPTG served to bind to the *lac* promoter region of the pFLAG-HRP, inducing protein expression. Final concentration of IPTG in the culture medium was 1mM. Induced flasks were incubated for a further 3.5 hours before harvesting the cells by centrifugation, following which the cell extract pellets were frozen at -20°C.

The three new variants all were successfully expressed using the standard culture growth and expression protocols. Growth yields varied but, as with previous chloroperoxidase-like variants, yields were lower than those produced by wild type HRP-C* expression. Although fermenter-scale growth may have given better yields of protein per litre of culture medium, this technique for protein expression was not carried out during this work.

3.2.5 Folding of new variants

Following protein expression, folding was carried out according to the procedure described in Chapter 2. Previously, HRP-C* peroxygenase variants had been reported to be more sensitive to the folding process, with precipitate forming during the process that potentially can cause further aggregation of protein. Measures were taken to remove precipitate where possible, in between steps of the folding process.

Variants EAA and AAD both folded correctly in the standard conditions so no change in the conditions was made, with the exception of the late haem addition step recommended by Cali (2009), following experimentation in this step of the procedure during his doctoral work. In this adjustment to the previously published method, the enzyme is allowed to fold under standard conditions (2.1M Urea, 50mM Tris pH 8.3, 8mM $CaCl_2$, 0.7mM oxidised glutathione), without any haem present in the folding solution, for 18 hours prior to a solution of haemin being added to the fold (final concentration of haemin was 6µM). Cali reported that the late haem addition step resulted in an improved yield of correctly folded protein (Cali, 2009). The late haem addition method was used exclusively during this work.

Whilst the variant AEA was found to grow normally in comparison to other previous variants in both small and large scale broth cultures, it appeared to fold exceptionally poorly during the standard, well established, HRP-C* folding process (based on Doyle and Smith, 1999). This led to high levels of aggregation of the protein despite attempts to modify the folding and purification protocol, for example by including an additional dialysis step at an intermediate pH before a short low pH dialysis prior to

cation exchange chromatography in order to reduce the stress on the folding protein caused by the changing pH. Completing the entire process at 4°C was also attempted with no discernable difference noted. No alteration to the folding protocol significantly reduced the precipitation of the AEA HRP-C*, resulting in a miniscule yield (<250µg) of protein from 5l growth. Repeated attempts with differing conditions gave similarly small yields.

It was concluded at this point that the AEA (R38A/F41E/H42A) variant, specifically with the introduced glutamate at position 41 causes instability in the protein in a way not seen previously seen for F41 variants (for example F41V by Smith *et al.*, 1992; F41L and F41T by Ozaki and Ortiz de Montellano, 1995). This instability resulted in a high proportion of protein folding incorrectly, despite protein being produced to the right expected molecular weight (as determined by SDS PAGE analysis).

At this stage no further work was carried out to probe the AEA variant and work was continued using EAA and AAD to facilitate comparison with the previous work. Variants AAE and AAEH which were known to be excellent in stereoselective sulphoxidation. In the future it would be reasonable to revisit the AEA variant to attempt to determine the causes of the instability in folding, particularly due to previous F41 variants not showing this apparent instability.

3.2.6 Purification of folded protein

Variants EAA and AAD were folded as described above, and were purified using a MonoS cation exchange chromatography column using an AKTA FPLC purifier system, the MonoS column being of negative charge compared to the net positive charge on the HRP-C* surface. Columns were pre-equilibrated using a buffer of 20mM sodium acetate, 1mM calcium chloride at pH 4.3 in order to maintain the structural Ca^{2+} ions necessary to keep the HRP-C* globular structure which can be lost at lower pH. Concentrated fold solution containing the correctly folded enzyme (as tested by crude ABTS assay in microcentrifuge tubes) was injected on to the column, with some unfolded or partially folded protein also binding to the charged column. Flow through from the protein injection to the column was retained to confirm that this did not

contain active enzyme – this was also tested using crude ABTS assays that would give an immediate indication as to whether the solution contained active HRP-C*. Flow through did not contain active HRP-C* and therefore the protocol proceeded.

Protein bound to the column was displaced by buffer containing sodium chloride which was applied at a steady gradient of between 0-100% 1M NaCl. Both variants eluted at approximately 20% 1M NaCl, in the region of the salt concentration resulting in the elution of wild type HRP-C*, giving a strong indication that variants were folded correctly and were essentially native-like in their fold. In some folds, there were smaller peaks that eluted at different salt concentrations. These were examined, both visually (as correctly folded HRP-C* variants have a characteristic red/brown colour resulting from the haem centre), UV/Visible spectra were taken and crude ABTS assays performed. None of the smaller peaks collected contained correctly folded HRP-C*, rather partially folded or misfolded variants.

A small proportion of the total protein for variants AAD and EAA, when loaded on to the ion exchange chromatography column, appeared to precipitate at the top of the MonoS column and did not enter the column correctly. This did not affect ability to isolate pure, correctly folded protein as expected at the 20% salt concentration, however the precipitate at the top of the column resulted in an increased pressure in the system. This build up of pressure required monitoring and compensating for by a reduction in the flow rate through the column depending on how much precipitate had formed. The flow rate was reduced by between 50% and 75% to maintain an appropriate column pressure without exceeding the pressure limit suggested by the column manufacturer. This precipitate appeared tightly bound to the column and three methods of removal were attempted to clear the column in preparation for further use. Firstly the column was run in reverse at low flow rate (less than 0.5ml/minute) using the standard buffer containing 1M NaCl, removing the loosely bound protein from the top of the column. This method had limited success, as much of the protein was tightly bound to the column bed. The second method involved the column also being run in reverse with a solution of 2M NaOH used instead of standard buffer. This resulted in the removal of more tightly bound enzyme, particularly when

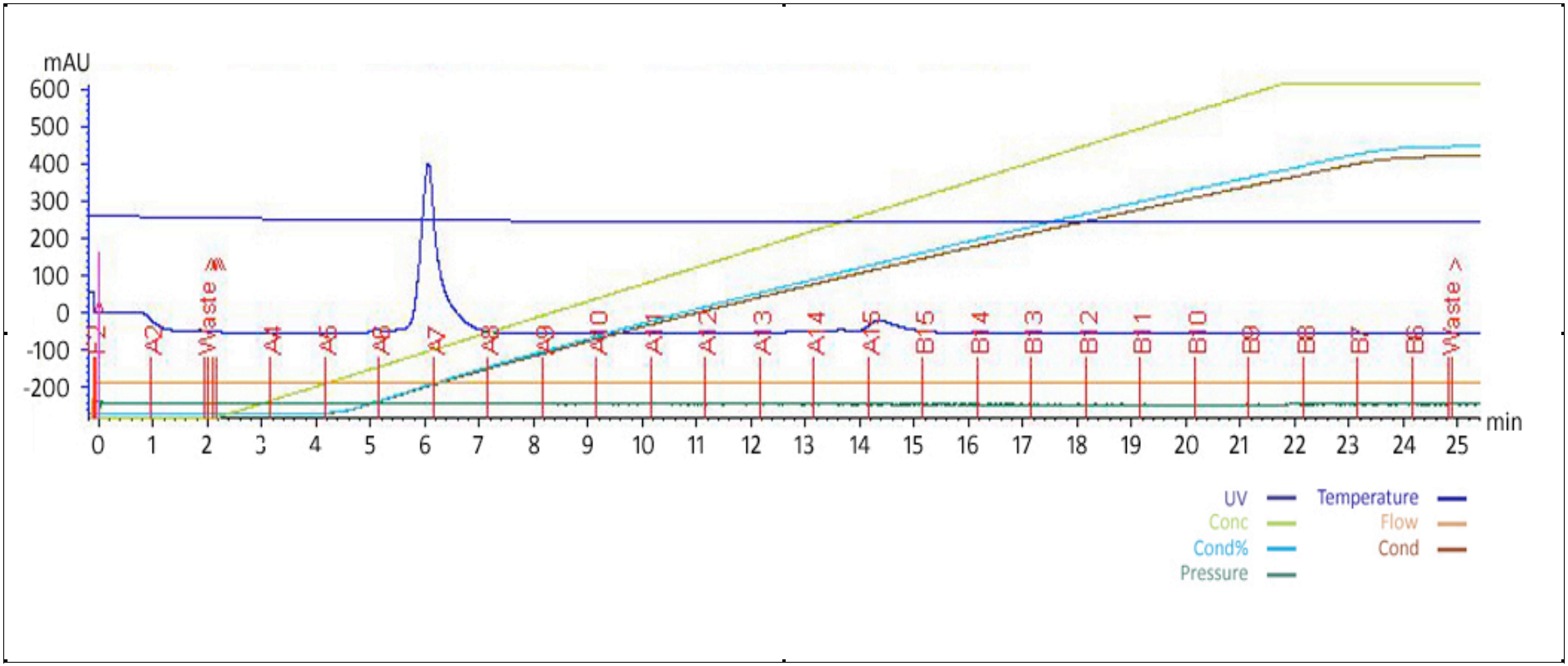
the column was injected with 2M NaOH and then incubated at room temperature for up to four hours and then run with standard buffer but without NaOH. If discolouration of the top of the column remained, indicating the presence of apparently misfolded protein since HRP-C* has a red/brown colouration due to the presence of the haem centre, the column was treated with a protein digest mixture. The column was injected, and then incubated over night at 37°C, with a solution of 1mg/ml pepsin, 0.1M acetic acid, 0.5M NaCl, as recommended by the manufacturer, to digest protein remaining on the column. After the overnight incubation, the column was run in reverse with standard buffer containing 1M NaCl to remove the pepsin digest mix. This procedure removed the precipitate at the top of the MonoS column and left it clear for future use. Whilst this procedure was not necessary following every purification protocol, it was repeated periodically to ensure the columns were cleared of any misfolded protein.

Although this characteristic of misfolding and subsequent precipitation is observed at a low level for wild type HRP, it was exacerbated in the case of the mutants. The implication that the samples were more highly aggregated with the most likely cause of this being instability caused by the change of residues in the active site.

Figure 3.5 shows a typical FPLC trace of HRP-C* variant AAD, with a salt gradient of 0-100% NaCl in solution of 20mM sodium acetate, 1mM calcium chloride at pH 4.3. A single symmetrical peak at 20% NaCl suggests that this is correctly folded native protein. The presence of a further small peak at 60% NaOH shows that, in this sample, a misfolded protein bound more tightly to the MonoS column than the correctly folded HRP-C* variant. This misfolded protein, displaced with 0.6M NaOH, represents just a small proportion of protein in solution. If this peak had been substantially larger then an error in the folding may have occurred to result in a high proportion of misfolded protein.

3.2.7 Confirmation of correct folding and purification of protein

The correct folding of the protein was subsequently confirmed by UV/Visible spectrophotometry, the spectra being compared against the wild type HRP-C*



Figre 3.5 FPLC chromatogram of AAD purification

This standard FPLC chromatogram is indicative of that seen on purification of HRP-C* variants AAD and EAA. Protein eluted from the MonoS column at approximately 20% salt (0.2M NaCl). The programme ran a gradual increase of salt from 0-100% 1M NaCl with a buffer of 1mM $CaCl_2$, 20mM sodium acetate pH 4.3. The column was pre-equilibrated with the buffer (in the absence of salt) and a steady baseline observed before the protein was loaded to the column.

A slight peak is shown at approximately 60% 1M NaCl, possibly indicative of somewhat misfolded protein which had bound more tightly to the column, hence requiring a higher concentration of salt to displace.

standard spectrum for Soret maxima and the presence of charge transfer bands (if applicable – variant AAEH does not display traditional charge transfer bands and has a shifted Soret peak due to its partial low spin character, a feature that will be remarked upon later). Figure 3.6 shows UV/Visible spectra of HRP-C* wild type and variants, against which new variants were compared in confirmation of protein folding.

Activity of the enzyme was measured on a crude basis during the folding process through the use of ABTS. This assay, as described in more detail later in the chapter, is an excellent peroxidase substrate and can be used to compare enzymes by establishing turnover number. In a crude form, the substrate, with a small amount of hydrogen peroxide, was used to establish whether the enzyme was capable of oxidising the substrate. This would give an indication of the correct folding of the enzyme and, during the first folding process, an early suggestion of the enzyme's activity. Wild type HRP-C* has a very high turnover of ABTS – in excess of $1000s^{-1}$ – whereas previous variants had much lower rates, thought to be due to the alteration of active site residues that promote electron transfer to the substrate at the haem edge.

Typical yields of protein, extracted from inclusion bodies and purified using the procedures detailed above and in Chapter 2, ranged from between 10-15mg from a growth of 5l. The concentration was determined using the UV/Visible spectrum of the enzyme using an extinction coefficient at the Soret peak calculated by the pyridine haemochrome method (described below).

3.2.8 Calculation of enzyme purity

Figure 3.6 shows sample UV/Visible spectra of the resting enzyme of EAA and AAD, normalised so that the value at the Soret peak = 1 absorbance unit. The spectra, along with peroxidase assays using hydrogen peroxide and the generated synthetic substrate ABTS (discussed in more detail later in this chapter), confirmed the correct folding of the enzymes and that the protein was active. The UV/Visible spectrum also gave indication of the purity of the enzyme, ensuring samples had not been contaminated during the folding process.

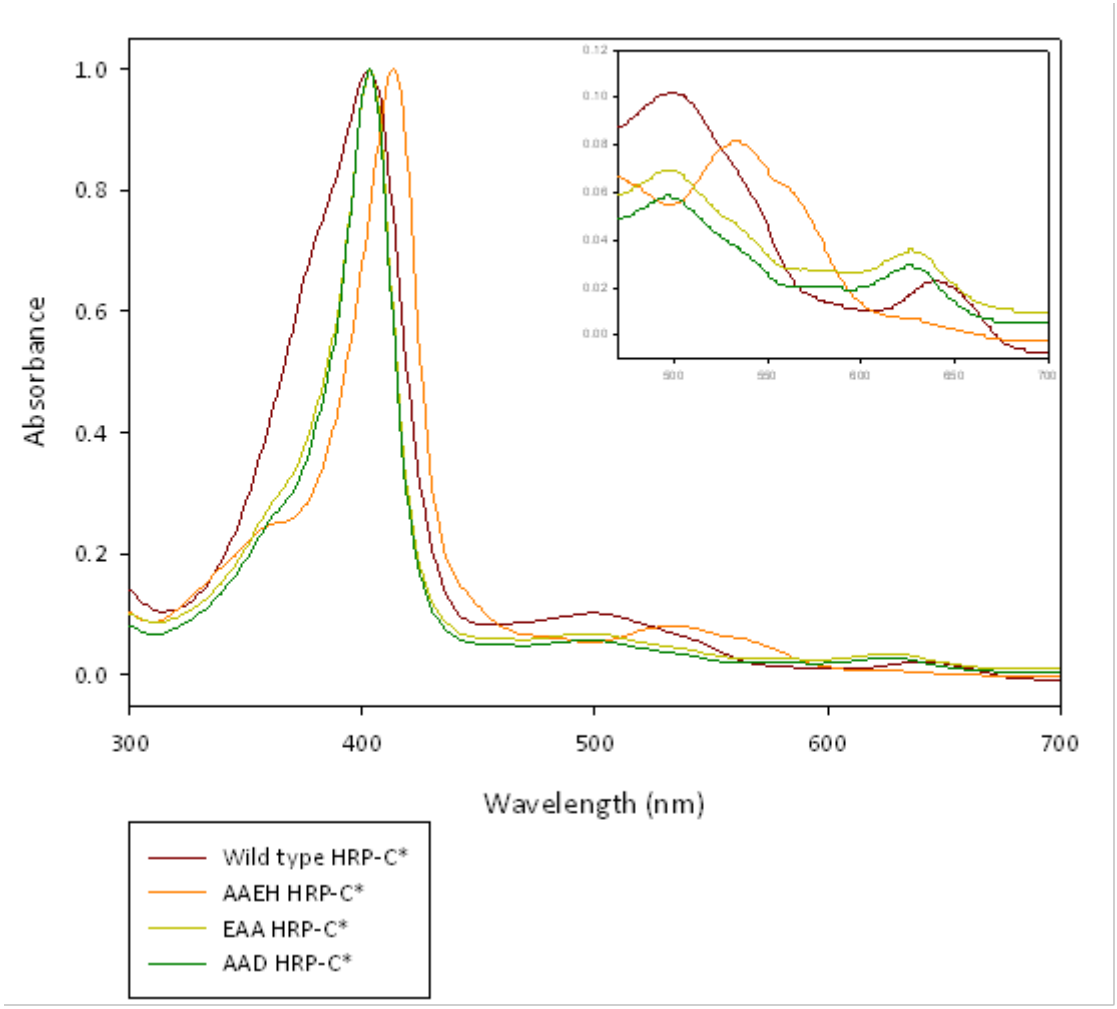


Figure 3.6 HRP-C* UV/Visible spectra
UV/Visible spectra of HRP and HRP variants AAEH, EAA and AAD adjusted such that Soret maxima equal 1. Variant AAEH shows a red shift in the Soret maxima to approximately 420nm (maxima documented in Table 3.x) with charge transfer bands replaced by single broad peak at 530-570nm. Variants AAD and EAA have UV/Visible spectra which resemble the wild type HRP-C*, however charge transfer band I is blue shifted to 620nm.

Variant	Soret maxima (nm)	R_z value	ε_{Soret} ($mM^{-1}cm^{-1}$)	Charge Transfer Band II	Charge Transfer Band I	β band (nm)	α band (nm)
Native HRP	403.0			498.0	~640nm		
WT HRP-C*	403.0	3.4	96±2.1	498.0	643.0	-	-
HAE	402.5	3.3	116±3.5	501.5	638.0	-	-
HAA	403.0	4.0	109±1.9	501.5	630.5	-	-
AAE	402.5	4.2	99±1.7	499.0	635.0	~535.0	~560.0
N70H	402.5	3.0	102±2.0	499.0	643.5	-	-
WT HRP-C* - OH, pH 12.0	415.5	Not determined	Not determined	-	-	545.0	572.0
AAEH	413.0	4.4	100±2.0	-	-	~535.0	~563.0
AAAH	414.0	4.0	104±2.1	-	-	~535.0	~563.0
EAA	403.0	4.1	103±1.8	497.0	626.0	~530.0	-
AAD	403.0	4.2	100±2.2	498.0	625.0	~530.0	-
Chloroperoxidase	403.0	-	-		515, 542, 650		

Table 3.3 - UV/Visible characteristics of HRP variants at pH 7.0 (unless stated)
Native HRP as reported by Schonbaum and Lo (1972), Wild type HRP-C*, WT HRP-C*-OH measured at pH 12.0 and variants HAE, HAA, AAE, N70H, AAEH and AAAH as reported by Ngo (2007). New variants EAA and AAD measured using a Shimadzu UV/Visible spectrophotometer in 5mM MOPS buffer at pH 7.0. Peaks in the spectrum of native chloroperoxidase are included as a comparison as reported by Morris and Hager (1965).
Wild type HRP-C* shows a characteristic 5-coordinated haem peroxidase spectrum, with a Soret maximum at 403nm and an Rz value of 3.4. Charge transfer bands are apparent at 498nm and 643nm. As seen in Figure 3.6, HRP variants have altered spectra compared to wild type. Variants that alter the catalytic site have increased Reinheitszhal values. Most distinct is the variant AAEH which has a red shifted Soret maximum, to 413.0 compared with 403.0 nm in the wild type HRP, as well as the loss of charge transfer bands, replaced by α and β bands at approximately 535.0 and 563.0nm.

Peroxidase purity can be measured using the ratio between the absorbance at the height of the Soret peak and the absorbance at 280nm (Dunford, 1999). This ratio is known as the Reinheitszhal number (R_z). Typical R_z values for wild type HRP-C* fall in the region of 3.0-3.4 for the 5-coordinated haem. 6-coordination of the haem, such as suggested from the spectrum of HRP variant AAEH, results in an increase in R_z value. Both new variants EAA and AAD displayed R_z values in excess of 4.0, as seen with previous variants AAEH and AAE. Although higher R_z values for HRP indicates

increased levels of purity in the enzyme, this does not necessarily correlate to increased activity (Dunford, 1999). Table 3.3 shows a comparison summary of UV/Visible characteristics between HRP variants discussed in this thesis.

3.2.9 Comparison of new variant spectra to previous work

Figure 3.6 also shows UV/Visible spectra of the wild type HRP-C*, previous variant AAEH along with the two correctly folded new variants. As described in the previous work (Ngo, 2007), the AAEH variant has considerable red shift seen in its Soret extinction peak. This is now seen at 414nm rather than the 403nm exhibited by the wild type enzyme and other previous iterations of peroxygenase variants. In addition, the AAEH variant also has no distinguishable charge transfer bands in the 500nm or 640nm regions. Instead the charge transfer bands have been replaced by two peaks at 535nm and 565nm, designated the α and β bands. The distinctive spectra of the AAEH HRP-C* variant suggests that it is a low spin 6-coordinated haem environment, quite different to the 5-coordinated high spin haem environment found in the wild type enzyme. This is thought to be due to the ligation of the introduced histidine residue at position 70 forming the sixth coordination position of the iron, either by extension of the residue to meet the haem, which would require substantial movement in the active site, or in supporting an OH^- charge in the sixth position.

The two new HRP-C* variants, EAA and AAD, both show very typical high spin spectra, though with a much more narrow Soret peak at 403 nm and loss of the Soret shoulder seen in the wild type HRP-C*. The charge transfer bands are retained, although there is some slight blue shifting of the band at 650nm to 630nm. The alterations to haem environment are likely to be the principal cause of these slight shifts in the spectra, however the overall spectrum is highly similar to the wild type 5-coordinate enzyme. Of additional note is the presence of a slight peak at 530nm, which may possibly be attributed to a small amount of 6-coordinate low spin haem being present in the sample. Although predominantly high spin haem enzymes, both variants appear to be more 6-coordinate than the wild type, particularly as judged by the loss of the Soret shoulder seen at 385nm in the wild type.

The UV/Visible spectra of the two new variants are similar to the AAE variant produced previously (Ngo, 2007), though the spectrum of AAE is not shown in Figure 3.5 for clarity in the figure. This variant showed a similar loss of Soret shoulder at 385nm when compared to the wild type HRP-C*, with the Soret peak appearing sharper as a consequence. It is clear from the spectra of the two new variants that they do not show the red shift associated with the 6-coordinate low spin haem as seen in variant AAEH. There is, therefore, no indication of strongly bound ligand at the sixth position in new variants EAA, AAD and previous variant AAE, although a weakly bound water molecule remains a possibility and would give reason for some of the shifts seen in the spectra compared to the wild type enzyme.

While variant AAE and the new variants show predominantly a five co-ordinate high spin haem, with no apparent water bound in the active site cavity, variant AAEH displays a six co-ordinate low spin haem. In the previous work (Ngo, 2007) it was shown that variant AAEH is able to resist high levels of peroxide that, in the wild type and previous peroxygenase variants, cause inactivation of the enzyme due to the 'bleaching' of the haem resulting in loss of activity and integrity in the enzyme. This protective effect in the AAEH variant is thought to have been a result of the low spin haem preventing hydrogen peroxide from gaining direct access to the haem and forming Compound I until the substrate has displaced the bound 6^{th} ligand (Ngo, 2007). When an aromatic substrate is present, it is hypothesised that the substrate binds more deeply in the engineered catalytic site, causing displacement of the ligand which consequently allows hydrogen peroxide to bind and Compound I formation to take place, the cycle thereby becoming concluded much like a P450 type enzyme.

Since new variants EAA and AAD showed UV/Visible spectra that appeared almost identical to previous variant AAE, it was expected that resistance to peroxide would be similar to that seen in variant AAE, rather than the increased resistance seen in the AAEH variant. As new variants EAA and AAD do not display substantial 6-coordinate behaviour it is expected that this ligand, and subsequent properties of peroxide resistance, are not present. This was shown to be true for both variants as when repeat spectra are taken with increasingly high levels of peroxide present, the Soret

peak drops beyond the expected level for Compound I formation and the enzyme loses function.

As shown in Table 3.3 and Figure 3.5, charge transfer band I in EAA and AAD both appeared to blue shift slightly, from 643nm to 626nm. This disruption to the electronic configuration of the haem is expected to be as a direct result of the alterations to the active site. The significance of the disruption to the hydrogen-bonding network present in this cavity is not fully known, however this is clearly a factor in the change in spectra seen for these enzymes. The shifts in this charge transfer band differ between variants and are more pronounced in the EAA variant which potentially has more disruption to the catalytic centre by the replacement of the arginine residue at position 38 with a glutamate.

Sensitivities in the catalytic site have profound implications to the activity of the enzyme, with apparently small changes having much larger effects. The comparison between the current and previous work show how acutely the HRP-C* active site responds to changes in its hydrogen bonding network, both in terms of functionality and stability.

3.3 Basic species characterisation

Basic species characterisation experiments were carried out to enable the new HRP-C* variants to be compared to previous work and to the wild type HRP-C*. The following results show new data from the HRP-C* variants EAA and AAD, with results from previous iterations of HRP-C* peroxygenase variants, wild type HRP-C* and chloroperoxidase, where appropriate.

The initial species characterisation experiments indicate that the new variants behave in a similar way to the previous variant AAE, with decreased peroxidase activity yet showing characteristics of a peroxygenase type enzyme.

3.3.1 Calculation of molar extinction (ε) coefficients

A pyridine haemochrome experiment (after Fuhrop and Smith, 1975) was carried out to determine the Soret extinction coefficient for the new enzyme variants, allowing precise measurement of enzyme concentration for further experiments. This is measured by comparing the difference spectra of oxidised and reduced haem following treatment of the enzyme samples with pyridine and alkali (NaOH).

The extinction coefficients determined are shown in Table 3.3, with additional extinction coefficients from variants AAE, AAEH, wild type HRP-C* and chloroperoxidase as comparisons. Wild type ε was calculated by Ngo to be 96±2.1 $mM^{-1}cm^{-1}$, with variants that display different characteristics with higher coefficient values likely to be caused by the small increases in 6-coordinate haem seen.

3.3.2 ABTS oxidation

A standard ABTS assay was carried out on new variants EAA and AAD as an initial gauge of conventional peroxidase activity and allows direct comparison to results obtained in earlier work. Childs and Bardsley first described ABTS as a potential peroxidase substrate in 1975, where the nature of the radical product of the peroxidative reaction is described in full. Since that time, ABTS has been used extensively in the detection of peroxidase activity, as the green radical product produced as a result of the oxidation is readily tracked at 414nm. Figure 3.7 shows ABTS and its radical product, as described by Childs and Bardsley (1975).

Wild type HRP-C* has a turnover number in excess of $1000s^{-1}$ (1605^{-1} as determined by Ngo) with respect to ABTS under optimum conditions. This rate has been shown to be greatly reduced in all variants that mimic peroxygenase activity, both previous to and including the current generation. This is expected as Compound I formation is impaired due to the absence of a strong base (the histidine from position 42 having been replaced with alanine, aspartate or glutamate) and the absence of a catalytic arginine residue, preventing proton assisted electron transfer by the model proposed by Henriksen *et al.*, 1999. The replacement of these residues by a glutamate still

allows the formation of Compound I, potentially with the glutamate residue assisting dioxygen bond cleavage.

Previous study of the AAEH variant had shown that these peroxygenase variants perform very poorly with standard peroxidase substrates such as ABTS and require higher concentrations of peroxidase to perform even at a low-level of activity. For this reason, hydrogen peroxide concentrations that are higher than those used in the wild type assays were used in the assays described here, to compensate for the poor rates

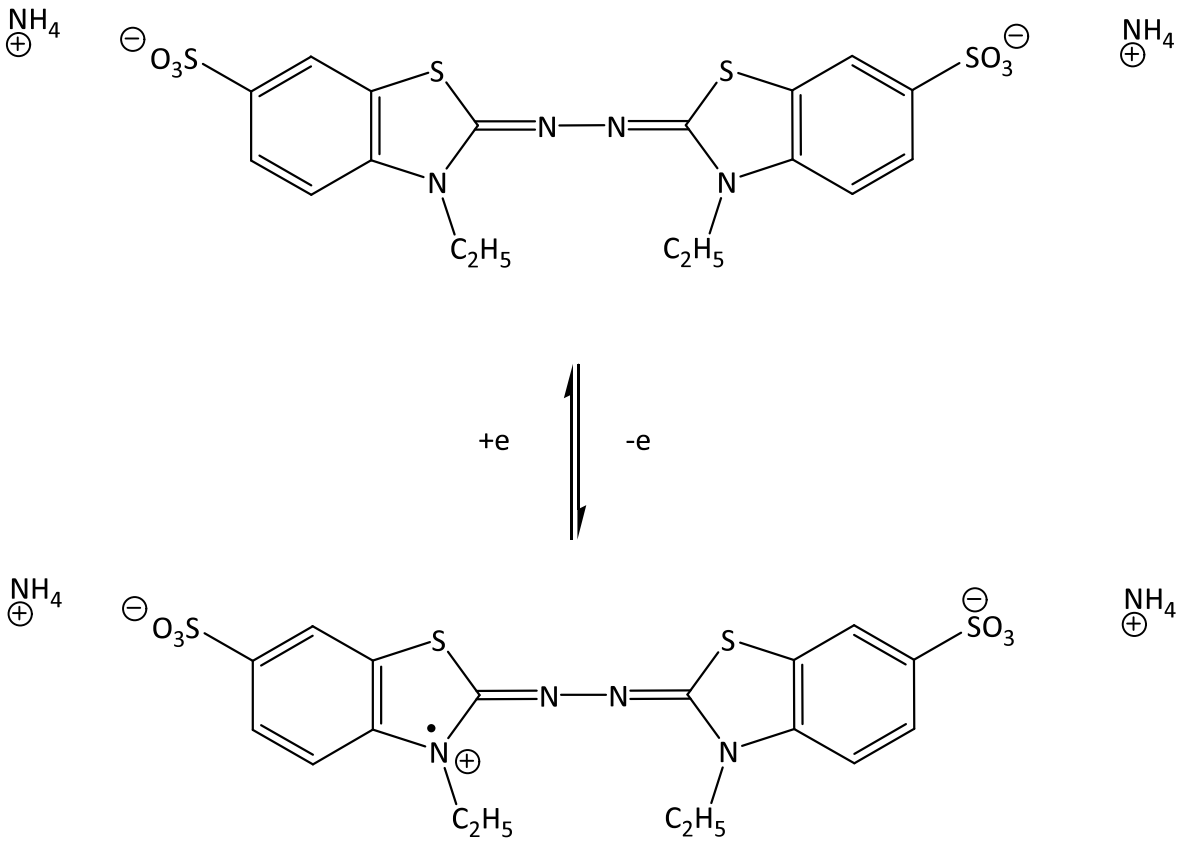


Figure 3.7 ABTS and radical product, reproduced from Childs and Bardsley, 1975
2,2'-Azino-di-(3-ethyl-benzthiazoline-6-sulphonic acid) (ABTS) is a common substrate for determining reaction rates in peroxidases. It has been widely used since its first identification as a chromogen by Childs and Bardsley in 1975. The radical product has a distinctive green colour, the production of which can be followed by an increase in absorbance at 414nm.

of Compound I formation. As shown later, in Chapter 4, high peroxide concentrations without the presence of substrate may result in loss of enzymatic activity due to degradation of the haem in a 'bleaching' effect, however with substrate present this appears not to be the case until very high peroxide concentrations are reached. Although impaired for peroxidase activity, the peroxygenase rates of these variants should be enhanced since peroxygenase activities are slower and are not influenced unduly by Compound I formation rates.

Table 3.4 shows the reactivities of wild type HRP-C* with ABTS in comparison to the HRP-C* variants mentioned previously in this chapter, at optimal pH and peroxide concentration as indicated (discussed below). Also included is data available for wild type chloroperoxidase as a comparative indication of ABTS activity in a peroxygenase type enzyme. Since chloroperoxidase is a peroxidase-cytochrome P450 functional hybrid, catalytic activity in both peroxidase and peroxygenase pathways are possible. This is a feature that has not been achieved with previous or current HRP-C* "peroxygenase" variants. Variants appear to perform poorly with traditional peroxidase substrates, though they do not reach the levels of activity of "peroxygenase" substrates as seen with chloroperoxidase.

New variant EAA, whilst still performing poorly compared to the wild type enzyme, with turnover numbers in excess of 60 fold lower than wild type HRP-C*, exhibits increased activity towards ABTS when compared with the previous variants, where turnover at optimum ABTS concentration did not reach $10s^{-1}$. The best functioning of the previous generation of variants (AAE) had a turnover of $8s^{-1}$, however displayed non Michaelis-Menten kinetics as this variant, and variant AAAH both appeared to be inhibited by high concentrations of ABTS (Ngo, 2007). This relative increase in activity in the EAA variant is likely to be due to the way in which the substrate and peroxide can access the haem. Variant AAD falls within the performance range of the previous generation of peroxygenase mimics, with a turnover of $4s^{-1}$. The poorest performing of the previous variants was variant AAAH, without the glutamate residue anywhere in the catalytic site, which had a turnover of approximately $2s^{-1}$. The AAD turnover is half

that of variant AAE, suggesting that the extension of the acidic residue into the catalytic site is contributing to the activity of the enzyme with ABTS.

Variant	H_2O_2 (mM)	pH	K_m (ABTS) (μM)	Turnover number (s^{-1})
Wild type HRP-C*	2.5	4.5	140 ± 1	1605
AAEH	60	5.5	264 ± 43	7 ± 0.2
AAE	8	8.5	Not determined	8 ± 0.3
EAA	10	7.0		26 ± 3.0
AAD	4	7.5		4 ± 0.3

Table 3.4 Summary of variant activities towards ABTS
Variant EAA showed increased activity towards the substrate ABTS when compared with previous peroxygenase-like variants in which the glutamate residue is found at position 42, whereas AAD performed half as well as its direct comparative variant AAE.

3.3.2.1 pH dependence of ABTS oxidation

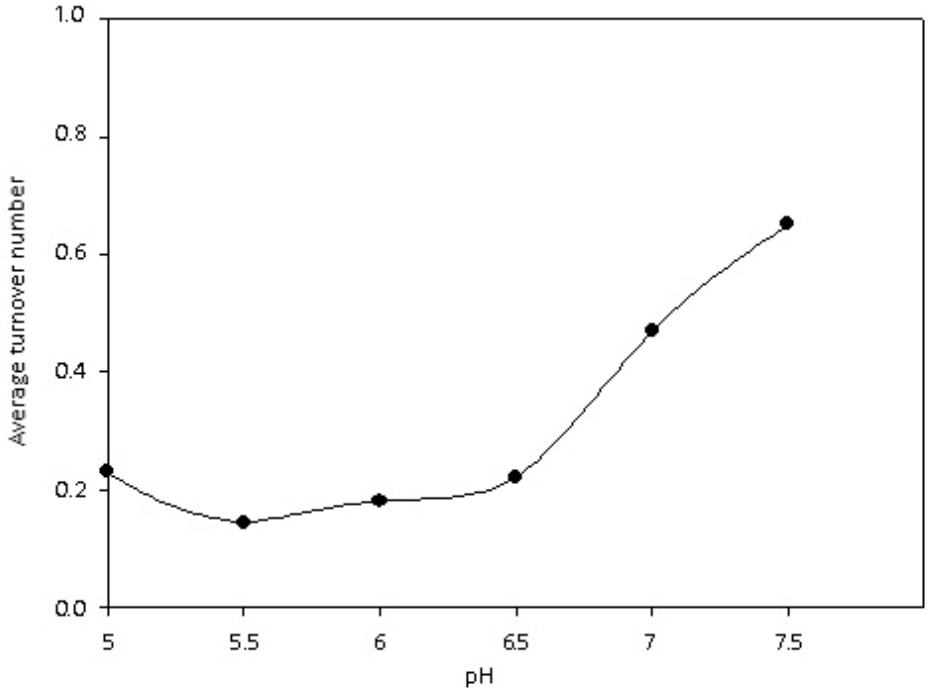
It has previously been reported that pH has a profound effect on the activity of peroxidase enzymes (Sanders *et al.,* 1994), and that optimum pH gives dramatically different activities with relatively small changes of pH. Ngo reported in his previous work (Ngo, 2007) that the pH curves for the new HRP-C* peroxygenase variants varied with the different active site residues. Particularly notable was the variant AAE, where the pH optimum was found to be 8.5, compared to the wild type HRP-C* optimum at pH 4.5.

Figure 3.8 shows curves corresponding to the pH dependence of enzymes AAD, EAA. These enzymes appear to perform better at increased pH, unlike the HRP-C* wild type which has a pH optimum of below 5.0. Subsequent reactions were carried out at pH 7.0 in a phosphate based buffer unless specified.

3.3.2.2 Hydrogen peroxide dependence on ABTS oxidation

As stated previously in this chapter, the new HRP-C* variants respond differently to the wild type HRP-C* when treated with increasing concentrations of hydrogen

a)



b)

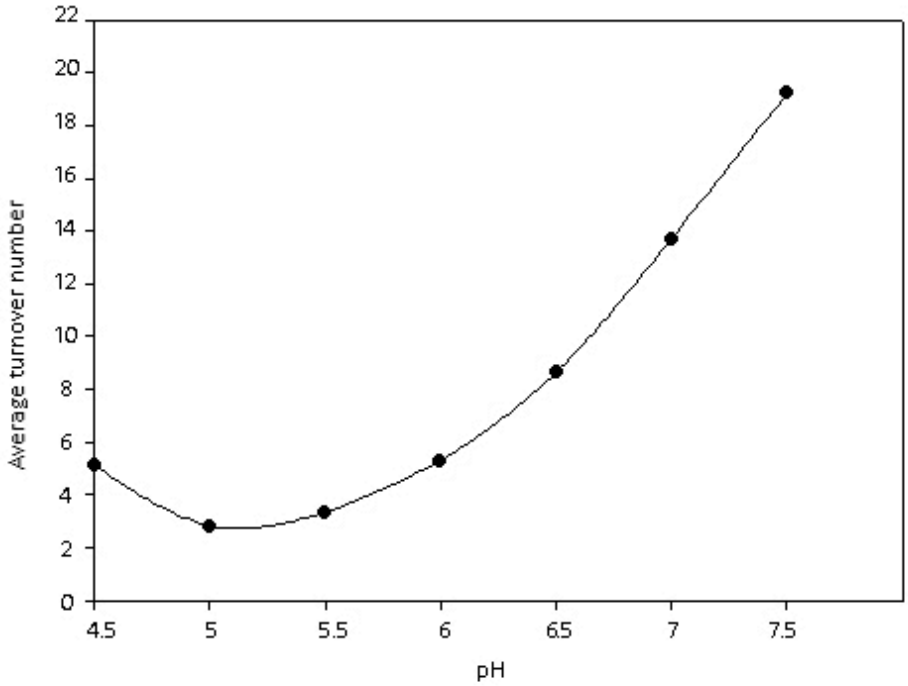


Figure 3.8 pH profiles of AAD (a) and EAA (b)

Performance of the peroxidase variants was improved at neutral pH, as shown in the graphs above, whereas the optimum pH for wild type HRP-C is 5.0.

Higher pH ranges in sodium borate buffer were trialled, however results were highly variable.

peroxide. In particular, variant AAEH demonstrated a significant preference towards higher concentrations of peroxide. Whilst variant AAE shows bell curves with respect to hydrogen peroxide concentration, AAEH and AAAH will withstand high concentrations of peroxide without the haem bleaching phenomenon observed with wild type HRP-C* and other variants. Whilst the turnover seen for AAAH is much lower, the levels of peroxide (up to 60mM has been recorded) reach levels that would cause other plant peroxidases to cease function. Ngo concluded that this activity was due to the presence of an additional ligand to the haem of these two variants, which required displacement prior to Compound I formation. This resistance to peroxide inactivation led to the suggestion that an additional haem ligand results in the prevention of formation of the intermediates that cause the bleaching effect.

Both new variants EAA and AAD showed similar peroxide sensitivity to variant AAE, resulting in an optimum range of peroxide concentration, with a bell shaped curve with respect to activity with increasing hydrogen peroxide concentration.

3.4 Summary

Three new HRP-C* variants were designed, the changes in their haem cavity architecture based on previous work that had successfully generated peroxygenase-like activity in HRP.

Two variants, AAD and EAA have been successfully expressed in *E.Coli*, the inclusion bodies purified and pure protein successfully refolded, generating two new active HRP-C* variants. UV/Visible spectra of the new variants suggest that they contain 5-coordinate, high spin haem groups, unlike previous variant AAEH and AAAH, which contain 6-coordinate, low spin, haem centres, however spectra also contain elements of 6-coordinate character, particularly evidence of a small peak at 530nm not seen in the wild type but present in variant AAEH.

The new variants were examined and activities shown to be similar to the previous variants in terms of reduced activity with the traditional peroxidase substrate ABTS. New variants also display differences in optimum pH and peroxide concentration

compared to the wild type enzyme, however behaving more like previous variant AAE than AAEH, with a limited tolerance to peroxide after which a bleaching effect is observed.

The variety of non-traditional HRP-C* substrates that have been examined with the HRP-C* peroxygenase variants is limited. Increasing the range of substrates that have been shown to be activated by these variants may give increased insight as to the role of the residues present in the active site. It may also give an increased industrial interest to this group of HRP-C* variants, since HRP-C* is a robust and relatively easily produced enzyme, unlike the more sensitive chloroperoxidase. Substrates examined during this work are discussed in Chapter 4.

Chapter 4

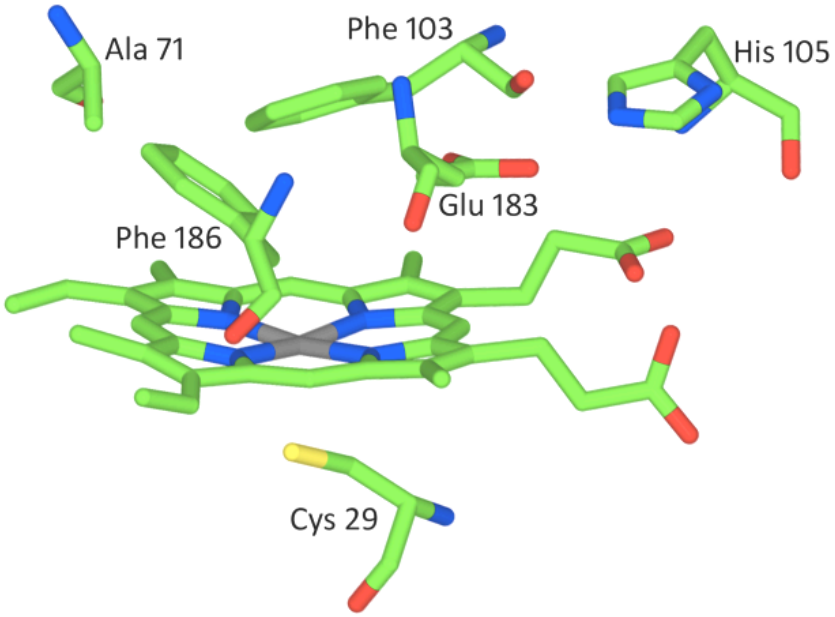
Peroxygenase variant activities

4.1 Introduction

The work described in this and the previous chapter is an extension of previous work carried out by Ngo and Smith, aimed at introducing alternative catalytic functionalities into horseradish peroxidase. This original work was the subject of a patent WO/2007/020428 (Ngo and Smith, 2007) and, in turn, built on the work of Ortiz de Montellano and coworkers (reviewed in van de Velde *et al*).

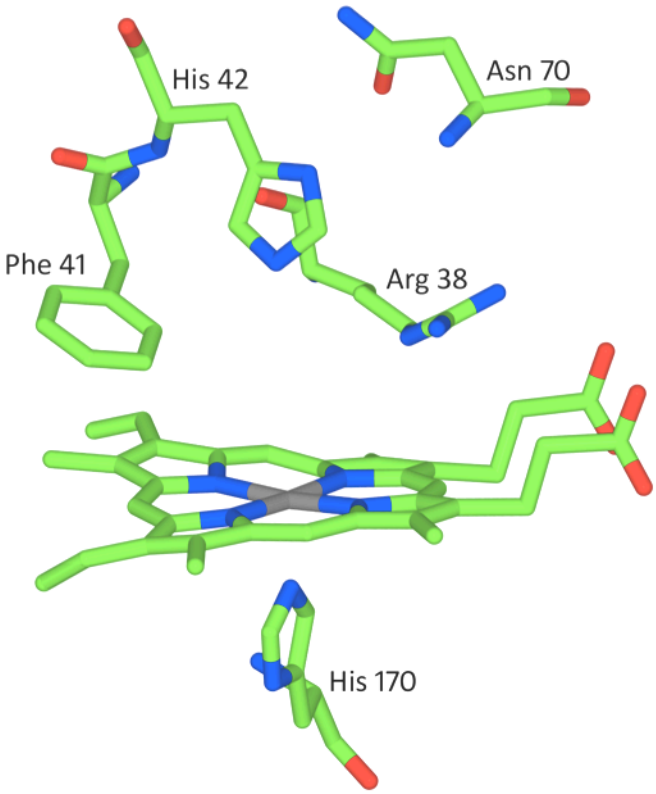
The aim of the work, as described in the patent (Ngo and Smith, 2007), was to engineer "robust" catalysts capable of stereoselective transfer of an oxygen atom into small aromatic substrates. These type of reactions are often difficult to achieve chemically, particularly with the high yield and enantioselectivity needed, for example in the pharmaceutical sector (Ngo and Smith, 2007). This extended a body of work which focussed the engineering of the active site of HRP to improve the accessibility of the reactive oxy-ferryl intermediate.

Peroxygenase activities are seen in nature in the cytochrome P450 enzymes. These monooxygenases, described in brief in Chapter 1, are efficient at oxygen transfer, using molecular oxygen as the source of oxidant. They can, however, also utilise hydrogen peroxide, although rather slowly and are hampered by rapid inactivation of the enzyme (Dunford, 1999). In contrast, chloroperoxidase is a unique enzyme that is able to perform both traditional peroxidase activities and true oxygen transfer (peroxygenase reactions), in addition to halogenation. The catalytic site of chloroperoxidase is unlike plant peroxidases, it is more hydrophobic in character with the critical acid-base mechanism needed for dioxygen bond cleavage provided by a unique glutamate residue seen at position 183 (Sundaramoorthy *et al*, 1995). This distal haem cavity is more open and is unlike plant peroxidases, where a combination of highly conserved arginine and histidine residues (at position 38 and 42 respectively



4.1.a Chloroperoxidase catalytic centre

Glutamate in the catalytic centre of CPO is unique amongst peroxidase enzymes and thought to confer some of the unique activities seen in CPO, particularly halogenation of substrates.



4.1.b HRP-C* catalytic centre

The HRP distal haem pocket is representative of plant peroxidases, with catalytic R38, F41 and H42 highly conserved in this class of enzyme. Histidine at position 170 provides the 5th co-ordination ligand for the iron centre of the haem.

in HRP-C) are responsible for the cleavage of the peroxide dioxygen bond in traditional peroxidases (Dunford, 1999). In chloroperoxidase, this role is played by glutamate (Glu183) and histidine residue that is slightly removed from the catalytic centre and stabilised by hydrogen bonding (Sundaramoorthy *et al.*, 1995). This glutamate residue is at a crucial central position in the active site of the chloroperoxidase molecule and is thought to be vital in the distinctive activity of chloroperoxidase, functioning as the base during peroxide cleavage but allowing a more open cavity.

Figure 4.1.a, also shown in Chapter 1, shows the distal haem cavity of chloroperoxidase. Of particular note is its open architecture, particularly in comparison with the closed distal haem cavity of HRP-C (Figure 4.1.b).

As discussed previously, enzymes in the cytochrome P450 class, as well as chloroperoxidase, have been shown to share a similar first catalytic intermediate to the oxyferryl intermediate of peroxidases, Compound I (Egawa *et al.*, 1994 for P450s and Stone *et al.*, 2005 for chloroperoxidase). However unlike more traditional peroxidases, the peroxygenase pathway is characterised by the stereoselective transfer of oxygen to the substrate in a manner similar to that of cytochromes P450.

Figure 4.2 shows the pathway of typical peroxidative action via Compound I to Compound II, returning to the resting state. After Compound I, the peroxygenase pathway allows direct transfer of oxygen to the substrate rather than the two sequential one-electron oxidation steps typical of the traditional peroxidase pathway (Sundaramoorthy *et al.*, 1995). This divergence of pathway is illustrated in Figure 4.3.

The recent description of the unique activities of the *Agrocybe aegerita* peroxidase/peroxygenase (AaP) and its similarity to the chloroperoxidase-cytochrome P450 hybrid system (Omura, 2005) has suggested the presence of an expanded range of peroxygenase-type enzymes in nature. Although their physiological role in Basidiomycetes remains unclear, they are capable of efficient direct hydroxylation, for example of naphthalene by AaP (Kluge *et al.*, 2007). C-H bond activitation remains a

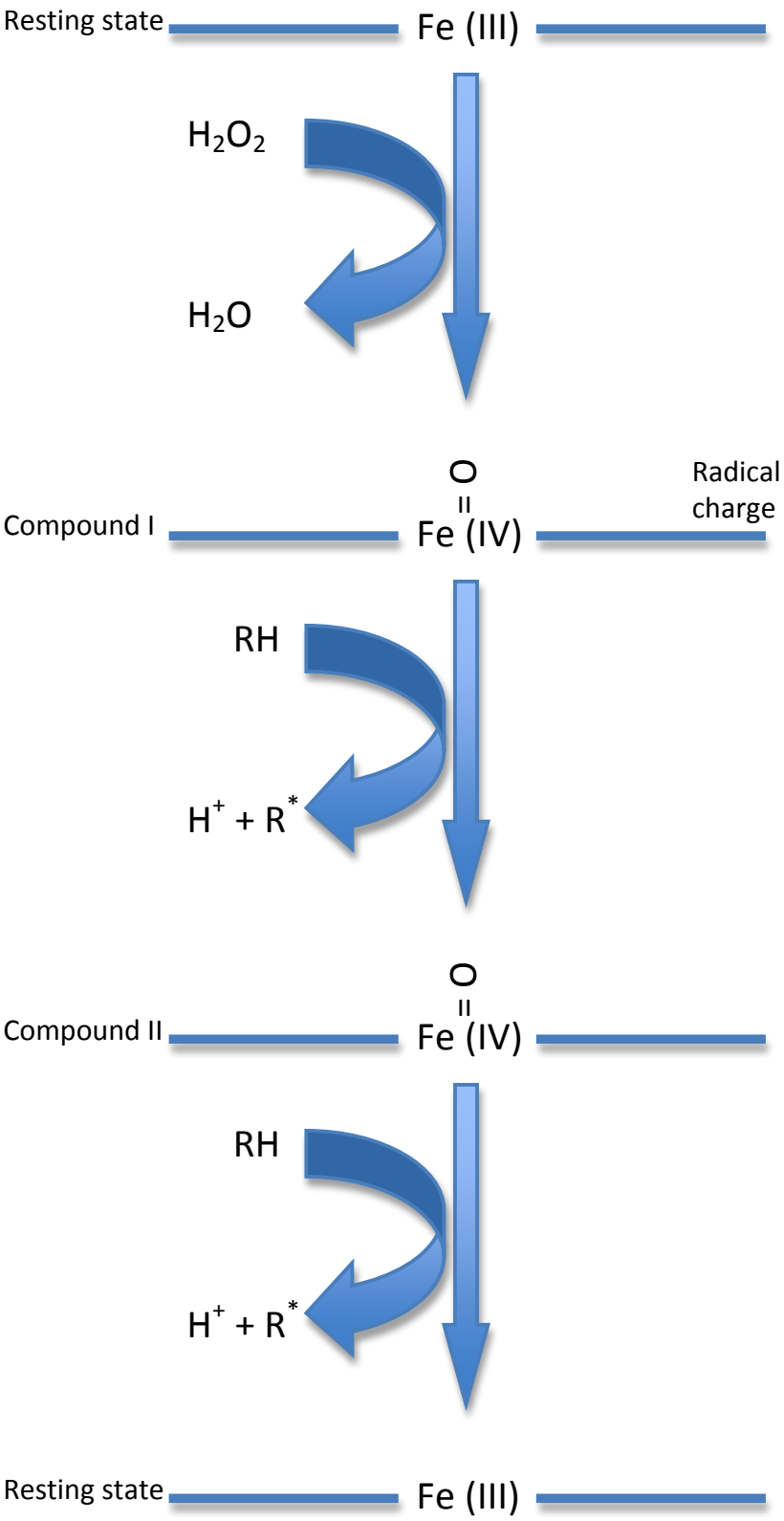


Figure 4.2 Basic steps of the peroxidase catalytic cycle
The traditional peroxidase catalytic cycle involves the two-electron oxidation of the haem group by hydrogen peroxide. The enzyme is returned to resting state via two one-electron reduction steps, oxidising small aromatic substrates at the haem edge. No direct access to the Fe^{IV}=O is possible due to the closed haem architecture.

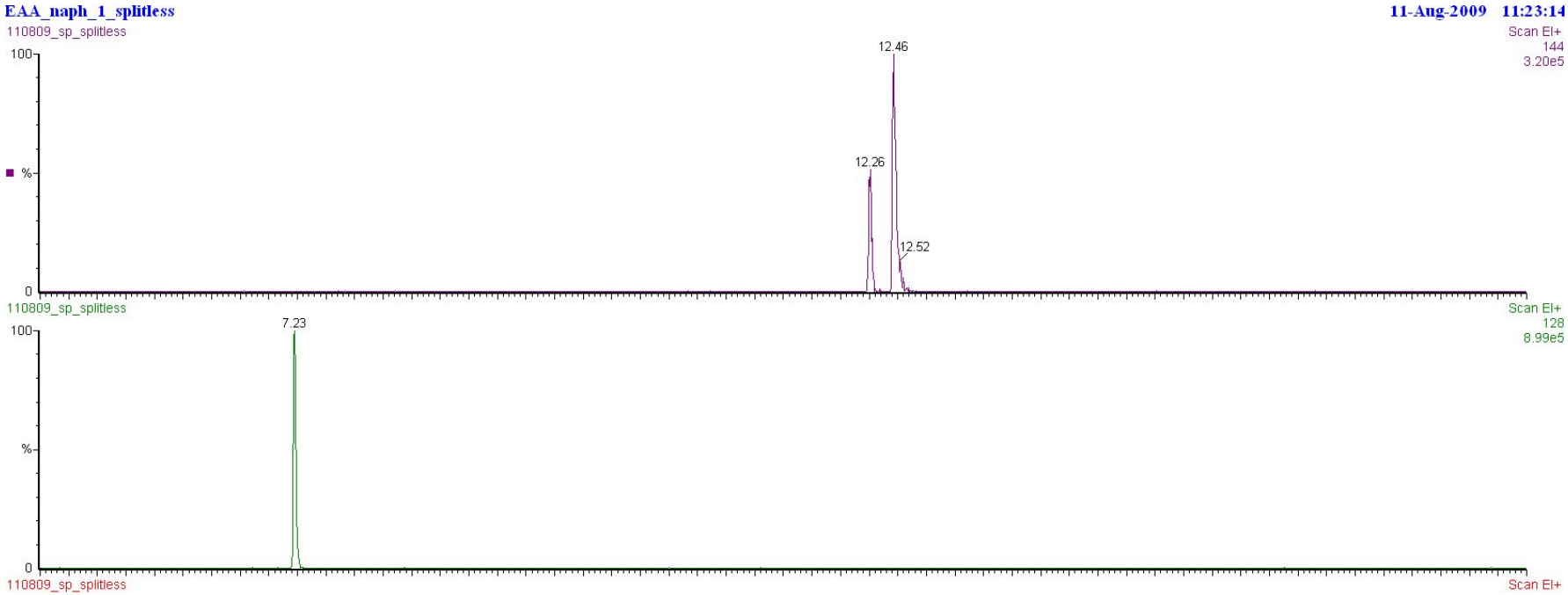


Figure 4.12 GC/MS trace of naphthalene and naphthols resulting from enzymatic oxygen transfer by EAA

The bottom trace represents the residual naphthalene in the sample (Mr = 128), with the top trace indicating the presence of both α and β naphthols (Mr = 144) (peaks not representative of proportion in solution).

highly desirable activity to perform under conditions which are not chemically harsh (Clark, 1999, in Ullrich 2005).

As described previously, work by Ngo, Schneider and others introduced elements of the chloroperoxidase active site into that of HRP in an attempt to create a more open active site architecture in order to reproduce the C-H bond activation activity more typical of chloroperoxidases or P450 enzymes. Both of these types of enzymes have been shown to act as "peroxygenases" in response to a limited range of activatable hydrocarbons in the presence of hydrogen peroxide as the oxidant.

Learning how to switch the reaction pathway between the peroxidase and peroxygenase routes would allow the development of new tailored enzyme systems for stereoselective oxygen insertion for high value synthetic purposes (Ngo and Smith, 2007).

As detailed in Table 3.1, a number of variants have been engineered to incorporate the unique distal glutamate residue seen in chloroperoxidase whilst simultaneously enlarging the catalytic site. Conclusions drawn from each generation of variants has allowed further specified substitutions as described in this work.

The HAE variant confirmed the importance of the reduction in the size of residue side chains in the HRP-C* active site. Other iterations in this phase of HRP-C* variants which maintained the introduced glutamate residue but varied the side chains did not give the rates of epoxidation of styrene seen in the R38H/F41A/H42E variant HRP. The specific activity of HAE in the production of styrene oxide was 126nmol.s^{-1}µmol^{-1} with the concomitant production of phenylacetylaldehyde of 257nmol.s^{-1}µmol^{-1}, in comparison with 0.01nmol.s^{-1}µmol^{-1} with trace production of phenylacetylaldehyde of the wild type HRP (Schneider *et al.*, unpublished data). Other variants examined alongside HAE, H42E and R38H/H42E, produced specific activities of styrene epoxidation in the region of 700-1200 fold higher than the wild type HRP (Schneider *et al.*, unpublished data). Another notable variant created and examined was variant AHE

which exhibited 4000 fold higher styrene epoxidation activity than wild type HRP, however still four fold lower than the HAE variant. This variant, like HAE, had the large hydrophobic phenylalanine at position 41 replaced and a more open active site due to the reduction in bulky side chains.

Although the rate of styrene epoxidation seen in both HAE and AHE had dramatically increased compared to wild type HRP, the rate was still in excess of 60 fold lower in recorded specific activity than for styrene oxide formation by chloroperoxidase (Schneider *et al.*, unpublished data). This generation of HRP variants showed that although engineering peroxygenation activity into a peroxidase was possible, it required a residue able to function as a surrogate base for peroxide activitation and seemed to favour a larger distal haem pocket.

The X-ray crystal structure of the variant HAE was solved and suggested how non-natural HRP substrates are able to access the haem directly, undergoing catalytic reactions not seen in the wild type enzyme. The distal cavity area was increased and it was postulated that substrates were able to reach the iron-oxo intermediate at the haem centre. This crystal structure showed how the replacement of the histidine at position 42 with a glutamate has withdrawn the bulk of this residue from the active site, potentially allowing increased substrate access but providing a weak base, as in chloroperoxidase, for peroxide activation. This structure shows a significant spatial increase within the active site when compared to the wild type HRP, as shown in Figure 4.4 by the overlaying of the HAE and wild type HRP structures (Berglund, Schneider and Smith, unpublished data). However the catalytic centre still retains a large histidine residue which has replaced the arginine at position 38. The cavity is not, therefore, as open as it could be.

The potential catalytic properties of engineered HRP variants was subsequently probed by Ngo in his doctoral work. Substitutions made within the haem cavity further enlarged the area directly around the haem centre, allowing substrates better access to the haem centre, with variants AAE, AAEH and AAAH. This increased substrate-access to the ferryl intermediate in a way similar to that seen in cytochrome P450

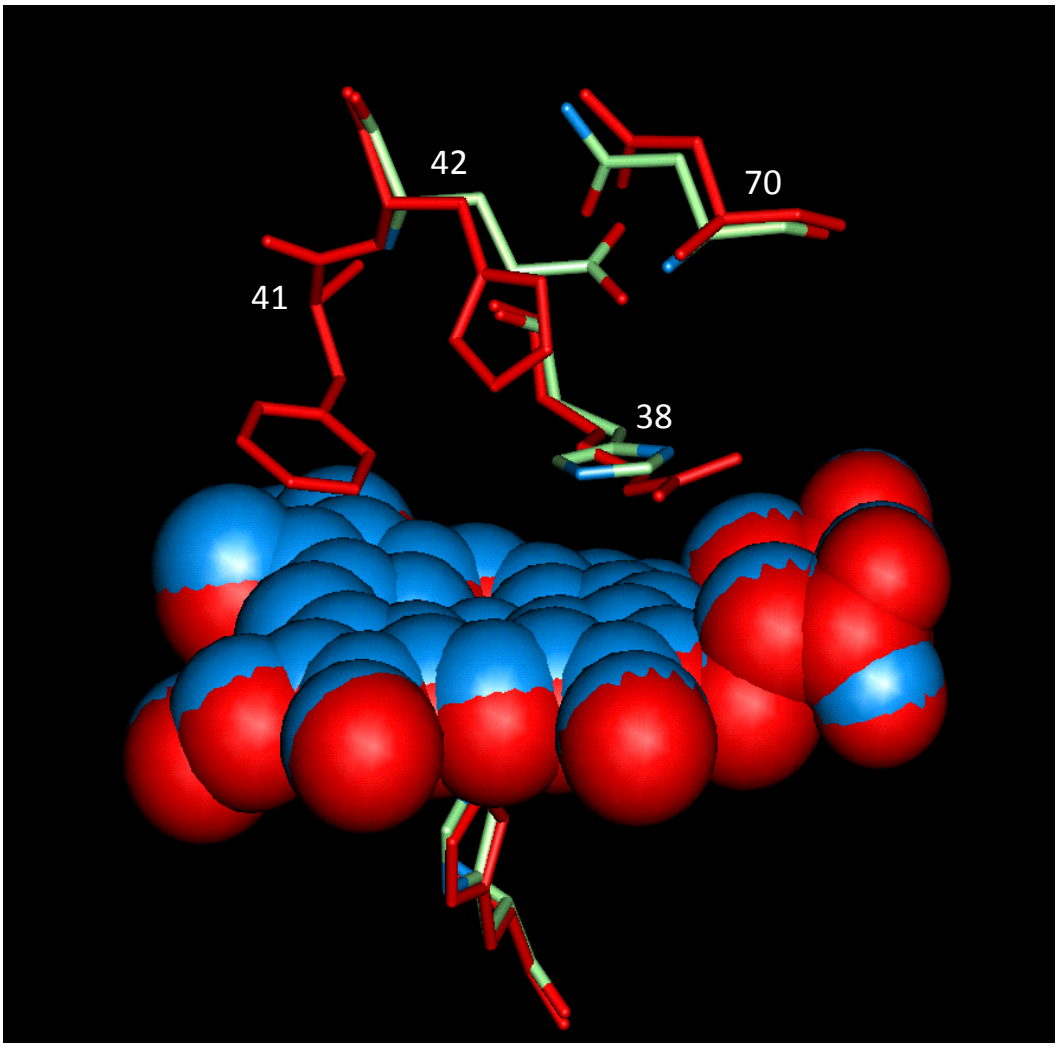


Figure 4.4 Overlay of HAE and wild type HRP crystal structures
The crystal structure of HAE showed the extent to which the distal haem cavity could be enlarged by substitution of key residues, particularly the substitution of F41 for an alanine residue. The enlargement of the site and introduction of the glutamate residue at position 42 enabled the engineered cavity to epoxidise styrene, a reaction not seen in the wild type enzyme.

enzymes (Poulos *et al.,* 1987). Ngo's variants continued the previous work with the introduced glutamate residue at position 42, but extended the active site further with alanine residues substituted for the bulky residues normally associated with the acid base mechanism that provides the hydrogen bonding network above the iron centre of HRP.

The substrate range examined with this generation of variants was extended. Not only did these variants show sulphoxidation activity towards thioanisole and methylthionaphthalene, in the case of the variants AAE and AAEH, these showed 100% stereospecificity in the sulphoxidation of thioanisole. This was shown to be a direct result of the glutamate at position 42, with variant AAAH not exhibiting this level of specificity in the reaction. Table 4.1 summarises this enantioselectivity of sulphoxidation exhibited, in comparison with previous variants.

Enzyme	S:R	Time (hours)	Yield (%)
HAA	51:49	1.45	100
F41A	37:63	1.50	54
HAE	48:52	2.00	100
AAEH	0:100	1.43	100
AAE	0:100	1.42	99
AAAH	19:81	1.35	80

Table 4.1 - Enantioselectivity of thioanisole sulphoxidation with HRP-C* enzyme variants.
HRP variants in which glutamate is retained at position 42, in combination with reduction in distal haem cavity residues, results in complete enantioselectivity towards the R form of thioanisole sulphoxide.

4.2 New generation HRP variants

Previous variants gave an indication of the necessity for the introduction of a glutamate at position 42, similar to that found in the active site of chloroperoxidase, upon which the variants were based. Earlier variants and their identifying codes are detailed in Chapter 3 and Table 3.1, along with the most recent generation of HRP

variants that are the subject of this chapter.

The new variants created in this work were chosen to probe the role of the glutamate residue introduced at position 42 and known to be important in previous work for a) providing a weak surrogate base for activation of peroxide without blocking the distal cavity and b) retention of stereoselectivity. A similar Glu residue appears to have a key role in chloroperoxidase. In this work, we address the sensitivity of peroxygenase activity, retaining the weak base (Glu), but moving it to an alternative position within the engineered active site. This should leave the haem cavity even more open than previous variants.

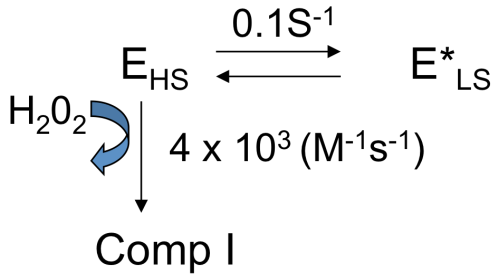
As described in the previous chapter, variants EAA, AEA and AAD were designed and protein expressed using an *E.coli* expression system, though only EAA and AAD were subsequently produced in reasonable yield. These new variants showed similar UV/visible spectroscopic properties to the wild type HRP, with Soret peaks at approximately the same wavelength and two clear charge transfer bands, indicating the haem is still in a high spin state. The spectral properties are similar to those seen for the AAE variant in the previous work, again similarly characterised and expected to contain a 5-coordinated high spin haem group.

In the previous chapter, I showed the greatly reduced rates of reaction with ABTS, a typical 'haem-edge' peroxidase substrate. ABTS is oxidised to give a green radical product which can be easily detectable. The wild type HRP-C* has a high activity for ABTS with a ~560s^{-1} rate limiting step (Smith *et al.*, 1992). The HRP variants AAE, AAEH and AAAH all had low activities, with rates of ~20s^{-1}. The AAD and EAA HRP variants exhibited similarly low activities with ABTS. This is to be expected as the mutations intended to make the cavity more open compromise the acid-base machinery and the Arg removed from position 38 has a key role in substrate reduction by phenols (Henriksen *et al.*, 1998). High rates of Compound I formation are not needed to sustain the much slower oxygen transfer reactions sought here.

4.2.1 Compound I formation by EAA, studied by stopped flow spectrophotometry

The new variant HRP EAA was studied using stopped flow spectrophotometry to calculate rates of Compound I formation.

Figure 4.5 follows the formation of Compound I at 400nm in the stopped flow over a period of 5 seconds. Data was fit to a double exponential curve, with a fast phase over the 0.75s, followed by a slower phase. Triple exponential fitting of the data was attempted, however the first (fastest) phase of this fit was a very small amplitude and therefore not reproducible. This may be due to oscillations in the mains background of the trace – this effect has been seen in other work of similar scale.



Scheme 4.1 Compound I formation by peroxygenase variant EAA
A two phase reaction is seen with Compound I formation. The fast phase is seen to be peroxide dependent, however the second is not, leading to the model proposed. A small proportion of 6 co-ordinate low spin enzyme is typically seen in UV/Visible spectra of this enzyme, therefore interconversion to the high spin state prior to reaction with peroxide must occur.

Scheme 4.1 suggests an explanation for this two phase Compound I formation reaction. In the UV/Visible traces of EAA, a small proportion of low spin character is seen, and was remarked upon previously. Unlike HRP peroxygenase variant AAEH, the predominant form of the enzyme is the high spin form, which is able to react directly with peroxide. This effect also supports the ability of variant AAEH to withstand high concentrations of peroxide – the low spin nature of the enzyme effectively "shields" the haem from peroxide unless a reducing substrate is present and binds to the active site.

Table 4.2 shows the rate of EAA Compound I formation in comparison to previous HRP

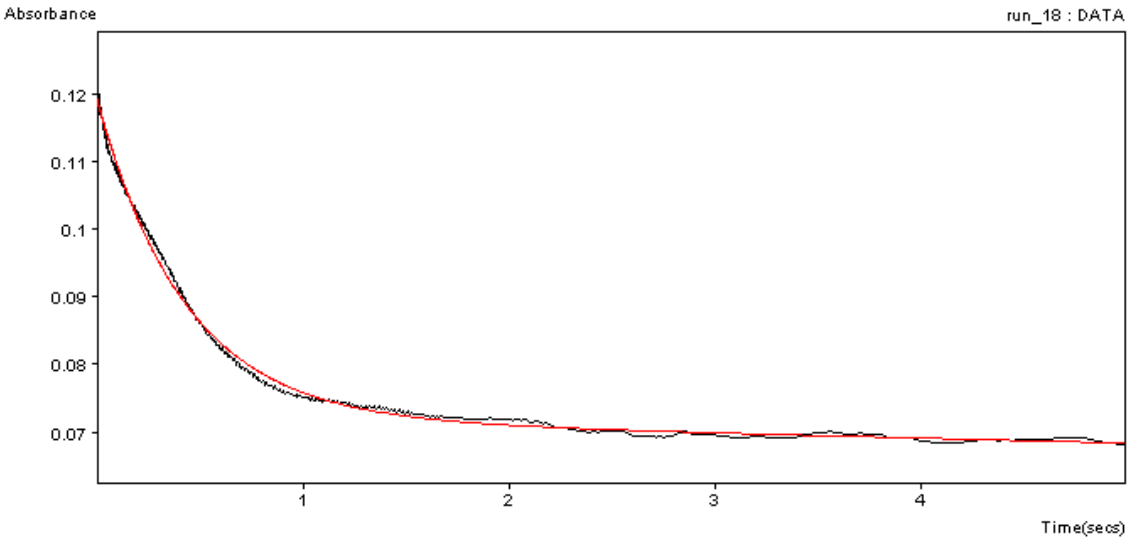


Figure 4.5 Compound I formation by new variant EAA measured at 400nm
Compound I formation was examined using stopped flow techniques. Plots were fit to a double exponential curve, providing a good model basis for the reaction course. An initial fast phase of approximately 0.75s was observed, followed by a slower phase.

catalytic site variants. Where a histidine is retained in the active site alongside an introduced glutamate (e.g. HAE, though other variants have similar rates), the acid-base mechanism of Compound I formation is retained and although a decrease in rate in comparison to wild type HRP is seen, these variants still retain an activity of >1 x 10^5. Where the histidine is removed entirely, the rate of Compound I formation drops dramatically. Previous work has shown that the histidine residue is required for the deprotonation of the peroxide and facilitates the subsequent dioxygen bond cleavage (Newmayer and Ortiz de Montellano, 1995). With the introduction of the single glutamate at position 42 (Tanaka et al., 1997), Compound I formation rates are seen that are lower than wild type HRP C*, however this result shows that, as in chloroperoxidase, a glutamate can provide support for Compound I formation as a general acid-base catalyst. Despite the surrogate proton acceptor, this substitution is not as efficient as the catalytic histidine within the active site (Tanaka et al., 1997).

Variant	K_1 (M^{-1} s^{-1})
R38E:F41A:H42A (EAA)	$4.0 \pm 0.3 \times 10^3$
R38H:F41A:H42E (HAE)	2.0×10^5
H42A	1.0×10^1
H42E	6.0×10^3
WT HRP C*	1.7×10^7

Table 4.2 Comparison of Compound I formation rates by HRP variants
Variants in which the acid base machinery is disrupted (e.g. H42A) show very poor rates of Compound I formation in comparison to wild type HRP, however this activity is partially restored by the introduction of a glutamate residue.

The removal of the catalytic histidine and replacement by alanine in variant EAA would be expected to cause a dramatic drop in Compound I formation, however the presence of the glutamate in the active site pocket serves to partially recover the ability to form Compound I, albeit at a rate that is around 500 fold lower than the wild type HRP C*. This suggests that the presence of the glutamate residue in the active site, rather than

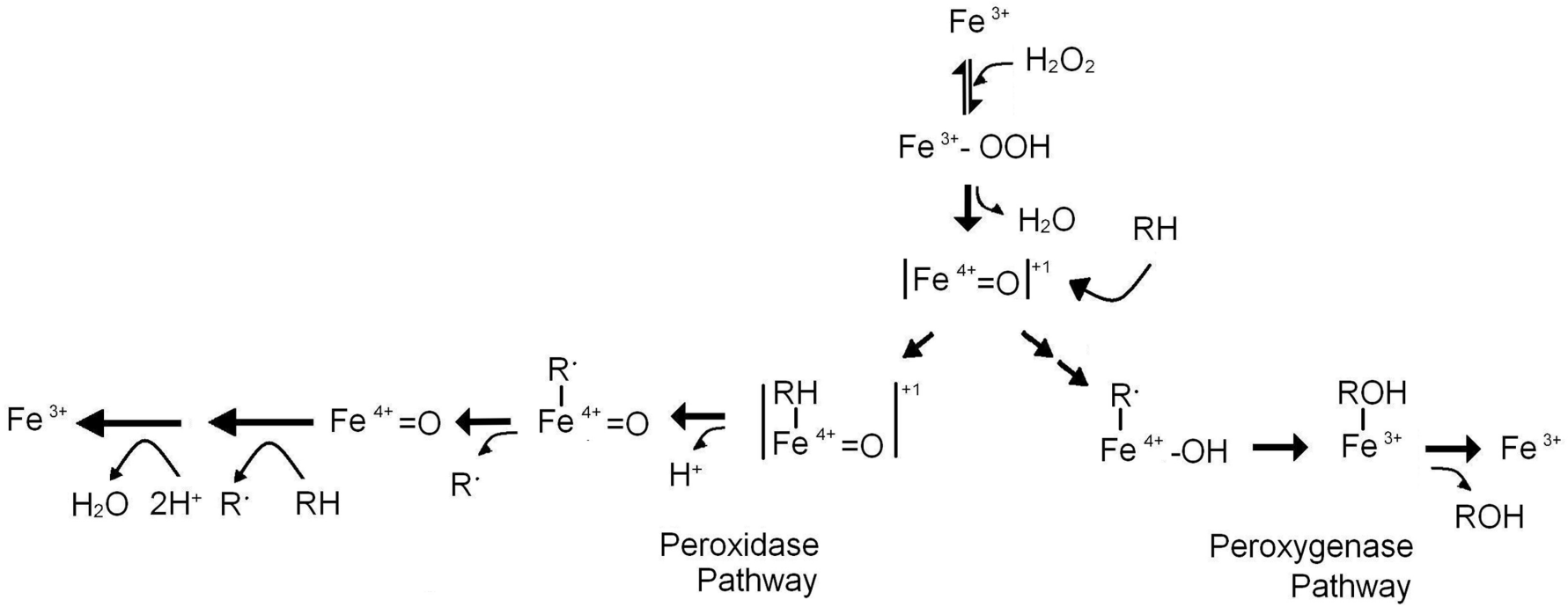


Figure 4.3 Peroxidase and peroxygenase catalytic pathways

At the point of Compound I formation, the pathways diverge, with peroxidase enzymes returning to the resting state as described in Figure 4.2. Peroxygenases, however, allow direct oxygen transfer to the substrates, thought to be facilitated by their more open catalytic site. Peroxygenases allow little or no reactivity at their haem edge.

its specific location in position 42, is sufficient to allow enzymatic activity. This would allow further examination of the active site residues in combination with this glutamate, particularly in giving a more chloroperoxidase-like nature to the site.

Over longer time courses, the higher concentration of hydrogen peroxide renders the variants sensitive to inactivation, as evidenced by the haem bleaching shown in Figure 4.6. Incubation of the EAA variant with 10mM hydrogen peroxide over a 15 minute period resulted in the near complete bleaching of the enzyme. Compound I is formed initially, however when there is no substrate present the Soret peak seen at 403nm, normally decreased by a factor of two in the formation of Compound I, gradually decreases, along with a corresponding reduction in the charge transfer bands seen in the visible region of the spectrum. Figure 4.7 indicates the effect that the bleaching reaction has on the defined charge transfer bands at 500nm and 630nm. In Figure 4.7.a and 4.7.b trace [A] depicts the resting state, trace [B] shows the effect of peroxide binding and shows the indicative loss of defined peaks in this region. Traces [C] and [D] demonstrate the bleaching reaction, modelled as a two phase reaction in which the trace [D] is the end point of 1000 seconds of data collection. In Figure 4.7.a and b, the change in electronic state of the haem is clear from the resting state to Compound I, with clear bands at 500nm, a slight shoulder at 530nm and a clear peak at 630nm being essentially flattened out as Compound I is formed.

It is also of note that, at concentrations of hydrogen peroxide in excess of 1mM, presence of the substrate does not protect the enzyme from this bleaching effect.

4.2.2 New oxygen transfer activities

As reported by Ullrich and Hofrichter (2007), hydroxylation of aromatic compounds, particularly consistent selectivity in hydroxylation, is of interest chemically due to the difficulty of obtaining pure products, such as those needed for pharmaceutical production. These reactions, when synthesised chemically, require multi-step processes that can be lengthy and complex. Ngo and Smith (2007) have already shown that engineered peroxidases (AAE, for example) were capable of 100% enantioselective sulphoxidation. The functionality and solvent compatibility of

peroxygenases has implications for future 'high value' chemical and synthetic steps.

Very recently, extracellular haem peroxygenases from *Agrocybe aegerita* have been shown to selectively oxidise toluene and 4-nitrotoluene to benzoic acids through a mechanism seen elsewhere in P450 enzymes (Kinne *et al*, 2010). This method, known as the "peroxide shunt", is a peroxide dependent transfer of oxygen to the substrate, allowing the P450 to hydroxylate substrates without additional electron-transferring proteins. In P450 enzymes this reaction requires a high concentration of peroxide and, despite the high concentration of peroxide, the turnover of the shunt reactions is slow (Matsunaga and Shiro, 2004) and often results in enzyme inactivation.

Previously, work by the Hofrichter group has shown that *Agrocybe aegerita* peroxidase (AaP) (possibly better classified as a peroxygenase) is capable of performing a suite of activities not otherwise seen naturally (reviewed in Hofrichter and Ullrich, 2006). Furthermore, a search for similar enzyme sequences has revealed AaP-like enzymes in other agaric basidiomycetes such as *Coprinellus radians*, leading to the description of this novel grouping as aromatic peroxygenases (APO) (Pecyna *et al.*, 2009). This group of enzymes, like chloroperoxidase, perform activities not seen elsewhere in the plant peroxidases, though it would appear that the halogenation activity of the APOs is not as pronounced as that seen in chloroperoxidase where it is the dominant activity in the presence of Cl^- ions (Ullrich and Hofrichter, 2005).

The sequence of the APO from *Agrocybe aegerita* has no homology to that of other plant peroxidases or P450 enzymes, though it shares 29% sequence identity with chloroperoxidase in its N-terminal half. The structure of AaP has recently been solved (Piontek *et al.*, 2010), though details of this are not yet published in the Protein Data Bank.

4.2.3 Naphthalene oxidation

Oxidation of naphthalene is a more difficult C-H bond activation reaction to perform chemically than epoxidation or sulphoxidation and is not a natural reaction for Class III plant peroxidases. The limited access to the haem centre precludes catalysis, the ferryl

oxygen is effectively 'hidden' from the substrate, with oxygen transfer sterically hindered. It is, however, a reaction performed by the recently characterised 'chloroperoxidase-like' enzymes of the newly classified group of APO fungal peroxygenases.

The expansion of the HRP active site in previous enzyme variants HAE, AAE, AAEH and AAAH resulted in the establishment of new epoxidation and sulphoxidation activities in HRP. It was thought that a more extensively modified cavity might allow the hydroxylation of naphthalene as seen for the APO group of enzymes. This would potentially lead to colorimetric screenable assay systems for directed evolution procedures towards tailored biocatalysts for industrial uses.

The assay used in this work was adapted from that published by Kluge *et al.*, detailed in Chapter 2. Naphthalene was dissolved in acetonitrile and made up fresh each day, this was kept in a sealed glass flask to reduce volatilisation of compounds.

4.2.4 Naphthalene binding assays

Initially, in order to confirm solvent compatibility, acetonitrile was titrated against enzyme in potassium phosphate buffer and activity checked in a standard ABTS assay. This did not show significant spectral variation in the enzyme, even with up to 20% acetonitrile in the enzyme solution, or any affect on ABTS activity. UV/Visible spectra of the haem are a sensitive indicator of local distal pocket environment, hence naphthalene was titrated against potassium phosphate buffer solutions containing enzyme variant. The difference spectrum of the enzyme in buffer with additional naphthalene compared with enzyme in buffer alone was collected for each concentration of naphthalene.

Figure 4.8 shows an example of the difference spectrum obtained on addition of naphthalene to EAA. A clear decrease in absorbance at 403nm is visible upon addition of naphthalene, through the large trough, with a slight increase in absorbance observed at 424nm.

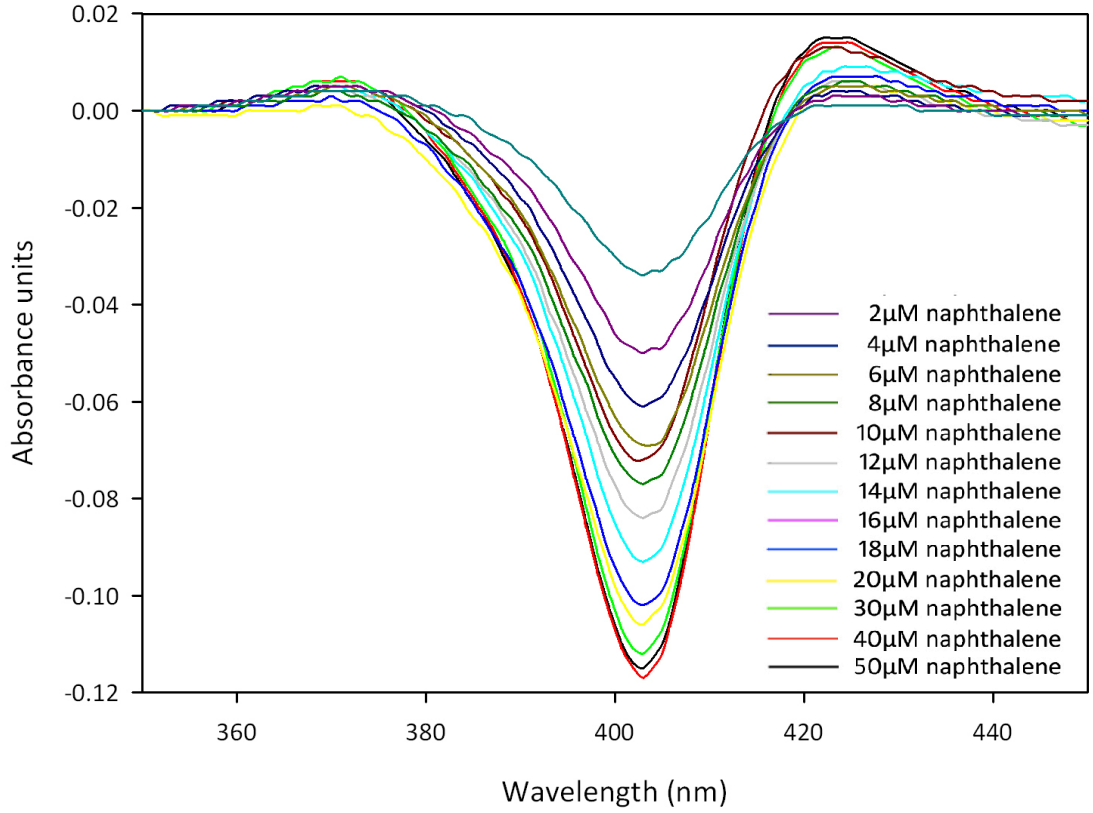


Figure 4.8 Naphthalene titration against 5µM EAA in potassium phosphate buffer
(Adjusted to 350nm = 0)
Changes in absorbance were seen upon addition of naphthalene to 5µM enzyme, with significant increases seen between 2µM and 10µM naphthalene added. Soret maxima dropped, and a small increase was seen at 424nm.

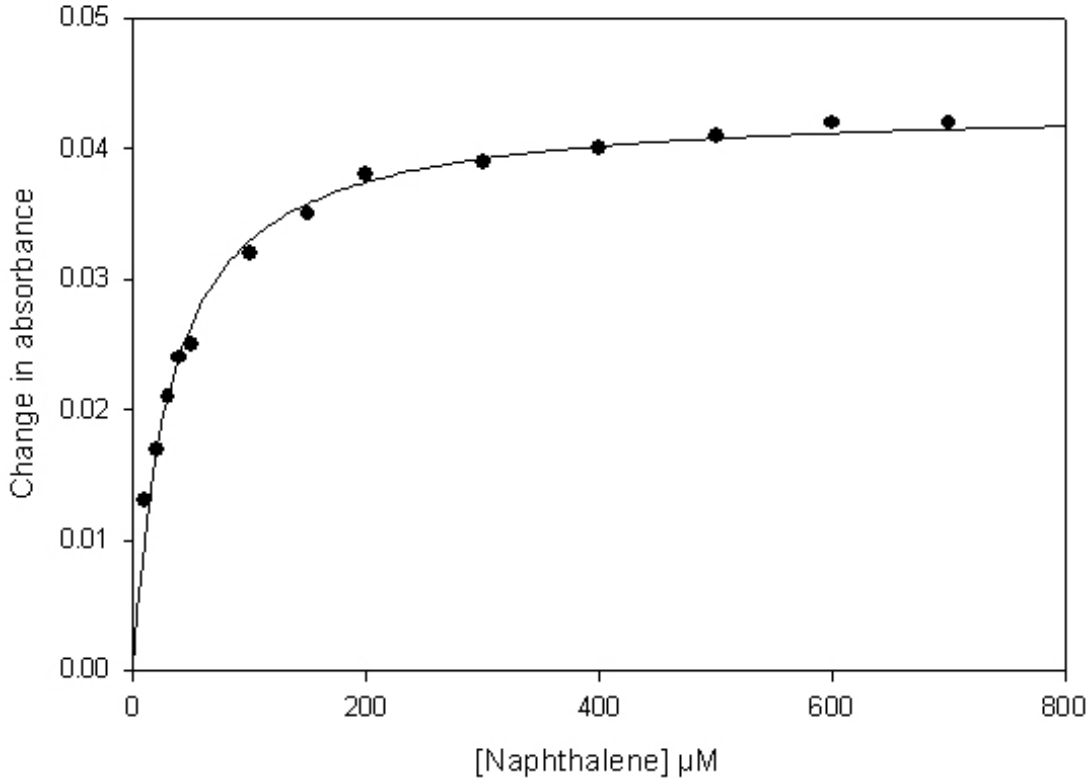


Figure 4.9 Change in absorbance recorded when naphthalene binds to new variant EAA
Changes in absorbance were calculated with 2µM enzyme and increasing concentrations of naphthalene.

The change in absorbance between peak and trough in the difference spectrum for increasing concentrations of naphthalene was measured, with the results plotted to determine binding constants for the enzyme. Figure 4.9 shows the binding curve plotted from the difference spectra of naphthalene binding to new variant EAA (enzyme concentration = 2μM). This allowed estimation of the binding constants to be made, with an estimated K_d of 31± 2μM and an estimated A_{max} = 0.04 calculated.

Also of note is the change in character of the enzyme upon binding naphthalene, as well as the drop in Soret peak. Figure 4.10 shows EAA in the native form and with naphthalene bound. A decrease in 6-coordinate character is seen, with a decrease in the Soret peak but increase in the Soret shoulder region (seen in the wild type HRP, in Chapter 1, Figure 1.10). This is more pronounced in variant AAEH, shown in Figure 4.11. The addition of 50μM naphthalene to the solution causes drop a in Soret maxima as seen with EAA, but additionally the loss of the α/β bands at 540nm, indicating that a more 5-coordinate species had formed as a result of the naphthalene binding.

Following UV/Visible evidence that the naphthalene binds to the enzyme, naphthalene oxidation assays were carried out as described in Kluge *et al.*, (2007) with minor modifications as noted.

4.2.5 Naphthalene oxidation assays

Assays were carried out using quartz cuvettes in a Shimadzu UV/Visible spectrophotometer, at 25°C. Reactions were started with the addition of enzyme, mixing was achieved by inverting the cuvettes before placing them into the spectrophotometer. This resulted in a short delay at the start of the reaction when no readings were taken, however the reaction was quite slow and therefore this delay was not significant.

Wild type HRP-C* showed no ability to oxidise naphthalene, even at concentrations in excess of 100μM enzyme, confirming previous reports that wild type horseradish

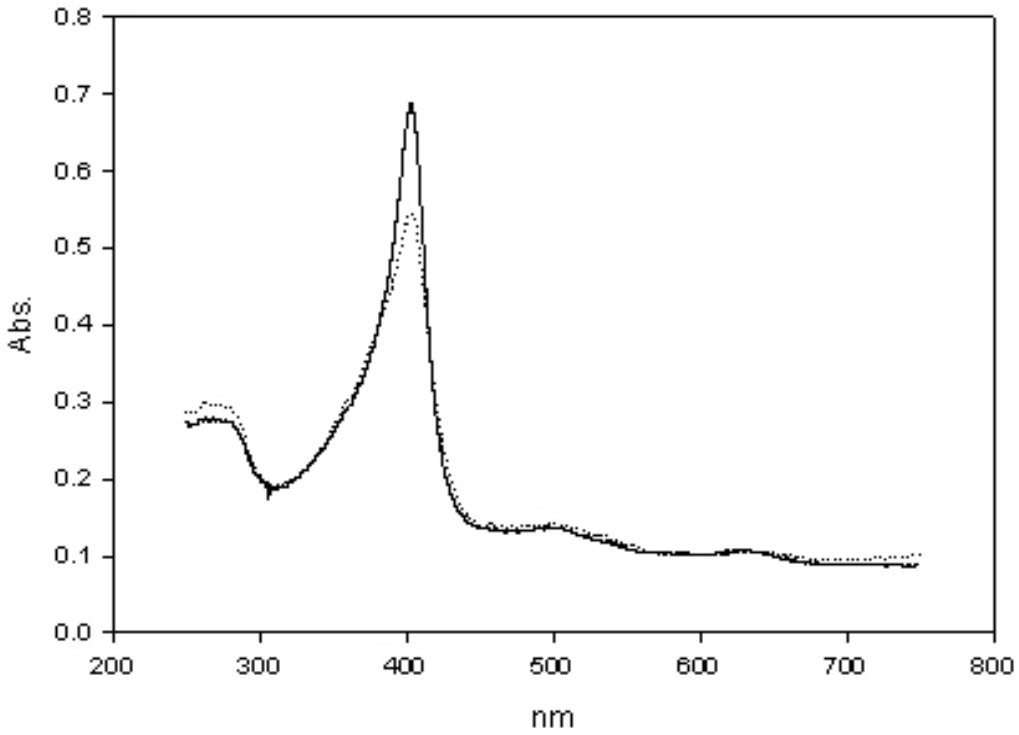


Figure 4.10 EAA in potassium phosphate buffer, with and without 50mM naphthalene added
Pure enzyme is shown in black, with the dashed line indicating the addition of naphthalene to the enzyme solution.
Addition of naphthalene results in a strong decrease in Soret maximum, with increase in Soret shoulder, also CT band has shifted to approximately 650nm. This may indicate an increase in 5 co-ordinate character.

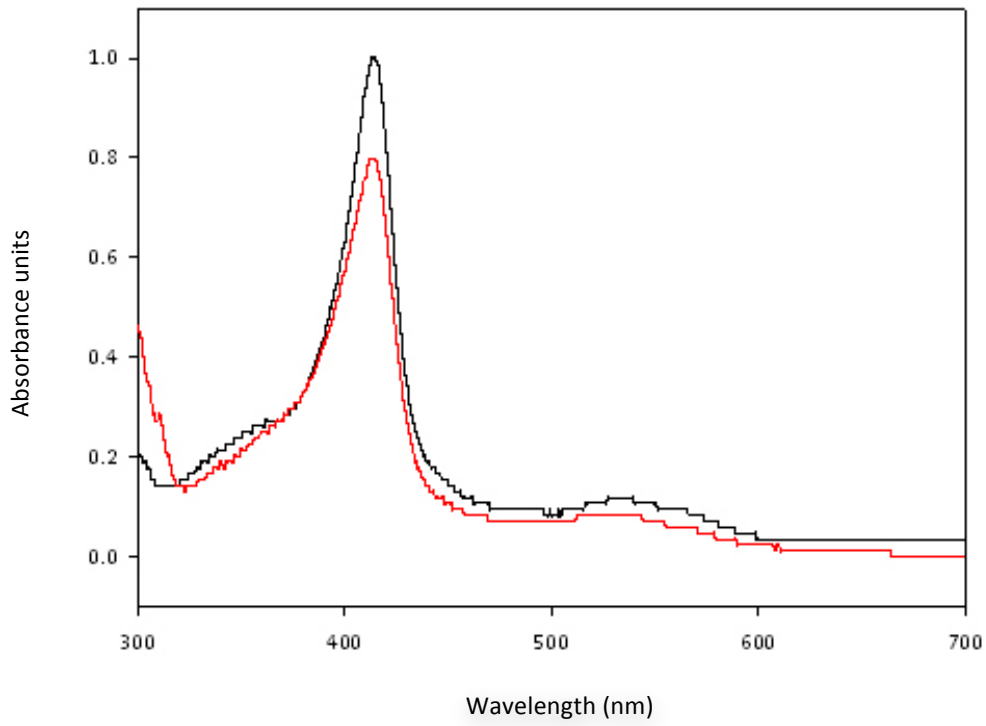


Figure 4.11 AAEH in potassium phosphate buffer, with and without 50mM naphthalene added.
Spectra adjusted so Soret = 1 absorbance unit for unbound AAEH. Unbound AAEH is shown in black, with the red line indicating addition of naphthalene to the solution.
Addition of naphthalene results in a strong decrease in the Soret maximum at 403.5nm, though not a shift in this peak. The slope of the peak is also altered, the appearance of a slight shoulder region at 380nm, somewhat resembling that of the wild type, though less pronounced.

peroxidase enzyme does not perform naphthalene oxidation. This is in line with findings from the previous work showing that wild type HRP does not carry out aromatic sulphoxidation or epoxidation of styrene due to the closed haem cavity and substrate access to the reactive centre only being possible through the haem edge.

Rates of naphthalene oxidation activity were also examined in a previous variant AAEH, characterised by Ngo (2006), in addition to the new variants AAD and EAA discussed in this thesis. It was predicted that the increase in size of the active site architecture which facilitated the epoxidation of styrene shown in the previous work (Ngo 2006 and Schneider, unpublished) would enable the oxidation of naphthalene as reported for the new peroxygenase enzyme *Agrocybe aegerita* (Kluge *et al.*, 2007).

Both variants AAD and AAEH, with the charged residue found at the 42 position in the active site showed very low activity towards naphthalene, despite the observed changes in the UV/Visible spectra upon binding. AAEH gave a rate of turnover of 3.1 min^{-1}, compared to the wild type rate of <0.7min^{-1}.

New HRP peroxygenase variant EAA however displayed significant naphthalene oxidation activity compared to the wild type. Production of naphthol products (α and β naphthols) was observed at a rate of 124 ± 4 min^{-1}. Whilst this rate is approximately 75 fold slower than the *Agrocybe aegerita* published rate (9960 min^{-1} Kluge *et al.,* 2007), it is a significant advance towards production of naphthols by an engineered peroxidase enzyme.

Naphtol products were confirmed by GC/MS analysis, based on the process published by Kluge and co-workers (2009) for naphthol identification. This GC/MS technique allowed identification of the naphthol product and that both α and β naphthols were produced, however this confirmation did not give an indication as to what proportion of each product was produced. Figure 4.12 shows the GC/MS trace confirming presence of naphthol product produced during oxidation of naphthalene by enzyme EAA. Bands on Figure 4.12.a are not indicative of the quantity of each product (α/β

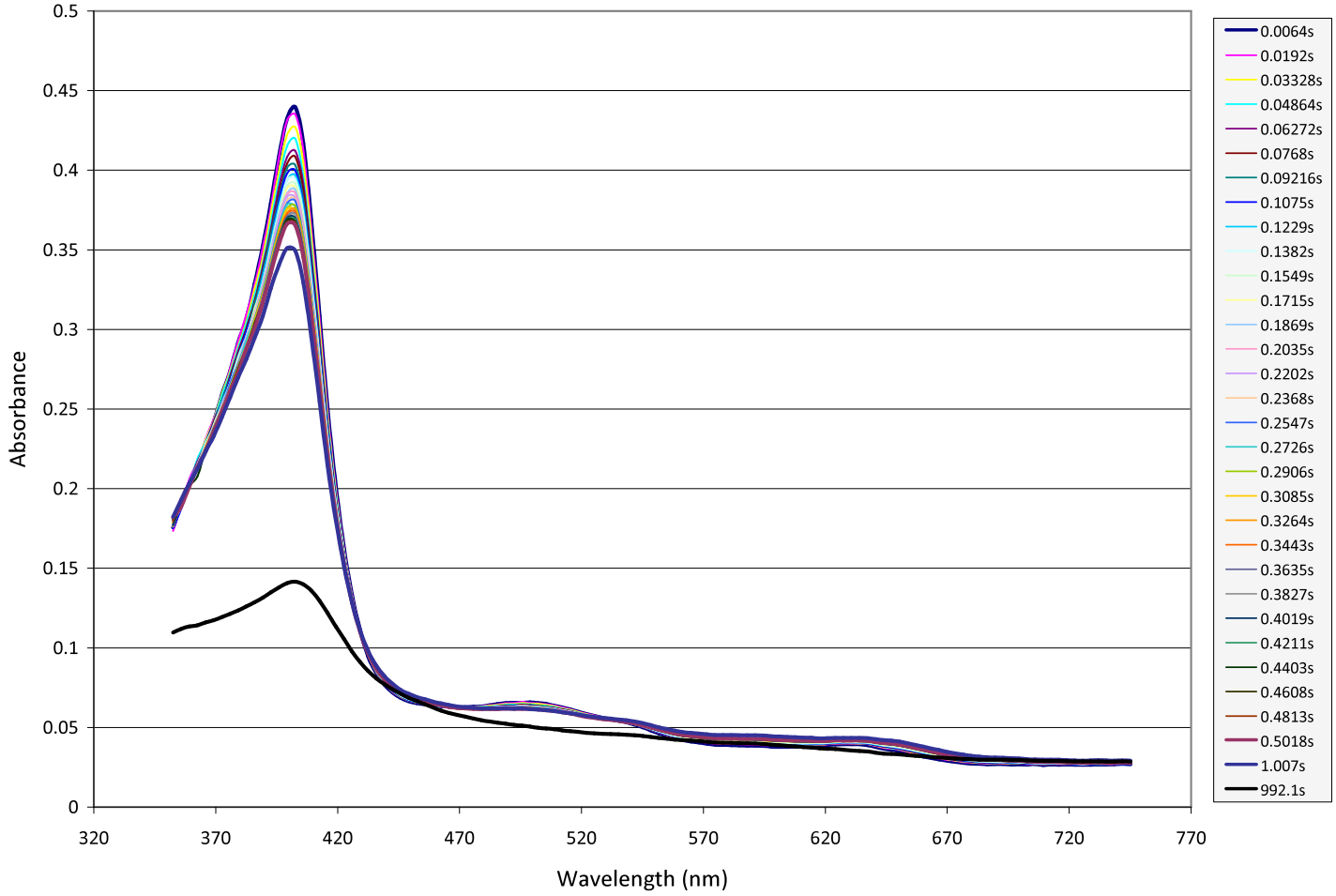


Figure 4.6 Haem bleaching of EAA shown using stopped flow spectrophotometry
Increasing concentrations of hydrogen peroxide resulted in a bleaching effect on the haem with variants EAA and AAD, with no apparent substrate shielding of the enzyme (as seen in variant AAEH with substrate bound where binding of substrate appears to prevent haem bleaching. Ngo 2007).

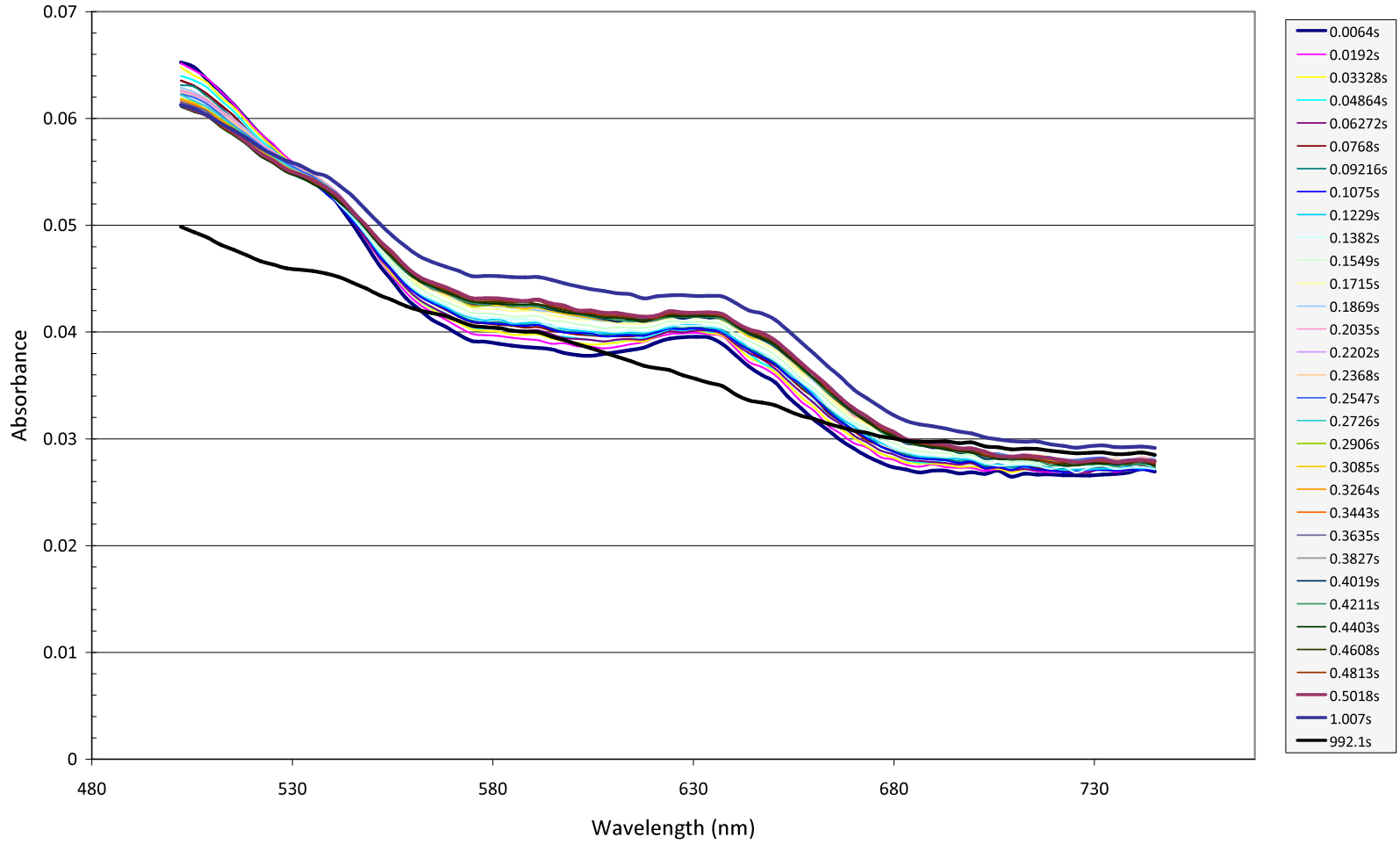


Figure 4.7 Haem bleaching effect on charge transfer band region (500-750 nm)
As seen with the Soret region (around 400nm), the bleaching causes loss of haem integrity and eventual complete degradation of the haem resulting in its loss from the enzyme. This is indicated in this region by the loss of charge transfer bands between 500 and 650nm.

naphthols) present, however two bands in Figure 4.12.a show elution at 12.26 and 12.46 minutes respectively, indicating two products with equal mass but slightly different elution times. Figure 4.12.b, showing the elution of product with M_r128, indicates that a quantity of unreacted naphthalene is still present in the mixture.

The GC/MS analysis confirmed the result shown by the UV/visible analysis of the reaction – that naphthols were forming when naphthalene was incubated with enzyme EAA in the presence of peroxide, but as seen in Kluge *et al*, a mix of α and β naphthols are formed.

Further GC/MS experiments were carried out with the products when naphthalene was incubated with enzymes AAD and AAEH in the presence of peroxide. These also supported the results seen in the UV/visible experiments – that there was no naphthol products detectable as the reaction proceeded at such a slow rate, particularly in comparison to the same reaction with EAA.

4.2.6 Naphthalene oxidation assays pre steady state

The reaction of variant EAA with peroxide in the presence of naphthalene was examined using stopped flow UV/Visible spectrophotometry. This enabled the pre steady state to be examined and gave indications into the complexity of this reaction in the steady state.

Figure 4.13 shows the time points taken through the first 10 seconds of the reaction. Initial enzyme-naphthalene complex is shown in dark blue, which has a lower peak at 403nm than observed with pure enzyme alone, as described previously (see Figure 4.8 showing how the presence of naphthalene affects the resting state spectrum of the enzyme). The reaction towards steady state of this complex with hydrogen peroxide can be followed. It is clear that in this reaction no Compound II is observed, this would be indicated by distinct red-shifting of the Soret peak and clear bands in the visible region. This Figure clearly shows the production of Compound I upon mixing of the

solution containing enzyme and naphthalene with hydrogen peroxide and seems to indicate Compound I as the steady state intermediate. The initial drop is represented by the pink trace (0.1s) and the steady state intermediate of Compound I by the yellow peak (1s). The drop in absorbance upon Compound I formation is not as pronounced as expected due to the decrease in absorbance noted on naphthalene binding previously observed. For wild type enzyme a reduction in the intensity of the Soret peak by 50% is anticipated upon Compound I formation, whereas this spectrum shows a decrease of ~20% as a consequence of the naphthalene complex being more 5-coordinate than the resting enzyme.

Figure 4.13 also shows evidence of the generation of the naphthol product, α-naphthol or β-naphthol, shown by an increase in absorbance at 350nm. An increase at the blue end of the spectrum corresponds with the predicted increase in naphthol formation as described in Kluge et al (2007), though the increase at 303nm tracked in the steady state reactions using a conventional rather than stopped flow spectrophotometer are not able to be observed reliably due to increased scattering.

The steady state is reached after approximately 0.2 seconds, after the observed formation of Compound I. The spectrum of the steady state intermediate is seen in Figure 4.13 prior to the observation of an increase in wavelength below 360 nm indicating product formation.

The cyan trace in Figure 4.13 also shows an increase in absorbance across the entire spectrum, symptomatic of the aggregation and precipitation of the protein in the light beam resulting in light scattering.

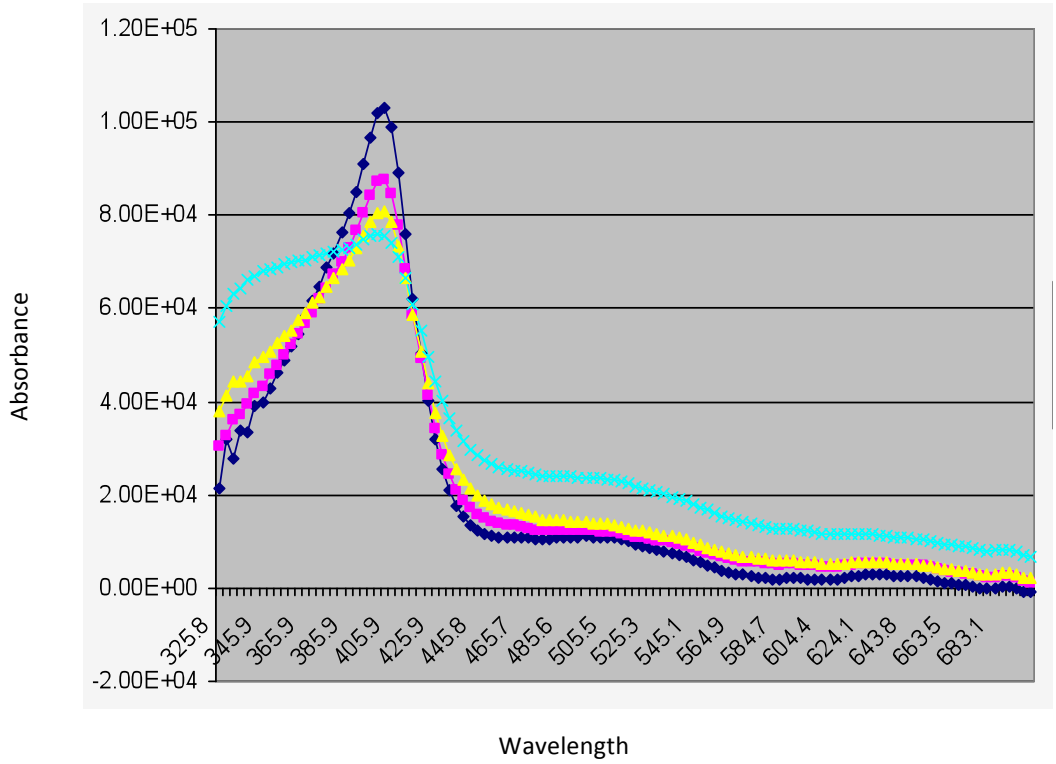


Figure 4.13 Naphthalene oxidation as recorded over 10s using stopped flow spectrophotometry
The initial mix, shown by the blue trace where t=0s, shows a decreased Soret peak when compared to unmixed enzyme. This means that the steady state Compound I (yellow trace, t = 1s) does not show the full characteristic decrease in the Soret peak, however levelling of the region between 500nm and 650nm occurs.
The cyan trace shows an increase across all wavelengths (t=10s), however this is more pronounced in the region around 350nm where naphthol product is formed. The increase across all wavelengths would be expected to be as a result of light damage to the protein causing random scatter, however the extra increase at 350nm would be expected as naphthols build up.

4.3 Crystallography of peroxygenase variants

Following the solution of a crystallographic structure for the first generation HRP peroxygenase variants, namely HAE, attempts were made in this project to crystallise AAE, AAEH and EAA, the subject of this thesis, since scope for the structure solutions of the new 'engineered' substrate complexes existed.

The HAE structure was solved using conditions similar to that used in the crystallisation of the wild type HRP by Henriksen *et al.* (1999). This structure was solved to 1.7Å and enabled comparison between the closed architecture of the wild type HRP-C active site. An overlay of these two structures is shown previously in Figure 4.4. This overlay indicates how the removal of the large arginine, phenylalanine and histidine residues in the active site of the enzyme opens up the haem cavity considerably, even with the introduction of the histidine moved from position 42 to 38 in an attempt to maintain the acid base components of the mechanism known to be vital in the formation of Compound I.

Whilst it is obvious that the further reduction in the size of amino acid side chains of residues in the active site will allow potential new substrates to access the oxy-ferryl haem more directly rather than at the haem edge as seen in other Class III plant peroxidases, understanding of the specific substrate interactions would enable further engineering to be specifically directed.

Previous work to crystallise AAE or AAEH, in collaboration with Dr. Nigel Brissett, produced plate-like crystals of the variant AAEH. These diffracted poorly, and whilst an attempt was made to solve the data using a molecular replacement solution with model 7ATJ, R values did not converge and resolution was poor, leading to further efforts on the data being abandoned.

In a renewed attempt to produce crystals suitable for structure solution from the engineered peroxygenase variants, a variety of crystal screens were set, initially based

around the previously published conditions (Henriksen *et al.,* 1999), however these did not result in formation of crystals. Pre-prepared screens (PACT Screens 1 & 2, Structure Screens 1 & 2, Clear Strategy Screens 1 & 2) were all prepared with varying concentrations of protein, however none of these screens proved successful in producing diffraction quality crystals. Microcrystals of AAEH were observed in drops containing sodium potassium phosphate (0.02M Na/KPO_4 20% w/v PEG 3350 from PACT Screen 2, well 10), with which optimisation of conditions was attempted, however only microcrystals or precipitate were seen.

It was thought that the process of freezing and thawing protein samples was disrupting the protein and causing aggregation of the protein which may have been disrupting crystal formation. This led to crystallisation trials being set in multiple conditions from protein that had not been subject to freezing. The protein was extracted from *E.coli* cell pellets and prepared in the way described in Chapter 2, then following purification by cation exchange chromatography was concentrated to approximately 8 mg/ml and crystallisation hanging drop trays set directly.

Unfortunately neither approach resulted in stable crystals that were suitable to subject to X-ray diffraction, either in-house or at a synchrotron source.

Figure 4.14 shows a photograph taken showing crystal growth along a hair that was accidentally found in a drop containing 20% PEG 8k, 0.1M cacodylate at pH 6.5 and 0.2M calcium acetate (in a screen around the wild type conditions). These were the largest crystals obtained during the attempts to crystallise variants AAE, AAEH and EAA as documented in this thesis and were thin, unstable crystals. When disrupted using a fine needle the crystals rapidly broke up and dissolved.

Heterologous seeding attempts were made using crystals of HRP variant S167M. The successful crystallisation of this variant is described in Chapter 5. These crystals were

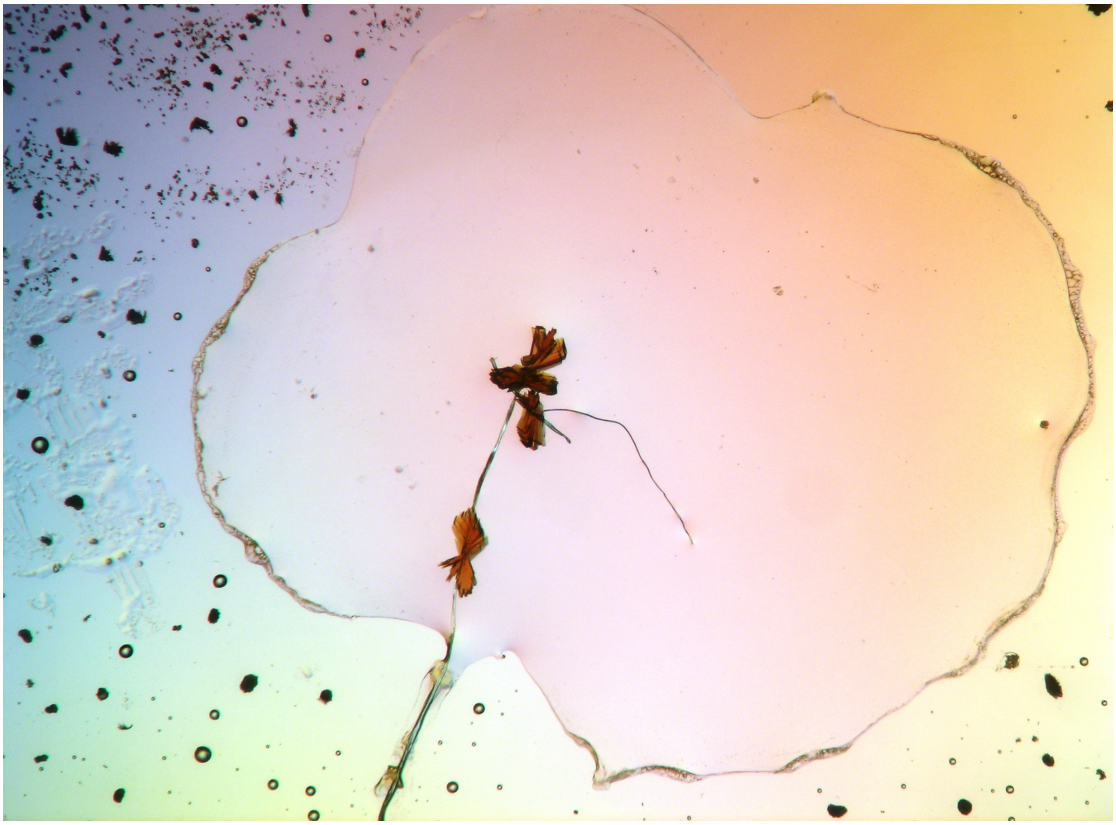


Figure 4.14 Crystals grown from variant AAEH
Crystals were observed growing along a hair laid across the protein-precipitant drop. These appeared to be three dimensional and not the same plate-like form as observed previously, however when interrogated with a whisker they were dislodged from the hair and subsequently dissolved back in to the drop. Further attempts were made around these conditions however none produced similar crystals.

disrupted using a needle and then the seeds drawn through supersaturated drops. Small needle-like crystals formed in seeded drops, however none of these resulted in crystals suitable for diffraction despite various optimisation attempts including altering of salt and PEG concentration.

The further attempts to crystallise these protein variants, described above, have confirmed what was thought previously – that although these variants are capable of carrying out new activities not previously seen in plant peroxidases, they are not stable and behave distinctly differently to the wild type HRP. This suggests a change in overall conformation of the protein aside from the obvious disruption of the haem cavity. It was previously suggested that the changes made to the haem cavity resulted in a shift in the F' loop region of the structure and, whilst this author cannot prove this fact, it would seem that the evidence shown suggests some disruption to crystallographic contacts resulting in unstable crystal formation. This may be caused by shifts in individual residues, collapse of the active site or, as suggested previously, by a substantial shift in part of the HRP tertiary structure.

Modelling of the catalytic site is possible and has been done on a limited scale based on the available structure of the wild type enzyme and HAE variant. This possibly represents an oversimplified picture as the catalytic evidence, particularly the stability of variant AAEH in high peroxide levels, strongly suggests that the overall structural changes are more significant than can be assumed by a simple molecular model.

4.4 Conclusions

Engineering of the active site of HRP has led to catalysis of non-native substrates which are of significant potential interest. Previous variants demonstrated the engineered ability to perform epoxidation and sulphoxidation reactions (Schneider *et al.*, unpublished data, Ngo and Smith, 2007).

Recent reports of the newly identified 'chloroperoxidase-like' enzyme in *Agrocybe aegerita* (Omura, 2005) with both peroxidase and peroxygenase function indicated a

further avenue of exploration, based on the publication (Kluge *et al.*, 2007) of results indicating that the AaP peroxidase/peroxygenase was capable of direct hydroxylation and C-H bond activation.

Direct oxidation of naphthalene to α- or β-naphthol is not a natural catalytic activity of wild type HRP, where substrates interact with the haem via the haem edge and the radical products are oxidised as the result of one-electron transfer steps (Dunford 1976, Poulos and Kraut, 1980). This reaction has been shown to occur both in AaP but, significantly, in engineered HRP variants in which features of the chloroperoxidase active site have been mimicked. This activity is most notable in variant EAA, in which the glutamate has been moved from the 42 position examined in previous variants AAE and AAEH. EAA, which is capable of direct oxygen transfer to naphthalene at a rate of 124 ± 4 min^{-1} , performs similarly to previous variants in terms of reduced activity towards the traditional peroxidase substrate ABTS yet shows an increased rate of naphthalene oxidation of approximately 40-fold over the best of the previous generation of peroxygenase variants (AAEH). Whilst this rate is still significantly slower than the rates reported for AaP (9,960 min^{-1}, Kluge *et al.*, 2007) it represents a new activity unable to be performed by the wild type HRP (<0.7 min^{-1}).

This engineering of the catalytic site of HRP represents a step towards the enzymatic C-H bond activation under conditions which are not considered chemically harsh (Clark, 1999, in Ullrich 2005).

Chapter 5

X-ray crystallographic analysis of an HRP variant, capable of autocatalytic cross-linking between the haem and the protein, S167M

5.1 Introduction

Production of the S167M HRP variant, its characterisation and functional analysis was carried out during the doctoral work of Dr. Khasim Cali (Cali, 2009). The results of this work will be remarked upon during the discussion of the crystal structure and the further investigation of the covalent linkage that is the subject of this chapter.

It has been shown previously, by Oxvig and co-workers, that the haem-protein linkages in eosinophil peroxidases are only partially formed during protein folding (Oxvig *et al.*, 1999). This work, along with the work of DePillis *et al* (1997), suggests a mechanism of processing for mammalian peroxidases, in which exposure to hydrogen peroxide induces formation of two fully formed ester linkages between the haem and protein. During this work it was found that the partial linkage results in a lower catalytic activity in the protein prior to exposure to hydrogen peroxide (DePillis *et al.*, 1997, Oxvig *et al*, 1999). The autocatalytic formation of the linkages leads to fully functionally active protein (DePillis *et al*, 1997).

Covalent linkages have been explored in plant peroxidases previously in an attempt to elucidate the mechanism of formation of these linkages in native mammalian peroxidases where they occur naturally but for which no mechanism of formation has been proposed in detail (Metcalfe *et al.*, 2004). In particular, the creation of the ester linkages in HRP by Colas and Ortiz de Montellano, and the sulphonium linkage in pea cytosolic ascorbate peroxidase by Metcalfe *et al.*, both in 2004, are of interest to this work.

It has been suggested more recently (Metcalfe *et al.*, 2007) that, under oxidising conditions, suitably located residues, particularly those that can be activated by radical

chemistry, in the protein backbone could undergo covalent attachment to the haem. Similarly, under reducing conditions covalent attachments can also be formed, exemplified by the S160C ascorbate peroxidase variant produced by Metcalfe *et al.* (2007).

To further explore the process of formation and stability of designed covalent haem-protein linkages, and affects on protein e.g. preventing the loss of the haem group, the S167M HRP variant was produced using site directed mutagenesis by Dr. Cali during his doctoral work. It has been suggested that the methionine to haem sulphonium linkage confers unique chlorination activity to myeloperoxidase, since variants in which the methionine of the sulphonium linkage is replaced by threonine or valine (the amino acids seen in the corresponding positions in other mammalian peroxidases) have a reduced affinity for Cl^- of up to 100-fold (Kooter *et al.*, 1999). The covalent linkage is also cited as contributing to the unusual properties of the myeloperoxidase enzyme.

Since mammalian peroxidases and plant peroxidases do not share a great deal of sequence homology and come from different fold families (Colas and Ortiz de Montellano, 2004), superposition of the myeloperoxidase structure (Fiedler *et al.*, 2000, PDB id: 1CPX) and HRP structure (Henriksen *et al.*, 1999, PDB id: 7ATJ) on the central haem was carried out to visually identify a residue for mutational targeting. Hence the serine at position 167 was found to be spatially equivalent in position to the methionine 243 in myeloperoxidase that forms the covalent sulphonium linkage to the haem prosthetic group. This residue is the equivalent in HRP to the serine at position 160 selected as a mutational target in ascorbate peroxidase by Metcalfe *et al.* in 2004, further confirming the likelihood of correct selection of the target residue. HRP was subjected to site directed mutagenesis by Dr. Cali, leading to the replacement of the serine at position 167 with a methionine (Cali, 2008).

The S167M variant was produced and folded by Dr. Cali and, through HPLC analysis and haem partitioning experiments, the purified and folded protein variant was shown not to contain a covalent attachment between the protein and the haem. This protein was therefore designated umS167M: unmodified S167M. When the umS167M variant

was subject to exposure to 10 stoichiometric equivalents of hydrogen peroxide for 2 hours, it was found that covalent attachment of the haem to the protein had occurred. The formation of the linkage was shown firstly by acid butanone extraction (described by Fuhrhop and Smith, 1975), where haem was unable to be extracted from the protein.

HPLC analysis, shown previously by other groups to be a reliable way of determining haem-protein linkage (Colas *et al.*, 2002), was carried out on both unmodified and modified S167M HRP. Co-elution of the protein and haem components is indicative of haem-protein linkages, as non protein-linked haem results in distinct peaks in the HPLC chromatogram. In both wild type HRP and unmodified S167M HRP, haem and protein components eluted in separate peaks. The mS167M variant showed co-elution of haem and protein peaks, with the disappearance of the separate haem peak from the elution trace.

Peroxidase (native)	**Soret maximum (nm)**	**Further maxima in resting state (nm)**	**References**
Myeloperoxidase	430	496, 570, 620, 690	Furtmüller *et al.*, 2006
Recombinant myeloperoxidase variant Met243Thr	413	500, 542, 590, 638	Furtmüller *et al.*, 2006
Lactoperoxidase	413	500, 550, 600, 650	Andersson *et al.,* 1996
Eosinophil peroxidase	413	500, 542, 580, 638	Furtmüller *et al.*, 2006
Horseradish peroxidase C*	403	380nm (Soret shoulder)	
umS167M	406		Cali, 2008
mS167M	417	540, 564	Cali, 2008

Table 5.1 Spectroscopic features of plant and mammalian peroxidases, with HRP C* variant S167M

The UV/Visible spectra of the haem centre of mammalian peroxidases is similar to that of the haem in HRP, however spectral changes are noticeable, thought to be a result of the distortion of the protoporphyrin IX haem environment caused by the covalent haem-protein linkages. Table 5.1 shows the key UV/Visible spectral features of plant and mammalian peroxidases, to which the umS167M and mS167M HRP variants were compared.

The UV/Visible spectrum of this variant was seen to be altered when compared to the spectrum of wild type HRP. In samples of umS167M HRP, the Soret maximum had red shifted slightly, appearing at 406nm rather than at 403nm seen in the wild type HRP. The molar extinction coefficient of this peak, determined by the pyridine haemochrome method, was shown to be 129±3$mM^{-1}cm^{-1}$, compared to the published wild type molar extinction coefficient of 102$mM^{-1}cm^{-1}$ (Dunford, 1999). This increase is thought to be caused by the replacement of the serine side chain with the methionine, causing an increase in the proportion of six-coordinate high spin character.

A considerable red shift in the Soret maximum was seen in the mS167M variant as a result of the modification, where the Soret maximum red shifted from 406nm to 417nm. Additionally, the charge transfer bands (CT I and CT II) are no longer seen and instead are replaced by peaks at 540nm and 564nm. The UV/Visible properties of the mS167M variant suggest a low-spin haem character.

The mS167M variant was seen to be labile and, when aged for a number of days at 4°C, the covalent sulphonium linkage was partially lost, releasing a modified haem species of mass 631.7 on MALDI analysis. Evidence for partial spontaneous formation of what appeared to be a covalent linkage between the haem and the protein was also obtained. When the unmodified S167M HRP sample was subject to acid butanone extraction of the haem, 14% of the expected haem mass was unrecoverable (Cali, unpublished data). This observation suggests that a small amount of the cross-linked enzyme is forming spontaneously, without exposure of the enzyme to hydrogen peroxide. This feature is progressively observed upon aging of the umS167M HRP

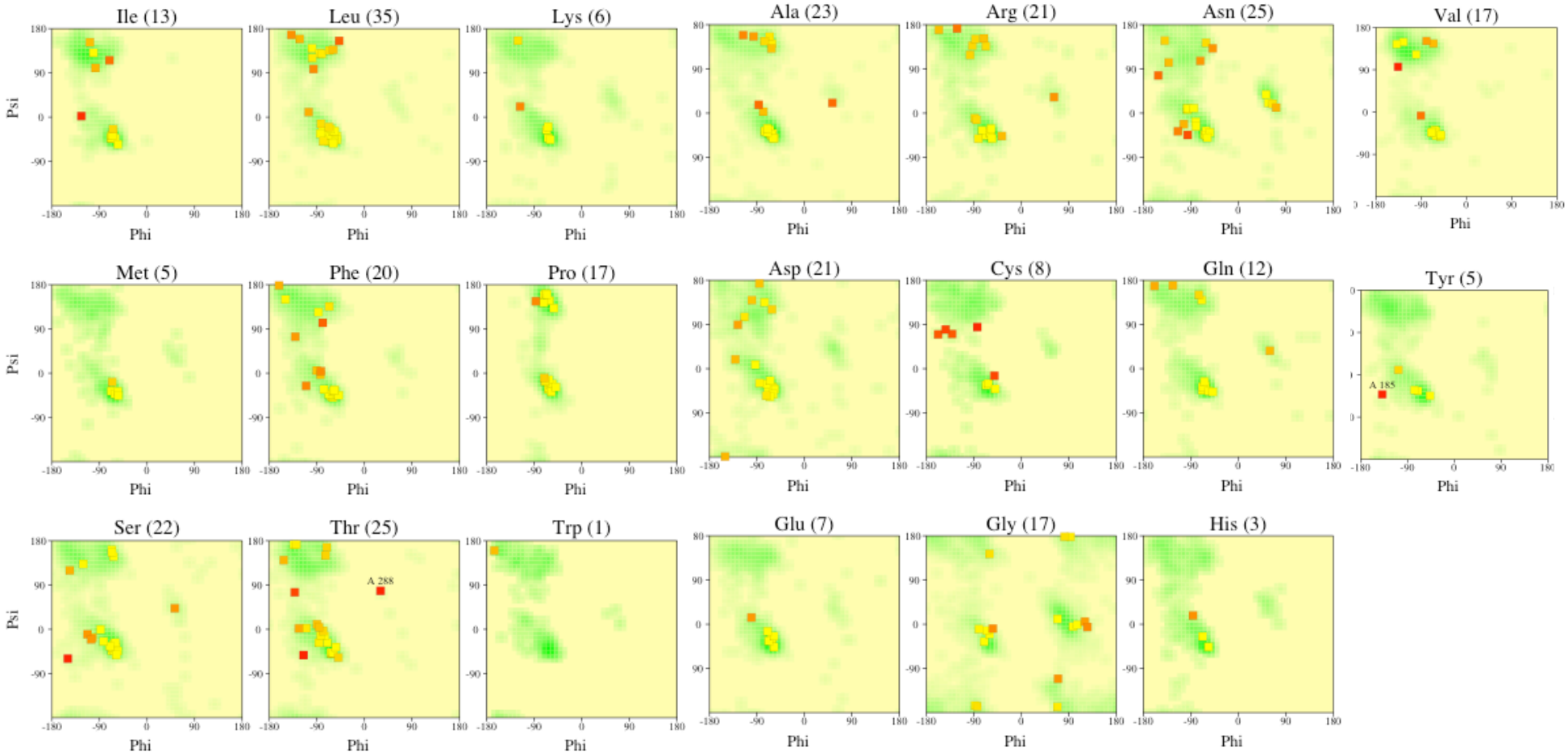


Figure 5.4 Ramachandran plots for individual residues in the umS167M structure

Examination of individual Ramachandran plots for each residue shows two residues highlighted (Thr 288 and Tyr 185) that caused alert due to their bond conformation. These residues were identified within the structure and each examined to ensure the model fit the electron density data correctly. Small changes were made, but the general fit to the data was good.

variant at 4°C for several hours (Cali, unpublished data). It was not in the scope of the previous work to investigate this spontaneous linkage formation fully, however it was noted that this linkage was not stable and there appears to be slow breakage of the covalent bond. The focus of the previous work was the autocatalytic formation of the sulphonium linkage in the S167M HRP variant. This chapter examines work carried out in probing the structure of the variant.

5.2 Results and discussion

5.2.1 Protein crystallisation

A pure sample of unmodified S167M HRP (umS167M), concentrated to 10mg/ml, was used in crystallisation trials. Screens were constructed that were based on the wild type HRP crystallisation conditions published in Henriksen *et al.* (1999) of 20% w/v polyethylene glycol (PEG) 8000, 0.2M $Ca(CH_3COO)_2$, 0.1M cacodylate buffer pH 6.5. Percentage PEG, concentration of $Ca(CH_3COO)_2$ and the concentration of cacodylate buffer were changed in a methodical pattern over the screen ensuring a range of conditions around the previously published conditions were checked. The screens were carried out in 24 well Linbro plates (Molecular Dimensions Ltd.) with cover slips sealed using high vacuum grease (Dow Corning) (see Appendix 2 for full details of individual well conditions). The hanging drops were set using 1µl of an isopropanol solution saturated with ferulic acid, which was allowed to evaporate on the cover slip before 1µl protein solution and 1µl reservoir solution were added. This ensured that the protein and precipitant solution was maintained in a globular drop.

Trays were stored at 4˚C in a Rumed 3000 incubator, and drops were checked daily for crystal formation under a Nikon light stereoscope at room temperature.

After 72 hours crystals were observed in many of the drops. Figure 5.1 shows a grid of the screen conditions with some examples of the crystals seen. The fine crystals that appear in the background of the drops are likely to be ferulic acid crystals. These are observed to be dried, in crystalline form, on the cover slip but soluble in the drop solution. Ferulic acid was employed as a co-crystallisation additive (Henriksen *et al.,*

1999) in the wild type crystallisation trials. As a typical HRP substrate, it was found to sit in the active site access channel in the wild type crystals and thought to potentially stabilise the crystalline form. Ferulic acid was utilised in the same way during the crystallisation of umS167M HRP, as a co-crystallisation additive introduced to the protein on the cover slip. Later in this chapter the orientation of the ferulic acid in the active site access channel will be discussed as this was not clear from the original crystal structures.

5.2.2 X-ray diffraction

Directly prior to mounting the crystals in the X-ray beam, each crystal was immersed in a cryoprotective solution of 40% v/v polyethylene glycol 8000, 0.1M $Ca(CH_3COO)_2$, 0.1M cacodylate buffer, pH 6.5 for a short time. This 'quick dip' allowed the well solution surrounding the crystal to be at least partially replaced by the cryoprotective solution without risking the conditions changing substantially and causing the crystals to distort. Although distortion of the crystalline form was unlikely due to the similarity between the cryoprotective solution and the well solutions, crystals were treated as gently as possible to minimise the disruptive stress of changing solutions. The cryoprotective solution selected, similar to the cryoprotectant used in Henriksen *et al.* (1999), had previously been tested without protein crystals present to ensure that vitreous glass, rather than crystalline ice, formed when placed in the nitrogen cryostream at 100K. Crystals were visually examined under the microscope in the cryoprotective solution to ensure that transfer to this solution had not caused obvious defects to form in the crystal lattice.

The number of crystals grown in the screen around the wild type crystallisation conditions allowed a considerable degree of choice when identifying crystals to be X-rayed. Crystals were selected subjectively after examination under the microscope, based on clarity of the crystal (no visible cracking in the surface of the crystal), singularity (crystals which were at all twinned were not examined further), size and structure (crystals that appeared too thin in any dimension were discounted). Initial X-ray images were taken to determine the potential X-ray quality of the crystal and the judgement was made as to whether to proceed with full data collection at that point.

A crystal was mounted from drop A2 (10% PEG 8000, 0.1M $Ca(CH_3COO)_2$, 0.1M cacodylate buffer, pH 6.5). Data were collected on an R-axis IV++ imaging plate using a Rigaku RU-H3RHB copper rotating anode generator, with X-rays of wavelength 1.542Å, focussed using Osmine Confocal Max-flux mirrors. Two initial diffraction images with 5-minute exposure times, 1˚ oscillation angle and 90˚ separation were obtained with an initial crystal to film distance of 200mm. Although diffraction was recorded from this crystal, the resolution was in excess of 2.4Å. In this circumstance, with a large number of crystals available, this crystal was discarded in favour of trying others of the available crystals to find a crystal that diffracted at high resolution.

Drop B3 (Figure 5.2), with a well solution of 15% PEG 8000, 0.2M $Ca(CH_3COO)_2$, 0.1M cacodylate pH 6.5, was examined and thought to contain potential candidates for high resolution data collection. The large crystals, labelled (a) and (b_1) in Figure 5.2, were rejected as, though visual inspection suggested few obvious imperfections in the crystal surface, the size of these crystals was prohibitive. As shown in Figure 5.2, the crystals (a) and (b_1) are in excess of 500µm in length. The largest available mounting loop was only approximately 250µm in diameter, therefore the crystal would be much too large for the loop. In addition, the larger the crystal, the increase in likelihood of flaws occurring in the crystal formation, for example an increase in disorder, subtle twinning of the crystal, or other unseen defects that would affect the quality of X-ray data collected. This means that potentially high-resolution crystals can produce poor resolution data. Also, as visible in Figure 5.2, crystal (b_1) has a small crystal labelled (b_2) which appears to be associated with the larger crystal. This could potentially cause twinned data if that area of the crystal was X-rayed.

Due to the various reasons detailed above, crystal (c) shown in Figure 5.2 was chosen as a likely candidate for high-resolution data collection. The crystal appeared singular, with no visible faults in the crystal surface. This crystal was immersed in the cryoprotectant as above and mounted as previously reported in Chapter 2.5.2. Again two initial images were collected with 5-minute exposure times, 1˚ oscillation angle, 90˚ separation and a crystal to film distance of 200mm. The images collected

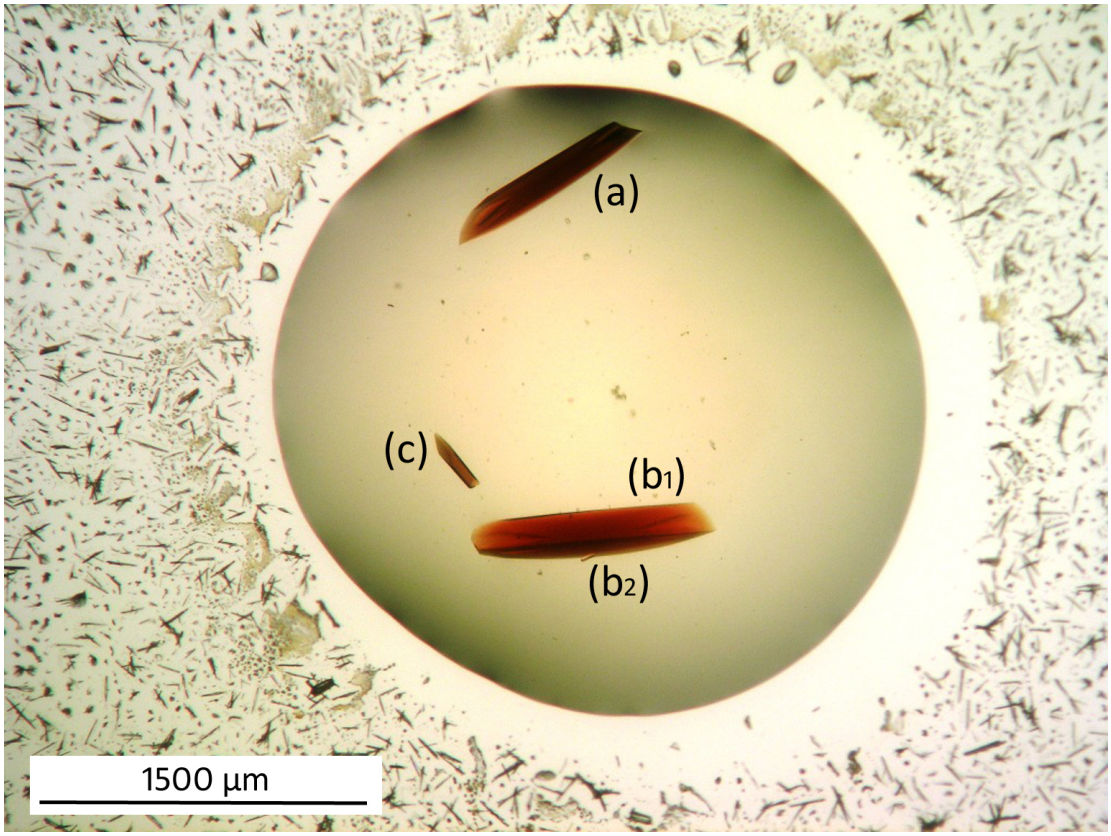


Figure 5.2 Crystallisation drop B3
Crystals grown in 15% polyethylene glycol 8000, 0.2M calcium acetate, 0.1M cacodylate pH6.5. 1µl isopropanol saturated with ferulic acid was placed on the cover slip and allowed to evaporate before adding 1µl well solution and 1µl of 10mg/ml protein to the cover slip. Crystal (c) was chosen to collect diffraction data from.

displayed high resolution diffraction so the image plate was moved closer to the crystal and another two images with 5 minute exposure times, 1˚ oscillation angle, 90˚ separation and a crystal to film distance of 160mm were taken. These images only showed clear diffraction to 1.7Å, so the image plate was set to a distance of 170mm and another set of two images collected as previously to confirm that the distance was appropriate.

Initial images were analysed using MOSFLM where a putative point group was assigned. This assignment allowed for rational data collection, ensuring that the optimal amount of data were collected giving the highest level of completeness with the minimum quantity data collected. A great deal of work has been carried out to establish the effect of ionising X-ray radiation on cryo-cooled protein crystals, for example a general overview in Ravelli and Garman 2006 and specifically on peroxidases in Berglund *et al*, 2002. This work has shown that collecting the minimum amount of data required to give a complete data set is important to avoid causing undue damage to the crystal and prevent unnecessary distortion to the dataset. The space group assignment for the umS167M crystal was P222, an assignment made based on both the diffraction images collected from this crystal, the crystal morphology and previous assignment of wild type HRP crystals (Henriksen *et al.*,1999). From the symmetry in the P222 space group, only 90° of data needed to be collected. MOSFLM's STRATEGY command was used to obtain data of the optimum angular range to collect to obtain the most complete data set: in one segment between 184° and 274°.

X-ray images were taken of the umS167M crystal from 184° to 274° with an oscillation angle of 1° per image. 90 images were taken, which gave a data set with 84.27% completeness. Each oscillation was exposed to the X-ray beam for 10 minutes, with a final crystal to film distance of 170mm.

A second, low resolution, data set was collected to attempt to improve the low-resolution data from the first set of data. These data could be used, if needed, to

supplement and improve the high-resolution data set. The same rotation was used in collection of the data as the high-resolution data, with oscillation angle of 1° per image. Again 90 images were taken, though the crystal to film distance was increased to 200mm to capture the lower resolution data with larger spacings.

5.2.3 Data processing

When collected, images were initially processed using MOSFLM. Firstly, the high resolution images were input to MOSFLM, spots assigned and then an attempt was made to autoindex the images, indexing each spot on each image. This standard process uses a complex algorithm to predict space group from a small number (<700) pre-assigned spots over a number of images, normally up to 90° apart (described in Powell, 1999). A list of potential space group assignments was given, with associated penalty factors for each. The highest symmetry option with the lowest penalty rating was chosen: P222, with a penalty factor of 5. Using this data an automatic assignment can be made to "find" spots on each of the other images in the data set.

The high resolution data set of the umS167M crystal did not autoindex correctly using spots taken from three images collected 45° apart. Predictions for spot locations were correct for some images but were distorted for others. Adjusting settings across the whole data set would allow different ranges of images to be predicted correctly. This problem was circumvented by splitting the dataset in to six segments and autoindexing these segments separately. This gave accurate spot assignments for the separated datasets, which were then processed individually, and the datasets merged prior to scaling.

Autoindexing each segment of data produced unrefined unit cell dimensions of: a=40.30Å, b=66.94Å, c=117.75Å, with angles α=90°, β=90°, γ=90°. This confirmed the previous assignment of an orthorhombic crystal system, with options for higher degrees of symmetry in space groups P222, $P222_1$, $P2_12_12$ or $P2_12_12_1$, although this could not be determined until systematic absences had been accounted for on the images. These unit cell dimensions were very close to the dimensions reported by Henriksen, Smith and Gajhede for the structure of wild type HRP in complex with

ferulic acid. In their reported crystal structure, the space group is shown to be $P2_12_12_1$, with unit cell dimensions of a=40.3Å, b=67.3Å, and c=117.6Å. With this knowledge, the umS167M HRP crystal symmetry was assigned as $P2_12_12_1$, and data processing proceeded. Cell dimensions were auto-refined using the REFINE CELL command, using two pairs of images with maximum separation for each segment.

Each of the segmented datasets were integrated individually using **MOSFLM** as part of the CCP4i suite (Leslie, 1992), running in "Batch mode". The integrated data were merged using **Sortmtz**. The integrated and merged data were processed using programmes of the CCP4 suite (Leslie, 1992). The Matthews coefficient was calculated using **matthews_coef** and a solvent content of the unit cell was established to be 41.35%, with one protein molecule per asymmetric unit. This was in agreement with the previous work in which wild type HRP had been crystallised with ferulic acid (Henriksen *et al.,* 1999). Data were scaled using **Scala** and **Truncate**, insuring that 5% of data were flagged in order to calculate R_{free} ratios during refinement. Scala was used to scale and merge intensity data from the MOSFLM output file. Truncate was used to convert intensity data in to structure factor amplitudes that could then be used in molecular replacement.

Data were examined in the light of the log files produced from the scaling and merging which led to a cut-off point being established for the data processing of 1.7Å from the 1.6Å collected, since completeness level decreased significantly towards the higher resolution range. This increased overall completeness of the data to >80% during refinement. The low-resolution data set was determined to be not of significant use to the model building process so this data was discarded and not examined further.

Molecular replacement was carried out initially using **MolRep**, with the co-ordinates of the wild type HRP structure, PDB id: 7ATJ solved by Henriksen *et al.* (1999), as the molecular replacement model co-ordinates. Table 5.2 below summarises the details of the model structure. The 7ATJ structure shows a number of alternative options for the specific rotamers of disulphide bonds and suggests three possible options for orientation of the ferulic acid molecule in the haem access channel. This will be

commented on later in the chapter, as the data collected in this work suggested only one orientation for this molecule.

Non-protein residues (modelled water molecules, haem moiety and the ferulic acid co-crystallisation additive) were removed from the co-ordinate file before molecular replacement took place, with the intention of modelling these in at a later point. In addition, the region around the substituted methionine residue, from residue 160 to residue 175, was removed. This removal of residues around residue 167 aimed to reduce model bias produced when using the model co-ordinates from the 7ATJ structure, a structure with only a single residue difference in amino acid sequence.

7ATJ Model Structure		
Cell parameters	a: 40.868Å b: 66.93Å c: 119.017Å	α: 90.00° β: 90.00° γ: 90.00°
Space group	$P2_12_12_1$	
Model		
Total atoms	2944 of which 440 are water molecules and 75 form the haem centre and additives to they crystallisation process	
Solvent content of model	48.6%	
Resolution range and data quality		
Resolution range	34.0Å – 1.47Å	
Completeness	91.2%	
Unique reflections	51544	

Table 5.2 Summary of data quality from HRP-C* model structure
Crystallographic modelling was carried out using the horseradish peroxidase C1a (recombinant) structure previously published by Henriksen, Smith and Gajhede (1999), known by PDB accession code 7ATJ.

The MolRep solution showed higher than expected R values with the initial model and a first round of rigid body refinement using **Refmac5**. For data of the quality obtained, it would be expected that R_{factor} and R_{free} would drop during the first round of automatic refinement, even before any manual refinement had taken place. As this was not seen with the MolRep structure solution this solution was treated suspiciously

and further programmes were used to produce a choice of models.

Amore and **Phaser** were used to search for alternative molecular replacement solutions, again a structural model with water molecules and residues around the site of mutagenesis removed was input. The solution from Amore was chosen to be the most reliable model structure to pursue with manual and automated refinement. The initial Rfactor and Rfree of this model were 34% and 39% respectively. After a single round of automatic rigid body refinement, the Rfactor and Rfree dropped to 28% and 32%, indicating that this model was a reasonable starting point for manual refinement of the structure. This initial round of automatic model refinement enabled the first $2F_o$-F_c to be produced, indicating regions of the model that needed initial attention where the model differed significantly from the data collected from the umS167M crystal.

Manual model refinement was carried out using **Coot**, allowing visualisation of the protein model in conjunction with the electron density map ($2F_o$-F_c and F_o-F_c). Residues were individually examined and manipulated to the best possible fit in the electron density map. Rotamer preferences for residues were viewed and rotomers that fit the electron density data without excessive extra manipulation were chosen where possible. Where adjustments needed to be made to fit the model to the electron density, these were made with consideration made to likely conformation of atoms.

Coot also enabled the modelling of the residues that had been removed in the molecular replacement step. Whilst the removal of these residues may have reduced model bias in the significant region of the molecule during the molecular replacement process, the reintroduction of these residues should further increase the accuracy of the model and hence reduce the R_{factor} and R_{free} accordingly. The primary amino acid sequence of the S167M HRP molecule was examined to ensure that the residues to be reintroduced were correctly identified. These were individually introduced by addition to the N-terminal of the gap. Coot suggested likely rotomers for the introduced residues that were examined in relation to the electron density map. In the case of

residues where the rotamer did not fit the density displayed, manual adjustment of each residue was carried out to allow best fitting to observed density.

There was clear density observed in the region of the serine to methionine mutation, allowing good modelling of the introduced residue. This density shows the methionine side chain stretching towards the haem group, though it was not as close as previously expected. The significance of the placement of this residue will be remarked upon later in this chapter. R_{factor} and R_{free} decreased on inclusion of the amino acid region around the S167M residue, to values of 25% and 29% upon restrained refinement.

Refmac5 (Version 5.2.0019, Murshudov *et al.*, 1997) was used in refinement mode to auto-refine the model with a restrained refinement against the original observed data after manual refinement in Coot had taken place. Restrained refinement, the default setting for refinement in Refmac5, maintains acceptable geometries within the model whilst allowing movement to better fit the electron density maps provided. 10 cycles of refinement were carried out each time Refmac5 was run on the data. The resolution range for the refined data was 1.7 Å – 22.22Å, with a completeness of 84.27% for the range. In total, 2685 non-hydrogen atoms were used in the refinement, including 258 water molecules, 85 atoms of the haem group and 13 atoms of the ferulate molecule.

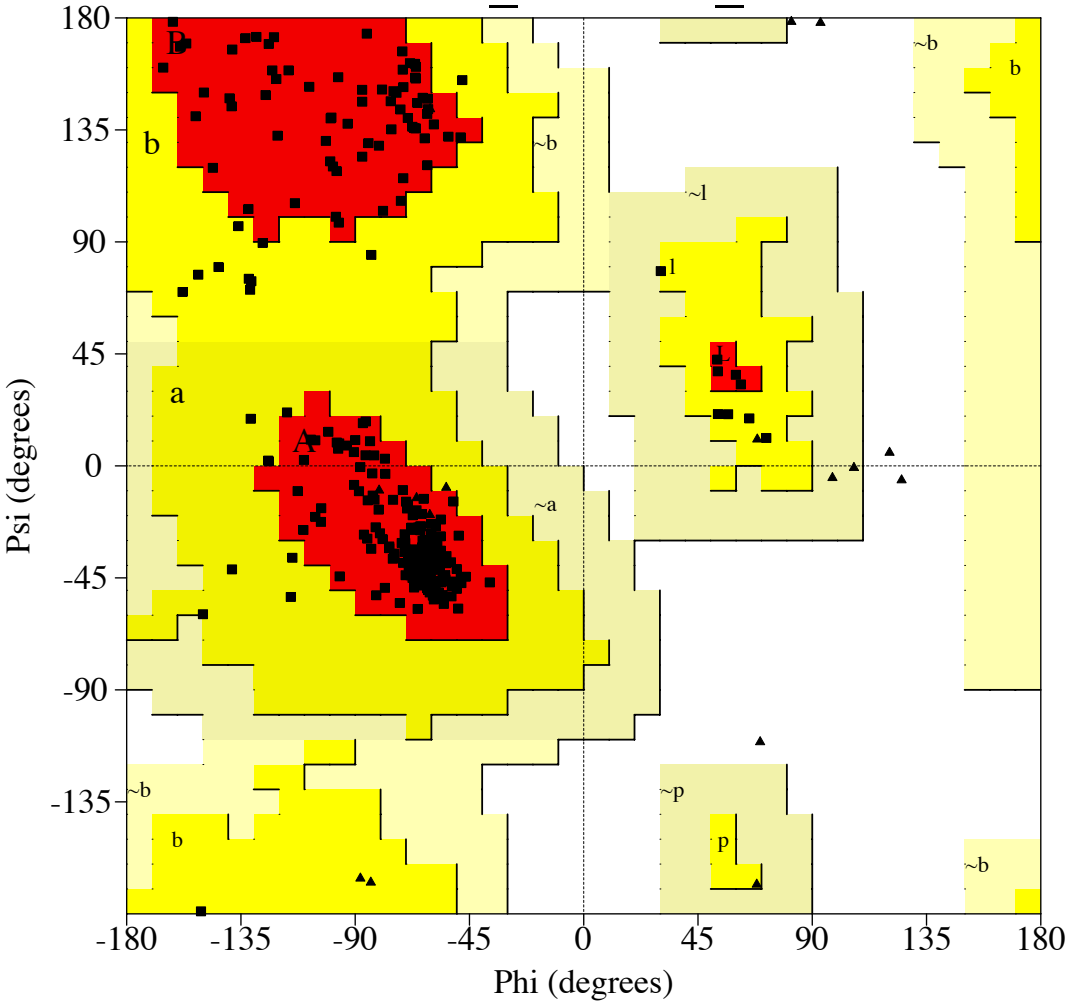
Refinement in Refmac5 produced new electron density maps and allowed the effectiveness of manual refinement to be followed with reduction in R_{factor} and R_{free} at the start of each Refmac5 run. R_{free} test set size was 5% of the original data selected randomly from throughout the data. Log graphs produced as a result of the refinement cycles of Refmac5 showed the decrease in R_{factor} and R_{free} per refinement cycle. Manual adjustment of the model for refinement purposes was ceased at a point where the R_{factor} and R_{free} did not significantly decrease after manual and automatic refinement cycles. At this point, the R_{factor} value was 17.4% and the R_{free} value was 23.8%. This was deemed adequate for the resolution of X-ray data initially collected and, though small reductions in R_{factor} could possibly have been made by further manual refinement, the effort required to achieve these minimal differences was

judged to be excessive and the refinement process was stopped. The correlation of the values of R_{free} to the R_{factor} throughout the refinement process shows that the model fits the data that was isolated at the start of the model building process.

Validation of the accuracy of the model structure was carried out using **ProCheck** (Laskowski *et al.*, 1993), a program that validates bond lengths and geometries from the model PDB file. This checking ensured that the model was within chemically likely parameters. Bond lengths and angles that fell outside these parameters were checked and rebuilt manually using Coot, ensuring that these alterations were consistent with the electron density maps for the refined structure. Figure 5.3 shows the result of the final ProCheck Ramachandran plot of the umS167M model structure and indicates that all residues, excluding glycine and proline, are within the allowable regions and that 91.1% of those residues are found within the most favourable regions. No non-glycine or proline residues are found in the generously allowed regions.

Figure 5.4 shows Ramachandran plots for each residue type in the polypeptide chain. Only two residues are found to be outside the allowed values for favourable conformations in the individual Ramachandran plots. Tyrosine at position 185 is on the cusp of an allowed region and analysis of the model structure along with the electron density map shows a high level of density around this region. Visual examination of the density confirms the correct positioning of Tyrosine185 within the density and therefore justifies the positioning of this residue as just outside the typically allowed regions of the Ramachandran plot for tyrosine residues. Threonine at position 288 was found to be outside the allowed regions, however examination of the model structure and associated electron density maps showed that the assignment of this residue was also correct and the residue position fits accurately with the density seen.

Analysis of main-chain parameters by ProCheck shows that over all the parameters analysed, this structure is found to be inside the allowable regions for the resolution of data obtained. Appendix 3 shows the full detail of the main-chain parameter data from the ProCheck analysis. Similarly, side chain parameter analysis by ProCheck



Plot statistics

Residues in most favoured regions [A,B,L]	245	91.1%
Residues in additional allowed regions [a,b,l,p]	24	8.9%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
	----	------
Number of non-glycine and non-proline residues	269	100.0%
Number of end-residues (excl. Gly and Pro)	267	
Number of glycine residues (shown as triangles)	17	
Number of proline residues	17	

Total number of residues	570	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Figure 5.3 Overall Ramachandran plot for umS167M structure

Residues all fall within the allowed regions, with 24 residues (8/9% of the total) found in the 'additional allowed regions'. No residues in the final structure were in generously allowed or disallowed regions, indicating that the overall topology of the enzyme is within biological constraints.

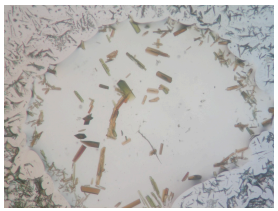
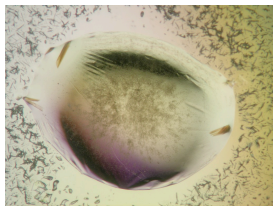
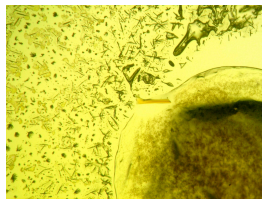
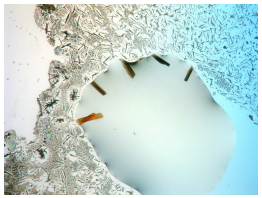
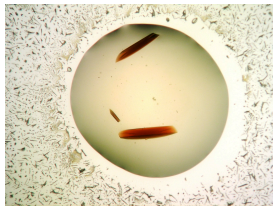
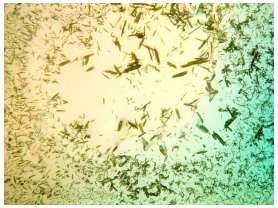
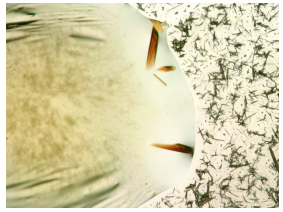
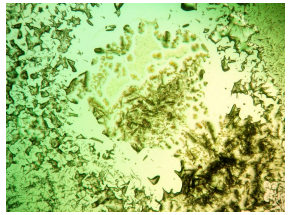
A1 	A2 	A3 	A4 	A5	A6
B1 	B2	B3 	B4	B5	B6
C1 	C2	C3	C4 	C5 	C6
D1	D2	D3	D4	D5	D6

Figure 5.1 Examples of crystals seen in umS167M crystallisation screen.
A crystal screen around the published wild type crystallisation conditions (Henriksen *et al.*, 1999) was carried out. After 48 hours of incubation at 4, crystals were seen in many of the drops. Also observed were dried crystals of ferulate (visible as microcrystals around the protien drops above) where the saturated isopropanol had evaporated, leaving ferulate crystals on the cover slip.

shows that side chain parameters are found to be inside or better than the allowable regions for the resolution of data collected. All residues expected to exhibit planarity within the side chain (arginine, asparagine, aspartate, glutamine, glutamate, histidine, phenylalanine, tryptophan and tyrosine) have been shown to have levels of planarity within acceptable root mean squared distances of planar atoms from their best-fit plane. Appendix 3 shows the full detail of the side chain parameter data from ProCheck analysis with additional annotations.

Crystal		
Cell parameters	a: 40.30Å	α: 90.00°
	b: 66.94Å	β: 90.00°
	c: 117.75Å	γ: 90.00°
Space group	$P2_12_12_1$	
Model		
Total atoms	2684 of which 260 are water molecules	
Solvent content of model	41.35%	
Structure Factors		
Resolution range	22.2Å – 1.6Å	
Unique reflections	36204	
% unique reflections above 3σ	87.8%	
Model vs. Structure Factors		
R_{factor}, for F > 2σ	21.3%	
R_{free}	24.1%	
Correlation factor	92.7%	

Table 5.3 – Summary of unmodified S167M crystal detail following data collection
Unit cell parameters closely matched the previous HRP crystals, which was consistent with the visual similarity between the crystalline forms since wild type HRP crystals had previously grown in rod-like form. Final space group of $P2_12_12_1$ was assigned after molecular replacement solutions had suggested this was the best fit for symmetry. Final R values were considered reasonable for the resolution range collected.

Additionally, the final structure was subject to analysis using **Structure Factor Check (SFCheck)**, (Vagin *et al.*, 1999). Table 5.3 shows a summation of data collected from SFCheck. Cell parameters are consistent with a crystal in space group $P2_12_12_1$.

Completeness of the data used in the crystal processing is high, this was not improved by the addition of the low-resolution data set collected on the same crystal.

Figure 5.5.a shows a Wilson Plot, calculated during the SFCheck analysis, which shows the validity of the overall data refinement process. Figure 5.5.b shows a high consistency of completeness within the data set, though it is clear that the level of completeness of the collected data reduces considerably at the higher resolution range of the data. Figure 5.5.c shows R_{factor} analysis through data resolution, with a clear correlation between R_{factor} and R_{free}. Figure 5.5.b and 5.5.c also show that the decision to limit the resolution range to 1.7Å was valid, since completeness drops and R_{factor}/R_{free} increases dramatically at the lowest resolution of the data collected.

Validation of the umS167M model structure using SFCheck and ProCheck has allowed this model structure to reach a state of completeness at which it is ready to be deposited in the Protein Data Bank. Confidence in the structure solution is high, with high level of completeness and acceptable R_{factor} and R_{free} for the resolution of data collected. **WhatCheck** was used as a final validation tool, since this programme is utilised by the PDB to ensure data quality is maintained throughout all files deposited.

The high level of confidence in the umS167M HRP model structure allows us to draw conclusions from this data in addition to the previous characterisation. Figure 5.6 shows a portion of the umS167M HRP structure with electron density added from the $2F_o$-F_c map at 1.3 rms. This confirms the expected high resolution of the data as the model can be fitted accurately to the electron density data, including the introduced methionine residue at position 167, with consistent improvement in R_{free} throughout alongside R_{factor} improvement during the refinement process.

5.3 Analysis of the crystal structure

5.3.1 Analysis of the crystal structure – S to M mutation

The umS167M HRP crystal structure revealed that the introduced methionine at position 167 in HRP was in a position similar to that anticipated in previous models,

however the distance seen between the sulphur in the side chain of the introduced methionine residue and the β carbon of the vinyl group is larger than expected and there was no evidence of electron density between the sulphur of the methionine group and the haem vinyl β carbon.

Comparison with the myeloperoxidase structure (Fenna *et al.*, 1995) shows that the bond distance between the sulphur and the haem β vinyl group in myeloperoxidase is 1.63Å. This structure, as described previously in Chapter 1.2, is the structure on which the choice of S167M was made. It has been shown that in lactoperoxidase, the formation of the covalent haem-protein ester linkages occurs autocatalytically upon exposure to hydrogen peroxide as a result of Compound I formation, this is also expected to be true in the formation of the ester and sulphonium linkages of myeloperoxidase (Zederbauer *et al.*, 2007).

The distance between the sulphur of S167M HRP and the haem β vinyl group was shown to be 3.15Å, as seen in Figure 5.6. Although the introduced methionine is located in a position where the linkage to the haem group is plausible, there would have to be significant movement of the protein structure and backbone to accommodate the autocatalytic modification. This structural evidence suggests that although structurally possible, through movement of the protein backbone and rotation of the vinyl α-β sigma bond, it is very likely that secondary conformational changes result from the associated crosslink. The overall catalytic modification is slow and presumably requires thermal motion of the protein to bring a transitory methionine sulphur radical into atom bonding distance with the carbon of the β vinyl. Presumably, this non-optimal distance makes the overall process of crosslinking very slow. Addition of reducing substrates is known to compete with the autocatalytic modification process, effectively preventing it from occurring (Cali, 2009). The low spin nature of the cross-linked enzyme supports the proposition that significant conformational change results from the crosslink.

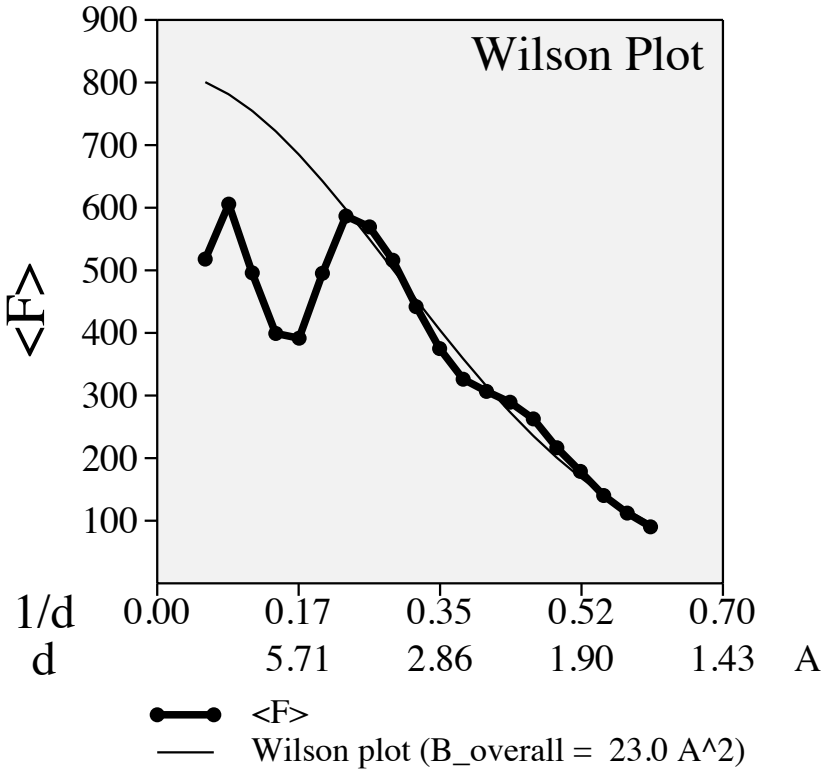


Figure 5.5.a Wilson plot for umS167M crystal structure

A plot of the structure factor against resolution bin shows the normal distribution that would be expected for an untwined data set.

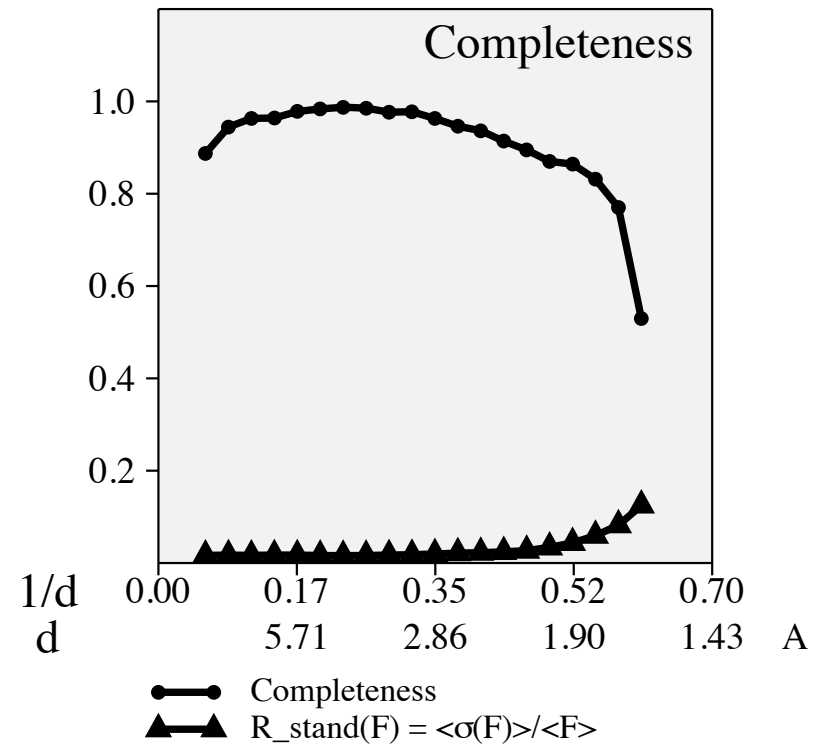


Figure 5.5.b Completeness of data collected for umS167m crystal

Data completeness exceeded 80% for >1.7Å data, showing that the correct cut off point was selected when limiting data used for the final structural model of the protein.

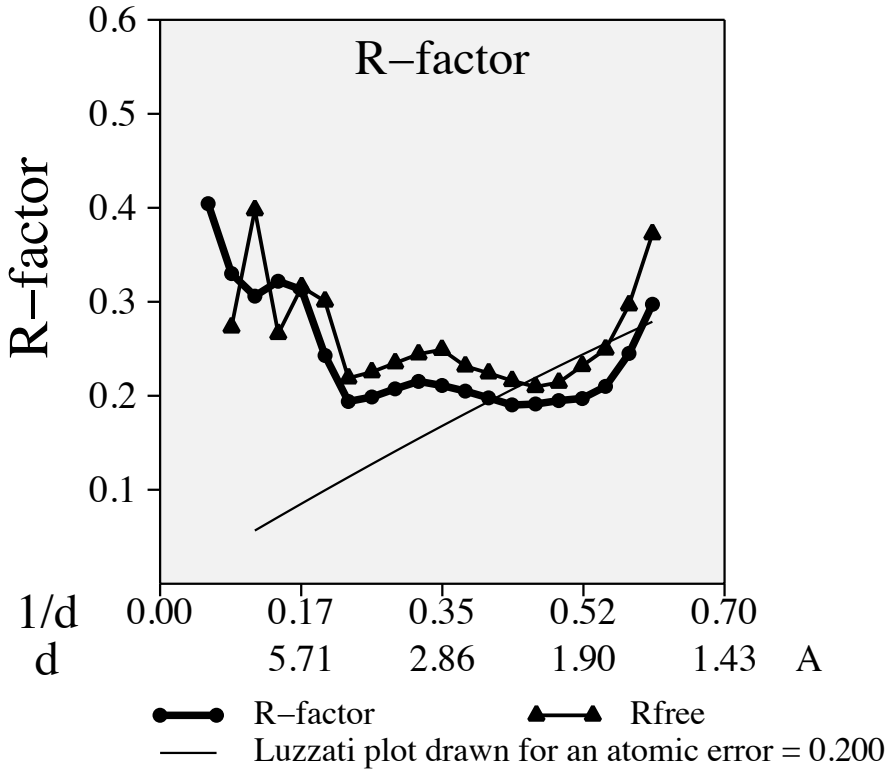


Figure 5.5.c R-factor analysis by resolution for umS167M crystal data
A plot of model R factor against resolution bin shows stable R factor up to a resolution of 1.8A with the expected increase that would follow at high resolution.

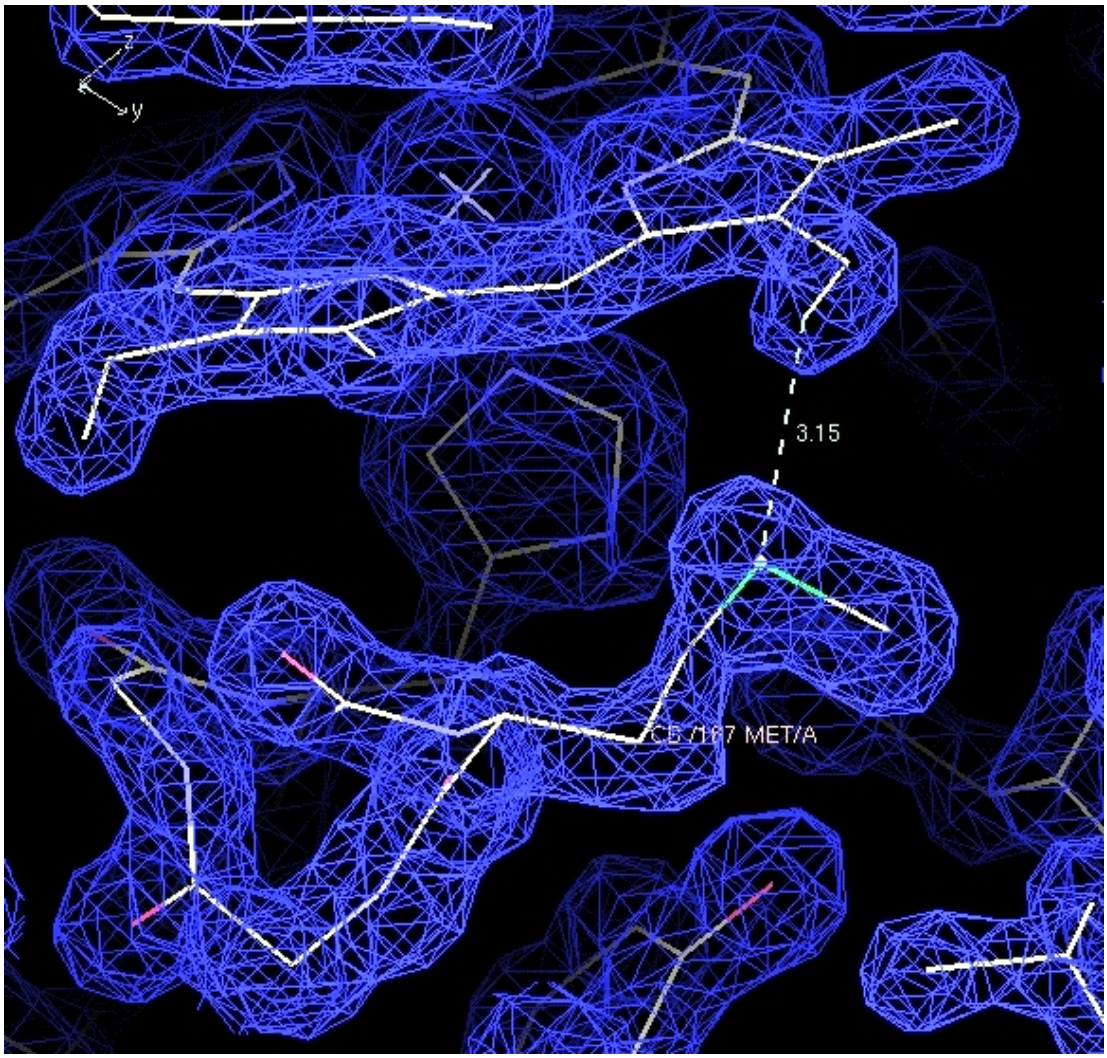


Figure 5.6 umS167M model and electron density map
Excellent fitting of the model to the electron density map allowed fitting of the introduced methionine in to the data. In the initial modelling, the area around and including M167 was excluded from the 7ATJ model to reduce any model bias when creating the electron density map. This region of amino acids were subsequently built in to the model and further refined.

Figure 5.7 shows a superposition of the new S167M structure with the HRP C* structure (PDB code 7ATJ) used in the molecular replacement solution. The electron density surrounding the introduced methionine residue is such that it is clear that the substitution is correct and that the serine residue has been replaced by a larger residue, with the side chain density directed towards the haem ring.

No density was seen in the region where the covalent bond would be expected, nor was any extra density observed around the β carbon of the β-vinyl group of the haem. This suggests that the sample of S167M utilised in setting up the crystallisation trays was truly umS167M, and the process of crystallisation occurred in such a time that no significant spontaneous cross linking was able to occur. An alternative hypothesis is that the cross linking alters the overall three dimensional fold such that cross linked protein is not incorporated in to growing crystal, supported by the proposed structural changes in the enzyme when a crosslink is formed. It would also be possible that the complex additives in the protein-precipitant solution function as reducing agents and therefore facilitate the breaking of the unstable covalent linkage.

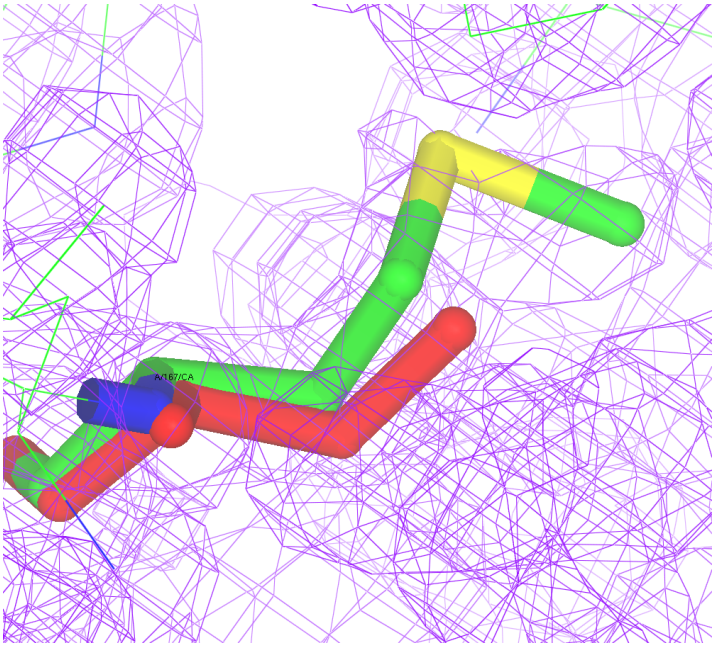
Figure 5.8 shows the haem from four angles, showing no residues in appropriate locations to form a covalent linkage, other than through the introduced methionine. This serves as further proof that the substitution was successful and that the cross linking seen in the earlier work was as a result of the introduced methionine rather than as a result of movement elsewhere in the protein chain caused by the site directed mutagenesis carried out.

5.3.2 Ferulic acid orientation

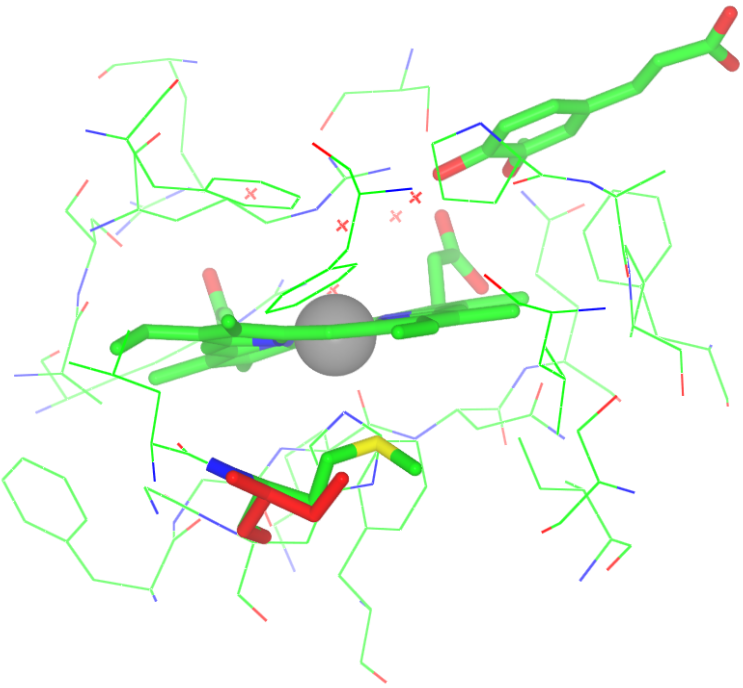
In previous HRP-C* models, the orientation of the ferulic acid monomer has often been unclear, with previous authors suggesting several different possible options for the way in which the ferulic acid monomer is arranged in the active site access channel.

Figure 5.7 Serine to Methionine single mutation
The new crystal structure of S167M is shown coloured by atom, the serine at position 167 in the wild type is highlighted here in red.

5.7.a Electron density map
The S167M electron density map clearly shows that the serine at position 167 has been replaced by a methionine, as additional density is seen in the side chain area which could not be assigned to another residue.



5.7.b Superposition of S167M structure on wild type HRP-C* model (7ATJ)
Superposition of the two crystallographic model structures and examination of the haem cavity shows how the addition of the methionine could allow the covalent haem-protein linkage to take place, although the distances involved suggest significant movement of the protein would be required.



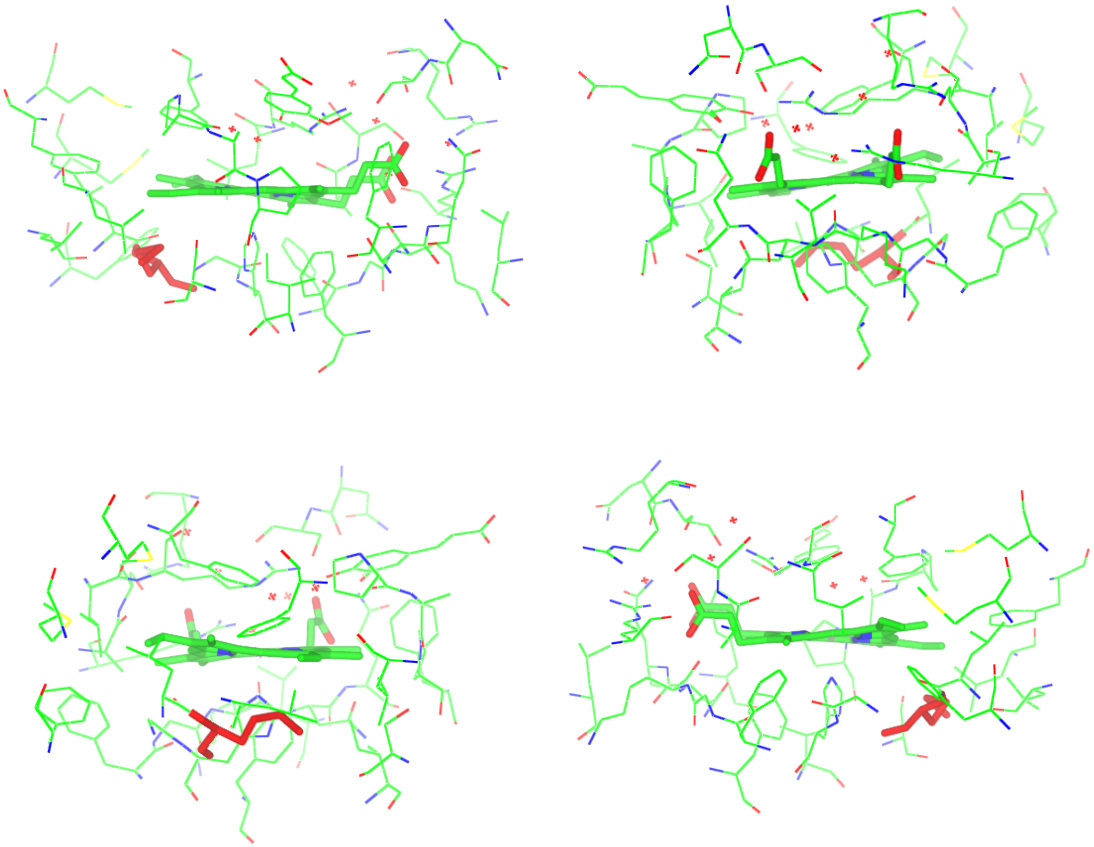


Figure 5.8 Views of the haem cavity in the unmodified S167M HRP-C* variant
Viewing the variant from four angles shows that the introduced methionine is the only residue located such that it is capable of forming the covalent bond seen when the S167M variant undergoes catalytic turnover with hydrogen peroxide. Other residues in the haem cavity are as seen in wild type HRP and, since no linkage occurs between protein and haem in the wild type, it can be safely assumed that the introduced residue is the cause of this linkage, though proteolytic analysis would show this conclusively.

In this work, prior to molecular replacement solutions being sought, the heteroatoms were removed from the model protein, along with the region around the mutation and all water molecules. This served to reduce model bias from the template molecule, with these additional atoms and small molecules being built into the new model structure at a later point.

When the ferulic acid molecule was incorporated into the S167M structure, the density suggested one single orientation, rather than the three possible options suggested in the 7ATJ structure. The amino acids at the entrance to the access channel have potential stabilising roles in the HRP-ferulic acid complex, as they would with other substrates. The role of these residues is discussed more fully in Chapter 6.

5.4 Modified S167M crystallisation

Attempts were made to obtain data from crystals containing mS167M HRP. Two methods were undertaken to obtain crystals of similar diffraction quality to the umS167M HRP crystals. The first strategy was to treat crystals of umS167M with a low level of hydrogen peroxide, resulting in possible cross-linking *in situ* within the crystal. The protein concentration in the crystals was unknown, and an excess of hydrogen peroxide was needed to ensure that as much of the protein as possible would become modified. In order to decrease oxidative stress on the crystals as much as possible, crystals were moved from their growth solution into an equivalent drop solution but with no soluble protein. Successive amounts of hydrogen peroxide were added to this drop, up to a final concentration of 50µM. This was left overnight at 4°C, as with the initial production of the modified S167M HRP in solution. Crystals were observed the next day and visually appeared to be altered in colour to the crystals observed prior to treatment with peroxide (slightly green). Unfortunately, severe cracking of the crystalline lattice was also observed, rendering a large proportion of these crystals unsuitable for data collection. Figure 5.9 shows a crystal following treatment with hydrogen peroxide and incubation overnight at 4°C. A number of crystals, which appeared to have the least visible disruption of the crystalline lattice, were mounted and initial diffraction patterns were taken, however these were seen to be very low-

resolution poor quality crystals, with high levels of apparent twinning within the crystal. Data were not collected from these crystals as in order to reliably see any sulphonium linkage formed, resolution would have to be sufficiently high to define bond electron densities accurately. The crystals which had been treated with hydrogen peroxide showed diffraction in excess of 5Å, a level that would not be expected give accurate bond densities to be able to examine the area around the introduced methionine residue. Furthermore, the quality of crystals was such that data collected would be exceedingly difficult to process, particularly with the damage sustained by the crystals during the exposure to hydrogen peroxide.

These treated crystals were dissolved in HPLC buffer and subject to the same treatment that Cali performed in previous work when showing the production of the covalent haem protein attachment. The dissolved crystal solution was very dilute due to the nature of the crystalline sample and the quantity of crystals available that had been treated, however the solution was subject to HPLC analysis. The chromatogram trace showed some population of enzyme that appeared to be covalently linked, with elution times matching that of more concentrated enzyme when treated with peroxide in solution, however since the signal from the dissolved crystals was so low, a repeat would be recommended to confirm the cross-linking.

Attempts were also made to grow crystals using solution containing S167M HRP known to be modified in solution. Attempts were made with an identical screen around the wild type conditions to that detailed previously (and in Appendix 2), with S167M that had been treated with 10 equivalents of hydrogen peroxide overnight and then exchanged into the correct crystallisation buffer. This method proved entirely unsuccessful in producing crystals of any size or quality, again implying that crystal contacts may have been lost through secondary conformational changes.

Subsequent attempts to grow crystals from modified S167M HRP samples also proved unsuccessful, despite a variety of conditions and techniques being attempted.

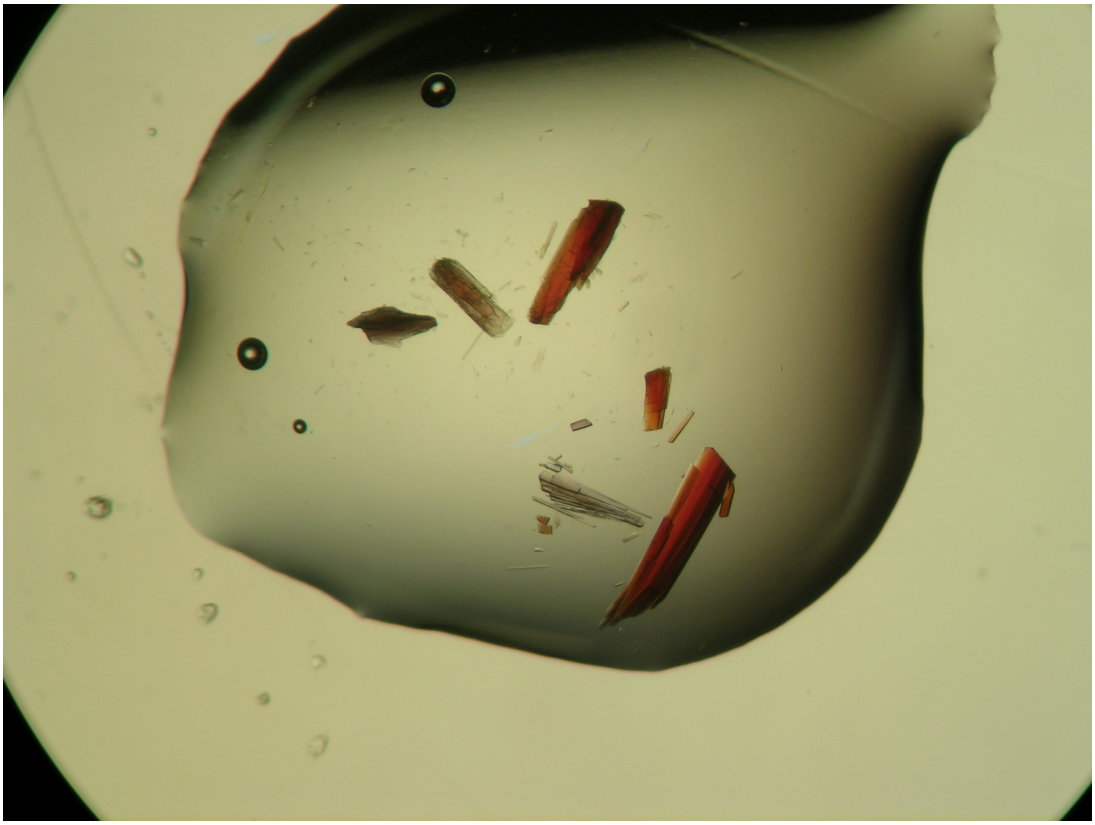


Figure 5.9 Crystals of umS167M cleaned and treated with hydrogen peroxide
Crystals from well A2 (see Figure 5.1) were removed from the drop in which they had grown and placed in a solution containing precipitant well solution but no protein or ferulic acid. These were subsequently treated with hydrogen peroxide up to a final concentration of 50µM and left over night at 4°C. This photograph shows the disruption in the crystals caused as a result of the peroxide addition.

Heterologous seeding resulted in microcrystalline formations being seen, however no large crystals of diffraction quality were produced.

5.5 Conclusions

Crystallisation of unmodified S167M HRP has led to a crystallographic model structure being solved to a resolution of 1.7Å using data collected in-house using a Rigaku rotating anode generator and the CCP4 suite of programmes for structure solution. This structure was solved using molecular replacement, with model co-ordinates taken from the structure of wild type HRP published by Henriksen *et al.* in 1999. Portions of this model, including the haem, the co-crystallisation additive ferulic acid and the region containing the single amino acid substitution of serine to methionine, were removed prior to molecular replacement. These were able to be accurately fitted to the electron density data, whilst their removal during the molecular replacement process served to reduce model bias.

The crystal structure revealed the distance between the sulphur constituent of the introduced methionine residue and the β carbon of the 2-vinyl group of the haem to be 3.15Å, a figure that is approaching double the bond length seen in the sulphonium linkage in myeloperoxidase. This suggests that during both the autocatalytic and spontaneous formation of this bond, movement must occur within the protein molecule to allow this bond to form. To reduce this distance, rotation of the vinyl C_α-C_β bond and some backbone perturbation must take place.

In previous work it was seen that the S167M HRP covalent sulphonium linkage formed spontaneously in a small proportion (14%) of the protein when aged at 4°C for several days. Although no specific time scale was proposed for this aging process, it did not disrupt the formation of unmodified S167M crystals. It may be that only umS167M protein molecules aggregated in the crystalline form, with any modified protein remaining in solution. Substrate (ferulic acid) was also present in the crystallisation medium that may have quenched any autocatalytic crosslinking chemistry. No additional density is seen on the haem or the methionine residue, suggesting that, if

this process is occurring, the protein appears to be unmodified rather than either of the two forms of the broken linkage suggested in the previous work (Cali, 2009).

Also of note, was that the ferulic acid co-crystallisation additive was clearly defined in the entrance to the haem access channel. In the previous work, three potential conformations of this molecule were proposed in the native HRP-C, as the substrate density observed was disordered and did not allow for unambiguous substrate modelling (Henriksen *et al.*, 1999). The crystal structure of S167M HRP obtained density data around the ferulic acid molecule that showed clearly the ferulic acid with the phenolic group towards the haem centre of the enzyme. The conformation of the ferulic acid in the active site matches the density described in the 1.45Å structure of the HRP-C cyanide complex described by Henriksen and co-workers (1999).

Despite varied strategies, crystallisation of the modified form of the S167M HRP variant was unsuccessful. This suggests that the linkage has subtly altered the overall tertiary structure of the molecule, with the result that key crystallographic contacts are no longer able to form and therefore stable crystals do not form using the previously determined conditions. A fuller screen of general crystallisation conditions was carried out, however this also proved unsuccessful.

5.6 Discussion

Haem protein linkages in mammalian peroxidases confer stability and impose conformational distortion of the normally planar haem centre, with this link formed as part of an autocatalytic process (Kooter *et al.,* 1999), involving initial exposure of the enzyme to hydrogen peroxide, and is thought to be required for the chlorination activity of myeloperoxidase (Devarajan *et al.,* 2008).

Previous work has modelled linkages from mammalian peroxidases in plant peroxidases, where covalent links between the haem and protein are not normally seen. This work attempted to influence the stability of plant peroxidases and probe the importance of the linkages in the function of mammalian peroxidases.

The work conducted by Cali described in his doctoral thesis examined the sulphonium linkage seen uniquely in myeloperoxidase. Mimicking this link in HRP gave a protein that, under conditions where a 10-fold excess of hydrogen peroxide was present, formed a covalent link. This link was also noted to form, at least partially, upon aging of the protein.

Cali found that this link was labile under some conditions, with the autocatalytically formed linkage partially breaking upon aging of the modified protein. This suggests that a slow nucleophilic hydrolysis may be taking place.

The crystal structure of the unmodified form of S167M has been described in the above chapter. This was attempted so as to further understand the nature of the site of bond formation and to determine the nature of the covalent bond distances and possible geometries.

As described above, the crystal structure of umS167M was solved to a resolution of 1.7Å, with the distance between the sulphur group of the methionine residue and the haem vinyl group measured to be 3.15Å. This distance is relatively substantial, particularly when compared with the 1.67Å of the sulphur to vinyl carbon in the covalently linked myeloperoxidase. This distance suggests a possible reason why the protein is not linked to the haem during the folding process, with the active site folded prior to the haem being added in the 'late haem addition' modification to the HRP refolding method described by Cali to produce a higher yield of correctly folded enzyme.

This distance also shows that some movement in the protein would plausibly cause the side chain of residue 167 to come within bond distance of the haem vinyl group, allowing mS167M to be formed autocatalytically in the presence of hydrogen peroxide, and spontaneously *albeit* slowly upon aging under air.

Whilst the crystal structure of the umS167M HRP described above does not show any evidence of a covalent linkage between protein and haem, it is possible to speculate

about the nature of this bond. The structure indicates that the sulphur atom in the methionine side chain is in a position such that bond formation is possible but with some disruption of the overall global structure. Bond formation between the protein and haem would cause some rearrangement of local structure, with, in its current location, torsion angles becoming stretched out of plausibility. It is expected that with formation of the covalent linkage, local structure would be disrupted, with the protein put under strain. This strain supports the earlier observation that both the autocatalytically and spontaneously formed linkage decay upon aging, with the thermodynamic stability of the protein under stress meaning that a broken linkage is favoured if local chemistry allow it.

The formation of the covalent linkage may also distort the planarity of the haem group, a feature seen in mammalian peroxidases and enhanced in myeloperoxidase (Fenna *et al.*, 1995).

Chapter 6

HRP variants with a more open haem edge

6.1 Introduction

As has been discussed in the previous chapters, the residues in the core peroxidase fold, particularly those surrounding the haem moiety, are well conserved between the different classes of the plant peroxidase superfamily. The substrate range of these enzymes, however, is highly variable, with particular members of the different classes having very specific molecules they are able to oxidise.

The specificity of certain enzyme groups to particular substrate molecules is thought to be influenced by the residues surrounding the entrance to the active site, with a high degree of variation between the nature of these residues between enzymes of different classes.

Class II plant peroxidase enzymes *Coprinus cinereus* peroxidase (CiP) and Lignin peroxidase (LiP) have a more open haem access channel compared to HRP and other Class III plant peroxidases. Residues, particularly at position 140 and 176 in HRP (at opposite sides of the haem access channel) are considered important in facilitating the catalysis of substrates.

Previously, Tanaka *et al.*, (1999) attempted to mimic the haem edge environment of *Arthromyces ramosus* peroxidase (ARP), a peroxidase essentially identical to CiP with variation only seen in the surface glycosylation (Smith and Veitch, 1998), by introducing charged residues that were seen in ARP but not in HRP. Lysine found at position 49 and glutamate at position 190 in ARP were suggested to modulate the activity towards different substrates, particularly that of luminol, which shows a 500-fold increase in activity when compared with HRP (discussed later in this chapter). These charged residues, however, when introduced in to equivalent sites in HRP by site directed mutagenesis (S35K, Q176E and the double mutant S35K/Q176E) did not

significantly improve the activity of HRP towards the substrate luminol. S35K gave a 1.5-fold improvement in catalytic efficiency (V_{max}/K_m(luminol)) compared to wild type HRP whereas Q176E had a slightly reduced catalytic efficiency. The double mutant S35K/Q176E had a 2-fold increase in efficiency towards luminol, indicating that the positive charge at position 35 allowed a low level of stabilisation with the polar side chains of the luminol molecule associating with the charged residue on the protein (Tanaka *et al.*, 1999). The substitutions for charged residues, however, clearly do not give the CiP/ARP-like activity towards luminol that was expected, showing that the introduction of charges to stabilise the luminol molecule in the haem access channel alone are not sufficient to create high turnover of luminol in HRP.

This led to the hypothesis that it was the size of residues surrounding the entrance of the haem access channel, rather than a specific charge, that was a major influence in the substrate accessibility and hence catalytic activity between plant peroxidases of different classes.

Previous work by Dr Gareth Williams during his doctoral studies aimed to alter the conformation of the haem access channel entrance by reducing the size of side chains in the vicinity of the channel entrance. This would give an indication of the role of residue size and conformation, as opposed to charge, in controlling substrate access to catalytic centres. Single mutant variants, in which residues around the haem access channel were replaced with alternative residues aiming to affect enzyme specificity, were created and the substrate accessibility analysed.

Whilst structural studies indicated that both the A140G and F179A single mutants demonstrated very little in the way of structural change, other than the apparent reduction in side chain diameter, kinetic studies showed that the subtle amino acid substitutions had a more profound effect on the activity of the enzyme. Compound I formation rates in the A140G variant was unchanged in comparison to wild type HRP-C*, however reduction of Compound II was increased with certain substrates (Williams, 2005). Subsequent binding studies also showed that enzyme-substrate affinity was altered, indicating that the two residues play a central role in modulation

of substrates access to the enzyme and hence enzymatic activity towards different substrates (Williams, 2005 and Cali, 2009).

The A140G/F179A HRP variant was constructed and characterised by Dr. Khasim Cali during his doctoral work, based on the previous work on single mutant variants. Also created was the double mutant F68A/F179A. The combination of the single HRP variants was anticipated to increase the activity towards substrates that are traditionally catalysed by Class II peroxidases.

As with the S167M HRP variant discussed in Chapter 5, the species characterisation and analysis carried out by Dr. Cali and Dr. Williams will be remarked upon during the discussion of the crystal structure.

As detailed in Chapter 1, residues at the entrance to the active site access channel of HRP do not contribute directly to the type of catalytic activity performed by the enzyme. These residues do not participate directly in the hydrogen bonding network that controls the formation of Compounds I and II at the catalytic centre, however the subtle changes to the haem environment have been shown to influence the apparent second order rate constants that affect Compound I formation (Williams, 2005).

Previous single mutant variants showed considerable differences in the ability of the enzyme to form Compound I. Whilst the A140G variant has the same apparent second order rate constant as the wild type HRP, more disruptive variants, such as A140Q, (where the small, non-polar alanine side chain of the wild type is replaced by the large, polar, though uncharged, glutamine) were shown to decrease the enzyme's ability to form Compound I (Williams, 2005).

A range of variants were created during the previous work, as detailed in Table 6.1, including both single and double site directed mutants. These reduced the bulky side chains of residues around the entrance to the haem access channel in order to increase substrate access to the haem cavity.

Variant	K_{cat}/K_m (Luminol)
Wild type HRP	3.3×10^4
Q176D HRP	2.2×10^4
A140G HRP	8.6×10^4
F179A HRP	3.5×10^4
F68A HRP	6.0×10^4
A140G/F179A	8.4×10^5
F68A/F179A	7.9×10^5

Table 6.1 HRP site directed mutants with variations to the haem access channel entrance.
Variants created by Williams and Cali altered residues at the entrance to the haem access channel, potentially giving substrates greater access to the oxyferryl haem and hence enhancing turnover rates with respect to certain substrates. Single site directed variants had similar orders of rates to wild type HRP, though enhanced rates were seen in A140G and F68A. Combinations of these variants in double mutants resulted in significant increase in turnover with respect to luminol.

The A140G/F179A variant was designed to have increased the substrate access to the haem by reducing the side chain dimensions of two residues that are found at the entrance to the haem access channel of HRP. These residues are shown in Figure 6.1, along with other residues highlighted in Table 6.1 that alter the efficacy towards oxidation of luminol. The purpose of increasing the accessibility of the haem access channel was to increase the affinity of the HRP to some key substrate molecules, particularly the substrate luminol, as discussed in Chapter 1. Increasing the affinity of HRP to this substrate could reduce the need for co-factor molecules in the reaction of HRP with luminol, something that has been shown to affect the reliability of enzymatic assays that use the chemiluminescence produced in the oxidation of luminol as an indicator.

Figure 6.1 shows the entrance to the haem access channel of wild type HRP, with key residues that modulate the access channel entrance shown as spheres. The steric hindrance of the residues at the entrance to the haem access channel is clear, in particular F68 and F179. The suitability for site directed mutagenesis to reduce these residues has been demonstrated in previous variants with single amino acid mutations,

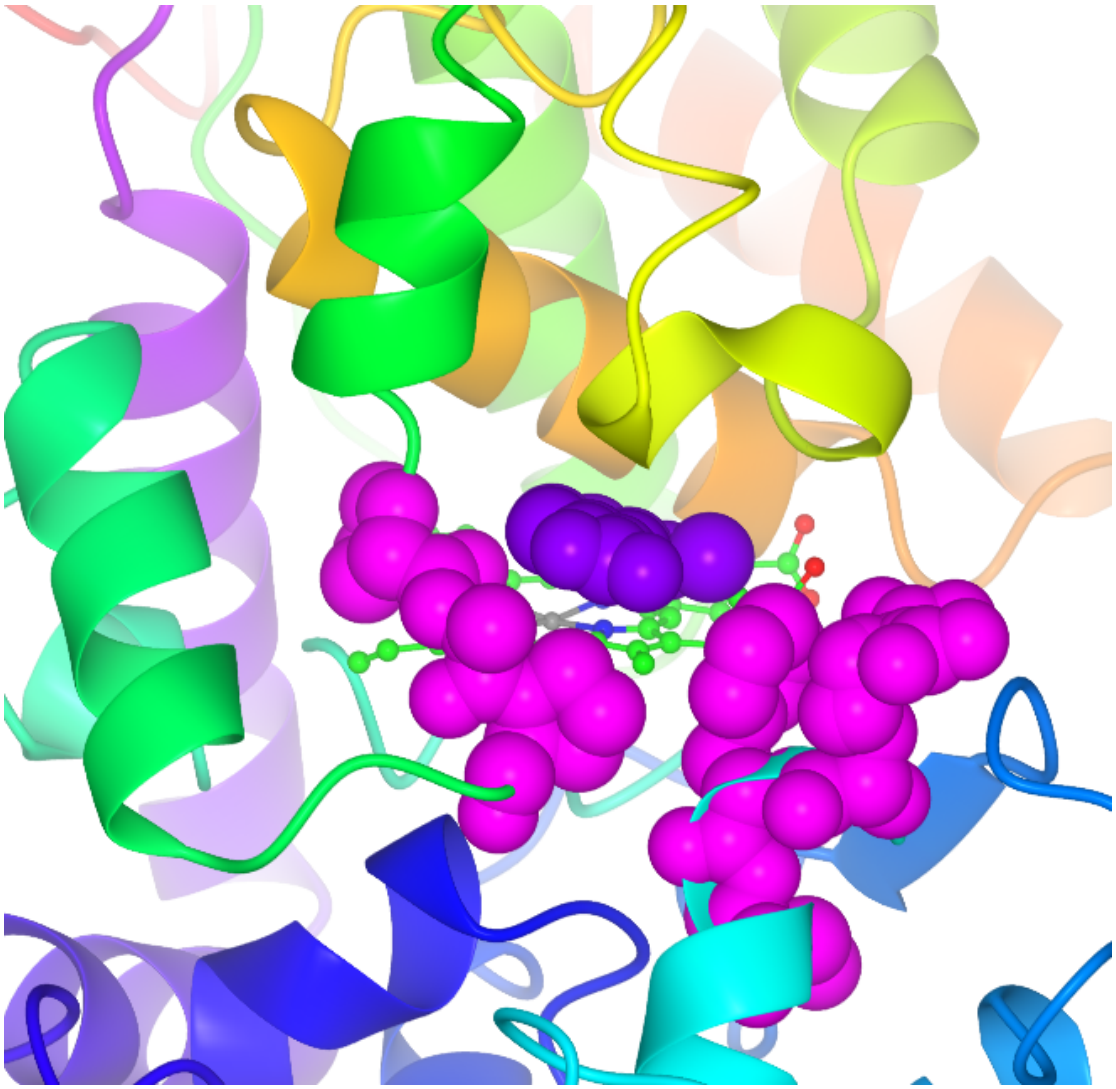


Figure 6.1 Wild type HRP haem access channel
HRP structure 7ATJ (Henriksen *et al*, 1999) shown with haem access channel entrance residues shown as spheres. Also shown is the co-crystallisation substrate ferulic acid (purple) in the entrance channel. Sterically hindering residues, particularly F179, restrict substrate access to the haem and hence catalytic activity.

however the double mutant variants show increased affinity, presumably as they are able to combine effects of reducing residue bulk.

Previous work by Cali showed that replacement of selected haem access channel residues with equivalently charged residues with smaller side chains can considerably increase the rates of reaction with wild type HRP substrates. These variants did not substantially alter the UV/visible spectrum of the protein, indicating that the substitutions in amino acid did not have a significant effect on protein structure nor haem environment. With variant A140G/F179A, the Soret peak appears at 403nm, with the Soret shoulder, seen at 70% height of the Soret peak in both the wild type and A140G/F179A variant HRP, at approximately 380nm. This shoulder is lost in the EAA "peroxygenase" variant discussed previously, indicative of changes in the haem environment not seen in the A140G/F179A variant. The A140G/F179A HRP variant also shows the characteristic charge transfer bands at 500nm and 640nm, which, in combination with the Soret peak at 403nm, confirms that the haem is in a 5-coordinated high spin form. This is as would be expected for this variant as the entrance to the haem cavity has been altered but the haem cavity remains unaltered and hence the haem unchanged from the wild type.

As shown in Table 6.1 there is a significant increase in the oxidation of luminol, with the A140G/F179A displaying an increase of 84 fold above wild type HRP. This achievement is a remarkable increase for replacement of two residues, with the F68A/F179A double mutant showing a similar increase of 79 fold higher than the wild type. These values are only around 5 fold lower than the CiP/ARP rates of luminol oxidation, showing that the replacement of combinations of the more bulky residues at the entrance to the haem access channel has a significant effect on the substrate accessibility, allowing HRP to better mimic CiP (Cali, 2009).

An understanding of the three dimensional structure of this variant would support the conclusions drawn in the previous work and may enable modelling of the luminol substrate in the active site of the new crystal structure. It would also enable more precise targeting of residues for further rounds of site directed mutagenesis in

combination with the A140G/F179A variant by choosing residues which would support the changes previously made, without too much disruption of the haem access channel overall structure.

Whilst the next obvious variant to attempt is the F68A/A140G/F179A, knowledge of the structure of the most effective of the double mutants could indicate the reasons for such a dramatic increase in oxidative capability over the individual single mutants.

6.2 Results and discussion

6.2.1 Protein crystallisation

A pure sample of A140G/F179A HRP C* was obtained from stocks of protein stored at -80°C. The purity of the sample was checked by examination of the UV/visible spectrum of the sample, ensuring that the freezing process had not led to denaturation of the protein which would disrupt any crystallisation attempt. The sample was confirmed to be of an acceptable purity and so concentrated to 8mg/ml using an Amicon concentrator pre-washed with ddH_2O.

This sample was set up in a hanging drop crystallisation screen around the conditions that produced wild type HRP crystals, according to Henriksen *et al.* (1999). Details of the individual well conditions are recorded in Appendix 2, these are the same conditions as used for the S167M variant. The screens were carried out in 24 well Linbro plates (Molecular Dimensions Ltd.) with cover slips sealed using high vacuum grease (Dow Corning). The drops were set, as with the S167M variant, at room temperature using 1µl of an isopropanol solution saturated with ferulic acid, which was allowed to evaporate on each cover slip before 1µl protein solution and 1µl reservoir solution were added to the slip, forming the "hanging drop". Evaporation of the ferulic acid saturated isopropanol allowed protein and reservoir solution to be maintained in a globular drop. Addition of the isopropanol after protein and precipitant solutions was placed on the cover slip resulted in dispersal of the protein drop across the cover slip, thought not to be conducive to protein crystal formation. The tray was incubated at 4˚C in a Rumed 3000 incubator, and drops were checked

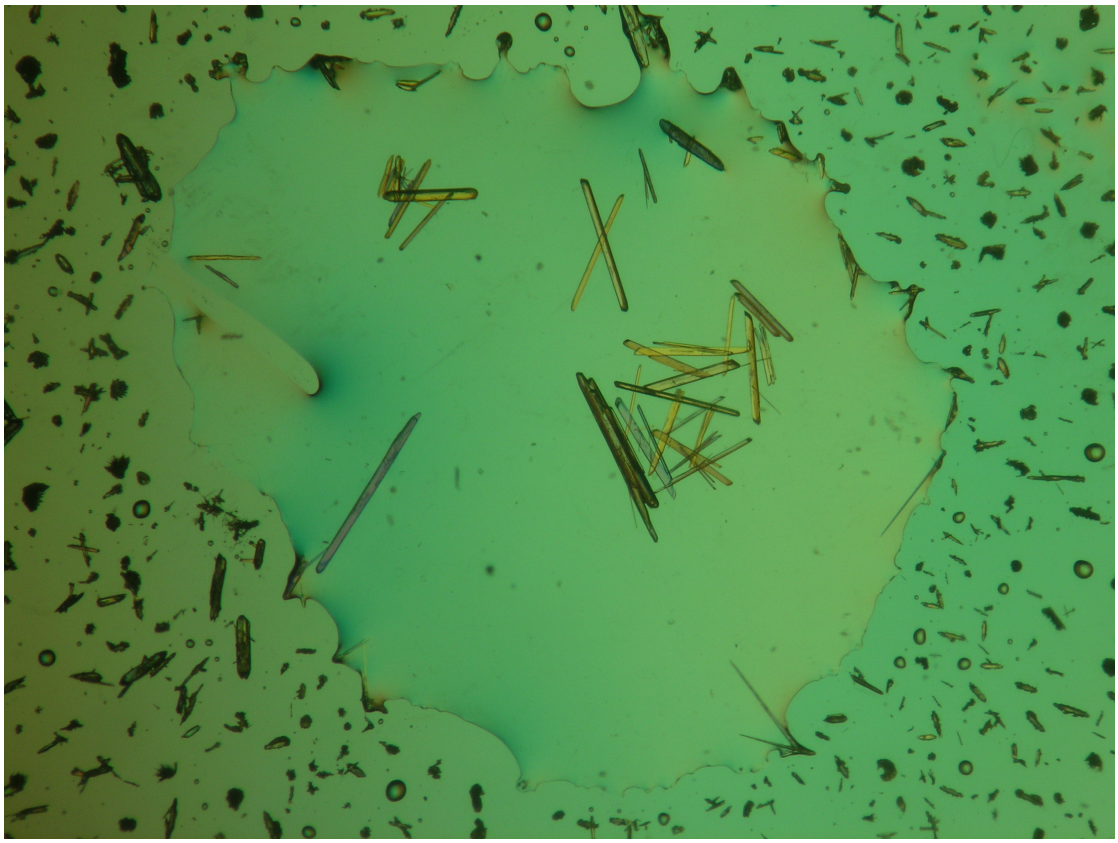


Figure 6.2 A140G/F179A crystals
Crystals of A140G/F179A HRP C* grew in a screen around the previously published conditions (identical to that used in crystallisation of the S167M HRP variant described in Chapter 5). This figure shows crystals from drop C4, and is representative of the crystals seen for this screen. The thickest, cleanest crystals were selected for diffraction.

regularly for crystal formation under a Nikon light stereoscope at room temperature. Crystals of the A140G/F179A HRP variant were seen in the screen within seven days. A variety of different sized crystals were seen, with some taking the form of apparently more two-dimensional "plates" rather than three-dimensional crystals needed for high-resolution data collection. A number of crystals were seen that were thought of as suitable for data collection, allowing a degree of choice and more ability to apply specific selectivity when choosing crystals to collect data from. Previous experience with thin crystals of the appearance of these plates suggested that little poor diffraction would be obtained from these crystal plates. Had suitable crystals not been seen, the plate-like crystals would have been used to seed in to drops where no crystal formation was observed. Figure 6.2 shows a photograph of an example of the crystals of A140G/F179A HRP grown in the screen.

6.2.2 X-ray diffraction

Crystals selected for data collection were first transferred to a cryoprotectant solution which had previously been tested to ensure that vitreous ice, rather than crystalline ice, formed when the sample loop was placed in the nitrogen stream at 100K. Cryoprotectant used was identical to that used for the X-ray data collection of the umS167M HRP variant and similar to that used by Henriksen *et al.* in the collection of data for wild type HRP crystals, with a solution of 40% polyethylene glycol 8000, 0.1M Ca $(CH_3COO)_2$, 0.1M cacodylate buffer, pH 6.5.

Crystals were selected from drop C4, as these appeared to be the best crystals based on visual inspection, discounting crystals which appeared joined to others or that had any apparent flaws in the surface. Crystals were placed in the cryoprotectant solution and again examined under the light microscope for any flaws in the crystal appearance. No major flaws were visible the crystal chosen, so it was mounted in the loop and placed in the X-ray path for initial images to be taken. To minimise stress on the crystal, the liquid nitrogen stream was covered while the crystal was mounted and then released. This aimed to prevent freezing and thawing of the cryoprotectant solution and hence decrease the possibility of ice crystals forming.

Data collection was carried out in-house, with images collected on an R-axis IV++ imaging plate using X-rays generated by a Rigaku RU-H3RHB copper rotating anode generator, with X-rays of wavelength 1.542Å, focused using Osmine Confocal Max-flux mirrors.

Initial images were taken with 1° oscillation at 0° and 90°, with 200mm crystal to image plate distance, with exposures of 5 minutes. These images allowed assessment of the crystal and decisions were taken whether to continue with the data collection. The first crystal chosen from drop A3 diffracted only weakly. This led to selection of another crystal as, although the diffraction of 3.0Å would be acceptable in some circumstances, the fact that other crystals were available meant that alternative crystals were chosen in the hope that higher resolution diffraction could be obtained.

A second crystal, from drop B2, was chosen and initial images recorded, again with 1° oscillation and at 0° and 90°, 200mm crystal to film distance, with 5 minutes exposure per image. This crystal gave more promising initial images so the image collection plate was moved closer to the crystal, to 160mm distance, and the exposure time increased to 15 minutes per image. Images were taken at the same oscillation angles. These showed diffraction to a maximum resolution of 1.9Å.

Initial images were exported to **MOSFLM** for analysis of unit cell calculations and putative space group predictions to be made. Unit cell dimensions were calculated by MOSFLM to be a=40.1Å, b=66.2Å, c=117.1Å, with angles of α=90.0°, β=90.0°, γ=90.0°. This gives a space group allocation of P2 symmetry, with the predicted space group to be P222. This assignment correlated with previous data collected from both the S167M HRP crystals and previous collection of data from HRP crystals with the same appearance (long rod-shaped crystals).

The P222 assignment allowed for rational data collection, ensuring that the optimal amount of data was collected giving the highest level of completeness with the minimum quantity data collected. Using the STRATEGY command, MOSFLM predicted a 90° segment of data that would give maximum completeness with minimal angle,

reducing damage to the crystal due to the power of the X-ray beam.

6.2.3 Data processing

90° of data was collected, with exposure time of 15 minutes and final image plate to crystal distance of 190mm.

Collected data was exported to MOSFLM where unit cell dimensions were confirmed as being a=40.103Å, b=66.229Å, c=117.114Å, with angles of α=90.0°, β=90.0°, γ=90.0°. Images were analysed for spot location using the FIND SPOTS command over multiple images. Spots were excluded which were deemed to be too large in x or y, or too faint. This ensured that "real" spots were chosen for data analysis and prediction, excluding spots with any overlap or ones that may not have been easily distinguishable from the background of the image. There were no ice rings present in the data, however had there been ice rings present, spots near to the ice rings could have been excluded from the prediction to eliminate the possibility of distortion of data as a result of the ice rings.

Autoindexing of the data was carried out with two images with maximum separation. Unlike the umS167M data, the A140G/F179A diffraction images did not show any anomalies when autoindexing the data, with spot predictions accurate when refined through adjustment of individual settings, therefore the images were correctly autoindexed as a complete dataset rather than in separate segments. This also eliminated the need for the data to be sorted in to one file prior to scaling and merging the data using **Scala** and **Truncate**.

Space group was assigned during the autoindexing process as $P2_12_12_1$, as this had the highest symmetry with, relatively, the smallest associated penalty factor (Penalty factor = 2). This confirms the previous assignment made prior to X-ray data collection and that the correct amount of data was collected to produce the highest level of completeness from the dataset. Spot predictions were then visualised and examined to ensure that each spot was predicted correctly. Subsequently, the REFINE CELL command was used in MOSFLM to improve the accuracy of the spot assignment and to

increase the number of spots recognised as valid data. Some spots were now assigned as being seen in multiple images. Table 6.2 shows a summary of the refinement data for the A140G/F179A crystal.

A140G/F179A Refinement Data			
	Overall	Inner shell	Outer shell
Low resolution limit	32.92	32.92	1.83
High resolution limit	1.74	5.49	1.74
R_{merge}	0.199	0.064	1.993
R_{meas} (within I+/I-)	0.236	0.077	2.706
R_{meas} (all I+ & I-)	0.236	0.077	2.706
R_{pim} (within I+/I-)	0.125	0.041	1.818
R_{pim} (all I+ & I-)	0.125	0.041	1.818
Fractional partial bias	0.019	-0.085	-0.166
Total number of observations	74518	3714	1198
Total number unique	24943	1175	916
Mean ((I)/sd(I))	6.5	27.7	0.3
Completeness	75.7	99.3	19.4
Multiplicity	3.0	3.2	1.3

Table 6.2 Refinement data for the A140G/F179A crystal
Data was collected to a maximum resolution of 1.74Å, however at this resolution the quality of the data deteriorated, leading to a cut off point being taken at 2.0Å.

Data were scaled using **Scala** and merged using **Truncate**, and unit cell capacity was calculated, with two molecules predicted per unit cell. This was expected and in line with the previous HRP data collected for crystals of this form and therefore accepted as correct.

Molecular replacement was carried out using **MolRep**, **Phaser** and **Amore**. Both MolRep and Phaser produced acceptable molecular replacement solutions, however Amore repeatedly failed to give a solution with failure error "AMORE: Error in label assignments". This was not pursued further since successful molecular replacement solutions had been produced from alternative programmes, however if alternative options had not been available data would have been analysed from previous steps

and repeated if necessary.

The molecular replacement solution from Phaser was taken forward and automatic refinement was carried out prior to any manual refinement. **Refmac5** was used for automatic refinement and initially restrained refinement was carried out to gain starting R values. Ten cycles of restrained refinement were carried out per iteration using the molecular replacement solution produced by Phaser. Initial R_{factor} was 34.4% with corresponding R_{free} of 36.6%. The R_{factor} had reduced to 28.7% after one iteration of restrained refinement, with corresponding reduction in R_{free} suggesting that this was a valid model to take through to the maunal refinement stage.

Crystal		
Cell parameters	a: 40.10Å b: 66.23Å c: 117.11Å	α: 90.00° β: 90.00° γ: 90.00°
Space group	$P2_12_12_1$	
Model		
Total atoms	2400 excluding hydrogens	
Solvent content of model	47.8%	
Structure Factors		
Resolution range	2.0Å – 31.86Å	
Unique reflections	18998	
Completeness for range	91.77%	
Model vs. Structure Factors		
R_{factor}, for F > 2σ	0.249	
R_{free}	0.30	

Table 6.3 Summary of crystallographic data for structure of HRP A140G/F179A

Manual refinement was carried out using **Coot** with the model and maps produced from the restrained refinement of Refmac5. Individual residues were examined by eye and were manually fitted to the electron density maps. Restrained refinement was carried out using Refmac5 after each round of manual refinement. Ten cycles of refinement were carried out and the resulting graphs examined to ensure that R_{factor} and R_{free} were decreasing as a result of the manual refinement. At the stage where no further significant decrease in R_{factor} was seen, manual refinement was stopped and the model taken forward to internal structural geometry checks.

Structural geometry was analysed using **ProCheck** as part of the CCP4 package (Leslie, 1992). Internal angles and bond lengths were examined and those which fell outside the acceptable range for bonds within protein molecules were manually adjusted. This process was followed by further checks using Refmac5 to ensure R_{factor} was maintained or decreased. Adjustments were made until a satisfactory ProCheck result was achieved.

Overall structure was evaluated using **SFCheck**, which produced visual representations of the quality of the structure. Examination of the results of the structure factor check revealed that several residues fell outside the "allowed" regions for protein structural geometry. These were individually examined and their bond angles adjusted accordingly.

6.3 Analysis of the crystal structure

Analysis of the crystal structure shows that the steric hindrance of large residues around the entrance to the haem access channel has been dramatically reduced by the substitution of alanine at position 140 by glycine and phenylalanine at position 179 by alanine. The reduction in the bulk of these residues allows larger substrates better access to the haem, facilitating the more rapid oxidation of these substrates without the need for mediators in the reaction. This result supports the previous work by Williams and Cali (unpublished data) that show increased reaction rates in this variant.

Figure 6.3 shows the haem access channel entrance, with the residues in pink as seen in the wild type crystal and those in green as seen in the A140G/F179A variant. Figure 6.4 shows the same view of wild type and variant HRP access channel, including other local residues which play a part in the modulation of substrate access to the haem. Residues 139-141 and 178-180 are shown in colour, as is the haem, indicating the haem access channel entrance. From these comparative images it is clear how much of a difference the reduction in key residues can make to the accessibility of the haem of HRP, with obvious implications for the binding of substrates which are larger than the small phenolic compounds typically oxidised by HRP.

Figure 6.3 HRP structure
Wild type and A140G/F179A structural overlay. Wild type residues A140 and F179 are shown in pink, with G140 and A179 of the variant shown in green.

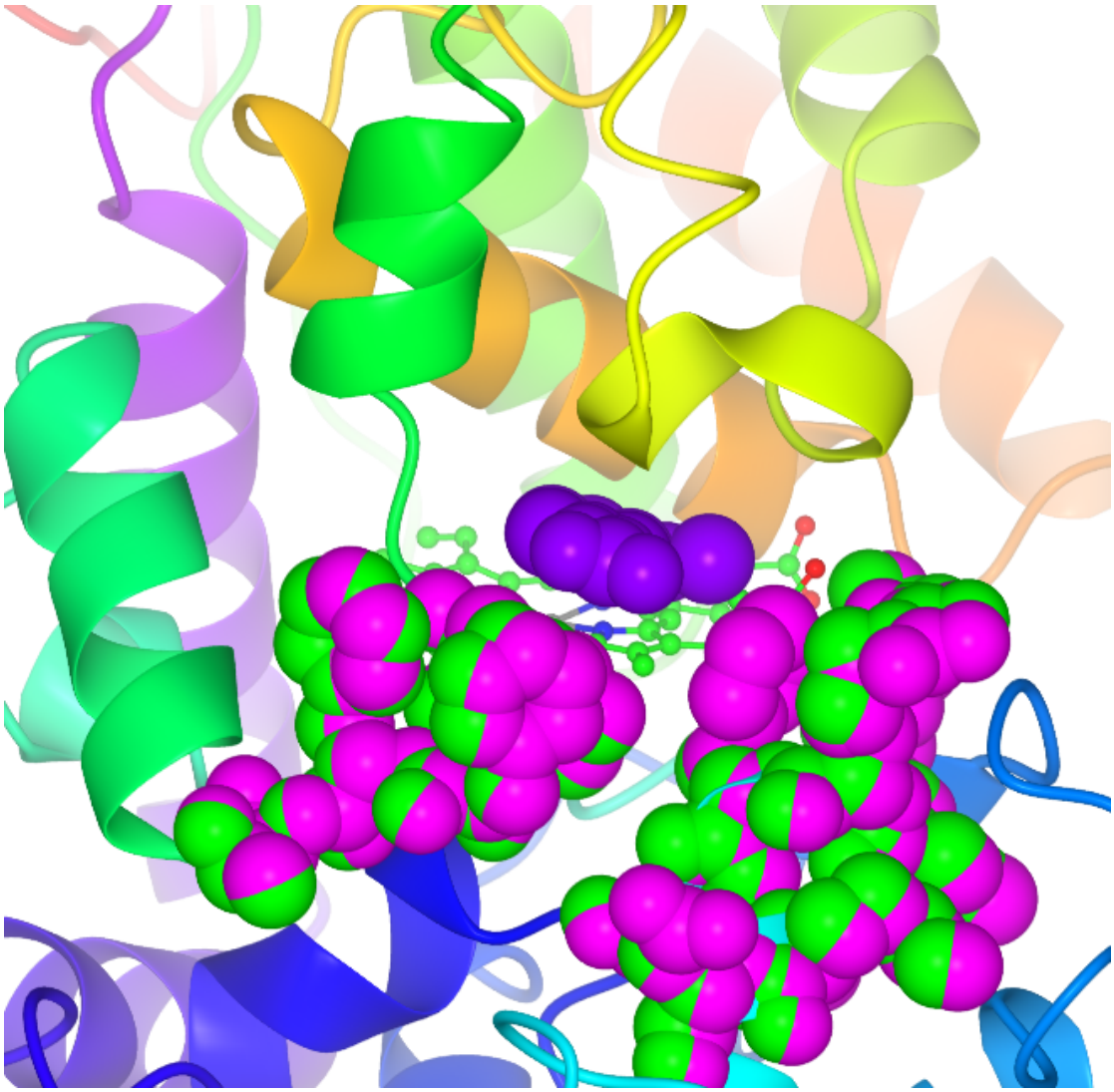


Figure 6.4 Haem access channel
Wild type residues shown in pink, A140G/F179A in green. Residues A140 and F179, in particular the bulky F179 are seen in the catalytic site access channel. The proximity of these residues to the ferulic acid molecule in the site show how they may hinder access to the haem.

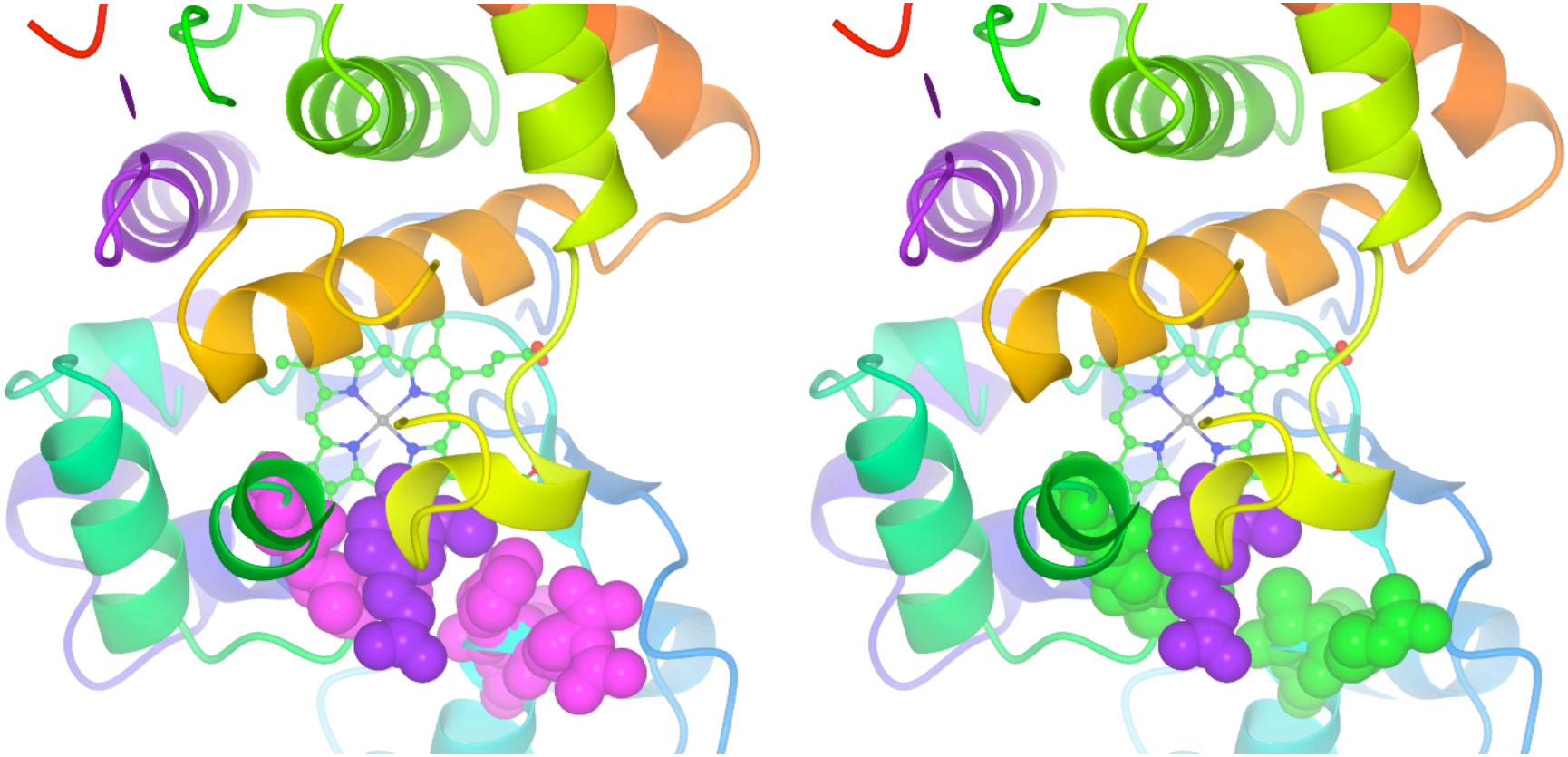


Figure 6.5 HRP access channel residues, view from distal haem side
Wild type HRP (left) and A140G/F179A (right) as viewed from the distal domain. Sterically hindering residues A140 and F179 are greatly reduced with the substitution of glycine and alanine respectively. This allows better access to the haem for bulky substrate molecules.

Figure 6.5 shows the view from the above the distal side of the haem, with Figure 6.5.a showing the wild type HRP and 6.5.b showing the A140G/F179A variant. Again this figure illustrates the way in which reduction in side chain bulk of large residues in the entrance to the haem access channel can open up this channel for substrate binding.

Comparison of the access to the catalytic cavity between wild type HRP and the A140G/F179A variant shows clearly how the reduction in dimensions of the side chains has a profound impact on the accessibility of the active site.

While the previous work by Tanaka *et al.* showed that altering the charge of residues around the entrance site to the haem cavity had little or negative effect on the ability of HRP to oxidise luminol, reduction in the size of these residues has achieved an increase in CiP-like behaviour.

6.4 Conclusions and discussion

As remarked on in Chapter 1, the role of luminol in clinical assays is key, with current reliance on additional substrates to enhance the luminescent signal a source of error in these assays. A method of increasing the luminescence produced by the HRP molecules in assay conditions without the use of enhancer molecules could prove of commercial significance. The previous catalytic analysis, combined with this new crystal structure of this enzyme variant, show a method by which the activity of this versatile and robust enzyme can be enhanced with relevant application.

This crystal structure has led to the clear identification of residues which could be targeted for further site directed mutagenesis of the HRP access channel to expand this channel, increasing substrate accessibility to the oxidising centre of the enzyme. These, while playing little role in the catalytic activity of the enzyme, play a significant role in the accessibility of the haem edge for substrate binding. As seen in the case of luminol binding, creation of more CiP-like residues at the entrance to the access channel increases the catalytic turnover by HRP by 84-fold compared to the wild type enzyme.

6.5 Future work

Future variants in which the two double mutants described above and in the previous work are combined, creating a F68A/A140G/F179A variant would give an increasingly more accessible haem access channel, potentially increasing the substrate range of the enzyme and increasing the rate of reactions with known substrates.

It might be pertinent to combine the triple mutant variant, if they were to prove as stable as the double mutants, with the variants created in the earlier studies of Tanaka *et al.*, since their hypothesis that the charge on residues in this location could stabilise the polar side chains seen in the luminol molecule. Whilst the over-riding factor appears to be the steric hindrance of the associated residues at the access channel entrance, the charge stabilisations may now play a role in the orientation of the luminol molecule and hence further enhance the efficacy of the enzyme.

Chapter 7

Conclusions and final discussion

It has long been known how closely a protein's structure relates to its function, and particularly so in the case of peroxidases, since these enzymes have been studied extensively and their structure-function relationships probed in great detail.

HRP, as shown throughout this work, is a robust enzyme with commercial applications as an immunodiagnostic reporter enzyme and in the catalysis of difficult chemical transformations. It is also a target for site directed mutagenesis and directed evolution as it has been studied as a model haem enzyme system since the early 1940's and, as such, the relationship between enzyme structure and function is relatively well understood.

Following the development of methods to produce the active recombinant enzyme in *E.coli* (Smith *et al.*, 1990) and completion of the crystallographic structure in 1997 (Gajhede *et al*., 1997) it has been possible to identify the structural requirements for activity and to extend these activities by protein engineering techniques. Three aspects relating to the enhancement of the 'normal' and 'designed' activities of selected variants have been explored at the structure function level in this work.

Earlier work (Gajhede *et al*., 1997; White *et al*., 1997) highlighted residues that interact with aromatic substrates but that also potentially occlude access to the reactive haem edge by larger bulky substrates of potential commercial interest. The X-ray structure of the HRP variant A140G/F179A was solved to 2.0Å. A larger engineered cavity at the haem edge was observed consistent with the ability of the variant to oxidise luminol directly, a property not seen in the wild-type enzyme. The structure factors for other residues in the haem access channel were not significantly affected, despite the enlarged haem access cavity.

The structure of the HRP variant S167M was also solved, a variant of significant interest due to its ability to form a novel sulphonium linkage to the haem vinyl group (K. Cali, DPhil thesis, University of Sussex). The sulphur to β vinyl distance was found to be 3.15Å, compared to 1.7Å in the natural sulphonium linkage of myeloperoxidase. This implies that significant thermal motion in the structure is required for the haem-protein crosslink to form, accounting for the relatively slow autocatalytic modification process observed in the presence of hydrogen peroxide.

Previous work (Ngo and Smith, Int. Pat. No. WO/2007/020428) has shown that HRP engineered with a more open distal haem pocket (mimicking that of chloroperoxidase or cytochrome P450s) with a weak surrogate base (provided by a Glu residue as in chloroperoxidase) was capable of both entantioselective sulphoxidation and epoxidation. Building on this work, an alternative variant was designed in which the location of the weak base, provided by a Glu or Asp residue, has been varied within the active site. In particular, the HRP variant R38E:F41A:H42A (EAA) catalyses the production of the 1 and 2 naphthol from naphthalene at a rate of 124±4 min^{-1}, suggesting the generation of an epoxide intermediate in the active site. The wild-type enzyme does not catalyse this reaction or the sulphoxidation reactions described for earlier variants. This is believed to be the first report of aromatic C-H bond activation by an engineered plant peroxidase and is unusual in that C-H bond activation of this type normally requires a P450-type thiolate ligated haem system. Equilibrium binding studies show that naphthalene binds to the engineered haem cavity with an estimated K_d of 30±2 μM. Unfortunately, crystals of HRP variants described by Ngo and Smith (2007) and of the new EAA variant described here could not be obtained, despite many crystallisation attempts under a wide range of conditions.

This present work has shown how what may seem like relatively insignificant changes to the amino acid sequence of a protein, with the substitution of a very few residues, can have dramatic effects on the structure, function and stability of that protein.

In the case of the mutations that alter the distal haem cavity, the adjustment of these residues to mimic the active site of the chloroperoxidase molecule has resulted in the

HRP variants being capable of performing new activities. This has been shown in previous work, but the current work has shown that the adjustment of residue position in the active site has a profound effect. The addition of a glutamate residue at position 38 rather than argenine in the wild type HRP, in combination with alanine residues substituted in positions 41 and 42, results in the HRP variant EAA being able to catalyse the direct transfer of oxygen on to a molecule of naphthalene and the formation of α- and β- naphthols. This extends the range of substrates that HRP-C* "peroxygenase" variants have been shown to act upon. Although these rates are not yet at a level where the products are useful commercially, previous variants have been transferred to a *Coprinus cinereus* expression system for larger scale production (Doyle and Smith, unpublished data).

The peroxygenase variants described in this work provide a good basis for introduction in to a yeast expression system. This would form a basis for a primary screen using directed evolution techniques to introduce novel catalytic activities into a peroxidase. Preliminary work on this has been started though with low yields of protein for the CiP-peroxygenase variants. Work is starting on CiP-EAA which might allow a direct screening assay with the introduced naphthalene activity.

This work has shown the feasibility of the HRP variant system in the development of selective oxygenation catalysts, as seen by the specific oxygenation of naphthalene by the EAA variant. The role of catalytic site residues is further highlighted by the comparison between the oxygenation capabilities of the EAA variant when compared with the previous AAE and AAEH variants. This leaves the potential for future work to further explore the haem cavity with an extended range of substrates.

References

Abdel-Meguid, S.S., in Sanderson, M.R., Skelly, J., (eds.) (2007) Macromolecular crystallography: conventional and high-throughput methods. *Oxford University Press*

Andersson, L.A., Bylkas, S.A., Wilson, A.E., (1996) Spectral analysis of lactoperoxidase. Evidence for a common heme in mammalian peroxidases. *The Journal of Biological Chemistry*, **271**, 7, 3406–3412,

Baumeister, W., Steven, A.C., (2000) Macromolecular electron microscopy in the era of structural genomics. *Trends in Biochemical Sciences*, **25**, 625-631

Berglund, G.I., Carlsson, G.H., Smith, A.T., Szoke, H., Henriksen, A., Hajdu, J., (2002) The catalytic pathway of horseradish peroxidase at high resolution. *Letters to Nature*, **417**, 463-468

Berman, H.M., Bhat, T.N., Bourne, P.E., Feng, Z., Gilliand, G., Weissig, H., Westbrook, J., (2000) The Protein Data Bank and the challenge of structural genomics, *Nature Structural & Molecular Biology,* 7, 957-959

Berman, H.M., (2007) The Protein Data Bank: a historical perspective. *Acta Crystallographica Section A*, A64, 88-95

Bernhardt, R., (2006) Cytochromes P450 as versatile biocatalysts. *Journal of Biotechnology,* 124, 128-145

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F. Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M., (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *The Journal of Molecular Biology,* 112, 535-542

Bertini, I., Gray, H.B., Stiefel, E.I., Valentine, J.S., (eds) (2007) Biological Inorganic Chemistry: Structure and Reactivity. University Science Books, U.S.

Biertümpfel, C., Basquin, J., Suck, D., Sauter, C., (2002). Crystallisation of biological macromolecules using agarose gel. *Acta Crystallographica Section D*, D58, 1657–1659

Blodig W., Smith, A.T., Winterhalter, K., Piontek, K., (1999) Evidence from Spin-Trapping for a Transient Radical on Tryptophan Residue 171 of Lignin Peroxidase. *Archives of Biochemistry and Biophysics,* **370**, 1, 86-92

Brzozowski, A.M., Walton, J., (2001) Clear strategy screens for macromolecule crystallisation. *Journal of Applied Crystallography*, 34, 97-101

Carpena, X., Vidossich, P., Schroettner, K., Calisto, B.M., Banerjee, S., Stampler, J., Soudi, M., Furtmüller, P.G., Rovira, C., Fita, I., Obinger, C.,(2009) Essential role of proximal histidine-asparagine interaction in mammalian peroxidases. *The Journal of Biological Chemistry,* **284**, 38, 25929-25937

Cali K., (2009) Doctoral Thesis, University of Sussex

Chayen, N.E. (2004). Turning protein crystallization from an art into a science. *Current Opinion in Structural Biology*, 5, 14:577-583

Chayen, N.E., in Sanderson, M.R., Skelly, J., (eds.) (2007) Macromolecular crystallography: conventional and high-throughput methods. *Oxford University Press*

Chayen, N.E., Saridakis, E., (2008) Protein crystallisation: from purified protein to diffraction-quality crystal. *Nature Methods,* **5**, 2, 147-153

Colas, C., Kuo, J.M., Ortiz de Montellano, P.R., (2002) Asp-225 and Glu-375 in autocatalytic attachment of the prosthetic heme group of lactoperoxidase. *The Journal of Biological Chemistry,* **277**, 9, 7191-7200

Colas, C., Ortiz de Montellano, P.R., (2004) Horseradish peroxidase mutants that autocatalytically modify their prosthetic heme group. *The Journal of Biological Chemistry,* **279**, 23, 24131-24140

DePillis, G. D., Ozaki, S., Kuo, J. M., Maltby, D. A., Ortiz de Montellano, P. R., (1997) Autocatalytic Processing of Heme by Lactoperoxidase Produces the Native Protein-Bound Prosthetic Group. *Journal of Biological Chemistry*, 272, 8857-8860

Devarajan, A., Gaenko, A.V., Ryde, U., (2008) Effect of covalent links on the structure, spectra, and redox properties of myeloperoxidase – A density functional study. *Journal of Inorganic Biochemistry*, 102, 1549-1557

Doyle, W.A., Blodig, W., Veitch, N.C., Piontek, K., Smith, A.T., (1998) Two substrate interaction sites in lignin peroxidase revealed by site-directed mutagenesis. *Biochemistry, 37,* 15097-15105

Doyle, W.A., Smith, A.T., (1996) Expression of lignin peroxidase H8 in *Escherichia coli*: folding and activation of the recombinant enzyme with Ca2+ and haem. *Biochemical Journal,* 315, 15-19

Dunford, H.B., (1999) Heme Peroxidases. *Wiley-Blackwell*

Dunford, H.B., Stillman, J.S., (1976) On the function and mechanism of action of peroxidases. *Coordination Chemistry Reviews,* 19, 187-251

Dunford, H.B., in Everse, J., Everse, K.E., Grisham, M.B., (eds) (1990) Peroxidases in Chemistry and Biology. *CRC Press*

Dunford, H.B., Stillman, J.S., (1976) On the function and mechanism of action of peroxidases. *Coordination Chemistry Reviews,* 19, 187-251

Duroux, L., Welinder, K., (2003) The peroxidase gene family in plants: a phylogenetic overview. *Journal of Molecular Evolution,* 57, 397-407

Egawa T., Shimada H., Ishimura Y., (1994) Evidence for Compound I Formation in the Reaction of Cytochrome-P450cam with *m*-Chloroperbenzoic Acid. *Biochemical and Biophysical Research Communications*, **201**, 3, 1464-1469

Emsley, P., Cowtan, K., (2004) *Coot*: model-building tools for molecular graphics. *Acta Crystallographic Section D,* D60, 2126-2132

Ewetz, L., Thore, A., (1976) Factors affecting the specificity of the luminol reaction with hematin compounds. *Analytical Biochemistry,* 71, 564-570

Everse, J., Everse, K.E., Grisham, M.B., (eds) (1990) Peroxidases in Chemistry and Biology. *CRC Press*

Fenna, R.E., Zeng, J., Davey, C.D., (1994) Structure of the green heme in myeloperoxidase. *Archives of Biochemistry and Biophysics*, **316**, 1, 653-656

Fiedler, T.J., Davey, C.D., Fenna, R.E., (2000) X-ray crystal structure and characterization of halide-binding sites of human myeloperoxidase at 1.8 A resolution. *The Journal of Biological Chemistry,* **275**, 16, 11964-11971

Furtmüller, P.G., Jürgen, A., Walter, J., Zederbauer, M., Christa, J., Obinger, C., (2005) Standard reduction potentials of all couples of the peroxidase cycle of lactoperoxidase. *Journal of Inorganic Biochemistry*, 99, 5, 1220-9.

Furtmüller, P.G., Zederbauer, M., Jantschko, W., Helm, J., Bogner, M., Jakopitsch, C., Obinger, C., (2006) The structure and catalytic mechanisms of human peroxidases. *Archives of Biochemistry and Biophysics,* 445, 199-213

Gajhede, M., (2001) Protein engineering of peroxidases and cytochrome P450.

Biochemical Society Transactions, 29, 91-99

Gajhede, M., Schuller, D.J., Henriksen, A., Smith., A.T & Poulos, T.L., (1997) Crystal structure determination of classical horseradish peroxidase at 2.15Å resolution. *Nature Structural Biology*, 4 1032-1039

Gonzalez, A., (2003) Faster data-collection strategies for structure determination using anomalous dispersion. *Acta Crystallographica Section D*, **D59**, 2, 315-322

Guss, J.M., Merritt, E.A., Phizackerley, R.P., Hedman, B., Murata, M., Hodgson, K.O., Freeman, H.C., (1988) Phase determination by multiple-wavelength x-ray diffraction: crystal structure of a basic "blue" copper protein from cucumbers. *Science* 241, 806-11

Hager, L.P., Morris, D.R., (1966) Chloroperoxidase, Isolation and properties of the crystalline glycoprotein. *Journal of Biological Chemistry*, **241**, 8, 1763-1768

Henisch, H.K., (1988) Crystals in gels and Liesegang rings. *Cambridge University Press*

Henne, K.R., Kunze, K.L., Zheng, Y.-M., Christmas, P., Soberman, R.J., Rettie, A.E., (2001) Covalent linkage of prosthetic heme to CYP4 family of P450 enzymes. *Biochemistry,* 40, 12925-12931

Henriksen, A., Smith, A.T., Gajhede, M., (1999) The structures of the horseradish peroxidase C-ferulic acid complex and the ternary complex with cyanide suggest how peroxidases oxidise small phenolic substrates. *The Journal of Biological Chemistry,* **274**, 49, 35005-35011

Henriksen, A., Schuller, D.J., Meno, K., Welinder, K.G., Smith, A.T., Gajhede, M., (1998) Structural interactions between HRPC and the substrate benzhydroxamic acid determined by X-ray crystallography. *Biochemistry,* **37**, 22, 8054-8060

Hoffrichter, M., Ullrich, R., (2006) Heme-thiolate haloperoxidases: versatile

biocatalysts with biotechnological and environmental significance. *Applied Microbiology and Biotechnology,* 71, 276-288

Houborg, K., Harris, P., Petersen, J., Rowland, P., Poulsen, J.-C.N., Schneider, P., Vind, J., Larsen, S., (2003) Impact of the physical and chemical environment on the molecular structure of *Coprinus cinereus* peroxidase. *Acta Crystallographica Section D,* D59, 989-996

Howes, B.D., Rodriguez-Lopez, J.N., Smith, A.T., Smulevich, G., (1997) Mutation of distal residues of horseradish peroxidase: influence on substrate binding and cavity properties. *Biochemistry,* 36, 1532-1543

Huang, L., Wojciechowski, G., Ortiz de Montellano, P.R., (2006) Role of heme-protein covalent bonds in mammalian peroxidases. *The Journal of Biological Chemistry,* **281**, 28, 18983-18988

Jancarik, J., Kim, S.-H., (1991) Sparse matrix sampling: a screening method for crystallization of proteins. *Journal of Applied Crystallography,* **24**, 4, 409-411

Jones, P., (2001) Roles of water in heme peroxidase and catalase mechanisms. *The Journal of Biological Chemistry,* **276**, 17, 13791-13796

Jones, P., Dunford, H.B., (2005) The mechanism of Compound I formation revisited. *Journal of Inorganic Biochemistry,* 99, 2292-2298

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., Phillips, D.C., (1958) A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature,* **181**, 4610, 662-666

Klibanov, A.M., Tu, T.-M., Scott, K.P., (1983) Peroxidase-catalysed removal of phenols from coal-conversion waste waters. *Science,* 221, 259-261

Kluge, M.G., Ullrich, R., Scheibner, K., Hofrichter, M., (2007) Spectrophotometric assay for detection of aromatic hydroxylation catalyzed by fungal haloperoxidase–peroxygenase. *Applied Microbiology and Biotechnology* , 75, 1473–1478

Kluge, M.G., Ullrich, R., Dolge, C., Schneibner, K., Hofrichter, M., (2009) Hydroxylation of naphthalene by aromatic peroxygenase from Agrocybe aegerita proceeds via oxygen transfer from H2O2 and intermediary epoxidation. *Applied Microbiology and Biotechnology,* 81, 1071-1076

Kooter I.M., Koehler, B.P., Moguilevsky, N., Bollen, A., Wever, R., Johnson, M.K., (1999 a) The Met243 sulfonium ion linkage is responsible for the anomalous magnetic circular dichroism and optical spectral properties of myeloperoxidase. *Journal of Biological Inorganic Chemistry*, 4, 684-691

Kooter, I.M., Moguilevsky, N., Bollen, A., Sijtsema, N.M., Otto, C., Dekker, H.L., Wever, R., (1999 b) Characterization of the Asp94 and Glu242 mutants in myeloperoxidase, the residues linking the heme group via ester bonds. *European Journal of Biochemistry,* 264, 211-217

Koua, D., Cerutti, L., Falquet, L., Sigrist, C.J., Theiler, G., Hulo, N., Dunand. C., (2009) PeroxiBase: a database with new tools for peroxidase family classification. *Nucleic Acids Research,* 37, D261-D266

Kunishima, N. Fukuyama, K., Matsubara, H., Hatanaka, H., Shibano, Y., Amachi, T., (1994) Crystal structure of the fungal peroxidase from *Arthromyces ramosus* at 1.9 A resolution. Structural comparisons with the lignin and cytochrome c peroxidases. *Journal of Molecular Biology*, 235, 331-344

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., (1993) Procheck – programs to check the stereochemical quality of protein structures. *Journal of Applied Crystallography,* 26, 283

LeBrun, L.A., Xu, F., Kroetz, D.L., Ortiz de Montellano, P.R., (2002) Covalent attachment of the heme prosthetic group in the CYP4F cytochrome P450 family. *Biochemistry*, 41, 5931-5937

LeBrun, L.A., Hoch, U., Ortiz de Montellano, P.R., (2002) Autocatalytic mechanism and consequences of covalent heme attachment in the cytochrome P4504A family. *The Journal of Biological Chemistry,* **277**, 15, 12755-12761

Leslie, A.G.W., (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 & ESF-EAMCB Newsletter on Protein Crystallography*, 26

Loughran, N.B., O'Connor, B., O'Fágáin, C., O'Connell, M.J., (2008) The phylogeny of the mammalian heme peroxidases and the evolution of their diverse functions. *BMC Evolutionary Biology,* **8**, 101

Markwalder, H.U., Neukom, H., (1976) Diferulic acid as a possible crosslink in hemicelluloses from wheat germ. *Phytochemistry,* 15, 836-837

Marquette, C.A., Blum, L.J., (2006) Applications of the luminol chemiluminescent reaction in analytical chemistry. *Bioanalytical Chemistry,* 385, 546-554

Matsunaga, I., Shiro, Y., (2004) Peroxide-utilizing biocatalysts: structural and functional diversity of heme-containing enzymes. *Current Opinion in Chemical Biology,* 8, 127-132

McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C., Read, R.J., (2005) Likelihood-enhanced fast translation functions. *Acta Crystallographica D,* **D61**, 458-464

Metcalfe, C.L., Ott, M., Patel, N., Singh, K., Mistry, S.C., Goff, H.M., Raven, E.L., (2004) Autocatalytic formation of green heme: evidence for H 2 O 2-dependent formation of a covalent methionine-heme linkage in ascorbate peroxidase. *Journal of the American Chemical Society,* 126, 16242-16248

Nagababu, E., Rifkind, J.M., (2000) Heme degradation during autooxidation of oxyhemoglobin. *Biochemical and biophysical research communications,* 273, 839-845

Nagano, S., Tanaka, M., Watanabe, Y., Morishima, I., (1995) Putative hydrogen bond network in the heme distal site of horseradish peroxidase. *Biochemical and Biophysical Research Communications,* **207**, 1, 417-423

Newman, J., Egan, D., Walter, T.S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J.L., Stuart, D.I., Perrakis, A., (2005) Towards rationalization of crystallization screening for small- to medium- sized laboratories: the PACT/JCSG+ strategy. *Acta Crystallographica Section D,* **D61**, 10, 1426

Newmyer, S.L., Ortiz de Montellano, P.R., (1995) Rescue of the Catalytic Activity of an H42A Mutant of Horseradish Peroxidase by Exogenous Imidazoles. *Journal of Biological Chemistry*, **271**, 25, 14891–14896

Ngo, E., (2006) Doctoral Thesis, University of Sussex

Ngo, E., Smith, A.T., (2007) (WO/2007/020428) Novel peroxidases and uses. *World Intellectual Property Organization*

Omura T., (2005) Heme-thiolate proteins. *Biochemical and Biophysical Research Communications*, 338, 404–409

Oxvig, C., Thomsen, A.R., Overgaard, M.T., Sørensen, E.S., Højrup, P., Bjerrum, M.J., Gleich, G.J., Sottrup-Jensen, L., (1999) Biochemical evidence for heme linkage through esters with Asp-93 and Glu-241 in human eosinophil peroxidase. The ester with Asp-93 is only partially formed in vivo. *The Journal of Biological Chemistry,* **274**, 24, 16953-16958

Ozaki, S., Ortiz de Montellano, P.R., (1995) Molecular engineering of horseradish peroxidase: thioether sulfoxidation and styrene epoxidation by Phe-41 leucine and

threonine mutants. *Journal of the American Chemical Society,* 117, 7056-7064

Passardi, F., Theiler, G., Zamocky, M., Cosio, C., Rouhier, N., Teixera, F., Margis-Pinheiro, M., Ioannidis, V., Penel, C., Falquet, L., Dunand, C., (2007) PeroxiBase: The peroxidase database. *Phytochemistry*, **68**, 12, 1605-11

Patterson, W.R., Poulos, T.L., Goodin, D.B., (1995) Identification of a porphyrin pi cation radical in ascorbate peroxidase Compound I, *Biochemistry,* 34, 4342-4345.

Pecyna, M.J., Ullrich, R., Bittner, B., Clemens, A., Scheibner, K., Schubert, R., Hofrichter, M., (2009) Molecular characterization of aromatic peroxygenase from *Agrocybe aegerita*. *Applied Microbiology and Biotechnology*, 84, 885–897

Petersen, K.H., (2009) Novel horseradish peroxidase substrates for use in immunohistochemistry. *Journal of Immunological Methods,* 340, 86-89

Piontek, K., Glumoff, T., Winterhalter, K., (1993). Low pH crystal structure of glycosylated lignin peroxidase from *Phanerochaete chrysosporium* at 2.5 Å resolution. *FEBS Letters*, 315, 119–124.

Piontek, K., Ullrich, R., Liers, C., Diederichs, K., Plattner, D.A., Hofrichter, M., (2010) Crystallization of a 45 kDa peroxygenase/peroxidase from the mushroom *Agrocybe aegerita* and structure determination by SAD utilizing only the haem iron. *Acta Crystallographica Section F,* **F66**, 6, 693-698

Piripou, Z., Bottrill, A.R., Svistunenko, D.A., Efimov, I., Basran, J., Mistry, S.C., Cooper, C.E., Raven, E.L., (2007 a) The reactivity of heme in biological systems: autocatalytic formation of both tyrosine-heme and tryptophan-heme covalent links in a single protein architecture. *Biochemistry,* 46, 13269-13278

Piripou, Z., Bottrill, A.R., Metcalfe, C.M., Mistry, S.C., Badyal, S.K., Rawlings, B.J., Raven, E.L., (2007 b) Autocatalytic formation of a covalent link between tryptophan 41 and

the heme in ascorbate peroxidase. *Biochemistry*, 46, **8**, 2174-2180

Potterton, E., McNicholas, S., Krissinel, E., Cowtan, K., Noble, M., (2002) The CCP4 molecular-graphics project. *Acta Crystallographica Section D,* **D58**, 1955-1957

Poulos, T.L., Kraut, J., (1980) The stereochemistry of peroxidase catalysis. *The Journal of Biological Chemistry,* **255**, 17, 8199-8205

Poulos, T.L., Finzel, B.C., Howard, A.J. (1987). High-resolution crystal structure of cytochrome P450cam. Journal of *Molecular Biology,* 195, 687-700

Powell, H.R., (1999) The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallographica Section D,* D10, 1690-1695

Rose, J.P., Wu, C.-K., Lieu, Z.-J., Newton, M.G., Wang, B.C., (2001) Using single wavelength anomalous scattering data for in-house protein structure determination. *Rigaku Journal,* 18, 4-12

Rhodes, G., (2000 (reprint)) Crystallography made crystal clear (second edition). *Academic Press*

Rodiguez-Lopez, J.N., Smith, A.T., Thorneley, R.N.F., (1996) Role of arginine 38 in horseradish peroxidase. A critical residue for substrate binding and catalysis. *The Journal of Biological Chemistry,* **271**, 8, 4023-4030

Ruf, J., Carrayon, P., (2005) Structural and functional aspects of thyroid peroxidase. *Archives of Biochemistry and Biophysics,* 445, 269-277

Sambrook, J., Fritsch, E.F., Maniatis, T., (1989) Molecular cloning, a laboratory manual. *Cold Spring Harbour Press*

Sanders, S.A., Bray, R.C., Smith, A.T., (1994) pH-dependent properties of a mutant

horseradish peroxidase isoenzyme C in which Arg38 has been replaced by lysine. *European Journal of Biochemistry,* 224, 1029-1037

Sanderson, M.R., Skelly, J., (eds.) (2007) Macromolecular crystallography: conventional and high-throughput methods. *Oxford University Press*

Schaefer, W.H., Harris, T.M., Guengerich, F.P., (1995) Characterizaion of the enzymatic and nonenzymatic peroxidative degredation of iron porphyrins and cytochrome P-450 heme. *Biochemistry,* 24, 2354-3263

Schonbaum, G.R., Lo, S., (1972) Interaction of peroxidases with aromatic peracids and alkyl peroxides. *The Journal of Biological Chemistry,* **247**, 10, 3353-3360

Schuller, D.J., Ban, N., van Huystee, R.B., McPherson, A., Poulos, T.L., (1996) The crystal structure of peanut peroxidase. *Structure,* 4, 311-321

Shaw, P.D., Hager, L.P., (1959) *Journal of the American Chemical Society*, 81, 1011

Smith, A.T., Sanders, S.A., Thorneley, R.N.F., Burke, J.F., Bray, R.C., (1992) Characterisation of a haem active-site mutant of horseradish peroxidase, Phe41 → Val, with altered reactivity towards hydrogen peroxide and reducing substrates. *European Journal of Biochemistry,* 207, 507-519

Smith, A.T., Santama, N., Dacey, S., Edwards, M., Bray, R.C., Thorneley, R.N.F., Burke, J.F., (1990) Expression of a synthetic gene for horseradish peroxidase C in Escherichia coli and folding and activation of the recombinant enzyme with Ca^{2+} and heme. *The Journal of Biological Chemistry,* **265**, 22, 13335-13343

Smith, A.T., Veitch, N.C., (1998) Substrate binding and catalysis in heme peroxidases. *Current Opinion in Chemical Biology,* 2, 269-278

Smulevich, G., Feis, A., Focardi, C., Tams, J., Welinder, K.G., (1994) Resonance raman

study of the active site of *Coprinus cinereus* peroxidase. *Biochemistry,* 33, 15425-15432

Smulevich, G., (1998) Understanding heme cavity structure of peroxidases: comparison of electronic absorption and resonance raman spectra with crystallographic results. *Biospectroscopy,* 4, S3-S17

Smulevich, G., Jakopitsch, C., Droghetti, E., Obinger, C., (2006) Probing the structure and bifunctionality of catalase-peroxidase (KatG). *Journal of Inorganic Biochemistry,* 100, 568-585

Sono, M., Roach, M.P., Coulter, E.D., Dawson, J.H., 1996. Heme- containing oxygenases. *Chemical Reviews*, 96, 2841–2888

Stone, K.L., Behan, R.K., Green, M.G., (2005) X-ray absorption spectroscopy of chloroperoxidase Compound I: Insight into the reactive intermediate of P450 chemistry. *Proceedings of the National Academy of Sciences,* 102, 46, 16563–16565

Stura E.A., Nemerow G.R., Wilson I.A (1992) Strategies in the crystallization of glycoproteiens and protein complexes. *Journal of Crystal Growth*, 122, 273-285

Sundaramoorthy, M., Terner, J., Poulos, T.L., (1995) The crystal structure of chloroperoxidase: a heme peroxidase-cytochrome P450 functional hybrid. *Structure,* 3, 1367-1377

Tanaka, M., Ishimori, K., Mukai, M., Kitagawa, T., Morishima, I., (1997) Catalytic activities and structural properties of horseradish peroxidase distal His42 → Glu or Gln mutant. *Biochemistry,* 36, 9889-9898

Taurog, A., Wall, M., (1998) Proximal and Distal Histidines in Thyroid Peroxidase: Relation to the Alternatively Spliced Form, TPO-2. *Thyroid*, 8, **2**, 185-191

Ullrich, R., Dolge, C., Kluge, M., Hofrichter, M., (2008) Pyridine as novel substrate for

regioselective oxygenation with aromatic peroxygenase from *Agrocybe aegerita*. *FEBS Letters,* 582, 4100-4106

Ullrich, R., Hofrichter, M., (2005) The haloperoxidase of the agaric fungus *Agrocybe aegerita* hydroxylates toluene and naphthalene. *FEBS Letters,* 579, 6247-6250

Ullrich, R., Hofrichter, M., (2007) Enzymatic hydroxylation of aromatic compounds. *Cellular and Molecular Life Sciences,* 64, 271-293

Vagin, A., Teplyakov, A., (1997) MOLREP: an automated program for molecular replacement. *Journal of Applied Crystallography*, 30, 1022-1025

Vagin, A.A., Richelle, J., Wodak, S.J.,(1999) SFCHECK: a unified set of procedure for evaluating the quality of macromolecular structure-factor data and their agreement with atomic model. *Acta Crystallographica Section D*, D55, 191-205

Veitch, N.C., (2004 a) Horseradish peroxidase: a modern view of a classic enzyme. *Phytochemistry,* 65, 249-259

Veitch, N.C., (2004 b) Structural determinants of plant peroxidase function. *Phytochemistry Reviews,* 3, 3-18

Veitch N.C., Smith, A.T., (2001) Horseradish Peroxidase. *Advances in Inorganic Chemistry*, 51, 107–162

Veitch, N.C., Gao, Y., Smith, A.T., White, C.G., (1997) Identification of a critical phenylalanine residue in horseradish peroxidase, Phe 179, by site-directed mutagenesis and 1H-NMR: Implications for complex formation with aromatic donor molecules. *Biochemistry,* 36, 14751-14761

van de Velde, F., van Rantwijk, F., Sheldon, R.A., (2001) Improving the catalytic performance of peroxidases in organic synthesis. *Trends in Biochemistry*, **19**, 2, 73-80

van Wart, H.E., Zimmer, J., (1985) Resonance Raman evidence for the activation of dioxygen in horseradish oxyperoxidase. *The Journal of Biological Chemistry,* **260**, 14, 8372-8377

David Watkin, British Crystallographic Association annual meeting, Loughborough, (2009)

Weber, B.H., Goodkin, P.E., (1970) A modified microdiffusion proceedure for the growth of single protein crystals by concentration-gradient equilibrium dialysis. *Archives of Biochemistry and Biophysics, 141, 489-498*

Welinder, K.G., Mazza, G., (1977) Amino-acid sequences of heme-linked, histidine-containing peptides of five peroxidases from horseradish and turnip. *European Journal of Biochemistry,* 73, 353-358

Welinder, K.G., (1979) Amino acid sequences studies of horseradish peroxidase. *European Journal of Biochemistry,* 96, 483-502

Welinder, K.G., (1992) Superfamily of plant, fungal and bacterial peroxidases. *Current Opinion in Structural Biology,* 2, 388-393

Whitehead, T.P., Thorpe, G.H.G., Carter, T.J.N., Groucutt, C., Kricka, L.J., Enhanced luminescence proceedure for sensitive determination of peroxidase-labelled conjugates in immunoassay. *Nature,* 305, 158-159

Whitford, D., (2005) Proteins: Structure and Function, *Wiley*

Wojaczyńska, E., Wojaczyńska, J., (2010) Enantioselective synthesis of sulfoxides: 2000-2009. *Chemistry Reviews,* 110, 4303-4356

Zámocký, M., Jakopitsch, C., Furtmüller, P.G., Dunand, C., Obinger, C., (2004) The

peroxidase-cyclooxygenase superfamily: reconstructed evolution of critical enzyme of the innate immune system. *Proteins,* 72, 589-605

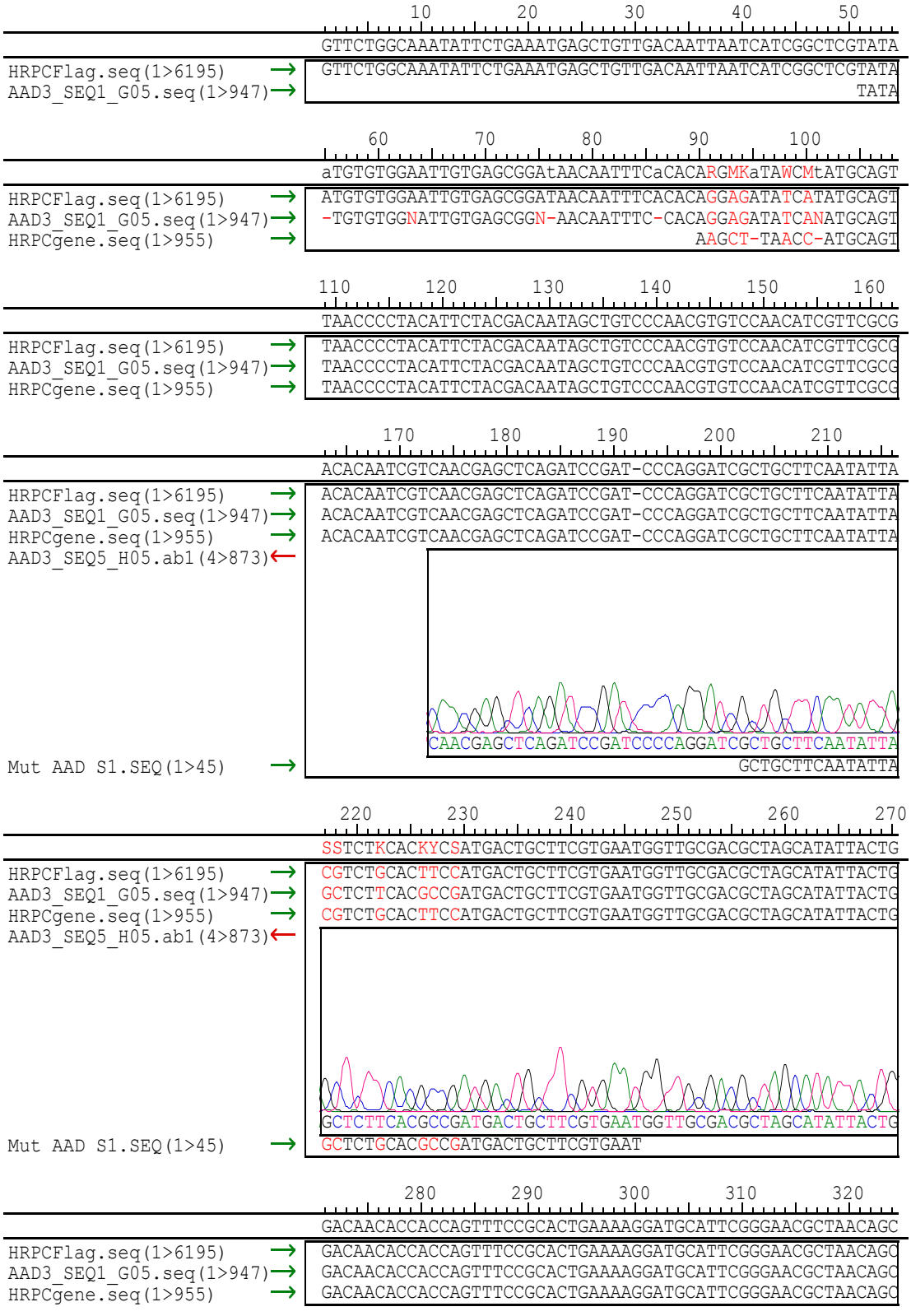
Zámocký, M., (2008) Phylogenetic relationships in class I of the superfamily of bacterial, fungal, and plant peroxidases. *European Journal of Biochemistry,* 271, 3297-3309

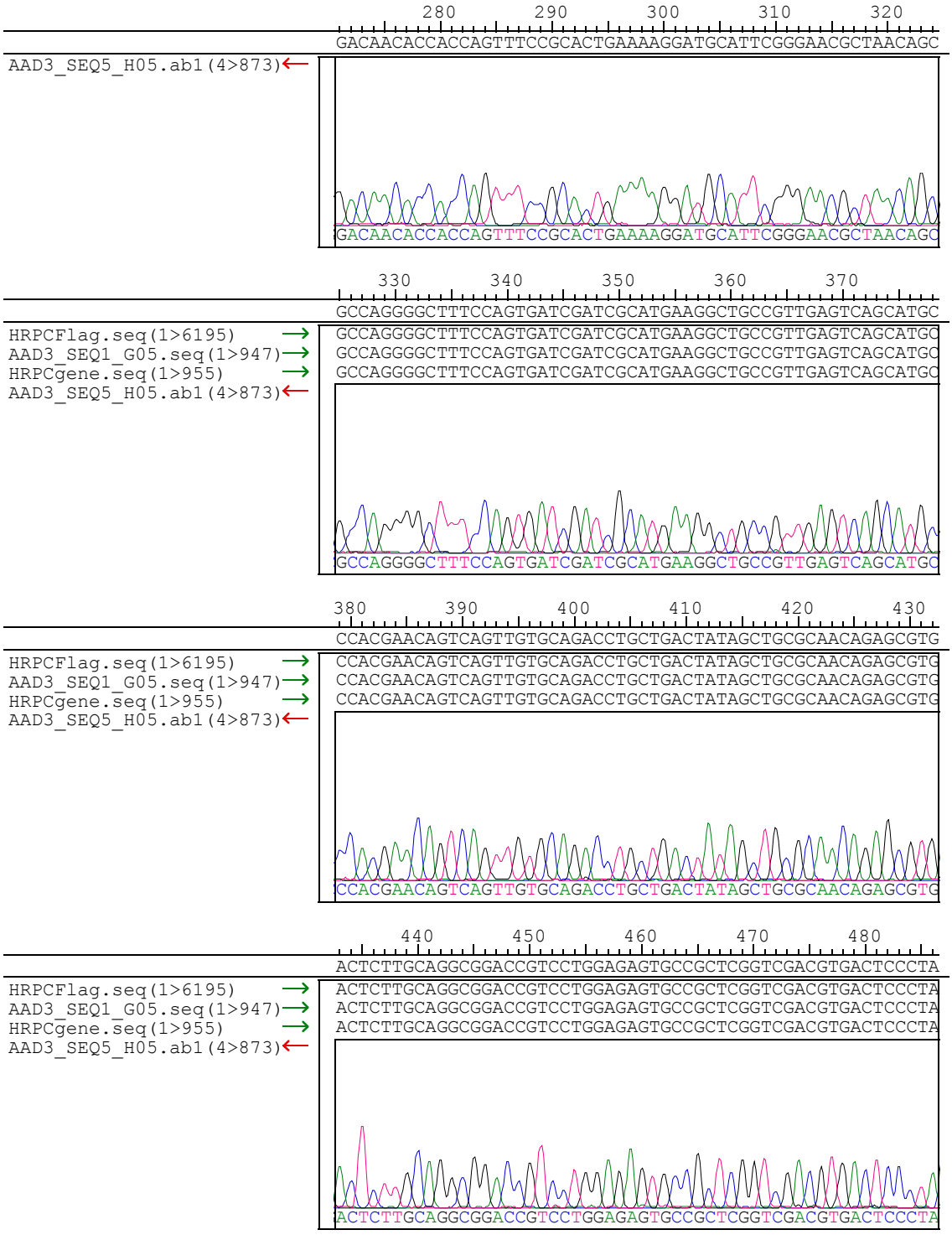
Zederbauer, M, Furtmüller, P.G., Brogioni, S., Jakopitsch, C., Smulevich, G., Obinger, C., (2007) Heme to protein linkages in mammalian peroxidases: impact on spectroscopic, redox and catalytic properties. *Natural Product Reports,* 24, 571-584

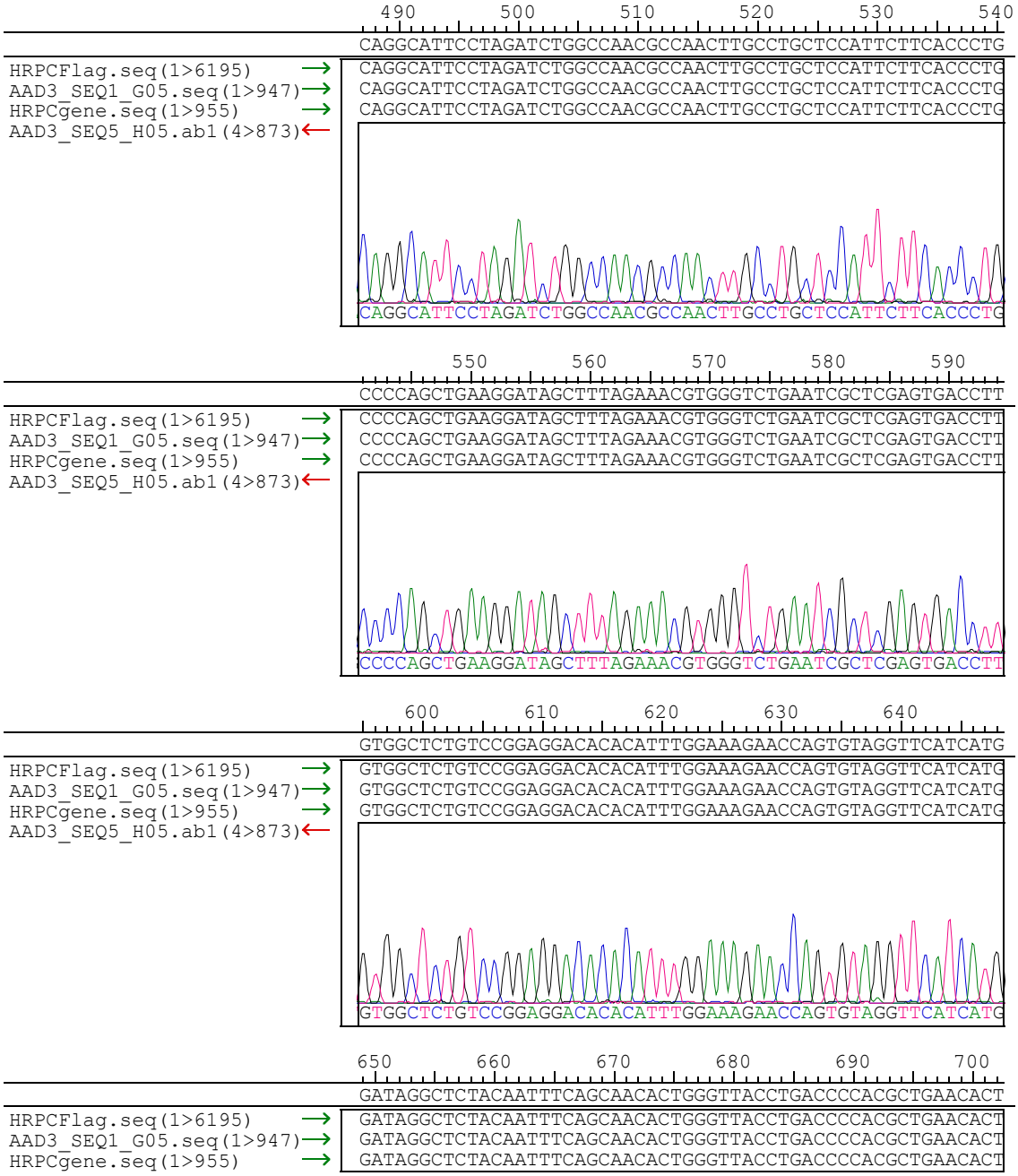
Zeng J., Fenna R.E., (1992) X-ray crystal structure of canine myeloperoxidase at 3 Å resolution. *Journal of Molecular Biology*, 226, **1**, 185-207

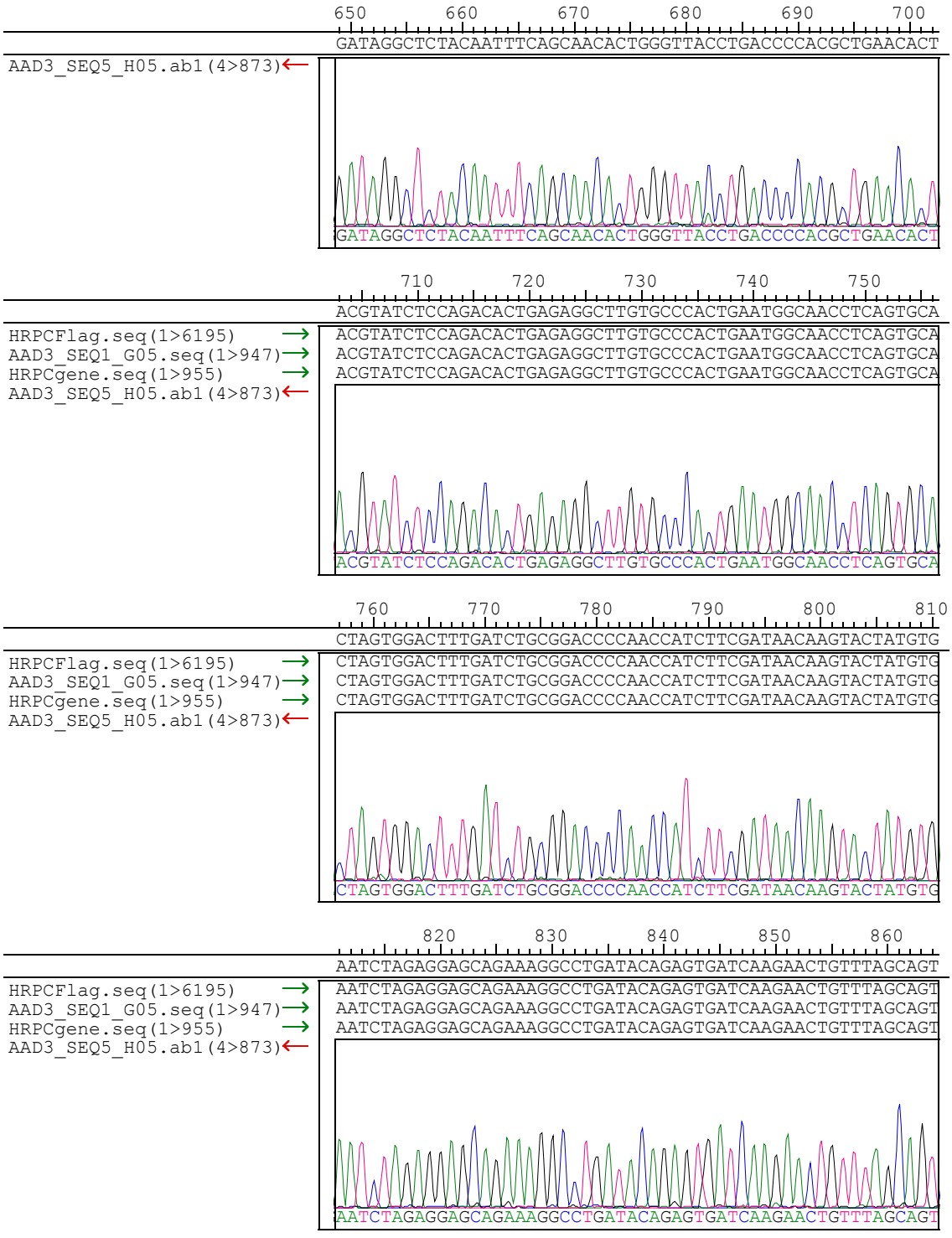
Appendix 1

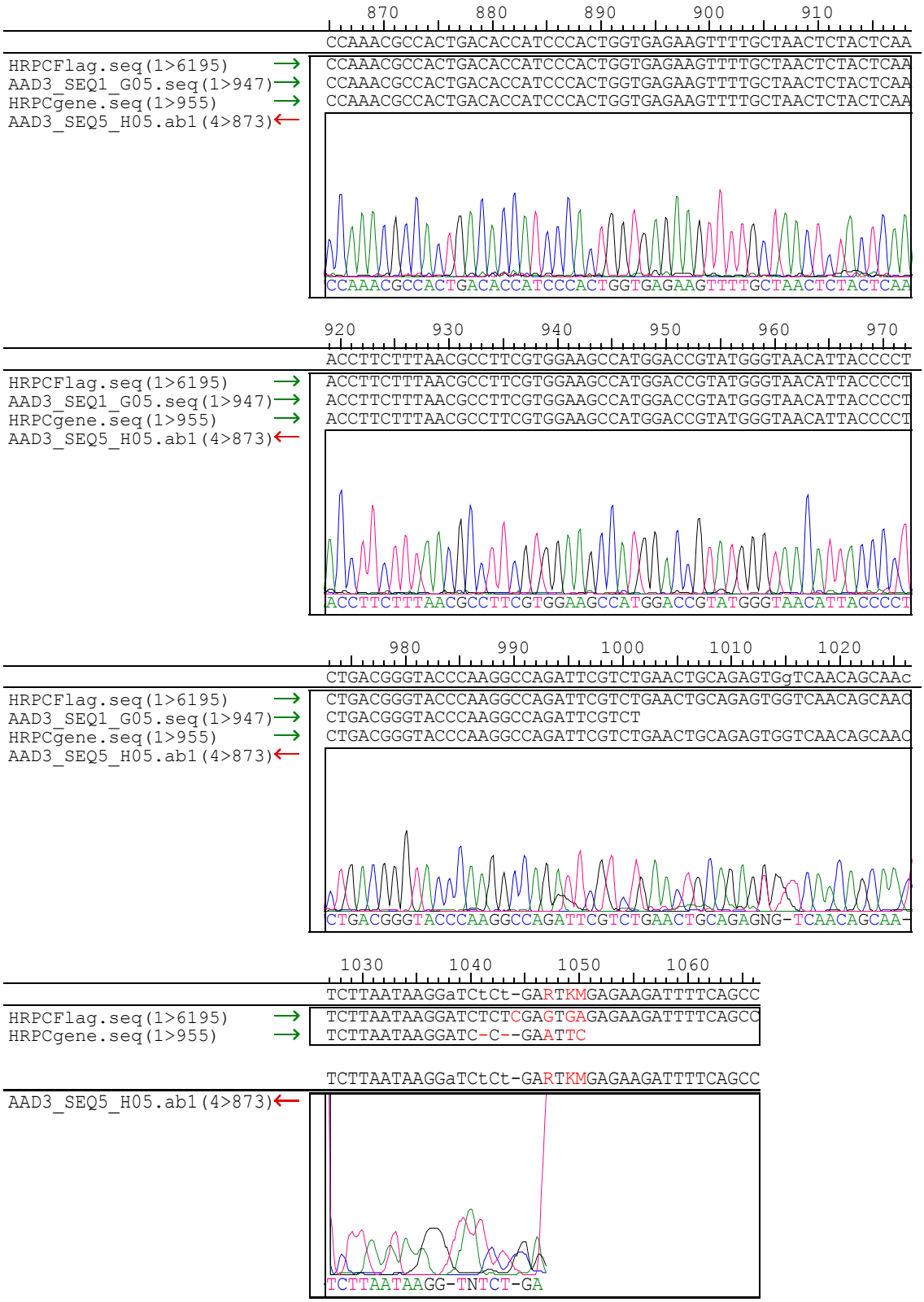
Sequence Contig of HRP-C* gene and AAD variant. Silent mutations that did not alter the amino acid sequence of the resultant protein were ignored.




280 290 300 310 320
GACAACACCACCAGTTTCCGCACTGAAAAGGATGCATTCGGGAACGCTAACAGC
AAD3_SEQ5_H05.ab1(4>873)
330 340 350 360 370
GCCAGGGGCTTTCCAGTGATCGATCGCATGAAGGCTGCCGTTGAGTCAGCATGC
HRPCFlag.seq(1>6195)
AAD3_SEQ1_G05.seq(1>947)
HRPCgene.seq(1>955)
AAD3_SEQ5_H05.ab1(4>873)
380 390 400 410 420 430
CCACGAACAGTCAGTTGTGCAGACCTGCTGACTATAGCTGCGCAACAGAGCGTG
HRPCFlag.seq(1>6195)
AAD3_SEQ1_G05.seq(1>947)
HRPCgene.seq(1>955)
AAD3_SEQ5_H05.ab1(4>873)
440 450 460 470 480
ACTCTTGCAGGCGGACCGTCCTGGAGAGTGCCGCTCGGTCGACGTGACTCCCTA
HRPCFlag.seq(1>6195)
AAD3_SEQ1_G05.seq(1>947)
HRPCgene.seq(1>955)
AAD3_SEQ5_H05.ab1(4>873)


490 500 510 520 530 540
CAGGCATTCCTAGATCTGGCCAACGCCAACTTGCCTGCTCCATTCTTCACCCTG
HRPCFlag.seq(1>6195)
AAD3_SEQ1_G05.seq(1>947)
HRPCgene.seq(1>955)
AAD3_SEQ5_H05.ab1(4>873)
550 560 570 580 590
CCCCAGCTGAAGGATAGCTTTAGAAACGTGGGTCTGAATCGCTCGAGTGACCTT
600 610 620 630 640
GTGGCTCTGTCCGGAGGACACACATTTGGAAAGAACCAGTGTAGGTTCATCATG
650 660 670 680 690 700
GATAGGCTCTACAATTTCAGCAACACTGGGTTACCTGACCCCACGCTGAACACT


650 660 670 680 690 700
GATAGGCTCTACAATTTCAGCAACACTGGGTTACCTGACCCCACGCTGAACACT
AAD3_SEQ5_H05.ab1(4>873)
710 720 730 740 750
ACGTATCTCCAGACACTGAGAGGCTTGTGCCCACTGAATGGCAACCTCAGTGCA
HRPCFlag.seq(1>6195)
AAD3_SEQ1_G05.seq(1>947)
HRPCgene.seq(1>955)
AAD3_SEQ5_H05.ab1(4>873)
760 770 780 790 800 810
CTAGTGGACTTTGATCTGCGGACCCCAACCATCTTCGATAACAAGTACTATGTG
HRPCFlag.seq(1>6195)
AAD3_SEQ1_G05.seq(1>947)
HRPCgene.seq(1>955)
AAD3_SEQ5_H05.ab1(4>873)
820 830 840 850 860
AATCTAGAGGAGCAGAAAGGCCTGATACAGAGTGATCAAGAACTGTTTAGCAGT
HRPCFlag.seq(1>6195)
AAD3_SEQ1_G05.seq(1>947)
HRPCgene.seq(1>955)
AAD3_SEQ5_H05.ab1(4>873)


870 880 890 900 910
CCAAACGCCACTGACACCATCCCACTGGTGAGAAGTTTTGCTAACTCTACTCAA
HRPCFlag.seq(1>6195)
AAD3_SEQ1_G05.seq(1>947)
HRPCgene.seq(1>955)
AAD3_SEQ5_H05.ab1(4>873)
920 930 940 950 960 970
ACCTTCTTTAACGCCTTCGTGGAAGCCATGGACCGTATGGGTAACATTACCCCT
980 990 1000 1010 1020
CTGACGGGTACCCAAGGCCAGATTCGTCTGAACTGCAGAGTGgTCAACAGCAAc
1030 1040 1050 1060
TCTTAATAAGGaTCtCt-GARTKMGAGAAGATTTTCAGCC
TCTTAATAAGGATC-C--GAATTC

Appendix 2

Sample of screen conditions around HRP-C* wild type (Henriksen *et al*., 1999), used with new HRP-C* 'peroxygenase' variant crystallisation attempts and for successful crystallisation of S167M and A140G/F179A HRP-C* variants. Scoring of drop contents allowed rationalisation of further screens if crystals were not observed. All crystallisation trays were incubated at 4°C when not being examined.

A1 5% PEG 6K 0.1M Cacodylate pH 6.5 0.1M Calcium Acetate	**A2** 10% PEG 6K 0.1M Cacodylate pH 6.5 0.1M Calcium Acetate	**A3** 15% PEG 6K 0.1M Cacodylate pH 6.5 0.1M Calcium Acetate	**A4** 20% PEG 6K 0.1M Cacodylate pH 6.5 0.1M Calcium Acetate	**A5** 25% PEG 6K 0.1M Cacodylate pH 6.5 0.1M Calcium Acetate	**A6** 30% PEG 6K 0.1M Cacodylate pH 6.5 0.1M Calcium Acetate
B1 5% PEG 6K 0.2M Cacodylate pH 6.5 0.1M Calcium Acetate	**B2** 10% PEG 6K 0.2M Cacodylate pH 6.5 0.1M Calcium Acetate	**B3** 15% PEG 6K 0.2M Cacodylate pH 6.5 0.1M Calcium Acetate	**B4** 20% PEG 6K 0.2M Cacodylate pH 6.5 0.1M Calcium Acetate	**B5** 25% PEG 6K 0.2M Cacodylate pH 6.5 0.1M Calcium Acetate	**B6** 30% PEG 6K 0.2M Cacodylate pH 6.5 0.1M Calcium Acetate
C1 5% PEG 6K 0.2M Cacodylate pH 6.5 0.1M Calcium Acetate	**C2** 10% PEG 6K 0.2M Cacodylate pH 6.5 0.1M Calcium Acetate	**C3** 15% PEG 6K 0.2M Cacodylate pH 6.5 0.1M Calcium Acetate	**C4** 20% PEG 6K 0.2M Cacodylate pH 6.5 0.1M Calcium Acetate	**C5** 25% PEG 6K 0.2M Cacodylate pH 6.5 0.1M Calcium Acetate	**C6** 30% PEG 6K 0.2M Cacodylate pH 6.5 0.1M Calcium Acetate
D1 5% PEG 6K 0.2M Cacodylate pH 6.5 0.2M Calcium Acetate	**D2** 10% PEG 6K 0.2M Cacodylate pH 6.5 0.2M Calcium Acetate	**D3** 15% PEG 6K 0.2M Cacodylate pH 6.5 0.2M Calcium Acetate	**D4** 20% PEG 6K 0.2M Cacodylate pH 6.5 0.2M Calcium Acetate	**D5** 25% PEG 6K 0.2M Cacodylate pH 6.5 0.2M Calcium Acetate	**D6** 30% PEG 6K 0.2M Cacodylate pH 6.5 0.2M Calcium Acetate

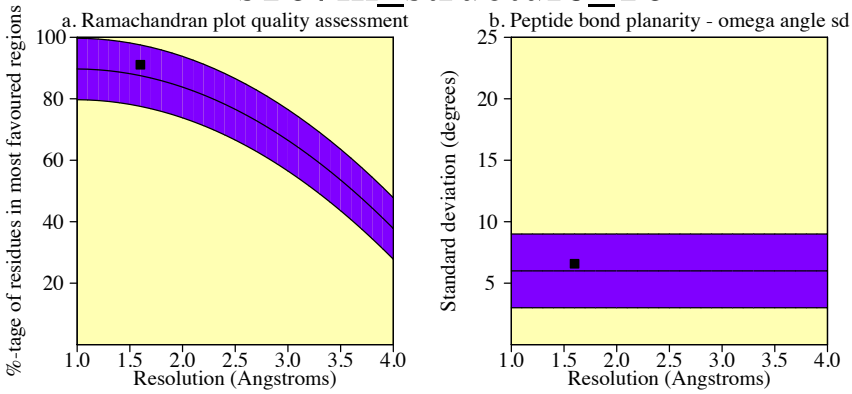
Appendix 3

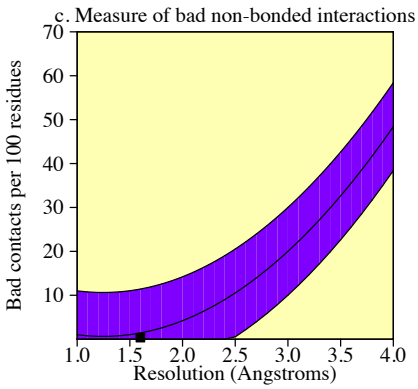
Data from ProCheck validating S167M structure – main chain parameters.

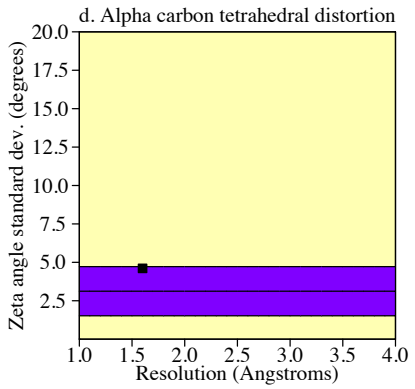
PROCHECK

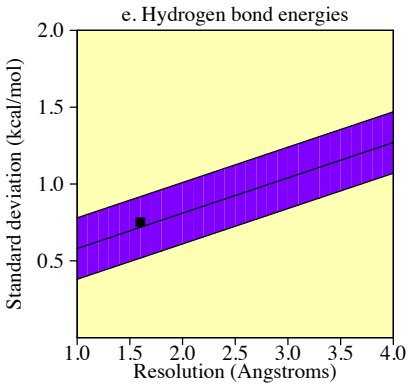
Main-chain parameters

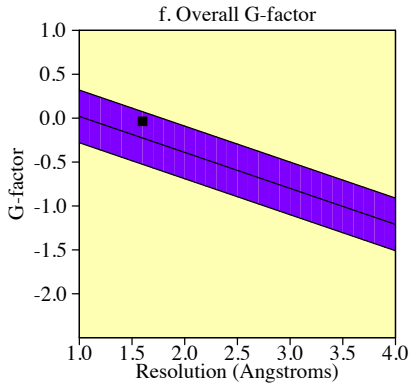
s167m_structure_16











Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	269	91.1	87.5	10.0	0.4	Inside
b. Omega angle st dev	304	6.6	6.0	3.0	0.2	Inside
c. Bad contacts / 100 residues	2	0.4	1.4	10.0	-0.1	Inside
d. Zeta angle st dev	288	4.6	3.1	1.6	0.9	Inside
e. H-bond energy st dev	194	0.8	0.7	0.2	0.2	Inside
f. Overall G-factor	570	0.0	-0.2	0.3	0.6	Inside

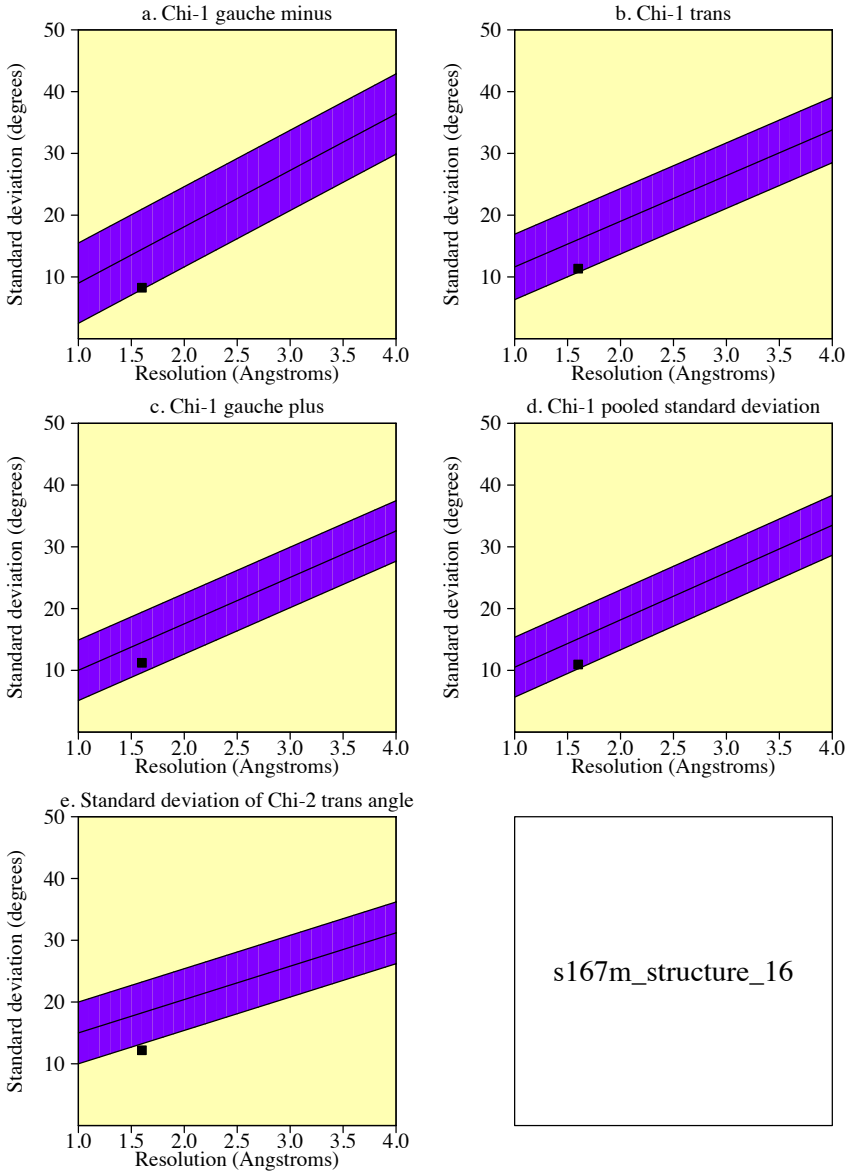
s167m_structure_16_04.ps

Data from ProCheck validating S167M structure – side chain parameters.

PROCHECK

Side-chain parameters

s167m_structure_16



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. Chi-1 gauche minus st dev	41	8.3	14.5	6.5	-1.0	Inside
b. Chi-1 trans st dev	75	11.3	16.1	5.3	-0.9	Inside
c. Chi-1 gauche plus st dev	132	11.2	14.5	4.9	-0.7	Inside
d. Chi-1 pooled st dev	248	10.9	15.1	4.8	-0.9	Inside
e. Chi-2 trans st dev	76	12.2	18.2	5.0	-1.2	BETTER

s167m_structure_16_05.ps