BLANK PAGE

Novel Synthetic Biology Applications of P450 BM3

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Science and Engineering

2019

Dominic R. Whittall

Department of Chemistry
School of Natural Sciences

Contents

Word Count: 58,299

List of Figures

List of Tables

Abbreviations

Å	Angstrom
Abs	Absorbance
ABTS	2,2'-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid)
ADME	Absorption, distribution, metabolism and excretion
AdR	Adrenodoxin reductase
Adx	Adrenodoxin
AIM	Auto induction media
AKTA	Fast protein liquid chromatography (FPLC)
B. megaterium	*Bacillus megaterium*
BCM	Barycentric mean
BM3	Cytochrome P450 BM3 (CYP102A1) from *Bacillus megaterium*
BSTFA	N,O-*bis*(trimethylsilyl)trifluoroacetamide
CO	Carbon monoxide
CPR	Cytochrome P450 reductase
CYP	Cytochrome P450
Da	Dalton
DEAE	Diethylaminoethyl
DM	P450 BM3 Double Mutant (I263P/F87V)
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DSC	Differential scanning calorimetry
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EPR	Electron paramagnetic resonance
ER	Endoplasmic reticulum
ESI	Electrospray ionisation
EtOH	Ethanol
FAD	Flavin adenine dinucleotide

FDA	Food and Drug Administration (United States of America)
FDR	Ferredoxin reductase
Fdx	Ferredoxin
FMN	Flavin mononucleotide
Gatekeeper	P450 BM3 variant containing substitutions F87V and A82F
GC	Gas chromatography
GC-MS	Gas chromatography mass spectrometry
HA	Hydroxyapatite
H_2O_2	Hydrogen peroxide
HS	High-spin
k_{cat}	Enzyme turnover rate constant
k_{cat}/K_M	Catalytic efficiency
KCl	Potassium chloride
K_2CO_3	Potassium carbonate
K_d	Dissociation constant
kDa	Kilo Dalton
K_M	Michalis constant
KPi	Potassium phosphate buffer
LC-MS	Liquid chromatography mass spectrometry
LIM	Limonene
Log *P*	Partition coefficient
LS	Low-spin
MDMA	3,4-methylenedioxy-methamphetamine
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MW	Molecular weight
m/z	Mass-to-charge ratio
NaCl	Sodium chloride
NAD (NAD^+)	Nicotinamide adenine dinucleotide
NADH	Reduced NAD

NADP (NADP$^+$)	Nicotinamide adenine dinucleotide phosphate
NADPH	Reduced NADP
Ni-IDA	Nickel iminodiacetic acid
NIST	National Institute of Standards and Technology (United States of America)
NMR	Nuclear magnetic resonance
NOS	Nitric oxide synthase
NPG	N-palmitoylglycine
OD	Optical density
OECD	The Organisation for Economic Co-operation and Development
OH	Hydroxyl group
P450	Cytochrome P450
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PdR	Putidaredoxin reductase
PdX	Putidaredoxin
PEG	Polyethylene glycol
pK_a	Acid dissociation constant
PPI	Proton pump inhibitor
PPMCC	Pearson Product-Moment Correlation Coefficient
Q-TOF	Quadrupole time-of-flight
ROH	Hydroxylated substrate
RPM	Revolutions per minute
SB	Substrate-bound
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SF	Substrate-free
TB	Terrific broth media
TM1	P450 BM3 Triple mutant 1 (I263P/F87V/V78A)
TM2	P450 BM3 Triple mutant 2 (I263P/F87V/A184K)
T_m	Melting temperature

TMCS	Trichloromethyl silane
Tris	Trisaminomethane
UV	Ultraviolet
UV-vis	UV-visible
WT	Wild-type
YASARA	Yet Another Scientific Artificial Reality Application
+AP	(+)-α-pinene
-BP	(-)-β-pinene
Δ-ALA	delta-aminolevulinic acid
4-PIM	4-phenylimidazole

Abstract

Oxidative transformation of organic substrates, particularly the insertion of a single oxygen atom across a C-H bond, is challenging to achieve with high levels of specificity using traditional methods of synthetic chemistry. These challenges can be circumvented by the use of mono-oxygenase enzymes as industrial biocatalysts. Cytochrome P450s, a superfamily of heme-containing mono-oxygenase enzymes, are capable of performing a variety of oxidation reactions, often with unparalleled levels of regio and stereoselectivity. A fatty acid hydroxylase, P450 BM3 (CYP102A1), isolated from the bacterium *Bacillus megaterium* demonstrates the highest rates of mono-oxygenase activity yet reported for a P450. The active site of BM3 has proven highly amenable to mutagenesis – single point mutations are sufficient to induce radical alterations in substrate selectivity, catalytic activity and oxidation specificity. Using a rational approach to mutagenesis, a novel point mutation situated within the I-helix structural motif caused dramatic alterations to substrate selectivity. In addition, the single variant demonstrated significant deviations from the wild-type product profile against fatty acid substrates – generating fatty acid products oxidised at novel positions. Finally, the single variant enabled the enantioselective epoxidation of the non-natural substrate styrene to the product styrene oxide. By developing the single mutant in rational, sequential manner – a collection of novel BM3 variants were developed. The variants successfully refocused hydroxylase activity with fatty acids towards further novel positions with higher degrees of regioselectivity, in addition to inverting the enantioselective epoxidation of styrene from *R*(+) to *S*(-) styrene oxide. Fatty acids oxidised at these novel, mid-chain positions serve as precursors to a variety of pharmaceutical, flavouring and industrial compounds. Moreover, biocatalysts capable of producing compounds, such as styrene oxide, with high levels of optical purity are desirable to the pharmaceutical industry. This thesis also presents the development of a high-throughput screening technique that combines a high activity BM3 variant, an FDA-approved drug library and native mass spectrometry. The BM3 variant in question, nicknamed the “gatekeeper” (F87V/A82F) is capable of producing metabolites akin to those produced by a human P450 with a plethora of drug compounds. This native-MS based screening technique identified a collection of putative substrates, including a compound capable of oxidative transformation by the gatekeeper variant that had been overlooked by a prior, alternative screening method. To the best of my knowledge, this thesis presents the first reported use of native-MS as a high-throughput compound screening technique with a BM3 enzyme.

Acknowledgements

Firstly, I would like to sincerely thank my supervisor, Professor Andrew Munro, for providing me with the opportunity to pursue a PhD. Whether discussing the direction of my project, the nature of my results or the merits of the Beatles' early material versus their later efforts, your guidance has been crucial throughout. I would also like to express my gratitude to Agilent Technologies for providing the funding for this project as part of a BBSRC iCASE studentship.

The work I have completed over the last 4 years would not have been possible without the help and support of all members of the Munro group, both past and present. In particular, I would like to thank Dr Hazel Girvan, Dr Kirsty McLean, Dr Harshwardhan Poddar and Mrs Marina Golonova for their seemingly infinite reserves of time and patience. Thanks to your collective efforts, I never had a practical question that went unanswered, I was never left without assistance and I never felt lost.

In addition, I would like to thank Dr Richard Blankley (Agilent Technologies) for his support and contributions to the body of work presented in chapter 4. I would also like to thank both Dr Colin Levy and Dr Richard Tunnicliffe for their help regarding all areas of X-ray crystallography.

Working in the Munro lab has led me to meet some of the most hardworking, decent and interesting people I have ever known. To those who I have known since the beginning (Alessia, Charlie, Laura, Mark and Sarah) to those who joined the lab during my studies (Emily, Irwin, James, Matt and Sian) each one of you has helped me along - via the pipette, the pub, or both.

I would also like to extend my heartfelt thanks to Izzy. The last four years would not have been nearly as happy without your care, support and understanding. If the last four years of my life had an alternative title to that given to this thesis, it would be "Hedonism and European Culture on a Budget".

It is said that all academic discoveries are achieved by standing on the shoulders of giants. The two giants whose shoulders on which I have principally stood are those of my parents, David and Rosemary. Above all else, this thesis is dedicated to them.

"*Don't look back in anger*"

Declaration

I declare that that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

Preface to the Journal Format Thesis

This thesis in presented in the University of Manchester journal style of PhD thesis. This allows the incorporation of data already published and/or sections that are in a format suitable for submission for publication in a peer-reviewed journal. The structure of each chapter follows that of the journal in which it is prepared for publications, meaning a separate methods chapter is not included as they are in the individual results chapters, and heading labels for each chapter may differ. The layout of the journal format thesis is as follows; an abstract, introduction, three results chapters (papers), and a summary, conclusions and future work chapter. Each chapter contains self-contained references. The general formatting of these papers has been kept consistent throughout this thesis, along with the page numbering. The journal format thesis was introduced in the University of Manchester to help students develop their skills in writing papers for peer-reviewed submission and to be overall more relevant to scientific research. Successful scientific publications are often collaborative and therefore as part of the journal format the contributions of each co-author are stated below.

Papers included as results chapters:

Chapter 2

Whittall, D. R., Girvan, H., Poddar, H., McLean, K., Golovanova, M., Levy, C., Leys, D., and Munro, A.W. (2019). Profound Alteration of P450 BM3 Regio- And Enantio-Selectivity via a Single Point-Mutation. This paper is currently unpublished but it is hoped that it will be published in The Journal of Biological Chemistry.

Chapter 3

Whittall, D. R., Girvan, H., McLean, K., Tunnicliffe, R., Golovanova, M., Levy, C., and Munro, A.W. (2019). Refocusing the Regio- And Inverting the Enantio-Selectivity of P450 BM3: Refashioning the Activity of CYP102A1 towards Fatty Acid Hydroxylation and Styrene Epoxidation. This paper is currently unpublished but it is hoped that it will be published in The Journal of Biotechnology.

Chapter 4

Whittall, D. R., Jeffreys, L. N., Selvam, I., Blankley, R., and Munro, A.W. (2019). Application of Native-MS Mediated Compound Screening To a Highly Promiscuous Variant P450. This paper is currently unpublished but it is hoped that it will be published in The Journal of the American Chemical Society.

Author Contributions

As PhD supervisor, Professor Andrew Munro contributed to the manuscript preparation for all papers.

Chapter 2

Dr Colin Levy collected X-ray diffraction data from variant BM3 crystals at the Diamond synchrotron.

Dr David Leys assisted with interpretation of the resulting structural data.

Dr Harshwardhan Poddar assisted with the refinement of crystal structures and preparation of the corresponding figures.

Dr Hazel Girvan collected EPR data.

Dr Kirsty McLean assisted with data interpretation of Resonance Raman and steady state kinetics.

Mrs Marina Golonova assisted with protein purification and use of the AKTA equipment.

Chapter 3

Dr Colin Levy collected X-ray diffraction data from variant BM3 crystals at the Diamond synchrotron.

Dr Richard Tunnicliffe assisted with the refinement of crystal structures and preparation of the corresponding figures.

Dr Hazel Girvan collected EPR data.

Dr Kirsty McLean assisted with data interpretation of Resonance Raman and steady state kinetics.

Mrs Marina Golonova assisted with protein purification and use of the AKTA equipment.

Chapter 4

Dr Laura Jeffreys provided the purified “gatekeeper” protein used throughout the chapter, in addition to performing the initial compound screen via UV-visible spectroscopy.

Dr Richard Blankley assisted with native-MS method development and product analysis.

B.Sc Irwin Selvam assisted in sample preparation and native-MS method development.

Chapter 1: General Introduction

1.1. Cytochromes P450

1.1.1. A Brief Introduction to Cytochromes P450

Cytochrome P450s (known as P450s or CYPs) are a superfamily of heme containing mono-oxygenase enzymes, first reported by Klingenberg and Garfinkel in 1958 (Klingenberg, 1958; Garfinkel, 1958). Since their discovery, P450s have been found to be ubiquitous in all forms of life from simple bacteria, through to archaea, viruses and complex eukarya. The title P450 arises from the absorption maximum (at ~450 nm) produced when the central heme iron (in the ferrous state) is saturated with CO (Figure 1.1) (Omura and Sato, 1962; Estabrook, 2003; Munro et al., 2007).

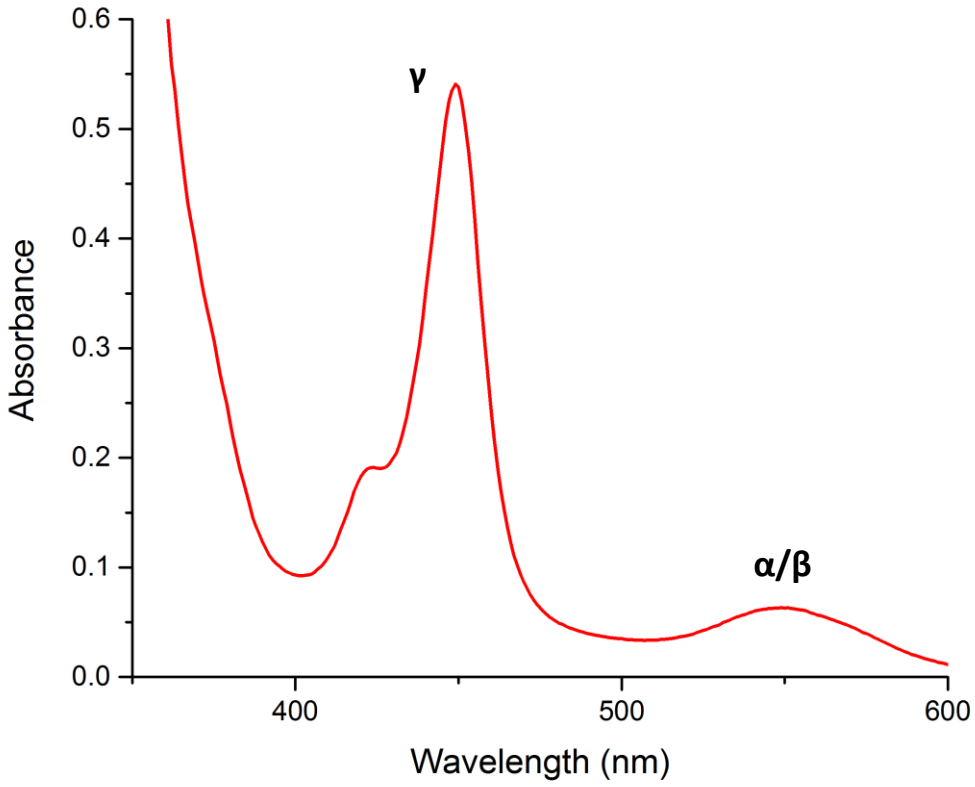

Figure 1.1 Carbon-monoxide binding spectra of a wild-type P450 BM3 (CYP102A1) in the ferrous state.

Intact P450 BM3 protein was prepared to a concentration of ~6 μM in the assay buffer. Following reduction to the ferrous state via the addition of sodium dithionite, CO gas (~1 cm^3) was gently added into the cuvette using a syringe. The α- and β- bands occur within the 500 - 600 nm region. The γ-band, or Soret maximum, is visible at ~450 nm.

P450s perform numerous roles in the oxidative transformation of organic substrates; notable functions include xenobiotic detoxification and steroid biosynthesis (Rendic et al., 2015). As mono-oxygenase enzymes, P450s incorporate a single oxygen atom at a specific position on a given substrate via the reductive activation and cleavage of molecular oxygen. This commonly results in the substrate becoming hydroxylated, along with the production of a water molecule through reduction of the remaining oxygen atom (Munro et al., 2007). Over 95% of the enzyme-mediated oxidations reported in the literature are performed by P450s (Rendic et al., 2015). Generally, the pyridine nucleotide cofactors NADPH and NADH serve as electron donors for reactions catalysed by P450s, although other sources have been reported (Schwaneberg et al., 2000). Electrons pass via redox partner proteins prior to being transferred to the heme prosthetic group. The arrangement of these redox partners differs between prokaryotic and eukaryotic P450s. The majority of bacterial P450s employ a ferredoxin reductase/ferredoxin system - a system also used by mitochondrial P450s. In contrast, eukaryotic P450s use a membrane-bound, flavin adenine dinicleotide (FAD) and flavin adenine mononucleotide (FMN) binding NADPH-cytochrome P450 reductase (CPR). Other redox systems exist and serve as means of P450 redox partner classification (Munro et al., 2007). After nearly 6 decades of intensive research into the structure and functioning of P450s, interest has centred on the wealth of potential biotechnological applications for these multifaceted biocatalysts. Rational design by site directed mutagenesis and forced evolution have focused on generating mutant P450s which insert oxygen at alternative positions on the substrate, alongside mutants with altered substrate selectivity (Di Nardo and Gilardi, 2012). This thesis is centred on the study of a unique P450-CPR fusion system with the highest reported rate of electron transfer of all P450s – cytochrome P450 BM3.

1.1.2. History of P450 Research

The genesis of cytochrome P450s research can be said to have occurred in the 1940s, with studies into drug and steroid metabolism conducted by Williams, Mueller and Miller (Williams 1947; Mueller and Miller, 1949). The earliest examined P450 enzymes were initially reported as carbon monoxide-binding pigments by Klingenberg and Garfinkel (Klingenburg, 1958; Garfinkel, 1958) and discovered as a consequence of research into mammalian liver microsomes. These pigments were then named as P450s by Omura and Sato (Omura and Sato, 1964) on account of the absorption maxima produced at ~450 nm following saturation with carbon monoxide, an assay that remains in use to this day. In 1968, research conducted by Gunsalus on a camphor oxidation system from the bacterium

Pseudomonas putida led to the discovery of the now model P450 system, P450cam (Katagiri et al., 1968; Guengerich, 2019). Early studies on mammalian P450 enzymes faced an additional layer of difficulty on account of their membrane-bound state and their (predominant) localisation within the endoplasmic reticulum (Guengerich, 2019). Nevertheless, Lu and Coon demonstrated successful separation of the phospholipid, P450 and reductase components from rabbit liver microsomes through the use of detergents in 1968 (Lu and Coon, 1968). The primary physiological functions of P450s were discovered to be the hydroxylation of steroids and the oxidation of xenobiotics by Estabrook and Cooper, respectively (Estabrook et al., 1963; Cooper et al., 1965).

The following decade saw improvements in purification techniques in addition to early characterisation of the reductase domain (Guengerich, 2019). Moreover, 1974 saw the first report of a "cell-free hydroxylating system" isolated from the bacterium *Bacillus megaterium* by Miura and Fulco (Miura and Fulco, 1974). The third P450 enzyme to be isolated from *B. megaterium* later came to be known as P450 BM3, the rational redesign of which is the focus of this thesis (Fulco, 1991). During the 1980s, the first crystal structure for a P450, that of P450cam, was solved to a resolution of 2.6 Å by Poulos (Poulos et al, 1985). As of June 2019, over 900 crystal structures relating to P450 enzymes have been deposited in the Protein Data Bank. These structures represent over 140 different P450s (Guengerich, 2018). Developments in recombinant DNA technology enabled understanding of the evolutionary relationships between members of the P450 gene superfamily to widen in scope and complexity (Nebert et al., 1989). As of June 2019, 57 P450 genes have been identified in humans alone (Guengerich, 2019). In stark contrast, 1476 P450-encoding genes have been identified in wheat (Nelson, 2009). The total number of P450 sequences known currently numbers over 18,000, with the functions of many yet to be elucidated (Guengerich and Cheng, 2011). The second P450 crystal structure to be solved, that of the P450 BM3 heme domain, was obtained in 1993 (PDB code 2HPD; Ravichandran et al., 1993).

Into the 1990s and 2000s, advances in both genetic and structural characterisation technologies paved the way for a proliferation of site-directed mutagenesis studies. The majority of these studies proceeded with the aim of engineering variant bacterial P450 enzymes for the production of valuable drug metabolites, analogous to those produced by human P450s (Fasan, 2012). This approach has been met with some success, with variants of the bacterial P450s CYP505X (*Aspergillus* sp.) and CYP102A7 (*Bacillus* sp.) having been

shown to produce human metabolites of diclofenac and chlorzoxazone, respectively (Fasan, 2012; Weis et al., 2009). The structure of the first membrane bound P450, that of the human CYP3A4, was solved by Yano and colleagues in 2004 (Yano et al., 2004). This structure served to confirm that both prokaryotic and eukaryotic P450s share a similar overall protein structure, including several highly conserved secondary motifs (Yano et al., 2004; Lamb and Waterman, 2013).

During the 2000s, research began into targeting the inhibition of bacterial P450s implicated in diseases such as tuberculosis (Leys et al., 2003). The discovery and development of drugs capable of inhibiting P450s may prove essential in combatting increasingly drug-resistant strains of pathogenic bacteria and fungi (Leys et al., 2003; Hargrove et al., 2017). In humans, targeted inhibition of P450 19A1, a steroid aromatase, is now being applied in the treatment of breast cancer (McNulty et al., 2012). In 2010, the previously elusive P450 catalytic intermediate, compound I, was captured and characterised by Rittle and Green (Rittle and Green, 2010; Guengerich, 2013b).

P450 research trends emerging within the current decade (2010s) include the identification of P450s with novel functionality, the engineering of well characterised P450 systems (such as CYP101A1 and CYP102A1) towards the production of specialty chemicals, and the optimisation of P450s for use in whole-cell biocatalysis (Urlacher and Girhard, 2012 and 2019). The manipulation of wild-type P450s via rational design and directed evolution approaches have yielded variant P450 enzymes capable of performing hydroxylation of non-natural substrates with high degrees of regio- and stereospecificity (Behrendorff et al., 2015). Furthermore, the development of P450s capable of performing non-natural reactions – notably reactions utilising carbene, nitrene and oxene species has progressed significantly in recent years (Coelho et al., 2013; Wei et al., 2018). A body of work relating to the development, via directed evolution, of variant P450s capable of performing novel chemistry was recognised by the awarding of the 2018 Nobel Prize in Chemistry to Frances Arnold (Cirino and Arnold, 2003; Coelho et al., 2013; Arnold, 2018).

P450s have become firmly established as indispensable tools of biotechnology. As of June 2019, a search of the Web of Science Core Collection for the term “P450” returns over 63,000 records, with the majority of the results being contained within the fields of pharmacology, biochemistry and toxicology. The continuing study of both the functions and applications of P450s will surely provide further breakthroughs in the fields of medicine, drug development and synthetic biology (Guengerich et al., 2005; Guengerich, 2013b).

1.1.3. Classification of P450s

P450s can be broadly classified into two major groups. Class I P450s are defined by a 3 component redox system consisting of a soluble P450, an iron-sulphur protein and a flavin-containing reductase protein (Oritz de Montellano, 2005). Class I P450s are associated with bacterial systems and this group is also referred to as "Class B" (Degtyarenko and Archakov, 1993). P450s within this group tend to favour NADH (over NADPH) as an electron donor. Class I P450s are significantly better characterised than their Class II counterparts on account of their solubility – a consequence of their cystolic localisation (De Mot and Parret, 2002). P450s that are part of the second group, known as class II, feature a 2 component redox system and are mainly associated with eukaryotic organisms. Both the P450 and the NADPH-dependent reductase characteristic of this system are embedded within cell membranes (due to a chain of hydrophobic residues at the N-terminus forming a membrane "anchor" region) (Munro et al., 2007). Owing to their membrane-bound location, class II P450s are usually insoluble in aqueous buffers. On account of this, the characterisation of class II P450s is often less developed in comparison to that of class I P450s. Most structural and mechanistic understanding of class II systems therefore comes from studies on model P450 systems, notably P450 BM3 (De Mot and Parret, 2002). A notable outlier of the class I/II system, P450 BM3 serves as an experimental model on account of its possessing a soluble, Class I-like P450 coupled to a soluble reductase protein (characteristic of a class II system, but missing the membrane anchor region) on a single polypeptide (Figure 1.2) (Munro et al., 2007). This group of catalytically self-sufficient P450 enzymes are referred to as being members of P450 class III (Ortiz de Montellano, 2005).

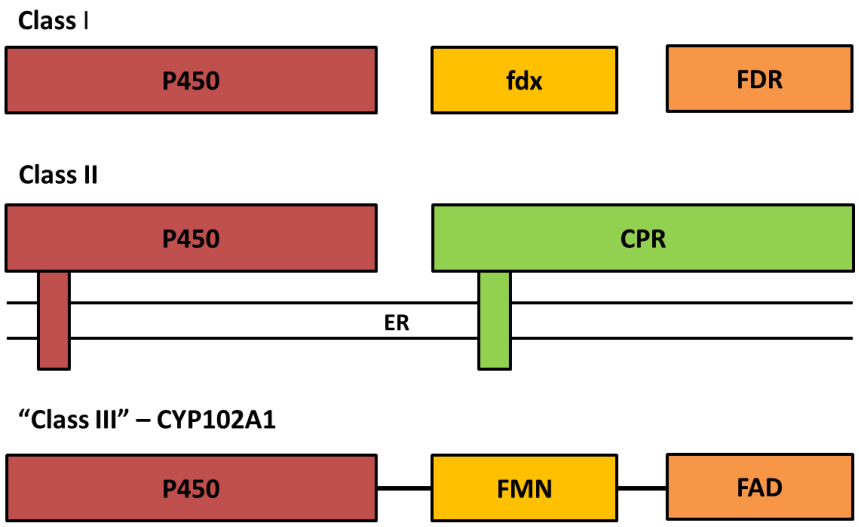

Figure 1.2 The three major classes of P450 enzymes and the arrangements of their redox partner proteins.
The heme-containing P450 domains are represented in maroon. The class I ferredoxin (fdx) and ferredoxin reductase (FDR) domains are represented in yellow and orange, respectively. The membrane bound P450 reductase domain (CPR) of the class II system is represented in green. The flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) domains are represented in yellow and orange, respectively. The domains of the class III system are linked to illustrate their natural fusion arrangement as part of a single polypeptide. Figure adapted from Lamb & Waterman, 2013.

P450 encoding genes are classified into families. Regarding formal P450 (CYP) identification nomenclature, the first number and letter indicate the family (40% sequence similarity) and subfamily (55% sequence similarity), respectively. Therefore, in the case of P450 BM3, CYP102A1 denotes family 102, subfamily A. The second number indicates genetic variants – i.e. variant 1, variant 2, etc. A (Nelson et al., 1996).

1.1.4. Hemoproteins and P450 Heme Structure

Heme proteins (hemoproteins) are a diverse group of proteins that possess heme as a prosthetic group. Hemoproteins have a variety of roles throughout numerous biological systems. Most notably, hemoproteins are involved in the transport and storage of diatomic gases and in electron transfer, as demonstrated by hemoglobin and cytochrome *c*, respectively (Li et al., 2011). In addition, hemoproteins have also been implicated in the processing of microRNA (Faller et al., 2007) and regulation of transcription (Sun et al., 2002). The heme prosthetic group can appear in several different forms, the most biologically abundant of which are presented in Figure 1.3.

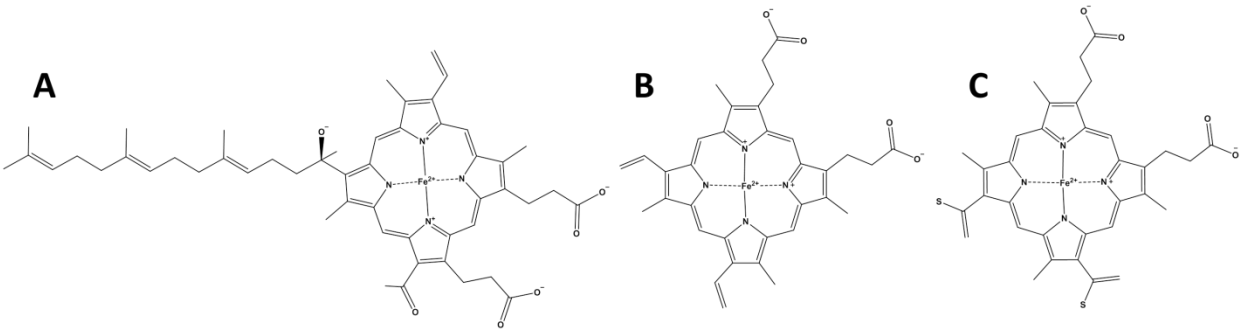

Figure 1.3 **Structures of the three most predominant heme prosthetic groups.** Inset A: Heme *a*, the least common of the three heme groups presented and containing both a farnesyl side chain and a formyl group; Inset B: Heme *b*, the most abundant heme group of those presented and the heme cofactor of P450s; Inset C: Heme *c*, a cofactor that covalently binds to the surrounding protein via formation of an ester linkage to a glutamate residue conserved in various CYP4 enzymes.

Of these presented heme structures, heme *b* is the most common group affiliated with proteins, comprising 53.5% of the heme prosthetic groups in the Heme Protein Database (Reed, et al., 2008). Heme *b*, also referred to as protoporphyrin IX, is the prosthetic group found naturally in the cytochromes P450, in addition to other oxygenase and oxidase enzymes (Gibney et al., 2000). Heme *b* associates with the surrounding protein structure (the apoprotein) via a noncovalent interaction between the heme iron and an amino acid side chain of the apoprotein (Li et al., 2011). Heme *c* is the second most commonly found heme cofactor, comprising 40.9% of the heme prosthetic groups in the Heme Protein Database (Reedy et al., 2008). Unlike heme *b*, heme *c* binds covalently to the apoprotein structure. This occurs via thioether bond formation between one or both of the sulfur atoms from the two cysteine residues of the apoprotein and the vinyl groups of the heme (Li et al., 2011). This covalent linkage enables heme *c* to span a larger range of reduction potentials than does heme *b* (Bowman et al., 2008). Consequently, proteins containing heme *c* (cytochromes c) function as excellent electron carriers (Bowman et al., 2008). The third most abundant heme type, Heme *a* (13.6% of the heme prosthetic groups in the Heme Protein Database) displays notable differences in the peripheral structure of the porphyrin ring in comparison to the heme *b* and *c* groups (Reedy et al., 2008). Heme *a* contains both a farnesyl side chain and a formyl group, neither of which exist in heme types *b* and *c* (Gibney et al., 2000). Heme *a* is the cofactor of Complex IV, the terminal complex of the mitochondrial respiratory chain, responsible for the transfer of electrons from reduced cytochrome *c* to oxygen (Valnot et al., 2000).

In the P450s, the heme iron is anchored to the surrounding protein structure through linkage by a proximally-coordinated thiolate bond to an absolutely conserved cysteine residue (in its cysteinate state), as shown in Figure 1.4. The electron donating character of this thiolate ligand is thought to underpin the ability of the P450s to reduce and activate the distally-bound molecular oxygen to facilitate oxidative catalysis (Munro et al., 1996). In the resting (ferric) state, a water ligand is distally coordinated to the heme at the 6th axial position (Ortiz de Montellano, 2005). This water ligand is displaced from the ferric heme iron upon the entry of substrate to the active site (Figure 1.4). This, in turn, results in reorganisation of the d-orbital electrons of the heme iron, producing a high-spin (HS) ferric state with a more positive reduction potential. This facilitates reduction by the redox partner and the first of two electron transfers from the CPR onto the heme iron. The reduction of the heme iron from the ferric to the ferrous state enables the binding of molecular oxygen at the distal face, triggering the P450 catalytic cycle (Munro et al., 2007). The gaseous inhibitors CO and NO can inhibit the P450 by binding to the distal position on the ferrous heme in lieu of molecular oxygen (Ortiz de Montellano, 2005).

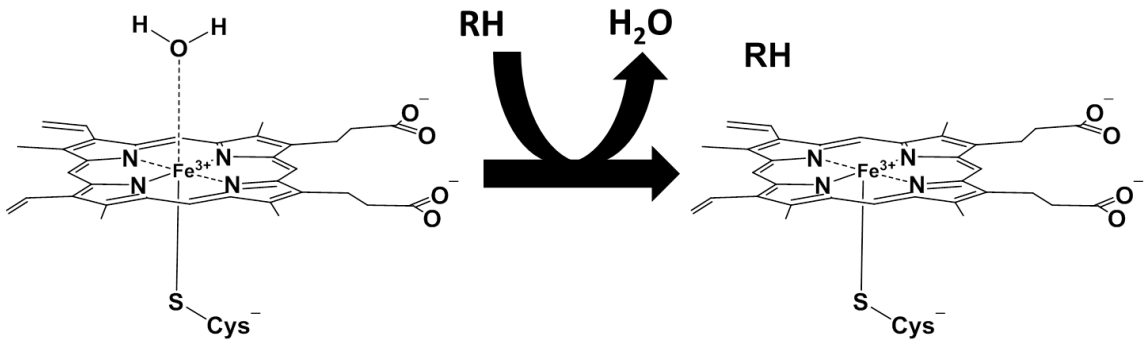

Figure 1.4 Displacement of the axially-ligated water molecule from the P450 heme domain following substrate binding.
LHS: Substrate-free, ferric heme in the 6-coordinate low-spin state, ligated to a water molecule at the 6th axial heme position. RHS: Substrate-bound, ferric heme in the 5-coordinate high-spin state, in which the water molecule has been displaced from the heme iron upon substrate (RH) binding. The proximal thiolate bond, which tethers the heme *b* cofactor to the apoprotein, is visible in both the low and high spin states.

1.1.5. Spin-State and Spectroscopic Properties of P450 Heme Groups

The electronic configurations of the heme iron 3d-orbital electrons in both the HS and LS states are displayed in Figure 1.5. Owing to the similar protein topology throughout the superfamily, the spectroscopic properties of P450s remain broadly constant (Luthra et al., 2011). In the resting state, the heme iron of the P450 exists in the ferric (Fe^{+3}) oxidation

state and is coordinated to a molecule of water at the distal face. In this oxidation state, the heme iron can exist in one of two spin states, depending mainly on the presence or absence of substrate. In the absence of substrate, the heme iron exists in the low-spin state (LS, S=1/2) and displays a heme Soret absorbance maximum at ~418 nm (black line, Figure 1.6). In the unbound, low-spin state, only a single d-orbital electron remains unpaired. The binding of substrate and displacement of the water ligand alters the splitting of 3d-orbitals between the lower energy t_{2g} and higher energy e_g sets and causes the ferric heme iron spin-state equilibrium to shift towards the high-spin state (HS, S=5/2), in which all 5 of the 3d-orbital electrons are unpaired (Shannon, 1976).

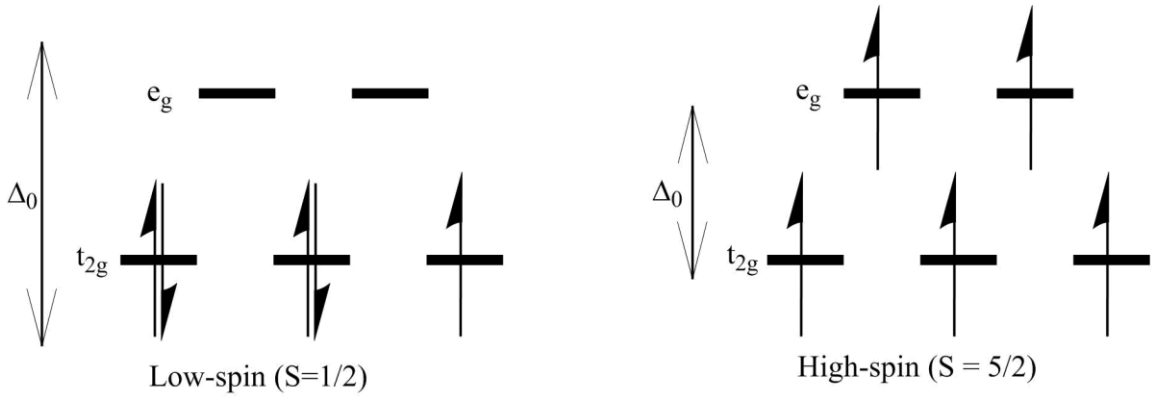

Figure 1.5 Configuration of d-orbital electrons of ferric P450 heme iron in the low-spin (LHS) and high-spin (RHS) states.
The octahedral splitting energy is indicated by Δ_o. Lower energy and higher energy sets are indicated by t_{2g} and e_g, respectively. In the low-spin state, all electrons occupy the t_{2g} set and S = ½. In the high-spin state, 2 electrons occupy the e_g set, 3 electrons occupy the t_{2g} set and S = 5/2.

High-spin heme iron can be characterised by a Soret absorbance maximum of ~390 nm (red line, Figure 1.6). On account of this Soret peak shift, transitions between substrate free (LS, ~418 nm) and substrate bound (HS, ~390 nm) heme iron can be characterised by way of UV-Visible spectroscopy (Schenkman and Jansson, 2006). In addition to the ~418 nm Soret peak, a low-spin P450 absorbance spectrum (Figure 1.5) features a standard protein peak at ~280 nm, a δ- band at ~360 nm and α- and β- bands in the 500 - 600 nm region. Tight binding inhibitor compounds, such as 4-phenylimidazole (4-PIM) (blue line, Figure 1.6) can displace the axial water molecule, consequently stabilising the low-spin state and causing a Soret peak to appear at ~424 nm.

P450 UV-Vis spectra can be classified into 3 groups, Type 1, Type 2 and Reverse Type 1 (Segall, 1997). Type 1 spectra are characterised by a minimum at ~420 nm and a maximum

at ~390 nm, are used to estimate the proportion of heme in LS and HS states, respectively. Structural characterisation using X-Ray crystallography has attributed this shift to the displacement of the distally bound water molecule from the heme iron (Poulos et al., 1987). Transitions from LS to HS displayed by type 1 spectra are also used as an initial test to check the ability of a P450 to bind to a given substrate. The isosbestic point for type 1 spectra occurs at ~407 nm (Poulos et al., 1987).

Type 2 spectra are evidenced by a maximum peak within the region of 425 – 440 nm, with a corresponding minimum peak at 415 nm (or lower) with an isosbestic point at ~419 nm (Schenkman, 1981). Type 2 spectra indicate the formation of a hexacoordinate complex with high affinity ligands, typically inhibitors (e.g. azole drugs and other molecules that bind directly to the heme iron), which shift the Soret band towards the infra-red end of the spectrum (Luthra et al., 2011).

Reverse type 1 spectra are characterised by the presence of a Soret band at ~420 nm and a decreased absorbance at ~390 nm, a “mirror image” of the peaks found in a type 1 spectrum. The occurrence of such a spectrum might suggest the displacement of an endogenous substrate (or substrate analogue) on titration with another molecule, as demonstrated by Sandhu et al. (1994).

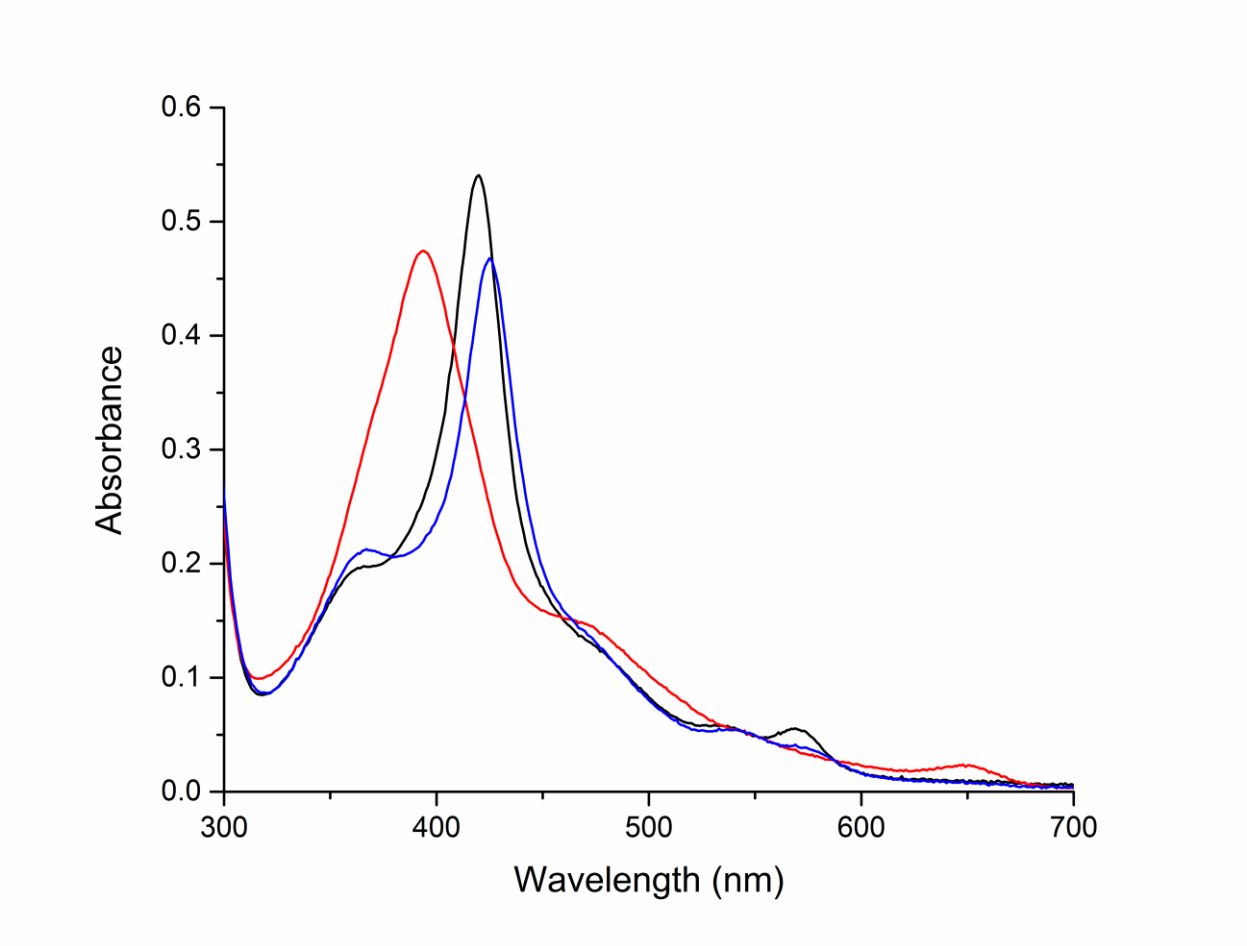

Figure 1.6 **UV-Vis spectra of ~6 µM wild-type P450 BM in the substrate-free (black line), N-palmitoylglycine-bound (red line) and 4-phenylimidazole-bound (blue line) states.**
In the presence of the substrate NPG, the Soret band shifts from ~420 to ~390 nm (type I). In the presence of the inhibitor 4-PIM, the substrate-free Soret band shifts from ~420 to ~425 nm (type II).

1.1.6. Structure of P450s

Identification of P450s throughout all domains of life may be determined by the recognition of two signature motifs, EXXR and CXG (Syed and Mashele, 2014). The EXXR motif is contained within the K-helix, and acts to stabilise the core structure of the protein (Graham and Peterson, 1999). The absolutely conserved CXG sequence is invariably located within the β-bulge or Cys-pocket region and contains the Cys residue that coordinates to the proximal face of the heme iron through a thiolate bond (Syed and Mashele, 2014; Hamdane et al., 2008; Graham and Peterson, 1999). Between distinct members of the P450 superfamily, amino acid sequence identities can be as little as 16%. Among members of each P450 family, protein sequence identity should be 40% or greater. When P450 amino

acid identity is greater than 50% the P450s are classed as members of the same P450 subfamily and will potentially act on the same substrates (Nelson, 1999; Syed and Mashele, 2014). Despite possessing less than 20% sequence identity across the superfamily, P450 enzymes share a common structure and topology (Denisov et al., 2005). P450 structural organisation (the "P450 fold" was first identified by Tom Poulos (Poulos et al., 1987) and assigned to the structure of the camphor hydroxylase P450cam.

The overall P450 structure itself can be likened to a trigonal prism, with the *b*-type heme prosthetic group situated between two major domains named, according to their prevailing secondary structural elements, alpha and beta. The alpha and beta domains account for 92% of the overall structure (Ravichandran et al., 1993). The alpha domain consists of a highly conserved four-helix bundle of 3 parallel helices known as D, L and I (Figure 1.7) (Mestres, 2005; Ravichandran et al., 1993). An antiparallel helix, the E-helix, completes this structure. The heme prosthetic group, situated at the centre of the P450 between the distal I- and the proximal L-helices, is also bound through its iron to an absolutely conserved cysteine residue at its proximal face. This cys ligand, situated in a loop region adjacent to the L-helix, attaches the heme to the protein structure via a thiolate bond (Munro et al., 2002). Within this loop region resides the signature amino acid sequence FxxGx(H/R)xCxG, which contains the proximal cysteine that coordinates the heme iron. Of these residues, only the cysteine remains absolutely conserved (Denisov et al., 2005). Notably, structural flexibility helps facilitate high levels of regio- and stereo-selectivity demonstrated by some P450s, as well as enabling a variety of different substrate sizes to bind to diverse P450s (Denisov et al., 2005; Mestres, 2005). The beta sheets, of which two are structurally conserved, help form the hydrophobic substrate access channel (Graham and Peterson, 1999).

The I-helix contains a highly conserved and catalytically essential acid-alcohol amino acid pair (e.g. Asp251-Thr252 in P450cam). The threonine residue (Thr268 in P450 BM3) couples proton transfer to the distal oxygen atom of the P450 in its ferrous-oxy state, forming a highly reactive ferric-hydroperoxy form that is rapidly converted to Compound I (a ferryl-oxo porphyrin radical cation species) which then catalyses substrate oxidation (Yeom et al., 1995; Munro et al., 2002). Additionally, the I-helix plays a key role in regulating access to the heme prosthetic group, and undergoes significant structural rearrangement following substrate binding.

The less structurally conserved regions, including the A-, B-, B'-, F- and G-helices and their associated loops, are associated with substrate binding and specificity (Mestres, 2005; Graham and Peterson, 1999). The F- and G-helices comprise part of the substrate binding channel, and together with the B'-helix, provide the canopy of the binding pocket (Graham and Peterson, 1999). Alignments between a number of P450 crystal structures have highlighted the flexibility of the F- and G-loop region between families (Hasemann et al., 1995). In a similar manner to the I-helix, both the F- and G-helices undergo significant structural realignment following substrate binding (Whitehouse et al., 2012).

While not directly exposed to the solvent, the heme active site is accessible via a long, hydrophobic channel lined with predominantly non-aromatic residues (Whitehouse et al., 2012). A number of hydrophobic residues located at the mouth of the substrate binding pocket were found to have key roles in substrate docking and selectivity. Key hydrophobic channel residues specific to cytochrome P450 BM3 are discussed later in this chapter.

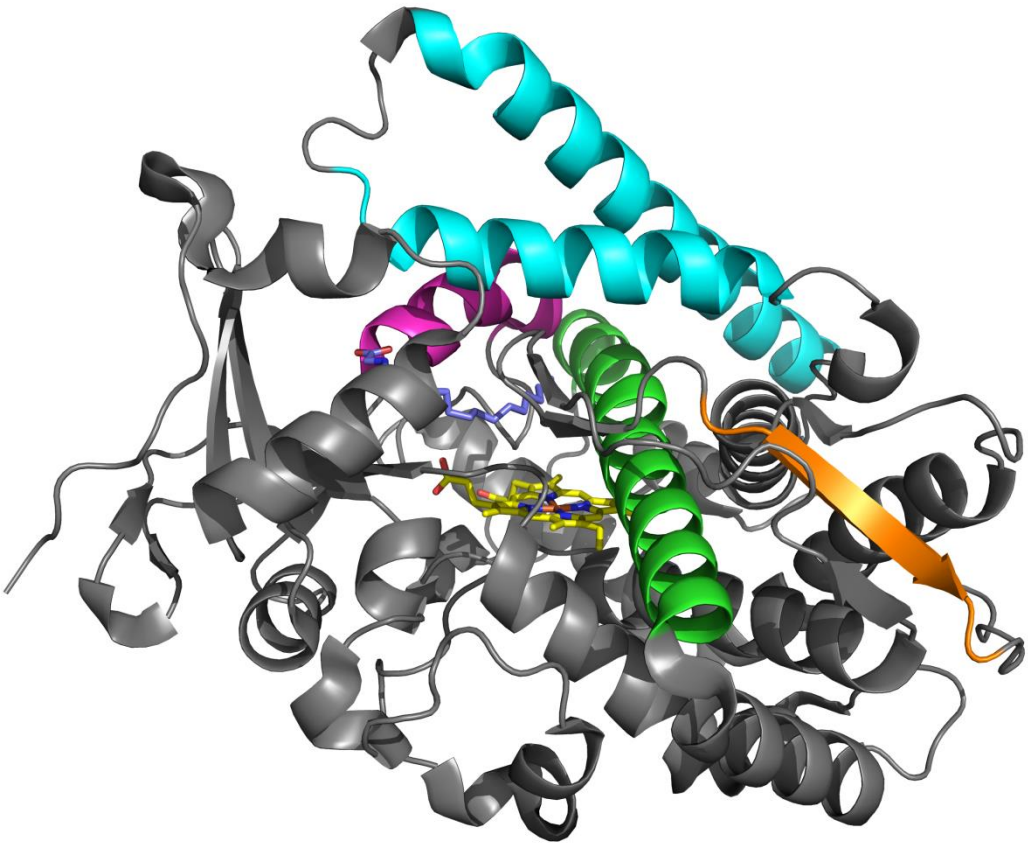

Figure 1.7 General structure and fold topography of a P450 heme domain. The heme domain of P450 BM3 (CYP102A1) is shown in the substrate (NPG)-bound state (PDB: 1JPZ) (Haines et al., 2001). Key structural motifs are highlighted: I-helix (green), F-G helices (cyan), B'-helix (pink), C-terminus (orange), heme prosthetic group (yellow). The substrate NPG is highlighted in purple. Helices were assigned according to data presented in Whitehouse et al. (2012) and using nomenclature described in Poulos et al. (1987).

1.1.7. Cytochrome P450 Reductase and Redox Partner Proteins

As previously detailed, the relationship between P450s and their redox systems differ throughout the superfamily. The vast majority of class I P450s (found in bacteria and in eukaryotic mitochondria) use two cofactor-containing proteins, typically an NAD(P)H-dependent FAD-binding oxidoreductase and one of either a 2Fe-2S cluster-binding ferredoxin or an FMN-binding (flavodoxin) domain. The most widely studied class I redox partner proteins are the 2Fe-2S cluster-containing putidaredoxin (Pdx) and its corresponding FAD-binding putidaredoxin reductase (PdR), which transfer electrons to P450cam (CYP101A1) (Schiffer and Bernhardt, 2003). PdR accepts 2 electrons from the

cofactor NADH, reducing the FAD to a hydroquinone state, before shuttling the first electron to Pdx and, in doing so, forming a FAD semiquinone (Sevrioukova et al., 2001). Pdx transfers this initial electron to the substrate-bound P450cam, reducing the heme iron to its ferrous form and enabling the binding of dioxygen (O_2) to the heme iron. This forms a ferrous-oxy state and a further electron transfer to the heme iron (from the remaining electron on the PdR FAD via the Pdx) produces the ferric-peroxy state. This species is protonated through proton transfer from Thr252 to form the ferric-hydroperoxy (Compound 0) state. Compound 0 is then further protonated, leading to the heterolytic cleavage of the O-O bond in Compound 0 and the release of a molecule of water. The highly-reactive ferryl-oxo porphyrin radical cation species (Compound I) is then formed and this species catalyses the insertion of an oxygen atom into the bound substrate of P450cam – leading to hydroxylation of the camphor (Estabrook et al., 1971). The 2Fe-2S cluster in Pdx thus serves to mediate the stepwise transfer of single electrons between the ~46 kDa PdR and P450cam. All components of this redox system are soluble, and have consequently been extensively characterised (Ortiz de Montellano, 2005).

Mitochondrial P450 redox systems, such as those found in CYP11A1, display considerable similarities to bacterial, class I systems (Schiffler and Bernhardt, 2003). Both systems contain 2Fe-2S ferredoxins, in addition to a FAD-containing reductase. However, both the P450 and adrenodoxin reductase (AdR) domains of mitochondrial systems remain membrane bound, while the 2Fe-2S containing ferredoxin, adrenodoxin (Adx), remains soluble (Schiffler and Bernhardt, 2003). Structures of the FAD and FMN cofactors and the 2Fe-2S ferredoxin cluster are displayed in Figure 1.8.

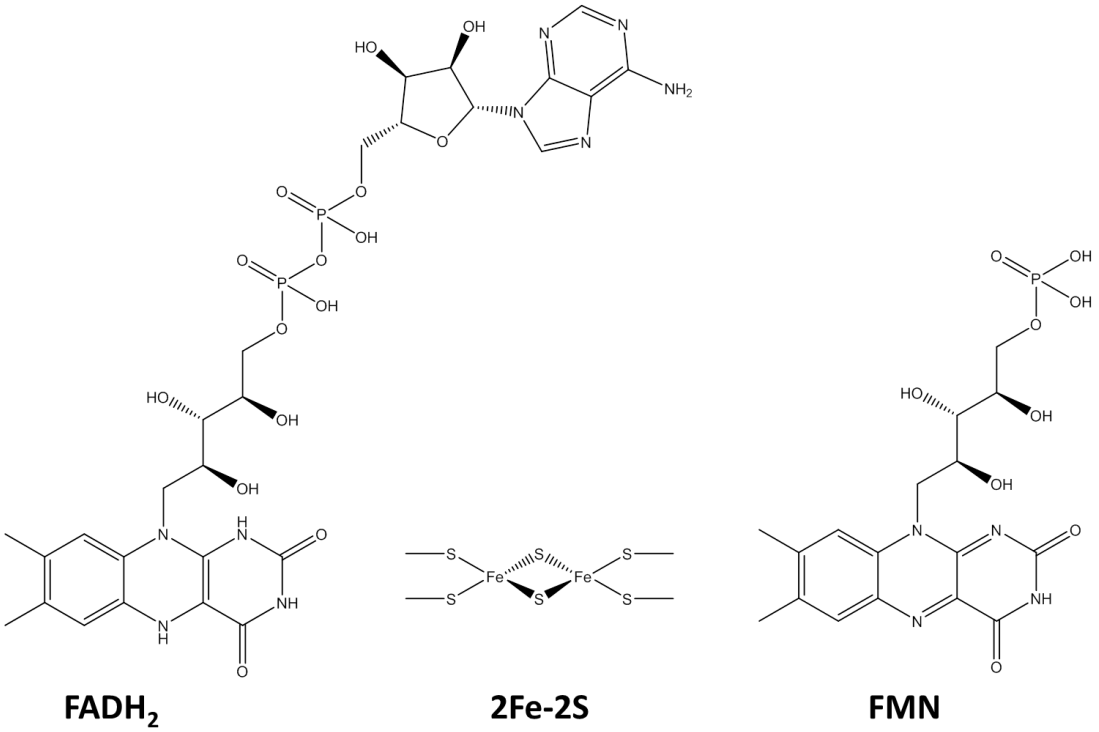

Figure 1.8 Structures of the cofactors associated with P450 reductase domains.

LHS: Flavin adenine dinucleotide (FAD); Centre: 2Fe-2S cluster; RHS: flavin mononucleotide (FMN). Other cofactors (e.g. 3Fe-4S) may also be used to reduce P450s

In contrast, class II P450s utilise a diflavin redox system – i.e. one that includes a molecule of FMN in addition to a molecule of FAD. Moreover, both redox cofactors are non-covalently attached to the redox protein in a single polypeptide bound to the endoplasmic reticulum (ER) of a eukaryotic cell, referred to as cytochrome P450 reductase (CPR). The majority of diflavin redox systems utilise NADPH as an electron donor, which binds to the FAD domain of the CPR. A linker region connects the FMN and FAD domains. This enables the close alignment of the two isoalloxazine rings of each flavin (at a distance of ~4 Å), enabling efficient electron transfer (Wang et al., 1997). In summary, the structure of a class II CPR consists of three distinct domains. From the N- to the C-terminus, these are the FMN binding domain, the linker region, and the FAD/NADPH-binding domain (Wang et al., 1997). It is the hydrophobic N-terminal domain that anchors the CPR to the ER membrane. As with class I redox partner systems, electrons are shuttled from the NADPH to the P450 (one at a time) via the components of the reductase. In mammals, only one other enzyme also contains both FAD and FMN cofactors, this being nitric-oxide synthase (Wang et al., 1997).

Sequence homology for the CPR domains between eukaryotic species is high, highlighting the evolutionary significance of the enzyme (Shen and Kasper, 1993).

P450 redox systems in class III also utilise a diflavin reductase, albeit fused as a single, soluble polypeptide, together with the P450 domain. The most notable example of this class is P450 BM3 (CYP102A1). Class III P450s are predominantly bacterial, with CYP102A1 homologues CYP102A2 and CYP102A3 being found in *Bacillus subtilis.* However, a eukaryotic analogue, CYP505A1, exists in the fungus *Fusarium oxysporum*, and is also known by the title P450 foxy (Nakayama et al., 1996).

Levels of catalytic turnover for both the class I and class II redox systems are often dependent on the rate of the first electron transfer (Lewis and Hlavica 2000). Levels of turnover in class III redox systems are high relative to classes I and II, a fact attributable to the highly efficient electron transfer facilitated by the fusion arrangement between the P450, FMN and FAD domains in enzymes such as P450 BM3 (Munro et al, 2002).

1.1.8. Catalytic Cycle of P450 Enzymes

In order for catalysis to occur, single electrons are required to be introduced at two separate points of the catalytic cycle. Both of these electrons are (commonly) supplied by the pyridine nucleotide cofactors NADH or NADPH. One at a time, each electron is transferred onto the heme through a reductase protein system (e.g. CPR). The first electron reduces the ferric iron atom to the ferrous state (Fe^{2+}), enabling the heme to bind molecular oxygen. The corresponding redox equation is displayed below (Equation 1.1) (Guengerich, 2018). The second electron is introduced at a later point in the cycle, leading to the production of a transient ferric-peroxy species (Ortiz de Montellano, 2005).

Heme reduction is further aided by the dissociation of the distal water ligand (a consequence of substrate binding) prompting an electronic reorganisation in the d-orbitals of the heme iron (Munro et al., 2007). Consequently, the ferric heme iron shifts into a high-spin state. Accordingly, the reduction potential of the heme undergoes a large increase of around +130 mV, vastly improving the heme iron's capacity to acquire electrons (Munro et al., 2002).

In summary, a single mono-oxygenase reaction performed by a P450 requires one molecule of molecular oxygen, two electrons and two protons; yielding a substrate usually oxidised at an aliphatic position and a single molecule of water.

Equation 1.1: $\mathbf{R\text{-}H + O_2 + NADPH + H^+ \rightarrow R\text{-}OH + NADP^+ + H_2O}$

Although contention has previously arisen surrounding the number of iron-oxygen intermediates, the catalytic cycle displayed by P450cam has been accepted as the exemplary model (Guengerich 2018; Meunier et al., 2004). The catalytic cycles of individual P450s may vary depending on both the specific substrate bound and the structural and chemical environment of the P450 active site in question. Notably, different redox partner proteins and other electron donors facilitate the most common departures from the canonical mechanism. The generally accepted P450 catalytic cycle is outlined below and illustrated in Figure 1.9.

1. The substrate molecule enters the active site, causing the displacement of the axial water ligand from the heme iron – resulting in the electronic reorganisation in the 3d orbital electrons of the heme iron.

2. Consequently, the heme iron shifts into a high-spin state, facilitating the transfer of a single electron from the redox partner protein onto the heme iron itself, causing the heme to be reduced from the ferric (Fe^{III}) to the ferrous (Fe^{II}) state.

3. Molecular oxygen subsequently binds to the ferrous heme, forming a transient ferrous-oxy intermediate. Delivery of the second electron to the heme iron via the redox partner reduces the ferrous-oxy intermediate to the second transient intermediate, the ferric-peroxy state.

4. Subsequent protonation of the ferric-peroxy state forms the short-lived ferric-hydroperoxo species known as compound 0.

5. A second, rapid protonation results in the heterolytic cleavage of the bound dioxygen, causing the production and release of a water molecule. Concurrently, the second protonation produces a highly-reactive ferryl-oxo porphyrin radical cation species – compound 1.

6. The highly reactive compound 1 electrophilically attacks the bound substrate – bringing about its hydroxylation (or otherwise catalysing another P450-specific oxidative reaction). On completion of the reaction, the oxidised product is released from the active site.

7. The product's departure enables water to reattach to the heme iron (now in its original ferric state). The enzyme returns to its original resting state - enabling the reaction cycle to start over.

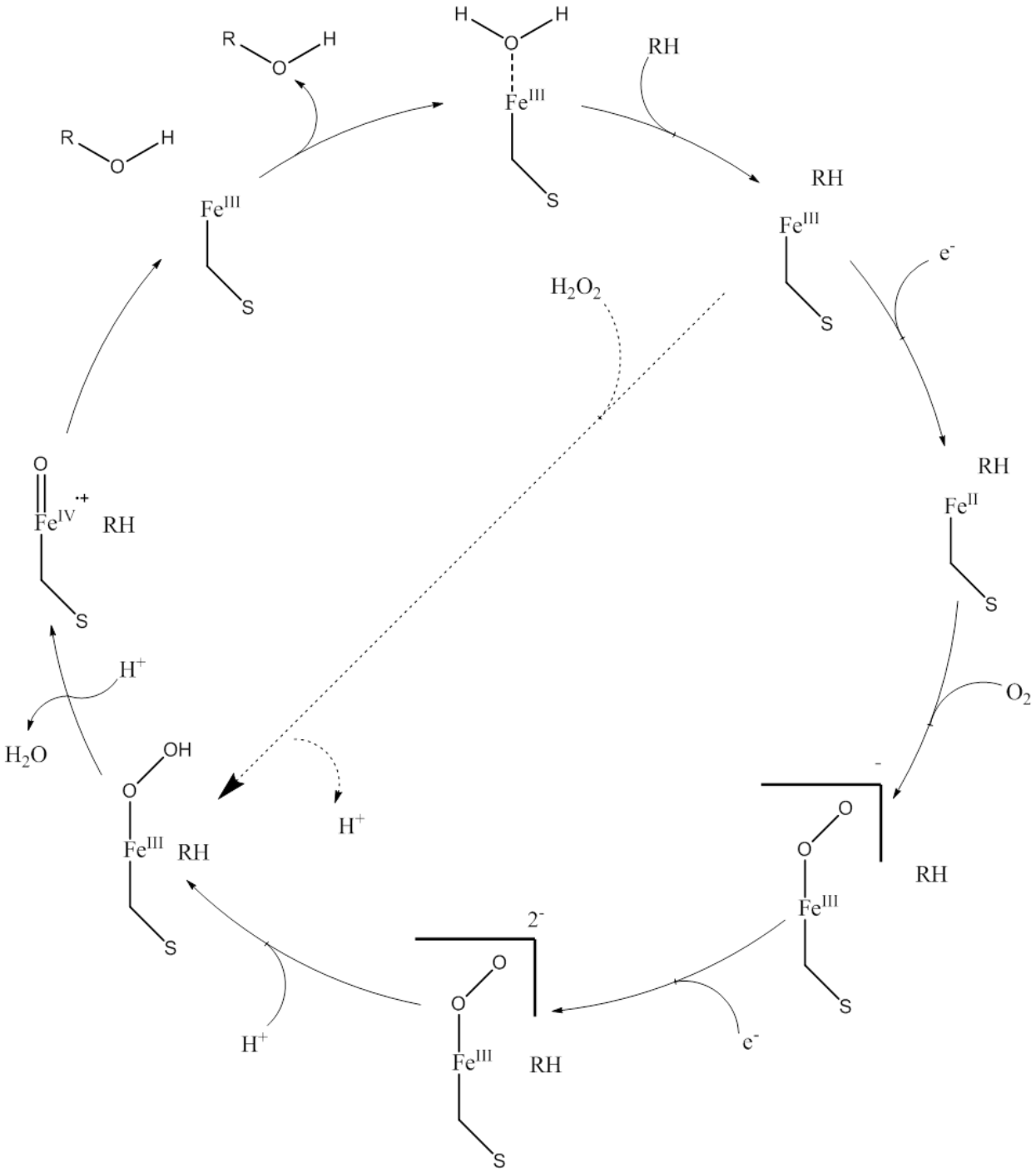

Figure 1.9 The catalytic cycle and "peroxide shunt" mechanism of P450 enzymes.

The proposed mechanism for the insertion of a single oxygen atom into a substrate (RH) performed by P450 enzymes is presented. The uncoupled, "peroxide shunt" mechanism is illustrated using a dashed line. Figure adapted from Meunier et al., 2004; Denisov et al., 2005; Munro et al., 2013 and Guengerich, 2018.

A "shortcut" to this traditional reaction mechanism, which produces compound 0 directly from the substrate-bound P450 in the ferric state, is possible when in the presence of

oxygen atom donor molecules such as hydrogen peroxide (Otey et al., 2006). This alternative route, dubbed the "peroxide shunt", is reversible – with protonation of compound 0 causing the P450 to "uncouple", regenerating the ferric, substrate bound P450 and producing a molecule of H_2O_2 (Figure 1.9) (Luthra et al., 2011; Ortiz de Montellano, 2005). However, owing to their highly oxidative nature, H_2O_2 and other peroxides can cause significant damage to both the heme cofactor and surrounding protein. This, when coupled with the seldom efficient nature of the peroxides, renders the "peroxide shunt" route inappropriate for industrial exploitation (Guengerich, 2013).

1.1.9. P450 Catalysed Reactions

As previously discussed, P450 enzymes are capable of catalysing a variety of reactions. The majority of reactions catalysed by P450s are oxidations, including carbon hydroxylation, epoxidation, group migration, dealkylation and heteroatom oxygenation (Figure 1.10) (Lamb and Waterman, 2013). These oxidations can be rationalised by the general mechanism of catalysis outlined in the previous section. Namely, oxidations involve the highly reactive catalytic intermediate compound I and the abstraction of a hydrogen atom (or electron), followed by a subsequent rebound of an oxygen atom from compound I to oxidise the bound substrate (Guengerich, 2007).

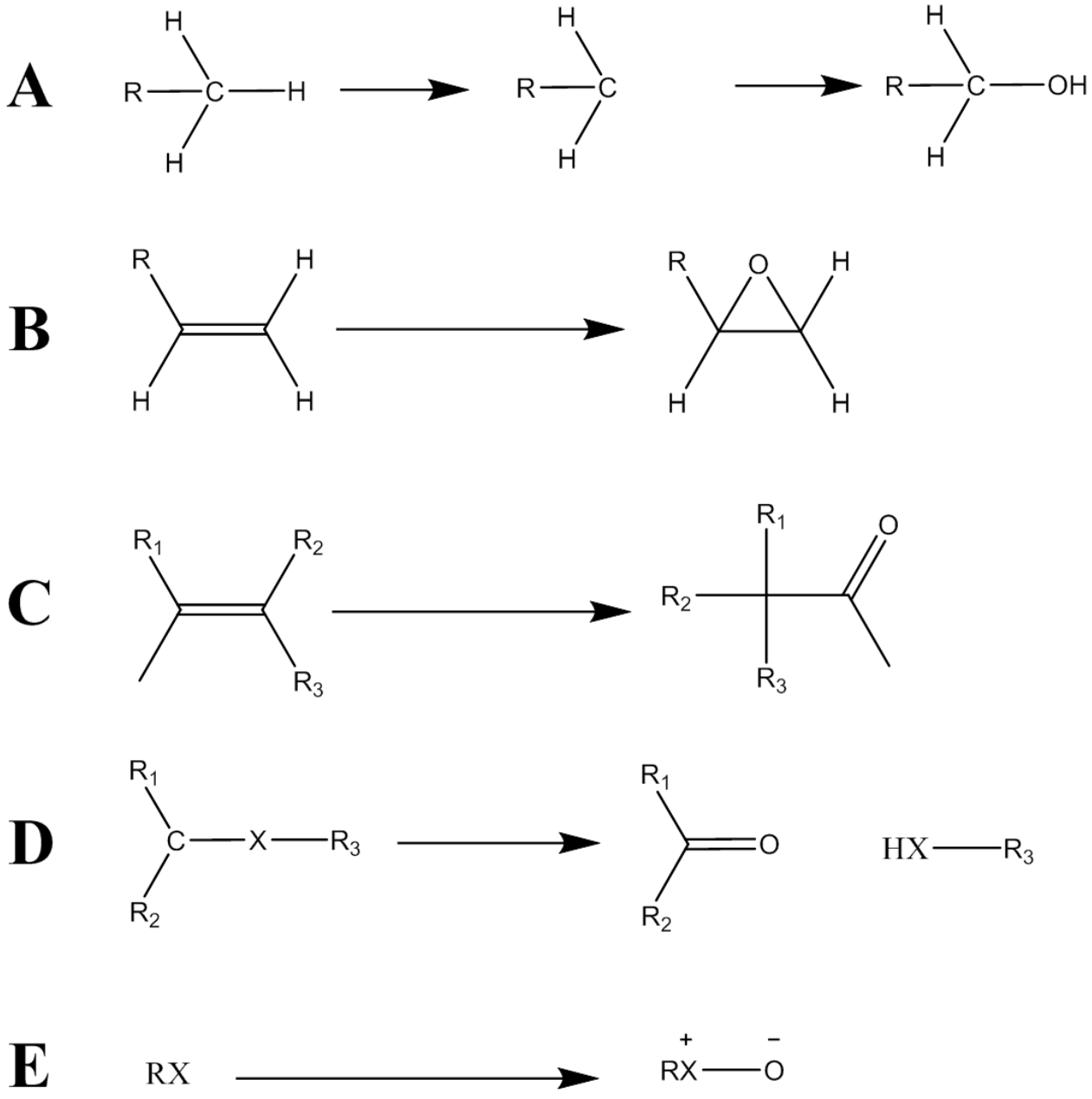

Figure 1.10 Selected reactions catalysed by P450 enzymes.
Inset A: C-H hydroxylation; inset B: C=C bond epoxidation; inset C: group migration; inset D: heteroatom release; inset E: heteroatom oxidation. Figure adapted from Lamb and Waterman (2013).

The type of the reaction performed is dependent on both the P450 and the substrate in question. However carbon hydroxylation, or C-oxidation, is perhaps the most common oxidative reaction performed by P450s (Figure 1.10, inset A). This reaction typically results in the formation of an alcohol. Further oxidation of the alcohol product will result in the formation of a carbonyl, and successive oxidation of a hydrated aldehyde may produce the corresponding carboxylic acid (Guengerich, 2007).

P450s are capable of performing epoxidation reactions with alkene substrates (Vaz et al., 1998). In the case of alkene epoxidation, compound I acts as an electrophile and attacks the carbon-carbon double bond of the substrate, generating an epoxide in addition to aldehyde side products (Meunier et al., 2004). The spin-state of the compound I heme, in addition to the oxidation state of the iron (ferric or ferryl) has a considerable impact on the mechanism of epoxidation and the amount of side products produced (Kumar et al., 2005). If the compound I heme is in the low-spin (doublet) state, epoxide formation proceeds through a concerted mechanism and side product formation is minimal. In comparison, reaction pathways followed by the high-spin (quartet) heme have higher energy barriers for ring closure, altering the stereochemistry of the epoxides produced and prompting group migration (Kumar et al., 2005). P450-mediated group migration is typically observed during the oxidation/epoxidation of aromatic compounds, with the most widely reported being the NIH shift (Meunier et al., 2004; Liebler and Guengerich, 1983; Guroff et al., 1967).

In addition, P450s have been shown to oxygenate heteroatoms with low-oxidation potentials, such as nitrogen, sulfur and phosphorus (Guengerich, 2007). Dealkylation, or heteroatom release, arises from the hydroxylation of a carbon atom adjacent to a heteroatom (commonly nitrogen or oxygen), resulting in the cleavage of the bond between the carbon atom and the heteroatom (Rydberg et al., 2008). As with hydroxylation, these reactions can be rationalised in the context of the transient intermediate, compound I, and the oxygen rebound mechanism (Guengerich, 2001; Brown et al., 2008). However, a single electron transfer mechanism has been proposed in the case of heteroatom oxidation, in which an electron from the heteroatom is transferred to the intermediate, prior to the transfer of oxygen to the heteroatom (Guengerich et al., 1996; Brown et al., 2008). P450s favour dealkylation over N-oxidation when metabolising compounds containing nitrogen atoms (Seto and Guengerich, 1993). For example, the opiate buprenorphine is extensively metabolised by the human P450 CYP3A4, yielding the N-dealkylated metabolite norbuprenorphine (Iribane et al., 1997).

In addition to these oxidation reactions, P450s have also been shown to catalyse reduction, desaturation, ring expansion, ring formation and desaturation reactions, amongst others (Guengerich, 2001). These transformations, oxidative or otherwise, underline the immense versatility of P450s in both the detoxification and bioactivation of xenobiotic compounds.

1.2. Cytochrome P450 BM3

1.2.1. Introduction to Cytochrome P450 BM3

In 1974, Miura and Fulco reported a "cell-free hydroxylating system" isolated from the soil dwelling bacterium *Bacillus megaterium* (Fulco, 1974). Although officially designated as CYP102A1 (Nelson, 1999) as the third such P450 identified in *B. megaterium*, the system was initially dubbed P450 BM3. Since its initial discovery, over 50 homologues of the *CYP102A1* gene have been identified throughout both bacteria and fungi (Whitehouse et al., 2012). BM3 is a noted exception to the established P450 classification system – it features a mammalian-like CPR, but remains soluble and without a membrane anchor region. A module akin to a class I P450 is found in the 55 kDa fatty acid hydroxylase P450 domain, whereas the 65 kDa reductase domain (consisting of both FAD and FMN flavins in an equimolar ratio) resembles a class II system (De Mot and Parret, 2002). However, it is the fusion arrangement between the two heme-binding and FAD/FMN-binding domains which distinguishes BM3 as both a model system and the first representative of a unique type of P450 (Munro et al., 2007). The fusion of the P450 domain to the CPR domain (via the C-terminus of the former) results in a single polypeptide of approximately 119 kDa (Munro et al., 2002; Whitehouse et al., 2012). This arrangement in effect renders BM3 as a soluble class II system, contained within and expressed as a single polypeptide, and is referred to as a class III redox system (Ortiz de Montellano, 2005). Functional as a dimer, P450 BM3 acts as a fatty acid hydroxylase, primarily oxidising mid-chain saturated fatty acids at the ω-1, ω-2 and ω-3 positions (Neeli et al., 2005). Perhaps unsurprisingly, the importance of BM3 from both a research and industrial perspective has led to its extensive study and characterisation.

1.2.2. Structure of P450 BM3

1.2.2.1. Heme Domain

The first crystal structure determined for a BM3 component, that of the P450 domain, was reported by Ravichandran et al. (1993). The structure (PDB code 2HPD), resolved to a resolution of 2 Å, revealed significant differences between BM3 and the only other P450 previously structurally characterised (at that time), P450cam. While the previous P450cam structures failed to show any means of accessing the active site (Poulos et al., 1987), the heme domain structure of BM3 showed a long, hydrophobic channel. Further studies postulated the occurrence of significant structural rearrangements upon substrate binding. This was confirmed when the first crystal structure of BM3 *in situ* with a fatty acid substrate bound, palmitoleic acid (PDB 1FAG; Li et al., 1997) demonstrated rearrangements in the

F/G loop, as well as in both the F- and G-helices. Disruption of these regions has been linked to destabilisation of surrounding residues, through loss or rearrangement of salt-bridges and hydrogen bonds within the active site (Whitehouse et al., 2012). High resolution crystal structures of the heme domain bound to the fatty acid-like substrate NPG (PDB 1JPZ; Haines et al., 2001) revealed the critical role of a water molecule in the transition between heme-iron spin states. Disruption of the I-helix resulting from substrate binding displaces this water molecule, causing it to dissociate from the heme iron. Consequently, the spin-state of the iron transitions from low-spin to high-spin (Haines et al., 2001). Disruption of the I-helix has also been implicated in the creation of a possible dioxygen binding site (Yeom et al., 1997). The role of the I-helix in regards to both substrate selectivity and regio- and enantioselectivity is detailed in chapter 2.

1.2.2.2. FMN and FAD domains

Unique amongst P450s of both classes, the P450 and CPR domains of BM3 and its homologues are linked via a polypeptide linker region. If this linker protein is cleaved, the P450 and CPR domains demonstrate a distinct lack of affinity for one another, rendering monooxygenase activity non-functional (Govindaraj et al. 1996). While both the P450 and CPR domains have been extensively studied and expressed and purified separately, the reductase domain has proven to be less stable than the reductase domain, with a lower thermal sensitivity (Jamakhandi, 2005). The two sub-domains of the reductase, the FMN and FAD/NADPH binding domains, are amenable to expression/purification and are stable in their dissected forms (Figure 1.11).

The structure of the P450 BM3 FMN-binding domain, as identified in a crystal structure formed from 2 molecules of the heme domain and 1 molecule of the FMN domain, was solved by Sevrioukova et al. (PDB code 1BVY) (Sevrioukova et al., 1999). In the ~2 Ångstrom structure, the single FMN domain was found to be positioned at ~18 Ångstroms from the heme of one of the heme domains, with the planes of the flavin and heme cofactors near perpendicular to one another. The binding site of the cleaved FMN domain is enveloped by hydrophobic or neutral residues, in contrast to the negatively charged residues that surround the FMN domain in the intact CPR. This serves to stabilise the BM3 FMN, increasing its reduction potential and consequently aiding the FAD to FMN electron transfer. Within the BM3 heme domain, residues Pro382 to Cys400 were hypothesised to serve as a direct electron transfer pathway from the FMN to one of the hemes. However, no experimental evidence was presented for this pathway (Sevrioukova et al., 1999).

The NADPH-binding FAD domain was the final domain of BM3 to have its crystal structure determined, in both the presence and absence of NADPH (PDB codes 4DQK and 4DQL, respectively) (Joyce et al., 2012). Mutagenesis of residues Cys773 and Cys999 to alanine, in order to prevent dimerisation of the FAD domain through formation of inter-domain disulphide bridges, was required for successful crystallisation of the monomeric FAD domain. The FAD domain of BM3 displays a strong preference for NADPH over NADH (K_M values of 6.5 μM and 3030 μM, respectively) (Neeli et al., 2005b). This preference can be attributed to the polar interactions formed between several charged residues at the cofactor binding site and the adenine ribosyl 2'-phosphate component of NADPH (Joyce et al., 2012). Superposition of the FAD and FMN domain revealed the close (~4 Angstrom) association of the isoalloxazine rings from each flavin, consistent with the direct and highly efficient electron transfer between the domains (Joyce et al., 2012; Munro et al., 1994). Structurally, the BM3 FAD domain resembles that of the FAD domain found in the CPR of the mammalian (rat) nitric oxide synthase (NOS) (Garcin et al., 2004).

Attempts to crystallise the full-length BM3 protein with all domains intact have so far proven unsuccessful (Joyce et al., 2012). This can be attributed to the complex, multi-domain arrangement of BM3 and the conformational heterogeneity of its domains (Girvan et al., 2007). The structure of a mutant BM3 FAD/NADPH-binding domain has been solved via rational design. This entailed the removal of a solvent-exposed cysteine residue that otherwise promoted the non-specific formation of disulfide bonds, resulting in a barrier for crystallisation (Girvan et al., 2007). The design of similar mutants for the full-length BM3 protein could prove key in obtaining a crystal structure of this unique fusion enzyme.

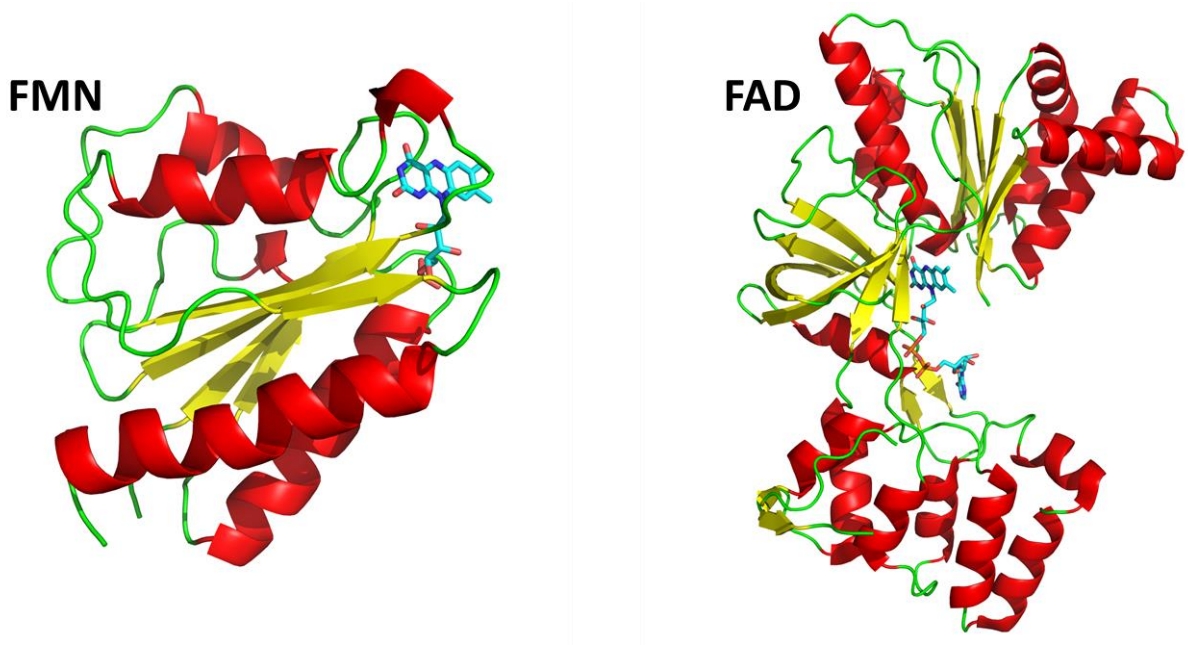

Figure 1.11 Structures of the P450 BM3 CPR redox partner domains containing flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD).
Alpha helices are represented in red. Beta sheets are represented in yellow. Random coils are represented in green. Isoalloxazine rings are coloured in cyan. The FMN domain structure is from Sevrioukova et al., 1999 (PDB 1BVY). The FAD domain structure is from Joyce et al., 2012 (PDB 4DQK).

1.2.3. Electron Transfer within Intact P450 BM3

The unusual fusion arrangement of BM3 makes it an attractive model enzyme in which to study electron transfer. A highly efficient monooxygenase, BM3 possesses the highest catalytic activity reported for P450s at ~17000 min^{-1} with arachidonic acid (Noble et al., 1999). Such activity has been attributed to the rapid electron transfer between the domains of BM3 via their constituent cofactors (Munro et al., 1996). Akin to class I systems, 2 electrons from NADPH (as a hydride ion) are channelled onto the FAD cofactor, before their being passed on to the FMN cofactor, which then shuttles the 2 single electrons to the heme iron. Upon reduction, the FAD becomes 2-electron reduced to form the hydroquinone. Thereafter, single electron transfer to the FMN results in the isoalloxazine rings of the FAD and FMN in both flavin sub-domains forming semiquinones. The anionic semiquinone of the FMN is the donor to the ferric heme iron, and after the first FMN-to-heme electron transfer the seconds electron from the FAD is transferred to the FMN to allow the anionic FMN semiquinone to reduce the ferric-superoxo species, thus forming the ferric-peroxo form (Whitehouse et al., 2012). The spatial positioning of the flavin cofactors

ensures a distance of less than 4 Angstroms between their respective isoalloxazine rings, thus facilitating efficient electron transfer (Munro et al., 2002). The flavin domains are attached via a "hinge" domain, postulated to switch between "open" and "closed" conformations (Munro et al., 2002). The closed conformation facilitates efficient inter-flavin electron transfer, and, following a ~10 Angstrom shift, switches to the open conformation to allow interaction (and subsequent reduction) of the P450 heme iron by the FMN domain (Munro et al., 2002). It is important to note that BM3 is active as a dimer, with the P450 domain of each monomer receiving electrons from the reductase of the opposing monomer, as presented in Figure 1.12 (Kitazume et al., 2007).

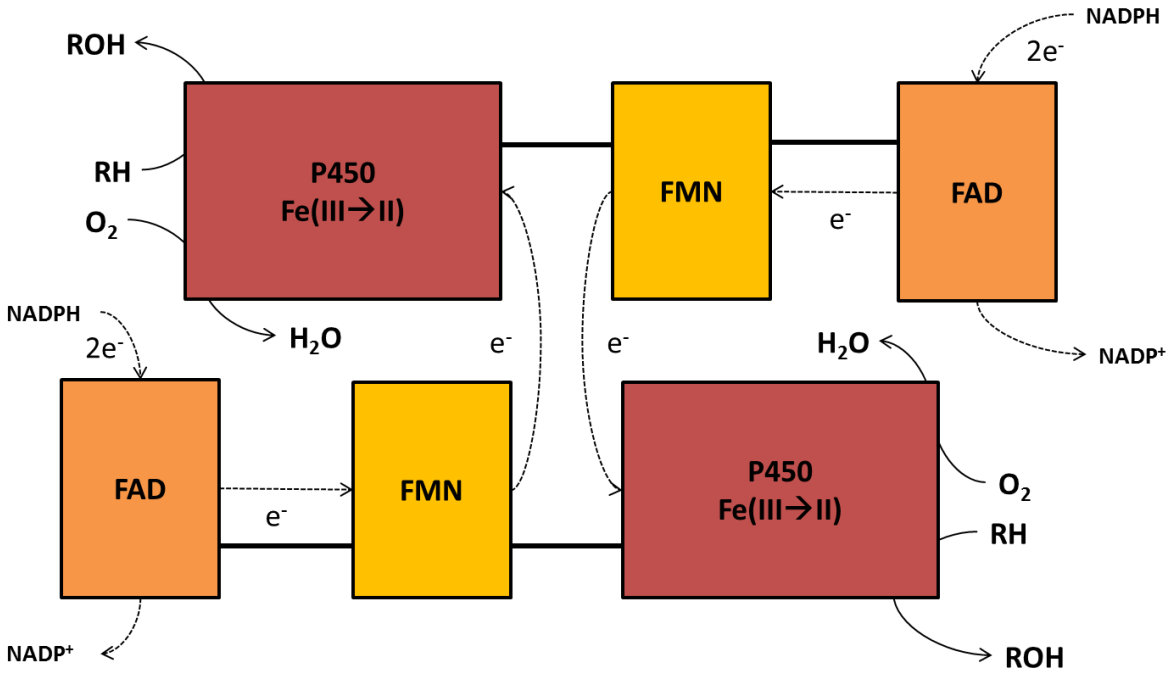

Figure 1.12 Electron transfer between monomers of intact, dimeric P450 BM3, as proposed by Munro et al. (2002).
The P450, FMN and FAD domains are represented in maroon, yellow and orange, respectively. Within each monomer, the P450 domain receives electrons from the FMN domain of the opposing monomer. Substrate is represented by RH.

Given the risk of oxidative damage to the heme/protein posed by the production of reactive oxidation species, control of the electron transfer process is essential. Modulation of the heme iron's reduction potential, determined by its spin-state, prevents the reduction of substrate-free, low-spin heme iron. This initial level of control is present in several members the P450 superfamily. In BM3, a second level of control is modulated via the reductase domain. If pre-incubated with NADPH in the absence of substrate, BM3

hydroxylase activity undergoes a significant time-dependent decline in activity (Daff et al., 1997). This occurrence is underpinned by the rapid accumulation and subsequent predomination of the reductase in a three-electron reduced state (with the FMN reduced to its hydroquinone form), which is far less efficient in reducing the heme iron (Munro et al., 1996). This adaptation most likely arose in order to maintain catalytic efficiency by preventing oxidation of NADPH in the absence of fatty acid substrate, whilst concurrently guarding against the formation of destructive oxygen radicals (Munro et al., 2002).

1.2.4. Physiology and Substrate Selectivity of P450 BM3

While known to function in fatty acid hydroxylation, the exact physiological role performed by BM3 remains unclear. In BM3's soil-dwelling parent organism *B. megaterium*, unsaturated fatty acids have proven toxic when applied exogenously (English et al., 1997). Owing to this, the primary function of BM3 has been postulated to be the detoxification of xenobiotic lipids produced by plants (Whitehouse et al., 2012). BM3 has been shown to act on a diverse range of substrates, including both saturated and unsaturated fatty acids, amides and alcohols, and catalyses an equally diverse array of reactions, notably substrate epoxidation and hydroxylation (Boddupalli et al., 1990). Indeed, BM3 has the highest reported catalytic activity observed for any P450 oxygenase – the hydroxylation of arachidonic acid produced a k_{cat} value of ~17000 min^{-1} (Noble et al., 1999). Hydroxylation of fatty acids typically occurs at the ω-1 to ω-3 positions, with an optimal aliphatic chain length of C_{12} to C_{20}, and with the highest activity reported for chain lengths of C_{15} and C_{16}. (Whitehouse et al., 2012; Narhi and Fulco, 1986). While capable of binding fatty acids with chain lengths fewer than 12 carbons, the resulting activity is low (Munro et al., 2002). Recent studies have pointed to branched-chain fatty acids as the natural substrates of BM3, including the high proportion of branched-fatty acids (~80%) present within the *B. megaterium* membrane (Kaneda, 1991). The high regio- and stereochemical purity of products resulting from BM3-mediated hydroxylation of branched-chain fatty acids has provided further evidence to support this claim (Cryle et al., 2006).

Unlike its fellow model system P450cam, in BM3 the consumption of NADPH is not directly coupled to product formation when saturated with substrate *in vitro* (Narhi and Fulco, 1986). In a further dissimilarity to P450cam, expression of BM3 cannot be induced in *B. megaterium* when saturated with aliphatic, saturated fatty acid substrates (Narhi and Fulco, 1982). In contrast, BM3 expression is induced when *B. megaterium* is grown in the presence of branched-chain, saturated fatty acids – lending further credence to the suggestion that these are the natural substrates of BM3 (English et al., 1997).

1.2.5. Key Amino Acid Residues for P450 BM3 Mutagenesis

Similar to other mammalian P450s, BM3 crystallises in a conformation corresponding to its pre-catalytic state (Whitehouse et al., 2012). Because of this, substrate is absent from the immediate vicinity of the heme iron, therefore making it difficult to determine which residues surrounding the hydrophobic access channel play roles in catalysis. Nevertheless, the roles of most residues surrounding the active site have been elucidated via site-directed mutagenesis studies, and assessed by correlating the oxidation rates and product range produced by these mutants to those of wild-type BM3. Using this method in partnership with structural data, several "key" residues, those which have the greatest influence on the substrate selectivity and overall activity of BM3, have been identified. Residues that have been the subject of intense study within BM3 with the aim of altering substrate selectivity include Phe87, Ala82, Val78, Tyr51 and Arg47 (Butler et al., 2013; Whitehouse et al., 2012; Noble et al., 1999). In addition, the residue Ile401, adjacent to the highly conserved Cys400 residue at the proximal face of the heme, has been shown to have significant influence over rates of catalysis in variant BM3 enzymes (Whitehouse et al., 2009). The positions of these key residues, relative to one another, the heme domain and substrate NPG, are presented in Figure 1.13.

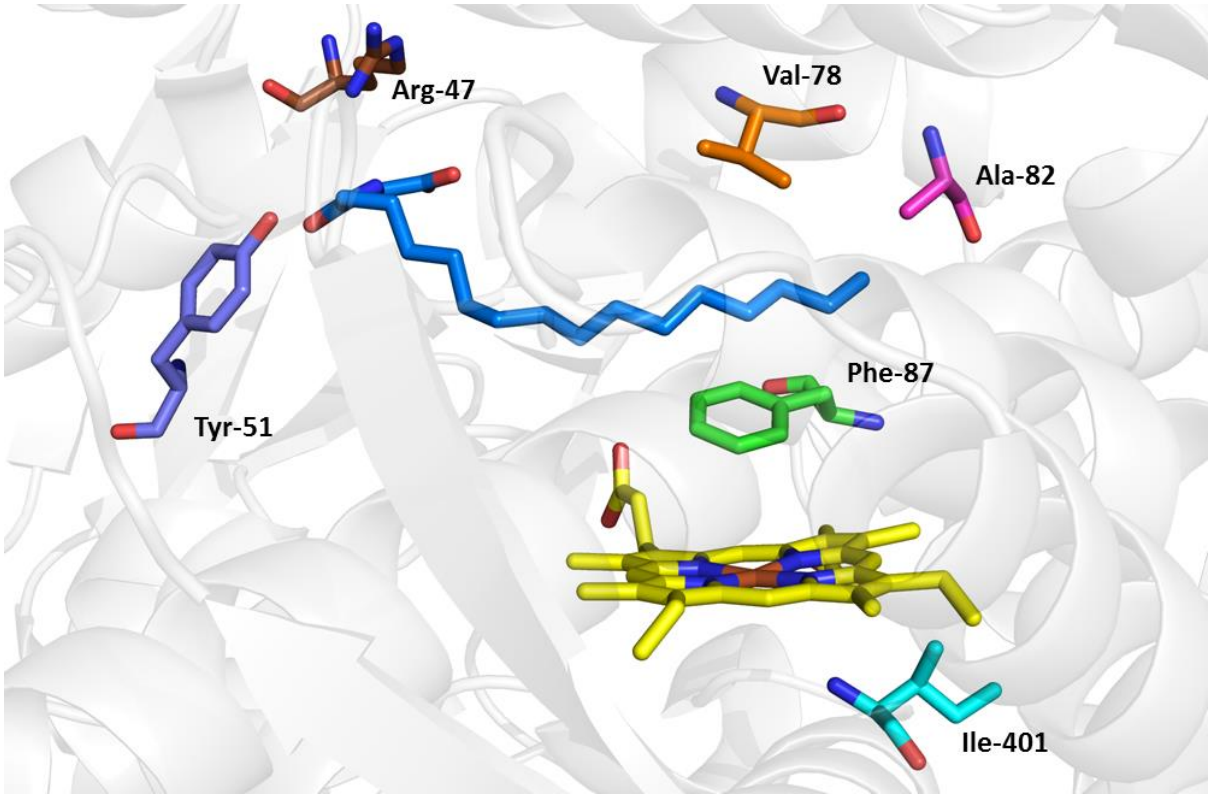

Figure 1.13 Key targets for mutagenesis in P450 BM3.
At the mouth of the active site, Arg47 (brown) and Tyr51 (lilac) are involved in substrate recognition and orientation. Val78 (orange) is involved in substrate orientation. Ala82 (pink) has a role in substrate selectivity. Phe87 (green), the most widely mutated residue in the protein, undergoes significant rearrangement following substrate binding and is integral in regulating substrate access to the heme. Proline-substituted variants of Ile401 (cyan), have demonstrated significant increases in NADPH consumption and rates of catalysis (Whitehouse et al., 2012). The heme cofactor is highlighted in yellow, with substrate NPG highlighted in blue. Secondary structural elements are presented in grey. Figure generated from PDB file 1JPZ (Haines et al., 2001).

1.2.5.1. Arg47 and Tyr51

Following the determination of a wild-type BM3 heme domain crystal structure in the fatty acid-bound state, two hydrophilic residues, Arg47 and Tyr51, were found to be integral to fatty acid substrate binding, confirming Fulco's initial prognosis (Ravichandran et al. 1993). Situated at the mouth of the hydrophobic binding channel, Arg47 and Tyr51 interact with the carboxylate group of fatty acid substrates, in effect tethering them in the immediate vicinity of the heme iron. Removal or substitution of either residue by site-directed mutagenesis has been found to raise substrate binding constants whilst lowering hydroxylase activity, with a double substitution mutant (R47A/Y51F) causing a two-thirds decrease in catalytic efficiency (Ost et al., 2000). Despite this reduction in efficiency with

fatty acid substrates, double mutations of Arg47 and Tyr51 have also been found to decrease substrate specificity, improving recognition of aromatic hydrocarbon substrates (Carmichael et al., 2001). Kinetic studies conducted by Noble et al. (1999) have shown that, of the two residues, Arg47 has a greater impact on fatty acid binding by acting to stabilise the fatty acid carboxylate.

1.2.5.2. Val78

Located within the binding channel of P450 BM3, Val78 is a residue vital to controlling the binding orientation of fatty acid substrates within the active site (Zheng et al., 2016). Crystal structures of BM3 heme domains bound to fatty acid substrates have revealed the proximity of Val78 to the terminal ω-carbon, with V78A variants demonstrating significant departures from the wild-type product profile with fatty acid substrates (Li and Poulos, 1997; Dietrich et al., 2009). In tandem with residues Ala82, Phe87, Thr88, Ile263 and Leu264, Val78 comprises a subpocket within the active site postulated to be essential in regulating regio- and stereoselective hydroxylation of fatty acid substrates (Chiang et al., 2013).

1.2.5.3. Ala82 and Phe87

A phenylalanine residue at position 87 undergoes a 90 degree structural rotation upon substrate binding. Residue Phe87 forms part of the hydrophobic access channel in close vicinity to the heme iron (Whitehouse et al., 2012). This rotation is postulated to be responsible for the displacement of the axial water ligand from the heme iron, and is thought to trigger further reorganisations within the active site structure (Haines et al., 2006). Phe87 is probably the most widely mutated residue in the whole protein, and the role of Phe87 has been extensively studied. Phe87 functions primarily in regulating the binding of the ω-terminal region of fatty acids, preventing hydroxylation at the ω-methyl position. Early studies suggested that mutations that removed the Phe87 aromatic side chain resulted in ω-hydroxylation. However, later studies showed instead that the effect was to introduce hydroxylation at a position further from the ω-terminal (ω-4, ω-5 etc) (Whitehouse et al., 2012). Owing to its bulky, hydrophobic structure, Phe87 prevents water (and other solvent) molecules from accessing the immediate vicinity of the heme iron. In doing so it helps to maintain the hydrophobic character of the active site (Kuper et al., 2012). Mutations which feature Phe87 replaced by more “svelte” residues tend to display sharp increases in uncoupling, albeit with concurrent increases in kinetic efficiency when compared to the activity of wild-type BM3 with the same substrates (Cowart et al., 2001). Because of this reduction in substrate specificity, substitution of the Phe87 residue is

favourable for the oxidation of non-natural, large and hydrophilic substitutes. Consequently, such mutations – notably F87V and F87A – are commonly found in BM3 variants designed for the metabolism of drug-like compounds, as well as for numerous other industrial applications (Jeffreys et al., 2019; Whitehouse et al., 2012; Sowden et al., 2005). These Phe87 variants achieve this by expanding the dimensions of the active site, enabling these larger substrates to get closer to the heme iron for catalysis (Jeffreys et al., 2019).

Located at a position peripheral to that of the heme cofactor, it is perhaps surprising that residue Ala82 has a significant impact on BM3 substrate selectivity. However, the A82F substitution induces a significant structural readjustment in the F- and G-helices, altering the conformational landscape of the BM3 active site (Butler et al., 2013). In doing so, A82F variants are able to accommodate novel substrates, such as the proton-pump inhibitor omeprazole (Butler et al., 2013). This increased active site plasticity can be attributed to the destabilising impact of the A82F substitution, which lowers the free energy barrier required for the active site to transition to the substrate-bound state (Bloom et al., 2006; Butler et al., 2013).

A BM3 double mutant incorporating F87V/A82F point mutations, known as the "gatekeeper", displays greater flexibility relative to the wild-type enzyme (Jeffreys et al., 2019). Consequently, the gatekeeper BM3 variant is able to bind to and oxidise a structurally diverse array of drug compounds that do not productively interact with wild-type P450 BM3. Notably, the gatekeeper variant has been shown to produce oxidised metabolites analogous to those produced by a human P450 with esomeprazole, lansoprazole sulfone and rabeprazole desmethyl ether (Jeffreys et al., 2019; Butler et al., 2014; Butler et al., 2013).

1.2.5.4. Ile401

Unlike other targets for mutagenesis discussed within this section, Ile-401 is located on the proximal side of the heme (Whitehouse et al., 2012). Adjacent to the C-terminal face of the highly conserved Cys400 residue, Ile401 has been the subject of a number of recent substitution studies with a focus on increasing BM3 activity with non-natural substrates (Whitehouse et al., 2009). When substituted with a proline residue, a naturally occurring residue at this position in a number of P450s, consumption rates of NADPH with non-natural substrates increased by a minimum of 400% (Whitehouse et al., 2009). The I401P mutant has a particularly dramatic effect on substrate oxidation rates across a broad range

of hydrophobic, and in some cases aromatic, hydrocarbons – substrates with which the wild-type enzyme is virtually inactive. This is a feat unmatched by any other known single-mutation BM3 variants (Whitehouse et al., 2010).

1.2.6. Biotechnological and Industrial Engineering of P450 BM3

The Organisation for Economic Co-operation and Development (OECD) defined the term "biotechnology" as: "the application of science and technology to living organisms as well as parts, products and models thereof, to alter living or non-living materials for the production of knowledge, goods and services." (OECD, 2005). It can be argued that, in utilising microorganisms for the purposes of brewing and baking, humankind has been practicing biotechnology for thousands of years. Both practices, in addition to the use of fertiliser within agriculture, were widespread throughout several ancient societies such as those of Mesopotamia and Egypt (Bud, 1994). Regarding the use of enzymes to produce high-value hydroxylated compounds, the earliest examples can be found in the use of bacterial fermentations of *Rhizopus* cultures to produce cortisone from progesterone via 11-α-hydroxylation (Peterson, 1952). However, within these early examples, the chemical process or enzymes involved were either poorly understood or uncharacterised (Guengerich, 2002). The birth of modern biotechnology can therefore be traced to the work performed by Cohen, Boyer and colleagues (Cohen et al., 1973), in which recombinant plasmids were successfully inserted into *E. coli* and shown to be functional..

As previously detailed, members of the P450 family catalyse the incorporation of a single oxygen atom into a variety of substrates, often with high regio- and stereo-specificity (Guengerich, 2001). Throughout nature, P450s are the primary catalysts involved in the detoxification of xenobiotic substrates. Consequently, there has been considerable interest from the pharmaceutical industry towards the applications of P450s in drug development (Guengerich and Ortiz de Montellano, 1995). To reduce the risk of adverse drug reactions, the metabolic profile of a drug must be known prior to commercialisation (Di Nardo and Gilardi, 2012). Consequently, toxicity testing must be performed on all metabolites of the drug candidate, the synthesis of which may prove challenging and/or expensive via traditional chemical methods (Di Nardo and Gilardi, 2012). In light of this, metabolite synthesis performed by P450 enzymes may offer an attractive, cost effective alternative.

Engineering of human P450s for industrial applications has been performed (Gillam and Guengerich, 2001). However, the difficulty and expense involved in obtaining human P450s, coupled to their often low levels of catalysis and stability, may render them

unsuitable for use in industry (Julsing et al., 2008). In general, the membrane-bound nature of eukaryotic P450 enzymes makes them unstable and difficult to purify. The additional requirement for a membrane-bound CPR partner further highlights the problems associated with applications of eukaryotic P450s for use as industrial biocatalysts. Because of these issues, bacterial P450s, notably P450-redox partner fusion systems, have distinct advantages over the eukaryotic P450s and are increasingly used for industrial applications.

The production of human metabolites and the study of pro-drugs have been the traditional applications of P450 within industry. However, numerous studies have focused on the uses of P450s in the production of fine or speciality chemicals, such as those of interest to the agricultural, fragrance and flavouring industries (Urlacher and Eiben, 2006). Bacterial P450s have also been utilised in the synthesis of high-value drug compounds. CYP105AS1, from *Amycolatopsis orientalis*, has been engineered to produce the cholesterol-lowering drug pravastatin via the stereo-specific hydroxylation of compactin (McLean et al., 2015). In the same bacterium, P450s OxyA, OxyB and OxyC catalyse hydroxylation steps in the synthesis of glycopeptide antibiotics (Bischoff et al., 2005). CYP107A1, found in *Saccharopolyspora erythraea*, is implicated in the biosynthesis of erythromycin, a macrolide antibiotic (Andersen et al., 1993).

As a highly active, self-sufficient and soluble redox system, P450 BM3 is an attractive proposition for industrial exploitation (Munro et al., 2002). Highly amenable to *in vivo* expression in recombinant *E. coli*, and more easily purified and cheaper than its more "temperamental" mammalian expression system counterparts, BM3 is particularly suited for use in the pharmaceutical industry. Wild-type BM3 has demonstrated activity towards substrates metabolised by the human P450s CYP3A4, CYP2E1 and CYP1A2 (Di Nardo et al., 2007). Notably, wild-type BM3 was able to catalyse the hydroxylation of aniline, chlorzoxazone and *p*-nitrophenol, in addition to the N-dealkylation of nifedipine (Di Nardo et al., 2007). Moreover, dealkylation of fluorescent substrates such as alkyl-resorufins, a test often used in high-throughput screening or toxicity assays, is a reaction capable of being performed by wild-type P450 BM3 (Urlacher and Eiben, 2006). Owing to its ability to mimic the activities of a number of mammalian P450s in its engineered forms (including CYP3A4), BM3 is useful for the generation of drug metabolites analogous to those produced by human P450s, fulfilling the essential, toxicity-testing step in drug development (Di Nardo and Gilardi, 2012).

As previously discussed, the ability of P450 BM3 to metabolise a wide range of compounds is in part attributable to the conformational flexibility of its active site (Ost et al., 2000). However, as the plastic nature of the active site affords the substrates a high degree of mobility, product output tends towards a mix of compounds (Whitehouse et al., 2012). Because of this, research tends to focus on redesign of the active site in order to induce activity towards novel substrates with the aim of yielding a single high-value compound.

Two broad approaches exist in the industrial development of BM3; rational design and random mutagenesis. While rational design of the BM3 active site has produced variants capable of acting upon non-natural substrates, it can often be difficult to predict the effects that a mutation would have on the overall substrate specificity and functionality of the protein. For example – a mutation which can successfully alter substrate selectivity may significantly weaken structural stability (Weissman, 2004). Rational mutagenesis approaches have been widely applied to P450 BM3 (Whitehouse et al., 2012). Residues Phe87, Thr268 and Ala328 were the earliest targeted for mutagenesis, with the aim of oxidising the non-natural substrates 1,1,2,2-tetrachloroethane and 2-phenylpropanal (Tsotsou et al., 2002; Whitehouse et al., 2012). Rational BM3 mutagenesis has yielded variants capable of producing (+)-nootkatone, a valuable grapefruit flavouring, from the sesquiterpene valencene (Sowden et al., 2005). A polar couplet modelled on the Arg47/Tyr51 pairing, introduced further down the substrate access channel (L75T/L181K), produced BM3 variants capable of accepting short-chain fatty acids (Ost et al., 2000). Regioselectivity of fatty acid hydroxylation has also been altered in BM3, with the rationally designed S72Y/V78A/F87V triple variant hydroxylating fatty acids at positions unreachable by the wild-type enzyme (Dietrich et al., 2009).

The forced/directed evolution approach subjects wild-type enzymes to repeated rounds of random mutagenesis. Mutants with desirable attributes are then selected by high-throughput screening methods (McLean et al., 2007). This approach has a number of advantages. For example, neither detailed structural data, such as that provided by crystal structures of the active site, nor a thorough knowledge of the reaction mechanism, are required for directed evolution (Urlacher and Eiben, 2006). However, this can present difficulties in reconciling the structural positions of some randomly introduced mutations with their impacts on catalytic activity evolution (Urlacher and Eiben, 2006).

Several random mutagenesis approaches have been successfully applied to P450 BM3. The first randomly generated BM3 variant reported, P25Q, demonstrated altered binding

affinities and product profiles with fatty acid substrates (Maves et al., 1997). Using error prone PCR, a library of BM3 variants capable of producing indigo from indole was produced. All variants contained mutations at the positions Ala74, Phe97 and Leu188 – and a subsequent site-saturation study produced the variant A74G/F87V/L188Q (GVQ) (Li et al., 2000). The GVQ variant is capable of oxidising a plethora of non-natural substrates, including alkanes, cycloalkanes, arenes and heteroarenes (Appel et al., 2001).

Random mutagenesis can also help identify residues within the protein that impact activity but have been previously unexplored, usually on account of their distance from the active site. A triple BM3 variant (A74E/F87V/P386S) demonstrated an 80-fold increase in activity with the rose ketone β-ionone and produced a single product, a high value flavour compound (R)-4-hydroxy-β-ionone (Urlacher et al., 2005). Located on the surface of the protein and some distance from the active site, residue P386 was identified via random mutagenesis. The additional mutations were then added to variant P386S in a rational manner (Urlacher et al., 2005). This combinatorial approach has produced variants demonstrating efficient catalysis and unparalleled substrate selectivity with non-natural substrates such as indole, 3,4-methylenedioxy-methamphetamine (MDMA) and various alkanes (Peters et al., 2003; Vugt-Lussenburg et al., 2007; McLean et al., 2007).

An iterative approach to site-saturation mutagenesis (ISM), which utilises randomization of DNA codons to produce all possible canonical amino acids at a given position, has been applied as an effective means of directed evolution by Reetz and co-workers (Parra et al., 2013; Acevedo-Rocha et al., 2018). In combination with mutability landscape data and molecular dynamics simulations, ISM yielded high-activity P450 BM3 variants with impeccable regio- and stereoselective hydroxylation of testosterone at the commercially important C16-position (Acevedo-Rocha et al., 2018).

Semi-rational approaches to BM3 engineering, based around mutant libraries developed from well-known first-sphere residues (residues that interact directly with the substrate) have also yielded variants with high levels of regio and stereoselectivity across a broad range of compounds (Capoferri et al., 2016; You et al., 2018; Urlacher and Girhard, 2019). P450 BM3 variants constructed in this manner by Urlacher and co-workers displayed 100% regioselective hydroxylation at the C9 position of terpenoid β-cembrenediol, a potential neuroprotective drug (Le-Huu et al., 2015). A novel “fingerprint method”, in which the active site configurations of a variant BM3 library were mapped using a set of semisynthetic chromogenic probes, enabled time-effective assessment of the

regioselective properties of a large number of BM3 variants towards a target substrate (Zhang et al., 2011). This approach generated a high-activity variant capable of selective hydroxylation of an anti-malarial agent, artemisinin, with perfect enantioselectivity and on a preparative scale (Zhang et al., 2011).

In recent years, a variety of BM3 variants engineered to catalyse non-natural reactions have been reported (Wei et al., 2018). Within the lab of Frances Arnold, cyclopropanation of styrene, intramolecular C-H amination and carbene transfer have all been achieved by variant BM3 enzymes (Coelho et al., 2013; McIntosh et al., 2013; Wang et al., 2014). These reactions were previously only possible through non-enzymatic means of synthetic chemistry (McIntosh et al., 2014).

The use of P450 BM3 as a biocatalyst on an industrial scale is limited by the requirement of the costly electron-donating cofactor, NADPH – although NADPH-regenerating systems can be used. The most widely used is the glucose-6-phosphate dehydrogenase system, which converts the cheaper $NADP^+$ to NADPH, consequently enabling longer reaction times (Waltham et al., 2011). Use of the "peroxide shunt" mechanism (illustrated in Figure 1.9) in which H_2O_2 replaces NADPH as the oxidant, is unfeasible on a large scale due to the oxidative damage sustained by the heme in the presence of excess H_2O_2 (Cirino and Arnold, 2003). Rational mutagenesis on residues surrounding the NADPH binding site has yielded P450 variants capable of functioning using the much cheaper and more stable NADH cofactor (Girvan et al., 2011; Neeli et al., 2005). A BM3 variant containing the substitution W1046A in the reductase domain was found to remove an aromatic "shield" surrounding the isoalloxazine ring of the FAD domain (Girvan et al., 2011). This enabled the variant to use the cofactor NADH in preference to NADPH, in addition to displaying improved levels of product formation compared to the wild-type enzyme (Girvan et al., 2011). Variant P450s which employ zinc compounds as an alternative electron source in lieu of NADPH have also been created, albeit with modest success (Schwaneberg et al., 2000). In addition, hybrid BM3 enzymes in which the heme domain has been covalently bound to a ruthenium(II) photosensitiser have demonstrated effective, light-driven catalysis with the substrate lauric acid (Tran et al., 2012).

Conditions required for industry, such as those encountered during production, recovery and storage are often not hospitable for P450s (Salazar et al., 2003). The high temperatures required for many industrial processes can cause significant reductions in enzyme activity, including denaturation. As a result, the thermostabillity of P450 BM3 is

another limiting factor for its current use within industry. A H_2O_2-driven BM3 heme domain variant (generated by directed evolution) with a T_m of 57.5 °C demonstrated a 250-fold increase in its half-life relative to the wild-type P450 BM3 heme domain (Salazar et al., 2003).

Poor expression, lack of substrate selectivity and low levels of activity are all additional roadblocks that serve to limit the wider exploitation of P450s within industry (Lundemo and Woodley, 2015). As previously discussed, improvements to catalytic activity and substrate selectivity are often achieved using a rational approach to mutagenesis of the P450 BM3 heme domain (Whiehouse et al., 2011; Butler et al., 2013). Both protein stability and coupling efficiency of P450 BM3 has been shown to improve when whole-cell formulations, as opposed to the use of cell-free or partially-purified enzyme, have been used in industrial processes (Kaluzna et al., 2016). Furthermore, heterologous co-expression of glucose dehydrogenase alongside wild-type P450 BM3 has also been shown to significantly improve regeneration of the costly NADPH cofactor, negating the requirement for further additions (Schewe et al., 2008). Building on the work of Schewe (et al., 2008), a group based within DSM (Geleen, Netherlands) were able to successfully obtain kilogram quantities of a high value oxidised compound (4-hydroxy-α-isophorone) using a P450 BM3 variant at a scale of 100 L (Kaluzna et al., 2016). Despite these advances, a number of issues remain – the most pressing being the impact of poor coupling efficiency on the stability of the enzyme, which in turn leads to compromised levels of turnover (Kaluzna et al., 2016).

Although several attempts have been made to mimic the high-activity fusion arrangement of BM3 by attaching bacterial P450s to eukaryotic reductases, there has been limited success (Deeni et al., 2001). Artificial self-sufficient constructs, modelled on that of P450 BM3, tend to suffer from low enzyme activity and high levels of uncoupling (Urlacher and Girhard, 2012).

Lack of success is likely due to the difficulty in replicating elements perfected by the wild-type enzyme in synthetic constructs. Chief among these is dimerisation, which provides the structural arrangement essential for efficient electron transfer in wild-type BM3 and, as of yet, has not been artificially replicated (Munro et al., 2007).

1.3. Project Aims

In summary, P450 BM3 has proven to be a highly versatile platform for bioengineering – with variants capable of producing drug metabolites analogous to those produced by human P450s, through to catalysing reactions previously only possible via synthetic chemistry. The work presented in this thesis concerns the characterisation and further development of a collection of novel BM3 variants, all of which display altered regio- and enantioselective oxidative activity relative to the wild-type enzyme. From the basis of a single, novel point mutation located within the I-helix motif (I263P), additional point mutations were introduced in a sequential manner with the aim of producing oxidised products of value to industry. Variant enzyme activity was characterised using binding titrations and steady-state kinetics. Variant product profiles, in addition to those of the wild-type enzyme, were investigated using GC-MS. Structures of the variants were examined using a variety of biophysical techniques, including EPR, Resonance Raman and X-ray crystallography.

The second goal of this thesis was to apply the high-throughput screening technique of native-MS to the "gatekeeper" (F87V/A82F) variant of BM3. The development of orthogonal high-throughput screening techniques is essential to accelerating the characterisation, optimisation and engineering of BM3 variants with desirable properties. An optimised high-throughput compound screen performed using native-MS will enable the binding activity of the gatekeeper variant with thousands of FDA-approved drug compounds to be assessed with much greater efficiency than alternative techniques. This will ideally yield a collection of new substrates and substrate classes, some of which may have been overlooked by other screening methods, for further investigation with the gatekeeper variant. Finally, prospective substrates will be subjected to turnover studies. Subsequent metabolite characterisation via LC-MS and MS/MS will indicate if the drug metabolites produced by the gatekeeper enzyme are also those produced by human P450s.

1.4. References

Acevedo-Rocha, C.G., Gamble, C.G., Lonsdale, R., Li, A., Nett, N., Hoebenreich, S., Lingnau, J.B., Wirtz, C., Fares, C., Hinrichs, H. and Deege, A., 2018. P450-catalyzed regio-and diastereoselective steroid hydroxylation: efficient directed evolution enabled by mutability landscaping. *ACS Catalysis*, 8, 3395-3410.

Andersen, J. F., Tatsuta, K., Gunji, H., Ishiyama, T. and Hutchinson, C. R. 1993. Substrate specificity of 6-deoxyerythronolide B hydroxylase, a bacterial cytochrome P450 of erythromycin A biosynthesis. *Biochemistry,* 32**,** 1905-13.

Appel, D., Lutz-Wahl, S., Fischer, P., Schwaneberg, U. and Schmid, R. D. 2001. A P450 BM-3 mutant hydroxylates alkanes, cycloalkanes, arenes and heteroarenes. *J Biotechnol,* 88, 167-71.

Arnold, F.H., 2018. Directed evolution: bringing new chemistry to life. *Angewandte Chemie International Edition*, 57, 4143-4148.

Behrendorff, J.B., Huang, W. and Gillam, E.M., 2015. Directed evolution of cytochrome P450 enzymes for biocatalysis: exploiting the catalytic versatility of enzymes with relaxed substrate specificity. *Biochemical Journal*, 467, 1-15.

Bloom, J. D., Labthavikul, S. T., Otey, C. R. and Arnold, F. H. 2006. Protein stability promotes evolvability. *Proc Natl Acad Sci U S A,* 103, 5869-74.

Boddupalli, S. S., Estabrook, R. W. and Peterson, J. A. 1990. Fatty acid monooxygenation by cytochrome P-450BM-3. *J Biol Chem,* 265, 4233-9.

Bowman, S. E. and Bren, K. L. 2008. The chemistry and biochemistry of heme c: functional bases for covalent attachment. *Nat Prod Rep,* 25, 1118-30.

Brown, C. M., Reisfeld, B. and Mayeno, A. N. 2008. Cytochromes P450: a structure-based summary of biotransformations using representative substrates. *Drug Metab Rev,* 40, 1-100.

Bud, R., 1994. *The uses of life: a history of biotechnology*. Cambridge University Press.

Butler, C. F., Peet, C., Mason, A. E., Voice, M. W., Leys, D. and Munro, A. W. 2013. Key mutations alter the cytochrome P450 BM3 conformational landscape and remove inherent substrate bias. *J Biol Chem,* 288, 25387-99.

Butler, C. F., Peet, C., Mclean, K. J., Baynham, M. T., Blankley, R. T., Fisher, K., Rigby, S. E., Leys, D., Voice, M. W. and Munro, A. W. 2014. Human P450-like oxidation of diverse proton pump inhibitor drugs by 'gatekeeper' mutants of flavocytochrome P450 BM3. *Biochem J,* 460, 247-59.

Carmichael, A. B. and Wong, L. L. 2001. Protein engineering of *Bacillus megaterium* CYP102. The oxidation of polycyclic aromatic hydrocarbons. *Eur J Biochem,* 268, 3117-25.

Capoferri, L., Leth, R., Ter Haar, E., Mohanty, A.K., Grootenhuis, P.D., Vottero, E., Commandeur, J.N., Vermeulen, N.P., Jørgensen, F.S., Olsen, L. and Geerke, D.P., 2016. Insights into regioselective metabolism of mefenamic acid by cytochrome P 450 BM 3 mutants through crystallography, docking, molecular dynamics, and free energy calculations. *Proteins: Structure, Function, and Bioinformatics*, 84, 383-396.

Chiang, C. H., Ramu, R., Tu, Y. J., Yang, C. L., Ng, K. Y., Luo, W. I., Chen, C. H., Lu, Y. Y., Liu, C. L. and Yu, S. S. 2013. Regioselective hydroxylation of C(12)-C(15) fatty acids with fluorinated substituents by cytochrome P450 BM3. *Chemistry,* 19, 13680-91.

Cirino, P. C. & Arnold, F. H. 2003. A self-sufficient peroxide-driven hydroxylation biocatalyst. *Angew Chem Int Ed Engl,* 42, 3299-301.

Coelho, P. S., Brustad, E. M., Kannan, A. and Arnold, F. H. 2013. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science,* 339, 307-10.

Cohen, S. N., Chang, A. C., Boyer, H. W. and Helling, R. B. 1973. Construction of biologically functional bacterial plasmids in vitro. *Proc Natl Acad Sci U S A,* 70, 3240-4.

Cooper, D. Y., Estabrook, R. W. and Rosenthal, O. 1963. The stoichiometry of C21 hydroxylation of steroids by adrenocortical microsomes. *J Biol Chem,* 238, 1320-3.

Cowart, L. A., Falck, J. R. And Capdevila, J. H. 2001. Structural determinants of active site binding affinity and metabolism by cytochrome P450 BM-3. *Arch Biochem Biophys,* 387, 117-24.

Cryle, M. J., Espinoza, R. D., Smith, S. J., Matovic, N. J. and DE VOSS, J. J. 2006. Are branched chain fatty acids the natural substrates for P450(BM3)? *Chem Commun (Camb)*, 2353-5.

Daff, S. N., Chapman, S. K., Turner, K. L., Holt, R. A., Govindaraj, S., Poulos, T. L. And Munro, A. W. 1997. Redox control of the catalytic cycle of flavocytochrome P-450 BM3. *Biochemistry,* 36, 13816-23.

Ortiz De Montellano, P. R. ed., 2005. *Cytochrome P450: structure, mechanism, and biochemistry*. Springer Science & Business Media.

De Mot, R. And Parret, A. H. 2002. A novel class of self-sufficient cytochrome P450 monooxygenases in prokaryotes. *Trends Microbiol,* 10, 502-8.

Deeni, Y. Y., Paine, M. J., Ayrton, A. D., Clarke, S. E., Chenery, R. And Wolf, C. R. 2001. Expression, purification, and biochemical characterization of a human cytochrome P450 CYP2D6-NADPH cytochrome P450 reductase fusion protein. *Arch Biochem Biophys,* 396, 16-24.

Degtyarenko, K. N. And Archakov, A. I. 1993. Molecular evolution of P450 superfamily and P450-containing monooxygenase systems. *FEBS Lett,* 332, 1-8.

Denisov, I. G., Makris, T. M., Sligar, S. G. And Schlichting, I. 2005. Structure and chemistry of cytochrome P450. *Chem Rev,* 105, 2253-77.

Di Nardo, G., Fantuzzi, A., Sideri, A., Panicco, P., Sassone, C., Giunta, C. And Gilardi, G. 2007. Wild-type CYP102A1 as a biocatalyst: turnover of drugs usually metabolised by human liver enzymes. *J Biol Inorg Chem,* 12, 313-23.

Di Nardo, G. And Gilardi, G. 2012. Optimization of the bacterial cytochrome P450 BM3 system for the production of human drug metabolites. *Int J Mol Sci,* 13, 15901-24.

Dietrich, M., Do, T. A., Schmid, R. D., Pleiss, J. And Urlacher, V. B. 2009. Altering the regioselectivity of the subterminal fatty acid hydroxylase P450 BM-3 towards gamma- and delta-positions. *J Biotechnol,* 139, 115-7.

English, N., Palmer, C. N., Alworth, W. L., Kang, L., Hughes, V. And Wolf, C. R. 1997. Fatty acid signals in *Bacillus megaterium* are attenuated by cytochrome P-450-mediated hydroxylation. *Biochem J,* 327 (Pt 2), 363-8.

Estabrook, R. W. 2003. A passion for P450s (remembrances of the early history of research on cytochrome P450). *Drug Metab Dispos,* 31, 1461-73.

Estabrook, R. W., Cooper, D. Y. And Rosenthal, O. 1963. The light reversible carbon monoxide inhibition of the steroid C21-hydroxylase system of the adrenal cortex. *Biochem Z,* 338, 741-55.

Faller, M., Matsunaga, M., Yin, S., Loo, J. A. And Guo, F. 2007. Heme is involved in microRNA processing. *Nat Struct Mol Biol,* 14, 23-9.

Fasan, R., 2012. Tuning P450 enzymes as oxidation catalysts. *ACS Catalysis*, *2*(4), pp.647-66.

Fulco, A. J. 1991. P450BM-3 and other inducible bacterial P450 cytochromes: biochemistry and regulation. *Annu Rev Pharmacol Toxicol,* 31, 177-203.

Fulco, A. J. And Miura, Y. 1974. (Omega -2) hydroxylation of fatty acids by a soluble system from *Bacillus megaterium*. *J Biol Chem,* 249, 1880-8.

Garfinkel, D. 1958. Studies on pig liver microsomes. I. Enzymic and pigment composition of different microsomal fractions. *Arch Biochem Biophys,* 77, 493-509.

Gibney, B. R., Isogai, Y., Rabanal, F., Reddy, K. S., Grosset, A. M., Moser, C. C. And Dutton, P. L. 2000. Self-assembly of heme A and heme B in a designed four-helix bundle: implications for a cytochrome *c* oxidase maquette. *Biochemistry,* 39, 11041-9.

Gillam, E. M. And Guengerich, F. P. 2001. Exploiting the versatility of human cytochrome P450 enzymes: the promise of blue roses from biotechnology. *IUBMB Life,* 52, 271-7.

Girvan, H.M., Seward, H.E., Toogood, H.S., Cheesman, M.R., Leys, D. And Munro, A.W. 2007. Structural and spectroscopic characterization of P450 BM3 mutants with unprecedented P450 heme iron ligand sets. New heme ligation states influence conformational equilibria in P450 BM3. *J Biol Chem,* 282, 564-72.

Girvan, H. M., Dunford, A. J., Neeli, R., Ekanem, I. S., Waltham, T. N., Joyce, M. G., Leys, D., Curtis, R. A., Williams, P., Fisher, K., Voice, M. W. And Munro, A. W. 2011. Flavocytochrome P450 BM3 mutant W1046A is a NADH-dependent fatty acid hydroxylase: implications for the mechanism of electron transfer in the P450 BM3 dimer. *Arch Biochem Biophys,* 507, 75-85.

Govindaraj, S. And Poulos, T. L. 1996. Probing the structure of the linker connecting the reductase and heme domains of cytochrome P450BM-3 using site-directed mutagenesis. *Protein Sci,* 5, 1389-93.

Graham, S. E. And Peterson, J. A. 1999. How similar are P450s and what can their differences teach us? *Arch Biochem Biophys,* 369, 24-29.

Guengerich, F. P. 2001. Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem Res Toxicol,* 14, 611-50.

Guengerich, F. P. 2002. Cytochrome P450 enzymes in the generation of commercial products. *Nat Rev Drug Discov,* 1, 359-66.

Guengerich, F. P. 2007. Mechanisms of cytochrome P450 substrate oxidation: MiniReview. *J Biochem Mol Toxicol,* 21**,** 163-8.

Guengerich, F. P. 2013. New trends in cytochrome P450 research at the half-century mark. *J Biol Chem,* 288, 17063-4.

Guengerich, F. P. 2018. Mechanisms of Cytochrome P450-Catalyzed Oxidations. *ACS Catal,* 8**,** 10964-76.

Guengerich, F.P., 2019. Cytochrome P450 research and The Journal of Biological Chemistry*. J Biol Chem*, *294*, 1671-80.

Guengerich, F. P., Martin, M. V., Sohl, C. D. And Cheng, Q. 2009. Measurement of cytochrome P450 and NADPH-cytochrome P450 reductase. *Nat Protoc,* 4, 1245-51.

Guengerich, F. P. And Munro, A. W. 2013. Unusual cytochrome P450 enzymes and reactions. *J Biol Chem,* 288, 17065-73.

Guengerich, F. P., Wu, Z. L. And Bartleson, C. J. 2005. Function of human cytochrome P450s: characterization of the orphans. *Biochem Biophys Res Commun,* 338, 465-9.

Guengerich, F. P., Yun, C. H. & Macdonald, T. L. 1996. Evidence for a 1-electron oxidation mechanism in N-dealkylation of N,N-dialkylanilines by cytochrome P450 2B1. Kinetic hydrogen isotope effects, linear free energy relationships, comparisons with horseradish peroxidase, and studies with oxygen surrogates. *J Biol Chem,* 271, 27321-9.

Guroff, G., Daly, J. W., Jerina, D. M., Renson, J., Witkop, B. And Udenfriend, S. 1967. Hydroxylation-induced migration: the NIH shift. Recent experiments reveal an unexpected and general result of enzymatic hydroxylation of aromatic compounds. *Science,* 157, 1524-30.

Haines, D. C. 2006. A role for the strained phenylalanine ring rotation induced by substrate binding to cytochrome CYP102A1. *Protein Pept Lett,* 13, 977-80.

Hamdane, D., Zhang, H. And Hollenberg, P. 2008. Oxygen activation by cytochrome P450 monooxygenase. *Photosynth Res,* 98, 657-66.

Hargrove, T. Y., Friggeri, L., Wawrzak, Z., Qi, A., Hoekstra, W. J., Schotzinger, R. J., York, J. D., Guengerich, F. P. And Lepesheva, G. I. 2017. Structural analyses of *Candida albicans* sterol 14α-demethylase complexed with azole drugs address the molecular basis of azole-mediated inhibition of fungal sterol biosynthesis. *J Biol Chem*, 292, 6728-43.

Hasemann, C. A., Kurumbail, R. G., Boddupalli, S. S., Peterson, J. A. And Deisenhofer, J. 1995. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure,* 3, 41-62.

Iribarne, C., Picart, D., Dréano, Y., Bail, J. P. And Berthou, F. 1997. Involvement of cytochrome P450 3A4 in N-dealkylation of buprenorphine in human liver microsomes. *Life Sci,* 60, 1953-64.

Jamakhandi, A. P., Jeffus, B. C., Dass, V. R. And Miller, G. P. 2005. Thermal inactivation of the reductase domain of cytochrome P450 BM3. *Arch Biochem Biophys,* 439, 165-74.

Jeffreys, L. N., Poddar, H., Golovanova, M., Levy, C. W., Girvan, H. M., Mclean, K. J., Voice, M. W., Leys, D. And Munro, A. W. 2019. Novel insights into P450 BM3 interactions with FDA-approved antifungal azole drugs. *Sci Rep,* 9, 1577.

Joyce, M. G., Ekanem, I. S., Roitel, O., Dunford, A. J., Neeli, R., Girvan, H. M., Baker, G. J., Curtis, R. A., Munro, A. W. And Leys, D. 2012. The crystal structure of the FAD/NADPH-binding domain of flavocytochrome P450 BM3. *FEBS J,* 279, 1694-706.

Julsing, M. K., Cornelissen, S., Bühler, B. And Schmid, A. 2008. Heme-iron oxygenases: powerful industrial biocatalysts? *Curr Opin Chem Biol,* 12, 177-86.

Kaluzna, I., Schmitges, T., Straatman, H., van Tegelen, D., Müller, M., Schürmann, M. and Mink, D., 2016. Enabling selective and sustainable P450 oxygenation technology. Production of 4-hydroxy-α-isophorone on kilogram scale. *Organic Process Research & Development*, 20, 814-819.

Kaneda, T. 1991. Iso- And anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. *Microbiol Rev,* 55, 288-302.

Katagiri, M., Ganguli, B. N. And Gunsalus, I. C. 1968. A soluble cytochrome P-450 functional in methylene hydroxylation. *J Biol Chem,* 243, 3543-6.

Kitazume, T., Haines, D. C., Estabrook, R. W., Chen, B. And Peterson, J. A. 2007. Obligatory intermolecular electron-transfer from FAD to FMN in dimeric P450BM-3. *Biochemistry,* 46, 11892-901.

Klingenberg, M. 1958. Pigments of rat liver microsomes. *Arch Biochem Biophys,* 75, 376-86.

Kumar, D., De Visser, S. P. And Shaik, S. 2005. Multistate reactivity in styrene epoxidation by compound I of cytochrome P450: mechanisms of products and side products formation. *Chemistry,* 11, 2825-35.

Kuper, J., Tee, K. L., Wilmanns, M., Roccatano, D., Schwaneberg, U. And Wong, T. S. 2012. The role of active-site Phe87 in modulating the organic co-solvent tolerance of cytochrome P450 BM3 monooxygenase. *Acta Crystallogr Sect F Struct Biol Cryst Commun,* 68, 1013-7.

Lamb, D. C. And Waterman, M. R. 2013. Unusual properties of the cytochrome P450 superfamily. *Philos Trans R Soc Lond B Biol Sci,* 368, 20120434.

Le-Huu, P., Heidt, T., Claasen, B., Laschat, S. and Urlacher, V.B., 2015. Chemo-, regio-, and stereoselective oxidation of the monocyclic diterpenoid β-cembrenediol by P450 BM3. *ACS Catalysis*, 5, 1772-1780.

Lewis, D. F. And Hlavica, P. 2000. Interactions between redox partners in various cytochrome P450 systems: functional and structural aspects. *Biochim Biophys Acta,* 1460, 353-74.

Leys, D., Mowat, C. G., Mclean, K. J., Richmond, A., Chapman, S. K., Walkinshaw, M. D. And Munro, A. W. 2003. Atomic structure of *Mycobacterium tuberculosis* CYP121 to 1.06 Å reveals novel features of cytochrome P450. *J Biol Chem,* 278, 5141-7.

Li, Q. S., Schwaneberg, U., Fischer, P. And Schmid, R. D. 2000. Directed evolution of the fatty-acid hydroxylase P450 BM-3 into an indole-hydroxylating catalyst. *Chemistry,* 6, 1531-6.

Li, T., Bonkovsky, H. L. And Guo, J. T. 2011. Structural analysis of heme proteins: implications for design and prediction. *BMC Struct Biol,* 11, 13.

Li, Y., Qin, B., Li, X., Tang, J., Chen, Y., Zhou, L. and You, S., 2018. Selective oxidations of cyperenoic acid by slightly reshaping the binding pocket of cytochrome P450 BM3. ChemCatChem, 10, 559-565.

Liebler, D. C. And Guengerich, F. P. 1983. Olefin oxidation by cytochrome P-450: evidence for group migration in catalytic intermediates formed with vinylidene chloride and trans-1-phenyl-1-butene. *Biochemistry,* 22, 5482-9.

Lu, A. Y. And Coon, M. J. 1968. Role of hemoprotein P-450 in fatty acid omega-hydroxylation in a soluble enzyme system from liver microsomes. *J Biol Chem,* 243, 1331-2.

Lundemo, M.T. and Woodley, J.M., 2015. Guidelines for development and implementation of biocatalytic P450 processes. *Applied microbiology and biotechnology*, 99, 2465-2483.

Lussenburg, B. M., Babel, L. C., Vermeulen, N. P. And Commandeur, J. N. 2005. Evaluation of alkoxyresorufins as fluorescent substrates for cytochrome P450 BM3 and site-directed mutants. *Anal Biochem,* 341, 148-55.

Luthra, A., Denisov, I. G. And Sligar, S. G. 2011. Spectroscopic features of cytochrome P450 reaction intermediates. *Arch Biochem Biophys,* 507, 26-35.

Maves, S. A., Yeom, H., Mclean, M. A. And Sligar, S. G. 1997. Decreased substrate affinity upon alteration of the substrate-docking region in cytochrome P450(BM-3). *FEBS Lett,* 414, 213-8.

Mcintosh, J. A., Coelho, P. S., Farwell, C. C., Wang, Z. J., Lewis, J. C., Brown, T. R. And Arnold, F. H. 2013. Enantioselective intramolecular C-H amination catalyzed by engineered cytochrome P450 enzymes in vitro and in vivo. *Angew Chem Int Ed Engl,* 52, 9309-12.

Mclean, K. J., Girvan, H. M. And Munro, A. W. 2007. Cytochrome P450/redox partner fusion enzymes: biotechnological and toxicological prospects. *Expert Opin Drug Metab Toxicol,* 3, 847-63.

Mclean, K. J., Hans, M., Meijrink, B., Van Scheppingen, W. B., Vollebregt, A., Tee, K. L., Van Der Laan, J. M., Leys, D., Munro, A. W. And Van Den Berg, M. A. 2015. Single-step fermentative production of the cholesterol-lowering drug pravastatin via reprogramming of *Penicillium chrysogenum*. *Proc Natl Acad Sci U S A,* 112, 2847-52.

Mcnulty, J., Nair, J. J., Vurgun, N., Difrancesco, B. R., Brown, C. E., Tsoi, B., Crankshaw, D. J. And Holloway, A. C. 2012. Discovery of a novel class of aldol-derived 1,2,3-triazoles: potent and selective inhibitors of human cytochrome P450 19A1 (aromatase). *Bioorg Med Chem Lett,* 22, 718-22.

Mestres, J. 2005. Structure conservation in cytochromes P450. *Proteins,* 58, 596-609.

Meunier, B., De Visser, S. P. And Shaik, S. 2004. Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. *Chem Rev,* 104, 3947-80.

Miura, Y. And Fulco, A. J. 1974. (Omega -2) hydroxylation of fatty acids by a soluble system from *Bacillus megaterium*. *J Biol Chem,* 249, 1880-8.

Mueller, G. C. And Miller, J. A. 1949. The reductive cleavage of 4-dimethylaminoazobenzene by rat liver; the intracellular distribution of the enzyme system and its requirement for triphosphopyridine nucleotide. *J Biol Chem,* 180, 1125-36.

Munro, A. W., Daff, S., Coggins, J. R., Lindsay, J. G. And Chapman, S. K. 1996. Probing electron transfer in flavocytochrome P-450 BM3 and its component domains. *Eur J Biochem,* 239, 403-9.

Munro, A. W., Girvan, H. M. And Mclean, K. J. 2007a. Cytochrome P450-redox partner fusion enzymes. *Biochim Biophys Acta,* 1770, 345-59.

Munro, A. W., Girvan, H. M. And Mclean, K. J. 2007b. Variations on a (t)heme - novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily. *Nat Prod Rep,* 24, 585-609.

Munro, A. W., Leys, D. G., Mclean, K. J., Marshall, K. R., Ost, T. W., Daff, S., Miles, C. S., Chapman, S. K., Lysek, D. A., Moser, C. C., Page, C. C. And Dutton, P. L. 2002. P450 BM3: the very model of a modern flavocytochrome. *Trends Biochem Sci,* 27, 250-7.

Munro, A. W., Noble, M. A., Miles, C. S., Daff, S. N., Green, A. J., Quaroni, L., Rivers, S., Ost, T. W., Reid, G. A. And Chapman, S. K. 1999. Flavocytochrome P-450 BM3: a paradigm for the analysis of electron transfer and its control in the P-450s. *Biochem Soc Trans,* 27, 190-6.

Nakayama, N., Takemae, A. And Shoun, H. 1996. Cytochrome P450foxy, a catalytically self-sufficient fatty acid hydroxylase of the fungus *Fusarium oxysporum*. *J Biochem,* 119, 435-40.

Narhi, L. O. And Fulco, A. J. 1982. Phenobarbital induction of a soluble cytochrome P-450-dependent fatty acid monooxygenase in *Bacillus megaterium*. *J Biol Chem,* 257, 2147-50.

Narhi, L. O. And Fulco, A. J. 1986. Characterization of a catalytically self-sufficient 119,000-dalton cytochrome P-450 monooxygenase induced by barbiturates in *Bacillus megaterium*. *J Biol Chem,* 261, 7160-9.

Nebert, D. W., Nelson, D. R. And Feyereisen, R. 1989. Evolution of the cytochrome P450 genes. *Xenobiotica,* 19, 1149-60.

Neeli, R., Roitel, O., Scrutton, N. S. And Munro, A. W. 2005. Switching pyridine nucleotide specificity in P450 BM3: mechanistic analysis of the W1046H and W1046A enzymes. *J Biol Chem,* 280, 17634-44.

Nelson, D. R. 1999. Cytochrome P450 and the individuality of species. *Arch Biochem Biophys,* 369, 1-10.

Nelson, D. R., Koymans, L., Kamataki, T., Stegeman, J. J., Feyereisen, R., Waxman, D. J., Waterman, M. R., Gotoh, O., Coon, M. J., Estabrook, R. W., Gunsalus, I. C. And Nebert, D. W. 1996. P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics,* 6, 1-42.

Nelson, D.R., 2009. The cytochrome P450 homepage. *Human genomics*, *4*(1), p.59. http://drnelson.uthsc.edu/cytochromeP450.html (Last accessed 17/06/19)

Noble, M. A., Miles, C. S., Chapman, S. K., Lysek, D. A., Mackay, A. C., Reid, G. A., Hanzlik, R. P. And Munro, A. W. 1999. Roles of key active-site residues in flavocytochrome P450 BM3. *Biochem J,* 339, 371-9.

Omura, T. And Sato, R. 1964a. THE CARBON MONOXIDE-BINDING PIGMENT OF LIVER MICROSOMES. I. EVIDENCE FOR ITS HEMOPROTEIN NATURE. *J Biol Chem,* 239**,** 2370-8.

Omura, T. And Sato, R. 1964b. THE CARBON MONOXIDE-BINDING PIGMENT OF LIVER MICROSOMES. II. SOLUBILIZATION, PURIFICATION, AND PROPERTIES. *J Biol Chem,* 239, 2379-85.

Organisation For Economic Co-Operation And Development (OECD), 2005. A framework for biotechnology statistics. https://www.oecd.org/sti/inno/34935605.pdf (last accessed 17/06/19)

Ost, T. W., Miles, C. S., Murdoch, J., Cheung, Y., Reid, G. A., Chapman, S. K. And Munro, A. W. 2000. Rational re-design of the substrate binding site of flavocytochrome P450 BM3. *FEBS Lett,* 486, 173-7.

Otey, C. R., Bandara, G., Lalonde, J., Takahashi, K. And Arnold, F. H. 2006. Preparation of human metabolites of propranolol using laboratory-evolved bacterial cytochromes P450. *Biotechnol Bioeng,* 93, 494-9.

Peters, M. W., Meinhold, P., Glieder, A. And Arnold, F. H. 2003. Regio- and enantioselective alkane hydroxylation with engineered cytochromes P450 BM-3. *J Am Chem Soc,* 125, 13442-50.

Parra, L.P., Agudo, R. and Reetz, M.T., 2013. Directed evolution by using iterative saturation mutagenesis based on multiresidue sites. *ChemBioChem*, 14, 2301-2309.

Peterson, D.H., Murray, H.C., Eppstein, S.H., Reineke, L.M., Weintraub, A., Meister, P.D. And Leigh, H.M., 1952. Microbiological Transformations of Steroids. 1 I. Introduction of Oxygen at Carbon-11 of Progesterone. *J Am Chem Soc*, *74*, 5933-36.

Poulos, T. L., Finzel, B. C., Gunsalus, I. C., Wagner, G. C. And Kraut, J. 1985. The 2.6-Å crystal structure of *Pseudomonas putida* cytochrome P-450. *J Biol Chem,* 260, 16122-30.

Poulos, T. L., Finzel, B. C. And Howard, A. J. 1987. High-resolution crystal structure of cytochrome P450cam. *J Mol Biol,* 195, 687-700.

Ravichandran, K. G., Boddupalli, S. S., Hasermann, C. A., Peterson, J. A. And Deisenhofer, J. 1993. Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450s. *Science,* 261, 731-6.

Reedy, C. J., Elvekrog, M. M. And Gibney, B. R. 2008. Development of a heme protein structure-electrochemical function database. *Nucleic Acids Res,* 36, D307-13.

Rendic, S. And Guengerich, F. P. 2015. Survey of Human Oxidoreductases and Cytochrome P450 Enzymes Involved in the Metabolism of Xenobiotic and Natural Chemicals. *Chem Res Toxicol,* 28, 38-42.

Rittle, J. And Green, M. T. 2010. Cytochrome P450 compound I: capture, characterization, and C-H bond activation kinetics. *Science,* 330, 933-7.

Rydberg, P., Ryde, U. And Olsen, L. 2008. Sulfoxide, Sulfur, and Nitrogen Oxidation and Dealkylation by Cytochrome P450. *J Chem Theory Comput,* 4, 1369-77.

Salazar, O., Cirino, P. C. And Arnold, F. H. 2003. Thermostabilization of a cytochrome P450 peroxygenase. *Chembiochem,* 4, 891-3.

Sandhu, P., Guo, Z., Baba, T., Martin, M. V., Tukey, R. H. And Guengerich, F. P. 1994. Expression of modified human cytochrome P450 1A2 in *Escherichia coli*: stabilization, purification, spectral characterization, and catalytic activities of the enzyme. *Arch Biochem Biophys,* 309, 168-77.

Schenkman, J. B. And Jansson, I. 2006. Spectral analyses of cytochromes P450. *Methods Mol Biol,* 320, 11-18.

Schenkman, J. B., Sligar, S. G. And Cinti, D. L. 1981. Substrate interaction with cytochrome P-450. *Pharmacol Ther,* 12, 43-71.

Schewe, H., Kaup, B.A. and Schrader, J., 2008. Improvement of P450 BM-3 whole-cell biocatalysis by integrating heterologous cofactor regeneration combining glucose facilitator and dehydrogenase in E. coli. *Applied microbiology and biotechnology*, 78, 55-65.

Schiffler, B. And Bernhardt, R. 2003. Bacterial (CYP101) and mitochondrial P450 systems-how comparable are they? *Biochem Biophys Res Commun,* 312, 223-8.

Schwaneberg, U., Appel, D., Schmitt, J. And Schmid, R. D. 2000. P450 in biotechnology: zinc driven omega-hydroxylation of p-nitrophenoxydodecanoic acid using P450 BM-3 F87A as a catalyst. *J Biotechnol,* 84, 249-57.

Segall, M. 1997. Physical Properties of P450. *http://www.tcm.phy.cam.ac.uk/~mds21/thesis/node50.html* (Last Accessed 17/06/19).

Seto, Y. And Guengerich, F. P. 1993. Partitioning between N-dealkylation and N-oxygenation in the oxidation of N,N-dialkylarylamines catalyzed by cytochrome P450 2B1. *J Biol Chem,* 268, 9986-97.

Sevrioukova, I. F., Hazzard, J. T., Tollin, G. And Poulos, T. L. 2001. Laser flash induced electron transfer in P450cam monooxygenase: putidaredoxin reductase-putidaredoxin interaction. *Biochemistry,* 40, 10592-600.

Sevrioukova, I. F., Li, H., Zhang, H., Peterson, J. A. And Poulos, T. L. 1999. Structure of a cytochrome P450-redox partner electron-transfer complex. *Proc Natl Acad Sci U S A,* 96, 1863-8.

Shannon, R. T. 1976. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 32, 751-67.

SHEN, A.L. And KASPER, C.B., 1993. Protein and gene structure and regulation of NADPH-cytochrome P450 oxidoreductase. In *Cytochrome P450* (pp. 35-58). Springer, Berlin, Heidelberg.

Sowden, R. J., Yasmin, S., Rees, N. H., Bell, S. G. And Wong, L. L. 2005a. Biotransformation of the sesquiterpene (+)-valencene by cytochrome P450cam and P450BM-3. *Org Biomol Chem,* 3, 57-64.

Sun, J., Hoshino, H., Takaku, K., Nakajima, O., Muto, A., Suzuki, H., Tashiro, S., Takahashi, S., Shibahara, S., Alam, J., Taketo, M. M., Yamamoto, M. And Igarashi, K. 2002. Hemoprotein Bach1 regulates enhancer availability of heme oxygenase-1 gene. *EMBO J,* 21, 5216-24.

Syed, K. And Mashele, S. S. 2014. Comparative analysis of P450 signature motifs EXXR and CXG in the large and diverse kingdom of fungi: identification of evolutionarily conserved amino acid patterns characteristic of P450 family. *PLoS One,* 9, e95616.

Tsotsou, G.E., Cass, A.E.G. And Gilardi, G., 2002. High throughput assay for cytochrome P450 BM3 for screening libraries of substrates and combinatorial mutants. *BiosensBioelectron*, 17, 119-31.

Tran, N. H., Nguyen, D., Dwaraknath, S., Mahadevan, S., Chavez, G., Nguyen, A., Dao, T., Mullen, S., Nguyen, T. A. And Cheruzel, L. E. 2013. An efficient light-driven P450 BM3 biocatalyst. *J Am Chem Soc,* 135, 14484-7.

Urlacher, V. B. And Eiben, S. 2006. Cytochrome P450 monooxygenases: perspectives for synthetic application. *Trends Biotechnol,* 24, 324-30.

Urlacher, V. B. And Girhard, M. 2012. Cytochrome P450 monooxygenases: an update on perspectives for synthetic application. *Trends Biotechnol,* 30, 26-36.

Urlacher, V.B. and Girhard, M., 2019. Cytochrome P450 monooxygenases in biotechnology and synthetic biology. *Trends in biotechnology*.

Valnot, I., Von Kleist-Retzow, J. C., Barrientos, A., Gorbatyuk, M., Taanman, J. W., Mehaye, B., Rustin, P., Tzagoloff, A., Munnich, A. And Rötig, A. 2000. A mutation in the human heme A:farnesyltransferase gene (COX10) causes cytochrome *c* oxidase deficiency. *Hum Mol Genet,* 9, 1245-9.

Vaz, A. D., Mcginnity, D. F. And Coon, M. J. 1998. Epoxidation of olefins by cytochrome P450: evidence from site-specific mutagenesis for hydroperoxo-iron as an electrophilic oxidant. *Proc Natl Acad Sci U S A,* 95**,** 3555-60.

Waltham, T. N., Girvan, H. M., Butler, C. F., Rigby, S. R., Dunford, A. J., Holt, R. A. And Munro, A. W. 2011. Analysis of the oxidation of short chain alkynes by flavocytochrome P450 BM3. *Metallomics,* 3, 369-78.

Wang, M., Roberts, D. L., Paschke, R., Shea, T. M., Masters, B. S. And Kim, J. J. 1997. Three-dimensional structure of NADPH-cytochrome P450 reductase: prototype for FMN- and FAD-containing enzymes. *Proc Natl Acad Sci U S A,* 94, 8411-6.

Wang, Z. J., Peck, N. E., Renata, H. And Arnold, F. H. 2014. Cytochrome P450-Catalyzed Insertion of Carbenoids into N-H Bonds. *Chem Sci,* 5, 598-601.

Wei, Y., Ang, E. L. And Zhao, H. 2018. Recent developments in the application of P450 based biocatalysts. *Curr Opin Chem Biol,* 43, 1-7.

Weis, R., Winkler, M., Schittmayer, M., Kambourakis, S., Vink, M., Rozzell, J.D. And Glieder, A., 2009. A diversified library of bacterial and fungal bifunctional cytochrome P450 enzymes for drug metabolite synthesis. *Adv Syn Catal*, 351, 2140-46.

Weissman, K. 2004. Rational or random? *http://www.rsc.org/chemistryworld/Issues/2004/July/rational.asp (*Last Accessed 17/06/19).

Whitehouse, C. J., Bell, S. G. And Wong, L. L. 2012. P450(BM3) (CYP102A1): connecting the dots. *Chem Soc Rev,* 41, 1218-60.

Whitehouse, C. J., Bell, S. G., Yang, W., Yorke, J. A., Blanford, C. F., Strong, A. J., Morse, E. J., Bartlam, M., Rao, Z. And Wong, L. L. 2009. A highly active single-mutation variant of P450BM3 (CYP102A1). *Chembiochem,* 10, 1654-6.

Whitehouse, C. J., Yang, W., Yorke, J. A., Rowlatt, B. C., Strong, A. J., Blanford, C. F., Bell, S. G., Bartlam, M., Wong, L. L. And Rao, Z. 2010. Structural basis for the properties of two single-site proline mutants of CYP102A1 (P450BM3). *Chembiochem,* 11, 2549-56.

Williams, R. T. 1947 *Detoxication Mechanisms,* 1st Ed., John Wiley & Sons, Inc., London

Yano, J. K., Wester, M. R., Schoch, G. A., Griffin, K. J., Stout, C. D. And Johnson, E. F. 2004. The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-A resolution. *J Biol Chem,* 279, 38091-4.

Zhang, K., El Damaty, S. and Fasan, R., 2011. P450 fingerprinting method for rapid discovery of terpene hydroxylating P450 catalysts with diversified regioselectivity. *Journal of the American Chemical Society*, 133, 3242-3245.

Zheng, Y., Li, L., Liu, Q., Zhang, H., Cao, Y., Xian, M. And Liu, H. 2016. High-specificity synthesis of novel monomers by remodeled alcohol hydroxylase. *BMC Biotechnol,* 16, 61.

Chapter 2: Profound Alteration of P450 BM3 Regio- And Enantio-Selectivity via a Single Point-Mutation

2.1. Abstract

Cytochrome P450 BM3 (CYP102A1) is a highly efficient and soluble fatty acid hydroxylase, consisting of a natural fusion between a eukaryotic-like cytochrome P450 reductase and a bacterial fatty acid-binding P450 domain. Its amenability to engineering for novel functionality, ease of expression and rapid rate of turnover, the highest yet reported for a mono-oxygenase P450, make P450 BM3 an attractive target for industrial exploitation (Munro, et al., 2002). A highly efficient fatty acid hydroxylase, wild-type BM3 almost exclusively produces ω-1, ω-2 and ω-3 mono-hydroxylated products from medium to long chain fatty acids, with a product distribution of 37%:27%:36% in the case of dodecanoate (Cirino et al., 2002).

A disruptive point mutation (I263P) was introduced into the catalytically important and highly conserved I-helix region (Girvan, 2006) to examine its effect on substrate binding and catalysis. While the I263P mutation appeared to cause less structural disruption than anticipated, the impact of the point mutation on the oxidation profiles of unsaturated fatty acid substrates, namely dodecanoate (C12:0), tetradecanoate (C14:0) and hexadecanoate (C16:0) was profound. In addition, variant I263P displays significant alteration to the enantioselectivity of styrene oxide production, producing both *R*(+) and *S*(-) enantiomers in a near-racemic ratio. Furthermore, variant I263P demonstrated altered hydroxylation and oxidative activity towards a variety of monoterpene compounds widely used within the fragrance industry.

Variant I263P provides an ideal basis for further development of rationally designed BM3 variants with altered oxidative functionality. Additional studies of the BM3 variants featuring an I263P mutation may serve to further the understanding of how conformational properties of the I-helix regulate fatty acid binding modes and control regio- and enantio-selectivity of substrate oxidation by P450 BM3.

2.2. Introduction

Cytochrome P450 BM3 (CYP102A1) is a highly efficient and soluble fatty acid hydroxylase, consisting of a natural fusion between a eukaryotic-like cytochrome P450 reductase and a bacterial fatty acid-binding P450 domain (Munro et al., 2013). A prodigious P450 catalyst, wild-type BM3 produces predominantly ω-1, ω-2 and ω-3 hydroxylated products from medium to long chain fatty acids (Miura, et al., 1974). In addition, BM3 displays the highest catalytic turnover rate yet observed in a P450 (~17,000 min^{-1} with arachidonate), attributable to highly efficient electron transfer through the cofactors enabled by the fusion arrangement of BM3 (Munro et al., 2002). Unsurprisingly, BM3 has generated great interest in both academic and industrial circles, and has hence been the subject of extensive study and characterisation (Whitehouse et al., 2012). Research from both perspectives has focused on the rational redesign of BM3 via site directed mutagenesis and directed evolution, with the aim of generating BM3 variants capable of acting on diverse substrates to produce high-value oxidised products (Fasan 2012).

The focus of this study is the catalytically important I-helix (Figure 2.1). Spanning a length of 50 Å and consisting of residues Asp250 to Lys282 within BM3, the I-helix is a highly conserved structural motif present throughout the P450 superfamily (Ravichandran et al., 1993).

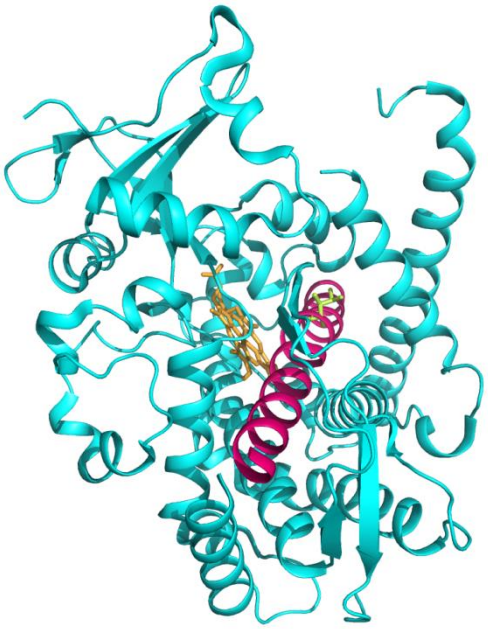

Figure 2.1 Structure of P450 BM3 wild-type heme domain in its substrate-free state.
The I-helix motif (pink), residue Ile263 (red) and the heme prosthetic group (orange) are highlighted to illustrate their positions relative to one another. Figure adapted from PDB structure 2IJ2 (Girvan et al., 2007).

The I-helix is not as highly conserved in P450 enzymes that do not require oxygen for catalysis, highlighting the importance of the I-helix in oxygen binding (Nelson, 2006). The I-helix itself contains a comprehensive network of hydrogen bonds, many of which undergo extensive rearrangement following substrate-binding (Haines et al., 2001). In the substrate-free state, the presence of a water molecule disrupts the hydrogen bond network in the region of Ile263 to Thr269. This interruption creates a kink in the structure of the I-helix, which is postulated to serve as the binding site for molecular oxygen (Whitehouse et al., 2012). This water molecule is expelled upon substrate binding, enabling hydrogen bonds to form between the carbonyl groups of Ile263, Ala264 and Gly265 to the corresponding amide groups of Glu267, Thr268 and Thr269, respectively (Ravichandran et al., 1993). This in turn prompts the angle of the structural kink to decrease, and the C-terminal region pivots away from the distal face of the heme. This rearrangement causes another water molecule, one distally-coordinated to the heme iron, to be displaced – priming the heme iron for catalysis and switching the spin-state of the iron from low- to high-spin (Haines, et al., 2001).

On account of its close proximity to the distal face of the heme (~9 Å), the I-helix is believed to influence both substrate selectivity and specificity of P450 enzymes (Munro et al., 1993).

Consequently, residues forming part of the I-helix offer attractive sites for mutagenesis studies with the aiming being to alter the substrate selectivity of wild-type BM3.

Residue Ile263 (I263), a key I-helix residue implicated in the hydrogen bonding network of the I-helix, was selected for mutagenesis. Residue I263 interacts with a neighbouring glutamate residue, Glu267 (E267). Previous mutagenesis studies conducted by Yeom et al. (1997) on E267 have demonstrated that removal of this interaction enables fatty acid substrates to get closer to the heme iron prosthetic group. On account of this, E267 variants demonstrated altered substrate activity and gave novel product profiles relative to the wild-type enzyme. Therefore, it can be postulated that modifications to E267's hydrogen-bonding partner, I263, can produce similar departures from the wild-type enzyme in regards to both activity and regioselectivity. The effects of mutagenesis at position 263 have been studied previously. Dietrich and colleagues reported that variants incorporating I263A/G point mutations display altered regioselectivity towards hydroxylation of saturated fatty acid substrate lauric acid (Dietrich et al., 2009). In addition, Sowden and co-workers incorporated an I263A point mutation within a BM3 variant capable of oxidising the industrially desirable sesquiterpene compound (+)-valencene (Sowden et al., 2005). Owing to its prime location within the active site, position 263 has also been included in number of site saturation mutagenesis studies concerned with altering substrate selectivity and oxidative functionality of P450 BM3 (Whitehouse et al., 2012; Meinhold et al., 2006). However, residue I263 has not been widely studied in isolation.

Residue I263 lies in close proximity to residues Ala264 and Thr268 – both of which influence the low-to-high spin state conversion and the coupling of NADPH oxidation to product formation (Vugt-Lussenburg et al., 2007). Any alterations to I263 are therefore highly likely to disrupt both the heme environment and catalysis of the wild-type enzyme.

The cyclic amino acid proline (Pro) was selected as a substitute for Ile at position 263, and was introduced via a point-mutation using site-directed mutagenesis. The resulting BM3 variant was titled variant I263P. Pro was selected as the substitute residue due to its conformational rigidity, its high likelihood of causing disruption within the I-helix and its compact structure relative to Ile. The substitution of Ile with Pro at position 263 could ideally serve to increase the area of the active site directly above the heme iron. This will hopefully enable substrates to get closer to the heme iron, impacting both regioselective substrate hydroxylation and substrate selectivity of the variant. The presence of Pro within

the I-helix could also perturb the kink angle of the I-helix in both the substrate-free and substrate-bound states of the variant enzyme, dramatically impacting both the structure and the functionality of the protein. Modifications to the structure of the I-helix brought about by the I263P substitution will be assessed via X-ray crystallography, in tandem with EPR and Resonance Raman spectroscopy.

Alterations to regioselective mono-hydroxylation will be assessed via the comparison of product profiles displayed by the wild-type enzyme and variant I263P against medium-chain fatty acid substrates. In addition, steady-state kinetic data will be collected for both wild-type and variant I263P against the same fatty acids substrates. The role of Ile263 within enantio-selective substrate hydroxylation will also be examined via the ratio of R(+) and S(-) enantiomers of enzyme-functionalised styrene and styrene oxide, produced by the variant I263P and the wild-type enzyme. Finally, changes to BM3 substrate specificity will be assessed via the activity of variant I263P towards a small library of monoterpene compounds relative to wild-type BM3.

2.3. Methods

2.3.1. Mutagenesis and Expression of Wild-Type and Variant I263P BM3

Intact BM3 variant I263P and its corresponding heme domain were produced using site-directed mutagenesis via PCR. Mutagenesis was performed using a QuikChange Lightning site-directed mutagenesis kit (Agilent Technologies, UK). PCR was performed using a Techne ^{3}Prime thermocycler (Bibby Scientific, UK). Oligonucleotide primers were supplied by Eurofins, the sequences of which are displayed in Table 2.1.

Table 2.1 Oligonucleotide Primers for Mutagenesis

Primer name	5’ > 3’
I263P_Fwd	ttgtttcgtgtcccgcaggtaagaatgtaataatttgatagcgaatgttctc
I263P_Rev	gagaacattcgctatcaaattattacattcttacctgcgggacacgaaacaa

For the intact BM3 variant I263P, wild-type BM3 (WT) within a pET15b construct served as template DNA. In the case of the heme domains, a pET20b plasmid construct containing WT BM3 DNA served as a template for the generation of the variant heme domain, I263P. The pET15b vector itself was selected to ensure the intact WT and BM3 variant I263P were expressed with an N-terminal hexa-histidine tag for ease of purification. The pET20b vector did not contain a tag – and consequently the I263P heme domain was expressed untagged. pET15b and pET 20b constructs were transformed into highly competent Novablue cells (Novagen, UK) as per the method described in Sambrook *et al*., (1989). Plasmid DNA from

successfully transformed colonies was extracted and purified using a Qjagen MiniPrep kit. Successful mutagenesis was confirmed via Sanger sequencing, conducted by Source Bioscience (Rochdale, UK). Primers SeqPR1, SeqPR2, SeqPR3 and VecPR, in addition to T7Forward, were used for sequencing all variant BM3 DNA samples. Sequences for all primers are displayed in Table 2.2.

Table 2.2 Oligonucleotide primers used for DNA sequencing

Primer name	5' > 3'
SeqPR1	tgagccgcttgatgacgagaac
SeqPR2	tcgcaacgcttgattcac
SeqPR3	gaatttatcgcccttctgcc
VecPR	atacccacgccgaaacaag
T7Forward	taatacgactcactataggg

Constructs were transformed into ultracompetent BL21 (DE3) *E. coli* cells (Stratagene-Agilent, UK) for protein expression. Cells transformed with pET15b constructs (intact BM3) were cultured in 500 ml Terrific-broth based auto-induction medium, whereas cells transformed with pET20b constructs (BM3 heme domain) were cultured in 500 ml of Terrific Broth alone (Formedium, UK). Cell cultures were grown at 37 °C and at 200 RPM in a shaking platform incubator until OD_{600} of ~0.6 was reached in each flask, at which point gene expression in the cell cultures was induced by the addition of 100 µM of the heme cofactor precursor δ-aminolevulinic acid (δ-ALA, Sigma-Aldrich, UK). Following addition of the heme precursor δ-ALA, the temperature of the shaking incubator was lowered to 25 °C. Both the addition of the δ-ALA and the reduction in incubator temperature were done in order to improve protein folding and heme incorporation within the P450 enzymes. Cell cultures were then left to grow for a further 24 hours. After the growth period, cells were harvested by centrifugation (6,000 RPM, 10 min, 4 °C) using a JLA 8.1 rotor and a Beckman-Coulter Avanti J-26 XP centrifuge. The cell pellet containing intact BM3 protein was re-suspended in Ni-column equilibration buffer (50 mM potassium phosphate (KPi), 250 mM NaCl, 10 mM imidazole, pH 8.0). The cell pellet containing the BM3 heme domain was re-suspended in 50 mM Tris + 1 mM EDTA pH 7.2. Both types of cell pellet were stored at -20 °C.

2.3.2. Purification of Heme Domains and Intact Wild-Type and Variant I263P BM3

Both intact and heme domain cell pellets were re-suspended on ice in a mixture of buffer (Ni –column equilibration or 50 mM Tris + 1 mM EDTA pH 7.2, respectively) and DNAse (10 mg/ml) (Thermo Fisher Scientific, UK). Protease inhibitor cocktail tablets (Sigma-Aldrich,

UK) were added at a ratio of 1 x tablet per 100 ml re-suspended cell pellet in order to prevent protease-mediated protein degradation. Cells were lysed via sonication using a Bandolin Sonopuls probe (40% intensity, 10 seconds pulse on and 50 seconds pulse off) and centrifuged (19,500 RPM, 4 °C, 45 min) using a JA 25.5 rotor in a Beckman centrifuge. A 30% ammonium sulphate cut was applied to the resulting protein supernatant and the mixture was left to stir at 4 °C for 3 hours. The mixture was then centrifuged (19,500 RPM, 4 °C, 45 min) to pellet additional cell debris.

In the case of intact BM3 constructs, protein supernatant was then added to Ni-IDA resin (Ni-Sepharose, GE Healthcare, UK) equilibrated with Ni-column equilibration buffer. The supernatant containing intact BM3 was stored at 4 °C and left overnight to bind to the resin. The Ni-IDA resin batched with the intact BM3 supernatant was loaded into a column (23 cm x 5 cm) and washed with Ni-column buffer containing increasing concentrations (50 – 100 mM) of imidazole. This was performed in order to remove unbound protein contaminants from the resin. Ni-column buffer containing 200 mM imidazole was sufficient to elute the intact BM3 protein from the column resin. The purity and concentration of the eluted protein sample was then assessed by way of SDS-PAGE and UV-visible spectroscopy, respectively. Fractions which displayed sufficient purity were pooled and subjected to extensive dialysis in 50 mM KPi (pH ~7.5) in order to remove ammonium sulphate and imidazole. The resulting intact BM3 protein was concentrated to a sufficient volume using a VivaSpin ultrafiltration column (MW cut-off of 100 kDa) (Sartorius Vivascience, UK). As a final purification step, any contaminant fatty acids bound to the active site of the intact BM3 enzymes were removed by passing the samples through Lipidex 1000 resin (Perkin Elmer, UK). Following this, protein samples were flash-frozen in liquid nitrogen and stored at -80 °C.

Cell extract containing BM3 heme domain protein was extensively dialysed into 50 mM Tris + 1 mM EDTA pH 7.2. The crude heme domain protein was then loaded on to a DEAE Sepharose Fast Flow column which, as with all subsequent columns, was mounted to an ÄKTA-pure system (GE Healthcare, UK). Protein was eluted in 50 mM Tris + 1 mM EDTA against a gradient of 0-1 M KCl. Protein fractions deemed sufficiently pure were pooled for dialysis into 50 mM KPi (pH ~7.5) and applied to a HiPrep Q-Sepharose Fast Flow anion exchange column (GE Life Sciences, UK) and eluted using a 0 – 500 mM KCl gradient. Fractions deemed pure enough were pooled and dialysed into 25 mM KPi (pH 7.5) for further purification using a hydroxyapatite column with an elution gradient of 25 – 500 mM

KPi (pH ~6.5). The fractions deemed purest via SDS-PAGE and UV-Vis analysis were pooled and concentrated to a volume of <1.5 mL using VivaSpin ultrafiltration columns (MW cut-off of 30 kDa). The concentrated BM3 heme domain was then injected onto a HiLoad 16/600 Superdex 200 pg gel filtration column (GE Life Sciences, UK) and run with 25 mM KPi + 150 mM NaCl (pH ~7.5). Post-elution, the purest fractions were determined, pooled and concentrated (as before) and passed through Lipidex 1000 resin (Perkin Elmer, UK) for instant use in protein crystallography.

2.3.3. Calculation of Protein Concentration and CO-Binding Titrations

Concentrations of BM3 heme domains in the low-spin ferric-state were derived using the Beer-Lambert law, as described by Omura and Sato (1964) and presented in Equation 2.1.

$$A = \varepsilon c l$$

Equation 2.1 **The Beer-Lambert law for concentration calculation of light-absorbing compounds.**
A refers to absorption, ε the sample-specific extinction coefficient, c the sample concentration and l the path-length of the sample.

The Soret extinction coefficient for ferric, substrate-free P450 BM3 was determined with reference to the spectrum of the Fe^{2+}-CO form of the enzyme. Absorption was recorded at the Soret maximum of 418 nm. For wild type and variant BM3 heme domains, the extinction coefficient of ε_{418} = 95 mM^{-1} cm^{-1} was used. Protein samples of ~5 μM were prepared in 1 ml Assay Buffer (100 mM KPi, pH 7.5) and placed in a quartz cuvette. Following the recording of their UV-visible spectra between the wavelengths of 250 – 700 nm, proteins were reduced by the addition of sodium dithionite. The cuvettes were stoppered, and ~ 1 cm^3 of CO gas was bubbled into the samples via a syringe. Spectra were collected once per minute over the course of an hour until complete formation of the ferrous-CO complex was observed. Difference spectra were generated using Origin Pro software (OriginLab, USA) via subtraction of the spectra for the partially reduced species from the ferrous/CO-bound spectrum.

2.3.4. Determination of Substrate Dissociation Constants for Wild-Type and Variant I263P BM3 Heme Domains

Substrate dissociation constants (K_d values) for wild-type and variant I263P BM3 enzymes against fatty acid, styrene and monoterpene substrates were determined by way of UV-vis binding titrations. Concentrated stocks of all substrate classes were prepared in absolute ethanol (EtOH). Binding studies were performed with the corresponding BM3 heme domain in all cases. Purified heme domain was selected on account of its increased stability

in solution, ease of preparation, and absence of flavin spectral signal in comparison to the intact enzyme. It is important to state that when in solution, intact BM3 is an obligate dimer, whereas the heme domain behaves as a monomer. However, no significant discrepancies in binding constant data, cooperativity or allostery have been observed between the two preparations of wild-type BM3 when binding either substrates or inhibitors in previous experiments conducted within the Munro lab (data not shown). Samples of BM3 heme domain were prepared in a quartz cuvette (path length 1 cm^3) and diluted to a concentration of 2 – 5 µM in 1 ml of Assay Buffer (100 mM KPi, pH 7.5). The substrate of interest was then titrated into the cuvette from the concentrated stock in steps varying between 0.1-1 µl in volume. Substrates were added to the cuvette using a Hamilton syringe. Spectra were recorded within the region from 200-700 nm in the absence of substrate and following sequential additions of substrates. Titrations continued until no further observable shift was detected in the Soret spectral region. Using Origin Pro software, difference spectra were generated by subtracting the substrate-free spectrum from those spectra recorded following each step in the various substrate titrations. The maxima (peak) and minima (trough) of the difference spectra were identified in each case and the difference in absorbance between the two points (using the same wavelength pair for each titration) was plotted against substrate concentration. The resulting data were fitted using the hyperbolic Michaelis-Menten function, adapted for determination of K_d values (Equation 2.2).

$$A_{obs} = \frac{A_{max}[S]}{K_d + [S]}$$

Equation 2.2 The Michaelis-Menten function adapted for K_d determination.
A_{max} is the maximal absorbance reached at apparent saturation of the P450 with its substrate. A_{obs} is the observed absorbance change at substrate concentration S, and K_d is the dissociation constant for the substrate-enzyme complex.

However, If the K_d value was found to be <5-fold that of the protein concentration, the data were instead fitted using the (quadratic) Morrison function presented in Equation 2.3 (Morrison et al., 1969).

$$A_{obs} = A_{max} \frac{([E_T] + [S] + K_d) - \sqrt{([E_T] + [S] + K_d)^2 - 4[E_T][S]}}{2[E_T]}$$

Equation 2.3 The Morrison (quadratic) function.
The Morrison equation is used to fit data for tight-binding substrates that demonstrate a K_d value <5-fold that of the protein concentration used. A_{max} is the absorbance change at substrate saturation, A_{obs} is the observed absorbance change at ligand concentration S, E_T is the P450 concentration, and K_d is the binding constant for the substrate-enzyme complex.

2.3.5. Kinetic Measurements in the Steady-State for Wild-Type and Variant BM3 Intact Enzymes

Steady-state kinetic analyses of the WT and variant BM3 enzymes were performed using a Cary 50 UV-vis spectrophotometer (Agilent, UK) with dodecanoic (C_{12}), tetradecanoic (C_{14}) and hexadecanoic (C_{16}) fatty acid substrates. Substrate-dependent NADPH oxidation was recorded at 340 nm, using the extinction coefficient ε_{340} = 6.21 mM^{-1} cm^{-1} for determination of the reaction rates. NADPH was added at a constant, near-saturating concentration of 100 µM. Protein concentration was kept as a constant in each assay. Protein concentration varied between the ranges of 80-200 nM dependent on the assay, present in 1 ml of assay buffer (100 mM KPi pH 7.5). Samples were prepared in disposable cuvettes with path lengths of 1 cm^3 and kept at a constant temperature using a Peltier set at 30 °C across all assays. Data points for the initial rate of NADPH oxidation were taken as the average of 3 concordant repeat $\Delta A_{340\ nm}$ values and plotted against the corresponding substrate concentration. Data were fitted using the Michalis-Menten function in Origin Pro software in order to obtain the kinetic parameters of K_M and k_{cat} for every BM3 variant against each fatty acid substrate.

2.3.6. Crystallographic Studies to Obtain a Structure of BM3 Variant I263P Heme Domain in the Substrate-Bound State

The substituted fatty acid n-palmitoyl glycine (NPG) was selected as the substrate most likely to generate ligand-bound crystal structures with all BM3 variant heme domains. NPG stocks were prepared in 50 mM K_2CO_3 buffer to a concentration of 10 mM. Purified BM3 variant heme domain proteins were complexed with NPG at a near-saturating concentration of 1 mM (well in excess of the determined K_d values for each BM3 variant). Protein samples were examined via UV-Vis spectroscopy to ensure the variant BM3 heme domains were in the substrate-bound, high-spin state.

The NPG-bound proteins were applied to a selection of 96 well crystal screening plates (Molecular Dimensions, Newmarket, United Kingdom) as per the sitting drop crystallisation method. Protein concentration was set at 3 conditions of 9, 12 and 15 mg/mL across all plates. 200 nl of protein were applied to each plate using a Mosquito LCP instrument (TTP Labtech, United Kingdom) at a 1:1 ratio with the screening plate mother liquor. Following application of protein, crystal trays were stored at 4 °C and the protein was left to crystallise for a minimum period of 7 days.

Upon examination, protein crystals suitable for seeding, but not diffraction quality, were obtained across a range of screening conditions. Optimal results were obtained using the BCS screening plate (condition 1-36: 22.5% PEG smear, 0.1 M PIPES 7, 0.1 M calcium chloride dehydrate, 0.1 M magnesium chloride hexahydrate) and a protein concentration of 12 mg/mL.

Following a second round of protein purification, diffraction quality crystals were obtained by seeding (10:1 protein/seed stock volume ratio). Selected crystals were sent to the Diamond light source (Oxford, United Kingdom) for X-ray diffraction studies, performed by Prof. David Leys (University of Manchester). A resulting, 1.6 Å resolution structure of the I263P BM3 variant bound to the NPG ligand were solved using molecular replacement with the assistance of Dr Colin Levy and Dr Harshwardhan Poddar (University of Manchester, United Kingdom). The structure underwent a process of iterative refinement using the Coot (Emsley et al., 2010) and PHENIX (Adams et al., 2010) programs, in addition to the online applications PDBRedo (Joosten et al., 2009) and PDBeFold (Krissinel et al., 2004).

2.3.7. Product Analysis by GC-MS

All turnover reaction mixes contained 2 µM of enzyme in the presence of 500 µM NADPH. Fatty acid turnover reactions contained a final substrate concentration of 200 µM. In the case of styrene and monoterpene turnover reactions, the final substrate concentration was 2 mM. All turnover reactions contained a NADPH regeneration system consisting of 0.6 mM $NADP^+$, 7.76 mM glucose-6-phosphate and 1.5 units of glucose-6-phosphate dehydrogenase. All reactions were made up to a final volume of 1 mL in 25 mM KPi (pH 7.2) and were placed in a 37 °C shaking incubator. Samples were incubated for 60 min for reactions containing fatty acids and 20 hours for reactions containing styrene or monoterpenes. Functionalised products from all samples were separated from the reaction mixture by solvent extraction. A 2:1 ratio of solvent to reaction mixture was employed in all cases, using the solvents dichloromethane (fatty acid reactions) and ethyl acetate (styrene

and monoterpene reactions). In all cases, samples were dried using a few grains of magnesium sulphate. In order to differentiate between various hydroxylated fatty acid products, samples were derivatised in an Eppendorf ThermoMixer (60 °C, 60 min, 800 RPM) (Fisher Scientific, UK) using the silylating reagent N,O-*bis*(trimethylsilyl)trifluoroacetamide (BSTFA) plus 1% trichloromethyl silane (BSTFA + 1% TMCS), added to each sample at a 1:1 ratio. Post derivatisation, samples were cooled at 4 °C for 30 min. Turnover reactions containing either styrene or monoterpenes were not subjected to derivatisation.

Fatty acid-containing samples were run on a non-polar Agilent VF-5ms column (30 M x 0.25 mM x 0.25 µM) attached to a 7890 GC system coupled to a 5975 Series MSD mass spectrometer (Agilent, Cheadle, United Kingdom) with the assistance of Dr Katherine Hollywood (University of Manchester). Initial temperature was set at 50 °C, with two temperature ramps to ensure good separation of the hydroxylated fatty acid products. Ramp 1 commenced at a rate of 20 °C/min until 220 °C and ramp 2 at a rate of 2 °C/min until the final post-injection temperature of 310 °C was reached. MS data were collected at a split ratio of 10:1. Silylated fragment pairs of corresponding m/z values served as "fingerprints" – enabling the identification of hydroxylated fatty acid products using the MassHunter software (Agilent, UK). Following identification, chromatogram peaks corresponding to each mono-hydroxylated fatty acid product were integrated using the Qualitative Analysis program (Agilent, UK). The resulting peak areas of each mono-hydroxylated fatty acid product were expressed as percentages of the total peak areas corresponding to mono-hydroxylated fatty acid products.

Monoterpene samples were run on a polar Agilent DB-WAX column (30 m x 0.25 mm x 0.25 µm) mounted to a 7890-B series GC system. Resulting fragments were detected using a 7890A series MSD mass spectrometer (Agilent, Cheadle, United Kingdom). The initial column temperature was 50 °C. Product separation was achieved using two temperature ramps. Ramp 1 proceeded at a rate of 5 °C/min until 68 °C and ramp 2 at 25 °C/min until reaching the temperature maximum of 240 °C. Hydroxylated monoterpene products were identified using NIST fragment identification software (NIST, Maryland, USA). Samples containing styrene were subjected to two distinct rounds of GC analysis, done with the assistance of Dr Helen Toogood (University of Manchester). The first round was to confirm the presence of styrene epoxide product and was conducted in an identical manner to the monoterpene samples. Following the initial round of GC analysis, styrene epoxide was identified using NIST software and an internal standard.

The second round was to determine the enantiomeric ratio of the styrene epoxide product. In the second round of GC analysis, enantiomer separation was achieved using a CP Chirasil-Dex CB column (25 m x 0.25 mm x 0.25 µm) attached to a 7890A series GC (Agilent, Cheadle, United Kingdom). Initial temperature was set at 40 °C. Ramp 1 increased at a rate of 5 °C/min until 100 °C was reached. Ramp 2 increased at a rate of 2 °C/min until the maximum temperature, 150 °C, was reached. Proportions of each enantiomer were determined by corresponding internal standards. Concentrations of each enantiomer produced by each enzyme were determined by using calibration curves.

2.3.8. Temperature Dependent Unfolding Studies of Wild-Type and Variant I263P BM3 Heme Domains

The impact of the amino acid substitution at position 263 on the thermal stability of the variant I263P heme domain was investigated using an "UNcle" thermodynamic instrument (Unchained Labs, Pleasanton, USA). Wild-type and variant I263P BM3 heme domain sample stocks were prepared to a concentration of 1 mg/mL in degassed 25 mM KPi buffer (pH ~7.5). Samples were repeated in at least triplicate and run with a thermal ramp of 20 – 90 °C and a ramp rate of 1 °C /minute. The thermal stability parameter (T_m) was calculated from the first derivative of the barycentric mean (BCM) of fluorescence intensity using UNcle Software. T_m parameters were successfully collected for the variant I263P and wild-type BM3 heme domains in the substrate-free state. Samples were run with the assistance of Dr Tom Jowitt (University of Manchester).

2.3.9. Estimation of Catalytic Uncoupling via Production of Hydrogen Peroxide

Levels of uncoupled turnover were quantified by the measurement of hydrogen peroxide formation in the absence and presence of 150 µM NADPH. This was performed using horseradish peroxidase (5 U/ml) and 1 mM 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid) (ABTS) based assay. A peroxidase substrate, oxidised ABTS produces a soluble end product that can be measured spectrophotometrically at 405 nm. Uncoupling was measured in both the absence of substrate and in the presence of 200 µM lauric acid. Intact BM3 protein samples were prepared to a concentration of 2 µM in 100 mM KPi, pH 7.5 within disposable cuvettes with a path length of 1 cm^3. Uncoupling was expressed as the proportion of NADPH converted to hydrogen peroxide.

2.3.10. EPR Studies of Ligand Binding in the Wild-Type and Variant I263P BM3 Heme Domains

Heme domain protein samples were prepared to a concentration of ~300 µM in 25 mM KPi, pH 7.5, in both the absence of substrate and in the presence of absolute EtOH solvent and

dodecanoate. Substrate-bound samples were incubated overnight with a saturating concentration of ~700 μM dodecanoate, itself prepared in a solvent of absolute EtOH. Final ethanol volume did not exceed 5% v/v. X-band EPR spectra were collected by Dr Hazel Girvan using a Bruker ER-300D instrument operated at 10 K.

2.3.11. Resonance Raman Studies of Ligand Binding to Wild-Type and Variant I263P BM3 Heme Domains

Resonance Raman spectra were collected with the assistance of Ms Bethan McAvan (University of Manchester). Spectra were collected for wild-type and variant I263P BM3 heme domains in both the substrate-free and substrate (NPG, 25 μM) bound states. Samples were prepared to a protein concentration of 50 μM in 25 mM KPi, pH 7.5. Samples were subjected to irradiation at 405 nm and data were acquired with a Renishaw InVia Raman microscope. Samples were held in a 96-well plate and exposed for 100 accumulations of 1 second each at 10% laser intensity. Spectra were recorded in the range of 200 to 2000 cm^{-1}. The 405 nm laser was calibrated using silicon as a reference at 520 cm^{-1} with an x50 objective.

2.4. Results

2.4.1. Binding Dissociation Constants (K_d) for C_{10} – C_{16} Fatty Acid Substrates

Binding dissociation constants were determined for both wild-type and BM3 variant I263P against C_{10} – C_{16} fatty acid substrates, in addition to the fatty acid analogue n-palmitoyl glycine (NPG), and are displayed in Table 2.3. As with the wild-type heme domain, binding affinities for variant I263P towards saturated fatty acid substrates increases with substrate chain length. However, variant I263P consistently displays tighter binding to fatty acid substrates compared to the wild-type. In the case of dodecanoic acid (C_{12}), variant I263P displays approximately 4-fold tighter binding compared to that of the wild-type. In fatty acids containing 14 carbons or more, variant I263P consistently displays a minimum of a 5-fold tighter binding affinity relative to that of the wild-type.

Table 2.3 Binding dissociation constants for wild-type and variant I263P heme domains against fatty acid substrates.

Substrate (C_n)	Binding dissociation constant (K_d) (μM)	
	Wild-type heme domain	I263P heme domain
Decanoic acid (C_{10})	479 ± 55.9	414 ± 64.4
Undecanoic acid (C_{11})	405 ± 25.4	203 ± 12.6
Dodecanoic acid (C_{12})	108 ± 3.51	27.2 ± 1.61
Tridecanoic acid (C_{13})	62.2 ± 2.20	19.5 ± 0.710
Tetradecanoic acid (C_{14})	12.6 ± 1.02	2.33 ± 0.171
Pentadecanoic acid (C_{15})	6.84 ± 0.250	1.09 ± 0.100
Hexadecanoic acid (C_{16})	2.98 ± 0.150	0.470 ± 0.090
N-Palmitoyl Glycine (C_{18})	0.474 ± 0.0870	0.0661 ± 0.0170

Despite consistently displaying tighter binding compared to the wild-type, variant I263P does not produce a fully high spin spectral shift against any tested substrate. In the case of the fatty acid analogue NPG, both the wild-type and variant I263P display a binding constant within the nanomolar range. The addition of saturating concentrations of NPG to the wild-type heme domain produces a near complete conversion of the protein to the high-spin state (Figure 2.2). In contrast, saturating the I263P variant with NPG substrate does not produce a fully high-spin spectral shift (Figure 2.3).

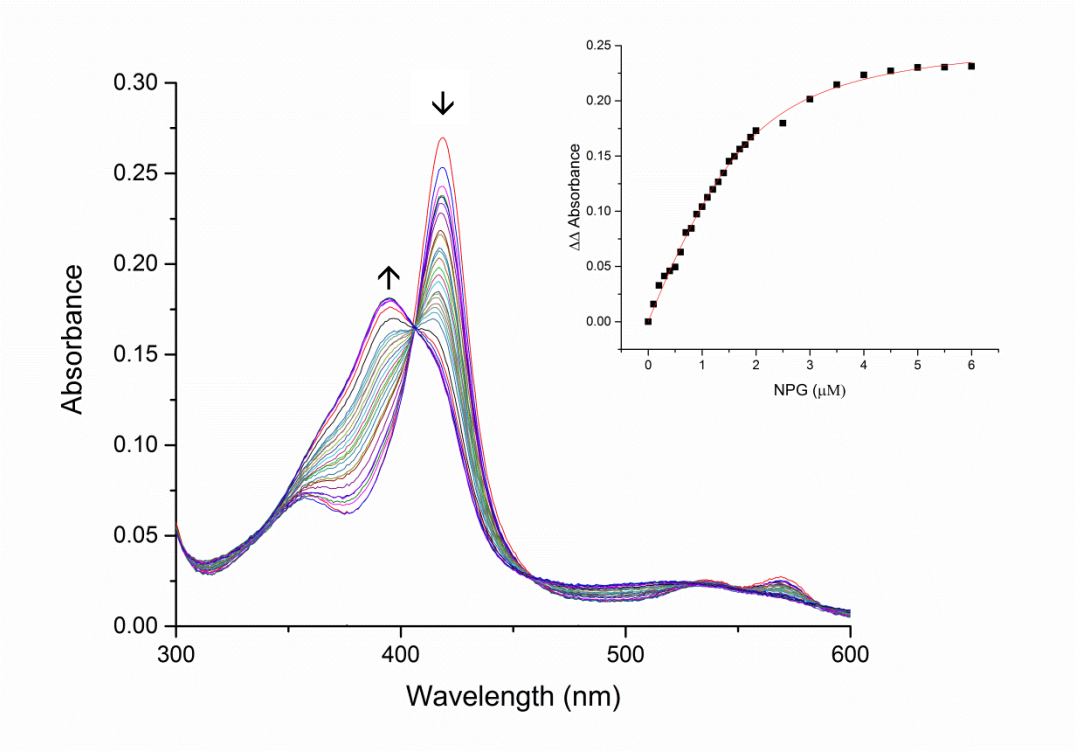

Figure 2.2 **Binding of NPG to the wild-type BM3 heme domain**

0.1 μl aliquots from a 5 mM stock of NPG substrate were titrated against a ~2.8 μM solution of wild-type BM3 heme domain. NPG was present at a near-saturating final concentration of 50 μM. A low (~418 nm) to high (~390 nm) spin-state shift is clearly visible following successive additions of NPG. Data shown within the insert were generated by subtracting trough data from peak data obtained from each of the difference spectra and by plotting the ΔΔAbsorbance data against corresponding concentrations of NPG. The rectangular hyperbola was obtained from fitting the data using the Michaelis-Menten function (Equation 2.2).

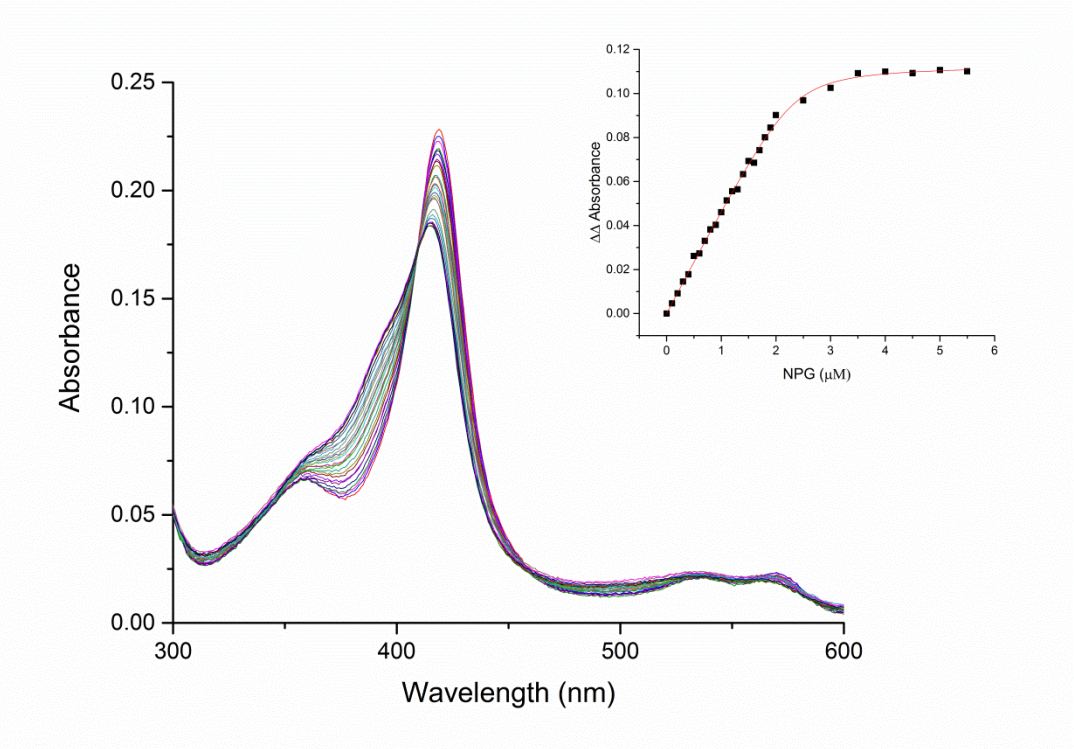

Figure 2.3 **Binding of NPG to the I263P BM3 variant heme domain**
0.1 µl aliquots from a 5 mM stock of NPG substrate were titrated against a ~2.3 µM solution of the I263P variant BM3 heme domain. The final concentration of NPG was a near-saturating 50 µM. A low (~417 nm) to high (~390 nm) spin spectral shift is occurring. Data shown within the insert were generated as outlined in Figure 2.2.

A similar effect can be observed in titrations with dodecanoic acid. While a proportion of high-spin heme is evident within the wild-type spectrum (Figure 2.4), low-spin heme species predominate within the spectra for the I263P variant (Figure 2.5). These data suggest that in BM3 variant I263P the axially ligated water molecule is not being fully displaced from the heme following substrate binding.

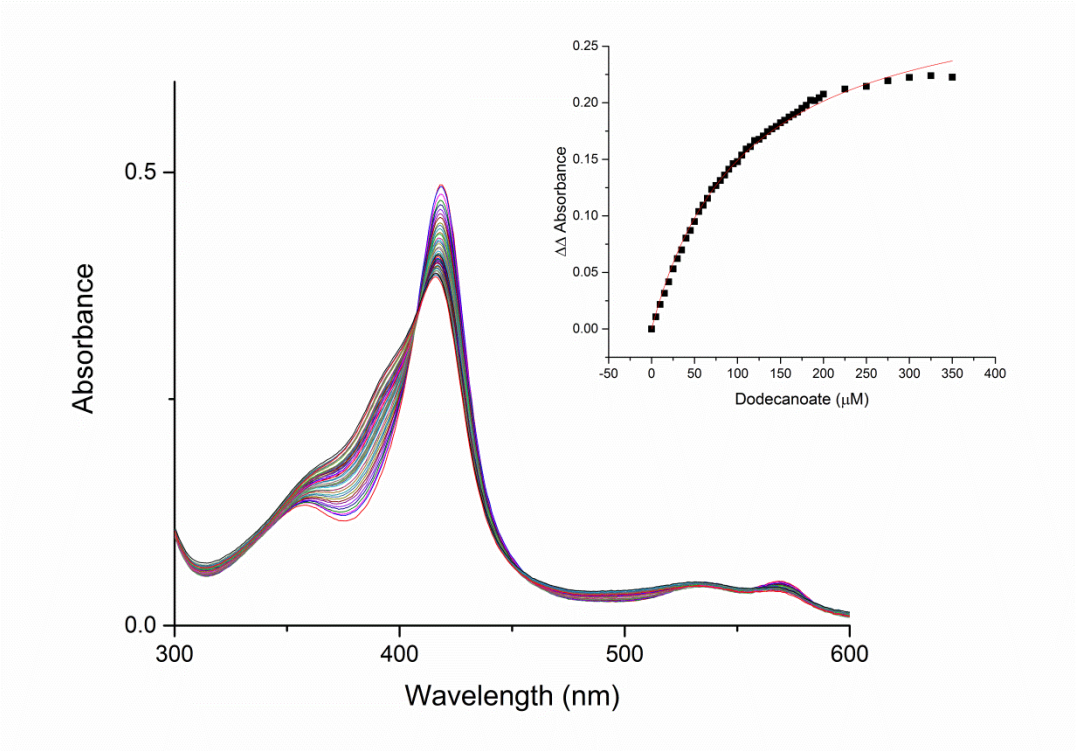

Figure 2.4 **Binding of dodecanoic acid to the wild-type BM3 heme domain**
0.1 μl aliquots from a 50 mM stock of fatty acid substrate dodecanoic acid were titrated against a ~5 μM solution of wild-type BM3 heme domain. The final concentration of dodecanoic acid was 350 μM. A significant proportion of high-spin heme is evident at ~390 nm following successive additions of substrate. Data shown within the insert were generated as outlined in Figure 2.2.

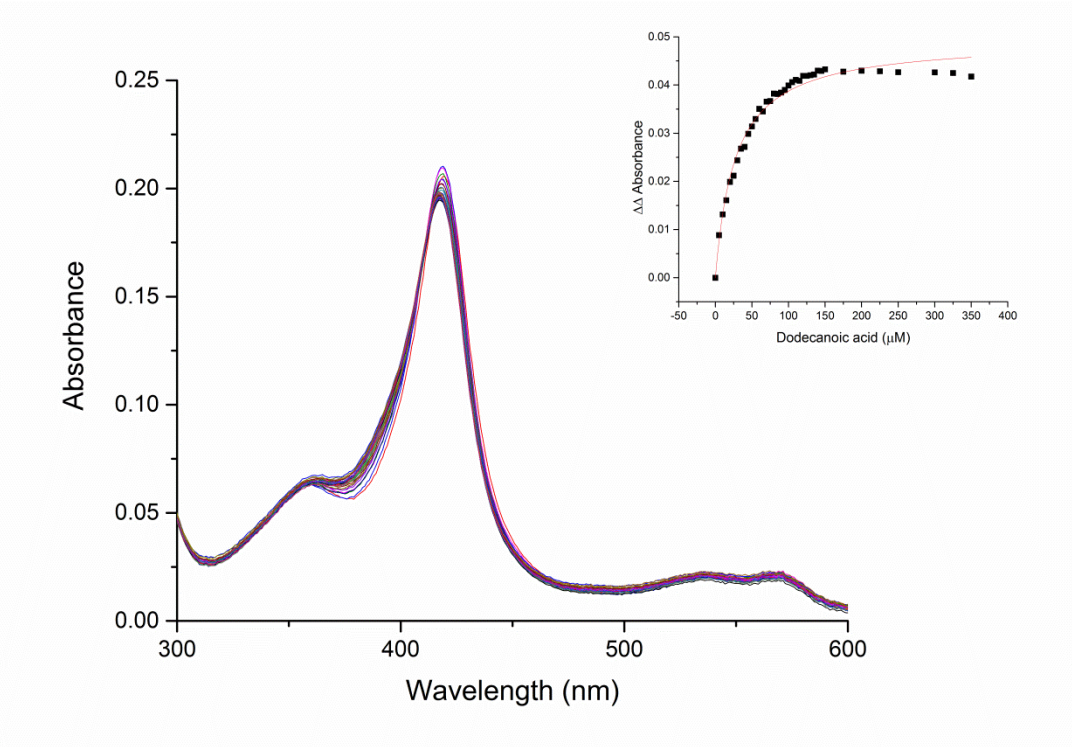

Figure 2.5 Binding of dodecanoic acid to the I263P BM3 variant heme domain.

0.1 μl aliquots from a 50 mM stock of fatty acid substrate dodecanoic acid were titrated against a ~2.5 μM solution of the I263P BM3 variant heme domain. The final concentration of dodecanoic acid was 350 μM. High-spin heme species are visible at ~390 nm. The I263P BM3 variant species in the low-spin state is evident at ~417 nm. Data shown within the insert were generated as outlined in Figure 2.2.

2.4.2. Steady-State Kinetic Data for Wild-Type and Variant I263P

Steady-state kinetic measurements were taken for intact wild-type BM3 and intact BM3 variant I263P. These studies were performed in order to assess the potentially disruptive impact of the proline substitution on the catalytic activity of intact BM3 variant I263P relative to the intact-wild-type enzyme. Substrate-dependent NADPH oxidation was recorded for both intact wild-type (Figure 2.6) and intact BM3 variant I263P (Figure 2.7) using the fatty acid substrates dodecanoic acid, tetradecanoic acid and hexadecanoic acid. The Michaelis constant (K_M), turnover number (k_{cat}) and catalytic efficiency (k_{cat}/K_M) parameters for both wild-type and variant I263P BM3 enzymes are presented in Table 2.4.

Table 2.4 **Steady-state kinetic parameters for intact wild-type and intact BM3 variant I263P enzymes with C12, C14 and C16 fatty acid substrates.**

Intact BM3 variant	Dodecanoate (C_{12})			Tetradecanoate (C_{14})			Hexadecanoate (C_{16})		
	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (μM s^{-1})	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (μM s^{-1})	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (μM s^{-1})
Wild-type	281 ± 43.9	57.6 ± 3.90	0.205	31.1 ± 3.01	63.9 ± 1.63	2.06	10.3 ± 1.40	86.2 ± 2.27	8.37
I263P	111 ± 20.3	42.1 ± 2.20	0.379	24.3 ± 3.50	53.8 ± 1.90	2.21	11.5 ± 1.80	67.4 ± 2.10	5.86

As indicated in Table 2.4, turnover number (k_{cat}) of BM3 variant I263P was decreased against all substrates when compared to the wild-type enzyme. Despite this, these data clearly suggest that the proline substitution at position 263 has not rendered the enzyme inactive. In line with the binding dissociation constant data displayed in Table 2.2, K_M values for variant I263P are lower for the dodecanoic and tetradecanoic acid substrates compared to the wild-type enzyme. The K_M value of 111 ± 20.3 μM displayed by intact I263P variant indicates a ~2.5 fold greater affinity for the dodecanoic acid substrate relative to the wild type enzyme. This mutation-induced increase in binding affinity for dodecanoate shown by the I263P variant K_M value is less pronounced in the case of tetradecanoic acid and is not observed in the case of hexadecanoic acid. The I263P BM3 variant displays marginally higher catalytic efficiencies (k_{cat}/K_M) with dodecanoic (0.379 μM s^{-1}) and tetradecanoic acids (2.21 μM s^{-1}) when compared to the wild-type values (0.205 and 2.06 μM s^{-1}, respectively). Both the I263P variant and the wild-type BM3 display the highest catalytic efficiencies (5.86 and 8.37 μM s^{-1}, respectively) for the substrate hexadecanoic acid – indicating that for both enzymes, this is the fatty acid most amenable to oxidation.

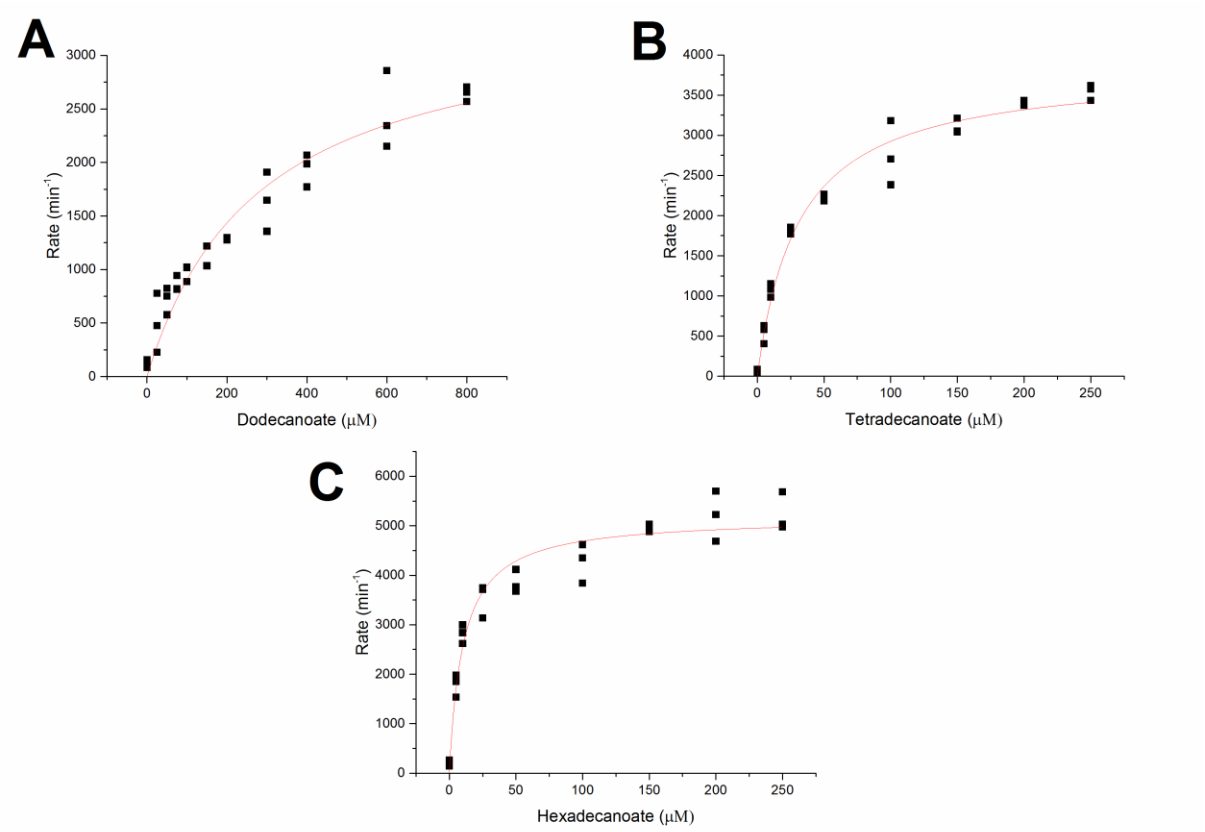

Figure 2.6 **Steady-state kinetic data for wild-type BM3-dependent NADPH oxidation with substrates dodecanoic acid (A), tetradecanoic acid (B) and hexadecanoic acid (C).**

The following kinetic parameters were obtained: A) K_M = 281 ± 43.9 μM, k_{cat} = 57.6 ± 3.90 s^{-1}, k_{cat}/K_M = 0.205 μM^{-1} s^{-1}; B) K_M = 31.1 ± 3.01 μM, k_{cat} = 63.9 ± 1.63 s^{-1}, k_{cat}/K_M = 2.06 μM^{-1} s^{-1}; C) K_M = 10.3 ± 1.40 μM, k_{cat} = 86.2 ± 2.27 s^{-1}, k_{cat}/K_M = 8.37 μM^{-1} s^{-1}. All data were fitted using the Michaelis-Menten function.

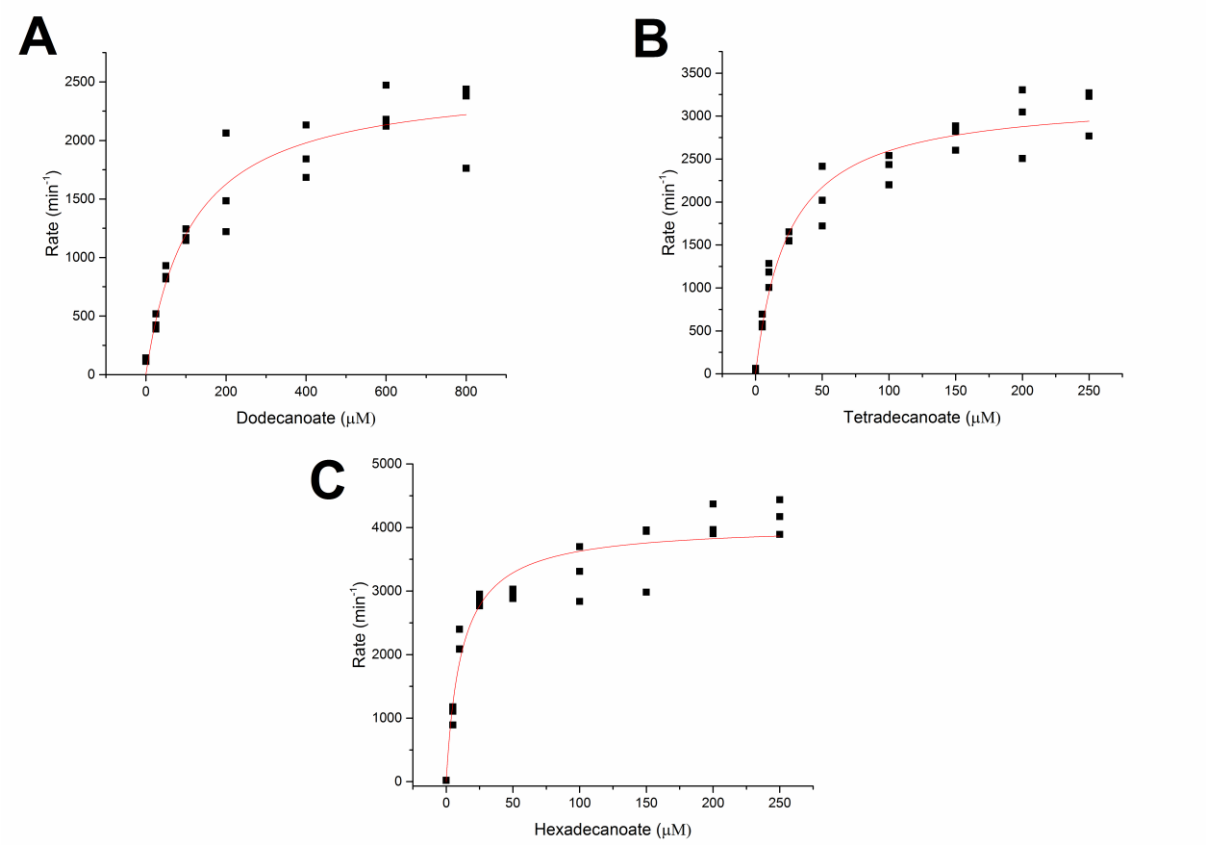

Figure 2.7 **Steady-state kinetic data for BM3 I263P variant-dependent NADPH oxidation with substrates dodecanoic acid (A), tetradecanoic acid (B) and hexadecanoic acid (C).**

The following kinetic parameters were obtained: A) K_M = 111 ± 20.3 μM, k_{cat} = 42.1 ± 2.20 s^{-1}, k_{cat}/K_M = 0.379 μM^{-1} s^{-1}; B) K_M = 24.3 ± 3.5 μM, k_{cat} = 53.8 ± 1.90 s^{-1}, k_{cat}/K_M = 2.21 μM^{-1} s^{-1}; C) K_M = 11.5 ± 1.80 μM, k_{cat} = 67.4 ± 2.10 s^{-1}, k_{cat}/K_M = 5.86 μM^{-1} s^{-1}. All data were fitted using the Michaelis-Menten function.

2.4.3. Uncoupling of NADPH Oxidation from Product Formation

With the wild-type enzyme, only 2.10% of NADPH oxidation is uncoupled from the hydroxylation of the fatty acid substrate dodecanoate (Table 2.5). In contrast, for the BM3 variant I263P, 12.9% of reducing equivalents were used to generate hydrogen peroxide (H_2O_2) in lieu of hydroxylated product. In the absence of substrate, both enzymes oxidise some NADPH and consequently produce small amounts of H_2O_2. Experiments to measure uncoupling reactions proceeding through the alternative oxidase or superoxide pathways were not conducted.

Table 2.5 Percentage of NADPH converted to H_2O_2 by uncoupling of the catalytic cycle

Intact wild-type and BM3 variant enzymes (~2 μM) were incubated with 150 μM NADPH in the absence and presence of 200 μM fatty acid substrate, dodecanoic acid.

	Wild-type	Wild-type + dodecanoate	I263P	I263P + dodecanoate
% NADPH converted to H_2O_2	1.80	2.10	3.20	12.9

2.4.4. Carbon Monoxide-Bound Spectra of Wild-Type and I263P Variant Heme Domains

Coordination states of both wild-type and BM3 variant I263P were determined by the formation of ferrous-CO adducts using CO titrations (Figure 2.8). Concentrations of both proteins were determined using the established BM3 heme domain extinction coefficient of 95 $mM^{-1}\,cm^{-1}$. Spectral properties of the I263P variant are broadly similar to those of the wild-type heme domain. Both proteins exhibit Soret maxima at ~418 nm in the substrate-free state, in addition to α and β bands present at 569 nm and 537 nm, respectively. Upon the addition of sodium dithionite, the heme prosthetic groups of both proteins are partially reduced to the ferric state and primed for the addition of CO. Subsequent addition of CO gas prompts the formation of ferrous-CO complexes within both proteins, indicated by the Soret maximum red-shifting to ~450 nm. This Soret maximum is characteristic of a P450 enzyme, and confirms that proximal coordination to the cysteine thiolate is retained in the I263P variant. However, the slight shoulder exhibited by the I263P variant at ~420 nm (Figure 2.8, right hand insert) indicates a population of ferrous-CO species in the P420 form. These P420 adducts reveal that a small proportion of the thiolate ligands in the I263P variant have become protonated, forming the P420 state.

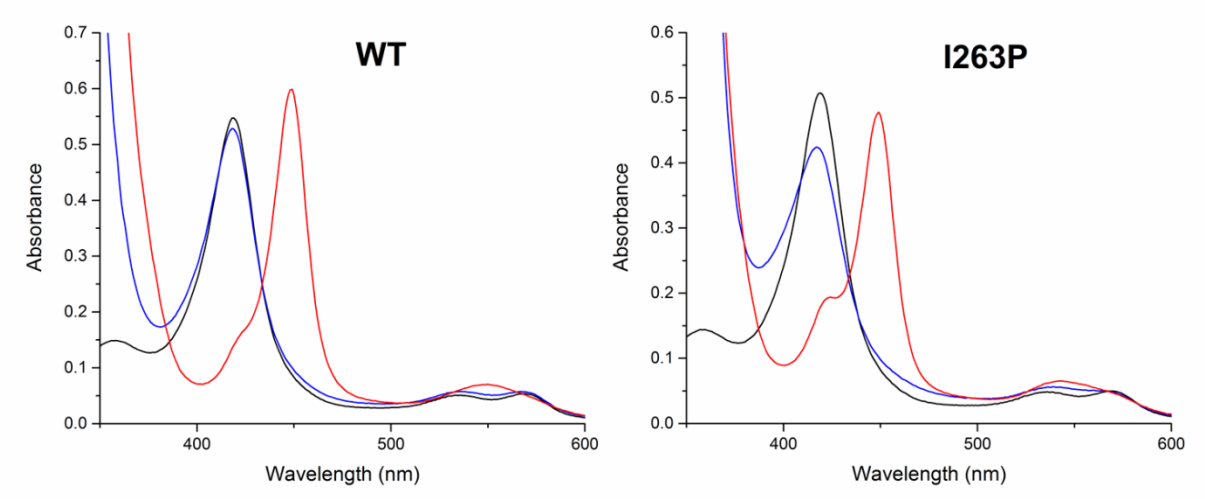

Figure 2.8 **UV-vis spectroscopic characterisation of wild-type (~6 µM, left hand insert) and I263P variant (~5 µM, right hand insert) heme domains, present in the substrate-free (black), partially reduced (blue) and the reduced/ferrous-CO (red) states.**

Spectra were collected aerobically using 100 mM KPi, pH ~7.5. In the low-spin, substrate-free, ferric state, Soret bands are at 418 nm, with α and β bands at 569 nm and 537 nm, respectively. The partially reduced, ferrous spectra have Soret bands at ~416 nm. If fully reduced, this band would be present at ~410 nm. The ferrous-CO spectra show characteristic P450 Soret peaks at 450 nm, with single bands at 550 nm corresponding to the peak of the α and β peaks. The 420 nm adduct in the ferrous-CO spectrum of I263P reveals formation of a small amount of the thiol-coordinated P420 complex (most prominent in the right hand insert).

2.4.5. Thermostability of Wild-Type and I263P Variant Heme Domains

Thermostability of the I263P variant heme domain was determined using a 20 – 90 °C temperature ramp modulated by an UNcle instrument. Melting point data (T_m) collected for both wild-type and I263P heme domains are displayed in Table 2.6.

Table 2.6 Melting point (T_m) data for wild-type and I263P variant heme domains.

Protein samples were prepared to a concentration of 1 mg/ml. All samples were substrate-free. Thermostability measurements were made using an UNcle instrument. Data were collected using a ramped temperature gradient between 20 – 90 °C at a rate of 1 °C / minute.

BM3 heme variant	Average T_m1 (°C)	Average T_m2 (°C)
Wild-type	58.4	67.1
I263P	56.0	N/A

Wild-type T_m values collected using UNcle correspond well to previous T_m data collected using DSC (58.7 °C and 64.9 °C, Munro et al., 1996). Relative to the wild-type, the I263P mutation has led to a decrease in heme stability by ~2.40 °C. In addition, the I263P variant only displays an apparent single protein unfolding event, in contrast to the two apparent T_m values displayed by the wild-type BM3 heme domain. As expected, the inclusion of a Pro residue within the I-helix has destabilised the I263P heme domain relative to the wild-type protein. However, other BM3 variants containing a single point mutation in which a comparatively small residue has been substituted by a bulkier one have displayed greater destabilisation of the heme domain relative to the wild-type form. Of particular note is the variant A82F, which produced a T_m reduction of ~6.00 °C relative to the wild-type heme domain (Butler et al., 2013). In light of this, the Pro substitution has not proven as structurally destabilising as might initially have been predicted. T_m values were calculated by recording temperature-dependent changes to sample fluorescence. The resulting curve was differentiated, with the peak of this differential corresponding to the T_m value (Figure 2.9).

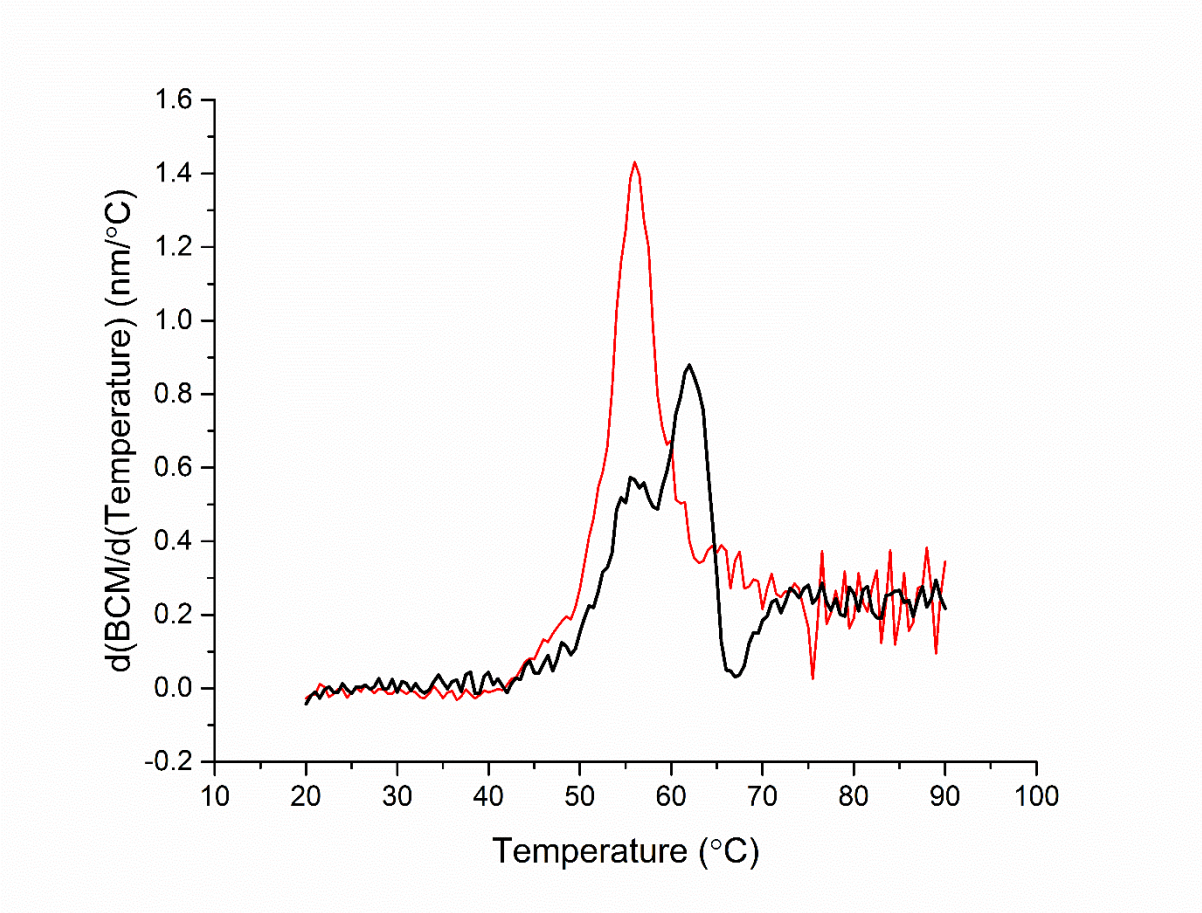

Figure 2.9 **Temperature dependent unfolding of wild-type (black) and I263P variant heme (red) domains.**
Substrate-free protein samples were prepared to a concentration of 1 mg/ml in assay buffer. Melting points (T_m) were calculated using an UNcle instrument. T_m values were determined to be 58.4 °C and 67.1 °C for wild-type and 56.0 °C for the I263P variant heme domain, respectively.

2.4.6. Resonance Raman Characterisation of Wild-Type and I263P Variant Heme Domains

Resonance Raman spectra for WT and I263P heme domains in the substrate-free and NPG-bound states are presented in Figures 2.10 and 2.11, respectively. All major features are assigned according to nomenclature presented in Spiro et al (1979). The most intense feature across all spectra is the oxidation state marker ν_4, its appearance at 1375 cm^{-1} indicating the presence of ferric heme iron in all samples. In both WT and I263P proteins, the position of the ν_3 marker band at 1502 cm^{-1} in the substrate-free spectra (Figure 2.10) is indicative of 6-coordinate, low-spin heme iron. The ν_3 species present at 1489 cm^{-1} indicates the presence of 5-coordinate, high-spin heme. In addition, 6-coordinate low-spin heme is further indicated by the presence of the ν_2 band at 1584 cm^{-1}. For the 5-coordinate, fully high-spin heme, the ν_2 band is at 1568 cm^{-1}. The ν_3 band corresponding to a high-spin

heme species is present in the NPG-bound spectra (Figure 2.11) for both the WT and I263P heme domains. However, the 1489 cm^{-1} band is much less pronounced in the I263P NPG-bound spectrum when compared to that of the WT, which displays a ν_3 high-spin feature at 1489 cm^{-1} that accounts for a significant portion of the overall ν_3 peak intensity. Other significant differences between the WT and I263P NPG-bound spectra include the features of the ν_2 band. In the WT NPG-bound spectrum, a 5-coordinate ν_2 shoulder at 1573 cm^{-1} is present alongside the 6-coordinate ν_3 parent band at 1584 cm^{-1}. This 5-coordinate ν_3 shoulder is notably less pronounced in the I263P NPG-bound spectrum, in which the 6-coordinate ν_3 peak at 1584 cm^{-1} is dominant. The vinyl stretching ($\nu_{C=C}$) vibrational mode of the NPG-bound I263P heme domains is observed at 1632 cm^{-1}, a 2 cm^{-1} shift from the 1630 cm^{-1} $\nu_{C=C}$ band displayed by the corresponding WT spectrum. The $\nu_{C=C}$ vibrational mode serves as an important structural indicator. These data, in tandem with a decrease in the $\nu_{C=C}$ band intensity, suggest that the heme vinyl groups of the I263P variant are present in a different conformation relative to the wild-type heme domain following substrate binding. In addition, the I263P variant displays a more prominent ν_{11} band at 1566 cm^{-1} in the NPG-bound state, indicating a domed heme ring caused by the presence of a distal water molecule. This, in turn, suggests that the distally ligated water molecule is not being displaced from the heme in the I263P variant on binding the NPG substrate. In the WT NPG-bound spectrum, the ν_{11} band is notably less intense, strongly implying the absence of the distal water molecule and thus a more planar heme.

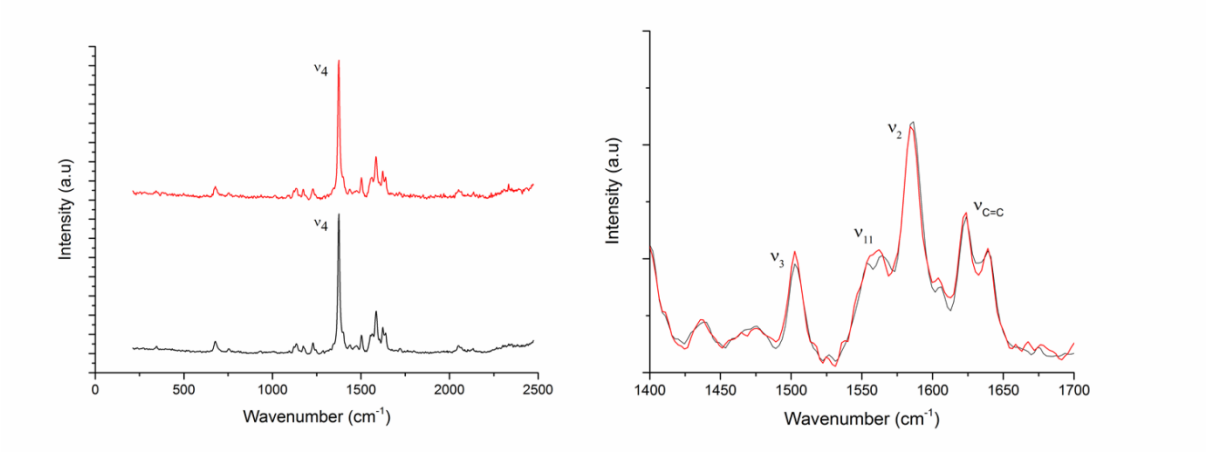

Figure 2.10 Resonance Raman spectra of substrate-free wild-type (black) and variant I263P (red) heme domains.

Protein samples were prepared to a concentration of 50 µM in 25 mM KPi, pH ~7.5. Data were collected using laser excitation at 405 nm. The left hand spectrum displays the oxidation state marker ν_4; showing that ferric heme is present in both proteins. The right hand spectrum displays bands in the high-frequency region with greater clarity. Band assignments were made according to Spiro *et al*. (1979).

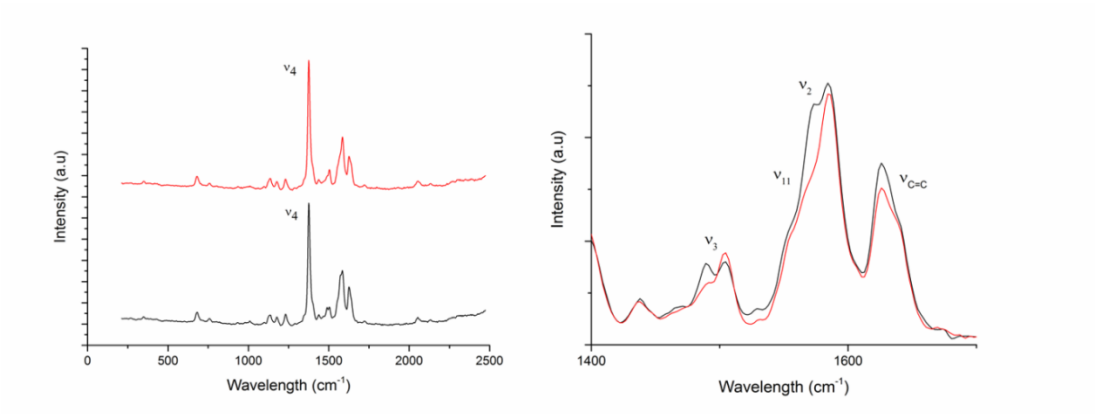

Figure 2.11 Resonance Raman spectra of NPG-bound wild-type (black) and I263P (red) heme domains.

Protein was present at a concentration of 50 µM. NPG substrate was present at a final concentration of 25 µM. The left hand spectrum displays both heme domains in the ferric oxidation state, as indicated by the ν_4 marker band. The right hand spectrum displays a close-up of the high-frequency wavelength region. Marker bands were assigned according to Spiro *et al*. (1979).

2.4.7. Alteration of Regioselective Fatty Acid Hydroxylation

The substitution of isoleucine for proline at position 263 has produced a profound effect on fatty acid oxidation profiles. Intact BM3 variant I263P produces a more diverse range of hydroxylated fatty acid compounds in addition to the products usually formed by the wild-type enzyme (typically ω-1 to ω-3 hydroxylated products). As displayed in Table 2.7, the I263P variant produces hydroxylated fatty acids in the range from ω-1 to ω-6 with the substrate dodecanoate.

Table 2.7 Product profile of dodecanoate mono-hydroxylation catalysed by wild-type and the I263P variant intact BM3 enzymes.

Samples were determined in triplicate, averaged and rounded to the nearest percentage point. Substrate concentration was 200 µM. Protein concentration was 2 µM. Turnover was performed at 37 °C for the duration of 60 min in the presence of 500 µM NADPH and a cofactor regeneration system was used as described in the Methods section.

Intact BM3 Variant	Proportion of mono-hydroxylation corresponding to ω-carbon position of dodecanoate (C:12)					
	ω-6	*ω-5*	*ω-4*	*ω-3*	*ω-2*	*ω-1*
WT	0	0	0	37	29	34
I263P	7	1	1	27	33	31

However, the vast majority (~90%) of hydroxylated products produced by variant I263P, i.e. hydroxylated dodecanoate products in the range ω-1 to ω-3, are also produced by the wild-type enzyme. In contrast, more significant departures from the product profile of the wild-type enzyme occur when variant I263P metabolises fatty acids with chain lengths greater than dodecanoate. In the case of the 14-carbon long substrate tetradecanoate (Table 2.8), variant I263P is capable of reaching the ω-7 position and generates 33% of products hydroxylated at positions different to that of wild-type BM3.

Table 2.8 **Product profile of tetradecanoate mono-hydroxylation catalysed by wild-type and the variant intact I263P BM3 enzymes.**

Samples were turned over for 60 min at 37 °C in the presence of 500 µM NADPH. Protein and substrate concentration were 2 µM and 200 µM, respectively. Percentage points were rounded to the nearest whole number. All samples were repeated in triplicate and averaged within a suitable standard deviation. A NADPH regeneration system was present as outlined in the Methods section.

Intact BM3 Variant	Proportion of mono-hydroxylation corresponding to ω-carbon position of tetradecanoate (C:14)						
	ω-7	*ω-6*	*ω-5*	*ω-4*	*ω-3*	*ω-2*	*ω-1*
WT	0	0	0	1	27	30	42
I263P	7	6	20	10	23	17	17

The trend of product profiles becoming more diverse with increasing fatty acid chain length continues with the 16-carbon chain substrate hexadecanoate (Table 2.9).

Table 2.9 **Product profile of hexadecanoate mono-hydroxylation performed by wild-type and I263P variant intact BM3 enzymes**

Samples were turned over at 37 °C for 60 min. Protein concentration was 2 µM. Substrate was present at a ~100 fold greater concentration than the protein. NADPH concentration was near-saturating at 500 µM. All samples were repeated in triplicates and averaged. Percentage points were rounded to the nearest whole number. A cofactor regeneration system was used as described in the Methods section.

Intact BM3 Variant	Proportion of mono-hydroxylation corresponding to ω-carbon position of hexadecanoate (C:16)							
	ω-8	*ω-7*	*ω-6*	*ω-5*	*ω-4*	*ω-3*	*ω-2*	*ω-1*
WT	0	0	0	0	0	13	56	31
I263P	6	13	26	18	4	6	18	9

In addition to reaching the ω-8 position, the majority of mono-hydroxylations (~66%) occur at positions unreachable by the wild-type enzyme. Moreover, the most abundant mono-hydroxylated product, ω-6 hydroxy-hexadecanoate (26%), is a novel product not produced by the wild-type enzyme. Against all tested substrates, variant I263P consistently hydroxylates fatty acids further away from the terminal ω-carbon and moving towards the α-carbon. This dramatic alteration of BM3 regioselectivity suggests that fatty acid substrates bound to variant I263P occupy the active site in a different conformation relative to that of the wild type enzyme.

GC-MS chromatograms for the intact wild-type and intact variant I263P enzymes against fatty acids dodecanoate, tetradecanoate and hexadecanoate are displayed within the supplementay information (Figures 2.S2, 2.S3 and 2.S4, respectively). Masses corresponding to silyated fragment pairs used to identify mono-hydroxylated fatty acid products are displayed within Table 2.S1 within the supplementary information.

Total percentage turnover of fatty acid substrates by intact wild-type and variant I263P BM3 enzymes are displayed in Table 2.10.

Table 2.10 Total turnover of fatty acid substrates by wild-type and I263P variant intact BM3 enzymes.

Turnover time was 60 min at 37 °C. Protein and substrate concentrations were 2 µM and 200 µM, respectively. NADPH was present at 500 µM, in addition to a cofactor regeneration system as described in the Methods section. Samples were repeated in triplicate and averaged to the nearest percentage point.

Intact BM3 variant	% Turnover of fatty acid substrate		
	Dodecanoate (C_{12})	Tetradecanoate (C_{14})	Hexadecanoate (C_{16})
WT	88	99	91
I263P	67	99	88

Against dodecanoate, the I263P variant displays an approximately ~25% decrease in turnover activity compared to the wild-type BM3 (i.e. a drop of 21 percentage points relative to the wild-type). Additionally, the I263P variant displays a marginal reduction in turnover against hexadecanoate and near-identical levels of turnover against tetradecanoate relative to intact, wild-type BM3. Levels of turnover were consistent within each repeat sample, with all standard deviations being less than 15% of the average.

2.4.8. Alteration of the Enantioselective Epoxidation of Styrene

The binding dissociation constant of styrene to variant I263Pis relatively unchanged to that of the wild-type enzyme (Table 2.11).

Table 2.11 Enantioselective hydroxylation of styrene performed by wild-type and I263P variant BM3 enzymes.

[1]Calculated using UV-visible spectroscopic binding titrations with heme domains at a concentration of ~5 µM. [2]Estimate for total concentration of styrene oxide produced by each enzyme, calculated with calibration curves. Reactions contained 2 µM of intact protein and 2 mM of styrene substrate in all cases, turned over for a 20 hour period. [3]Calculated using peak ratios from GC chromatograms. %enantiomer = {{peak area of enantiomer / total peak area of enantiomers} x 100}. [4] R(+) *%e.e* = {%R(+) styrene oxide - %S(-) styrene oxide}.

Intact BM3 variant	Binding constant[1] (K_d) (µM)	Styrene oxide formed[2] (µM)	R(+)-Styrene oxide[3] (%)	S(-)-Styrene oxide[3] (%)	R(+)-Styrene oxide[4] (% e.e.)
WT	148 ± 13.8	16.9 ± 6.83	58.7 ± 2.02	41.3 ± 2.02	17.4 ± 4.04
I263P	167 ± 16.8	85.2 ± 9.24	49.9 ± 0.10	50.1 ± 0.10	-0.20 ± 0.20

Despite this, styrene epoxidation, resulting in the formation of styrene oxide product, has more than quadrupled in the intact I263P variant in comparison to the wild-type BM3. Moreover, the ratio of R(+) and S(-) enantiomers of styrene oxide produced by the I263P variant has undergone significant changes following the point mutation at position 263. These data, in addition to the retention times of R(+) and S(-) styrene oxide (~16.4 and ~16.7 min, respectively), are illustrated in Figure 2.12.

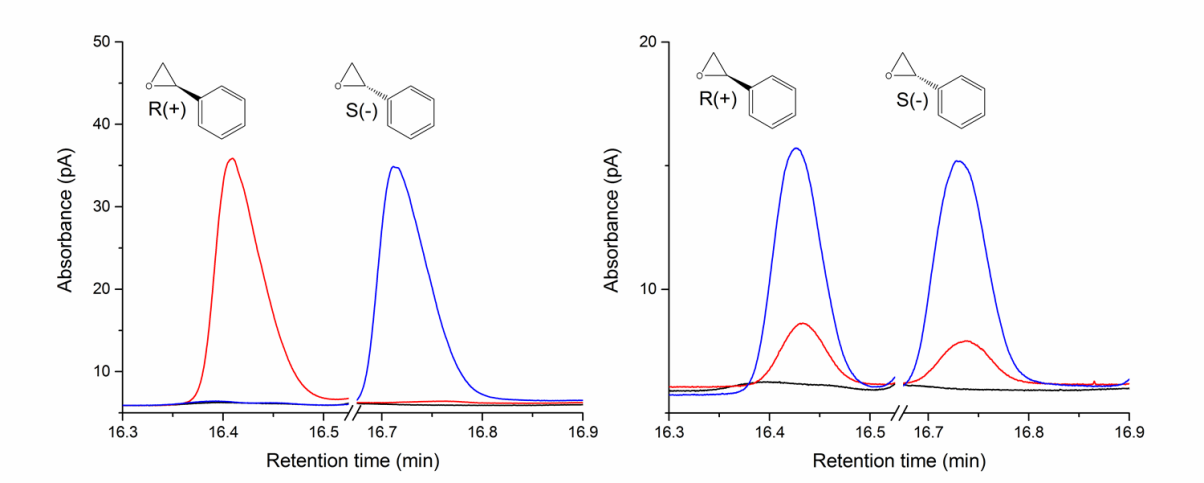

Figure 2.12 GC chromatogram of retention times of 125 µM R(+) (~16.43 min) and S(-) (~16.75 min) styrene oxide standards (left-hand insert). Production of R(+) and S(-) styrene oxide in intact wild-type (red line) and variant I263P (blue line) enzymes (right-hand insert). Control samples containing 2 mM styrene are represented by black lines in both inserts. Chromatograms were normalised to the internal standard.

Wild-type BM3 was shown to favour production of the R(+) enantiomer of styrene oxide by an enantiomeric excess of 17.4 ± 4.04 %. In contrast, the I263P variant produces R(+) and S(-) styrene oxide at near equal proportions. These data suggest that production of the S(-) enantiomer of styrene oxide has increased by approximately 10% in the I263P variant. In addition, these data suggest that in the I263P variant, the pro-S(-) face of styrene has a near equal amount of exposure to the heme-iron as the pro-R(+) face. This, in turn, implies that the I263P variant undergoes conformational changes to the I-helix that are distinct from those of the wild-type upon substrate binding. It can be suggested that any steric hindrance in the wild-type BM3 that hampers the production of S(-)-styrene oxide has been decreased in the I263P variant.

2.4.9. Altered Functionality of Variant I263P towards Monoterpene Compounds

When compared to fatty acid (Table 2.3) and styrene (Table 2.11) substrates, both wild-type and the I263P variant heme domains display low binding affinities towards the monoterpene compounds presented in Table 2.12. The substitution of Ile263 for proline has not made a significant impact on the K_d values of the variant towards the substrates (+)-α-pinene (+AP) and (–)-β-pinene (-BP) relative to the wild-type heme domain. Differences between the K_d values displayed by wild-type and I263P variant heme domains against both pinene substrates fall well within the margin of error. The other monoterpene substrate, (*R*)-limonene (LIM), displays greater affinity towards both the wild-type (60.6 ±

6.50 µM) and the I263P variant (74.5 ± 6.09 µM) when compared to the two pinene substrates. However, the I263P variant displays a marginally weaker binding affinity towards LIM than the wild-type enzyme. These binding data appear concordant with the data presented in Table 2.11, indicating the lack of impact that the I263P substitution has on binding affinities for bulkier substrates, namely those incorporating 6-carbon rings.

Table 2.12 Binding dissociation constants (K_d) of wild-type and I263P variant heme domains against monoterpene compounds.

Values were determined using UV-visible substrate binding titrations (~5 µM heme domain, diluted in assay buffer). Binding data were fitted using the hyperbolic function (see Methods, Equation 2.2).

BM3 Heme variant	Substrate binding dissociation constant (K_d) (µM)		
	(+)-α-pinene	(-)-β-pinene	(*R*)-limonene
WT	222 ± 62.4	200 ± 27.1	60.6 ± 6.50
I263P	246 ± 32.0	183 ± 25.2	74.5 ± 6.09

Wild-type BM3 displays little hydroxylase activity towards the selected monoterpenoid compounds. With substrates +AP and -BP, no clear products predominate, and turnover appears minimal. Against LIM, wild-type BM3 was shown to produce predominantly the mono-hydroxylated product isopiperitenol – albeit in small amounts.

In a stark contrast, functionalised monoterpene products were produced for all 3 substrates when incubated with the I263P variant enzyme. Furthermore, in the case of each substrate, a single product was shown to predominate. Products were identified according to level 2 annotation via comparison of product spectra to reference spectra contained within the MS NIST library (Sumner et al., 2007). Peak heights of all chromatogram traces were normalised to the peaks of the internal standard (IS), *sec*-butyl-benzene. Wild-type and variant BM3 turnover of +AP is displayed in Figure 2.13.

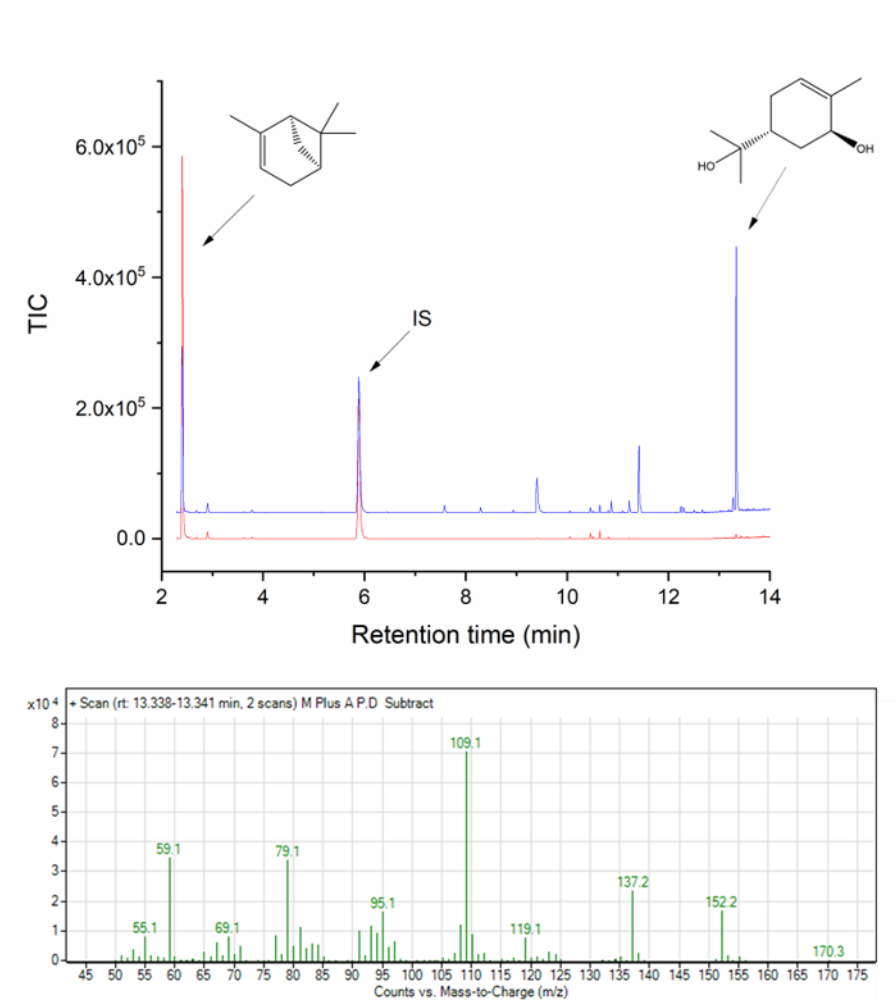

Figure 2.13 GC chromatogram of wild-type (red line) and I263P variant (blue line)-dependent turnover of (+)-α-pinene (top insert). The m/z spectrum of the predominant product produced by variant I263P (sobrerol) is shown in the lower insert.

The (+)-α-pinene peak is present at ~2.41 min, with the sobrerol product at ~13.34 min (top insert). Samples contained 2 µM intact BM3 protein and 2 mM substrate, turned over for 20 hours at 37°C.

As illustrated by Figure 2.13, the peak at ~2.41 min, corresponding to unmetabolised +AP, undergoes a significant decrease in intensity in the I263P variant trace relative to the wild-type trace - clearly indicating an increase in +AP metabolism by the I263P variant BM3 enzyme in comparison to the wild-type BM3. In tandem with a decrease in the +AP peak, a peak appears at ~13.34 min for the I263P variant spectrum. This peak becomes the predominant form in the I263P variant spectrum. From the corresponding m/z data (Figure 2.13, bottom insert), the product peak at ~13.34 was putatively identified as the functionalised monoterpene compound sobrerol (NIST match paramters: Match 908; R. Match 912; % probabillity 65.0; RI 1380). These data suggest that variant I263P is capable

of performing additional oxidations to initial products – i.e. producing secondary metabolites.

Turnover data for wild-type and the I263P variant with substrate –BP are displayed in Figure 2.14.

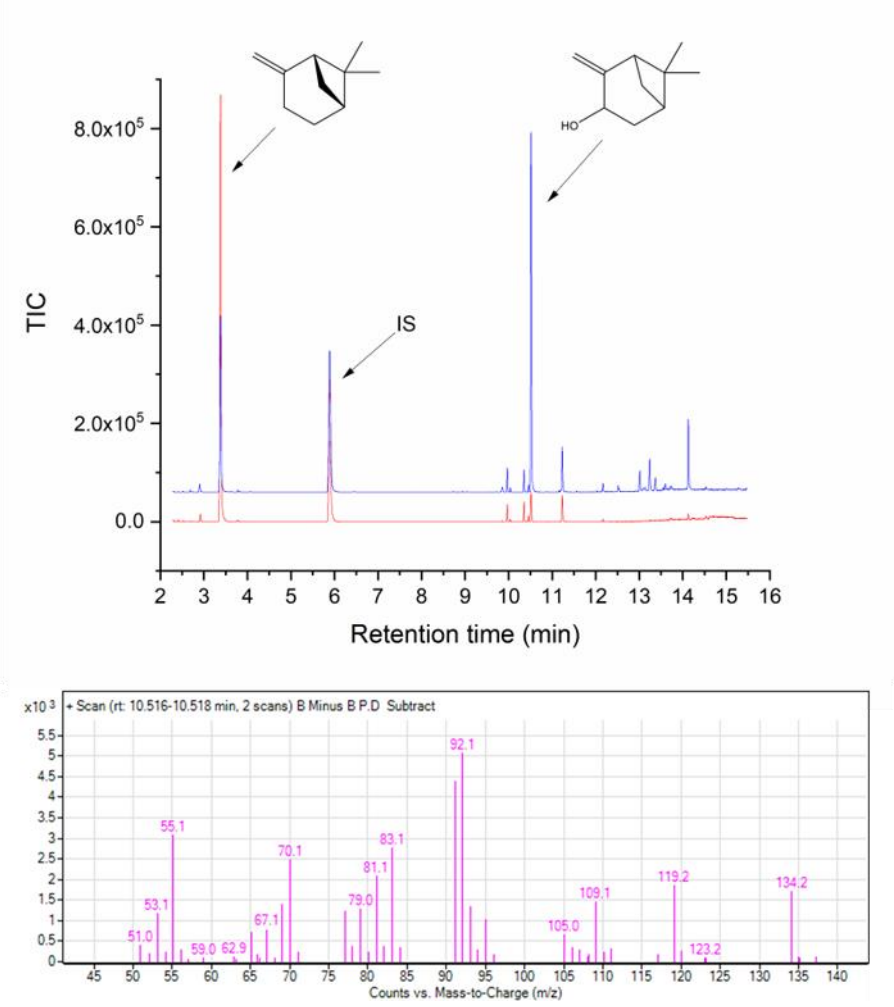

Figure 2.14 GC chromatogram of wild-type (red line) and I263P variant (blue line) turnover of (-)-β-pinene (top insert). The m/z spectrum of the predominant product produced by variant I263P (pinocarvol) is shown in the lower insert.

The (-)-β-pinene peak is present at ~3.38 min, with the pinocarveol product at ~10.52 min (top insert). Samples were turned over at 37 °C for a 20 hour period. Both wild-type and the I263P variant samples contained 2 µM intact BM3 protein and 2 mM substrate.

In Figure 2.14, peak heights at ~3.38 and ~10.52 retention times correspond to –BP and pinocarveol, respectively. As with +AP, the I263P variant displays increased metabolic activity with –BP when compared to the wild-type enzyme. Both the wild-type and the

I263P variant putatively produce the mono-hydroxylated terpene product pinocarveol, evidenced by the m/z spectra displayed in the bottom insert of Figure 2.14 (NIST match paramters: Match 930; R. Match 952; % probabillity 55.5; RI 1139). However, according to peak area - variant I263P demonstrates a near 15-fold increase in pinocarveol production relative to the wild-type enzyme. In addition, the I263P variant produces 3 further turnover products that are not evident in the wild-type spectrum. However, these product peaks were of insufficient intensity to allow accurate identification. Hydroxylase activity of wild-type and the I263P variant towards LIM is illustrated in Figure 2.15.

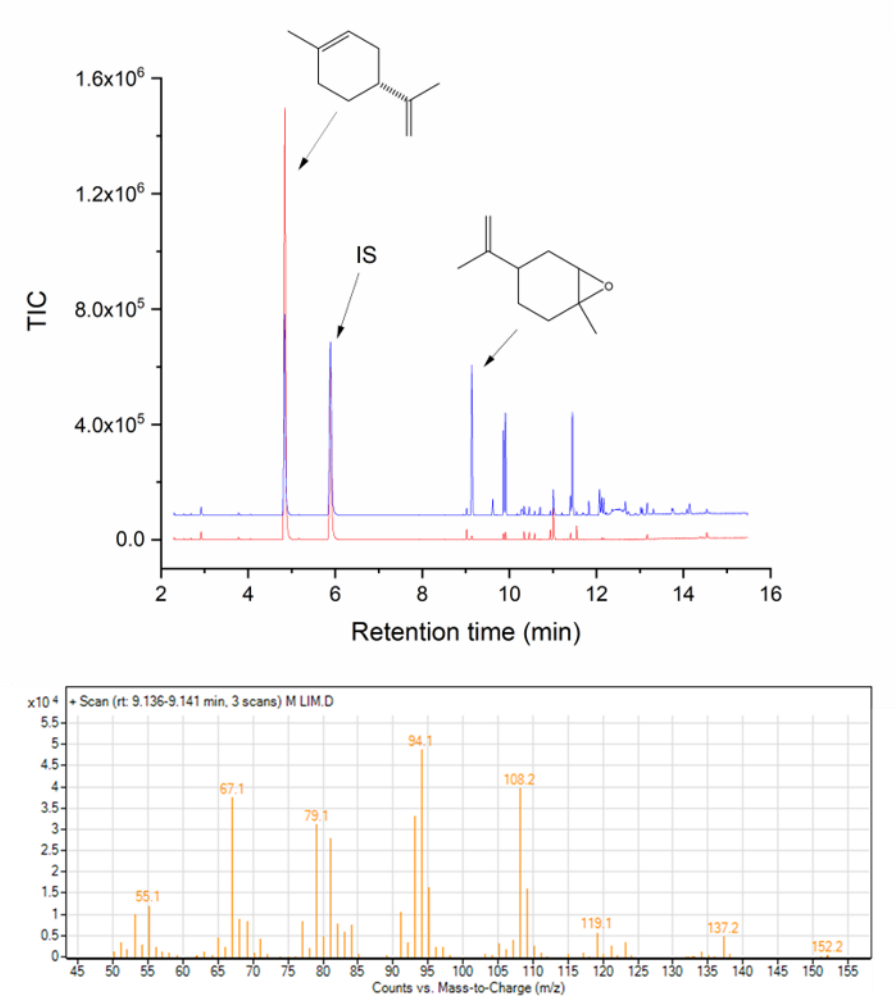

Figure 2.15 GC chromatogram of wild-type (red line) and the I263P variant (blue line) turnover of limonene (top insert). The m/z spectrum of the predominant product produced by variant I263P identifies the product as limonene oxide (bottom insert).

The peak at ~4.84 min corresponds to limonene. The limonene oxide product peak occurs at ~9.14 min (top insert). Samples were turned over at 37 °C for a period of 20 hours. Both wild-type and the I263P variant samples contained 2 µM intact BM3 protein and 2 mM substrate.

Unmetabolised LIM elutes in a single peak with a retention time of ~4.84 min. The predominant product produced by variant I263P, limonene oxide, elutes at a retention time of ~9.14 min. The corresponding m/z spectra of the ~4.84 peak produced by variant I263P (Figure 2.15, bottom insert) enabled the putative identification of the limonene oxide product (NIST match paramters: Match 949; R. Match 960; % probabillity 56.1; RI 1138). Epoxidation of the carbon-carbon double bond at positons 1-2 on LIM is not apparent in significant amounts for the wild-type enzyme. The corresponding diol of limonene oxide (1,2-epoxylimonene diol) was not observed in either reaction. In addition to limonene oxide, the I263P variant produced 5 other functionalised products of LIM not observed within the control or wild-type samples. However, the low abundance of these peaks prevented accurate product identification. The I263P variant also demonstrated a ~1.5 fold decrease in isopiperitenol production relative to the wild-type enzyme.

2.4.10. The Crystal Structure of the I263P Variant Heme Domain Bound to Substrate NPG

As outlined in the Methods section, a 1.6 Å resolution crystal structure was obtained for the I263P variant heme domain bound to NPG (I263P-SB). The structural similarity of the I263P-SB heme domain to that of the wild-type in the NPG-bound state (WT-SB) was assessed using PDBeFold software (Krissinel *et al.,* 2004). The NPG-bound wild-type structure was obtained from PDB structure 1JPZ (Haines *et al*., 2001). This comparison generated an RMSD score of 0.320 for I263P-SB relative to the WT-SB, indicating minimal deviation between the two structures. Topological differences between WT-SB and I263P-SB heme domains were assessed using a structural overlay (supplementary information, Figure 2.S1). As indicated by the RMSD value, minimal differences to the overall protein fold, including all main secondary structural elements, with the notable exception of the I-helix, were observed between the WT-SB and I263P-SB structures. To visualise any dramatic changes to the architecture of the I-helix, the I262P-SB structure was overlaid with wild-type BM3 heme domain structures in both the substrate free (WT-SF), (PDB: 1BU7) and NPG-bound (WT-SB) (PDB: 1JPZ) states (Figure 2.16).

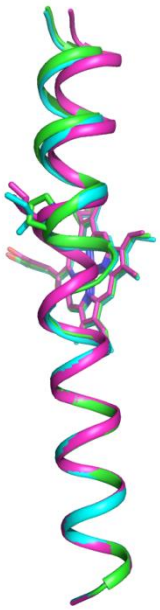

Figure 2.16 I-helical alignments of substrate-free wild-type (purple – PDB 1BU7), NPG-bound wild-type (blue- PBD 1JPZ) and NPG-bound I263P (green) heme domains.
Upon NPG binding, the I-helix kink angle of the wild-type BM3 heme domain decreases from 8.90° to 4.30°. Following the addition of NPG, the kink angle of the I263P variant decreases to 4.60°. Alignment of the wild-type NPG-bound and I263P NPG-bound I-helical structures.

Despite the potentially disruptive nature of the proline substitution, the I-helix remains intact within I263P-SB. Following substrate binding, the I-helix of the wild-type heme domain undergoes a conformational shift, resulting in a hydrogen bonding interaction forming between the side-chains of residues Ile263 and Glu267 (Haines et al., 2001). This in turn reduces the angle of the I-helix "kink" centred on residue Gly265. I-helices of each structure were aligned between residues Leu273 and Lys282. The kink angle was measured between residues Leu273, Gly265 and Lys282 in each structure. Within the wild-type BM3 heme domain, substrate binding prompts the kink angle of the I-helix to decrease, from 8.90° (WT-SF) to 4.30° upon substrate binding (WT-SB). Structure I263P-SB displays an I-helix kink angle of 4.60°– a marginal reduction of 0.30° relative to the WT-SB structure. Despite this change to the kink angle, the turn pattern of the I263P-SB I-helix remains conserved. The inter-helical interactions responsible for the "kink" in WT-SB, hydrogen bonds between I263(CO) and E267(NH) and between Ala264(CO) and Thr268(NH) remain preserved within I263P-SB. The imidazole containing ring of residue His266 exists in a near-identical rotational confirmation to WT-SB within I263P-SB, as do the positions of residues Phe261 and Phe87.

As illustrated in Figure 2.17, the binding mode of NPG to the I263P variant appears very similar to that of the wild-type BM3 heme domain.

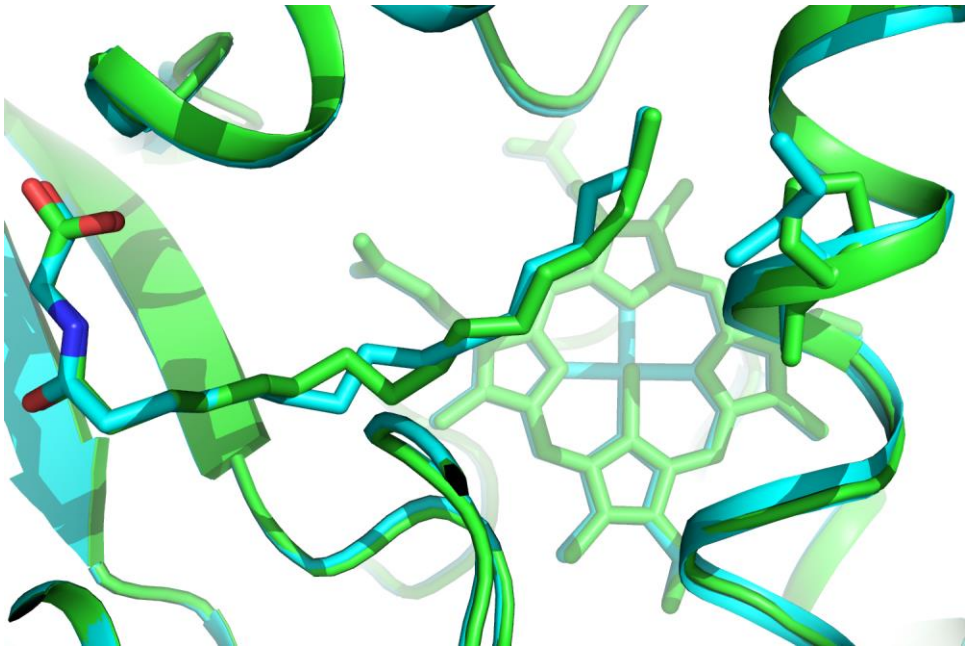

Figure 2.17 Alignment of NPG-bound I263P heme domain (green) to NPG-bound wild-type heme domain (blue) (PDB 1JPZ).
NPG substrate and residues at position 236 are highlighted.

The distance between the C_α atoms of the residue at position 263 and that of Leu75, located on the opposite side of the binding channel, was measured in each structure. Within the I263P-SB structure, the distance between L75 and P263 was 13.3 Å. In comparison, the distance between L75 and I263 in the WT-SB structure was 13.0 Å, indicating that the active site channel has been widened by 0.30 Å in the I263P variant heme domain. Finally, the position of the axial water molecule in the I263P-SB structure could not be accurately assigned. This was due to the presence of difference density (Fo-Fc density) directly above the heme cofactor. These data suggest that a component of the crystal screening condition, likely polyethylene glycol (PEG), is interacting with the heme iron.

2.4.11. EPR Spectra of Wild-Type and I263P Variant Heme Domains

EPR spectra for the wild-type and I263P variant heme domains were collected in order to examine the effects of substrate binding on the environment and coordination state of the heme iron. Spectra were collected for both species in the substrate-free and dodecanoate-bound states, in addition to samples with an EtOH solvent control (5% v/v). All collected

EPR spectra displayed rhombic signals typical of cysteine thiolate-coordinated P450 enzymes. Spectra for the wild-type heme domain are presented in Figure 2.18.

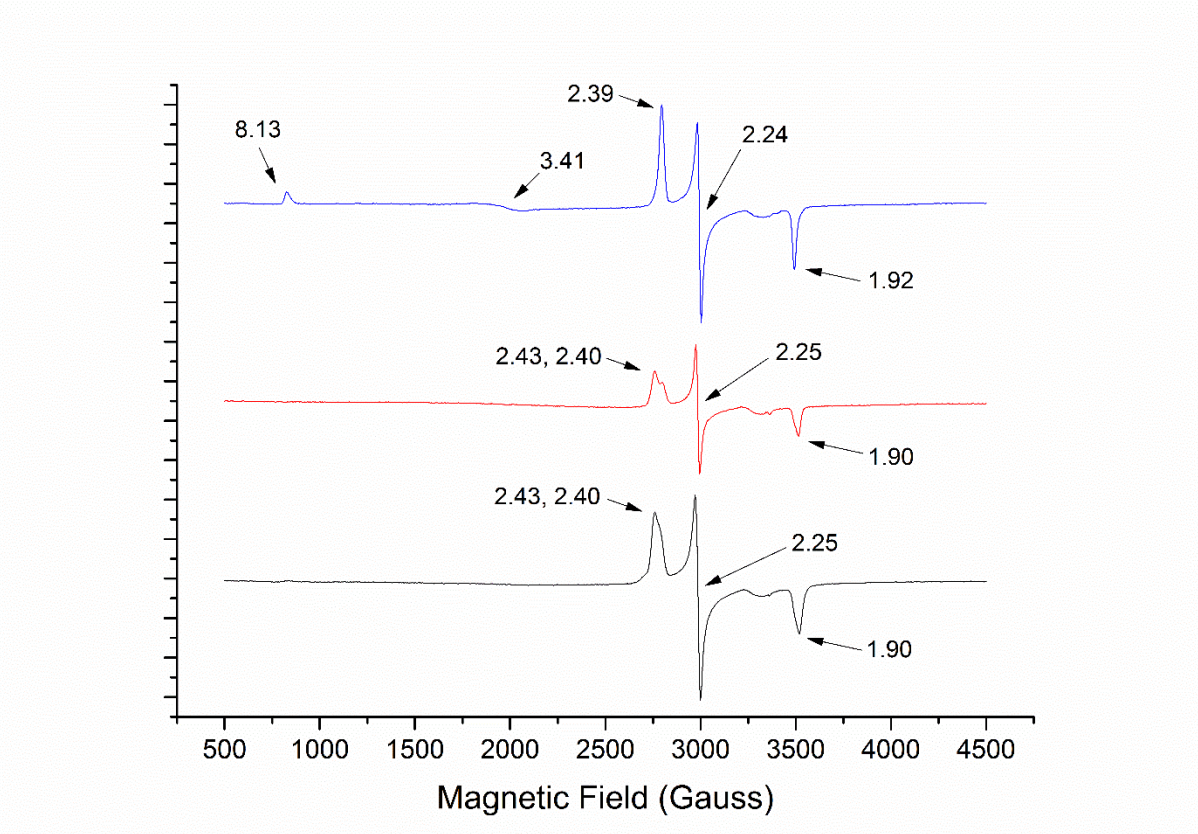

Figure 2.18 EPR spectra of wild-type BM3 heme domain in the substrate-free (black) state, and in the presence of EtOH (red) and dodecanoate (blue).

Spectra were collected at a protein concentration of 300 µM and a temperature of 10K. Dodecanoate was present at a concentration of 700 µM and at a final volume of 5% v/v. Absolute EtOH was present at a final volume of 5% v/v. Annotations correspond to g-values.

In the substrate-free state (Figure 2.18, black spectrum) the wild-type BM3 displays g-values of 2.43 (g_z), 2.25 (g_y) and 1.90 (g_x) – corresponding to a low-spin (S = ½), ferric heme. A g_z peak shoulder present at 2.40 suggests the presence of a low-spin species with different distal water coordination. The addition of EtOH solvent to substrate-free wild-type heme domain (Figure 2.18, red spectrum) does not cause significant changes to the heme environment, displaying g-values of 2.43/2.25/1.90. Once again, this corresponds to a low-spin heme in the ferric oxidation state. However, the addition of EtOH has prompted an increase in the shoulder present at g = 2.40, indicating increased perturbation of the distal water molecule. The presence of a near-saturating concentration of the substrate

dodecanoate (Figure 2.18, blue spectrum) induces a shift in the wild-type BM3 g-values to 8.13/3.41/1.92. These peaks indicate that a high-spin, ferric (S = 5/2) iron species is formed following the addition of dodecanoate.

EPR spectra collected for variant I263P heme domain are presented in Figure 2.19.

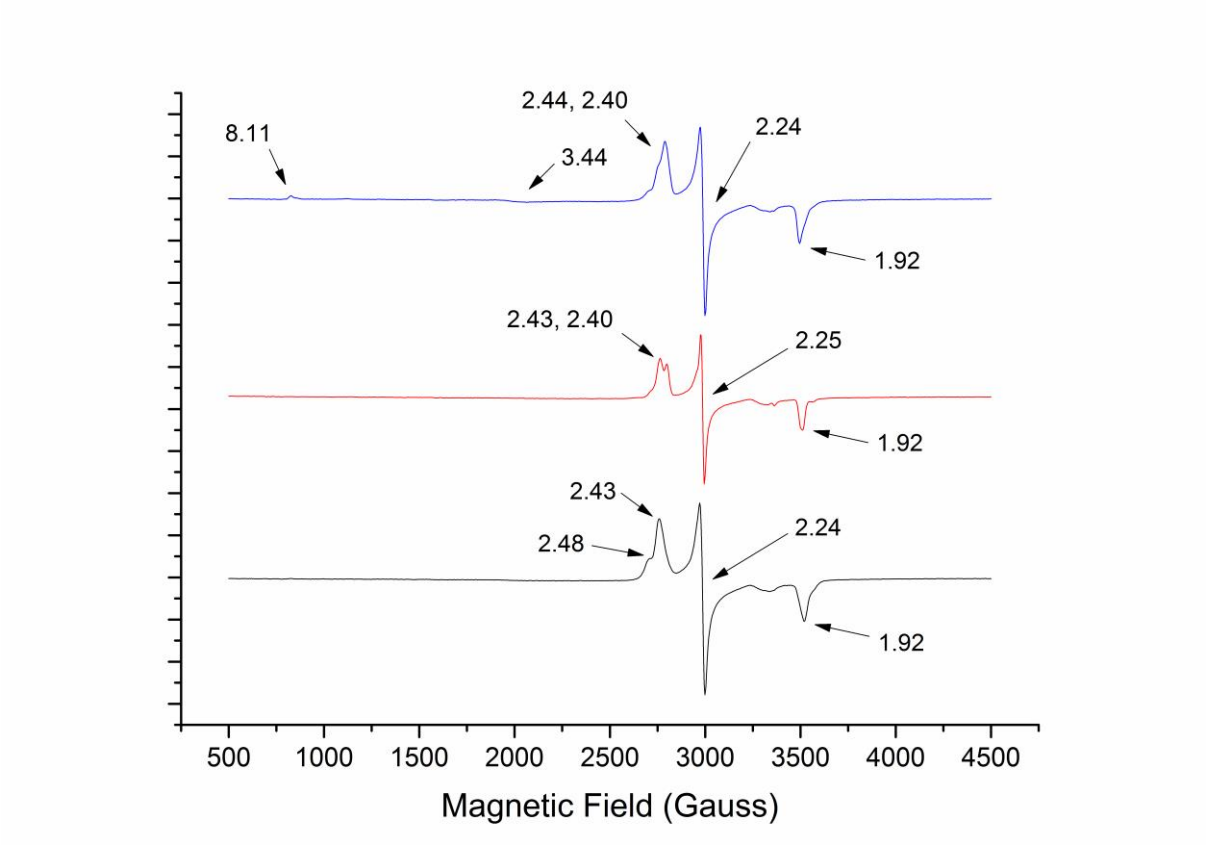

Figure 2.19 EPR spectra of the I263P BM3 variant heme domain in the substrate free (black) state, and in the presence of EtOH (red) and docdecanoate (blue).

Protein concentration was 300 µM across all spectra. Spectra were collected at a temperature of 10 K. Absolute EtOH was present at a final volume of 5% v/v, as was dodecanoate. Dodecanoate was present at a 700 µM concentration. Annotations correspond to g-values.

In the absence of both EtOH and substrate, the I263P variant displays g-values of 2.48/2.24/1.92 – suggesting a low-spin, ferric heme environment (Figure 2.19, black spectrum). However, the change in the g_z value relative to the substrate-free wild-type spectra indicates an altered heme environment in the substrate-free I263P variant heme domain. Specifically, this g_z value suggests that significant perturbation of the distal water molecules within the plane of the heme iron has occurred in the I263P variant. The I263P variant heme domain displays g-values of 2.43/2.24/1.92 following the addition of EtOH. As

with the wild-type heme domain, this spectrum indicates the presence of low-spin, ferric heme in which the distal water molecule exists in two different confirmations (Figure 2.19, red spectrum). Upon the addition of dodecanoate, the I263P variant heme domain displays g-values corresponding to a high-spin, ferric heme species of with g-values of 8.11/3.44/1.92 (Figure 2.19, blue spectrum).

The comparatively low proportion of high-spin heme detected in both the wild-type and I263P variant heme domain can be attributed to the cryogenic temperature at which the spectra were collected (10 K). This low temperature has a significant negative impact on ligand binding within BM3 (as ligand binding is endothermic) – but is a necessary condition for the collection of hemeprotein EPR spectra.

2.5. Discussion

In wild-type BM3, substrate binding prompts significant structural rearrangements within the active site (Whitehouse, et al., 2012). Notably, the extensive hydrogen bonding network of the I-helix is re-organised, causing the I-helix to pivot away from the distal face of the heme. This, in turn, enables the distally-ligated water ligand to be displaced from the heme iron, resulting in a spectral shift corresponding to the high-spin ferric state (Haines et al., 2001).

As indicated by the UV-vis binding titrations, the I263P heme domain was unable to fully transition to the high-spin state against all tested substrates, a feature also observed in another Ile263 variant (Ost et al., 2000). These data suggest that alterations to Ile263 hinder the expulsion of the distally-ligated water molecule from the distal face of the heme, potentially impacting the reorganisation of the H-bonding network within the I-helix. This hypothesis is supported by the prominent ν_{11} band found in the NPG-bound I263P variant using resonance Raman spectroscopy. The intensity of this band relative to the corresponding wild-type spectrum is indicative of a domed heme ring – offering further evidence to suggest that the distal water molecule remains in the plane of the heme following substrate binding (Smith et al., 2003).

Furthermore, the differing set of g-values provided by the substrate-free EPR spectrum of the I263P variant heme domain indicates that changes to the heme iron environment have occurred prior to the addition of substrate. These spectra further suggest that alterations to the organisation of water molecules have occurred as a result of structural changes to the I-helix caused by the proline substitution in variant I263P.

However, the crystal structure of the NPG-bound I263P variant heme domain does not provide accurate insight into the organisation of water molecules within the active site. While the resolution is relatively high (1.6 Å), the presence of an unidentified ligand, most likely a PEG compound, binding to the heme iron prevents accurate water molecule assignment.

As illustrated by the NPG-bound crystal structure, the I-helix remains intact in the I263P variant heme domain. Aside from a reduction in thermostability of ~2.40 °C relative to the wild-type heme domain, the proline substitution does not appear to have made a significant structural impact on the I263P variant heme domain. As with the wild-type, the I-helix "kink" angle decreases in the I263P variant following substrate binding, albeit with a reduction of 0.30 degrees. The reduction of this "kink" angle confirms that, while the hydrogen-bonding interaction between residue 263 and Glu267 has been preserved, the dimensions of the active site in the substrate-bound conformation have been altered relative to the wild-type heme domain.

Moreover, the width of the active site channel directly above the heme has also been widened by 0.3 Å in the I263P variant heme domain, as evidenced by the NPG-bound structure. This mutation appears to enable fatty acid substrates to get closer to the heme iron, as demonstrated by the concurrent increases in binding affinities. However, the increase in space close to the heme has not been enlarged to an extent that significantly alters binding affinities for bulky substrates such as styrene and the monoterpene compounds. Despite this, the I263P variant BM3 enzyme demonstrates radical alterations to both fatty acid regioselective mono-hydroxylation and the enantioselective oxidation of styrene.

Wild-type BM3 was shown to produce predominantly ω-1, ω-2 and ω-3 hydroxylated dodecanoate, in accordance with the initial studies by Miura and Fulco (1975). With tetradecanoate, the wild-type BM3 was shown to produce ω-1 as the predominant hydroxylated fatty acid product, in addition to forming ω-2 and ω-3 hydroxylated fatty acids in agreement with (Cryle et al., 2007). Proportions of hydroxylated products recorded for wild-type BM3 with hexadecanoic acid were also in broad agreement with those presented by Truan et al. (1999), albeit with a greater proportion of the ω-2 product produced (56% mono-hydroxylated primary product) in this study by comparison to the proportion (43%) of the ω-2 product reported by Truan et al. As a result, the proportion of ω-1 hydroxylated products fell by ~10 percentage points (to 13%) in this study relative to

that reported by Truan et al. (1999) at ~23%. The discrepancy between these data can likely be attributed to wild-type BM3 performing secondary metabolism on the ω-1 and ω-2 hydroxylated products within the study by Traun, et al (1999), as this would impact the ratio of mono-hydroxylated products. In this study, no products of secondary metabolism were evident. Wild-type BM3 can only perform secondary metabolism on ω-1 and ω-2 hydroxylated products, with both reactions being highly energetically unfavourable for the enzyme (Sevrioukova et al., 1996). As such, data analysis for mono-hydroxylase product profiles of both wild-type and the I263P variant against hexadecanoic acid should not be drastically affected.

For all fatty acid substrates tested, the I263P variant BM3 enzyme consistently hydroxylates carbon positions distant from the terminal ω-carbon and towards the α-carbon of the carboxylate group. In the case of the C12, C14 and C16 substrates used with the I263P BM3 enzyme, fatty acid hydroxylation extends further towards the α-carbon end of the fatty acids in the order of C16>C14>C12 (see Tables 2.7-9). It appears that the I263P substitution alters the binding mode of the tested fatty acid substrates, enabling greater substrate flexibility within the active site. In so doing, the fatty acids substrates have been positioned further down the binding channel, facilitating greater access to the heme iron with capacity to oxidise positions closer to the α-carbon ends of the fatty acids. Consequently, the product profile of the I263P variant enzyme against these fatty acid substrates has been altered. However, the NPG-bound structure of the I263P variant did not display significant differences in the positioning of the substrate relative to the heme iron. Novel oxidation reactions, such as epoxidation or multiple hydroxylations were not observed for the I263P variant enzyme against dodecanoate and tetradecanoate. However, minor amounts of secondary metabolite products of hexadecanoic acid were identified, as with the wild-type, for the I263P variant enzyme. However, these products could not be accurately assigned, and were excluded from profile ratios of mono-hydroxylated products.

Previous mutagenic studies on position Ile263 have shown that substitution with an alanine residue of creates additional space for the ω-terminus of fatty acid substrates, prompting changes to regioselectivity (Dietrich et al., 2009). Dietrich et al. demonstrated an increase in regioselective hydroxylation of dodecanoate at the ω-7 position within variants containing I263A/G mutations. Hydroxylations occurring at the δ carbon of fatty acids offer sustainable, enzymatic routes to lactone precursors of commercial interest (Dietrich et al., 2009; Manning et al., 2019). Currently, these positions are out of reach of the I263P single

mutant variant – with hydroxylation reaching the ω-6 (dodecanoate), ω-7 (tetradecanoate) and ω-8 (hexadecanoate) positions. However, additional mutagenesis to variant I263P, notably the replacement of the selectivity-influencing Phe87 with a smaller residue, could position the fatty acid substrates further into the active site (Noble et al., 1999), enabling these commercially viable hydroxylation positions to be accessed.

While the I263P substitution does not significantly affect the binding affinity of styrene relative to the wild-type, the variant displays alterations in enantiopreference of the epoxidised product, styrene oxide. There is some debate over the exact enantiomeric ratios of styrene oxide produced by wild-type BM3. Eiben et al. (2006) reported that wild-type BM3 produces a 25% ee of the S(-) enantiomer, whereas Tee et al. (2006) observed a 20% ee in favour of the R(+) enantiomer. Furthermore, Huang et al. (2011) reported that wild-type BM3 does not produce styrene oxide. In contrast to the data presented by Huang et al., data collected for both the wild-type and I263P variant enzymes gave evidence for the turnover of styrene to styrene oxide in both cases.

The wild-type P450 BM3 enzyme was shown to produce both enantiomers of styrene oxide, with R(+) styrene oxide predominating at an ee of ~18%, in broad agreement with the data reported by Tee et al. (2006). Interestingly, variant I263P displayed a near-racemic product profile, producing both enantiomers at a near identical ratio. These data suggest that the I263P BM3 variant has prompted a change in the binding mode of styrene, consequently improving access for styrene to the the pro-S face of the heme iron. Shehzad et al. (2013) postulated that that wild-type BM3 favours production of the R(+) due to a neighbouring I-helix residue, Ala264 part covering the heme pyrrole ring, providing sufficient steric hindrance to make exposure of the pro-S face of styrene unfavourable. Within variants that favoured the S(-) enantiomer, further structural data presented by Shehzad et al. suggested that the I-helix residues Ala264 and Glu265 were located further from the heme on account of a reduced “kink” angle. In the I263P variant, the I-helical kink angle was shown to decrease by 0.30 degrees following substrate binding. These data suggest that even a minor alteration to the kink angle of the I-helix, such as that displayed by the I263P variant, is sufficient to alter the position of styrene within the active site. This enables the pro-S face of the styrene greater access to the heme iron, altering the enantio-specific oxidation of styrene accordingly. These data suggest that position 263 likely has a greater role in determining entantioselectivity than has previously been reported.

In light of the alterations to both regio- and enantio-selectivity demonstrated by the I263P variant enzyme, a small monoterpene library consisting of (+)-α-pinene (+AP), (-)-β-pinene (-BP) and limonene (LIM) were selected for turnover studies. Monoterpenoids are common throughout nature, occurring in essential oils derived from plant leaves, fruit and bark (Banthorpe et al., 1971). Oxidised monterpene compounds often have valuable fragrance, flavour and pharmaceutical properties (Rozenbaum et al., 2006). Rationally designed P450 enzymes with altered substrate selectivity may offer an enzymatic, environmentally sustainable route to these valuable compounds by negating the need for direct extraction. Development of variant BM3 enzymes with enhanced functionality towards monoterpene compounds has been attempted in numerous studies (Schewe et al., 2008; Seifert et al., 2009; Silva et al., 2012). Additionally, BM3 variants containing substitutions at position 263 have demonstrated increased hydroxylase activity towards the sesquiterpene (+)-valencene (Sowden et al., 2005).

As with their structurally comparable counterpart, styrene, the I263P variant enzyme did not display significantly altered binding affinities for the monoterpene compounds. However, the I263P variant enzyme displayed increased oxidative activity towards each of the monoterpene compounds by comparison to the wild-type enzyme. Against +AP, the I263P variant enzyme demonstrated a marked increase in turnover compared to the wild-type BM3, predominantly producing the di-hydroxylated product, sobrerol. Sobrerol has been shown to have a variety of clinical benefits, from promoting mucus production (Yuta et al., 2005) to the chemoprevention of breast cancer (Crowell, 1997). Production of sobrerol from +AP using a BM3 variant containing multiple point mutations (A74G/L188Q) has previously been reported by Stamm et al. (2019). However, as postulated by Hernandez-Ortega et al. (2018), sobrerol formation most likely proceeds via water-mediated, non-enzymatic degradation of an epoxide intermediate, rather than successive oxidation performed by the I263P variant BM3 enzyme.

The I263P Bm3 variant demonstrated a near 15-fold increase in the production of the mono-hydroxylated –BP product pinocarveol when compared to the wild-type enzyme. Pinocarveol occurs naturally within the common gum seed, and is capable of acting as a $GABA_A$ receptor modifier (Kessler et al., 2014). As indicated by the comparative lack of turnover, the heme iron of the wild-type BM3 proved to be near inaccessible to –BP. The increase in –BP turnover within variant I263P once again suggests that a substitution at position 263 has created space directly above the heme iron, enabling –BP to access the

heme for subsequent oxidation (Sowden et al., 2005). Production of pinocarveol has previously been displayed by a BM3 triple-mutant in a study by Branco et al. (2008). In this case, the authors suggested that substitutions at positions A74 and L188, with the purpose of widening the substrate access channel, were prerequisites for BM3 to display any hydroxylase activity towards pinene compounds. However, data collected for I263P clearly demonstrates that a degree of oxidative activity towards pinene compounds can be achieved through fewer structural modifications to the wild-type enzyme.

Monoterpene epoxidation performed by the I263P variant enzyme was also evident against the substrate LIM, in which limonene oxide was the most predominant product. Limonene oxide serves as a valuable intermediate in the manufacture of numerous commercial products (Wróblewska, 2014). The I263P variant demonstrated a significant increase in limonene oxide production relative to the wild-type enzyme. However, the predominant product produced by the wild-type enzyme, isopiperitenol, is decreased by 1.5-fold in the I263P variant enzyme.

Inclusion of an I263A point mutation in BM3 variants has previously been shown to induce a significant increase in epoxidation of alkenes (Kubo et al., 2006). These data, together with the turnover data collected for styrene and +AP, suggest that the I263P variant enzyme may preferentially epoxidise compounds containing carbon-carbon double bonds exposed to the heme, in lieu of performing a hydroxylation reaction elsewhere on the compound. Increased rates of substrate epoxidation over hydroxylation have been observed in variants of CYP2E1 (Vaz et al., 1998). The authors postulated that this is due to the increased presence of a ferric hydroperoxy catalytic intermediate, arising from alterations to the proton transfer network within the catalytic cycle. In the case of the I263P variant, it could be suggested that formation of a ferryl species, itself responsible for the majority of oxidation reactions performed by BM3, has been impacted (Ortiz de Montellano et al., 2002). However, Ortiz de Montellano's group cited a lack of evidence for the role of the hydroperoxy species in catalysis, with EPR and ENDOR studies of CYP101A1 strongly suggesting the intermediate has no role in P450 catalysis, but instead is a precursor to the the ferryl species Compound I – which is the major oxidising species in the P450 catalytic cycle (Davydov et al., 2001).

The intact I263P variant clearly retains catalytic activity – as confirmed by kinetic and turnover studies. Total turnover of fatty acid substrates did not collapse in the I263P variant, and was only significantly reduced in the case of the substrate dodecanoate.

However, the I263P variant displays decreased turnover rates relative to the wild-type enzyme in addition to a notable increase in uncoupling. The increase in peroxide uncoupling in the I263P enzyme may well arise from the altered topology of the I-helix, causing either a decrease in dioxygen binding efficiency or an increase in the solvation of the active site (Loida et al., 1993). The decrease in catalytic activity of the I263P variant with fatty acid substrates could also be attributed to the increased uncoupling of NADPH oxidation from product formation. Previous alterations to residues in the I-helix have resulted in similar decreases in catalytic activity together with a comparable rise in uncoupling (Yeom et al., 1995). In addition, even fairly unadventurous substitutions of I-helix residues (e.g. Glu to Gln) have been shown to lead to large decreases in catalytic efficiency (Yeom et al., 1997). In this context, it is perhaps surprising that the I263P variant has retained a semblance of wild-type catalytic activity at all.

Additional studies with the I263P variant may offer further insights into the roles of both Ile263 and the I-helix in regards to regio-, enantio- and substrate-specific oxidation performed by BM3. Notably, collection of additional crystal structures of the I263P heme domain in the substrate-bound state should provide more conclusive evidence regarding the impact of I263P on the binding orientation of the styrene and monoterpene substrates. Further structural data would also ideally enable more thorough analysis of the I263P heme environment, notably the organisation of key water molecules within the active site, and any role these perturbations may have on the altered functionality of variant I263P.

To conclude, a single point mutation in a catalytically essential part of the I-helix has enabled profound alterations to the regio-, enantio- and substrate specificity of P450 BM3. With further development and rational evolution, variant BM3 enzymes incorporating I263P could offer novel routes to a variety of pharmaceutical, fragrance and flavour compounds of industrial value.

2.6. References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). 'PHENIX: a comprehensive Python-based system for macromolecular structure solution', *Acta Crystallogr D Biol Crystallogr,* 66(Pt 2), pp. 213-21.

Banthorpe, D. V., Baxendale, D., Gatford, C. & Williams, S. R. (1971). 'Monoterpenes of some *Artemisia* and *Tanacetum* species grown in England', *Planta Med,* 20(2), pp. 147-52.

Branco, R. J., Seifert, A., Budde, M., Urlacher, V. B., Ramos, M. J. & Pleiss, J. (2008). 'Anchoring effects in a wide binding pocket: the molecular basis of regioselectivity in engineered cytochrome P450 monooxygenase from *B. megaterium*', *Proteins,* 73(3), pp. 597-607.

Butler, C. F., Peet, C., Mason, A. E., Voice, M. W., Leys, D. & Munro, A. W. (2013). 'Key mutations alter the cytochrome P450 BM3 conformational landscape and remove inherent substrate bias', *J Biol Chem,* 288(35), pp. 25387-99.

Cirino, P.C. and Arnold, F.H., 2002. Regioselectivity and activity of cytochrome P450 BM-3 and mutant F87A in reactions driven by hydrogen peroxide. *Adv Synth Catal*, 344(9), pp.932-937.

Crowell, P. L. (1997). 'Monoterpenes in breast cancer chemoprevention', *Breast Cancer Res Treat,* 46(2-3), pp. 191-7.

Cryle, M.J. and De Voss, J.J., 2007. Facile determination of the absolute stereochemistry of hydroxy fatty acids by GC: application to the analysis of fatty acid oxidation by a P450BM3 mutant. *Tetrahedron: Asymmetry*, 18(4), pp.547-551.

Davydov, R., Makris, T. M., Kofman, V., Werst, D. E., Sligar, S. G. & Hoffman, B. M. (2001). 'Hydroxylation of camphor by reduced oxy-cytochrome P450cam: mechanistic implications of EPR and ENDOR studies of catalytic intermediates in native and mutant enzymes', *J Am Chem Soc,* 123(7), pp. 1403-15.

Dietrich, M., Do, T. A., Schmid, R. D., Pleiss, J. & Urlacher, V. B. (2009). 'Altering the regioselectivity of the subterminal fatty acid hydroxylase P450 BM-3 towards gamma- and delta-positions', *J Biotechnol,* 139(1), pp. 115-7.

Eiben, S., Kaysser, L., Maurer, S., Kühnel, K., Urlacher, V. B. & Schmid, R. D. (2006). 'Preparative use of isolated CYP102 monooxygenases - a critical appraisal', *J Biotechnol,* 124(4), pp. 662-9.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). 'Features and development of Coot', *Acta Crystallogr D Biol Crystallogr,* 66(4), pp. 486-501.

Fasan, R., 2012. Tuning P450 enzymes as oxidation catalysts. *ACS Catal*, *2*(4), pp.647-66

Girvan, H. M., Waltham, T. N., Neeli, R., Collins, H. F., McLean, K. J., Scrutton, N. S., Leys, D. & Munro, A. W. (2006). 'Flavocytochrome P450 BM3 and the origin of CYP102 fusion species', *Biochem Soc Trans,* 34(6), pp. 1173-7.

Haines, D. C., Tomchick, D. R., Machius, M. & Peterson, J. A. (2001). 'Pivotal role of water in the mechanism of P450BM-3', *Biochemistry,* 40(45), pp. 13456-65.

Hernandez-Ortega, A., Vinaixa, M., Zebec, Z., Takano, E. & Scrutton, N. S. (2018). 'A Toolbox for Diverse Oxyfunctionalisation of Monoterpenes', *Sci Rep,* 8(1), p. 14396.

Huang, W. C., Cullis, P. M., Raven, E. L. & Roberts, G. C. (2011). 'Control of the stereo-selectivity of styrene epoxidation by cytochrome P450 BM3 using structure-based mutagenesis', *Metallomics,* 3(4), pp. 410-6.

Joosten, R. P., Salzemann, J., Bloch, V., Stockinger, H., Berglund, A. C., Blanchet, C., Bongcam-Rudloff, E., Combet, C., Da Costa, A. L., Deleage, G., Diarena, M., Fabbretti, R., Fettahi, G., Flegel, V., Gisel, A., Kasam, V., Kervinen, T., Korpelainen, E., Mattila, K., Pagni, M., Reichstadt, M., Breton, V., Tickle, I. J. & Vriend, G. (2009). 'PDB_REDO: automated re-refinement of X-ray structure models in the PDB', *J Appl Crystallogr,* 423), pp. 376-84.

Kessler, A., Sahin-Nadeem, H., Lummis, S. C., Weigel, I., Pischetsrieder, M., Buettner, A. & Villmann, C. (2014). 'GABA(A) receptor modulation by terpenoids from Sideritis extracts', *Mol Nutr Food Res,* 58(4), pp. 851-62.

Krissinel, E. & Henrick, K. (2004). 'Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions', *Acta Crystallogr D Biol Crystallogr,* 60 (Pt 12 Pt 1), pp. 2256-68.

Kubo, T., Peters, M. W., Meinhold, P. & Arnold, F. H. (2006). 'Enantioselective epoxidation of terminal alkenes to (R)- and (S)-epoxides by engineered cytochromes P450 BM-3', *Chemistry,* 12(4), pp. 1216-20.

Lewis, D. F. & Pratt, J. M. (1998). 'The P450 catalytic cycle and oxygenation mechanism', *Drug Metab Rev,* 30(4), pp. 739-86.

Loida, P. J. & Sligar, S. G. (1993). 'Molecular recognition in cytochrome P-450: mechanism for the control of uncoupling reactions', *Biochemistry,* 32(43), pp. 11530-8.

Manning, J., Tavanti, M., Porter, J., Kress, N., De Visser, S., Turner, N. and Flitsch, S., 2019. Regio-and Enantio-selective Chemo-enzymatic C-H-lactonization of Decanoic Acid to (S)-δ-Decalactone. *Angew Chem Intl Ed Engl*. 58(17):5668-71.

Meinhold, P., Peters, M.W., Hartwick, A., Hernandez, A.R. and Arnold, F.H., 2006. Engineering cytochrome P450 BM3 for terminal alkane hydroxylation. *Advanced Synthesis & Catalysis,* 348(6), pp.763-772.

Miura, Y. & Fulco, A. J. (1974). '(Omega -2) hydroxylation of fatty acids by a soluble system from *Bacillus megaterium'*, *J Biol Chem,* 249(6), pp. 1880-8.

Miura, Y. & Fulco, A. J. (1975). 'Omega-1, Omega-2 and Omega-3 hydroxylation of long-chain fatty acids, amides and alcohols by a soluble enzyme system from *Bacillus megaterium*', *Biochim Biophys Acta,* 388(3), pp. 305-17.

Morrison, J. F. (1969). 'Kinetics of the reversible inhibition of enzyme-catalysed reactions by tight-binding inhibitors', *Biochim Biophys Acta,* 185(2), pp. 269-86.

Munro, A. W., Girvan, H. M., Mason, A. E., Dunford, A. J. & McLean, K. J. (2013). 'What makes a P450 tick?', *Trends Biochem Sci,* 38(3), pp. 140-50.

Munro, A. W., Leys, D. G., McLean, K. J., Marshall, K. R., Ost, T. W., Daff, S., Miles, C. S., Chapman, S. K., Lysek, D. A., Moser, C. C., Page, C. C. & Dutton, P. L. (2002). 'P450 BM3: the very model of a modern flavocytochrome', *Trends Biochem Sci,* 27(5), pp. 250-7.

Munro, A. W., Lindsay, J. G., Coggins, J. R., Kelly, S. M. & Price, N. C. (1996). 'Analysis of the structural stability of the multidomain enzyme flavocytochrome P-450 BM3', *Biochim Biophys Acta,* 1296(2), pp. 127-37.

Nelson, D. R. (2006). 'Cytochrome P450 nomenclature, 2004', *Methods Mol Biol,* 320, pp. 1-10.

Noble, M. A., Miles, C. S., Chapman, S. K., Lysek, D. A., MacKay, A. C., Reid, G. A., Hanzlik, R. P. & Munro, A. W. (1999). 'Roles of key active-site residues in flavocytochrome P450 BM3', *Biochem J,* 339 (2), pp. 371-9.

Omura, T. and Sato, R. (1964) The carbon monoxide-binding pigment of liver microsomes. I: evidence for its hemoprotein nature. *J. BioI. Chem*. 239, 2370-2378.

Ortiz de Montellano, P. R. & De Voss, J. J. (2002). 'Oxidizing species in the mechanism of cytochrome P450', *Nat Prod Rep,* 19(4), pp. 477-93.

Ost, T. W., Miles, C. S., Murdoch, J., Cheung, Y., Reid, G. A., Chapman, S. K. & Munro, A. W. (2000). 'Rational re-design of the substrate binding site of flavocytochrome P450 BM3', *FEBS Lett,* 486(2), pp. 173-7.

Ravichandran, K. G., Boddupalli, S. S., Hasermann, C. A., Peterson, J. A. & Deisenhofer, J. (1993). 'Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's', *Science,* 261(5122), pp. 731-6.

Rozenbaum, H.F., Patitucci, M.L., Antunes, O.A.C. and Pereira Jr, N., 2006. Production of aromas and fragrances through microbial oxidation of monoterpenes. *Braz J ChemEng*, *23*(3), pp.273-279.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press*. *New York*.

Schewe, H., Kaup, B. A. & Schrader, J. (2008). 'Improvement of P450(BM-3) whole-cell biocatalysis by integrating heterologous cofactor regeneration combining glucose facilitator and dehydrogenase in *E. coli*', *Appl Microbiol Biotechnol,* 78(1), pp. 55-65.

Seifert, A., Vomund, S., Grohmann, K., Kriening, S., Urlacher, V. B., Laschat, S. & Pleiss, J. (2009). 'Rational design of a minimal and highly enriched CYP102A1 mutant library with improved regio-, stereo- and chemoselectivity', *Chembiochem,* 10(5), pp. 853-61.

Sevrioukova, I., Truan, G. & Peterson, J. A. (1996). 'The flavoprotein domain of P450BM-3: expression, purification, and properties of the flavin adenine dinucleotide- and flavin mononucleotide-binding subdomains', *Biochemistry,* 35(23), pp. 7528-35.

Shehzad, A., Panneerselvam, S., Linow, M., Bocola, M., Roccatano, D., Mueller-Dieckmann, J., Wilmanns, M. & Schwaneberg, U. (2013). 'P450 BM3 crystal structures reveal the role of the charged surface residue Lys/Arg184 in inversion of enantioselective styrene epoxidation', *Chem Commun (Camb),* 49(41), pp. 4694-6.

Silva, A.C.R.D., Lopes, P.M., Azevedo, M.M.B.D., Costa, D.C.M., Alviano, C.S. and Alviano, D.S., 2012. Biological activities of a-pinene and β-pinene enantiomers. *Molecules*, 17(6), pp.6305-16.

Smith, S. J., Munro, A. W. & Smith, W. E. (2003). 'Resonance Raman scattering of cytochrome P450 BM3 and effect of imidazole inhibitors', *Biopolymers,* 70(4), pp. 620-7.

Sowden, R. J., Yasmin, S., Rees, N. H., Bell, S. G. & Wong, L. L. (2005). 'Biotransformation of the sesquiterpene (+)-valencene by cytochrome P450cam and P450BM-3', *Org Biomol Chem,* 3(1), pp. 57-64.

Spiro, T.G., Stong, J.D. and Stein, P., 1979. Porphyrin core expansion and doming in heme proteins. New evidence from resonance Raman spectra of six-coordinate high-spin iron (III) hemes. *J Am Chem Soc*, 101(10), pp. 2648-55.

Stamm, A., Tengdelius, M., Engström, J., Schmidt, B., Syrén, P.O., Fogelström, L. and Jonsson, E.M., 2019. Chemo-enzymatic pathways toward pinene-based renewable materials. *Green Chemistry*, 21, 2720-31

Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A., Fan, T.W.M., Fiehn, O., Goodacre, R., Griffin, J.L. and Hankemeier, T., 2007. Proposed minimum reporting standards for chemical analysis. Metabolomics, 3(3), pp.211-221.

Tee, K. L. & Schwaneberg, U. (2006). 'A screening system for the directed evolution of epoxygenases: importance of position 184 in P450 BM3 for stereoselective styrene epoxidation', *Angew Chem Int Ed Engl,* 45(32), pp. 5380-3.

Truan, G., Komandla, M.R., Falck, J.R. and Peterson, J.A., 1999. P450BM-3: absolute configuration of the primary metabolites of palmitic acid. *Arch Biochem Biophys*, 366(2), pp.192-8.

van Vugt-Lussenburg, B. M., Stjernschantz, E., Lastdrager, J., Oostenbrink, C., Vermeulen, N. P. & Commandeur, J. N. (2007). 'Identification of critical residues in novel drug metabolizing mutants of cytochrome P450 BM3 using random mutagenesis', *J Med Chem,* 50(3), pp. 455-61.

Vaz, A. D., McGinnity, D. F. & Coon, M. J. (1998). 'Epoxidation of olefins by cytochrome P450: evidence from site-specific mutagenesis for hydroperoxo-iron as an electrophilic oxidant', *Proc Natl Acad Sci U S A, 95*(7), pp. 3555-60.

Whitehouse, C. J., Bell, S. G. & Wong, L. L. (2012). 'P450(BM3) (CYP102A1): connecting the dots', *Chem Soc Rev,* 41(3), pp. 1218-60.

Wróblewska, A. (2014). 'The epoxidation of limonene over the TS-1 and Ti-SBA-15 catalysts', *Molecules,* 19(12), pp. 19907-22.

Yeom, H. & Sligar, S. G. (1997). 'Oxygen activation by cytochrome P450BM-3: effects of mutating an active site acidic residue', *Arch Biochem Biophys,* 337(2), pp. 209-16.

Yeom, H., Sligar, S. G., Li, H., Poulos, T. L. & Fulco, A. J. (1995). 'The role of Thr268 in oxygen activation of cytochrome P450BM-3', *Biochemistry,* 34(45), pp. 14733-40.

Yuta, A. & Baraniuk, J. N. (2005). 'Therapeutic approaches to mucus hypersecretion', *Curr Allergy Asthma Rep,* 5(3), pp. 243-51.

2.7. Supplementary Information

Table 2.S1 Pairs of m/z fragments used to identify mono-hydroxylated fatty acid products.

Fatty acid (C:n)	*m/z* Fragment paris for corresponding mono-hydroxylated fatty acid products								
	ω -9	*ω -8*	*ω -7*	*ω -6*	*ω -5*	*ω -4*	*ω -3*	*ω -2*	*ω -1*
Dodecanoate C:12	233/229	247/215	261/201	275/187	289/173	303/159	317/145	331/131	345/117
Tetradecanoate C:14	261/229	275/215	289/201	303/187	317/173	331/159	345/145	359/131	373/117
Hexadecanoate C:16	289/229	303/215	317/201	331/187	345/173	359/159	373/145	387/131	401/117

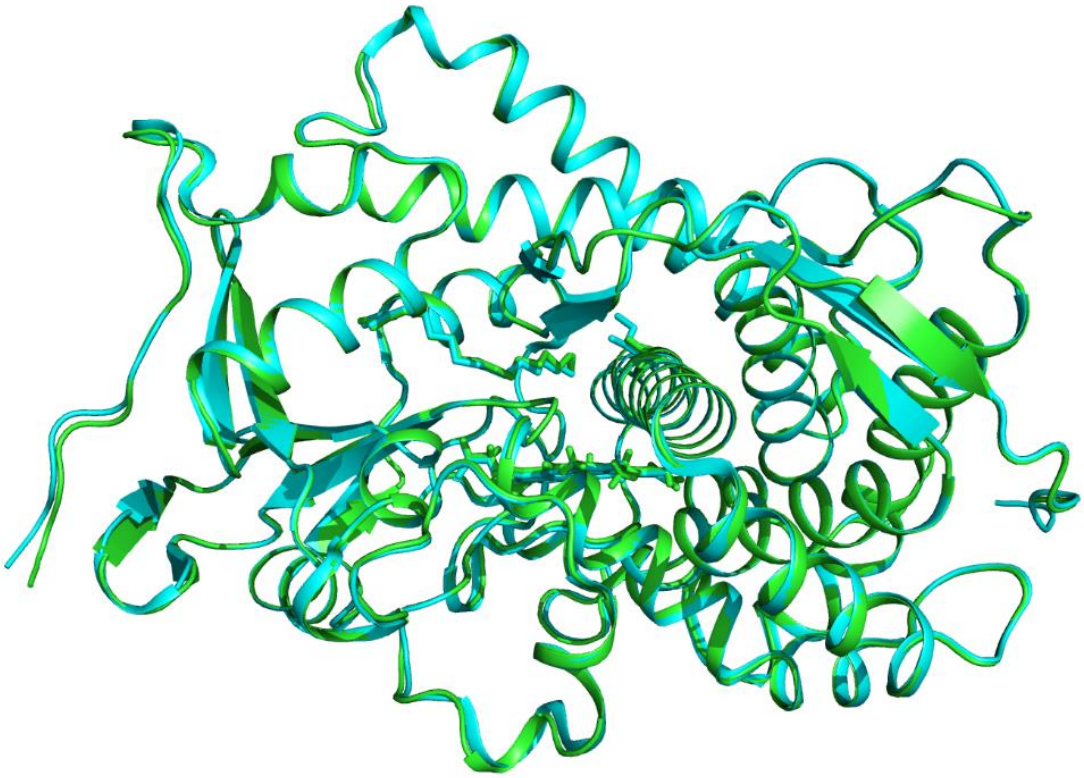

Figure 2.S1 **Secondary structural overlay of wild-type (PDB 1JPZ, cyan) and variant I263P (green) heme domains.** This overlay generated an RMSD score of 0.320, indicating minimal topological differences between the two heme domains. Residues Ile263 (WT) and Pro263 (I263P) are highlighted within the I-helices of both structures.

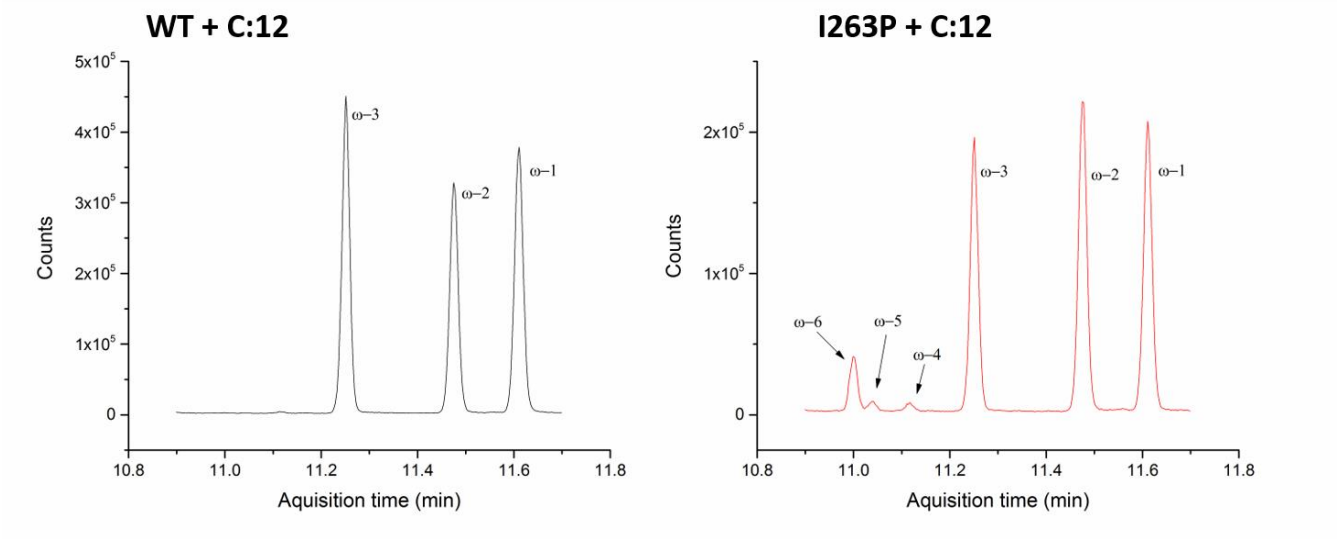

Figure 2.S2 **GC Chromatogram of the product profiles produced by intact WT BM3 (left hand insert) and intact variant I263P (right hand insert) against fatty acid substrate dodecanoic acid.** Mono-hydroxylated products are indicated**.** Positions of hydroxylation are given relative to the ω-carbon.

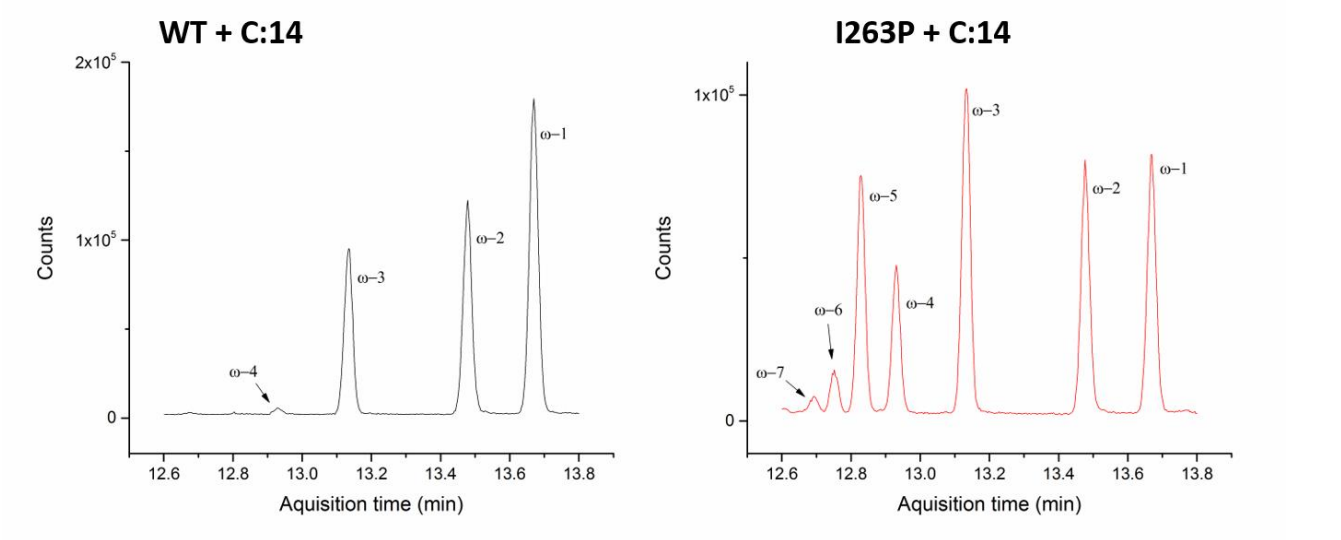

Figure 2.S3 **GC Chromatogram of the product profiles produced by intact WT BM3 (left hand insert) and intact variant I263P (right hand insert) against fatty acid substrate tetradecanoic acid.** Mono-hydroxylated products are indicated. Positions of hydroxylation are given relative to the ω-carbon.

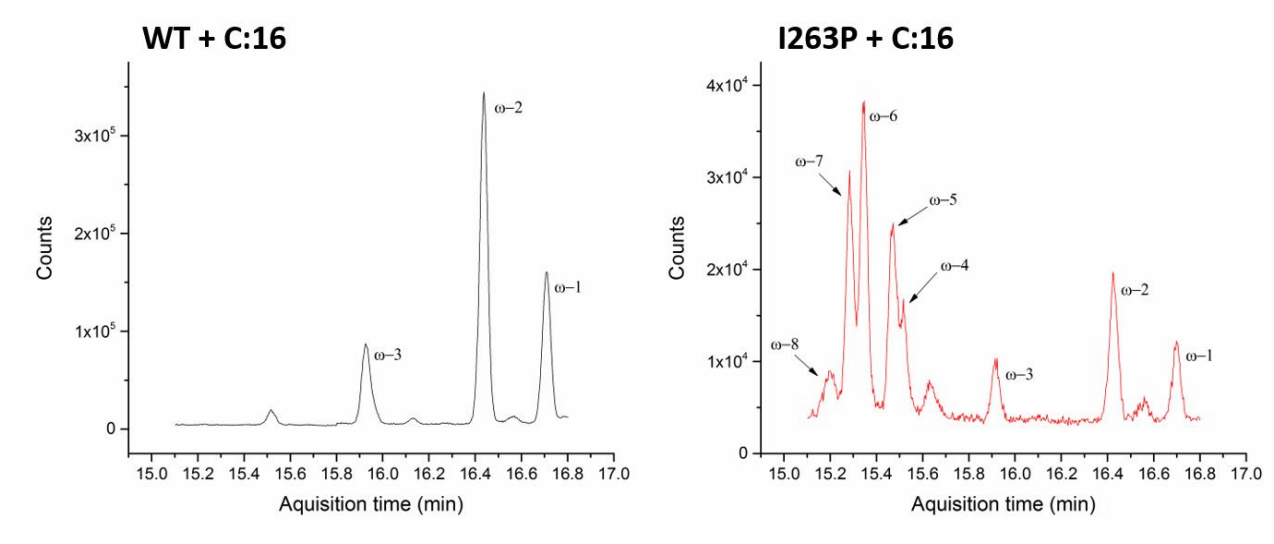

Figure 2.S4 **GC Chromatogram of the product profiles produced by intact WT BM3 (left hand insert) and intact variant I263P (right hand insert) against fatty acid substrate hexadecanoic acid.** Mono-hydroxylated products are indicated. Positions of hydroxylation are given relative to the ω-carbon.

Chapter 3: "Refocusing the Regio- And Inverting the Enantio-Selectivity of P450 BM3: Refashioning the Activity of CYP102A1 towards Fatty Acid Hydroxylation and Styrene Epoxidation."

3.1. Abstract

When attempted using traditional approaches to synthetic chemistry, oxidative functionalisation of C-H bonds with a high degree of regio- or enantioselectivity can often be challenging to achieve. Alternative enzymatic approaches to C-H bond oxidation, which can produce valuable oxidised products with high levels of both regio- and sterospecificity and in a single step, are therefore highly desirable. Cytochromes P450 (P450s), a *b*-heme containing superfamily of mono-oxygenase biocatalysts with an unparalleled number of accepted substrates, are well suited for use throughout synthetic biology. Cytochrome P450 BM3 (CYP102A1) a highly selective, high activity mono-oxygenase existing as a single, soluble catalytic system has proven a popular choice for industrial exploitation.

In a prior study, introduction of a proline residue within the I-helix motif of BM3 prompted significant diversification in the regioselective hydroxylation of fatty acid substrates relative to the wild-type enzyme (Whittall et al., unpublished data). In addition, the variant (I263P) displayed alterations in the enantioselective epoxidation of styrene, producing a near-racemic mix of *R*(+) and *S*(-) enantiomers of styrene oxide.

This study focused on the rational engineering of the novel CYP102A1 variant (I263P) towards the production of high value functionalised fatty acid compounds hydroxylated at mid-chain positions. This was successfully achieved with a triple BM3 variant (I263P/F87V/V78A), which refocused hydroxylation of C:12-C:16 fatty acids from the ω-1, ω-2 and ω-3 positions generated by the wild-type enzyme to the ω-5, ω-6 and ω-7 positions.

An additional aim of this study was to push production of the *S*(-)enantiomer of styrene oxide to an enantiomeric excess. A second triple BM3 variant (I263P/F87V/A184K) was found to be capable of producing the *S*(-)-styrene oxide enantiomer to an excess of 56.7 ± 0.306 % *ee***,** the highest yet reported for a BM3 variant.

3.2. Introduction

Oxidation of C-H bonds is an important goal within synthetic chemistry. Owing to their ubiquity throughout organic compounds, oxidative transformation of C-H bonds offers enormous potential in the synthesis of a broad array of compounds. However, achieving C-H bond oxidation with a high degree of region- and stereo-specificity often poses a significant challenge when attempted using traditional, often multistep methods of chemical synthesis (Godula and Sames, 2006). These reactions often require the use of strong oxidising agents, metal catalysts or directing groups, all of which possess significant limitations (Pazmino et al., 2010; Godula and Sames, 2006; Desai et al., 2004). Strong oxidising agents and metal catalysts can either be indiscriminate in their hydroxylation, or strongly favour only the most reactive C-H bond within the compound of interest (Fasan 2012; Pazmińo et al., 2010). In addition, directing groups can be difficult to install and remove from substrates depending on both of their respective structures, potentially restricting substrate selectivity (Rousseau and Breit, 2011).

Consequently, there is a demand for catalytic systems able to functionalise C-H bonds in a highly selective, single-step manner. Furthermore, these systems must ideally be capable of efficient operation under milder reaction conditions to those offered by chemical methods (Fasan, 2012). Cytochromes P450 (P450s), a highly versatile superfamily of mono-oxygenase enzymes containing prosthetic *b*-heme groups, are well suited to these requirements (Urlacher and Girhard, 2012). In addition to performing hydroxylation in a highly regio and stereospecific manner, P450s catalyse an extensive repertoire of reactions including epoxidation, demethylation and heteroatom abstraction (Lamb and Waterman, 2013; Munro et al., 2007). This extensive array of reactions renders P450s highly attractive for industrial exploitation (Urlacher and Girhard, 2019).

However, industrial application of many P450s is limited by the requirement of expensive electron donating cofactors and their corresponding regeneration systems, coupled to low rates of turnover and impaired stability (Sakaki, 2012). Industrial applications of P450s are further complicated by the highly specific relationship that underpins efficient electron transfer - that between the P450 and its corresponding redox partner proteins (typically cytochrome P450 reductases [CPRs] in eukaryotes and ferredoxin reductases/ferredoxins in prokaryotes). Many of these barriers can be overcome by the use of engineered variants of a unique, high-activity prokaryotic P450 that functions as part of a single soluble redox system – P450 BM3 (CYP102A1) (Eiben et al. 2006; Narhi and Fulco, 1986). Isolated from the soil-dwelling bacterium *Bacillius megaterium*, BM3 is a natural fusion enzyme in which

the N-terminal heme domain (P450 domain) is covalently linked to the C-terminal diflavin reductase domain (CPR domain) (Munro et al., 2002; Narhi and Fulco, 1986). This unusual arrangement is hypothesised to enable rapid electron transfer from the electron-donating cofactor (NADPH), via the reductase and on to the heme (Munro et al., 1996). Consequently, BM3 displays the highest rate of turnover yet observed for a mono-oxygenase enzyme (~17,000 min^{-1} with the substrate arachidonic acid) (Noble et al., 1999). The natural substrates and exact physiological role of BM3 are yet to be confirmed (Whitehouse, et al., 2012). However, as branched-chain fatty acids have been shown to induce expression of BM3 when applied exogenously, in addition to comprising approximately 90% of the fatty acid content within *B. megaterium*, it has been suggested that they are the natural substrates of BM3 (Cryle et al., 2006; English et al., 1997).

Fatty acids are the natural substrates for several P450 classes (Hammerer et al., 2018). Hydroxylated fatty acid products are used throughout the chemical, cosmetic and food industries – often as precursors in the synthesis of emulsifiers, lubricants and stabilisers (Kim and Oh, 2013). As a fatty acid hydroxylase, the activity of BM3 towards saturated and unsaturated fatty acids has been extensively studied (Whitehouse et al., 2012). Initial studies by Miura and Fulco (1986) demonstrated that BM3 hydroxylated a range of medium to long chain (C:12 – C:22) fatty acids at sub-terminal positions. With the substrate lauric acid (C:12), BM3 generates products hydroxylated at the ω-1, ω-2 and ω-3 positons in a relatively even distribution (Miura and Fulco, 1975).

Styrene oxide is a highly versatile chemical intermediate utilised in the synthesis of a variety of optically active pharmaceutical compounds (Shehzad et al., 2013; Panke et al., 1998; Bloom et al., 1992; Takahashi et al., 1989). In addition, styrene oxide serves as an important precursor in the production of five-membered cyclic carbonates, which have extensive applications throughout the agricultural, cosmetic and fuel industries (Sun et al., 2005). However, the production of optically pure styrene oxide can prove challenging (Katsuki 2001; Panke et al., 1998). Enzyme-mediated production of styrene oxide to a high degree of enantiomeric excess is therefore highly desirable (Tee and Schwaneberg, 2006; Panke et al., 1998).

A single point mutation (I263P) within the catalytically essential I-helix motif of BM3 was shown to broaden the regioselectivity of fatty acid hydroxylation relative to the wild-type enzyme (Whittall et al., unpublished data). The product profile of hydroxylated fatty acids produced by variant I263P tended towards novel positions away from the ω-terminus.

Hydroxylation of fatty acids occurring at mid-chain positions, such as the γ and δ carbons, serve as valuable precursors in the production of lactones, aroma compounds important in the formulation of apricot, peach and strawberry flavourings (Dietrich et al., 2009; Waché et al., 2003). In addition, the BM3 I263P variant inverted the enantioselectivity of styrene epoxidation from the production of an enantiomeric excess of (*R*)-styrene oxide to a near-racemic mix of both (*R*) and (*S*) enantiomers of styrene oxide (Whittall et al., unpublished data).

The aim of this study was to build upon the I263P mutation in order to achieve further alterations to regio and enantioselectivity of BM3. A systematic approach to mutagenesis led to the creation of two triple mutants, TM1 (I263P/F87V/V78A) and TM2 (I263P/F87V/A184K). The goal of the first triple mutant was to push regioselectivity of fatty acid hydroxylation towards valuable mid-chain positions. The objective of the second triple mutant was to further alter enantioselective epoxidation of styrene to favour production of the (*S*)-enantiomer of styrene oxide.

In tandem with I263P, the BM3 F87V mutation was incorporated into both triple variants. This was to enable greater flexibility of the fatty acid terminus within the active site, in addition to allowing the bulkier substrate styrene greater access to the heme, ideally increasing levels of styrene epoxidation (Hammerer et al., 2018; Whitehouse et al., 2012). Despite initial reports by Oliver et al. (1997), multiple studies of variants containing F87V or F87A substitutions indicate that the focus of hydroxylation is shifted away from the ω-terminus and towards the centre of the fatty acid substrate (Whitehouse et al., 2012; Dietrich et al., 2009; Cirino and Arnold, 2002; Cirino and Arnold, 2003).

In the TM1 variant, the active site was further enlarged through the incorporation of a V78A substitution. This was to enable greater substrate flexibility within the active site channel, with the aim of creating alternate substrate binding modes and hence produce novel patterns of hydroxylation with a range of fatty acid substrates (Hammerer et al., 2018; Dietrich et al., 2009).

In the TM2 variant, a charged residue, lysine, was substituted in lieu of an alanine residue at position 184 situated in the F-helix motif. Previous studies have reported that the inclusion of charged residues at positions distant from the active site can have dramatic effects on the enantiopreference of styrene oxide production in variants of BM3 (Shehzad et al., 2013; Huang et al., 2011; Tee and Schwanenberg, 2006). Notably, variants

incorporating A184K substitutions have been shown to push production of styrene oxide towards an excess of the (*S*)-enantiomer (Shehzad et al., 2013; Tee and Schwanenberg, 2006).

Fatty acid and styrene oxide product profiles will be assessed using gas chromatography-mass spectrometry (GC-MS). Structural rationale for alterations to regio and stereoselectivity displayed by each variant will be provided by Resonance Raman, X-ray crystallography and molecular docking simulations. In addition, catalytic and binding activity of the variants will be examined through steady-state kinetics and binding titrations, respectively.

3.3. Methods

3.3.1. Generation, Expression and Purification of Variant BM3 Enzymes

Variant BM3 constructs were generated by site-directed mutagenesis and amplified using PCR. An Agilent QuikChange Lightning site-directed mutagenesis kit was used for mutagenesis and a Techne ^{3}Prime thermocycler (Bibby Scientific, UK) was utilised for PCR. Oligonucleotide primers were designed using QuickChange software (Agilent Technologies, UK) and were produced by Eurofins. Oligonucleotide primer sequences are displayed in Table 3.1.

Table 3.1 Sequences of oligonucleotide primers used in site-directed mutagenesis by PCR

Primer	5' > 3'
F87V_Fwd	ttgcaggagacgggttagttacaagctggacgcatg
F87V_Rev	catgcgtccagcttgtaactaacccgtctcctgcaa
V78A_Fwd	agtcaagcgcttaaatttgcacgtgattttgcaggagac
V78A_Rev	gtctcctgcaaaatcacgtgcaaatttaagcgcttgact
A184K_Fwd	cgctgcagcttgttcatttttttcatccagtgcacggac
A184K_Rev	gtccgtgcactggatgaaaaaatgaacaagctgcagcg

DNA corresponding to the BM3 variant I263P was used as a template for both the intact BM3 variant constructs and their related heme domains. Intact BM3 variant constructs were generated in pET15b plasmids containing a hexa-histidine purification tag. Variant BM3 heme domains were generated in the pET20b vector and were expressed un-tagged. Subsequent mutagenesis of the BM3 variant I263P was performed in a sequential manner, with the I263P pET15b and pET20b constructs serving as the basis for the I263P/F87V double mutant (DM) variant. Likewise, the DM pET15b and pET20b constructs served as the basis for the triple mutant variant constructs I263P/F87V/V78A (TM1) and I263P/F87V/A184K (TM2). Plasmid constructs were transformed into Novablue competent

cells (Novagen, UK) according to the method presented in Sambrook et al. (1989). Subsequently, plasmid DNA was extracted and purified according to the Qiagen MiniPrep protocol. Successful amino acid substitutions were confirmed in each of the variant BM3 DNA samples via Sanger sequencing. Primers used in the sequencing reactions are presented in Table 3.2. All sequencing reactions were performed by Source Bioscience (Rochdale, UK).

Table 3.2 Oligonucleotide primers used in sequencing reactions

Primer	5' > 3'
SeqPR1	tgagccgcttgatgacgagaac
SeqPR2	tcgcaacgcttgattcac
SeqPR3	gaatttatcgcccttctgcc
VecPR	atacccacgccgaaacaag
T7Forward	taatacgactcactataggg

Subsequent expression and purification of intact variant BM3 enzymes and their corresponding heme domains followed the protocol detailed in Whittall et al. (unpublished data).

3.3.2. Generation of P450 CO-Bound Spectra and Calculation of the Concentrations of Variant BM3 Enzymes

Protein concentrations of all BM3 variants were calculated by application of the Beer-Lambert law as outlined by Omura and Sato (1964). Extinction coefficients of ε_{418} = 95 mM^{-1} cm^{-1} and ε_{418} = 105 mM^{-1} cm^{-1} were used for the heme domains and intact BM3 proteins, respectively.

CO-bound P450 spectra were collected for each variant BM3 heme domain. All spectra were collected using UV-visible Cary 50 spectrophotometers (Agilent, UK). Samples were prepared in 1 ml quartz cuvettes with path lengths of 1 cm. Protein samples were diluted to a final concentration of ~5 μM using 100 mM KPi, pH 7.5. Protein samples were reduced to the ferrous oxidation state by the addition of sparing amounts of sodium dithionite. Approximately 1 cm^3 of CO gas was gradually introduced into the stoppered cuvette via a syringe. Spectra were recorded in the wavelength range of 250 – 700 nm at 1 minute intervals over a 1 hour period until no further ferrous-CO adduct was formed. Data were subsequently processed using Origin Pro software (OriginLab, USA).

3.3.3. Binding Titration Determination of Substrate Dissociation Constants

Substrate binding affinity of variant BM3 heme domains against all tested substrates was assessed using UV-visible binding titrations. Binding affinities were expressed as

dissociation constants (K_d values). Protein samples were prepared to a concentration of 2 – 5 uM in 100 mM KPi, pH 7.5. Assays were conducted in quartz cuvettes with path lengths of 1 cm. Substrate stocks were prepared in absolute EtOH and were successively titrated into the cuvette using volumes of between 0.1 – 0.5 μl. Hamilton syringes (Hamilton, USA) were used in all titrations to ensure accuracy. Spectra were recorded between 200 – 700 nm following sequential substrate additions using Cary 50 UV-visible spectrophotometers. Resulting spectral data were processed and fitted using OriginPro software as outlined in Whittall et al. (unpublished data). The majority of data were fitted using the Michaelis-Menten function, while tight-binding substrates were fitted using the Morrison function (Whittall et al., unpublished data).

3.3.4. Steady-State Kinetic Activity of Intact BM3 Variants against Fatty Acid Substrates

Steady-state kinetic assays used 100 nM intact variant BM3 enzymes in the case of fatty acid substrates lauric (C_{12}) and myristic (C_{14}) acids. For the fatty acid substrate palmitic (C_{16}) acid, 80 nM intact BM3 variant was used. The concentration of intact BM3 protein relative to each substrate was kept constant across all variants. All assays were performed at 30 °C in 100 mM KPi, pH 7.5. The NADPH cofactor was used at the near-saturating concentration of 100 μM. Substrate-dependent oxidation of NADPH was recorded at 340 nm over a time-course of 30 seconds using a Cary 50 UV-vis spectrophotometer. The initial rate of NADPH oxidation was recorded at each substrate concentration using the NADPH extinction coefficient (ε_{340} = 6.21 mM^{-1} cm^{-1}). Data points represented the average initial rate (ΔA_{340}) of 3 assays plotted against the corresponding concentration of the fatty acid substrate. The kinetic parameters K_M and k_{cat} were obtained by fitting the data using the Michaelis-Menten hyperbolic function with OriginPro software.

3.3.5. Attempts to Obtain Crystal Structures of Substrate-Bound BM3 Variant Heme Domains

A crystal structure of the NPG-bound I263P variant heme domain was previously solved (Whittall et al., unpublished data). In light of this, attempts were made to obtain crystal structures of the DM, TM1 and TM2 variant BM3 heme domains bound to the substrate N-palmitoylglycine (NPG). Purified variant heme domains were near-saturated with 500 μM NPG, added from a 5 mM stock concentration prepared in 50 mM potassium carbonate. High-spin variant heme domain species were confirmed by UV-Vis spectroscopy prior to their addition to a variety of 96 well crystallisation screens (Molecular Dimensions,

Newmarket, United Kingdom) for protein crystallisation. Protein concentration was fixed at 9 mg/mL across all screening conditions.

Initial screening plates included the drop volume conditions of 200, 400 and 600 nl of protein at a 1:1 ratio with the screening plate mother-liquor. Plates were stored at 4 °C following protein application, and were left undisturbed for one week to facilitate protein crystallisation.

No seeding or diffraction quality crystals have yet been produced for the DM enzyme, despite alterations to both protein concentrations (increased to 12 and 15 mg/mL) and drop volume.

Protein crystals suitable for use for seeding were obtained for both TM1 and TM2. The SG1 screen yielded seed-quality crystals for NPG-bound TM1 in plate condition 1-33 (0.2 M magnesium chloride hexahydrate, 0.1 M Tris, pH 8.5, 25 % w/v PEG 3350). Seed-quality crystals were also obtained for NPG-bound TM2 in PACT screen condition 2-16 (0.2 M potassium thiocyanate, 0.1 M Bis-Tris propane, pH 6.5, 20 % w/v PEG 3350).

Following seeding at a 10:1 ratio of protein to seed stock, crystals suitable for diffraction were obtained for both the TM1 and TM2 heme domains. X-ray diffraction was performed by Dr Colin Levy at the Diamond light source synchrotron (Harwell, United Kingdom). Following molecular replacement and refinement using the programs Coot (Emsley et al., 2010) and PHENIX (Adams et al., 2010) 2.1 Å and 2.2 Å resolution structures were obtained for TM1 and TM2 variant heme domains, respectively. The resulting structure underwent a final refinement using the online application PDBRedo (Joosten et al., 2009). Refinement of both TM1 and TM2 structures was performed with the assistance of Dr Colin Levy and Dr Richard Tunnicliffe (University of Manchester).

However, despite the presence of ligand electron density within the active site pocket, the positioning of the NPG ligand could not be clearly resolved for either of the TM1 and TM2 structures. Despite this, the secondary structural topography of both the TM1 and TM2 active sites suggested that the enzymes had adopted substrate-bound conformations. Consequently, the ligand-free TM1 and TM2 heme domains were used for computational substrate docking in order to simulate the position of the substrates within the active sites.

3.3.6. Computational Docking Simulations Using YASARA Software

Docking simulations were performed using YASARA (version 18.2.7) with the assistance of Nico Kress (University of Manchester). Protein-ligand interactions were predicted using the

built in VINA software. The ligands palmitic acid (C:16) and styrene were docked into the substrate-free structures of the TM1 and TM2 heme domains. Contaminant molecules of polyethylene glycol (PEG), arising from the conditions of the screening plates, were removed prior to commencing docking studies. Docking simulations with the wild-type BM3 heme domain were performed using the substrate-bound heme domain (PDB: 1JPZ), in which the NPG substrate had been removed prior to docking. Across all simulations, a 12 Å simulation cell (12 x 12 x 12 cube) was constructed around residue A264, situated directly above the heme iron within the I-helix, in order to encompass the active site channel. All simulations were performed using the AMBER03 force field. In each simulation, a total of 25 docking runs were performed and scored according to the YASARA scoring function.

3.3.7. Assessment of Thermal Stability of Variant BM3 Heme Domains

Stability of substrate-free variant BM3 heme domains was assessed using a thermal ramp controlled by an UNcle instrument (Unchained Labs, Pleasanton, USA). Variant heme domain samples were diluted to a concentration of 1 mg/ml in 25 mM KPi buffer (pH ~7.5). Samples were subjected to a thermal ramp of 20 – 90 °C at a gradient of 1 °C / minute. Thermostability was expressed as the parameter T_m and obtained from dynamic light scattering (DLS) data generated by following changes to fluorescence intensity of the sample as a function of temperature.

3.3.8. Resonance Raman Analysis of Variant BM3 Heme Domains

Resonance Raman spectra of variant BM3 heme domains were collected in both the substrate-free and substrate-bound states. All samples were run at a protein concentration of 50 μM in 25 mM KPi buffer (pH 7.5). With substrate-bound samples, NPG (prepared to a stock concentration of 5 mM in 50 mM K_2CO_3) was present at a final saturating concentration of 25 μM. Samples were irradiated using a 405 nm laser and the resulting spectra were measured using an InVia Raman Renishaw microscope. The 405 nm laser was calibrated through a x50 objective at 520 cm^{-1} using a silicon reference chip. All spectra were collected between 200 to 2000 cm^{-1} with the assistance of Bethan McAvan (University of Manchester).

3.3.9. GC-MS Analysis of Products Generated by Intact Variant BM3 Enzymes

The concentration of intact P450 BM3 variant enzymes was fixed at 2 μM across all reactions. The final reaction volume was 1 ml. NADPH was used at a saturating concentration of 500 μM, in addition to a regeneration system comprising of 7.76 mM glucose-6-phosphate, 1.5 units of glucose-6-phosphate dehydrogenase and 0.6 mM $NADP^+$. All reagents were prepared in 25 mM KPi, pH 7.5, with the exception of substrate stocks,

which were prepared in absolute EtOH. Final substrate concentrations were 200 μM for fatty acid substrates and 2 mM for styrene.

Reactions were done in a shaking incubator set at 37 °C for either 60 minutes (fatty acid substrates) or ~20 hours (styrene). Products were separated from the reaction mixtures by a 2:1 solvent extraction. Hydroxylated fatty acid products were extracted in dichloromethane (DCM) and functionalised styrene products were extracted in ethyl acetate. Residual buffer was removed from each sample by the addition of magnesium sulfate.

To facilitate product identification, fatty acid turnover reactions were subjected to derivatisation using the silylating reagent N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) plus 1% trichloromethyl silane (BSTFA + 1% TMCS). Samples were derivatised at a 1:1 ratio of sample to reagent at 60 °C for 60 minutes within a shaking thermomixer (Fisher Scientific, UK). Styrene turnover reactions did not undergo derivatisation.

Samples containing fatty acids were run on an Agilent VF-5ms column (30 M x 0.25 mM x 0.25 μM) mounted to a 7890 GC system, itself paired with a 5975 Series MSD mass spectrometer (Agilent, Cheadle, United Kingdom). Mono-hydroxylated fatty acid products were separated by two temperature ramps, spanning 50 – 220 °C and 220 – 310 °C. The initial ramp proceeded at a 20 °C/min gradient, with the second ramp proceeding at a reduced gradient of 2 °C/min. Mono-hydroxylated fatty acid products were identified through pairs of silylated fragments specific to each product and analysed using MassHunter software (Agilent, UK) (Supplementary information, Table 3.S1). Product profiles of mono-hydroxylated fatty acids were calculated using chromatogram peak areas specific to each product as indicated by Qualitative Analysis software (Agilent, UK).

Samples containing functionalised styrene products were analysed using an Agilent DB-WAX column (30 M x 0.25 mM x 0.25 μM) attached to a 7890-B series GC system, itself coupled to a 7890A series MSD mass spectrometer (Agilent, UK). Products were separated by two ramps, spanning 50 – 68 °C and 68 – 240 °C. The temperature gradient of the initial ramp was 5 °C/min. The subsequent ramp proceeded at the steeper gradient of 25 °C/min.

Following the confirmed presence of styrene oxide product (identified via standards and NIST), the enantiomeric ratio of styrene oxide products was quantified using a second round of GC analysis. *R*(+) and *S*(-) styrene oxide enantiomers were separated using a CP Chirasil-Dex CB column (25 M x 0.25 mM x 0.25 μM) mounted to a 7890A series GC

(Agilent, UK). Once again, two temperature ramps were used to enable optimal product separation, with the initial ramp spanning 40 – 100 °C (5 °C/min) and the second 100 – 150 °C (2 °C/min). Ratios of *R*(+) and *S*(-) enantiomers of styrene oxide were assigned and quantified using standards and peak area integration (respectively), with concentrations being calculated using calibration curves.

Dr Katherine Hollywood and Dr Helen Toogood (University of Manchester) assisted with the running of samples and the use of the GC-MS instruments.

3.4. Results

3.4.1. Binding Dissociation Constants (K_d) of BM3 Variant Heme Domains Against Fatty Acid Substrates

Dissociation constants (K_d) for wild-type and variant BM3 heme domains with lauric, myristic and palmitic fatty acid substrates and the fatty acid analogue N-palmitoyl glycine (NPG) are presented in Table 3.3.

Table 3.3 Dissociation constants (K_d) of fatty acid substrates for wild-type[1] and variant BM3 heme domains

[1]Whittall et al. (unpublished data)

BM3 heme variant	Fatty acid substrate dissociation constant (K_d) (μM)			
	Lauric (C_{12})	Myristic (C_{14})	Palmitic (C_{16})	NPG (C_{18})
Wild-type[1]	108 ± 3.51	12.6 ± 1.02	2.98 ± 0.150	0.474 ± 0.0870
I263P[1]	27.2 ± 1.61	2.33 ± 0.171	0.470 ± 0.090	0.0661 ± 0.0170
DM (I263P/F87V)	16.9 ± 1.30	1.08 ± 0.110	0.103 ± 0.0380	0.174 ± 0.0600
TM1 (I263P/F87V/V78A)	75.5 ± 5.30	2.33 ± 0.0880	0.431 ± 0.0290	0.590 ± 0.0430
TM2 (I263P/F87V/A184K)	162 ± 9.51	12.6 ± 0.610	1.48 ± 0.110	1.30 ± 0.0700

As with the wild-type heme domain, binding affinities displayed by variants DM, TM1 and TM2 with saturated fatty acid substrates increased with the concurrent increase in their carbon chain length. Inclusion of the point mutation F87V in addition to I263P in the DM variant resulted in further increases in binding affinity for all of the saturated fatty acid substrates. However, addition of the further V78A mutation reverses this trend, with the TM1 variant displaying reduced binding affinities for saturated fatty acid substrates relative to the DM variant. Despite this, the TM1 variant still displays tighter binding affinities for all the saturated fatty acid substrates in comparison to the affinities for the wild-type BM3

heme domain. Conversely, the alterations to saturated fatty acid binding affinity displayed by the TM2 variant are markedly different compared to the DM and TM1 variants. Addition of an A184K substitution to the DM (TM2) variant led to a ~9.5-fold decrease in lauric acid binding affinity relative to the DM variant. Furthermore, the reduction in binding affinity is such that the TM2 variant displays lower affinity for lauric acid (K_d = 162 ± 9.51 μM) than the unmodified wild-type heme domain (K_d = 108 ± 3.51 μM). Interestingly, despite the presence of multiple amino acid substitutions, the TM2 variant displays a near identical binding affinity for myristic acid (12.6 ± 0.610 μM) compared to the wild-type enzyme (12.6 ± 1.02 μM). For the final saturated fatty acid examined, palmitic acid, TM2 displays a near 2-fold increase (1.48 ± 0.110 μM) in binding affinity in comparison to the wild-type (2.98 ± 0.150 μM). Despite this, the introduction of the A184K mutation to the DM (TM2) variant caused a dramatic decrease in binding affinity towards palmitic acid relative to the DM variant, as illustrated by the K_d values of 0.103 ± 0.0380 μM (DM) and 1.48 ± 0.110 μM (TM2), respectively.

All variant BM3 heme domains displayed tight binding affinities for the fatty acid analogue NPG. However, only the DM displayed a K_d value (0.174 ± 0.0600 μM) marginally tighter than that displayed by the wild-type heme domain (0.474 ± 0.870 μM). Both the TM1 and TM2 heme domains showed reduced binding affinity for NPG relative to the wild-type heme domain. However, the K_d value of 0.590 ± 0.0430 μM produced by TM1 with the NPG substrate lies within the margin of error of the corresponding wild-type K_d value (0.474 ± 0.0870 μM). Moreover, binding affinities for NPG decreased relative to those for palmitic acid in both the DM and TM1 mutants in a departure from the wild-type heme domain. The TM2 variant displayed comparable binding affinities for both palmitic acid and NPG, with both K_d values falling within the margin of error of the other.

In summary, all BM3 variants displayed tightest binding affinities for fatty acids substrates within the range of C_{16} – C_{18}, akin to the preferences of the wild-type heme domain. The DM and TM1 variants displayed consistently tighter binding affinities towards fatty acid substrates than the wild-type heme domain. Conversely, the TM2 variant displayed reduced binding affinities towards all tested fatty acid substrates relative to the wild-type heme domain.

3.4.2. Measurement of Intact Variant BM3 Enzyme Kinetics in the Steady-State

Steady state kinetic measurements were recorded for each of the intact variant BM3 enzymes. Studies examining substrate-dependent oxidation of NADPH were performed

against laurate, myristate and palmitate acid substrates for each BM3 variant. The resulting catalytic parameters K_M (Michaelis constant), k_{cat} (turnover number) and k_{cat}/K_M (catalytic efficiency) for each BM3 variant are presented in Table 3.4. Catalytic data collected for intact wild-type BM3 is included within Table 3.4 for the purposes of comparison (Whittall et al., unpublished data). Values for catalytic parameters presented in Table 3.4 were obtained from Figure 3.1 (DM), Figure 3.2 (TM1) and Figure 3.3 (TM2).

Table 3.4 Kinetic data for intact wild-type[1] and variant BM3 proteins in steady-state using saturated fatty acid substrates.
[1]Whittall et al. (unpublished data)

Intact BM3 variant	Laurate (C_{12})			Myristate (C_{14})			Palmitate (C_{16})		
	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\mu M^{-1} s^{-1}$)	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\mu M^{-1} s^{-1}$)	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\mu M^{-1} s^{-1}$)
Wild-type[1]	281 ± 43.9	57.6 ± 3.90	0.205	31.1 ± 3.01	63.9 ± 1.63	2.06	10.3 ± 1.40	86.2 ± 2.27	8.37
DM (I263P/F87V)	27.8 ± 6.29	9.55 ± 0.390	0.343	13.0 ± 2.14	17.9 ± 0.619	1.38	10.1 ± 1.88	31.6 ± 1.13	3.13
TM1 (I263P/F87V/V78A)	185 ± 19.9	32.5 ± 1.20	0.178	27.1 ± 3.44	35.6 ± 1.14	1.33	10.8 ± 2.19	35.0 ± 1.39	3.24
TM2 (I263P/F87V/A184K)	40.2 ± 0.170	9.05 ± 0.424	0.223	23.1 ± 5.10	11.7 ± 0.591	0.556	16.4 ± 3.47	15.6 ± 0.722	0.995

As indicated by the corresponding k_{cat} values, the enzyme turnover rate displayed by all the BM3 variants for each of the tested fatty acid substrates was decreased relative to the wild-type enzyme. k_{cat} values obtained for the DM, TM1 and TM2 variants against laurate were 9.55 ± 0.390, 32.5 ± 1.20 and 9.05 ± 0.424 s^{-1}, respectively. In comparison, the intact wild-type BM3 enzyme displayed a turnover number of 57.6 ± 3.90 s^{-1} with lauric acid as substrate (Whittall et al., unpublished data). A similar drop in activity was observed for each variant with the substrate myristate, with variants DM, TM1 and TM2 displaying k_{cat} values of 17.9 ± 0.619, 35.6 ± 1.14 and 11.7 ± 0.591 s^{-1}, respectively. These values stand in contrast to those obtained for the wild-type enzyme at 63.9 ± 1.63 s^{-1}. All variant enzymes display corresponding, albeit less severe, decreases in catalytic activity for palmitate, with

DM, TM1 and TM2 variants generating k_{cat} values of 31.6 ± 1.13, 35.0 ± 1.39 and 15.6 ± 0.722 s^{-1}, respectively, compared to the wild-type value of 86.2 ± 2.27 s^{-1}.

As indicated by the corresponding K_M values, all the variants display increased binding affinities towards laurate and myristate relative to the wild-type enzyme. Binding affinities for palmitate remained relatively consistent with the wild-type K_M data in the case of the DM and TM1 variants, with a marginal reduction in palmitate affinity displayed by the TM2 variant relative to the wild-type BM3 enzyme. These data stand are in sharp contrast with the binding dissociation constants (K_d values) displayed in Table 3.3.

As with the wild-type enzyme, palmitate was the fatty acid substrate that produced the highest catalytic rates for each of the variants, as indicated by the k_{cat}/K_M values of 3.13, 3.24 and 0.995 μM^{-1} s^{-1} produced by the DM, TM1 and TM2 BM3 variants, respectively. In general, the catalytic efficiency of all the variants was decreased relative to the wild-type enzyme. Exceptions were seen for the DM and TM2 variants, which displayed k_{cat}/K_M values of 0.343 and 0.223 μM^{-1} s^{-1} with laurate, giving an increase in catalytic efficiency compared to the wild-type enzyme (0.205 μM^{-1} s^{-1}).

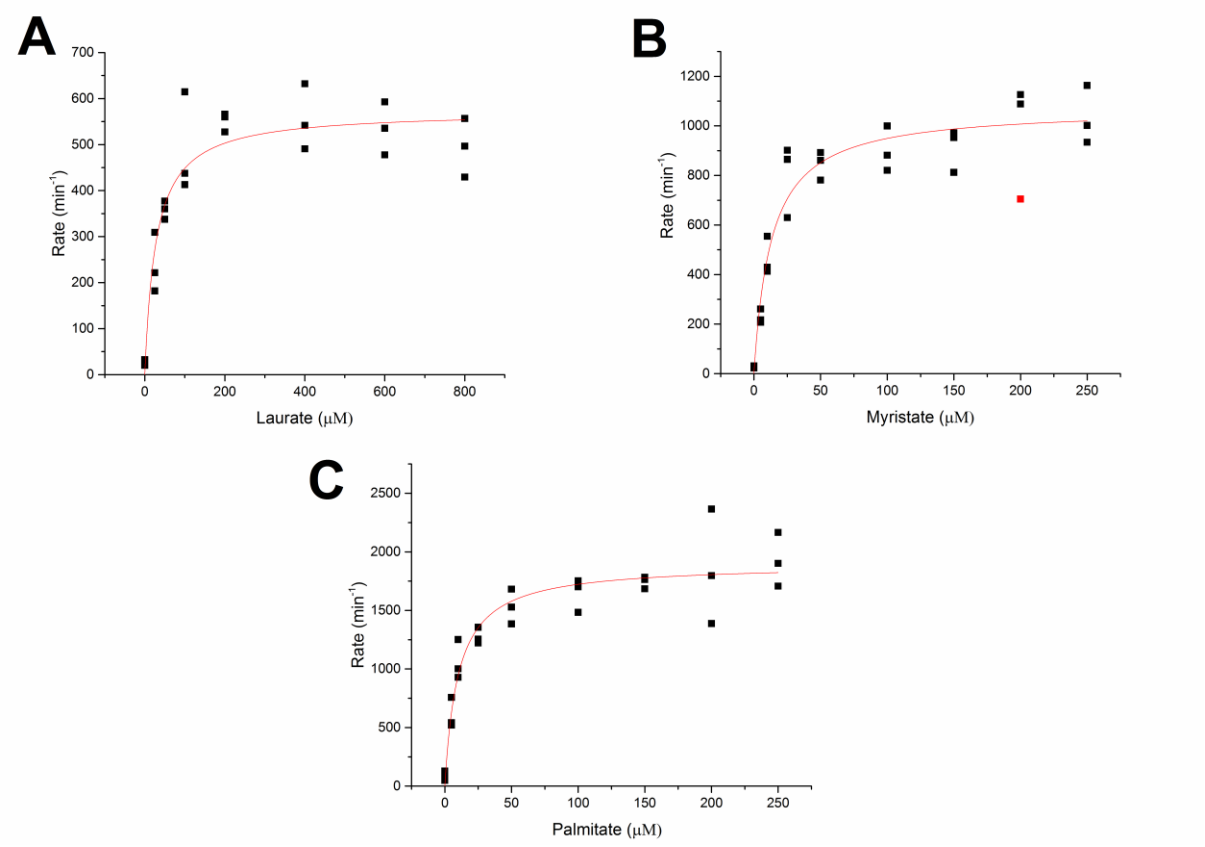

Figure 3.1 **Steady-state kinetic data for BM3 DM variant-dependent NADPH oxidation with lauric acid (A), myristic acid (B) and palmitic acid (C).**

The following kinetic parameters were obtained: A) K_M = 27.8 ± 6.29 µM, k_{cat} = 9.55 ± 0.390 s^{-1}, k_{cat}/K_M = 0.343 μM^{-1} s^{-1}; B) K_M = 13.0 ± 2.14 µM, k_{cat} = 17.9 ± 0.619 s^{-1}, k_{cat}/K_M = 1.38 μM^{-1} s^{-1}; C) K_M = 10.1 ± 1.88 µM, k_{cat} = 31.6 ± 1.13 s^{-1}, k_{cat}/K_M = 3.13 μM^{-1} s^{-1}. All data were fitted using the Michalis-Menten function. Outlier data points excluded from the final curve fit are highlighted in red.

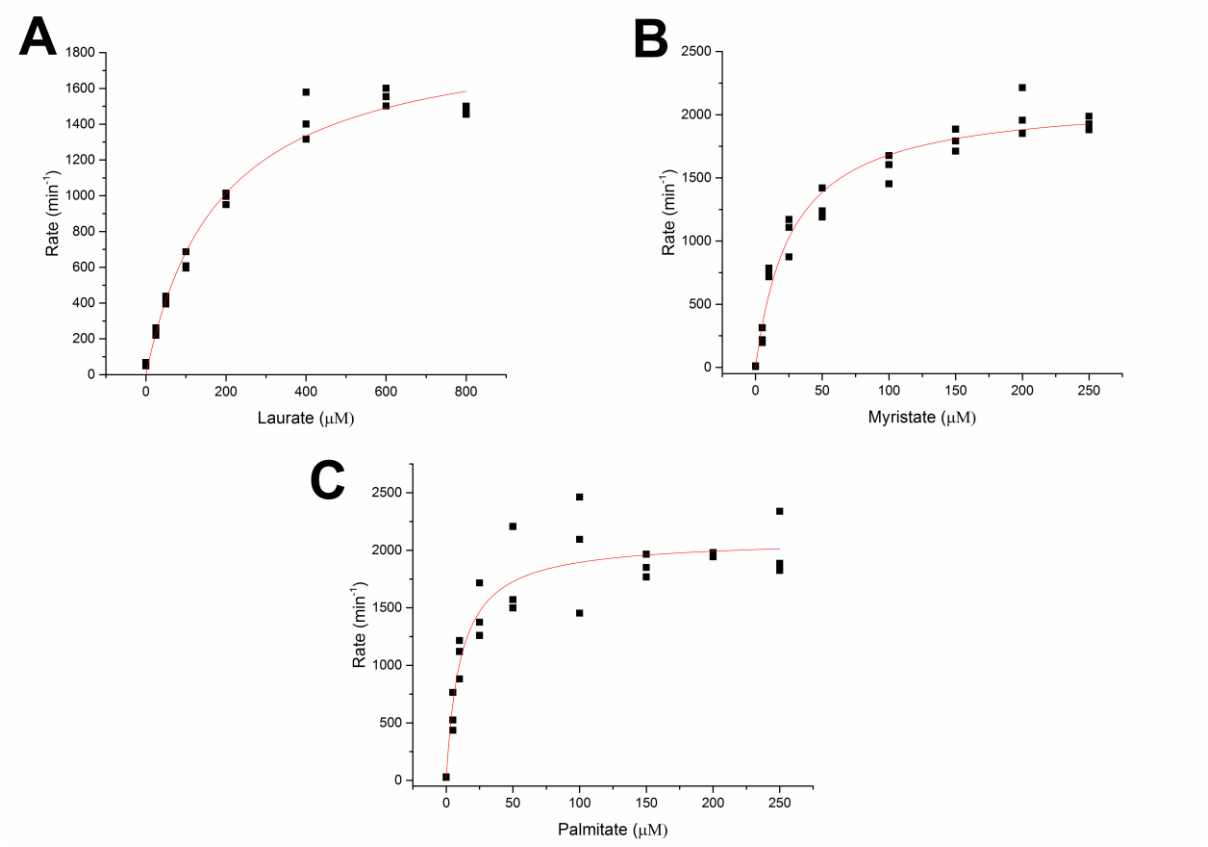

Figure 3.2 **Steady-state kinetic data for BM3 variant TM1-dependent NADPH oxidation with lauric acid (A), myristic acid (B) and palmitic acid (C).**

The following kinetic parameters were obtained: A) K_M = 185 ± 19.9 µM, k_{cat} = 32.5 ± 1.20 s^{-1}, k_{cat}/K_M = 0.178 μM^{-1} s^{-1}; B) K_M = 27.1 ± 3.44 µM, k_{cat} = 35.6 ± 1.14 s^{-1}, k_{cat}/K_M = 1.33 μM^{-1} s^{-1}; C) K_M = 10.8 ± 2.19 µM, k_{cat} = 35.0 ± 1.39 s^{-1}, k_{cat}/K_M = 3.24 μM^{-1} s^{-1}. All data were fitted using the Michaelis-Menten function.

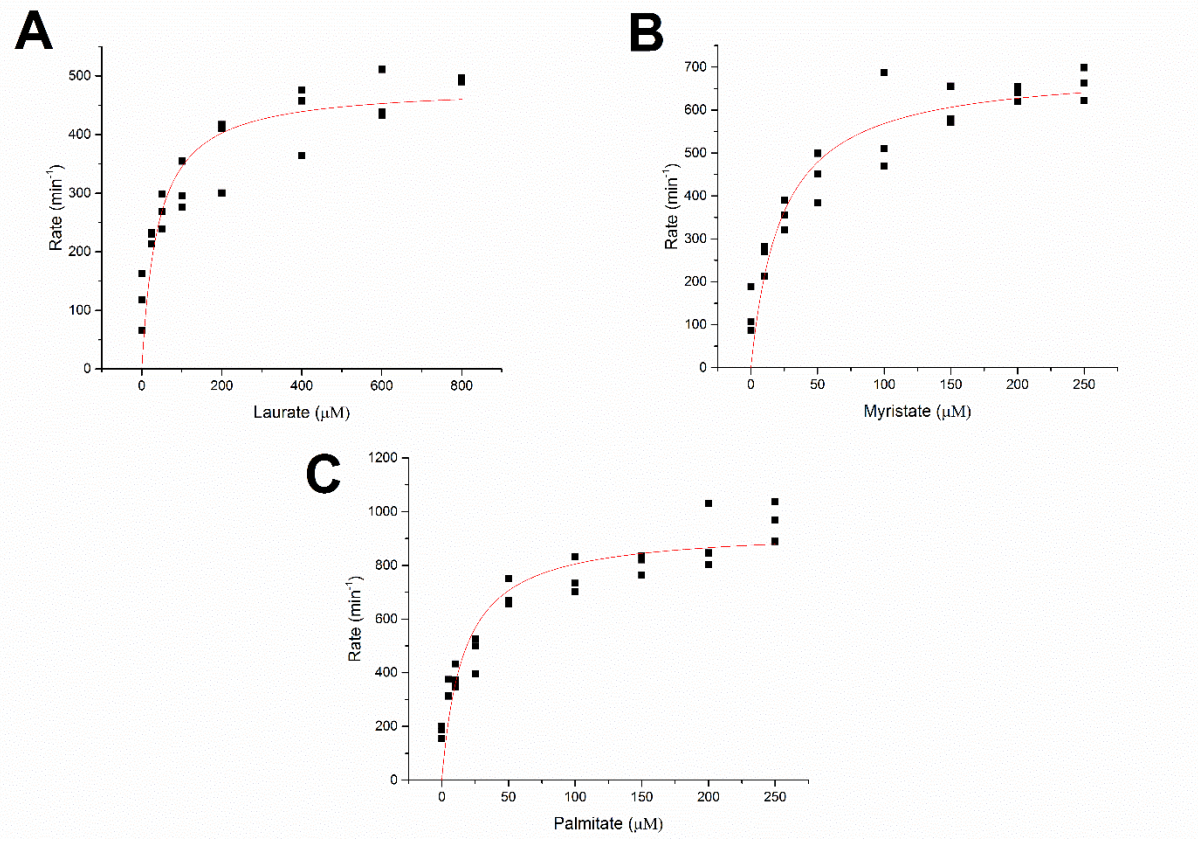

Figure 3.3 **Steady-state kinetic data for BM3 variant TM2-dependent NADPH oxidation with lauric acid (A), myristic acid (B) and palmitic acid (C).**

The following kinetic parameters were obtained: A) K_M = 40.2 ± 0.170 µM, k_{cat} = 9.05 ± 0.424 s^{-1}, k_{cat}/K_M = 0.223 μM^{-1} s^{-1}; B) K_M = 23.1 ± 5.10 µM, k_{cat} = 11.7 ± 0.591 s^{-1}, k_{cat}/K_M = 0.556 μM^{-1} s^{-1}; C) K_M = 16.4 ± 3.47 µM, k_{cat} = 15.6 ± 0.722 s^{-1}, k_{cat}/K_M = 0.995 μM^{-1} s^{-1}. All data were fitted using the Michaelis-Menten function.

3.4.3 Altered Spectral Properties of CO-Bound DM, TM1 and TM2 Variant BM3 Heme Domains

Spectral properties of the variant BM3 heme domains in their binding to carbon monoxide were characterised through analysis of the formation of ferrous-CO complexes, as generated by addition of carbon monoxide to samples of reduced wild-type and mutant BM3 heme domains. CO-binding spectra for each of the variant heme domains are presented in Figure 3.4.

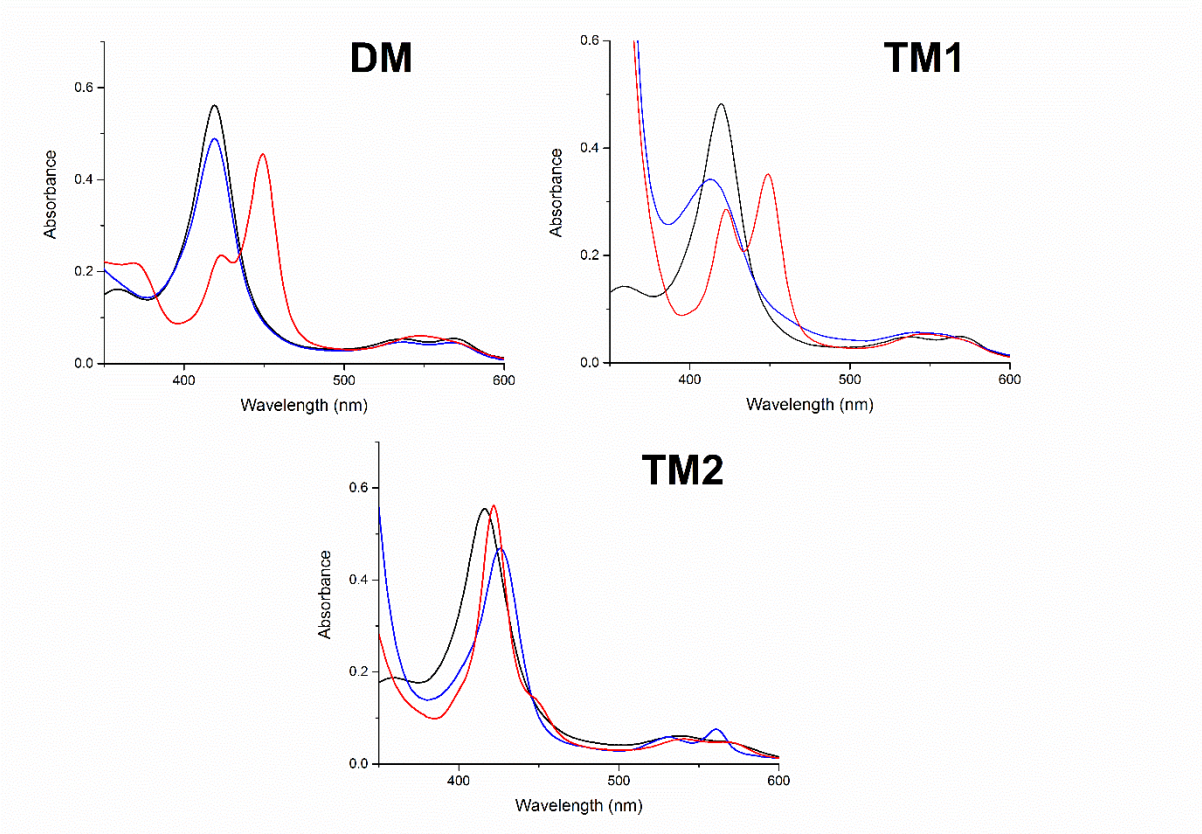

Figure 3.4 **UV-vis spectroscopic characterisation of DM (top-left inset), TM1 (top-right inset) and TM2 (bottom inset) variant heme domains, present in the substrate-free (black), partially dithionite-reduced (blue) and the reduced/ferrous CO-bound (red) states.**
All spectra were collected aerobically using 100 mM KPi, pH 7.5. Protein concentration was ~6 µM for all variants.

In the substrate-free state, both DM and TM1 have similar spectral properties to the wild-type heme domain, displaying Soret maxima at ~418 nm, accompanied by α and β bands at ~569 and ~537 nm, respectively. The substrate-free spectrum exhibited by the TM2 heme domain differs from both other variants, displaying a Soret maximum at ~416 nm. Furthermore, variant TM2 displays a ~569 nm α band partially obscured by a prominent β band at ~540 nm in the substrate-free state. The increased intensity and ~3 nm red shift of the TM2 β band suggests interaction between the A184K residue and the heme – however, this is not supported by a red shift of the Soret maximum to ~420 nm on binding CO.

Each variant was reduced to the ferrous state by the addition of a small amount of sodium dithionite. Formation of ferrous-CO complexes were achieved via the addition of CO gas to each sample, upon which P450 enzymes will characteristically produce Soret maxima at ~450 nm. In addition to a Soret maximum at 450 nm, the heme domain from the BM3

variant DM displayed a significant Soret band shoulder at 420 nm following saturation with carbon monoxide (Figure 3.4). Furthermore, the Soret maximum of the TM1 heme domain does not fully convert to 450 nm upon binding with carbon monoxide, and displays a more prominent 420 nm shoulder than the DM variant (Figure 3.4). The TM2 variant, which contains the A184K point mutation, appeared to fully convert to the P420 state upon formation of the ferrous-CO complex, as evidenced by the Soret maximum at 420 nm (Figure 3.4). These data suggest that all variants become partially protonated on the proximal cysteine (i.e. partially converted to a thiol form) upon heme reduction to the ferrous state. Protonation of the cysteinate ligand results in the formation of a thiol in place of a thiolate ligation. Production of the thiol-ligated (or P420) form of the enzyme hinders catalysis, as enzymes in the P420 state are unable to perform reductive cleavage of molecular oxygen during the P450 catalytic cycle (Dawson, 1988). Formation of P420 complexes is also a hallmark of low protein stability (Davydov et al., 2003).

3.4.4. Assessment of BM3 Variant Heme Domain Thermostablity

The stability of the substrate-free variant BM3 heme domains was assessed using an UNcle instrument (Unchained Labs, USA). Melting point data (T_m) for each of the DM, TM1 and TM2 variant heme domains are displayed in Table 3.5. Melting point data for the wild-type BM3 heme domain are displayed for the purposes of comparison (Whittall et al., unpublished data).

Table 3.5 Melting points (T_m values) for DM, TM1 and TM2 variant BM3 heme domains.

Samples were prepared to a concentration of 1 mg/ml. Thermostability measurements were made using a temperature ramp of 20 – 90 °C using an UNcle instrument. Protein samples were substrate-free and prepared to a concentration of ~1 mg/ml in 25 mM KPi buffer, pH ~7.5. [1]Wild-type data were taken from Whittall et al. (unpublished data).

BM3 heme variant	T_m1 (°C)	T_m2 (°C)
Wild-type[1]	58.4	67.1
DM (I263P/F87V)	53.0	-
TM1 (I263P/F87V/V78A)	49.4	-
TM2 (I263P/F87V/A184K)	46.4	-

Perhaps unsurprisingly, the thermostability of each BM3 variant heme domain was lower in comparison to that of the wild-type heme domain. Moreover, each BM3 variant displayed a single distinct unfolding event, compared to the two distinct unfolding events displayed by

wild-type heme domain. The thermostability of the DM variant heme domain decreased by 5.40 °C compared to the wild-type heme domain. The presence of a single I263P point mutation was shown to reduce thermostability of the wild-type heme domain by 2.40 °C (Whittall et al., unpublished data). Therefore, it can be suggested that the addition of the F87V mutation has itself caused a further drop in thermostability of ~3 °C. Addition of the V78A substitution to the DM heme domain (TM1) prompted a further reduction in thermostability of 3.60 °C, generating a variant (TM1) that was 9.00 °C less stable than the wild-type. However, the addition of the A184K mutation to the DM heme domain (TM2) led to a comparatively greater reduction of 6.60 °C, resulting in a variant heme domain 12.0 °C more unstable than the unmodified, wild-type heme domain. These data clearly indicate that successive point mutations have served to destabilise the heme domains of each BM3 variant relative to the wild-type heme domain, with the A184K substitution having the most destabilising impact on thermostability.

Melting point data were determined by measuring changes to the intrinsic fluorescence of tyrosine and tryptophan residues of each variant as the samples underwent temperature-induced conformational change (Figure 3.5) (Finan and Barco, 2017). Fluorescence intensity of each protein sample, expressed as the barycentric mean (BCM), altered significantly with increasing temperature (Figure 3.5, left-hand insets). T_m values were calculated from the peaks generated by the first derivative of each unfolding curve (Figure 3.5, right-hand insets).

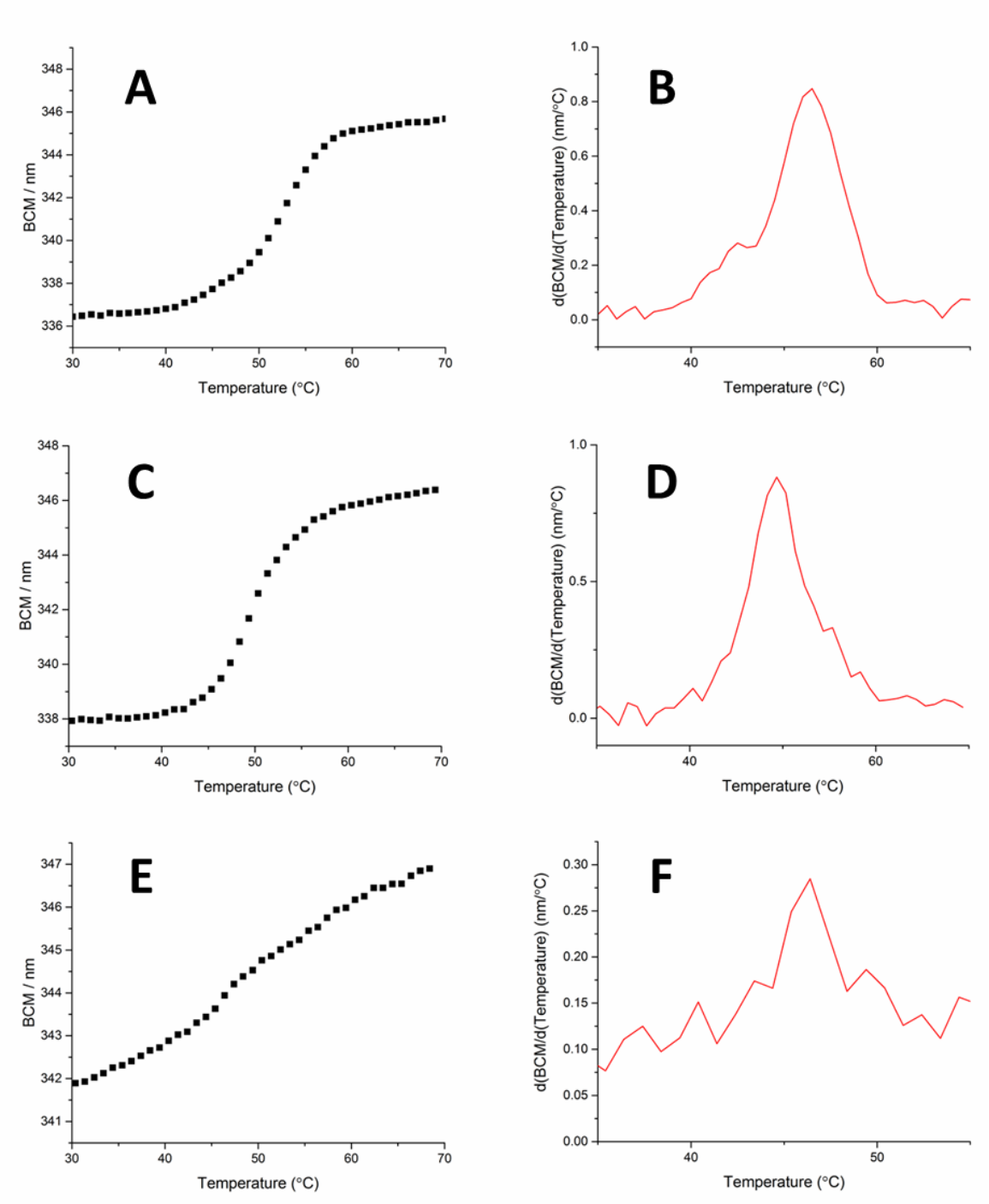

Figure 3.5 **Temperature-induced unfolding of DM (insets A and B), TM1 (insets C and D) and TM2 (insets E and F) variant heme domains.** Substrate-free heme domains were diluted to a concentration of 1 mg/ml in 25 mM KPi pH ~7.5. Temperature-induced changes to the intrinsic fluorescence of tyrosine and tryptophan residues of the variant heme domains are displayed in insets A (DM), C (TM1) and E (TM2). Melting points (T_m) were calculated from the peaks of the derivative curves (insets B, D, F) using an UNcle instrument. T_m values were determined to be 53.0 °C (DM), 49.4 °C (TM1) and 46.4 °C (TM2), respectively.

3.4.5. Resonance Raman Studies of BM3 Variant Heme Domains

Resonance Raman spectra were collected for each BM3 variant in the substrate-free and NPG-bound states. Spectra for DM, TM1 and TM2 are presented in Figures 3.6, 3.7 and 3.8, respectively. All prominent features were assigned according to nomenclature detailed in Spiro et al (1979) and are presented in Table 3.6.

Table 3.6 Band assignments for Resonance Raman spectra collected for variant BM3 heme domains

BM3 Variant	Resonance Raman band position (cm^{-1})				
	ν_4	ν_3	ν_{11}	ν_2	$\nu_{C=C}$
DM	1373	1500	1558	1582	1621
DM + NPG	1371	1500 1487	1560	1582	1621
TM1	1373	1500	1558	1580	1621
TM1 + NPG	1373	1503	1560	1582	1635
TM2	1371	1500	1560	1580	1624
TM2 + NPG	1369	1501 1489	1558	1578	1620

The oxidation state marker band, ν_4, is the most prominent within each spectrum (Figures 3.6, 3.7, 3.8, left-hand spectra). In all the spectra, the position of the ν_4 band is indicative of heme iron present in the ferric state. In all cases, the spectra show the ligand-free samples to be in the low-spin state. The positions of the ν_3 bands, the prime diagnostic of heme iron spin-state, at ~1500 cm^{-1}, is consistent with low-spin ferric heme iron. In contrast, high-spin ferric heme iron is indicated by the presence of the ν_3 band at ~1488 cm^{-1}. While the ν_3 band at ~1500 cm^{-1} remained prominent, shoulder peaks occurring at ~1488 cm^{-1} were observable for the NPG-bound DM and TM2 heme domains. No ν_3 species corresponding to high-spin heme were evident in the TM1 + NPG samples. In addition, the intensity of the low-spin ν_3 band was greater in both the substrate-free and NPG-bound samples of TM1 compared to all other spectra. The predominance of low-spin heme across all samples is corroborated by the presence of ν_2 marker bands at ~1580 cm^{-1} – again indicative of heme in the 6-coordinate, low-spin state. Despite notable perturbations in the TM2 + NPG sample, no ν_2 bands corresponding to high-spin, 5-coordinate protein (~1568 cm^{-1}) were evident in the protein + NPG spectrum. In the DM + NPG spectrum, the ν_2 marker (~1582 cm^{-1}) displayed a significantly increased intensity relative to the substrate-free spectrum. In wild-type BM3, the vinyl stretching vibration, $\nu_{C=C}$, typically appears as two modes, 1621 cm^{-1} and 1632 cm^{-1}, corresponding to porphyrin vinyl groups coplanar and out-of-plane with the heme iron, respectively (Chen et al., 2004). The predominant $\nu_{C=C}$ bands at ~1621 cm^{-1} in both substrate-free and NPG-containing samples of the DM variant suggest that the

vinyl groups of the porphyrin are predominantly in-plane with the heme. In contrast, the $\nu_{C=C}$ band at ~1635 cm^{-1} predominates in the TM1 + NPG spectra, suggesting that the majority of vinyl conformers are out-of-plane with the heme iron. In the substrate-free TM2 spectrum, the occurrence of a broad peak at ~1624 cm^{-1} suggests that the vinyl groups of this variant are broadly co-planar with that of the heme. In contrast, following the addition of NPG, the $\nu_{C=C}$ stretching mode of the TM2 variant shifts to ~1620 cm^{-1}, suggesting movement of the vinyl groups to a more co-planar conformation relative to the substrate-free spectrum. The position of the ν_{11} maker band, an additional indicator of the porphyrin ring planarity, remains consistent (1558-1560 cm^{-1}) across all samples. The intensity of the ν_{11} band increases in the DM and TM2 samples following the addition of NPG relative to the corresponding substrate-free samples. In contrast, the intensity of the ν_{11} band decreases in the TM2 spectrum following the addition of NPG.

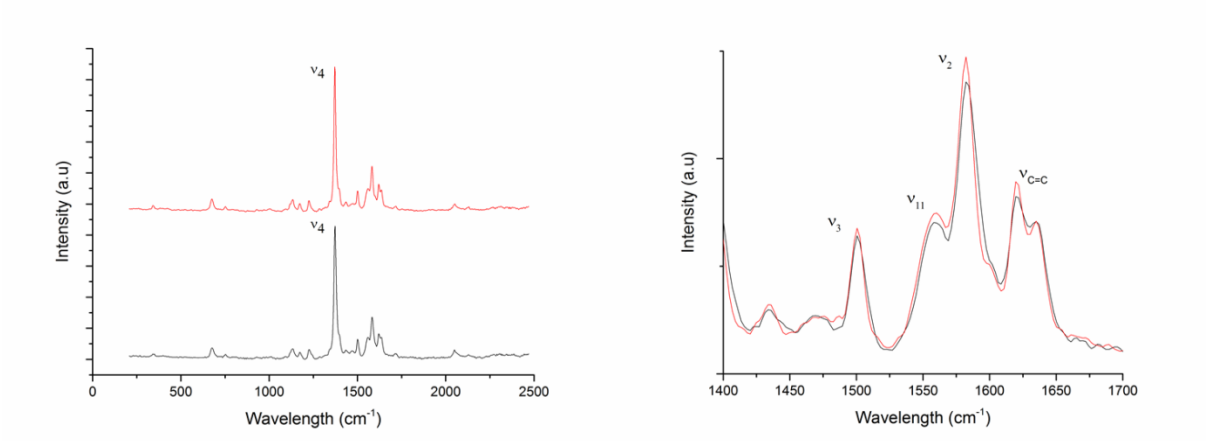

Figure 3.6 **Resonance Raman spectra of the DM (I263P/F87V) variant heme domain in the substrate-free (black trace) and NPG-bound states (red trace).**

Protein samples were prepared in 25 mM KPi, pH 7.5 to a concentration of 50 µM. The substrate NPG was present at a final concentration of 25 µM. Data were collected using a 405 nm laser. The oxidation marker band, ν_4, is highlighted in the left hand spectrum and is indicative of ferric heme iron. The right hand spectrum displays high-frequency bands in greater detail. Bands were assigned according to nomenclature presented in Spiro et al. (1979).

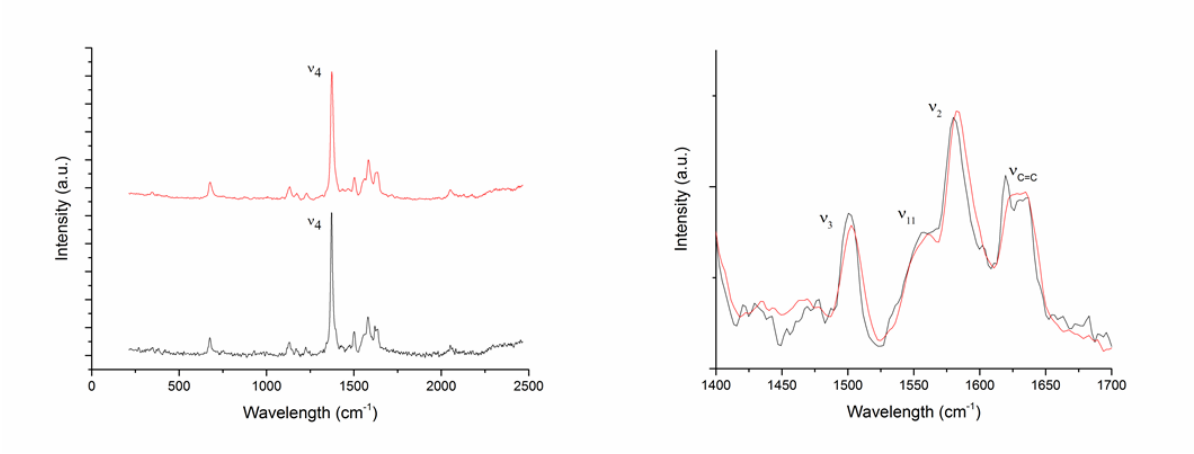

Figure 3.7 Resonance Raman spectra of the TM1 (I263P/F87V/V78A) variant heme domain in the substrate-free (black) and NPG-bound (red) states.

Data were collected using a laser with 405 nm excitation. The final concentration of the TM1 protein was 50 µM in both samples. The substrate NPG was present at a final concentration of 25 µM. The left hand spectra display the oxidation state marker band ν_4 at a position corresponding to ferric heme in both traces. The right hand spectra display bands occurring in the high-frequency region. Assignments were made according to nomenclature set out in Spiro et al. (1979).

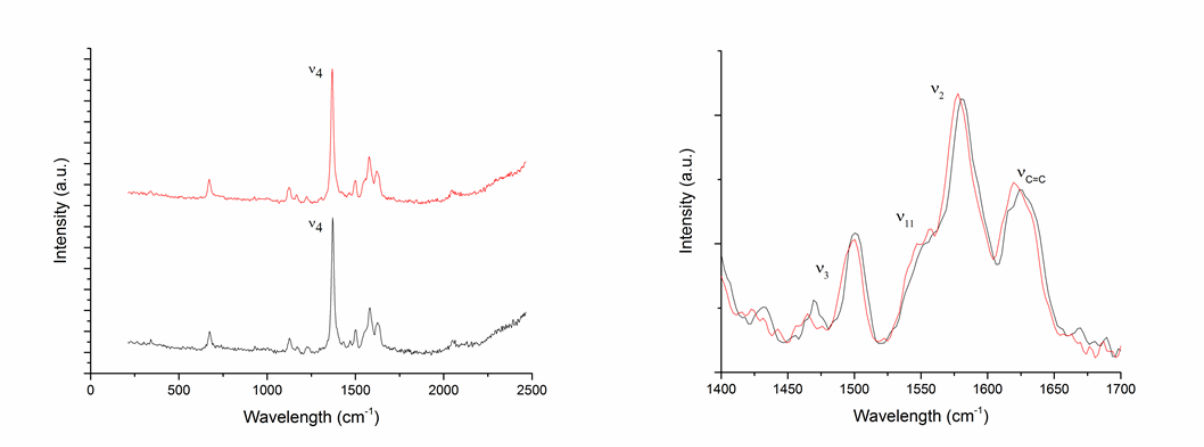

Figure 3.8 **Resonance Raman spectra of the TM2 BM3 variant (I263P/F87V/A184K) heme domain. Substrate-free: black trace; NPG-bound: red trace.**

Protein was present at a concentration of 50 µM in both samples. The substrate NPG (red trace) was present at a saturating concentration of 25 µM. Samples were prepared in 25 mM KPi buffer (pH 7.5). Samples were collected using a 405 nm laser. Left hand spectra: complete resonance Raman spectra for the TM2 heme domain present in both substrate-free and NPG-bound states. The oxidation marker band ν_4 indicates the presence of ferric heme iron. Right hand spectra: High-frequency region with structurally important bands highlighted. Band identities were assigned according to data outlined by Spiro et al. (1979).

3.4.6. Regioselective Hydroxylation of Fatty acid Substrates by Intact BM3 Variants

All examined intact BM3 variants displayed significant departures from the regioselective product profile of the wild-type P450 BM3 against laurate (C12:0), myristate (C14:0) and palmitate (C16:0) fatty acid substrates. In general, each successive residue substitution drove fatty acid hydroxylation towards the α carbon of each substrate and away from the ω -1, -2 and -3 positions favoured by the wild-type enzyme. In the case of laurate, both the DM and TM2 variants were able to access and hydroxylate the δ carbon (ω-7) position (Table 3.7).

Table 3.7 Product profile of laurate mono-hydroxylation performed by intact wild-type and BM3 variants DM, TM1 and TM2.

Turnover was performed at 37 °C for 60 min. Protein concentration used was 2 µM. All fatty acid substrates were present at a concentration of 200 µM. Samples were turned over in the presence of 500 µM NADPH and a cofactor regeneration system (outlined in the methods section). Triplicate samples were taken for each reaction, averaged and rounded to the nearest whole percentage point. [1]Wild-type turnover data are presented from Whittall et al. (unpublished data).

Intact BM3 Variant	Proportions of mono-hydroxylated laurate (C12:0) products formed by WT, DM, TM1 and TM2 BM3 enzymes								
	ω-9 (β)	*ω-8 (γ)*	*ω-7 (δ)*	*ω-6*	*ω-5*	*ω-4*	*ω-3*	*ω-2*	*ω-1*
WT[1]	0	0	0	0	0	0	37	29	34
DM (I263P/F87V)	0	0	2	5	9	7	39	21	17
TM1 (I263P/F87V/V78A)	0	0	0	4	8	7	34	25	22
TM2 (I263P/F87V/A184K)	2	2	2	4	15	6	25	23	21

The WT BM3 enzyme hydroxylates lauric acid at the ω-1, ω-2 and ω-3 positions only, consistent with previous data (Miura and Fulco, 1975). The DM BM3 also generates ω-4, ω-5, ω-6 and ω-7 hydroxylated species (to a total of 23% of the final products). The TM1 variant builds on the DM BM3 by the addition of the V78A mutation. The TM1 variant pushes regioselectivity back towards the ω carbon, but does produce ω-4, ω-5 and ω-6 hydroxylated lauric acid as 19% of the final products. Conversely, the TM2 variant, which incorporates an A184K substitution in addition to the DM substitutions, is capable of accessing positions on the fatty acid chain up to and including the ω-9 (or β-carbon) position of laurate. TM2 hydroxylates lauric acid at all positions from ω-1 to ω-9, with 31% of the hydroxylated fatty acid products being in the range from ω-4 to ω-9. However, despite the significant departures in product profile demonstrated by each variant towards laurate relative to the wild-type enzyme, the majority of products produced by each variant remain those also produced by the wild-type. Namely, products hydroxylated at the ω-1, ω-2 and ω-3 positions predominate within each variant, with the novel hydroxylated products comprising 23%, 19% and 31% of the total mono-hydroxylated products in variants DM, TM1 and TM2, respectively.

The profile of mono-hydroxylated products generated by each variant underwent considerable diversification in the case of the fatty acid substrate myristate (Table 3.8).

Table 3.8 Product profile of myristate mono-hydroxylation performed by intact wild-type and BM3 variants DM, TM1 and TM2.

All samples were repeated in triplicate, averaged and rounded to the nearest percentage point. Substrate and protein concentrations were 200 and 2 µM, respectively. Reactions were turned over at 37 °C for 60 min in the presence of 500 µM NADPH and a cofactor regeneration system, as described in the methods. [1]Wild-type data taken from Whittall et al (unpublished data).

Intact BM3 Variant	Proportions of mono-hydroxylated myristate (C14:0) products formed by WT, DM, TM1 and TM2 BM3 enzymes						
	ω-7	*ω-6*	*ω-5*	*ω-4*	*ω-3*	*ω-2*	*ω-1*
WT[1]	0	0	0	1	27	30	42
DM (I263P/F87V)	18	18	39	5	4	6	10
TM1 (I263P/F87V/V78A)	36	17	25	4	4	6	8
TM2 (I263P/F87V/A184K)	23	18	34	7	5	6	7

The majority of mono-hydroxylated products produced from myristate by each variant were outside the product range generated by the wild-type BM3; i.e. hydroxylations occurring at positions other than the ω-4, ω-3, ω-2 and ω-1 carbons. This was most pronounced in the TM1 variant, in which 78% of mono-hydroxylated fatty acids generated by the variant were outside of the wild-type range. Novel mono-hydroxylated products accounted for 75% of the total product generated by both the DM and TM2 variants. Despite this substantial increase in the formation of hydroxylated myristic acid products at positions from ω-4 to ω-7, hydroxylation occurring in the direction of the α carbon was not evident past the ω-7 position.

As with myristate, the majority of mono-hydroxylations performed by each variant occurred at positions inaccessible to the wild-type enzyme for the final fatty acid substrate examined, palmitate (Table 3.9).

Table 3.9 Product profile of palmitate mono-hydroxylation performed by intact wild-type and DM, TM1 and TM2 variant BM3 enzymes. Samples were turned over at 37 °C for 60 min. Substrate concentration was 200 µM. Protein concentration was 2 µM. NADPH concentration was at a near-saturating 500 µM. A cofactor regeneration system was present, as described in the methods. All samples were performed in triplicate, averaged and rounded to the nearest whole number. [1]Wild-type data adapted from Whittall et al (unpublished data).

Intact BM3 Variant	Proportions of mono-hydroxylated palmitate (C16:0) products formed by WT, DM, TM1 and TM2 BM3 enzymes								
	ω-9	*ω-8*	*ω-7*	*ω-6*	*ω-5*	*ω-4*	*ω-3*	*ω-2*	*ω-1*
WT[1]	0	0	0	0	0	0	13	56	31
DM (I263P/F87V)	0	10	48	12	3	2	4	14	7
TM1 (I263P/F87V/V78A)	0	6	63	11	4	3	3	7	4
TM2 (I263P/F87V/A184K)	5	5	43	15	15	2	4	7	4

In contrast to the diverse product profiles displayed by each variant against the substrates laurate and myristate, in the case of palmitate, the predominant mono-hydroxylation occurred at position ω-7 across all variant profiles. In the case of the TM1 variant, ω-7 hydroxylated palmitate accounted for 63% of the mono-hydroxylated palmitate product profile – the highest proportion generated by any variant against all tested fatty acids. All variants were capable of reaching the ω-8 positions, with the TM2 variant reaching the furthest away from the palmitate ω-carbon, generating a product hydroxylated at the ω-9 position.

Despite the significant alterations to regioselectivity, each variant displayed a decreased total percentage turnover of substrate relative to the wild-type enzyme (Table 3.10). The intact DM variant retained the greatest semblance of wild-type turnover activity, with total turnover dropping by 16, 1 and 3 percentages points relative to the wild-type BM3 against laurate, myristate and palmitate, respectively. In contrast, the TM1 variant underwent dramatic drops in turnover against all tested fatty acids, metabolising only 9% of laurate, a near 10-fold decrease relative to the wild-type enzyme. Percentage turnover displayed by TM2 was also shown to decrease relative to both the wild-type and the DM variant, albeit in a manner less severe than that of TM1. Against laurate, the TM2 variant displayed a near 5-fold decrease in percentage turnover compared to the wild-type. As with the wild-type

enzyme, all variants achieved their highest and lowest percentage turnover against myristate and laurate, respectively.

Table 3.10 Total percentage turnover of fatty acid substrates by intact wild-type and BM3 variant enzymes DM, TM1 and TM2.
Samples were turned over for 60 min at 37 °C. Protein and substrate concentrations were 2 µM and 200 µM, respectively. The NADPH cofactor was present at 500 µM, in addition to a regeneration system as described in the methods section. Data presented are the average of three samples rounded to the nearest percentage point. [1]Presented from Whittall et al. (unpublished data).

Intact BM3 variant	% Turnover of fatty acid substrate		
	Laurate (C_{12})	Myristate (C_{14})	Palmitate (C_{16})
WT[1]	88	99	91
DM (I263P/F87V)	72	98	88
TM1 (I263P/F87V/V78A)	9	63	35
TM2 (I263P/F87V/A184K)	18	80	71

3.4.7. Enantioselective Functionalisation of Styrene to Styrene Oxide in Wild-Type and Variant BM3 Enzymes

Production levels and enantio-preference of styrene oxide formation underwent significant alterations in each intact BM3 variant relative to the wild-type enzyme (Table 3.11).

Table 3.11 Enantioselective production of styrene oxide performed by intact wild-type and DM, TM1 and TM2 variant enzymes.

[1]All wild-type data presented from Whittall, et al. (unpublished data). [2]Calculated by spectroscopic binding titrations. Protein concentration was ~5 µM. [3]Estimates for total concentration of styrene oxide generated by each variant were calculated using appropriate calibration curves. Turnover reactions contained 2 µM and 2 mM of intact protein and styrene substrate, respectively. All reactions were turned over for 20 hours. [4]Calculated via peak ratios of GC chromatograms. %enantiomer = {{peak area of enantiomer / total peak area of enantiomers} x 100}. [5] R(+) *%e.e* = {%R(+) styrene oxide - %S(-) styrene oxide}.

Intact BM3 variant	Styrene Binding constant[2] (K_d, µM)	Styrene oxide formed[3] (µM)	*R*(+)-Styrene oxide[4] (%)	*S*(-)-Styrene oxide[4] (%)	Styrene oxide[5] (% *e.e.)*
WT[1]	148 ± 13.8	16.9 ± 6.83	58.7 ± 2.02	41.3 ± 2.02	R(+): 17.4 ± 4.04
DM (I263P/F87V)	100 ± 7.47	166 ± 17.0	25.3 ± 0.100	74.7 ± 0.100	S(-): 49.4 ± 0.200
TM1 (I263P/F87V/V78A)	73.3 ± 4.60	34.1 ± 2.59	21.7 ± 0.874	78.3 ± 0.874	S(-): 56.7 ± 1.75
TM2 (I263P/F87V/A184K)	156 ± 14.8	161 ± 14.2	21.7 ± 0.153	78.3 ± 153	S(-): 56.7 ± 0.306

In contrast to the wild-type enzyme, each variant favoured the production of the *S*(-) enantiomer of styrene oxide. Inclusion of substitutions I263P and F87V in the DM variant appeared to have the greatest impact on the inversion of enantiomer production from *R*(+) to *S*(-). In variant DM, the production of *S*(-)-styrene oxide increased by ~33 percentage points relative to the wild-type enzyme, with a corresponding decrease in production of the *R*(+) enantiomer by the same proportion (Figure 3.9).

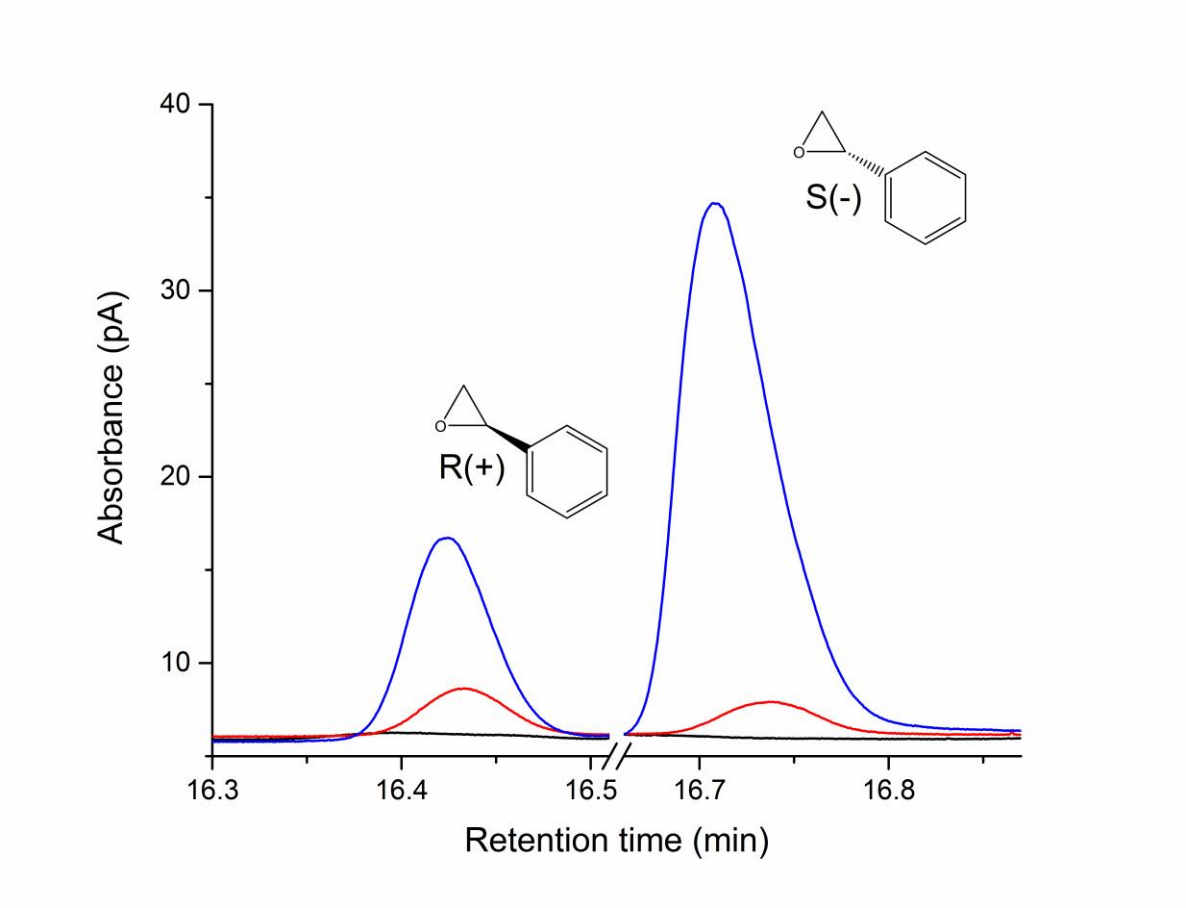

Figure 3.9 **Production of R(+) and S(-) enantiomers of styrene oxide from styrene substrate by wild-type (red chromatogram) and variant DM (blue chromatogram) BM3 intact enzymes.**
The black chromatogram corresponds to a 2 mM styrene-only control sample. Chromatograms were normalised to an internal standard (not shown).

Consequently, the DM variant produced a 49.4 ± 0.200 % enantiomeric excess of the *S*(-) enantiomer of styrene oxide. The DM variant also demonstrated a near 10-fold increase in styrene oxide production (166 ± 17.0 µM) in comparison to the wild-type enzyme (16.9 ± 6.83 µM). In addition, the DM variant heme domain displays an increased binding affinity towards the styrene substrate (100 ± 7.47 µM) relative to the wild type heme domain (148 ± 13.8 µM).

The TM1 variant displayed an even greater binding affinity towards the styrene substrate (73.3 ± 4.60 µM). However, the total amount of styrene oxide produced fell to 34.1 ± 2.59 µM, a near 5-fold decrease on the amount produced by the DM variant. Enantiopreference of styrene oxide production shifted towards a greater excess of the *S*(-) enantiomer, with variant TM1 producing a 56.7 ± 1.75 *e.e* of *S*(-)-styrene oxide (Figure 3.10).

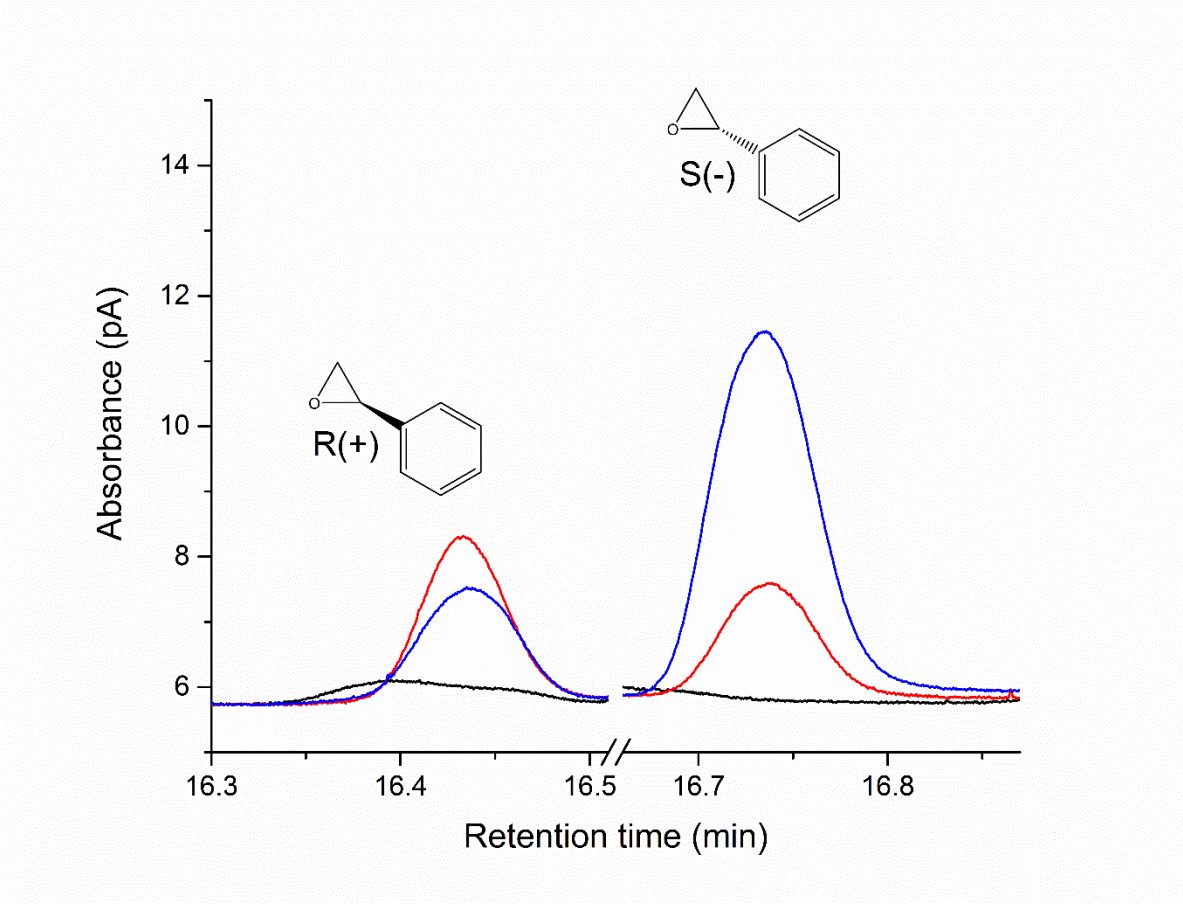

Figure 3.10 **GC chromatograms of R(+) and S(-) enantiomers of styrene oxide produced by intact wild-type (red line) and TM1 variant (blue line) BM3 enzymes.**
Control samples containing 2 mM styrene are indicated by the black line. All chromatograms were normalised to the internal standard (not shown).

In the case of the TM2 variant, inclusion of the A184K substitution prompted the variant to display a decreased affinity towards the styrene relative to both of the DM and TM1 variants, and comparable to that of the wild-type heme domain (TM2: 156 ± 14.8 µM, WT: 148 ± 13.8 µM). Despite this, the TM2 variant produced 161 ± 14.2 µM styrene oxide product, an amount similar to that produced by the DM variant (166 ± 17.0 µM). Enantiopreference of styrene oxide production was relatively unaltered with reference to the TM1 variant, with the TM2 mutant producing a 56.7 ± 0.306 % enantiomeric excess of *S*(-)-styrene oxide, a proportion within the margin of error of that of the TM1 variant (Figure 3.11).

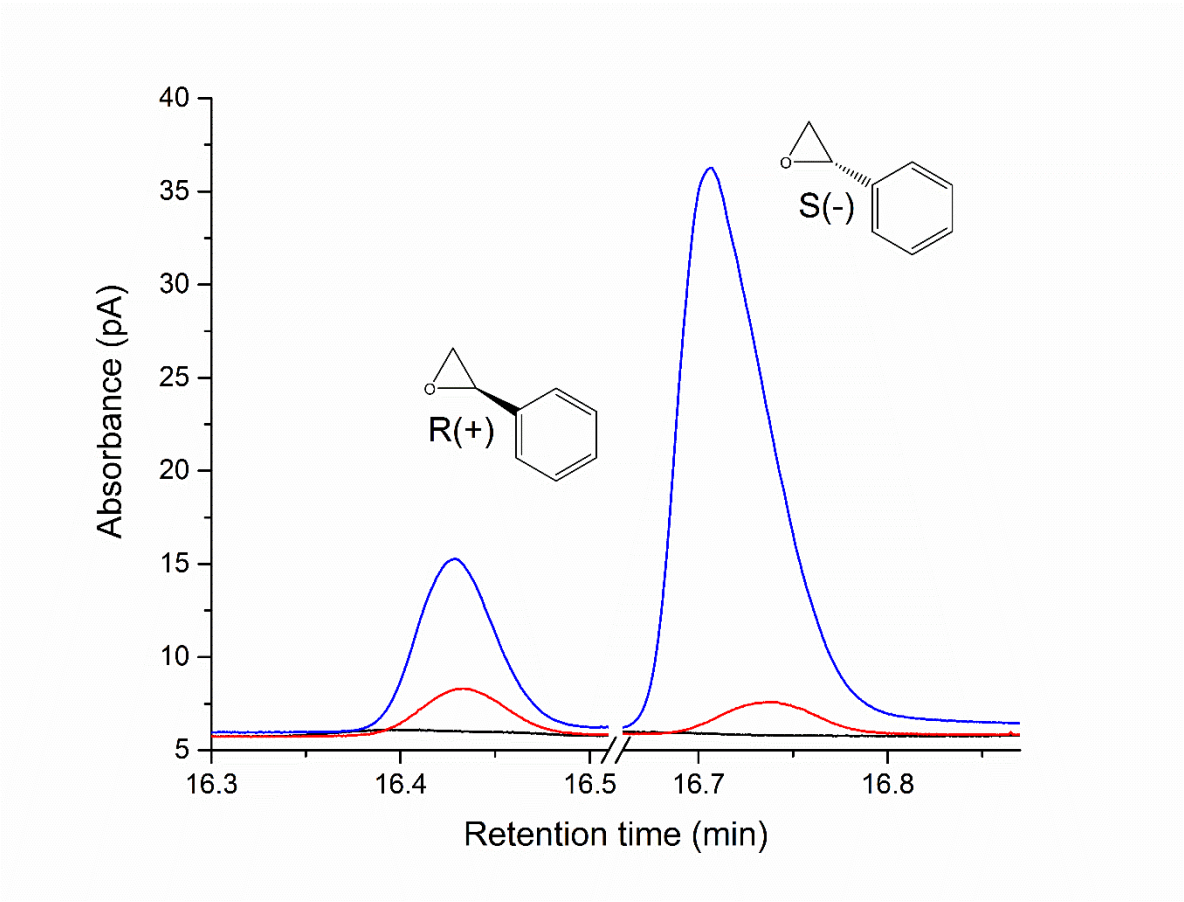

Figure 3.11 GC chromatogram of R(+) and S(-) styrene oxide production in intact wild-type (red line) and TM2 variant (blue line) BM3 enzymes.
Control samples containing 2 mM styrene are represented by the black lines. Chromatograms were normalised to the internal standard.

3.4.8. Docking Studies with Substrate-Free Variant Heme Domain Crystal Structures

In order to rationalise alterations to regio and enantioselectivity displayed by the variants, docking studies were performed using crystal structures collected for the TM1 and TM2 heme domains. In the TM1 and TM2 structures, all contaminants, including molecules of PEG, were removed before docking. For the purpose of comparison, docking studies were performed on a wild-type heme domain structure (PDB 1JPZ) from which the ligand (NPG) had been removed prior to docking. In addition, following energy minimisation of the target structures and ligand, all water molecules were removed prior to running docking simulations.

Docking simulations were performed using YASARA software, as detailed in the methods section. In each simulation, 25 docking poses were generated and scored according to the

YASARA scoring function (Krieger and Vriend, 2015). Poses were not selected for further analysis based on score alone. Many high-scoring poses displayed substrate conformations that were deemed implausible or were unsubstantiated by other data, i.e. the carboxylate of the fatty acid was facing the heme. Poses selected for further analysis were therefore decided upon using a combination of YASARA score, product profile data and previous structural information presented in the literature.

A model of the binding of palmitic acid to the wild-type heme domain is presented in Figure 3.12.

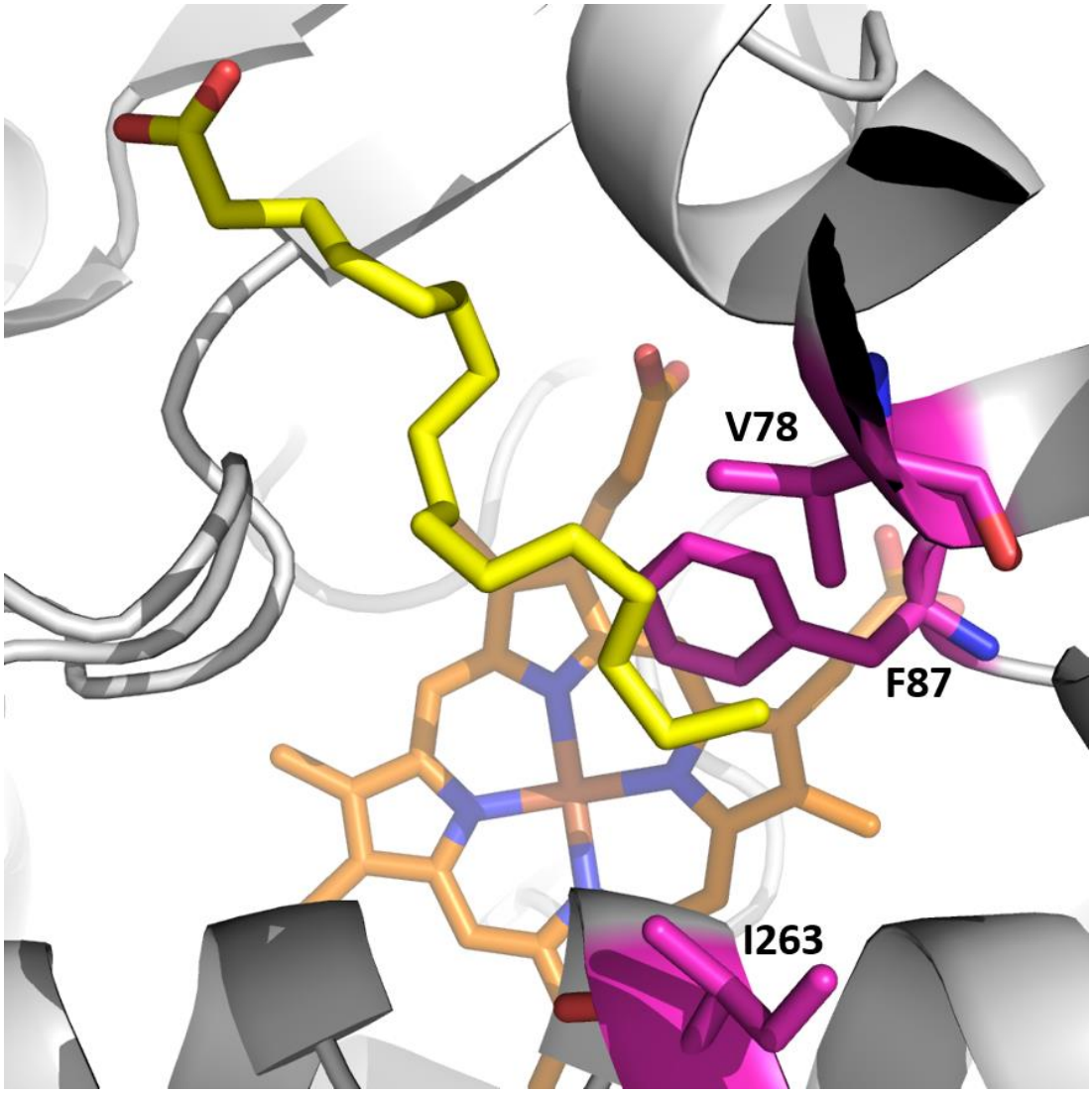

Figure 3.12 Model of palmitate binding in the heme domain of wild-type BM3. The structure of the wild-type heme domain (grey) was adapted from PDB file 1JPZ (Haines et al., 2001). The position of substrate palmitic acid (yellow) is shown in relation to the heme domain (orange). Residues of interest are highlighted in magenta. Modelling was performed using YASARA software (version 18.2.7) (Krieger and Vriend, 2015).

No structure of wild-type BM3 heme domain bound to palmitic acid has yet been reported. However, the manner of binding displayed by the model (Figure 3.12) is comparable to that displayed by NPG – a fatty acid analogue, 18 carbons in length – in the PDB structure 1JPZ (Haines et al., 2001). The position of the ω terminus relative to the heme iron suggests hydroxylation occurring at the ω-2 and ω-1 positions – consistent with the product profile reported for the intact wild-type BM3 detailed earlier in this section (ω-2: 56%; ω-1: 31%). These data affirm that docking studies provide suitable models for examining the binding of fatty acid substrates within the BM3 heme domain, and provide a reasonable basis for comment regarding product profile of substrate hydroxylation.

A model of palmitic acid binding within the heme domain of the TM1 variant is presented in Figure 3.13.

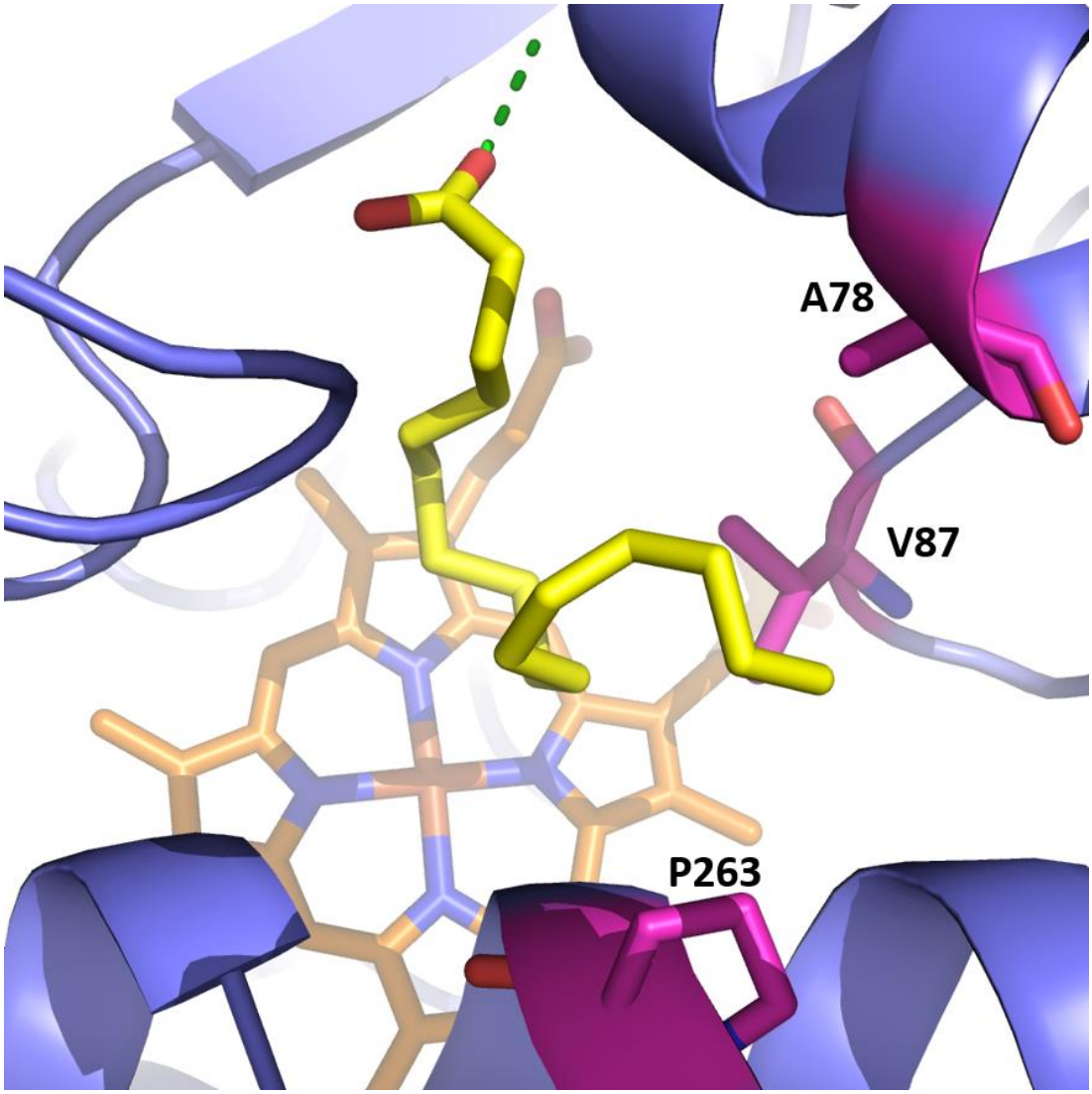

Figure 3.13 Postulated binding mode of palmitic acid within the heme domain of the BM3 TM1 variant (I263P/F87V/V78A).
The pose was obtained by docking studies performed using YASARA software (version 18.2.7) (Krieger and Vriend, 2015). The secondary structure of the TM1 variant is presented in blue. The palmitic acid substrate is highlighted in yellow and framed relative to the heme domain, highlighted in orange. Residues of interest are highlighted in magenta. A hydrogen bonding interaction suggested by the model, occurring between residue S72 and the carboxylate group of palmitic acid, is highlighted in green.

Relative to the wild-type model (Figure 3.12), the palmitic acid substrate appears to approach closer to the heme prosthetic group. Moreover, the model suggests the formation of a hydrogen bonding interaction between the unmodified S72 residue and the carboxylate group of palmitic acid. This interaction is not observed in the wild-type model. The ω-7 carbon atom of palmitic acid is positioned directly above the heme iron, with the

model also suggesting that the ω terminus of the substrate is sequestered into the pocket created by the V78A and I263P substitutions. Carbon atoms from the ω-9 position onwards, moving in the direction towards the carboxylate group, appear inaccessible to the heme iron.

A model suggesting the position of palmitic acid within the heme domain of the TM2 variant is displayed in Figure 3.14.

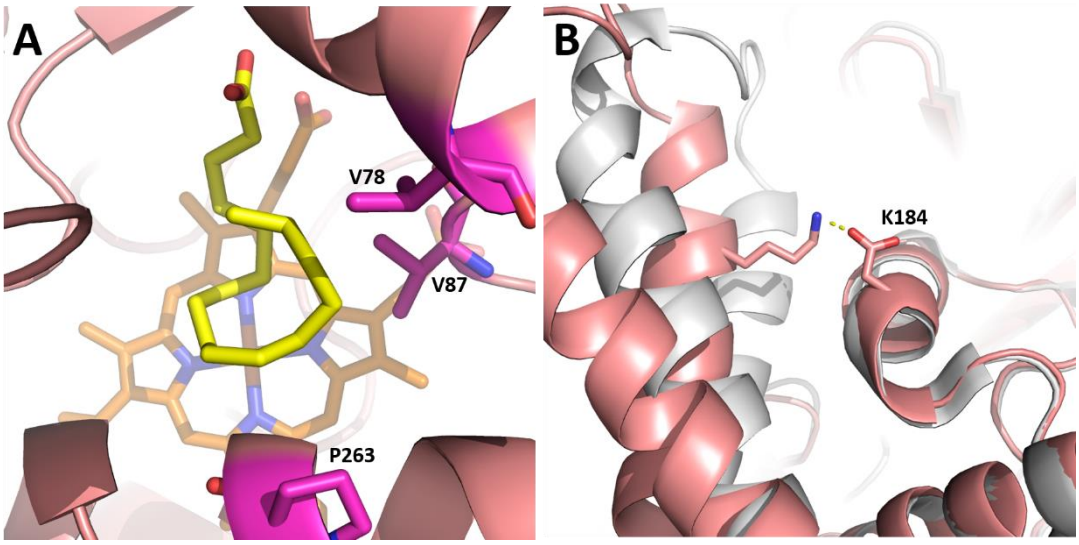

Figure 3.14 Hypothesised binding conformation of the fatty acid substrate palmitic acid in the active site of the TM2 BM3 variant (I263P/F87V/A184K).
Panel A) The mode of palmitate (yellow) bound in the TM2 variant heme domain (pink) suggested by docking studies performed using YASARA software (version 18.2.7) (Krieger and Vriend, 2015). The substrate pose is presented relative to the heme cofactor (orange). Residues of interest are highlighted in magenta. Panel B) F- and G- helical translocation via salt bridge formation between residues K184 and D80 within variant TM2 (pink) relative to wild-type BM3 (grey).

Akin to the binding pose displayed in the TM1 variant model (Figure 3.13), the model generated using the structure of the TM2 variant suggests palmitic acid is afforded greater access to the heme iron when compared to the wild-type model. However, unlike the binding mode suggested by the model for variant TM1, no hydrogen bonding interactions between any active site residues are evident in the model corresponding to the TM2 variant. As with TM1, the TM2 model suggests that the carbon atom at the ω-7 position of the substrate is situated directly above the heme iron.

In addition to modelling the binding of palmitic acid, docking simulations were conducted using styrene for the wild-type, TM1 and TM2 variants with YASARA. However, the data provided by these models did not provide sufficient structural rationale to explain the alterations to enantio-preference of styrene epoxidation in the TM1 and TM2 variants. Multiple binding conformations, corresponding to positions favouring the production of both *R(*+) and *S*(-) enantiomers of styrene oxide, were evident in all simulations. Models for the binding of styrene to wild-type, TM1 and TM2 variant heme domains are presented in Figures 3.S1, 3.S2 and 3.S3 in the supplementary information.

3.5. Discussion

This study was concerned with the rational development of a highly active, single residue mutant (variant I263P) into a versatile biocatalyst capable of performing regio and stereospecific oxidation reactions outside the range of the wild-type enzyme. Alterations to both fatty acid regioselectivity and styrene enantioselectivity were successfully achieved across all variants. However, these changes were achieved at the expense of significant reductions in both stability and catalytic activity of the variants relative to the wild-type enzyme.

Perhaps unsurprisingly, each successive mutation led to further reductions in heme domain thermostability. However, decreases in enzyme stability can often suggest a concurrent increase in structural flexibility, prompting beneficial alterations to substrate selectivity (Butler et al., 2013; Rentmeister et al., 2011). This is evident in variants DM and TM1, which display increased binding affinities for all tested fatty acids, in addition to styrene, when compared to the wild-type BM3. However, the variant most destabilised relative to the wild-type, TM2, did not conform to this pattern. The TM2 variant displayed comparable binding affinities for styrene (156 ± 14.8 μM) to the wild-type (148 ± 13.8 μM). In the case of the fatty acid substrates, TM2 displayed reduced affinities for lauric and palmitic acids, and displayed an affinity towards myristate comparable to that of the wild-type. As the preceding variant, the DM mutant, displayed dramatically increased affinity for all substrates tested relative to the wild-type, the comparative decrease in binding affinity for the same substrates displayed by variant TM2 can be directly attributed to the A184K substitution.

Despite the decreases in stability indicated by the thermal unfolding data, the resonance Raman spectra obtained for all variants were characteristic of native, cysteine-ligated P450 enzymes (McLean et al., 2002). The P420 species observed following CO binding, most

evident in the TM2 variant, were not observed in the substrate-free or bound resonance Raman spectra.

Located within the F-helix (P172 – Q189), the substitution of a lysine residue at position 184 resulted in significant structural rearrangements in the TM2 variant compared to the wild-type enzyme. When overlaid with a substrate-bound structure of the wild-type (PBD 1JPZ; Haines et al., 2001), the architecture of the TM2 active site retains (with the exceptions of the I263P and F87V substitutions) a high degree of similarity. However, due to the formation of a salt-bridge interaction between the substituted K184 residue and D80 of the B'-helix (S72-G83) both the F- and G- (G200 – S226) helices of the TM2 variant pivot away from the active site (Figure 3.19, Panel B). The translocation of these helices serves to widen both the substrate access channel and raise the canopy of the active site. These structural findings are consistent with those reported for other A184K containing variants by Shehzad et al., (2013). The presence of the salt bridge between residues K184 and D80, and the concurrent repositioning of the F- and G-helices, likely underpins the reduction in substrate binding affinities displayed by the TM2 variant. This additional polar interaction may serve to reduce the flexibility of the F- and G-helices, hindering the access of smaller substrates, namely lauric acid and styrene, to the active site.

While showing tighter binding affinities against the fatty acid substrates relative to the wild-type enzyme, the V78A substitution in the TM1 variant led to decreases in fatty acid binding affinities compared to the DM variant. This decrease in binding affinity was not seen in the TM1 variant for styrene. Furthermore, the TM1 variant displays the lowest percentage turnover of any of the mutants against all fatty acid substrates. When compared to the more modest turnover reductions displayed by the DM variant, the loss of activity in the TM1 variant can be attributed to the inclusion of the V78A substitution. These data suggest that residue V78 has a role in fatty acid orientation within the active site of BM3, as outlined by Zheng et al. (2016).

Apparent K_M and k_{cat} values were determined for each BM3 variant against lauric, myristic and palmitic fatty acids. The overall catalytic activity (k_{cat}) decreased across all variants and against all fatty acids, with the TM2 variant displaying the greatest loss of activity relative to the wild-type enzyme. As with the wild-type BM3 enzyme, catalytic activity increased with substrate chain length in the case of the DM and TM2 variants. The TM1 variant displayed remarkably consistent catalytic activity across the chain lengths, albeit at rates that paled in comparison to the wild-type enzyme. This may suggest that, in the TM1

variant, fatty acids occupy similar conformations within the active site upon binding, facilitating a consistent rate of electron transfer to the heme iron. However, K_M values displayed by the intact TM1 enzyme vary depending on the chain length of the fatty acid - suggesting a degree of substrate-induced conformational change within the TM1 active site (Koshland, 2002). Consequently, the similar k_{cat} values obtained for each fatty acid substrate could simply represent the near maximum rate of catalysis achievable by the TM1 variant. The tendency of the TM1 and TM2 variants to form inactive P420 complexes (through protonation of the thiolate proximal ligand), as indicated by the carbon-monoxide binding spectra may further serve to explain the compromised catalytic activity (Manna and Mazumdar, 2008).

The I263P and F87V substitutions present throughout all the BM3 variants, in addition to the V78A substitution in the TM1 variant, have increased the active site space accessible to the fatty acid substrates. However, this increased space has come at the cost of maintaining a sufficiently hydrophobic active site environment required for effective catalysis (Doyle et al., 2000). With the exception of the modest increases displayed by the DM and TM2 variants against laurate relative to the wild-type, all variants displayed decreased catalytic efficiencies (k_{cat}/K_M) with the fatty acids examined. Generally, K_M values recorded for each intact variant with the fatty acid substrates followed the trend established by the K_d values for each variant heme domain, i.e. binding affinity increased in tandem with chain length. However, the K_d value determined for the TM2 heme domain variant against lauric acid (162 ± 9.51 μM) is near 4-fold greater than the K_M value collected for the same variant using the intact protein following kinetic analysis in the steady-state (40.2 ± 0.170 μM). These data infer significant differences between the activity of the TM2 heme domain in isolation and the activity of the TM2 heme domain as part of an intact BM3 protein.

In light of this, it is important to state that NADPH consumption by intact wild-type BM3 is not directly linked to production of hydroxylated product (Loida and Sligar, 1993). This is particularly evident for laurate – which has been reported to display low levels of coupling (34%) when the reaction is catalysed by the wild-type enzyme (Cryle and De Voss, 2007). Moreover, the TM2 variant displayed low levels of laurate turnover (18%) – in contrast to displaying comparatively high levels of turnover against myristate (80%) and palmitate (71%). These data suggest that, with laurate, the TM2 variant may exhibit high levels of uncoupling. Uncoupling of electron transfer from product formation, which can arise when

substrates only loosely fit into the active site, will produce high amounts of either hydrogen peroxide or water in lieu of hydroxylated laurate (Whitehouse et al., 2012; Cryle and De Voss, 2008). Consequently, the apparent K_M value generated from the measurement of NADPH consumption may not accurately reflect the binding affinity of lauric acid for the TM2 variant.

Across all variants, the reductions in catalytic activity were accompanied by significant alterations to the regioselectivity of fatty acid hydroxylation. Both the DM and TM2 variants successfully produced laurate hydroxylated at the δ position, with the TM2 variant producing gamma and β hydroxylated laurate. As previously discussed, fatty acids hydroxylated at these positions serve as valuable precursors in the production of lactones, which are commercially important flavouring agents (Manning et al., 2019). However, these precursor compounds only accounted for 2% and 6% of the total mono-hydroxylated product profile in both the DM and TM2 mutants, respectively. Disappointingly, these valuable positions were not reached by the TM1 variant – which displayed a mere 9% total turnover of laurate – the lowest displayed by all variants across all the tested fatty acids. Severe reductions in both the turnover and catalytic ability of variants which displayed radically altered product profiles against laurate were also reported by Dietrich et al. (2009).

Turnover data were more encouraging in the case of myristate and palmitate. Across all variants, mono-hydroxylations occurring at novel positions on myristate accounted for over 75% of all products. Hydroxylation occurred predominantly in the mid-chain positions of ω-5, ω-6 and ω-7, with the TM1 variant favouring the production of ω-7 hydroxy-myristate and the DM and TM2 variants producing an excess of ω-5 hydroxy-myristate. In addition, these alterations to regioselectivity were not accompanied by a collapse in percentage turnover as observed for the triple variants with laurate substrate. Mid-chain hydroxylation was also apparent for palmitate – with the most abundant product produced by all variants being ω-7 hydroxy-palmitate. Once again, the TM2 variant was able to push hydroxylation further from the ω terminus, reaching the ω-9 carbon of palmitate.

In nature, hydroxylation of fatty acid substrates occurring at these mid-chain positions seldom occurs outside of eukaryotic P450s. Mid-chain hydroxylation has been demonstrated by members of the CYP1 and CYP2 classes, often with a high degree of regioselectivity (Hammerer et al., 2018). One notable example, CYP1M1 (isolated from the Rainbow trout *Oncorhynchus mykiss*), exclusively produces laurate hydroxylated at the ω-6

position (Yang, et al., 1998). Within plants, CYP703A2 (isolated from *Arabidopsis thaliana*) catalyses laurate hydroxylation between the ω-3 and ω-6 carbons, chiefly producing ω-5 hydroxy-laurate (Morant et al., 2007). CYP70A2 also performs mid-chain hydroxylation on myristate and palmitate, albeit at much lower rates of conversion in comparison to laurate (Morant et al., 2007). Mid-chain hydroxylation has also been observed in plant P450s belonging to the CYP81, CYP703 and CYP709 classes with a variety of fatty acids (Kandel et al., 2006). Human CYP4Z1, overexpression of which correlates with poor prognosis of human cancers, produces hydroxylated fatty acid products in the range of ω-2 to ω-5 with laurate and myristate (Zöllner, et al., 2009). CYP4Z1 is currently the only human CYP4 enzyme for which this pattern of mid-chain hydroxylation has been reported (Zöllner et al., 2009).

Within prokaryotes, mid-chain hydroxylation is performed by CYP107H1 (P450-BioI) – a gene present within the biotin biosynthetic operon of *Bacillus subtilis* (Greule et al., 2018). P450-BioI hydroxylates myristate at the ω-7 position as the initial step in the production of pimelic acid, an intermediate in the synthesis pathway of biotin (vitamin B_7) (De Voss and Cryle, 2007; Lawson et al., 2004). The binding pocket of P450-BioI extends beyond the heme domain, providing a space in which the ω terminus of fatty acids can be sequestered (Cryle and Schlichting, 2008). As revealed by substrate-bound P450-BioI crystal structures (PDB 3EJB) – this space enables fatty acids to adopt a U-shape position above the heme group, with the ω-6 and ω-7 carbons exposed directly over the heme iron (Cryle and Schlichting, 2008).

Docking simulations conducted using the crystal structures of the TM1 and TM2 variant heme domains - in tandem with product profile data - suggest that fatty acid substrates adopt a similar U-shaped conformation in both structures. In wild-type P450 BM3, residues V78, A82, F87, T88, I263 and L264 comprise an asymmetric subpocket that is postulated to influence regioselectivity (Chiang et al., 2013). The subpocket of the wild-type enzyme is sterically constricted and can only accommodate alkane groups of 3 or 4 carbons in length, restricting hydroxylation to the ω-1, ω-2 and ω-3 positions (Chiang et al., 2015; Haines et al., 2001). In the TM1 variant, docking simulations indicate that the fatty acid palmitate occupies a subpocket widened by the V78A, F87V and I263P substitutions. The ω terminus of palmitate points towards the B'-helix, residing in the space created by the V78A substitution, while the sub-terminal carbons up to and including the ω-6 position occupy the subpocket widened by the F87V and I263P mutations. The ω-7 carbon, comprising the

bottom of the U-shape, directly faces the heme iron. Docking simulations involving the TM2 variant also display palmitate occupying a deeper position within the active site relative to the wild-type simulations - consequently enabling mid-chain carbons greater access to the heme iron. The TM2 variant lacks a V78A substitution within the active site and, consequently, the ω-terminus of palmitate bends back on itself, facing away from the B'-helix and in the direction of its carboxylate group. These docking simulations therefore provide a rational basis for the alterations to regioselective fatty acid hydroxylation displayed by the DM, TM1 and TM2 variants. Namely, expansion of the binding subpocket removes steric hindrance present within the wild-type BM3 heme domain and enables fatty acid substrates to slip further into the active site, refocusing hydroxylation at mid-chain positions as a consequence.

Moreover, docking studies on the TM1 variant heme domain suggested the formation of a hydrogen bond between residue S72 of the B'-helix and the carboxylate group of the fatty acid substrate. This interaction may serve to anchor the substrate – focusing hydroxylation at the ω-7 position and limiting access to carbons closer to the α carbon. This hypothesis is supported by product profiles collected for myristate and palmitate with the TM1 variant, in which products hydroxylated at the ω-7 positon predominate (36% and 63% for myristate and palmitate, respectively). Interestingly, despite the immediate area around the active site being larger in the TM1 variant than in the DM and TM2 variants, the product profile generated with laurate was less diverse. The comparative lack of product diversity displayed by the TM1 variant with laurate suggests that the propensity of fatty acids to access the widened subpocket and interact with residue S72 is dependent on chain length. Introduction of a larger residue at positon S72, such as the S72Y substitution introduced as part of a wider mutant by Dietrich et al. (2009) may serve to stabilise fatty acids of less than 14 carbons deeper within the binding pocket. In the case of laurate, the introduction of a new anchor point via a S72Y substitution was shown to be essential to reaching positions close to the α carbon.

In contrast, the ability of TM2 to hydroxylate positions inaccessible to both DM and TM1 with laurate and palmitate suggests that the A184K substitution, in combination with F87V and I263P mutations, affords the substrate greatest fluidity within the active site of all variants examined. This flexibility can likely be attributed to the lifted active site canopy caused by the A184K substitution. In addition, the lack of a tethering hydrogen bonding

interaction enables the fatty acids to adopt multiple binding conformations within the active site, facilitating a diverse product profile.

Active site expansion is also likely to underpin the increases in styrene oxide production displayed by all the variants. Increases in the production of styrene oxide are particularly pronounced in the DM and TM1 variants, which both display near 10-fold increases relative to the wild-type enzyme. Despite possessing the tightest binding affinity for styrene (73.3 ± 4.60 μM), the TM1 variant generates much smaller amounts of styrene oxide (34.1 ± 2.59 μM) when compared to the DM (166 ± 17.0 μM) and TM2 (161 ± 14.2 μM) variants. With the exception of the TM1 variant, which features a V78A substitution, both the DM and TM2 mutants contain the same residue substitutions in the binding subpocket close to the heme iron. It can be postulated that, akin to the suggested binding modes of the fatty acids, the vinyl group of the styrene is sequestered into the expanded subpocket. This pose would consequently disfavour the production of styrene oxide, as the vinyl group would be positioned away from the heme iron in a manner that would prevent efficient epoxidation.

Production of the *S*-enantiomer of styrene oxide was favoured by all variants, an inversion of the preference reported for the wild-type enzyme – which produces *R*(+)-styrene oxide in excess (Tee and Schwanenberg, 2006; Shehzad et al., 2013; Whittall et al., unpublished data). The I263P mutation in isolation was shown to produce a near racemic mix of both *R*(+) and *S*(-) enantiomers of styrene oxide (Whittall et al., unpublished data). Following the addition of the F87V mutation, the enantiomeric excess of *S*(-)-styrene oxide produced by the DM variant climbed to 49.4 ± 0.200 %. This excess was followed by more modest increases to 56.7 ± 1.75 % (TM1 variant) and 56.7 ± 0.306 % (TM2 variant) following V78A and A184K substitutions to the DM variant, respectively. These data suggest that the F87V substitution, in combination with I263P, has the greatest impact on inverting the enantiopreference of styrene epoxidation.

Docking simulations were performed using the TM1 and TM2 heme domain structures (in addition to the wild-type) in an attempt to rationalise the alterations in styrene enantioselectivity (supplementary information, Figures 3.S1, 3.S2 and 3.S3). However, unlike fatty acids, polar interactions are not involved with the orientation of styrene in the active site (Fruetel et al., 1992). As the active site channel of BM3 is hydrophobic, docking studies performed using styrene suggest the substrate possesses a high degree of mobility within the active site. As a result of this, docking simulations across both triple variants and

the wild-type enzyme proved inconclusive, as simulations generated poses corresponding to both the pro-*R* and pro-*S* positons.

Despite the lack of structural rationale provided by the docking studies, previously reported crystal structures of *S*-selective BM3 variants (PDB 4HGF, 4HGG) suggest that the kink angle of the I-helix plays a significant role in the orientation of styrene within the active site (Shehzad et al., 2013). As outlined in Whittall et al. (unpublished data), the inclusion of a proline residue within the I-helix (I263P) marginally reduces the kink of the I-helix. While unconfirmed, this reduction could be on account of a water molecule interrupting the hydrogen bonding network of the I-helix (Whittall et al. unpublished data; Haines et al., 2001). Moreover, the inclusion of an F87A substitution in the structures collected by Shehzad et al. (2013) indicated that the substitution of F87 was a prerequisite for providing sufficient space for styrene binding. In addition, the removal of the sterically restrictive F87 residue also orientated styrene in a position that favoured production of the *S*-enantiomer. In contrast, studies conducted by Huang et al. (2011) demonstrated that constricting the volume of the active site produced variants that favoured the production of the *R*-enantiomer of styrene oxide. These data suggest that translocation of the I-helix, coupled to the additional space provided by substitution of residue F87, are the key factors that control production of the *S*-enantiomer of styrene oxide in P450 BM3.

While the addition of charge at position 184, (A184K) has a significant impact on the reorganisation of the F and G helices, in addition to the translocation of the I-helix, it alone does not command as significant a role in governing enantioselectivity, as previously reported (Tee and Schwannenberg, 2006; Shehzad et al., 2013). Greater levels of enantioselectivity were produced by directly altering the I-helix and expanding active site environment. The DM and TM1 variants, both without A184K substitutions, display much higher enantioselectivity towards the production of *S*(-)-styrene oxide (49.4 ± 0.200 and 56.7 ± 1.75% *ee*, respectively) compared to the data reported by Shehzad et al (2013) (27.3% *ee* for a F87A/A184K/T325A variant). However, crystal structures of all variant heme domains bound to styrene must be obtained before any substantial conclusions can be drawn regarding the exact roles that the I-helix, the expanded active site and the A184K substitution perform regarding styrene enantioselectivity.

To conclude, the single I263P mutant was successfully utilised to develop a collection of variants with significantly altered regio and enantioselectivity relative to the wild-type enzyme. Hydroxylase activity of BM3 was refocused towards mid-chain positions across a

range of fatty acids, yielding product profiles analogous to those of eukaryotic P450s. While hydroxylation at positions corresponding to lactone precursors was not achieved in significant amounts, BM3 was successfully re-engineered to oxidise myristate and palmitate at mid-chain positions with high degrees of regioselectivity. In addition to producing fatty acid metabolites akin to a plethora of eukaryotic P450s, the TM1 variant produced a biotin precursor, ω-7 hydroxymyristate, with a high level of specificity. This product is also produced by wild-type P450-BioI (Lawson et al., 2004). However, the majority of products produced by P450-BioI are, as with wild-type BM3, hydroxylated at the ω-1, ω-2 and ω-3 positions (Cryle et al., 2003). Variant TM1 may therefore provide a more efficient alternative in the initial stages of industrial synthesis of biotin. In addition to being an important B vitamin, biotin is widely used in labelling via covalent attachment to a protein of interest – a process known as biotinylation (Bayer and Wilchek, 1990).

Future studies will seek to incorporate additional point mutations to the TM1 variant at position S72. These substitutions will seek to replicate the tethering hydrogen bonding interaction between S72 and the carboxylate of medium size (C:14 – C:16) fatty acids with smaller fatty acids (<C:14). In addition, activity of variants towards fatty acids greater than C:16 will be examined. Mid-chain hydroxylation occurring at the ω-6 and ω-8 positons of octadecanoate (C:18) produces 10- and 12-hydroxystearic acids, respectively; both serve as important industrial lubricants (Hou, 2009; Hou, 2000).

The enantiomeric excess of *S*(-)-styrene oxide (56.7 ± 1.86 % and 56.7 ± 0.306 *ee*) generated by variants TM1 and TM2 are, to our knowledge, the highest yet reported for BM3 variants. Future work regarding the enantioselectivity of styrene epoxidation will focus on further modifications to the I-helix. Identification of further substitutions and residues of interest will be facilitated by the determination of styrene-bound crystal structures using the existing DM, TM1 and TM2 variant heme domains.

3.6. References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. and Zwart, P. H. (2010). 'PHENIX: a comprehensive Python-based system for macromolecular structure solution', *Acta Crystallogr D Biol Crystallogr,* 66(Pt 2), pp. 213-21.

Bayer, E. A. and Wilchek, M. (1990). 'Protein biotinylation', *Methods Enzymol,* 184, pp. 138-60.

Bloom, J. D., Dutia, M. D., Johnson, B. D., Wissner, A., Burns, M. G., Largis, E. E., Dolan, J. A. and Claus, T. H. (1992). 'Disodium (R,R)-5-[2-[[2-(3-chlorophenyl)-2-hydroxyethyl]-amino] propyl]-1,3-benzodioxole-2,2-dicarboxylate (CL 316,243). A potent beta-adrenergic agonist virtually specific for beta 3 receptors. A promising antidiabetic and antiobesity agent', *J Med Chem,* 35(16), pp. 3081-4.

Butler, C. F., Peet, C., Mason, A. E., Voice, M. W., Leys, D. and Munro, A. W. (2013). 'Key mutations alter the cytochrome P450 BM3 conformational landscape and remove inherent substrate bias', *J Biol Chem,* 288(35), pp. 25387-99.

Chen, Z., Ost, T. W. and Schelvis, J. P. (2004). 'Phe393 mutants of cytochrome P450 BM3 with modified heme redox potentials have altered heme vinyl and propionate conformations', *Biochemistry*, 43(7), pp.1798-1808.

Chiang, C. H., Ramu, R., Tu, Y. J., Yang, C. L., Ng, K. Y., Luo, W. I., Chen, C. H., Lu, Y. Y., Liu, C. L. and Yu, S. S. (2013). 'Regioselective hydroxylation of C(12)-C(15) fatty acids with fluorinated substituents by cytochrome P450 BM3', *Chemistry,* 19(41), pp. 13680-91.

Cirino, P. C. and Arnold, F. H. (2002). 'Regioselectivity and activity of cytochrome P450 BM-3 and mutant F87A in reactions driven by hydrogen peroxide', *Adv Syn Catal*, *344*(9), pp. 932-37.

Cirino, P. C. and Arnold, F. H. (2003). 'A self-sufficient peroxide-driven hydroxylation biocatalyst', *Angew Chem Int Ed Engl,* 42(28), pp. 3299-301.

Cryle, M. J., Matovic, N. J. and De Voss, J. J. (2003). 'Products of cytochrome P450BioI (CYP107H1)-catalyzed oxidation of fatty acids', *Org Lett*, 5(18), pp.3341-44.

Cryle, M. J., Espinoza, R. D., Smith, S. J., Matovic, N. J. and De Voss, J. J. (2006). 'Are branched chain fatty acids the natural substrates for P450(BM3)?', *Chem Commun (Camb),* (22), pp. 2353-5.

Cryle, M. J. and De Voss, J. J. (2007). 'Facile determination of the absolute stereochemistry of hydroxy fatty acids by GC: application to the analysis of fatty acid oxidation by a P450BM3 mutant', Tetrahedron: Asymmetry, 18(4), pp.547-51.

Cryle, M. J. and De Voss, J. J. (2008). 'The Role of the Conserved Threonine in P450BM3 Oxygen Activation: Substrate-Determined Hydroxylation Activity of the Thr268Ala Mutant', ChemBioChem, 9(2), pp. 261-66.

Cryle, M. J. and Schlichting, I. (2008). 'Structural insights from a P450 Carrier Protein complex reveal how specificity is achieved in the P450(BioI) ACP complex', *Proc Natl Acad Sci U. S. A.,* 105(41), pp. 15696-701.

Davydov, D.R., Halpert, J.R., Renaud, J.P. and Hoa, G.H.B. (2003). Conformational heterogeneity of cytochrome P450 3A4 revealed by high pressure spectroscopy. *Biochemical and biophysical research communications*, 312(1), pp.121-130.

Dawson, J. H. (1988). 'Probing structure-function relations in heme-containing oxygenases and peroxidases', *Science*, 240(4851), pp. 433-39.

Desai, L. V., Hull, K. L. & Sanford, M. S. (2004). 'Palladium-catalyzed oxygenation of unactivated sp3 C-H bonds', *J Am Chem Soc,* 126(31), pp. 9542-3.

De Voss, J. J. and Cryle, M. J. (2007). 'Carbon-carbon bond cleavage by P450 systems', *Met Ions Life Sci*, 3, pp.397-435.

Dietrich, M., Do, T. A., Schmid, R. D., Pleiss, J. and Urlacher, V. B. (2009). 'Altering the regioselectivity of the subterminal fatty acid hydroxylase P450 BM-3 towards gamma- and delta-positions', *J Biotechnol,* 139(1), pp. 115-7.

Doyle, S. A., Fung, S. Y. and Koshland, D. E. (2000). 'Redesigning the substrate specificity of an enzyme: isocitrate dehydrogenase', *Biochemistry,* 39(46), pp. 14348-55.

Eiben, S., Kaysser, L., Maurer, S., Kühnel, K., Urlacher, V. B. and Schmid, R. D. (2006). 'Preparative use of isolated CYP102 monooxygenases - a critical appraisal', *J Biotechnol,* 124(4), pp. 662-9.

Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (2010). 'Features and development of Coot', *Acta Crystallogr D Biol Crystallogr,* 66(Pt 4), pp. 486-501.

English, N., Palmer, C. N., Alworth, W. L., Kang, L., Hughes, V. and Wolf, C. R. (1997). 'Fatty acid signals in *Bacillus megaterium* are attenuated by cytochrome P-450-mediated hydroxylation', *Biochem J,* 327 (Pt 2), pp. 363-8.

Fasan, R., 2012. 'Tuning P450 enzymes as oxidation catalysts', *ACS Catal*, *2*(4), pp.647-66

Finan, D. and Barco, J., 2017. 'Protein Characterization Turned Up to 11: A Flexible Platform for Protein Stability Measurements', *Genet Eng Biotech N*, 37(3), pp.22-23.

Fruetel, J. A., Collins, J. R., Camper, D. L., Loew, G. H. and Ortiz de Montellano, P. R. (1992). 'Calculated and experimental absolute stereochemistry of the styrene and beta-methylstyrene epoxides formed by cytochrome P 450cam', *J Am Chem Soc*, 114(18), pp.6987-93.

Godula, K. & Sames, D. (2006). 'C-H bond functionalization in complex organic synthesis', *Science,* 312(5770), pp. 67-72.

Greule, A., Stok, J. E., De Voss, J. J. and Cryle, M. J. (2018). 'Unrivalled diversity: the many roles and reactions of bacterial cytochromes P450 in secondary metabolism', *Nat Prod Rep,* 35(8), pp. 757-91.

Haines, D. C., Tomchick, D. R., Machius, M. and Peterson, J. A. (2001). 'Pivotal role of water in the mechanism of P450BM-3', *Biochemistry,* 40(45), pp. 13456-65.

Hammerer, L., Winkler, C. K. and Kroutil, W., 2018. 'Regioselective biocatalytic hydroxylation of fatty acids by cytochrome P450s', *Catal Lett*, 148(3), pp.787-812.

Hou, C.T. (2000) 'Biotransformation of unsaturated fatty acids to industrial products', *Adv Appl Microbiol* 47, 201–20

Hou, C.T. (2009). 'Biotechnology for fats and oils: new oxygenated fatty acids', *New Biotechnol*, *26*(1-2), pp.2-10.

Huang, W. C., Cullis, P. M., Raven, E. L. and Roberts, G. C. (2011). 'Control of the stereo-selectivity of styrene epoxidation by cytochrome P450 BM3 using structure-based mutagenesis', *Metallomics,* 3(4), pp. 410-16.

Joosten, R. P., Salzemann, J., Bloch, V., Stockinger, H., Berglund, A. C., Blanchet, C., Bongcam-Rudloff, E., Combet, C., Da Costa, A. L., Deleage, G., Diarena, M., Fabbretti, R., Fettahi, G., Flegel, V., Gisel, A., Kasam, V., Kervinen, T., Korpelainen, E., Mattila, K., Pagni, M., Reichstadt, M., Breton, V., Tickle, I. J. and Vriend, G. (2009). 'PDB_REDO: automated re-refinement of X-ray structure models in the PDB', *J Appl Crystallogr,* 42(Pt 3), pp. 376-84.

Kandel, S., Sauveplane, V., Olry, A., Diss, L., Benveniste, I. and Pinot, F. (2006). 'Cytochrome P450-dependent fatty acid hydroxylases in plants', *Phytochem Rev*, 5(2-3), pp.359-72.

Katsuki, T., 2001. 'The Catalytic Enantioselective Synthesis of Optically Active Epoxides and Tetrahydrofurans. Asymmetric Epoxidation, the Desymmetrization of meso-Tetrahydrofurans, and Enantiospecific Ring-Enlargement', *Curr Org Chem*, 5(6), pp.663-78.

Kim, K. R. and Oh, D. K. (2013). 'Production of hydroxy fatty acids by microbial fatty acid-hydroxylation enzymes', *Biotechnol Adv,* 31(8), pp. 1473-85.

Koshland Jr, D.E., 2002. 'The application and usefulness of the ratio k_{cat}/K_M', *Bioorg Chem*, 30(3), pp. 211-13.

Krieger, E. and Vriend, G. (2015). 'New ways to boost molecular dynamics simulations', *J Comput Chem,* 36(13), pp. 996-1007.

Lamb, D. C. and Waterman, M. R. (2013). 'Unusual properties of the cytochrome P450 superfamily', *Philos Trans R Soc Lond B Biol Sci,* 368(1612), p. 20120434.

Lawson, R. J., Leys, D., Sutcliffe, M. J., Kemp, C. A., Cheesman, M. R., Smith, S. J., Clarkson, J., Smith, W. E., Haq, I., Perkins, J. B. and Munro, A. W. (2004). 'Thermodynamic and biophysical characterization of cytochrome P450 BioI from *Bacillus subtilis*', *Biochemistry,*43(39), pp. 12410-26.

Loida, P. J. and Sligar, S. G. (1993). 'Molecular recognition in cytochrome P-450: mechanism for the control of uncoupling reactions', *Biochemistry,* 32(43), pp. 11530-8.

Manna, S. K. and Mazumdar, S. (2008). 'Reversible inactivation of cytochrome P450 by alkaline earth metal ions: auxiliary metal ion induced conformation change and formation of inactive P420 species in CYP101', *J Inorg Biochem,* 102(5-6), pp. 1312-21.

Manning, J., Tavanti, M., Porter, J. L., Kress, N., De Visser, S. P., Turner, N. J. and Flitsch, S. L., 2019. 'Regio-and Enantio-selective Chemo-enzymatic C-H-lactonization of Decanoic Acid to (S)-δ-Decalactone', *Angew Chem Int Ed Engl*. 58(17), pp. 5668-71

McLean, K. J., Cheesman, M. R., Rivers, S. L., Richmond, A., Leys, D., Chapman, S. K., Reid, G. A., Price, N. C., Kelly, S. M., Clarkson, J., Smith, W. E. and Munro, A. W. (2002). 'Expression, purification and spectroscopic characterization of the cytochrome P450 CYP121 from *Mycobacterium tuberculosis*', *J Inorg Biochem,* 91(4), pp. 527-41.

Miura, Y. and Fulco, A. J. (1975). 'Omega-1, Omega-2 and Omega-3 hydroxylation of long-chain fatty acids, amides and alcohols by a soluble enzyme system from *Bacillus megaterium*', *Biochim Biophys Acta,* 388(3), pp. 305-17.

Morant, M., Jørgensen, K., Schaller, H., Pinot, F., Møller, B. L., Werck-Reichhart, D. and Bak, S. (2007). 'CYP703 is an ancient cytochrome P450 in land plants catalyzing in-chain hydroxylation of lauric acid to provide building blocks for sporopollenin synthesis in pollen', *Plant Cell,* 19(5), pp. 1473-87.

Munro, A. W., Daff, S., Coggins, J. R., Lindsay, J. G. & Chapman, S. K. (1996). 'Probing electron transfer in flavocytochrome P-450 BM3 and its component domains', *Eur J Biochem,* 239(2), pp. 403-9.

Munro, A. W., Girvan, H. M. and McLean, K. J. (2007). 'Variations on a (t)heme - novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily', *Nat Prod Rep,* 24(3), pp. 585-609.

Munro, A. W., Leys, D. G., McLean, K. J., Marshall, K. R., Ost, T. W., Daff, S., Miles, C. S., Chapman, S. K., Lysek, D. A., Moser, C. C., Page, C. C. and Dutton, P. L. (2002). 'P450 BM3: the very model of a modern flavocytochrome', *Trends Biochem Sci,* 27(5), pp. 250-7.

Narhi, L. O. and Fulco, A. J. (1986). 'Characterization of a catalytically self-sufficient 119,000-dalton cytochrome P-450 monooxygenase induced by barbiturates in *Bacillus megaterium*', *J Biol Chem,* 261(16), pp. 7160-9.

Noble, M. A., Miles, C. S., Chapman, S. K., Lysek, D. A., MacKay, A. C., Reid, G. A., Hanzlik, R. P. and Munro, A. W. (1999). 'Roles of key active-site residues in flavocytochrome P450 BM3', *Biochem J,* 339 (Pt 2), pp. 371-9.

Oliver, C. F., Modi, S., Sutcliffe, M. J., Primrose, W. U., Lian, L. Y. and Roberts, G. C. (1997). 'A single mutation in cytochrome P450 BM3 changes substrate orientation in a catalytic intermediate and the regiospecificity of hydroxylation', *Biochemistry,* 36(7), pp. 1567-72.

Omura, T. and Sato, R. (1964) 'The carbon monoxide-binding pigment of liver microsomes. I: evidence for its hemoprotein nature'*, J Biol Chem*. 239, 2370-78.

Panke, S., Witholt, B., Schmid, A. and Wubbolts, M. G. (1998). 'Towards a biocatalyst for (S)-styrene oxide production: characterization of the styrene degradation pathway of *Pseudomonas* sp. strain VLB120', *Appl Environ Microbiol,* 64(6), pp. 2032-43.

Pazmino, D.T., Winkler, M., Glieder, A. and Fraaije, M.W., 2010. 'Monooxygenases as biocatalysts: classification, mechanistic aspects and biotechnological applications', *J Botechnol*, *146*(1-2), pp. 9-24.

Rentmeister, A., Brown, T.R., Snow, C.D., Carbone, M.N. and Arnold, F.H., 2011. 'Engineered bacterial mimics of human drug metabolizing enzyme CYP2C9'. *ChemCatChem*, 3(6), pp.1065-71.

Rousseau, G. and Breit, B. (2011). 'Removable directing groups in organic synthesis and catalysis', *Angew Chem Int Ed Engl,* 50(11), pp. 2450-94.

Sakaki, T. (2012). 'Practical application of cytochrome P450', *Biol Pharm Bull,* 35(6), pp. 844-9.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual. Cold Spring Harbor laboratory press*. *New York*.

Shehzad, A., Panneerselvam, S., Linow, M., Bocola, M., Roccatano, D., Mueller-Dieckmann, J., Wilmanns, M. and Schwaneberg, U. (2013). 'P450 BM3 crystal structures reveal the role of the charged surface residue Lys/Arg184 in inversion of enantioselective styrene epoxidation', *Chem Commun (Camb),* 49(41), pp. 4694-6.

Smith, S. J., Munro, A. W. & Smith, W. E. (2003). 'Resonance Raman scattering of cytochrome P450 BM3 and effect of imidazole inhibitors', *Biopolymers,* 70(4), pp. 620-7.

Spiro, T.G., Stong, J.D. and Stein, P. (1979). 'Porphyrin core expansion and doming in heme proteins. New evidence from resonance Raman spectra of six-coordinate high-spin iron (III) hemes', *J Am Chem Soc*, *101*(10), pp.2648-55.

Sun, J., Fujita, S.I., Zhao, F. and Arai, M. (2005). 'A highly efficient catalyst system of ZnBr2/n-Bu4NI for the synthesis of styrene carbonate from styrene oxide and supercritical carbon dioxide', *Appl Catal A: General*, 287(2), pp.221-6.

Takahashi, O., Umezawa, J., Furuhashi, K. and Takagi, M. (1989). 'Stereocontrol of a tertiary hydroxyl group via microbial epoxidation: A facile synthesis of prostaglandin ω-chains', *Tetrahedron Lett*, 30(12), pp.1583-4

Tee, K. L. and Schwaneberg, U. (2006). 'A screening system for the directed evolution of epoxygenases: importance of position 184 in P450 BM3 for stereoselective styrene epoxidation', *Angew Chem Int Ed Engl,* 45(32), pp. 5380-3.

Torres Pazmiño, D. E., Winkler, M., Glieder, A. and Fraaije, M. W. (2010). 'Monooxygenases as biocatalysts: Classification, mechanistic aspects and biotechnological applications', *J Biotechnol,* 146(1-2), pp. 9-24.

Urlacher, V. B. and Girhard, M. (2012). 'Cytochrome P450 monooxygenases: an update on perspectives for synthetic application', *Trends Biotechnol,* 30(1), pp. 26-36.

Urlacher, V. B. and Girhard, M. (2019). 'Cytochrome P450 Monooxygenases in Biotechnology and Synthetic Biology', *Trends Biotechnol,* 37(8), pp. 882-97.

Waché, Y., Aguedo, M., Nicaud, J. M. and Belin, J. M. (2003). 'Catabolism of hydroxyacids and biotechnological production of lactones by *Yarrowia lipolytica*', *Appl Microbiol Biotechnol,* 61(5-6), pp. 393-404.

Whitehouse, C. J., Bell, S. G. and Wong, L. L. (2012). 'P450(BM3) (CYP102A1): connecting the dots', *Chem Soc Rev,* 41(3), pp. 1218-60.

Yang, Y. H., Wang, J. L., Miranda, C. L. and Buhler, D. R. (1998). 'CYP2M1: cloning, sequencing, and expression of a new cytochrome P450 from rainbow trout liver with fatty acid (omega-6)-hydroxylation activity', *Arch Biochem Biophys,* 352(2), pp. 271-80.

Zheng, Y., Li, L., Liu, Q., Zhang, H., Cao, Y., Xian, M. and Liu, H. (2016). 'High-specificity synthesis of novel monomers by remodeled alcohol hydroxylase', *BMC Biotechnol,* 16(1), p. 61.

Zöllner, A., Dragan, C. A., Pistorius, D., Müller, R., Bode, H. B., Peters, F. T., Maurer, H. H. and Bureik, M. (2009). 'Human CYP4Z1 catalyzes the in-chain hydroxylation of lauric acid and myristic acid', *Biol Chem,* 390(4), pp. 313-7.

3.7. Supplementary Information

Table 3.S1 Pairs of m/z fragments used to identify mono-hydroxylated fatty acid products

Fatty acid (C:n)	*m/z* Fragment "fingerprints" for mono-hydroxylated fatty acid products								
	ω -9	*ω -8*	*ω -7*	*ω -6*	*ω -5*	*ω -4*	*ω -3*	*ω -2*	*ω -1*
Laurate C:12	233/229	247/215	261/201	275/187	289/173	303/159	317/145	331/131	345/117
Myristate C:14	261/229	275/215	289/201	303/187	317/173	331/159	345/145	359/131	373/117
Palmitate C:16	289/229	303/215	317/201	331/187	345/173	359/159	373/145	387/131	401/117

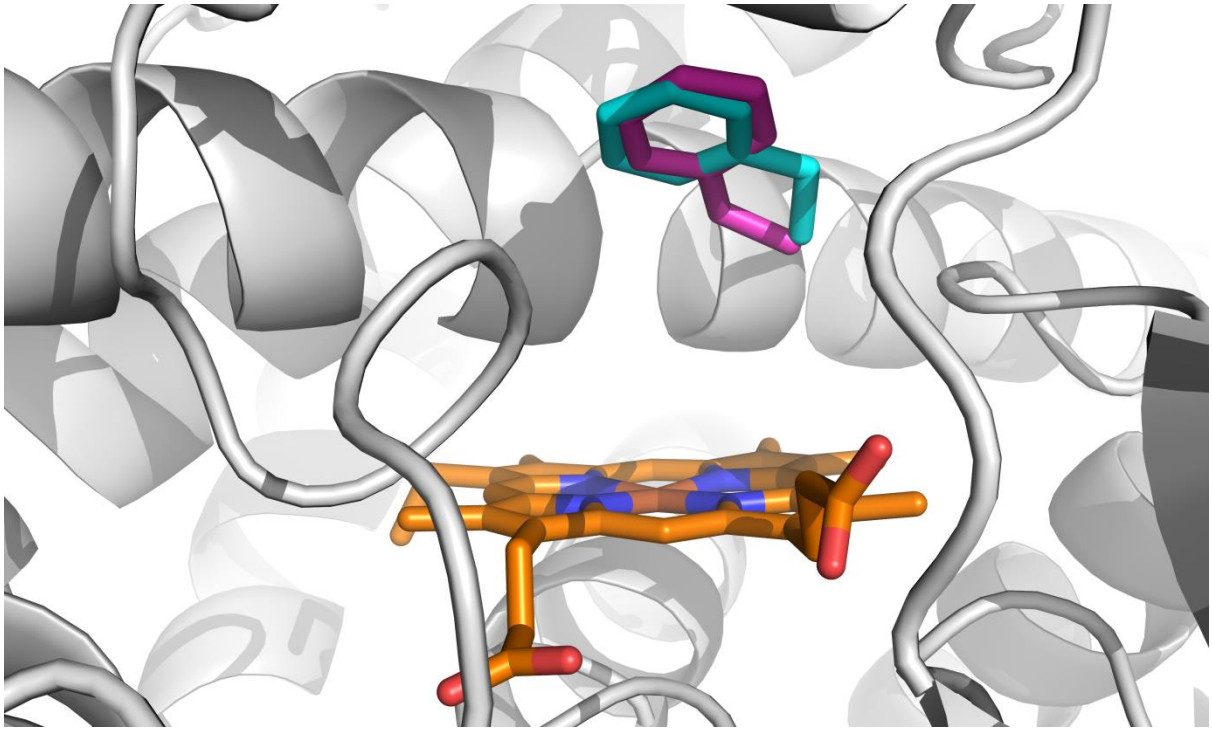

Figure 3.S1 Model of styrene binding within the active site of wild-type BM3. Model generated using YASARA mediated docking simulations. Wild-type structure adapted from PDB 1JPZ (Haines et al., 2001). The docking pose of styrene favouring production of the (*R*)-enantiomer of styrene oxide is highlighted in cyan. The pose favouring the production of (*S*)-styrene oxide is presented in magenta. Docking poses are framed relative to the heme cofactor, highlighted in orange.

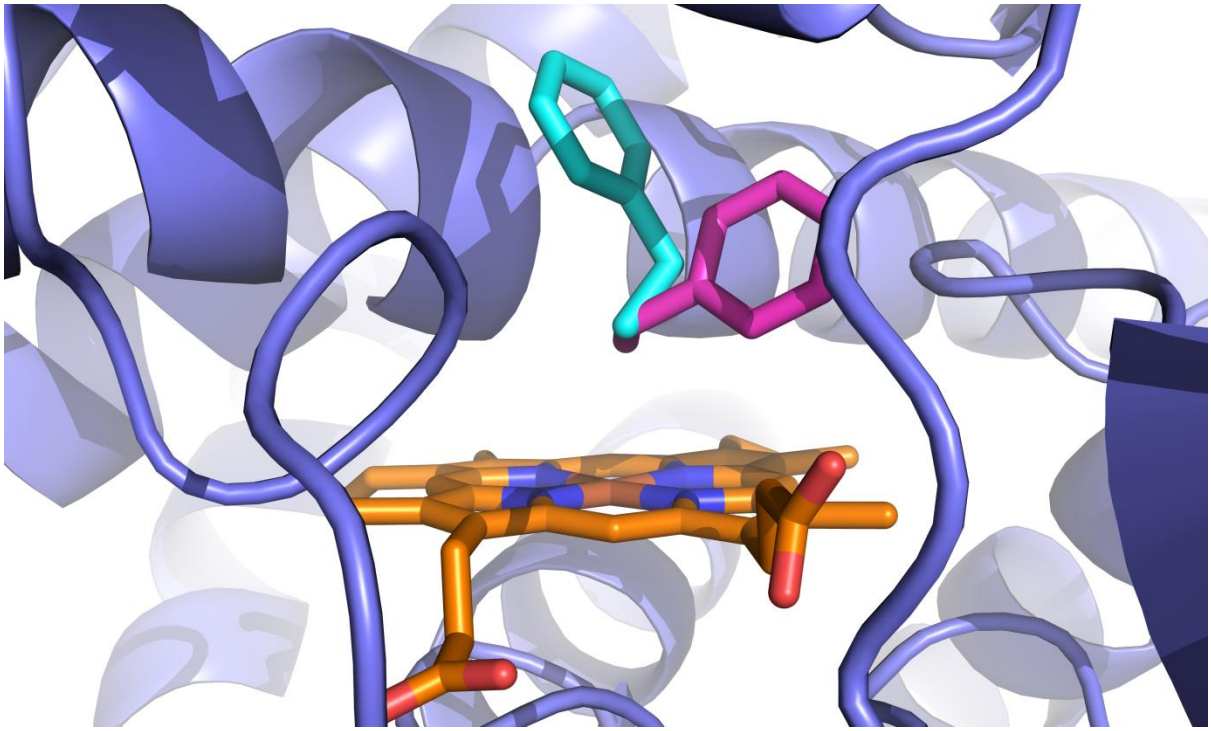

Figure 3.S2 Simulated docking poses of styrene within the active site of BM3 variant TM1 (I263P/F87V/V78A).
The secondary structure of the TM1 heme domain is presented in blue. The styrene pose highlighted in cyan will lead to the production of the (*R*)-enantiomer of styrene oxide. The pose highlighted in magenta will lead to the formation of the (*S*)-enantiomer of styrene oxide. Poses were generated by docking simulations performed using YASARA software, as detailed in the methods.

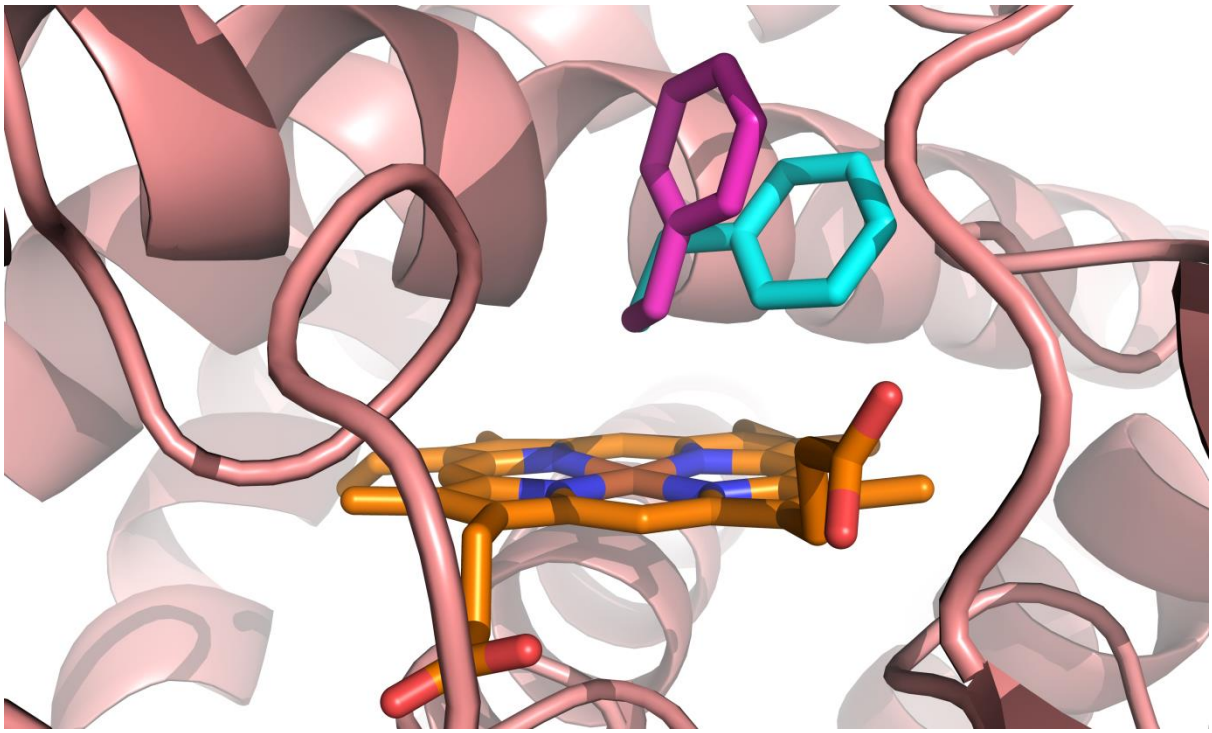

Figure 3.S3 Model of styrene binding within the heme domain of BM3 variant TM2 (I263P/F87V/A184K).
Secondary structural motifs of TM2 are presented in pink, with the heme cofactor being presented in orange. Styrene substrate is represented in two poses. The pose highlighted in cyan will lead to the formation of (*R*)-styrene oxide, with the magenta pose leading to the formation of (*S*)-styrene oxide. Docking studies were performed using YASARA software.

Chapter 4: Application of Native-MS Mediated Compound Screening To a Highly Promiscuous Variant P450.

4.1. Abstract

Native protein mass spectrometry (native-MS) can be utilised as a high-throughput, resource-saving screening technique, capable of identifying protein-compound interactions that could be overlooked using alternative methods. In this study, we have optimised a technique that couples characterisation of a high-activity bacterial P450 enzyme via native-MS to a commercially important drug library. The enzyme of interest is a variant of CYP102A1 (P450 BM3). The P450 BM3 variant in question, nicknamed the "gatekeeper" (F87V/A82F), performs biotechnologically relevant oxidations with a wide range of substrates, and has been shown previously to produce human drug metabolites from several important pharmaceutical compounds. Production of human drug metabolites is of considerable interest to the pharmaceutical industry. However, the use of human P450s in metabolite production is expensive and often challenging. Activity of the gatekeeper variant towards these compounds offers a cheaper bacterial route to the production of high-value hydroxylated or otherwise oxidised drug metabolites, analogous to those from human P450s. The drug library selected for screening against the gatekeeper variant consists of a diverse array of pharmaceutical compounds approved for use by the American Food and Drugs Administration (FDA). This study demonstrates that the BM3 gatekeeper variant is a well suited for compound screening using native-MS. The technique has been successfully optimised and requires less protein and compound compared with an alternative screening method. In addition, the technique can be run in a matter of minutes whilst still obtaining high-quality data. Finally, screening by native-MS was able to identify novel drug compounds unnoticed by a prior screening method as substrates for the gatekeeper enzyme. In future, we hope to apply this technique to further compound screens in order to facilitate the discovery of high-value drug metabolites.

4.2. Introduction

Throughout all stages of drug research, mass spectrometry (MS) analysis is both widely used and highly effective at identifying hit compounds or lead fragments for further development (Özbal et al., 2004). Following its introduction in the late 1980s and its development during the 1990s, the soft ionisation technique of electrospray ionisation (ESI) has become an established method of monitoring proteins and protein complexes in their native states using MS (Loo et al., 1997; Ganem et al., 1991; Fenn et al., 1989). The gentle manner of ionisation afforded by ESI preserves the weak non-covalent interactions (i.e. van der Waals interactions, hydrogen bonds and hydrophobic effects) of the protein-ligand complexes with little to no fragmentation following transfer to the gas phase (Sanglier et al., 2008; Loo et al., 1997; Ganem et al., 1991). Coupling ESI to MS analysis can therefore enable proteins and protein-ligand interactions to be characterised by MS in their "native" conformations, as opposed to their denatured or unfolded states.

While native-MS does not provide detailed structural information in a manner comparable to X-ray crystallography or NMR, native-MS nevertheless possesses several advantages compared to these other, more established biophysical techniques (Heck, 2008). As a technique, native-MS is highly sensitive, selective and rapid – picomolar amounts of compound can be clearly identified and measured to a high degree of accuracy, with samples taking mere minutes to run and analyse (Vivat Hannah et al., 2010). In addition, native-MS screening assays require much smaller quantities of both protein target and ligands required compared to alternative methods. The intrinsic masses of both the target protein and ligands serve as "labels" – with the mass-to-charge ratio and peak abundances providing information regarding binding affinity and stoichiometry (Sanglier et al., 2008; Hofstadler and Sannes-Lowery, 2006). The lack of need to label the sample components presents a significant advantage – potential adverse effects on binding properties on account of labelling modifications is avoided in native-MS (Vivat Hannah et al., 2010; Hofstadler and Sannes-Lowery, 2006). Finally, from sample mixing and injection to mass analysis, the technique is highly amenable to automation – enabling the assessment of larger compound screens in shorter spaces of time. Consequently, native-MS is a tool well suited for an orthogonal approach to high-throughput screening of drug and compound libraries, and has been successfully incorporated into the early stages of drug-discovery (Maple et al., 2012; Vivat Hannah et al., 2010).

Cytochrome P450s, a superfamily of heme containing mono-oxygenase enzymes – catalyse a plethora of physiologically important biosynthesis reactions (notably those of steroids

and antibiotics) and perform vital roles in the metabolism of xenobiotic compounds (Munro et al., 2007; Bernhardt, 2006; Ortiz de Montellano, 2005). Members of the P450 superfamily have attracted considerable interest for use in industry – many display high regio- and stereoselectivity, in addition to being highly amenable to both mutagenesis and overexpression (Urlacher et al., 2004). Regio- and stereoselective oxygenation of organic compounds is often challenging to perform using standard approaches to organic synthesis (Schaus et al., 2002). Moreover, these synthesis reactions are not well suited for use at industrial scales, and often require hazardous reagents, multiple reaction steps and significant fractionation of product mixes (Munro et al., 2007). In contrast, enzymatic approaches towards organic synthesis are often less hazardous, can be performed in a single step and yield less diverse product profiles. For example, engineering of CYP105AS1 generated a variant P450 capable of producing the "blockbuster" cholesterol-lowering drug pravastatin in a single-step from the precursor compactin (McLean, et al., 2015).

A P450 enzyme which displays one of the greatest potentials for industrial exploitation is CYP102A1 (P450 BM3), isolated from the soil-dwelling bacterium *Bacillus megaterium* (Munro et al., 2002). A model P450 system, BM3 is notable for its unique fusion arrangement between its constituent domains. The reductase domain of BM3 incorporates FAD and FMN, characteristic of a mammalian redox system, albeit with the reductase fused to the C-terminus of the P450 as part of a single, soluble polypeptide of 119 kDa (Munro et al., 2002; Narhi and Fulco, 1987). A highly efficient mono-oxygenase, BM3 has demonstrated a catalytic activity of ~17,000 min^{-1} against the fatty acid substrate arachidonate, the highest reported for a mono-oxygenase P450 (Noble et al., 1999). This high rate of catalysis is underpinned by the unique fusion arrangement of BM3, which enables efficient electron transfer from the cofactor, NADPH, through the components of the reductase domain and on to the heme of the P450 (Munro et al., 1996). While its exact physiological role is unknown, BM3 has been shown to function as a fatty acid hydroxylase, oxidising medium to long-chain length fatty acids at positons approaching the ω terminus (Munro et al., 2002; Narhi and Fulco, 1987).

Each domain of BM3 has been well characterised, and the roles of several residues in both the active site and throughout the protein are well documented (Whitehouse et al., 2012). Engineering of BM3 has generated a wide range of variants capable of producing oxidised steroids, terpenes and sundry pharmaceutical compounds (Acevedo-Rocha et al., 2018; van Vugt-Lussenberg et al., 2007; Sowden et al., 2005). Structures of the substrate-free and -

bound heme domain collected using X-ray crystallography, often combined with molecular modelling tools, inform the rational approach to BM3 mutagenesis (Haines et al., 2001; Ravichandran et al., 1993). Random and directed mutagenesis approaches have also generated BM3 variants with altered oxidative activity towards non-natural substrates with some success (Acevedo-Rocha et al., 2018; van Vugt-Lussenberg et al., 2007).

However, all of these approaches tend to alight on a selection of key residues, typically located within or at the mouth of the active site. The most widely studied residue in the protein, Phe-87, is located within the immediate vicinity of the heme iron (Whitehouse et al., 2012). The phenyl ring of Phe-87 serves as something of a selectivity filter, regulating substrate access to the heme and shielding the ω-terminus of fatty acids bound within the active site, preventing hydroxylation at this position (Oliver et al., 1997). Substitution of Phe-87 with residues of Ala or Val removes this steric hindrance and enables BM3 to accommodate a wider array of substrates. Consequently, F87V/F87A substitutions are common components of drug-metabolising BM3 variants (Whitehouse et al., 2012).

Variants which contain residue substitutions designed to expand the dimensions of the hydrophobic binding pocket are not exclusive in altering the substrate selectivity or activity of BM3. Substitutions which act to fill the hydrophobic pocket by replacing smaller residues with residues with larger, more hydrophobic side-chains can also prompt increases in catalytic efficiency and substrate selectivity (Huang et al., 2007). The most notable example of this is variant A82F, which in addition to displaying enhanced affinity for fatty acids, also induces significant conformational rearrangements within key structural motifs outside of the active site (Butler et al., 2013; Huang et al., 2007). Furthermore, the enhanced flexibility of the A82F variant imparts dramatic changes to substrate selectivity. A BM3 variant containing the A82F substitution in isolation was shown to bind to a human pharmaceutical compound, the proton-pump inhibitor (PPI) omeprazole (Butler et al., 2013).

A recent body of work conducted by the Munro group has combined both F87V and A82F substitutions into a double (F87V/A82F) variant (Jeffreys et al., unpublished; Jeffreys et al., 2019; Butler et al., 2014; Butler et al., 2013). On account of its ability to influence substrate selectivity, this F87V/A82F double mutant is referred to as the “gatekeeper” variant (Butler et al., 2014). The gatekeeper variant has demonstrated a propensity to bind to numerous human drugs, including members of the azole and fibrate classes, with considerably higher affinity than wild-type BM3 (Jeffreys et al., unpublished; Jeffreys et al., 2019). Moreover,

gatekeeper-mediated turnover was evident against a selection of compounds belonging to a wide array of drug classes, including fibrates, glitazones and statins (Jeffreys et al., unpublished). Against the proton pump inhibitor (PPI) compounds esomeprazole and lansoprazole, the gatekeeper was shown to produce oxidised metabolites akin to those produced by the human P450s CYP2C19 and CYP3A4, respectively (Butler et al., 2014). Further studies conducted by Jeffreys et al., (unpublished) revealed evidence of gatekeeper-mediated production of human drug metabolites from compounds belonging to the glitazone class of drugs.

The production of human drug metabolites is highly lucrative. During the development of new pharmaceuticals, the absorption, distribution, metabolism and excretion (ADME) of a compound must be intensely studied prior to it reaching the market (Caldwell et al., 2009). Toxicity testing of drug metabolites is therefore an essential step in drug development. As previously discussed, highly regio- or stereospecific chemical synthesis of oxidised compounds can often prove challenging and expensive. As P450s are responsible for the phase I metabolism of 70-80% of all human drugs currently in clinical use, they are well-suited for the production of human drug metabolites (Zanger and Schwab, 2013). However, human P450s are often challenging to express in high yields on account of their membrane bound states, and can display poor turnover and limited substrate specificity (O'Reilly et al., 2011). As a result, there is a demand for engineered bacterial P450s that are capable of producing metabolites analogous to those produced by human P450s.

On account of its promiscuity, a recent study conducted by Jeffreys et al., (2019; unpublished) sought to identify novel pharmaceutical compounds recognisable to the gatekeeper heme domain. Drugs taken from an FDA-approved compound library were screened against the gatekeeper heme domain by way of UV-Visible spectroscopic titrations in order to determine binding. While accurate, this method of screening is both time and resource intensive. In this study, we have applied native-MS as a high-throughput screening technique in order to assess binding interactions between compounds from the FDA library and the gatekeeper heme domain. Our study proceeded with two objectives. The first was to demonstrate the viability of native-MS as an effective screen of binding activity between the gatekeeper variant and the FDA library using known substrates. The second was to validate the method by identifying new potential substrates from the FDA library that had been previously overlooked via the screening method of UV-visible spectroscopy. Finally, turnover of a selection of substrates identified via native-MS, but

overlooked by the UV-visible spectroscopy screen, was assessed using LC-MS and tandem mass spectrometry (MS/MS) – with the goal of identifying human-like drug metabolites.

4.3. Materials and Methods

4.3.1. Sample Preparation for Native-ESI Mass Spectrometry

Purified BM3"Gatekeeper" heme domain was expressed and purification was performed by Dr Laura Jefferys as described in Jeffreys et al. (2019). Prior to native-ESI mass spectrometry, protein samples were subjected to two rounds of gel filtration in 100 mM ammonium acetate (pH ~7.5) using a HiLoad 16/600 Superdex 200 pg column (GE Life Sciences, UK) attached to an ÄKTA-Pure system (GE Healthcare, United Kingdom). Following gel filtration, protein samples were concentrated to a volume of <2 mL using VivaSpin Ultrafiltration columns with MW cut-offs of 30,000 Da (Sartorius AG, Germany). Concentrated samples were applied to columns packed with Lipidex 1000 resin (Perkin Elmer, UK) and eluted in 100 mM ammonium acetate identical to that used in gel filtration. Protein concentration was determined by UV-visible spectroscopy and samples were diluted to a concentration of 1 mg/mL (18.5 µM) using ammonium acetate buffer identical to that used in the prior purification steps. Owing to the sensitivity of the technique, it was essential to obtain high levels of protein purity prior to sample preparation. Protein purity was determined using UV-visible spectroscopy and assessed using the Reinheitszahl (Rz) purity ratio ($A_{418\ nm}/A_{280\ nm}$). Protein samples with a Rz value of 1.2 or greater were deemed suitable for native-ESI mass spectrometry studies.

Drug compounds were taken from an FDA-approved library supplied by Selleck Chemicals (L1300) (Houston, Texas, USA). Compound stocks were present in absolute DMSO at a concentration of 10 mM. Aliquots of 10 µl were removed from the compound stocks and placed into a 96-well plate (0.5 mL, 14 mm, round wells, polypropylene) (Agilent Technologies, UK). 20 µl of 1 mg/mL gatekeeper heme domain protein was added to the 96-well plate in rows alternative to the drug compounds (i.e. rows A, C, E, G contained compound, rows B, D, F, H contained protein). In total, each 96-well plate contained 48 different drug compounds.

Within the initial screening plate, drug compounds were selected according to percentage high-spin shifts calculated from binding titrations performed with the gatekeeper heme domain. Within each titration assay, 150 µM of each compound (present in absolute DMSO) were titrated against 3 µM of gatekeeper heme domain. Binding titrations were performed by Dr. Laura Jeffreys via UV-visible spectroscopy and are detailed in Jeffreys et

al., (2019). The 48 library compounds that displayed the greatest high-spin shift, i.e. those that have previously demonstrated binding affinity towards the gatekeeper heme domain, were placed into the first plate. The second plate consisted of 48 "coloured" drug compounds, the majority of which absorbed visible light within the region of 400 – 550 nm (i.e. compounds perceived to be yellow, orange and red). These compounds were selected regardless of their previously demonstrated spin-shift with the gatekeeper heme domain. Identities of the drug compounds contained within each plate, their masses, their drug target and their percentage high-spin shift recorded against the gatekeeper heme domain are presented in the supplementary material of this chapter (Tables 4.S1 and 4.S2).

4.3.2. Instrument Parameters used for Native-ESI Mass Spectrometry

The compound screening assay was performed using a 1290 Infinity II LC AutoSampler coupled to a 6545XT AdvanceBio LC/Q-TOF mass spectrometer. A 0.1 µL volume of drug compound was drawn from wells containing 10 mM stocks, and the needle was washed for 10 seconds in absolute isopropanol. The compound was subsequently injected into a well containing protein and thoroughly mixed within the protein well. The final concentration of ligand was therefore 50 µM within each well. A 1 µL volume of the protein + ligand solution was then nebulized using a Dual AJS ESI source. Sheath temperature and gas flow rate were set to 50 °C and 3 L/min to ensure native protein conditions. Nozzle voltage was set to 2000 V. Nebulizer pressure was set to 30 psig. Ion polarity of the Q-TOF instrument was positive and fragment masses within the range of 50 to 6000 m/z were acquired at a scan rate of 1 spectrum/second. A gradient of 0.1% formic acid present in water to 0.1% formic acid present in acetonitrile was utilised at a flow rate of 40 µL/min. Different charge states present in the resulting spectra were deconvoluted using Agilent BioConfirm software. Percentages of compound-bound gatekeeper protein were calculated using the integrated peak areas of substrate-bound protein peaks divided by the sum of the substrate-free and substrate-bound peak areas. Integrated peak areas were calculated using Agilent BioConfirm software.

4.3.3. Sample Preparation for Metabolite Studies via LC-MS

Turnover studies were conducted using intact BM3 gatekeeper enzyme, provided by Dr. Laura Jeffreys. Intact protein samples were subjected to successive rounds of Lipidex column purification. This was performed in order to ensure total removal of fatty acid compounds from the active site of the enzyme. Intact protein samples were prepared to a concentration of 2 µM in 100 mM ammonium acetate buffer. Reaction volumes were 500 µL. Substrates were present in absolute DMSO at a final concentration of 100 µM,

representing 1% of the total reaction volume (v/v). Samples were turned over in the presence of 500 µM NADPH. An NADPH regeneration system was utilised, and consisted of the following components and final respective concentrations: $NADP^+$ (0.6 mM), glucose-6-phosphate (7.76 mM), glucose-6-phosphate dehydrogenase (1.5 units). All samples were placed into a shaking incubator set at 37 °C and were turned over for a period of 3 hours. Reactions were stopped by the addition of 500 µL of cold acetonitrile. Samples were vortexed for 20 seconds and precipitated protein was pelleted by centrifugation (13 K rpm, 10 minutes). Samples were filtered into mass spectrometry vials using an Impact protein precipitation plate (Phenomenex, Macclesfield, UK).

4.3.4. Product Analysis via LC-MS and LC-MS/MS

Samples were injected onto a C18 HPLC column mounted to an Agilent 1290 Infinity II LC system, prior to being run on an Agilent MS 6545 QTOF instrument (Agilent Technologies, UK). Samples were ionised using an Agilent Dual AJS ESI ion source. As before, the Q-TOF instrument was run in positive mode and used a gradient of 0.1% formic acid present in water to 0.1% formic acid present in acetonitrile. Collison energy was 30 V. Targeted masses were 399.076 and 415.075, corresponding to non-metabolised and mono-hydroxylated azulfidine, respectively. The presence of hydroxylated azulfidine product was confirmed using Agilent MassHunter Qualitative Analysis software. Identity of the hydroxylated azulfidine product was postulated using LC-MS/MS fragmentation and an Agilent Metabolome LC-MS/MS compound library.

4.4. Results

4.4.1. Optimisation of Sample Preparation and Instrument Parameters

Sample preparation and instrument parameters both required optimisation before compound screening could proceed. Obtaining spectra with a low signal-to-noise ratio was a priority. High signal-to-noise ratios within compound-binding spectra prevent accurate detection of protein-ligand complexes, and significantly hinder the detection of low-affinity compounds. Selection of an appropriate buffer and protein concentration were therefore essential steps in sample preparation. Buffers typically used within protein purification, namely those containing inorganic salts (potassium phosphate, sodium chloride etc.) or non-volatile compounds (Tris, Hepes etc), are unsuitable for use in native-MS experiments. Non-volatile components cause prolific formation of non-specific adducts during ionisation. These non-specific adducts can reduce mass measuring accuracy and lower signal resolution – consequently, analytes of interest may be supressed (Xia, et al., 2019). The volatile buffer ammonium acetate, at a concentration of 100 mM and a pH of ~7.1, was

therefore selected for use. Buffer exchange was performed using gel filtration as outlined in the methods. A protein concentration of 18.5 µM, corresponding to 1 mg/ml of gatekeeper BM3 heme domain, was selected for use. Trials of lower protein concentrations of 5 and 10 µM, respectively, yielded spectra with unacceptably high signal-to-noise ratios (data not shown). Moreover, as concentration of the drug compound was fixed at 50 µM, higher protein concentrations would not offer an acceptable ratio of protein to ligand. In addition, on account of the high-throughput goal of the assay development, use of a relatively low protein concentration is desirable.

Instrument parameters (notably voltages, pressures and flow rates) underwent a process of iterative optimisation. This was to ensure noncovalent interactions between the protein and ligands were preserved for detection by the mass spectrometer. The predominant thermal unfolding events of the gatekeeper heme domain in the substrate-free and substrate-bound states occur at ~58 °C and ~60 °C, respectively (Butler, et al. 2013). Consequently, the sheath gas temperature was kept constant at 50 °C, the lowest temperature possible for successful ionisation using ESI. Instrument parameters are presented in full within the methods section.

4.4.2. Native-MS of Substrate-Free Gatekeeper Heme Domain

Prior to running the compound assay, substrate-free BM3 "gatekeeper" heme domain was subjected to ionisation under native conditions. Native-MS spectra for the BM3 gatekeeper heme domain in the substrate-free state is presented in Figure 4.1.

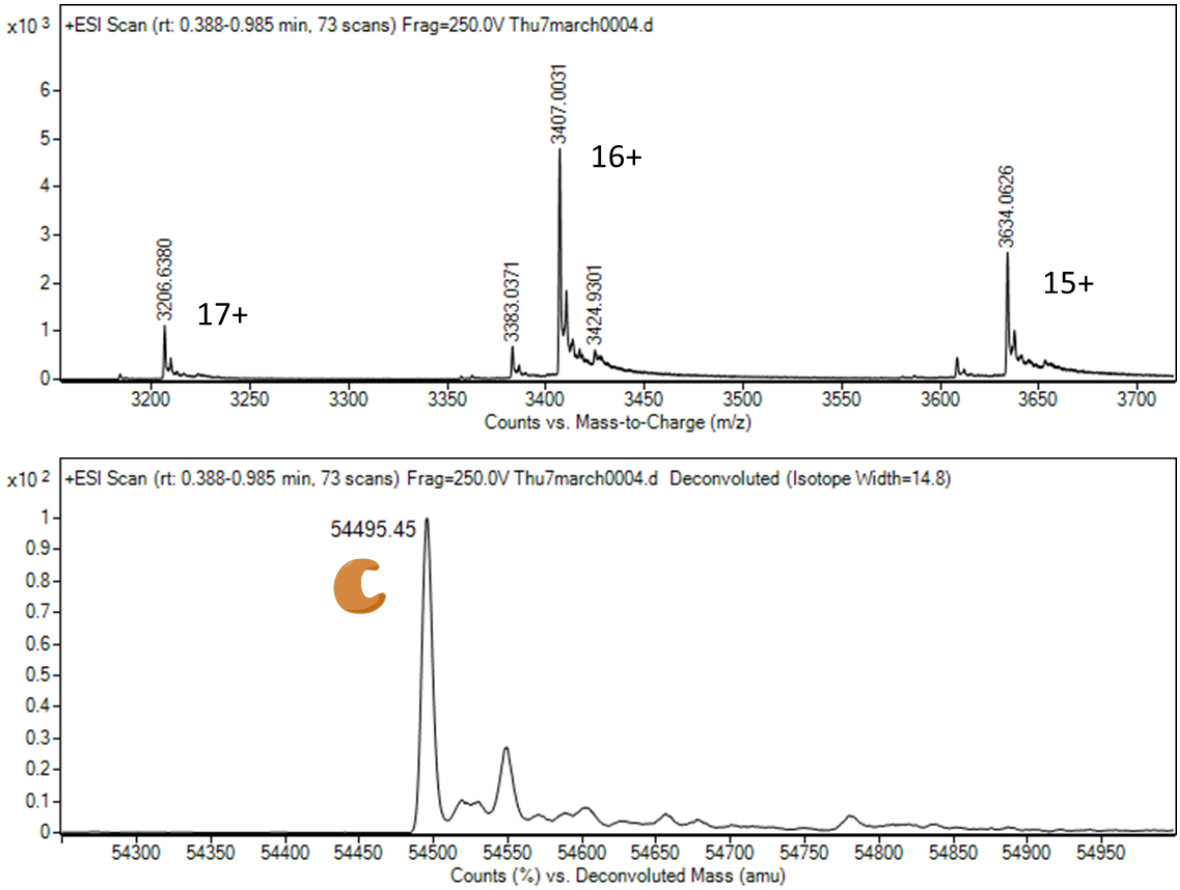

Figure 4.1 Native-MS spectra of the BM3 gatekeeper heme domain in the substrate-free state.

Protein was prepared to a concentration of 1 mg/ml in 100 mM ammonium acetate buffer (pH ~7.1). Protein was ionised via ESI in non-denaturing (native) conditions. Initial MS data featuring distinct protein charge states are presented in the top panel. Deconvoluted MS data are presented within the bottom panel. The mass of the substrate-free BM3 gatekeeper heme domain was determined to be 54495.45 Da, consistent with the theoretical mass of 54496.67 Da.

Three distinct charge states (17+/16+/15+) were observed for the substrate-free gatekeeper heme domain. This relatively small distribution of charge states indicates that the protein is close to its native conformation. The multiple charge states presented within the top insert of Figure 4.1 were deconvoluted into single spectra, displayed in the bottom insert of Figure 4.1, enabling the mass of the protein to be determined. Two species of gatekeeper protein appeared to be present in the samples used. The predominant species was confirmed to have a mass of 54495.45 Da. The second species, present in negligible amounts, had a mass of 54112.01 Da. Following deconvolution, only the predominant species (54495.45 Da) was visible within the spectrum. The mass of 54495.45 is consistent with the average mass calculated from the known amino acid sequence (54496.67 Da,

calculated from PDB: 4KF2 using Agilent MassHunter Sequence Manager. Full sequence information is presented in Figure 4.S1 within the supplementary information). These data indicated that gatekeeper heme domain remained stable following the conditions required for ionisation and was a suitable candidate for compound screening. In addition, these data confirmed that salt removal via buffer exchange had been successful, indicated by the presence of narrow peaks suitable for accurate mass determination within the deconvoluted spectrum. The theoretical isoelectric point (pI) of the gatekeeper heme domain was calculated to be 5.56 (Calculated using ExPASy and sequence data from PDB: 4KF2).

4.4.3. Initial Compound Screen

Following the establishment of the exact mass of the substrate-free heme domain, the compound assay began. The initial screening plate contained the top 48 compounds from the library that induced the greatest percentage high-spin shift against the gatekeeper heme domain. These known binders were selected in order to test the efficacy of the method, and ensure that the non-covalent interactions between the protein and ligands had been preserved for detection by mass spectrometry. The names of the compounds within the initial screening plate are presented within Table 4.S1 within the supplementary information.

Of the 48 compounds examined, 40 displayed a binding interaction with the gatekeeper heme domain. A successful interaction was confirmed if a peak corresponding to the mass of the protein, plus the mass of at least one substrate compound, was evident within the deconvoluted spectrum. Within each spectrum, the mass of substrate-free protein remained consistent at 54495 ± 1 Da.

The binding spectra of three library substrates, representing a variety of molecular weights and conformations, have been selected for presentation in order to highlight the effectiveness of the screening technique.

Figure 4.2 illustrates the formation of a 1:1 protein-substrate complex between the gatekeeper heme domain and the vasodilator Cyclandelate. Three distinct charged states (14+/13+/12+) were visible within the mass spectrum (Figure 4.2, top inset). Substrate-free gatekeeper heme domain is indicated by the peak at 54495.44 Da within the deconvoluted spectrum (Figure 4.2, lower inset). Cyclandelate has a molecular weight of 276.37 Da. Therefore, a peak present at 54771.81 Da would correspond to a 1:1 protein+ Cyclandelate complex. However, a broad peak occurs at 54773.34, a discrepancy of 1.53 Da –

attributable to the width of the peak. The peak at 54773.34 Da therefore corresponds to a protein+ Cyclandelate complex. According to peak areas corresponding to substrate-free and substrate-bound protein, following incubation with 50 µM of the compound, 37.4% of gatekeeper heme domain was present in the Cyclandelate-bound state following ionisation. For comparison, titration of 150 µM Cyclandelate prompted a high-spin shift of 80.1% against 3 µM of gatekeeper heme domain (supplementary data, Table 4.S1).

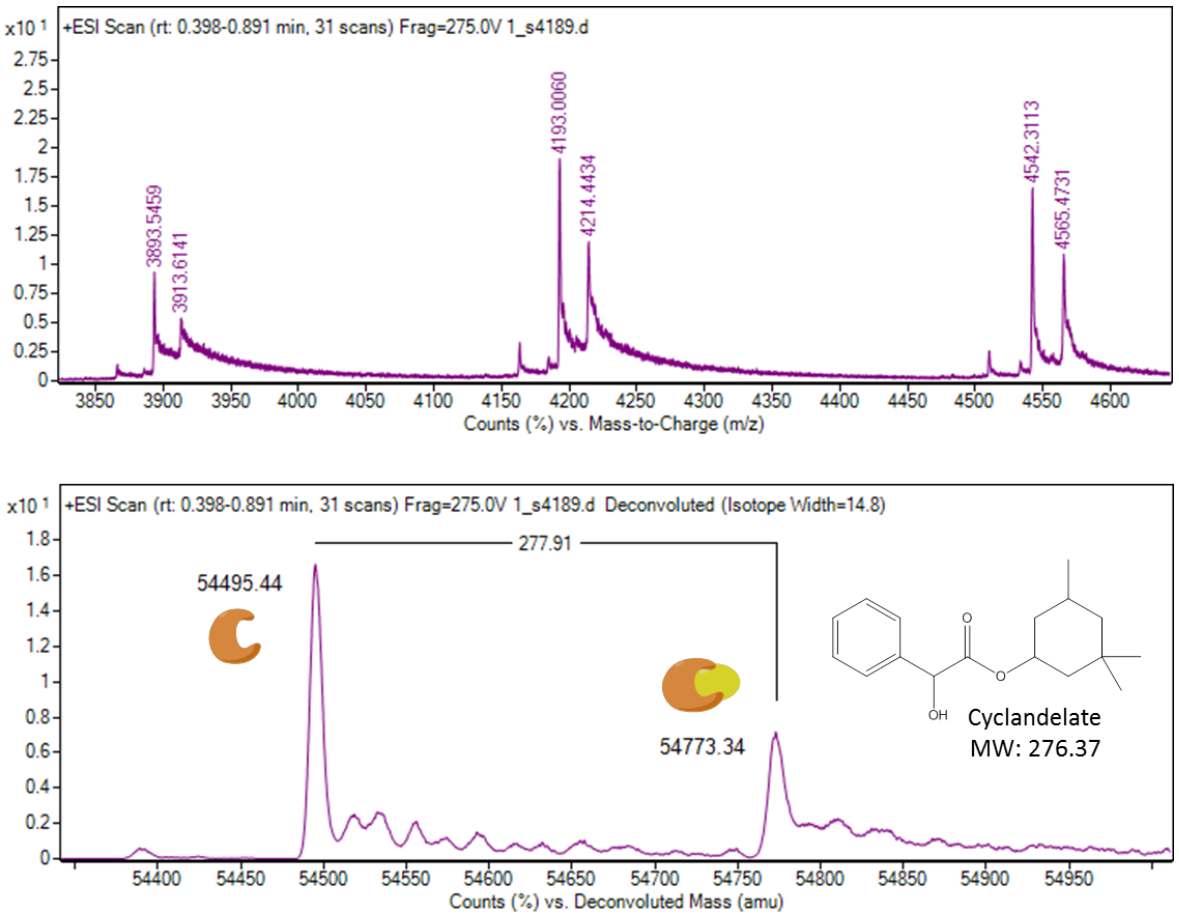

Figure 4.2 **Native-MS spectra of the BM3 gatekeeper heme domain in the substrate-free and Cyclandelate-bound state**

Protein was present at a concentration of 18.5 µM (1 mg/ml) within volatile ammonium acetate buffer (100 mM, pH ~7.1). Drug compound Cyclandelate was prepared in absolute DMSO. Final Cyclandate concentration was 50 µM within the sample, comprising 0.5% of the final sample volume. The sample was ionised for MS analysis via ESI. The native MS spectrum, displaying protein in the substrate-free and Cyclandelate-bound conformation is present in the top panel. Presence of protein in both the Cyclandelate-free and -bound states was confirmed via deconvolution of the multiple charge states (bottom panel). A peak broadly corresponding to the mass of protein plus the mass of a single Cyclandelate molecule is evident at 54773.34 Da – confirming the presence of a 1:1 protein-ligand complex.

Against the anti-obesity drug Rimonabant (molecular weight: 463.76 Da), 56.7% of the gatekeeper heme domain was present in the substrate-bound state at a 1:1 stoichiometry (Figure 4.3). Four charge states (14+/13+/12+/11+) were visible within the mass spectra prior to deconvolution. Following deconvolution, the mass of the substrate-free protein was 54495.51 Da. A well resolved, narrow peak corresponding to a protein (54495 ± 1 Da) +

Rimonabant (463.68 Da) complex is evident at 54959.19 Da. Binding titration data showed 150 µM of Rimonabant generated a 74.4% high-spin shift against 3 µM of gatekeeper heme domain (supplementary data, Table 4.S1).

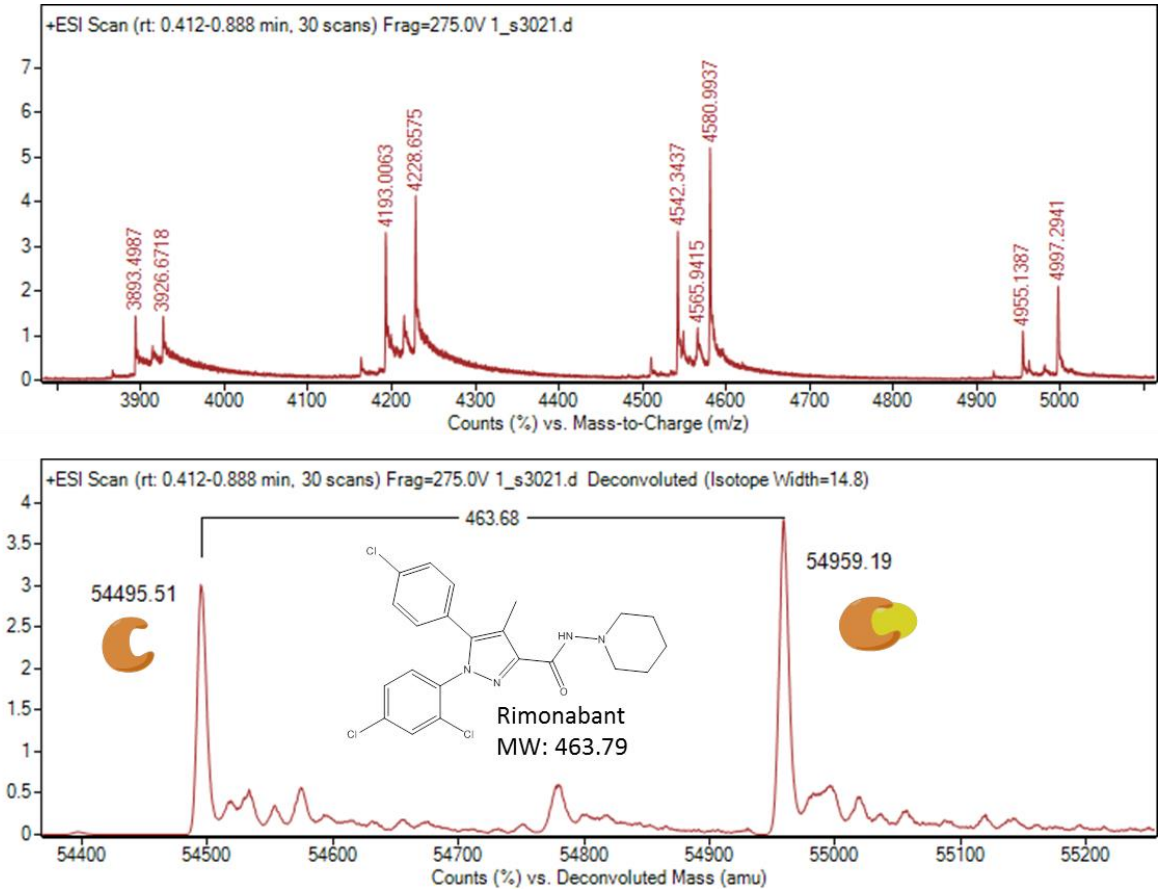

Figure 4.3 Native-MS spectra of the BM3 gatekeeper heme domain in the Rimonabant-bound and substrate-free states.

Samples were ionised using ESI. Protein was prepared to a concentration of 1 mg/ml within volatile buffer (100 mM ammonium acetate, pH ~7.1). The drug compound Rimonabant was present within the sample at a final concentration of 50 µM. Four charge states were evident within the initial mass spectrum (top panel). Deconvolution of these charge states enabled accurate mass measurement of the substrate-free protein (54495.51 Da) and the 1:1 protein+Rimonabant complex (54959.19 Da) (bottom panel).

Five distinct charge states (15+/14+/13+/12+/11+) were evident within the protein + Rosiglitazone spectrum (Figure 4.4). An antidiabetic drug, Rosiglitazone has a molecular weight of 357.45 Da. Once again, within the deconvoluted spectrum, substrate-free gatekeeper heme domain has a recorded mass of within the range of 54495 ± 1 Da (54495.39 Da). A peak present at 54851.43 Da therefore corresponds to a 1:1 protein + Rosiglitazone complex (with a ~1.39 Da discrepancy once again attributable to the peak

width of the protein + Rosiglitazone complex). Binding titration studies (supplementary data, Table 4.S1) demonstrated that 150 µM of Rosiglitazone was capable of converting 55.7% of 3 µM of gatekeeper heme domain to the high-spin state. Binding data collected via Native-MS and presented in Figure 4.4 indicate that 27.4% of the gatekeeper protein detected forms a 1:1 non-covalent complex with 50 µM of Rosiglitazone.

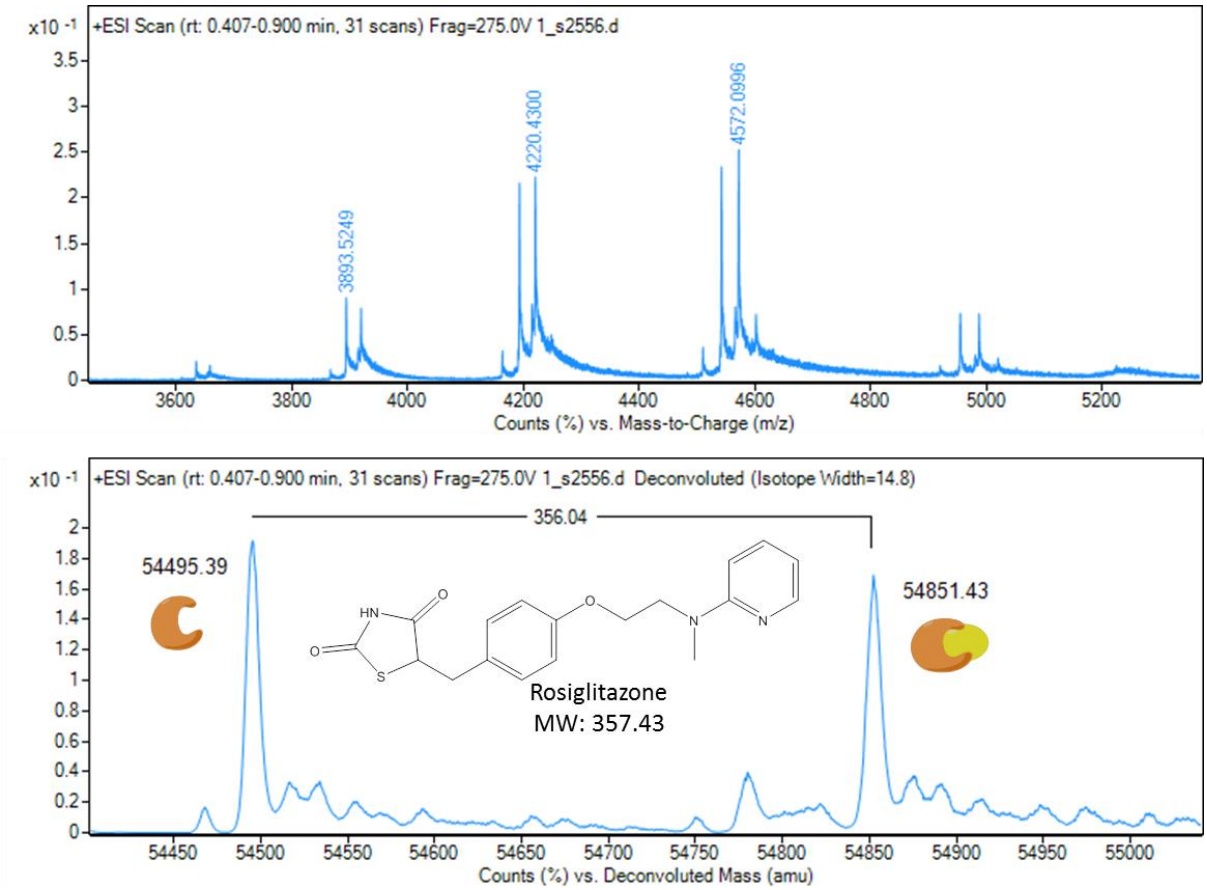

Figure 4.4 **Native-MS spectra of the BM3 gatekeeper variant heme domain in the substrate-free and Rosiglitazone-bound states.**
Gatekeeper heme domain was prepared in volatile ammonium acetate buffer (100 mM, pH ~7.1) for electrospray ionisation (ESI). The drug compound Rosiglitazone was taken from a drug-library stock at a concentration of 10 mM and prepared within absolute DMSO. The final concentration of Rosiglitazone within the protein+ligand sample was 50 µM. Following ionisation, multiple charge states, corresponding to ligand-free and Rosiglitazone-bound protein, were detected (top panel). Deconvolution of these charge states enabled accurate mass measurement of the substrate-free (54495.39 Da) and Rosiglitazone-bound (54851.43 Da) species (bottom panel). The presence of a single protein-Rosaglitazone peak within the deconvoluted spectra suggested a protein:ligand binding stoichiometry of 1:1.

Of the 48 compounds examined in the initial screen, non-specific or multiple ligand binding to the protein was displayed by 4 compounds: Bezafibrate, Omeprazole, Betamethanasone valerate and Pimobendan. Structures of compounds for which multiple or non-specific interactions were measured via native-MS are presented in Figure 4.5.

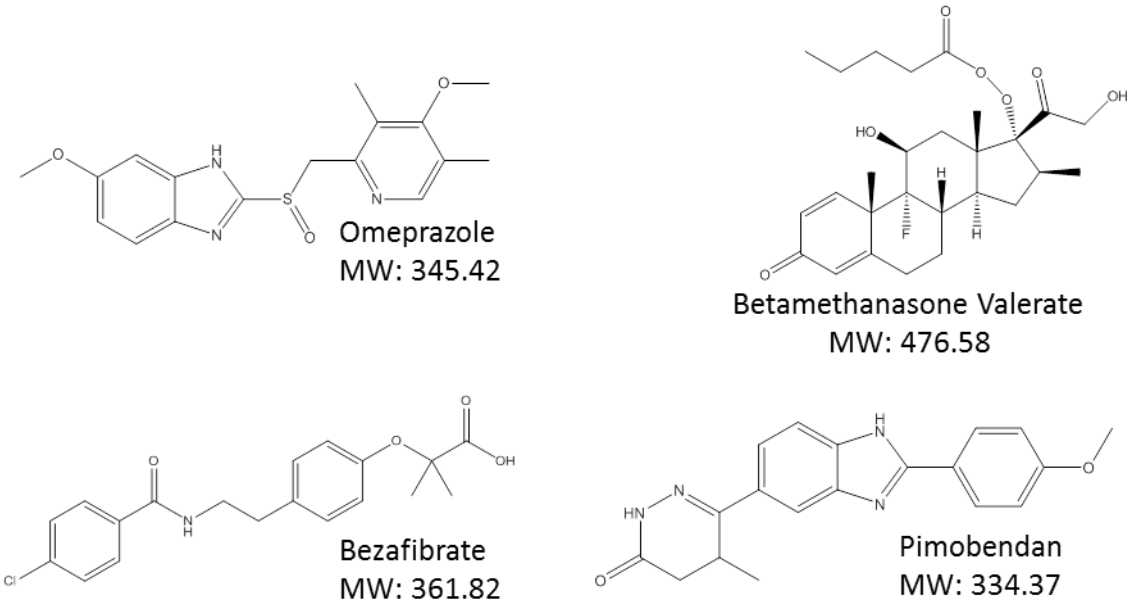

Figure 4.5 Structures of drug compounds included within the initial 48 compound screen for which multiple or non-specific binding was evident within deconvoluted native-MS spectra
Molecular weights are given in Daltons (Da).

Figure 4.6 displays an example of non-specific ligand binding between the protein and the cholesterol-lowering drug Bezafibrate (molecular weight 361.82 Da). Five charge states are apparent within the non-deconvoluted spectra (top insert), spanning charges of +15 to +11. Following deconvolution (bottom insert), three narrow peaks are evident. The initial peak at 54495.36 Da corresponds to substrate-free protein. The second and most intense peak occurs at 54857.17 Da, and corresponds to the mass of one molecule of Bezafibrate bound to the protein at a 1:1 stoichometric ratio. The third most intense peak, indicating the presence of a compound with a mass of 55218.91 Da, corresponds to the mass of the gatekeeper heme domain plus two molecules of Bezafibrate. This mass could correspond to one of two potential protein+ligand complexes. Specifically, this peak represents either a protein+ligand complex in which Bezafibrate is bound at a stoichiometric ratio of 2:1 or is indicative of non-specific binding of Bezafibrate to the protein. Titration of Bezafibrate against the gatekeeper heme domain produced a high-spin shift of 87.0%. Of the

gatekeeper protein complexes observed by Native-MS, 73.3% of the protein was bound, in some manner, to Bezafibrate.

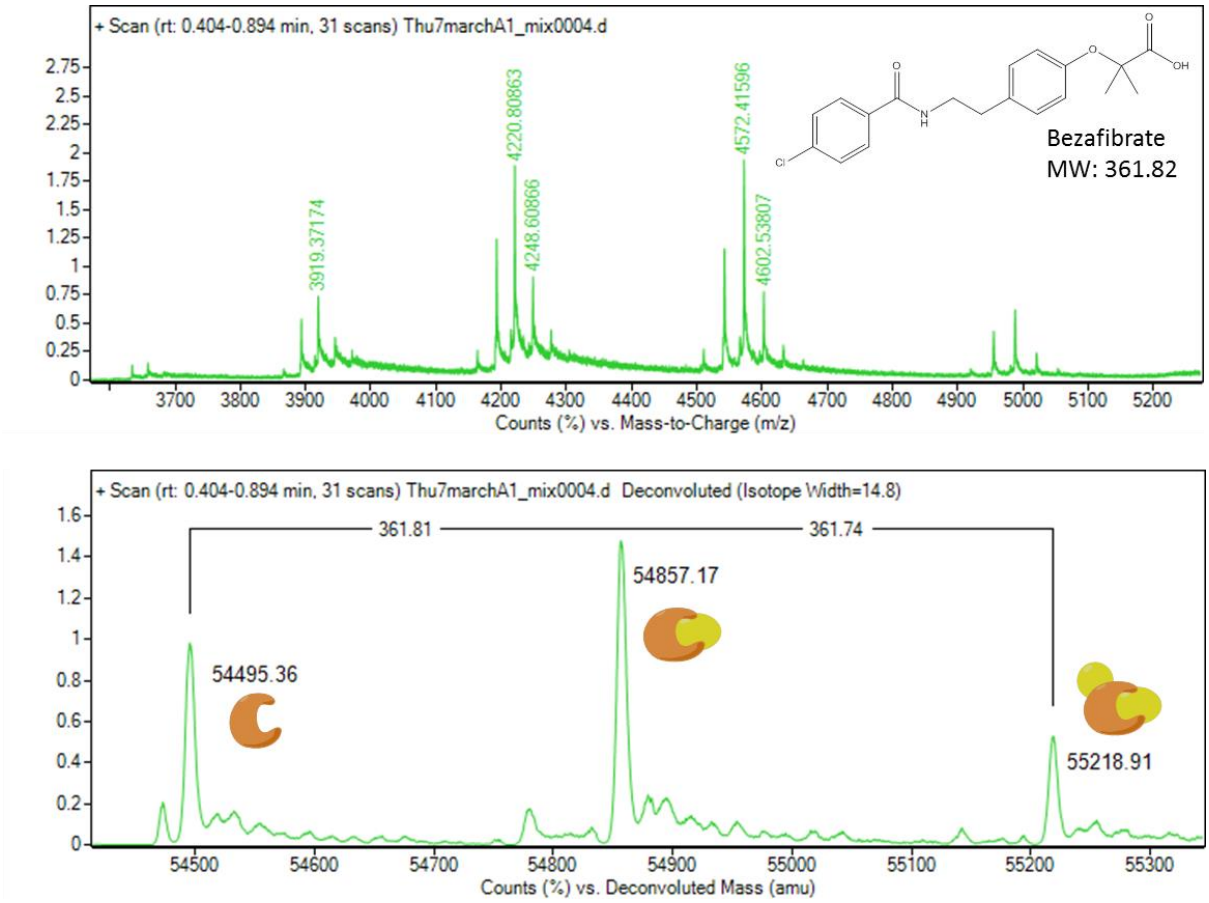

Figure 4.6 Native-MS spectra of the BM3 gatekeeper variant heme domain in the substrate-free and Bezafibrate-bound states.
Protein was prepared within 100 mM ammonium acetate buffer (100 mM, pH ~7.1). Bezafibrate (molecular weight: 361.92 Da) was present within absolute DMSO. Final concentrations of protein and Bezafibrate were 18.5 µM and 50 µM, respectively. Samples were ionised using ESI. Following ionisation, five charge states were detected via mass spectrometry (top panel). Upon deconvolution, gatekeeper protein appeared to be present in three states: substrate free (54495.36 Da), bound to one molecule of Bezafibrate (54857.17 Da) and bound to two molecules of Bezafibrate (55218.91 Da) (bottom panel).

Within the remaining spectra that exhibited non-specific or multiple ligand binding to the protein (not shown), peaks corresponding to complexes containing three or more ligand molecules were not evident. In all cases of multiple binding, peaks corresponding to complexes containing the protein and two molecules of ligand were present.

Of the 8 compounds that did not interact with the protein, 2 (Cholecalciferol and Nystatin) were not visible as intact ligands within the mass spectra. The remaining 6 compounds were visible within the corresponding mass spectra, but did not display peaks corresponding to a binding interaction. The formation of a protein+ligand complex either did not occur prior to ionisation, or did not remain intact following ionisation. There was no clear mass, structural or UV-Vis high-spin shift correlation between the 6 compounds for which no binding activity was observed by native-MS. An example of native-MS spectra which does not display a binding interaction between the gatekeeper protein and a compound from the dug library, in this case Pimozide, is presented in Figure 4.7.

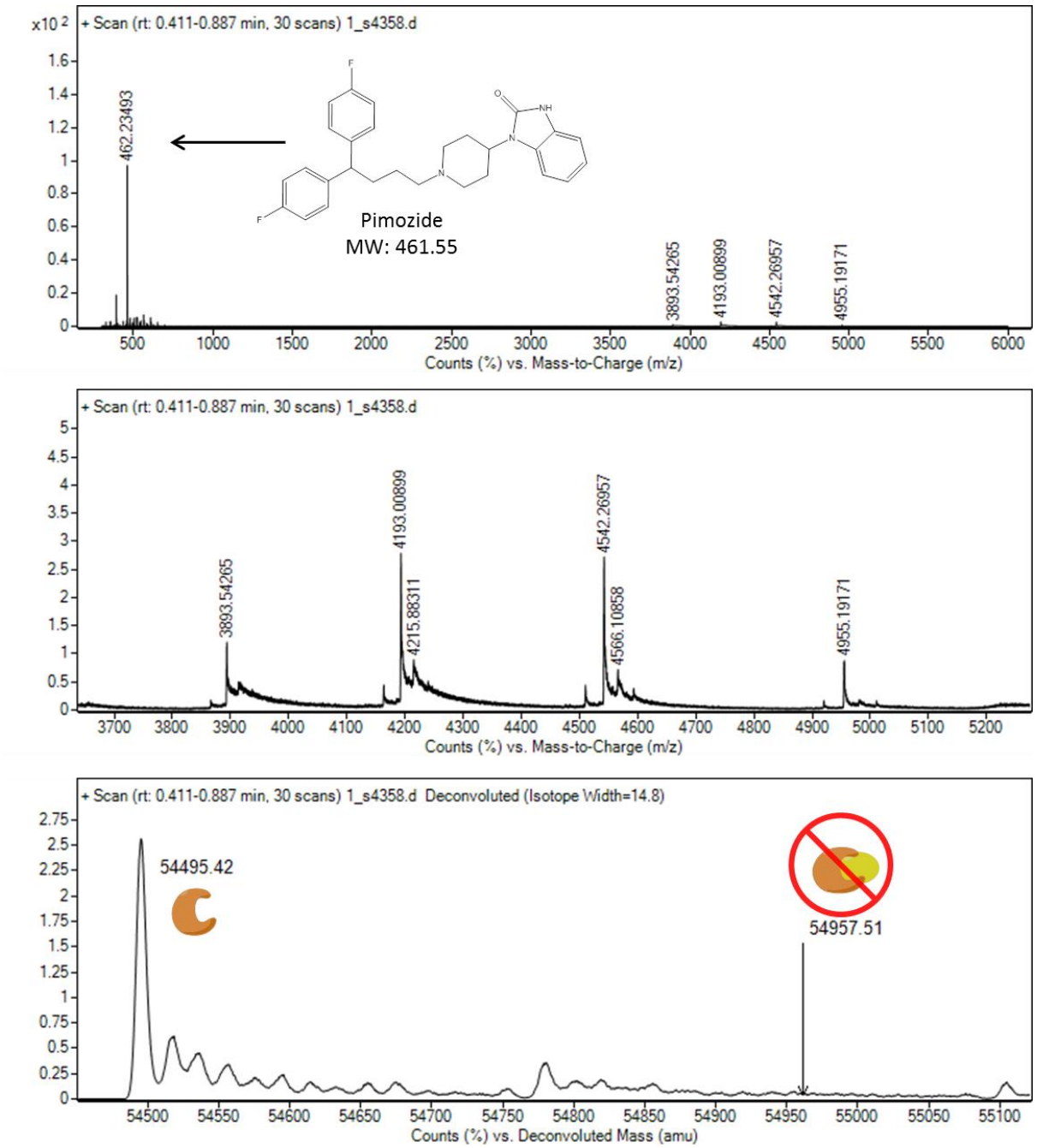

Figure 4.7 **Native-MS spectra of BM3 gatekeeper variant heme domain in the presence of drug compound Pimozide.**

Gatekeeper heme domain protein was prepared within 100 mM volatile ammonium acetate buffer (pH ~7.1) to a final concentration of 1 mg/ml. Drug compound Pimozide was present within absolute DMSO to a final concentration of 50 µM within the sample. Following ionisation via ESI and detection via mass spectrometry, no protein+ Pimozide binding interactions were visible. Recorded masses of substrate-free protein and Pimozide were 54495.42 Da and 462.23 Da, respectively.

Four charge states (14+/13+/12+/11+) were evident within the non-deconvoluted mass spectrum (top panel). The drug compound Pimozide (theoretical mass 461.55 Da, recorded mass 462.23 Da) and the substrate-free protein (recorded mass 54495.42 Da) are both evident within the mass spectra displayed in the top and middle panels of Figure 4.6. However, no 1:1 protein+ligand complex, estimated to occur at 54957.51 Da, was observed within the deconvoluted mass spectrum (bottom panel). Prior titration studies of Pimozide against the gatekeeper heme domain displayed a high-spin shift of 39.9% - the lowest shift of all compounds tested in the initial 48 compound assay (supplementary information, Table 4.S1).

For each compound, the high-spin shift recorded by binding titrations was plotted against the percentage of compound-bound protein measured using native-MS (variable set 1). In addition, the molecular weight of each compound was plotted against the corresponding percentage of compound-bound protein measured using the deconvoluted native-MS spectra (variable set 2). The final linear relationship examined, that between the compound partition coefficient (Log *P*) and the percentage of compound-bound protein measured using native-MS, was designated variable set 3. Correlations between each variable set are presented as figures within the supplementary information (Figures 4.S2, 4.S3 and 4.S4). Pearson product moment correlation coefficients (PPMCC) were calculated in order to assess the relationships between each set of variables. PPMCC data are presented within Table 4.1.

Table 4.1 **Pearson product moment correlation coefficients expressing the statistical relationship between two sets of variables.**
Set 1: %HS Shift calculated by UV-Vis binding titrations against % compound-bound protein recorded using native-MS. Set 2: Drug compound molecular weight against % compound-bound protein recorded using native-MS. Set 3: Drug compound Log *P* values against % compound-bound protein recorded using native-MS.

Variables	Pearson's r	N	Significance
Set 1	0.471	48	<0.001
Set 2	-0.134	48	0.366
Set 3	0.0190	48	0.899

These data indicate a significant moderate positive correlation between the first set of variables. Therefore, the propensity of drug compounds to induce high-spin state shifts

within the gatekeeper is moderately related to the same compounds ability to form a protein+compound complex measurable by native-MS. The PPMCC data relating to the second set of variables suggests a weak negative correlation, albeit one with little significance, between the mass of the drug compound and the ability of the compound to form a protein+compound complex detectable by Native-MS. Little or no linear correlation is apparent between the variables of set 3, indicating that the Log *P* of the drug compound has no impact on the formation of a protein+compound complex detectable by Native-MS.

In total, binding interactions were detected via native-MS for 40 of the 48 compounds within the initial assay. Excluding the 2 compounds for which degradation was apparent, protein+compound complexes were detected for 87.0% of drug compounds within the initial assay, confirming the suitability of the native-MS technique for use in compound screening.

4.4.4. Second Compound Screen

The second screening plate contained 48 "coloured" compounds taken from the drug library. These compounds were selected on account of their intense colours, and not their ability to produce high-spin heme protein as recorded in the prior UV-Vis binding studies. The compounds within this assay were capable of prompting high-spin shifts against the gatekeeper protein within the range of 56.4% to 0.0%. The majority of compounds included in this screen absorbed within the region of 400 – 550 nm, consequently appearing yellow, orange or red. The drug content of the second screening plate is presented in Table 4.S2 within the supplementary information.

Of the 48 compounds tested, 15 displayed a clear binding interaction with the protein. All 15 compounds that displayed a binding interaction with the protein did so at a stoichiometric ratio of 1:1, with no multiple or non-specific binding recorded. Within 24 deconvoluted spectra, protein+ligand complexes were not evident, despite the presence of both intact protein and compound within the mass spectra. Consequently, these 24 compounds were recorded as having no interaction with the gatekeeper heme domain. The mass spectra corresponding to the remaining 9 compounds did not produce conclusive data. In these cases, either intact protein or the compounds in question were not visible within the mass spectra. As a result of this, the presence or absence of a binding interaction could not be accurately determined. Excluding the 9 compounds for which accurate data could not be determined, 38.5% of the compounds within the second assay displayed a binding interaction with the gatekeeper heme domain (supplementary data, Table 4.S2).

Five of the compounds for which binding interactions were detected using native-MS were recorded as not prompting any (0.0%) high-spin shift when titrated against the protein and measured by UV-visible spectroscopy. Structures and molecular weights of these compounds ("0.0% shift" compounds) are presented in Figure 4.8.

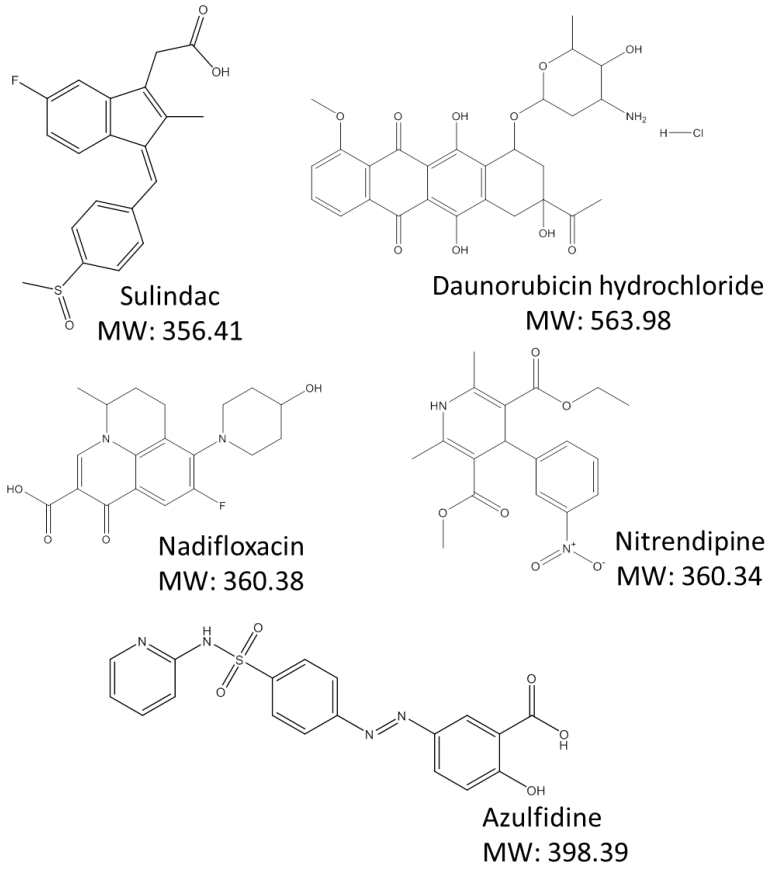

Figure 4.8 Structures of drug compounds that did not induce a high-spin shift in the gatekeeper heme domain when measured using UV-Vis, but displayed protein+substrate complexes when binding was examined using native-MS.
Molecular weights are presented in Daltons (Da).

The native-MS obtained for one of these "0.0% shift" compounds, the antirheumatic drug azulfidine, is presented in Figure 4.9.

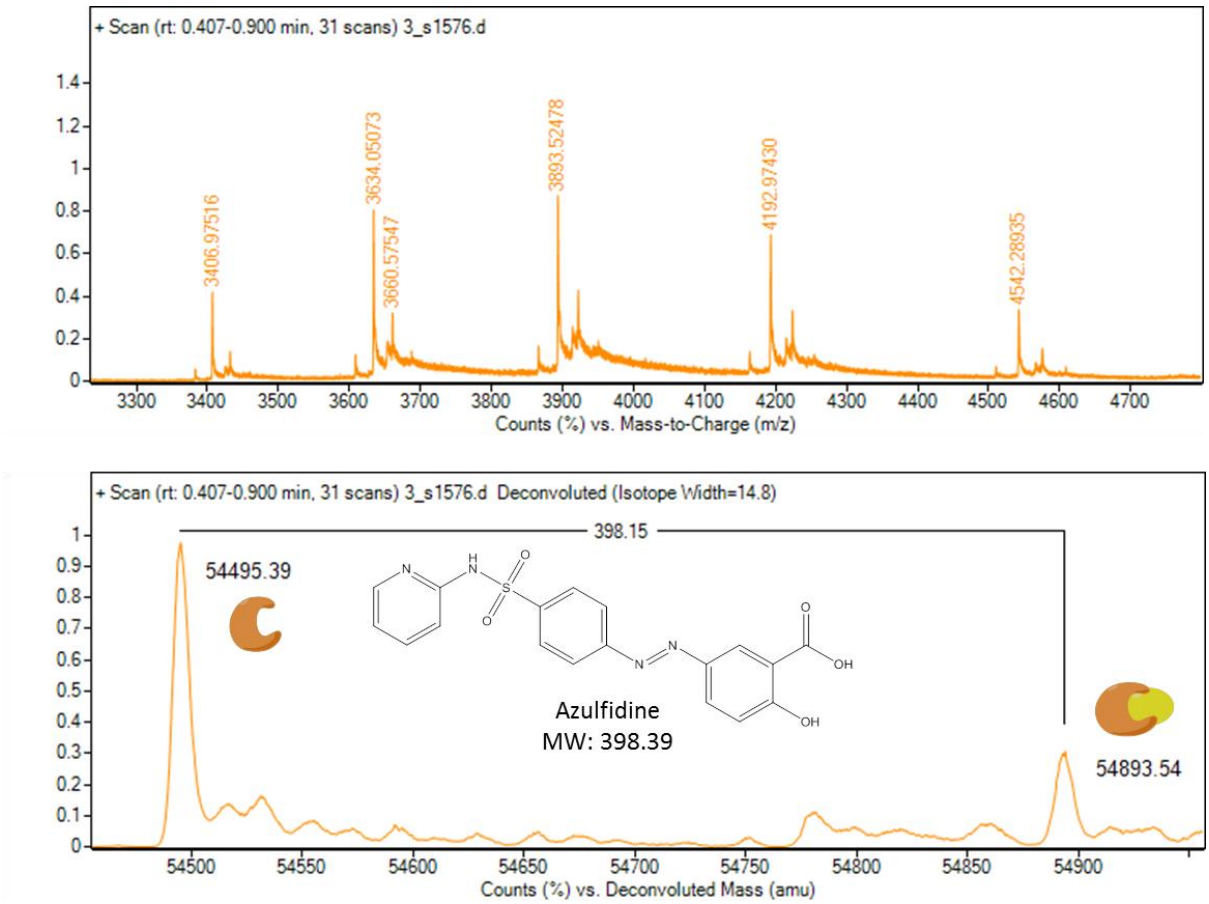

Figure 4.9 **Native-MS spectra of the BM3 gatekeeper variant heme domain in the presence of drug compound Azulfidine**

Protein was prepared to a concentration of 18.5 µM within 100 mM ammonium acetate (pH ~7.1). Azulfidine (molecular weight ~398 Da) was suspended within absolute DMSO and was present within the sample at a final concentration of 50 µM. Sample was ionised using ESI. The top panel displays the number of charge states adopted by the protein and protein+azulfidine complex. These charge states were deconvoluted into a single spectrum (bottom panel). The presence of substrate-free protein was indicated by the peak at 54495.39 Da. The peak highlighted at 54839.54 Da corresponds to a protein+azulfidine complex at a 1:1 stoichiometric ratio.

Within the non-deconvoluted native-MS for the gatekeeper protein and drug compound azulfidine (398.39 Da), five charge states (16+/15+/14+/13+/12+) were detected. Following deconvolution, substrate-free gatekeeper protein was measured to a mass of 54495.39 Da. A 1:1 protein+azulfidine complex, corresponding to the mass of substrate-free protein plus the mass of the compound (54893.54 Da) was detected. Of the total amount of gatekeeper protein detected using native-MS, 26.1% was in complex with azulfidine – a stark contrast

to the data collected via UV-visible binding titrations, in which no high-spin shift was evident.

Azulfidine, in addition to the four other "0.0% shift" compounds, were selected for turnover studies. LC-MS was used to examine the samples for the presence of metabolites generated by the intact gatekeeper variant. The percentage of bound protein detected via native-MS with each of the "0.0% shift" compounds is presented within Table 4.2.

Table 4.2 Drug componds selected for turnover analysis studies using LC-MS.

Library compounds with an absorbance at ~400-450 nm selected for LC-MS metabolite analysis

Drug Name	Molecular weight (Da)	UV-Visible titration % HS shift	% Ligand-bound protein detected in MS
Sulfasalazine (Azulfidine)	398.39	0.0	26.1
Sulindac (Clinoril)	356.41	0.0	12.5
Nitrendipine	360.3699	0.0	17.3
Daunorubicin HCl (Daunomycin HCl)	563.98	0.0	13.4
Nadifloxacin	360.38	0.0	19.3

4.4.5. LC-MS Turnover Studies and Metabolite Analysis

The five compounds presented in Figure 4.8 and Table 4.2 were subjected to metabolite analysis via LC-MS. All samples were prepared in triplicate. Following completion of the sample run, resulting chromatograms and their corresponding mass spectra were examined for evidence of turnover. Of the five compounds examined, substrate turnover was apparent only within samples containing azulfidine (C18H14N4O5S). Gatekeeper mediated turnover of azulfidine was initially inferred by the significant reductions in the peak intensities of chromatograms representing non-metabolised substrate. Mono-hydroxylated azulfidine product was identified within the sample chromatograms by searching for the formula C18H14N4O6S within Agilent MassHunter Qualitative analysis software (Agilent, UK). A compound with a mass of 415.07 m/z was identified using tandem mass spectrometry (MS/MS). The compound generated a Tgt score (a 0-100 score reflecting how well the compound matches the retention time, isotope pattern and mass of the queried formula) of 98.55 against the formula of mono-hydroxylated azulfidine

product. Formulas of predominant fragments and their respective masses of the compound are presented within Table 4.3. Analysis of the mass fragmentation pattern (Figure 4.10) enabled the identity of the mono-hydroxylated azulfidine compound generated by the intact gatekeeper enzyme to be suggested as ((E)-2-hydroxy-5-((3-hydroxy-4-(N-(pyridin-2-yl)sulfamoyl)phenyl)diazenyl)benzoic acid. Unmetabolised azulfidine compound has a mass of 398.39 Da. The molecular ion peak at 415.07 m/z suggests the addition of a hydroxyly group to the azulfidine substrate. The base peak at 144.99 m/z, in addition to fragments occurring at 315.06, 224.03 and 158.05 m/z, suggests that this hydroxylation is situated on the central of the three 6-membered rings of the azulfidine compound. Finally, the fragment present at 118.01 m/z, in tandem with the base peak (144.99 m/z) suggests that azulfidine oxidation has occurred at the third carbon within this ring, carbon atom number 14.

Table 4.3 Formulas and masses of predominant fragment peaks present within the mass spectrum of mono-hydroxylated azulfidine product

Fragment formula	Fragment mass (m/z)
$C_{18}H_{14}N_4O_6S$	415.07
$C_{17}H_{12}N_4O_3S_2^{\bullet}$	352.06
$C_{14}H_{11}N_4O_3S_3^{\bullet}$	315.06
$C_9H_8N_2O_3S_2^{\bullet}$	224.03
$C_8H_7N_2O_2S_3^{\bullet}$	195.02
$C_9H_6N_2O_4^{\bullet}$	158.05
$C_4H_4NO_3S_2^{\bullet}$	145.99
$C_7H_5N_27^{\bullet}$	117.05

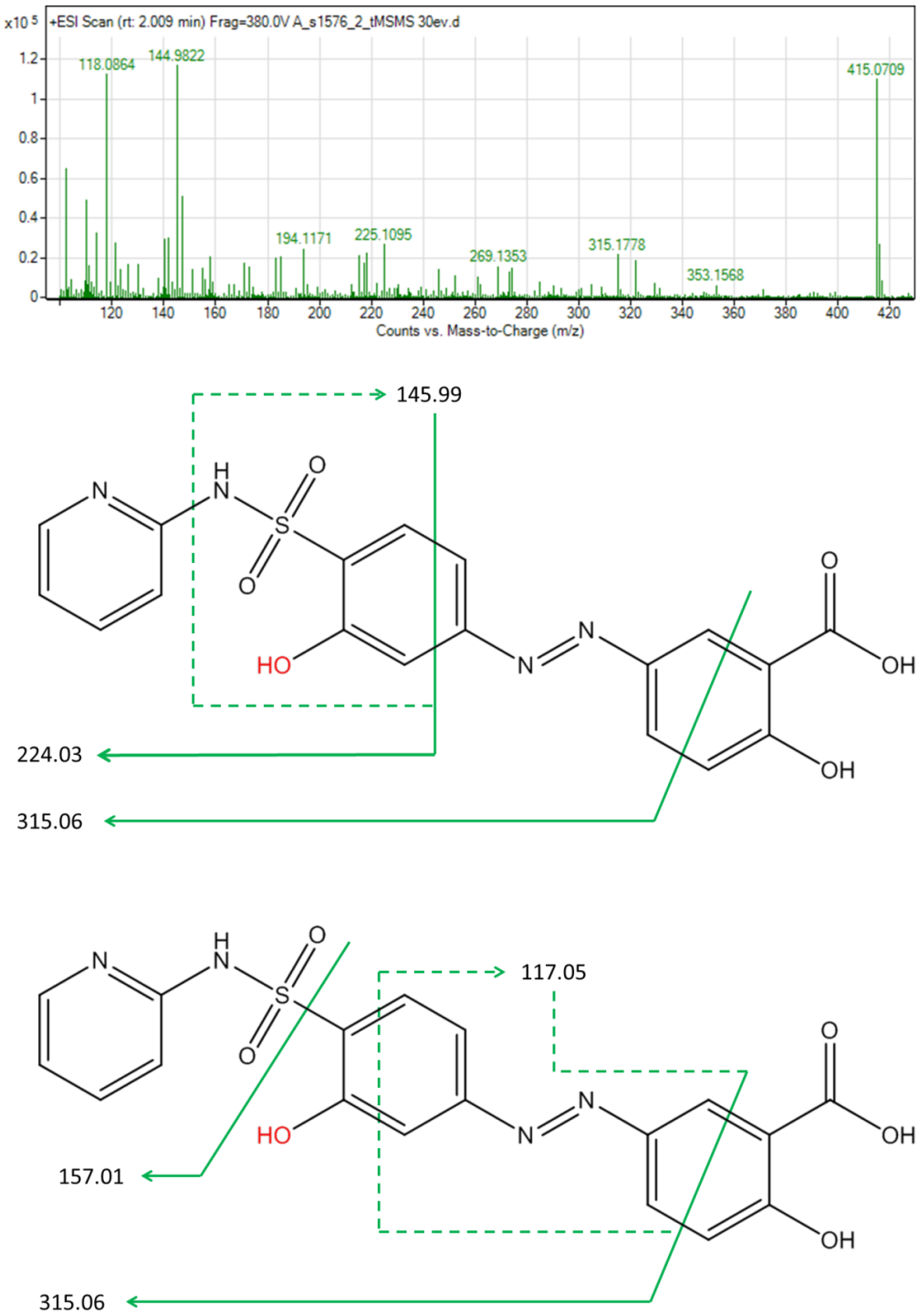

Figure 4.10 Mass spectra (top panel) and fragmentation patterns (mid and bottom panels) of mono-hydroxylated azulfidine product generated by the intact Gatekeeper BM3 variant.

Fragmentation patterns were assessed using Agilent Masshunter Qualtitative analysis and ChemDraw Professional software. Site of enzyme-mediated hydroxylation is highlighted in red. Total mass of intact product was recorded to be 415.08 m/z. Identity of product (displayed within middle and bottom panels) is postulated to be ((E)-2-hydroxy-5-((3-hydroxy-4-(N-(pyridin-2-yl)sulfamoyl)phenyl)diazenyl)benzoic acid.

4.5. Discussion

To our knowledge, this is the first time native-MS has been applied as a drug-screening technique against the P450 BM3 protein. The mass of the BM3 "gatekeeper" variant heme domain could be measured with a high degree of accuracy following ionisation via ESI. Within all deconvoluted spectra, the mass of the heme domain in the substrate-free state was consistently recorded to be 54495 ± 1 Da, a difference of only ~1.22 Da from the theoretical mass (54496.67 Da, calculated using PDB: 4KF2). This slight mass deviation can be attributed to instrument resolution and does not correspond to either solvation or adduct formation, both of which produce more dramatic deviations from the expected compound masses (Tu et al., 2005). These data confirm that the gatekeeper heme domain remains intact following ionisation and that the ESI parameters have been successfully optimised for use with the protein.

The degree to which protein structures in the gas phase accurately reflect those found in solution is perhaps the most debated aspect of native-MS (Hall and Robinson, 2012).However, there is a substantial body of work that supports the notion that the native conformations of proteins can be retained in the gas phase (Breuker and McLafferty 2008; Benesch et al., 2007). In addition to the conformation of the protein, the buffer used, instrument parameters and surface tension of the ESI droplet all contribute to the resulting charge state of the protein ions (Susa et al., 2017; Ivarone and Williams, 2003; Thomson, 1997). High charge states do not always imply unfolded protein; the distribution of charge states provides a more accurate insight into protein conformation (Hall and Robinson, 2012; Ashcroft; 2005). The narrow distribution of charge states, in addition to the occurrence of ion peaks at the higher end of the m/z region (~3200 – 3700 m/z) within the mass spectrum of the substrate-free gatekeeper protein are indicative of a protein folded in its native conformation (Kaltashov and Aabzalimov, 2008). In contrast, unfolded or denatured protein would produce a much broader distribution of charge states, as the multiple conformations adopted by the disordered protein each provide different steric accessibility for protonation (Ashcroft, 2005; Konermann and Douglas, 1997).

Data presented by Susa et al. (2017), confirm that the average protein charge state is proportional to both molecular weight and pI, with average charge increasing in tandem with molecular weight and a more acidic pI. Within ammonium acetate, a 51.7 kDa protein with a pI of 6.0 (analogous to the gatekeeper heme domain, pI = 5.56; ~54.5 kDa) displayed an average charge state of 15+ (Susa et al., 2017). Within the substrate-free mass spectra

for the gatekeeper heme domain, charge states ranged from 17+ to 15+, with 16+ being the most predominant – in good accordance with Susa et al. (2017).

Within substrate-bound mass spectra, charge states corresponding to protien+ligand complexes ranged from 15+ to 11+, with the most predominant charge state varying depending on the compound. The number of charge states, a good indicator of protein+ligand complex stability, also varied between compounds. The general reduction in the average charge state of the protein+ligand complexes relative to the average charge of the substrate-free protein indicates a reduction in protonation sites within the substrate-bound protein (Kaltashov and Aabzalimov, 2008). This, in turn, suggests conformational changes, such as reduced flexibility, occur when the gatekeeper protein is in the substrate-bound state (Kaltashov and Aabzalimov, 2008).

The Gatekeeper heme domain has previously been shown to undergo a major thermal unfolding event (T_m) at ~58 °C, with the binding of substrate (n-palmitoyl glycine) prompting a concurrent increase in thermostabillity by ~2 °C (Butler, et al., 2013). Ionisation by ESI was performed at 50 °C – a temperature which demonstrably preserved the native conformation of the gatekeeper heme domain.

Compound stocks from the drug library were prepared to a concentration of 10 mM within absolute DMSO. Use of DMSO as a solvent is ubiquitous throughout commercial drug libraries due to its highly favourable physicochemical properties (Göth et al., 2017; Cubrilovic and Zenobi, 2013). However, the presence of DMSO at 2% (v/v) within native-MS samples has been reported to cause protein denaturation following ionisation, in addition to increasing protein charge states (Lu et al., 2015; Sterling et al., 2011). Moreover, modest v/v proportions of DMSO have been found to disrupt secondary structural motifs and coordinate to the heme iron of wild-type and variant BM3 heme domains (Roccatano et al., 2005; Clodfelter et al., 2006; Whitehouse et al., 2012). Consequently, even small amounts of DMSO can impact protein-ligand binding affinities (Göth et al., 2017; Cubrilovic and Zenobi, 2013). As a result of this, concentrations of DMSO were fixed at 50 µM, corresponding to 0.5% sample v/v, throughout all assays.

Within the initial screening assay, protein+ligand interactions were visible for 40 of the 48 drug compounds. Substrate-free, folded gatekeeper protein was observed within every spectrum, regardless of the detection/proportion of protein in the substrate-bound state. There was no discernible trend for the 8 compounds for which no binding interaction was

recorded by native-MS. High-spin shifts, in the range of 75.1% - 39.9% were recorded using UV-visible spectroscopy for all 8 compounds for which no binding was observed within the native-MS assay. The distribution of compounds throughout the screening plate, in tandem with the detection substrate-free protein within each spectrum, suggests that blockages of the ESI equipment did not occur during the assay. The presence of intact compound within the mass spectra of 6 of these 8 non-binding assays also suggests that these compounds remained intact following ionisation via ESI.

The linear relationship between molecular weight of the compounds within the initial assay and the percentage of compound-bound protein detected by native-MS displays little significant correlation (Table 4.1, variable set 2: r = -0.134; N = 48; sig = 0.366). The 6 compounds for which no-binding was detected, despite evidence of intact compound within the mass spectra, cover a range of masses between 228.07 – 461.55 Da. In comparison, compounds which do form a complex with the gatekeeper protein measureable via native-MS cover a mass range of 252.27 – 713.12 Da. These data suggest that compound mass has no significant impact on native-MS detection of protein-ligand complexes and does therefore not impede the accuracy of the technique for compound screening. Two compounds, Cholecalciferol (384.64 Da) and Nystatin (926.09 Da), were not visible intact within their respective protein+compound mass spectra – suggesting either degradation of the library stocks or fragmentation following ionisation. Consequently, binding information could not be accurately determined for both Cholecalciferol and Nystatin.

Excluding the two compounds for which binding could not be accurately measured, protein+ligand complexes were detected for 87.0% of the drug compounds contained within the initial compound screen. A moderate positive linear correlation was apparent between the sets of binding data collected via UV-Visible binding titrations and binding data measured using native-MS (Table 4.1, variable set 1: r = 0.471; N=48; <0.001). That is to say, the propensity of a compound to induce a high percentage high-spin shift as recorded by UV-vis is correlated to the percentage of protein+ligand complexes measured via native-MS. The lack of a more distinctive positive correlation between the two variables is most likely attributable to the discrepancy in the ratio of protein to compound used within each technique. On account of the previously discussed sample limitations required for effective Native-MS, the protein to compound ratio is 18.5 µM protein: 50 µM compound. In contrast, the binding data collected using UV-Vis spectroscopy was 3 µM

protein: 150 µM compound. Binding constant data, K_d, was not determined for compounds using either technique. However, it could be suggested that the 6 compounds for which no binding was detected via Native-MS all possess K_d values well in excess of 50 µM. Consequently, compounds with relatively low binding affinities (K_d > 50 µM) may be have overlooked within the assay.

The transfer of the protein, or protein+ligand complex, from the liquid to the gas phase may also have had significant impact on the nature of the non-covalent interactions between the protein and the ligand (Pedro and Quinn, 2016). If the protein-ligand interactions are predominantly hydrophobic within solution, the chance of complete protein-ligand complex dissociation within the gas phase has been reported to be high (Pedro and Quinn, 2016; Bich, et al., 2010; Bovet, et al., 2007). In contrast, hydrophobic interactions existing between a protein and ligand have been shown to stabilise following the transition from solution to the gas phase (Yin, et al., 2008).

Despite this, no linear correlation, positive or negative, was evident between the Log *P* values of drug compounds within the initial assay and the percentage of compound-bound protein recorded using native-MS (Table 4.1, variable set 3, r= 0.0190; N= 48; sig= 0.899). In addition, the Log *P* values for the 6 compounds for which no binding interaction was detected via native-MS ranged from 6.3 to 2.4 i.e. all were hydrophobic in character. In addition, it is important to state that the binding pocket of BM3 heme domain is hydrophobic in nature, with the interactions formed between the residues of the active site and the substrate being hydrophobic in nature (Munro, et al., 2002; Ost, et al., 2000). Therefore total ionisation induced dissolution of any potential complexes between these 6 "non-binding" compounds and the gatekeeper heme domain is unlikely.

Within the initial assay, 4 compounds displayed multiple or non-specific binding. As BM3 heme domain possesses only a single ligand binding site, binding of ligands at a stoichiometric ratio of anything other than 1:1 is unlikely (Modi, et al., 1995). Non-specific binding is a common phenomenon within native-MS and typically occurs during ionisation (Kitova, et al., 2012; Göth, et al., 2017). The likelihood of non-specific binding increases with high concentrations of ligand, and is compounded if the ligand binds the target protein with a high affinity. Therefore, a decrease in ligand concentration can minimise the risk of non-specific binding (Smith and Light-Wahl, 1993; Göth et al., 2017). However, as part of a compound screening approach, reduction in compound concentration may result in failure to detect weaker interactions between a target protein and low-affinity ligands –

resulting in potentially costly false-negatives (Kitova et al., 2012). As calculation of the percentage of protein+ligand complexes detected via native-MS took into account all instances of binding, specific or otherwise – the four substrates bound in a non-specific manner registered the top four percentages of compound-bound protein in the assay. Non-specific binding can therefore complicate data analysis, potentially skewing interpretation towards the identification of false-positive compounds. These data highlight the importance of compromise between the ratio of protein-to-compound, their respective concentrations and the final volume of DMSO used within the native-MS screening assays.

The second compound assay consisted of 48 "coloured" compounds, typically those displaying fluorescence maxima within the range of 400 – 550 nm (i.e. those predominantly appearing yellow, orange or red). For example, a compound that appears yellow in colour absorbs light within the visible region of 400-450 nm. Within BM3, transition from the low-to-high spin state occurs between 418 - 390 nm (Whitehouse et al., 2012). Therefore, compounds with intense absorbance within this region may obscure this spin state transition, potentially resulting in either a false negative or an inaccurate determination of binding affinity. Compound screening via native-MS may therefore identify compounds from the library previously overlooked by the screening assays conducted using UV-visible titrations.

Within the second screening assay, 15 of the 48 compounds formed a complex with the gatekeeper heme domain. Of these 15 compounds, 12 displayed a high-spin shift of less than ~10% against the gatekeeper enzyme following binding titrations recorded using UV-vis spectroscopy. Most interestingly, 5 of these compounds did not display any (0.0%) high-spin shift with the gatekeeper protein. In contrast, these compounds generated protein+ligand complexes (1:1 stoichiometric ratio) at a prevalence of 12.5 – 26.1% when measured using native-MS. Of these compounds, 4 were yellow in colour, while 1 (Daunorubicin HCl) appeared bright orange. According to data from the UV-Vis screen, Soret bands were either not visible or partially obscured due to high absorbance at wavelengths shorter than 400 nm within the corresponding spectra of each compound. These data appear to validate the claim that native-MS can identify binding interactions which cannot be accurately measured using the more traditional method of UV-visible spectroscopy.

As with UV-visible spectroscopy, indication of a binding interaction in native-MS does not imply substrate turnover. Consequently, each of the 5 compounds were incubated with the

intact gatekeeper enzyme, with the resulting samples being examined for metabolites using LC-MS. Of the compounds examined, turnover by the intact gatekeeper enzyme was observed for azulfidine only. Azulfidine, the trade name of sulfasalazine, is a commercially available anti-inflammatory, antiheumatic and anti-ulcer prodrug predominantly used for treating rheumatoid arthritis and inflammatory bowel disorders (Cheng et al., 2011; Blake et al., 2004; Pullar, 1989). Appearing an intense yellow in colour, azulfidine generates an absorbance maximum at 360 nm (Cheng et al., 2011). *In vivo*, azulfidine is activated through reduction and successive cleavage of the azo-bond to produce the active component 5-aminosalicylic acid and additional metabolite sulfapyridine (Jung et al., 2000). Within humans, azulfidine metabolism is performed by intestinal bacteria and is not meditated by P450 enzymes (Peppercorn, 1984).

LC-MS data and subsequent studies using tandem mass spectra (MS/MS) indicated that intact gatekeeper protein generated a single, mono-hydroxylated product against the azulfidine substrate. Based on the MS/MS fragmentation pattern, the identity of this compound was postulated to be ((E)-2-hydroxy-5-((3-hydroxy-4-(N-(pyridin-2-yl)sulfamoyl)phenyl)diazenyl)benzoic acid. While LC-MS and MS/MS data can establish the presence of gatekeeper generated, mono-hydroxylated product with confidence, further metabolite studies using NMR are essential in order to establish the identity of the product beyond doubt.

To conclude, this study has demonstrated that native-MS is an effective method for the high-throughput screening of drug compounds against the highly promiscuous “gatekeeper” BM3 variant. Moreover, an important drug compound, one previously overlooked by an alternative screening method, has been identified as a substrate for the gatekeeper enzyme. Now that native-MS conditions have been successfully optimised for P450 BM3, the technique is open for application to a variety of future studies. In addition to further compound screening, future work will focus on applying the technique towards obtaining substrate binding constants in a similarly high-throughput manner.

4.6. References

Acevedo-Rocha, C.G., Gamble, C.G., Lonsdale, R., Li, A., Nett, N., Hoebenreich, S., Lingnau, J.B., Wirtz, C., Fares, C., Hinrichs, H. and Deege, A., 2018. 'P450-catalyzed regio- and diastereoselective steroid hydroxylation: efficient directed evolution enabled by mutability landscaping'. *ACS Catalysis*, 8(4), pp.3395-410.

Ashcroft, A. E. (2005). 'Recent developments in electrospray ionisation mass spectrometry: noncovalently bound protein complexes', *Nat Prod Rep,* 22(4), pp. 452-64.

Benesch, J. L., Ruotolo, B. T., Simmons, D. A. & Robinson, C. V. (2007). 'Protein complexes in the gas phase: technology for structural genomics and proteomics', *Chem Rev,* 107(8), pp. 3544-67.

Bernhardt, R. (2006). 'Cytochromes P450 as versatile biocatalysts', *J Biotechnol,* 124(1), pp. 128-45.

Bich, C., Baer, S., Jecklin, M. C. & Zenobi, R. (2010). 'Probing the hydrophobic effect of noncovalent complexes by mass spectrometry', *J Am Soc Mass Spectrom,* 21(2), pp. 286-9.

Blake, A. J., Lin, X., Schröder, M., Wilson, C. & Yuan, R. X. (2004). 'The imide tautomer of sulfasalazine', *Acta Crystallogr C,* 60(Pt 4), pp. o226-8.

Bovet, C., Wortmann, A., Eiler, S., Granger, F., Ruff, M., Gerrits, B., Moras, D. & Zenobi, R. (2007). 'Estrogen receptor-ligand complexes measured by chip-based nanoelectrospray mass spectrometry: an approach for the screening of endocrine disruptors', *Protein Sci,*16(5), pp. 938-46.

Breuker, K. & McLafferty, F. W. (2008). 'Stepwise evolution of protein native structure with electrospray into the gas phase, 10(-12) to 10(2) s', *Proc Natl Acad Sci U S A,* 105(47), pp. 18145-52.

Butler, C. F., Peet, C., Mason, A. E., Voice, M. W., Leys, D. & Munro, A. W. (2013). 'Key mutations alter the cytochrome P450 BM3 conformational landscape and remove inherent substrate bias', *J Biol Chem,* 288(35), pp. 25387-99.

Butler, C. F., Peet, C., McLean, K. J., Baynham, M. T., Blankley, R. T., Fisher, K., Rigby, S. E., Leys, D., Voice, M. W. & Munro, A. W. (2014). 'Human P450-like oxidation of diverse proton pump inhibitor drugs by 'gatekeeper' mutants of flavocytochrome P450 BM3', *Biochem J,* 460(2), pp. 247-59.

Caldwell, G. W., Yan, Z., Tang, W., Dasgupta, M. & Hasting, B. (2009). 'ADME optimization and toxicity assessment in early- and late-phase drug discovery', *Curr Top Med Chem,* 9(11), pp. 965-80.

Cheng, S.H., Liao, W.N., Chen, L.M. and Lee, C.H., 2011. pH-controllable release using functionalized mesoporous silica nanoparticles as an oral drug delivery system. *J Mater Chem*, *21*(20), pp.7130-37.

Clodfelter, K. H., Waxman, D. J. & Vajda, S. (2006). 'Computational solvent mapping reveals the importance of local conformational changes for broad substrate specificity in mammalian cytochromes P450', *Biochemistry,* 45(31), pp. 9393-407.

Cubrilovic, D. & Zenobi, R. (2013). 'Influence of dimethyl sulfoxide on protein-ligand binding affinities', *Anal Chem,* 85(5), pp. 2724-30.

Ortiz De Montellano, P.R. ed., 2005. *Cytochrome P450: structure, mechanism, and biochemistry*. Springer Science & Business Media.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. (1989). 'Electrospray ionization for mass spectrometry of large biomolecules', *Science,* 246(4926), pp. 64-71.

Ganem, B., Li, Y.T. and Henion, J.D., 1991. Detection of noncovalent receptor-ligand complexes by mass spectrometry. *J Am Chem Soc*, 113(16), pp.6294-96.

Göth, M., Badock, V., Weiske, J., Pagel, K. & Kuropka, B. (2017). 'Critical Evaluation of Native Electrospray Ionization Mass Spectrometry for Fragment-Based Screening', *ChemMedChem,* 12(15), pp. 1201-11.

Haines, D. C., Tomchick, D. R., Machius, M. & Peterson, J. A. (2001). 'Pivotal role of water in the mechanism of P450BM-3', *Biochemistry,* 40(45), pp. 13456-65.

Hall, Z. & Robinson, C. V. (2012). 'Do charge state signatures guarantee protein conformations?', *J Am Soc Mass Spectrom,* 23(7), pp. 1161-8.

Heck, A. J. (2008). 'Native mass spectrometry: a bridge between interactomics and structural biology', *Nat Methods,* 5(11), pp. 927-33.

Hofstadler, S. A. & Sannes-Lowery, K. A. (2006). 'Applications of ESI-MS in drug discovery: interrogation of noncovalent complexes', *Nat Rev Drug Discov,* 5(7), pp. 585-95.

Huang, W. C., Westlake, A. C., Maréchal, J. D., Joyce, M. G., Moody, P. C. & Roberts, G. C. (2007). 'Filling a hole in cytochrome P450 BM3 improves substrate binding and catalytic efficiency', *J Mol Biol,* 373(3), pp. 633-51.

Iavarone, A. T. & Williams, E. R. (2003). 'Mechanism of charging and supercharging molecules in electrospray ionization', *J Am Chem Soc,* 125(8), pp. 2319-27.

Jeffreys, L. N., Poddar, H., Golovanova, M., Levy, C. W., Girvan, H. M., McLean, K. J., Voice, M. W., Leys, D. & Munro, A. W. (2019). 'Novel insights into P450 BM3 interactions with FDA-approved antifungal azole drugs', *Sci Rep,* 9(1):1577.

Jung, Y. J., Lee, J. S. & Kim, Y. M. (2000). 'Synthesis and in vitro/in vivo evaluation of 5-aminosalicyl-glycine as a colon-specific prodrug of 5-aminosalicylic acid', *J Pharm Sci,* 89(5), pp. 594-602.

Kaltashov, I. A. & Abzalimov, R. R. (2008). 'Do ionic charges in ESI MS provide useful information on macromolecular structure?', *J Am Soc Mass Spectrom,* 19(9), pp. 1239-46.

Kitova, E. N., El-Hawiet, A., Schnier, P. D. & Klassen, J. S. (2012). 'Reliable determinations of protein-ligand interactions by direct ESI-MS measurements. Are we there yet?', *J Am Soc Mass Spectrom,* 23(3), pp. 431-41.

Konermann, L. & Douglas, D. J. (1997). 'Acid-induced unfolding of cytochrome *c* at different methanol concentrations: electrospray ionization mass spectrometry specifically monitors changes in the tertiary structure', *Biochemistry,* 36(40), pp. 12296-302.

Loo, J. A. (1997). 'Studying noncovalent protein complexes by electrospray ionization mass spectrometry', *Mass Spectrom Rev,* 16(1), pp. 1-23.

Lu, J., Trnka, M. J., Roh, S. H., Robinson, P. J., Shiau, C., Fujimori, D. G., Chiu, W., Burlingame, A. L. & Guan, S. (2015). 'Improved Peak Detection and Deconvolution of Native Electrospray Mass Spectra from Large Protein Complexes', *J Am Soc Mass Spectrom,* 26(12), pp. 2141-51.

Maple, H. J., Garlish, R. A., Rigau-Roca, L., Porter, J., Whitcombe, I., Prosser, C. E., Kennedy, J., Henry, A. J., Taylor, R. J., Crump, M. P. & Crosby, J. (2012). 'Automated protein-ligand interaction screening by mass spectrometry', *J Med Chem,* 55(2), pp. 837-51.

McLean, K. J., Hans, M., Meijrink, B., van Scheppingen, W. B., Vollebregt, A., Tee, K. L., van der Laan, J. M., Leys, D., Munro, A. W. & van den Berg, M. A. (2015). 'Single-step fermentative production of the cholesterol-lowering drug pravastatin via reprogramming of *Penicillium chrysogenum*', *Proc Natl Acad Sci U S A,* 112(9), pp. 2847-52.

Modi, S., Primrose, W. U., Boyle, J. M., Gibson, C. F., Lian, L. Y. & Roberts, G. C. (1995). 'NMR studies of substrate binding to cytochrome P450 BM3: comparisons to cytochrome P450 cam', *Biochemistry,* 34(28), pp. 8982-8.

Munro, A. W., Daff, S., Coggins, J. R., Lindsay, J. G. & Chapman, S. K. (1996). 'Probing electron transfer in flavocytochrome P-450 BM3 and its component domains', *Eur J Biochem,* 239(2), pp. 403-9.

Munro, A. W., Girvan, H. M. & McLean, K. J. (2007). 'Variations on a (t)heme - novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily', *Nat Prod Rep,* 24(3), pp. 585-609.

Munro, A. W., Leys, D. G., McLean, K. J., Marshall, K. R., Ost, T. W., Daff, S., Miles, C. S., Chapman, S. K., Lysek, D. A., Moser, C. C., Page, C. C. & Dutton, P. L. (2002). 'P450 BM3: the very model of a modern flavocytochrome', *Trends Biochem Sci,* 27(5), pp. 250-7.

Narhi, L. O. & Fulco, A. J. (1987). 'Identification and characterization of two functional domains in cytochrome P-450BM-3, a catalytically self-sufficient monooxygenase induced by barbiturates in *Bacillus megaterium*', *J Biol Chem,* 262(14), pp. 6683-90.

Noble, M. A., Miles, C. S., Chapman, S. K., Lysek, D. A., MacKay, A. C., Reid, G. A., Hanzlik, R. P. & Munro, A. W. (1999). 'Roles of key active-site residues in flavocytochrome P450 BM3', *Biochem J,* 339 (2), pp. 371-9.

O'Reilly, E., Köhler, V., Flitsch, S. L. & Turner, N. J. (2011). 'Cytochromes P450 as useful biocatalysts: addressing the limitations', *Chem Commun (Camb),* 47(9), pp. 2490-501.

Oliver, C. F., Modi, S., Sutcliffe, M. J., Primrose, W. U., Lian, L. Y. & Roberts, G. C. (1997). 'A single mutation in cytochrome P450 BM3 changes substrate orientation in a catalytic intermediate and the regiospecificity of hydroxylation', *Biochemistry,* 36(7), pp. 1567-72.

Ost, T. W., Miles, C. S., Murdoch, J., Cheung, Y., Reid, G. A., Chapman, S. K. & Munro, A. W. (2000). 'Rational re-design of the substrate binding site of flavocytochrome P450 BM3', *FEBS Lett,* 486(2), pp. 173-7.

Ozbal, C. C., LaMarr, W. A., Linton, J. R., Green, D. F., Katz, A., Morrison, T. B. & Brenan, C. J. (2004). 'High throughput screening via mass spectrometry: a case study using acetylcholinesterase', *Assay Drug Dev Technol,* 2(4), pp. 373-81.

Pedro, L. & Quinn, R. J. (2016). 'Native Mass Spectrometry in Fragment-Based Drug Discovery', *Molecules,* 21(8). pii:E984.

Peppercorn, M. A. (1984). 'Sulfasalazine. Pharmacology, clinical use, toxicity, and related new drug development', *Ann Intern Med,* 101(3), pp. 377-86.

Pullar, T. (1989). 'Sulphasalazine and related drugs in rheumatoid arthritis', *Pharmacol Ther,* 42(3), pp. 459-68.

Ravichandran, K. G., Boddupalli, S. S., Hasemann, C. A., Peterson, J. A. & Deisenhofer, J. (1993). 'Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's', *Science,* 261(5122), pp. 731-6.

Roccatano, D., Wong, T. S., Schwaneberg, U. & Zacharias, M. (2005). 'Structural and dynamic properties of cytochrome P450 BM-3 in pure water and in a dimethylsulfoxide/water mixture', *Biopolymers,* 78(5), pp. 259-67.

Sanglier, S., Atmanene, C., Chevreux, G. & Dorsselaer, A. V. (2008). 'Nondenaturing mass spectrometry to study noncovalent protein/protein and protein/ligand complexes: technical aspects and application to the determination of binding stoichiometries', *Methods Mol Biol,* 484, pp. 217-43.

Schaus, S. E., Brandes, B. D., Larrow, J. F., Tokunaga, M., Hansen, K. B., Gould, A. E., Furrow, M. E. & Jacobsen, E. N. (2002). 'Highly selective hydrolytic kinetic resolution of terminal epoxides catalyzed by chiral (salen)Co(III) complexes. Practical synthesis of enantioenriched terminal epoxides and 1,2-diols', *J Am Chem Soc,* 124(7), pp. 1307-15.

Smith, R.D. and Light-Wahl, K.J. (1993). The observation of non-covalent interactions in solution by electrospray ionization mass spectrometry: promise, pitfalls and prognosis. *Biological Mass Spectrometry*, *22*(9), pp.493-501.

Sowden, R. J., Yasmin, S., Rees, N. H., Bell, S. G. & Wong, L. L. (2005). 'Biotransformation of the sesquiterpene (+)-valencene by cytochrome P450cam and P450BM-3', *Org Biomol Chem,* 3(1), pp. 57-64.

Sterling, H. J., Prell, J. S., Cassou, C. A. & Williams, E. R. (2011). 'Protein conformation and supercharging with DMSO from aqueous solution', *J Am Soc Mass Spectrom,* 22(7), pp. 1178-86.

Susa, A. C., Xia, Z., Tang, H. Y. H., Tainer, J. A. & Williams, E. R. (2017). 'Charging of Proteins in Native Mass Spectrometry', *J Am Soc Mass Spectrom,* 28(2), pp. 332-340.

Thomson, B.A., 1997. Declustering and fragmentation of protein ions from an electrospray ion source. *J Am Soc Mass Spectrom*, 8(10), pp.1053-58.

Tu, Y. P., He, L., Fitch, W. & Lam, M. (2005). 'Solvation in electrospray mass spectrometry: effects on the reaction kinetics of fragmentation mediated by ion-neutral complexes', *J Org Chem,* 70(13), pp. 5111-8.

Urlacher, V. B., Lutz-Wahl, S. & Schmid, R. D. (2004). 'Microbial P450 enzymes in biotechnology', *Appl Microbiol Biotechnol,* 64(3), pp. 317-25.

van Vugt-Lussenburg, B. M., Stjernschantz, E., Lastdrager, J., Oostenbrink, C., Vermeulen, N. P. & Commandeur, J. N. (2007). 'Identification of critical residues in novel drug metabolizing mutants of cytochrome P450 BM3 using random mutagenesis', *J Med Chem,*50(3), pp. 455-61.

Vivat Hannah, V., Atmanene, C., Zeyer, D., Van Dorsselaer, A. & Sanglier-Cianférani, S. (2010). 'Native MS: an 'ESI' way to support structure- and fragment-based drug discovery', *Future Med Chem,* 2(1), pp. 35-50.

Whitehouse, C. J., Bell, S. G. & Wong, L. L. (2012). 'P450(BM3) (CYP102A1): connecting the dots', *Chem Soc Rev,* 41(3), pp. 1218-60.

Xia, Z., DeGrandchamp, J. B. & Williams, E. R. (2019). 'Native mass spectrometry beyond ammonium acetate: effects of nonvolatile salts on protein stability and structure', *Analyst,* 144(8), pp. 2565-73.

Yin, S., Xie, Y. & Loo, J. A. (2008). 'Mass spectrometry of protein-ligand complexes: enhanced gas-phase stability of ribonuclease-nucleotide complexes', *J Am Soc Mass Spectrom,* 19(8), pp. 1199-208.

Zanger, U. M. & Schwab, M. (2013). 'Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation', *Pharmacol Ther,*138(1), pp. 103-41.

4.7. Supplementary Information

Table 4.S1 Content of initial drug compound screening plate.

The 48 compounds from the FDA approved library that generated the greatest high-spin shift against the gatekeeper protein, as determined by UV-vis spectroscopy, were selected for the initial screen by native-MS. [1]UV-Visible binding titrations were performed by Dr. Laura Jeffreys as detailed in Jeffreys et al. (unpublished data). [2]Native-MS sample information, parameters and method are presented within the methods section.

Initial screening assay: Top-binding 48 compounds identified via UV-Visible spectroscopy					
Plate position	**Drug Name**	**MW**	**Drug Target**	**UV–Visible titration % HS shift[1]**	**% Ligand-bound protein detected in MS[2]**
A1	Bezafibrate	361.82	Others	87.03%	73.28%
A2	Cephalomannine	831.9	Others	80.78%	23.45%
A3	Omeprazole (Prilosec)	345.42	Proton Pump, ATPase	80.50%	68.79%
A4	Cyclandelate	276.37	Others	80.12%	37.37%
A5	Rimonabant (SR141716)	463.79	Cannabinoid Receptor	76.40%	56.67%
A6	Norethindrone	298.42	Others	76.25%	52.21
A7	Amorolfine hydrochloride	353.97	Others	75.79%	35.02%
A8	Esomeprazole magnesium (Nexium)	713.12	5-alpha Reductase	75.50%	36.49%
A9	Adapalene	412.52	Others	75.05%	0.00%
A10	Pranlukast	481.5	Others	71.26%	35.73%
A11	Domperidone (Motilium)	425.91	Dopamine Receptor	70.89%	37.79%
A12	Betamethasone valerate (Betnovate)	476.58	Others	69.09%	84.80%
C1	Carmofur	257.26	Antimetabolites	68.01%	25.84%
C2	Esomeprazole sodium (Nexium)	367.4	ATPase	62.64%	30.96%
C3	Chenodeoxycholic acid	392.57	NULL	61.59%	33.67%
C4	Artemether (SM-224)	298.37	Others	61.17%	20.02%

C5	Doxazosin mesylate	547.58	Adrenergic Receptor	60.88%	16.91%
C6	Bexarotene	348	Others	57.29%	0.00%
C7	Lomustine (CeeNU)	233.7	Others	56.61%	0.00%
C8	Chlorquinaldol	228.07	Others	56.54%	0.00%
C9	Nystatin (Mycostatin)	926.09	Others	56.43%	**N/A**
C10	Rosiglitazone (Avandia)	357.43	PPAR	55.71%	27.45%
C11	Isotretinoin	300.44	Hydroxylase	55.43%	64.62%
C12	Rosiglitazone maleate	473.5	PPAR	54.76%	27.88%
E1	DAPT (GSI-IX)	432.46	Beta Amyloid	54.30%	34.16%
E2	Rosiglitazone HCl	393.89	PPAR	53.56%	27.06%
E3	Levonorgestrel (Levonelle)	312.45	Others	53.45%	34.22%
E4	Pioglitazone hydrochloride (Actos)	392.9	Others	52.55%	28.88%
E5	Lacidipine (Lacipil, Motens)	455.54	Calcium Channel	51.86%	11.25%
E6	Idoxuridine	354.1	Others	50.85%	12.88%
E7	Vemurafenib (PLX4032)	489.92	Raf	49.85%	19.25%
E8	Pioglitazone (Actos)	356.44	Others	49.05%	9.94%
E9	Pimobendan (Vetmedin)	334.37	PDE	48.26%	77.97%
E10	Phenytoin (Lepitoin)	252.27	Sodium Channel	47.92%	29.89%
E11	Lithocholic acid	376.57	Others	47.19%	34.48%
E12	Gestodene	310.43	Estrogen/progestogen Receptor	47.10%	29.20%
G1	Manidipine (Manyper)	610.7	Calcium Channel	46.78%	8.77%
G2	Repaglinide	452.59	Potassium Channel	46.56%	23.71%
G3	Drospirenone	366.49	Estrogen/progestogen Receptor	45.84%	25.11%
G4	Terbinafine hydrochloride (Lamisil)	327.89	Others	45.83%	0.00%

G5	Paclitaxel (Taxol)	853.91	Microtubule Associated	45.50%	17.14%
G6	Trilostane	329.43	Dehydrogenase	43.07%	30.03%
G7	AMG-073 HCl (Cinacalcet hydrochloride)	393.87	CaSR	42.77%	0.00%
G8	S-(+)-Rolipram	275.34	PDE	42.72%	32.43%
G9	Naftopidil (Flivas)	392.49	Adrenergic Receptor	42.19%	15.64%
G10	Vitamin D3 (Cholecalciferol)	384.64	Others	41.73%	**N/A**
G11	Artemisinin	282.33	Others	40.97%	20.35%
G12	Pimozide	461.55	Others	39.88%	0.00%

Table 4.S2 Content of secondary screening plate

Compounds were selected based on their intense colours (typically red, orange or yellow) regardless of previous shift data collected by binding titrations. [1]UV-Visible binding data were collected by Dr. Laura Jeffreys as presented in Jeffreys et al. (unpublished data). [2]Information regarding sample preparation, methodology and data processing of native-MS samples is outlined within the methods section.

Second screening assay: Coloured library compounds with absorbance between 400 – 550 nm					
Plate position	**Drug Name**	**MW**	**Drug Target**	**UV –Visible titration % HS shift**	**MS – Protein + Ligand Peak detected?**
A1	Nystatin (Mycostatin)	926.09	Others	56.43%	**N**
A2	Isotretinoin	300.44	Hydroxylase	55.43%	**Y**
A3	Manidipine (Manyper)	610.7	Calcium Channel	46.78%	**N**
A4	Benzoic acid	122.12	Others	33.68%	**N**
A5	Atovaquone (Atavaquone)	366.84	Free Base	32.37%	**Y**
A6	Idebenone	338.44	Others	30.94%	**N/A**
A7	Azelnidipine	582.65	Calcium Channel	27.35%	**N**
A8	Meloxicam (Mobic)	351.4	Others	25.68%	**N**

A9	Rifampin (Rifadin, Rimactane)	822.94	Others	19.86%	**N**
A10	Amiloride hydrochloride (Midamor)	266.09	Others	16.42%	**N/A**
A11	Terbinafine (Lamisil, Terbinex)	291.43	Others	14.14%	**N**
A12	Tetracycline HCl	480.9	Others	13.68%	**Y**
C1	Clozapine (Clozaril)	326.82	5-HT Receptor	12.87%	**N**
C2	Lornoxicam (Xefo)	371.82	Others	12.50%	**N/A**
C3	Menadione	172.18	Others	10.28%	**N**
C4	Doxorubicin (Adriamycin)	579.98	Topoisomerase	10.22%	**Y**
C5	Dimethyl fumarate	144.13	Others	9.73%	**N**
C6	Phenindione (Rectadione)	222.24	Others	8.96%	**N**
C7	Methacycline hydrochloride (Physiomycine)	478.88	Others	8.86%	**Y**
C8	Furaltadone HCl	360.75	Others	8.50%	**Y**
C9	Oxytetracycline (Terramycin)	460.43	NULL	7.83%	**N/A**
C10	Nifedipine (Adalat)	346.33	Others	6.73%	**Y**
C11	Nitrofurazone (Nitrofural)	198.14	Others	6.50%	**Y**
C12	Meropenem	383.46	Others	3.59%	**N**
E1	Epirubicin HCl	579.98	Topoisomerase	3.32%	**Y**
E2	Rifaximin (Xifaxan)	785.88	Others	3.20%	**N**
E3	Rifapentine (Priftin)	877.03	Others	3.06%	**N**
E4	Idarubicin HCl	533.95	Topoisomerase	2.99%	**N/A**
E5	Mesalamine (Lialda)	153.14	Others	2.89%	**N**
E6	Ethacridine lactate monohydrate	361.39	Others	2.16%	**N**
E7	Tizanidine HCl	290.17	Adrenergic Receptor	2.00%	**N**
E8	Sitafloxacin Hydrate	436.84	Others	1.91%	**Y**
E9	Amfenac Sodium (monohydrate)	295.27	Others	1.43%	**N/A**

E10	Nithiamide	187.18	Others	1.09%	**N**
E11	Indomethacin (Indocid, Indocin)	357.79	Others	0.77%	**N**
E12	Doripenem Hydrate	438.52	Others	0.00%	**N**
G1	Sulfasalazine (Azulfidine)	398.39	Others	0.00%	**Y**
G2	Zolmitriptan (Zomig)	287.36	Others	0.00%	**N/A**
G3	Sparfloxacin	392.4	Others	0.00%	**N**
G4	Sulindac (Clinoril)	356.41	NULL	0.00%	**Y**
G5	Nitrendipine	360.3699	Calcium Channel	0.00%	**Y**
G6	Amiloride hydrochloride dihydrate	302.12	Sodium Channel	0.00%	**N**
G7	Clobetasol propionate	466.97	Others	0.00%	**N**
G8	Daunorubicin HCl (Daunomycin HCl)	563.98	Telomerase	0.00%	**Y**
G9	Nadifloxacin	360.38	Others	0.00%	**Y**
G10	Diminazene aceturate	515.52	Others	0.00%	**N/A**
G11	Nifuroxazide	275.22	Others	0.00%	**N**
G12	Bephenium Hydroxynaphthoate	443.53	Others	N/A	**N/A**

```
1      TIKEMPQPKT FGELKNLPLL NTDKPVQALM KIADELGEIF KFEAPGRVTR
             GGGTTGGGG  SS HHHHHH HHHHHH SEE EEEETTEEEE

51     YLSSQRLIKE ACDESRFDKN LSQALKFVRD FFGDGLVTSW THEKNWKKAH
       EE  HHHHHH HT TTTEEEE   HHHHHHHH HHTTSGGGS  TTSHHHHHHH

101    NILLPSFSQQ AMKGYHAMMV DIAVQLVQKW ERLNADEHIE VPEDMTRLTL
       HHHGGGGSHH HHHHHHHHHH HHHHHHHHHH HT  TT  EE HHHHHHHHHH

151    DTIGLCGFNY RFNSFYRDQP HPFITSMVRA LDEAMNKLQR ANPDDPAYDE
       HHHHHHHHS    GGG SS  HHHHHHHHH HHHHHHHHT      GGGHH

201    NKRQFQEDIK VMNDLVDKII ADRKASGEQS DDLLTHMLNG KDPETGEPLD
       HHHHHHHHHH HHHHHHHHHH HHHHT      SHHHHHHH  B TTT  B

251    DENIRYQIIT FLIAGHETTS GLLSFALYFL VKNPHVLQKA AEEAARVLVD
       HHHHHHHHHH HHHHTTHHHH HHHHHHHHHH HH HHHHHHH HHHHHHH  S

301    PVPSYKQVKQ LKYVGMVLNE ALRLWPTAPA FSLYAKEDTV LGGEYPLEKG
       SS  HHHHHT  HHHHHHHHH HHHHS SS E EEEEESSSEE ETTTEEE TT

351    DELMVLIPQL HRDKTIWGDD VEEFRPERFE NPSAIPQHAF KPFGNGQRAC
        EEEEEHHHH TT HHHH S  TTS  GGGGS SGGGS TTS   TT  GGGS

401    IGQQFALHEA TLVLGMMLKH FDFEDHTNYE LDIKETLTLK PEGFVVKAKS
       TTHHHHHHHH HHHHHHHHHH EEEE TT      EEESSSEE ETT EEEEEE

451    KKIPLGG
```

Total average mass: 54496.6671

Sequence molecular formula: C2440H3814N650O727S19

Figure 4.S1 Sequence and secondary structure (DSSP) for BM3 "gatekeeper" variant heme domain (PDB: 4KF2) (chain A)

Masses are presented in Daltons (Da). Figure generated using PDB sequence and DSSP image tool. Red single letter code: α helix; blue single letter code: beta strand; green single letter code: turn; black single letter code: no secondary structure assigned. Average mass calculated to be 54496.76 Da.

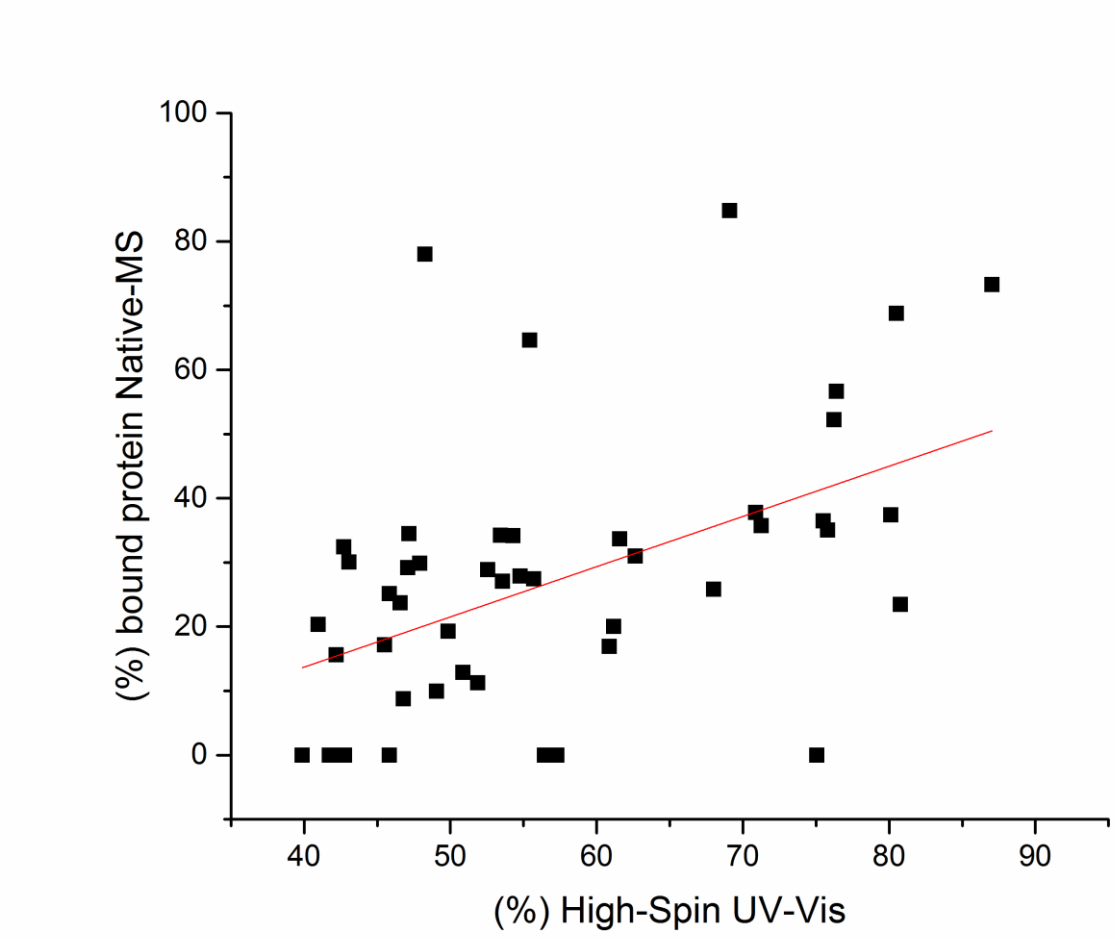

Figure 4.S2 **Correlation between %HS shift calculated by UV-Vis binding titrations and % compound-bound protein recorded using native-MS (variable set 1). Data fit using Pearson's r linear product moment correlation coefficient (PPMCC) (red line).**
PPMCC data: Pearson's r = 0.471; Sig = <0.001; N = 48.

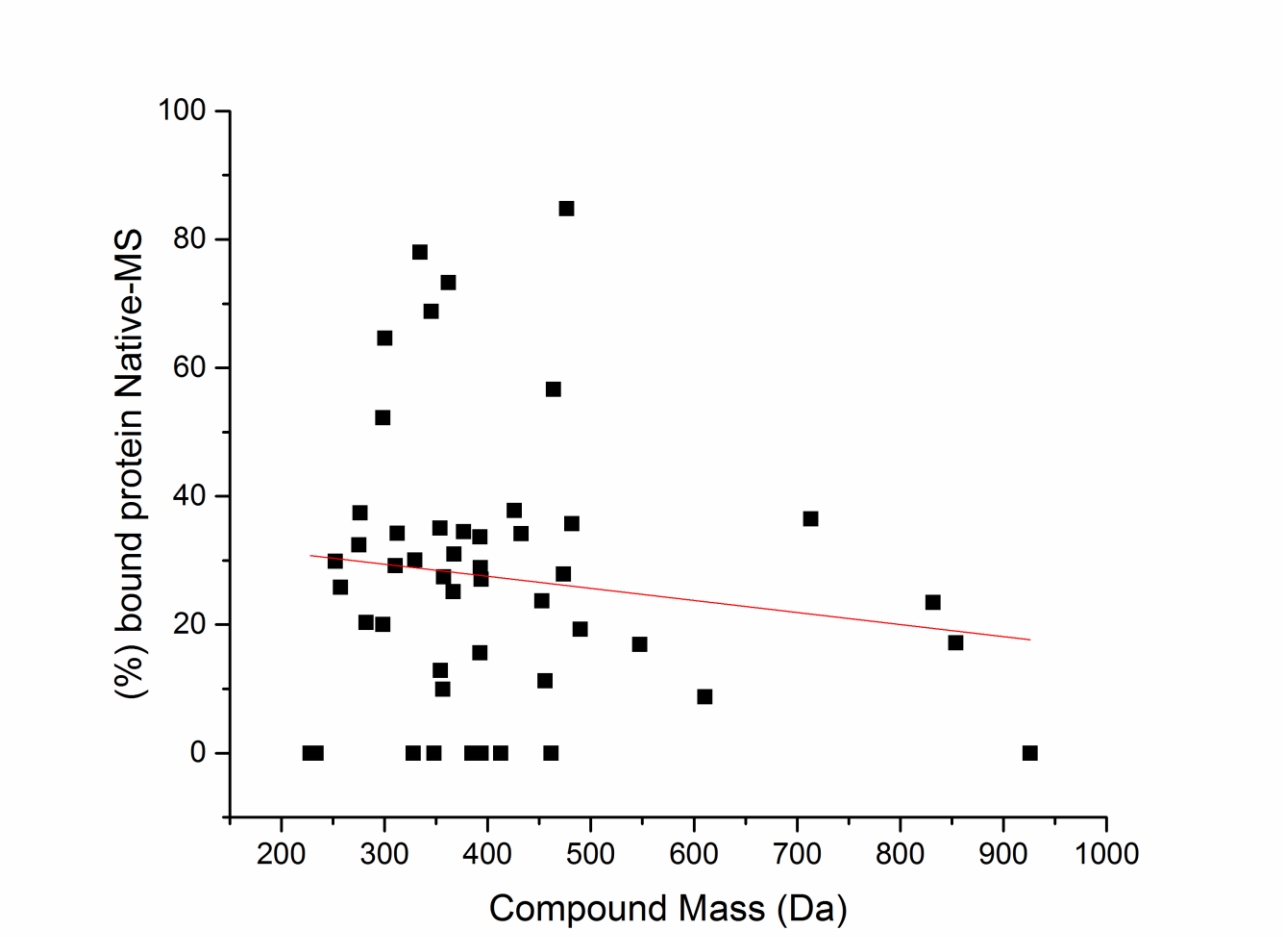

Figure 4.S3 **Correlation between drug compound molecular weight and % compound-bound protein recorded using native-MS (variable set 2). Data fit using Pearson's r linear product moment correlation coefficient (PPMCC) (red line).**

PPMCC data: Pearson's r = -0.134; Sig = 0.366; N = 48

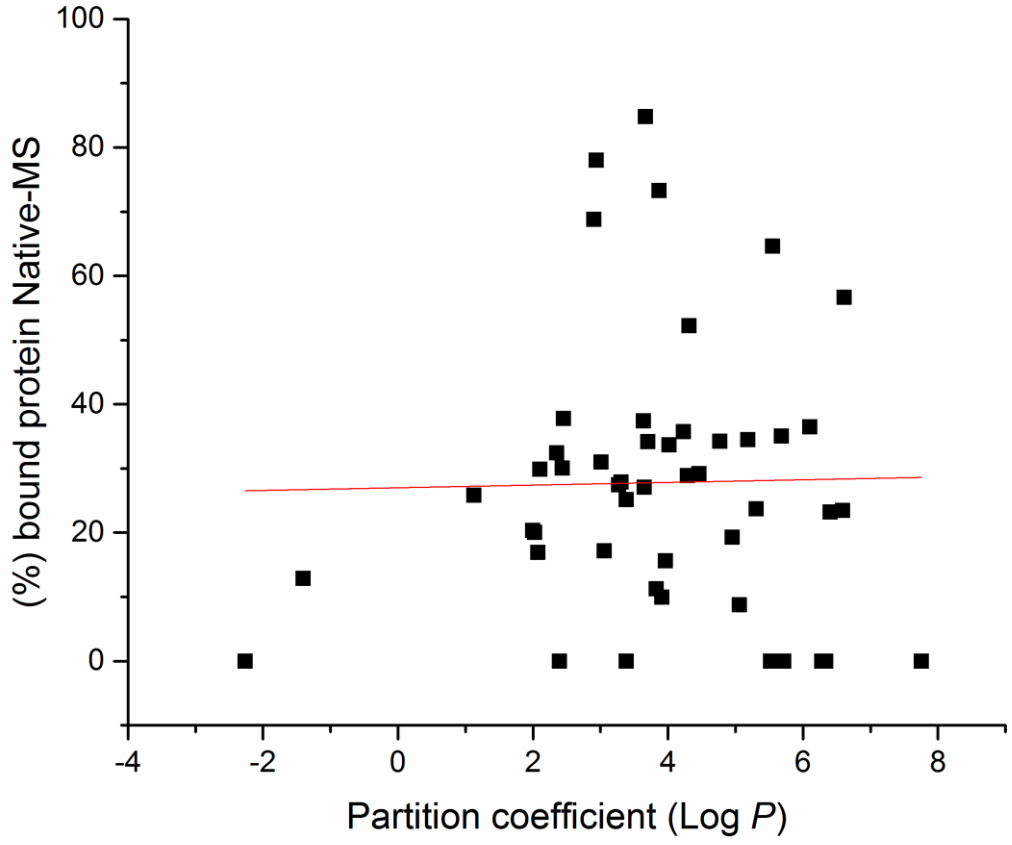

Figure 4.S4 **Correlation between drug compound partition coefficient (Log *P*) and % compound-bound protein measured via native-MS (variable set 3). Data fit using Pearson's r linear product moment correlation coefficient (PPMCC) (red line).**
PPMCC data: Pearson's r = 0.0190; Sig = 0.899; N = 48

Chapter 5: Summary, Conclusions and Future Work

5.1. Summary

The work presented in this thesis focuses on novel synthetic biology applications of P450 BM3 (CYP102A1), a fatty acid mono-oxygenase enzyme first isolated from the bacterium *Bacillus megaterium*. At the time of writing, BM3 has been a topic of research for 45 years, and has been the topic of over 500 published research papers (Web of Science Core Collection, search term "P450 BM3"; Miura and Fulco, 1974). While an already extensive body of research surrounds BM3, current research tends to focus on the wider industrial applications of BM3 as an effective tool of synthetic biology (Urlacher and Girhard, 2019; Fasan, 2012). Oxidation of C-H bonds with high levels of regio- and stereospecificity is a pressing issue in synthetic chemistry. Traditional methods of organic chemistry can often require several reaction steps, display indiscriminate hydroxylation (or other reactions) and produce low yields (Pazmino et al., 2010). Consequently, biocatalysts capable of performing highly regio- and stereo-selective oxidation reactions, ideally in a single step, are in high demand within industry.

The active site of BM3 has proven to be highly malleable, and is capable of being engineered to accept a wide range of non-natural substrates (Whitehouse et al., 2012). Numerous mutagenesis studies have successfully developed BM3 into a versatile biocatalyst, capable of functionalising an array of high value compounds as diverse as pharmaceuticals, steroids and terpenes (Girvan and Munro, 2016; van Vugt-Lussenberg et al., 2007). In addition, BM3 has also been utilised in the production of human drug metabolites, identical to those produced by human P450 enzymes (Di Nardo and Gilardi, 2012). A prokaryotic P450, BM3 serves as an ideal mimic to eukaryotic P450s, albeit one contained as a single, soluble polypeptide – forgoing the need for a complex purification protocol or costly redox partner proteins (Munro et al., 2007). Furthermore, the rate of activity reported for BM3, ~17,000 min^{-1} with the substrate arachidonic acid, remains unparalleled for a mono-oxygenase P450 (Noble, et al., 1999).

The work presented in this thesis is concerned with the manipulation of BM3 to produce high-value, hydroxylated products, difficult to obtain via non-enzymatic means. Chapter 2 of this thesis focused on the development and characterisation of a single-point mutation within a key, structurally conserved motif – the I-helix. Spanning residues Asp250 to Lys282, the I-helix lies close to the distal face of the heme cofactor (Ravichandran et al., 1993). On account of its close proximity to the heme cofactor, the I-helix motif has a

significant role in controlling access to the heme iron, and hence influences both substrate and regioselectivity. Previous studies have reported that point mutations at the position Ile263, situated within the I-helix, prompt significant departures in both substrate selectivity and regioselectivity (Dietrich et al., 2009; Sowden et al., 2005). The point mutation in question, I263P, was chosen to cause as much structural perturbation as possible – with the aim of "kinking" the I-helix away from the heme cofactor. The I-helix contains a complex network of hydrogen bonds that undergo significant rearrangement following substrate binding (Haines et al., 2001). A previous study conducted by Yeom et al (1997) on the hydrogen bonding partner of Ile263, i.e. Glu267, also yielded variants with altered oxidative activity relative to the wild-type enzyme. Proline was selected as the substitute residue for Ile at position 263. Proline was selected due to its comparatively small side chain and its high likelihood of inducing structural distortion within the I-helix.

The resulting variant, I263P was characterised using an array of biophysical techniques including X-ray crystallography. Alterations to oxidative activity relative to the wild-type enzyme were examined using a small library of monoterpenes. Changes to the pattern of regioselective hydroxylation of fatty acids were assessed using laurate, myristate and palmitate. The impact of I263P on the enantioselective epoxidation of styrene was also assessed.

While the proline substitution caused a drop in thermostability of ~2.4 C – structural data collected using X-ray crystallography revealed that the I-helix was not as substantially perturbed as anticipated. Moreover, while catalysis was impaired in the variant I263P relative to the wild-type enzyme, the loss of activity was not as significant as feared. Residue I263 lies in close proximity to the catalytically essential T268 – a residue conserved throughout P450s and implicated in the activation of molecular oxygen (Yeom et al., 1995). Despite this, the oxidative activity of variant I263P differed substantially to that of the unmodified, wild-type enzyme. The single variant consistently displayed increased binding affinities towards fatty acid substrates. Moreover, the product profile generated by the intact I263P mutant enzyme against fatty acid substrates laurate, myristate and palmitate was significantly more diverse than that of the wild-type. In the case of I263P-generated product profiles, hydroxylation of the fatty acids trended away from the sub-terminal ω-1, ω-2 and ω-3 positions towards positions closer to the α carbon.

Enantiopreference of styrene epoxidation was also shifted away from the production of an enantiomeric excess of the *R*(+)-styrene oxide enantiomer towards a near-racemic mix of

both the *R*(+) and *S*(-) enantiomers of styrene oxide. Moreover, production of styrene oxide increased near 5-fold relative to the wild-type enzyme. Finally, the I263P variant displayed increased levels of turnover, in addition to the formation of novel, oxidised products from a monoterpenoid library consisting of (*R*)-limonene, (+)-α-pinene and (-)-β-pinene. Particularly notable was the near 15-fold increase in the production of the functionalised terpene pinocarveol from the substrate (-)-β-pinene. In addition, the I263P variant produced the valuable fragrance compound limonene oxide from the substrate limonene – a product produced in negligible amounts by the wild-type enzyme.

As detailed in chapter 3, further development of the single I263P variant proceeded using a rational approach. From the literature, additional mutagenesis sites were identified: V78, F87 and A184 (Whitehouse et al., 2012; Tee and Schwaneberg, 2006). Successive point mutations generated three additional variants, named DM (I263P/F87V), TM1 (I263P/F87V/V78A) and TM2 (I263P/F87V/A184K). Triple mutants TM1 and TM2 were designed with two separate goals in mind. Variant TM1 was designed with the objective of further diversifying the regioselectivity of fatty acid hydroxylation towards mid-chain positions. Fatty acids hydroxylated at these positions can serve as important precursor compounds to flavouring agents known as lactones, in addition to a plethora of industrial lubricants (Dietrich et al., 2009; Hou, 2009). The second triple variant, TM2, was designed in order to favour the production of the *S*(-)-enantiomer of styrene oxide. Previous studies conducted by Huang et al. (2013) and Tee and Schwaneberg (2006) reported that the presence of a charged residue located at a position distant from the heme cofactor inverted regioselectivity in favour of *S*(-)-styrene oxide. Styrene oxide is a versatile chemical precursor required in the synthesis of numerous optically active pharmaceutical compounds (Panke et al., 1998). Consequently, enzymatic formation of optically pure styrene oxide is of value to industry. Alterations to regio- and enantioselectivity displayed by each triple variant were rationalised using molecular docking studies. Docking studies were performed using YASARA software and crystal structures of the TM1 and TM2 variant heme domains.

Across variants DM, TM1 and TM2, mono-hydroxylation of fatty acids was successfully focused on the mid-chain positions of the fatty acid substrates laurate, myristate and palmitate. Hydroxylation occurring at mid-chain positions, defined as those between the β carbon and the sub-terminal positions of ω-1, ω-2 and ω-3, became more pronounced with increasing fatty acid chain length. Hydroxylation performed by both triple variants focused

on the ω-6 and ω-7 positions with myristate and palmitate. Docking studies suggested that palmitate adopted a U-shaped conformation within both the TM1 and TM2 active sites, positioning the ω-7 carbon directly above the heme iron. In variant TM1 docking simulations, palmitate appeared to form a hydrogen bonding interaction with a residue of the B'-helix (S72), tethering the fatty acid at a position close to the heme iron. In the same variant, the ω terminus of palmitate appeared to be sequestered within an expanded active site sub-pocket, enabling the fatty acid to penetrate deeper into the active site. Consequently, the TM1 variant disfavoured oxidation at the ω-1, ω-2 and ω-3 sub-terminal positions that are extensively hydroxylated by the wild-type enzyme.

However, it was the TM2 variant that displayed the greatest diversity in product profile with fatty acid substrates across all the variants tested. The crystal structure of the TM2 variant revealed that the inclusion of the A184K substitution resulted in the formation of a salt-bridge interaction between residue K184 and D80 of the B'-helix. This interaction caused the F- and G-helices to translocate away from the heme cofactor, effectively raising the ceiling of the active site and increasing access to the substrate binding channel. This appears to have afforded fatty acid substrates greater mobility within the active site, leading to the production of hydroxylated laurate products oxidised at the β, γ and δ positions. Unlike the TM1 variant, the formation of a hydrogen bond between residue S72 and the substrate was not observed in the TM2 variant, implying that the fatty acid substrates adopt different binding modes within their respective active sites.

Building on the trend towards the production of the *S*(-) enantiomer of styrene oxide, as demonstrated by the single I263P mutant, all variants produced *S*(-)-styrene oxide to an enantiomeric excess. The TM2 variant produced the S(-) enantiomer in an excess (56.7 ± 0.306 % *ee*), the highest yet reported for a BM3 variant, and also generated a near 10-fold increase in styrene oxide production. The DM and TM1 variants also demonstrated a preference for the production of the *S*(-) enantiomer. However, overall formation of styrene oxide was less pronounced in the DM variant, which displayed only a 2-fold increase in production relative to the wild-type enzyme. Similar attempts were made to utilise docking simulations to rationalise the alterations in styrene enantioselectivity. However, due to the hydrophobic nature of both the active site and the styrene substrate, accurate models of styrene binding could not be produced for any of the variants, in addition to the wild-type enzyme.

In summary, chapter 3 successfully built upon the work began in chapter 2. Both the regio- and enantioselectivity of wild-type BM3 was successfully reengineered towards products of wide use throughout industry. These studies further emphasise the importance of BM3 as a multifaceted biocatalyst, in addition to highlighting the often dramatic impact that minimal residue substitutions can have on the activity of the enzyme. Purification and expression of the variant enzymes did not present any significant challenges, nor did the generation of protein crystals suitable for diffraction. Only the DM variant proved recalcitrant to crystallography – no protein structure was successfully collected for this variant, despite several attempts. Moreover, ligands present within the crystal structures of the TM1 and TM2 variants could not be successfully resolved. In both cases, it appeared that components present arising from the crystal screening plates, namely molecules of PEG, were present in the active sites of both heme domains. In addition, attempts to achieve styrene-bound crystal structures of all variants by crystal soaking proved fruitless. The absence of these structures complicated the analysis of the alterations in styrene enantioselectivity displayed by the variants. Quantification of the amount of each hydroxylated fatty acid product would have also been desirable. However, due to the large amount of standards this would have required (one for each carbon position per fatty acid) and the resulting expense (many of the in-chain hydroxylated compounds are costly to produce, hence the interest in doing so via enzymatic means) this was unfeasible. Additional biophysical data, in the form of EPR spectra collected for the DM, TM1 and TM2 variants, would provide further information regarding the respective heme environments, supplementing the data collected by resonance Raman spectroscopy.

The work presented in chapter 4 of this thesis details the application of native protein mass spectrometry (native-MS) for the screening of drug compounds against a promiscuous BM3 variant known as the “gatekeeper” (F87V/A82F). Prior research conducted by Butler et al. (2013; 2014) and Jeffreys et al. (2019; unpublished) have established the gatekeeper variant as capable of binding to and oxidising several human drug compounds. Moreover, against a selection of proton pump inhibitor, fibrate and statin compounds, the oxidised products generated by the gatekeeper were found to be analogous to those produced by human P450s (Jeffreys et al., unpublished data; Butler et al., 2014). Human drug metabolites are an integral component of ADME toxicity testing. All human drug metabolites must meet strict criteria relating to their manner of absorption, distribution, metabolism and excretion (ADME) prior to public consumption (Caldwell et al., 2009). In humans, P450 enzymes perform up to 80% of the phase I metabolism of all known human

drug compounds (Zanger and Schwab, 2013). Human P450s therefore appear to be the obvious choice when it comes to selecting catalysts for the production of human metabolites. However, as previously discussed in chapters 1 and 4, use of eukaryotic P450s in synthetic biology is hampered by a considerable set of issues. Consequently, engineered prokaryotic P450s capable of producing drug metabolites identical to those produced by human P450s are of high importance to the pharmaceutical industry.

The work detailed in chapter 4 builds on work initiated by Jeffreys et al. (unpublished data) in which an FDA approved compound library was screened using the gatekeeper variant. Affinity of the gatekeeper mutant towards these drug compounds was assessed using binding titrations measured using UV-visible spectroscopy. While this technique was successful, and yielded several "hit" compounds for further analysis, it was both time consuming and resource intensive. In addition, hit compounds were designated as such on their ability to induce a low-spin (~418 nm) to high-spin (~390 nm) shift in heme iron spin state. Several compounds in the FDA library were coloured yellow, orange or red – and consequently absorb at light within the region of 400-450 nm. As a result, low- to high-spin transitions may be difficult to record for these compounds using UV-visible spectroscopy. Compounds towards which the gatekeeper variant may display oxidative activity could therefore be overlooked by high throughput screening conducted solely using UV-visible spectroscopy.

These issues are overcome by the use of native-MS as a high throughput screening technique. Highly sensitive, quick to run and automatable, native-MS identifies the formation of and binding stoichiometry of protein+ligand complexes (Vivat Hannah et al., 2010). Using soft electrospray ionisation (ESI) conditions, proteins and protein+ligand complexes are able to retain their native conformations. The resulting mass-to-charge data enables protein+ligand complexes to be identified. Moreover, native-MS requires significantly smaller amounts of both protein and ligand when compared to UV-spectroscopic binding titrations – with data analysis possible in only a few minutes.

The work presented in chapter 4 began with two aims – both of which were successfully met. The first was to develop the native-MS screening method for use with the gatekeeper protein, and to ensure that it was an effective technique for the purposes of high throughput screening. This was achieved by screening a selection of FDA compounds previously identified as hits by UV-visible spectroscopy using native-MS. Protein+ligand complexes were detected for ~87% of the compounds previously identified as hits by UV-

visible spectroscopy, validating native-MS as a screening technique. The second aim was to identify new compounds from the library that could serve as potential substrates for the gatekeeper enzyme and that were previously overlooked by the spectroscopic screening approach. Five compounds for which no shift (0.0% high spin conversion) was detected using UV-visible spectroscopy were revealed to form protein+ligand complexes with the gatekeeper mutant when screening using native-MS (Sulindac, Nadifloxacin, Azulfidine, Nitrendipine and Daunorubicin hydrochloride).

These compounds were then subjected to LC-MS and MS/MS analysis following turnover with the intact gatekeeper enzyme. Of the five candidates, oxidised metabolites were only observed for the compound azulfidine, the commercial name of sulfasalazine – a prodrug used for the treatment of rheumatoid arthritis, amongst other ailments (Pullar, 1989). Subsequent MS/MS analysis revealed that while azulfidine had been oxidised by the gatekeeper enzyme, the postulated metabolite did not correspond to any metabolite produced by a human P450. Despite this, the data presented in chapter 4 demonstrate the viability of native-MS as suitable technique for high-throughput compound screening against a variant BM3 heme domain. To our knowledge, chapter 4 contains the first report of native-MS being utilised as a compound screening technique against any BM3 protein. Now that the technique has been optimised, native-MS will surely prove a useful tool in the discovery of future drug compounds for the voracious gatekeeper variant.

5.2. Conclusions

The aim of this thesis was to explore novel synthetic biology applications of P450 BM3. The work presented in chapter 2 contributes a versatile new point mutation, I263P, to the existing pantheon of single residue substitutions that have a profound impact on the activity of BM3. In chapter 3, the further engineering of the I263P variant produced a selection of biocatalysts capable of hydroxylating fatty acids at mid-chain positions, with high levels of regioselectivity. Hydroxylation at these mid-chain carbons seldom occurs within prokaryotic P450s. Therefore, this thesis demonstrates the successful re-engineering of BM3 into a mimic of a eukaryotic fatty acid hydroxylase. The resulting products act as valuable precursors in the production of a diverse range of compounds with applications in the agricultural, cosmetic and flavouring industries. In addition, this work demonstrated that a limited numbers of residue substitutions, including I263P, were sufficient to invert the enantioselective epoxidation of styrene in P450 BM3. Hopefully, the I263P variant will serve as the basis for future BM3 variants capable of the oxidative transformation of further high-value compounds.

The work presented in this thesis also demonstrated the successful pairing of a novel method of compound screening with a highly promiscuous variant of P450 BM3. The successful application of the native-MS technique to the gatekeeper variant (as detailed in chapter 4) will facilitate the screening of large numbers of FDA approved drug compounds in a vastly reduced timeframe. This orthogonal screening method will enable the identification of additional pharmaceutical compounds that may serve as potential substrates for the gatekeeper variant. Gatekeeper mediated oxidative transformation of these drug compounds may yield human drug metabolites, providing a more cost effective method of generating components essential to one of the most crucial stages of drug toxicity testing. To the best of our knowledge, this thesis also contains the first report of compound screening performed using native-MS applied to a variant of P450 BM3. The work presented in this thesis will therefore open the door to future applications of native-MS as a means of identifying activity between compounds and variant BM3 enzymes.

5.3. Future Work

The collection of ligand-bound crystal structures proved elusive for all variants but the single I263P mutant. While structures for the TM1 and TM2 variants were obtained, and while the secondary structures of these variants appeared consistent with that of a substrate-bound structure, the positions adopted by the NPG or styrene ligands could not be successfully resolved. This necessitated the use of docking studies. While these simulations provided suitable rationales for the alterations to hydroxylated fatty acid product profiles, they provided inadequate explanations for the changes to the enantiomeric ratios of styrene oxide production. Collection of styrene-bound structures will hopefully shed further light on the exact roles the I-helix and the expanded sub-pocket present within the DM, TM1 and TM2 variants perform with respect to the orientation of styrene within the active sites of each variant. These data would enable additional residue substitutions to be made to the variants in a rational manner with the aim of producing optically pure *S*(-)-styrene oxide.

A crystal structure of a fatty acid, or fatty acid analogue such as NPG, bound within the active site of the TM1 variant would be highly desirable. The structural data would serve to establish beyond doubt the existence of a hydrogen bonding interaction between residue S72 and the carboxylate group of the substrate as suggested by the docking studies. If confirmed, a point mutation at position S72, designed to introduce this interaction further up the substrate access channel, may serve to tether shorter chain fatty acids (<C:14) in poses favouring oxidation at the valuable mid-chain positons.

The exposure of the DM, TM1 and TM2 to the same monoterpenoid library as the single variant (limonene, (+)-α-pinene, (-)-β-pinene) was also attempted. While diversified product profiles were apparent for each variant, the products generated were frequently inconsistent, difficult to reproduce and hard to identify. A more focused approach, targeting the production of a single, high value oxidised terpene compound using variants built upon I263P was abandoned due to project time constraints. This idea was to be explored through the addition of the I263P variant to established BM3 variants, namely the GVQ (A74G/F87V/L188Q) and gatekeeper (F87V/A82F) mutants (Li et al., 2000; Butler et al., 2013). Nevertheless, depending on the selection of an appropriate target product, speculative "GVQP" (A74G/F87V/L188Q/I263P) and F87V/A82F/I263P variants could yet prove to be effective biocatalysts.

This thesis demonstrated the successful combination of high throughput native-MS, a commercially important drug library and a high activity BM3 variant. However, the total number of compounds currently tested represents less than 10 percent of the total compound library.

Full coverage of the compound library, in tandem with fine-tuning the method development, would be the obvious next steps. Hit compounds would be subjected to LC-MS analysis in order to confirm the discovery of new substrates or substrate groups amenable to oxidative transformation by the gatekeeper mutant. Following the completion of screening with the gatekeeper variant, screening of other variants of P450 BM3 against the FDA library should be conducted. Metabolite analysis using nuclear magnetic resonance (NMR) spectroscopy would be desirable in order to assign the precise sites of oxidation with confidence. Once identified, the production of relevant drug metabolites would be scaled up.

Finally, I see scope for directing future native-MS projects towards identification of inhibitor compounds for other P450 enzymes. Drug resistance in pathogenic bacteria drives a critical need for discovery of new drug compounds capable of inhibiting bacterial metabolism. The targeting of P450 enzymes present within pathogenic bacteria such as *Mycobacterium tuberculosis* would be of particular interest. Application of an optimised screening technique could identify antagonist compounds of interest to the pharmaceutical industry, in addition to wider public health authorities.

5.4. References

Butler, C. F., Peet, C., Mason, A. E., Voice, M. W., Leys, D. and Munro, A. W. (2013). 'Key mutations alter the cytochrome P450 BM3 conformational landscape and remove inherent substrate bias', *J Biol Chem,* 288(35), pp. 25387-99.

Butler, C. F., Peet, C., McLean, K. J., Baynham, M. T., Blankley, R. T., Fisher, K., Rigby, S. E., Leys, D., Voice, M. W. and Munro, A. W. (2014). 'Human P450-like oxidation of diverse proton pump inhibitor drugs by 'gatekeeper' mutants of flavocytochrome P450 BM3', *Biochem J,* 460(2), pp. 247-59.

Caldwell, G. W., Yan, Z., Tang, W., Dasgupta, M. and Hasting, B. (2009). 'ADME optimization and toxicity assessment in early- and late-phase drug discovery', *Curr Top Med Chem,*9(11), pp. 965-80.

Di Nardo, G. and Gilardi, G. (2012). 'Optimization of the bacterial cytochrome P450 BM3 system for the production of human drug metabolites', *Int J Mol Sci,* 13(12), pp. 15901-24.

Dietrich, M., Do, T. A., Schmid, R. D., Pleiss, J. and Urlacher, V. B. (2009). 'Altering the regioselectivity of the subterminal fatty acid hydroxylase P450 BM-3 towards gamma- and delta-positions', *J Biotechnol,* 139(1), pp. 115-7.

Fasan, R. (2012). 'Tuning P450 enzymes as oxidation catalysts.' *ACS Catalysis*, 2(4), pp.647-66.

Girvan, H. M. and Munro, A. W. (2016). 'Applications of microbial cytochrome P450 enzymes in biotechnology and synthetic biology', *Curr Opin Chem Biol,* 31, pp. 136-45.

Haines, D. C., Tomchick, D. R., Machius, M. and Peterson, J. A. (2001). 'Pivotal role of water in the mechanism of P450BM-3', *Biochemistry,* 40(45), pp. 13456-65.

Hou, C.T. (2009). 'Biotechnology for fats and oils: new oxygenated fatty acids.' *New Biotechnol*, 26(1-2), pp. 2-10.

Jeffreys, L. N., Poddar, H., Golovanova, M., Levy, C. W., Girvan, H. M., McLean, K. J., Voice, M. W., Leys, D. and Munro, A. W. (2019). 'Novel insights into P450 BM3 interactions with FDA-approved antifungal azole drugs', *Sci Rep,* 9(1):1577.

Li, Q. S., Schwaneberg, U., Fischer, P. and Schmid, R. D. (2000). 'Directed evolution of the fatty-acid hydroxylase P450 BM-3 into an indole-hydroxylating catalyst', *Chemistry,*6(9), pp. 1531-6.

Miura, Y. and Fulco, A. J. (1974). '(Omega -2) hydroxylation of fatty acids by a soluble system from bacillus megaterium', *J Biol Chem,* 249(6), pp. 1880-8.

Munro, A. W., Girvan, H. M. and McLean, K. J. (2007). 'Variations on a (t)heme - novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily', *Nat Prod Rep,* 24(3), pp. 585-609.

Noble, M. A., Miles, C. S., Chapman, S. K., Lysek, D. A., MacKay, A. C., Reid, G. A., Hanzlik, R. P. and Munro, A. W. (1999). 'Roles of key active-site residues in flavocytochrome P450 BM3', *Biochem J,* 339 (Pt 2), pp. 371-9.

Panke, S., Witholt, B., Schmid, A. and Wubbolts, M. G. (1998). 'Towards a biocatalyst for (S)-styrene oxide production: characterization of the styrene degradation pathway of Pseudomonas sp. strain VLB120', *Appl Environ Microbiol,* 64(6), pp. 2032-43.

Pazmino, D.T., Winkler, M., Glieder, A. and Fraaije, M.W. (2010). 'Monooxygenases as biocatalysts: classification, mechanistic aspects and biotechnological applications.' *J Biotechnol*, 146(1-2), pp. 9-24.

Pullar, T. (1989). 'Sulphasalazine and related drugs in rheumatoid arthritis', *Pharmacol Ther,*42(3), pp. 459-68.

Ravichandran, K. G., Boddupalli, S. S., Hasemann, C. A., Peterson, J. A. and Deisenhofer, J. (1993). 'Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's', *Science,* 261(5122), pp. 731-6.

Shehzad, A., Panneerselvam, S., Linow, M., Bocola, M., Roccatano, D., Mueller-Dieckmann, J., Wilmanns, M. and Schwaneberg, U. (2013). 'P450 BM3 crystal structures reveal the role of the charged surface residue Lys/Arg184 in inversion of enantioselective styrene epoxidation', *Chem Commun (Camb),* 49(41), pp. 4694-6.

Sowden, R. J., Yasmin, S., Rees, N. H., Bell, S. G. and Wong, L. L. (2005). 'Biotransformation of the sesquiterpene (+)-valencene by cytochrome P450cam and P450BM-3', *Org Biomol Chem,* 3(1), pp. 57-64.

Tee, K. L. and Schwaneberg, U. (2006). 'A screening system for the directed evolution of epoxygenases: importance of position 184 in P450 BM3 for stereoselective styrene epoxidation', *Angew Chem Int Ed Engl,* 45(32), pp. 5380-3.

Urlacher, V. B. and Girhard, M. (2019). 'Cytochrome P450 Monooxygenases in Biotechnology and Synthetic Biology', *Trends Biotechnol,* 37(8), pp. 882-97.

van Vugt-Lussenburg, B. M., Stjernschantz, E., Lastdrager, J., Oostenbrink, C., Vermeulen, N. P. and Commandeur, J. N. (2007). 'Identification of critical residues in novel drug metabolizing mutants of cytochrome P450 BM3 using random mutagenesis', *J Med Chem,* 50(3), pp. 455-61.

Vivat Hannah, V., Atmanene, C., Zeyer, D., Van Dorsselaer, A. and Sanglier-Cianférani, S. (2010). 'Native MS: an 'ESI' way to support structure- and fragment-based drug discovery', *Future Med Chem,* 2(1), pp. 35-50.

Yeom, H. and Sligar, S. G. (1997). 'Oxygen activation by cytochrome P450BM-3: effects of mutating an active site acidic residue', *Arch Biochem Biophys,* 337(2), pp. 209-16.

Yeom, H., Sligar, S. G., Li, H., Poulos, T. L. and Fulco, A. J. (1995). 'The role of Thr268 in oxygen activation of cytochrome P450BM-3', *Biochemistry,* 34(45), pp. 14733-40.

Zanger, U. M. and Schwab, M. (2013). 'Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation', *Pharmacol Ther,* 138(1), pp. 103-41.