

**Deciphering electron transfer and
cytochrome P450 activity in
*Mycobacterium marinum***

Stella Child



Thesis submitted for the degree of Doctor of Philosophy

Supervisor:

Dr. Stephen Bell

School of Physical Sciences

The University of Adelaide

23rd May 2018

Table of Contents

Table of Figures	v
List of Tables	ix
Citations	xi
Glossary and abbreviations	xii
Thesis declaration	xiv
Acknowledgements	xv
Abstract	1
Chapter 1: Introduction	3
1.1: Cytochrome P450s	3
1.2: The catalytic cycle	7
1.3: Electron transfer systems supporting cytochrome P450s	11
1.4: Ferredoxins in bacterial cytochrome P450 systems	14
1.5: Bacterial CYPs in natural product synthesis pathways	18
1.6: <i>Mycobacterial</i> CYPs: evolution and characterisation	22
1.7: Electron transfer in <i>Mycobacterium marinum</i>	31
1.8: Summary and project goals	36
References	38
Chapter 2:	50
Electron transfer ferredoxins with unusual cluster binding motifs support monooxygenase secondary metabolism in many bacteria	50
Statement of Authorship	51

Chapter 3:.....	63
Selective ω -1 oxidation of fatty acids by CYP147G1 from <i>Mycobacterium marinum</i>	63
Statement of Authorship.....	64
Highlights	65
Abstract.....	66
Key words.....	67
Graphical abstract	67
Abbreviations.....	68
1. Introduction	69
2. Experimental.....	72
3. Results and Discussion	76
4. Conclusions	92
References	94
Chapter 4:.....	98
The characterisation of two members of the cytochrome P450 CYP150 family: CYP150A5 and CYP150A6 from <i>Mycobacterium marinum</i>	98
Statement of Authorship.....	99
Highlights	100
Abstract.....	101
Abbreviations.....	103
1. Introduction	104
2. Experimental.....	107

3. Results/Discussion.....	112
4. Conclusions	126
References	127
Chapter 5:.....	132
Structural and functional characterisation of the cytochrome P450 enzyme CYP268A2 from <i>Mycobacterium marinum</i>	132
Statement of Authorship.....	133
Chapter 6:.....	152
A comparison of the steroid binding cytochrome P450s from <i>Mycobacterium marinum</i> and <i>Mycobacterium tuberculosis</i>	152
Statement of Authorship.....	153
Highlights	154
Abstract.....	155
1. Introduction	158
2. Experimental.....	161
3. Results/Discussion.....	166
4. Conclusions	180
References	182
Conclusions and future directions.....	185
Appendices.....	192
Chapter 2 Supplementary Information	192
Chapter 3 Supplementary Information.....	250

Chapter 4 Supplementary information	273
Chapter 5 Supplementary Information	295
Chapter 6 Supplementary Information	318

Table of Figures

Chapter 1:

Figure 1: The haem b cofactor at the centre of cytochrome P450s.	3
Figure 2: The main substrate class of CYPs from all kingdoms of life.....	5
Figure 3: The structure of P450cam (CYP101A1).	7
Figure 4: The catalytic cycle of cytochrome P450s (showing steps I → VII).....	8
Figure 5: The radical recombination mechanism of the ferryl intermediate (Compound I) of the P450 (expansion of Step VII → I).....	9
Figure 6: The effect of substrate binding on the coordination state of the Fe-haem complex and electron configurations for the 3d orbitals of the Fe(III).....	10
Figure 7: The binding modes of ligands to the haem.	11
Figure 8: Varying electron transfer systems of P450 enzymes.	12
Figure 9: The formal reduction potentials (E_0') of the species involved in the catalytic cycle of a Class 1 CYP electron transfer system.	13
Figure 10: The structures of ferredoxin iron-sulfur cluster variants.....	15
Figure 11: The structure of himastatin.....	19
Figure 12: Biosynthetic gene clusters of various Actinomycete species.....	21
Figure 13: Venn diagrams showing the number of coding sequences that are shared between five Mycobacterial species.	23
Figure 14: Overlay of the CYPs from <i>M. tuberculosis</i> that have been structurally characterised in the substrate-free state.	25
Figure 15: Products of the Mycobacterium CYP enzymes.....	26
Figure 16: Conserved amino acids from Mycobacterial CYP families.....	27
Figure 17: A proposed model of the cell wall of <i>M. smegmatis</i>	30
Figure 18: The novel ferredoxin sequences of <i>M. marinum</i> at the cluster binding motif.....	33

Figure 19: A comparison of the iron-sulfur clusters of two ferredoxins.	35
---	----

Chapter 2:

Figure 1: Sequence alignment of the eleven ferredoxins of <i>M. marinum</i>	56
--	----

Figure 2: A phylogenetic tree (phenogram) of the [3/4Fe-4S] ferredoxins from <i>M. marinum</i>	57
---	----

Figure 3: CYP147G1 oxidation of undecanoic acid to 10-hydroxyundecanoic acid and the oxidation of β -ionone to 4-hydroxy- β -ionone by CYP278A1 and CYP150A5.	57
---	----

Figure 4: CYP147G1 product formation is reduced when supported by the mutant Fdx partners.....	58
---	----

Figure 5: The UV/Vis absorbance spectra of aerobically purified Fdx4 and Fdx8 from <i>M. marinum</i>	59
---	----

Figure 6: Activity of CYP150A5 and CYP278A1 when supported by native electron transfer partners <i>in vivo</i>	59
---	----

Chapter 3:

Figure 1: A phylogenetic tree (phenogram) of CYP147G1 and analogous enzymes.....	77
---	----

Scheme 1: A selection of the substrates tested on CYP147G1	80
---	----

Figure 2: A selection of substrate dissociation constants for CYP147G1	82
---	----

Figure 3: GC-MS chromatogram of <i>in vivo</i> turnovers of CYP147G1	85
---	----

Scheme 2: CYP147G1 oxidation of octanoic acid, decanoic acid, undecanoic acid and dodecanoic acid.....	84
---	----

Figure 4: Expansion of the region where the products elute in the GC-MS chromatograms of the <i>in vivo</i> turnovers of CYP147G1	87
--	----

Scheme 3: CYP147G1 oxidation of methyl branched fatty acids.....	88
---	----

Chapter 4:

Figure 1: Phylogenetic tree (phenogram) of <i>M. marinum</i> enzymes CYP150A5 and CYP150A6.....	113
Figure 2: CYP150A6 (black, A418), the reduced ferrous (blue, A417) and the ferrous form bound with CO (red, A446) showing the characteristic ~450 nm absorbance.	115
Figure 3: Structures of selected substrates tested with CYP150A5 and CYP150A6.	117
Figure 4: Dissociation constant analysis of CYP150A5.....	118
Figure 5: Dissociation constants of CYP150A6.....	119
Figure 6: GC chromatogram of in vivo turnover of CYP150A5 with Fdx8 and FdR1 ..	121
Scheme 1: CYP150A5 conversion of β -ionol to 4-hydroxy- β -ionol.	121
Figure 7: CYP150A6 resolved to 1.5 Å.....	123

Chapter 5:

Figure 1: Phylogenetic tree of the CYP268 family.....	141
Figure 2: Ferric CYP268A2, the reduced ferrous form and the ferrous form bound with CO, showing the characteristic absorbance at ~450 nm..	142
Figure 3: Structures of compounds tested as potential CYP268A2 substrates.....	142
Figure 4: Dissociation constant analysis of CYP268A2 with selected substrates.....	145
Figure 5: GC chromatograms of CYP268A2 catalysed substrate turnovers..	146
Figure 6: Hydroxylation products of CYP268A2.....	147
Figure 7: X-ray crystal structure of CYP268A2.....	148
Figure 8: CYP268A2 with the inner surface cavity of the enzyme active site which encloses the pseudoionone substrate.....	149

Chapter 6:

Figure 1: Phylogenetic tree showing selected members of the 125, 142 and 124 families from <i>Mycobacterium</i> and other species	167
Figure 2: The structures of a variety of azoles that have been tested as inhibitors of the <i>M. tuberculosis</i> CYPs.	171
Figure 3: CO binding spectra for the <i>M. marinum</i> enzymes CYP124A1, CYP125A6, CYP125A7 and CYP142A3.	168
Figure 4: A selection of substrate and inhibitor dissociation constants.....	173
Figure 5: The overlay of X-ray crystal structures of the CYP124A1 enzymes from <i>M. marinum</i> and <i>M. tuberculosis</i>	177
Figure 6: The residues close to the haem in the CYP124A1 active site.....	178

List of Tables

Chapter 1

Table 1: Approximate ranges of reduction potentials for iron-sulfur cluster containing ferredoxins.....	16
--	----

Chapter 2

Table 1: The potential electron transfer partners of the CYPome of <i>M. marinum</i>	55
Table 2: Characterised gene clusters with a ferredoxin which has a ferredoxin motif similar to those identified in <i>M. marinum</i> M.....	56

Chapter 3

Table 1: Sequence identities of the CYP, Fdx and FdR equivalents of the CYP147G1/Fdx3/FdR1 operon from Mycobacterium species and elsewhere.....	78
Table 2: Substrate binding and product formation data for CYP147G1.....	81

Chapter 4

Table 1: Crystal refinement data for CYP150A6 from <i>M. marinum</i> (PDB: 6DCB).....	111
Table 2: Substrate binding data for both CYP150A5 and CYP150A6.....	116
Table 3: Binding data for possible inhibitors of both CYP150A5 and CYP150A6.	119
Table 4: The active site residues of CYP150A6 and the aligned residues of other CYP150 family members.....	125

Chapter 5

Table 1: Spin-state shift and dissociation constants of CYP268A2 with a variety of substrates.....	143
Table 2: Crystal refinement data for CYP268A2 from <i>M. marinum</i> (PDB: 6BLD).....	146

Chapter 6

Table 1: Crystal refinement data for CYP124A1 from <i>M. marinum</i> (PDB: 6CVC)	165
Table 2: Spin state shift and dissociation constant analysis of CYP125A1 from <i>M. tuberculosis</i> and CYP125A6 and CYP125A7 from <i>M. marinum</i>	170
Table 3: Spin state shift and dissociation constant analysis of CYP142A1 from <i>M. tuberculosis</i> and CYP142A3 <i>M. marinum</i>	175
Table 4: Spin state shift analysis of CYP124A1 from <i>M. tuberculosis</i> and <i>M. marinum</i>	176
Table 5: Comparison of the active site residues of CYP124A1 from <i>M. tuberculosis</i> and <i>M. marinum</i>	179

Citations

Child, S. A.; Bradley, J. M.; Pukala, T.; Svistunencko, D. A.; Le Brun, N. E.; Bell, S. G., Electron transfer ferredoxins with unusual cluster binding motifs support secondary metabolism in many bacteria, submitted for publication in *Chemical Science*.

Child, S. A.; Rossi, V.; Bell, S. G., Selective ω -1 oxidation of fatty acids by CYP147G1 from *Mycobacterium marinum*, manuscript in preparation.

Child, S. A.; Flint, K.; Bell, S. G., The characterisation of two members of the cytochrome P450 CYP150 family: CYP150A5 and CYP150A6 from *Mycobacterium marinum*, manuscript in preparation.

Child, S. A.; Naumann, E. F.; Bruning, J. B.; E. Bell, S. G., Structural and functional characterisation of the cytochrome P450 enzyme CYP268A2 from *Mycobacterium marinum*, *Biochem. J.*, 2018, 475(4), 705-722.

Child, S. A.; Bruning, J. B.; Bell, S. G., A comparison of the steroid binding cytochrome P450s from *Mycobacterium marinum* and *Mycobacterium tuberculosis*, manuscript in preparation.

Glossary and abbreviations

2xYT, 2 x concentration yeast extract tryptophan broth;

ArR, a ferredoxin reductase from *Novosphingobium aromaticivorans*;

BSTFA-TMSCl, N,O-bis(trimethylsilyl) trifluoroacetamide and trimethylsilylchloride;

CYP or P450, Cytochrome P450 enzyme;

DTT, dithiothreitol;

EMM, *E. coli* minimal media;

FAD, flavin adenine dinucleotide;

FdR, ferredoxin reductase;

Fdx, ferredoxin;

FFFAS, Fold and Function Assignment System;

GC-MS or MS, gas-chromatography mass spectrometry or mass spectrometry;

H β CD, hydroxyl- β -cyclodextrin;

IPTG, Isopropyl β -D-1-thiogalactopyranoside;

LB, Lysogeny broth (also known as Luria or Lennox Broth),

MAC, *Mycobacterium avium* complex;

MCAC, *Mycobacterium chelonae-abscessus* complex;

MDR, multi drug-resistant;

MM, mycomembrane;

MRSAD, Molecular Replacement Single Wavelength Anomalous Diffraction;

MTBC, *Mycobacterium tuberculosis* complex;

NAD(P)H reduced nicotinamide adenine dinucleotide (phosphate);

NRPS, non-ribosomal peptide synthetase;

NTM, non-tuberculous mycobacteria;

PCR, polymerase chain reaction;

PDB, Protein Data Bank; PDR, phthalate dioxygenase reductase;

PDIM, phthiocerol dimycocerosates.

PIM, phenylimidazole;

PKS, poly-ketide synthase;

RT, retention time;

SAP, saprophytes;

SOC, Super Optimal broth with Catabolite repression;

Tdx, terpredoxin a ferredoxin from *Pseudomonas sp*;

TMS, trimethylsilyl;

TOMM, thiol/oxazole-modified microcin;

XDR, extensively drug-resistant.

Thesis declaration

I certify that to the best of my knowledge and belief, this work contains no material previously published or written by another person, except where due reference has been made in the text. In addition, this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution, except where as a result of the Thesis by Publication format, it was necessary for the preparation of the manuscript. In these instances the work clearly indicated in the authorship statements that accompany the relevant publications and is **not** submitted for credit to this degree.

I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works. I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of a University of Adelaide Faculty of Sciences Divisional Scholarship.

Stella Child

Acknowledgements

Many thanks are owed for the support I received in producing this thesis and the work behind it. I would like to start by thanking my supervisor Stephen Bell, who has been endlessly patient, particularly in the last few months of drafting, and whose guidance throughout has made me the better, more thorough researcher that I (hopefully) am now.

Among the others that have contributed to or supported this project, I would like to thank our collaborators at the University of East Anglia, in particular Justin Bradley and Nick Le Brun, and John Bruning for all his help with the crystallography work in this project, and also the Bruning group members Andrew Marshall and Andrew Thompson for their patience with my questions. I am grateful to Tara Pukala for the help with mass spectrometry, Simon Pyke for advice on NMR and Kate Wegener for the effort that went into the NMR protein structure determination attempt. Thanks also go to Vanessa Rossi, Kate Flint and Elise Naumann who worked on parts of this project while on placement in our lab.

I am also thankful for help of Katherine Stevens, Tom Coleman, and Joel Lee in proof reading parts of this document, and to all other members of the Bell group over the last five years of my honours and now PhD work in this lab.

Finally, my family and friends deserve more thanks than I can give them. I can't list everyone but to name a few, my housemates past and present (who've had to live with me!), Steph, Kieran, Gabs, Robyn and Alice, plus Holly with our weekly lunches, the Glamping crew, and Galen with no-talking-about-work backgammon time, have done a lot to make this what it was. My brothers, Billy and Kit, and my mum, Henrietta, are the reason I have got to the end of it.

Abstract

Cytochrome P450s are haem-monooxygenase enzymes, responsible for the catalytic hydroxylation of a large variety of organic molecules. The bacterium *Mycobacterium marinum*, has a larger genome than its close relatives, the causative agents of human tuberculosis (*Mycobacterium tuberculosis*) and Buruli ulcer (*Mycobacterium ulcerans*), which have undergone substantial reductive evolution. The genome of *M. marinum* contains an unusually large number of P450 genes (47). Twelve ferredoxin genes are associated with the CYPome and eleven of these are uncharacterised ferredoxins of the 3/4Fe-4S type. In their iron-sulfur cluster binding motif (CXX?XXC(X)_nCP), these ferredoxins (Fdx1 – Fdx11) have non-standard residues at the ? position of the sequence. Instead of the cysteine residue expected of a [4Fe-4S] ferredoxin, or the alanine/glycine residue expected in a [3Fe-4S] ferredoxin, they contain histidine, asparagine, tyrosine, serine, threonine and phenylalanine residues. In the course of this work, they have been purified aerobically and anaerobically. When isolated anaerobically, three of these ferredoxins were determined, by non-denaturing ESI-MS and EPR to contain 3Fe-4S clusters. The reduction potentials for the three varied from +150 mV to -360 mV, which are highly anomalous for [3Fe-4S] ferredoxins. Similar ferredoxins were found to accompany P450s in the biosynthetic gene clusters of other bacteria, especially in Actinomycete species.

These ferredoxins were demonstrated to support the activity of a number of the *M. marinum* P450s, some of which were from previously uncharacterised families. CYP147G1, in combination with the electron transfer partners Fdx3 and FdR1 was demonstrated to act as a ω-1 fatty acid hydroxylase. CYP147G1 selectivity favoured the ω carbon when branched methyl substrates were used. The same ferredoxin reductase, FdR1, was also shown to support the activity of CYP278A1 (with Fdx2), and CYP150A5 (with Fdx8), both of which

were shown to regioselectively hydroxylate β -ionone. CYP150A5 binds terpenes and polycyclic substrates. An additional CYP150 enzyme, CYP150A6, was crystallised and structurally resolved to 1.6 Å in the substrate-free form.

CYP268A2, when reconstituted with a non-native electron transfer chain, hydroxylated the branched fatty acetate derivatives, pseudoionone and geranyl acetate, at the terminal position. The structure of CYP268A2 with *trans*-pseudoionone bound in the active site was solved by X-ray crystallography to a resolution of 2.0 Å and from this the selectivity of the enzyme was rationalised.

Several *M. marinum* P450s that have close counterparts in *M. tuberculosis* were selected for comparison, in order to investigate whether the substrate and inhibitor binding affinities were preserved between species. The P450s investigated were analogues of the steroid metabolising P450s in *M. tuberculosis*. CYP125A6 and CYP125A7 have a single counterpart in *M. tuberculosis* (CYP125A1). The sequence identity and cholesterol binding affinity of CYP125A7 indicates it more closely resembles CYP125A1. However, CYP125A7 interacts differently to CYP125A1 with a range of inhibitors. CYP142A3 bound sterols with similar affinities as the *M. tuberculosis* CYP142A1. CYP124A1 from *M. marinum* was structurally characterised by X-ray crystallography, and showed a very closely preserved active site when compared to the *M. tuberculosis* analogue. These results suggest that individual P450 enzymes have maintained similar substrate specificities and roles between *Mycobacterium* species. However, for effective inhibitor design cross-species differences should be noted.

Chapter 1: Introduction

1.1: Cytochrome P450s

Monooxygenases are a diverse class of enzymes, responsible for the catalytic selective hydroxylation of a wide range of organic molecules by molecular oxygen. Cytochrome P450s (CYP or P450), a family of monooxygenase enzymes, were first isolated from pig liver cells in 1958, by Garfinkel [1] and Klingenberg [2] and identified as a haem containing protein by Omura and Sato in 1964 [3]. They are named after the characteristic Soret absorbance band they exhibit at 450 nm when the enzyme's haem cofactor is in the ferrous form and bound to a CO ligand (Figure 1). The haem cofactor consists of an iron centre, surrounded by four nitrogen ligands of a porphyrin ring, with the distal position occupied by a cysteine amino acid coordinating via the sulfur.

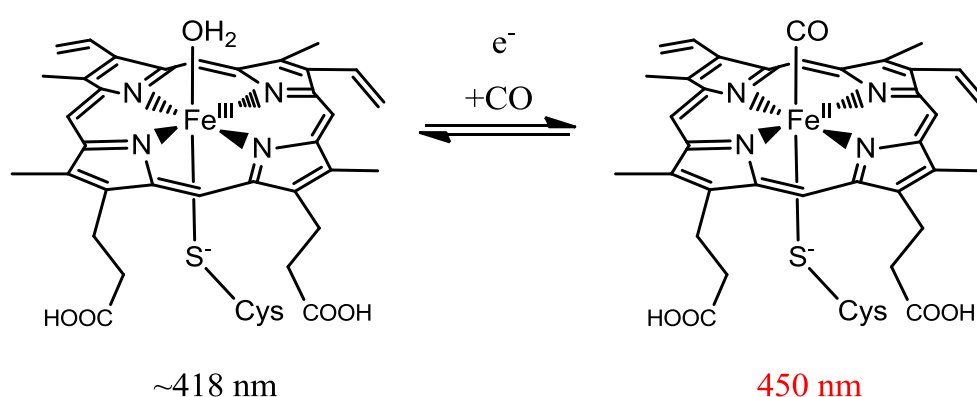


Figure 1: The haem b cofactor at the centre of cytochrome P450s. The six-coordinate resting state is the ferric thiolate form, where the sulfur atom of the cysteine residue is not protonated. The 450 nm shift of the Fe(II)-CO complex is due to the π -acceptor role of CO, which causes a shift in the Soret band that results from the porphyrin $\pi \rightarrow \pi^*$ transition [4].

Cytochrome P450s were gradually identified in many different organisms, including humans (which have 57 distinct P450 systems in the genome) [5], mice (102 different

P450s) [6] plants (457 in the tomato *Solanum lycopersicum*) [7], various bacteria (the antibiotic producing *Streptomyces coelicolor* has 18) [8], in some fungal strains (over 150 individual CYP sequences in some species) [9] and even viruses [10]. Their presence across all kingdoms of life indicates the enzyme family dates back to prokaryotic evolution [11]. Indeed, the most obvious evidence for this is the presence of analogues of a particular enzyme, CYP51, across kingdoms with a conserved function as a sterol α -demethylase [11-13]. The proliferation of diversity in the human xenobiotic metabolising CYP enzymes is thought to have arisen from the so-called animal plant warfare 400-800 million years ago [14]. There are now more than 350,000 unique CYP sequences known and the database of CYP names, maintained by Dr Nelson from University of Tennessee, contains over 41,000 sequences [15].

CYP enzymes are capable of very high substrate selectivity, and the family collectively oxidises a wide range of substrates (Figure 2). Within cells CYP enzymes hold a wide variety of key physiological roles, such as fatty acid hydroxylation [16], steroid synthesis [17] and drug metabolism [18]. Synthesis of complex natural products in bacterial cells is also often CYP mediated [19]. The importance of CYP enzymes to the evolution of life today is hard to overstate: CYP-mediated cholesterol synthesis indirectly made multicellularity possible, cutins and other plant-waterproofing molecules that allowed life to survive on land were synthesised by CYP fatty acid hydroxylases, which were also necessary, via lignin synthesis, for plants to support the weight of their own growth [20]. In plants especially, it is established that the diversity of secondary metabolites, including the vast number of different oxy-functionalised terpenes, is achieved in large part via CYPs from various families [21]. In humans a single enzyme, CYP3A4, interacts with an estimated 45-50% of therapeutic drugs, while CYP enzymes collectively perform 74% of all drug metabolism [22].

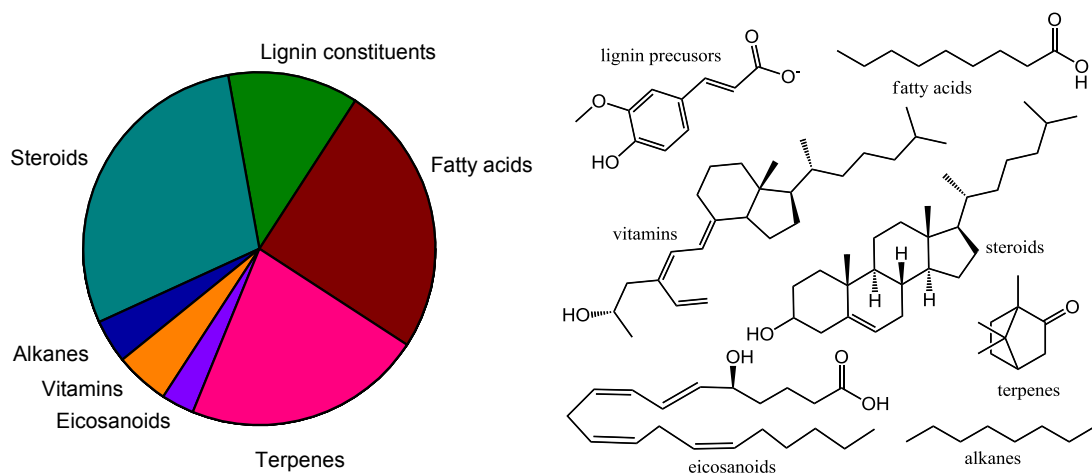


Figure 2: The main substrate class of CYPs from all kingdoms of life. Structural examples of each are included, showing the diversity of small molecule substrates. The analysis excludes xenobiotic metabolism as this function often overlaps with other substrate classes. Figure adapted from the analysis by Parvez *et al* [24] using substrate classes as described in [5].

In the early days of P450 study, individual enzymes were named based on the available information, which tended to be the substrate class or the species of origin. For example, the first bacterial CYP studied, P450_{cam}, was named for the known substrate, camphor [25]. As genome sequencing was developed and enzyme numbers grew, a more systematic nomenclature was developed [17]. Individual CYPs are now classified into families and sub-families. Based on amino acid sequence, enzymes with shared sequence homology of > 40% are classed into a family, which is given a number, and then, where sequence similarity is greater than 55%, into a subfamily, which is given a letter. CYP147G1, for example, is a member of the 147 family, and is the first member of the G sub-family. Similarity above 80% is sufficient for the two enzymes to share a systematic name [17]. Family numbers are based on the species of origin; for example bacterial families are numbered from 100 to 300 and then 1000 and above. Classifications do not necessarily indicate that enzyme function will be the same, as similarity is determined by the homology across the whole sequence, while substrate recognition is controlled by

active site residues [17]. It is therefore possible for the substrate binding site of two enzymes of the same family to be quite dissimilar, and functions within enzyme families are often quite diverse. For example, CYP107 family members have roles including fatty acid cleavage (CYP107H1) [26] and erythromycin biosynthesis (CYP107L) [27]. In contrast, all currently known members of the CYP153 family are alkane hydroxylases [28].

As the total known CYP sequences has grown, the number of conserved sequence elements has decreased. Highly conserved sequence elements such as the acid-alcohol pair in the I helix, the EXXR motif in the K helix and the haem binding motif (a Phe seven residues before the Cys) are all altered or missing in some CYP genes [29, 30]. This conserved Phe is thought to be necessary to maintain the reduction potential of the haem in the physiologically desirable range [31]. The last remaining conserved residue, the Cys residue that coordinates the haem, was found to be absent in the insect CYP408 family [29, 32]. The overall fold of the enzymes, however, is quite tightly conserved, even in the absence of these residues, consisting primarily of α -helices (40% α -helix and 10% β sheets in CYP101A1) [33, 34]. The individual helices referred to above are named based on a system developed using the crystal structure of CYP101A1 (Figure 3). A common and well-established feature of CYPs is the structural change they exhibit upon substrate binding [35-39]. In the 'open' form, the active site is solvent accessible; substrate binding triggers a change to the 'closed' state. The structural change is most significant in the regions of the F-G helices and the B-C loop, which together define the substrate binding channel. Changes to the position of these loops can alter the apparent shape of the active site significantly[35].

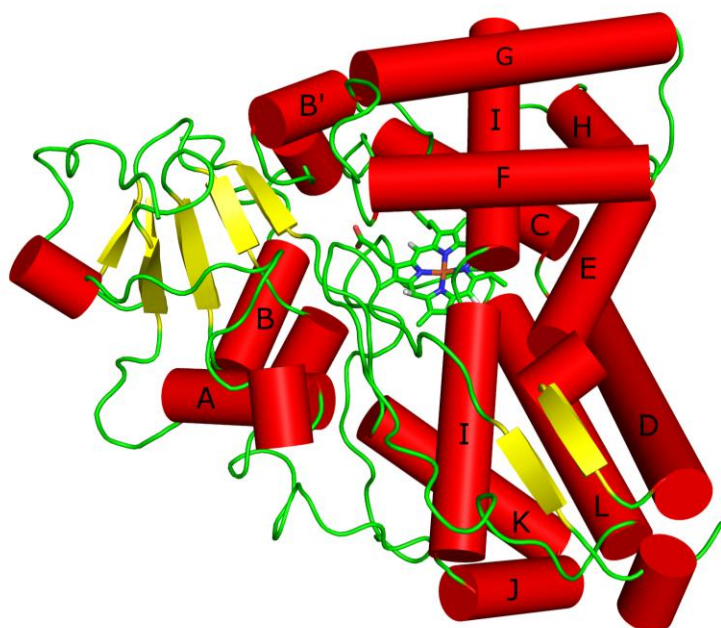


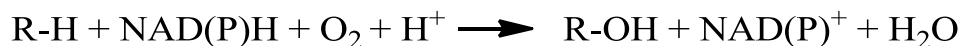
Figure 3: The structure of P450cam (CYP101A1). The enzyme was the first structurally characterised P450, demonstrating the overall fold of cytochrome P450s and the helix notation (PDB: 5CP4). The major helices are named A to L beginning from the N terminus. The prime labelled helices are those that in some P450s form a single helix (for example B and B'). The kink in the middle of the I helix near the haem is caused by a highly conserved alcohol residue (Thr252 in P450_{cam}) involved in oxygen activation [33].

1.2: The catalytic cycle

While P450s predominantly catalyse hydroxylation reactions, they are often also capable of further oxidation, converting the alcohol product to a ketone. They are known to catalyse reactions as wide ranging as C-C bond cleavage, epoxidation, dehydrogenation, and ring coupling [40]. This versatile metabolite production makes P450s desirable for enzyme-catalysed synthesis. The selective insertion of an oxygen atom into a carbon-hydrogen bond is difficult to achieve by standard synthetic methods [41].

In order to be catalytically active, P450s require the presence of the co-enzyme NADPH or NADH (henceforth NAD(P)H), which supplies its reducing equivalents to electron carrier proteins and ultimately to the P450.

The overall reaction of P450s is:



The catalytic cycle by which hydroxylation is achieved has been found to be consistent across CYP families with few exceptions (Figure 4)[42].

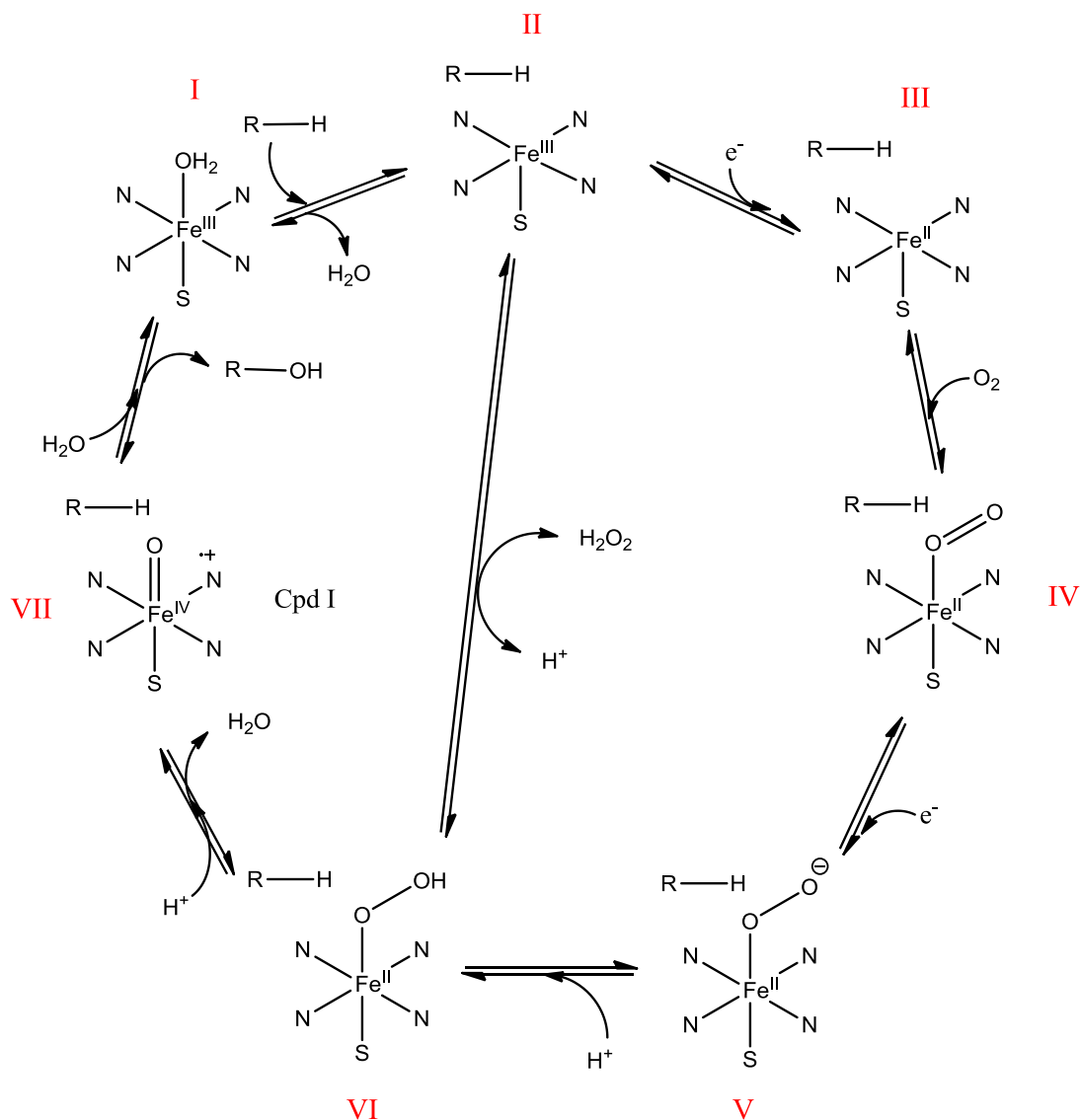


Figure 4: The catalytic cycle of cytochrome P450s (showing steps I → VII). The transition from VI → II is an uncoupling pathway. Some CYPs (known as peroxygenases) can achieve II → VI, known as the peroxide shunt.

In the resting state, a water molecule is the sixth ligand to the haem (I, Figure 4). The substrate displaces this and binds to the active site close to the ferric iron of the haem group (II). The first electron is then transferred to the haem, reducing Fe(III) to Fe(II), and enabling a dioxygen molecule to bind as the sixth ligand (IV). This is immediately followed by the second electron transfer step, which is accompanied by the protonation of the oxygen (VI). A second protonation then occurs, leading to the cleavage of the O-O bond and the release of water. This gives rise to the radical cation intermediate, Fe(IV) (Compound I, VII)[43], which undergoes the radical rebound mechanism (Figure 5), resulting in the insertion of the oxygen molecule into the C-H bond. The radical cation intermediate abstracts a hydrogen atom from the substrate (Figure 5). The released radical group reacts with the oxygen of Cpd II, transferring an electron back to the Fe(IV), which reduces the haem back to Fe(III). The newly formed product, R-OH, is then released and a water molecule binds to the ferric iron regenerating the resting state of the P450.

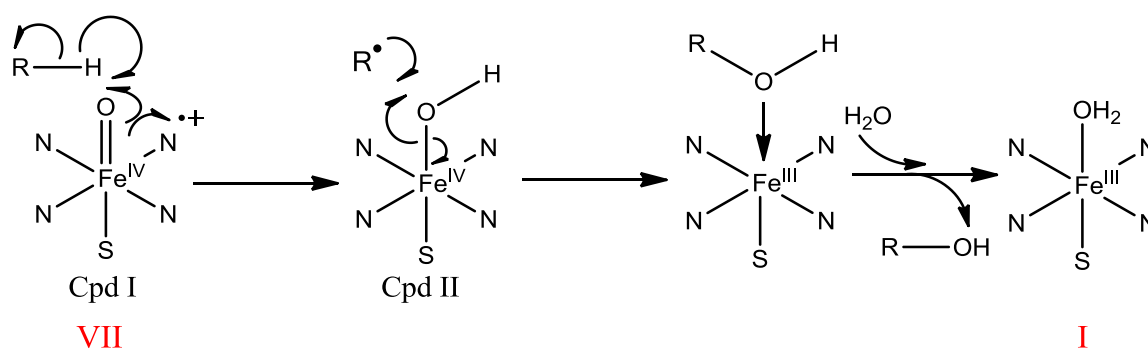


Figure 5: The radical recombination mechanism of the ferryl intermediate (Compound I) of the P450 (expansion of Step VII → I).

There are several pathways by which this cycle can breakdown, leading to so-called ‘uncoupling’, a lack of organic product and consumed reducing equivalents [44]. One such uncoupling pathway is shown in Figure 4, where the protonation event VI→VII occurs on the Fe-bound oxygen in the Fe(II)-OOH complex, leading to the release of

peroxide rather than water and returning the Fe to the ferric state. A subset of CYP enzymes can perform the reverse of this reaction, using H₂O₂ instead of NAD(P)H derived reducing equivalents to progress directly to the iron-oxo complex [45].

The substrate binding step I→II involves a transition from a six-coordinate Fe to a five coordinate state (Figure 6). The removal of the σ donating H₂O ligand reduces the ligand field splitting energy, and leads to a shift in spin state from low spin ($S = \frac{1}{2}$) to high spin ($S = \frac{5}{2}$) [46]. This shift can be monitored by UV-Visible absorbance spectroscopy as the Soret peak shifts (to 390 nm) [46].

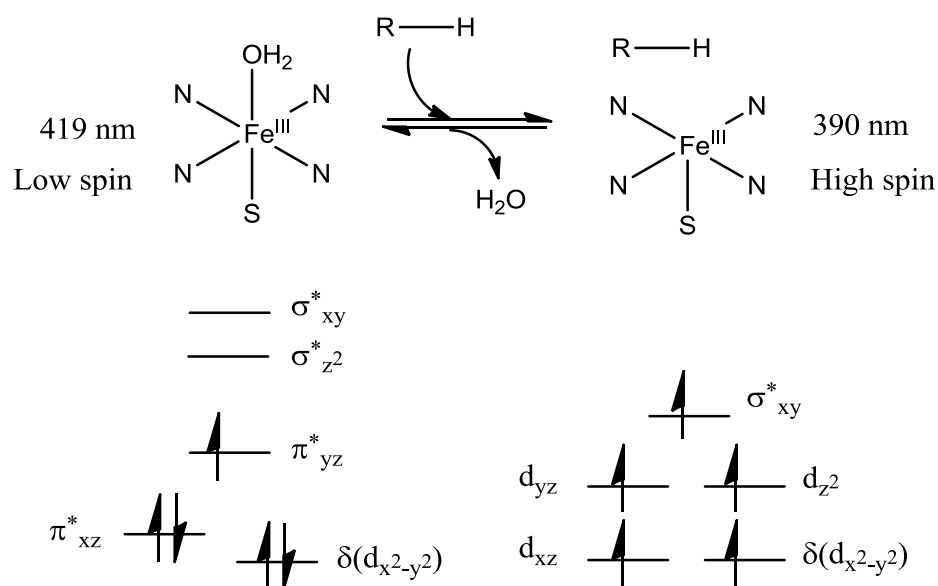


Figure 6: The effect of substrate binding on the coordination state of the Fe-haem complex and electron configurations for the 3d orbitals of the Fe(III). The resting state of the enzyme is low spin ($S = \frac{1}{2}$). The binding of the substrate in the active site displaces the sixth ligand, reducing the antibonding nature of the $\sigma_{z^2}^*$ and π_{yz}^* orbitals to give nonbonding d orbitals [47]. The weakened ligand field decreases the spacing of the d orbitals and leads to a transition to a high spin state ($S = \frac{5}{2}$) and a blue shift of the Soret peak to 390 nm.

Ligands such as nitrogen-containing azoles, cyanide or thiolates, many of which act as inhibitors to CYP enzymes, produce a difference spectra known as ‘Type II’ with

absorbance maxima between 422 – 435 nm (maintaining the low spin state, Figure 7)[48]. The precise wavelength is dependent on the extent to which the ligand acts as a σ -donor as well as π acceptor [49]. A third binding mode exists, where the substrate coordinates in the place of the water ligand via a hydroxyl group, maintaining the six-coordinate Fe. This results in a ‘reverse type I’ spectra, with an absorbance maxima of ~420 nm (Figure 7), which is again a low spin state [50]. With Type II and Reverse Type I binding modes the Fe remains in a low-spin state, which alters the reduction potential of the haem complex, and interrupts the subsequent electron transfer step [51]. Ligand interactions that interrupt critical steps in the catalytic cycle, such as the oxygen binding step (for example CO binding preferentially over O₂) or the reduction of Fe(III) to Fe(II) (an azole inhibitor), can competitively inhibit the enzyme [48].

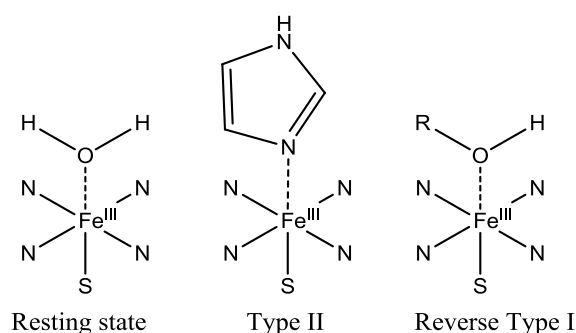


Figure 7: The binding modes of ligands to the haem. The examples include Type II (imidazole) and Reverse Type I (alcohol containing) ligands, as well as the resting state (water) with the Fe-porphyrin complex. Type I substrates do not coordinate the haem.

1.3: Electron transfer systems supporting cytochrome P450s

The electron transfer process is tightly controlled by the relative reduction potentials of the CYP enzyme to the electron transfer partner proteins as well as the complementary interactions between the electron transfer partner and the proximal face of the CYP enzyme. Class I ET systems are generally found in prokaryotes, where the P450 and its two electron transfer partners are soluble, while eukaryotic cells tend to use Class II

electron transfer systems, where the P450 and its electron transfer partners are membrane bound (Figure 8) [52]. Class I systems, in either bacterial or mammalian mitochondrial systems, have three components; a flavin adenine dinucleotide (FAD) dependent NAD(P)H ferredoxin reductase (FdR), an iron sulfur ferredoxin (often [2Fe-2S], Fdx), and the P450. The electron transfer chain begins with two electrons passing from the NAD(P)H to the FAD cofactor of the ferredoxin reductase, and then onward transfer, one electron at a time, to the iron-sulfur cluster of the ferredoxin. They are then shuttled to the P450 in two distinct steps, at the appropriate point in the catalytic cycle (Figure 4).

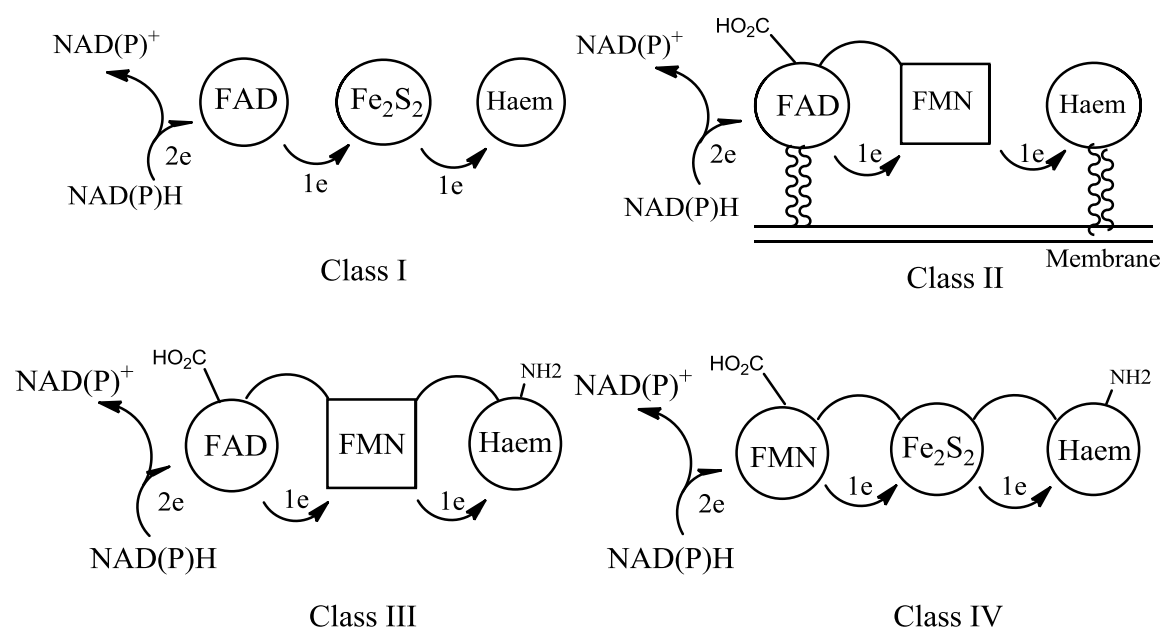


Figure 8: Varying electron transfer systems of P450 enzymes. Class I, III and IV are prokaryotic systems and Class II is eukaryotic.

Membrane bound Class II CYP systems feature a single membrane bound electron transfer partner, containing both FAD and FMN domains. Class III and IV are bacterial systems which feature fusion between the P450 and their ET system. In Class IV systems the equivalent of the reductase protein typically has two domains, with a flavin mononucleotide (FMN) and a [2Fe-2S] domain. Class III systems, the most studied

example of which is the enzyme P450_{BM3} (CYP102A1), are comprised of the same components as Class II systems, but all three are fused together via a polypeptide linker instead of being membrane bound.

In the CYP catalytic reaction cycle, binding of the substrate and the subsequent transition to the five-coordinate state triggers the first electron transfer. The thermodynamic explanation for the preferential reduction of the substrate-bound form is based on the change in formal potential of the haem when the coordination state changes from six to five (Figure 9) [51].

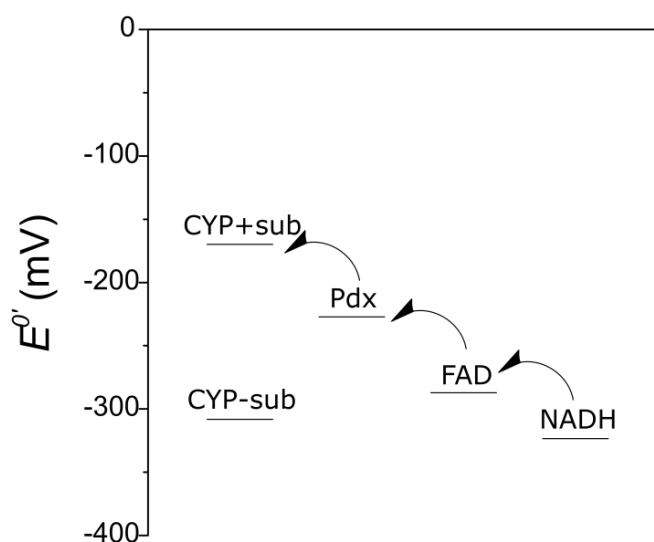


Figure 9: The formal reduction potentials (E°) of the species involved in the catalytic cycle of a Class 1 CYP electron transfer system. The change in reduction potential of CYP101A1 (P450_{cam}) in the substrate-free (-sub) and substrate-bound (+sub) forms dictates the favourability of the transition from the [2Fe-2S] ferredoxin Pdx to the substrate-bound form [53]. Arrows show the direction of electron transfer.

However, Honeychurch *et al* considered the thermodynamic barrier to be insufficient to prevent electron transfer to the substrate-free enzyme, as the speed of the subsequent O₂ binding step essentially couples the two reaction steps, leading to favourable reduction potentials for both the substrate and substrate-free forms in the presence of dioxygen [54].

Instead, they suggested an alternate kinetic model, in which the higher energy of reorganisation necessary for the reduction of the six-coordinate substrate-free Fe(III) to the five-coordinate Fe(II) is more likely to be the barrier to reduction in the absence of substrate [54]. In either model, it is the coordination state that dictates the favourability of the reduction step. This then explains the inhibitory effect of molecules that bind directly, in a Type II or Reverse Type I manner, as they maintain the same number of ligands to the Fe. Without the substrate binding step, the Fdx→CYP electron transfer is generally unfavourable *in vitro*, although both Guengerich and Johnson [55] and Munro *et al* [56] report reduction of CYP enzymes in the absence of substrate. There is evidence that futile cycling does occur *in vivo*, as demonstrated by Johnston *et al*, who found significant populations of reduced haem in intact cells in the absence of substrate [57].

1.4: Ferredoxins in bacterial cytochrome P450 systems

The electron transfer ability of a ferredoxin arises from the iron-sulfur cluster, which delocalises the electrons over the Fe and S atoms [58] before transferring them onwards to the P450. In each of the clusters, the iron atoms are ordinarily coordinated by the sulfur group of cysteine amino acids. The composition of the cluster (the variation in numbers of Fe and S atoms) and the ligands (in cases where they are not cysteines) can change the reduction potential of the protein. The cluster type varies between different ferredoxins, with [2Fe-2S] clusters being the most studied among those that support P450 activity (Figure 10) [59, 60]. [3Fe-4S] and [4Fe-4S] clusters are also known but less well characterised in the context of P450 electron transfer [61, 62].

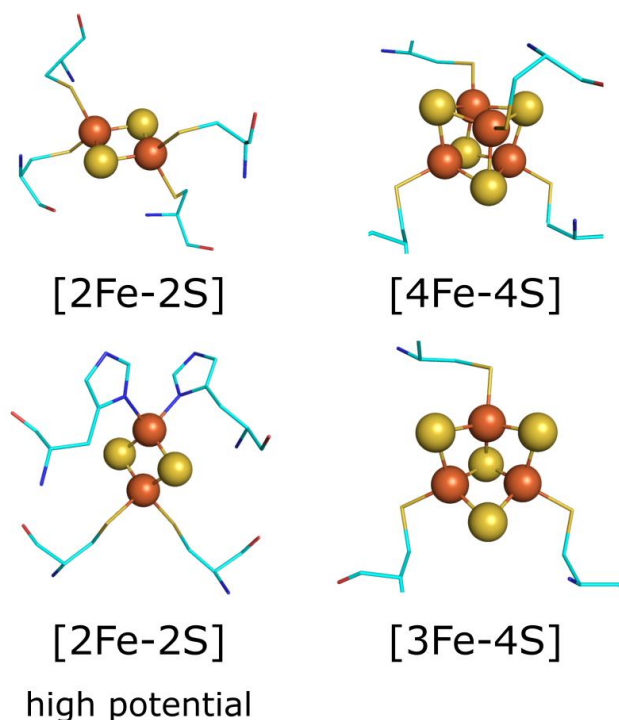
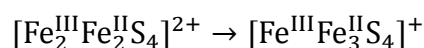


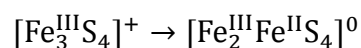
Figure 10: The structures of ferredoxin iron-sulfur cluster variants. They include [4Fe-4S], [3Fe-4S], [2Fe-2S] and Rieske high-potential [2Fe-2S], showing coordinating cysteine residues, and in the case of the Rieske [2Fe-2S], histidine residues. Iron: orange spheres; sulfide: yellow spheres.

The redox couple for each of the clusters differ, with $[2\text{Fe-2S}]^{2+/+}$, $[4\text{Fe-4S}]^{2+/+}$ and $[3\text{Fe-4S}]^{1+/0}$ being the physiologically relevant cluster oxidation state [63, 64]. In the [2Fe-2S] cluster this represents a transition where one of the two Fe(III) atoms is reduced to Fe(II).

The formal charge distribution over the atoms in a [4Fe-4S] cluster is:



with two mixed valent Fe(II)/Fe(III) pairs in the oxidised state and one homo-valent Fe(III) couple and a Fe(II)/Fe(III) couple in the reduced state ($S = 2$ in the oxidised state, $S = \frac{1}{2}$ when reduced). While in a [3Fe-4S] cluster it is:



with one high spin Fe(III) and a mixed valent Fe(II)/Fe(III) pair in the reduced state ($S = \frac{1}{2}$ in oxidised state, $S = 2$ when reduced). This unsurprisingly leads to differences in reduction potential between ferredoxins with different cluster types (Table 1). However, large differences in reduction potential can be seen even in clusters of the same type (see Table 1), which can only be a result of the protein architecture as the cluster geometry and compositions are conserved. It is not well understood how the protein environment affects this. Surface charge, hydrogen bonding and aromatic residues near the cluster have all been proposed as possible causes for the variation seen in reduction potential between ferredoxins of the same cluster type [64]. The reduction potential of the ferredoxin has an important effect on the rate of CYP reduction. The second electron transfer step has been found to be rate limiting for most CYP reactions under physiological conditions [53, 65]. Variation in the reduction potential of the physiological ferredoxins might be a mechanism of rate control or CYP-ferredoxin specificity in the native cell system.

Table 1: Approximate ranges of reduction potentials for iron-sulfur cluster containing ferredoxins. Rieske [2Fe-2S] high potential ferredoxins have histidine residues coordinating the Fe atoms in the cluster.

Cluster	Reduction Potential (mV)
[2Fe-2S]	-400 to -150
[3Fe-4S]	-203 to -85
[4Fe-4S]	-715 to -280
[2Fe-2S] high potential	+100 to +400 [66]

Ordinarily, the numbers of electron transfer partner proteins such as ferredoxins and ferredoxin reductases are lower than the number of CYPs in a given bacterial genome [67-69]. The lower numbers of electron transfer proteins infers that certain electron transfer systems must be able to support multiple CYPs. CYP enzymes, in particular

those from bacteria, often display high redox partner specificity [60]. The P450_{cam}/Pdx model system has been characterised in detail and serves as one example of this. The ferredoxin binding site has been determined to be the proximal face of the CYP, which is positively charged and can interact with the negative surface of the Fdx [70]. Pdx binding was first hypothesised to play an ‘effector role’ by Sligar *et al* who identified the Trp residue at the carboxylate terminus of Pdx as critical to the binding interaction to P450_{cam} [71]. Pochapsky *et al* showed that Pdx binding to the surface of P450_{cam} promoted a conformation change that prevented uncoupling [72]. Various spectral analyses demonstrated Pdx binding had an effect on the haem environment [73-76]. The theory has been further expanded on by the work of the Poulos group, who have demonstrated via crystal structures and analysis of the spin state that the binding of Pdx promotes the open form of the enzyme [70, 72]. This frees a key aspartate residue (Asp251) to participate in proton-coupled electron transfer. P450_{cam} is not active when supported with Arx, another ferredoxin from *Novosphingobium aromaticivorans* [77]. However, the structurally and functionally similar CYP101D1 (also from *N. aromaticivorans*) can be supported by both Arx and Pdx (albeit to a lesser extent with the latter) [60, 77], suggesting this strict ‘effector role’ may not be consistent across all CYPs [78]. There is even some suggestion that in rare cases the product formation of the CYP may be altered by the electron transfer protein pairing [79].

While bacterial CYP activity can sometimes be reconstituted with commercially available or alternative redox partners, this is almost always at low levels. Despite this, it is still often favoured over the process of testing large numbers of possible new CYP/Fdx/FdR pairings in Class I systems. Native ferredoxins in these systems are commonly used only when they are easily identified from the genome, for example when co-located with the CYP (eg. P450_{cam} with Pdx [80]). However co-location of the component proteins is not

universal, such as in *M. tuberculosis* [81] or *N. aromaticivorans* [68]. There have been few attempts to piece together the electron transfer systems of more complex CYPomes, as the method of expressing and measuring the activity of all possible proteins is labour-intensive. The work of Chun *et al* is one example, where the possible electron transfer partners of *S. coelicolor* A3(2) (six ferredoxins and four ferredoxin reductases) were tested for capacity to support CYP105D5 activity [82]. The breadth and variety of CYP electron transfer systems, including different components and electron sources [19], makes their study key to understanding P450s today as well as their evolution.

1.5: Bacterial CYPs in natural product synthesis pathways

Cytochrome P450s were initially thought of and studied primarily as xenobiotic metabolising enzymes, particularly common in animals and plants, and while P450s were found throughout all kingdoms of life, bacterial genomes commonly contained very few or none at all (for example *Escherichia coli* and *Salmonella typhimurium* both have none). However the sequencing of the first *Streptomyces* genomes involved in antibiotic production, *S. avermitilis* [83] and *S. coelicolor* [8], revealed larger numbers (33 and 18 CYPs, respectively). It was found that P450s are relatively abundant in Actinomycetes, and in *Streptomyces* species in particular. P450s have been frequently identified as part of biosynthetic operons, where they are generally responsible for late-stage functionalisation of natural products, commonly adding oxygen-containing or otherwise reactive groups to the scaffold of the product [84].

These natural products are sometimes referred to as secondary metabolites [85], which are more properly defined as any molecule that is not absolutely required for life. Secondary metabolites have functions such as signalling, competition, defence and nutrient transport. As a result of their inter- or intra-cellular roles, such natural products

often show biological activity when isolated, including antibiotic, antifungal and immunosuppressant activity [86, 87]. One example of this is himastatin, a natural product isolated from *Streptomyces hygroscopicus* which displays both antibiotic activity against gram-positive bacteria and antitumor activity in mice [88-90]. In the biosynthesis of himastatin, the CYP enzymes HmtT and HmtN both catalyse a stereoselective hydroxylation reaction on the cyclohexapeptide monomer (Figure 11). The HmtT hydroxylation triggers a subsequent cyclisation reaction promoted by neighbouring N group attack on the nearby C to the CYP-added electrophilic group. A third CYP, HmtS, catalyses the dimerisation reaction [91, 92]. The genes for these three enzymes form part of a 45 kbp gene cluster comprising 20 genes in total which encode a variety of other proteins that are involved with himastatin production [91]. The genomic location of such CYPs in known biosynthetic operons sometimes makes the discovery of the endogenous substrate more straightforward.

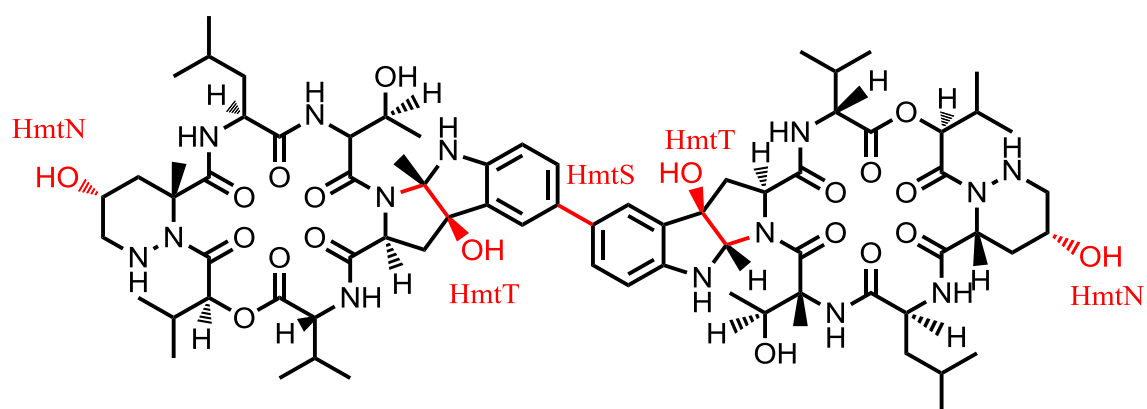


Figure 11: The structure of himastatin. CYP catalysed additions are shown in red.

Bacterial CYP enzymes typically perform C-H bond hydroxylations or epoxidations across alkene double bonds, but less common reactions have also been observed. In the example of himastatin, the C-N bond formation is achieved via the formation of an electrophilic group by HmtT hydroxylation. C-N bond formation has also been reported

where the catalytic cycle of the CYP is altered, as performed by TxtE from pathogenic *Streptomyces* spp. Instead of proceeding via hydroxylation, NO reacts with the Fe(III)-OO⁻ directly to form Fe(III)-OONO which is cleaved to release NO₂, followed by nitration mediated by the Fe(IV)=O species [93]. CYP enzymes have also been shown to be involved in C-S and C-C bond formation, often in the context of complexity adding reactions such as ring opening or closing. Uncommon P450 reactions vastly increase the variety of possible natural products [94] and often improve bioactivity [95].

In order to act on such a variety substrates as bacterial natural products, corresponding diversity in enzyme structure is required. The flexibility in the overall fold of CYP enzymes is sufficient to allow large differences in substrate size and chemistry. Bacterial CYP enzymes work on natural products with backbones including peptides, polyketides, saccharides and terpenoids [84, 95, 96]. The solved structures of bacterial P450s have offered examples of structural differences in CYPs such as additional recognition domains (the PCP-domain in P450_{sky} [97] or ACP recognition by P450_{Biol} [98]), 180° flipped haem group (in CYP121A1 [99], CYP154A1 [100] and several others) and even secondary active sites performing non-haem catalysed reactions (terpene synthase activity in CYP170A1 [101]) [84]. There are also examples of key sequence features missing from some enzymes, for example EXXR (CYP157C1 and similar enzymes) or the proximal Cys (CYP107AJ1 [102]). The majority of these bacterial systems are supported by Class I electron transfer systems (in an network analysis of *Streptomyces* CYPs, >2300 unique sequences were Class I while <400 were Class II/III [84]). The identification of the electron transfer partners has been more difficult as they are often remote from the biosynthetic gene cluster [84].

Consistent with their large numbers of CYPs, *Streptomyces* species are the source of many of the natural product biosynthetic gene clusters studied so far, although other Actinomycetes are also well represented (Figure 12) [84, 87, 103]. Natural products from the genus *Mycobacterium* include glycol-phospholipids [104, 105], non-ribosomal peptides [106] as well as polyketide [107] and terpene scaffolds [103]. Several of these synthesis pathways have been found to include CYP enzymes.

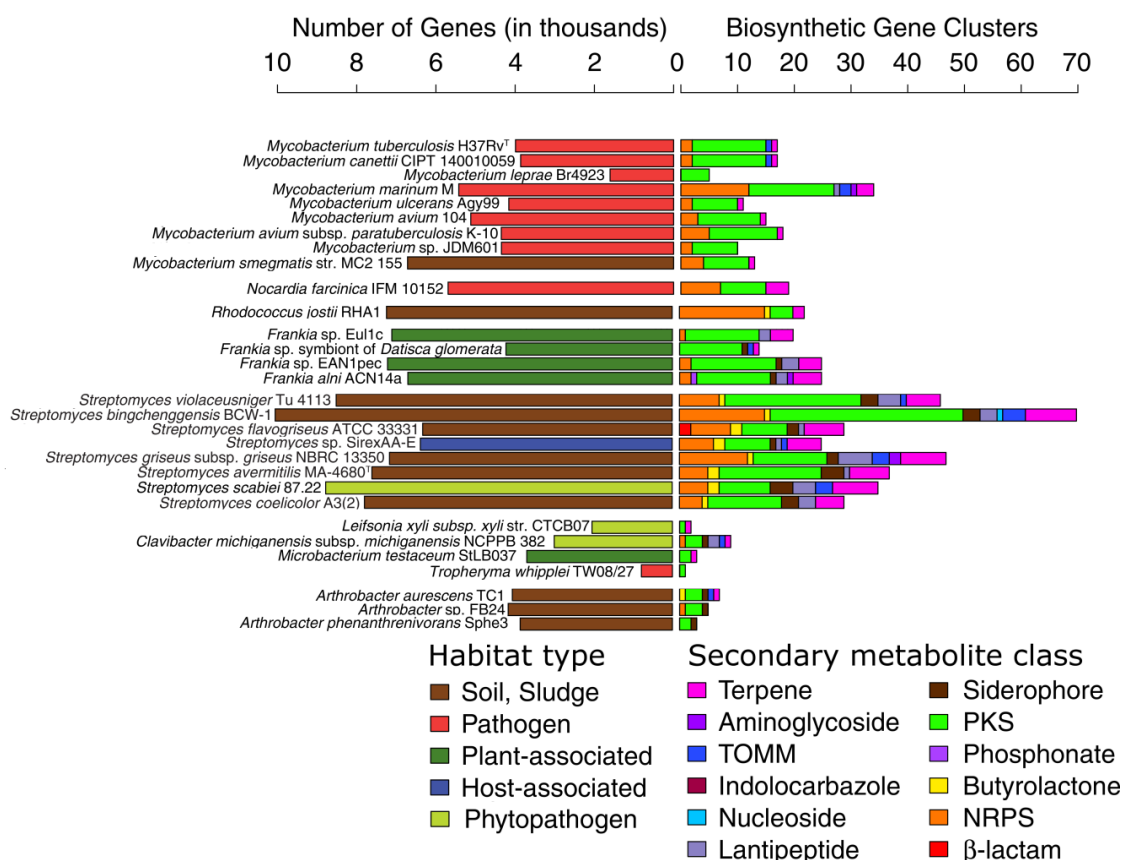


Figure 12: Biosynthetic gene clusters of various Actinomycete species. The analysis by Doroghazi and Metcalf shows the high numbers of biosynthetic gene clusters found in *Streptomyces* species, identified by the presence of a key biosynthetic protein. Note *Mycobacterium marinum* M has an unusually large number compared to the other *Mycobacterium* species, including the soil dwelling *Mycobacterium smegmatis* MC2-155. Colour of the bar on the left indicates habitat type, on the right indicates the class of secondary metabolite the gene cluster encodes. TOMM, thiol/oxazole-modified microcin; NRPS, non-ribosomal peptide synthetase; PKS, poly-ketide synthase. Reproduced with modifications from [103].

1.6: *Mycobacterial* CYPs: evolution and characterisation

Genome sequencing identified the species *M. marinum* as the closest genetic relative of *M. tuberculosis* outside of the *Mycobacterium tuberculosis* complex (MTBC), sharing 85% sequence similarity [106]. It is also a close relative of *Mycobacterium ulcerans*, with which it shares a genetic similarity of 97% [106, 108]. *M. marinum* is genetically similar to a hypothesised most recent common ancestor of pathogenic *Mycobacterium* species such as *M. tuberculosis* and *M. ulcerans* (Figure 13). It has undergone substantially less genome deletion than those species and retains the broad host range and ability of environmental survival (primarily a pathogen of frogs and fish causing the colloquially known ‘fish tuberculosis’, it is also capable of opportunistic infection in humans causing the disease aquarium granuloma) [106, 108]. *M. ulcerans* is the pathogen responsible for the Buruli ulcer [108] (also referred to as Bairnsdale or Daintree ulcer) which is a skin disease mostly located in tropical areas and common in central and western Africa and northern Australia [109]. *M. marinum* is commonly used as a model organism for *M. tuberculosis* as it is a much faster growing organism (doubling in ~4 hours rather than >20 leading to colony formation in a few days rather than 2-3 weeks) [110]. Importantly, it retains key *M. tuberculosis* cellular features such as the dense mycolic acid cell wall layer (discussed below) as well as virulence features such as asymptomatic latent infection and granuloma formation in the infected organism [110]. Zebra fish infections with *M. marinum* have advantages over the most common alternative models; offering faster growth than *M. bovis* and more pathogenic similarity to *M. tuberculosis* than the soil-growing *M. smegmatis* [111].

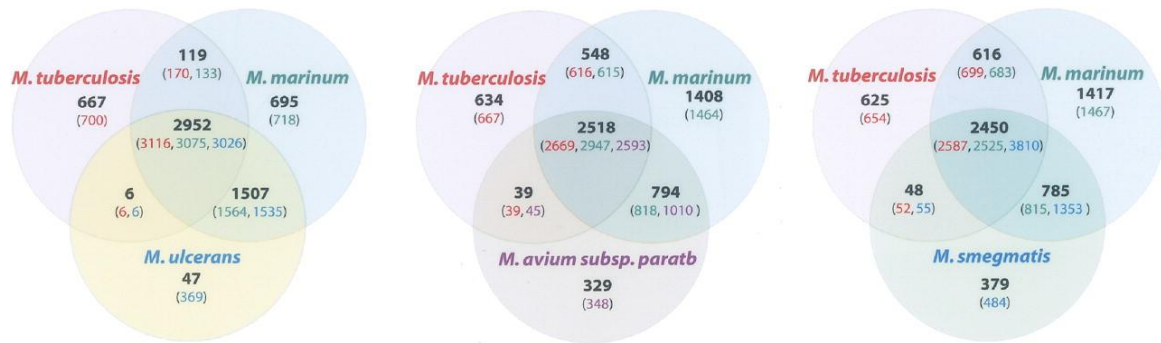


Figure 13: Venn diagrams showing the number of coding sequences that are shared between five Mycobacterial species. These are *M. tuberculosis*, *M. marinum*, *M. ulcerans*, *M. avium* subsp. *paratuberculosis* and *M. smegmatis*. Note that *M. marinum* has the highest number of shared genes with *M. tuberculosis*. Numbers in parenthesis also include paralogous (duplicated) sequences. Figure reproduced from Steinar *et al* [106] without modification.

As *M. marinum* is more flexible in host environment and has not relinquished the ability to survive outside the host, it is perhaps unsurprising that its genome offers more customisable virulence and survival mechanisms. As reviewed by Tobin and Ramakrishnan [110], genes such as the light induced β -carotene-producing *crtB*, which is important for protection from photo-oxidation damage and necessary only outside the host, are present in *M. marinum* but absent from *M. tuberculosis* [112, 113]. Additionally *M. marinum* has been shown to regulate polyketide lipids and sterols differently depending on host [114]. However, it is thought that the general mechanism of host colonisation is the same between *M. tuberculosis* and *M. marinum* and that pathogenicity evolved in the common ancestor prior to speciation [115]. Very few of the *M. tuberculosis* specific regions of the genome have been identified as fundamental to virulence [116, 117]. Instead, these genes, a large proportion of which were acquired via lateral transfer, are more likely relevant to host transmission and organ targeting [110]. *M. marinum* therefore contains the majority of genes of general importance to *M. tuberculosis* pathogenicity, as well as a large component of genetic material that was

discarded in specialisation (2.2 Mb) that is the key to understanding the speciation process.

The publication of the genome of *Mycobacterium tuberculosis* H37Rv, the pathogen responsible for human tuberculosis, revealed 20 distinct P450 genes sequences [118], which have subsequently been the object of much attention and research [81, 119]. The CYPs of *M. tuberculosis* belong to 19 different families and make up a total of 0.5% of the genome. *M. ulcerans* Agy99 contains 26, with an additional 9 pseudogenes. *M. marinum*, in contrast, contains 47 CYP encoding genes (0.9% of the genome), which belong to 36 different families. *Mycobacterium leprae* (the causal agent of human leprosy) contains only 1 CYP. Parvez *et al* report that the absolute number of CYPs in *Mycobacterial* genomes decreases in the following order; saprophytes (SAP), *Mycobacterium avium* complex (MAC), Non-tuberculous mycobacteria (NTM), *Mycobacterium chelonae-abscessus* complex (MCAC), and finally *Mycobacterium tuberculosis* complex (MTBC), from average of 50 CYPs down to 19 for MTBC species [24]. The same pattern is reflected in the CYP content in the genome, with some SAP and MAC species containing > 1% CYPs. MTBC species contain an average of 0.5% CYPs.

CYPs from *M. tuberculosis* and *M. ulcerans* are targets for inhibition of these pathogenic strains, as they are often involved in key metabolic roles [120, 121] rather than xenobiotic response. While 2018 marks 20 years since the publication of the genome of *M. tuberculosis*, progress on the characterisation of the CYPome has been relatively slow. Of the enzymes, five have now been structurally and functionally characterised (CYP51B1 [122, 123], CYP121A1 [107], CYP124A1 [124], CYP125A1 [125, 126], CYP142A1 [127]), while three (CYP126A1 [128], CYP130A1 [39] and CYP144A1 [129]) have been successfully crystallised but their substrates are still uncertain. Another, CYP128A1, has

a proposed role but no structural or biochemical information has been reported [130]. This leaves 11 ‘orphan’ enzymes for which there is still little to no information about.

Many of the CYP enzymes from *M. tuberculosis* have been crystallised in the substrate-free or inhibitor bound forms (Figure 14), but this has not lead directly to their biochemical characterisation or determination of their physiological function. Indeed, since the publication of the first structure of a P450, that of P450_{cam} in 1985 [34], there has been a proliferation in the number of structurally characterised CYPs. In recent years, the number of PDB entries for P450s has risen dramatically (numbering 855 in 2018). However, the conserved overall fold as well as the conformational change in the enzyme structure that follows substrate binding [131] often limits the use of this structural information.

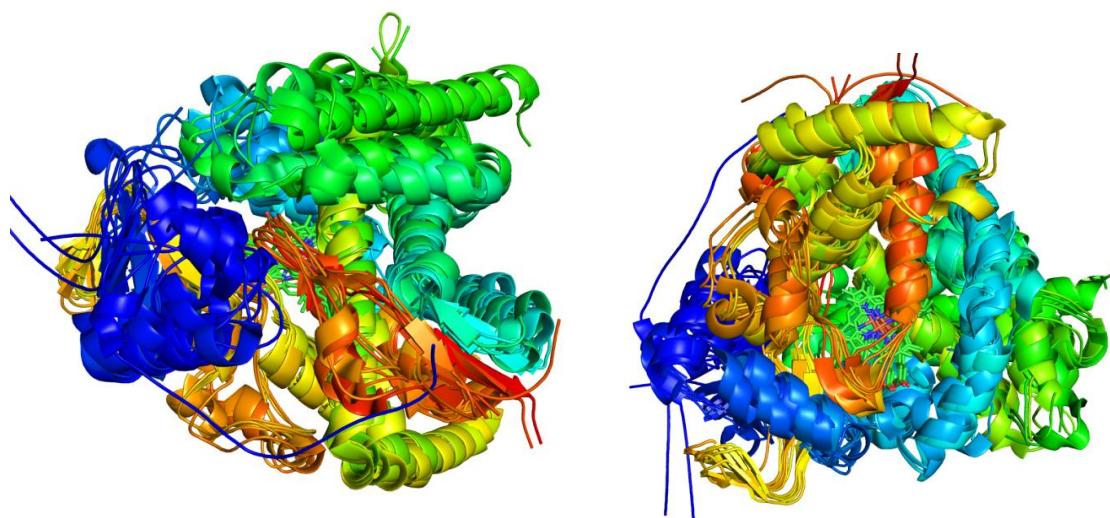


Figure 14: Overlay of the CYPs from *M. tuberculosis* that have been structurally characterised in the substrate-free state. Both distal and proximal views are shown, all coloured red to blue from N to C terminus. The conserved P450 fold is present in all. These enzymes catalyse distinct reactions such as fatty acid hydroxylation, cholesterol metabolism and carbon-carbon peptide cyclisation, and yet there are few obvious structural indicators of this diversity in the substrate-free state. PDB codes: 1H5Z, 1N40, 2WM5, 2X5L, 2XKR, 2UUQ and 5HDI [39, 99, 122, 124, 127, 129, 132].

Substrate library screening with common CYP substrates has produced some results, leading to the identification of a CYP130A1 catalysed N-demethylation reaction at low levels [133]. However, some of the enzymes catalyse highly specialised reactions, such as CYP121A1, which is present in *M. tuberculosis* and other members of the MTBC and forms an internal carbon-carbon bond in the cyclodipeptide cyclo(l-Tyr-l-Tyr) to generate a novel natural product (Figure 15) [107]. These highly specialised CYPs are difficult to characterise even by screening very large compound libraries. In the case of CYP121A1, the function was instead discovered via the characterisation of an upstream peptide synthetase, which produced the cyclodipeptide, after a period of initial study of the CYP enzyme had failed to identify a substrate [99, 134-136].

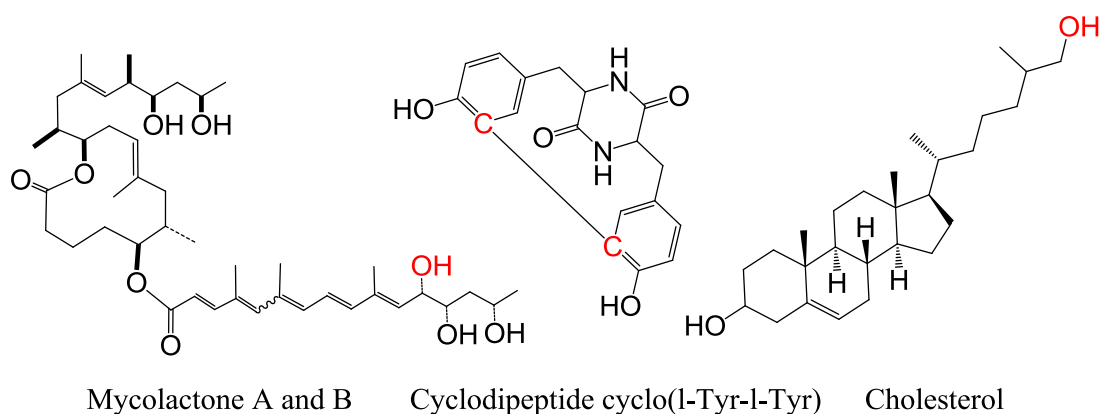


Figure 15: Products of the Mycobacterium CYP enzymes CYP140A7 from *M. ulcerans* [137], CYP121A1 [107] and CYP125A1 [125] from *M. tuberculosis*. Red indicates the bonds formed by the CYP.

Some indication as to whether a CYP enzyme might play a specialised role can be found via genome analysis. According to the analysis of Parvel *et al*, the members of the CYP121 family are very highly conserved between *Mycobacterial* species (with 391 amino acid residues conserved between 23 *Mycobacterial* family members, Figure 16), second only to the CYP141 family [24]. When non-*Mycobacterial* CYP sequences were also considered, expanding the analysis to over 17,000 CYP sequences from 113 families, the

CYP121 family (now with a total of 34 members) is still high among the most strictly conserved CYP families, after only the CYP141, CYP51 and CYP137 families. The degree of sequence homology between CYP sequences ordinarily correlates with the variation in substrate profile between the enzymes, so it would be expected that function would be more strictly conserved in enzyme families ranked highly in the list. Considered differently, in an enzyme with a broad substrate range an individual mutation is unlikely to destroy all substrate activity and so less strict amino acid sequence conservation is necessary. Hence, substrate variety in a given enzyme family allows more sequence diversity (and further proliferation). When the substrate profile is narrow, as with CYP121 and its single substrate, a greater number of the possibly occurring mutations would be deleterious to activity without corresponding gains in activity with other substrates.

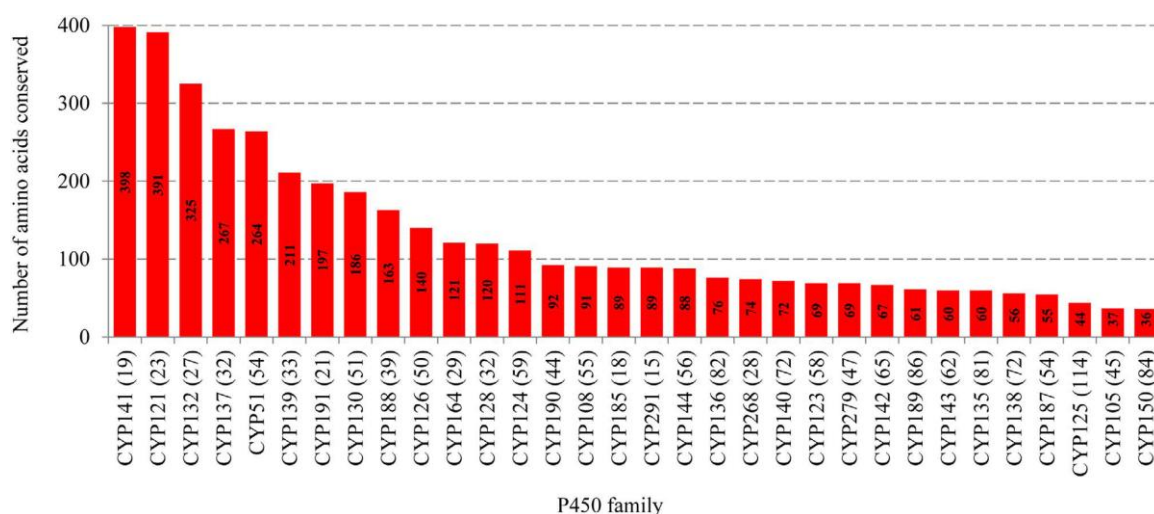


Figure 16: Conserved amino acids from Mycobacterial CYP families. The variation in the degree of conservation between families is shown. The number of P450s analysed from each family is in parenthesis next to the CYP family name. Reproduced from Parvez *et al* [24].

Another barrier to their study is the failure to produce several of the *Mycobacterial* CYPs in common *E. coli* expression systems. *Mycobacterial* genomes are renowned for the high G-C content which affects *E. coli* recombinant expression (as the bacteria has a relatively low G-C content), but even codon optimized gene sequences often produce little to no protein in such systems. Of the 20 enzymes, some 12 have never been reported as expressed and purified (despite the efforts of several laboratories).

Additionally, the reductive evolution that has decreased the number of CYPs in *M. tuberculosis* to 20 (instead of the 47 present in the NTB *M. marinum*) has likewise cut the number of associated electron transfer partners in the genome. Only two *M. tuberculosis* CYPs have closely located ferredoxins. These are CYP51, which has a [3Fe-4S] ferredoxin encoded by *Rv0763c* [123] and CYP143A1 which is close to *Rv1786* [138]. None have a closely located ferredoxin reductase. This has led to reliance on commercially available electron transfer partners or expressed non-physiological systems from other bacterial species. Although there have been some attempts to use *M. tuberculosis* ferredoxin systems for other CYPs (for instance Capyk *et al* with CYP125 [125]) it has met with limited success and there is still a general lack of understanding as to how electron transfer to *M. tuberculosis* CYPs is achieved *in vivo*. While many CYPs have low promiscuous activity with non-physiological ferredoxins, use of non-physiological electron transfer partners might nevertheless be the barrier to substrate oxidation in those that have thus far resisted characterisation.

Despite these difficulties, the study of *M. tuberculosis* CYPs has so far resulted in the identification of several viable enzyme drug targets [120, 139]. The above mentioned CYP121A1 has been subjected to structure-based inhibitor design as it has been found to be essential for organism survival *in vitro* [140-142]. Further, *M. tuberculosis* can survive

in vitro on cholesterol alone, which has been demonstrated to be due to the activity of the enzymes CYP125A1 and CYP142A1. Together these were found to be essential for cholesterol breakdown [125, 127]. In dual-knockout systems the intermediate 4-cholesten-3-one builds to toxic levels [126] and the organism can be inhibited by the addition of 4-cholesten-3-one synthesised with degradation resistant side chains [143].

CYP enzymes from other related *Mycobacterium* species have been studied to cast light on the evolutionary pathway to virulence, and in particular to explain the factors that promoted specialisation to the human host over the broad host-independent *Mycobacterium* ancestor. For example, the work of the Ortiz de Montellano group demonstrated that in soil living *M. smegmatis* the cholesterol breakdown system is mediated via three enzymes: two CYP125 family members and a CYP142 enzyme [144-146]. The system as a whole was conserved, but the third enzyme provided additional redundancy compared to the two in *M. tuberculosis*. It is particularly important to understand the roles of these CYPs in other *Mycobacterium* species, as conservation only in the MTBC is often taken to mean the enzyme plays a role in virulence or human infectivity, and vice versa [81]. However, if the roles of CYP enzymes conserved across species change between those species, then their conservation or lack thereof throughout *Mycobacteria* is less reliable as an indicator of their importance. Except for the *M. smegmatis* cholesterol system discussed above, CYP enzymes from non-MTBC species have thus far been investigated only incidentally or where they were known to belong to families of biocatalytic interest. The later includes the alkane hydroxylase CYP153A16 [147] which is the only *M. marinum* CYP to have been studied. From *M. vanbaalenii* PYR-1 CYP150 enzymes (polycyclic aromatic hydrolases) [148] and several members of the CYP151 family (saturated heterocyclic amine oxidation) have also been studied [149]. There are a number of entirely uncharacterised enzyme families (>20) present in *M.*

marinum that, given the variability of CYP enzymes studied from *Mycobacteria* so far, offer interesting possibilities for inhibitory, biocatalytic and mechanistic studies.

Many of the CYP families found in *Mycobacterium* can also be found in other actinobacterial species. Tracing this phylogenetic evolution of CYPs can explain physiological features of species, as well as the role of the CYPs themselves. For example, within Actinobacteria, the genera *Corynebacterium*, *Mycobacterium* and *Nocardia* form a taxon called CMN, which is characterised by the presence of the mycolic cell envelope (Figure 17) [150]. This is a layer of long chain branched fatty acids known as mycolic acids that covalently bind the peptidoglycan layer.

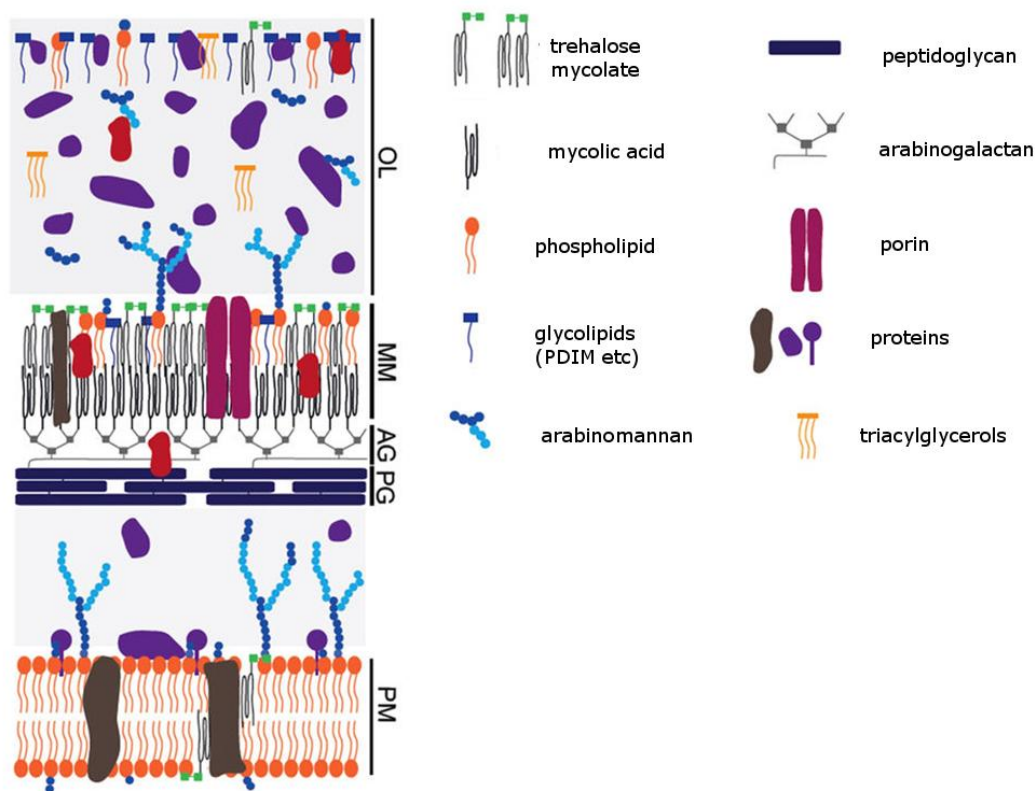


Figure 17: A proposed model of the cell wall of *M. smegmatis*. The mycolic acid layer (or mycomembrane, MM) is characteristic of all *Mycobacteria*. This is covalently attached to the peptidoglycan layer (PG) by arabinogalactan (AG). The phospholipid bilayer forms the plasma membrane (PM) and the outer leaflet (OL) is comprised of a variety of proteins and lipids. Note that glycolipids such as PDIM are present in the MM as well as the OL. Reproduced with modifications from Chiaradia *et al* [151].

The density of this wall in *M. tuberculosis* has been linked to the production of phthiocerol dimycoserates (PDIM), the production of which has in turn been linked to the cholesterol degradation activity of CYP125A1 [152]. Analogues of this enzyme are found in many *Mycobacterium* species, as well as *Nocardia* (at ~70% similarity) and even *Rhodococcus* but not *Corynebacterium*. Species to species analysis of the roles of these enzymes might rationalise such differences and explain their variant metabolisms, in particular those that act as human pathogens. Study of CYPs from *M. marinum* offers the possibility of insight into the evolution of the human-specific pathogens, the diversity of actinobacterial CYPs, as well as the identification of the roles of a number of as-yet uncharacterised CYP families.

1.7: Electron transfer in *Mycobacterium marinum*

The CYPome of *M. marinum* presents a further advantage with regards to the electron transfer systems. Due to the minimal number of gene deletions, *M. marinum* has a greater number of both ferredoxin and ferredoxin reductase proteins than either *M. tuberculosis* or *M. ulcerans*. This makes identifying the physiological electron transfer systems of these bacteria more achievable.

M. marinum has twelve ferredoxin gene sequences, of which one has been identified as encoding a [2Fe-2S]-type ferredoxin (*Mmar_3155*) [147]. This accompanies CYP153A16 (*Mmar_3154*) and a ferredoxin reductase gene (encoded by *Mmar_3153* and henceforth FdR2) [147]. A second apparent complete electron transfer system exists, with CYP147G1 (*Mmar_2930*) immediately beside both a ferredoxin and ferredoxin reductase FdR1 (*Mmar_2932* and *Mmar_2931*, respectively). This system contains one of the eleven sequences which have similarities to [3Fe-4S] or [4Fe-4S] ferredoxins. Nine of these, including *Mmar_2932*, are closely associated with a CYP in the genome (defined

here as within five genes distance). The co-location of these nine ferredoxin genes identifies them as the expected physiological ferredoxin for those systems and provides a first avenue for study. Given the larger number of CYP enzymes, it is likely that these ferredoxins will also support additional CYPs in the organism, although determining these relationships is a more substantial task.

Only three of these ferredoxins are conserved in *M. tuberculosis* (these are CYP51/*Rv0763c*, CYP143A1/*Rv1786* and a third, *Rv3503* which has no associated CYP). Importantly, neither of the ferredoxin reductase genes are conserved. The remainder of these uncharacterised, non-conserved ferredoxins in *M. marinum* are of greater importance as the electron transfer systems of CYPs from *M. tuberculosis* have proven difficult to establish in most cases. With the absence of closely located electron transfer partners, searching the *M. tuberculosis* genome has revealed other possible candidates. A *Mycobacterial* NADH-consuming FdR that couples *Rv0763c* and can support lanosterol demethylation by CYP51 has been reported [153]. There is also a NADPH-ferredoxin oxidoreductase with sequence homology to animal adrenodoxin reductase, FprA, which was shown to support both [2Fe-2S] and 7Fe ferredoxins from *M. smegmatis* [154] but it has not been established as capable of supporting CYP activity as yet. *M. tuberculosis* also has a number of other ferredoxins including the 7Fe FdxA type *Rv2007c* (conserved in *M. marinum* as *Mmar_2080*) and FdxC type *Rv1177* (conserved as *Mmar_4274*) however neither are near CYP genes. Additionally two apparent fusion FdR-ferredoxin proteins from *M. tuberculosis*, FprB and FdxB (*Rv0886* and *Rv3554* respectively) [155], have analogues in *M. marinum*.

The cluster type in each of the complement of 11 [3/4Fe-4S] ferredoxins in *M. marinum* is not immediately identifiable because they all contain an unusual variation on the

conserved sequence motif. Ordinarily, the iron-sulfur cluster binding sequence motif differs between [3Fe-4S] and [4Fe-4S] ferredoxins, being **CXXA/GXXC(X)_nCP** and **CXXCXXC(X)_nCP** respectively, and this is used to characterise the proteins [61]. The three cysteine residues in [3Fe-4S] ferredoxins and four in [4Fe-4S] ferredoxins each coordinate an iron residue in the iron-sulfur clusters. None of the eleven [3/4Fe-4S] ferredoxins of *M. marinum* have either a cysteine, alanine or glycine residue at the second conserved residue in the sequence motif (Figure 18). Rather, residues at that position include histidine (4), asparagine (3), tyrosine (1), serine (1), threonine (1) and phenylalanine (1) [156].

Fdx1	K	C	V	G	H	A	Q	C	Y	S	C	P	E	M	A
Fdx2	I	C	M	A	T	G	M	C	V	L	C	P	S	G	A
Fdx3	K	C	Q	Q	Y	A	Q	C	V	S	C	P	V	Q	A
Fdx4	L	C	E	A	N	G	F	C	E	Q	C	P	K	A	A
Fdx5	R	C	Q	Q	H	T	L	C	A	S	C	P	E	Q	A
Fdx6	V	C	D	G	F	G	L	C	A	D	C	P	A	H	A
Fdx7	L	C	E	S	N	G	I	C	M	Q	C	P	R	Q	A
Fdx8	R	C	R	G	H	G	M	C	L	C	C	P	E	Q	A
Fdx9	I	C	A	S	S	G	N	C	V	S	C	P	A	L	A
Fdx10	R	C	E	G	N	A	V	C	L	E	C	P	R	A	A
Fdx11	L	C	Q	Q	H	A	M	C	E	M	C	P	T	Q	A
[3Fe-4S] { S-coelicolor	R	C	V	G	A	G	Q	C	V	A	C	P	S	G	A
[4Fe-4S] { T-litoralis	A	C	I	G	C	G	V	C	A	S	C	P	T	G	A
C-thermoaceticum	L	C	I	A	C	G	T	C	I	E	C	P	T	E	A

CXX?XXC (X)_n CP

Figure 18: The novel ferredoxin sequences of *M. marinum* at the cluster binding motif. Typical [3Fe-4S] (*Streptomyces coelicolor* [67]) and [4Fe-4S] (*Thermococcus litoralis* [157] and *Clostridium thermoaceticum* [158]) ferredoxin sequences included for comparison. The non-standard amino acid is indicated in the blue box.

Of these residues, all bar phenylalanine could potentially coordinate an iron atom in the cluster. Precedent can be found for this in other systems. The [4Fe-4S] ferredoxin from the thermophilic *Pyrococcus furiosus* contains an aspartate residue that coordinates the fourth iron in the metal cluster and as demonstrated by Brereton *et al*, mutation of that residue to either cysteine or serine also leads to the isolation of a protein that maintained

the [4Fe-4S] cluster [159]. After cysteine, the most common residue coordinating iron in such clusters is histidine, as for example the histidine coordination in Rieske clusters (Figure 10). In the nickel-iron hydrogenase from *Desulfovibrio gigas*, one of the three iron-sulfur clusters is a [4Fe-4S] cluster with one Fe coordinated by a histidine residue [160]. In contrast, there are three examples of similar histidine-containing ferredoxins in CYP systems which have been characterised aerobically as containing [3Fe-4S] clusters. These are HaPuxC, from *Rhodopseudomonas palustris* (Figure 19) [156] and Rv0763c [123] and Rv1786 [138], both from *M. tuberculosis*. The crystal structure of HaPuxC shows the histidine residue (His17) lying above a [3Fe-4S] cluster. The imidazole ring of the histidine is directly above the sulfur indicating the position of this residue is quite similar even when a cysteine is not present. A comparison of the position of the residue can be made with the [4Fe-4S] ferredoxin from *P. furiosus* which has been structurally characterised with the aspartate residue mutated to a cysteine (D14C variant [161]).

Thus far histidine-containing [3Fe-4S] clusters have only been isolated under aerobic conditions. As Brereton *et al* demonstrated, the aspartate-iron coordination was highly sensitive to oxygen, isolated only by anaerobic purification [159]. Thus it is an open question whether the histidine-containing ferredoxins from *Mycobacterium* species may coordinate a fourth iron. The same applies for the other amino acids at that position in the cluster, where no comparative studies on similar enzymes are available.

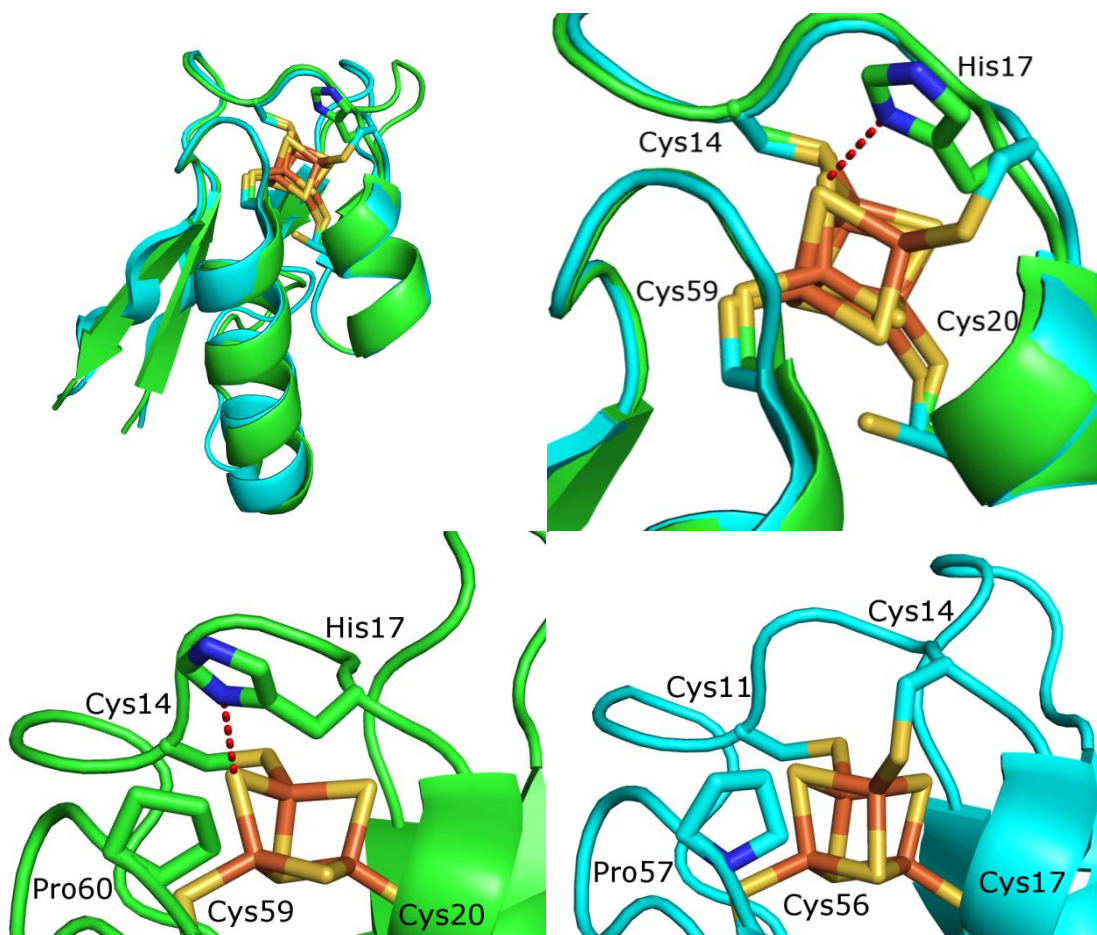


Figure 19: A comparison of the iron-sulfur clusters of two ferredoxins. The *R. palustris* ferredoxin HaPuxC, has a histidine residue (His17) at the CXX?XXC(X)_nCP position of the ferredoxin binding motif, and was characterised as containing [3Fe-4S] (green, PDB: 4OV1) [156] and the D14C (the equivalent position to His17 in HaPuxC) mutant of the [4Fe-4S] ferredoxin from *P. furiosus* (blue, PDB: 2Z8Q) [161] are overlaid. Sulfide ions and cysteine sulfurs are shown in gold and iron ions in orange. The distance between the N atom of the His17 residue and the sulfur atom is 3.6 Å (shown in red). Labels on the overlay of both structures are given for the coordinating residues of HaPuxC.

However, even if these residues do not directly bind a metal ion in the cluster, they may have a significant effect on the thermodynamics of the cluster. The closeness of the residue (as seen in Figure 19) means that hydrogen bonding, differences in charge or hydrophobic interactions near the cluster may change the reduction potential. Indeed, Rv0763c, the ferredoxin from *M. tuberculosis*, was found to have an unusually high

reduction potential of -31 mV [123], more positive than previously recorded for [3Fe-4S] clusters coupled with CYP enzymes. In theory, this may contribute a thermodynamic barrier to reduction of the coupled CYP, and thus a mechanism of modulation of the rate of monooxygenase activity. However, the effect of varying the residue at this position on the reduction potential, as well as the importance in electron transfer has not been determined. Characterisation of the role of the variable residue of the binding sequence, where it is not a cysteine, and identification of the proteins as either [3Fe-4S] or [4Fe-4S] ferredoxins would be beneficial to an overall understanding of the electron transfer mechanism and to the effect of reduction potential on P450 partner specificity. The *M. marinum* ferredoxins, presenting six different residues at this position, a large accompanying CYPome and two identified ferredoxin reductases, are therefore ideal candidates for study.

1.8: Summary and project goals

CYP enzymes are diverse in function and important in many biological systems. In bacteria they are responsible for catabolic and anabolic metabolism. Particularly in the synthesis of secondary metabolites they perform scaffold functionalisation and complexity addition with high specificity. In *Mycobacteria* they have been found to perform critical roles, cholesterol breakdown and fatty acid oxidation. This project aims to build on the existing knowledge of *Mycobacterium* CYP enzymes via study of those from the non-pathogenic close relative, *M. marinum*. In doing so, the electron transfer system of the species, containing as it does two full Class I systems and a set of as yet uncharacterised [3/4Fe-4S] ferredoxins, will also be studied. These ferredoxins have unusual cluster binding motifs that might affect their electron transfer ability. Hence the project has a dual focus, on uncharacterised *Mycobacterium* CYPs and their unusual physiological electron transfer systems.

The aims of the present study include:

- Purification of the eleven [3/4Fe-4S] ferredoxins of *Mycobacterium marinum* both aerobically and anaerobically.
- Biophysical characterisation of the ferredoxins, including determination of the composition of the iron-sulfur cluster in each of the ferredoxins, and the role of the non-standard amino acid at the cluster binding motif.
- Determination of the substrate and product range of *M. marinum* CYPs, as well as structural characterisation.
- Comparative study of conserved CYPs of *M. marinum* and *M. tuberculosis*, including substrate and inhibitor binding analysis.
- Phylogenic analysis of these CYP/ferredoxin systems across the bacterial kingdom.

References

- [1] D. Garfinkel, Studies on pig liver microsomes. I. Enzymic and pigment composition of different microsomal fractions, *Arch. Biochem. Biophys.*, 77 (1958) 493-509.
- [2] M. Klingenberg, Pigments of rat liver microsomes, *Arch. Biochem. Biophys.*, 75 (1958) 376-386.
- [3] T. Omura, R. Sato, The Carbon Monoxide-Binding Pigment of Liver Microsomes. I. Evidence for Its Hemoprotein Nature, *J. Biol. Chem.*, 239 (1964) 2370-2378.
- [4] L.K. Hanson, W.A. Eaton, S.G. Sligar, I.C. Gunsalus, M. Gouterman, C.R. Connell, Letter: Origin of the anomalous Soret spectra of carboxycytochrome P-450, *J. Am. Chem. Soc.*, 98 (1976) 2672-2674.
- [5] F.P. Guengerich, Human Cytochrome P450 Enzymes in: P.R. Ortiz de Montellano (Ed.) *Cytochrome P450: Structure, Mechanism, and Biochemistry*, Springer International Publishing, 2015, pp. 523-786.
- [6] D.R. Nelson, D.C. Zeldin, S.M. Hoffman, L.J. Maltais, H.M. Wain, D.W. Nebert, Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants, *Pharmacogenetics*, 14 (2004) 1-18.
- [7] The tomato genome sequence provides insights into fleshy fruit evolution, *Nature*, 485 (2012) 635-641.
- [8] S.D. Bentley, K.F. Chater, A.M. Cerdeno-Tarraga, G.L. Challis, N.R. Thomson, K.D. James, D.E. Harris, M.A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C.W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C.H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabinowitsch, M.A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B.G. Barrell, J. Parkhill, D.A. Hopwood, Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2), *Nature*, 417 (2002) 141-147.
- [9] H. Doddapaneni, R. Chakraborty, J.S. Yadav, Genome-wide structural and evolutionary analysis of the P450 monooxygenase genes (P450ome) in the white rot fungus *Phanerochaete chrysosporium*: evidence for gene duplications and extensive gene clustering, *BMC genomics*, 6 (2005) 92.
- [10] D.C. Lamb, L. Lei, A.G.S. Warrilow, G.I. Lepesheva, J.G.L. Mullins, M.R. Waterman, S.L. Kelly, The First Virally Encoded Cytochrome P450, *J. Virol*, 83 (2009) 8266-8269.
- [11] T. Rezen, N. Debeljak, D. Kordis, D. Rozman, New aspects on lanosterol 14 α -demethylase and cytochrome P450 evolution: lanosterol/cycloartenol diversification and lateral transfer, *J. Mol. Evol.*, 59 (2004) 51-58.
- [12] Y. Yoshida, M. Noshiro, Y. Aoyama, T. Kawamoto, T. Horiuchi, O. Gotoh, Structural and Evolutionary Studies on Sterol 14-Demethylase P450 (CYP51), the Most Conserved P450 Monooxygenase: II. Evolutionary Analysis of Protein and Gene Structures, *J. Biochemistry*, 122 (1997) 1122-1128.
- [13] D.R. Nelson, H.W. Strobel, Evolution of cytochrome P-450 proteins, *Mol. Biol. Evol.*, 4 (1987) 572-593.

- [14] D.F. Lewis, E. Watson, B.G. Lake, Evolution of the cytochrome P450 superfamily: sequence alignments and pharmacogenetics, *Mutation research*, 410 (1998) 245-270.
- [15] D.R. Nelson, Cytochrome P450 diversity in the tree of life, *Biochim. Biophys. Acta - Proteins and Proteomics*, 1866 (2018) 141-154.
- [16] S. Bhattarai, K. Liou, T.J. Oh, Hydroxylation of long chain fatty acids by CYP147F1, a new cytochrome P450 subfamily protein from *Streptomyces peucetius*, *Arch. Biochem. Biophys.*, 539 (2013) 63-69.
- [17] D.R. Nelson, L. Koymans, T. Kamataki, J.J. Stegeman, R. Feyereisen, D.J. Waxman, M.R. Waterman, O. Gotoh, M.J. Coon, R.W. Estabrook, I.C. Gunsalus, D.W. Nebert, P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature, *Pharmacogenetics*, 6 (1996) 1-42.
- [18] J.O. Miners, D.J. Birkett, Cytochrome P4502C9: an enzyme of major importance in human drug metabolism, *British journal of clinical pharmacology*, 45 (1998) 525-538.
- [19] S.L. Kelly, D.E. Kelly, Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us?, *Philos. Trans. Royal Soc. B*, 368 (2013) 20120476.
- [20] D.R. Nelson, A world of cytochrome P450s, *Philos. Trans. Royal Soc. B*, 368 (2013) 20120430.
- [21] B. Hamberger, S. Bak, Plant P450s as versatile drivers for evolution of species-specific chemical diversity, *Philos. Trans. Royal Soc. B*, 368 (2013) 20120426.
- [22] G. Luo, T. Guenther, L.S. Gan, W.G. Humphreys, CYP3A4 induction by xenobiotics: biochemistry, experimental methods and impact on drug discovery and development, *Curr. Drug Metab.*, 5 (2004) 483-505.
- [23] F.P. Guengerich, Human Cytochrome P450 Enzymes, in: P.R. Ortiz de Montellano (Ed.) *Cytochrome P450: Structure, Mechanism, and Biochemistry*, vol. 3, Kluwer Academic/Plenum Publishers, New York, 2005, pp. 377-530.
- [24] M. Parvez, L.B. Qhanya, N.T. Mthakathi, I.K.R. Kgosiemang, H.D. Bamal, N.S. Pagadala, T. Xie, H. Yang, H. Chen, C.W. Theron, R. Monyaki, S.C. Raseleman, V. Salewe, B.L. Mongale, R.G. Matowane, S.M.H. Abdalla, W.I. Booie, M. van Wyk, D. Olivier, C.E. Boucher, D.R. Nelson, J.A. Tuszyński, J.M. Blackburn, J.-H. Yu, S.S. Mashele, W. Chen, K. Syed, Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: Special focus on mycobacterial P450s, *Scientific Reports*, 6 (2016) 33099.
- [25] H. Koga, B. Rauchfuss, I.C. Gunsalus, P450cam gene cloning and expression in *Pseudomonas putida* and *Escherichia coli*, *Biochem. Biophys. Res. Commun.*, 130 (1985) 412-417.
- [26] M.J. Cryle, N.J. Matovic, J.J. De Voss, Products of cytochrome P450(BioI) (CYP107H1)-catalyzed oxidation of fatty acids, *Organic letters*, 5 (2003) 3341-3344.
- [27] J.R. Cupp-Vickery, T.L. Poulos, Structure of cytochrome P450eryF involved in erythromycin biosynthesis, *Nat. Struct. Biol.*, 2 (1995) 144-153.
- [28] S. Honda Malca, D. Scheps, L. Kuhnel, E. Venegas-Venegas, A. Seifert, B.M. Nestl, B. Hauer, Bacterial CYP153A monooxygenases for the synthesis of omega-hydroxylated fatty acids, *Chem. Commun.*, 48 (2012) 5115-5117.

- [29] D.C. Lamb, M.R. Waterman, Unusual properties of the cytochrome P450 superfamily, *Philos. Trans. Royal Soc. B*, 368 (2013) 20120434.
- [30] S. Rupasinghe, M.A. Schuler, N. Kagawa, H. Yuan, L. Lei, B. Zhao, S.L. Kelly, M.R. Waterman, D.C. Lamb, The cytochrome P450 gene family CYP157 does not contain EXXR in the K-helix reducing the absolute conserved P450 residues to a single cysteine, *FEBS letters*, 580 (2006) 6338-6342.
- [31] T.W. Ost, C.S. Miles, A.W. Munro, J. Murdoch, G.A. Reid, S.K. Chapman, Phenylalanine 393 exerts thermodynamic control over the heme of flavocytochrome P450 BM3, *Biochemistry*, 40 (2001) 13421-13429.
- [32] Y. Guo, J. Zhang, R. Yu, K.Y. Zhu, Y. Guo, E. Ma, Identification of two new cytochrome P450 genes and RNA interference to evaluate their roles in detoxification of commonly used insecticides in *Locusta migratoria*, *Chemosphere*, 87 (2012) 709-717.
- [33] T.L.J. Poulos, E. F., Structures of Cytochrome P450 Enzymes, in: P.R. Ortiz de Montellano (Ed.) *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 2015, pp. 3-32.
- [34] T.L. Poulos, B.C. Finzel, I.C. Gunsalus, G.C. Wagner, J. Kraut, The 2.6-Å crystal structure of *Pseudomonas putida* cytochrome P-450, *J. Biol. Chem.*, 260 (1985) 16122-16130.
- [35] H. Li, T.L. Poulos, The structure of the cytochrome P450Bm-3 haem domain complexed with the fatty acid substrate, palmitoleic acid, *Nat. Struct. Biol.*, 4 (1997) 140-146.
- [36] B. Zhao, F.P. Guengerich, A. Bellamine, D.C. Lamb, M. Izumikawa, L. Lei, L.M. Podust, M. Sundaramoorthy, J.A. Kalaitzis, L.M. Reddy, S.L. Kelly, B.S. Moore, D. Stec, M. Voehler, J.R. Falck, T. Shimada, M.R. Waterman, Binding of two flavin substrate molecules, oxidative coupling, and crystal structure of *Streptomyces coelicolor* A3(2) cytochrome P450 158A2, *J. Biol. Chem.*, 280 (2005) 11599-11607.
- [37] S.Y. Park, K. Yamane, S. Adachi, Y. Shiro, K.E. Weiss, S.A. Maves, S.G. Sligar, Thermophilic cytochrome P450 (CYP119) from *Sulfolobus solfataricus*: high resolution structure and functional properties, *J. Inorg. Biochem.*, 91 (2002) 491-501.
- [38] E.E. Scott, Y.A. He, M.R. Wester, M.A. White, C.C. Chin, J.R. Halpert, E.F. Johnson, C.D. Stout, An open conformation of mammalian cytochrome P450 2B4 at 1.6-Å resolution, *Proc. Natl. Acad. Sci. U. S. A.*, 100 (2003) 13196-13201.
- [39] H. Ouellet, L.M. Podust, P.R. de Montellano, *Mycobacterium tuberculosis* CYP130: crystal structure, biophysical characterization, and interactions with antifungal azole drugs, *J. Biol. Chem.*, 283 (2008) 5069-5080.
- [40] F.P. Guengerich, Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity, *Chem. Res. Toxicol.*, 14 (2001) 611-650.
- [41] F.P. Guengerich, Cytochrome P450 enzymes in the generation of commercial products, *Nature reviews. Drug discovery*, 1 (2002) 359-366.
- [42] B. Meunier, S.P. de Visser, S. Shaik, Mechanism of oxidation reactions catalyzed by cytochrome p450 enzymes, *Chemical reviews*, 104 (2004) 3947-3980.
- [43] J. Rittle, M.T. Green, Cytochrome P450 compound I: capture, characterization, and C-H bond activation kinetics, *Science*, 330 (2010) 933-937.

- [44] I.G. Denisov, S.G. Sligar, Activation of Molecular Oxygen in Cytochromes P450, in: P.R. Ortiz de Montellano (Ed.) *Cytochrome P450: Structure, Mechanism, and Biochemistry*, Springer International Publishing, Cham, 2015, pp. 69-109.
- [45] O. Shoji, Y. Watanabe, Peroxygenase reactions catalyzed by cytochromes P450, *J. Biol. Inorg. Chem.*, 19 (2014) 529-539.
- [46] A. Luthra, I.G. Denisov, S.G. Sligar, Spectroscopic features of cytochrome P450 reaction intermediates, *Arch. Biochem. Biophys.*, 507 (2011) 26-35.
- [47] S. Shaik, S. Cohen, Y. Wang, H. Chen, D. Kumar, W. Thiel, P450 Enzymes: Their Structure, Reactivity, and Selectivity—Modeled by QM/MM Calculations, *Chemical reviews*, 110 (2010) 949-1017.
- [48] M.A. Correia, P.F. Hollenberg, Inhibition of Cytochrome P450 Enzymes, in: P.R. Ortiz de Montellano (Ed.) *Cytochrome P450: Structure, Mechanism, and Biochemistry*, Springer International Publishing, Cham, 2015, pp. 177-259.
- [49] C.W. Locuson, J.M. Hutzler, T.S. Tracy, Visible spectra of type II cytochrome P450-drug complexes: evidence that "incomplete" heme coordination is common, *Drug metabolism and disposition: the biological fate of chemicals*, 35 (2007) 614-622.
- [50] J.B. Schenkman, D.L. Cinti, S. Orrenius, P. Moldeus, R. Kraschnitz, The nature of the reverse type I (modified type II) spectral change in liver microsomes, *Biochemistry*, 11 (1972) 4243-4251.
- [51] S.G. Sligar, D.L. Cinti, G.G. Gibson, J.B. Schenkman, Spin state control of the hepatic cytochrome P450 redox potential, *Biochem. Biophys. Res. Commun.*, 90 (1979) 925-932.
- [52] F. Hannemann, A. Bichet, K.M. Ewen, R. Bernhardt, Cytochrome P450 systems--biological variations of electron transport chains, *Biochim. Biophys. Acta*, 1770 (2007) 330-344.
- [53] D.F.V. Lewis, P. Hlavica, Interactions between redox partners in various cytochrome P450 systems: functional and structural aspects, *Biochim. Biophys. Acta (BBA) - Bioenergetics*, 1460 (2000) 353-374.
- [54] M.J. Honeychurch, H.A.O. Hill, L.-L. Wong, The thermodynamics and kinetics of electron transfer in the cytochrome P450cam enzyme system, *FEBS letters*, 451 (1999) 351-353.
- [55] F.P. Guengerich, W.W. Johnson, Kinetics of ferric cytochrome P450 reduction by NADPH-cytochrome P450 reductase: Rapid reduction in the absence of substrate and variations among cytochrome P450 systems, *Biochemistry*, 36 (1997) 14741-14750.
- [56] A.W. Munro, J.G. Lindsay, J.R. Coggins, S.M. Kelly, N.C. Price, NADPH oxidase activity of cytochrome P-450 BM3 and its constituent reductase domain, *BBA - Bioenergetics*, 1231 (1995) 255-264.
- [57] W.A. Johnston, D.J.B. Hunter, C.J. Noble, G.R. Hanson, J.E. Stok, M.A. Hayes, J.J. De Voss, E.M.J. Gillam, Cytochrome P450 Is Present in Both Ferrous and Ferric Forms in the Resting State within Intact *Escherichia coli* and Hepatocytes, *J. Biol. Chem.*, 286 (2011) 40750-40759.
- [58] D.C. Johnson, D.R. Dean, A.D. Smith, M.K. Johnson, Structure, function, and formation of biological iron-sulfur clusters, *Annual review of biochemistry*, 74 (2005) 247-281.

- [59] S.G. Bell, A. Dale, N.H. Rees, L.L. Wong, A cytochrome P450 class I electron transfer system from *Novosphingobium aromaticivorans*, *Appl. Microbiol. Biotechnol.*, 86 (2010) 163-175.
- [60] W. Yang, S.G. Bell, H. Wang, W. Zhou, N. Hoskins, A. Dale, M. Bartlam, L.L. Wong, Z. Rao, Molecular characterization of a class I P450 electron transfer system from *Novosphingobium aromaticivorans* DSM12444, *J. Biol. Chem.*, 285 (2010) 27372-27384.
- [61] W. Kaim, Iron-Sulfur and other non-heme iron proteins, in: B. Schwederski (Ed.) *Bioinorganic chemistry : inorganic elements in the chemistry of life* Wiley, Chichester ;, 1994.
- [62] K.J. McLean, D. Leys, A.W. Munro, Microbial Cytochromes P450, in: P.R. Ortiz de Montellano (Ed.) *Cytochrome P450: Structure, Mechanism, and Biochemistry*, Springer International Publishing, Cham, 2015, pp. 261-407.
- [63] F. Fabrizi de Biani, P. Zanello, The competition between chemistry and biology in assembling iron-sulfur derivatives. Molecular structures and electrochemistry. Part IV. {[Fe3S4](SγCys)3} proteins, *Inorganica Chimica Acta*, 455 (2017) 319-328.
- [64] P. Zanello, The competition between chemistry and biology in assembling iron-sulfur derivatives. Molecular structures and electrochemistry. Part V. {[Fe4S4](SCysγ)4} proteins, *Coordination Chemistry Reviews*, 335 (2017) 172-227.
- [65] Y. Imai, R. Sato, T. Iyanagi, Rate-Limiting Step in the Reconstituted Microsomal Drug Hydroxylase System1, *J. Biochem.*, 82 (1977) 1237-1246.
- [66] E.N. Brown, R. Friemann, A. Karlsson, J.V. Parales, M.M. Couture, L.D. Eltis, S. Ramaswamy, Determining Rieske cluster reduction potentials, *J. Biol. Inorg. Chem.*, 13 (2008) 1301-1313.
- [67] D.C. Lamb, T. Skaug, H.L. Song, C.J. Jackson, L.M. Podust, M.R. Waterman, D.B. Kell, D.E. Kelly, S.L. Kelly, The cytochrome P450 complement (CYPome) of *Streptomyces coelicolor* A3(2), *J. Biol. Chem.*, 277 (2002) 24000-24005.
- [68] S.G. Bell, A. Dale, N.H. Rees, L.-L. Wong, A cytochrome P450 class I electron transfer system from *Novosphingobium aromaticivorans*, *Appl. Microbiol. Biotechnol.*, 86 (2010) 163-175.
- [69] S.G. Bell, N. Hoskins, F. Xu, D. Caprotti, Z. Rao, L.L. Wong, Cytochrome P450 enzymes from the metabolically diverse bacterium *Rhodopseudomonas palustris*, *Biochem. Biophys. Res. Commun.*, 342 (2006) 191-196.
- [70] S. Tripathi, H. Li, T.L. Poulos, Structural basis for effector control and redox partner recognition in cytochrome P450, *Science*, 340 (2013) 1227-1230.
- [71] S.G. Sligar, P.G. Debrunner, J.D. Lipscomb, M.J. Namtvedt, I.C. Gunsalus, A Role of the Putidaredoxin COOH-terminus in P-450(cam) (Cytochrome m) Hydroxylations, *Proc. Natl. Acad. Sci. U. S. A.*, 71 (1974) 3906-3910.
- [72] S.S. Pochapsky, T.C. Pochapsky, J.W. Wei, A model for effector activity in a highly specific biological electron transfer complex: the cytochrome P450(cam)-putidaredoxin couple, *Biochemistry*, 42 (2003) 5649-5656.
- [73] M. Unno, J.F. Christian, T. Sjodin, D.E. Benson, I.D. Macdonald, S.G. Sligar, P.M. Champion, Complex formation of cytochrome P450cam with Putidaredoxin. Evidence for protein-specific interactions involving the proximal thiolate ligand, *J. Biol. Chem.*, 277 (2002) 2547-2553.

- [74] T. Tosha, S. Yoshioka, S. Takahashi, K. Ishimori, H. Shimada, I. Morishima, NMR study on the structural changes of cytochrome P450cam upon the complex formation with putidaredoxin. Functional significance of the putidaredoxin-induced structural changes, *J. Biol. Chem.*, 278 (2003) 39809-39821.
- [75] H. Shimada, S. Nagano, Y. Ariga, M. Unno, T. Egawa, T. Hishiki, Y. Ishimura, F. Masuya, T. Obata, H. Hori, Putidaredoxin-cytochrome p450cam interaction. Spin state of the heme iron modulates putidaredoxin structure, *J. Biol. Chem.*, 274 (1999) 9363-9369.
- [76] T. Tosha, S. Yoshioka, K. Ishimori, I. Morishima, L358P mutation on cytochrome P450cam simulates structural changes upon putidaredoxin binding: the structural changes trigger electron transfer to oxy-P450cam from electron donors, *J. Biol. Chem.*, 279 (2004) 42836-42843.
- [77] D. Batabyal, T.L. Poulos, Crystal structures and functional characterization of wild-type CYP101D1 and its active site mutants, *Biochemistry*, 52 (2013) 8898-8906.
- [78] D. Batabyal, A. Lewis-Ballester, S.R. Yeh, T.L. Poulos, A Comparative Analysis of the Effector Role of Redox Partner Binding in Bacterial P450s, *Biochemistry*, 55 (2016) 6517-6523.
- [79] W. Zhang, Y. Liu, J. Yan, S. Cao, F. Bai, Y. Yang, S. Huang, L. Yao, Y. Anzai, F. Kato, L.M. Podust, D.H. Sherman, S. Li, New reactions and products resulting from alternative interactions between the P450 enzyme and redox partners, *J. Am. Chem. Soc.*, 136 (2014) 3640-3646.
- [80] H. Koga, E. Yamaguchi, K. Matsunaga, H. Aramaki, T. Horiuchi, Cloning and Nucleotide Sequences of NADH-Putidaredoxin Reductase Gene (camA) and Putidaredoxin Gene (camB) Involved in Cytochrome P-450cam Hydroxylase of *Pseudomonas putida*, *J. Biochem.*, 106 (1989) 831-836.
- [81] H. Ouellet, J.B. Johnston, P.R. Ortiz de Montellano, The *Mycobacterium tuberculosis* cytochrome P450 system, *Arch. Biochem. Biophys.*, 493 (2010) 82-95.
- [82] Y.J. Chun, T. Shimada, R. Sanchez-Ponce, M.V. Martin, L. Lei, B. Zhao, S.L. Kelly, M.R. Waterman, D.C. Lamb, F.P. Guengerich, Electron transport pathway for a *Streptomyces* cytochrome P450: cytochrome P450 105D5-catalyzed fatty acid hydroxylation in *Streptomyces coelicolor* A3(2), *J. Biol. Chem.*, 282 (2007) 17486-17500.
- [83] H. Ikeda, J. Ishikawa, A. Hanamoto, M. Shinose, H. Kikuchi, T. Shiba, Y. Sakaki, M. Hattori, S. Ōmura, Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*, *Nature Biotechnology*, 21 (2003) 526.
- [84] J.D. Rudolf, C.Y. Chang, M. Ma, B. Shen, Cytochromes P450 for natural product biosynthesis in *Streptomyces*: sequence, structure, and function, *Nat. Prod. Rep.*, 34 (2017) 1141-1172.
- [85] D.H. Williams, M.J. Stone, P.R. Hauck, S.K. Rahman, Why Are Secondary Metabolites (Natural Products) Biosynthesized?, *J. Nat. Prod.*, 52 (1989) 1189-1208.
- [86] J. Masschelein, M. Jenner, G.L. Challis, Antibiotics from Gram-negative bacteria: a comprehensive overview and selected biosynthetic highlights, *Nat. Prod. Rep.*, 34 (2017) 712-783.
- [87] O. Genilloud, Actinomycetes: still a source of novel antibiotics, *Nat. Prod. Rep.*, 34 (2017) 1203-1232.

- [88] K.S. Lam, G.A. Hesler, J.M. Mattei, S.W. Mamber, S. Forenza, K. Tomita, Himastatin, a new antitumor antibiotic from *Streptomyces hygroscopicus*. I. Taxonomy of producing organism, fermentation and biological activity, *J Antibiotics*, 43 (1990) 956-960.
- [89] J.E. Leet, D.R. Schroeder, B.S. Krishnan, J.A. Matson, Himastatin, a new antitumor antibiotic from *Streptomyces hygroscopicus*. II. Isolation and characterization, *J. Antibiot.*, 43 (1990) 961-966.
- [90] J.E. Leet, D.R. Schroeder, J. Golik, J.A. Matson, T.W. Doyle, K.S. Lam, S.E. Hill, M.S. Lee, J.L. Whitney, B.S. Krishnan, Himastatin, a new antitumor antibiotic from *Streptomyces hygroscopicus*. III. Structural elucidation, *The J. Antibiot.*, 49 (1996) 299-311.
- [91] J. Ma, Z. Wang, H. Huang, M. Luo, D. Zuo, B. Wang, A. Sun, Y.Q. Cheng, C. Zhang, J. Ju, Biosynthesis of Himastatin: Assembly Line and Characterization of Three Cytochrome P450 Enzymes Involved in the Post-tailoring Oxidative Steps, *Angew. Chem. Int. Ed.*, 50 (2011) 7797-7802.
- [92] H. Zhang, J. Chen, H. Wang, Y. Xie, J. Ju, Y. Yan, H. Zhang, Structural analysis of HmtT and HmtN involved in the tailoring steps of himastatin biosynthesis, *FEBS letters*, 587 (2013) 1675-1680.
- [93] S.M. Barry, J.A. Kers, E.G. Johnson, L. Song, P.R. Aston, B. Patel, S.B. Krasnoff, B.R. Crane, D.M. Gibson, R. Loria, G.L. Challis, Cytochrome P450-catalyzed L-tryptophan nitration in thaxtomin phytotoxin biosynthesis, *Nat Chem Biol*, 8 (2012) 814-816.
- [94] X. Zhang, S. Li, Expansion of chemical space for natural products by uncommon P450 reactions, *Nat. Prod. Rep.*, 34 (2017) 1061-1089.
- [95] L.M. Podust, D.H. Sherman, Diversity of P450 enzymes in the biosynthesis of natural products, *Nat. Prod. Rep.*, 29 (2012) 1251-1266.
- [96] V. Ulrich, C. Brieke, M.J. Cryle, Biochemical and structural characterisation of the second oxidative crosslinking step during the biosynthesis of the glycopeptide antibiotic A47934, *Beilstein J. Org. Chem.*, 12 (2016) 2849-2864.
- [97] S. Uhlmann, R.D. Sussmuth, M.J. Cryle, Cytochrome p450sky interacts directly with the nonribosomal peptide synthetase to generate three amino acid precursors in skyllamycin biosynthesis, *ACS Chem. Biol.*, 8 (2013) 2586-2596.
- [98] M.J. Cryle, I. Schlichting, 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 (2008) 15696-15701.
- [99] D. Leys, C.G. Mowat, K.J. McLean, A. Richmond, S.K. Chapman, M.D. Walkinshaw, A.W. Munro, Atomic structure of *Mycobacterium tuberculosis* CYP121 to 1.06 Å reveals novel features of cytochrome P450, *J. Biol. Chem.*, 278 (2003) 5141-5147.
- [100] L.M. Podust, H. Bach, Y. Kim, D.C. Lamb, M. Arase, D.H. Sherman, S.L. Kelly, M.R. Waterman, Comparison of the 1.85 Å structure of CYP154A1 from *Streptomyces coelicolor* A3(2) with the closely related CYP154C1 and CYPs from antibiotic biosynthetic pathways, *Protein Sci.*, 13 (2004) 255-268.

- [101] B. Zhao, L. Lei, D.G. Vassylyev, X. Lin, D.E. Cane, S.L. Kelly, H. Yuan, D.C. Lamb, M.R. Waterman, Crystal Structure of Albaflavenone Monooxygenase Containing a Moonlighting Terpene Synthase Active Site, *J. Biol. Chem.*, 284 (2009) 36711-36719.
- [102] N.P. Niraula, B.K. Kanth, J.K. Sohng, T.J. Oh, Hydrogen peroxide-mediated dealkylation of 7-ethoxycoumarin by cytochrome P450 (CYP107AJ1) from *Streptomyces peucetius* ATCC27952, *Enzyme Microb Technol*, 48 (2011) 181-186.
- [103] J.R. Doroghazi, W.W. Metcalf, Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes, *BMC genomics*, 14 (2013) 611.
- [104] N.S. Li, L. Scharf, E.J. Adams, J.A. Piccirilli, Highly stereocontrolled total synthesis of beta-D-mannosyl phosphomycoketide: a natural product from *Mycobacterium tuberculosis*, *J. Org. Chem.*, 78 (2013) 5970-5986.
- [105] I. Matsunaga, A. Bhatt, D.C. Young, T.Y. Cheng, S.J. Eyles, G.S. Besra, V. Briken, S.A. Porcelli, C.E. Costello, W.R. Jacobs, Jr., D.B. Moody, *Mycobacterium tuberculosis* pks12 produces a novel polyketide presented by CD1c to T cells, *J. Exp. Med.*, 200 (2004) 1559-1569.
- [106] T.P. Stinear, T. Seemann, P.F. Harrison, G.A. Jenkin, J.K. Davies, P.D. Johnson, Z. Abdellah, C. Arrowsmith, T. Chillingworth, C. Churcher, K. Clarke, A. Cronin, P. Davis, I. Goodhead, N. Holroyd, K. Jagels, A. Lord, S. Moule, K. Mungall, H. Norbertczak, M.A. Quail, E. Rabbinowitsch, D. Walker, B. White, S. Whitehead, P.L. Small, R. Brosch, L. Ramakrishnan, M.A. Fischbach, J. Parkhill, S.T. Cole, Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*, *Genome research*, 18 (2008) 729-741.
- [107] P. Belin, M.H. Le Du, A. Fielding, O. Lequin, M. Jacquet, J.B. Charbonnier, A. Lecoq, R. Thai, M. Courcon, C. Masson, C. Dugave, R. Genet, J.L. Pernodet, M. Gondry, Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*, *Proc. Natl. Acad. Sci. U. S. A.*, 106 (2009) 7426-7431.
- [108] C. Demangel, T.P. Stinear, S.T. Cole, Buruli ulcer: reductive evolution enhances pathogenicity of *Mycobacterium ulcerans*, *Nature reviews. Microbiology*, 7 (2009) 50-60.
- [109] K. Nakanaga, R.R. Yotsu, Y. Hoshino, K. Suzuki, M. Makino, N. Ishii, Buruli ulcer and mycolactone-producing mycobacteria, *Jpn J Infect. Dis*, 66 (2013) 83-88.
- [110] D.M. Tobin, L. Ramakrishnan, Comparative pathogenesis of *Mycobacterium marinum* and *Mycobacterium tuberculosis*, *Cellular microbiology*, 10 (2008) 1027-1039.
- [111] M.U. Shiloh, P.A. Champion, To catch a killer. What can mycobacterial models teach us about *Mycobacterium tuberculosis* pathogenesis?, *Current opinion in microbiology*, 13 (2010) 86-92.
- [112] L. Ramakrishnan, H.T. Tran, N.A. Federspiel, S. Falkow, A crtB homolog essential for photochromogenicity in *Mycobacterium marinum*: isolation, characterization, and gene disruption via homologous recombination, *J. Bacteriol.*, 179 (1997) 5862-5868.
- [113] M.M. Mathews, Studies on the localization, function, and formation of the carotenoid pigments of a strain of *Mycobacterium marinum*, *Photochem. Photobiol.*, 2 (1963) 1-8.
- [114] E.M. Weerdenburg, A.M. Abdallah, F. Rangkuti, M. Abd El Ghany, T.D. Otto, S.A. Adroub, D. Molenaar, R. Ummels, K. Ter Veen, G. van Stempvoort, A.M. van der

Sar, S. Ali, G.C. Langridge, N.R. Thomson, A. Pain, W. Bitter, Genome-wide transposon mutagenesis indicates that *Mycobacterium marinum* customizes its virulence mechanisms for survival and replication in different hosts, *Infection and immunity*, 83 (2015) 1778-1788.

[115] T. Tonjum, D.B. Welty, E. Jantzen, P.L. Small, Differentiation of *Mycobacterium ulcerans*, *M. marinum*, and *M. haemophilum*: mapping of their relationships to *M. tuberculosis* by fatty acid profile analysis, DNA-DNA hybridization, and 16S rRNA gene sequence analysis, *J. Clin. Microbiol.*, 36 (1998) 918-925.

[116] J. Rengarajan, B.R. Bloom, E.J. Rubin, Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages, *Proc. Natl. Acad. Sci. U. S. A.*, 102 (2005) 8327-8332.

[117] C.M. Sasseti, E.J. Rubin, Genetic requirements for mycobacterial survival during infection, *Proc. Natl. Acad. Sci. U. S. A.*, 100 (2003) 12989-12994.

[118] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry, 3rd, F. Tekaiia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M.A. Quail, M.A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J.E. Sulston, K. Taylor, S. Whitehead, B.G. Barrell, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, *Nature*, 393 (1998) 537-544.

[119] K.J. McLean, D. Clift, D.G. Lewis, M. Sabri, P.R. Balding, M.J. Sutcliffe, D. Leys, A.W. Munro, The preponderance of P450s in the *Mycobacterium tuberculosis* genome, *Trends in microbiology*, 14 (2006) 220-228.

[120] S.A. Hudson, K.J. McLean, A.W. Munro, C. Abell, *Mycobacterium tuberculosis* cytochrome P450 enzymes: a cohort of novel TB drug targets, *Biochem. Soc. Trans.*, 40 (2012) 573-579.

[121] A.W. Munro, K.J. McLean, K.R. Marshall, A.J. Warman, G. Lewis, O. Roitel, M.J. Sutcliffe, C.A. Kemp, S. Modi, N.S. Scrutton, D. Leys, Cytochromes P450: novel drug targets in the war against multidrug-resistant *Mycobacterium tuberculosis*, *Biochem. Soc. Trans.*, 31 (2003) 625-630.

[122] L.M. Podust, L.V. Yermalitskaya, G.I. Lepesheva, V.N. Podust, E.A. Dalmasso, M.R. Waterman, Estriol bound and ligand-free structures of sterol 14 α -demethylase, *Structure (London, England : 1993)*, 12 (2004) 1937-1945.

[123] K.J. McLean, A.J. Warman, H.E. Seward, K.R. Marshall, H.M. Girvan, M.R. Cheesman, M.R. Waterman, A.W. Munro, Biophysical characterization of the sterol demethylase P450 from *Mycobacterium tuberculosis*, its cognate ferredoxin, and their interactions, *Biochemistry*, 45 (2006) 8427-8443.

[124] J.B. Johnston, P.M. Kells, L.M. Podust, P.R. Ortiz de Montellano, Biochemical and structural characterization of CYP124: a methyl-branched lipid omega-hydroxylase from *Mycobacterium tuberculosis*, *Proc. Natl. Acad. Sci. U. S. A.*, 106 (2009) 20687-20692.

[125] J.K. Capyk, R. Kalscheuer, G.R. Stewart, J. Liu, H. Kwon, R. Zhao, S. Okamoto, W.R. Jacobs, Jr., L.D. Eltis, W.W. Mohn, Mycobacterial cytochrome p450 125 (*cyp125*)

- catalyzes the terminal hydroxylation of c27 steroids, *J. Biol. Chem.*, 284 (2009) 35534-35542.
- [126] H. Ouellet, S. Guan, J.B. Johnston, E.D. Chow, P.M. Kells, A.L. Burlingame, J.S. Cox, L.M. Podust, P.R. de Montellano, *Mycobacterium tuberculosis* CYP125A1, a steroid C27 monooxygenase that detoxifies intracellularly generated cholest-4-en-3-one, *Molec. Microbiol.*, 77 (2010) 730-742.
- [127] M.D. Driscoll, K.J. McLean, C. Levy, N. Mast, I.A. Pikuleva, P. Lafite, S.E. Rigby, D. Leys, A.W. Munro, Structural and biochemical characterization of *Mycobacterium tuberculosis* CYP142: evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen, *J. Biol. Chem.*, 285 (2010) 38270-38282.
- [128] J.T. Chenge, L.V. Duyet, S. Swami, K.J. McLean, M.E. Kavanagh, A.G. Coyne, S.E. Rigby, M.R. Cheesman, H.M. Girvan, C.W. Levy, B. Rupp, J.P. von Kries, C. Abell, D. Leys, A.W. Munro, Structural characterization and ligand/inhibitor identification provide functional insights into the *Mycobacterium tuberculosis* cytochrome P450 CYP126A1, *J. Biol. Chem.*, 292 (2017) 1310-1329.
- [129] J. Chenge, M.E. Kavanagh, M.D. Driscoll, K.J. McLean, D.B. Young, T. Cortes, D. Matak-Vinkovic, C.W. Levy, S.E. Rigby, D. Leys, C. Abell, A.W. Munro, Structural characterization of CYP144A1 - a cytochrome P450 enzyme expressed from alternative transcripts in *Mycobacterium tuberculosis*, *Scientific reports*, 6 (2016) 26628.
- [130] K.M. Sogi, C.M. Holsclaw, G.K. Fragiadakis, D.K. Nomura, J.A. Leary, C.R. Bertozzi, Biosynthesis and regulation of sulfomenaquinone, a metabolite associated with virulence in *Mycobacterium tuberculosis*, *ACS infectious diseases*, 2 (2016) 800-806.
- [131] T.L. Poulos, B.C. Finzel, A.J. Howard, Crystal structure of substrate-free *Pseudomonas putida* cytochrome P-450, *Biochemistry*, 25 (1986) 5314-5322.
- [132] H. Ouellet, P.M. Kells, P.R. Ortiz de Montellano, L.M. Podust, Reverse type I inhibitor of *Mycobacterium tuberculosis* CYP125A1, *Bioorg. Med. Chem. Lett.*, 21 (2011) 332-337.
- [133] S. Ortega Ugalde, R.A. Luirink, D.P. Geerke, N.P.E. Vermeulen, W. Bitter, J.N.M. Commandeur, Engineering a self-sufficient *Mycobacterium tuberculosis* CYP130 by gene fusion with the reductase-domain of CYP102A1 from *Bacillus megaterium*, *J. Inorg. Biochem.*, 180 (2018) 47-53.
- [134] K.J. McLean, M.R. Cheesman, S.L. Rivers, A. Richmond, D. Leys, S.K. Chapman, G.A. Reid, N.C. Price, S.M. Kelly, J. Clarkson, W.E. Smith, A.W. Munro, Expression, purification and spectroscopic characterization of the cytochrome P450 CYP121 from *Mycobacterium tuberculosis*, *J. Inorg. Biochem.*, 91 (2002) 527-541.
- [135] C.G. Mowat, D. Leys, K.J. McLean, S.L. Rivers, A. Richmond, A.W. Munro, M. Ortiz Lombardia, P.M. Alzari, G.A. Reid, S.K. Chapman, M.D. Walkinshaw, Crystallization and preliminary crystallographic analysis of a novel cytochrome P450 from *Mycobacterium tuberculosis*, *Acta crystallographica. Section D*, 58 (2002) 704-705.
- [136] H.E. Seward, A. Roujeinikova, K.J. McLean, A.W. Munro, D. Leys, Crystal structure of the *Mycobacterium tuberculosis* P450 CYP121-fluconazole complex reveals new azole drug-P450 binding mode, *J. Biol. Chem.*, 281 (2006) 39437-39443.
- [137] H. Hong, T. Stinear, J. Porter, C. Demangel, P.F. Leadlay, A Novel Mycolactone Toxin Obtained by Biosynthetic Engineering, *Chembiochem*, 8 (2007) 2043-2047.

- [138] Y. Lu, F. Qiao, Y. Li, X.H. Sang, C.R. Li, J.D. Jiang, X.Y. Yang, X.F. You, Recombinant expression and biochemical characterization of *Mycobacterium tuberculosis* 3Fe-4S ferredoxin Rv1786, *Appl. Microbiol. Biotechnol.*, (2017).
- [139] P.R. Ortiz de Montellano, Potential drug targets in the *Mycobacterium tuberculosis* cytochrome P450 system, *J. Inorg. Biochem.*, 180 (2018) 235-245.
- [140] K.J. McLean, P. Carroll, D.G. Lewis, A.J. Dunford, H.E. Seward, R. Neeli, M.R. Cheesman, L. Marsollier, P. Douglas, W.E. Smith, I. Rosenkrands, S.T. Cole, D. Leys, T. Parish, A.W. Munro, Characterization of active site structure in CYP121. A cytochrome P450 essential for viability of *Mycobacterium tuberculosis* H37Rv, *J. Biol. Chem.*, 283 (2008) 33406-33416.
- [141] M.E. Kavanagh, A.G. Coyne, K.J. McLean, G.G. James, C.W. Levy, L.B. Marino, L.P.S. de Carvalho, D.S.H. Chan, S.A. Hudson, S. Surade, D. Leys, A.W. Munro, C. Abell, Fragment-Based Approaches to the Development of *Mycobacterium tuberculosis* CYP121 Inhibitors, *J. Med. Chem.*, 59 (2016) 3272-3302.
- [142] S.A. Hudson, K.J. McLean, S. Surade, Y.Q. Yang, D. Leys, A. Ciulli, A.W. Munro, C. Abell, Application of Fragment Screening and Merging to the Discovery of Inhibitors of the *Mycobacterium tuberculosis* Cytochrome P450 CYP121, *Angew. Chem. Int. Ed.*, 51 (2012) 9311-9316.
- [143] D.J. Frank, Y. Zhao, S.H. Wong, D. Basudhar, J.J. De Voss, P.R. Ortiz de Montellano, Cholesterol analogs with degradation-resistant alkyl side chains are effective *Mycobacterium tuberculosis* growth inhibitors, *J. Biol. Chem.*, 291 (2016) 7325-7333.
- [144] E. Garcia-Fernandez, D.J. Frank, B. Galan, P.M. Kells, L.M. Podust, J.L. Garcia, P.R. Ortiz de Montellano, A highly conserved mycobacterial cholesterol catabolic pathway, *Environ. Microbiol.*, 15 (2013) 2342-2359.
- [145] D.J. Frank, Y. Madrona, P.R. Ortiz de Montellano, Cholesterol ester oxidation by mycobacterial cytochrome P450, *J. Biol. Chem.*, 289 (2014) 30417-30425.
- [146] D.J. Frank, C.A. Waddling, M. La, P.R. Ortiz de Montellano, Cytochrome P450 125A4, the third cholesterol C-26 hydroxylase from *Mycobacterium smegmatis*, *Biochemistry*, 54 (2015) 6909-6916.
- [147] D. Scheps, S.H. Malca, H. Hoffmann, B.M. Nestl, B. Hauer, Regioselective omega-hydroxylation of medium-chain n-alkanes and primary alcohols by CYP153 enzymes from *Mycobacterium marinum* and *Polaromonas* sp. strain JS666, *Org. Biomol. Chem.*, 9 (2011) 6727-6733.
- [148] B. Brezna, O. Kweon, R.L. Stingley, J.P. Freeman, A.A. Khan, B. Polek, R.C. Jones, C.E. Cerniglia, Molecular characterization of cytochrome P450 genes in the polycyclic aromatic hydrocarbon degrading *Mycobacterium vanbaalenii* PYR-1, *Appl. Microbiol. Biotechnol.*, 71 (2006) 522-532.
- [149] P. Poupin, V. Ducrocq, S. Hallier-Soulier, N. Truffaut, Cloning and characterization of the genes encoding a cytochrome P450 (PipA) involved in piperidine and pyrrolidine utilization and its regulatory protein (PipR) in *Mycobacterium smegmatis* MC2155, *J. Bacteriology*, 181 (1999) 3419-3426.
- [150] M. Ventura, C. Canchaya, A. Tauch, G. Chandra, G.F. Fitzgerald, K.F. Chater, D. van Sinderen, Genomics of Actinobacteria: Tracing the Evolutionary History of an Ancient Phylum, *Microbiol. Mol. Biol. Rev.*, 71 (2007) 495-548.

- [151] L. Chiaradia, C. Lefebvre, J. Parra, J. Marcoux, O. Burlet-Schiltz, G. Etienne, M. Tropis, M. Daffe, Dissecting the mycobacterial cell envelope and defining the composition of the native mycomembrane, *Scientific reports*, 7 (2017) 12807.
- [152] H. Ouellet, E.D. Chow, S. Guan, J.S. Cox, A.L. Burlingame, P.R. de Montellano, Genetic and mass spectrometric tools for elucidating the physiological function(s) of cytochrome P450 enzymes from *Mycobacterium tuberculosis*, *Methods Mol. Biol.* (Clifton, N.J.), 987 (2013) 79-94.
- [153] A. Zanno, N. Kwiatkowski, A.D. Vaz, H.M. Guardiola-Diaz, MT FdR: a ferredoxin reductase from *M. tuberculosis* that couples to MT CYP51, *Biochim. Biophys. Acta*, 1707 (2005) 157-169.
- [154] F. Fischer, D. Raimondi, A. Aliverti, G. Zanetti, *Mycobacterium tuberculosis* FprA, a novel bacterial NADPH-ferredoxin reductase, *FEBS J.*, 269 (2002) 3005-3013.
- [155] K.J. McLean, M. Sabri, K.R. Marshall, R.J. Lawson, D.G. Lewis, D. Clift, P.R. Balding, A.J. Dunford, A.J. Warman, J.P. McVey, A.M. Quinn, M.J. Sutcliffe, N.S. Scrutton, A.W. Munro, Biodiversity of cytochrome P450 redox systems, *Biochem. Soc. Trans.*, 33 (2005) 796-801.
- [156] T. Zhang, A. Zhang, S.G. Bell, L.-L. Wong, W. Zhou, The structure of a novel electron-transfer ferredoxin from *Rhodospseudomonas palustris* HaA2 which contains a histidine residue in its iron-sulfur cluster-binding motif, *Acta Crystallogr. D*, 70 (2014) 1453-1464.
- [157] P.L. Wang, A. Donaire, Z.H. Zhou, M.W. Adams, G.N. La Mar, Molecular model of the solution structure for the paramagnetic four-iron ferredoxin from the hyperthermophilic archaeon *Thermococcus litoralis*, *Biochemistry*, 35 (1996) 11319-11328.
- [158] J.I. Elliott, S.S. Yang, L.G. Ljungdahl, J. Travis, C.F. Reilly, Complete amino acid sequence of the 4Fe-4S, thermostable ferredoxin from *Clostridium thermoaceticum*, *Biochemistry*, 21 (1982) 3294-3298.
- [159] P.S. Brereton, M.F. Verhagen, Z.H. Zhou, M.W. Adams, Effect of iron-sulfur cluster environment in modulating the thermodynamic properties and biological function of ferredoxin from *Pyrococcus furiosus*, *Biochemistry*, 37 (1998) 7351-7362.
- [160] A. Volbeda, M.-H. Charon, C. Piras, E.C. Hatchikian, M. Frey, J.C. Fontecilla-Camps, Crystal structure of the nickel-iron hydrogenase from *Desulfovibrio gigas*, *Nature*, 373 (1995) 580.
- [161] M.N. Lovgreen, M. Martic, M.S. Windahl, H.E. Christensen, P. Harris, Crystal structures of the all-cysteinyl-coordinated D14C variant of *Pyrococcus furiosus* ferredoxin: [4Fe-4S] <--> [3Fe-4S] cluster conversion, *J. Biol. Inorg. Chem.*, 16 (2011) 763-775.

Chapter 2:

Electron transfer ferredoxins with unusual cluster binding motifs support monooxygenase secondary metabolism in many bacteria

Submitted for publication in *Chemical Science*

Citation:

Child, S. A.; Bradley, J. M.; Pukala, T.; Svistunenko, D. A.; Le Brun, N. E.; Bell, S. G.,
Electron transfer ferredoxins with unusual cluster binding motifs support secondary
metabolism in many bacteria, submitted for publication in *Chemical Science*.

Statement of Authorship

Principal Author

Name	Stella Child		
Contribution to the Paper	Cloning and preparation of optimised affinity tagged gene sequences. Aerobic purification of ferredoxins and cytochrome P450s and anaerobic purification of ferredoxins. Sample preparation for mass spectrometry. Substrate characterisation and product formation studies of cytochrome P450s. Manuscript preparation. NB: CYP147G1 product formation with undecanoic acid, Fdx3 mutant turnovers and CYP278A1 turnover of β -ionone (Fig 4 and Fig 5B) were performed as part of prior degree and hence are not included in this assessment.		
Overall percentage (%)	40		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	23/5/18

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Justin Bradley		
Contribution	Anaerobic purification of ferredoxins, EPR sample preparation, native mass spectrometry, redox cycling and potential determination. Manuscript preparation.		
Signature		Date	02/05/2018

Name of Co-Author	Tara Pukala		
Contribution	Mass spectrometry of ferredoxins for molecular weight determination.		
Signature		Date	23/5/2018

Name of Co-Author	Dimitri Svistunenko		
Contribution	EPR of anaerobically purified ferredoxins		
Signature		Date	03/05/2018

Name of Co-Author	Nick Le Brun		
Contribution	Experimental design and supervision, manuscript preparation.		
Signature		Date	03/05/18

Name of Co-Author	Stephen Bell		
Contribution	Experimental design, supervision, manuscript preparation.		
Signature		Date	23/5/2018

Electron transfer ferredoxins with unusual cluster binding motifs support monooxygenase secondary metabolism in many bacteria

Stella A. Child,^a Justin M. Bradley,^b Tara L. Pukala,^a Dimitri A. Svistunenko,^c Nick E. Le Brun^b and Stephen G. Bell^{a,*}

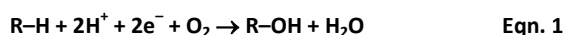
Received 00th January 20xx,
Accepted 00th January 20xx

www.rsc.org/

The proteins responsible for controlling electron transfer in bacterial secondary metabolism are not always known or characterised. Here we demonstrate that many bacteria contain a set of unfamiliar ferredoxin encoding genes which are associated with those of cytochrome P450 (CYP) monooxygenases and as such are involved in anabolic and catabolic metabolism. The model organism *Mycobacterium marinum* M contains eleven of these genes which encode [3/4Fe-4S] ferredoxins but which have unusual iron-sulfur cluster binding motif sequences, CXX?XXC(X)_nCP, where '?' indicates a variable amino acid residue. Rather than a cysteine residue, which is highly conserved in [4Fe-4S] clusters, or alanine or glycine residues, which are common in [3Fe-4S] ferredoxins, these genes encode at this position histidine, asparagine, tyrosine, serine, threonine or phenylalanine. We have purified, characterised and reconstituted the activity of several of these CYP/electron transfer partner systems and show all those examined contain a [3Fe-4S] cluster. Furthermore the identity of the variable residue affects the functionality of the monooxygenase system and has a significant influence on the redox properties of the ferredoxins. Similar ferredoxin encoding genes were also identified across *Mycobacterium* species, including in the pathogenic *M. tuberculosis* and *M. ulcerans*, as well as in a wide range of other bacteria such as *Rhodococcus* and *Streptomyces*. In the majority of instances these are associated with CYP genes. These ferredoxin systems are important in controlling electron transfer across bacterial secondary metabolite production processes which include antibiotic and pigment formation among others.

Introduction

Electron transfer is fundamental for all living organisms, being involved in respiratory processes to produce chemical energy within the cell, pathways to build large molecules from smaller substrates (anabolic) and the breakdown of molecules into smaller species for cellular metabolism (catabolic). The electron carrier proteins involved in these pathways tightly regulate the shuttling of electrons between the donor and acceptor.^{1, 2} Monooxygenases are an essential set of enzymes that are intrinsically involved in these anabolic and catabolic processes and require a supply of electrons in order to function. They typically catalyse the selective hydroxylation of a wide range of organic molecules using dioxygen (Eqn. 1).^{3, 4}



In Nature, monooxygenase enzymes, including cytochrome P450 enzymes (CYP), show exquisite selectivity. The bacterial

CYP family has a broad substrate spectrum yet each individual enzyme can be highly specific. For these reasons many bacteria, including *Mycobacteria*, have a large and highly evolved CYP complement (CYPome) which functions to break down or synthesise molecules as required by the organism.⁵ These enzymes are valuable resources for biocatalysis and key targets for antibacterial drug design.⁶⁻⁸

The two electrons required by CYP enzymes are usually derived from NAD(P)H and delivered by electron transfer proteins in two distinct steps. Bacterial CYP enzymes typically use class I electron transfer systems which normally consist of a NAD(P)H-dependent ferredoxin reductase (FdR) and an iron-sulfur ferredoxin (Fdx).⁹ The reductase most often contains an FAD cofactor. The best studied ferredoxins in these systems are [2Fe-2S] types but [3Fe-4S], [4Fe-4S] clusters, combinations thereof and exceptions such as the FMN containing cindoxin and self-sufficient P450Bm3 are known.^{7, 9-12} The CYP enzymes have evolved to be highly specific for their electron transfer partners from the same species.¹³ Yet within a given bacterium the electron transfer systems can support multiple CYP enzymes and accordingly the number of genes in a bacterium normally decreases in the order CYP > ferredoxin > ferredoxin reductase.¹⁴⁻¹⁷

One bacterium which contains many of these systems is *Mycobacterium marinum*, a ubiquitous pathogen of fish and amphibians, which can infect humans causing aquarium

^a Department of Chemistry, University of Adelaide, SA 5005, Australia.

^b Centre for Molecular and Structural Biochemistry, School of Chemistry, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK.

^c School of Biological Sciences, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK.

Electronic Supplementary Information (ESI) available: Details of *M. marinum* CYPome, Experimental methods and additional Data Tables and Figures. See DOI: 10.1039/x0xx00000x

granuloma.^{20, 21} *M. marinum* can survive in extracellular environmental conditions and is used as a model organism to plot the evolutionary pathway of other more specialist *Mycobacterium* pathogens.²² It is closely related to the slower growing human pathogens *Mycobacterium ulcerans* (97% nucleotide sequence identity) and *Mycobacterium tuberculosis* (85%) which cause Buruli ulcer (Bairnsdale or Daintree ulcer) and tuberculosis, respectively.²⁰⁻²³ Tuberculosis is a global epidemic and a major cause of human mortality and is a growing problem due to the evolution of drug resistant strains.^{24, 25} Buruli ulcer is a serious skin disease prevalent in Africa, Oceania and Asia.²⁶ The mycobacterial pathogens that cause these more serious conditions have adapted to the specific host environment in which they are found and are not viable in other settings. In comparison to *M. marinum*, both *M. tuberculosis* and *M. ulcerans* have undergone reductive evolution. This occurred via genome downsizing and pseudogene formation, through the introduction of mutations or insertion sequences.²⁷⁻²⁹ Another human pathogen, *Mycobacterium leprae*, responsible for leprosy, has undergone extreme reductive evolution and contains less than half the number of genes of *M. tuberculosis*. In addition these bacteria have also acquired certain unique genes which encode proteins critical to their survival.^{23, 29}

Here we report that there are substantially more CYP encoding genes in *M. marinum* than in *M. ulcerans* and *M. tuberculosis* as a direct result of processes of this reductive evolution. The genome of *M. marinum* also contains many genes which encode atypical ferredoxins. Most of these are located close to CYP genes and the majority are [3/4Fe-4S] type ferredoxins containing unusual residues in the iron-sulfur cluster binding motif. Rather than having the CXXCXXC(X)_nCP motif of a [4Fe-4S] cluster ferredoxin or the CXXA/GXXC(X)_nCP motif of a [3Fe-4S] cluster ferredoxin, a different amino acid such as histidine replaces the second cysteine of the [4Fe-4S] motif or the glycine/alanine residue in the [3Fe-4S] ferredoxin motif (CXXHXXC(X)_nCP).¹⁸ The majority of these have yet to be studied and characterised.^{18, 19}

By searching for similar electron transfer systems we find that other residues are commonly located at this position (CXX?XXC(X)_nCP) in ferredoxins of this type from other bacteria.^{15, 30-34} For example, the ferredoxin genes of *M. marinum* encode proteins which contain histidine, asparagine, serine, threonine, tyrosine and phenylalanine. We identify that these types of ferredoxin genes are prevalent across many other bacterial species and isolate and characterise several of the *M. marinum* suite. We demonstrate that, in combination with the same ferredoxin reductase, they can support monooxygenase activity of their associated CYP enzymes. We characterise the cluster type and importance of the variable residue by a variety of methods and show that they significantly alter the properties of the iron-sulfur cluster. Therefore this residue has a role in controlling the behaviour of these electron transfer proteins. These ferredoxins have a critical function in the metabolism of *M. marinum* and other bacteria through their support of electron transfer to monooxygenase systems.

Results

Analysis of the CYPome and the potential electron transfer partners of *M. marinum*

There are forty seven CYP enzyme encoding genes in *M. marinum* and these belong to thirty six different P450 families and thirty nine subfamilies (see Supplementary Information for further details; Table S1, S2 and Fig. S1). Comparatively there are only twenty CYP enzymes in *M. tuberculosis* and twenty four in *M. ulcerans*. This follows from the smaller gene complement of these bacteria due to reductive evolution (Table S1 and S3). There is one unique cytochrome P450 encoding gene in *M. ulcerans*, CYP140A7, which is involved in the biosynthesis of Mycolactone A.³⁵ There are five unique cytochrome P450s in *M. tuberculosis*, of which one, CYP121A1, is essential for the bacterium's survival and is responsible for the formation of an intramolecular C-C bond in the cyclodipeptide cyclo(I-Tyr-I-Tyr).³⁶ In contrast to these species, *M. leprae* contains only one CYP enzyme encoding gene, CYP164A1.³⁷

The CYPome of *M. marinum* plays a role in the bacterium's anabolic and catabolic metabolism. The substrate profiles of several of the cytochrome P450 enzymes can be understood from those that have been previously studied in other mycobacterial species, e.g. CYP51B1 (sterol demethylase), CYP124A1 (phytanic acid ω -hydroxylation), CYP153A16 (alkane oxidation) and CYP142A3 and CYP125A7 (both cholesterol monooxygenases).³⁸⁻⁴³ The potential functions of other CYP enzymes from *M. marinum*, which are hypothesised to include terpene and polyketide synthesis, are discussed in the Supplementary Information.

Of the proposed electron transfer partners for the CYPome of *M. tuberculosis*, (ferredoxin and ferredoxin-NAD(P)H reductase-like proteins) all but two ferredoxin genes, Rv0763 and Rv1786, are located remotely from the CYP genes (Table S4 and S5).^{5, 30, 44-46} The two ferredoxins are located next to the genes of CYP51 and CYP143 (Table 1). Equivalent genes to all of the electron transfer partners of *M. tuberculosis* are also present in the genome of *M. marinum* (Table S5). In addition there are several other ferredoxin and ferredoxin NAD(P)H reductase-like proteins which are described below.

The gene for the CYP153A16 enzyme (Mmar_3154) is in a cluster with genes encoding a [2Fe-2S] ferredoxin (Mmar_3155) and a ferredoxin reductase (FdR2, Mmar_3153) which completes a class I electron transfer system in *M. marinum* (FdR2/[2Fe-2S] ferredoxin/CYP153A16).⁴³ A second gene encoding the CYP147G1 (Mmar_2930) enzyme is in a cluster with genes encoding a ferredoxin reductase (FdR1, Mmar_2931) and a ferredoxin (Fdx3, Mmar_2932) There are therefore two obvious complete electron transfer systems in *M. marinum* (FdR1/Fdx3/CYP147G1 and FdR2/[2Fe-2S] ferredoxin/CYP153A16). By analogy with other bacterial systems these two ferredoxin reductases are likely to be responsible for the reduction of ferredoxins and therefore support the CYP enzymes in *M. marinum*.^{11, 13, 47, 48}

Gene name	ID	Iron Sulfur cluster		AA	<i>M. ul</i>	<i>M. tb</i>	CYP enzyme
		binding motif	pI				
Mmar_2667	Fdx1	CXXHXXC(X) _n CP	3.9	63	Mul_5	Rv1786	CYP143A4
Mmar_2879	Fdx2	CXXTXXC(X) _n CP	4.0	63	Mul_4	-	CYP278A1
Mmar_2932	Fdx3	CXXYXXC(X) _n CP	4.3	70	-	-	CYP147G1
Mmar_3973	Fdx4	CXXNXXC(X) _n CP	3.7	62	Mul_6	-	CYP269A1 & CYP138A4
Mmar_4716	Fdx5	CXXHXXC(X) _n CP	3.7	65	Mul_1	-	CYP188A3
Mmar_4730	Fdx6	CXXFXXC(X) _n CP	4.3	97	-	-	CYP190A3
Mmar_4734	Fdx7	CXXNXXC(X) _n CP	3.8	62	-	-	CYP190A3 & CYP150A5
Mmar_4736	Fdx8	CXXHXXC(X) _n CP	4.0	61	-	-	CYP150A5
Mmar_4763	Fdx9	CXXSXXC(X) _n CP	3.9	63	Mul_2	-	CYP105Q4
Mmar_4933	Fdx10	CXXHXXC(X) _n CP	4.4	67	Mul_3	Rv0763	CYP51B1
Mmar_4991	Fdx11	CXXNXXC(X) _n CP	3.9	81	Mul_7	Rv3503c	-
Mmar_3155	[2Fe-2S]		4.0	105	-	-	CYP153A16
Mmar_2931	FdR1	n.a.	5.8	466	-	-	CYP147G1
Mmar_3153	FdR2	n.a.	4.9	400	-	-	CYP153A16

Table 1 The potential electron transfer partners of the CYPome of *M. marinum* (which are located close to the CYP genes). The gene name as per the databases at the National Center for Biotechnology Information is provided. The sequence of the iron-sulfur cluster binding motif of the [3/4Fe-4S] ferredoxins as well as the predicted pI and length of the amino acid chain are provided. The equivalent ferredoxin genes in *M. ulcerans* and *M. tuberculosis* are given. The neighbouring CYP genes (1-4 genes away) are also shown (Table S1).

There are also several additional ferredoxin genes associated with the CYPome of *M. marinum*. Overall *M. marinum* contains a set of eleven genes encoding ferredoxins of the [3Fe-4S] type (Fdx1-Fdx11), which include the motif CXX?XXC(X)_nCP without a cysteine at the '?' position (Fig. 1). These include the equivalent ferredoxins to Rv0763 and Rv1786 of *M. tuberculosis*, Fdx1 and Fdx10 (CXXHXXC(X)_nCP, Table 1) which are also associated with gene for CYP143 and CYP51, respectively (Table S3, Fig. S2). The ferredoxins, Fdx5 and Fdx8, also contain a histidine residue at the variable position. This set of ferredoxins also encompasses Fdx3, which has a tyrosine in its cluster binding motif (CXXYXXC(X)_nCP) and is associated with the genes for CYP147G1 and FdR1. The other ferredoxins of *M. marinum* contain asparagine (Fdx4, Fdx7 and Fdx11), serine (Fdx9), threonine (Fdx2) or phenylalanine (Fdx6) residues at this position of the cluster binding motif (Fig. 1). All but phenylalanine could potentially act as a ligand to a metal ion in the cluster. The *M. marinum* ferredoxins of this type range in size from 62 to 97 amino acid residues in length and all have predicted pI values lower than 7.0 (Table S1). Ten of the potential ferredoxins of *M. marinum*, Fdx11 being the exception, are located next to or close to a CYP gene (one to four genes away, Fig. S2). Therefore these ferredoxins are likely electron transfer partners which deliver reducing equivalents to the CYP enzymes.³³

Of these eleven ferredoxin genes, seven are conserved in *M. ulcerans* and only three in *M. tuberculosis* (Table 1 and Fig. 3).^{22, 24, 26} The ferredoxins which are conserved in both *M. marinum* and *M. ulcerans* are very similar or identical showing 97-100% sequence identity, while those of *M. tuberculosis* show more sequence divergence (78-92% identity, Fig. 2). Neither of the ferredoxin reductase genes (FdR1 and FdR2) are conserved in *M. ulcerans* or *M. tuberculosis*.

Extending the search to include more diverse species of *Mycobacteria* and other metabolically varied bacteria revealed the prevalence of analogous ferredoxins (Table S6 & S7, Fig. S2). With the exception of the threonine containing Fdx2, many equivalent ferredoxin genes are found across the *Mycobacterium* genus. These are, more often than not, located in gene clusters with associated CYP genes (Table S6, Fig. S2). They are also found across other bacteria including the antibiotic synthesising *Streptomyces* (where variants in the motif at '?' include histidine, serine and tyrosine; Table S7) and *Rhodococcus* (where variants in the motif at '?' include histidine, asparagine and tyrosine; Table S7).¹⁴ They are also associated with the CYPome of other diverse bacterial species including those of *R. palustris* strains mentioned earlier (Table S7). To provide insight into the important CYP enzyme catalysed reactions these ferredoxins support we analysed biosynthetic gene cluster databases for these types of electron transfer partners (Table 2). The synthesis of a range of complex secondary metabolites was found to be supported by these species mainly in strains of *Streptomyces* bacteria (Table 2). This gives a small snapshot of the widespread and critical role these types of ferredoxins have in bacterial metabolism.

Reconstitution of the Activity of CYP147G1 using FdR1 and Fdx3 in *E. coli*

Fdx3 (Mmar_2932) is the only ferredoxin of this type from *M. marinum* to have both a reductase gene (FdR1, Mmar_2931) as well as a CYP gene (CYP147G1, Mmar_2930) co-located in the genome. Of particular interest is that bacterial operon predictions (operondb.cbcb.umd.edu) show that the CYP147 genes in other bacteria including *Methanosarcina barkeri* (CYP147E1), *Myxococcus xanthus* (CYP147A1), *M. vanaabelinii*

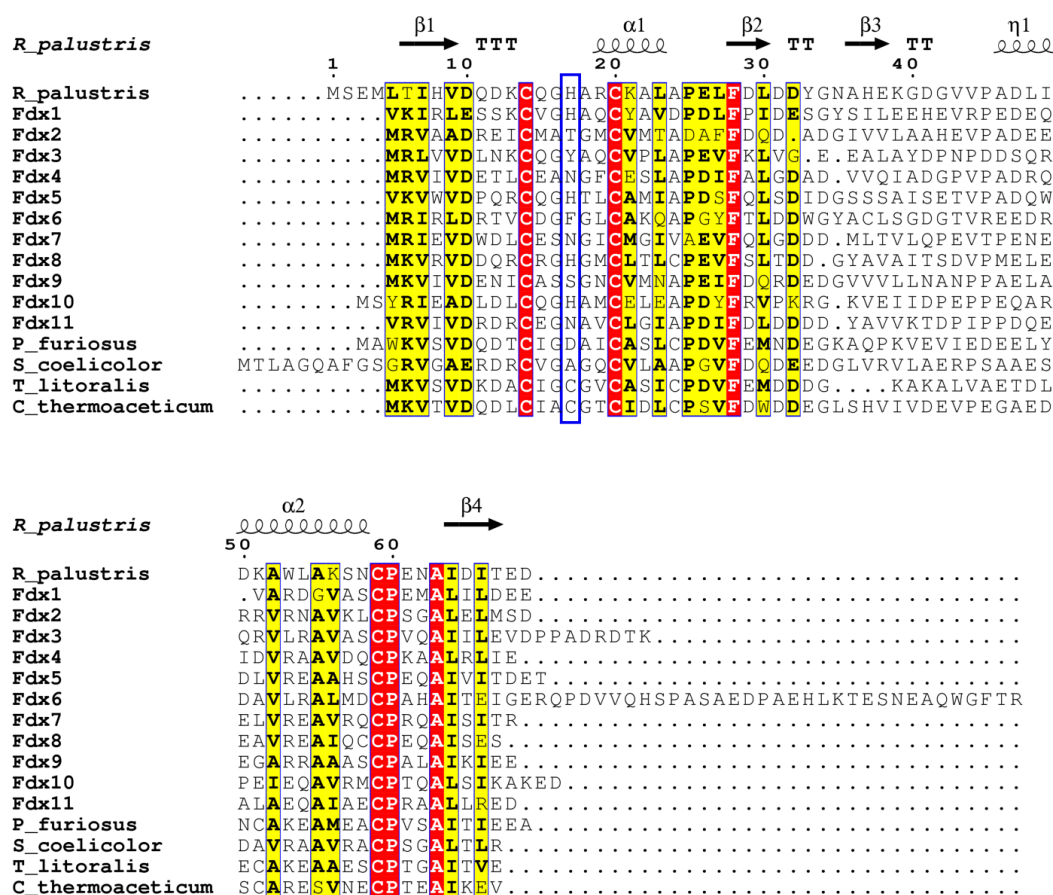


Figure 1. Sequence alignment of the eleven ferredoxins of *M. marinum* (Fdx1 – Fdx11). A structurally characterised histidine containing ferredoxin from *R. palustris* HaA2 (PDB: 4ID8) is shown for reference. The three conserved cysteine residues are highlighted in red while the location of the variable residue (CXX?XXC(X)_nCP) is highlighted in the blue box. For comparison the [3Fe-4S] ferredoxins from *Streptomyces coelicolor*¹⁴ and [4Fe-4S] ferredoxins from *Clostridium thermoacetum*⁴⁹ and *Thermococcus litoralis*^{50,51} are also included.

Natural product	Organism	Fdx Gene name	Fdx Motif	CYP gene name
Griseorhodin A	<i>Streptomyces</i> sp. JP95	<i>grhO4</i>	CXXSXXC(X) _n CP	<i>grhO4</i>
Salinomycin2	<i>S. albus</i> subsp. <i>albus</i>	<i>salF</i>	CXXTXXC(X) _n CP	<i>salD</i>
Salinomycin3	<i>S. albus</i>	<i>slnE</i>	CXXTXXC(X) _n CP	<i>slnF</i>
Tetronomycin	<i>Streptomyces</i> sp. NRRL 11266	<i>tmn14a</i>	CXXSXXC(X) _n CP	<i>tmn14</i>
Rapamycin	<i>S. rapamycinicus</i>	<i>rapO</i>	CXXSXXC(X) _n CP	<i>rapN</i>
Phoslactomycin	<i>Streptomyces</i> sp. HK803	<i>plmT4</i>	CXXTXXC(X) _n CP	<i>plmT6</i>
Filipin	<i>S. avermitilis</i> MA-4680	<i>pteE</i>	CXXSXXC(X) _n CP	<i>pteD</i>
BE-7585A[a]	<i>Amycolatopsis orientalis</i> subsp. <i>vinearia</i>	<i>bexO</i>	CXXSXXC(X) _n CP	<i>bexK</i>
Chrysomycin	<i>S. albaduncus</i>	<i>chryY</i>	CXXHXXC(X) _n CP	<i>chryOIII</i>
Enterocin	<i>S. maritimus</i>	<i>EncQ</i>	CXXSXXC(X) _n CP	<i>EncR</i>
Lysolipin	<i>S. tendae</i>	<i>llpK</i>	CXXSXXC(X) _n CP	<i>llpOIV</i>
Xantholipin	<i>S. flavogriseus</i>	<i>xanK</i>	CXXSXXC(X) _n CP	<i>xanO2</i>
Cinnabaramide	<i>Streptomyces</i> sp. JS360	<i>ORF11</i>	CXXSXXC(X) _n CP	<i>cinL</i>
Leinamycin	<i>S. atroolivaceus</i> S-140	<i>lnmB</i>	CXXSXXC(X) _n CP	<i>lnmA</i>

Table 2. Characterised gene clusters with a ferredoxin (located close to the CYP genes) which has a ferredoxin motif similar to those identified in *M. marinum* M. The gene name as per the databases at the National Center for Biotechnology Information or the gene cluster is provided. The sequence of the iron-sulfur cluster binding motif of the [3/4Fe-4S] ferredoxin is provided. The neighbouring CYP genes (all next to each other with one exception which is 2 genes away) are also shown. The data was obtained from the Minimum Information about a Biosynthetic Gene cluster (<https://mibig.secondarymetabolites.org/index.html>) and DoBiscuit (<http://www.bio.nite.go.jp/pks/tutorial/view>). Most species are from strains of *Streptomyces*. [a] a 2-Thiosugar-Containing Angucycline-Type Natural Product

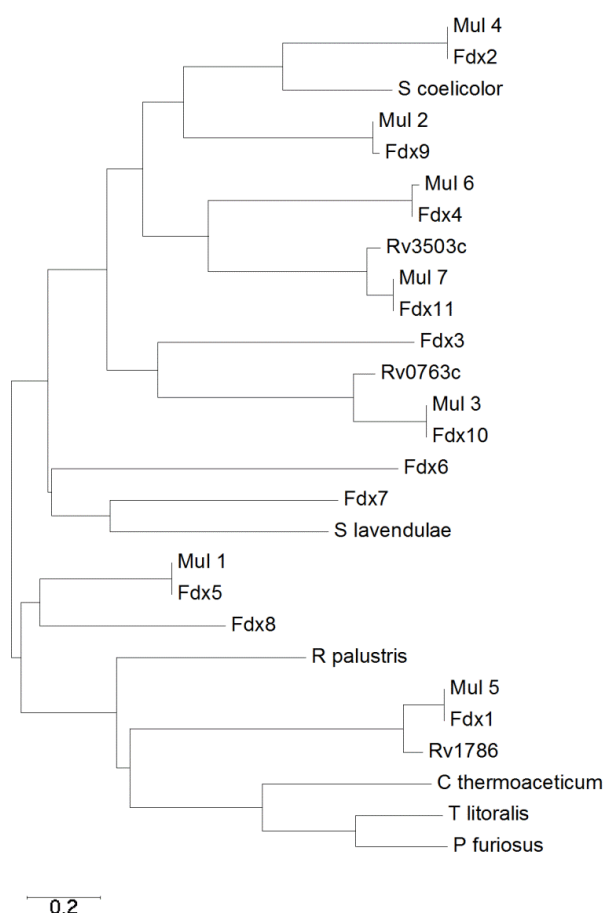


Figure 2. A phylogenetic tree (phenogram) of the [3/4Fe-4S] ferredoxins from *M. marinum* (Fdx1 – Fdx11), *M. ulcerans* (Mul_1 – Mul_7), and *M. tuberculosis* (Rv0763c, Rv1786 and Rv3503c). The ferredoxins from *S. coelicolor*, *S. lavendulae* and the structurally characterised ferredoxins from *R. palustris* HaA2, *P. furiosus* C. thermoaceticum and *T. litoralis* are included for comparison (see Fig. 2). The grouping of the ferredoxins from *M. marinum* and *M. ulcerans* show they are closely related (97–100% sequence identity). There is a lower yet significant similarity to the ferredoxins from *M. tuberculosis* (78–92% sequence identity, note the low 78% value is unusual and arises as the gene Rv3503c is shorter than Fdx11 by the equivalent of nineteen amino acids). For the majority of the ferredoxins there is a low similarity to those from other bacterial species, for example Fdx1 has only 35% sequence identity with the structurally characterised *R. palustris* HaA2 ferredoxin (PDB: 4ID8). The threonine containing Fdx2 has the closest relationship with the [3Fe-4S] ferredoxins from *Streptomyces* species while the [4Fe-4S] ferredoxins from the thermophiles *P. furiosus*, *C. thermoaceticum* and *T. litoralis* cluster together.^{11, 14}

(CYP147G2, 68% sequence identity to CYP147G1) and *Streptomyces avermitilis* (CYP147B1), are all found next to a ferredoxin reductase encoding gene.

All of these also have a gene present which encodes a similar ferredoxin to Fdx3 (Table S8) which, in the case of *M. xanthus*, is a ferredoxin-ferredoxin reductase fusion protein. All of these ferredoxins contain a CXXYXXC(X)_nCP iron-sulfur cluster binding motif. Overall the data suggest that these three genes may form part of an operon with a similar function across many different species. In order to assess if the ferredoxin reductase (FdxR1) and ferredoxin (Fdx3) electron transfer proteins could support the activity of the CYP147G1 enzyme, we constructed a whole-cell oxidation system.^{33, 52} The FdxR1 and Fdx3 genes were cloned in pETDuet to generate pETDuetFdx3/FdxR1 and the CYP147G1 and Fdx3 genes were

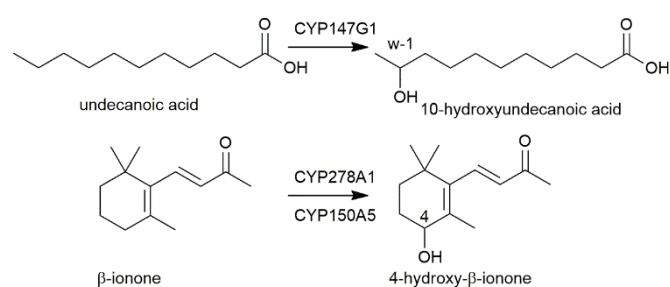


Figure 3. CYP147G1 oxidation of undecanoic acid to 10-hydroxyundecanoic acid and the oxidation of β -ionone to 4-hydroxy- β -ionone by CYP278A1 and CYP150A5.

combined with pRSFDuet to construct pRSFDuetFdx3/CYP147G1. By transforming both vectors into *E. coli* we were able to produce all three proteins together and support substrate turnover by CYP147G1 utilising intracellular NAD(P)H. When the cells were grown and protein produced, the culture media took on a blue colouration indicating that indole generated in the growth media from tryptophan breakdown was being oxidised to indigo.⁵³ By adding indole to the growth we were able to generate more indigo (Fig. S3). When the CYP147G1 was produced in *E. coli* in the absence of FdxR1 and Fdx3 no indigo formation was observed, suggesting that the two electron transfer partners from *M. marinum* are required to support CYP147G1 activity. The CYP147G1 enzyme was produced in *E. coli* and purified by two ion exchange chromatography steps. After purification, CYP147G1 was tested for the characteristic absorbance of a P450, the Soret absorbance (Fig. S4). The binding of CO to the reduced ferrous form of the heme centre results in the almost complete shift of the Soret peak to 450 nm (~95%), indicating the viability of the CYP147G1 enzyme.⁵⁴

Investigation of the substrate range of the purified CYP147G1 enzyme was undertaken by UV/Vis analysis of the spin state of the heme iron.⁵⁵ Undecanoic acid resulted in the largest spin state shift (40% high-spin, compared to indole, <5%) and the dissociation constant was determined to be $25 \pm 4 \mu\text{M}$ (Fig. S5). These results indicated that undecanoic acid is complementary to the active site of CYP147G1 and this substrate was chosen for product formation studies.

Attempts at purifying the electron transfer partners, Fdx3 and FdxR1, in a soluble form have not been successful. The yield of both proteins after cell lysis was insufficient for further workup or detailed *in vitro* analysis of activity. As Fdx3 and FdxR1 are required for the creation of a native-like electron transfer system *in vitro* it was necessary to use whole-cell oxidation systems to investigate product formation. Undecanoic acid was chosen as a substrate and after extraction and derivatisation with BSTFA/TCMS, the turnovers were analysed via GC-MS (Fig. S6). The CYP147G1 turnover of undecanoic acid showed one peak in addition to that of the substrate. Analysis of the mass spectrum fragmentation pattern for the hydroxyundecanoic acid products displayed an increase in the base peak at 117.1 *m/z* (Fig. S6). This is consistent with cleavage next to the CHOSiMe₃ group on the ω -1 carbon (forming a CH₃CHOSiMe₃⁺ fragment). NMR analysis confirmed the product of undecanoic acid turnover was the ω -1

hydroxylated acid (10-hydroxyundecanoic acid) from several diagnostic signals (Fig. S7).

The importance of the Tyr residue in the iron-sulfur cluster motif of Fdx3

In order to investigate the role of this tyrosine residue of Fdx3, mutant versions of the genes were generated in which the polar aromatic tyrosine was replaced with glycine or cysteine. These two mutations were chosen to mimic the iron-sulfur cluster binding motifs of a [3Fe-4S] and a [4Fe-4S] ferredoxin, respectively (Fig. S8). The mutant ferredoxin enzymes were cloned into both plasmids of the whole-cell oxidation system, pETDuetFdx3/FdR1 and pRSFDuetFdx3/CYP147G1, and used to test the activity of the CYP147G1 enzyme coupled with the mutant and WT forms of the ferredoxin. Samples of the turnovers taken at 4 hours contained a product (Fig. 4). The level of product formation when coupled with the WT Fdx partner was almost double that when coupled with either the Y12C or Y12G ferredoxins. Both mutant ferredoxins performed similarly (Fig. 4). These results show that the tyrosine amino acid in the ferredoxin binding sequence is important for the regulation of electron transfer to the CYP147G1 enzyme or for the stability of Fdx3.

Assessing the activity of selected other ferredoxin electron transfer partners of *M. marinum*

In order to assess an entire CYP electron transfer chain the ferredoxin reductase, ferredoxin and CYP enzyme have to be isolated and the likely substrate for the CYP enzyme has to be identified. For a bacterium such as *M. marinum* with so many potential combinations of CYP enzymes and electron transfer partners this would be impractical. We therefore chose pairings of CYP enzymes and ferredoxins which showed promising levels of protein production in *E. coli* and were coupled to ferredoxin genes with different residues at the variable position (?) of the iron-sulfur cluster motif (data not shown). Based on the protein production data the combinations chosen were Fdx2(Thr)/CYP278A1, Fdx4(Asn)/CYP269A1, Fdx8(His)/CYP150A5 and Fdx9(Ser)/CYP105Q4. All four CYP enzymes were produced in good yields.

CYP105Q4, CYP278A1 and CYP150A5 displayed the expected substrate-free low spin CYP UV/Vis spectra with a Soret maximum at 419 nm (Fig. S4). CYP269A1 was more unusual in that it had a spectrum that resembled a high spin ferric heme spectrum with a Soret maximum at 390 nm (Fig. S5). The addition of the imidazole antifungal agent miconazole shifted the heme Soret absorbance maximum of CYP269A1 to 423 nm, and bound with a K_d of $0.050 \pm 0.007 \mu\text{M}$ (Fig. S5). CYP105Q4, CYP278A1 and CYP150A5 showed the characteristic shift to 450 nm for the ferrous-CO bound forms. In the absence of substrate CYP269A1 had a large peak at 420 nm with a shoulder at 450 nm (Fig. S4). When miconazole was added the peak at 450 nm, indicating the ferrous-CO bound form of the enzyme, was generated in greater quantity (Fig. S4).

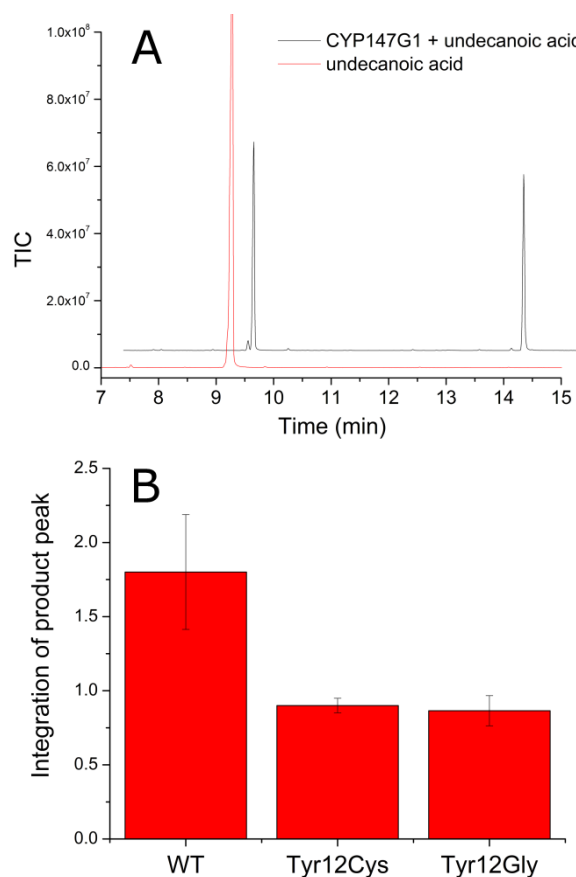


Figure 4. CYP147G1 product formation is reduced when supported by the mutant Fdx partners. (a) GC-MS chromatogram of the CYP147G1 turnover of undecanoic acid (black trace) after derivatisation with BSTFA/TMSCl. Derivatised undecanoic acid (RT 9.2 min, control red trace) and the 10-hydroxyundecanoic acid (RT 13.9 min) are shown. The chromatogram has been offset along the x and y axes for clarity. (b) Quantitation of the 10-hydroxyundecanoic acid product from variant Fdx3-CYP147G1 whole-cell turnovers of undecanoic acid. The axis shows the triplicate average of the area of integrated product peak divided by the area of the internal standard peak. Error bars show one standard deviation.

CYP278A1 has significant sequence overlap with certain members of CYP109 family including CYP109D1 (*Sorangium cellulosum*, 42% sequence identity) and CYP109B1.⁵⁶ Norisoprenoids and fatty acids have been found to be efficiently oxidised by CYP109 family monooxygenases and the addition of β -ionone was found to shift the spin state of CYP278A1 to the high spin form (50%, Fig. S5). The binding affinity of CYP278A1 for β -ionone was also tight; K_d , $5.1 \pm 1.5 \mu\text{M}$, Fig. S5).

The CYP150 enzyme from *Mycobacterium vanbalaanii* PYR1, which is in the same family as CYP150A5, has been reported to oxidise hydrophobic aromatic compounds.⁵⁷ Screening a range of aromatic compounds for their ability to bind to CYP150A5 produced low Type I shifts in the Soret peak absorption (Fig. S5). However we found that addition of β -ionone induced a 60% shift to the high spin form and the binding affinity was also reasonably high; K_d of $41 \pm 3 \mu\text{M}$ (Fig. S5). None of the extensive range of substrates tested with CYP105Q4 altered the spin state from the low spin form. Having identified viable substrates for CYP150A5 and CYP278A1 and an inhibitor of CYP269A1 we attempted to produce and purify the associated

ferredoxins, Fdx8, Fdx2 and Fdx4 respectively, as well as Fdx9. The codon optimised genes of each ferredoxin were obtained and a 6xHis tag was added to the C-terminus by PCR (Supplementary Information). Under aerobic conditions Fdx2 and Fdx9 (associated with CYP278A1 and CYP105Q4, respectively) did not produce significant levels of folded ferredoxin after cell lysis. However Fdx4 and Fdx8 were purified in significant quantities using an ion exchange step followed by affinity chromatography (~1 mg of purified protein per litre of broth, Fig. S9). The UV/Vis spectra of Fdx4 and Fdx8 showed characteristic absorbances of [3/4Fe-4S] cluster containing ferredoxins (Fig. 5).^{18,30}

We used whole-cell oxidation systems to study the monooxygenase activity of the FdR1/Fdx2/CYP278A1 and FdR1/Fdx8/CYP150A5 systems. GC and HPLC analysis of the turnovers of both systems showed a single product was formed from β -ionone oxidation. The product eluted at the same retention time for both systems. Co-elution experiments with turnovers of β -ionone using the CYP101B1 and P450Bm3 which generate the 3- or 4-hydroxy products, respectively, revealed that the sole product from both CYP278A1 and CYP150A5 systems was 4-hydroxy- β -ionone (Fig. 6). This demonstrates FdR1 is able to support the activities of the Fdx2/CYP278A1 and Fdx8/CYP150A5 systems (Fig. 3).

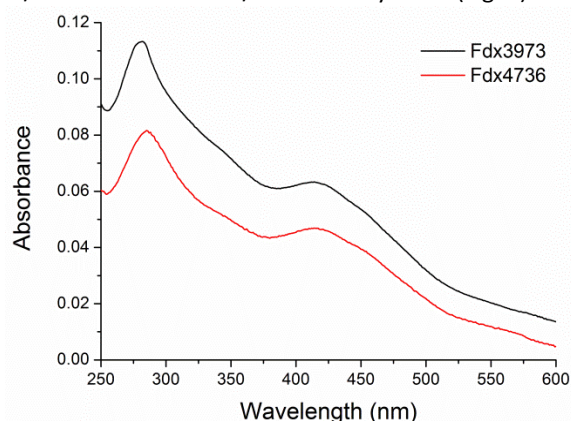


Figure 5. The UV/Vis absorbance spectra of aerobically purified Fdx4 (Mmar3963, black) and Fdx8 (Mmar4736, red) from *M. marinum*. Other spectra are included in the Supplementary Information.

Further characterisation of the ferredoxins generated under anaerobic conditions

The purification of a range of these ferredoxins was also undertaken under anaerobic conditions to assess their redox activity and stability to oxygen. Fdx2 and Fdx3 yielded very low levels of folded protein while Fdx4 (Asn), Fdx5 (His) and Fdx9 (Ser) were isolated in good yields (Supplementary Information).

The ferredoxins were characterised by UV/visible absorbance and CD spectroscopy (Fig. S11). Fdx4, Fdx5 and Fdx9 were shown to bind a [3Fe-4S] cluster by a combination of non-denaturing ESI-MS and standard LC-MS (Fig. S12). They also exhibited EPR spectra characteristic of [3Fe-4S] centres (Fig. S13) meaning that the cluster observed by mass spectrometry is unlikely to be the result of degradation during ionisation.⁵⁸ A

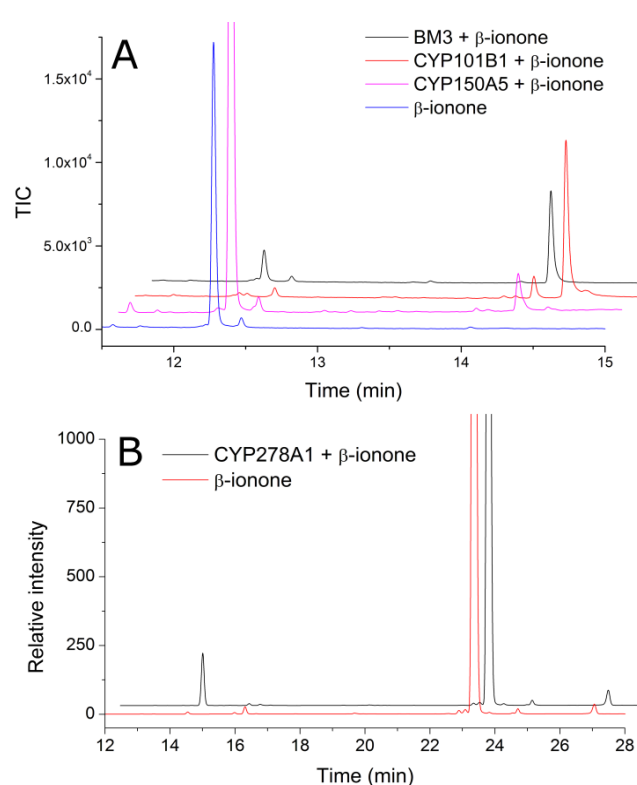


Figure 6. Activity of CYP150A5 and CYP278A1 when supported by native electron transfer partners *in vivo* (a) GC analysis of the whole-cell oxidation turnover of CYP150A5 and β -ionone: blue, β -ionone control (RT 12.3 min); magenta, CYP150A5 turnover supported by FdR1 and Fdx8 (product RT 14.3 min); red, *in vitro* turnover of β -ionone by CYP101B1 (major product 3-hydroxy- β -ionone, RT 14.4 min; minor product 4-hydroxy- β -ionone, RT 14.3 min) and black, *in vitro* turnover of P450Bm3 (CYP102A1) and β -ionone (sole product, 4-hydroxy- β -ionone). (b) Reverse phase HPLC analysis of the whole-cell turnover of CYP278A1 and β -ionone (black, RT 14.5 min); red, β -ionone control (RT 23.4 min). The chromatograms have been offset along the x- and y-axes for clarity.

sample of Fdx4 purified by ion exchange and size exclusion methods had optical properties identical to those purified exploiting the poly-His tag, demonstrating that IMAC chromatography does not result in loss of a loosely bound Fe ion from the cluster. Furthermore, the CD and absorbance spectra of Fdx4 were unaffected by incubation with ferrous ion (Fig. S14). We therefore conclude that each of the ferredoxins binds a [3Fe-4S] cluster following expression in *E. coli* and that *in vitro* incubation with Fe^{2+} does not lead to incorporation of a fourth metal ion and reconstitution of a [4Fe-4S] cluster. The stability of Fdx2 and Fdx4 to oxygen was also assessed by CD spectroscopy with no significant cluster degradation occurring on exposure to 120 μM O_2 for 40 minutes (Fig. S15).

Attempts to cycle the oxidation state of the clusters bound to anaerobically purified Fdx4, Fdx5 and Fdx9 were suggestive of large differences in their midpoint potentials on variation of the residue at position '?' in the binding motif. The CD and electronic absorbance spectra of Fdx4 were altered by incubation with either a stoichiometric equivalent of $\text{K}_3\text{Fe}(\text{CN})_6$ or excess EuCl_2 as oxidant and reductant, respectively (Fig. S11). This suggests a mixed oxidation state of the cluster in the protein isolated under anaerobic conditions with the cluster being readily oxidised by $\text{K}_3\text{Fe}(\text{CN})_6$ and reduced by EuCl_2 . EPR

spectra of the chemically poised samples (Fig S13) were also consistent with oxidation and reduction to the EPR active and EPR silent forms of the cluster, respectively. Equilibration of Fdx4 with a 1.5 mM solution of sodium ascorbate resulted in a CD spectrum readily interpreted as a 50:50 sum of those of the oxidised and reduced clusters, suggesting a midpoint potential closely matched to the ascorbate solution potential (+60 mV vs. SHE). The optical spectra of Fdx5 were invariant following incubation with EuCl_2 whilst incubation with a stoichiometric equivalent of $\text{K}_3\text{Fe}(\text{CN})_6$ produced significant changes that were reversed upon subsequent incubation with excess EuCl_2 (Fig. S11). Therefore Fdx5 isolated under identical conditions to Fdx4 contains clusters predominantly in the reduced state. Incubation of reduced Fdx5 with 1.5 mM sodium ascorbate had no significant effect on the CD spectrum indicating that the cluster remained reduced at +60 mV, suggesting a midpoint potential greater than +150 mV. In contrast, incubation with $\text{K}_3\text{Fe}(\text{CN})_6$ had no effect on the optical spectra of Fdx9 whilst EuCl_2 led to a reversible loss of CD intensity demonstrating that the protein was isolated with the cluster in the oxidised state (Fig S11). The EPR spectra of chemically poised Fdx9 samples indicated that 50% of the sample retained oxidised clusters following equilibration with excess EuCl_2 (Fig. S13) suggesting a midpoint potential similar to that of the $\text{Eu}^{3+}/\text{Eu}^{2+}$ couple (-360 mV vs. SHE).

Discussion

The CYPome of *M. marinum* is larger than that of *M. tuberculosis* and the related *M. ulcerans*. There are also more electron transfer partner genes in *M. marinum* and these are closely associated with the genes of the CYPome. The majority of the ferredoxins are of the [3/4Fe-4S] type and all have a non-cysteine residue in their iron-sulfur cluster binding motif, but not an alanine or glycine residue that commonly replaces the second cysteine of the motif in [3Fe-4S] ferredoxins. The identity of this residue varies among the ferredoxins of *M. marinum* and includes histidine, asparagine, threonine, tyrosine, serine and phenylalanine. The character of the non-cysteine residues would modify the environment of the iron-sulfur cluster and would be expected to have significant effects on the properties and function of the ferredoxin. Many of these non-cysteine amino acid substitutions have the potential to act as ligands to metal ions and determine the cluster type ([3Fe-4S] or [4Fe-4S]) as well as control the properties and redox potential of the protein. With the exception of the histidine containing version, these types of ferredoxins have not been characterised previously. Anaerobic isolation of and characterisation of the ferredoxins showed that all contained a [3Fe-4S] cluster.

The most studied small ferredoxin of this type is that from the thermophile *Pyrococcus furiosus* in which one of the iron ions of the cluster coordinates to an aspartate residue (CXXDXXC(X)_nCP).^{19, 59, 60} Replacement of the more usual cysteine residue with aspartate alters the properties of this ferredoxin, most notably the reduction potential, compared to typical cysteine-ligated [4Fe-4S] ferredoxins.⁵¹ This

ferredoxin is isolated as a [3Fe-4S] ferredoxin under aerobic conditions but can be isolated as a [4Fe-4S] ferredoxin when oxygen is excluded. In common with other thermophilic ferredoxins, it also contains an additional disulphide bond which is thought to take part in the redox cycling of the enzyme. In contrast to the ferredoxins we have identified, the redox partner for the aspartate containing ferredoxin from *P. furiosus* is as yet unknown. This makes an analysis of the role of this residue in physiological electron transfer reactions difficult.

The histidine containing ferredoxins from *Mycobacterium* and *Rhodopseudomonas* species have also been isolated aerobically as [3Fe-4S] ferredoxins.^{18, 30-32} By way of contrast, histidine coordination to [4Fe-4S] iron-sulfur clusters has been observed in Ni-Fe and Fe-only hydrogenases and in [2Fe-2S] Rieske proteins.⁶¹⁻⁶³ Previously, the best characterised mycobacterial electron transfer ferredoxin of this type is encoded by the gene Rv0763c and is associated with CYP51 of *M. tuberculosis*.³⁰ This ferredoxin has been shown to support the first electron transfer step of CYP51 from *M. tuberculosis* (but not the monooxygenase activity). The low activity observed was rationalised by the high reduction potential of the ferredoxin which is reported to be unfavourable for the reduction of the substrate bound CYP51.

The identity of the iron-sulfur cluster, the ligands which coordinate to the metal ions and the surrounding environment can have a profound effect on the reduction potential of the ferredoxins. Cysteine-coordinated [2Fe-2S] containing ferredoxins (reduction potential -150 to -400 mV versus SHE) and the histidine Rieske equivalents (reduction potential +100 to +400 mV) contain two Fe(III) ions in the oxidised form with one of these being reduced to Fe(II) in the reduced form.⁶³ It is usual for [4Fe-4S] clusters to shuttle between the [4Fe-4S]^{2+/+} state and have reduction potentials ranging from -280 to -715 mV, though high potential iron-sulfur clusters with potentials of +90 to +450 mV are known ([4Fe-4S]^{3+/2+}).⁶³ Previously characterised [3Fe-4S] ferredoxins have redox potentials ranging from -203 to -85 mV.⁵¹ The histidine containing [3Fe-4S] ferredoxin from *M. tuberculosis* (-31 mV) and the aspartate version from *P. furiosus* (-160 mV for 3Fe form and -375 mV for [4Fe-4S] form) both have significantly different reduction potentials compared to standard [3Fe-4S] and [4Fe-4S] ferredoxins. Our results show that the ferredoxins from *M. marinum* are isolated in the [3Fe-4S] form and their reduction potentials vary between -360 to +150 mV. Several, including the neutral histidine and asparagine containing ferredoxins, are on the positive side compared to other proteins of this type. The serine containing ferredoxin (Fdx9) had a significantly lower redox potential, similar to those of the cysteine containing species. One interesting observation is that Fdx4 is associated with an unusual P450 with substrate free spectra indicating it exists in the high-spin form. It seems probable that bacteria may use these different motifs to tune the redox potential in order to control electron transfer to the different monooxygenases present. As the relative number of CYP and electron transfer partner genes it is expected that

each electron transfer ferredoxin could support multiple CYP enzymes.

Given that the majority of the types of ferredoxins found in *M. marinum* have not been reported previously, we were surprised to discover that they are prevalent across a range of bacteria in particular in *Mycobacterium*, *Rhodococcus*, *Streptomyces* and other species of actinobacteria. They are also found in other bacteria. For example the tyrosine containing ferredoxins are found in bacteria of the phylum *Chloroflexi*. It is telling that in many instances these ferredoxins are associated with CYP genes. It seems likely that they are involved in controlling the electron transfer pathways to enable the monooxygenase enzymes to synthesise complex natural products with a diverse array of biological function.

Conclusions

Overall we have identified, isolated and characterised these unusual ferredoxins from *M. marinum* as being [3Fe-4S] cluster containing proteins. We have used several of them in conjunction with a ferredoxin reductase to reconstitute the activity of their associated P450 enzyme. Similar ferredoxins are co-located with the CYPomes of *Mycobacteria*, as well as in many other types of bacteria. The identity of the altered residue of the motif was found to alter the redox potential. The diversity of these genes in *M. marinum* make it an excellent model organism for investigating electron transfer and its role in bacterial secondary metabolism. The observation of similar CYP systems in pathogenic bacteria such as *M. ulcerans* means these could be targets for drug design resulting in inhibition. The prevalence of these types of ferredoxins across the bacterial kingdom and their presence in the gene clusters of complex secondary metabolites highlights their importance in prokaryote secondary metabolism. Further study will lead to a better understanding of the role of the electron transfer partner proteins for efficient metabolism and natural product synthesis which in turn will allow the design of improved monooxygenase systems for applications in synthetic biology.

Experimental

Phylogenetic analysis of the P450s and ferredoxins was carried out using standard methodologies as described in the Supplementary information. The cloning, protein purification steps, whole-cell turnovers, metabolite characterisation and the aerobic and anaerobic protein analysis are described in full in the Supplementary information.

Acknowledgements

This work was supported by ARC through a Future Fellowship (FT140100355 to SGB) and by BBSRC grants BB/R002363/1 and BB/R003203/1. The authors also acknowledge the award of University of Adelaide Faculty of Sciences Divisional Scholarship (PhD to SAC). We thank the Faculty of Sciences,

University of Adelaide, for a Mid-Career Travel grant and Dr Myles Cheesman (Centre for Molecular and Structural Biochemistry, University of East Anglia) for access to the Circular Dichroism spectrophotometer.

Conflicts of Interest

The authors confirm there are no conflicts of interest to declare.

References

1. C. C. Page, C. C. Moser and P. L. Dutton, *Curr. Opin. Chem. Biol.*, 2003, **7**, 551-556.
2. C. M. Paquete and R. O. Louro, *Acc Chem Res*, 2014, **47**, 56-65.
3. P. R. Ortiz de Montellano, ed., *Cytochrome P450: Structure, Mechanism, and Biochemistry*, Kluwer Academic/Plenum Publishers, New York, 2005.
4. A. Sigel, H. Sigel and R. Sigel, *The Ubiquitous Roles of Cytochrome P450 Proteins*, John Wiley & Sons, 1st edn., 2007.
5. H. Ouellet, J. B. Johnston and P. R. Ortiz de Montellano, *Arch. Biochem. Biophys.*, 2010, **493**, 82-95.
6. K. J. McLean, A. J. Dunford, R. Neeli, M. D. Driscoll and A. W. Munro, *Arch. Biochem. Biophys.*, 2007, **464**, 228-240.
7. C. J. Whitehouse, S. G. Bell and L. L. Wong, *Chem. Soc. Rev.*, 2012, **41**, 1218-1260.
8. V. B. Urlacher and M. Girhard, *Trends Biotechnol.*, 2012, **30**, 26-36.
9. F. Hannemann, A. Bichet, K. M. Ewen and R. Bernhardt, *Biochim. Biophys. Acta*, 2007, **1770**, 330-344.
10. T. Mandai, S. Fujiwara and S. Imaoka, *FEBS J.*, 2009, **276**, 2416-2429.
11. Y. J. Chun, T. Shimada, R. Sanchez-Ponce, M. V. Martin, L. Lei, B. Zhao, S. L. Kelly, M. R. Waterman, D. C. Lamb and F. P. Guengerich, *J. Biol. Chem.*, 2007, **282**, 17486-17500.
12. D. B. Hawkes, G. W. Adams, A. L. Burlingame, P. R. Ortiz de Montellano and J. J. de Voss, *J. Biol. Chem.*, 2002, **277**, 27725-27732.
13. W. Yang, S. G. Bell, H. Wang, W. Zhou, N. Hoskins, A. Dale, M. Bartlam, L. L. Wong and Z. Rao, *J. Biol. Chem.*, 2010, **285**, 27372-27384.
14. D. C. Lamb, T. Skaug, H. L. Song, C. J. Jackson, L. M. Podust, M. R. Waterman, D. B. Kell, D. E. Kelly and S. L. Kelly, *J. Biol. Chem.*, 2002, **277**, 24000-24005.
15. S. G. Bell, N. Hoskins, F. Xu, D. Caprotti, Z. Rao and L. L. Wong, *Biochem. Biophys. Res. Commun.*, 2006, **342**, 191-196.
16. S. G. Bell and L. L. Wong, *Biochem. Biophys. Res. Commun.*, 2007, **360**, 666-672.
17. K. M. Ewen, F. Hannemann, Y. Khatri, O. Perlova, R. Kappl, D. Krug, J. Huttermann, R. Muller and R. Bernhardt, *J. Biol. Chem.*, 2009, **284**, 28590-28598.
18. T. Zhang, A. Zhang, S. G. Bell, L. L. Wong and W. Zhou, *Acta Crystallogr. D Biol. Crystallogr.*, 2014, **70**, 1453-1464.
19. R. C. Conover, A. T. Kowal, W. G. Fu, J. B. Park, S. Aono, M. W. Adams and M. K. Johnson, *J. Biol. Chem.*, 1990, **265**, 8533-8541.
20. T. P. Stinear, G. A. Jenkin, P. D. Johnson and J. K. Davies, *J. Bacteriol.*, 2000, **182**, 6322-6330.
21. D. M. Tobin and L. Ramakrishnan, *Cell. Microbiol.*, 2008, **10**, 1027-1039.

22. T. P. Stinear, T. Seemann, P. F. Harrison, G. A. Jenkin, J. K. Davies, P. D. Johnson, Z. Abdellah, C. Arrowsmith, T. Chillingworth, C. Churcher, K. Clarke, A. Cronin, P. Davis, I. Goodhead, N. Holroyd, K. Jagels, A. Lord, S. Moule, K. Mungall, H. Norbertczak, M. A. Quail, E. Rabinowitsch, D. Walker, B. White, S. Whitehead, P. L. Small, R. Brosch, L. Ramakrishnan, M. A. Fischbach, J. Parkhill and S. T. Cole, *Genome Res.*, 2008, **18**, 729-741.
23. K. Roltgen, T. P. Stinear and G. Pluschke, *Infect. Genet. Evol.*, 2012, **12**, 522-529.
24. S. T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eglmeier, S. Gas, C. E. Barry, 3rd, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead and B. G. Barrell, *Nature*, 1998, **393**, 537-544.
25. S. T. Cole and G. Riccardi, *Curr. Opin. Microbiol.*, 2011, **14**, 570-576.
26. T. P. Stinear, T. Seemann, S. Pidot, W. Frigui, G. Reysset, T. Garnier, G. Meurice, D. Simon, C. Bouchier, L. Ma, M. Tichit, J. L. Porter, J. Ryan, P. D. Johnson, J. K. Davies, G. A. Jenkin, P. L. Small, L. M. Jones, F. Tekaia, F. Laval, M. Daffe, J. Parkhill and S. T. Cole, *Genome Res.*, 2007, **17**, 192-200.
27. C. Demangel, T. P. Stinear and S. T. Cole, *Nat. Rev. Microbiol.*, 2009, **7**, 50-60.
28. P. Singh and S. T. Cole, *Future Microbiol.*, 2011, **6**, 57-71.
29. F. J. Veyrier, A. Dufort and M. A. Behr, *Trends Microbiol.*, 2011, **19**, 156-161.
30. K. J. McLean, A. J. Warman, H. E. Seward, K. R. Marshall, H. M. Girvan, M. R. Cheesman, M. R. Waterman and A. W. Munro, *Biochemistry*, 2006, **45**, 8427-8443.
31. P. Poupin, V. Ducrocq, S. Hallier-Soulier and N. Truffaut, *J. Bacteriol.*, 1999, **181**, 3419-3426.
32. B. Sielaff, J. R. Andreesen and T. Schrader, *Appl. Microbiol. Biotechnol.*, 2001, **56**, 458-464.
33. S. G. Bell, A. Dale, N. H. Rees and L. L. Wong, *Appl. Microbiol. Biotechnol.*, 2010, **86**, 163-175.
34. Y. Lu, F. Qiao, Y. Li, X. H. Sang, C. R. Li, J. D. Jiang, X. Y. Yang and X. F. You, *Appl. Microbiol. Biotechnol.*, 2017, DOI: 10.1007/s00253-017-8454-7.
35. T. P. Stinear, A. Mve-Obiang, P. L. Small, W. Frigui, M. J. Pryor, R. Brosch, G. A. Jenkin, P. D. Johnson, J. K. Davies, R. E. Lee, S. Adusumilli, T. Garnier, S. F. Haydock, P. F. Leadlay and S. T. Cole, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 1345-1349.
36. P. Belin, M. H. Le Du, A. Fielding, O. Lequin, M. Jacquet, J. B. Charbonnier, A. Lecoq, R. Thai, M. Courcon, C. Masson, C. Dugave, R. Genet, J. L. Pernodet and M. Gondry, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 7426-7431.
37. A. G. Warrilow, C. J. Jackson, J. E. Parker, T. H. Marczylo, D. E. Kelly, D. C. Lamb and S. L. Kelly, *Antimicrob. Agents Chemother.*, 2009, **53**, 1157-1164.
38. A. Bellamine, A. T. Mangla, W. D. Nes and M. R. Waterman, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 8937-8942.
39. J. K. Capyk, R. Kalscheuer, G. R. Stewart, J. Liu, H. Kwon, R. Zhao, S. Okamoto, W. R. Jacobs, Jr., L. D. Eltis and W. W. Mohn, *J. Biol. Chem.*, 2009, **284**, 35534-35542.
40. M. D. Driscoll, K. J. McLean, C. Levy, N. Mast, I. A. Pikuleva, P. Lafite, S. E. Rigby, D. Leys and A. W. Munro, *J. Biol. Chem.*, 2010, **285**, 38270-38282.
41. J. B. Johnston, P. M. Kells, L. M. Podust and P. R. Ortiz de Montellano, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 20687-20692.
42. J. B. Johnston, H. Ouellet and P. R. Ortiz de Montellano, *J. Biol. Chem.*, 2010, **285**, 36352-36360.
43. D. Scheps, S. H. Malca, H. Hoffmann, B. M. Nestl and B. Hauer, *Org. Biomol. Chem.*, 2011, **9**, 6727-6733.
44. F. Fischer, D. Raimondi, A. Aliverti and G. Zanetti, *Eur. J. Biochem.*, 2002, **269**, 3005-3013.
45. A. Zanno, N. Kwiatkowski, A. D. Vaz and H. M. Guardiola-Diaz, *Biochim. Biophys. Acta*, 2005, **1707**, 157-169.
46. S. Ricagno, M. de Rosa, A. Aliverti, G. Zanetti and M. Bolognesi, *Biochem. Biophys. Res. Commun.*, 2007, **360**, 97-102.
47. Y. Peng, F. Xu, S. G. Bell, L. L. Wong and Z. Rao, *Acta Crystallograph Sect F Struct Biol Cryst Commun*, 2007, **63**, 422-425.
48. I. F. Sevrioukova, H. Li and T. L. Poulos, *J. Mol. Biol.*, 2004, **336**, 889-902.
49. J. I. Elliott, S. S. Yang, L. G. Ljungdahl, J. Travis and C. F. Reilly, *Biochemistry*, 1982, **21**, 3294-3298.
50. P. L. Wang, A. Donaire, Z. H. Zhou, M. W. Adams and G. N. La Mar, *Biochemistry*, 1996, **35**, 11319-11328.
51. P. S. Brereton, M. F. Verhagen, Z. H. Zhou and M. W. Adams, *Biochemistry*, 1998, **37**, 7351-7362.
52. S. G. Bell, A. B. Tan, E. O. Johnson and L. L. Wong, *Molecular bioSystems*, 2010, **6**, 196-204.
53. E. M. Gillam, L. M. Notley, H. Cai, J. J. De Voss and F. P. Guengerich, *Biochemistry*, 2000, **39**, 13817-13824.
54. T. Omura and R. Sato, *J. Biol. Chem.*, 1964, **239**, 2370-2378.
55. S. Bhattarai, K. Liou and T. J. Oh, *Arch. Biochem. Biophys.*, 2013, **539**, 63-69.
56. A. Zhang, T. Zhang, E. A. Hall, S. Hutchinson, M. J. Cryle, L. L. Wong, W. Zhou and S. G. Bell, *Molecular bioSystems*, 2015, **11**, 869-881.
57. B. Brezna, O. Kweon, R. L. Stingley, J. P. Freeman, A. A. Khan, B. Polek, R. C. Jones and C. E. Cerniglia, *Appl. Microbiol. Biotechnol.*, 2006, **71**, 522-532.
58. R. Bache, P. M. H. Kroneck, H. Merkle and H. Beinert, *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1983, **722**, 417-426.
59. L. Calzolari, C. M. Gorst, Z. H. Zhao, Q. Teng, M. W. Adams and G. N. La Mar, *Biochemistry*, 1995, **34**, 11373-11384.
60. R. E. Duderstadt, C. R. Staples, P. S. Brereton, M. W. Adams and M. K. Johnson, *Biochemistry*, 1999, **38**, 10585-10593.
61. A. Volbeda, M. H. Charon, C. Piras, E. C. Hatchikian, M. Frey and J. C. Fontecilla-Camps, *Nature*, 1995, **373**, 580-587.
62. J. W. Peters, W. N. Lanzilotta, B. J. Lemon and L. C. Seefeldt, *Science*, 1998, **282**, 1853-1858.
63. J. Meyer, *J. Biol. Inorg. Chem.*, 2008, **13**, 157-170.

Chapter 3:

Selective ω -1 oxidation of fatty acids by CYP147G1 from *Mycobacterium marinum*

Prepared in the style of *Biochimica et Biophysica Acta* General Subjects in preparation for submission.

Citation:

Child, S. A.; Rossi, V.; Bell, S. G., Selective ω -1 oxidation of fatty acids by CYP147G1 from *Mycobacterium marinum*, manuscript in preparation.

Statement of Authorship

Principal Author

Name	Stella Child		
Contribution to the Paper	Enzyme purification, substrate characterisation, product formation, manuscript preparation and review. NB: CYP147G1 product formation and binding analysis of the linear fatty acids C10-C12 were performed as part of prior degree and hence are not included in this assessment.		
Overall percentage (%)	70		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	23/5/18

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Vanessa Rossi		
Contribution	Enzyme purification and substrate characterisation		
Signature		Date	16/5/18

Name of Co-Author	Stephen Bell		
Contribution	Cloning of gene sequences, experimental design, supervision, manuscript preparation.		
Signature		Date	23/5/2018

Selective ω -1 oxidation of fatty acids by CYP147G1 from *Mycobacterium marinum*

Stella A. Child, Vanessa P. Rossi, and Stephen G. Bell*

Department of Chemistry, University of Adelaide, SA 5005, Australia

* To whom correspondence should be addressed.

Stephen G. Bell (stephen.bell@adelaide.edu.au)

Highlights

- Cytochrome P450 CYP147G1, from *Mycobacterium marinum*, binds a range of fatty acids
- Activity was reconstituted *in vivo* with physiological electron transfer partners
- The enzyme is regioselective for the ω -1 position for linear and ω -2 branched acids
- Branched ω -1 methyl fatty acids alter enzyme selectivity toward the ω position
- Closely related complete CYP monooxygenase systems were identified in other bacteria

Abstract

Background

Cyp147G1 is one of 47 cytochrome P450 encoding genes in *Mycobacterium marinum* M, a pathogenic bacterium with a high degree of sequence similarity to *Mycobacterium tuberculosis* and *Mycobacterium ulcerans*. *Cyp147G1* is one of only two of these *cyp* genes which are closely associated with a complete electron transfer system.

Methods

The substrate range of the enzyme was tested *in vitro* and the activity of CYP147G1 was reconstituted *in vivo* by co-producing the P450 with the ferredoxin and ferredoxin reductase.

Results

The substrate range of CYP147G1 includes fatty acids ranging from octanoic to hexadecanoic acid. CYP147G1 catalysed the selective hydroxylation of linear and ω -2 methyl branched fatty acids at the ω -1 position ($\geq 98\%$). Oxidation of ω -1 methyl branched fatty acids generated the ω and ω -1 hydroxylation products in almost equal proportions, indicating altered position of hydrogen abstraction.

Conclusions

This selectivity of fatty acid hydroxylation inferred that linear species must bind in the active site of the enzyme with the terminal methyl group sequestered so that abstraction at the C-H bonds of the ω -1 position is favoured. With branched substrates, one of the methyl groups must be close to the compound I oxygen atom and enable hydroxylation at the terminal methyl group to compete with the reaction at the ω -1 C-H bond.

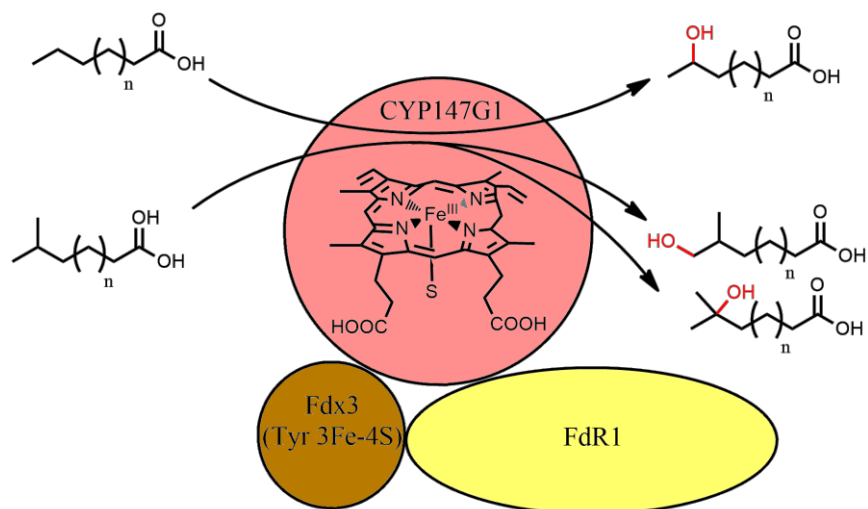
General Significance

Hydroxy fatty acids are widely used for industrial, food and medical purposes. CYP147G1 demonstrates high regioselectivity for hydroxylation at a sub-terminal position on a broad range of linear fatty acids, not seen in other CYP enzymes.

Key words

enzymology, fatty acid metabolism, electron transfer, biocatalysis, cytochrome P450

Graphical abstract



Abbreviations

2xYT, 2 x concentration yeast extract tryptophan broth; BSTFA-TMSCl, N,O-bis(trimethylsilyl) trifluoroacetamide and trimethylsilylchloride; CYP or P450, Cytochrome P450 enzyme; DTT, dithiothreitol; EMM, *E. coli* minimal media; FAD, flavin adenine dinucleotide; FdR, ferredoxin reductase; Fdx, ferredoxin; GC-MS or MS, gas-chromatography mass spectrometry or mass spectrometry; IPTG, Isopropyl β -D-1-thiogalactopyranoside; LB, Lysogeny broth (also known as Luria or Lennox Broth), NAD(P)H reduced nicotinamide adenine dinucleotide (phosphate); RT, retention time; SOC, Super Optimal broth with Catabolite repression; TMS, trimethylsilyl; PCR, polymerase chain reaction.

1. Introduction

Monoxygenases are a diverse class of enzymes, involved in the selective hydroxylation of organic molecules using molecular dioxygen. Cytochrome P450 enzymes hold key roles in the physiological functioning of cells, such as fatty acid hydroxylation, [1] steroid synthesis [2, 3] and drug metabolism [4, 5]. Such versatility of metabolite production, combined with their potential for high specificity, makes them desirable catalysts for synthetic processes [6, 7]. Various bacterial CYPs have been studied for their ability to activate fatty acids, which are readily available from natural sources, into more commercially useful starting materials by hydroxylation [8-11]. For example, the hydroxylation of fatty acids at the ω terminus allows further oxidation to a dicarboxylic acid, and from there to a wide range of materials including fragrances, polymers and adhesives [12]. Hydroxylation at any position increases the reactivity, viscosity and miscibility of the fatty acid, and as a result hydroxy fatty acids are utilised widely for both industrial and medical purposes as well as in food [13]. Fatty acid hydroxylases are commonly divided into carboxyl-terminal (α and β), terminal (ω) and sub-terminal ($\omega-1$ through γ) [14]. For example, the CYP152 family are known α/β hydroxylases [15], while the CYP153 family, including CYP153A16 from *Mycobacterium marinum*, are ω hydroxylases [16]. P450BM3 (CYP102A1) is a highly efficient sub-terminal fatty acid hydroxylase ($\omega-1$ to $\omega-3$ are hydroxylated in roughly equal proportion) but high regioselectivity at sub-terminal positions is very rare. Enzymes such as P450BM3 have been the subject of many mutagenesis studies aiming to improve regioselectivity [14, 17].

In order to catalyse their reactions, the majority of CYPs require a supply of electrons derived from the co-enzymes NADPH or NADH and delivered via electron carrier proteins. Bacterial CYP enzymes often display high redox partner specificity with non-physiological electron transfer partners only able to support low levels of activity, if at all [18]. Identifying the natural electron transfer partners often increases product formation. Additionally, it

allows detailed investigation into the mechanisms by which CYP enzymes interact with the ferredoxins [18, 19].

The *Mycobacterium marinum* M genome encodes 47 cytochrome P450 enzymes. It is a pathogenic bacterium which affects frogs and fish and in humans is the cause of aquarium granuloma. It has a high degree of genetic similarity to both *Mycobacterium ulcerans* (97% nucleotide identity) and *Mycobacterium tuberculosis* (85% nucleotide identity) the human pathogens responsible for the Buruli ulcer (a tropical skin disease) [20] and tuberculosis, respectively [21-25]. *M. marinum* is a less specialised pathogen than either *M. tuberculosis* or *M. ulcerans*, having undergone less genome reduction. Consequently more *cyp* gene sequences are present in its genome [21, 26, 27]. Due to the minimal number of gene deletions, *M. marinum* also has a greater number of ferredoxin and ferredoxin reductase electron transfer partner genes than either *M. tuberculosis* or *M. ulcerans* and several of these are located next to or close to the *cyp* genes [23, 28]. The larger number of CYPs present, similar to the average number found in soil-living *Mycobacteria* (50 CYPs) [29], likely support a more varied range of reactions than those in the specialised pathogens and presumably are important for the ability of the organism to survive inside and outside of its host. Previous study has revealed the substrate range of one such enzyme from *M. marinum* M, CYP268A2 [30]. CYPs from *M. tuberculosis* and *M. ulcerans* have been identified as targets for inhibition as they are often involved in key metabolic roles [31, 32].

The number of electron transfer partner genes such as ferredoxin and ferredoxin reductase proteins in a given genome tends to be smaller than the number of CYPs, inferring they are able to support multiple CYP enzymes [33, 34]. There are 12 gene sequences encoding small ferredoxin proteins, each containing a single metal cluster, in the genome of *M. marinum* M. One of these encodes a [2Fe-2S] ferredoxin and is associated with genes for the alkane hydroxylase, CYP153A16 and a ferredoxin reductase [35]. The other 11

ferredoxins have sequence similarities to [3Fe-4S] or [4Fe-4S] cluster ferredoxins but have alternate amino acids at one position of the iron sulfur cluster motif (which is ordinarily an alanine or glycine residue in [3Fe-4S] clusters, CXXA/GXXC(X)_nCP). As we have reported previously, several of the *M. marinum* ferredoxins have been determined to contain [3Fe-4S] clusters with highly variant reduction potentials [36]. In the genome, nine of the CYP genes from *M. marinum* M are associated with one of these ferredoxins. CYP147G1 (*Mmar_2930*) is the only CYP gene to have both a ferredoxin of this type (Fdx3, *Mmar_2932*) as well as a reductase gene (FdR1, *Mmar_2931*) co-located.

As a result of the sequence similarity between the strains of *Mycobacterium* and the more easily interpretable electron transfer protein network, study of *M. marinum* CYPs could give insight into the metabolism of the human specific pathogens. We have previously shown CYP147G1 can oxidise undecanoic to 10-hydroxyundecanoic acid when supported by FdR1 and Fdx3 [36]. This was in contrast to the only other characterised family member (CYP147F1), from *Streptomyces peucetius*, which unselectively hydroxylated dodecanoic acid at the ω -1, ω -2 and ω -3 positions [11]. The established native electron transfer chain makes CYP147G1 an ideal target for further analysis. Here we report the widespread nature of this system across different bacteria, characterisation of the substrate range of the enzyme and the investigation of its selectivity for C–H bond abstraction with fatty acids.

2. Experimental

2.1 General

All organic substrates, derivatisation agents and other general reagents, except where otherwise noted were purchased from Sigma-Aldrich, Alfa-Aesar, VWR International or Tokyo Chemical Industry. Branched fatty acids were from Larodan (Sweden). Antibiotics, detergents, DTT and IPTG were from Astral Scientific. The media for cell growth and maintenance (LB, 2xYT, SOC, EMM and trace elements) were prepared as reported previously [30]. Antibiotics were added to the following working concentrations; ampicillin, $100 \mu\text{g mL}^{-1}$ and kanamycin, $30 \mu\text{g mL}^{-1}$. UV-Visible spectra were recorded on a Varian Cary 5000 or Cary 60 spectrophotometer at $30 \pm 0.5 \text{ }^\circ\text{C}$. GC-MS analysis was performed using a Shimadzu GC-17A equipped with a QP5050A MS detector and DB-5 MS fused silica column (30 m x 0.25 mm, 0.25 μm) or a Shimadzu GC-2010 equipped with a QP2010S GC-MS detector, AOC-20i autoinjector, AOC-20s autosampler and DB-5 MS fused silica column (30 m x 0.25 mm, 0.25 μm). For both, the injector was held at $250 \text{ }^\circ\text{C}$ and the interface at $280 \text{ }^\circ\text{C}$. Column flow was set at 1.5 mL min^{-1} and the split ratio was 24. For fatty acid substrates, the initial oven temperature was $120 \text{ }^\circ\text{C}$ which was held for 3 min, before increasing to $220 \text{ }^\circ\text{C}$ at $7 \text{ }^\circ\text{C min}^{-1}$, where it was held for 7 min. NMR was performed using an Agilent DD2 spectrometer at 500 MHz for ^1H and 126 MHz for ^{13}C .

2.2 CYP147G1 production and purification

CYP147G1 was produced and purified from *E. coli* as reported previously [36] and stored at $-20 \text{ }^\circ\text{C}$ after addition of an equal volume of 80% glycerol. Before use, glycerol in stored protein samples was removed via buffer exchange into 50 mM Tris (pH 7.4) using a PD-10 desalting column (5 mL, GE Healthcare). The CYP147G1 concentration was determined using $\epsilon_{419} = 111 \pm 4 \text{ mM}^{-1} \text{ cm}^{-1}$ [36].

2.3 Spin-state shift assays and dissociation constant determination

CYP147G1 was diluted to approximately 1 μM using 50 mM Tris buffer (pH 7.4) and the spectrum was recorded between 600 and 250 nm on a UV-Vis spectrophotometer at 30 $^{\circ}\text{C}$. Aliquots (1 to 5 μL) of substrate stock solutions (50 mM to 100 mM, DMSO or EtOH) were added and the spectrum recorded until the shift from 420 nm to 390 nm reached a stable point. The ratio of high spin to low spin CYP (390 nm peak to 420 nm peak) was estimated (to $\pm 5\%$) as described previously [36].

To measure the binding affinity, CYP147G1 was diluted to 2 μM in a volume of 2.5 mL in the same buffer and this sample was used to baseline the spectrophotometer. Varying aliquots (1 to 3 μL) of substrate stock solutions (1 mM, 10 mM or 100 mM concentrations) were added via a Hamilton syringe. The sample was mixed and the difference spectrum was recorded between 300 nm and 600 nm. Aliquots of substrate were added until no further change in the peak-to-trough ratio at 420 nm and 390 nm for a Type I spectrum (or 410 and 430 nm for a Type II spectrum) was observed. The difference in absorbance versus substrate concentration was fitted to the hyperbolic function (Equation 1):

$$\Delta A = \frac{\Delta A_{max} \times [S]}{K_d + [S]}$$

where K_d is the dissociation constant, $[S]$ is the substrate concentration, ΔA the peak-to-trough ratio, and ΔA_{max} the maximum peak-to-trough absorbance. In the instances where the substrate exhibited tight binding ($K_d < 10 \mu\text{M}$, less than five times the concentration of the enzyme), the data were instead fitted to the tight-binding quadratic equation:

$$\Delta A = \Delta A_{max} \times \frac{[E] + [S] + K_d - \sqrt{([E] + [S] + K_d)^2 - 4[E][S]}}{2[E]}$$

where K_d is the dissociation constant, $[S]$ is the substrate concentration, ΔA the peak-to-trough ratio, ΔA_{max} the maximum peak-to-trough absorbance and $[E]$ is the enzyme concentration [37].

2.4 Whole-cell oxidation turnovers

Whole-cell turnovers with the enzymes CYP147G1, Fdx3 and FdR1 were performed in *E. coli* with the substrates added to a final concentration of 1 mM over 24 hours, as per the method previously reported [36]. After extraction, the fatty acid samples were derivatised with BSTFA/TMSCl before analysis by GC-MS. For the larger scale extraction of dodecanoic acid, 200 mL of the supernatant was acidified with 3 M HCl to pH 2, extracted three times with an equal volume of ethyl acetate. Extracts were washed with water and saturated brine solutions, combined and dried over MgSO₄ and the solvent was removed under reduced pressure.

2.5 Product analysis

Enzyme turnover analysis was performed by GC-MS. Where GC-MS indicated a product:substrate ratio of $\geq 95\%$, reverse phase solid phase extraction (SPE) DSC-18 SPE tubes (Supelco) were used to isolate all fatty acid compounds, using the method described by Horak *et al* with minor modifications [38]. SPE columns were activated with methanol (3 mL), washed with water (5 mL) and the extract was dissolved in the minimum amount of methanol and water (200 μ L) and loaded onto the column. The column was washed with 5 mL 10% v/v methanol solution followed by 5 mL of 20% v/v methanol:water. The acid products were eluted with 600 μ L methanol and the elution was dried under a flow of nitrogen and dissolved in deuterated chloroform, 0.8 mL, before characterisation by NMR.

2.6 Phylogenetic analysis

Sequences were obtained from the databases at the National Centre for Biotechnology Information (NCBI) or Dr Nelson P450 homepage for bacterial P450s [39]. Sequence alignments were performed using ClustalW [40]. The evolutionary history was inferred by using the Maximum Likelihood method based on the Jones-Taylor-Thornton (JTT) matrix-based model [41]. Initial tree(s) for the heuristic search were obtained automatically by

applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA6 [42].

3. Results and Discussion

3.1 Phylogenetic analysis of CYP147G1

The CYP147 family of enzymes is prevalent in bacteria with CYP147A1 being first identified in *Myxococcus xanthus* and the subsequent discovery of enzymes from other subfamilies; CYP147B1 (*Streptomyces avermitilis*), CYP147C1 (*Streptomyces tubercidus*), CYP147D1 (*Magnetospirillum magnetotacticum*), CYP147E1 (*Methanosarcina barkeri*), CYP147F1 (*Streptomyces peucetius*) and CYP147G1, the subject of this manuscript from *M. marinum* [39]. The only CYP147 family member to be characterised so far is CYP147F1, from *Streptomyces peucetius* which was reported to be a fatty acid hydroxylase [11].

CYP147G1 shares 43% sequence identity with CYP147F1. A BLAST search revealed several CYP enzymes from the CYP147G subfamily, mostly from other strains of *Mycobacteria* including *M. vanbaalenii* PYR-1 (68% sequence identity), *M. kansasii* (84%), *M. rhodesiae* (72%), and *M. liflandii* (97%). While no copy of the gene is present in *M. ulcerans* Agy99, an enzyme of high similarity is present in the strain *M. ulcerans* subsp. *shinshuense* (98%). However no subfamily homologues are present in *M. tuberculosis* or *M. smegmatis* and CYP147G1 shares no more than 34% similarity with any enzyme from *M. tuberculosis*. Many of the CYP147 enzymes found in the BLAST search were from *Nocardia* (all had sequence similarity between 70 and 65% e.g. *Nocardia vinacea* 70%, Fig. 1 and Table S1) and *Streptomyces* species (all had 50% similarity and below, e.g. *S. tubercidus* 50%, Fig. 1). A phylogenetic tree revealed CYP147G1 clustered mostly with other *Mycobacterium* enzymes, with the CYP147C subfamily the closest of all the others (Fig. 1). The closest structurally characterised CYP homologue to CYP147G1 is CYP164A2 from *M. smegmatis* (35%) which is a homologue of the only CYP gene retained in *M. leprae* [43]. The next closest structurally determined analogue is P450BioI, a fatty acid cleaving CYP enzyme (32%) [44].

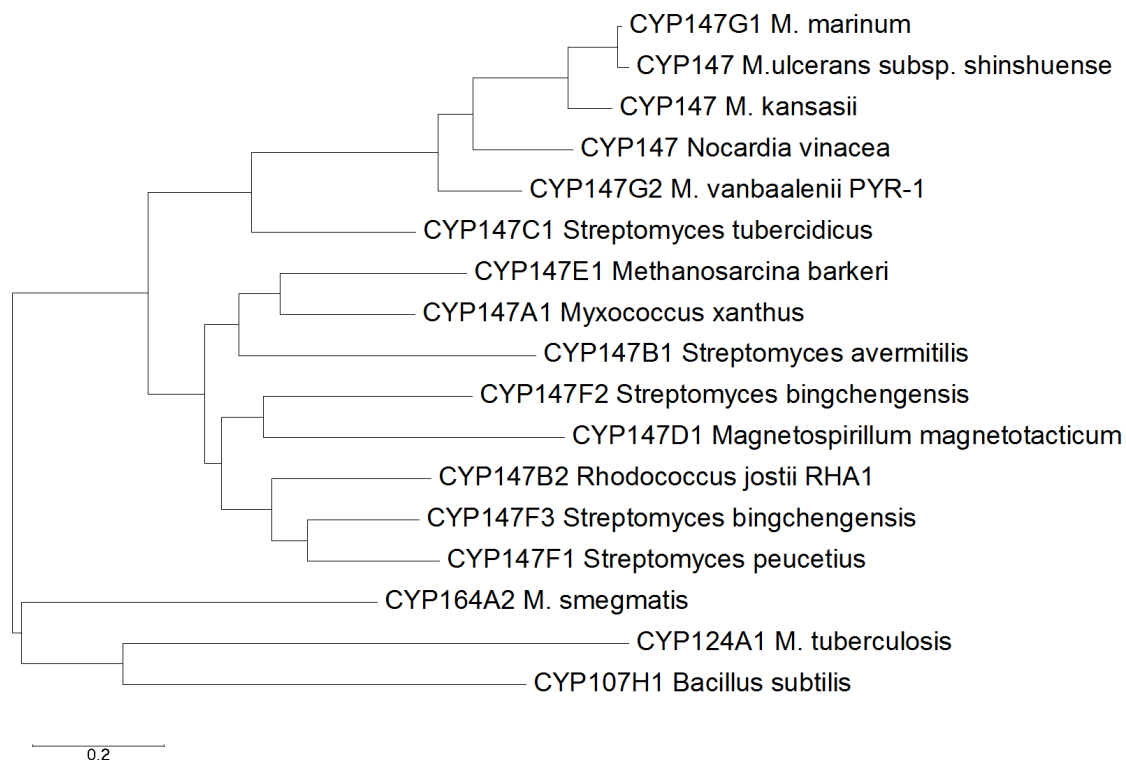


Figure 1: A phylogenetic tree (phenogram) of CYP147G1 and analogous enzymes, including the CYP147 enzymes from *M. vanbaalenii* PYR-1 (Mvan_0401) and *M. ulcerans* subsp. *shinshuense* (Shtp_2364), and other members of the CYP147 family (Table 1, see also Fig. S1 and Table S1). For comparison the closest structural homologues from the Protein Data Bank CYP164A2 from *M. smegmatis* and P450BioI (CYP107H1) from *Bacillus subtilis* and CYP124A1, a branched fatty acid binding CYP enzyme from *M. tuberculosis* have been included (Table S1). The scale shows number of substitutions per site.

Like CYP147G1, the CYP147 genes in *M. barkeri*, *M. xanthus*, *M. vanbaalenii* and *S. avermitilis* are all found next to a ferredoxin reductase encoding gene (Fig. S2) [45]. All of these, with the exception of *M. xanthus*, also have a gene which encodes a similar ferredoxin to Fdx3 present, containing a **CXXYXXC(X)_nCP** iron sulfur cluster binding motif. Many *Nocardia* and other *Streptomyces* species also have a similar arrangement of genes (Table 1, see also Table S2 and Fig. S3). The *M. xanthus* ferredoxin reductase gene is fused to a ferredoxin domain, which retains the Tyr residue in the cluster binding motif. The other exception is in *M. liflandii* and *M. ulcerans* subsp. *shinshuense*, where the FdR1-like gene has undergone a frame-shift and does not encode a full length protein (Table 1, Fig. S2). The retention of this operon across species, including several instances where the two

accompanying electron transfer proteins have fused together, indicates that it is likely that the CYP enzyme will demonstrate high specificity for the native electron transfer chain.

Table 1: Sequence identities of the CYP, Fdx and FdR equivalents of the CYP147G1/Fdx3/FdR1 operon from Mycobacterium species and elsewhere. All the ferredoxins listed contain the Tyr residue replacing the second Cys of the iron-sulfur cluster motif (CXXYXXC(X)_nC).

Species	CYP % identity	CYP name ^a	Associated Fdx % identity	Associated FdR % identity	Fusion Fdx/FdR
<i>M. ulcerans</i> subsp. <i>shinshuense</i>	98	-	99	- ^b	
<i>M. liflandii</i> 128FXT	97	-	99	- ^b	
<i>M. bohemicum</i>	85	-	94	90	
<i>M. saskatchewanense</i>	85	-	93	90	
<i>M. kansasii</i> ATCC 12478	84	-	94	94	
<i>M. gastri</i>	84	-	94	91	
<i>Nocardia vinacea</i>	70	-	77	79	
<i>M. rhodesiae</i> DSM44223	69	-	94	82	
<i>M. aromaticivorans</i>	69	-	93	83	
<i>M. vanbaalenii</i> PYR-1	68	147G2	84	81	
<i>Nocardia fusca</i>	67	-	78	75	
<i>Streptomyces tubercidus</i>	50	147C1	62	74	
<i>Myxococcus xanthus</i> DK 1622	47	147A1	62	61	Y
<i>Streptomyces avermitilis</i>	45	147B1	66	63	
<i>Streptomyces bingchenggensis</i>	45	147F2	66	59	
<i>Rhodococcus jostii</i> RHA1	45	147B2	61	61	
<i>Ktedonobacter racemifer</i>	44	-	70	63	
<i>Streptomyces bingchenggensis</i>	44	147F3	69	60	
<i>Frankia</i> sp. CN3	44	-	67	60	Y
<i>Streptomyces peucetius</i>	43	147F1	60	61	
<i>M. kansasii</i> ATCC 12478	42	-	69	60	
<i>Methanosarcina barkeri</i>	42	147E1	62	57	
<i>M. gastri</i>	41	-	72	64	

^a CYP name given in accordance with the NCBI database and Dr Nelson P450 homepage for bacterial P450s where listed [39] ^b pseudogene is present in both species, sharing 99% similarity with the FdR1 gene from *M. marinum* M.

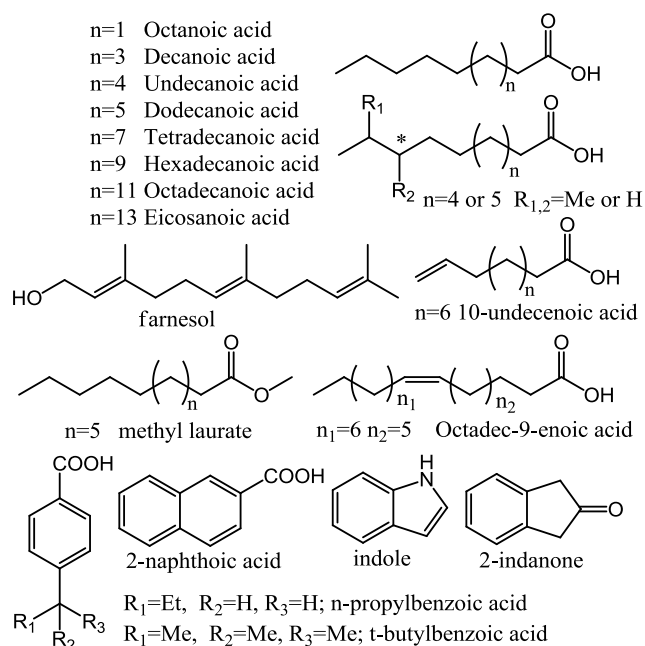
In *M. marinum*, *M. liflandii* and *M. kansasii*, the CYP operon is just downstream of a PE-PGRS protein (a family of *Mycobacterial* proteins with possible roles in fibronectin binding [46] or as antigens [47]), and a cutinase (α/β hydrolases of the plant polymer cutin, but also other substrates such as triacylglycerol [48, 49]) (Fig. S2). The CYP147 family members retain

the heme binding residues (EXXR) (Table S4). In CYP147G1, CYP147G2 and other *Mycobacterium* CYP147 enzymes, the acid alcohol pair is an aspartate and serine (D261 and S262), instead of the glutamate and threonine residues in the rest of the CYP147 family analysed (Fig. S1, Table S4). The CYP147G1 enzyme does not retain the highly conserved phenylalanine located seven residues before the proximal cysteine, instead encoding a tryptophan (W361) as does CYP147G2. This change in residue is of interest as the broad sub-terminal hydroxylase CYP267A1 from *Sorangium cellulosum*, contains a leucine at this position. Mutation of this residue to a phenylalanine shifted the hydroxylation pattern of the enzyme with fatty acids, and increased the selectivity for the ω -1 to ω -3 positions [50].

3.2 Substrate binding studies

We have previously shown that CYP147G1 can oxidise indole to indigo and undecanoic acid to 10-hydroxyundecanoic acid. However, the addition of indole and related substances such as indanone did not result in any significant shift in the spin-state of CYP147G1 (Table 2). Bhattarai *et al.* demonstrated that dodecanoic acid, myrsitic acid and palmitic acid bound to and were oxidised by CYP147F1. CYP147F1 displayed the strongest binding affinity to dodecanoic acid (with a reported spin state shift of $\geq 95\%$ and a K_d value of 0.22 μM) [11]. The CYP147G1 enzyme was produced in *E. coli* and tested with a range of fatty acids from C_{22:0} to C_{8:0} (Table 2 and Scheme 1). Addition of dodecanoic acid resulted in a spin state shift equal to that of undecanoic (40% high spin for both). A lower shift was observed for acids with successively longer or shorter chain lengths. Decanoic acid induced a 30% high spin state, whereas only a 15% shift was observed with octanoic acid. Similarly, tetradecanoic acid and hexadecanoic acid, resulted in 35% and 30% shifts, respectively. Unsaturated substrates were equally as effective at inducing the high spin state as their saturated counterparts. 10-Undecenoic acid showed a spin state shift of 40%, equal to that of

undecanoic acid. Octadec-9-enoic acid (C_{18:1}) and octadecanoic acid (C_{18:0}) also recorded the same spin state shift (20%).



Scheme 1: A selection of the substrates tested on CYP147G1. * Indicates a chiral centre in the substrate when R₂ = Me.

A range of methyl-branched fatty acids and derivatives were also tested with the enzyme, as they are known substrates of other Mycobacterial CYPs such as CYP124A1 and CYP268A2 [30, 51]. 10-Methylundecanoic and 11-methyldodecanoic acids both gave spin state shifts equal to that of the non-branched acid of the same length (40%). A methyl branch at the ω-2 position, as in 9-methylundecanoic and 10-methyldodecanoic acids, reduced the proportion of the high spin state of the enzyme induced on substrate addition (15% for both) relative to undecanoic and dodecanoic acid. 3,7-Dimethyloctanoic induced almost no shift (~5%). Longer multiply-branched substrates such as farnesol (55%) and phytol (60%) were also effective at displacing the coordinating water ligand, despite not having an acid terminus. Similarly, farnesyl acetate induced a spin state shift of 60% in CYP147G1. Additionally, dodecyl acetate gave a shift of 35% but dodecanol, 1-dodecene and dodecamide all failed to generate any spin state shift in CYP147G1. Dodecyl amine gave a red shifted

Soret absorbance (417 to 421 nm) and Type II difference spectrum (Fig. 2). We also tested aromatic benzoic and naphthoic acid substrates such as 4-isopropylbenzoic and 2-naphthoic acid but the binding of these substrates resulted in low shifts of CYP147G1 to the high spin state (Table 2). The hydrophilic terminal moiety appears to have a strong effect on the enzyme's substrate affinity.

Table 2: Substrate binding and product formation data for CYP147G1. The results are presented in descending order of the magnitude of the induced shift to the high spin state. Spin state shifts recorded as the percentage of CYP in the high spin (HS) state. The dissociation constants, K_d , are also reported where measured.

Substrate	% HS	K_d (μ M)	Major product(s) ^a
Farnesyl acetate	60	1.2 ± 0.3	-
Farnesol	55	2.0 ± 0.3	-
Phytol	50	*	-
10-Methylundecanoic acid	40	55 ± 18	ω , ω -1
10-Undecenoic acid	40	25 ± 4	-
11-Methyldodecanoic acid	40	57 ± 6	ω , ω -1
Dodecanoic acid	40	19 ± 2	ω -1 ^b
Undecanoic acid	40	25 ± 3	ω -1 ^b
Dodecyl acetate	35	2.5 ± 0.3	-
Tetradecanoic acid	35	*	ω -1
Decanoic acid	30	32 ± 6	ω -1
Hexadecanoic acid	30	*	ω -1
Methyl laurate	20	-	ω -1 [†]
Octadecanoic acid	20	-	ω -1
10-Methyldodecanoic acid	15	-	ω -1
9-Methylundecanoic acid	15	-	ω -1
Octanoic acid	15	-	ω -1
3,7-Dimethyloctanoic acid	~ 5	-	ω , ω -1
Dodecyl amine	Type II	1.8 ± 0.1	-
Econazole	Type II	1.0 ± 0.1	-
Miconazole	Type II	0.3 ± 0.08	-

(*) indicates a dissociation constant could not be accurately determined. (-) not determined (†) uncertain if this product arose from enzyme oxidation of this substrate, see main text for details ^a ω -1 is present at $\geq 98\%$; ω , ω -1 together make up $\geq 98\%$ of the total product in roughly equal amounts (see main text for details) ^b ω -2 is present as the minor product (<2%).

Significant spin state shifts (>30%) indicated the substrate may be complementary to the active site of CYP147G1, and hence binding of these substrates was further investigated by determining the dissociation constant (Fig. 2).

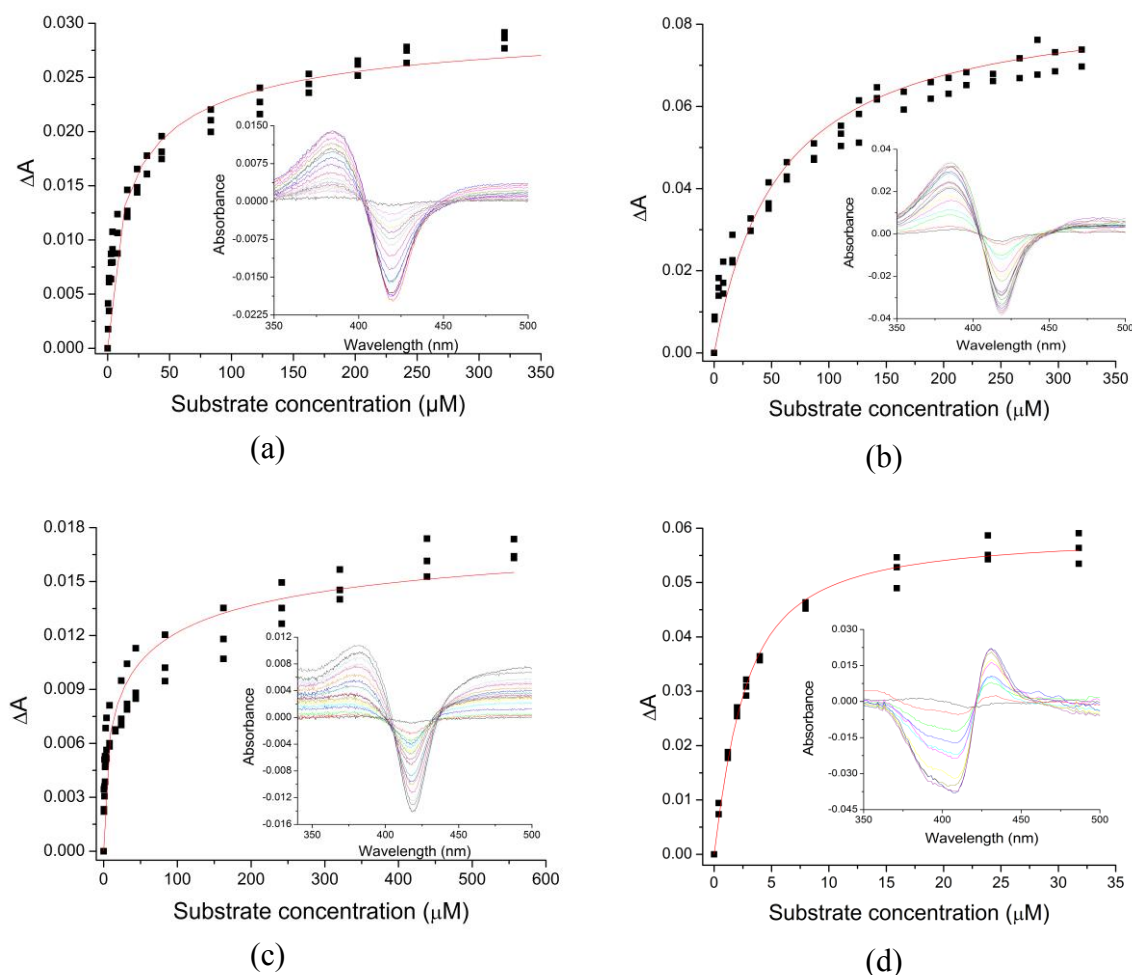


Figure 2: A selection of substrate dissociation constants for CYP147G1. The inset represents a typical substrate binding titration. Substrates shown are (a) dodecanoic acid (b) 11-methyldodecanoic acid (c) decanoic acid (d) dodecyl amine. The peak to trough difference in absorbance was measured from 420 nm to 390 nm in each, except with dodecyl amine (410 to 430 nm) where a Type II shift was recorded.

Dodecanoic acid ($K_d = 19 \pm 2 \mu\text{M}$) had a higher binding affinity than the C_{11:0} and C_{11:1} acids, undecanoic and 10-undecenoic ($25 \pm 3 \mu\text{M}$ [36] and $25 \pm 4 \mu\text{M}$ respectively). All three had higher binding affinity than decanoic acid ($32 \pm 6 \mu\text{M}$). The determination of an accurate dissociation constant for the longer hexadecanoic and tetradecanoic acids was prevented by

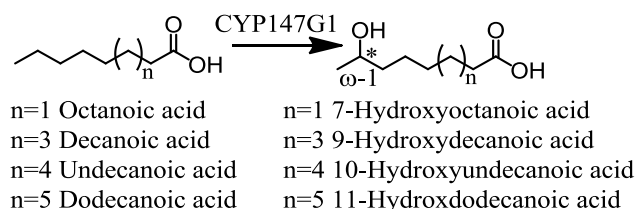
substrate insolubility at the higher concentrations necessary for an accurate fit (suggesting low affinity binding). Dodecyl acetate ($2.5 \pm 0.3 \mu\text{M}$) bound tightly as did farnesol ($2.0 \pm 0.3 \mu\text{M}$) and farnesyl acetate ($1.2 \pm 0.3 \mu\text{M}$). 11-Methyldodecanoic acid ($57 \pm 6 \mu\text{M}$) bound less tightly than its linear C_{12} counterpart, as did 10-methylundecanoic acid ($54 \pm 18 \mu\text{M}$).

In addition to dodecyl amine, knownazole inhibitors econazole, ketoconazole and miconazole were tested. While ketoconazole did not appear to bind to CYP147G1, both econazole and miconazole generated Type II shifts with the enzyme (the Soret absorbance was shifted to 423 nm and 420 nm, respectively), and were further investigated as potential inhibitors. Miconazole bound more tightly than econazole ($0.2 \pm 0.06 \mu\text{M}$ and $1 \pm 0.2 \mu\text{M}$, respectively), while dodecyl amine also bound with high affinity ($1.8 \pm 0.1 \mu\text{M}$).

3.3 Product characterisation

As the electron partner proteins Fdx3 and FdR1 are not able to be produced and purified in significant quantities by standard procedures using *E. coli*, it was necessary to use whole-cell oxidation systems to investigate product formation. As previously reported, all three enzymes were produced in *E. coli* using Duet vectors [36]. CYP147G1 product formation was undertaken with the same method used previously [36]. Fatty acids varying from 8 to 16 carbons in length were tested with the enzyme. We also tested several non-physiological electron transfer partners with CYP147G1 and all displayed significantly reduced or no metabolite formation (Fig. S4). Octanoic, decanoic and dodecanoic acids all generated a single major metabolite in good yield (Fig. 3). In all instances the MS indicated that this arose from hydroxylation (a mass of doubly-derivatised TMS metabolite -15, -Me, was observed, Fig. S7). As with undecanoic acid, the major product in each case can be assigned as the ω -1 hydroxylation metabolite due to the presence of the mass spectrum peak at 117 m/z (corresponding to the loss of a $\text{CH}_3\text{CO}(\text{SiMe}_3)^+$ cation). Furthermore, the major product of the turnover of dodecanoic acid was purified and characterised by NMR to confirm the

assignment (Fig. S14). A minor product (< 2%) was detected in the dodecanoic acid turnover which could be assigned as the ω -2 hydroxylation product, due to the peak at 131.05 m/z, indicating the loss of a terminal $\text{CH}_3\text{CH}_2\text{CO}(\text{SiMe}_3)^+$ group. A small peak with a similar mass spectrum was also present in the turnover of undecanoic acid (Fig 3). Turnovers of tetradecanoic and hexadecanoic acid formed only a small amount of product. However GC-MS analysis showed that tetradecanoic (RT 14.1 min, Fig. S9) was converted to 13-hydroxytetradecanoic acid (18.2 min) and hexadecanoic (17.0 min, Fig. S10) to 15-hydroxyhexadecanoic acid (20.7 min). CYP147G1 hence maintains its selectivity for the ω -1 position no matter the length of the fatty acid, and is capable of oxidising acids of chain lengths from 8 to 16 carbon atoms (Scheme 2). There was a preference shown for those which were 10-12 carbons long. The level of product generated decreased with fatty acids above 14 carbons in size in line with the reduced spin state shift, though this could be due to decreased solubility and uptake by the cells during the whole-cell turnovers.



Scheme 2: CYP147G1 oxidation of octanoic acid, decanoic acid, undecanoic acid and dodecanoic acid to 7-hydroxyoctanoic acid, 9-hydroxydecanoic acid, 10-hydroxyundecanoic acid and 11-hydroxydodecanoic acid, respectively. * Indicates a new chiral centre from the reaction.

The selectivity of the enzyme for the ω -1 position was greatest for the shorter fatty acids (8 – 10 carbons long), where no minor product peaks were detectable. The only other hydroxyl metabolite detected with any of the linear fatty acids was ω -2 (\leq 2%), with no ω , ω - 3 or ω -4 hydroxylation products observed in any of the turnovers.

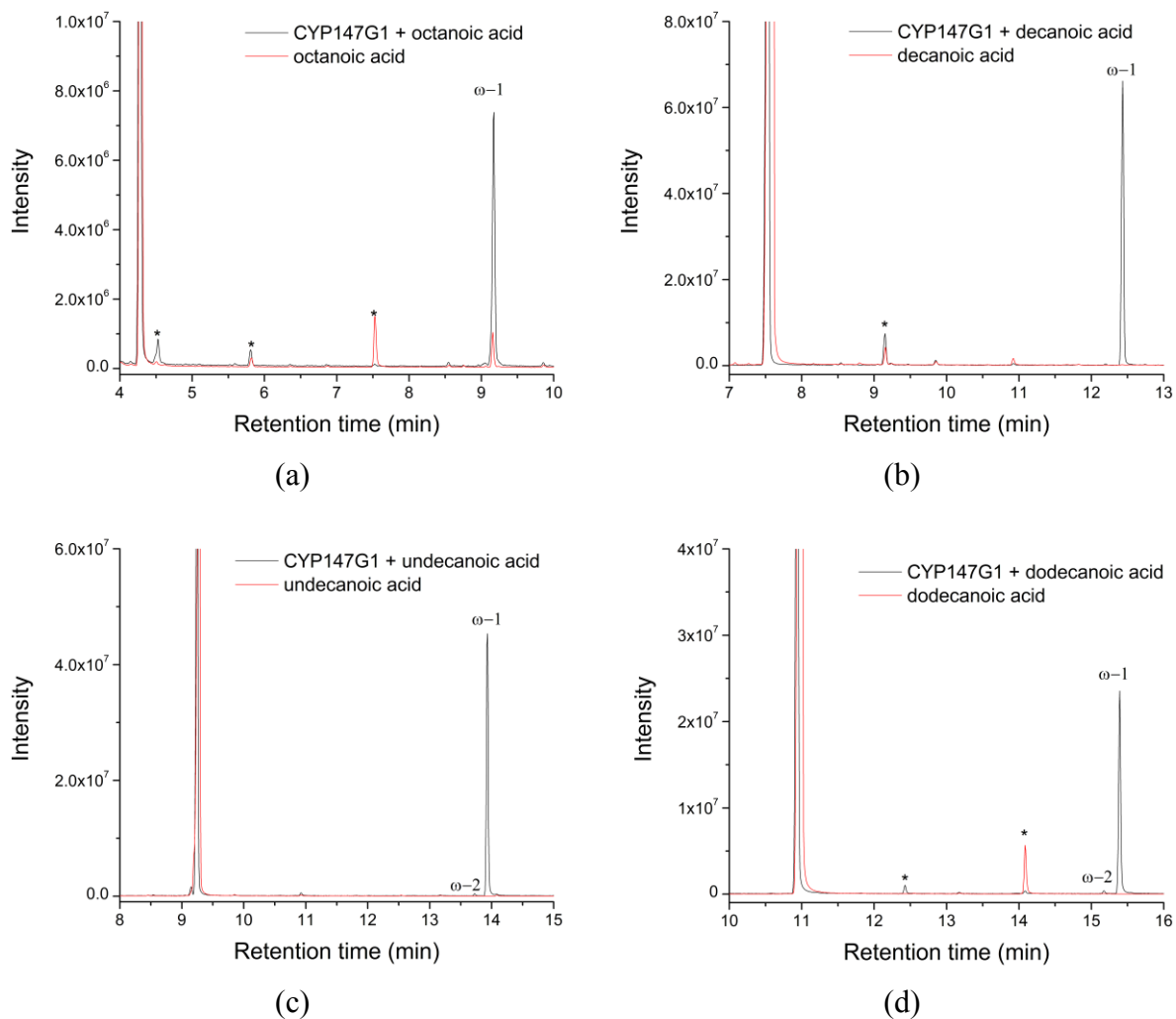


Figure 3: GC-MS chromatogram of in vivo turnovers of CYP147G1 after 4 hours with (a) octanoic acid, (b) decanoic acid, (c) undecanoic acid and (d) dodecanoic acid with substrate-only controls for each. Retention times are as follows: octanoic acid substrate 4.3 min, ω -1 product 9.1 min; decanoic acid substrate 7.5 min, ω -1 product 12.4 min; undecanoic acid substrate 9.25 min, ω -2 minor product 13.7 min, ω -1 major product 13.9 min; dodecanoic acid substrate 11.0 min, ω -2 minor product 14.5 min, ω -1 major product 15.4 min. * Indicates an impurity as determined by GC-MS as (a) decanoic acid (RT 7.5 min) (b) undecanoic acid (RT 9.25 min) (d) decanoic acid hydroxylation product (RT 12.4 min) and tetradecanoic acid (RT 14.1 min, Fig. S9) or unidentified.

CYP147F1, in contrast, was found to hydroxylate fatty acids at multiple sub-terminal positions and the product distribution varied with the chain length. For example, dodecanoic acid was oxidised at ω -1, ω -2 and ω -3 (favouring the ω -2), while a greater number of metabolites were observed for tetradecanoic acid (ω -1 through to ω -5 in approximately equal

proportions). The site of oxidation changed and moved away from the ω -terminus with hexadecanoic acid (ω -5, ω -6 and ω -7) [11].

To further probe the selectivity of the enzyme for C–H bond abstraction, the methyl branched fatty acids were tested as substrates. 11-Methyldodecanoic and 10-methyldodecanoic acid (ω -1 and ω -2 methyl groups) were both hydroxylated by CYP147G1 (Fig. 4(a)). 10-Methyldodecanoic acid gave only one product, the ω -1 hydroxy metabolite, which was identified by GC-MS (RT 15.4 min, Fig. 4 and S7). 11-Methyldodecanoic acid was turned over to two products in almost equal amounts (a ratio of 47% to 53%; RT 15.35 and 16.0 min, Fig. 4 and S7). These were determined to be the ω -1 (the characteristic fragment is 131.05 m/z from a $\text{CH}_3\text{CCH}_2\text{OSiMe}_3^+$ fragment) and the ω (characterised by a 103.0 m/z fragment of CHOSiMe_3^+), respectively. Similar results were recorded with 10-methylundecanoic and 9-methylundecanoic acid (Fig. 4(b) and S7). The major product of 9-methylundecanoic acid (RT 14.1 min) was the ω -1 hydroxylation product, although a small peak (RT 14.4 min, < 3% total product) corresponding to a methyl hydroxy metabolite could be seen (note that it is not possible to determine if this is hydroxylation at the methyl branch or the ω terminus, which are undistinguishable by MS alone). 10-Methylundecanoic acid gave two products, at RT 14.0 and 14.75 min, being the ω -1 and the ω hydroxy metabolite. The product ratio was comparable to that found for 11-methyldodecanoic acid, with a slight preference for the ω position (47% to 53%). With 3,7-dimethyloctanoic acid, two product peaks were observed by GC-MS (RT 10.1 min and 11.0 min) compared to the sole peak arising from the turnover of octanoic acid. These were also characterised as the ω and ω -1 products, respectively, with a ratio of 41% to 59% again in favour of ω -hydroxylation. Collectively the results demonstrate that the additional methyl group at the ω -1 position promoted the formation of the ω product (Scheme 3).

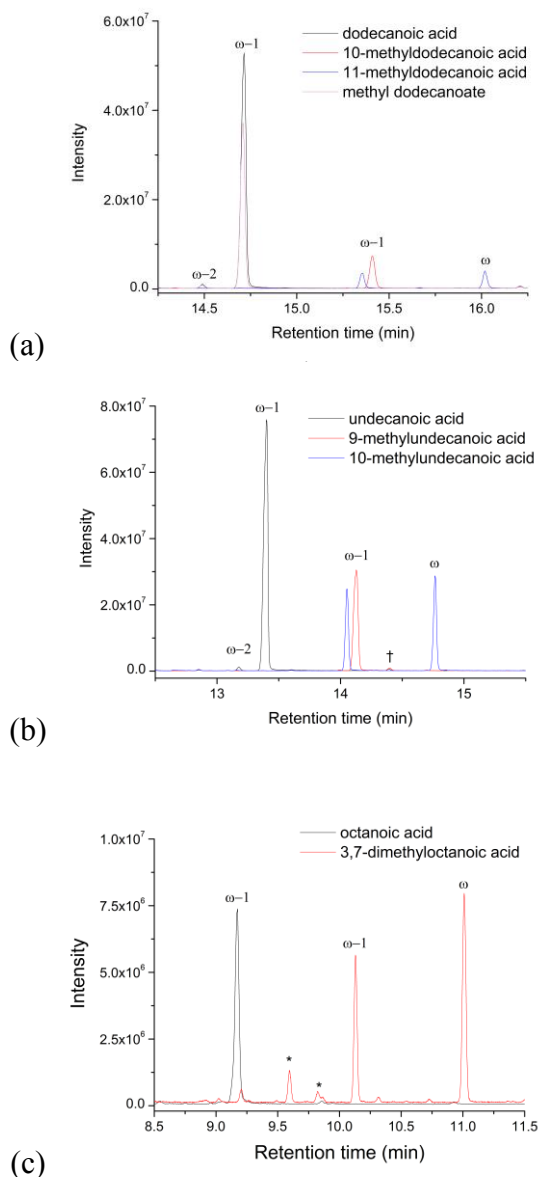
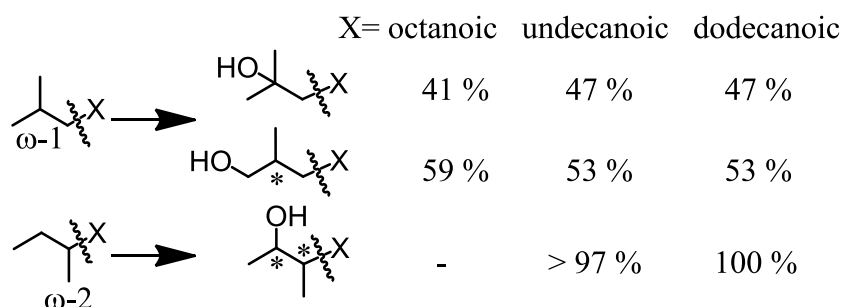


Figure 4: Expansion of the region where the products elute in the GC-MS chromatograms of the in vivo turnovers of CYP147G1 with (a) dodecanoic acid ($\omega-1$, RT 14.7 min), 10-methyldodecanoic acid ($\omega-1$, RT 15.4 min) and 11-methyldodecanoic acid ($\omega-1$ and ω , RT 15.35 and 16.0 min), and methyl dodecanoate ($\omega-1$, RT 14.7 min) (b) undecanoic acid ($\omega-2$ and $\omega-1$, RT 13.2 and 13.4 min), 9-methylundecanoic acid ($\omega-1$ RT 14.1, \dagger is probable ω , 14.4 min), 10-methylundecanoic acid ($\omega-1$ and ω , RT 14.0 and 14.75 min) and (c) octanoic acid ($\omega-1$, RT 9.2 min) and 3,7-dimethyloctanoic acid ($\omega-1$ and ω , RT 10.1 and 11.0 min). * Indicates an unidentified impurity as determined by MS.

From the regioselectivity of the enzyme with the linear fatty acids, it can be inferred that the ω terminus of the substrate is held close to the heme. The methyl group may be sequestered in a cleft, preventing ω -hydroxylation, and arranging the rest of the fatty acid chain to promote C-H bond abstraction at the $\omega-1$ carbon. The lack of hydroxylation products

at the energetically equivalent secondary carbons along the chain (ω -2, ω -3 etc) infers that these are further from the heme-iron. For the methyl branched (ω -1) fatty acids, if the terminal methyl group is sequestered in a cleft near the Fe of the heme, the other methyl group would be positioned close enough to compete with the ω -1 carbon. The product distribution with the ω -1 methyl branched fatty acids shows a slight preference for CH_3 hydroxylation, despite the decreased reactivity of this position owing to the higher C-H bond strength of the primary carbon ($104 \text{ kcal mol}^{-1}$ compared to 92 kcal mol^{-1} for a tertiary carbon [52]). It is possible that the product distribution for these prochiral ω -1 branched fatty acids could be rationalised by the binding of the different terminal methyl groups in the active site cleft as each would place either the C-H or C- CH_3 bonds of the ω -1 carbon closer to the heme. The ω hydroxylation products would be chiral and the determination of the enantioselectivity would be informative.



Scheme 3: CYP147G1 oxidation of 3,7-dimethyloctanoic acid, 10-methylundecanoic acid, 9-methylundecanoic acid, 11-methyldodecanoic acid and 10-methylundecanoic acid to their respective products. A methyl group at the ω -1 position on the substrate promotes the formation of the ω hydroxylation product by CYP147G1 while a methyl group at ω -2 does not alter the regioselectivity of the enzyme. * Indicates a chiral centre either in the substrate or introduced after hydroxylation by the enzyme. The stereoselectivity of these fatty acid metabolites have not yet been determined.

In general, the yield of product formation decreased with the methyl-substrates. The combined integrated product peaks of 10-methyl and 9-methylundecanoic acid amounted to 53% and 32%, respectively, of the products of undecanoic acid, while the 11-methyl and 10-methyldodecanoic acid products totalled only 11% and 13% of the dodecanoic acid product

(all run on equivalent aliquots of the same cell growths and presented relative to an internal standard, Fig. 4).¹ The reduced product formation is consistent with the increased steric imposition of the additional methyl group in the substrate, at what appears to be a tightly controlled position near the heme. The binding analysis of the ω -1 branched fatty acids supports this, showing weaker binding than with the linear chain acids.

The ω -2 branched fatty acid substrates contain a stereocentre and were supplied as a racemic mixture. Therefore hydroxylation of these substrates at the ω -1 carbon could generate diastereomers (Scheme 3). The binding affinity and oxidation of one of the enantiomers of the ω -2 methyl branched fatty acids could be preferred over the other. Only one peak was visible in the GC-MS analysis of the turnovers, although it is possible that multiple diastereomers were present but not separated.

A range of other substrates were tested for activity with the enzyme whole-cell system. Methyl dodecanoate (the methyl ester of dodecanoic acid), gave a major product from the *in vivo* turnover with the same retention time and mass spectrum as that of dodecanoic acid (Fig. 4(a) and S7), suggesting the methyl has been lost from the acetate group. However, it is not clear if it was cleaved *in vivo* before hydroxylation or after, during extraction or derivatisation. Additionally, the whole-cell system reactions with substrates such as 3,7,11-trimethyldodecanoic acid and dodecyl acetate (the dodecyl ester of acetic acid) resulted in no product formation. The turnovers of the other tightly binding substrates, including farnesol, farnesyl acetate and phytanic acid were attempted but no product could be detected. When tested *in vivo* there was also no evidence of any mono-oxygenation product from 10-undecenoic acid. Undecanoic acid and 10-undecenoic acid displayed almost

¹ The reduced product formation of the enzyme with the methyl branched substrates was calculated by comparing the branched substrates to the non-branched of the same chain length (ie. 10-methyldodecanoic to dodecanoic acid). To account for possible decreased solubility of the branched substrates given their increased mass, 9-methylundecanoic and 10-methylundecanoic acid were additionally compared to dodecanoic acid. They generated 48% and 81% of the total products of dodecanoic acid, respectively.

identical binding affinities to CYP147G1 and epoxidation or allylic hydroxylation products often result from the oxidation of alkenes by CYPs. However, terminal alkenes have also been reported to act as mechanism-based inhibitors of monooxygenases such as CYPs [53].

These data strongly suggest the ω end of the fatty acid is sequestered close by the heme of CYP147G1, regardless of chain length, promoting the regioselective hydroxylation at the ω -1 position. In the case of a tertiary ω -1 methyl branched carbon, one of the methyl groups remains sequestered, while the other must be positioned close to the iron-oxo complex, allowing hydrogen abstraction from the primary carbon and formation of the ω product. The strict regioselectivity of the enzyme may limit the substrate range. However, such selectivity for a single sub-terminal position on a saturated fatty acid is unusual, with few equivalents in the literature [14]. CYP2M1 and CYP2M17 hydroxylate dodecanoic acid solely at the ω -6 position [54]. Fungal CYP505A30 from *Myceliophthora thermophila*, which is a CPR-fused CYP similar to P450BM3, has demonstrated selectivity for the ω -1 position of the fatty acid chain (88% and 63% ω -1 product with dodecanoic acid and tetradecanoic acid, respectively) [55]. Selective oxidation of unsaturated fatty acids is more common. Two CYP102 enzymes from *Ktedonobacter racemifer* selectively hydroxylate unsaturated fatty acids. Krac_9955 produces only the ω -2 hydroxy metabolite of 10-undecenoic acid, while Krac_9936 strongly favours the ω -1 position of *cis*-9-hexadecenoic acid (93%) [56]. Several CYPs can mediate the selective hydroxylation of *cis*-9-octadecenoic acid to 12-hydroxy-*cis*-9-octadecenoic acid, and similar unsaturated substrates [14, 57]. Terminal ω hydroxylases (such as the CYP153 family) are often very regioselective. CYP153A33 from *Marinobacter aquaeolei* hydroxylates C₁₀ to C₁₈ fatty acids with >95% selectivity for the ω position. The substrate range of this enzyme was increased to incorporate octanoic acid by the G307A variant while the L354I mutant shifted the selectivity to 76% ω -1 [16]. The equivalent residues in CYP147G1 are A257 and V304, respectively. P450BM3 and other CYP102

enzymes have been targeted with a range of methyl-branched fatty acids, which improve regioselectivity up to 85% for ω -1 in the case of 12-methyltetradecanoic acid with P450BM3 [58]. CYP102A2 oxidised 13-methyltetradecanoic acid at the ω -2 position with 91% selectivity [59]. However, to our knowledge no CYP enzyme with selectivity for the same sub-terminal position across such a range of substrates as CYP147G1 demonstrates has been characterised previously.

4. Conclusions

Members of the CYP147 family are found as part of a highly conserved operon containing a ferredoxin and ferredoxin reductase in many *Mycobacterium* and other bacterial species (CYP147G1, Fdx3 and FdR1). The tyrosine residue in the cluster binding motif of Fdx3 is highly conserved in the ferredoxins found alongside other CYP147 family members. In the instances where the equivalent *cyp147G1* and ferredoxin genes had not been removed from the genome of the pathogenic *Mycobacterium* species, the gene for the ferredoxin reductase was compromised. The substrate range of CYP147G1 was determined to include fatty acids. The CYP147G1 enzyme has an optimal substrate chain length of ~11-12 carbons. Branching methyl groups at the ω -2 position of the substrate disfavoured substrate binding more than when at the ω -1 position. The activity of fatty acid oxidation by CYP147G1 was reconstituted *in vivo* using an electron transfer system consisting of the native partners. Dodecanoic acid, undecanoic acid, decanoic acid and octanoic acid were selectively hydroxylated at the ω -1 position by CYP147G1. The regioselectivity of the enzyme differed with ω -1 methyl-branched substrates, with hydroxylation occurring on one of the terminal CH₃ groups in addition to the ω -1 C-H bond. These data suggest the preferred orientation of the substrate in the CYP147G1 active site is with an ω methyl group sequestered near the heme promoting hydrogen abstraction at the ω -1 position. The hydrophilic group at the other terminus appears to be less rigidly bound as the enzyme can accommodate substrates ranging in length from 8 to 16 carbons long and with different functional groups (acid, ester, alcohol). In contrast to CYP147F1, the substrate range reported here for CYP147G1 is broader, and the demonstrated regioselectivity is not seen in the related enzyme. In addition, dodecyl amine, 10-undecenoic acid and the azoles econazole and miconazole were all identified as potential inhibitors of CYP147G1.

Acknowledgements

This work was supported by the Australian Research Council through a Future Fellowship (FT140100355 to S.G.B.) and the award of a University of Adelaide Faculty of Sciences Divisional Scholarship (PhD to S.A.C.). We are grateful to Professor Tim Stinear from the University of Melbourne and Professor Lalita Ramakrishnan of the University of Cambridge for providing the genomic DNA of *M. marinum*.

References

- [1] I.N. Van Bogaert, S. Groeneboer, K. Saerens, W. Soetaert, The role of cytochrome P450 monooxygenases in microbial fatty acid metabolism, *FEBS J*, 278 (2011) 206-221.
- [2] H. Sugimoto, Y. Shiro, Diversity and substrate specificity in the structures of steroidogenic cytochrome P450 enzymes, *Biol. Pharm. Bull.*, 35 (2012) 818-823.
- [3] S.L. Kelly, D.E. Kelly, Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us?, *Philos. Trans. Royal Soc. B*, 368 (2013) 20120476.
- [4] F.P. Guengerich, Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity, *Chem. Res. Toxicol.*, 14 (2001) 611-650.
- [5] F.P. Guengerich, Cytochrome p450 and chemical toxicology, *Chem. Res. Toxicol.*, 21 (2008) 70-83.
- [6] S.G. Bell, N. Hoskins, C.J.C. Whitehouse, L.L. Wong, Design and engineering of cytochrome P450 systems, in: *Metal Ions in Life Sciences*, vol. 3, 2007, pp. 437-476.
- [7] V.B. Urlacher, M. Girhard, Cytochrome P450 monooxygenases: an update on perspectives for synthetic application, *Trends Biotechnol.*, 30 (2012) 26-36.
- [8] J.H. Capdevila, S. Wei, C. Helvig, J.R. Falck, Y. Belosludtsev, G. Truan, S.E. Graham-Lorence, J.A. Peterson, The highly stereoselective oxidation of polyunsaturated fatty acids by cytochrome P450BM-3, *J. Biol. Chem.*, 271 (1996) 22663-22671.
- [9] M.J. Cryle, N.J. Matovic, J.J. De Voss, Products of cytochrome P450(Biol) (CYP107H1)-catalyzed oxidation of fatty acids, *Organic Letters*, 5 (2003) 3341-3344.
- [10] Y.J. Chun, T. Shimada, R. Sanchez-Ponce, M.V. Martin, L. Lei, B. Zhao, S.L. Kelly, M.R. Waterman, D.C. Lamb, F.P. Guengerich, Electron transport pathway for a *Streptomyces* cytochrome P450: cytochrome P450 105D5-catalyzed fatty acid hydroxylation in *Streptomyces coelicolor* A3(2), *J. Biol. Chem.*, 282 (2007) 17486-17500.
- [11] S. Bhattarai, K. Liou, T.J. Oh, Hydroxylation of long chain fatty acids by CYP147F1, a new cytochrome P450 subfamily protein from *Streptomyces peucetius*, *Arch. Biochem. Biophys.*, 539 (2013) 63-69.
- [12] M. Akiyama, Y. Doi, Production of poly(3-hydroxyalkanoates) from α , ω -alkanedioic acids and hydroxylated fatty acids by *Alcaligenes* sp, *Biotechnology Letters*, 15 (1993) 163-168.
- [13] U.T. Bornscheuer, *Enzymes in Lipid Modification*, *Annu. Rev. Food Sci. T.*, (2017).
- [14] L. Hammerer, C.K. Winkler, W. Kroutil, Regioselective biocatalytic hydroxylation of fatty acids by cytochrome P450s, *Catalysis Letters*, (2017).
- [15] I. Matsunaga, T. Sumimoto, A. Ueda, E. Kusunose, K. Ichihara, Fatty acid-specific, regiospecific, and stereospecific hydroxylation by cytochrome P450 (CYP152B1) from *Sphingomonas paucimobilis*: Substrate structure required for α -hydroxylation, *Lipids*, 35 (2000) 365-371.
- [16] S. Honda Malca, D. Scheps, L. Kuhnel, E. Venegas-Venegas, A. Seifert, B.M. Nestl, B. Hauer, Bacterial CYP153A monooxygenases for the synthesis of omega-hydroxylated fatty acids, *Chem. Commun.*, 48 (2012) 5115-5117.
- [17] C.J. Whitehouse, S.G. Bell, L.L. Wong, P450(BM3) (CYP102A1): connecting the dots, *Chem. Soc. Rev.*, 41 (2012) 1218-1260.
- [18] W. Yang, S.G. Bell, H. Wang, W. Zhou, N. Hoskins, A. Dale, M. Bartlam, L.L. Wong, Z. Rao, Molecular characterization of a class I P450 electron transfer system from *Novosphingobium aromaticivorans* DSM12444, *J. Biol. Chem.*, 285 (2010) 27372-27384.
- [19] S. Tripathi, H. Li, T.L. Poulos, Structural basis for effector control and redox partner recognition in cytochrome P450, *Science*, 340 (2013) 1227-1230.

- [20] K. Nakanaga, R.R. Yotsu, Y. Hoshino, K. Suzuki, M. Makino, N. Ishii, Buruli ulcer and mycolactone-producing mycobacteria, *Jpn J. Infect. Dis.*, 66 (2013) 83-88.
- [21] C. Demangel, T.P. Stinear, S.T. Cole, Buruli ulcer: reductive evolution enhances pathogenicity of *Mycobacterium ulcerans*, *Nat. Rev. Microbiol.*, 7 (2009) 50-60.
- [22] K. Roltgen, T.P. Stinear, G. Pluschke, The genome, evolution and diversity of *Mycobacterium ulcerans*, *Infect. Genet. Evol.*, (2012).
- [23] T.P. Stinear, G.A. Jenkin, P.D. Johnson, J.K. Davies, Comparative genetic analysis of *Mycobacterium ulcerans* and *Mycobacterium marinum* reveals evidence of recent divergence, *J. Bacteriol.*, 182 (2000) 6322-6330.
- [24] T.P. Stinear, T. Seemann, P.F. Harrison, G.A. Jenkin, J.K. Davies, P.D. Johnson, Z. Abdellah, C. Arrowsmith, T. Chillingworth, C. Churcher, K. Clarke, A. Cronin, P. Davis, I. Goodhead, N. Holroyd, K. Jagels, A. Lord, S. Moule, K. Mungall, H. Norbertczak, M.A. Quail, E. Rabinowitsch, D. Walker, B. White, S. Whitehead, P.L. Small, R. Brosch, L. Ramakrishnan, M.A. Fischbach, J. Parkhill, S.T. Cole, Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*, *Genome Res.*, 18 (2008) 729-741.
- [25] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry, 3rd, F. Tekaiia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M.A. Quail, M.A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J.E. Sulston, K. Taylor, S. Whitehead, B.G. Barrell, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, *Nature*, 393 (1998) 537-544.
- [26] P. Singh, S.T. Cole, *Mycobacterium leprae*: genes, pseudogenes and genetic diversity, *Future Microbiol.*, 6 (2011) 57-71.
- [27] F.J. Veyrier, A. Dufort, M.A. Behr, The rise and fall of the *Mycobacterium tuberculosis* genome, *Trends Microbiol.*, 19 (2011) 156-161.
- [28] T.P. Stinear, T. Seemann, P.F. Harrison, G.A. Jenkin, J.K. Davies, P.D. Johnson, Z. Abdellah, C. Arrowsmith, T. Chillingworth, C. Churcher, K. Clarke, A. Cronin, P. Davis, I. Goodhead, N. Holroyd, K. Jagels, A. Lord, S. Moule, K. Mungall, H. Norbertczak, M.A. Quail, E. Rabinowitsch, D. Walker, B. White, S. Whitehead, P.L. Small, R. Brosch, L. Ramakrishnan, M.A. Fischbach, J. Parkhill, S.T. Cole, Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*, *Genome Research*, 18 (2008) 729-741.
- [29] M. Parvez, L.B. Qhanya, N.T. Mthakathi, I.K.R. Kgosiemang, H.D. Bamal, N.S. Pagadala, T. Xie, H. Yang, H. Chen, C.W. Theron, R. Monyaki, S.C. Raselemane, V. Salewe, B.L. Mongale, R.G. Matowane, S.M.H. Abdalla, W.I. Booii, M. van Wyk, D. Olivier, C.E. Boucher, D.R. Nelson, J.A. Tuszyński, J.M. Blackburn, J.-H. Yu, S.S. Mashele, W. Chen, K. Syed, Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: Special focus on mycobacterial P450s, *Scientific Reports*, 6 (2016) 33099.
- [30] S.A. Child, E.F. Naumann, J.B. Bruning, S.G. Bell, Structural and functional characterisation of the cytochrome P450 enzyme CYP268A2 from *Mycobacterium marinum*, *Biochem. J.*, (2018).
- [31] S.A. Hudson, K.J. McLean, A.W. Munro, C. Abell, *Mycobacterium tuberculosis* cytochrome P450 enzymes: a cohort of novel TB drug targets, *Biochem. Soc. Trans.*, 40 (2012) 573-579.
- [32] P. Belin, M.H. Le Du, A. Fielding, O. Lequin, M. Jacquet, J.B. Charbonnier, A. Lecoq, R. Thai, M. Courcon, C. Masson, C. Dugave, R. Genet, J.L. Pernodet, M. Gondry, Identification and structural basis of the reaction catalyzed by CYP121, an essential

- cytochrome P450 in *Mycobacterium tuberculosis*, Proc. Natl. Acad. Sci. U. S. A., 106 (2009) 7426-7431.
- [33] S.G. Bell, N. Hoskins, F. Xu, D. Caprotti, Z. Rao, L.L. Wong, Cytochrome P450 enzymes from the metabolically diverse bacterium *Rhodopseudomonas palustris*, Biochem. Biophys. Res. Commun., 342 (2006) 191-196.
- [34] D.C. Lamb, T. Skaug, H.L. Song, C.J. Jackson, L.M. Podust, M.R. Waterman, D.B. Kell, D.E. Kelly, S.L. Kelly, The cytochrome P450 complement (CYPome) of *Streptomyces coelicolor* A3(2), J. Biol. Chem., 277 (2002) 24000-24005.
- [35] D. Scheps, S.H. Malca, H. Hoffmann, B.M. Nestl, B. Hauer, Regioselective omega-hydroxylation of medium-chain n-alkanes and primary alcohols by CYP153 enzymes from *Mycobacterium marinum* and *Polaromonas* sp. strain JS666, Org. Biomol. Chem., 9 (2011) 6727-6733.
- [36] S.A. Child, J.M. Bradley, T.L. Pukala, D.A. Svistunenko, N.E. Le Brun, S.G. Bell, Electron transfer ferredoxins with unusual cluster binding motifs support secondary metabolism in *Mycobacteria* and are prevalent in many other bacteria, (*Unpublished*).
- [37] J.M. Dogne, X. de Leval, P. Benoit, S. Rolin, B. Pirotte, B. Masereel, Therapeutic potential of thromboxane inhibitors in asthma, Expert Opin. Investig. Drugs, 11 (2002) 275-281.
- [38] T. Horak, J. Culik, P. Cejka, M. Jurkova, V. Kellner, J. Dvorak, D. Haskova, Analysis of free fatty acids in beer: comparison of solid-phase extraction, solid-phase microextraction, and stir bar sorptive extraction, J. Agric. Food Chem., 57 (2009) 11081-11085.
- [39] D.R. Nelson, The cytochrome p450 homepage, Human genomics, 4 (2009) 59-65.
- [40] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol. Sys. Biol., 7 (2011) 539.
- [41] D.T. Jones, W.R. Taylor, J.M. Thornton, The rapid generation of mutation data matrices from protein sequences, Comput. Appl. Biosci., 8 (1992) 275-282.
- [42] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: Molecular Evolutionary Genetics Analysis version 6.0, Mol Biol Evol, 30 (2013) 2725-2729.
- [43] A.G. Warrilow, C.J. Jackson, J.E. Parker, T.H. Marczylo, D.E. Kelly, D.C. Lamb, S.L. Kelly, Identification, characterization, and azole-binding properties of *Mycobacterium smegmatis* CYP164A2, a homolog of ML2088, the sole cytochrome P450 gene of *Mycobacterium leprae*, Antimicrob. Agents Chemother., 53 (2009) 1157-1164.
- [44] M.J. Cryle, I. Schlichting, 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 (2008) 15696-15701.
- [45] L. Overmars, R. Kerkhoven, R.J. Siezen, C. Francke, MGcV: the microbial genomic context viewer for comparative genome analysis, BMC genomics, 14 (2013) 209.
- [46] C. Espitia, J.P. Lacleite, M. Mondragon-Palomino, A. Amador, J. Campuzano, A. Martens, M. Singh, R. Cicero, Y. Zhang, C. Moreno, The PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis*: a new family of fibronectin-binding proteins?, Microbiol., 145 (Pt 12) (1999) 3487-3495.
- [47] S. Banu, N. Honore, B. Saint-Joanis, D. Philpott, M.C. Prevost, S.T. Cole, Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens?, Mol. Microbiol., 44 (2002) 9-19.
- [48] S. Chen, L. Su, J. Chen, J. Wu, Cutinase: characteristics, preparation, and application, Biotech. Adv., 31 (2013) 1754-1767.
- [49] S. Chen, X. Tong, R.W. Woodard, G. Du, J. Wu, J. Chen, Identification and Characterization of Bacterial Cutinase, J. Biol. Chem., 283 (2008) 25854-25862.

- [50] Y. Khatri, F. Hannemann, M. Girhard, R. Kappl, M. Hutter, V.B. Urlacher, R. Bernhardt, A natural heme-signature variant of CYP267A1 from *Sorangium cellulosum* So ce56 executes diverse ω -hydroxylation, *FEBS J.*, 282 (2015) 74-88.
- [51] J.B. Johnston, P.M. Kells, L.M. Podust, P.R. Ortiz de Montellano, Biochemical and structural characterization of CYP124: a methyl-branched lipid omega-hydroxylase from *Mycobacterium tuberculosis*, *Proc. Natl. Acad. Sci. U. S. A.* , 106 (2009) 20687-20692.
- [52] J.B. Johnston, H. Ouellet, L.M. Podust, P.R. Ortiz de Montellano, Structural control of cytochrome P450-catalyzed ω -hydroxylation, *Arch. Biochem. Biophys.*, 507 (2011) 86-94.
- [53] P.R. Ortiz de Montellano, B.L. Mangold, C. Wheeler, K.L. Kunze, N.O. Reich, Stereochemistry of cytochrome P-450-catalyzed epoxidation and prosthetic heme alkylation, *J. Biol. Chem.*, 258 (1983) 4208-4213.
- [54] Y.H. Yang, J.L. Wang, C.L. Miranda, D.R. Buhler, 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 (1998) 271-280.
- [55] G.J. Baker, H.M. Girvan, S. Matthews, K.J. McLean, M. Golovanova, T.N. Waltham, S.E.J. Rigby, D.R. Nelson, R.T. Blankley, A.W. Munro, Expression, Purification, and Biochemical Characterization of the Flavocytochrome P450 CYP505A30 from *Myceliophthora thermophila*, *Acs Omega*, 2 (2017) 4705-4724.
- [56] S.D. Munday, N.K. Maddigan, R.J. Young, S.G. Bell, Characterisation of two self-sufficient CYP102 family monooxygenases from *Ktedonobacter racemifer* DSM44963 which have new fatty acid alcohol product profiles, *Biochim. Biophys. Acta*, 1860 (2016) 1149-1162.
- [57] R. Holic, H. Yazawa, H. Kumagai, H. Uemura, Engineered high content of ricinoleic acid in fission yeast *Schizosaccharomyces pombe*, *Appl. Microbiol. Biotechnol.*, 95 (2012) 179-187.
- [58] M.J. Cryle, R.D. Espinoza, S.J. Smith, N.J. Matovic, J.J. De Voss, Are branched chain fatty acids the natural substrates for P450BM3?, *Chem. Comms.*, (2006) 2353-2355.
- [59] M. Budde, M. Morr, R.D. Schmid, V.B. Urlacher, Selective hydroxylation of highly branched fatty acids and their derivatives by CYP102A1 from *Bacillus megaterium*, *Chembiochem*, 7 (2006) 789-794.

Chapter 4:

The characterisation of two members of the cytochrome P450 CYP150 family: CYP150A5 and CYP150A6 from *Mycobacterium marinum*

Prepared in the style of *Biochimica et Biophysica Acta* General Subjects in preparation for submission.

Citation:

Child, S. A.; Flint, K.; Bruning, J. B.; Bell, S. G., The characterisation of two members of the cytochrome P450 CYP150 family: CYP150A5 and CYP150A6 from *Mycobacterium marinum*, manuscript in preparation.

Statement of Authorship

Principal Author

Name	Stella Child		
Contribution to the Paper	Enzyme purification and substrate characterisation, product formation studies, analysis and discussion, wrote manuscript		
Overall percentage (%)	75		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	23/5/18

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Kate Flint		
Contribution	CYP150A6 enzyme purification and substrate characterisation		
Signature		Date	23/5/18

Name of Co-Author	John Bruning		
Contribution	Supervision and advice on X-ray crystallography		
Signature		Date	23/5/2018

Name of Co-Author	Stephen Bell		
Contribution	Experimental design, supervision, manuscript preparation.		
Signature		Date	23/5/2018

The characterisation of two members of the cytochrome P450 CYP150 family:

CYP150A5 and CYP150A6 from *Mycobacterium marinum*

Stella A. Child¹, Kate L. Flint¹, John B. Bruning², and Stephen G. Bell^{1*}

¹Department of Chemistry, University of Adelaide, SA 5005, Australia

²School of Biological Sciences, University of Adelaide, SA 5005, Australia

* To whom correspondence should be addressed.

Stephen G. Bell (stephen.bell@adelaide.edu.au)

Highlights

- Two *Mycobacterium marinum* cytochrome P450s of the CYP150A subfamily were characterised
- CYP150A5 binds a broad range of terpenoids and selectively oxidised β -ionol
- CYP150A6 binds a narrower range of substrates and was structurally characterised to 1.5 Å
- Members of the CYP150A subfamily were discovered across many bacteria, including in pathogens
- Azole inhibitors that bind with different affinities to CYP150A5 and CYP150A6 were identified

Abstract

Background

Actinobacteria, including the *Mycobacteria*, have a large component of cytochrome P450 family monooxygenases. Examples include the pathogens *Mycobacterium tuberculosis*, *M. ulcerans* and *M. marinum*, and the soil-dwelling *M. vanbaalenii*. The ability of P450s to support hydroxylation of unactivated C–H bonds result in them having important roles in natural product biosynthesis and therefore being desirable biocatalysts.

Methods

Two members of the bacterial CYP150 family, CYP150A5 and CYP150A6 from *M. marinum*, were produced, purified and characterised. The substrate range of both enzymes were analysed and the monooxygenase activity of CYP150A5 was reconstituted using a physiological electron transfer partner system. CYP150A6 was structurally characterised by X-ray crystallography.

Results

CYP150A5 was shown to bind various norisoprenoids and terpenoid substrates. It could regioselectively hydroxylate β -ionol. The X-ray crystal structure of substrate-free CYP150A6 was solved to 1.5 Å. The enzyme displayed an open conformation with short F and G helices, an unresolved F-G loop and exposed active site pocket. The active site residues could be identified and important variations were found with other CYP150A enzymes. Haem-bindingazole inhibitors were identified for both enzymes.

Conclusions

Cyclic terpenoid compounds were identified as substrates for CYP150A5. The structure of CYP150A6 will facilitate the identification of physiological substrates and the design of better inhibitors for members of this P450 family. Based on the observed differences in

substrate preference and sequence variations among the active site residues, their roles are predicted to be different.

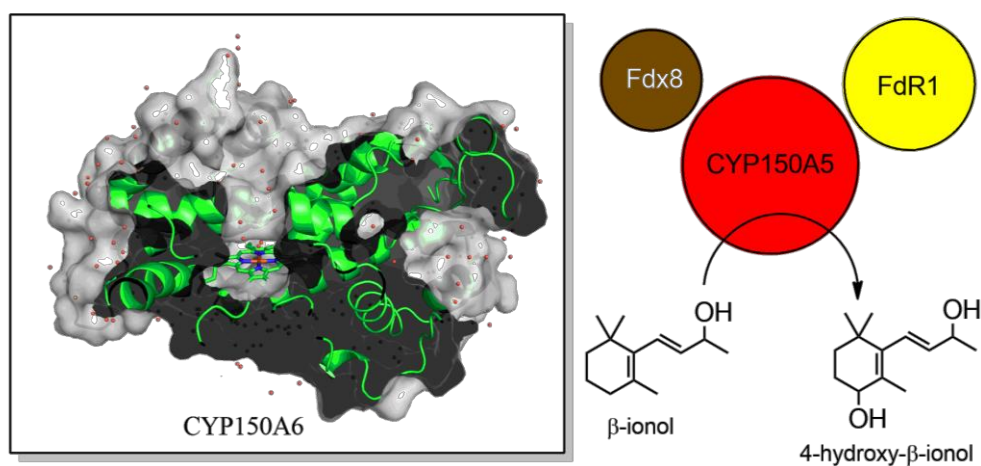
General Significance

CYP150 family members were found across many bacteria and are prevalent in the *Mycobacteria* including several human pathogens. Inhibition and structural data are reported here for these enzymes for the first time.

Key words

enzymology, electron transfer, biocatalysis, cytochrome P450, *Mycobacterium*

Graphical abstract



Abbreviations

2xYT, 2 x concentration yeast extract tryptophan broth; CYP or P450, Cytochrome P450 enzyme; DTT, dithiothreitol; EMM, *E. coli* minimal media; FAD, flavin adenosine dinucleotide; FdR, ferredoxin reductase; Fdx, ferredoxin; GC-MS or MS, gas-chromatography mass spectrometry or mass spectrometry; IPTG, Isopropyl β -D-1-thiogalactopyranoside; LB, Lysogeny broth (also known as Luria or Lennox Broth), NAD(P)H reduced nicotinamide adenine dinucleotide (phosphate); PDB, Protein Data Bank; PDR, phthalate dioxygenase reductase; RT, retention time; SOC, Super Optimal broth with Catabolite repression.

1. Introduction

Cytochrome P450s (CYP or P450) are a family of haem monooxygenase enzymes. They act with a conserved mechanism to selectively insert an oxygen atom from molecular dioxygen into a C-H bond of the substrate, forming an alcohol product [1]. Various CYP enzymes have also been found to catalyse other oxidation reactions, such as epoxidation, sulfoxidation, decarboxylation, hydrogenation, and carbon-carbon bond formation [2]. The catalytic cycle of the enzymes requires the transfer of electrons, ultimately derived from NADH or NADPH, to the haem via electron transfer proteins. Bacterial CYPs most frequently utilise a two component electron transfer system, comprised of an iron-sulfur ferredoxin and a FAD-containing ferredoxin reductase, which is termed a Class 1 system [3]. In a given genome, the number of genes tends to decrease in the order CYP > ferredoxin > ferredoxin reductase [4, 5]. The reconstruction of the native electron transfer chain for a given CYP enzyme is often required for optimal activity, as individual CYPs are often highly selective for a preferred ferredoxin [6, 7].

In humans, CYP enzymes are responsible for a large proportion of drug interactions but they are also widely found across plants, bacteria and fungi, performing both anabolic and catabolic roles [8, 9]. Due to their diversity, CYP enzymes are categorised into families, given a number, and sub-families, given a letter code, based on sequence similarities [10]. Members of the same family share >40% sequence identity, while sub-family members share >55%. Above 80% identity is sufficient for two enzymes to share a name. Bacterial CYPs have generally been investigated either as inhibition targets in pathogenic species, or as biocatalysts, as they frequently catalyse the formation of synthetically valuable compounds [9]. C-H functionalisation is particularly difficult to achieve by chemical methods, often requiring harsh conditions and resulting in poor selectivity. In contrast, CYPs often display

very high selectivity of oxidation, and regio- and stereo- selectivity in product formation is widespread [11].

Mycobacterium tuberculosis, the pathogen responsible for human tuberculosis, has been the target of ongoing drug development efforts as increasing levels of resistance to first and second-line anti-tuberculosis drugs is seen globally [12]. There are 20 CYP enzymes in *Mycobacterium tuberculosis*, several of which have been found to be essential for cell function or infectivity in the pathogen [13-15]. One of these, CYP121A1, which forms a carbon-carbon bond in the cyclic dipetide cyclo(l-Tyr-l-Tyr) has been subject to targeted inhibitor design [16-18]. Another, CYP128A1 has been implicated in virulence via the mediation of the metabolite sulfomenaquinone [19]. CYP125A1 and CYP142A1 are together necessary for *M. tuberculosis* cholesterol breakdown [20, 21]. In many other *Mycobacterium* species, the number of *cyp* genes is higher [22, 23]. The human pathogen *Mycobacterium ulcerans*, which causes the Buruli ulcer common in tropical areas of West Africa and Australia, has 24 while *Mycobacterium marinum*, a marine pathogen capable of opportunistic infection of humans, has 47 *cyp* genes [24, 25]. The lower number of CYPs in *M. ulcerans* and *M. tuberculosis* is thought to be caused by reductive evolution as they specialised towards pathogenicity [22-24]. *M. marinum*, in contrast to the two human pathogens, retains its ability to survive outside the host and is hypothesised to resemble the most recent common ancestor of *M. tuberculosis* and *M. ulcerans*, with which it shares a high degree of sequence identity (97% and 85%, respectively) [25]. Study of the CYP complement of *M. marinum* offers the understanding of the role and function of the additional CYPs, and hence the altered metabolism of the more pathogenic strains [26]. In addition, the CYPome of *M. marinum* is accompanied by a greater number of ferredoxins, with 12 single cluster containing species being associated with CYP enzymes in the genome [27]. Only three of these are retained in *M. tuberculosis*, while *M. ulcerans* retains seven.

M. marinum M contains two members of the CYP150 family, CYP150A5 and CYP150A6, which are the fifth and sixth members of the same sub-family. An analogue of CYP150A6 is present in *M. ulcerans* Agy99, but neither enzyme is conserved in *M. tuberculosis*. Other members of the CYP150 family have been identified in *Mycobacteria* including in *M. vanbaalenii* PYR-1 (CYP150A7). This enzyme has been characterised as having the potential to oxidise polycyclic aromatic hydrocarbons and is therefore of interest for its potential application in environmental remediation [28]. The CYP150A5 gene is closely associated in the genome with three ferredoxin genes, and we have previously reported the enzyme is able to hydroxylate the fragrance compound β -ionone when coupled with a native electron transfer chain [27]. Here we report the functional and structural characterisation of the two enzymes CYP150A5 and CYP150A6 to gain insight into their function in the bacterial kingdom.

2. Experimental

2.1 General

All organic substrates, derivatisation agents and other general reagents, except where otherwise noted were purchased from Sigma-Aldrich, Alfa-Aesar, VWR International or Tokyo Chemical Industry. Antibiotics, detergents, DTT and IPTG were from Astral Scientific. The media for cell growth and maintenance (LB, 2xYT, SOC, EMM and trace elements) were prepared as reported previously [26]. Antibiotics were added to the working concentrations listed here; ampicillin, 100 $\mu\text{g mL}^{-1}$ and kanamycin, 30 $\mu\text{g mL}^{-1}$.

UV-Visible spectra were recorded on a Varian Cary 5000 at 30 ± 0.5 °C. GC-MS analysis was performed using a Shimadzu GC-17A equipped with a QP5050A MS detector and DB-5 MS fused silica column (30 m x 0.25 mm, 0.25 μm) or a Shimadzu GC-2010 equipped with a QP2010S GC-MS detector, AOC-20i autoinjector, AOC-20s autosampler and DB-5 MS fused silica column (30 m x 0.25 mm, 0.25 μm). On both instruments, the injector was held at 250 °C and the interface at 280 °C. Column flow was set at 1.5 mL min^{-1} and the split ratio was 24. For norisoprenoid substrates, the initial oven temperature was 120 °C (held for 3 min), before increasing to 220 °C at 10 °C min^{-1} , where it was maintained for 7 min.

2.2 CYP150A6 production and purification

CYP150A6 was purified according to the same method as reported previously for CYP150A5, using two ion-exchange steps [27]. Before each use, the stored protein samples were buffer exchanged into 50 mM Tris (pH 7.4) using a PD-10 desalting column (5 mL, GE Healthcare) to remove glycerol. The CYP150A5 concentration was determined using $\epsilon_{419} = 111 \pm 4 \text{ mM}^{-1} \text{ cm}^{-1}$ [27]. The extinction coefficient of CYP150A6 was determined by the CO binding assay initially developed by Omura and Sato [29] and performed as reported previously [26].

2.3 Spin-state shift assays and dissociation constant determination

The CYP was diluted to ~1 μM using 50 mM Tris (pH 7.4) buffer and the UV/Vis spectrum was recorded between 600 and 250 nm, while held at 30 °C. Aliquots (1 to 5 μL) of substrate stock solutions (50 mM or 100 mM, DMSO or EtOH) were added. The spectra were recorded until the shift reached a stable point. The ratio of high spin to low spin CYP (390 nm peak to 420 nm peak) was estimated to $\pm 5\%$ by comparison to the P450_{cam}-camphor bound substrate spectra [26].

To measure the binding affinity, CYP147G1 was diluted to ~2 μM in a volume of 2.5 mL in 50 mM Tris (pH 7.4) buffer and used to baseline the spectrophotometer. Varying aliquots (1 to 3 μL) of substrate stock solutions of increasing concentrations (1 mM, 10 mM or 100 mM) were added via a Hamilton syringe and mixed. The difference spectrum was recorded between 300 nm and 600 nm. Further aliquots of substrate were added until no change in the peak-to-trough ratio at 420 nm and 390 nm (for a Type I spectrum) or 410 and 430 nm (for a Type II spectrum) was observed. The difference in absorbance versus substrate concentration was fitted to the hyperbolic function (Equation 1):

$$\Delta A = \frac{\Delta A_{max} \times [S]}{K_d + [S]}$$

where K_d is the dissociation constant, $[S]$ is the substrate concentration, ΔA the peak-to-trough ratio, and ΔA_{max} the maximum peak-to-trough absorbance. Where a particular substrate exhibited tight binding (K_d equalling less than five times the concentration of the enzyme), the data were instead fitted to the tight-binding quadratic equation:

$$\Delta A = \Delta A_{max} \times \frac{[E] + [S] + K_d - \sqrt{([E] + [S] + K_d)^2 - 4[E][S]}}{2[E]}$$

where K_d is the dissociation constant, $[S]$ is the substrate concentration, ΔA the peak-to-trough ratio, ΔA_{max} the maximum peak-to-trough absorbance and $[E]$ is the enzyme concentration [30].

2.4 Whole-cell oxidation turnovers

Whole-cell turnovers with CYP150A5 were performed as per the method previously reported [26]. Briefly, a pRSF Duet vector containing the CYP and Fdx8, and a pET Duet vector containing the Fdx8 and FdR1 were expressed in *E. coli* BL21 cells and grown in LB with the appropriate antibiotics and trace elements (3 mL L^{-1}) at $37 \text{ }^{\circ}\text{C}$. Alternative pET Duet vectors were used containing the electron transfer partners listed in Table S1. Once the cells reached late log phase, the temperature was reduced to $18 \text{ }^{\circ}\text{C}$. Benzyl alcohol (0.02% v/v), ethanol (2% v/v) were added and protein expression was induced after a further 30 min with IPTG (0.1 mM). After 16 hrs, the cells were resuspended in *E. coli* minimal media (EMM) and the substrates added to a final concentration of 1 mM and shaken for a further 24 hrs. Aliquots of these growths (including cells) were extracted into ethyl acetate, and analysed by GC-MS.

2.5 Phylogenetic analysis

Sequences were obtained either from the National Centre for Biotechnology Information (NCBI) database or from the Dr Nelson P450 homepage for bacterial P450s [31]. Sequence alignments were performed using ClustalW [32]. The evolutionary history was inferred by using the Maximum Likelihood method based on the Jones-Taylor-Thornton (JTT) matrix-based model [33]. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA6 [34].

2.6 Crystallography

CYP150A5 and CYP150A6 underwent a further purification step by size exclusion (Enrich SEC Column, $650 \times 10 \times 300 \text{ mm}$, 1 mL min^{-1}) before concentration to $\sim 30 \text{ mg mL}^{-1}$ in 50

mM Tris (pH 7.4). Commercially available screening conditions (Hampton Research) were used for initial screening in 96 well sitting drop trays, using 1 μ L of both protein and reservoir solution. Crystal conditions were then refined using the hanging drop vapour diffusion method, again using both 1 μ L of protein and reservoir solution with a 500 μ L reservoir. Diffraction quality crystals of CYP150A6 were obtained after 2 weeks at 16 $^{\circ}$ C from the condition containing 0.2 M ammonium phosphate, 20% w/v polyethylene glycol 3,350, pH 4.7. They were harvested using a Micromount (MiTeGen) and cryo-protected by immersion in Parabar 10312 (Paratone-N, Hampton Research) before flash cooling in liquid N₂. Data were collected by X-ray diffraction at the Australian Synchrotron MX1 beamline (360 exposures using 1 $^{\circ}$ oscillations at a wavelength of 0.9357 \AA). The data were processed into the space group P3121 using iMosflm [35], followed by truncation and addition of R_{free} flags using Aimless [36], as part of the CCP4 package [37]. Molecular replacement phasing was carried out using the MrBump pipeline [38], also part of CCP4, comprising PhaserMR [39] and one round of Buccaneer [40] model building and refinement. The search model used was 3EJD, prepared for molecular replacement by MOLREP [41]. The model was rebuilt using Coot [42] based on initial electron density maps and refined using phenix.refine [43] over several iterations. The structure was deposited in the PDB (Accession code: 6DCB) and the data collection and refinement statistics are presented in Table 1.

Table 1: Crystal data collection and refinement statistics for CYP150A6 from *M. marinum* (PDB: 6DCB)

Data collection statistics^a	
Wavelength	0.95370
Unit cell	a = 80.57 b = 80.57 c = 134.76 $\alpha = 90 \beta = 90 \gamma = 120$
Space group	P 3 1 2 1
Mol. in asym. unit	1
Resolution	48.47 – 1.55 (1.57 – 1.55) ^b
Unique reflections	74838 (3655)
Completeness	100.00 (100.00)
Redundancy	22.0 (20.7)
(I)/[σ (I)]	11.7
R _{merge} (all I+ and I-)	0.262 (4.716)
R _{pim} (all I+ and I-)	0.057 (1.056)
CC(1/2)	0.998 (0.362)
R _{work}	0.2234 (0.3423)
R _{free}	0.2474 (0.3489)
% solvent	53.46
Residues modelled	406
RMS deviation from restraint values	
Bond lengths	0.004
Bond angles	0.80
Ramachandran analysis	
Most favoured	97.76
Additionally allowed	2.24

^aData collected from one crystal

^bValues in parenthesis are for highest resolution shell.

3. Results/Discussion

3.1 Phylogenetic analysis

The CYP150 family has only one named subfamily at present (according to the Cytochrome P450 Homepage maintained by Dr Nelson of the University of Tennessee [31]). There are 15 members of the subfamily that have been assigned CYP names so far, all of which are found in either *Mycobacterium* or *Frankia* species. CYP150A5 and CYP150A6 are both found in *M. marinum* M, with a CYP150A6 species also found in *M. ulcerans* Agy99 (98% sequence identity). Other *Mycobacterium* species such as *M. vanbaalenii* PYR-1, *M. smegmatis* MC2 155, *M. avium* sp. *paratuberculosis* and *M. kansasii* ATCC 12478 contain more than one CYP150 family member. As such, the CYP150 family appears prolific in *Mycobacterium* species, including human pathogens *M. ulcerans* and *M. colombiense* (Table S2). A BLAST search revealed there are a large number (~1000 entries with >55% identity) of similar proteins to both CYP150A5 and CYP150A6, the vast majority of which are from *Mycobacterium* (>90% of entries found). Enzymes of this family are also found in species of *Frankia*, *Streptomyces*, *Nocardia*, and *Rhodococcus* (see Table S2 for more details). This distribution of analogous enzymes in other Actinobacteria is similar to that observed in other *M. marinum* CYP enzymes CYP268A2 and CYP147G1 [26, 27]. The analysis of Parvez *et al* placed the CYP150 family in a clan with the other *Mycobacterium* P450 families of CYP278, CYP1016, CYP269, CYP121, CYP1120, CYP1126, and CYP144 [23].

A sequence alignment of CYP150 family members (selected enzymes shown in Fig. S1) showed the proximal cysteine is conserved as Cys363 (the residues numbering is given for CYP150A6 which coincidentally is the same for CYP150A5) as well as the nearby phenylalanine (Phe356). The EXXR motif is conserved in all CYP150 enzymes (Fig. S1). All contain an acid-alcohol pair; a glutamate (Glu255) and a threonine (Thr256). Phylogenetic

analysis revealed CYP150A6 clusters closer to the previously studied *M. vanbaalenii* PYR-1 CYP150A7 (the polycyclic aromatic hydrolase [28]) and A8, while CYP150A9 and A10 are closer to CYP150A5 (Fig. 1). There appears to be two distinct CYP150 family groups; one containing CYP150A5 and the other with CYP150A6. However, all the CYP150 family members cluster together compared to the closest structurally characterised enzymes CYP144A1 from *M. tuberculosis* (32% identity to CYP150A5) and P450_{Biol} from *Bacillus subtilis* (30% identity to CYP150A6).

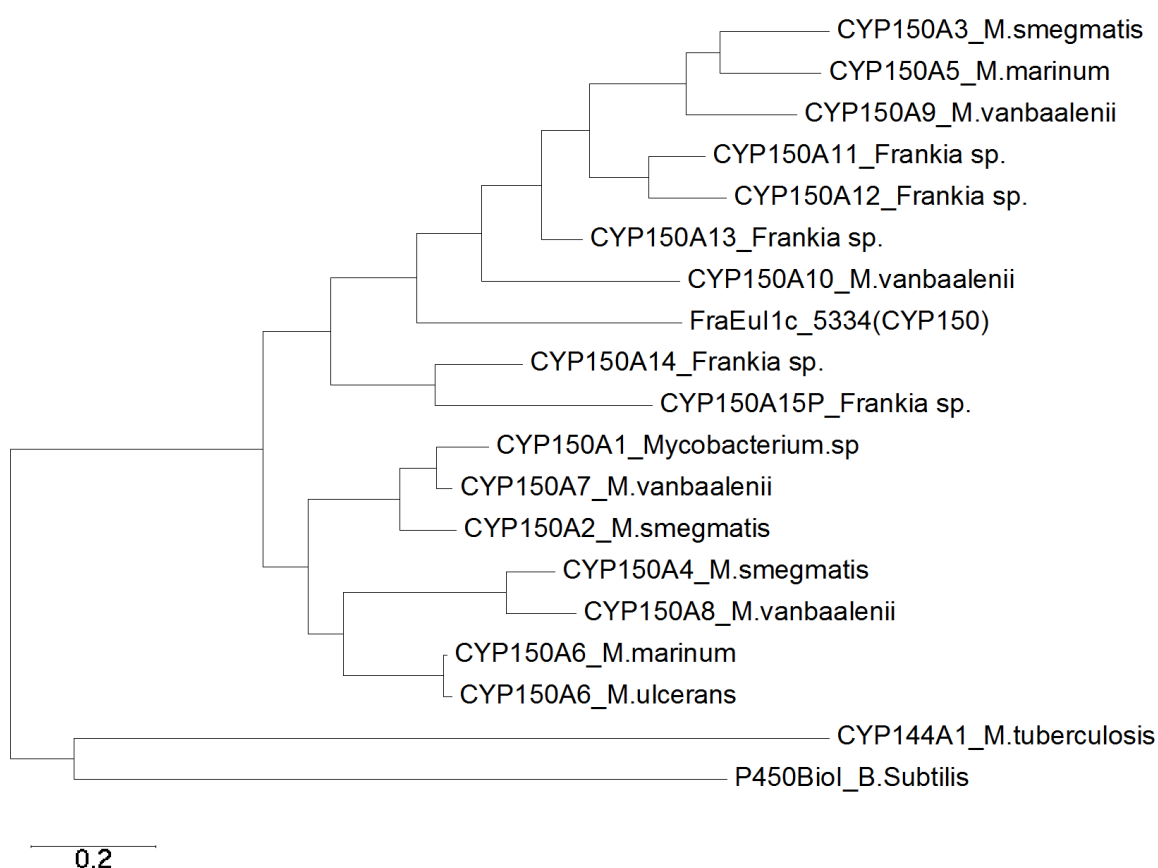


Figure 1: Phylogenetic tree (phenogram) of *M. marinum* enzymes CYP150A5 and CYP150A6 alongside other members of the CYP150 family from various *Mycobacterium* and other species. CYP150A7 from *M. vanbaalenii* has been reported to bind polyaromatic hydrocarbons. The CYP150 family member from *Frankia* sp. Eu1c (encoded by the gene *FraEul1c_5334*) is included. The scale shows number of substitutions per site.

The genomic context of the two CYP enzymes in *M. marinum* M is dissimilar (Fig. S2). CYP150A5 (*Mmar_4737*) lies near CYP190A3 (*Mmar_4733*) and there are three [3Fe-

4S] ferredoxins nearby: *Mmar_4730*, *Mmar_4734* and *Mmar_4736*. *M. liflandii* 128FXT, *M. avium* sp. *paratuberculosis*, *M. smegmatis* MC2 155 and other *Mycobacterium* species show similar gene neighbours in the region of their CYP150A subfamily analogues. These also include a TetR transcriptional regulator, an acyl-CoA dehydrogenase, an acyl-CoA acetyltransferase and an amidohydrolase. CYP150A9 from *M. vanbaalenii* PYR-1 has the most similar genomic region to that of CYP150A5 of the four CYP150 members from that species. In contrast, while CYP150A6 is present in *M. ulcerans* Agy99 (the sequence identity is 98%), comparison of the surrounding genomic region shows the enzyme is not part of a conserved operon. In *M. ulcerans* Agy99 CYP188A3 is nearby (three genes upstream) accompanied by a ferredoxin, while in *M. marinum* M the equivalent pair of genes are 23 genes downstream from CYP150A6. *Frankia* species that contain CYP150 family members show very little regional genomic similarity to *M. marinum* M (Fig. S2).

3.2 Substrate characterisation of CYP150 enzymes

CYP150A6 was expressed in *E. coli*, purified and the ferrous CO bound spectrum was recorded (mass confirmed by SDS-PAGE, Fig. S3). The absorbance of the enzyme's Soret peak shifts (>95%) to 450 nm upon reduction and CO exposure, indicating the reconstituted protein is a viable P450 enzyme (Fig. 2). The CYP150A6 enzyme demonstrated a comparable shift to CYP150A5 and the extinction coefficient was determined to be $\epsilon_{418}=110 \pm 6 \text{ mM}^{-1} \text{ cm}^{-1}$.

In order to determine the substrate range, other members of the CYP150 family were considered. CYP150A7 member from *M. vanbaalenii* oxidises polyaromatic hydrocarbons, such as pyrene, dibenzothiophene, and 7-methylbenz- α -anthracene [28]. However, as Brezna *et al* reported, there is no correlation between presence of a member of the CYP150 family and the polycyclic aromatic degradation ability of a particular strain, suggesting this may be a promiscuous activity of the enzyme [28].

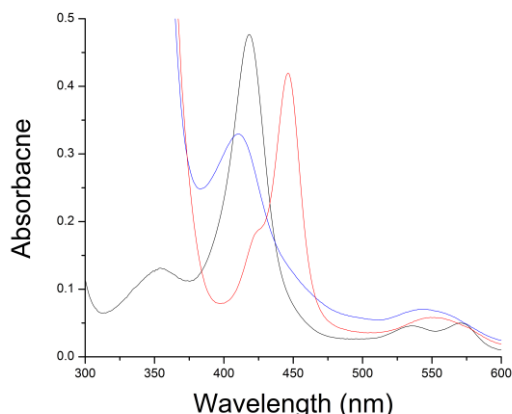


Figure 2: CYP150A6 (black, A418), the reduced ferrous (blue, A417) and the ferrous form bound with CO (red, A446) showing the characteristic ~450 nm absorbance. The shoulder at ~420 nm comprises < 5% of the total area. The extinction coefficient for the enzyme was determined to be $\epsilon_{418} = 110 \pm 6 \text{ mM}^{-1} \text{ cm}^{-1}$.

The first substrates tested with both enzymes were aromatic and polycyclic substrates. These included bicyclic aromatics such as naphthalene, naphthol, phenylphenol, tetralin, and quinoline. None disturbed the spin state of the CYP150A5 enzyme by more than 15% (see Table S3). Ionone derivatives such as α - and β -ionol, camphor and cineole as well as larger molecules such as sclareol and sclareolide were then attempted as substrates (Fig. 3). A large range of norisoprenoid substrates were capable of shifting the spin state shift of CYP150A5 to the high spin state (Table 2). These substrates included α -ionol (70% HS), methyl- α -ionone (65%), and α -ionone (45%). Interestingly, β -damascone, which differs in the location of the ketone group by only two carbons to the ionone substrates, shifted only 15% of the haem to the high spin form (compared to 60% high spin with β -ionone). (S)-(-)-Camphor (60%) and other monoterpenoid substrates such as bornyl acetate and isobornyl acetate (75% and 70% respectively) fit into the active site of the enzyme in this initial test. The diterpene sclareol induced the highest spin state shift, at 90%, compared to 50% by sclareolide, which has an additional ring in the structure (Fig. 3). The bicyclic sesquiterpene guaiazulene shifted

60% of the haem to the high spin form. Both guaiazulene and 2-naphthol (70% high spin) were better substrates than 2-methylnaphthalene (10%) suggesting the substituents around the aromatic ring are important to substrate recognition in CYP150A5.

Table 2: Substrate binding data for both CYP150A5 and CYP150A6. See Figure 4 for dissociation constant analysis. The Supplementary Information contains binding data of additional substrates (Table S3).

Substrates	CYP150A5 Spin		CYP150A6
	state shift (%)	K_D (μM)	Spin state shift (%)
Sclareol	90	0.8 ± 0.08	10
Bornyl acetate	75	45 ± 6	10
Isobornyl acetate	70	34 ± 2	20
α -Ionol	70	17 ± 2	10
2-Naphthol	70	1900 ± 280	<5
Methyl- α -ionone	65	23 ± 4	10
Guaiazulene	60	1.9 ± 0.4	<5
β -Ionone	60	41 ± 2 [27]	15
1,8-Cineole	60	331 ± 45	<5
Fenchyl acetate	60	88 ± 13	10
β -Ionol	55	64 ± 6	20
Sclareolide	50	-	25

(-) indicates dissociation constant was not determined due to limitations in substrate solubility preventing the endpoint being reached.

Dissociation constant analyses were performed for compounds that, from the spin state shift, indicated they were complementary to the active site of this enzyme (defined here as where the shift to the high spin form was greater than 60%, Fig. 4). CYP150A5 with sclareol had the highest binding affinity, at $0.8 \pm 0.08 \mu\text{M}$, followed by guaiazulene with $1.9 \pm 0.4 \mu\text{M}$, which was then almost an order of magnitude tighter than the next best, α -ionol ($17 \pm 2 \mu\text{M}$). The previously reported result with β -ionone ($41 \pm 2 \mu\text{M}$ [27]) was similar to those recorded for bornyl acetate ($45 \pm 6 \mu\text{M}$) and slightly weaker than with isobornyl acetate ($34 \pm 2 \mu\text{M}$). β -Ionol was less tightly bound ($64 \pm 6 \mu\text{M}$). Despite the high spin state shifts recorded for naphthol and cineole with the enzyme, the dissociation constants indicated they

bound weakly ($1900 \pm 280 \mu\text{M}$ and $331 \pm 45 \mu\text{M}$ for 2-naphthol and 1,8-cineole, respectively). The results of this binding analysis provide evidence for a substrate range for CYP150A5 that includes both polycyclic hydrocarbons and terpenoid substrates.

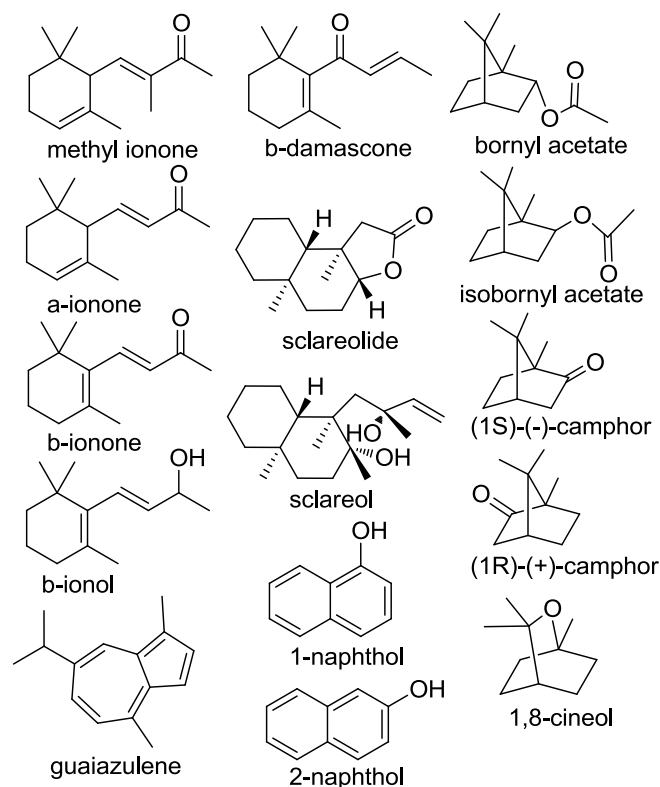


Figure 3: Structures of selected substrates tested with CYP150A5 and CYP150A6.

CYP150A6, in contrast, did not demonstrate any significant shifts in the spin state (all $\leq 25\%$) with any substrate tested. Sclareolide (25%), 1,4-cineole (20%), isobornyl acetate (20%) and β -ionol (20%) were the substrates that induced the greatest spin state shifts. A wide range of substrates were tested, including fatty acids, benzoic acids, terpenoids, aromatics and steroids, however, no significant spin state shift was recorded (all $\leq 20\%$, Table S3). The CYP150A6 enzyme requires further characterisation as this screening method did not reveal many substrates which were able to modify the spin state. This suggests a more specific physiological substrate for this enzyme.

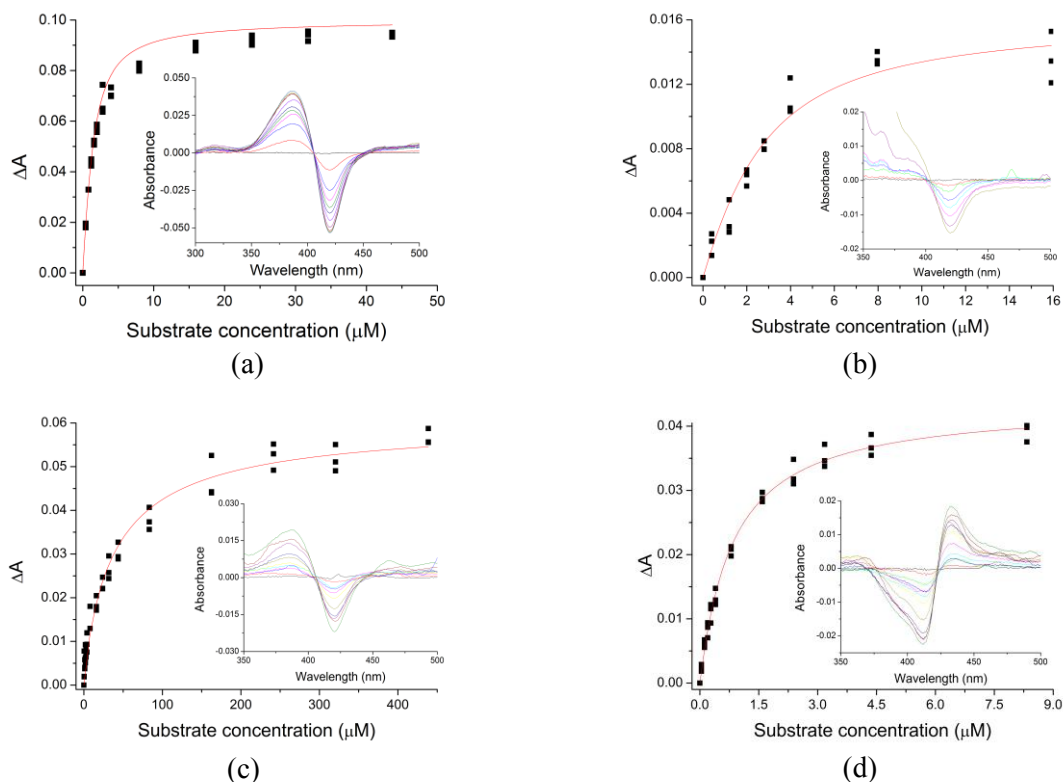


Figure 4: Dissociation constant analysis of CYP150A5 with (a) sclareol, (b) guaiazulene, (c) β -ionol and (d) ketoconazole. The inset represents a typical substrate titration. The peak to trough difference in absorbance was measured for (a) and (c), from 420 to 390 nm, for (b) trough to baseline (420 to 600 nm) due to interfering substrate absorbance and for (d) 432 to 411 nm. For additional dissociation constant analyses see Figure S5 and S6.

Both enzymes were tested with a range of azole compounds, known competitive inhibitors of CYPs (Fig. 5). Econazole, ketoconazole and miconazole among others, generated Type II shifts in both enzymes indicative of N binding directly (or indirectly via a bridging H_2O ligand) to the haem Fe (Table 3). Both 1- and 4-phenylimidazole gave inhibitory shifts in both CYPs, while 2-phenylimidazole did not, similar to the results obtained with CYP268A2 and other enzymes [26, 44, 45]. Fluconazole bound to CYP150A5, shifting the absorbance in a Type II manner, while with CYP150A6 a small (HS ~5%) Type I shift was observed suggesting it bound the enzyme in a substrate-like manner. In general, the phenylimidazoles bound less tightly, with dissociation constants in the tens of μM for both enzymes in contrast to the nM affinity of the larger azoles.

Table 3: Binding data for possible inhibitors of both CYP150A5 and CYP150A6.

Possible inhibitors	CYP150A5 Spin		CYP150A6 Spin	
	state shift	K_D (μM)	state shift	K_D (μM)
1-Phenylimidazole	type II, 423 nm	19.2 ± 5.1	type II, 423 nm	29.9 ± 13.2
2-Phenylimidazole	~ 0%	-	~ 0%	-
Clotrimazole	type II, 424 nm	0.016 ± 0.01	type II, 425 nm	0.05 ± 0.02
Econazole	type II, 424 nm	0.01 ± 0.01	type II, 424 nm	1.1 ± 0.06
Fluconazole	type II, 419 nm	-	Type 1, ~5%	-
Ketoconazole	type II, 423 nm	0.80 ± 0.14	type II, 422 nm	6.6 ± 1
Miconazole	type II, 424 nm	0.045 ± 0.01	type II, 423 nm	1.0 ± 0.09

(-) indicates dissociation constant was not determined.

The affinity of the azoles were higher for CYP150A5 than CYP150A6 (for example, with econazole $K_D = 0.01 \pm 0.01 \mu\text{M}$ compared to $1.1 \pm 0.06 \mu\text{M}$ for the respective enzymes). The affinity of clotrimazole for CYP150A6 was high in comparison to the other tested azoles with that enzyme ($K_D = 0.05 \pm 0.02 \mu\text{M}$, two orders of magnitude tighter than the next best) although the CYP150A5 affinity ($0.016 \pm 0.01 \mu\text{M}$) was still tighter.

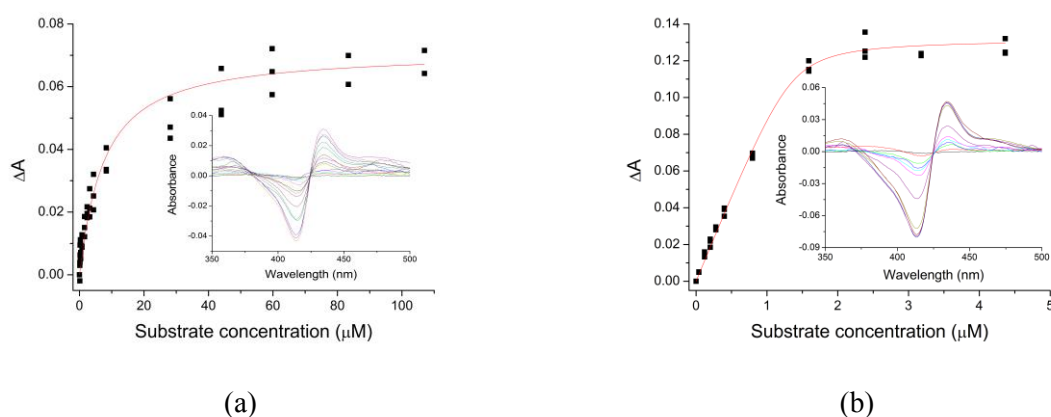


Figure 5: Dissociation constants of CYP150A6 with (a) ketoconazole and (b) clotrimazole. The inset represents a typical substrate titration. The peak to trough difference in absorbance was measured from was (a) 435 to 414 nm and (b) 434 to 413 nm. For additional dissociations constants see Figure S8.

3.3 Product formation

Reconstitution of the *in vivo* activity of CYP150A5 was attempted with the native ferredoxin Fdx8 (*Mmar_4736*) as it is the adjacent gene to the CYP. No reductase gene is located nearby so the ferredoxin reductase FdR1 (*Mmar_2931*), which has been demonstrated to support CYP147G1 activity with Fdx3, was used [27]. Additional electron transfer systems including Tdx/ArR [7, 46], the fused phthalate dioxygenase reductase (PDR) domain from *Pseudomonas putida* pp_1957 [47], and other native ferredoxins from *M. marinum* were also tested (see Table S1). This included the ferredoxin seven genes away from CYP150A5, Fdx6 (*Mmar_4730*). The product formation of each electron transfer system was tested with the substrate β -ionone. The native Fdx8 in combination with FdR1 was the best system tested. However product formation levels were still low. Given the [3Fe-4S] ferredoxin is located nearby in the genome, the most likely barrier to efficient electron transfer is the ferredoxin reductase.

A range of substrates were tested with the CYP150A5 enzyme, including β -ionol and other terpenoids. A single hydroxylation product of β -ionol with CYP150A5 was visible by GC-MS analysis (Fig. 6). The mass spectrum and retention time corresponded to the 4-hydroxy- β -ionol product previously identified from CYP101B1 (Fig. S9)[48]. This is consistent with the 4-hydroxy product of β -ionone with the same enzyme. Metabolites could not be detected with other substrates using the *in vivo* systems, possibly due to solubility and cell uptake difficulties, and further investigation is required to establish the products generated.

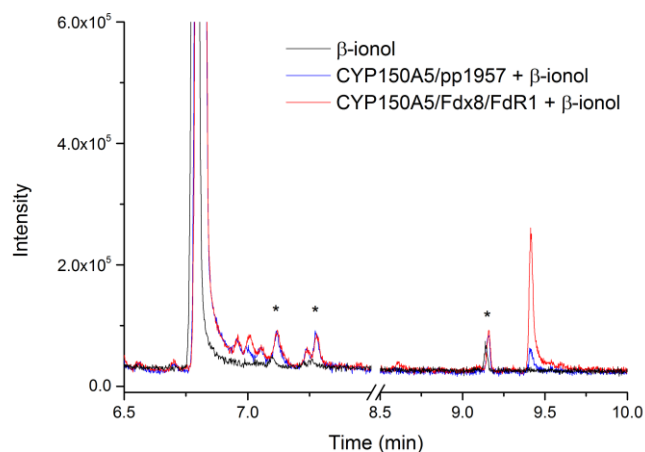
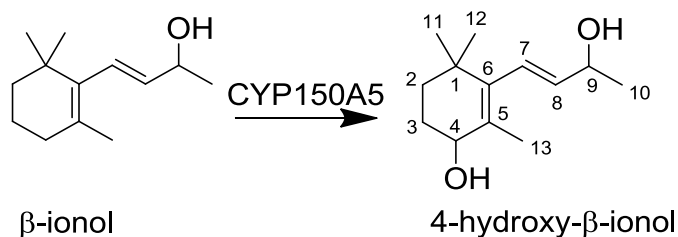


Figure 6: GC chromatogram of *in vivo* turnover of CYP150A5 with Fdx8 and FdR1 (red) and the alternative electron transfer system of *pp_1957* from *Pseudomonas putida* (blue) [47] both with the substrate β -ionol. β -Ionol (RT 6.9 min) was turned over to single hydroxylation product, 4-hydroxy- β -ionol (RT 9.4 min). The physiological electron transfer partners Fdx8 and FdR1 supported greater CYP activity than the alternative systems tested, including *pp_1957* (shown above, see Table S1 for others). * indicates a substrate impurity.



Scheme 1: CYP150A5 conversion of β -ionol to 4-hydroxy- β -ionol.

3.4 Crystal structure

In order to better understand the substrate range of both enzymes we attempted crystallisation of both CYP150A5 and CYP150A6 in a range of conditions. No crystals were obtained for CYP150A5, but diffraction quality crystals were obtained for substrate-free CYP150A6. Data for CYP150A6 were collected from one crystal and the structure solved to 1.5 Å (Table 1, PDB: 6DCB). The structure presented difficulties in phasing that were resolved by using a truncated model of P450_{BioI} prepared by Sculptor (as CYP150A6 is lacking the Acyl-carrier

domain of the P450_{BioI} structure). Several regions required full rebuilding in Coot before refinement (residues 67 to 76 and 350 to 360). The final structure modelled the enzyme from residues 3 to 424 with the exception of the residues from 180 to 195 of the F-G loop for which the electron density was unresolved (Fig. 7). Despite the missing 16 residues, the overall structure of the enzyme clearly displays the conserved fold of a P450 with an open structure when compared to other structures in the PDB (Fig. S10 and Fig. S11). The disorder in the F-G loop is relatively common in previously characterised P450 enzymes in the open form [49, 50]. Both the F and G helices in CYP150A6 appear to be shortened (Fig. 7(b)), leaving a probable 18 residue long F-G loop region (in contrast *M. tuberculosis* CYP144A1 has only 3 residues between the F and G helices, Fig S11). The proline residue at the end of the G helix (P199) is conserved in all *Mycobacterium* CYP150 members, while the glycine at the end of the F helix (G180) is conserved in *M. ulcerans* CYP150A6 and *M. vanbaalenii* CYP150A7, but not in CYP150A5 and some others, inferring that in these proteins the F helix may be longer. The length of the F and G helices is similar to those in mammalian CYP3A4, which has a large degree of substrate promiscuity [51, 52]. However the F-G loop region in CYP3A4 contains two further helices (F' and G') and is 35 residues long. As a result of the unresolved residues in CYP150A6, however, interpreting the precise nature of the active site and access channel is more challenging. As modelled, the active site appears to be very open to solvent (Fig. S12). The F-G loop may be resolved or the F and G helices may reform to a length similar to those in other systems when the enzyme is crystallised in the presence of a substrate.

The structure was compared to that of P450_{BioI} (3EJD) which it was solved against, but the active site of the two structures present few similarities apart from the overall fold when overlaid (Fig. S11). An overlay with the closest structural model from *M. tuberculosis*, CYP144A1 (as identified using MrBump and also the Dali server [53]) showed two regions

of dissimilarity, including small α -helices (between the α -B and B' region, and the other between α -L and β -5) that are not present in CYP144A1 [54]. These are referred to in this text as L' and B'' (Fig. 7(a)). The remaining best matches as identified by the Dali server were *Bacillus megaterium* CYP109A2 (5OFQ) and *M. tuberculosis* CYP142A1 (2XKR). The β -sheet region in the B-C loop of CYP144A1 was absent in CYP150A6 (Fig. S11). The B-C loop area was more similar to that of the *M. smegmatis* CYP142A2 enzyme (Fig. S11) [55].

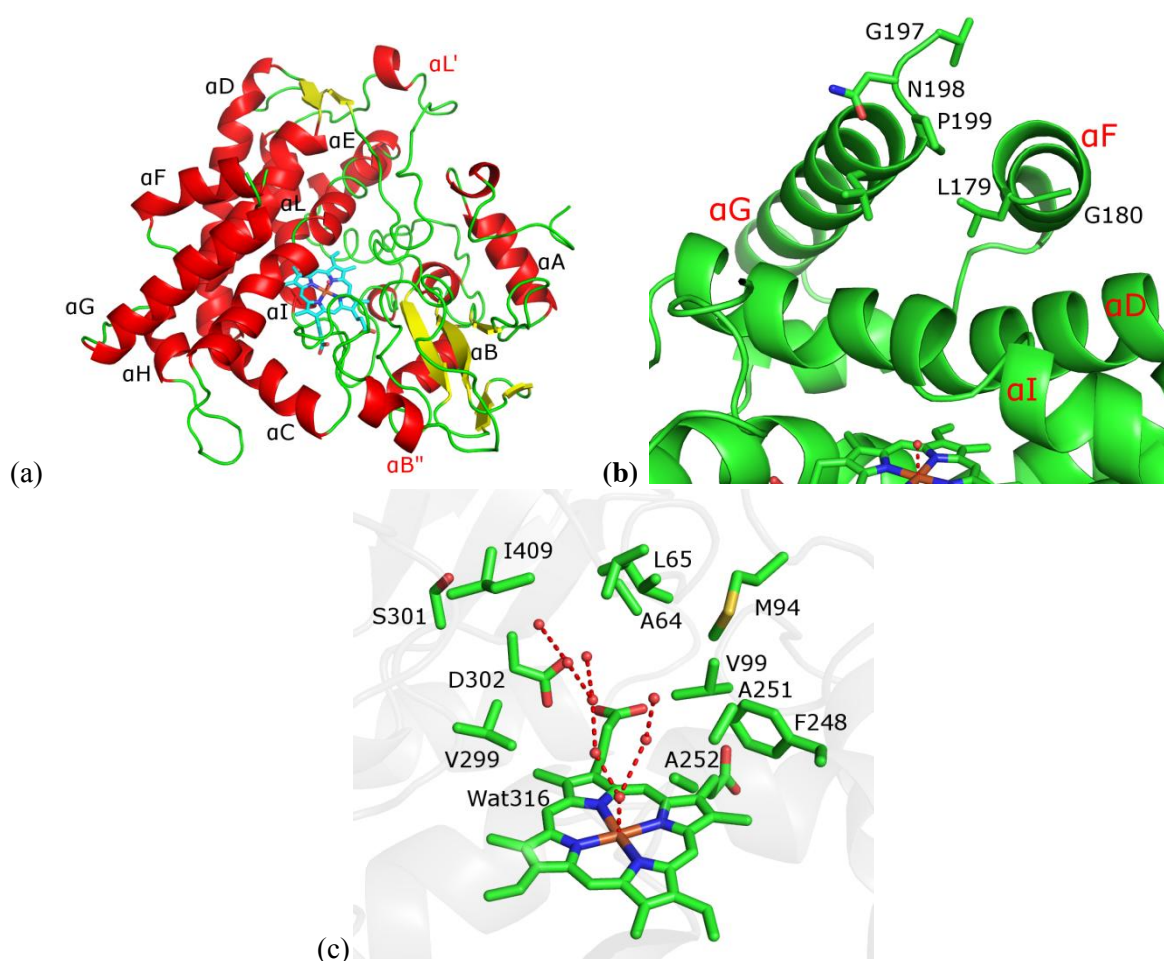


Figure 7: (a) CYP150A6 resolved to 1.5 Å with labelled α -helices according to the nomenclature developed in P450_{cam} (CYP101A1) [56]. Helices are coloured in red, β -sheets in yellow and the remaining backbone in green. The haem is in blue. The two additional helices of CYP150A6 are labelled in red. In the absence of substrate a water molecule binds to the distal site of the haem Fe. (b) The key residues P199 and G180 at the end of the F and G helices showing the short F and G loops (for comparison to other P450 structures see Figures S10 and S11). (c) The residues as well as the network of water molecules (hydrogen bonds in red) in the active site of the substrate-free CYP150A6 (green) enzyme.

The distal side of the haem contains an iron-bound water molecule (Wat316, Fe-O 2.2Å). Wat316 is stabilised by a network of hydrogen bonded waters in the active site (Fig. 7(c)). Two further water molecules (Wat110 and Wat10) are found bound in the kink of the I helix, forming hydrogen bonds between the A251 and T256 residues (where the T256 and E255 are the acid-alcohol pair of the enzyme, Fig. S13). The substrate binding pocket is primarily non-polar although the polar residues S301, D302 and M94 are present. The D302 residue, however, interacts with the carboxylate group of the haem (analogous to D297 in P450_{cam} [57]). These, along with A64 and L65, V99, F248, A251 and A252, V299, and I409 can be identified as probable active site residues. The residues between 248 and 252 belong to the I helix (as seen in Fig. S13) while the B-C loop provides M94 and V99. As the residues of the F-G loop that may have formed the cap of the active site are not modelled, their probable interactions with a ligand cannot be seen. However, there is only limited evidence for direct substrate interaction with the F-G loop [58].

Applying the sequence alignment of CYP150A5 and other CYP150 enzymes to the structure of CYP150A6 allows the identification of the probable active site residues of those species (Table 4). The residues are very dissimilar between CYP150A6 and CYP150A5. Instead of the A251, CYP150A5 has a serine residue (the residue numbering in both enzymes is aligned so will not be given) while the S301 residue of CYP150A6 is changed to a glycine. F248 is changed to asparagine, while both the V99 and M94 residues are replaced with a phenylalanine. A64 and L65 are a serine and valine respectively in CYP150A5. Of the residues identified above, only the acid alcohol pair (E255 and T256), V299, D302 (which interacts with a haem carboxylate) and I409 are conserved in both enzymes. Together the changes represent significant alteration of the polarity and shape of the active site between the enzymes, increasing the likelihood of significantly different physiological roles for the two.

Table 4: The active site residues of CYP150A6 and the aligned residues of other CYP150 family members, including CYP150A6 from *M. ulcerans*, CYP150A5 from *M. marinum* and all four enzymes from *M. vanbaalenii* PYR-1. Bold indicates the residue matches that of CYP150A6, underlined indicates it matches CYP150A5. For emphasis those that are conserved in all are given in red.^a Additional CYP150 family members are listed in Table S4.

Mmar 150A6	Mulc 150A6	Mmar 150A5	Msmeg 150A3	Mvan 150A7	Mvan 150A8	Mvan 150A9	Mvan 150A10
A64	A64	<u>S64</u>	<u>S65</u>	A64	A64	<u>S65</u>	A70
L65	L65	<u>V65</u>	<u>V66</u>	V65	I65	<u>V66</u>	<u>V71</u>
M94	M94	<u>F94</u>	<u>F96</u>	M94	I94	<u>F98</u>	<u>F100</u>
V99	V99	<u>P99</u>	<u>P101</u>	V99	V99	<u>P103</u>	<u>P105</u>
F248	F248	<u>N248</u>	<u>N250</u>	F253	F250	<u>N252</u>	<u>N253</u>
A251	A251	<u>S251</u>	<u>S253</u>	A256	A253	<u>S255</u>	A256
<u>A252</u>	<u>A252</u>	<u>A252</u>	<u>A254</u>	<u>A257</u>	<u>A254</u>	<u>A256</u>	<u>A257</u>
<u>V299</u>	<u>V299</u>	<u>V299</u>	<u>V301</u>	<u>V304</u>	T301	<u>V303</u>	I304
<u>I409</u>	<u>I409</u>	<u>I409</u>	<u>I411</u>	<u>I414</u>	L411	<u>I413</u>	<u>I414</u>

^a E255 and T256 (the acid alcohol pair) are also conserved in all but are not listed here.

This supports the phylogenetic analysis of the whole sequence and the substrate binding data of the CYP150A5 and CYP150A6 enzymes. The CYP150A6 enzyme in *M. ulcerans* Agy99, however, preserves all of these active site residues, so it would be expected to both perform a similar function and the same inhibitors would be effective. Of the *M. vanbaalenii* PYR-1 enzymes, the most similar in the active site region is CYP150A7 (in which all the CYP150A6 active site residues are conserved) while CYP150A9 is the least. However CYP150A9 shares the residues of CYP150A5 at these positions. CYP150A8 and CYP150A10 share some active site residues with both *M. marinum* enzymes, and potentially would have distinct roles again.

4. Conclusions

Members of the CYP150 family are widely found in *Mycobacterium* species, as well as in related Actinomycetes. The only previously studied member of the family, CYP150A7 from *M. vanbaalenii* PYR-1, was found to oxidise certain polycyclic aromatic hydrocarbons which was linked to the activity of this species to degrade this class of compound. Here, the enzyme CYP150A5 is shown to bind and hydroxylate cyclic terpenes, with sclareol having the best binding parameter, rather than polycyclic aromatics. A physiological ferredoxin was used to support CYP activity in combination with a ferredoxin reductase also from *M. marinum*. The structural characterisation of CYP150A6 will facilitate future substrate identification for this enzyme through use of modelling and high-throughput screening methods. As both enzymes are found in a range of *Mycobacteria*, including human pathogens, the inhibitors which were determined for both enzymes could be used to stop CYP-related metabolism in these species which could form the basis of future drug design.

Acknowledgements

This work was supported by the Australian Research Council through a Future Fellowship (FT140100355 to S.G.B.). The authors also acknowledge the award of University of Adelaide Faculty of Sciences Divisional Scholarship (PhD to S.A.C.). X-ray diffraction data collection was undertaken on the MX1 beamline at the Australian Synchrotron, part of the Australian Nuclear Science and Technology Organisation. We acknowledge financial support from the Australian Synchrotron in the form of MXCAP9674. We are grateful to Professor Tim Stinear from the University of Melbourne and Professor Lalita Ramakrishnan of the University of Cambridge for providing the genomic DNA of *M. marinum*.

References

- [1] P.R. Ortiz de Montellano, Substrate Oxidation by Cytochrome P450 Enzymes in: P.R. Ortiz de Montellano (Ed.) *Cytochrome P450: Structure, Mechanism, and Biochemistry*, Springer International Publishing, 2015, pp. 111-176.
- [2] F.P. Guengerich, Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity, *Chem. Res. Toxicol.*, 14 (2001) 611-650.
- [3] F. Hannemann, A. Bichet, K.M. Ewen, R. Bernhardt, Cytochrome P450 systems--biological variations of electron transport chains, *Biochim. Biophys. Acta*, 1770 (2007) 330-344.
- [4] S.G. Bell, N. Hoskins, F. Xu, D. Caprotti, Z. Rao, L.L. Wong, Cytochrome P450 enzymes from the metabolically diverse bacterium *Rhodospseudomonas palustris*, *Biochem. Biophys. Res. Commun.*, 342 (2006) 191-196.
- [5] D.C. Lamb, T. Skaug, H.L. Song, C.J. Jackson, L.M. Podust, M.R. Waterman, D.B. Kell, D.E. Kelly, S.L. Kelly, The cytochrome P450 complement (CYPome) of *Streptomyces coelicolor* A3(2), *J. Biol. Chem.*, 277 (2002) 24000-24005.
- [6] D. Batabyal, T.L. Poulos, Crystal structures and functional characterization of wild-type CYP101D1 and its active site mutants, *Biochemistry*, 52 (2013) 8898-8906.
- [7] S.G. Bell, A. Dale, N.H. Rees, L.-L. Wong, A cytochrome P450 class I electron transfer system from *Novosphingobium aromaticivorans*, *Appl. Microbiol. Biotechnol.*, 86 (2010) 163-175.
- [8] F.P. Guengerich, Human Cytochrome P450 Enzymes in: P.R. Ortiz de Montellano (Ed.) *Cytochrome P450: Structure, Mechanism, and Biochemistry*, Springer International Publishing, 2015, pp. 523-786.
- [9] S.L. Kelly, D.E. Kelly, Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us?, *Philos. Trans. Royal Soc. B*, 368 (2013) 20120476.
- [10] D.R. Nelson, L. Koymans, T. Kamataki, J.J. Stegeman, R. Feyereisen, D.J. Waxman, M.R. Waterman, O. Gotoh, M.J. Coon, R.W. Estabrook, I.C. Gunsalus, D.W. Nebert, P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature, *Pharmacogenetics*, 6 (1996) 1-42.
- [11] V.B. Urlacher, S. Eiben, Cytochrome P450 monooxygenases: perspectives for synthetic application, *Trends in Biotechnology*, 24 (2006) 324-330.
- [12] World Health Organisation: Global Tuberculosis Report, in, http://www.who.int/tb/publications/global_report/en/, 2017.
- [13] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry, 3rd, F. Tekaiia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M.A. Quail, M.A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J.E. Sulston, K. Taylor, S. Whitehead, B.G. Barrell, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, *Nature*, 393 (1998) 537-544.

- [14] K.J. McLean, A.J. Dunford, R. Neeli, M.D. Driscoll, A.W. Munro, Structure, function and drug targeting in *Mycobacterium tuberculosis* cytochrome P450 systems, *Arch. Biochem. Biophys.*, 464 (2007) 228-240.
- [15] H. Ouellet, J.B. Johnston, P.R. Ortiz de Montellano, The *Mycobacterium tuberculosis* cytochrome P450 system, *Arch. Biochem. Biophys.*, 493 (2010) 82-95.
- [16] K.J. McLean, P. Carroll, D.G. Lewis, A.J. Dunford, H.E. Seward, R. Neeli, M.R. Cheesman, L. Marsollier, P. Douglas, W.E. Smith, I. Rosenkrands, S.T. Cole, D. Leys, T. Parish, A.W. Munro, Characterization of active site structure in CYP121. A cytochrome P450 essential for viability of *Mycobacterium tuberculosis* H37Rv, *J. Biol. Chem.*, 283 (2008) 33406-33416.
- [17] P. Belin, M.H. Le Du, A. Fielding, O. Lequin, M. Jacquet, J.B. Charbonnier, A. Lecoq, R. Thai, M. Courcon, C. Masson, C. Dugave, R. Genet, J.L. Pernodet, M. Gondry, Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*, *Proc. Natl. Acad. Sci. U. S. A.*, 106 (2009) 7426-7431.
- [18] S.A. Hudson, K.J. McLean, S. Surade, Y.Q. Yang, D. Leys, A. Ciulli, A.W. Munro, C. Abell, Application of Fragment Screening and Merging to the Discovery of Inhibitors of the *Mycobacterium tuberculosis* Cytochrome P450 CYP121, *Angew. Chem. Int. Ed.*, 51 (2012) 9311-9316.
- [19] K.M. Sogi, C.M. Holsclaw, G.K. Fragiadakis, D.K. Nomura, J.A. Leary, C.R. Bertozzi, Biosynthesis and regulation of sulfomenaquinone, a metabolite associated with virulence in *Mycobacterium tuberculosis*, *ACS Infect. Dis.*, 2 (2016) 800-806.
- [20] J.K. Capyk, R. Kalscheuer, G.R. Stewart, J. Liu, H. Kwon, R. Zhao, S. Okamoto, W.R. Jacobs, Jr., L.D. Eltis, W.W. Mohn, Mycobacterial cytochrome p450 125 (cyp125) catalyzes the terminal hydroxylation of c27 steroids, *J. Biol. Chem.*, 284 (2009) 35534-35542.
- [21] M.D. Driscoll, K.J. McLean, C. Levy, N. Mast, I.A. Pikuleva, P. Lafite, S.E. Rigby, D. Leys, A.W. Munro, Structural and biochemical characterization of *Mycobacterium tuberculosis* CYP142: evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen, *J. Biol. Chem.*, 285 (2010) 38270-38282.
- [22] D.A. Stahl, J.W. Urbance, The division between fast- and slow-growing species corresponds to natural relationships among the mycobacteria, *J. Bacteriol.*, 172 (1990) 116-124.
- [23] M. Parvez, L.B. Qhanya, N.T. Mthakathi, I.K.R. Kgosiemang, H.D. Bamal, N.S. Pagadala, T. Xie, H. Yang, H. Chen, C.W. Theron, R. Monyaki, S.C. Raselemane, V. Salewe, B.L. Mongale, R.G. Matowane, S.M.H. Abdalla, W.I. Booi, M. van Wyk, D. Olivier, C.E. Boucher, D.R. Nelson, J.A. Tuszynski, J.M. Blackburn, J.-H. Yu, S.S. Mashele, W. Chen, K. Syed, Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: Special focus on mycobacterial P450s, *Scientific Reports*, 6 (2016) 33099.
- [24] C. Demangel, T.P. Stinear, S.T. Cole, Buruli ulcer: reductive evolution enhances pathogenicity of *Mycobacterium ulcerans*, *Nature reviews. Microbiology*, 7 (2009) 50-60.
- [25] T.P. Stinear, T. Seemann, P.F. Harrison, G.A. Jenkin, J.K. Davies, P.D. Johnson, Z. Abdallah, C. Arrowsmith, T. Chillingworth, C. Churcher, K. Clarke, A. Cronin, P. Davis, I. Goodhead, N. Holroyd, K. Jagels, A. Lord, S. Moule, K. Mungall, H. Norbertczak, M.A.

- Quail, E. Rabbinowitsch, D. Walker, B. White, S. Whitehead, P.L. Small, R. Brosch, L. Ramakrishnan, M.A. Fischbach, J. Parkhill, S.T. Cole, Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*, *Genome Research*, 18 (2008) 729-741.
- [26] S.A. Child, E.F. Naumann, J.B. Bruning, S.G. Bell, Structural and functional characterisation of the cytochrome P450 enzyme CYP268A2 from *Mycobacterium marinum*, *Biochem. J.*, 475 (2018) 705-722.
- [27] S.A. Child, J.M. Bradley, T.L. Pukala, D.A. Svistunenko, N.E. Le Brun, S.G. Bell, Electron transfer ferredoxins with unusual cluster binding motifs support secondary metabolism in *Mycobacteria* and are prevalent in many other bacteria, (*Unpublished*).
- [28] B. Brezna, O. Kweon, R.L. Stingley, J.P. Freeman, A.A. Khan, B. Polek, R.C. Jones, C.E. Cerniglia, Molecular characterization of cytochrome P450 genes in the polycyclic aromatic hydrocarbon degrading *Mycobacterium vanbaalenii* PYR-1, *Appl. Microbiol. Biotechnol.*, 71 (2006) 522-532.
- [29] T. Omura, R. Sato, The Carbon Monoxide-Binding Pigment of Liver Microsomes. II. Solubilization, Purification, and Properties, *J. Biol. Chem.*, 239 (1964) 2379-2385.
- [30] J.M. Dogne, X. de Leval, P. Benoit, S. Rolin, B. Pirotte, B. Masereel, Therapeutic potential of thromboxane inhibitors in asthma, *Expert Opin. Investig. Drugs*, 11 (2002) 275-281.
- [31] D.R. Nelson, The cytochrome p450 homepage, *Human genomics*, 4 (2009) 59-65.
- [32] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol. Syst. Biol.*, 7 (2011) 539.
- [33] D.T. Jones, W.R. Taylor, J.M. Thornton, The rapid generation of mutation data matrices from protein sequences, *Comput. Appl. Biosci.*, 8 (1992) 275-282.
- [34] K. Tamura, G. Stecher, D. Peterson, A. FilipSKI, S. Kumar, MEGA6: Molecular Evolutionary Genetics Analysis version 6.0, *Mol. Biol. Evol.*, 30 (2013) 2725-2729.
- [35] T.G. Battye, L. Kontogiannis, O. Johnson, H.R. Powell, A.G. Leslie, iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM, *Acta Crystallogr. D*, 67 (2011) 271-281.
- [36] P.R. Evans, G.N. Murshudov, How good are my data and what is the resolution?, *Acta Crystallogr. D*, 69 (2013) 1204-1214.
- [37] M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R.M. Keegan, E.B. Krissinel, A.G. Leslie, A. McCoy, S.J. McNicholas, G.N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin, K.S. Wilson, Overview of the CCP4 suite and current developments, *Acta Crystallogr. D*, 67 (2011) 235-242.
- [38] R.M. Keegan, M.D. Winn, Automated search-model discovery and preparation for structure solution by molecular replacement, *Acta Crystallogr. D*, 63 (2007) 447-457.
- [39] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, Phaser crystallographic software, *J. Appl. Crystallogr.*, 40 (2007) 658-674.
- [40] K. Cowtan, The Buccaneer software for automated model building. 1. Tracing protein chains, *Acta Crystallogr. D*, 62 (2006) 1002-1011.

- [41] A. Vagin, A. Teplyakov, MOLREP: an Automated Program for Molecular Replacement, *J. Appl. Crystallogr.*, 30 (1997) 1022-1025.
- [42] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, *Acta Crystallogr. D*, 66 (2010) 486-501.
- [43] P.D. Adams, P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr. D*, 66 (2010) 213-221.
- [44] K.J. McLean, A.J. Warman, H.E. Seward, K.R. Marshall, H.M. Girvan, M.R. Cheesman, M.R. Waterman, A.W. Munro, Biophysical characterization of the sterol demethylase P450 from *Mycobacterium tuberculosis*, its cognate ferredoxin, and their interactions, *Biochemistry*, 45 (2006) 8427-8443.
- [45] D.L. Harris, Y.-T. Chang, G.H. Loew, Molecular dynamics simulations of phenylimidazole inhibitor complexes of cytochrome P450 cam, *Molec. Eng.* 5 (1995) 143-156.
- [46] S.G. Bell, W. Yang, J.A. Yorke, W. Zhou, H. Wang, J. Harmer, R. Copley, A. Zhang, R. Zhou, M. Bartlam, Z. Rao, L.L. Wong, Structure and function of CYP108D1 from *Novosphingobium aromaticivorans* DSM12444: an aromatic hydrocarbon-binding P450 enzyme, *Acta Crystallogr. D*, 68 (2012) 277-291.
- [47] S.G. Bell, L. French, N.H. Rees, S.S. Cheng, G. Preston, L.L. Wong, A phthalate family oxygenase reductase supports terpene alcohol oxidation by CYP238A1 from *Pseudomonas putida* KT2440, *Biotechnol. Appl. Biochem.*, 60 (2013) 9-17.
- [48] E.A. Hall, M.R. Sarkar, J.H.Z. Lee, S.D. Munday, S.G. Bell, Improving the Monooxygenase Activity and the Regio- and Stereoselectivity of Terpenoid Hydroxylation Using Ester Directing Groups, *ACS Catalysis*, 6 (2016) 6306-6317.
- [49] T.L. Poulos, Cytochrome P450 flexibility, *Proc. Natl. Acad. Sci. U. S. A.*, 100 (2003) 13121-13122.
- [50] T.C. Pochapsky, S. Kazanis, M. Dang, Conformational Plasticity and Structure/Function Relationships in Cytochromes P450, *Antioxidants & Redox Signaling*, 13 (2010) 1273-1296.
- [51] J.K. Yano, M.R. Wester, G.A. Schoch, K.J. Griffin, C.D. Stout, E.F. Johnson, The Structure of Human Microsomal Cytochrome P450 3A4 Determined by X-ray Crystallography to 2.05-Å Resolution, *J. Biol. Chem.*, 279 (2004) 38091-38094.
- [52] P.A. Williams, J. Cosme, D.M. Vinković, A. Ward, H.C. Angove, P.J. Day, C. Vonrhein, I.J. Tickle, H. Jhoti, Crystal Structures of Human Cytochrome P450 3A4 Bound to Metyrapone and Progesterone, *Science*, 305 (2004) 683-686.
- [53] L. Holm, L.M. Laakso, Dali server update, *Nucleic Acids Res*, 44 (2016) W351-355.
- [54] J. Cheng, M.E. Kavanagh, M.D. Driscoll, K.J. McLean, D.B. Young, T. Cortes, D. Matak-Vinkovic, C.W. Levy, S.E. Rigby, D. Leys, C. Abell, A.W. Munro, Structural characterization of CYP144A1 - a cytochrome P450 enzyme expressed from alternative transcripts in *Mycobacterium tuberculosis*, *Scientific Reports*, 6 (2016) 26628.

- [55] E. Garcia-Fernandez, D.J. Frank, B. Galan, P.M. Kells, L.M. Podust, J.L. Garcia, P.R. Ortiz de Montellano, A highly conserved mycobacterial cholesterol catabolic pathway, *Environ. Microbiol.*, 15 (2013) 2342-2359.
- [56] T.L. Poulos, B.C. Finzel, I.C. Gunsalus, G.C. Wagner, J. Kraut, The 2.6-Å crystal structure of *Pseudomonas putida* cytochrome P-450, *J. Biol. Chem.*, 260 (1985) 16122-16130.
- [57] T.L. Poulos, B.C. Finzel, A.J. Howard, High-resolution crystal structure of cytochrome P450cam, *J. Mol. Biol.*, 195 (1987) 687-700.
- [58] S.C. Dodani, G. Kiss, J.K.B. Cahn, Y. Su, V.S. Pande, F.H. Arnold, Discovery of a regioselectivity switch in nitrating P450s guided by molecular dynamics simulations and Markov models, *Nat. Chem.*, 8 (2016) 419.

Chapter 5:

Structural and functional characterisation of the cytochrome P450 enzyme CYP268A2 from *Mycobacterium marinum*

Published in Biochemical Journal, 2018.

Citation:

Child, S. A.; Naumann, E. F.; Bruning, J. B.; E. Bell, S. G., Structural and functional characterisation of the cytochrome P450 enzyme CYP268A2 from *Mycobacterium marinum*, *Biochem. J.*, 2018, 475(4), 705-722.

Statement of Authorship

Principal Author

Name	Stella Child		
Contribution to the Paper	Enzyme purification, substrate characterisation, product formation, crystallisation, structural refinement, manuscript preparation and review.		
Overall percentage (%)	80		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	23/5/18

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Elise Naumann		
Contribution	Enzyme purification and substrate characterisation		
Signature		Date	2/05/2018

Name of Co-Author	John Bruning		
Contribution	Supervision and advice on X-ray crystal refinement		
Signature		Date	23/5/18

Name of Co-Author	Stephen Bell		
Contribution	Experimental design, supervision, manuscript preparation.		
Signature		Date	23/5/2018

Research Article

Structural and functional characterisation of the cytochrome P450 enzyme CYP268A2 from *Mycobacterium marinum*

Stella A. Child¹, Elise F. Naumann¹, John B. Bruning² and  Stephen G. Bell¹

¹Department of Chemistry, School of Physical Sciences, University of Adelaide, Adelaide, SA 5005, Australia; ²School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia

Correspondence: Stephen G. Bell (stephen.bell@adelaide.edu.au)

Members of the cytochrome P450 monooxygenase family CYP268 are found across a broad range of *Mycobacterium* species including the pathogens *Mycobacterium avium*, *M. colombiense*, *M. kansasii*, and *M. marinum*. CYP268A2, from *M. marinum*, which is the first member of this family to be studied, was purified and characterised. CYP268A2 was found to bind a variety of substrates with high affinity, including branched and straight chain fatty acids (C10–C12), acetate esters, and aromatic compounds. The enzyme was also found to bind phenylimidazole inhibitors but not larger azoles, such as ketoconazole. The monooxygenase activity of CYP268A2 was efficiently reconstituted using heterologous electron transfer partner proteins. CYP268A2 hydroxylated geranyl acetate and *trans*-pseudoionone at a terminal methyl group to yield (2*E*,6*E*)-8-hydroxy-3,7-dimethylocta-2,6-dien-1-yl acetate and (3*E*,5*E*,9*E*)-11-hydroxy-6,10-dimethylundeca-3,5,9-trien-2-one, respectively. The X-ray crystal structure of CYP268A2 was solved to a resolution of 2.0 Å with *trans*-pseudoionone bound in the active site. The overall structure was similar to that of the related phytanic acid monooxygenase CYP124A1 enzyme from *Mycobacterium tuberculosis*, which shares 41% sequence identity. The active site is predominantly hydrophobic, but includes the Ser99 and Gln209 residues which form hydrogen bonds with the terminal carbonyl group of the pseudoionone. The structure provided an explanation on why CYP268A2 shows a preference for shorter substrates over the longer chain fatty acids which bind to CYP124A1 and the selective nature of the catalysed monooxygenase activity.

Introduction

Cytochrome P450 (CYP) enzymes are a family of haem monooxygenases, capable of catalysing the insertion of a single oxygen atom, derived from molecular oxygen, into an inert carbon–hydrogen bond of a wide range of organic substrates [1]. Cytochrome P450 enzymes are ubiquitous in nature, with genes in humans [2], other animals [3], plants, and many fungal and bacterial species. CYP enzymes perform both anabolic (building up metabolites) and catabolic (breaking them down) processes, making them key enzymes in secondary and xenobiotic metabolism [2] and targets for antibacterial drug design [4]. They have also been shown to perform a variety of reactions, most commonly hydroxylation but also further oxidation, C–C bond formation, desaturation, and epoxidation, using electrons ultimately sourced from NAD(P)H [reduced nicotinamide adenine dinucleotide (phosphate)] [5]. Individual CYPs often show high specificity for their substrate and selectivity in the site of the target molecule where the oxidation takes place [6,7]. The catalytic activity of CYP enzymes is dependent on the delivery of the two electrons to the haem in two separate, highly regulated steps. In bacterial species, this is most often achieved by the combination of two cytosolic electron transfer partners, a flavin adenine dinucleotide (FAD) containing ferredoxin reductase and an iron–sulphur

Received: 13 December 2017
Revised: 10 January 2018
Accepted: 16 January 2018

Accepted Manuscript online:
17 January 2018
Version of Record published:
16 February 2018

ferredoxin, together known as a Class 1 electron transfer system [8]. CYPs are classified based on amino acid sequence similarity, where members of a family share >40% sequence homology and subfamily members >55% [9].

Upon sequencing of the *Mycobacterium tuberculosis* genome [10], the high number of CYP genes found (20) was unprecedented for a bacterial species at the time. (In contrast, *Escherichia coli* has none.) As a result, the CYPome of *M. tuberculosis* has been a target for inhibitory drug design in the years since [11]. Owing to increasing numbers of drug-resistant and multi-drug-resistant strains, and a disease profile that overlaps with that of HIV/AIDS, *M. tuberculosis* continues to be responsible for large-scale loss of life [12,13]. Members of the *Mycobacterium* family are widespread and range from soil bacteria to human pathogens. *Mycobacterium marinum* M, a pathogen of frogs and fish, is a close relative of both *M. tuberculosis* H37Rv (85% nucleotide identity) and also the pathogenic species *Mycobacterium ulcerans* Agy99 (97%) [14]. *M. ulcerans* is responsible for the Buruli ulcer (also referred to as the Daintree or Bairnsdale ulcer), which is a skin disease primarily found in tropical areas, most often in central and western Africa [15]. *M. marinum* has 47 individual CYP encoding genes in its genome, while *M. ulcerans* has 24. *M. marinum* is thought to resemble a common ancestor of the more pathogenic *Mycobacterium* species, with a genome that has not undergone the extensive reduction by gene deletion and pseudogene formation that characterises the genome of *M. tuberculosis* and, to a lesser extent, *M. ulcerans* [14,15]. *M. marinum* is a more versatile pathogen, primarily affecting aquatic life but also capable of surviving in a human host as the causal agent of aquarium granuloma, and, unlike the human pathogens, of persisting outside of its host entirely [14,16]. It has been shown to adapt to a variety of hosts, altering virulence mechanisms to suit, including the differential regulation of polyketide lipids, and sterol utilisation [16]. Thus, the larger genome of *M. marinum* provides both increased redundancy, with a smaller percentage of essential genes than *M. tuberculosis*, but also increased adaptability. This is part of a common trend, where the number of CYP genes in *Mycobacterium* species decreases as the organism environment changes from soil living mycobacteria (average of 50 CYPs) to a human pathogen (average of 19) [17].

Where there are direct counterparts for *M. marinum* CYPs in *M. tuberculosis*, the roles of the majority are still unknown. Several of the enzymes are reported as cholesterol oxidases, including some that are essential for viability *in vitro*. The deletion of the CYP125A1 enzyme together with CYP124A1 leads to a build-up of the intermediate cholest-4-en-3-one and the inhibition of the organism [18]. The cholesterol degradation activity of CYP125A1 has been linked to the density of the mycobacterial cell wall, increasing the mass of phthiocerol dimycocerosates (PDIM) [19]. The Mycobacterial cell envelope is distinguished by several features, most prominently the additional layer of long-chain fatty acids known as mycolic acid covalently bound to the peptidoglycan of the cell wall [20]. PDIM and other multiple methyl-branched long-chain fatty acids further increase the density and thickness of this layer. Lipid metabolism is another common role of bacterial CYPs, and in *M. tuberculosis*, CYP124A1 has been shown to hydroxylate phytanic acid and similar fatty acid compounds [21]. The majority of *M. tuberculosis* CYPs, however, have either resisted recombinant expression efforts, or have not shown activity when screened against libraries of common substrates, leaving uncertainty about the roles of the Mycobacterial CYPome as a whole. Even where the substrate and product is known, it is not well understood how these play into the metabolism of the organism. For example, while *M. tuberculosis* is known to have no sterol synthesis pathway, CYP51B1, a highly conserved sterol- α -demethylase, is present [22]. The current understanding of *M. tuberculosis* virulence points to specialised areas such as mycolic acid synthesis, other lipid metabolism, and cholesterol catabolism as critical [23], in the second two of which there is evidence for CYP involvement.

Analogues of many of these characterised CYP proteins are present in other species of *Mycobacteria*. For example, CYP51B1 is found in a highly conserved operon in *Mycobacterium smegmatis* MC2 155 and *M. tuberculosis*, containing a CYP123 enzyme, a ferredoxin, a TetR regulator, and an ORF of unknown function [24]. The same operon is conserved in *M. marinum* M and *M. ulcerans* Agy99. The conservation of various CYPs outside of the *M. tuberculosis* complex (MTBC) has been taken as evidence that they perform general or housekeeping roles, while conservation only in *Mycobacterium bovis* or mammalian pathogens has led to a suggested role in virulence or human infectivity [25]. The characterisation of CYP enzymes present only in non-MTBC species has been attempted for various reasons. Some were potential catalysts for reactions of bio-synthetic interest, such as CYP153A16, from *M. marinum*, which was found like other members of the CYP153 family to oxidise medium-chain alkanes [7]. CYP151A1 in *M. smegmatis* MC2 155 and CYP151A2 from *Mycobacterium* sp. strain RP1 were identified in the effort to understand the ability of the respective organisms to utilise pyrrolidine and piperidine as sole carbon sources [26,27]. Both can oxidise secondary

amines, catalysing ring opening and allowing further catabolism. In addition, CYP150 family members from *Mycobacterium vanbaalenii* PYR-1 have been hypothesised to oxidise polycyclic aromatic hydrocarbons, making them a catalyst of interest for environmental remediation [28].

As a result of its larger CYPome, many CYPs are found in *M. marinum* M that do not have direct counterparts in either of *M. tuberculosis* or *M. ulcerans*, and which may play a role in the different pathogenicity of the organism. The enzyme CYP268A2 is one instance of this, with only a truncated pseudogene (*cyp268A2P*) remaining in the *M. ulcerans* Agy99 genome, and no relative in *M. tuberculosis*. It is highly conserved across a broad range of other *Mycobacterium* species and other bacteria. We have performed preliminary structural and functional characterisation of this enzyme, with a view to elucidating its role in the metabolism of *M. marinum* and across the *Mycobacteria* (Mycobacteriaceae) as a whole.

Experimental

General

All organic substrates, derivatisation agents, and general laboratory reagents, except where otherwise noted, were purchased from Sigma–Aldrich, Alfa-Aesar, VWR International, or Tokyo Chemical Industry. Antibiotics, detergents, DTT (dithiothreitol), and IPTG (isopropyl β -D-1-thiogalactopyranoside) were from Astral Scientific. Restriction enzymes used for cloning were purchased from New England Biolabs. KOD polymerase, used for the PCR (polymerase chain reaction) steps, and the expression vectors were from Merck-Millipore.

The following were used as media for cell growth and maintenance: LB (lysogeny broth); tryptone (10 g), yeast extract (5 g), and NaCl (10 g) per litre; 2 \times YT contains tryptone (16 g), yeast extract (10 g), and NaCl (5 g) per litre; SOC (Super Optimal broth with Catabolite repression); tryptone (20 g), yeast extract (5 g), NaCl (0.5 g), KCl (0.2 g), MgCl₂ (1 g), and 5 ml of 40% glucose per litre; EMM (*E. coli* minimal media); K₂HPO₄ (7 g), KH₂PO₄ (3 g), Na₃citrate (0.5 g), (NH₄)₂SO₄ (1 g), MgSO₄ (0.1 g), 20% glucose (20 ml), and glycerol (1%, v/v) in 1 l; trace elements: 0.74 g CaCl₂·H₂O, 0.18 g ZnSO₄·7H₂O, 0.132 g MnSO₄·4H₂O, 20.1 g Na₂EDTA, 16.7 g FeCl₃·6H₂O, 0.10 g CuSO₄·5H₂O, 0.25 g CoCl₂·6H₂O. Antibiotics were added to the following working concentrations: ampicillin, 100 μ g ml⁻¹ and kanamycin, 30 μ g ml⁻¹.

UV-visible spectra were recorded on a Varian Cary 5000 at 30 \pm 0.5°C. GC–MS analysis was performed using a Shimadzu GC-17A equipped with a QP5050A GC–MS detector and DB-5 MS-fused silica column (30 m \times 0.25 mm, 0.25 μ m) or a Shimadzu GC-2010 equipped with a QP2010S GC–MS detector, an AOC-20i autoinjector, an AOC-20s autosampler, and a DB-5 MS-fused silica column (30 m \times 0.25 mm, 0.25 μ m). The injector was held at 250°C and the interface at 280°C. Column flow was set at 1.5 ml min⁻¹ and the split ratio was 24. Solvent cut time was set to 3 min. For ionone and acetate substrates, the oven was held at 80°C for 3 min followed by an increase of 10°C min⁻¹ up to 220°C and a final hold for 3 min. For 4-phenyltoluene, the initial temperature and hold time were the same but the rate of increase was 12°C min⁻¹ to a maximum of 230°C, where it was held for 5 min. Preparative HPLC analysis was carried out on a Shimadzu system equipped with a DGU-20A5R degasser, 2 \times LC-20AR pumps, SIL-20AC HT autosampler, an SPD-M20A photodiode array detector, and a CT0-20AC column oven. Separation was performed using an Ascentis Si HPLC column (25 cm \times 10 mm \times 5 μ m; Sigma–Aldrich). NMR was performed using an Agilent DD2 spectrometer at 500 MHz for ¹H and 126 MHz for ¹³C.

Sequence alignment performed by ClustalW. Phylogenetic tree drawn using the maximum-likelihood method based on the Jones–Taylor–Thornton (JTT) model with complete deletion of missing data. Initial tree (s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model and then selecting the topology with a superior log-likelihood value.

Recombinant protein expression and purification

The CYP268A2 gene (*Mmar_3761*) was amplified by PCR using oligonucleotide primers (*vide supra*). The gene was amplified by 30 cycles of strand separation at 95°C for 45 s followed by annealing at 55°C for 30 s and extension at 68°C for 80 s. The genes were cloned into the pET26 vector using the appropriate NdeI and HindIII restriction enzymes, and the correct insertion was checked by Sanger sequencing performed by Australian Genome Research Facility Ltd (AGRF). The plasmid was then transformed into *E. coli* BL21(DE3). The transformed *E. coli* cells were grown on an LB_{kan} plate and incubated in 2YT_{kan} (1.2 l total, in six flasks) at 37°C for 5 h at 160 rpm. Following this, the temperature was reduced to 18°C, the speed to 90 rpm. Benzyl

alcohol (0.02%, v/v), ethanol (2%, v/v) and, after 30 min, IPTG (0.1 mM) were added to induce protein expression [29]. The growths were continued for a further 16 h before harvesting of the cell pellet by centrifugation (5000 g, 15 min). The cells were then resuspended in 50 mM Tris (pH 7.4) with 1 mM DTT (henceforth Buffer T), with β -mercaptoethanol (1 ml), TWEEN (1 ml), and glycerol (20%, v/v), to a total volume of 200 ml. The resuspended cells were then lysed by sonication (25 cycles of 20:40 s on/off, 70%, 19 mm probe, Sonics Vibra-Cell) while kept on ice. The supernatant was isolated from the cell debris by centrifugation (40 000 g, 30 min) and then loaded onto a DEAE Sepharose column (XK50, 200 mm \times 40 mm, GE Healthcare) and eluted with a linear gradient of 100–300 mM KCl in Buffer T. The fractions containing the desired P450 were identified by their red colour, and pooled, concentrated using a Vivacell 100 (Sartorius Stedim, 10 KD membrane), and desalted using a Sephadex G-25 medium grain column (250 mm \times 40 mm) with elution using Buffer T. The protein sample was then loaded onto a further Source-Q anion exchange column (XK26, 80 mm \times 30 mm, GE Healthcare) and eluted with a gradient of 0–1 M KCl in Buffer T. Collected fractions were concentrated again by ultrafiltration and stored at -20°C in an equal volume of glycerol, after filtration with a 0.22 μM syringe filter. The final A_{420}/A_{280} ratio was 1.6.

Before use, the glycerol was removed from CYP268A2 by gel filtration, using a 5 ml PD-10 column (GE Healthcare) and elution with Buffer T (without DTT). The extinction coefficient for CYP268A2 was determined by CO difference spectra using $\epsilon_{450} = 91 \text{ mM}^{-1} \text{ cm}^{-1}$ for the reduced CO-bound form [30]. The CYP268A2 concentration was determined using $\epsilon_{419} = 108 \pm 7 \text{ mM}^{-1} \text{ cm}^{-1}$.

Spin-state shifts and substrate-binding titrations

To determine the extent of substrate binding, CYP268A2 was diluted to $\sim 1 \mu\text{M}$ in 50 mM Tris (pH 7.4) to a volume of 500 μl , and various substrates (100 mM, EtOH or DMSO) were added. The absorbance between 600 and 250 nm was recorded on the UV spectrophotometer until no further spectral change was observed. The high spin percentage was estimated ($\pm 5\%$) by comparison to a set of spectra, generated by the sum of substrate-free ($>95\%$ low spin, 418 nm Soret maximum) and camphor-bound ($>95\%$ high spin, 392 nm Soret maximum) CYP101A1 to the appropriate percentages.

For substrate-binding titrations, CYP268A2 was diluted to 2 μM in 50 mM Tris (pH 7.4) to a volume of 2.5 ml, and 1–3 μl of substrate was added via a Hamilton syringe from either 1, 10, and 100 mM (EtOH or DMSO) stock solution, starting from the lowest concentration. The peak-to-trough difference in absorbance, between 600 and 250 nm, was recorded until additional aliquots caused no further spectral change in the Soret band. The dissociation constant for that substrate was obtained by fitting the difference in absorbance against the substrate concentration to the hyperbolic function:

$$\Delta A = \frac{\Delta A_{\max} \times [S]}{K_d + [S]}$$

where K_d is the binding constant, $[S]$ is the substrate concentration, ΔA the peak-to-trough ratio, and ΔA_{\max} the maximum peak-to-trough absorbance. In the instances where the substrate exhibited tight binding ($K_d < 10 \mu\text{M}$, less than five times the concentration of the enzyme), the data were instead fitted to the tight-binding quadratic equation:

$$\Delta A = \Delta A_{\max} \times \frac{[E] + [S] + K_d - \sqrt{([E] + [S] + K_d)^2 - 4[E][S]}}{2[E]}$$

where K_d is the binding constant, $[S]$ is the substrate concentration, ΔA the peak-to-trough ratio, ΔA_{\max} the maximum peak-to-trough absorbance, and $[E]$ is the enzyme concentration [31].

Construction of *in vivo* systems to support product formation

To construct whole-cell turnover systems, CYP268A2 was cloned into a pRSFDuet vector using the NdeI and KpnI sites introduced by PCR. ArR (a ferredoxin reductase from *Novosphingobium aromaticivorans*) has been cloned into pETDuet previously [32]. The terpredoxin gene was purchased (gblock; Integrated DNA Technologies) and cloned into pETDuetArR using NcoI and HindIII sites (Supplementary Material). The two vectors, pRSFDuet-containing CYP268A2, and pETDuet, containing Tdx [terpredoxin (a ferredoxin from

Pseudomonas sp.]) and ArR, were transformed into *E. coli* BL21 cells. The transformed cells were grown overnight on an LB_{amp/kan} plate. A colony was picked and grown in 50 ml LB_{amp/kan} for 6 h at 37°C and 110 rpm. It was then cooled to 18°C and slowed to 90 rpm, with the addition of benzyl alcohol (0.02%, v/v) and ethanol (2%, v/v), followed after 30 min by IPTG (0.1 mM). The culture was then left overnight. The cell pellet was harvested by centrifugation (5000 g, 10 min) before being resuspended in EMM_{amp/kan} (100 ml). The substrate was then added to 1 mM final concentration before shaking at 150 rpm at 30°C. After 24 h, the turnover was then centrifuged (15 min, 5000 g) and the supernatant was isolated. Samples (1 ml) of the turnover were taken for initial testing via GC–MS at various time points. The samples were extracted into ethyl acetate, dried over MgSO₄ before resuspension in anhydrous acetonitrile (200 µl). Where GC–MS showed product formation, larger scale growths by the same method (200 ml EMM) were performed and extracted. The extract was dissolved in 5% isopropanol:hexane and purified by semi-preparative HPLC using an elution gradient of 5–10% isopropanol. The chromatogram was monitored at 254 nm. Relevant peaks (as confirmed by GC–MS) were collected, pooled, and resuspended in deuterated chloroform to allow analysis by NMR.

In vitro NADH activity assays

Purified CYP268A2 (0.5 µM) together with the ferredoxin Tdx (5 µM) and ferredoxin reductase ArR (0.5 µM) (the purification method of these are reported elsewhere [32,33]) were mixed to a total volume of 1.2 ml in oxygenated 50 mM Tris (pH 7.4), with added catalase (120 µg). The mixture was equilibrated for 2 min at 30°C before the addition of 320 µM NADH ($A_{340} \approx 2.0$). Substrate was then added to a concentration of 0.25 mM. The reaction was monitored at 340 nm for the duration. The rate of NADH turnover was calculated by plotting the A_{340} against time, using $\epsilon_{340} = 6.22 \mu\text{M}^{-1}\text{cm}^{-1}$. Once the reaction was at completion, 1 ml of the turnover was extracted into ethyl acetate and analysed by GC–MS as above.

Crystallography, data collection, data processing, and structural determination

CYP268A2 was further purified by size-exclusion chromatography (Enrich SEC Column, 650 10 × 300 mm, 1 ml min⁻¹ flow rate) before being concentrated to 30–35 mg ml⁻¹ in Buffer T. The substrate pseudoionone (100 mM EtOH, mixture of isomers) was added to the protein to a final concentration of 1 mM immediately before crystallisation. The initial screening of crystallisation conditions was performed by the sitting-drop method in 96-well plates with commercially available screening conditions (Hampton Research) using 1 µl of both the protein solution and the reservoir solution. Crystal conditions were refined using the hanging drop vapour diffusion method, and again with 1 µl of both the protein solution and the reservoir solution, equilibrated with a 500 µl reservoir. Diffraction-quality crystals (plates with dimensions ~150 × 140 × 20 µm, Supplementary Figure S4) were obtained after 2 weeks at 16°C from the condition containing 0.96 M ammonium phosphate and 0.3 M sodium citrate (pH 7.0). They were harvested with a Micromount (MiTeGen) and cryo-protected by immersion in Parabar 10312 (Paratone-N, Hampton Research) before flash cooling in liquid N₂. Data were collected by X-ray diffraction at 100 K on the Australian Synchrotron MX1 beamline (360 exposures using 1° oscillations at a wavelength of 0.9537 Å) [34]. The data were processed in the space group C2 using XDS [35,36]. Molecular replacement phasing was carried out using the modified search model (residues 1–23, 145–152, 322–337, and 244–253 were removed from the surface to improve the search model clashes) from PDB entry 2WM4 [21] (CYP124A1 with phytanic acid bound, found by searching the FFAS server) with the MRSAD (Molecular Replacement Single Wavelength Anomalous Diffraction) protocol of Auto-Rickshaw [37,38] using the native dataset (however, SAD failed and phasing were obtained by MR only). Within the pipeline, various programmes from the CCP4 program suite [39] were used and the model phases were improved by model refinement using CNS [40,41] and REFMAC5 [42], density modification using PIRATE [43], and rebuilding of the model using SHELXE [44], RESOLVE [45], and Buccaneer [46], finally refinement of the resulting model using Phenix [47] and REFMAC5. The structure was further rebuilt using Coot [48] based on the initial electron density maps, with multiple structural refinement iterations using phenix.refine. Composite omit maps were generated using the Composite Omit Maps program in Phenix.

Results/discussion

Phylogenetic and sequence analysis

Alongside CYP268A2, genes encoding members of the cytochrome P450 monooxygenase CYP268 family are primarily found in other *Mycobacterium* species. These include CYP268A1 from *M. avium* subsp.

paratuberculosis, A3 and B2 from *M. smegmatis*, and B1 and C1 from *M. vanbaalenii* (accession numbers and similarities are listed in Supplementary Table S1). A single member, CYP268A4 is found in *Streptomyces bingchengensis*. A BLAST search found a large number of proteins with high sequence identity (250 results >70%), almost exclusively from other *Mycobacterium* species (none of the first 250 results were from non-*Mycobacteria* species). CYP268A2 is conserved in *Mycobacterium liflandii* (98%) and as a pseudogene in *M. ulcerans* Agy99 (95%, truncated after 195 residues). A similar gene without truncation is found in *M. ulcerans* subsp. *shinshuense* (97%), a clinical isolate from Japan [49]. Close analogues are also found in *M. avium* (78%), *M. colombiense* (77%), and *M. kansasii* (85%), among others. A *M. tuberculosis* strain TKK-01-0051 contained an analogous protein with high sequence similarity (78%), although there is some suggestion that this strain may be misclassified and is better referred to as *M. colombiense* [50]. *M. colombiense* is a member of the *M. avium* complex and is known to opportunistically infect HIV-positive immuno-compromised patients [51].

The CYP268 family also shares high sequence similarity with the CYP124, CYP125, and CYP142 families, of which members from *M. tuberculosis*, *M. marinum*, *M. smegmatis*, and *M. vanbaalenii* are known (Supplementary Table S1). In particular, CYP268A2 has a sequence similarity above 40% with many members of the CYP124 family. The CYP142B subfamily members cluster together (Figure 1), more closely to the 124 and 268 families than to the 142A subfamily, suggesting a high degree of overlap between the three families. CYP124A1 from *M. tuberculosis* (with 41% similarity to CYP268A2) has been characterised as a lipid hydroxylase [21] and additionally can oxidise vitamin D3 and other analogues [52]. CYP142A1, also from *M. tuberculosis* (33% similarity to CYP268A2), can oxidise cholesterol esters [53], and CYP125A1 (40%) is similarly a cholesterol hydroxylase [54].

A sequence alignment (Supplementary Figure S1) revealed that of the cytochrome P450 commonly conserved sequence elements, CYP268A2 retains the glutamate and arginine pair (Glu300, Arg303), as well as the phenylalanine residue (Phe364) in the K helix, seven residues before the conserved proximal cysteine at Cys372 [55]. The acid alcohol pair in the enzyme is an aspartate (Asp263) and threonine (Thr264). The CYP124A1 family members contain a glutamate (Glu270 in CYP124A1) rather than the aspartate found in the CYP268 enzymes (Supplementary Figure S1).

The area surrounding putative *cyp268* genes is marked by the presence of a highly conserved operon (Supplementary Figure S2), containing a GTPase, a large ribonuclease, two regulatory proteins (AcrR and Sir2-like), and a downstream NAD synthetase. The *cyp268A2* gene is flanked on both sides by a PE-PGRS (proline glutamate-polymorphic GC-rich sequence proteins) gene (glycine-rich proteins detected across the *Mycobacteria* with possible roles as antigens [56] or fibronectin-binding [57]). The environments of the *cyp124* genes (*M. marinum* and *M. tuberculosis*) do not have any of these genes, but share some similarity with each other (Supplementary Figure S2).

Characterisation of CYP268A2 and its substrate range

The CYP268A2 enzyme was produced in *E. coli* and purified by two ion exchange chromatography steps. The protein was tested for the characteristic Soret absorbance, occurring when the ferrous form of the CYP enzyme binds CO [58]. CYP268A2 after reduction by sodium dithionite and gentle bubbling with CO shifted almost completely from the resting state absorbance at 419–450 nm, with only a small shoulder (<5%) at 420 nm (Figure 2). The extinction coefficient for the Soret absorbance [58], $\epsilon_{450} = 91 \text{ mM cm}^{-1}$, was used to determine the extinction coefficient for CYP268A2 $\epsilon_{419} = 108 \pm 7 \text{ mM cm}^{-1}$, which was henceforth used to determine the concentration of the enzyme.

As CYP268A2 is the first member of its P450 family to be studied, previously characterised family members are not available to give an indication as to the role of the enzyme. Based on the similarity to members of the CYP124 lipid-hydroxylase family [21], the initial substrates tested on CYP268A2 were branched and straight chain fatty acids and esters (Figure 3). Many of these were successful in shifting the majority of the enzyme into the high spin form, indicating that they were accommodated by the active site of the enzyme (Supplementary Figure S5). Geranyl acetate (80% HS) and farnesyl acetate (75% HS) both induced higher spin-state shifts in CYP268A2 than undecanoic acid (70% HS), which was the best performing acid substrate (Table 1). Additionally, it was discovered that a range of aromatic compounds could bind to the enzyme. 4-Phenyltoluene (55% HS), phenyl acetate (50% HS), and phenylcyclohexane (50% HS) all successfully induced a Type I Soret shift. The binding affinity of CYP268A2 for the substrates that gave the highest spin-state shifts was then assessed (Figure 4). Farnesol (75% HS) bound to the enzyme with high affinity, K_d , $0.8 \pm 0.2 \mu\text{M}$. Undecanoic acid gave K_d , $1.1 \pm 0.5 \mu\text{M}$. Pseudoionone, a linear ionone precursor, gave 80% HS and K_d , $3.6 \pm$

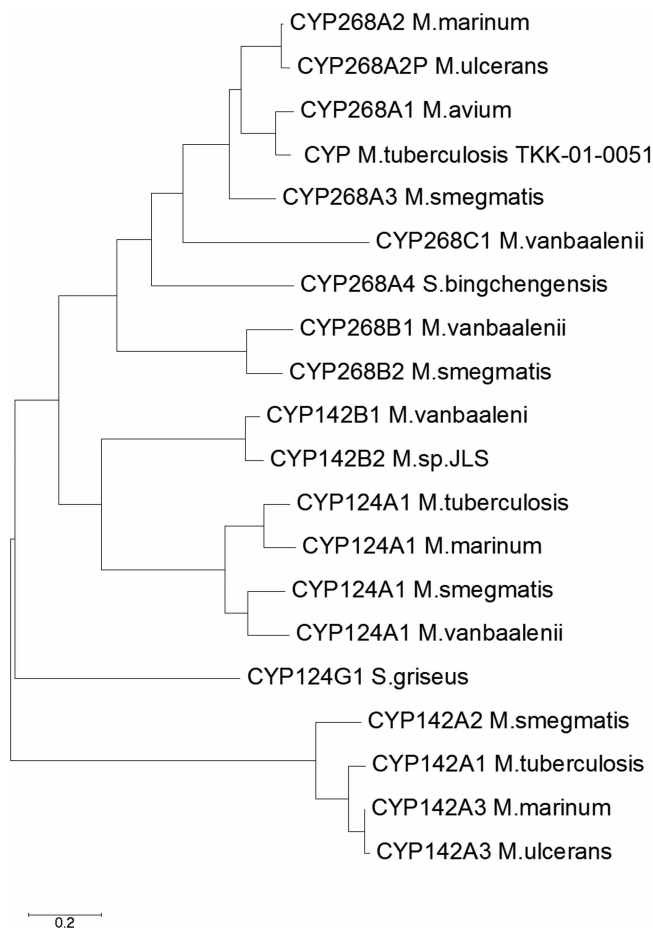


Figure 1. Phylogenetic tree of the CYP268 family.

Included are some members of the CYP124 and CYP142 families, both of which have been predominately found in *Mycobacterium* species. The grouping shows that CYP268A2 is closely related to members of the 268 family, with significant sequence similarity to the 124 family (all above 40%) and the 142B subfamily (43% to 142B1). CYP124G1, which does not cluster with the remainder of the 124 family, has 41% similarity to CYP268A2. Percentage identities can be found in Supplementary Table S1. The tree was drawn to scale, with branch lengths and scale measured in the number of substitutions per site.

0.6 μM (as a mixture of isomers). 4-Phenyltoluene also bound tightly, K_d , $13 \pm 2.4 \mu\text{M}$. Notably, the longer substrates preferred by CYP124A1 such as phytanic acid failed to induce a significant spin-state shift in CYP268A2 (Supplementary Table S2).

The enzyme did not appear to have the strict requirement for branching methyl groups at the terminus of the substrate as shown by CYP124A1, with lauric acid (55% HS) outperforming both 11-methylauric acid (30% HS) and 10-methylauric acid (20% HS), although 15-methylhexadecanoic acid (15%) was more effective than palmitic acid (0%). The binding data suggested that the enzyme active site could better accommodate straight chain substrates over those with longer bent chains, preferring the *trans*-isomer, geranyl acetate (K_d , $8.5 \pm 1.9 \mu\text{M}$), over the *cis*-form, neryl acetate (30% HS, K_d , $106 \pm 29 \mu\text{M}$, Supplementary Figure S7). CYP124A1 has additionally been characterised as having cholesterol and vitamin D binding activity [52] (as do CYP125 and CYP142 family members, [54,59]). As a result, CYP268A2 was tested with a variety of cholesterol and vitamin D analogues, but those gave no or very little (0–5% high spin form) indication of binding. The substrate-binding data demonstrate that the active site of CYP268A2 is versatile and can accommodate a range of linear and aromatic hydrocarbons, terpenes, and fatty acids. Some substrates such as farnesol that have been reported for CYP124A1 bind well with CYP268A2, but it appears to support a wider range of substrate binding.

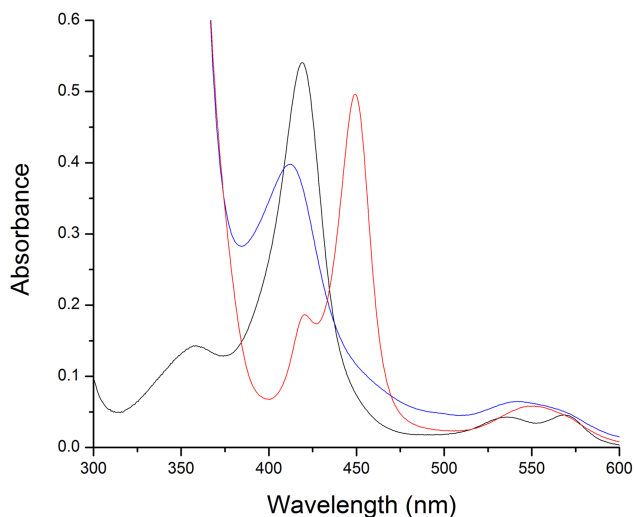


Figure 2. Ferric CYP268A2 (black, A_{419}), the reduced ferrous form (blue, A_{412}), and the ferrous form bound with CO (red, A_{449}), showing the characteristic absorbance at ~450 nm.

The shoulder at 420 nm in the ferrous–CO form comprises ~5% of the total area.

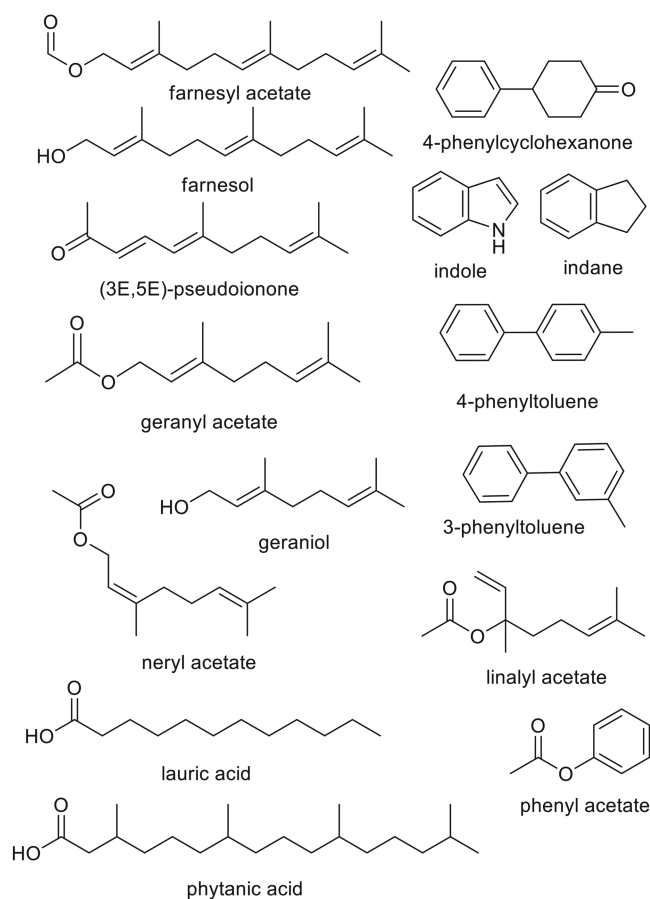


Figure 3. Structures of compounds tested as potential CYP268A2 substrates.

Table 1 Spin-state shift and dissociation constants of CYP268A2 with a variety of substrates, presented in descending order of magnitude of the type I spin-state shift.

The spin-state shifts of a range of additional substrates are listed in Supplementary Table S2.

CYP268A2 substrates	Spin-state shift (% HS)	K_d (μM)
Geranyl acetate	80	8.5 ± 1.9
Pseudoionone	80	3.6 ± 0.6
Farnesol	75	0.8 ± 0.2
Farnesyl acetate	75	5.1 ± 1.9
10-Undecenoic acid	70	1.6 ± 0.2
Undecanoic acid	70	1.1 ± 0.5
Capric acid	65	1.1 ± 0.5
4-Phenyltoluene	55	13 ± 2.4
Geraniol	55	7.9 ± 1.6
Lauric acid	55	270 ± 36
Linalyl acetate	55	110 ± 25
Neryl acetate	30	106 ± 29

Manyazole drugs have been reported to bind CYPs as competitive inhibitors, coordinating directly to the Fe atom of the haem and generating Type II spectral shifts, shifting the Soret band to a higher wavelength rather than to 390 nm which is characteristic of displacement of the distal water ligand [60,61]. These have been proposed as methods for inhibiting bacterial growth, particularly in species where CYPs play essential roles [62]. Many possibleazole inhibitors were tested with CYP268A2 (Table 2). CYP268A2 gave Type II shifts with two of these, 1-phenylimidazole (shifted the Soret maximum to 421 nm) and 4-phenylimidazole (shifted the Soret maximum to 423 nm). The binding affinity of these to CYP268A2 was assessed by determining the dissociation constant (Figure 4). 1-Phenylimidazole bound more tightly (K_d , $0.9 \pm 0.3 \mu\text{M}$) than 4-phenylimidazole (K_d , $4.5 \pm 0.6 \mu\text{M}$). The addition of 2-phenylimidazole and other largerazole inhibitors yielded a small Type I shift.

Product characterisation

As CYP268A2 has no closely located ferredoxin gene in the *M. marinum* genome, it was expressed in *E. coli* with a variety of electron transfer partners from different organisms (Supplementary Table S3). The pair terpre-doxin (2Fe–2S ferredoxin, from a *Pseudomonas* sp.) and an FAD-containing ferredoxin reductase ArR (from *N. aromaticivorans*) were selected (based on levels of product formation with geranyl acetate, Supplementary Material) to enable characterisation of the products of the enzyme. *In vitro* turnovers were used to analyse the product formation rate and where product was formed, substrates were scaled up with *in vivo* turnovers to allow characterisation.

The addition of geranyl acetate to CYP268A2 *in vitro* gave an NADH consumption rate of $\sim 60 \text{ nmol P450}^{-1} \text{ min}^{-1}$ and when extracted and analysed by GC–MS, it showed a major product (9.5 min, Figure 5a) with a mass peak of 152.15 (expected mass of hydroxylation product is 214.16, 152.12 with loss of the acetate) after 24 h. To characterise the product, large-scale *in vivo* turnovers were performed (Figure 5a). The *in vivo* major product had a slightly different retention time at 9.4 min and the mass peaks (154.25, 136.20, and others) were two mass units higher than those of the *in vitro* product. The pattern of products in the GC trace of the whole-cell turnovers over time combined with mass spectrum peaks suggests that the major product *in vivo* was also hydrogenated in addition to the CYP-mediated hydroxylation (Supplementary Figure S8). HPLC purification of the major product followed by NMR analysis identified the metabolite as the terminal hydroxylation product (*E*)-8-hydroxy-3,7-dimethyloct-2-en-1-yl acetate (Figure 6 and Supplementary Figure S15). The *in vitro* turnover experiment with geranyl acetate confirms that the hydroxylation is

CYP-mediated while it is likely that the hydrogenation was performed by an endogenous *E. coli* ene reductase [63,64].

The *in vitro* turnover of pseudoionone by CYP268A2 had an NADH consumption rate of 48 nmol nmolP450⁻¹ min⁻¹ and generated a single major product at 16.1 min with a mass peak at 208.1 (expected hydroxylation product at 208.15). The addition of pseudoionone to the *in vivo* CYP/Tdx/ArR system generated a single major product after 24 h with a retention time of 15.9 min (Figure 5b). Mass peaks of the *in vivo* product (210.2, 195.20, and others) were again two mass units higher than the *in vitro* product (Supplementary Figure S9). The major product from the whole-cell turnover was isolated by HPLC and characterised by NMR to be (3*E*,5*E*)-11-hydroxy-6,10-dimethylundeca-3,5-dien-2-one (Figure 6 and Supplementary Figure S16). Similarly to geranyl acetate, the *in vivo* product has been hydrogenated at the ω-2 alkene in addition to ω hydroxylation. The pseudoionone was a mixture of isomers, and there are two substrate peaks at 12.0 and 12.7 min (192.15 and 192.20 *m/z* compared with expected molecular ion peak at 192.15), which are the *cis* (3*E*,5*Z*) and *trans* (3*E*,5*E*) isomers, respectively. In both the *in vitro* and *in vivo* turnovers, the enzyme showed a preference for the *trans*-pseudoionone isomer, which was consumed over the course of the turnover, whereas the *cis* was not.

A product from the CYP268A2 catalysed oxidation of 4-phenyltoluene was also identified by GC–MS (Supplementary Figure S10). This was assigned as 4-biphenylmethanol by MS comparison and GC co-elution experiment with other P450 turnovers [65]. Other substrates, including undecanoic acid, neryl acetate, farnesol, farnesyl acetate, and geraniol, were tried but no product was detected. The selectivity for *trans*- over *cis*-pseudoionone and geranyl over neryl acetate, which aligns with the binding data, indicates that the enzyme prefers the straight chain isomer over the bent form. The absence of product for geraniol suggests that substrates with the acetate ester moiety are favoured. The catalytic turnover rate reported is expected to be limited by the use of non-native electron transfer partners, which generally support lower levels of oxygenase activity [66]. However, the ability of Tdx to support CYP268A2 activity is a strong indication that the physiological electron transfer partner may be a [2Fe–2S] ferredoxin. Ideally, the native transfer partners of CYP268A2 from *M. marinum* would be identified. The availability of known substrates of CYP268A2 which have demonstrated product formation will facilitate this process.

Crystal structure of pseudoionone-bound CYP268A2

Crystallisation of both substrate-bound and substrate-free CYP268A2 was attempted. When pseudoionone was added before crystallisation, CYP268A2 formed sharp-edged single crystals after 2 weeks and diffraction data were collected to 2.0 Å at the Australian Synchrotron. No suitable crystals of the substrate-free form were obtained. The solved structure consists of a single polymer chain in the asymmetric unit and the *trans* (3*E*,5*E*) form of the substrate pseudoionone in the active site (PDB code: 6BLD). All residues were modelled except the first five of the N-terminus. Refinement statistics for the structure are located in Table 2. The overall fold conforms closely to the canonical P450 fold (Figure 7b). A loop of residues (Pro46 to Phe61) between two β-sheet regions forms an active site ‘cap’ similar to that of CYP124A1 [21]. The electron density map shows that the pseudoionone is arranged with one of the ω methyl groups held directly over the haem (4.3 Å away from the Fe atom). The carbonyl group of the substrate interacts with residues from the G helix and the B–C loop, towards the apparent exit of the active site (Figure 7a). The residues of the active site that interact directly with the carbonyl of pseudoionone, Gln209 and Ser99, sit close enough at 2.8 and 3.4 Å, respectively, to form hydrogen bonds (Figure 7c). This arrangement provides structural justification for the substrate-binding preference for oxygen-containing groups at one end of the substrate (acetates, acids, or alcohols). Geranyl acetate, undecanoic acid, and other similar substrates would, in theory, have a functional group that could also interact in a similar manner and position.

The pseudoionone molecule appears to be completely enclosed in the active site (Figure 8), suggesting that CYP268A2 has crystallised in the ‘closed form’ of the P450 [67]. Indeed, the enzyme completely encloses the substrate, showing no access to the substrate channel from the surface (Supplementary Figure S11a). The active site residues (12 within 4 Å of the substrate, Figure 7c) together create a linear substrate-binding pocket which seems likely to preferentially exclude the *cis* form of pseudoionone, which was also present in the crystallisation conditions but not observed binding in the solved structure. This further supports the product formation data, which have shown that the enzyme preferentially hydroxylates the *trans* isomer of pseudoionone at the ω terminus from the carbonyl group. The only significant vacant space in the active site is near the haem, where

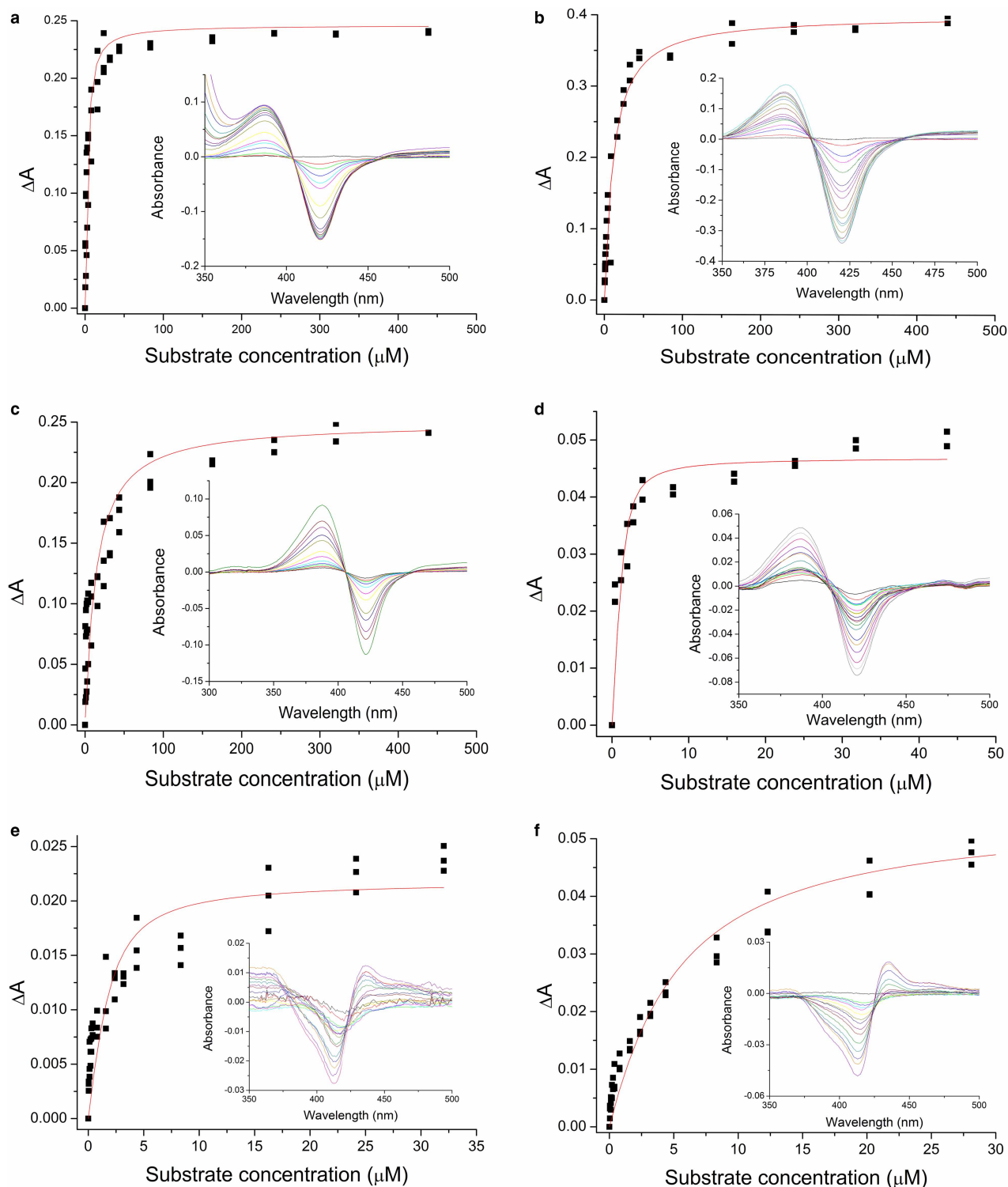


Figure 4. Dissociation constant analysis of CYP268A2 with selected substrates.

(a) Pseudoionone, (b) geranyl acetate, (c) 4-phenyltoluene, (d) undecanoic acid, (e) 1-phenylimidazole, and (f) 4-phenylimidazole. The inset shows the spectral change upon titration with substrate. The peak-to-trough difference in absorbance was measured from 420 to 390 nm, except with 1-phenylimidazole (peak-to-trough 413–435 nm) and 4-phenylimidazole (414–434 nm), where Type II spectral shifts were recorded.

Table 2 Crystal refinement data for CYP268A2 from *M. marinum* (PDB code: 6BLD).

Data collection statistics ¹	
Wavelength	0.95370
Unit cell	$a = 154.76, b = 44.726, c = 57.727$ $\alpha = 90, \beta = 100.84, \gamma = 90$
Space group	C2
Number of molecules in asymmetric unit	1
Resolution	2.00–42.91 (2.00–2.05) ²
Number of unique reflections	26 452 (1743)
Completeness	99.2 (89.4)
Redundancy	7.4 (7.0)
$\langle I \rangle / \langle \sigma(I) \rangle$	9.7 (2.0)
R_{merge} (all $I+$ and $I-$)	0.130 (0.742)
R_{pim} (all $I+$ and $I-$)	0.071 (0.411)
CC(1/2)	0.997 (0.768)
R_{work}	19.96%
R_{free}	24.79%
% solvent	42.89
Number of residues modelled	413
RMS deviation from restraint values	
Bond lengths	0.002
Bond angles	0.551
Ramachandran analysis	
Most favoured	98.06%
Additionally allowed	1.94%

¹Data collected from one crystal.
²Values in parenthesis are for highest resolution shell.

there is room to accommodate a larger group which would rationalise the binding of phenyltoluene-like molecules and the phenylimidazole inhibitors.

Two obvious water channels from the surface to the haem are observed, and these are similar to those found in other CYPs. One, referred to as the ‘water channel’, approaches the coordinating cysteine from the proximal face of the enzyme, beginning from the base of the B–C loop (Supplementary Figure S12), while the other, on the distal side between the E, F, and I helices, is the ‘solvent channel’ [68]. Residues of the I helix, including the acid Asp263, participate in the hydrogen-bonding network, by interacting directly with the water molecules that comprise the solvent channel (Supplementary Figure S13). The water channel is thought to be involved in active site solvation, and the solvent channel to be responsible for proton relay [68].

Comparison of the structure of the enzyme to the close relative CYP124A1 was made (Figure 7b). The active site of CYP124A1 accommodates larger substrates (~C16), and a phytanic acid-bound structure is available (PDB code: 2WM4) [21]. This structure shows the substrate bound in a similar manner to that of pseudoionone close to the haem, but the carboxylate end of the phytanic acid curves into a pocket that is not available in the CYP268A2 enzyme, as a result of the presence of a tryptophan residue, Trp90 (Figure 7d). Both the Gln209 and Ser99 residues are not present in CYP124, replaced by a serine group (Ser216, which is shifted further away from the active site) and a phenylalanine (Phe107), respectively, neither of which interact electrostatically with the longer phytanic acid. Furthermore, the hydrophilic Gln100 is not present in CYP124A1 (replaced by Gly108) and the Thr264 (of the acid–alcohol pair) is flipped (Thr271 in CYP124A1). The position of Thr264 also affects the commonly conserved hydrogen bond between the alcohol and the nearby alanine

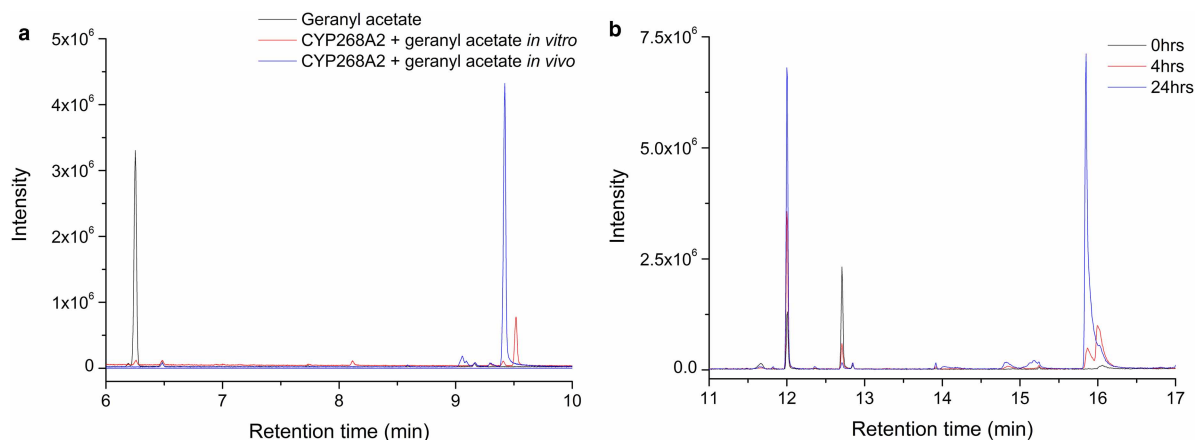


Figure 5. GC chromatograms of CYP268A2 catalysed substrate turnovers.

(a) GC chromatogram of the *in vivo* and *in vitro* turnovers of CYP268A2 with geranyl acetate. The retention times are as follows: geranyl acetate 6.3 min, *in vivo* product 9.4 min, *in vitro* product 9.5 min. (b) *In vivo* turnover with pseudoionone monitored over 24 h. Identification of isomer peaks was done by retention time comparison with neryl acetate (*cis*) and geranyl acetate (*trans*), where the *cis* form had the shorter retention time. The retention times are: *cis*-pseudoionone 12.0 min, *trans*-pseudoionone 12.7 min, 4 h products 15.9 min and 16.1 min, 24 h product 15.9 min.

carbonyl (Ala260 in CYP268A2 and Ala267 in CYP124A1), with the bond distance in CYP268A2 2.7 Å compared with 3.7 Å in CYP124A1 (Supplementary Figure S14). The position of the threonine in CYP268A2 matches the orientation of the equivalent residue in camphor-bound P450_{cam}, which has a distance of 2.5 Å [69]. The tightness of this hydrogen bond is hypothesised to affect solvent access to the active site of the enzyme: in the closed position, solvent access is restricted, in the open form allowed [70].

The crystal structure of pseudoionone-bound CYP268A2 provides important insights into the substrate selectivity of these enzymes. The comparison with related CYP124A1 from *M. tuberculosis* provides an understanding on how these enzymes have evolved to modify their substrate selectivity. *M. marinum* also contains a gene encoding a CYP124A1 protein (84% similarity to CYP124A1 from *M. tuberculosis*), and both could have evolved from a common ancestor. At some point in time, the progenitors of the CYP268A2 and CYP124A1 enzymes may have had an overlapping function, providing genetic redundancy, and hence CYP268A2 may only be maintained in the larger genome of the less immuno-challenged organism. However, our data suggest that CYP268A2 would support a more varied range of substrate hydroxylation in the native system than CYP124A1. Along with other protein-encoding genes, the broad substrate range of the CYP268A2 system potentially aids *M. marinum* in surviving in more diverse environments.

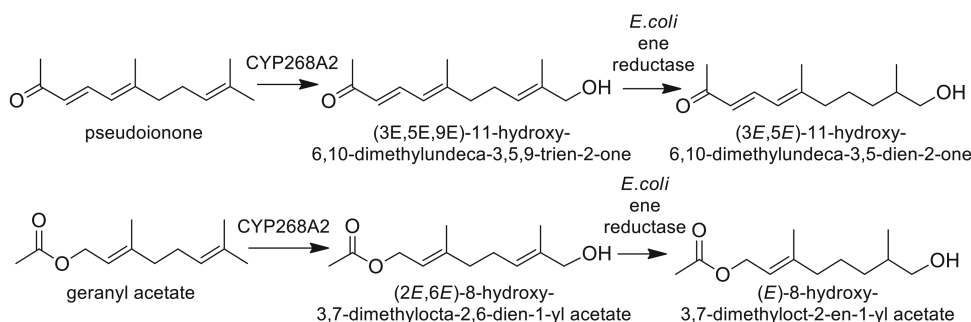


Figure 6. Hydroxylation products of CYP268A2.

*In each, the hydroxylation at ω -terminus would generate a stereo-centre at the ω -1 carbon. However, *in vivo* the hydrogenation occurred after the hydroxylation, and as a result, the stereo-selectivity of the isolated product would be dependent on the *E. coli* ene reductase enzyme.

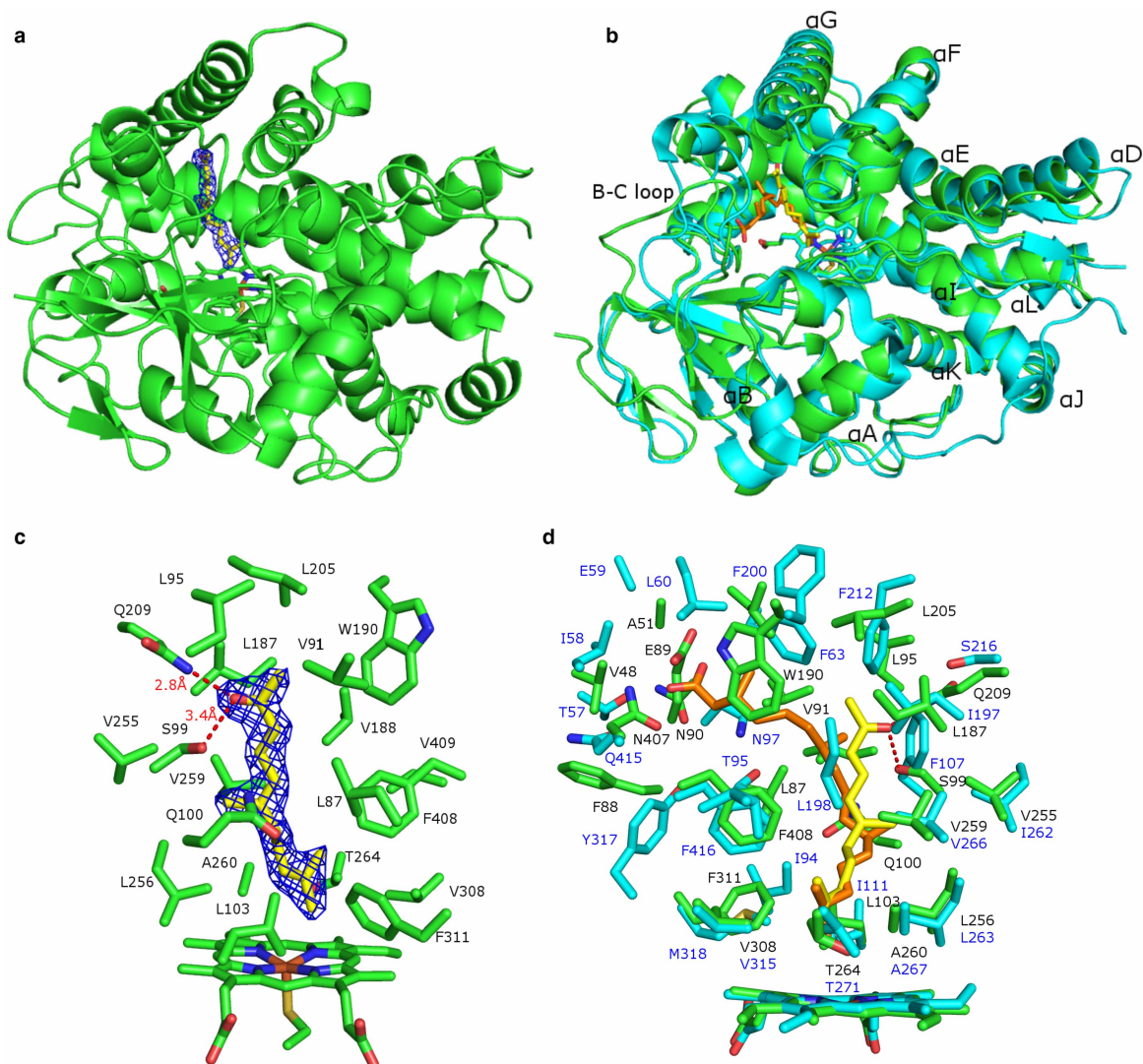


Figure 7. X-ray crystal structure of CYP268A2.

(a) The overall structure of CYP268A2 (green) from *M. marinum* with pseudoionone bound (yellow), showing the composite omit $2F_o - F_c$ electron density at $\sigma = 1.5$ (blue mesh) of the pseudoionone in the active site and (b) CYP268A2 (green) with pseudoionone (yellow) bound overlaid with *M. tb* CYP124A1 (blue) with phytanic acid (orange) bound, showing the conserved P450 helices [55]. (c) The active site region of CYP268A2 showing side chains of amino acids within 5 Å of the pseudoionone molecule. There are 12 residues within 4 Å: S99, Q100, L103, L187, L205, Q209, L256, V259, A260, T264, F311, and F408. The terminal carbonyl of the pseudoionone has polar contacts (red) with the amine group of Q209 and the hydroxyl group of S99. (d) An overlay of the active site regions of CYP268A2 and CYP124A1 showing side chains of amino acids from both enzymes within 5 Å of the pseudoionone and phytanic acid molecules (CYP268A2 labels in black and CYP124A1 blue).

Conclusion

The larger CYP complement of *M. marinum* contains several CYP families, including CYP268, which are widespread across *Mycobacterium* species but are absent in *M. tuberculosis*. The characterisation of CYP268A2 demonstrates that the selectivity and versatility of these enzymes can vary significantly even where there is structural and sequence similarity in the active site. CYP268A2 is the first member of this family to be characterised, and the substrate range of the enzyme reported here is broad with several molecules binding with high affinity. The successful hydroxylation of a long-chain branched acetate and ketone demonstrates that activity can be efficiently reconstituted with non-native electron transfer proteins including a [2Fe–2S] ferredoxin. This

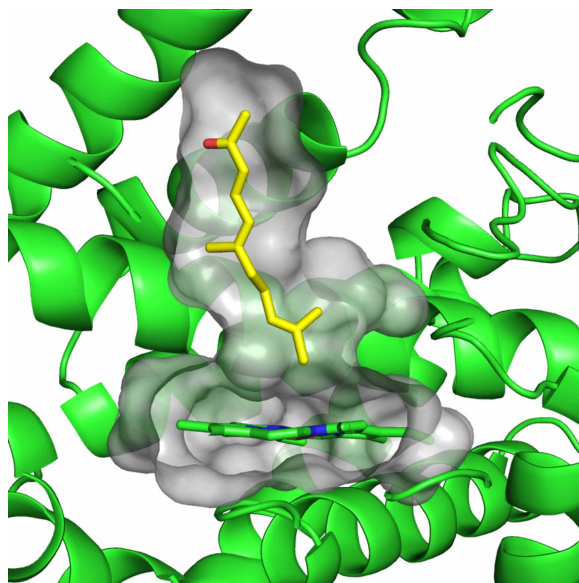


Figure 8. CYP268A2 (green) with the inner surface cavity of the enzyme active site (transparent grey) which encloses the pseudoionone substrate (yellow).

There is a small empty cavity near the haem.

will assist with future attempts to identify the native electron transfer proteins. The substrate binding and turnover data presented here may be only the first of a broader range of molecules that can be oxidised by CYP268 family members. Furthermore, the crystallisation and structural analysis rationalises the observed catalytic activity and forms the basis for any future attempt to improve it. Finally, potential azole inhibitors were identified for the enzyme, which bind to the haem iron. As CYP268 family members are present across a range of *Mycobacterium* species including human pathogens, this could form the basis of future inhibitor design against bacterial infection.

Abbreviations

ArR, a ferredoxin reductase from *Novosphingobium aromaticivorans*; CYP, Cytochrome P450 enzyme; DTT, dithiothreitol; EMM, *E. coli* minimal media; FAD, flavin adenine dinucleotide; IPTG, Isopropyl β -D-1-thiogalactopyranoside; JTT, Jones–Taylor–Thornton; LB, Lysogeny broth (also known as Luria or Lennox Broth); MAC, *Mycobacterium avium* complex; MTBC, *Mycobacterium tuberculosis* complex; Tdx, terpredoxin (a ferredoxin from *Pseudomonas* sp.); PCR, polymerase chain reaction; PDIM, phthiocerol dimycocerosates; PE-PGRS, proline glutamate-polymorphic GC-rich sequence proteins.

Author Contribution

S.A.C., E.F.N., and S.G.B. performed molecular biology experiments. S.A.C. and J.B.B. determined the crystal structure. S.A.C. and S.G.B. wrote the manuscript.

Funding

This work was supported by the Australian Research Council through a Future Fellowship [FT140100355 to S.G.B.]. The authors also acknowledge the award of University of Adelaide Faculty of Sciences Divisional Scholarship (PhD to S.A.C.). X-ray diffraction data collection was undertaken on the MX1 beamline at the Australian Synchrotron, part of the Australian Nuclear Science and Technology Organisation. We acknowledge financial support from the Australian Synchrotron in the form of MXCAP9674.

Acknowledgements

The authors thank Santosh Panjikar for his help with onsite phasing at the Australian Synchrotron and Associate Professor Tara Pukala (University of Adelaide) for assistance with mass spectrometry. We acknowledge and are

grateful to Professor Tim Stinear from the University of Melbourne and Professor Lalita Ramakrishnan of the University of Cambridge (formerly University of Washington) for providing the genomic DNA of *M. marinum*.

Competing Interests

The Authors declare that there are no competing interests associated with the manuscript.

References

- Ortiz de Montellano, P.R. (2015) Substrate oxidation by cytochrome P450 enzymes. In *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Ortiz de Montellano, P.R., ed.), pp. 111–176. Springer International Publishing
- Guengerich, F.P. (2015) Human cytochrome P450 enzymes. In *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Ortiz de Montellano, P.R., ed.), pp. 523–786. Springer International Publishing
- Nelson, D.R., Zeldin, D.C., Hoffman, S.M.G., Maltais, L.J., Wain, H.M. and Nebert, D.W. (2004) Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. *Pharmacogenetics* **14**, 1–18 <https://doi.org/10.1097/00008571-200401000-00001>
- Munro, A.W., McLean, K.J., Marshall, K.R., Warman, A.J., Lewis, G., Roitler, O. et al. (2003) Cytochromes P450: novel drug targets in the war against multidrug-resistant *Mycobacterium tuberculosis*. *Biochem. Soc. Trans.* **31**, 625–630 <https://doi.org/10.1042/bst0310625>
- Guengerich, F.P. (2001) Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem. Res. Toxicol.* **14**, 611–650 <https://doi.org/10.1021/bx0002583>
- Khatiri, Y., Girhard, M., Romankiewicz, A., Ringle, M., Hannemann, F., Urlacher, V.B. et al. (2010) Regioselective hydroxylation of norisoprenoids by CYP109D1 from *Sorangium cellulosum* So ce56. *Appl. Microbiol. Biotechnol.* **88**, 485–495 <https://doi.org/10.1007/s00253-010-2756-3>
- Scheps, D., Malca, S.H., Hoffmann, H., Nestl, B.M. and Hauer, B. (2011) Regioselective ω -hydroxylation of medium-chain n-alkanes and primary alcohols by CYP153 enzymes from *Mycobacterium marinum* and *Polaromonas* sp. strain JS666. *Org. Biomol. Chem.* **9**, 6727–6733 <https://doi.org/10.1039/c1ob05565h>
- Hannemann, F., Bichet, A., Ewen, K.M. and Bernhardt, R. (2007) Cytochrome P450 systems — biological variations of electron transport chains. *Biochim. Biophys. Acta, Gen. Subj.* **1770**, 330–344 <https://doi.org/10.1016/j.bbagen.2006.07.017>
- Nelson, D.R., Koymans, L., Kamataki, T., Stegeman, J.J., Feyereisen, R., Waxman, D.J. et al. (1996) P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* **6**, 1–42 <https://doi.org/10.1097/00008571-199602000-00002>
- Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D. et al. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544 <https://doi.org/10.1038/31159>
- Hudson, S.A., McLean, K.J., Munro, A.W. and Abell, C. (2012) *Mycobacterium tuberculosis* cytochrome P450 enzymes: a cohort of novel TB drug targets. *Biochem. Soc. Trans.* **40**, 573–579 <https://doi.org/10.1042/BST20120062>
- Young, D.B. and Cole, S.T. (1993) Leprosy, tuberculosis, and the new genetics. *J. Bacteriol.* **175**, 1–6 <https://doi.org/10.1128/jb.175.1.1-6.1993>
- Bloom, B.R. and Murray, C.J.L. (1992) Tuberculosis: commentary on a reemerging killer. *Science* **257**, 1055–1064 <https://doi.org/10.1126/science.257.5073.1055>
- Stinear, T.P., Seemann, T., Harrison, P.F., Jenkin, G.A., Davies, J.K., Johnson, P.D.R. et al. (2008) Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*. *Genome Res.* **18**, 729–741 <https://doi.org/10.1101/gr.075069.107>
- Demangel, C., Stinear, T.P. and Cole, S.T. (2009) Buruli ulcer: reductive evolution enhances pathogenicity of *Mycobacterium ulcerans*. *Nat. Rev. Microbiol.* **7**, 50–60 <https://doi.org/10.1038/nrmicro2077>
- Weerdenburg, E.M., Abdallah, A.M., Rangkuti, F., Abdel Ghany, M., Otto, T.D., Adroub, S.A. et al. (2015) Genome-wide transposon mutagenesis indicates that *Mycobacterium marinum* customizes its virulence mechanisms for survival and replication in different hosts. *Infect. Immun.* **83**, 1778–1788 <https://doi.org/10.1128/IAI.03050-14>
- Parvez, M., Qhanya, L.B., Mthakathi, N.T., Kgosiemang, I.K.R., Bamal, H.D., Pagadala, N.S. et al. (2016) Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: special focus on mycobacterial P450s. *Sci. Rep.* **6**, 33099 <https://doi.org/10.1038/srep33099>
- Frank, D.J., Zhao, Y., Wong, S.H., Basudhar, D., De Voss, J.J. and Ortiz de Montellano, P.R. (2016) Cholesterol analogs with degradation-resistant alkyl side chains are effective *Mycobacterium tuberculosis* growth inhibitors. *J. Biol. Chem.* **291**, 7325–7333 <https://doi.org/10.1074/jbc.M115.708172>
- Ouellet, H., Chow, E.D., Guan, S., Cox, J.S., Burlingame, A.L. and Ortiz de Montellano, P.R. (2013) Genetic and mass spectrometric tools for elucidating the physiological function(s) of cytochrome P450 enzymes from *Mycobacterium tuberculosis*. *Methods Mol. Biol.* **987**, 79–94 https://doi.org/10.1007/978-1-62703-321-3_7
- Barry, III, C.E., Lee, R.E., Mdluli, K., Sampson, A.E., Schroeder, B.G., Slayden, R.A. et al. (1998) Mycolic acids: structure, biosynthesis and physiological functions. *Prog. Lipid Res.* **37**, 143–179 [https://doi.org/10.1016/S0163-7827\(98\)00008-3](https://doi.org/10.1016/S0163-7827(98)00008-3)
- Johnston, J.B., Kells, P.M., Podust, L.M. and Ortiz de Montellano, P.R. (2009) Biochemical and structural characterization of CYP124: a methyl-branched lipid ω -hydroxylase from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. U.S.A.* **106**, 20687–20692 <https://doi.org/10.1073/pnas.0907398106>
- McLean, K.J. and Munro, A.W. (2008) Structural biology and biochemistry of cytochrome P450 systems in *Mycobacterium tuberculosis*. *Drug Metab. Rev.* **40**, 427–446 <https://doi.org/10.1080/03602530802186389>
- Forrellad, M.A., Klepp, L.I., Gioffre, A., Sabio y Garcia, J., Morbidoni, H.R., de la Paz Santangelo, M. et al. (2013) Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence* **4**, 3–66 <https://doi.org/10.4161/viru.22329>
- Jackson, C.J., Lamb, D.C., Marczylo, T.H., Parker, J.E., Manning, N.L., Kelly, D.E. et al. (2003) Conservation and cloning of CYP51: a sterol 14 α -demethylase from *Mycobacterium smegmatis*. *Biochem. Biophys. Res. Commun.* **301**, 558–563 [https://doi.org/10.1016/S0006-291X\(02\)03078-4](https://doi.org/10.1016/S0006-291X(02)03078-4)
- Ouellet, H., Johnston, J.B. and Ortiz de Montellano, P.R. (2010) The *Mycobacterium tuberculosis* cytochrome P450 system. *Arch. Biochem. Biophys.* **493**, 82–95 <https://doi.org/10.1016/j.abb.2009.07.011>
- Poupin, P., Ducrocq, V., Hallier-Soulier, S. and Truffaut, N. (1999) Cloning and characterization of the genes encoding a cytochrome P450 (PipA) involved in piperidine and pyrrolidine utilization and its regulatory protein (PipR) in *Mycobacterium smegmatis* mc2155. *J. Bacteriol.* **181**, 3419–3426 PMID:10348853

- 27 Trigui, M., Pulvin, S., Truffaut, N., Thomas, D. and Poupin, P. (2004) Molecular cloning, nucleotide sequencing and expression of genes encoding a cytochrome P450 system involved in secondary amine utilization in *Mycobacterium* sp. strain RP1. *Res. Microbiol.* **155**, 1–9 <https://doi.org/10.1016/j.resmic.2003.09.008>
- 28 Brezna, B., Kweon, O., Stingley, R.L., Freeman, J.P., Khan, A.A., Polek, B. et al. (2006) Molecular characterization of cytochrome P450 genes in the polycyclic aromatic hydrocarbon degrading *Mycobacterium vanbaalenii* PYR-1. *Appl. Microbiol. Biotechnol.* **71**, 522–532 <https://doi.org/10.1007/s00253-005-0190-8>
- 29 de Marco, A., Vigh, L., Diamant, S. and Goloubinoff, P. (2005) Native folding of aggregation-prone recombinant proteins in *Escherichia coli* by osmolytes, plasmid- or benzyl alcohol-overexpressed molecular chaperones. *Cell Stress Chaperones* **10**, 329–339 <https://doi.org/10.1379/CSC-139R.1>
- 30 Omura, T. and Sato, R. (1964) The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. *J. Biol. Chem.* **239**, 2370–2378 PMID:14209971
- 31 Pirote, B., Benoit, P., Dogné, J.-M., Masereel, B., Rolin, S. and de Leval, X. (2002) Therapeutic potential of thromboxane inhibitors in asthma. *Expert Opin. Investig. Drugs* **11**, 275–281 <https://doi.org/10.1517/13543784.11.2.275>
- 32 Bell, S.G., Dale, A., Rees, N.H. and Wong, L.-L. (2010) A cytochrome P450 class I electron transfer system from *Novosphingobium aromaticivorans*. *Appl. Microbiol. Biotechnol.* **86**, 163–175 <https://doi.org/10.1007/s00253-009-2234-y>
- 33 Bell, S.G., Yang, W., Yorke, J.A., Zhou, W., Wang, H., Harmer, J. et al. (2012) Structure and function of CYP108D1 from *Novosphingobium aromaticivorans* DSM12444: an aromatic hydrocarbon-binding P450 enzyme. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **68**, 277–291 <https://doi.org/10.1107/S090744491200145X>
- 34 McPhillips, T.M., McPhillips, S.E., Chiu, H.J., Cohen, A.E., Deacon, A.M., Ellis, P.J. et al. (2002) *Blu-ice* and the *Distributed Control System*: software for data acquisition and instrument control at macromolecular crystallography beamlines. *J. Synchrotron Radiat.* **9**, 401–406 <https://doi.org/10.1107/S0909049502015170>
- 35 Kabsch, W. (2010) *XDS*. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **66**, 125–132 <https://doi.org/10.1107/S0907444909047337>
- 36 Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **66**, 133–144 <https://doi.org/10.1107/S0907444909047374>
- 37 Panjikar, S., Parthasarathy, V., Lamzin, V.S., Weiss, M.S. and Tucker, P.A. (2005) *Auto-rickshaw*: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **61**, 449–457 <https://doi.org/10.1107/S0907444905001307>
- 38 Panjikar, S., Parthasarathy, V., Lamzin, V.S., Weiss, M.S. and Tucker, P.A. (2009) On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **65**, 1089–1097 <https://doi.org/10.1107/S09074449090029643>
- 39 Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R. et al. (2011) Overview of the *CCP4* suite and current developments. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **67**, 235–242 <https://doi.org/10.1107/S0907444910045749>
- 40 Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W. et al. (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **54**, 905–921 <https://doi.org/10.1107/S0907444998003254>
- 41 Brünger, A.T. (2007) Version 1.2 of the crystallography and NMR system. *Nat. Protoc.* **2**, 2728–2733 <https://doi.org/10.1038/nprot.2007.406>
- 42 Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A. et al. (2011) *REFMAC5* for the refinement of macromolecular crystal structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **67**, 355–367 <https://doi.org/10.1107/S0907444911001314>
- 43 Cowtan, K. (2000) General quadratic functions in real and reciprocal space and their application to likelihood phasing. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **56**, 1612–1621 <https://doi.org/10.1107/S0907444900013263>
- 44 Sheldrick, G.M. (2002) Macromolecular phasing with SHELXE. *Z. Kristallogr.* **217**, 644–650 <https://doi.org/10.1524/zkri.217.12.644.20662>
- 45 Terwilliger, T.C. (2000) Maximum-likelihood density modification. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **56**, 965–972 <https://doi.org/10.1107/S0907444900005072>
- 46 Cowtan, K. (2006) The *Buccaneer* software for automated model building. 1. Tracing protein chains. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **62**, 1002–1011 <https://doi.org/10.1107/S0907444906022116>
- 47 Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N. et al. (2010) *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **66**, 213–221 <https://doi.org/10.1107/S0907444909052925>
- 48 Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of *Coot*. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **66**, 486–501 <https://doi.org/10.1107/S0907444910007493>
- 49 Yoshida, M., Nakanaga, K., Ogura, Y., Toyoda, A., Ooka, T., Kazumi, Y. et al. (2016) Complete genome sequence of *Mycobacterium ulcerans* subsp. *shinshuense*. *Genome Announcements* **4**, e01050-16 <https://doi.org/10.1128/genomeA.01050-16>
- 50 Riojas, M.A., McGough, K. and Hazbon, M.H. *NGS-Based Phylogenomic Analysis Supports Reclassification of All Species Within the Mycobacterium tuberculosis Complex as Mycobacterium Tuberculosis*. Bei Resources. <https://www.beiresources.org/Portals/2/PDFS/NGS-Based%20Phylogenomic%20Analysis%20of%20MTB%20Complex.pdf>
- 51 Maya-Hoyos, M., Leguizamón, J., Mariño-Ramírez, L. and Soto, C.Y. (2015) Sliding motility, biofilm formation, and glycopeptidolipid production in *Mycobacterium colombiense* strains. *BioMed Res. Int.* **2015**, 1–11 <https://doi.org/10.1155/2015/419549>
- 52 Vasilevskaya, A.V., Yantsevich, A.V., Sergeev, G.V., Lemish, A.P., Usanov, S.A. and Gilep, A.A. (2017) Identification of *Mycobacterium tuberculosis* enzyme involved in vitamin D and 7-dehydrocholesterol metabolism. *J. Steroid Biochem. Mol. Biol.* **169**, 202–209 <https://doi.org/10.1016/j.jsbmb.2016.05.021>
- 53 Frank, D.J., Madrona, Y. and Ortizde Montellano, P. R. (2014) Cholesterol ester oxidation by mycobacterial cytochrome P450. *J. Biol. Chem.* **289**, 30417–30425 <https://doi.org/10.1074/jbc.M114.602771>
- 54 Capyk, J.K., Kalscheuer, R., Stewart, G.R., Liu, J., Kwon, H., Zhao, R. et al. (2009) Mycobacterial cytochrome p450 125 (*cyp125*) catalyzes the terminal hydroxylation of c27 steroids. *J. Biol. Chem.* **284**, 35534–35542 <https://doi.org/10.1074/jbc.M109.072132>
- 55 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**, 16122–16130 PMID:4066706

- 56 Banu, S., Honoré, N., Saint-Joanis, B., Philpott, D., Prévost, M.-C. and Cole, S.T. (2002) Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? *Mol. Microbiol.* **44**, 9–19 <https://doi.org/10.1046/j.1365-2958.2002.02813.x>
- 57 Espitia, C., Lacleste, J.P., Mondragón-Palomino, M., Amador, A., Campuzano, J., Martens, A. et al. (1999) The PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis*: a new family of fibronectin-binding proteins? *Microbiology* **145**(Pt 12), 3487–3495 <https://doi.org/10.1099/00221287-145-12-3487>
- 58 Omura, T. and Sato, R. (1964) The carbon monoxide-binding pigment of liver microsomes. II. Solubilization, purification, and properties. *J. Biol. Chem.* **239**, 2379–2385 PMID:14209972
- 59 Driscoll, M.D., McLean, K.J., Levy, C., Mast, N., Pikuleva, I.A., Lafite, P. et al. (2010) Structural and biochemical characterization of *Mycobacterium tuberculosis* CYP142: evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen. *J. Biol. Chem.* **285**, 38270–38282 <https://doi.org/10.1074/jbc.M110.164293>
- 60 Frank, D.J., Waddling, C.A., La, M. and Ortiz de Montellano, P.R. (2015) Cytochrome P450 125A4, the third cholesterol C-26 hydroxylase from *Mycobacterium smegmatis*. *Biochemistry* **54**, 6909–6916 <https://doi.org/10.1021/acs.biochem.5b01029>
- 61 Ouellet, H., Podust, L.M. and Ortiz de Montellano, P.R. (2008) *Mycobacterium tuberculosis* CYP130: crystal structure, biophysical characterization, and interactions with antifungal azole drugs. *J. Biol. Chem.* **283**, 5069–5080 <https://doi.org/10.1074/jbc.M708734200>
- 62 McLean, K.J., Dunford, A.J., Neeli, R., Driscoll, M.D. and Munro, A.W. (2007) Structure, function and drug targeting in *Mycobacterium tuberculosis* cytochrome P450 systems. *Arch. Biochem. Biophys.* **464**, 228–240 <https://doi.org/10.1016/j.abb.2007.03.026>
- 63 Mueller, N.J., Stueckler, C., Hauer, B., Baudendistel, N., Housden, H., Bruce, N.C. et al. (2010) The substrate spectra of pentaerythritol tetranitrate reductase, morphinone reductase, *N*-ethylmaleimide reductase and estrogen-binding protein in the asymmetric bioreduction of activated alkenes. *Adv. Synth. Catal.* **352**, 387–394 <https://doi.org/10.1002/adsc.200900832>
- 64 Grau, M.M., van der Toorn, J.C., Otten, L.G., Macheroux, P., Taglieber, A., Zilly, F.E. et al. (2009) Photoenzymatic reduction of C=C double bonds. *Adv. Synth. Catal.* **351**, 3279–3286 <https://doi.org/10.1002/adsc.200900560>
- 65 Hall, E.A., Sarkar, M.R. and Bell, S.G. (2017) The selective oxidation of substituted aromatic hydrocarbons and the observation of uncoupling via redox cycling during naphthalene oxidation by the CYP101B1 system. *Catal. Sci. Technol.* **7**, 1537–1548 <https://doi.org/10.1039/C7CY00088J>
- 66 Yang, W., Bell, S.G., Wang, H., Zhou, W., Hoskins, N., Dale, A. et al. (2010) Molecular characterization of a class I P450 electron transfer system from *Novosphingobium aromaticivorans* DSM12444. *J. Biol. Chem.* **285**, 27372–27384 <https://doi.org/10.1074/jbc.M110.118349>
- 67 Poulos, T.L., Finzel, B.C. and Howard, A.J. (1986) Crystal structure of substrate-free *Pseudomonas putida* cytochrome P-450. *Biochemistry* **25**, 5314–5322 <https://doi.org/10.1021/bi00366a049>
- 68 Cojocaru, V., Winn, P.J. and Wade, R.C. (2007) The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta, Gen. Subj.* **1770**, 390–401 <https://doi.org/10.1016/j.bbagen.2006.07.005>
- 69 Poulos, T.L., Finzel, B.C. and Howard, A.J. (1987) High-resolution crystal structure of cytochrome P450cam. *J. Mol. Biol.* **195**, 687–700 [https://doi.org/10.1016/0022-2836\(87\)90190-2](https://doi.org/10.1016/0022-2836(87)90190-2)
- 70 Poulos, T.L. (2014) Heme enzyme structure and function. *Chem. Rev.* **114**, 3919–3962 <https://doi.org/10.1021/cr400415k>

Chapter 6:

A comparison of the steroid binding cytochrome P450s from *Mycobacterium marinum* and *Mycobacterium tuberculosis*


Prepared in the style of *Biochimica et Biophysica Acta* General Subjects in preparation for submission.

Citation:

Child, S. A.; Bruning, J. B.; Bell, S. G., A comparison of the steroid binding cytochrome P450s from *Mycobacterium marinum* and *Mycobacterium tuberculosis*, manuscript in preparation.

Statement of Authorship

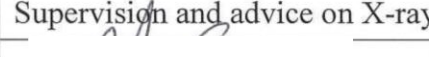
Principal Author


Name	Stella Child		
Contribution to the Paper	Enzyme purification, substrate characterisation, manuscript preparation and review.		
Overall percentage (%)	80		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	23/5/18

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	John Bruning		
Contribution	Supervision and advice on X-ray crystallography		
Signature		Date	23/5/2018

Name of Co-Author	Stephen Bell		
Contribution	Cloning of gene sequences, experimental design, supervision, manuscript preparation.		
Signature		Date	23/5/2018

**A comparison of the steroid binding cytochrome P450s from *Mycobacterium marinum*
and *Mycobacterium tuberculosis***

Stella A. Child¹, John B. Bruning², and Stephen G. Bell^{1*},

¹Department of Chemistry, University of Adelaide, SA 5005, Australia

²School of Biological Sciences, University of Adelaide, SA 5005, Australia

* To whom correspondence should be addressed.

Stephen G. Bell (stephen.bell@adelaide.edu.au)

Highlights

- The CYP125A, CYP142A and CYP125A enzymes of *M. marinum* and *M. tuberculosis* were produced and their substrate range compared
- Significant differences were observed with CYP125 subfamily members in each species
- CYP142A3 closely mimics the substrate range of CYP142A1 as did the equivalent CYP124A enzymes
- CYP124A1 from *M. marinum* was structurally characterised and found to be similar to the *M. tuberculosis* enzyme
- Changes in the binding affinity of inhibitors for these enzymes were observed between species

Abstract

Background

The steroid binding cytochrome P450 enzymes of *Mycobacterium tuberculosis* are essential for organism survival through metabolism of cholesterol and its derivatives. The counterparts to these enzymes from *Mycobacterium marinum* were studied to determine the degree of functional conservation between them.

Methods

Spectral analysis of the substrate and inhibitor binding for the four *M. marinum* enzymes CYP125A6, CYP125A7, CYP142A3 and CYP124A1 were performed and compared to those of *M. tuberculosis*. *M. marinum* CYP124 was characterised by X-ray crystallography.

Results

CYP125A7 of *M. marinum* was more similar sequentially to CYP125A1 from *M. tuberculosis* than CYP125A6, but showed biochemical differences in resting haem spin state and azole binding mode and affinity. CYP142A3 demonstrated similar affinity for the substrates of CYP142A1. The two CYP124A1 enzymes displayed strong conservation of active site residues, except near where the carboxylate terminal of the phytanic acid binds.

Conclusions

The steroid binding CYP enzymes in *M. marinum* demonstrate some key differences in substrate and inhibitor binding. The *M. marinum* CYP142 had 10-fold higher affinities for azole inhibitors, whereas CYP125A7 did not show a Type II inhibitor-like shift with any azoles.

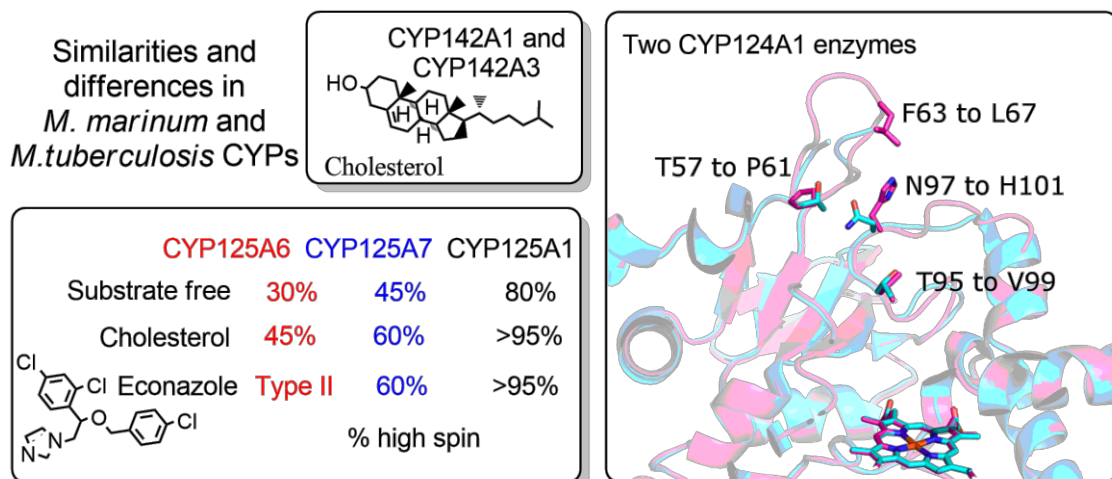
General Significance

These enzymes in *M. tuberculosis* have been identified as candidates for inhibition and the data here demonstrates that difference in inhibitor design may be required to target CYP family members from different pathogenic *Mycobacterium* species.

Key words

enzymology, electron transfer, cholesterol metabolism, cytochrome P450, *Mycobacterium*

Graphical abstract



Abbreviations

CYP or P450, cytochrome P450 enzyme; DTT, dithiothreitol; H β CD, hydroxyl- β -cyclodextrin; IPTG, isopropyl β -D-1-thiogalactopyranoside; LB, lysogeny broth (also known as Luria or Lennox Broth), NAD(P)H reduced nicotinamide adenine dinucleotide (phosphate); MDR, multi drug-resistant; MTBC, *Mycobacterium tuberculosis* complex; RT, retention time; SOC, Super Optimal broth with Catabolite repression; PCR, polymerase chain reaction; PIM, phenylimidazole; XDR, extensively drug-resistant.

1. Introduction

Mycobacterium tuberculosis is the pathogen responsible for human tuberculosis [1]. Tuberculosis continues to be a major human health concern, resulting in the deaths of 1.6 million people in 2016, and over six million new infections [2]. The startling rise in multi-drug (MDR) and extensively-drug resistant (XDR) strains of this species makes tuberculosis a growing, rather than diminishing threat to human health, despite a yearly decrease in incidence of 2% [2]. There is both a lack of effective new drugs on the market and difficulties with existing treatment regimes (the standard treatment continues for six months), all of which increases the risk of substantial resistance developing, and emphasises the necessity of further research into this bacterial species. Many other *Mycobacterium* species are also pathogenic; for example *Mycobacterium leprae* is the causal agent of leprosy. *Mycobacterium ulcerans* infection results in the skin disease known as the Buruli or Daintree ulcer, which is common in Africa and increasingly so in northern and eastern Australia, and is the third most common human *Mycobacterium* pathogen after tuberculosis and leprosy [3, 4]. *Mycobacterium marinum* is a primarily a pathogen in frogs and fish (although it can opportunistically infect humans, causing aquarium granuloma) [5]. It bears high sequence similarity to the human pathogens (85% and 97% to *M. tuberculosis* and *M. ulcerans*, respectively). This species is often used as a model organism for *M. tuberculosis* as the disease progression and features (such as granuloma formation) are well conserved [6]. *M. marinum* has been hypothesised to resemble a common ancestor of the more pathogenic *Mycobacterium* species [5]. *Mycobacterium smegmatis* is also sometimes used as a model species. However, this species is a non-pathogenic soil-dwelling bacteria and thus has limitations in usefulness [7].

The sequencing of *Mycobacterium* genomes has revealed, among other things, that the number of cytochrome P450 genes in many *Mycobacteria* is unusually high for bacterial

species [8, 9]. Cytochrome P450s (CYPs) are haem-thiolate monooxygenases, which perform the catalytic insertion of a single oxygen atom across a C-H bond. CYPs are named based on shared sequence similarity, where >40% similarity indicates a shared family group (denoted by number), greater than >55% indicates a shared subfamily (denoted by letters) [10]. Bacterial family groups are numbered from 100 to 300 and then from 1000 and above. For instance, CYP101A1 is the first member of the A subfamily in the 101 family of bacterial CYPs. In *M. tuberculosis*, there are 20 unique CYP sequences, which make up almost 0.5% of the genome [1]. In contrast, *Escherichia coli* has none. *M. marinum*, which has a larger genome that is less affected by the reductive evolution that characterises that of *M. tuberculosis*, has 47 CYP sequences (comprising almost 1% of the genome) [5]. Among the 20 CYPs of *M. tuberculosis*, several have been found to be critical for *in vitro* or *in vivo* function. These include CYP121A1, which performs a dipeptide cyclisation reaction that appears to be highly specific to species in the *M. tuberculosis* complex (MTBC), and the CYP125A1 and CYP142A1 pair, which both perform cholesterol hydroxylation at the C27 position [11, 12]. Dual knockout of the latter two CYPs produces cholesterol build-up and cytotoxicity [13]. An additional non-essential enzyme, CYP124A1, was first shown to bind branched fatty acids [14], but additional activity with cholesterol analogues has also been demonstrated by this enzyme [15]. The physiological relevance of this catalytic ability has not been demonstrated, however, this finding does serve to highlight the similarities in the reactions catalysed by the three; CYP125A1, CYP142A1 and CYP124A1 which all perform hydroxylations at the ω -terminus of branched aliphatic hydrocarbons.

Cholesterol breakdown in *M. tuberculosis* is necessary for the optimal growth of *M. tuberculosis* both *in vitro* and in animal models [16-20]. A large portion of the *M. tuberculosis* genome is associated with cholesterol breakdown, from the transport of cholesterol into the cell, transcriptional regulators of this process, and the catabolic enzymes

themselves [1]. Many of the genes are essential for organism function [16, 21]. For example, the release of either acetyl-CoA or propionyl-CoA accompanies many of the steps in the cholesterol breakdown process [16]. One of the cellular fates of propionyl-CoA is incorporation into methyl-branched polyketide lipids such as phthiocerol dimycoserates (PDIM) that are present in the mycomembrane layer of the *M. tuberculosis* cell wall [16]. CYP125A1 function has been connected to the density of the cell wall through PDIM via cholesterol breakdown [22].

These steroid binding CYP enzymes are conserved across diverse *Mycobacterium* species. Small changes in sequence can significantly alter substrate recognition in CYP enzymes [10], which suggests that investigation into the degree of conserved function between enzymes of the same family species is appropriate. Work on *M. smegmatis* identified a second CYP125 family member, making at least three cholesterol metabolising enzymes in that species: CYP125A3, CYP125A4 and CYP142A2 [23]. *M. marinum* also contains two CYP125 isoforms, CYP125A6 and CYP125A7, as well as a CYP142 and CYP124 family member. However, the *M. marinum* CYP125A7 enzyme is closer in sequence to the *M. tuberculosis* CYP125A6. Its system of potential steroid binding enzymes offers insight into the recent evolution of the *M. tuberculosis* CYP system. Here, the steroid and fatty acid binding CYP enzymes of *M. marinum* (CYP125, CYP142 and CYP124) were purified and characterised and compared to the equivalent systems in *M. tuberculosis*. The X-ray crystal structure of CYP124A1 from *M. marinum* was determined.

2. Experimental

General

Except where otherwise noted, all organic substrates, derivatisation agents and other general reagents were purchased from Sigma-Aldrich, Alfa-Aesar, VWR International or Tokyo Chemical Industry. Antibiotics, detergents, DTT and IPTG were purchased from Astral Scientific. The media for cell growth and maintenance (LB, 2xYT, SOC and trace elements) were prepared as reported previously [24]. Appropriate antibiotics were added to working concentrations; 100 $\mu\text{g mL}^{-1}$ and 30 $\mu\text{g mL}^{-1}$ for ampicillin and kanamycin, respectively. UV-Visible spectra were acquired using a Varian Cary 5000 at 30 ± 0.5 °C.

Recombinant protein expression and purification

The *M. marinum* CYP124A1, CYP125A6, CYP125A7 and CYP142A3 genes were amplified by PCR using oligonucleotide primers (see Supplementary Information for details). PCR was performed by the method reported for CYP268A2 [25]. The genes were cloned into the pET26 vector using the appropriate restriction enzymes and the correct insertion checked by Sanger sequencing performed by Australian Genome Research Facility (AGRF). The *M. tuberculosis* CYP124A1, CYP125A1 and CYP142A1 were each obtained as a gblock (IDT) with NdeI and HindIII restriction enzyme sites incorporated at the 5' and 3' termini, respectively. *M. tuberculosis* CYP124A1 (MtbCYP124A1), *M. marinum* CYP124A1 (MmarCYP124A1), CYP125A1, CYP125A6, CYP125A7, CYP142A1 and CYP142A3 were each cloned in pET26 vectors and transformed into *E. coli* BL21 (DE3) cells. The growth and purification of each was performed using the method reported previously for CYP268A2 [25]. After two ion-exchange steps, the purity was analysed SDS-PAGE (Fig. S1). The protein was stored in glycerol (50% v/v at -20°C), which was removed before use by a 5 mL PD-10 column (GE Healthcare) in 50 mM Tris pH 7.4. CO assays were performed according to the method developed by Omura and Sato [26]. Extinction coefficients were determined

using $\epsilon_{450} = 91 \text{ mM}^{-1} \text{ cm}^{-1}$ for the reduced CO bound form. Concentration was determined for CYP142A3 concentration using $\epsilon_{419} = 108 \pm 11 \text{ mM}^{-1} \text{ cm}^{-1}$ and for MmarCYP124A1 using $\epsilon_{419} = 139 \pm 10 \text{ mM}^{-1} \text{ cm}^{-1}$. CYP125A6 and CYP125A7 concentrations were estimated using the reported extinction coefficients for CYP125A3 and A4 [23, 27].

Spin state shifts and substrate binding titrations

To determine the extent of substrate binding for a particular enzyme, the purified protein was diluted to $\sim 1 \text{ }\mu\text{M}$ in 50 mM Tris pH 7.4, to a volume of 500 μL , followed by the addition of various substrates. Substrate stocks were made up to 10 mM in DMSO, with the exception of steroids, for which stock solutions were at a concentration of 1-2 mM in 50 mM Tris pH 7.4, with 10% hydroxyl- β -cyclodextrin (H β CD). The absorbance between 600 nm and 250 nm was recorded using a UV-Visible spectrophotometer until no further spectral change was observed. The high spin percentage ($\pm 5\%$) was estimated by comparison to a set of spectra, generated by the sum of substrate-free ($>95\%$ low spin) and substrate-bound (camphor, $\geq 95\%$ high spin) CYP101A1 to the appropriate percentages as described previously [25]. For substrate binding titrations, the P450 was diluted to $\sim 1 \text{ }\mu\text{M}$ in 50 mM Tris pH 7.4 to a volume of 2.5 mL. 1-3 μL of substrate was added via a Hamilton syringe from either 0.1, 1 and 10 mM (in DMSO or 50 mM Tris pH 7.4, 10% H β CD) stock solution, starting from the lowest concentration. The peak-to-trough difference in absorbance, between 600 nm and 250 nm, was recorded until additional aliquots caused no further spectral change in the Soret band. In instances where the substrate exhibited tight binding ($K_d < 5 \text{ }\mu\text{M}$, less than five times the concentration of the enzyme), the data were fitted to the tight-binding quadratic equation:

$$\Delta A = \Delta A_{max} \times \frac{[E] + [S] + K_d - \sqrt{([E] + [S] + K_d)^2 - 4[E][S]}}{2[E]}$$

where K_d denotes the binding constant, $[S]$ the substrate concentration, ΔA the peak-to-trough ratio, ΔA_{max} the maximum peak-to-trough absorbance and $[E]$ is the enzyme

concentration [28]. When substrates did not exhibit tight binding, the dissociation constant was obtained by fitting the difference in absorbance against the substrate concentration to the hyperbolic function:

$$\Delta A = \frac{\Delta A_{max} \times [S]}{K_d + [S]}$$

where K_d denotes the binding constant, $[S]$ the substrate concentration, ΔA the peak-to-trough ratio, and ΔA_{max} the maximum peak-to-trough absorbance.

Crystallography, data collection, data processing and structural determination

M. marinum CYP124A1 was further purified by size exclusion (Enrich SEC Column, 650 x 10 x 300 mm, 1 mL min⁻¹) before concentration to ~30 mg mL⁻¹ in 50 mM Tris pH 7.4. Crystallisation was attempted in the presence and absence of 1 mM phytanic acid. Commercially available screening conditions (Hampton Research) were used for initial screening in 96 well sitting drop trays, using 1 μ L of both protein and reservoir solution. Crystal conditions were then refined using the hanging drop vapour diffusion method, again using both 1 μ L of protein and reservoir solution with a 500 μ L reservoir. Diffraction quality crystals of CYP124A1 were obtained after 2 weeks at 16 °C from both substrate and substrate-free trays. The selected substrate-free crystals were grown in 0.26 M ammonium sulphate, 20% w/v polyethylene glycol 3,350, pH 6.1. The crystals were harvested using a Micromount (MiTeGen) and cryo-protected by immersion in Parabar 10312 (Paratone-N, Hampton Research) before flash cooling in liquid N₂. Data was collected by X-ray diffraction at the Australian Synchrotron MX1 beamline (360 exposures using 1° oscillations at a wavelength of 0.9357 Å) [29]. The data were processed into the space group C2 using XDS, followed by truncation and addition of R_{free} flags using Aimless [30], as part of the CCP4 package [31]. Molecular replacement phasing was carried out by Phaser [32], also part of CCP4. The model was rebuilt using Coot [33] based on initial electron density maps and

refined using phenix.refine over several iterations. Composite omit maps were generated using the Composite Omit Maps program in Phenix. Data collection and refinement statistics are presented in Table 1 and the structure was deposited to the PDB (6CVC). Crystals from substrate bound conditions were solved but contained no substrate in the active site. Similar experiments were carried out with CYP125A6, A7 and CYP142A3 but as yet no suitable diffraction-quality crystals have been obtained.

Table 1: Crystal data collection and refinement statistics for CYP124A1 from *M. marinum* (PDB: 6CVC)

Data collection statistics^a	
Wavelength	0.95370
Unit cell	a = 98.78 b = 72.36 c = 65.60 $\alpha = 90 \beta = 109.72 \gamma = 90$
Space group	C2
Molecules in asymm. unit	1
Resolution	2.20 – 46.49 (2.20 – 2.27) ^b
Unique reflections	21918 (2122)
Completeness	98.8 (97.1)
Redundancy	7.4(7.1)
(I)/[σ (I)]	7.7 (3.1)
R _{merge} (all I+ and I-)	0.298 (1.135)
R _{pim} (all I+ and I-)	0.118 (0.460)
CC(1/2)	0.989 (0.855)
R _{work}	20.43%
R _{free}	25.37%
% solvent	46.07
Residues modelled	422
RMS deviation from restraint values	
Bond lengths	0.002
Bond angles	0.46
Ramachandran analysis	
Most favoured	98.33%
Additionally allowed	1.67%

^a Data collected from one crystal.

^b Values in parenthesis are for highest resolution shell.

3. Results/Discussion

The CYP125A6 and A7 enzymes of *M. marinum*

The *M. marinum* species CYP125A6 shares a sequence identity of 75% with CYP125A1, while CYP125A7 shares 90%, making it the closest named relative to the *M. tuberculosis* enzyme (see Fig. 1).² Other bacteria contain CYP125 family members including *M. smegmatis*, *M. ulcerans*, and various *Streptomyces*, *Nocardia* and *Rhodococcus* strains. *M. smegmatis* encodes two CYP125 enzymes, CYP125A3 and A4, which share 78% and 71% identity to CYP125A1, respectively [23]. CYP125A3 is more closely related to CYP125A7 than CYP125A6 (78% versus 69%), and vice versa (CYP125A4 is closer to CYP125A6 than CYP125A7, 82% compared to 71%). *M. ulcerans* contains CYP125A7 (99% identity with CYP125A7 from *M. marinum*) and a pseudogene of CYP125A6 (truncated after 200 residues, 98% sequence identity to CYP125A6). The *Rhodococcus* sp. RHA1 CYP125 enzyme shares 69% identity with the *M. tuberculosis* and 79% with an analogue from *Nocardia farcinia* [34]. Some species, such as *M. vanbaalenii* PYR-1, contain 3 or more CYP125 enzymes. The analysis of Parvez *et al* suggested that CYP125 is conserved in the MTBC but also in the *M. chelonae-abscessus* complex (MCAC), the *M. avium* complex (MAC), nontuberculous *Mycobacteria* (NTM), which includes *M. marinum*, and Saprophytes (SAP), such as *M. smegmatis* [35].

In order to investigate the CYP125A family in more detail we cloned the genes of, and produced and purified CYP125A6 and A7 from *M. marinum* and CYP125A1 from *M. tuberculosis* for comparison. The first obvious difference between CYP125A6 and CYP125A7 from *M. marinum* is that the resting state of both enzymes has a less high spin character than that observed for CYP125A1.

² According to the nomenclature system for CYP names, given sequence identity is above 80%, it seems this should more properly be CYP125A1. It is referred to here as CYP125A7 in accordance with Dr Nelson's Cytochrome P450 homepage (<http://drnelson.uthsc.edu/bacterial.P450s.2011.htm>). This also applies to CYP142A3 with respect to CYP142A1.

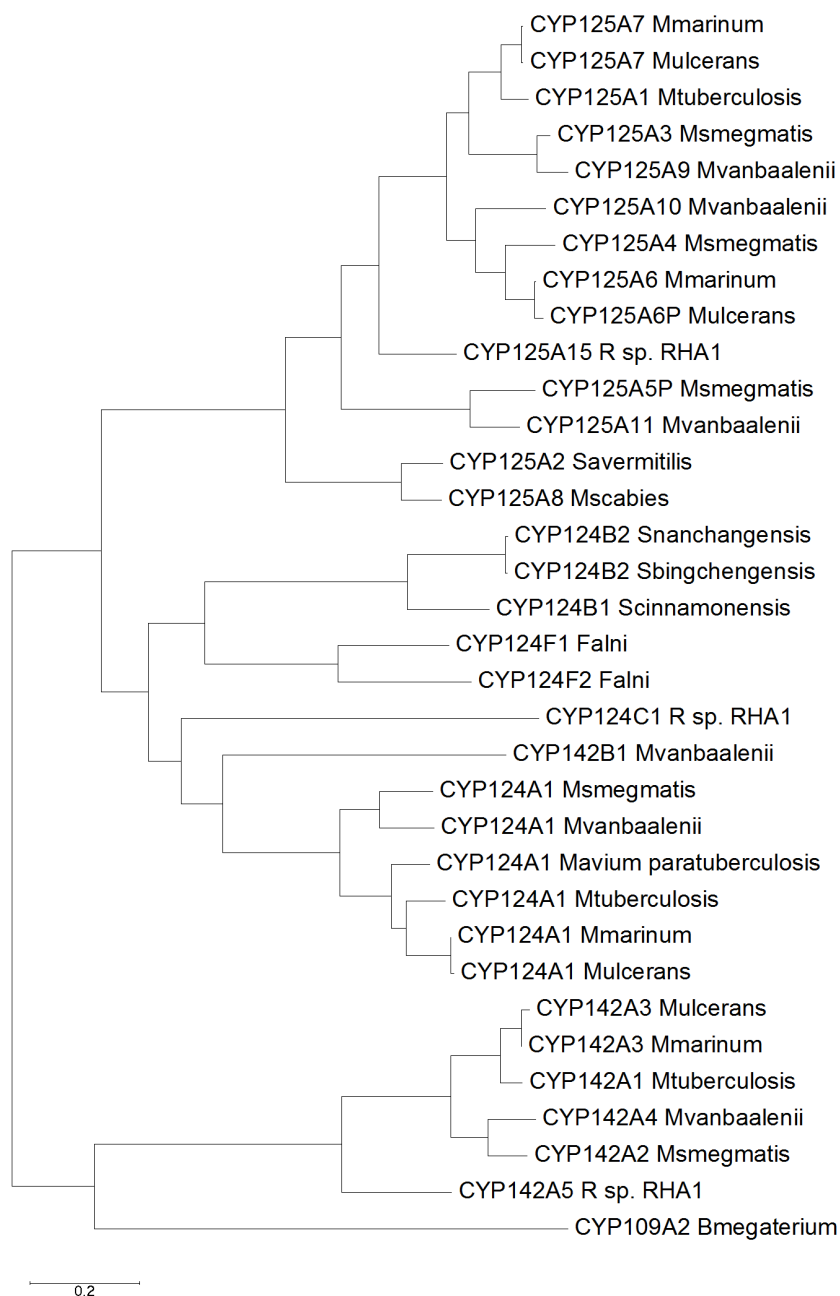


Figure 1: Phylogenetic tree showing selected members of the 125, 142 and 124 families from *Mycobacterium* and other species including *Streptomyces*, *Frankia*, and *Rhodococcus*. The *Mycobacterium* members of each family cluster together, with the exception of CYP142B1 from *M. vanbaalenii*. The closest relative of the *M. tuberculosis* enzyme in the all three families is from *M. marinum*, which is also closest to those from *M. ulcerans*. The closest structurally characterised homologue, CYP109A2 from *Bacillus megaterium*, is included.

In the resting state, CYP125A6 is ~30% high spin, whereas CYP125A7 is 45% high spin (Fig. 2). CYP125A1 was purified using the same method and showed ~80% high spin

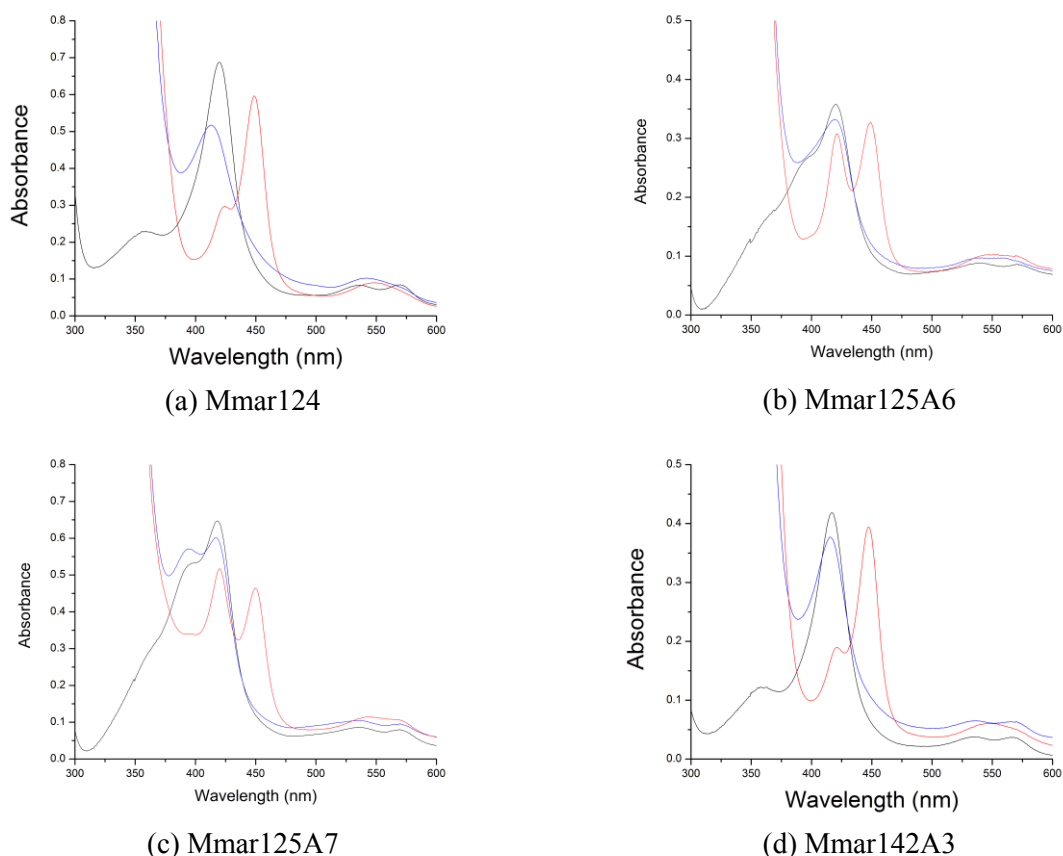


Figure 2: CO-binding spectra for the *M. marinum* enzymes CYP124A1, CYP125A6, CYP125A7 and CYP142A3. For each, the resting state (black), reduced enzyme (blue) and reduced CO bound (red) spectra are shown. Both CYP125 enzymes had split A_{450}/A_{420} absorbances in the CO-bound state.

character (Fig. S5) in agreement with what is reported by others for this enzyme and similar to that reported for the CYP125A3 member from *M. smegmatis* [23]. In addition, CYP125_{RHA} has been reported as >90% high spin in the resting state while CYP125A4 was 60% high spin [27, 34]. Both the CYP125A6 and CYP125A7 pair display more low spin character than any CYP125 enzymes thus far reported. Interestingly, the resting coordinate state of CYP125A1 was connected to the position of the V267 residue of the I-helix in the crystal structure [19]. Two alternate conformations of this valine were observed, corresponding to the presence or absence of the coordinating water ligand. However this residue is highly conserved in all these CYP125 species including those from *M. marinum* (as are nearby residues, and those of the proximal loop around the haem iron coordinating cysteine residue, Fig. S2).

All three enzymes showed a significant 420 nm peak upon reduction and CO binding. This is in line with what was reported previously with CYP125A1, due to facile protonation of the proximal cysteine thiolate [19, 36]. For all three enzymes, the Reverse Type I inhibitor LP10 was able to fully generate a low spin type spectra, as previously demonstrated for CYP125A1 (Fig. 3)[37]. As a substrate, cholesterol is known to shift CYP125A1 to the high spin state (80% to $\geq 95\%$). Likewise, for CYP125A6 and A7, cholesterol induced shifts (from 30% and 45% resting states) to 45% and 60% respectively (Table 2). 4-Cholesten-3-one binding generated slightly greater shifts, to 55% and 70% high spin, with the same enzymes. The binding affinity of CYP125A7 for 4-cholesten-3-one was on par with that of CYP125A1 ($K_d = 0.7 \pm 0.4$ compared to 1.2 ± 0.1 μM [19]).

The inhibition of the *M. tuberculosis* CYP125 enzyme by azole derivatives has been investigated previously, as the function of CYP125A1 is essential for *M. tuberculosis* growth in some conditions [19, 21]. The range of inhibitors which bound to CYP125A7 was significantly altered from that of CYP125A1. While CYP125A1 displayed inhibitory shifts with a range of azoles, including econazole, miconazole and both 4- and 2-phenylimidazole, CYP125A7 did not display a Type II shift with any of these. Econazole induced a high spin state of 60% ($K_d = 0.3 \pm 0.1$ μM). Surprisingly, given its lower sequence identity, CYP125A6 bound a range of these inhibitors and so was more similar to CYP125A1. Both 1- and 4-phenylimidazole (PIM), econazole and clotrimazole induced a Type II binding shift (although not 2-PIM). The binding of 2-PIM in a Type II manner to CYP125A1 is unusual given its structure (Fig. 3). More frequently both the 1- and 4-PIM generate Type II shifts but 2-PIM does not, as presumably the position of the nitrogen group on the ring is less favourable for haem coordination [38].

Table 2: Spin state shift and dissociation constant analysis of CYP125A1 from *M. tuberculosis* and CYP125A6 and CYP125A7 from *M. marinum*. Data from elsewhere is noted, otherwise analyses were performed as recorded in the experimental section. Type I shifts are listed by high spin percentage. Type II is listed as the nm shift, unless not reported. Reverse Type I is listed as RT1.

	<i>M.mar</i> CYP125A6		<i>M.mar</i> CYP125A7		<i>M.tb</i> CYP125A1		<i>M.smeg</i> CYP125A3	
	Spin state	K_d (μ M)	Spin state	K_d (μ M)	Spin state	K_d (μ M)	Spin state	K_d (μ M)
Resting state	30%		45%		>80%		>80%	
Cholesterol	45%	ND	60%	ND	$\geq 95\%$	0.11 ± 0.06^c		1.01 ± 0.5^b
4-cholesten-3-one	55%	ND	70%	0.7 ± 0.4	$\geq 95\%$	1.2 ± 0.1^c		2.3 ± 1.4^b
Econazole	422nm	4.7 ± 0.4	60%	0.3 ± 0.1	$\geq 95\%$	11.7 ± 0.7^a	Type II [†]	7.38 ± 0.71^b
Clotrimazole	423nm	1.5 ± 0.35	0%	-	421nm	5.3 ± 0.6^a	Type II	14.53 ± 1.58^b
Miconazole	0%	-	0%	-	421nm	4.6 ± 0.4^a	Type II	1.66 ± 0.21^b
Ketoconazole	-	-	0%	-	419 nm	27.1 ± 0.9^a	-	-
Fluconazole	0%	-	0%	-	418 nm	43.1 ± 3.8^a	-	-
4-phenylimidazole	423nm	47 ± 23	0%	-	424nm	216 ± 5^a	-	-
2-phenylimidazole	0%	-	0%	-	422nm	345 ± 4^a	-	-
1-phenylimidazole	422nm	34 ± 5	0%	-	-	-	-	-
LP-10	RT1	0.36 ± 0.15	RT1	0.01 ± 0.002	RT1	$9.86 \pm 1.56^{\ddagger}$	-	-

(ND) indicates dissociation constant was not determined due to insufficient alteration of the enzyme spin state shift by the substrate. [†] CYP125A4 from *M. smegmatis* also was reported to undergo a Type II shift with this substrate. Its resting state was reported as 60% high spin [27]. [‡] Ouellet *et al* give as $K_d = 1.68 \pm 0.08$ [37].

^a McLean *et al*, 2009 [19].

^b Garcia-Fernandez *et al*, 2013 [23].

^c as reported by Ouellet *et al*, 2011 [37] but similar results were obtained by Capyk *et al*, 2009 [12].

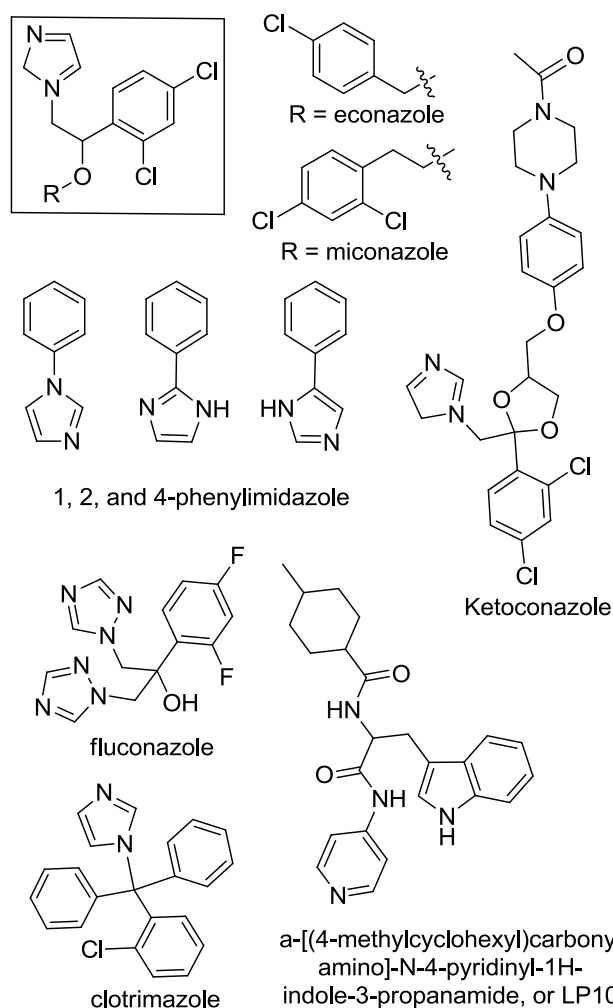


Figure 3: The structures of a variety of azoles which have been tested as inhibitors of the *M. tuberculosis* CYPs, as well as the Reverse Type I inhibitor LP10.

Crystal structures of CYP125A1 with androstendione and the inhibitor econazole bound have been solved previously[19]. When the residues of the active site of CYP125A1 were compared across these enzymes, CYP125A7 was found to have all these residues conserved except P213 (Table S2). These conserved residues include the D108-K214 salt bridge, which is absent in CYP125A3, and is hypothesised to govern substrate entry to the active site (see Fig. S8)[19, 23]. The conservation of the majority of these residues makes the altered interactions with econazole and other substrates and inhibitors suggestive of a conformational change in the structure of the *M. marinum* CYP125s. This could result from the change of the P213 residue to an alanine in CYP125A6 and A7, or alterations elsewhere

in the enzyme. The P213 residue is conserved in both *M. smegmatis* CYP125A3 and CYP125_{RHA}. *M. ulcerans* CYP125A7 shares the active site residues of *M. marinum* CYP125A7, and so also has an equivalent proline to alanine substitution, while CYP125A4 has a glycine at the corresponding position. The presence of this proline at the end of the G helix in CYP125A1 [19] disrupts the helix structure as the backbone of proline residues are not conducive to the α -helix secondary structure (Fig. S8). An alanine in its place would likely not have the same effect, and thus it is plausible that the G helix could be extended in CYP125A7 until the P206 residue (conserved in all CYP125A members) that forms the end of the small helix in the loop between the F and G helices. P213 is similarly replaced with an alanine in CYP125A6. The remaining active site residues of CYP125A6 are less conserved than in CYP125A7, with a leucine at the position of K214 in the salt bridge, among other changes (Table S2). The large number of active site changes makes binding differences that were observed more expected in CYP125A6.

CYP142A3 and CYP142A1 compared

CYP142 members are similarly prolific in *Mycobacteria*. *M. tuberculosis*, *M. smegmatis*, *M. marinum* and *M. ulcerans* all contain an enzyme of this family while *M. vanbaalenii* contains two, although the second of these, CYP142B1, clusters more closely with CYP124 family members in the phylogenetic tree (Fig. 1). *Rhodococcus* sp. RHA1 encodes CYP142A5. Species from the MCAC, interestingly, appear to lack CYP142 members, but routinely contain 3 to 4 CYP125 enzymes [35]. CYP142A3 from *M. marinum* has a sequence identity of 91% with CYP142A1 from *M. tuberculosis*. CYP142A3 is also present in *M. ulcerans*, with 98% similarity to the analogue in *M. marinum*. CYP142A2 from *M. smegmatis* shares 78% identity with both the *M. tuberculosis* and *M. marinum* CYP142 enzymes.

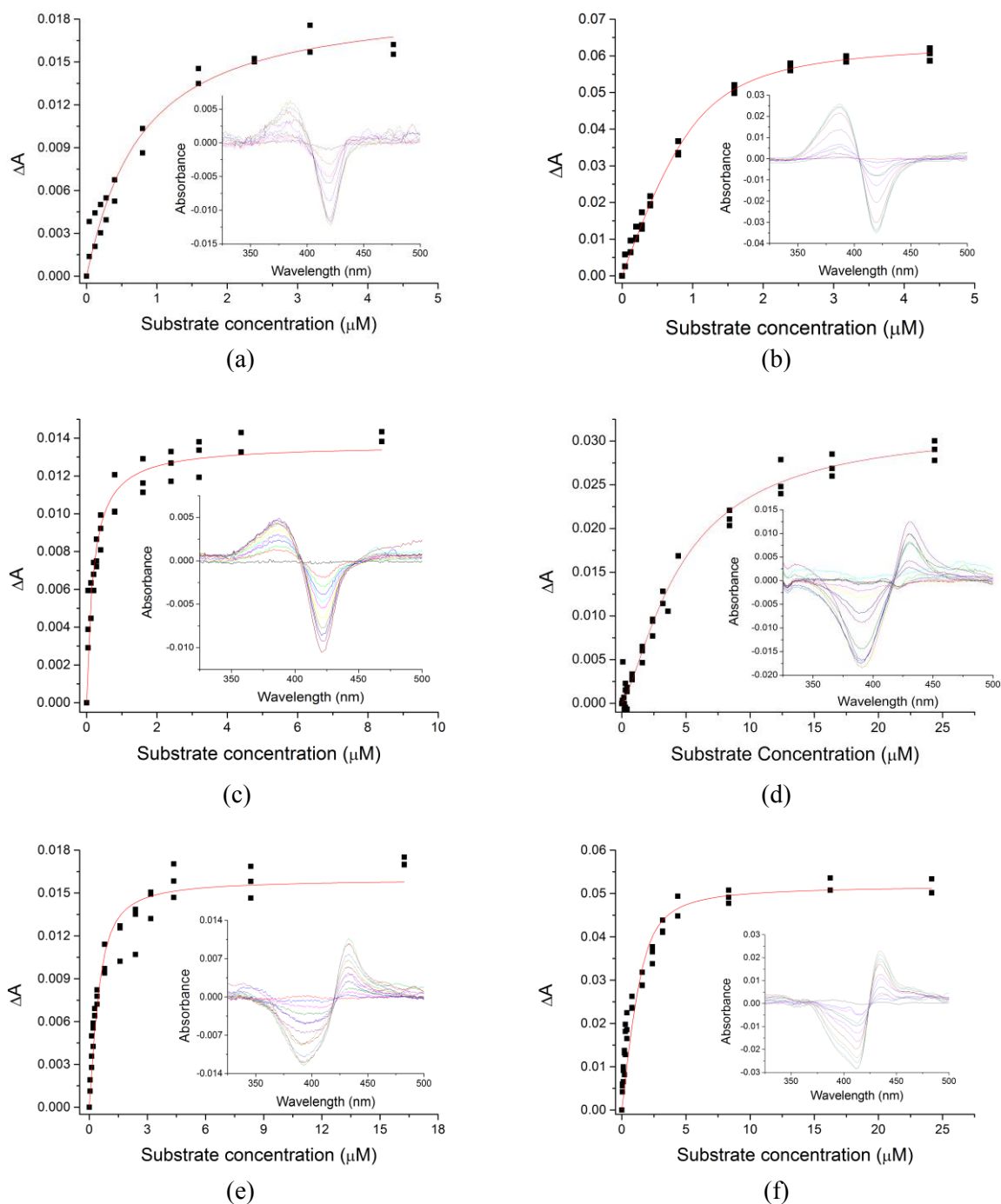


Figure 4: A selection of substrate and inhibitor dissociation constants for (a) CYP125A7 with 4-cholesten-3-one, (b) CYP142A3 with 4-cholesten-3-one, (c) CYP125A7 with econazole, (d) CYP125A6 with econazole, (e) MmarCYP124A1 with clotrimazole and (f) MtbCYP124A1 with clotrimazole. The peak to trough absorbance for (a), (b) and (c) was measured at 390 and 420 nm, while for (d), (e) and (f) it was 430 to 390 nm, 434 to 412 nm, and 433 to 393 nm respectively.

When cloned, produced and purified, the *M. marinum* CYP142A3 and *M. tuberculosis* CYP142A1 both demonstrated low spin resting states, and the Soret peak of both enzymes shifted almost completely to 450 nm upon binding of CO to the ferrous state (Fig. 3 and Fig. S5). The spin state shifts of *M. tuberculosis* CYP142A1 with known substrates matched literature values (Table 3)[11]. When the substrates cholesterol and 4-cholesten-3-one were tested with the *M. marinum* CYP142A3 enzyme, both induced smaller high spin shifts (55% compared to $\geq 95\%$ for CYP142A3 versus CYP142A1, see Table 3). However, the affinity for cholesterol was comparable, $K_d = 0.40 \pm 0.1 \mu\text{M}$ for CYP142A3 and $0.34 \pm 0.20 \mu\text{M}$ [11] for CYP142A1. No significant difference in the range of effective inhibitors for both enzymes was observed. Miconazole and clotrimazole induced Type II shifts in both enzymes, while ketoconazole did not for either. Both 1- and 4-PIM also gave Type II shifts while 2-PIM did not. A minor difference was observed for econazole with CYP142A3, where Type I binding was observed to approximately 20% high spin, but further additions induced a Type II shift. Affinities with the azole inhibitors were consistently higher for CYP142A3 than those reported for CYP142A1 (for example, for miconazole, $K_d = 0.39 \pm 0.1 \mu\text{M}$ compared to $4.0 \pm 0.5 \mu\text{M}$ [11]). A comparison of the active site residues of CYP142 enzymes was made, using the CYP142A1 crystal structure with 4-cholesten-3-one bound as a reference (Fig. S3)[11]. It has been previously noted that CYP142A1 has an active site which shares similarities (both residues and their orientation) with both CYP125A1 as well as CYP124A1 [11]. The residues that make up the CYP142A active site are highly conserved in CYP142A3 from *M. marinum* and *M. ulcerans* and CYP142A2 from *M. smegmatis* (Table S3). This supports the binding data, suggesting the behaviour of this enzyme would likely be similar across species. The small changes observed may be due to subtle differences in the conformation of the enzyme caused by mutations remote from the active site. Structural studies of CYP142A3 with substrate bound may help resolve this.

Table 3: Spin state shift and dissociation constant analysis of CYP142A1 from *M. tuberculosis* and CYP142A3 *M. marinum*. Data from elsewhere is noted, otherwise was performed as recorded in the experimental section.

	<i>M.mar142A3</i>		<i>M.tb142A1</i>	
	Spin state (HS%)	K_d (μ M)	Spin state (HS%)	K_d (μ M)
Cholesterol	55%	0.40 ± 0.1	$\geq 95\%$	0.34 ± 0.20^a
4-cholesten-3-one	65%	0.2 ± 0.1	$\geq 95\%$	0.36 ± 0.04^a
Miconazole	Type II, 424 nm	0.39 ± 0.1	Type II, 424 nm	4.0 ± 0.5^a
Clotrimazole	Type II, 423 nm	0.34 ± 0.12	Type II, 424 nm	3.8 ± 0.9^a
Fluconazole	<5%	-	<5%	860 ± 108^a
Econazole	Type II, 424 nm [†]	0.9 ± 0.1	Type II, 423 nm	4.6 ± 0.2^a
Ketoconazole	<5%	-	<5%	21 ± 4^a
1-Phenylimidazole	Type II, 422 nm	4.6 ± 0.2	Type II	-
2-Phenylimidazole	<5%	-	<5%	-
4-Phenylimidazole	Type II, 423 nm	7.1 ± 1.0	Type II, 421 nm	12.0 ± 1.5^a

^a Driscoll *et al*, 2010 [11]

[†] Initial 20% Type I shift before Type II shift

CYP124 in *M. marinum* and *M. tuberculosis*

CYP124A enzymes are also widely conserved in the *Mycobacteria*. Additionally, there are members of other subfamilies in *Frankia*, *Rhodococcus* and *Streptomyces* species. *M. tuberculosis* CYP124A1 is conserved in *M. marinum* and has a sequence identity of 84% (and has the same nomenclature CYP142A1; from here on these will be named MtbCYP124A1 and MmarCYP124A1). CYP124 ranks among the highest conserved CYP families when members were compared across species from different kingdoms by Parvez *et al* (above both CYP125 and CYP142)[35]. An analogue of MmarCYP124 exists in *M. ulcerans* (99% sequence identity). The *M. smegmatis* CYP124 enzyme, following the trend established with the two families discussed previously, shares 74% identity with both the *M. marinum* and *M. tuberculosis* members.

MtbCYP124 has been reported to bind branched fatty acids such as phytanic acid [14]. MmarCYP124 was purified and when CO was added to the ferrous form the Soret peak shifted almost entirely to the 450 nm absorbance typical of a viable P450 (Fig. 3).

MmarCYP124 demonstrated similar spin state shifts to MtbCYP124 with a range of substrates. Both farnesol and phytanic acid shifted both enzymes to a $\geq 95\%$ high spin state (Table 4). Spectral differences between the enzymes were observed with certain sub-terminally branched fatty acids, 14-methylhexadecanoic acid and 15-methylhexadecanoic acid as well as dodecyl acetate. Both fatty acids shifted MtbCYP124A1 to a greater extent than MmarCYP124A1 (85% and 90% for MtbCYP142A1 compared to 55% and 70% for MmarCYP124A1 with 14-methyl and 15-methyl hexadecanoic acid, respectively). Dodecyl acetate also shifted MtbCYP142A1 to a greater extent (90%) than MmarCYP124A1 (55%). Differences were observed when testing inhibitors with the two enzymes. Where no shift was recorded with MtbCYP124 for econazole and miconazole, both induced a Type I shift in CYP124 (45% and 40%, respectively). Clotrimazole was the only recorded azole to induce a Type II shift for both (MmarCYP124, $K_d = 0.34 \pm 0.12 \mu\text{M}$ and MtbCYP124, $0.22 \pm 0.05 \mu\text{M}$).

Table 4: Spin state shift analysis of CYP124A1 from *M. tuberculosis* and *M. marinum*.

	<i>M.mar124</i> Spin state (HS%)	<i>M.tb124</i> Spin state (HS%)
Farnesol	≥ 95	≥ 95
Phytanic acid	≥ 95	≥ 95
4,7,3-Trimethyldodecanoic acid	≥ 95	≥ 95
Farnesyl acetate	≥ 95	≥ 95
Linalyl acetate	≥ 95	≥ 95
15-Methylhexadecanoic acid	70	90
Undecanoic acid	70	75
Methyl laurate	60	70
3,7-Dimethyloctanoic acid	60	60-70
14-Methylhexadecanoic acid	55	85
Palmitic acid	55	60
Dodecyl acetate	55	90-95
Palmitelaidic acid	50	55
Methyl palmitate	30	55
Econazole	45	0
Miconazole	40	0
Clotrimazole	Type II, 426 nm	Type II

NB: Based on the substrate range determined by Johnston *et al*, 2009 [14]

MmarCYP124A1 was crystallised in the presence of the substrate phytanic acid, in order to structurally compare it to the *M. tuberculosis* enzyme. However, while crystals developed from both substrate-bound and substrate-free conditions, no substrate was present in the solved structures from substrate-bound conditions. Accordingly, only the substrate-free form of the enzyme was structurally resolved, to 2.2 Å (PDB: 6CVC). Of 433 residues, 424 were modelled, including the β -sheet loop region which was not modelled in the substrate-free Mtb structure (2WM5) [14]. When overlaid with the Mtb124, the overall fold of the two enzymes are very similar (Fig. 5). Using Chimera, the RMSD for the 419 atoms which overlapped was calculated to be 0.663 Å [39].

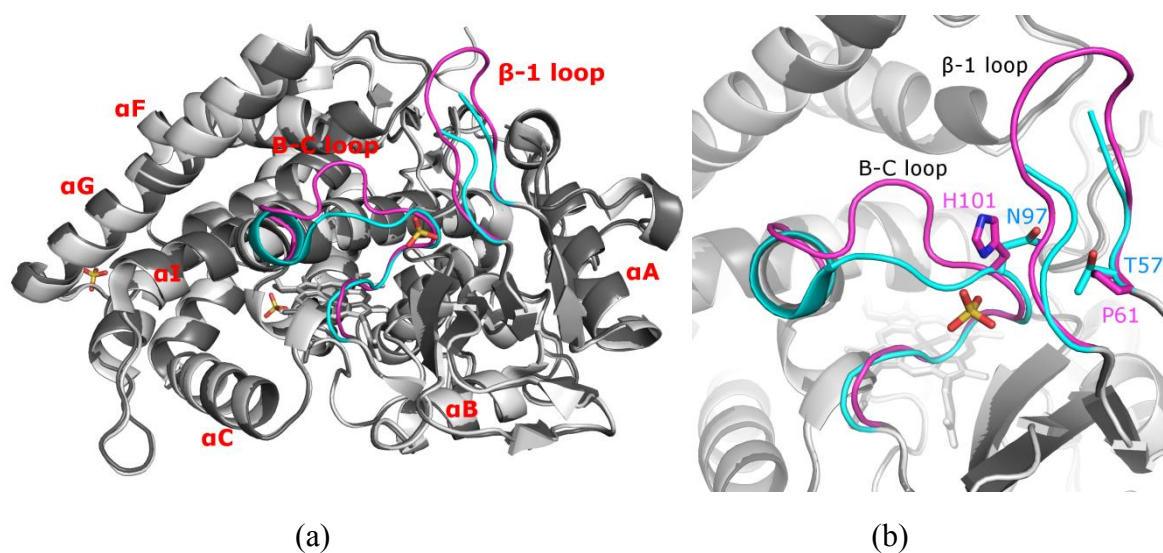


Figure 5: The overlay of X-ray crystal structures of the CYP124A1 enzymes from *M. marinum* (grey, highlights in purple) and *M. tuberculosis* (white, highlights in blue, PDB code 2WM5 [14]). (a) The overall fold of the two proteins is almost indistinguishable: the biggest deviation is in a small unstructured region between helix B and B' (also known as the B-C loop). The full loop turn region in the $\beta-1$ sheet modelled in the *M. marinum* enzyme but not in *M. tuberculosis*. The RMSD of the two structures is 0.663 Å across all 419 paired residues. (b) Zoom of the two altered regions showing the active site residues N97 and T57, altered to H101 and P61.

The active site residues are conserved for the most part. There are three active site residues that are altered: the MtbCYP124 residue N97 is changed for a histidine, T95 for a valine and T57 for a proline (Table 5). These changes are also present in the *M. ulcerans* analogue. The proline that replaces T57 (in MmarCYP124 numbering, P61) does not seem to substantially alter the position of the β -1 loop. The entire loop is not modelled in the MtbCYP124A1 structure, however, so comparison of the loop position cannot be made (Fig. 5). The effect of the T95 alteration to a valine (and hence increased hydrophobicity of the active site) would be better observed in a substrate-bound structure, so substrate-free MmarCYP124A1 was overlaid with phytanic acid bound MtbCYP124A1 (Fig 7).

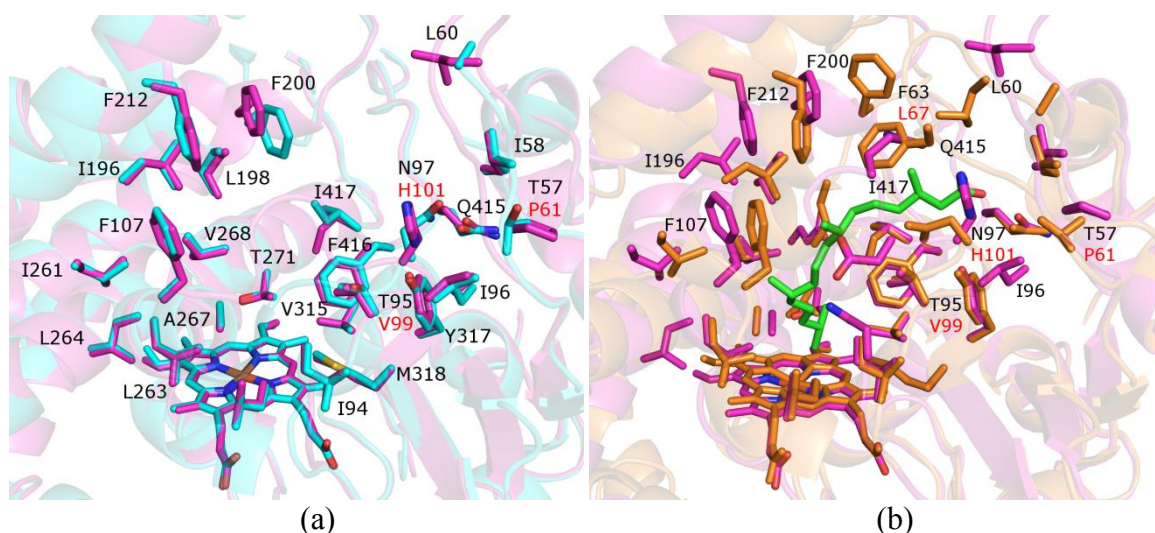


Figure 6: The residues close to the haem in the CYP124A1 active site as identified by Johnston *et al* show very tight retention in the *M. marinum* analogue. (a) MmarCYP124A1 overlaid with substrate-free MtbCYP124A1 (2WM5, blue) and (b) with phytanic acid bound MtbCYP124A1, which shows the conformational changes due to substrate binding (2WM4, orange) [14]. Labels in black are for MtbCYP124 (and where they are conserved in both) and red indicates a MmarCYP124 residue.

T95 and N97 were the two hydrophilic residues, identified as interacting with the carboxy terminus of the acid [14]. This may reduce enzyme affinity or recognition of the polar terminus of substrates such as fatty acids, although it does not appear to have significantly affected phytanic acid binding. The remaining change, N97 to H101 (in the respective

enzyme residue numbering), also falls within the region of most the significant structural change between the two enzymes (near the MtbCYP124 residues 97 – 101, shown in Fig. 5). This structural movement, which includes the premature end to the B' helix, may be due to the alteration of N97 to H101. It is close to the active site and as a result could alter substrate binding if it reflects a genuine structural alteration between the *M. marinum* and *M. tuberculosis* enzymes rather than a crystallisation artefact (see Fig. S9 and S10).

Table 5: Comparison of the active site residues of CYP124A1 from *M. tuberculosis* and *M. marinum* based on the residues identified as within 5Å of phytanic acid in the substrate-bound form of CYP124A1 from *M. tuberculosis* [14]. CYP124A1 from *M. smegmatis* is included for comparison. Residues altered from MtbCYP124A1 are in red.

Mtb 124A1	Mmar [†] 124A1	Msmeg 124A1
T57	P61	A53
I58	I62	F54
L60	L64	T56
F63	L67	F60
I94	I98	I91
T95	V99	V92
I96	I100	I93
N97	H101	G94
F107	F111	F104
I111	I115	I108
I197	I201	I194
L198	L202	L195
F200	F204	F197
F212	F216	F209
I262	I266	I259
L263	L267	L260
V266	V270	V263
A267	A271	A264
T271	T275	T268
V315	V319	V312
Y317	Y321	Y314
M318	M322	M315
Q415	Q419	A412
F416	F420	F413
I417	I421	I414

[†] *M. ulcerans* CYP124A1 is the same as MmarCYP124A1 for every residue listed (residue numbering between the two is also identical).

4. Conclusions

In each of the three CYP families, CYP125, CYP142 and CYP124, the *M. marinum* analogue of the enzyme was the most closely related to the *M. tuberculosis*, outside of the MTBC. The *M. marinum* enzymes were also closely related ($\geq 98\%$ identity) to the equivalent enzymes from *M. ulcerans*. In the CYP125 system, the additional redundancy present in *M. smegmatis* was replicated in *M. marinum*, but the sequence identities of the enzymes showed one, CYP125A7, offered a clearer direct analogue to CYP125A1, and thus similar activity was expected. However, differences in resting spin state, substrate induced haem spin state shifts and inhibitor binding affinities were observed with CYP125A7 compared to CYP125A1. The variations in sequence that might affect these changes were identified but need to be confirmed. These were remote from the active site and may either affect substrate recognition or alter the dynamics of the enzyme to modify substrate binding.

A greater level of conservation in the substrate binding properties of the CYP142 family isoforms was observed. CYP142A3 had similar affinity for the substrates which interact with CYP142A1. The recorded affinity for azole inhibitors with the *M. marinum* enzyme was higher by an order of magnitude for most of those tested. The active site residues, which were identical between the enzymes, did not reveal the probable cause of this change.

CYP124A1 from *M. marinum* was also subjected to substrate and inhibitor binding studies, and generally displayed a high degree of conserved substrate range with the *M. tuberculosis* enzyme. The substrate-free structure of MmarCYP124A1 was solved to 2.2 Å, and was shown to have a very similar overall fold to the MtbCYP124A1. From the structure, we identified residue changes in the region of substrate binding. The N97→H and T57→V changes (present in both *M. marinum* and *M. ulcerans*) are close to where the carboxylate

terminus of phytanic acid binds in the *M. tuberculosis* substrate-bound structure. However, despite these active site changes, the *M. marinum* CYP124A1 binding properties closely mirror that of the *M. tuberculosis* analogue.

The three enzymes together offer a better model for the cholesterol metabolising monooxygenases *M. ulcerans* than any previously studied, and highlight some possible changes in substrate binding and therefore potentially activity and role of these enzymes in that species (and *M. marinum*) and *M. tuberculosis*. The cholesterol metabolising pathway is a target for *Mycobacterium* inhibition studies. Therefore, of particular relevance is the demonstration that azole inhibitor affinities and the binding spectra change significantly between enzymes of these different *Mycobacterium* species and fine tuning of these azole molecules may be required to optimise the effects between bacteria.

Acknowledgements

This work was supported by the Australian Research Council through a Future Fellowship (FT140100355 to S.G.B.) and the award of a University of Adelaide Faculty of Sciences Divisional Scholarship (PhD to S.A.C.). X-ray diffraction data collection was undertaken on the MX1 beamline at the Australian Synchrotron, part of the Australian Nuclear Science and Technology Organisation. We acknowledge financial support from the Australian Synchrotron in the form of MXCAP9674. We are grateful to Professor Tim Stinear from the University of Melbourne and Professor Lalita Ramakrishnan of the University of Cambridge for providing the genomic DNA of *M. marinum*.

References

- [1] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry, 3rd, F. Tekaiia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M.A. Quail, M.A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J.E. Sulston, K. Taylor, S. Whitehead, B.G. Barrell, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, *Nature*, 393 (1998) 537-544.
- [2] World Health Organisation: Global Tuberculosis Report, in, http://www.who.int/tb/publications/global_report/en/, 2017.
- [3] C. Demangel, T.P. Stinear, S.T. Cole, Buruli ulcer: reductive evolution enhances pathogenicity of *Mycobacterium ulcerans*, *Nature reviews. Microbiology*, 7 (2009) 50-60.
- [4] K. Nakanaga, R.R. Yotsu, Y. Hoshino, K. Suzuki, M. Makino, N. Ishii, Buruli ulcer and mycolactone-producing mycobacteria, *Jpn J Infect Dis*, 66 (2013) 83-88.
- [5] T.P. Stinear, T. Seemann, P.F. Harrison, G.A. Jenkin, J.K. Davies, P.D. Johnson, Z. Abdallah, C. Arrowsmith, T. Chillingworth, C. Churcher, K. Clarke, A. Cronin, P. Davis, I. Goodhead, N. Holroyd, K. Jagels, A. Lord, S. Moule, K. Mungall, H. Norbertczak, M.A. Quail, E. Rabinowitsch, D. Walker, B. White, S. Whitehead, P.L. Small, R. Brosch, L. Ramakrishnan, M.A. Fischbach, J. Parkhill, S.T. Cole, Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*, *Genome research*, 18 (2008) 729-741.
- [6] D.M. Tobin, L. Ramakrishnan, Comparative pathogenesis of *Mycobacterium marinum* and *Mycobacterium tuberculosis*, *Cellular microbiology*, 10 (2008) 1027-1039.
- [7] M.U. Shiloh, P.A. Champion, To catch a killer. What can mycobacterial models teach us about *Mycobacterium tuberculosis* pathogenesis?, *Current opinion in microbiology*, 13 (2010) 86-92.
- [8] K.J. McLean, D. Clift, D.G. Lewis, M. Sabri, P.R. Balding, M.J. Sutcliffe, D. Leys, A.W. Munro, The preponderance of P450s in the *Mycobacterium tuberculosis* genome, *Trends in microbiology*, 14 (2006) 220-228.
- [9] H. Ouellet, J.B. Johnston, P.R. Ortiz de Montellano, The *Mycobacterium tuberculosis* cytochrome P450 system, *Arch. Biochem. Biophys.*, 493 (2010) 82-95.
- [10] D.R. Nelson, L. Koymans, T. Kamataki, J.J. Stegeman, R. Feyereisen, D.J. Waxman, M.R. Waterman, O. Gotoh, M.J. Coon, R.W. Estabrook, I.C. Gunsalus, D.W. Nebert, P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature, *Pharmacogenetics*, 6 (1996) 1-42.
- [11] M.D. Driscoll, K.J. McLean, C. Levy, N. Mast, I.A. Pikuleva, P. Lafite, S.E. Rigby, D. Leys, A.W. Munro, Structural and biochemical characterization of *Mycobacterium tuberculosis* CYP142: evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen, *J. Biol. Chem.*, 285 (2010) 38270-38282.
- [12] J.K. Capyk, R. Kalscheuer, G.R. Stewart, J. Liu, H. Kwon, R. Zhao, S. Okamoto, W.R. Jacobs, Jr., L.D. Eltis, W.W. Mohn, Mycobacterial cytochrome p450 125 (*cyp125*) catalyzes the terminal hydroxylation of c27 steroids, *J. Biol. Chem.*, 284 (2009) 35534-35542.
- [13] H. Ouellet, S. Guan, J.B. Johnston, E.D. Chow, P.M. Kells, A.L. Burlingame, J.S. Cox, L.M. Podust, P.R. de Montellano, *Mycobacterium tuberculosis* CYP125A1, a steroid C27 monooxygenase that detoxifies intracellularly generated cholest-4-en-3-one, *Molecular microbiology*, 77 (2010) 730-742.

- [14] J.B. Johnston, P.M. Kells, L.M. Podust, P.R. Ortiz de Montellano, Biochemical and structural characterization of CYP124: a methyl-branched lipid omega-hydroxylase from *Mycobacterium tuberculosis*, Proc. Natl. Acad. Sci. U. S. A., 106 (2009) 20687-20692.
- [15] A.V. Vasilevskaya, A.V. Yantsevich, G.V. Sergeev, A.P. Lemish, S.A. Usanov, A.A. Gilep, Identification of *Mycobacterium tuberculosis* enzyme involved in vitamin D and 7-dehydrocholesterol metabolism, J. Steroid Biochem. Mol. Biol., 169 (2017) 202-209.
- [16] K.M. Wilburn, R.A. Fieweger, B.C. VanderVen, Cholesterol and fatty acids grease the wheels of *Mycobacterium tuberculosis* pathogenesis, Pathogens and Disease, 76 (2018) fty021-fty021.
- [17] B.C. VanderVen, R.J. Fahey, W. Lee, Y. Liu, R.B. Abramovitch, C. Memmott, A.M. Crowe, L.D. Eltis, E. Perola, D.D. Deiningner, T. Wang, C.P. Locher, D.G. Russell, Novel Inhibitors of Cholesterol Degradation in *Mycobacterium tuberculosis* Reveal How the Bacterium's Metabolism Is Constrained by the Intracellular Environment, PLOS Pathogens, 11 (2015) e1004679.
- [18] R. Van der Geize, K. Yam, T. Heuser, M.H. Wilbrink, H. Hara, M.C. Anderton, E. Sim, L. Dijkhuizen, J.E. Davies, W.W. Mohn, L.D. Eltis, A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages, Proc. Natl. Acad. Sci. U. S. A., 104 (2007) 1947-1952.
- [19] K.J. McLean, P. Lafite, C. Levy, M.R. Cheesman, N. Mast, I.A. Pikuleva, D. Leys, A.W. Munro, The Structure of *Mycobacterium tuberculosis* CYP125: MOLECULAR BASIS FOR CHOLESTEROL BINDING IN A P450 NEEDED FOR HOST INFECTION, J. Biol. Chem., 284 (2009) 35524-35533.
- [20] J. Pieters, Entry and survival of pathogenic mycobacteria in macrophages, Microbes and infection, 3 (2001) 249-255.
- [21] S.L. Kendall, S.C. Rison, F. Movahedzadeh, R. Frita, N.G. Stoker, What do microarrays really tell us about *M. tuberculosis*?, Trends in microbiology, 12 (2004) 537-544.
- [22] H. Ouellet, E.D. Chow, S. Guan, J.S. Cox, A.L. Burlingame, P.R. de Montellano, Genetic and mass spectrometric tools for elucidating the physiological function(s) of cytochrome P450 enzymes from *Mycobacterium tuberculosis*, Methods Mol. Biol. (Clifton, N.J.), 987 (2013) 79-94.
- [23] E. Garcia-Fernandez, D.J. Frank, B. Galan, P.M. Kells, L.M. Podust, J.L. Garcia, P.R. Ortiz de Montellano, A highly conserved mycobacterial cholesterol catabolic pathway, Environ. Microbiol., 15 (2013) 2342-2359.
- [24] S.A. Child, E.F. Naumann, J.B. Bruning, S.G. Bell, Structural and functional characterisation of the cytochrome P450 enzyme CYP268A2 from *Mycobacterium marinum*, Biochem. J., (2018).
- [25] S.A. Child, E.F. Naumann, J.B. Bruning, S.G. Bell, Structural and functional characterisation of the cytochrome P450 enzyme CYP268A2 from *Mycobacterium marinum*, Biochemical J., 475 (2018) 705-722.
- [26] T. Omura, R. Sato, The Carbon Monoxide-Binding Pigment of Liver Microsomes. II. Solubilization, Purification, and Properties, J. Biol. Chem., 239 (1964) 2379-2385.
- [27] D.J. Frank, C.A. Waddling, M. La, P.R. Ortiz de Montellano, Cytochrome P450 125A4, the third cholesterol C-26 hydroxylase from *Mycobacterium smegmatis*, Biochemistry, 54 (2015) 6909-6916.
- [28] J.M. Dogne, X. de Leval, P. Benoit, S. Rolin, B. Pirotte, B. Masereel, Therapeutic potential of thromboxane inhibitors in asthma, Expert Opin. Investig. Drugs, 11 (2002) 275-281.
- [29] T.M. McPhillips, S.E. McPhillips, H.J. Chiu, A.E. Cohen, A.M. Deacon, P.J. Ellis, E. Garman, A. Gonzalez, N.K. Sauter, R.P. Phizackerley, S.M. Soltis, P. Kuhn, Blu-Ice and the

Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines, *J. Synchrotron Radiat.*, 9 (2002) 401-406.

[30] P.R. Evans, G.N. Murshudov, How good are my data and what is the resolution?, *Acta Crystallogr. D*, 69 (2013) 1204-1214.

[31] M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R.M. Keegan, E.B. Krissinel, A.G. Leslie, A. McCoy, S.J. McNicholas, G.N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin, K.S. Wilson, Overview of the CCP4 suite and current developments, *Acta Crystallogr. D*, 67 (2011) 235-242.

[32] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, Phaser crystallographic software, *J. Appl. Crystallogr.*, 40 (2007) 658-674.

[33] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, *Acta Crystallogr. D*, 66 (2010) 486-501.

[34] R.K. Z., W.M. H., C.J. K., M.W. W., O. Martin, V.D.G. Robert, D. Lubbert, E.L. D., Cytochrome P450 125 (CYP125) catalyses C26-hydroxylation to initiate sterol side-chain degradation in *Rhodococcus jostii* RHA1, *Molecular microbiology*, 74 (2009) 1031-1043.

[35] M. Parvez, L.B. Qhanya, N.T. Mthakathi, I.K.R. Kgosiemang, H.D. Bamal, N.S. Pagadala, T. Xie, H. Yang, H. Chen, C.W. Theron, R. Monyaki, S.C. Raselemane, V. Salewe, B.L. Mongale, R.G. Matowane, S.M.H. Abdalla, W.I. Booi, M. van Wyk, D. Olivier, C.E. Boucher, D.R. Nelson, J.A. Tuszyński, J.M. Blackburn, J.-H. Yu, S.S. Mashele, W. Chen, K. Syed, Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: Special focus on mycobacterial P450s, *Scientific Reports*, 6 (2016) 33099.

[36] K.J. McLean, A.J. Warman, H.E. Seward, K.R. Marshall, H.M. Girvan, M.R. Cheesman, M.R. Waterman, A.W. Munro, Biophysical characterization of the sterol demethylase P450 from *Mycobacterium tuberculosis*, its cognate ferredoxin, and their interactions, *Biochemistry*, 45 (2006) 8427-8443.

[37] H. Ouellet, P.M. Kells, P.R. Ortiz de Montellano, L.M. Podust, Reverse type I inhibitor of *Mycobacterium tuberculosis* CYP125A1, *Bioorg. Med. Chem. Lett.*, 21 (2011) 332-337.

[38] D.L. Harris, Y.-T. Chang, G.H. Loew, Molecular dynamics simulations of phenylimidazole inhibitor complexes of cytochrome P450 cam, *Molec. Eng.*, 5 (1995) 143-156.

[39] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera--a visualization system for exploratory research and analysis, *J. Comp. Chem.*, 25 (2004) 1605-1612.

Conclusions and future directions

The work contained in this PhD thesis represents the beginning of a substantial task. The *M. marinum* CYPome is interesting as, being larger than that of *M. tuberculosis* and indeed most bacteria studied so far, it offers greater diversity in P450s and the single iron-sulfur cluster ferredoxins that support them. Nevertheless, in the purification and characterisation of nine of the eleven ferredoxins (Fdx1 - Fdx11) and ten P450s from this *Mycobacterium* (these being CYP147G1, CYP278A1, CYP268A2, CYP269A1, CYP150A5 and A6, CYP125A6 and A7, CYP124A1 and CYP142A3), substantial progress has been made towards the understanding of the CYPome. The first of this thesis' objectives was the purification of the uncharacterised [3/4Fe-4S] ferredoxins. The aerobic purification of eight of the ferredoxins was straightforward and gave reasonable yields. These were the ferredoxins which contained histidine, phenylalanine and asparagine residues at the non-conserved position of the cluster binding motif. The remaining three, however, which had the hydroxyl-containing residues tyrosine, threonine and serine at that position, could not be purified aerobically. The well-known oxygen sensitivity of iron-sulfur clusters led to the attempted anaerobic purification of several, including the successful purification of the serine-containing Fdx9. That the difficulties in expression and purification should be confined to those with some of the most unusual residues at that particular position is noteworthy. Phylogenetic analysis revealed ferredoxins containing the CXXTXXC(X)_nCP motif were not found outside of *Mycobacterium* while histidine-containing ferredoxins were distributed most broadly. The phenylalanine, serine, asparagine and tyrosine containing ferredoxins were also widespread in *Mycobacteria* and other Actinomycetes.

EPR and non-denaturing ESI-MS confirmed Fdx4, 5 and 9 (with an asparagine, histidine, and serine residue, respectively) contain [3Fe-4S] clusters after isolation under anaerobic conditions. While the cluster composition of the entire suite has not been determined, it

seems likely that these ferredoxins do not contain a 4Fe-4S cluster through the direct coordination of a fourth Fe atom by the residue that replaces cysteine in the cluster binding motif. The redox properties of these ferredoxins are unusual, encompassing reduction potentials (between -360 and +150 mV) which fall outside the range of potentials observed previously with CYP-coupled ferredoxins. Earlier mutant studies with CYP147G1/Fdx3 had shown partial, but not complete, loss of function upon mutation of the tyrosine residue of that ferredoxin. This leaves the role of the residue at this position unclear. In this sense, the second aim of this project, the biophysical characterisation of the ferredoxins, has been partially achieved. The highly positive reduction potentials of Fdx4 (+60 mV) and Fdx5 (+150 mV), in particular, raise the question of how these ferredoxins and those similar to them are able to support P450 activity, given that favourable thermodynamics are hypothesised to facilitate electron transfer between the ferredoxin and the substrate-bound P450 (the reduction potential of which is ordinarily below -100 mV).

To fully test this required the achievement of the third goal of this project, the characterisation of the substrate range of the accompanying CYPs in the *M. marinum* genome. Among the 47 *Cyp* genes in the genome, those studied were selected based on their location near to ferredoxins of interest, coupled with the results of small-scale test heterologous protein production in *Escherichia coli*. It was apparent that the results of the characterisation of those as yet uncharacterised enzymes were of interest in their own right. Some of these P450 enzymes belong to families that are found across other bacteria, beyond *Mycobacterium* species. One of these, the CYP147G1 enzyme, demonstrated very high selectivity for ω -1 hydroxylation of fatty acids, and the physiological Fdx3/FdR1 full electron transfer system supported efficient turnover. The high expression levels of CYP268A2 and early indications of the substrate range made it a good candidate as a broad activity CYP, hindered only by the lack of a clear physiological Fdx/FdR pair. P450 activity was

reconstituted with a non-native system. The crystal structure of the substrate-bound enzyme revealed features connected with substrate selectivity and will be the basis of future investigation into this CYP family. CYP150A5, which has three ferredoxins nearby in the genome, and a family member in *M. vanbaalenii* reported as a possible polyaromatic hydroxylase was also an obvious target. It was studied in conjunction with CYP150A6, as the substrate range of the two was expected to be similar. However, only CYP150A5 activity was reconstituted as the substrate range of the CYP150A6 enzyme proved difficult to characterise. CYP150A6 was, however, characterised by X-ray crystallography. The regioselective hydroxylation of the terpenoids β -ionone and β -ionol by CYP150A5 was achieved. The electron transfer ability of Fdx4 and 5 could not be studied directly, as the activity of the CYPs that accompany them in the genome (CYP269A1/CYP138A4 and CYP188A3, respectively) could not be produced. Fdx8 also contains a histidine residue at that position and can support CYP150A5 activity. The other two ferredoxins that accompany CYP150A5 in the genome, Fdx6 and Fdx7 demonstrated little to no activity with CYP150A5. Alongside the CYP147G1/Fdx3 and CYP278A1/Fdx2 pairs, this makes three of the eleven ferredoxins (with tyrosine, serine and histidine residues at that position) that have been demonstrated to support their accompanying CYPs. It supports the conjecture based on their genome position, that these ferredoxins are one of the electron transfer partners of the CYP enzymes, despite the unfavourable reduction potentials.

The final standalone project goal was the comparative analysis of CYPs that are conserved between *Mycobacterium* species. The *M. tuberculosis* steroid binding enzymes, CYP125, CYP124 and CYP142, are conserved in *M. marinum* and many other *Mycobacteria* and bacteria. This led to the purification and analysis of the four equivalent *M. marinum* enzymes from the same families (it contains two CYP125 family members), and the crystallisation of the *M. marinum* CYP124A1. The phylogenetic, binding and structural characterisation

revealed the *M. marinum* counterparts had key similarities and differences to those of *M. tuberculosis*. They were expected to be a closer match to the analogous enzymes from *M. ulcerans*. The altered resting state between the two *M. marinum* CYP125 members and the *M. tuberculosis* CYP125 enzyme infers a possible structural change despite the high levels of sequence identity while the CYP124 and CYP142 analogues from both species were very similar in activity. The differences observed here have particular relevance to the targeting of CYP enzymes from *M. ulcerans* and *M. tuberculosis* for inhibition as a methodology for developing antimicrobial medicine.

The importance of cholesterol and fatty acids to the *Mycobacterium* physiology has been born out in the course of this study. Many of the CYP enzymes studied demonstrated high selectivity for such substrates. With the majority of *M. tuberculosis* enzymes conserved in *M. marinum* (15 of the 20), the effect of the larger *M. marinum* genome was reflected in the additional breadth of the P450 enzyme activities. For example, in addition to the fatty acid hydroxylase CYP124A1, *M. marinum* also encodes CYP147G1, which accepts shorter chain-length fatty acid substrates, and CYP268A2, which favours branched fatty acids and acetates. In addition, the physiological ferredoxins allowed the characterisation of the activities of CYPs that appear highly selective for their electron transfer partner. While commercially available or non-physiological electron transfer systems have been used successfully elsewhere, and were used for some reactions in this project, the native system where known consistently outperformed these. The distribution of ferredoxins such as these in other bacterial secondary metabolism reactions, supporting the biosynthesis of many biologically active compounds, increases the relevance of their effective reconstitution here. Thus the suggested advantages highlighted at the beginning of this project have been confirmed; the breadth and interpretability of the *M. marinum* CYPome has led to the most relevant findings.

There are still several open questions about the *M. marinum* system studied here and found in other bacteria however, which should be addressed in future work. While regioselectivity was demonstrated with several of the studied enzymes, the stereoselectivity of the reaction products remains unknown. Particularly with CYP147G1 and CYP268A2, this is of potentially significant interest. The biggest hurdles to the commercial application of P450s remain achieving binding for the desired substrate, product selectivity and the identification of appropriate electron transfer partners. The determination of product stereoselectivity, if any is present, also adds to the sum of knowledge available for engineering these enzymes, especially if accompanied by structural information about substrate binding as in CYP268A2.

The second of these is whether FdR1 is the physiological reductase for all of these systems. Where the CYP147G1/Fdx3/FdR1 system is present in other bacteria, the Fdx3/FdR1 pair are often found as a single, fused protein. This seems to decrease the likelihood that FdR1 functions as a general reductase for multiple P450 systems. Indeed, it seems possible that in the CYP150A5 system, the low product formation observed was due to the choice of ferredoxin reductase, rather than the ferredoxin or the substrate affinity. The importance of determining the physiological reductase is high as it has likely evolved to optimise their interaction with the ferredoxins and the reduction potential for electron transfer. The purification of the known ferredoxin reductases in the *M. marinum* genome, while not successful as yet, would allow the more precise *in vitro* characterisation of protein-protein interactions than *in vivo* product formation studies can achieve. However, given neither FdR1 nor FdR2 (the latter supports the [2Fe-2S] Fdx and CYP153A16) are conserved in *M. tuberculosis*, the hypothesis that the majority of the P450s in both species are supported by yet another ferredoxin reductase or electron transfer partner is credible.

The interaction between a ferredoxin and CYP pair or even between the ferredoxin and multiple supported CYPs has not been analysed quantitatively here. The given reduction

potentials of the ferredoxins are estimations by titration with oxidising or reducing agents (rather than by electrochemistry) and the CYP reduction potentials have not been measured as yet. Further, the relative importance of protein-protein interactions versus reduction potential in controlling the specificity and efficiency of the transfer is still uncertain. This could be achieved by a combination of structural analysis and additional biophysical characterisation.

Another unanswered question regards the essentiality of these enzymes. A greater overlap of function has been observed with the enzymes from *M. marinum* (CYP278A1 and CYP150A5 with terpenes, CYP268A2 and CYP147G1 with fatty acids, and the additional CYP125 family member), but this leads to less clarity about their intracellular roles. Here gene knock-out, transcription or proteomic studies in *M. marinum* would be informative. While activity for fatty acids and steroid oxidation and metabolism is apparent with several enzymes, this is currently without cellular context, as the final metabolite and its function are unknown. As many of these P450s are closely related to analogues in *M. ulcerans*, these roles could be expected to be conserved in the pathogenic metabolism. Systematic analysis of the biosynthetic gene clusters in the organism, with a focus on the presence of P450s, might reveal more information about some of these enzymes. Similar ferredoxins to those that accompany *M. marinum* P450s were frequently found associated with P450s in natural product biosynthetic gene clusters in other organisms. The biosynthetic gene clusters of many antibiotics, anticancer and antifungal compounds, such as the *Streptomyces* products salinomycin, cinnabaramide and filipin, all contain ferredoxins of this type.

In attempting the first analysis of such a large P450 complement as is presented by *M. marinum*, it was not expected that complete understanding could be achieved in the course of this study. However, despite encountering the common issues of protein expression, stability and unclear function, the work presented here has relevance both to the theoretical

understanding of cytochrome P450 systems, as well as the practical applications of their study, inhibition and biocatalysis. By our characterisation, the nature and activity of a series of CYP-supporting [3Fe-4S] ferredoxins has been elucidated. These were found to be particularly common supporting P450s in Actinomycete biosynthetic gene clusters, which are still one of the predominant sources of new natural products today. Significant progress has been made towards comprehending the diverse CYPome of *M. marinum*, with findings that are sure to have relevance to the ongoing study of the more pathogenic strains of *Mycobacteria*. The Buruli ulcer continues to be difficult to treat and contain, and binding and inhibition data is given here for several enzymes that are very closely related to their *M. ulcerans* counterparts. The unique pathogenicity of many strains of *Mycobacteria* gives the study of the contribution by P450s to the physiology of *M. marinum* applications in human health and disease, especially with respect to the closely related strain *M. ulcerans*. The abundance of cytochrome P450s throughout all the kingdoms of life and the selectivity of their metallo-enzyme catalysis makes their study pertinent to the entire field of biological chemistry.

Appendices

Chapter 2 Supplementary Information

The CYPome of *Mycobacterium marinum*

There are forty seven CYP enzyme encoding genes in *M. marinum* and these belong to thirty six different P450 families and thirty nine subfamilies (Table S1 and Fig. S1). Those members of the same family but different subfamilies are CYP123A3 and CYP123B1 (43%; sequence identity), CYP136A2 and CYP136B2 (46%) and CYP138A3 and A4 and CYP138B1 (47% and 45%, respectively). There are multiple members of several subfamilies including CYP125A6 and A7 (75%), CYP135B3, B4 (71%) and B6 (66% to B3 and 65% to B4), CYP138A3 and A4 (62%), CYP143A3 and A4 (58%), CYP150A5 and A6 (55%), CYP187A4 and A5 (57%) and CYP189A6 and A7 (58%). The proteins range from 389 to 491 amino acids in length and fourteen of the CYPs encoded by these genes are predicted to be positively charged at pH 7.0 with the remainder being negatively charged (Table S1).

The glutamate and arginine pair (EXXR) heme binding residues, which are highly conserved in the K-helix of the majority of CYP enzymes, was retained in all of those from *M. marinum* (Table S1).¹ The acid alcohol pair in the I helix was also conserved across most CYP members, the exceptions being CYP276A1 (GT) CYP135B3 (DN) and CYP51B1 (HT). Where it was conserved the acid residue was a mixture of aspartate and glutamate residues (18:27). The alcohol of this pair was predominantly threonine with only CYP144A4, CYP147G1 and CYP269A1 containing a serine residue (Table S1).² The phenylalanine residue which is often found seven residues before the conserved proximal cysteine (or eight if an additional glycine residue is inserted) was also mostly retained with only CYP183B1 (glutamate), CYP278A1 (leucine) and CYP147G1 (tryptophan) deviating from the norm (Table S1).³

In addition to the CYP51B1, CYP124A1, CYP153A16, CYP142A3, CYP125A6 and CYP125A7 enzymes mentioned in the main text, several other CYP enzymes in *M. marinum* have homologues from *M. tuberculosis* which have been structurally characterised and tested with azole inhibitors (e.g. CYP130 and CYP144).^{4,5} The likely substrates for some of the other CYP enzymes can be predicted based on their homology with other CYP enzymes e.g. CYP150 family members are reported to support aromatic hydrocarbon oxidation.⁶ Finally the function of some can be inferred from their neighbouring genes as they are clustered with terpene and polyketide synthases and peptide synthetases. For example CYP226B1, CYP271A1, CYP183B1 and CYP274A1 are clustered together with enzymes involved in isoprenoid synthesis, CYP185A4 is found next to a gene with modules predicted to be responsible for the synthesis of an eight amino acid metabolite and a thioesterase domain, CYP139A3 is surrounded by genes involved in polyketide synthases and macrolide transport and CYP108B4, CYP187A4 and CYP187A5 are found together with genes involved in fatty acid and lipid metabolism. CYP genes such as CYP123A3 and CYP51B1 and CYP142A3 and CYP125A7, are found close together in a similar environment to those which have been described for other *Mycobacterium*.⁷ For example the CYP125 and CYP142 genes are part of a conserved cholesterol metabolism gene cluster.⁸

There are only twenty CYP enzymes in *M. tuberculosis* and twenty four in *M. ulcerans* (plus ten pseudogenes) compared to forty seven in *M. marinum* (Table S3). This follows from the smaller gene complement of these bacteria due to reductive evolution (Table S1). Twelve are common to all three *Mycobacteria*, twenty one are unique to *M. marinum* with eleven being common to *M. marinum* and *M. ulcerans* strains and three shared by *M. marinum* and *M. tuberculosis* (Table S3).

All of the CYP enzymes most closely associated with these ferredoxins have low pI values though other mycobacterial CYP enzymes have pI values higher than 7.0 (Table S1). The significance of the pI of the ferredoxin and CYP enzymes has not yet been fully established.

However, based on the number of published studies on mycobacterial CYP enzymes those with low pI appear to be more readily produced in a soluble form in *E. coli*. This trend seems to extend across other bacterial families where CYP enzymes with a high pI (> 7.0) are not produced as readily.^{5,9-16}

Experimental

CYP assignments and alignments

CYP family and subfamily assignments were made by Dr. David Nelson according to the P450 nomenclature^{17,18} and are used as given in the National Centre for Biotechnology Information (NCBI) database. The genes and proteins from *M. marinum* M were compared to *M. ulcerans* Agy99, *M. tuberculosis* H37Rv and *M. leprae* TN.^{19,20}

Sequence alignments were performed using ClustalW. The evolutionary history was inferred by using the Maximum Likelihood method based on the Jones-Taylor-Thornton (JTT) matrix-based model.²¹ The tree with the highest log likelihood is shown (Fig. S1). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis of the CYPs involved the amino acid sequences of all forty seven enzymes from *M. marinum* and selected others for comparison. The analysis of the ferredoxins included all similar species from *M. marinum*, *M. ulcerans*, *M. tuberculosis* and other structurally characterised ferredoxins of this type. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA6.²²

Construction of whole-cell oxidation system for CYP147G1/Fdx3/FdR1

General DNA manipulations and microbiological experiments were carried out by standard methods. The pETDuet and pRSFDuet vectors and the KOD Hot start polymerase used for the PCR

steps were from Merck-Millipore. Enzymes for molecular biology were from New England Biolabs. The genomic DNA of *M. marinum* (ATCC 700278D-5) was obtained from Prof. Tim Stinear (University of Melbourne, Australia) and Prof. Lalita Ramakrishnan (University of Washington, USA, now at Cambridge University, UK). The genes encoding the P450 enzyme CYP147G1 (MMar_2930), CYP150A5 (MMar_4737), CYP269A1 (MMar_3969) CYP278A1 (MMar_2877) and the ferredoxin reductase FdR1 (MMar_2931) were amplified by PCR using oligonucleotide primers (*vide infra*). The FdR1 and CYP genes were amplified by 30 cycles of strand separation at 95 °C for 45 s followed by annealing at 55 °C for 30 s and extension at 68 °C for 80 s. The genes were cloned into the pET26 vector using the appropriate restriction enzymes. Codon optimised genes for Fdx2 (Mmar_2879), Fdx3 (Mmar_2932), Fdx4 (Mmar_3973), Fdx8 (Mmar_4736), Fdx9 (Mmar_4763) and the mutants of Fdx3 encoding the putative 3/4Fe-4S ferredoxin were obtained from IDT in the form of a g-block with NcoI and HindIII restriction sites incorporated at the 5' and 3' termini, respectively (for primer details, *vide infra*). The sequence encoding a 6xHis tag was inserted at the 3' end of the gene using PCR (amplification for 30 cycles with strand separation at 95 °C for 30 s followed by annealing at 55 °C for 30 s and extension at 68 °C for 20 s, for primers, *vide infra*).

The CYP genes (CYP147G1, CYP150A5, CYP269A1 and CYP278A1) were incorporated into the pET26 vector (Merck-Millipore) between NdeI and HindIII (XhoI for CYP150A5) restriction sites. The ferredoxin genes were cut using the NcoI and HindIII restriction enzymes introduced by PCR and cloned into the pETDuet and pRSFDuet vectors using the same restriction sites to yield pETDuetFdx and pRSFDuetFdx. The pETDuetFdx constructs were used to produce the encoded proteins using *E. coli*. To generate the whole-cell oxidation systems the FdR1 and CYP genes were cut using NdeI and KpnI (XhoI for CYP150A5) restriction enzymes. The FdR1 gene was cloned into the pETDuetFdx vectors to yield the plasmid pETDuetFdx/FdR1. The CYP genes were cloned into the appropriate pRSFDuetFdx (NdeI/KpnI) to yield the plasmid pRSFDuetFdx/CYP. Successful incorporation of the genes and mutants into each vector was

confirmed by restriction enzyme digest followed by DNA sequencing (AGRF, Adelaide node) using the primers; T7 promoter, T7 terminator, pET Upstream, ACYDuetUP1, DuetUP2, DuetDOWN1 appropriate for each parent vector (Merck-Millipore).

CYP enzyme production and purification

The pET26 plasmid containing the appropriate CYP gene was transformed into BL21(DE3) competent *E. coli* and plated onto LB_{kan} plates and left overnight. A colony was picked and grown in 5 mL LB_{kan} at 37 °C and this starter culture was used to grow 2L of 2xYT_{kan} (4 x 500 mL) at 37 °C for 5 hours at 150 rpm. The incubator temperature was lowered to 25 °C, followed by the addition of IPTG (0.1 mM) and further incubation at this temperature for 16 hours at 120 rpm. The cells were harvested by centrifugation (5,000 g, 15 min) and resuspended in 200 mL of 50 mM Tris, pH 7.4, containing 1 mM DTT (henceforth Buffer T). The resuspended cells were kept on ice and lysed by sonication (25 cycles at 20:40 seconds on:off, 70%, 19 mm probe, Sonics Vibra-Cell). The supernatant, containing the desired protein, was isolated from cell debris by centrifugation (40,000 g, 30 min). The protein was then loaded onto a DEAE Sepharose column, (XK50, 200 mm x 40 mm, GE Healthcare) and eluted using a linear salt gradient of 100 mM to 400 mM KCl in Buffer T. The fractions containing the desired protein (identified by red colour of the P450 enzymes) were combined and the volume reduced using a Vivacell 100 (Sartorius Stedim, 10 kD membrane) aided by centrifugation (1,500 g). The protein was desalted using a Sephadex G-25 medium grain column (250 mm x 40 mm) eluted with Buffer T. The desalted protein was then further purified by loading it onto a Source-Q ion-exchange column (XK26, 80 mm x 30 mm, GE Healthcare) and eluted with a gradient of 0 to 300 mM KCl in Buffer T. The fractions were selected for purity of protein by measurement of A_{418}/A_{280} ratio. Fractions with the highest A_{418}/A_{280} ratio were retained. The protein was concentrated via ultrafiltration and an equal volume of 80% glycerol was added before filtration through a 0.22 µm syringe filter and storage at -20 °C.

To assess the viability of the P450 enzyme in the purified samples and determine the protein extinction coefficient, the enzyme was diluted to approximately 2 μM in Buffer T and the spectrum recorded between 300 and 700 nm on a UV/Vis spectrometer. A few flakes of sodium dithionite were added to reduce the iron and the spectrum recorded, then CO was gently bubbled through the cell and the spectrum was recorded. The extinction coefficient was determined using $\epsilon_{450} = 91 \text{ mM}^{-1} \text{ cm}^{-1}$ for the reduced CO-bound form. For CYP enzymes that did not fully shift to 450 nm, the concentration of the heme was determined by the pyridine hemochromagen assay as reported by Barr and Guo.²³

Ferredoxin purification

The pETDuet vectors with the incorporated genes for Fdx2 (Thr; Mmar_2879), Fdx3 (Tyr; Mmar_2932), Fdx4 (Asn; Mmar_3973), Fdx5 (His; Mmar_4716), Fdx8 (His; Mmar_4736), Fdx9 (Ser; Mmar_4763) were transformed into competent *E. coli* BL21(DE3) and plated onto LB_{amp} plates and left overnight at 37 °C. A colony was picked and grown in 5 mL LB_{amp} at 37 °C overnight and this starter culture was then added across 2 L of 2xYT_{amp} (4 x 500 mL). After growing at 37°C for 5 hours at 150 rpm the incubator temperature was lowered to 25 °C, followed by the addition of benzyl alcohol (0.02% v/v) and ethanol (2% v/v) and then IPTG (0.1 mM) after a further 30 min. The cultures were incubated at this temperature for 16 hours with shaking at 120 rpm. The cells were harvested by centrifugation (5,000 g, 15 min) and resuspended in 200 mL of 50 mM Tris (pH 7.4) with DTT (1 mM) plus 30 mL glycerol, 3 mL β-mercaptoethanol and 1 mL TWEEN. Lysozyme (300 mg) was added and the resuspended cells were stirred on ice for 30 min, before sonication. The supernatant, containing the desired protein, was isolated from cell debris by centrifugation (40,000 g, 30 min). The protein was then loaded onto a DEAE Sepharose column, (XK50, 200 mm x 40 mm, GE Healthcare) and eluted using a linear salt gradient of 150 mM to 400 mM KCl in Buffer T. The fractions containing the desired protein (identified by brown colour of the Fdx enzyme) were combined and the volume reduced using a Vivacell 100 (Sartorius Stedim, 3 kD membrane) aided by centrifugation (1,500 g). The protein was loaded onto a His-trap column (GE Healthcare) equilibrated with Buffer T plus 300 mM NaCl and 20 mM imidazole. The column was washed with a further 5 column volumes before elution of the His-tagged protein with Buffer T containing 300 mM NaCl and 300 mM imidazole. The protein was concentrated via ultrafiltration and an equal volume of 80% glycerol was added before filtration through a 0.22 µm syringe filter and storage at -20 °C.

Spin state shift assays and binding constant determination

Glycerol in stored protein samples was removed via buffer exchange into Buffer T using a PD-10 desalting column (GE Healthcare). The P450 was diluted to approximately 1 - 2 μM using the same buffer and the spectrum was recorded on a UV/Vis spectrophotometer. Aliquots (1 to 10 μL) of substrate stock solutions in DMSO or EtOH (50 mM - 100 mM) were added and the spectrum recorded until the shift from 420 nm to 390 nm reached a stable point. The ratio of high-spin P450 to low-spin P450 (390 nm peak to 420 nm peak) was estimated (to $\pm 5\%$) by comparison to camphor-bound P450_{cam} spectra.

For binding assays varying aliquots (1 to 3 μL) of substrate stock solutions (1 mM, 10 mM or 100 mM concentrations) were added to 2.5 mL of protein (0.5 - 1.5 μM). The sample was mixed and the absorbance difference spectrum was recorded between 300 nm and 600 nm on a UV/Vis spectrophotometer. Further substrate was added until no difference in peak-to-trough ratio at 420 nm and 390 nm was observed. The difference in absorbance versus substrate concentration was fitted to the hyperbolic function (Equation 1):

$$\Delta A = \frac{\Delta A_{max} \times [S]}{K_d + [S]}$$

where K_d is the binding constant, $[S]$ is the substrate concentration, ΔA the peak-to-trough ratio, and ΔA_{max} the maximum peak-to-trough absorbance. Miconazole binding to CYP269A1 exhibited tight binding, with $K_d < 5 \mu\text{M}$ and the data were fitted to the tight binding quadratic equation:²⁴

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

where ΔA_{max} is the maximum absorbance difference and $[E]$ is the enzyme concentration.

Whole-cell oxidation turnovers

The appropriate DNA vectors pETDuetFdx/FdR1 and pRSFDuetFdx/CYP were mixed with competent BL21(DE3) *E. coli* cells, and grown on LB_{amp/kan} plates for 16 hours at 37 °C. Colonies were inoculated into 2 mL LB_{amp/kan} and grown at 37 °C for 4 hours at 200 rpm. This starter culture was then added to 200 mL 2xYT_{amp/kan} and grown at 120 rpm and 37 °C for a further 4 hours. The culture was cooled to room temperature, IPTG was added (to a final concentration of 1 mM) to induce protein expression and the growth was incubated overnight at 120 rpm. The cells were harvested via centrifugation (5,000 g, 10 min). The resulting cell pellet was resuspended in EMM (200 mL, 2% DMSO), added to a baffled flask for increased aeration and the substrate was then added before shaking at 150 rpm at 30 °C.

The ability of CYP147G1 to oxidise indole (0.1 mM) to indigo was qualitatively determined colorimetrically via the formation of an intense blue colour when compared to the pET26CYP147G1 cultures as a control (Fig. S4). The cell pellet of this control was red in colour with no indication of any indigo formation in the cell pellet or the supernatant (Fig. S4).

After 16 hours the turnovers with other substrates were centrifuged (15 min, 5,000 g) and the supernatant isolated. Samples (1 mL) of the turnover were taken for initial testing and extracted with ethyl acetate (400 µL). For fatty acid substrates 3 M HCl (20 µL) was added to turnovers and the samples were extracted into ethyl acetate, dried over MgSO₄ before resuspension in anhydrous acetonitrile (200 µL). The acid samples were derivatised with TMCS/BSTFA before analysis by GC-MS. For other substrates the samples were extracted using ethyl acetate and the organic layer used directly for analysis.

For larger scale extractions 200 mL of the supernatant was acidified with 3 M HCl to pH 2, extracted three times with an equal volume of ethyl acetate. Extracts were washed with water and

saturated brine solutions, combined and dried over MgSO₄ and the solvent was removed under reduced pressure.

To generate sufficient fatty acid product for characterisation the whole-cell oxidation system (200 mL) was used to convert 1 mM of the fatty acid substrates to product. The sample was extracted as described previously. Where GC-MS indicated a product:substrate ratio of $\geq 95\%$ reverse phase solid phase extraction (SPE) DSC-18 SPE tubes (Supelco) were used to isolate all fatty acid compounds, using the method described by Horak *et al* with minor modifications.²⁵ SPE columns were activated with methanol (3 mL), washed with water (5 mL) and the extract was dissolved in methanol and water (200 μ L) and loaded onto the column. The column was washed with 5 mL 10% v/v methanol solution followed by 5 mL of 20% v/v methanol:water. The acid products were eluted with 600 μ L methanol and the elution was dried under a flow of nitrogen and dissolved in deuterated chloroform, 0.8 mL, before characterisation. Full product characterisation was performed by NMR. ¹H NMR and 2D COSY spectra were recorded at 500 MHz (Agilent DD2 500MHz NMR).

Product analysis

The activity of the mutant ferredoxins of Fdx3 with CYP147G1 was tested *in vivo* by performing turnovers in triplicate according to the procedure outlined above using 1 mM undecanoic acid as the substrate. Comparison of product formation by the wild type Fdx3 to the mutant Tyr→Cys and Tyr→Gly was analysed by GC-MS as before except that an internal standard (octanoic acid) was added to the samples (final concentration 0.5 mM) before extraction. The product yield was compared using the ratio of the area of the product peak versus that of the internal standard.

GC-MS analysis was used for enzyme turnover analysis and was performed on a Shimadzu GC-17A with a DB-5 MS fused silica column (30 m x 0.25 mm, 0.25 μ m) and a QP5050A GC-MS detector. The injector was held at 250 °C and the interface at 280 °C. For fatty acid substrates, the

initial oven temperature was 120 °C which was held for 3 min, before increasing to 220 °C at 7 °C per minute, where it was held for 7 min. The total ion count was monitored and the MS fragmentation pattern was recorded. For β -ionone, the initial oven temperature was 120 °C which was held for 3 min, before increasing to 220 °C at 10 °C per minute, where it was held for 7 min. The total ion count was monitored and the MS fragmentation pattern was recorded. Analytical liquid chromatography was performed using an Agilent 1260 Infinity pump equipped with an Agilent Eclipse Plus C18 column (250 mm x 4.6 mm, 5 μ m), an autoinjector and UV detector. A gradient, 20 – 95%, of acetonitrile (with trifluoroacetic acid, 0.1%) in water (TFA, 0.1%) was used.

Protein mass spectrometry

Protein mass measurements were carried out under denaturing conditions using an Agilent 6560 ion mobility quadrupole time-of-flight instrument with Dual AJS electrospray ionisation source, coupled to an Agilent 1290 Infinity II LC System. The protein was buffer exchanged into 250 mM ammonium acetate, concentrated to ~10 mM, then diluted 1:1 with acetonitrile. Protein sample (3 μ L) was injected and electrosprayed using 50% aqueous acetonitrile/0.01% formic acid at a flow rate of 0.1 mL min⁻¹, without chromatographic separation. ESI-MS conditions were: positive-ion mode; capillary voltage, 3500 V; nozzle voltage 1000 V; fragmentor, 400 V; gas 8 L min⁻¹; gas temperature, 300 °C; sheath gas 11 L min⁻¹; and sheath gas temperature, 350 °C. Spectra were deconvoluted using BioConfirm software (Agilent).

Characterisation of ferredoxins after anaerobic purification

Protein was generated, using the same plasmid vectors as described above, from *E. coli* BL21(DE3) cells as follows. Culture was grown in LB to OD₆₀₀ of 0.8 at 37 degrees 200 rpm shaking and cold shocked on ice for 18 minutes prior to induction with 50 μ M IPTG. Post induction the cultures were supplemented with 20 μ M ammonium ferric citrate and grown for a further 20 hours at 25 °C, 90 rpm shaking, to a final OD₆₀₀ ~1.5.

After harvesting the pellets were re-suspended in anaerobic buffer (20 mM HEPES, 100 mM NaCl, 20 mM imidazole, pH 7.4), ruptured by sonication and centrifuged under nitrogen in gas tight tubes to remove debris. All subsequent steps were carried out anaerobically ($[O_2] < 10$ ppm). Supernatant was loaded onto a Ni^{2+} charged IMAC column equilibrated with the buffer above and eluted using a 50 mL gradient from 0 to 100% buffer containing 500 mM imidazole. The sample was then exchanged into imidazole free buffer containing 1.5 M NaCl as cryoprotectant using a PD-10 column and stored in an anaerobic chamber at -5 °C prior to use.

In order to discount stripping of ferrous iron from the sample during IMAC purification a duplicate sample of Fdx4 was produced using the identical expression protocol as above. Pellets were re-suspended in 20 mM HEPES, 50 mM NaCl, pH 7.4. Following sonication and centrifugation the sample was loaded on to a 5 ml FF Hi-Trap Q-sepharose column (GE Healthcare) and eluted using 20 mM HEPES, 500 mM NaCl, pH 7.4. The cluster containing fractions were pooled, concentrated and loaded on to a Sephacryl S-100 size exclusion column equilibrated with 20 mM HEPES, 500 mM, NaCl pH 7.4 and eluted using this buffer. Sample purity as judged from SDS-PAGE was as least equivalent to samples purified using an IMAC column (Fig. S9). Sample purified in this way gave an almost identical CD spectrum demonstrating that Fdx4 binds a $[3Fe-4S]$ cluster regardless of the purification method used. The minor differences in the CD spectra obtained following the two preparation methods are consistent with slightly altered ratios of oxidised to reduced cluster in the samples obtained (Fig. S13).

Native- and LC-MS spectra of Fdx4, Fdx5 and Fdx9 were recorded on a Bruker micrOTOF-QIII mass spectrometer (Bruker Daltonics) equipped with an UltiMate 3000 HPLC system (Dionex). Samples for native MS were de-salted by dilution with an equal volume of 50 mM ammonium acetate pH 7.4 followed by exchange into the same buffer using a PD-10 column. Desalted samples were infused directly into the ionisation chamber at a flow rate of $300 \mu\text{L min}^{-1}$. For LC-MS $50 \mu\text{L}$ samples were diluted with $450 \mu\text{L}$ of 2% acetonitrile, 0.1% formic acid solution.

1 μL of the diluted sample was then injected onto a ProSwift reversed phase RF-1S column at 25 °C and eluted using a gradient of 2 – 100% acetonitrile, 0.1% formic acid with a flow rate of 0.2 mL min^{-1} (15 min). Instrument parameters were: dry gas flow 8 L min^{-1} , nebuliser gas pressure 1.8 bar, dry gas temperature 240 °C and capillary voltage 4500 V (LC-MS) or dry gas flow 4 L min^{-1} , nebuliser gas pressure 0.8 bar, dry gas temperature 180 °C and capillary voltage 3000 V (native MS). Collision cell energy was 5.0 eV in all cases. Neutral mass spectra were calculated by maximum entropy deconvolution over the mass range 6 - 9 kDa using ESI Compass version 1.3 (Bruker Daltonics). The mass of the bound cluster was calculated from the difference in mass between the apo protein (deduced from LC-MS) and the cluster bound form (deduced from native MS) and found to be 296 Da (corresponding to the predicted mass of a [3Fe-4S] centre) in all cases.

Redox activities of the [3Fe-4S] clusters of Fdx4, Fdx5 and Fdx9 were probed spectroscopically following anaerobic chemical poisoning. Electronic absorbance spectra were recorded on a Jasco V550 spectrophotometer and CD spectra on a Jasco J-810 spectropolarimeter. Sample concentration was estimated assuming an extinction coefficient $\epsilon_{410\text{nm}} = 9000 \text{ M}^{-1}\text{cm}^{-1}$. 250 μL samples of ferredoxin in 1 mm pathlength anaerobic cuvettes were equilibrated with either: 5 μL of an approximately 100 mM EuCl_2 solution in 20 mM HEPES, pH 7.4 containing 1.5 M NaCl (reduction), an appropriate volume of a 14 mM $\text{K}_3\text{Fe}(\text{CN})_6$ solution in water to provide a 1:1 ratio with estimated cluster concentration (oxidation) or a 1.5 mM solution of sodium ascorbate (to set a defined solution potential of +60 mV).²⁶ Following characterisation by absorbance and CD spectroscopy, samples of as isolated and chemically poised proteins were loaded into EPR tubes and flash frozen by plunging into liquid N_2 . EPR spectra were recorded at 10 K using a Bruker EMX (X-band) EPR spectrometer equipped with an Oxford Instruments liquid helium system and a spherical high-quality ER 4122 SP 9703 resonator (Bruker). Instrument parameters were as follows: microwave frequency $\nu_{\text{MW}} = 9.4657 \text{ GHz}$, modulation frequency $\nu_{\text{M}} = 100 \text{ kHz}$, time constant $\tau = 82 \text{ ms}$, microwave power = 3.19 mW, modulation amplitude $A_{\text{M}} = 5 \text{ G}$, scan rate $\nu = 22.6 \text{ G s}^{-1}$.

Estimated concentrations of [3Fe-4S] clusters prior to addition of chemical reagents were as follows; 180 μM (Fdx4 and Fdx9) and 330 μM (Fdx5). An equivalent sample of Fdx4 was anaerobically incubated with 180 μM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, subsequent characterisation (CD spectroscopy and redox activity) showed no evidence of incorporation of metal ion to yield a [4Fe-4S] centre.

Gene fragments and oligonucleotides used in this work.

Sequences of codon optimised ferredoxin genes; NcoI sites and HindIII sites were incorporated at the 5' and 3' ends respectively. The restriction sites NcoI and HindIII are underlined, start and stop codons highlighted in bold. Note that an additional codon GTG encoding a valine has been added to the N-terminus to allow for the incorporation of the NcoI site. A double stop codon was incorporated at the C-terminus. In the mutants the modified triplet codon and the amino acid are highlighted in red.

Fdx3 (MMar_2932)

ttaattccatggtgcgtctggtggtgatttaacaaatgccaaaggctacgctcagtcgctaccactggcaccggaagtttcaagctggttggtg
aggaagctctggcttatgatccgaaccggacgactctcagcgtcagcgtgtactgcgcgcggtagcatcctgtccggtcaagcaattattctg
gaagtagaccgccggccgatcgcgacactaaat**aatagaagct**taattaat

The amino acid sequence of Fdx3 with residues important in the ferredoxin binding motif highlighted in bold and underlined

MVRLVVDLNKCQGYAQCVPLAPEVFKLVGEEALAYDPNPDDSQRQRVLRAVASCPVQAI
ILEVDPPADRDTK

Fdx3 Tyr-Gly mutant

ttaattccatggtgcgtctggtggtgatttaacaaatgccaaaggc**ggc**gctcagtcgctaccactggcaccggaagtttcaagctggttggt
gaggaagctctggcttatgatccgaaccggacgactctcagcgtcagcgtgtactgcgcgcggtagcatcctgtccggtcaagcaattattct
ggaagtagaccgccggccgatcgcgacactaaat**aatagaagct**taattaat

The amino acid sequence of Fdx3 with residues important in the ferredoxin binding motif highlighted in bold and underlined

MVRLVVDLNKCQG**G**AQCVPLAPEVFKLVGEEALAYDPNPDDSQRQRVLRAVASCPVQAI
ILEVDPPADRDTK

Fdx3 Tyr-Cys mutant

ttaatt**catg**gtgcgtctggtggtgatttaacaaatgccaaggc**tcg**ctcagtcgctaccactggcaccggaagtttcaagctggtggtg
aggaagctctggcttatgatccgaaccggacgactctcagcgtcagcgtgtactgcgcgcggtagcatcctgtccggtcaagcaattattctg
gaagtagaccgccgcccgatcgcgacactaaat**aatagaagct**taattaat

The amino acid sequence of Fdx3 with residues important in the ferredoxin binding motif highlighted in bold and underlined

MVRLVVDLNL**C****Q****G****C**A**Q****C**VPLAPEVFKLVGEEALAYDPNPDDSQRQRLRAVAS**C****P****V****Q****A****I**
ILEVDPPADRDTK

Fdx2 (MMar_2879)

ttaatt**catg**gtccgcgtggctgcccaccgcgagatctgtatggccaccggcatgtgtgtgatgaccgctgacgcattcttcgaccaggacgc
cgacggcattgtgtgctggctgcgcacgaagtccggcggacgaagagcgtagagttcgtaatgcagtgaaactgtcccgtccggcgcct
ggaactgatgtccgatt**aatagaagct**taattaat

The amino acid sequence of Fdx2 with residues important in the ferredoxin binding motif highlighted in bold and underlined

MVRVAADREI**C****M****A****T****G****M****C****V****M****T****A****D****A****F****D****Q****D****A****D****G****I****V****V****L****A****A****H****E****V****P****A****D****E****R****R****V****R****N****A****V****K****L****C****P****S****G**
ALELMSD

Fdx4 (MMar_3973)

ttaatt**catg**gttcgcgttatcgttgacgaaactctgtgcgaggcgaacggttctgtgaaagtctggcaccagacatcttctgctctgggcgatg
ctgatgtagttcagatcgtgacggcccggttcctgccgaccgccagatcagcgtgcgtgccgctggtgatcagtgcccgaaggccgcctgc
gtctgatcag**taatagaagct**taattaat

The amino acid sequence of Fdx4 with residues important in the ferredoxin binding motif highlighted in bold and underlined

MVRVIVDETLCEANGFCESLAPDIFALGDADVQIADGPVPADRQIDVRAAVDQCPKAAL
RLIE

Fdx5 (MMar_4716)

ttaattcatggtaaaagttgggtgatccgcagcgtgtcaaggtcacaccctgtgtgctatgatcgcgccggacagctccagctctctgaca
tcgacggttctagctcggcgatcagcgaaactgtccggctgatcagtgggacctggtgcgtgaagcggcgcatagctgtccggagcagggc
atcgatcactatagaagcttaattaat

The amino acid sequence of Fdx5 with residues important in the ferredoxin binding motif highlighted in bold and underlined

MVKVWVDPQRCQGHTLCAMIAPDSFQLSDIDGSSSAISETVPADQWDLVREAAHSCPEQA
IVITDET

Fdx8 (MMar_4736)

ttaattcatggtcaaagtagctgtgacgatcagcgttgcccgccacggatgtgcctgacctgtgtccagaagtgtctctctgacggac
gatggttacgcagtggtatcactagcgacgtaccgatggaactggaagaggctgtgcgtgaagcgatccagtgtgcccggagcagggcat
ctccgaatcttatagaagcttaattaat

The amino acid sequence of Fdx8 with residues important in the ferredoxin binding motif highlighted in bold and underlined

MVKVRVDDQRCRGHGMCLTLCPVFLTDDGYAVAITSDVPMELEEAVREAIQCPEQAI
SES

Fdx9 (Mmar_4763)

ttaattcatggtgaaagtgatcgtagatgaaaatatctgcgcgtccagcggcaactgtgtgatgaatgcgccggaatttcgaccagcgcga
cgaggacggcgtggtagtgtgcttaacgaaatccgccagcggaaactgccgaaggtgcccgcgtgtgctgcttcttgcggccctgg
caattaaaatcgaggagtatagaagcttaattaat

The amino acid sequence of Fdx9 with residues important in the ferredoxin binding motif highlighted in bold and underlined

MVKVIVDENIC**ASSGNC**VMNAPEIFDQRDEDGVVLLNANPPAELAEGARRAAAS**CP**ALA
IKIEE

The primers used to insert a 6xHis Tag at the C-terminus of the ferredoxins. The NcoI and HindIII restriction sites and the sequence of the 6xHis are underlined and the start and stop codons are highlighted in bold.

Mmar_2879 5' tttctatccatggtccgcgtggctgc

Mmar_2879 3' attaattaagctt**ctatta**atgatggtggtgatgatgatcggacatcagttccagggc

Mmar_2932 5' tttctatccatggtcgctctggtggtg

Mmar_2932 3' attaattaagctt**ctatta**atgatggtggtgatgatgtagtgcgcgatcggccgg

Mmar_3973 5' tttctatccatggttcgcgttatcgttgac

Mmar_3973 3' attaattaagctt**ctatta**atgatggtggtgatgatgctcgcacagacgcagggcg

Mmar_4716 5' tttctatccatggtaaaagtgtgggtggatc

Mmar_4716 3' attaattaagctt**ctatta**atgatggtggtgatgatgggttcatcgggtgatgacgac

Mmar_4736 5' tttctatccatggtcaaagtacgtgttgac

Mmar_4736 3' attaattaagctt**ctatta**atgatggtggtgatgatgagattcggagatggcctgctc

Mmar_4763 5' tttctatccatggtgaaagtgatcgtatgatg

Mmar_4763 3' attaattaagctt**ctatta**atgatggtggtgatgatgctcctcgattttaattgccaggg

The primers used to clone the cytochrome P450 genes of CYP278A1 (*Mmar_2877*), CYP147G1 (*Mmar_2930*), CYP269A1 (*Mmar_3969*), CYP150A5 (*Mmar_4737*), CYP105Q4 and the ferredoxin reductase gene (*Mmar_2931*). The restriction sites are underlined and the start and stop codons are highlighted in bold. An additional KpnI site was added to all the genes to allow incorporation in the Duet vectors for the whole-cell oxidation system with the exception of CYP150A5 where the HindIII site at the 3' end was replaced with XhoI.

CYP278A1 NdeI 5' ttaattcatatgtcaacagagaccgtttcagg

CYP278A1 KpnI 3' ttaattaagcttggtaccctattatgacaggtgcaggggtagc

CYP147G1 NdeI 5' ttaattcatatgaatgccgaaaccgcttgggc

CYP147G1 KpnI 3' ttaattaagcttggtaccctattattcggtgatcgctgcgaaatc

FdR2931 NdeI 5' ttaattcatatgaaccggggtcgttggtcg

FdR2931 KpnI 3' ttaattaagcttggtaccctattagcctcggcggggccggaattc

CYP269A1 NdeI 5' ttaattcatatggcctatcctgaaaccaatac

CYP269A1 KpnI 3' ttaattaagcttggtaccctattaccaacgcactggcagcgacag

CYP150A5 NdeI 5' ttaattcatatgaatgattgtgccgagccgg

CYP150A5 XhoI 3' ttaattctcgagctattatgcgggctgaaatccaaatg

CYP105Q4 NdeI 5' ttaattcatatgtccgacacgctcgcaagcc

CYP105Q4 KpnI 3' ttaattaagcttggtaccctattaccggtcacgggtagttcatag

Table S1 Chromosome features of *M. marinum* M compared with three other *Mycobacteria*.

	<i>M. marinum</i>	<i>M. ulcerans</i>	<i>M. tuberculosis</i>	<i>M. leprae</i>
	<i>M</i>	<i>Agy99</i>	<i>H37Rv</i>	<i>TN</i>
Size of Chromosome (bp)	6,636,827	5,631,606	4,411,532	3,268,203
Coding sequences	5424	4160	3974	1605
Pseudogenes	65	771	17	1115
Number of P450 genes	47	24	20	1
Associated Fdx genes	11	6	2	0
Associated FdR genes	2	0	0	0

Table S2 The CYPome of *M. marinum*. The gene and P450 name are assigned as per the databases at the National Center for Biotechnology Information. The sequences of conserved regions of the I-helix, K-helix and the heme binding motif as well as the predicted pI and length of the amino acid chain are also provided.

Gene name	Accession number	ID	I-helix	K-helix	Heme binding motif	pI	AA
Mmar_0122	ACC38592.1	CYP279A2	GTDTT	ETMR	IQT F GAGMHY C LG	4.9	411
Mmar_0272	ACC38739.1	CYP226B1	ATDTT	EGER	HAT F GFGTHI C SG	5.1	422
Mmar_0274	ACC38741.1	CYP271A1	GLDTV	ELMR	HLA F GSGIHR C LG	5.9	430
Mmar_0281	ACC38748.1	CYP183B1	GTETT	ETLR	YIP E GGGARK C IG	10.5	461
Mmar_0283	ACC38750.1	CYP274A1	ATETS	ETLR	FIP F GMGKHK C IG	9.6	450
Mmar_0346	ACC38813.1	CYP138A3	GHETT	EVQR	WIP F GGGTRR C VG	10.2	440
Mmar_0399	ACC38866.1	CYP191A3	GTETV	EMIR	SLA F GRGQH F CIG	5.2	401
Mmar_0852	ACC39310.1	CYP185A4	GEDTT	EAMR	YLP F GGGGRS C LG	8.8	473
Mmar_0928	ACC39385.1	CYP189A6	GNETT	ETLR	HLT F GKGVHY C LG	5.8	405
Mmar_0938	ACC39395.1	CYP135B4	GHDTT	ETLR	WLP F GGGNRR C LG	9.8	462
Mmar_1564	ACC40016.1	CYP276A1	AH G TT	ESLR	AVM F GAGIH Y CLG	7.6	410
Mmar_1634	ACC40084.1	CYP136A2	AHDTS	ESIR	FTP F GGGAH K CLG	6.6	491
Mmar_2475	ACC40925.1	CYP139A3	GYETT	ETLR	FIP F SGLLHR C IG	9.8	432
Mmar_2631	ACC41077.1	CYP143A3	GLDTV	EIVR	HWG F GGGTHR C LG	5.1	389
Mmar_2654	ACC41098.1	CYP144A4	GGE S T	ETLR	HIS F GKGAH F CVG	4.8	403
Mmar_2666	ACC41109.1	CYP143A4	GLDTV	EIVR	HWG F GGGPHR C LG	6.1	390
Mmar_2768	ACC41210.1	CYP140A5	GFETT	EILR	HLA F STGRH F CLG	8.4	437
Mmar_2783	ACC41225.1	CYP125A6	GNETT	EIVR	VG F GGTGAH Y CIG	4.8	427
Mmar_2877	ACC41317.1	CYP278A1	GSETT	ETLR	HLS L GHGLH F CLG	4.8	426
Mmar_2930	ACC41369.1	CYP147G1	GHD S T	EVQR	H F GW S GSIHT C MG	5.0	421
Mmar_2978	ACC41416.1	CYP135B6	GYDTS	ETLR	WLP F GGGARR C LG	10.3	472
Mmar_3135	ACC41569.1	CYP136B2	AHDTS	EALR	WVP F GGGAH K CIG	7.2	484
Mmar_3154	ACC41588.1	CYP153A16	GNDTT	EIIR	HIS F GFGVHR C MG	6.0	462
Mmar_3361	ACC41787.1	CYP124A1	GNETT	EIVR	VG F GGGGAH F CLG	4.9	432
Mmar_3761	ACC42177.1	CYP268A2	GNDTT	ELVR	VG F GGGGVH F CLG	5.1	418
Mmar_3969	ACC42377.1	CYP269A1	GID S T	EVLR	HVS F GHGR F LCPG	4.8	402
Mmar_3976	ACC42384.1	CYP138A4	GHETT	EVQR	WIP F GGGTRR C IG	9.1	441
Mmar_3996	ACC42404.1	CYP187A4	GLETT	EGLR	HIA F AGGIH M CLG	4.8	407
Mmar_3999	ACC42407.1	CYP108B4	GHDTT	EMIR	VA F GYGVH F CMG	4.9	409
Mmar_4008	ACC42416.1	CYP187A5	GLETT	EGLR	HIS F AAGEHT C LG	6.3	403
Mmar_4184	ACC42591.1	CYP130A4	GNDTV	ELLR	ILT F SHGAH H CLG	5.7	412
Mmar_4430	ACC42837.1	CYP138B1	GHETT	EVQR	WIP F GGGIHR C IG	8.4	455
Mmar_4483	ACC42889.1	CYP135B3	GHD N T	ETLR	WLP F GGGSRR C LG	10.1	456
Mmar_4694	ACC43098.1	CYP150A6	GQETT	ESLR	HMA F ARGVH S CPG	5.0	423
Mmar_4717	ACC43121.1	CYP188A3	GFDTT	EFLR	HFS F GIGVHR C IG	4.9	453
Mmar_4733	ACC43137.1	CYP190A3	GAETV	ELLR	NTL G FGYGIH S CLG	4.8	399
Mmar_4737	ACC43141.1	CYP150A5	GQETT	EALR	HLS F GRGIH S CPG	4.9	422
Mmar_4753	ACC43157.1	CYP189A7	GNETT	ELLR	HLT F SVGTH Y CLG	4.8	399
Mmar_4762	ACC43166.1	CYP105Q5	GHETT	ELLR	NVA F GYGRH Q CVG	5.6	413
Mmar_4833	ACC43238.1	CYP123B1	GHETT	ELLR	VA F GRGIH F CLG	4.8	402
Mmar_4915	ACC43319.1	CYP126A3	GAETT	EMVR	LG F GQGVH Y CLG	4.8	417
Mmar_4930	ACC43334.1	CYP123A3	GNETT	ETLR	LS F GSGAH F CLG	5.0	405
Mmar_4932	ACC43336.1	CYP51B1	GH H TS	ETLR	WIP F GAGRHR C VG	5.4	455
Mmar_5002	ACC43406.1	CYP142A3	GDETT	EMLR	LA F GFGTH F CMG	4.5	400
Mmar_5032	ACC43436.1	CYP125A7	GNETT	EIVR	VG F GGTGAH Y CIG	4.6	416
Mmar_5175	ACC43581.1	CYP137A2	GHETT	ETLR	WVP F GGGAKR C LG	10.4	455
Mmar_5268	ACC43672.1	CYP164A3	GHETT	ETMR	HLG F GRGAH Y CLG	4.7	441

Table S3 Analysis of the CYPomes of *M. marinum*, *M. ulcerans* and *M. tuberculosis*. The subfamily name of the *M. marinum* gene is given. If a pseudogene is present in *M. ulcerans* it has been highlighted in red and underlined. Ferredoxin (Fdx) and ferredoxin reductase (FdR) genes that are associated with CYP genes are highlighted in blue and italics. Neither of the ferredoxin reductase genes in *M. marinum* are present in *M. ulcerans* or *M. tuberculosis*. (A pseudogene is a dysfunctional relative of a gene which contains stop codon, frame shifts or insertions). Fdx11, which is not associated with a CYP gene, is conserved in all three bacterium.

	M. marinum only	M. marinum and M. ulcerans	Conserved in all three	M. marinum and M. tuberculosis
CYP123B1	<u>CYP164A3</u>	CYP105Q4 Fdx9	CYP51B1 Fdx10	CYP135B4
<u>CYP125A6</u>	CYP183B1	CYP108B4	CYP123A3	<u>CYP137A2</u>
CYP135B3	<u>CYP189A6</u>	CYP143A3	CYP124A1	<u>CYP139A3</u>
CYP135B6	CYP190A3Fdx6/Fdx7	CYP150A6	CYP125A7	
CYP136B2	<u>CYP226B1</u>	CYP185A4	CYP126A3	M. tuberculosis only
<u>CYP138A4</u>	<u>CYP268A2</u>	CYP187A4	CYP130A4	CYP121A1
CYP138B1	CYP271A1	CYP187A5	CYP136A2	CYP128A1
CYP147G1	<u>CYP274A1</u>	CYP188A3 Fdx5	CYP138A3	CYP132A1
Fdx3/FdR1	CYP276A1	CYP189A7	CYP140A5	CYP135A1
CYP150A5 Fdx8	<u>CYP278A1</u> Fdx2	CYP191A3	CYP142A3	CYP141A1
CYP153A16n [2Fe-2S]/FdR2	CYP279A2	CYP269A1 Fdx4	CYP143A4 Fdx1 CYP144A4	
		Fdx11		M. ulcerans only CYP140A7

Table S4 The genes encoding the [3/4Fe-4S] ferredoxins of *M. ulcerans* and *M. tuberculosis* which have equivalents in *M. marinum*. The sequences of iron sulphur cluster binding motif of the [3/4Fe-4S] ferredoxins as well as the predicted pI and length of the amino acid chain are provided.

Gene name	Accession		<i>Mycobacterium ulcerans</i>		
	number	ID	Iron Sulfur cluster binding motif	pI	AA
Mul_0316	ABL03046.1	Mul_1	CXXHXXC(X) _n CP	3.7	65
Mul_0334	ABL03060.1	Mul_2	CXXSXXC(X) _n CP	3.9	63
Mul_0472	ABL03175.1	Mul_3	CXXHXXC(X) _n CP	4.4	67
Mul_2873	ABL05156.1	Mul_4	CXXTXXC(X) _n CP	4.0	63
Mul_3090	ABL05334.1	Mul_5	CXXHXXC(X) _n CP	3.9	63
Mul_3830	ABL05923.1	Mul_6	CXXNXXC(X) _n CP	3.6	62
Mul_4066	ABL06117.1	Mul_7	CXXNXXC(X) _n CP	4.0	81
<i>Mycobacterium tuberculosis</i>					
Gene name		ID	Iron Sulfur cluster binding motif	pI	AA
Rv0763c	CCP43510.1	-	CXXHXXC(X) _n CP	4.6	67
Rv1786	CCP44552.1	-	CXXHXXC(X) _n CP	3.7	66
Rv3503c	CCP46325.1	-	CXXNXXC(X) _n CP	3.6	62

Table S5 The genes encoding the other potential electron transfer proteins of *M. marinum* which are not closely associated with CYP enzyme genes. The gene name as per the databases at the National Center for Biotechnology Information is provided. The predicted pI and length of the amino acid chain are provided as are the names of the equivalent genes in *M. ulcerans* and *M. tuberculosis*.^a Mmar_2994 is 16 genes away from CYP135B6^b

Mmar_5043 is 11 genes away from CYP125A7. No others are within 30 genes of members of the CYPome of *M. marinum*. While this does not rule out their ability to function as electron transfer partners for a P450 enzyme they could also be involved in other metabolic processes which require these types of proteins. We also cannot rule out that they may have evolved to support P450 electron transfer in these other bacteria.

Mycobacterium Marinum						
Gene name	ID	Iron Sulfur cluster binding motif	pI	AA	M. ul	M. tb
Mmar_2080 ACC40530.1	FdxA	7Fe Ferredoxin	3.9	114	Mul_3264 ABL05469.1	Rv2007c CCP44779.1
Mmar_2994 ^a ACC41432.1	FdxA	7Fe Ferredoxin	4.1	113	-	-
Mmar_3421 ACC41846.1	FdxC	7Fe Ferredoxin	3.8	107	Mul_2700 ABL05017.1	-
Mmar_4274 ACC42683.1	FdxC	7Fe Ferredoxin	3.4	107	Mul_1025 ABL03633.1	Rv1177 CCP43933.1
Mmar_4794 ACC43198.1	2Fe-2S	2Fe-2S	3.9	93	Mul_0363 ABL03085.1	-
Mmar_1017 ACC39474.1	FdR3	FdR3	4.6	411	Mul_0769 ABL03414.1	Rv0688 CCP43431.1
Mmar_1526 ACC39977.1	FprA	FprA	5.1	455	Mul_2413 ABL04766.1	-
Mmar_3420 ACC41845.1	FprA	FprA	5.3	453	Mul_2699 ABL05016.1	Rv3106 CCP45916.1
Mmar_4646 ACC43049.1	FprB	FprB	7.1	560	Mul_0264 ABL02999.1	Rv0886 CCP43634.1
Mmar_5043 ^b ACC43447.1	FdxB	FdxB	5.9	673	Mul_4117 ABL06161.1	Rv3554 CCP46376.1

Table S6 A list of other ferredoxin genes from strains of *Mycobacterium* which share the same ferredoxin cluster binding motif as those from *M. marinum*. Note the list is not exhaustive and more than one ferredoxin with the stated motif may be present in the strain listed. Ferredoxin genes with the CXXTXXC(X)_nCP motif were only found in a few strains closely related to *M. marinum* e.g. *M. ulcerans* and *M. liflandii*.

^a may be a pseudogene. ^b a ferredoxin reductase gene is also located by the ferredoxin. Entries in red have a CYP enzyme clustered with the ferredoxin and those in blue have a CYP gene close by.

cluster binding motif				
CXXHXXC(X) _n CP	CXXNXXC(X) _n CP	CXXYXXC(X) _n CP ^b	CXXFXXC(X) _n CP	CXXSXXC(X) _n CP
<i>M. simiae</i>	<i>M. xenopi</i>	<i>M. gastri</i>	<i>M. kansasii</i>	<i>M. asiaticum</i>
WP_044508369.1	WP_039890166.1	WP_036416246.1	WP_023367649.1	WP_036365249.1
<i>M. colombiense</i>	<i>M. heckeshornense</i>	<i>M. kansasii</i>	<i>M. gastri</i>	<i>M. lentiflavum</i>
WP_007771429.1	WP_048893299.1	WP_023364435.1	WP_036414200.1	CQD20859.1
<i>M. kyorinense</i>	<i>M. intracellulare</i>	<i>M. aromaticivorans</i>	<i>M. asiaticum</i>	<i>M. xenopi</i>
WP_045384657.1	WP_014382133.1	WP_036346035.1	WP_036365153.1	WP_039891214.1
<i>M. kansasii</i>	<i>M. lentiflavum</i>	<i>M. rhodesiae</i>	<i>M. parascrofulaceum</i>	<i>M. simiae</i>
WP_023367619.1	CQD18585.1	WP_005140322.1	WP_007168606.1	WP_044512409.1
<i>M. lentiflavum</i>	<i>M. asiaticum</i>	<i>M. chlorophenicum</i>	<i>M. avium</i>	<i>M. avium</i>
CQD20693.1	WP_036358226.1	WP_048472356.1	WP_003877260.1	WP_023876861.1
<i>M. xenopi</i>	<i>M. simiae</i>	<i>M. vanbaalenii</i>	<i>M. lentiflavum</i>	<i>M. sinense</i>
WP_003919555.1	WP_044507197.1	ABM11247.1	CQD20744.1	WP_013830364.1
<i>M. gastri</i>	<i>M. genavense</i>		<i>M. nebraskense</i>	<i>M. heckeshornense</i>
WP_036414692.1	WP_025738432.1 ^a		WP_046181754.1	WP_048890125.1
<i>M. parascrofulaceum</i>	<i>M. kyorinense</i>		<i>M. europaeum</i>	<i>M. genavense</i>
WP_007166396.1	WP_045383051.1		CQD03229.1	WP_025737053.1
<i>M. sinense</i>	<i>M. avium</i>		<i>M. heckeshornense</i>	<i>M. heraklionense</i>
WP_013830530.1	WP_009975858.1		WP_048890092.1	WP_047318276.1
<i>M. nebraskense</i>	<i>M. parascrofulaceum</i>		<i>M. xenopi</i>	<i>M. kansasii</i>
WP_046186374.1	WP_007168596.1		WP_003922631.1	WP_036393982.1
<i>M. haemophilum</i>	<i>M. europaeum</i>		<i>M. vanbaalenii</i>	
WP_047316334.1	CQD03261.1		WP_041308028.1	
<i>M. europaeum</i>	<i>M. nebraskense</i>		<i>M. gilvum</i>	
CQD04192.1	WP_046181777.1		WP_041799998.1	
<i>M. phlei</i>	<i>M. thermoresistibile</i>		<i>M. rufum</i>	
WP_003889324.1	WP_003927083.1		WP_043411031.1	

Continued overleaf

		cluster binding motif		
CXXHXXC(X) _n CP	CXXNXXC(X) _n CP	CXXYXXC(X) _n CP	CXXFXXC(X) _n CP	CXXSXXC(X) _n CP
<i>M. hassiacum</i>	<i>M. iranikum</i>		<i>M. vaccae</i>	
WP_005630344.1	WP_024448284.1		WP_040539836.1	
<i>M. tusciae</i>	<i>M. phlei</i>		<i>M. chubuense</i>	
WP_006241645.1	WP_003889603.1		WP_048418201.1	
<i>M. canettii</i>	<i>M. heraklionense</i>		<i>M. chlorophenicum</i>	
WP_042914619.1	WP_047319309.1		WP_048473402.1	
<i>M. smegmatis</i>	<i>M. rhodesiae</i>		<i>M. iranikum</i>	
WP_015309433.1	WP_014212260.1		WP_036465671.1	
<i>M. thermoresistibile</i>	<i>M. rufum</i>		<i>M. phlei</i>	
WP_003926283.1	WP_043411043.1		WP_003889597.1	
<i>M. elephantis</i>	<i>M. gilvum</i>		<i>M. elephantis</i>	
WP_046752094.1	WP_011893358.1		WP_046750099.1	
<i>M. avium</i>	<i>M. aurum</i>		<i>M. aurum</i>	
WP_019734709.1	WP_048631967.1		WP_048631973.1	
<i>M. heckeshornense</i>	<i>M. chubuense</i>		<i>M. tusciae</i>	
WP_048889694.1	WP_014816820.1		WP_006241377.1	
<i>M. heraklionense</i>	<i>M. tusciae</i>		<i>M. rhodesiae</i>	
WP_047320532.1	WP_006241383.1		WP_014212266.1	
<i>M. goodie</i>	<i>M. vaccae</i>		<i>M. thermoresistibile</i>	
WP_049744806.1	WP_003928240.1		WP_040547386.1	
<i>M. rufum</i>	<i>M. vulneris</i>		<i>M. smegmatis</i>	
WP_043412685.1	WP_036448623.1		WP_015308293.1	
<i>M. setense</i>	<i>M. smegmatis</i>			
WP_039382128.1	WP_015308299.1			
<i>M. mageritense</i>				
WP_036434743.1				
<i>M. aromaticivorans</i>				
WP_036341668.1				

Table S7 A list of other ferredoxin genes from other bacteria which share the same ferredoxin cluster binding motif as those from *M. marinum*. Entries in red have a CYP enzyme clustered with the ferredoxin. Note the list is not exhaustive and more than one ferredoxin with the stated motif may

be present in the strain listed. Ferredoxin genes with the CXXTXXC(X)_nCP motif were not found in other bacteria using BLAST searches of the NCBI database.

^a clustered with a dioxygenase gene. ^b clustered close to a monooxygenase gene. ^c Part of a larger steroid degrading biosynthetic cluster which contains a P450 encoding gene. ^d a ferredoxin reductase gene is also located by the ferredoxin. ^e the ferredoxin gene is fused to the ferredoxin reductase gene.

		cluster binding motif		
CXXHXXC(X) _n CP	CXXNXXC(X) _n CP	CXXYXXC(X) _n CP ^d	CXXFXXC(X) _n CP	CXXSXXC(X) _n CP
<i>Nocardia higoensis</i> WP_040801623.1	<i>Nocardia pneumonia</i> WP_040773457.1 ^{a,b,c}	<i>Nocardia jiangxiensis</i> WP_040830166.1	<i>Nocardia jiangxiensis</i> WP_040825470.1	<i>Saccharopolyspora spinosa</i> WP_010693449.1 ^e
<i>Nocardia farcinica</i> WP_011209115.1	<i>Nocardia farcinica</i> WP_011207065.1 ^c	<i>Streptomyces avermitilis</i> WP_037650459.1	<i>Microbacterium ketosireducens</i> KJL44014.1 ^a	<i>Saccharomonospora cyanea</i> WP_005457293.1 ^e
<i>Gordonia amicalis</i> WP_024497964.1	<i>Rhodococcus opacus</i> WP_005244549.1 ^c	<i>Streptomyces natalensis</i> WP_030064645.1	<i>Rhodococcus rhodochrous</i> WP_033237125.1	<i>Streptomyces rapamycinicus</i> AGP60579.1 ^e
<i>Rhodococcus fascians</i> WP_037176148.1	<i>Rhodococcus jostii</i> RHA1 ABG96480.1	<i>Streptomyces collinus</i> WP_020940041.1	<i>Saccharopolyspora spinosa</i> WP_010314919.1	<i>Streptomyces violaceusniger</i> WP_014058289.1 ^e
<i>Rhodococcus opacus</i> WP_015888533.1	<i>Nocardioides luteus</i> WP_045549377.1	<i>Streptomyces antibioticus</i> WP_053212137.1		<i>Streptosporangium amethystogenes</i> WP_030915565.1
<i>Frankia</i> sp. EAN1pec WP_020462084.1	<i>Haliangium ochraceum</i> WP_012827286.1	<i>Streptomyces virginiae</i> WP_033214178.1		<i>Streptomyces durhamensis</i> WP_031160704.1 ^e
<i>Songiibacter tropicus</i> WP_051151222.1	<i>Frankia</i> sp. EAN1pec WP_020459245.1 ^a	<i>Streptosporangium roseum</i> WP_012895195.1		<i>Saccharomonospora glauca</i> WP_005465046.1 ^e
<i>Novosphingobium malaysiense</i> WP_039287162.1 ^a	<i>Saccharomonospora halophila</i> WP_019811873.1	<i>Ktedonobacter racemifer</i> WP_052569463.1	<i>Continued overleaf</i>	<i>Streptacidiphilus albus</i> WP_034086903.1 ^e

CXXHXXC(X) _n CP	CXXNXXC(X) _n CP	cluster binding motif	CXXFXXC(X) _n CP	CXXSXXC(X) _n CP
<i>Streptomyces mirabilis</i> WP_037736688.1	<i>Herbidospora cretacea</i> WP_030456498.1	<i>Actinopolymorpha alba</i> WP_020578491.1 ^e		<i>Frankia</i> sp. CN3 WP_007510855.1 ^e
<i>Spirillospora albida</i> WP_030164411.1	<i>Gordonia amarae</i> WP_005182707.1 ^b	<i>Rhodococcus opacus</i> WP_012689438.1		
<i>Saccharomonospora marina</i> WP_009157031.1	<i>Actinomadura atramentaria</i> WP_019633183.1	<i>Methanosarcina barkeri</i> WP_048110266.1		
<i>Actinomadura oligospora</i> WP_026414661.1				
<i>Tomitella biformata</i> WP_024794878.1				
<i>Amycolatopsis halophile</i> WP_034277668.1				
<i>Streptomyces griseus</i> WP_037679114.1				
<i>Kineosporia aurantiaca</i> WP_052531168.1				
<i>Methyloferula stellate</i> WP_020176888.1				
<i>Rhodopseudomonas palustris</i> BisB5 WP_011502257.1				
<i>Novosphingobium aromaticivorans</i> ABP64540.1				

Table S8 Conservation of the CYP147/Fdr1/Fdx3 operon across different bacterial families. Fusion is where the ferredoxin gene is fused to the ferredoxin reductase gene.

^a This ferredoxin gene is associated with the ferredoxin reductase and CYP147 genes. ^b There are three other ferredoxin genes with the CXXYXXC(X)_nCP motif in *Streptomyces avermitilis* MA-4680.

CYP147G1 operon			
CYP147	FdR	Ferredoxin	CXX?XXC(X) _n CP
	<i>Frankia</i> sp. CN3		
WP_007510563.1	WP_007510561.1	fusion	Y
<i>Methanosarcina barkeri</i> str. <i>Fusaro</i>			
WP_011306930.1	WP_011306931.1	WP_011306932.1	Y
<i>Methylobacterium extorquens</i> CM4			
WP_015952233.1	WP_015952234.1	WP_003597514.1	Y
<i>Methylobacterium radiotolerans</i> JCM 2831			
WP_012318808.1	WP_012318807.1	WP_012318806.1 ^a	Y
<i>Mycobacterium vanbaalenii</i> PYR-1			
ABM11249.1	ABM11248.1	ABM11247.1	Y
<i>Myxococcus xanthus</i> DK 1622			
ABF88234.1	ABF90123.1	fusion	Y
<i>Rhodococcus jostii</i> RHA1			
WP_011595234.1	WP_009475197.1	WP_011595232.1	Y
<i>Rhodococcus opacus</i> B4			
WP_012689440.1	WP_012689439.1	WP_012689438.1	Y
<i>Streptomyces avermitilis</i> MA-4680 ^b			
WP_010982022.1	WP_010982021.1	WP_010982020.1	Y
<i>Streptosporangium roseum</i> DSM 43021			
ACZ91470.1	ACZ91469.1	ACZ91468.1	Y

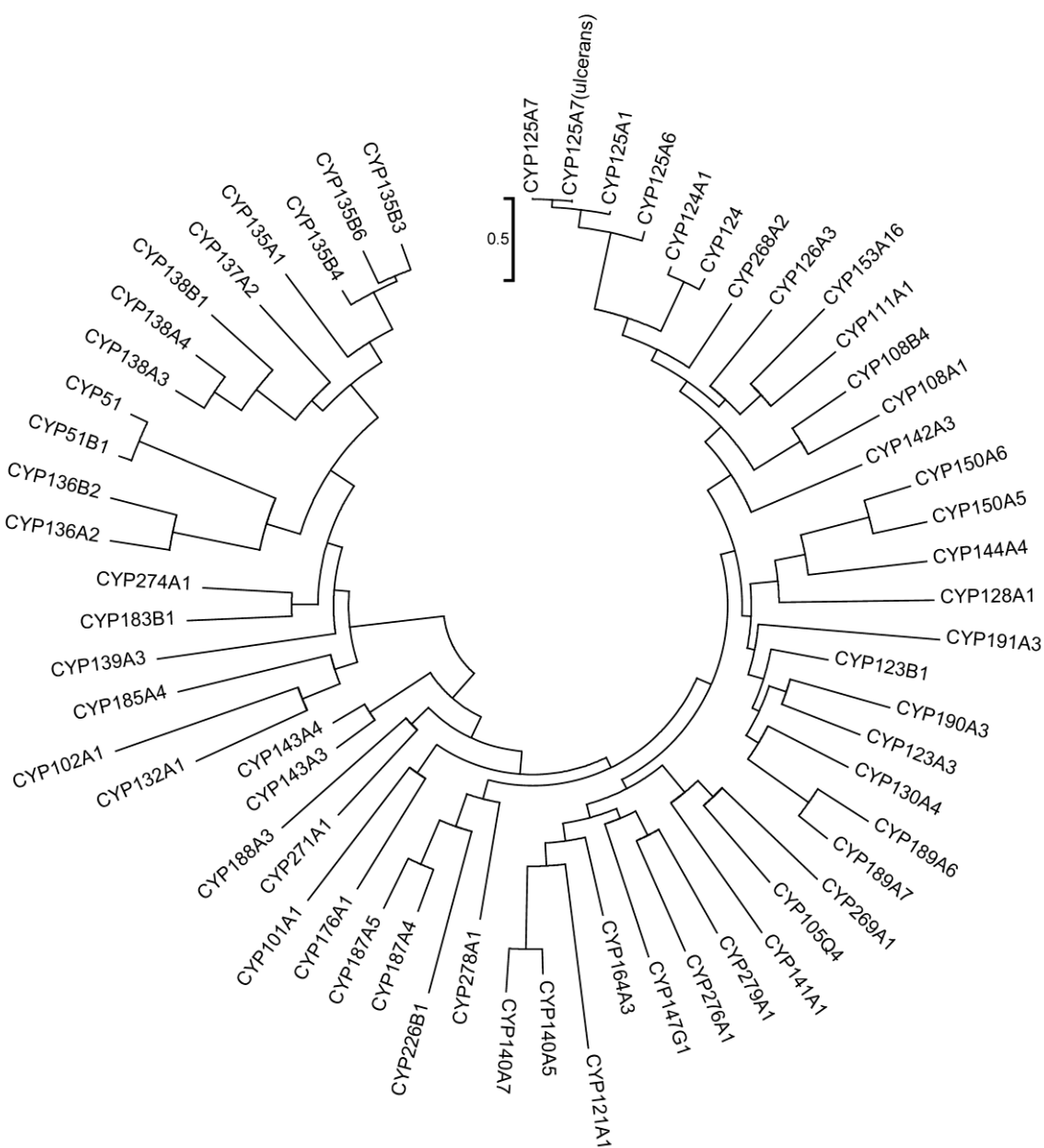
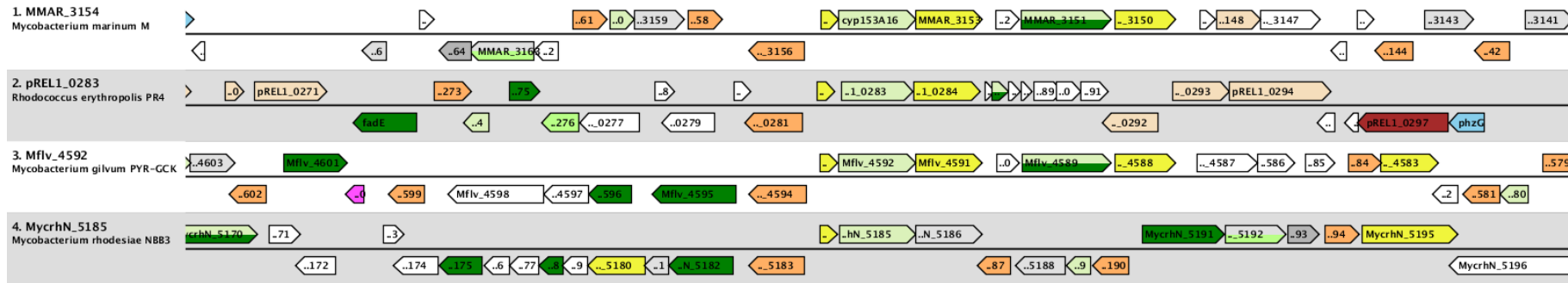


Figure S1 A phylogenetic tree of the forty seven CYP enzymes from *M. marinum*. Also included for comparison are the mycobacterial CYP enzymes; CYP125A1, CYP51, CYP124, CYP121A1, CYP141A1, CYP128A1, CYP132A1 and CYP135A1 from *M. tuberculosis* and CYP125A7 and CYP140A7 from *M. ulcerans*. CYP101A1 (P450cam), CYP102A1 (P450Bm3), CYP108A1 (P450terp), CYP111A1 (P450lin) and CYP176A1 (P450cin) are also included.

(c) Mmar_3154 (CYP153A16), associated with the [2Fe-2S] ferredoxin (Mmar_3155) and the ferredoxin reductase (FdR2, Mmar_3153), comparison to *Rhodococcus erythropolis* PR4 (pREL1_0283), *M. gilvum* PYR-GCK (Mflv_4592), *M. rhodesiae* NB83 (MycrhN_5185). The ferredoxin reductase (FdR2) is conserved in *M. gilvum* PYR-GCK but not *M. rhodesiae* NB83 or *R. erythropolis* PR4. Mmar_3150 encodes a NAD dependent zinc-containing alcohol dehydrogenase, Mmar_3151, a medium chain fatty-acid-CoA ligase, Mmar_3152, a protein of unknown function. MMar_3156 encodes a transcriptional regulatory protein.

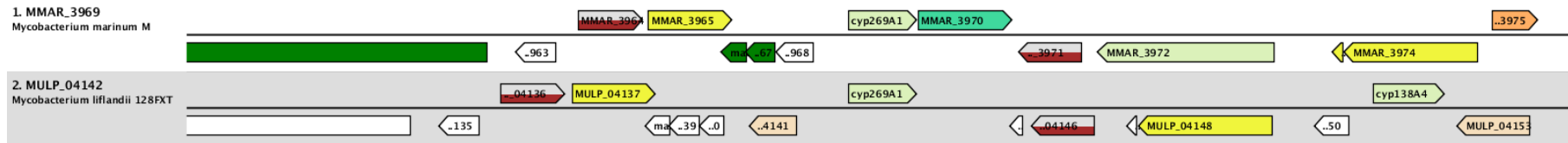


(d) Mmar_3969 (CYP269A1), associated with the Fdx4 (Mmar_3973), comparison to *M. liflandii* 128FXT (MulP_04142). The ferredoxin is not conserved in *M. liflandii* 128FXT.

Mmar_3962 encodes a fatty acid synthase, Mmar_3964 a subunit of a ring hydroxylating dioxygenase, Mmar_3965 a NAD-dependent aldehyde dehydrogenase, Mmar_3966 an acyl dehydratase.

Mmar_3971 encodes a probable oxidoreductase and Mmar_3972 an enzyme involved in non-ribosomal peptide synthetase.

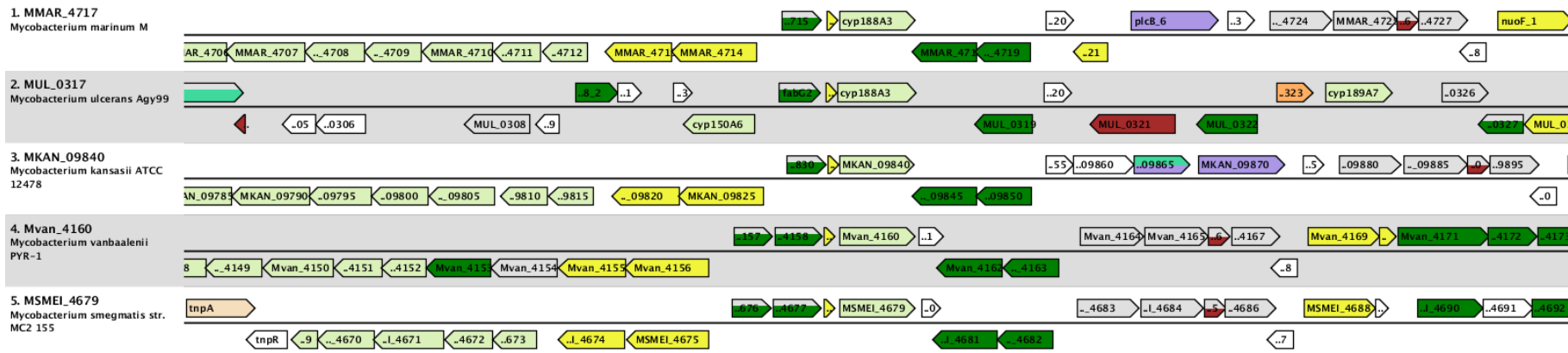
Mmar_3974 encodes an acyl-CoA transferase.



(e) Mmar_4717 (CYP188A3), associated with Fdx5 (Mmar_4716), comparison to *M. ulcerans* Agy99 (Mul_0317), *M. kansasii* ATCC 12478 (MKAN_09840), *M. vanbaalenii* PYR-1 (Mvan_4160) and *M. smegmatis* str. Mc2 155 (155 (MSMEI_4679)). Several of the surrounding genes are highly conserved across the *Mycobacterium*. Of particular note is the close proximity of the CYP150A6 gene in *M. ulcerans* Agy99 (the equivalent gene in *M. marinum* M is Mmar_4694).

Mmar_4705 to Mmar_4710 encode proteins involved in MCE (mammalian cell entry), Mmar_4711 and Mmar_4712 conserved hypothetical proteins of unknown function, Mmar_4713 a proposed dehydratase, Mmar_4714 a NAD-dependent aldehyde dehydrogenase and Mmar_4715 a NAD-dependent aldehyde dehydrogenase.

Mmar_4718 and Mmar_4719 encode acyl-CoA dehydrogenases and Mmar_4722 a membrane-associated phospholipase plcB_6 possibly involved in sphingomyelin and phosphatidylcholine hydrolysis.



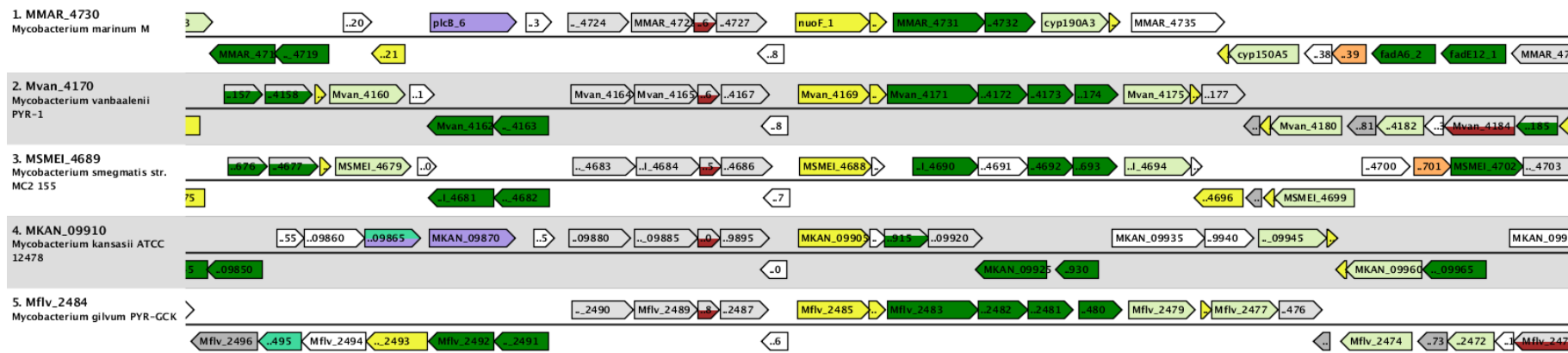
(f) Mmar_4730 (Fdx6), associated with CYP190A3 (Mmar_4733), Fdx7 (Mmar_4734), Fdx8 (Mmar_4736) and CYP150A5 (Mmar_4737), comparison to *M. vanbaalenii* PYR-1 (Mvan_4170), *M. smegmatis* str. Mc2 155 (MSMEI_4689), *M. kansasii* ATCC 12478 (MKAN_09910) and *M. gilvum* PYR-GCK (Mflv_2484). Many of the surrounding genes are highly conserved across the *Mycobacterium*. Of particular note is the close proximity of the genes equivalent to CYP188A3 and its associated ferredoxin in *M. vanbaalenii* PYR-1 and *M. smegmatis* str. Mc2 155 (see Fig S2(e)).

Mmar_4724 and Mmar_4725 encode hypothetical metal-dependent hydrolase, Mmar_4726, similar to a Rieske ferredoxin subunit of certain proteins, a Mmar_4727, a hydrolase, Mmar_4729 a protein with similarity to NADH dehydrogenase I (chain F).

Mmar_4732 and Mmar_4733 encode a 3-ketoacyl-CoA thiolase and a thioesterase, respectively.

Mmar_4735 encodes a PE-PGRS family protein.

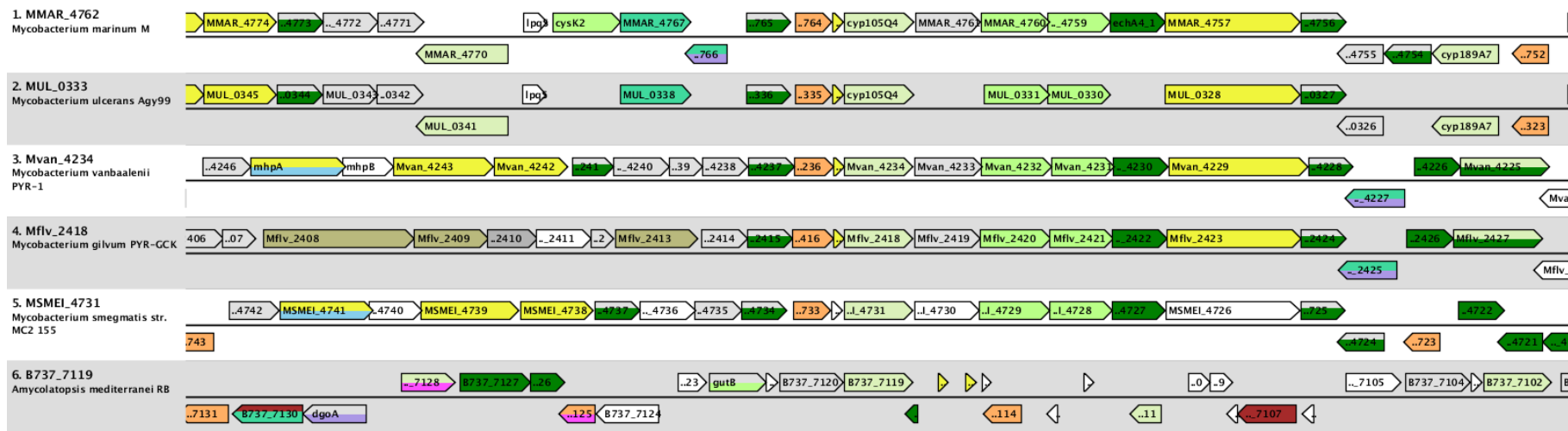
Mmar_4739 encodes a transcriptional regulatory protein, Mmar_4740 and Mmar_4741, an acetyl-CoA acetyltransferase FadA6_2 and an acyl-CoA dehydrogenase FadE12_1, respectively and Mmar_4742 a metal-dependent amidohydrolase.



(g) Mmar_4762 (CYP105Q4), associated with Fdx9 (Mmar_47633), comparison to *M. ulcerans* Agy99 (Mul_0333), *M. vanbaalenii* PYR-1 (Mvan_4234), *M. gilvum* PYR-GCK (Mflv_2418), *M. smegmatis* str. Mc2 155 (MSMEI_4731) and *Amycolatopsis mediterranei* RB (B737_7119). Many of the surrounding genes are highly conserved across the *Mycobacterium*. A ferredoxin gene is also found in close proximity to the P450 gene in *A. mediterranei* RB.

Mmar_4756 to Mmar_4760 encodes an oxidoreductase, an acyl-CoA transferase, an enoyl-CoA hydratase, EchA4_1, an aminopeptidase and a dipeptidase, respectively. The function of Mmar_4761 is unknown.

Mmar_4764 may encode a transcription regulator, Mmar_4765 encodes a short chain dehydrogenase, Mmar_4766 a hypothetical protein, Mmar_4767 a membrane transport protein, Mmar_4768 a cysteine synthase a CysK2, Mmar_4769 a lipoprotein LpqS, Mmar_ Mmar_4770 an oxidase.

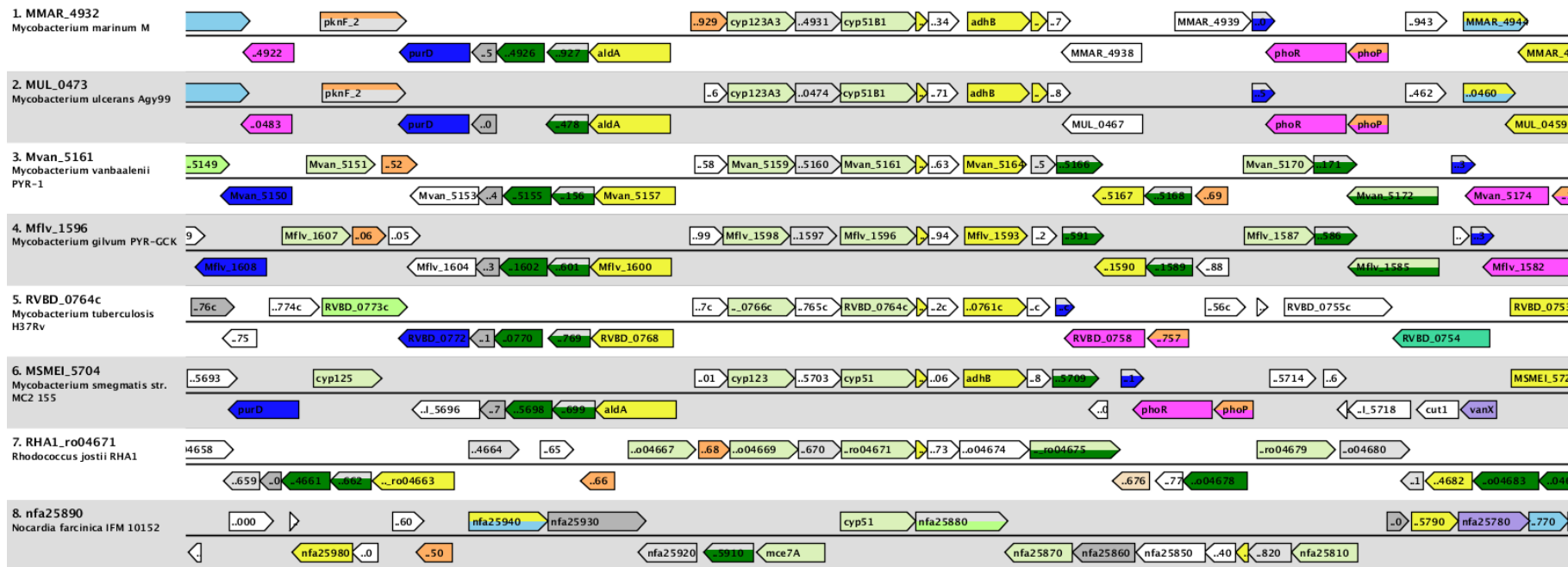


(h) Mmar_4932 (CYP51B1), associated with Fdx10 (Mmar_4933) and CYP123A1 (Mmar_4930), comparison to *M. ulcerans* Agy99 (Mul_0473), *M. vanbaalenii* PYR-1 (Mvan_5161), *M. gilvum* PYR-GCK (Mflv_1596), *M. tuberculosis* H37Rv (RVBD_0764c), *M. smegmatis* str. Mc2 155 (MSMEI_5704), *Rhodococcus jostii* RHA1 (RHA1_ro04671) and *Nocardia farcinica* IFM 10152. Many of the surrounding genes are highly conserved across the *Mycobacterium*. Several of the genes are also conserved in the *R. jostii* RHA1 species but the genes in the vicinity of the CYP51 gene in *N. farcinica* IFM 10152 are significantly different.

Mmar_4925 though to Mmar_4928 encode a phosphoribosylamine-glycine ligase, PurD, a 4-carboxymuconolactone decarboxylase, two dehydrogenase/reductases and a NAD-dependent aldehyde dehydrogenase, AldA. Mmar_4929 encodes a protein of unknown function.

Mmar_4931 encodes a short-chain alcohol dehydrogenase.

Mmar_4934 to Mmar_4937 encodes a conserved/hypothetical protein, a zinc-containing alcohol dehydrogenase NAD-dependent AdhB, a possible NADH:flavin oxidoreductase and an unknown conserved hypothetical protein containing a nuclear transport factor 2 (NTF2) domain.



(i) Mmar_4991 (Fdx11, which is not associated with a P450 gene), comparison to *M. liflandii* 128FXT (MulP_05239), *M. ulcerans* Agy99 (Mul_4066), *M. tuberculosis* H37Rv (RVBD_3503c), *M. smegmatis* str. Mc2 155 (MSMEI_5744), *M. kansasii* ATCC 12478 (MKAN_11680), *M. rhodesiae* NB83 (MycrhN_2333) and *M. gilvum* PYR-GCK (Mflv_1596). Many of the surrounding genes are highly conserved across the *Mycobacterium* species.

Mmar_4980 to Mmar_4989 encodes a conserved MCE-associated protein, a conserved hypothetical alanine and valine rich MCE-associated protein, a MCE-family protein Mce4F, a MCE-family lipoprotein LprN, a MCE-family protein Mce4D, a MCE-family protein Mce4C, a MCE-family protein Mce4B, a MCE-family protein Mce4A, a conserved hypothetical membrane protein YrbE4B and a conserved membrane protein YrbE4A, respectively.

Mmar_4990 encodes a short-chain type dehydrogenase/reductase.

Mmar_4992 to Mmar_4996 encodes an acyl-CoA dehydrogenase FadE26, an acyl-CoA dehydrogenase FadE27, a fatty-acid-CoA synthetase FadD17, a PE-PGRS family protein and a conserved protein of unknown function which contains an eta-lactamase/transpeptidase-like superfamily domain

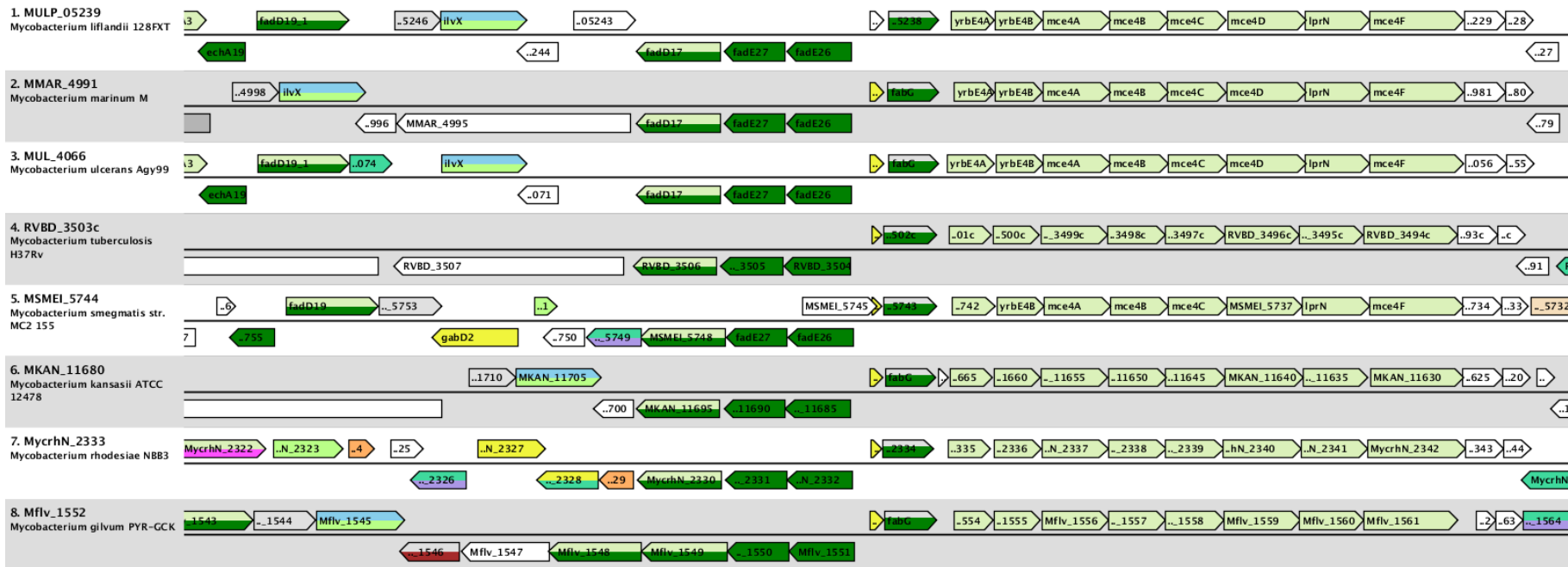
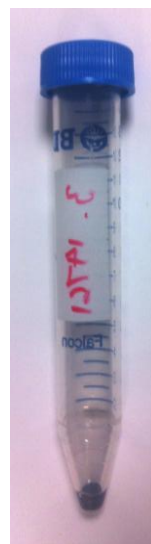
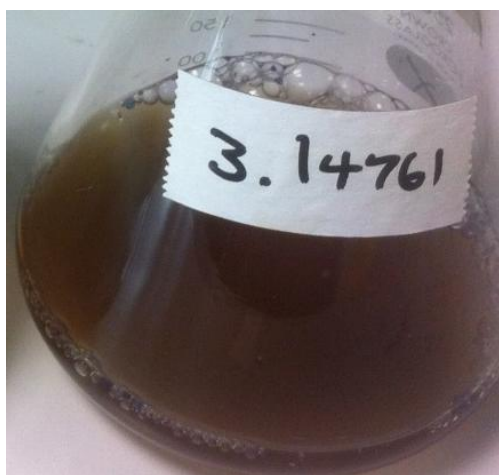
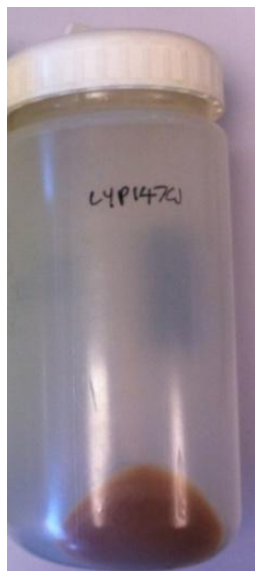


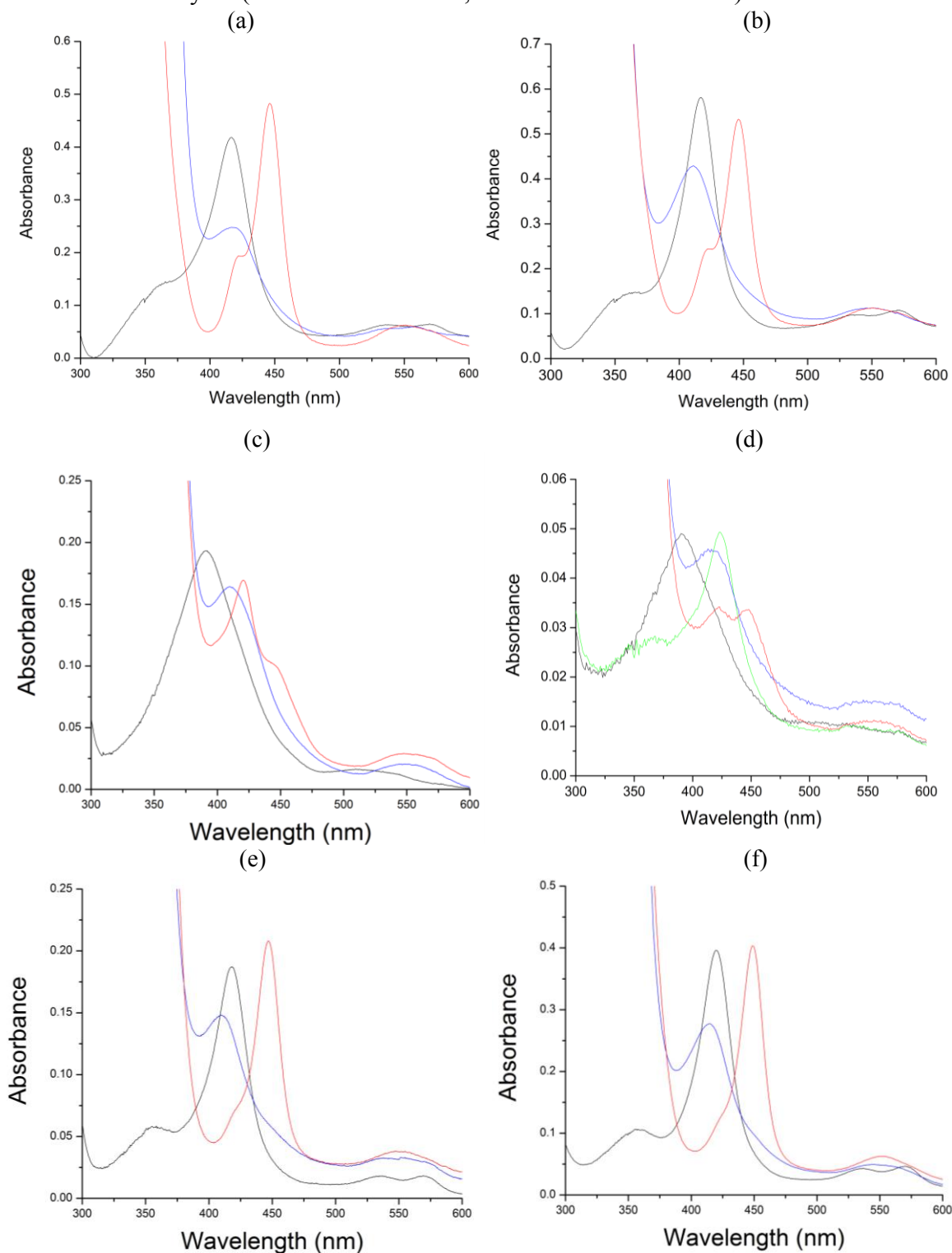
Figure S3 (a) Cell pellet of *E. coli* containing pET26CYP147G1 showing a red colouration due to the CYP147G1 expression but no blue colouration due to indigo formation. (b) Cell culture of the whole-cell oxidation system consisting of pETDuetFdx3/FdR1 and pRSFDuetFdx3/CYP147G1 and the cell pellet showing the formation of a blue dye.



(a)

(b)

Figure S4 The spectra of the ferrous, carbon monoxide bound forms of (a) CYP147G1, (b) CYP278A1, (c) CYP269A1, (d) CYP269A1 in the presence of miconazole, (e) CYP150A5 and (f) CYP105Q4. With the exception of CYP269A1 approximately 95% of the proteins shifted from 418 nm (black trace, ferric state) to 450 nm (red trace, reduced CO-bound). In (d), the addition of miconazole to CYP269A1 (black trace before addition, green trace after) stabilised the reduced CO-bound form of the enzyme (blue is reduced state, red is ferrous CO-bound).



The absorbances of the ferric, ferrous and ferrous-CO bound forms and the extinction coefficients are provided overleaf.

Absorbance maxima of the Soret band for the ferric, ferrous and ferrous-CO species of each P450 enzyme:

CYP147G1 420 nm, 420 nm and 446 nm

CYP278A1 418 nm, 411 nm and 446 nm

CYP269A1 392 nm, 410 nm and 423/446 nm split peak

CYP150A5 418 nm, 409 nm and 447 nm

CYP105Q4 420 nm, 415 nm and 449 nm

Extinction coefficients for the P450s (calculated on CO difference spectra and confirmed by the Pyridine hemochromagen method) * CYP269A1 was only determined by the pyridine hemochromagen method.

CYP147G1

$$\epsilon_{417} = 111 \text{ mM cm}^{-1}$$

CYP150A5

$$\epsilon_{418} = 103 \text{ mM cm}^{-1}$$

CYP278A1

$$\epsilon_{417} = 126 \text{ mM cm}^{-1}$$

CYP269A1*

$$\epsilon_{392} = 114 \text{ mM cm}^{-1}$$

CYP105Q4

$$\epsilon_{420} = 110 \text{ mM cm}^{-1}$$

Figure S5 Spin-state shift and binding constants assays for (a) CYP147G1 with undecanoic acid, (b) CYP278A1 with β -ionone and (c) CYP150A5 with β -ionone.

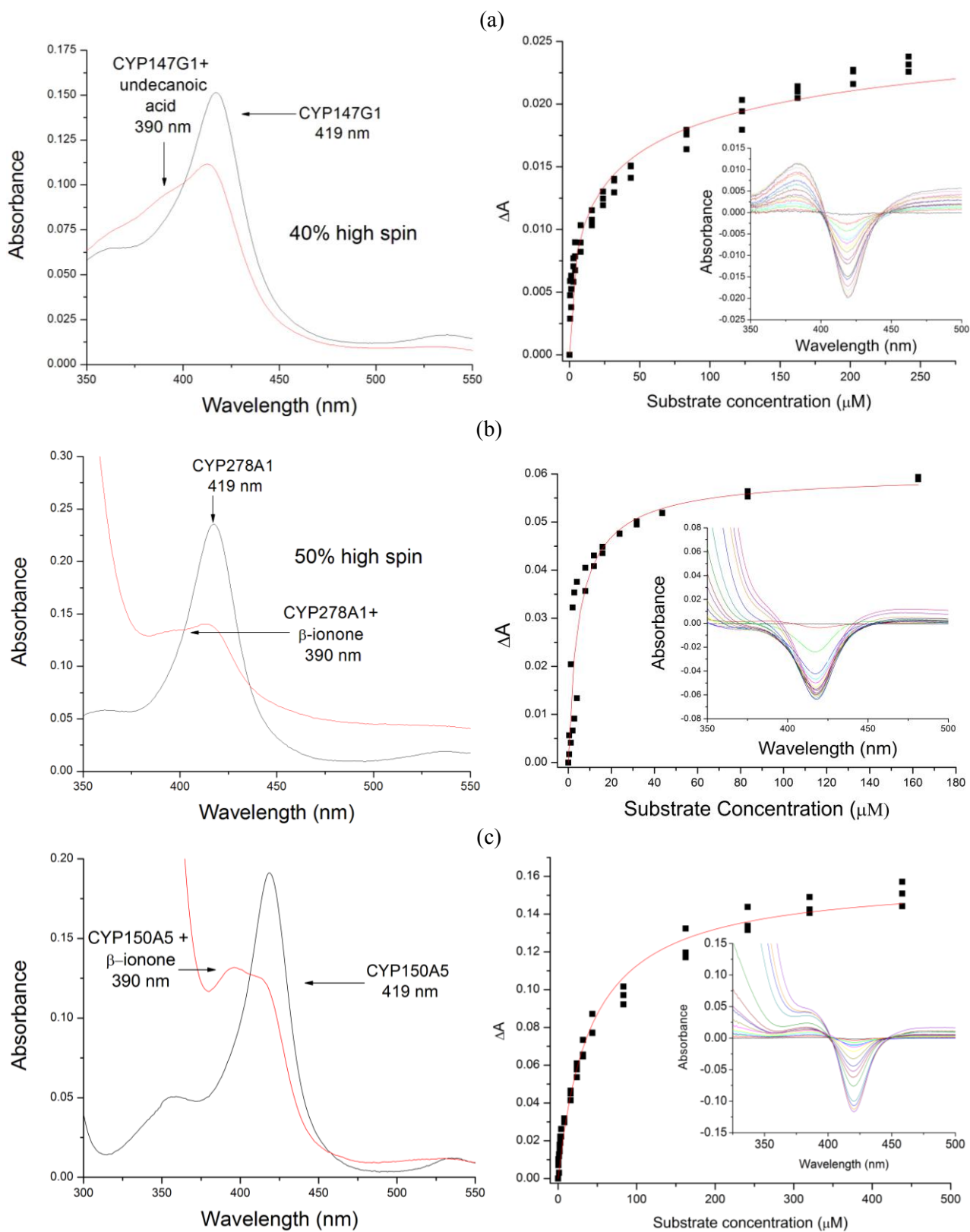
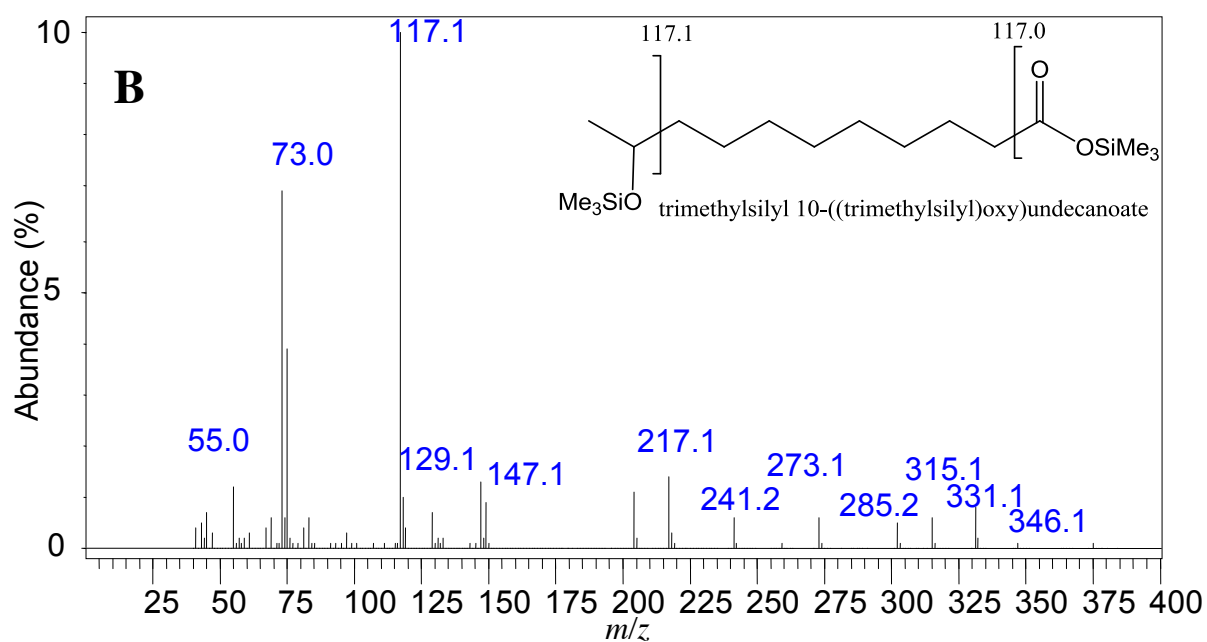
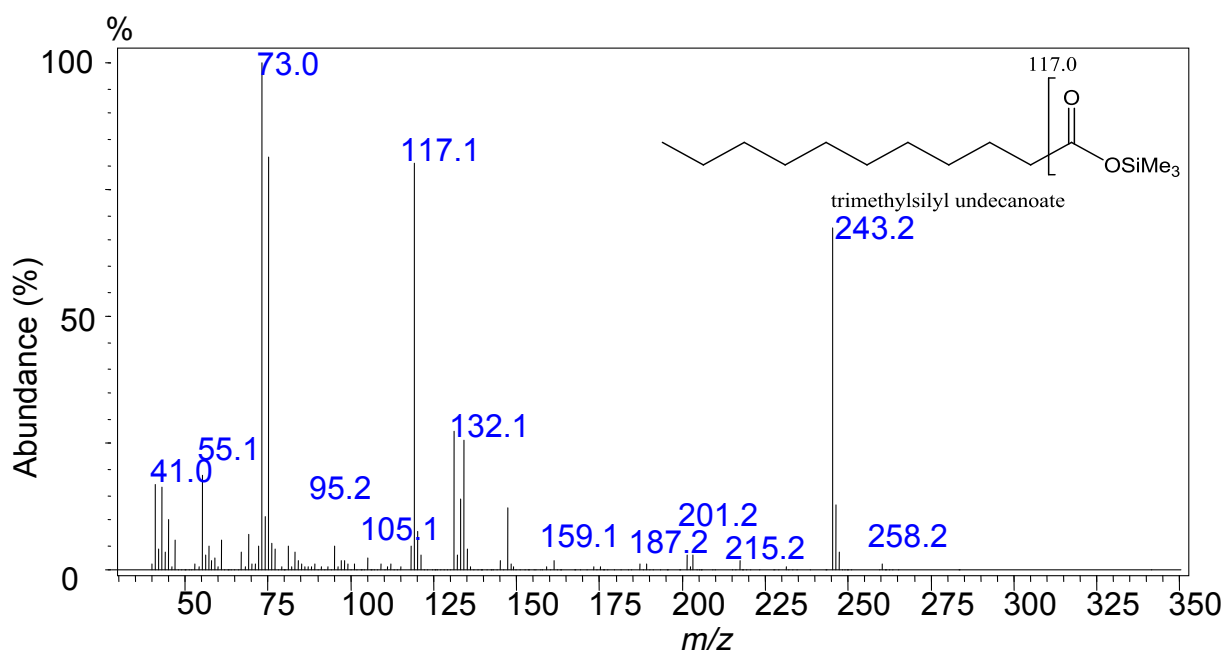


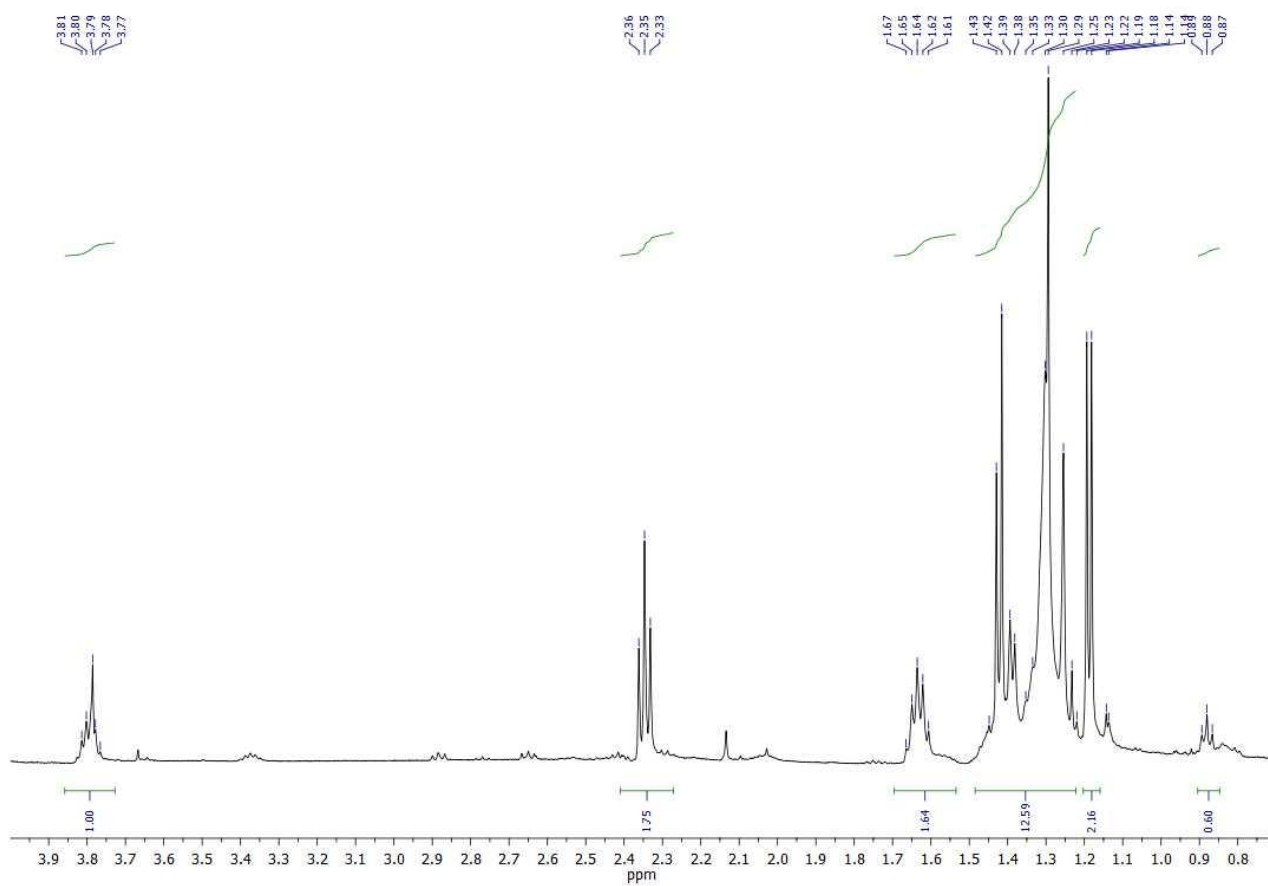
Figure S6 Mass spectrum of GC peaks from CYP147G1 turnovers with undecanoic acid (A) undecanoic acid substrate (retention time 9.2 min), (B) hydroxyundecanoic acid (13.9 min).



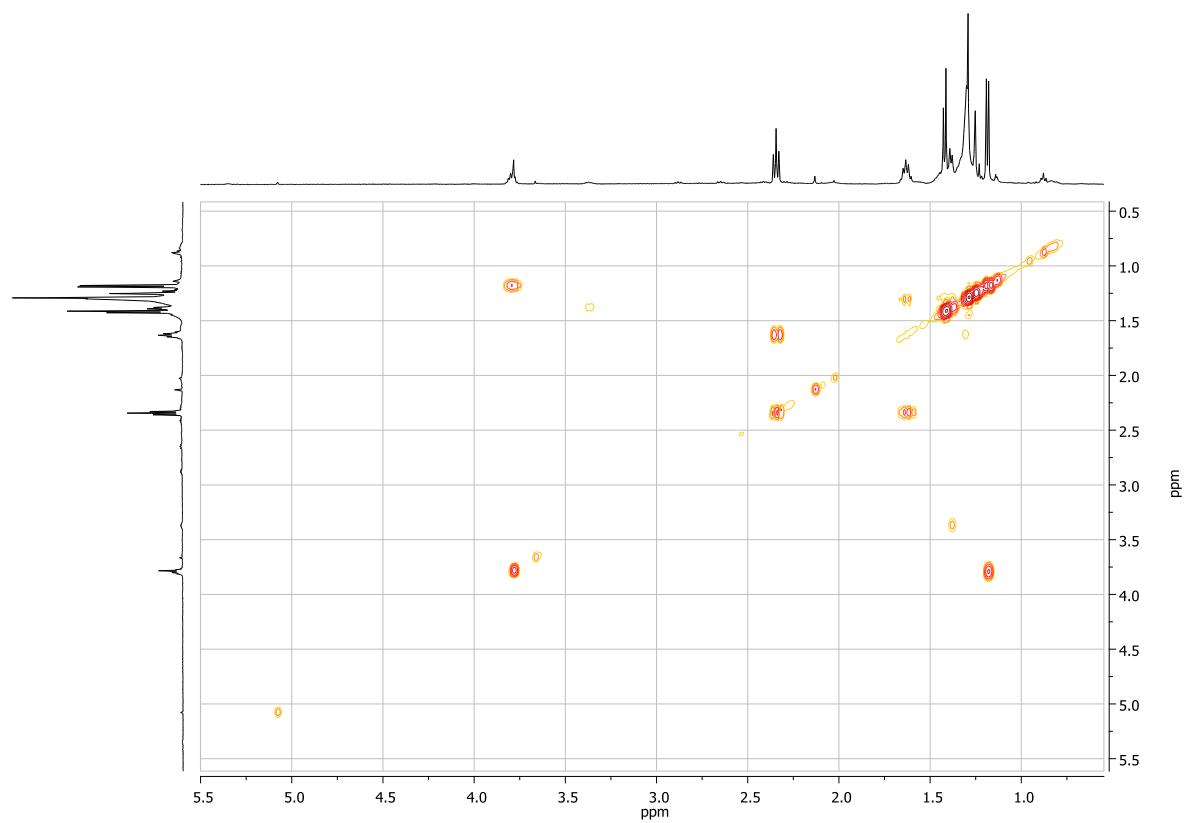
The peak at 9.2 min showed an m/z peak at 243.2 in comparison to the expected m/z of 258.2 for trimethylsilyl undecanoic acid (a mass loss of 15 m/z) and coeluted with the substrate standard (Appendix D). The product of the undecanoic acid turnover (with two trimethylsilyl protecting groups) has an expected m/z of 346.24. The peak at 13.9 min showed an m/z of 346.1 and a secondary peak of 331.1 (a mass difference of 15 m/z from the apparent molecular ion peak).

10-hydroxyundecanoic acid NMR data. ^1H NMR (500 MHz, CDCl_3) δ 3.80 (tq, $J = 6.3$ Hz, 1H), 2.35 (t, $J = 7.4$ Hz, 2H), 1.64 (tt, $J = 7.4$ Hz, 2H), 1.46 – 1.23 (m, 12H), 1.19 (d, $J = 6.2$ Hz, 3H). ^{13}C NMR (126 MHz, CDCl_3) δ 136.81, 132.71, 68.17, 39.26, 33.66, 31.58, 30.26, 29.66, 29.63, 29.45, 29.27, 29.06, 28.95, 25.63, 24.64, 23.45, 19.18.

The overlapping triplet of quartets (tq) at approximately δ 3.8 ppm is distinctive for ω -1 hydroxylation. If the acid was hydroxylated at any other sub-terminal position, the H on the hydroxylated carbon would be a triplet of triplets, as it would not be split by the terminal CH_3 group. Terminal hydroxylation would also be distinctive, as the CH_2OH group signal would be observed as a triplet of 2H intensity. The 3H doublet peak at δ 1.19 ppm can be assigned to the terminal CH_3 group, split by the single H on the CHOH group, thereby confirming the formation of 10-hydroxyundecanoic acid.

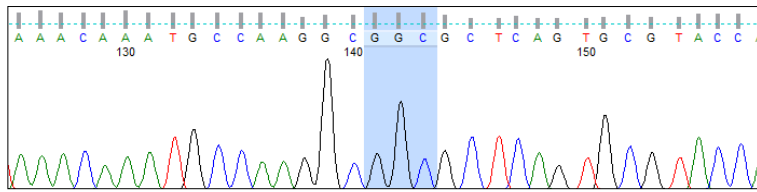


^1H NMR of 10-hydroxyundecanoic acid

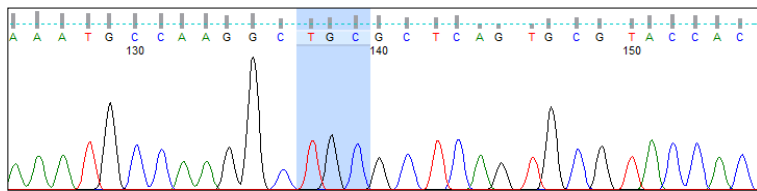


^1H COSY NMR of 10-undecanoic acid

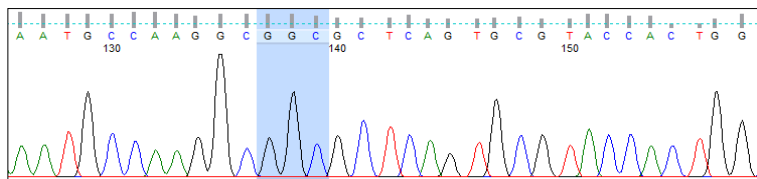
Figure S7 Excerpts of sequence chromatograms for Fdx3 mutants, with the successful codon change highlighted in blue.



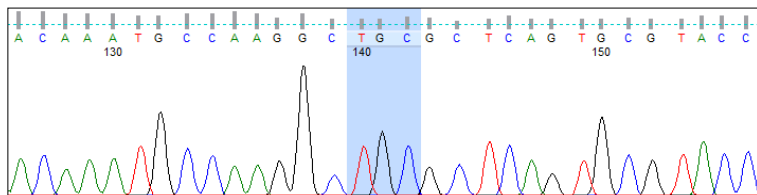
pETDuet with FdR2931 (FdR1) and Fdx2932 (Fdx3) Tyr12Gly – with codon GGC replacing TAC



pRSFDuet with CYP147G1 and Fdx2932 (Fdx3) Tyr12Cys – with codon TGC replacing TAC



pRSFDuet with CYP147G1 and Fdx2932 (Fdx3) Tyr12Gly – with codon GGC replacing TAC



pETDuet with Fdr2931 (FdR1) and Fdx2932 (Fdx3) Try12Cys – with codon TGC replacing TAC

Figure S8 Type II spin-state shift and binding constants assays for CYP269A1 with miconazole.

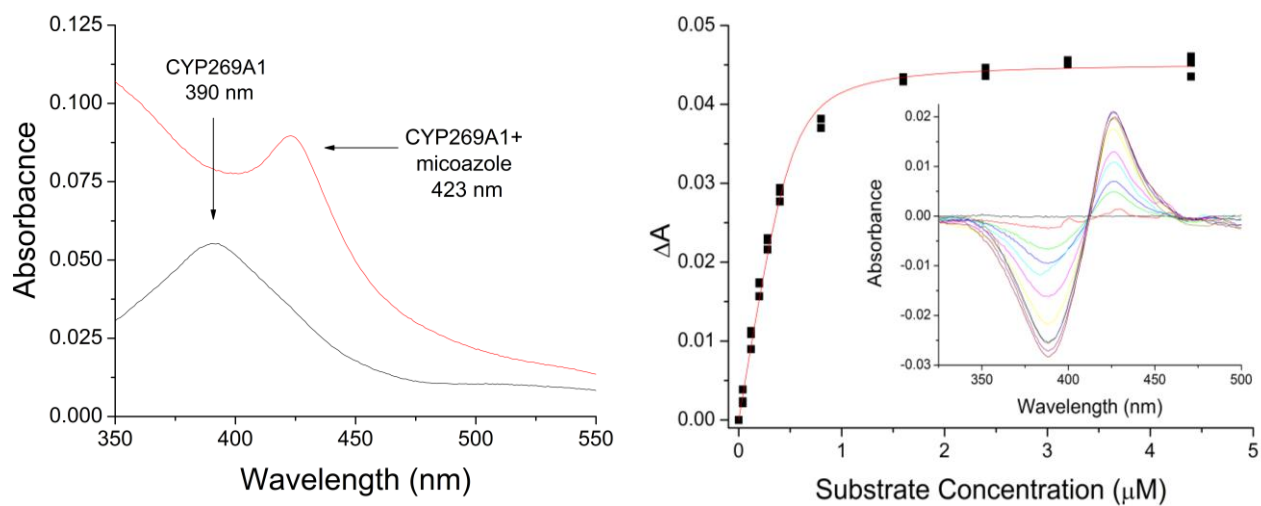
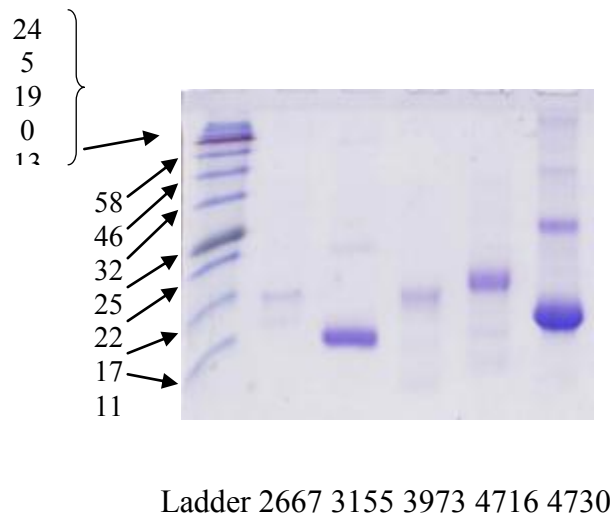


Figure S9 (a) SDS page of ferredoxin proteins. Note that the small highly acidic ferredoxins are difficult to quantitate in terms of their molecular weight via SDS page.²⁸



Protein MS data of selected ferredoxins

The average mass of the proteins was measured to be:

2Fe-2S (Mmar_3155)

12093.13 Da (theoretical average mass, 12216.65 Da)

Fdx4 (Mmar_3973)

7647.67 Da (theoretical average mass, 7642.71 Da)

Fdx5 (Mmar_4716)

8288 Da (theoretical average mass, 8118.77 Da)

Fdx6 (Mmar_4730)

11817.81 Da (theoretical average mass, 11781.52 Da)

Fdx8 (Mmar_4736)

7736.88 Da (theoretical average mass 7764.57 Da)

Fdx9 (Mmar_4763)

7662 Da (theoretical average mass, 7657.71 Da)

Note terminal methionine cleavage (-131) was observed in the [2Fe-2S] ferredoxin and Fdx4.

Figure S9 (b) SDS page analysis of Fdx4 purified by IMAC or Ion exchange followed by size exclusion (S100)

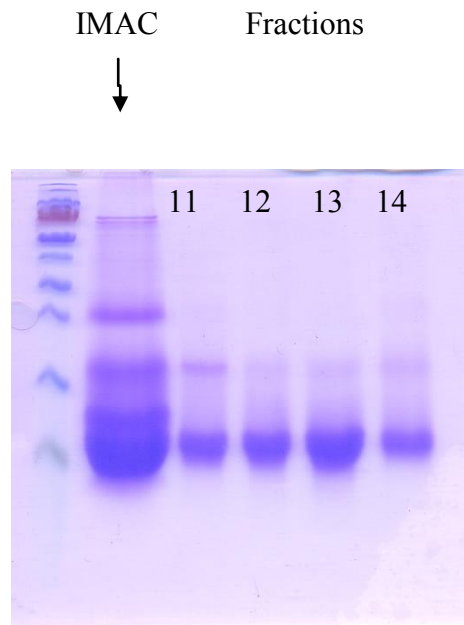


Figure S9 (c) SDS page analysis of associated CYP proteins.

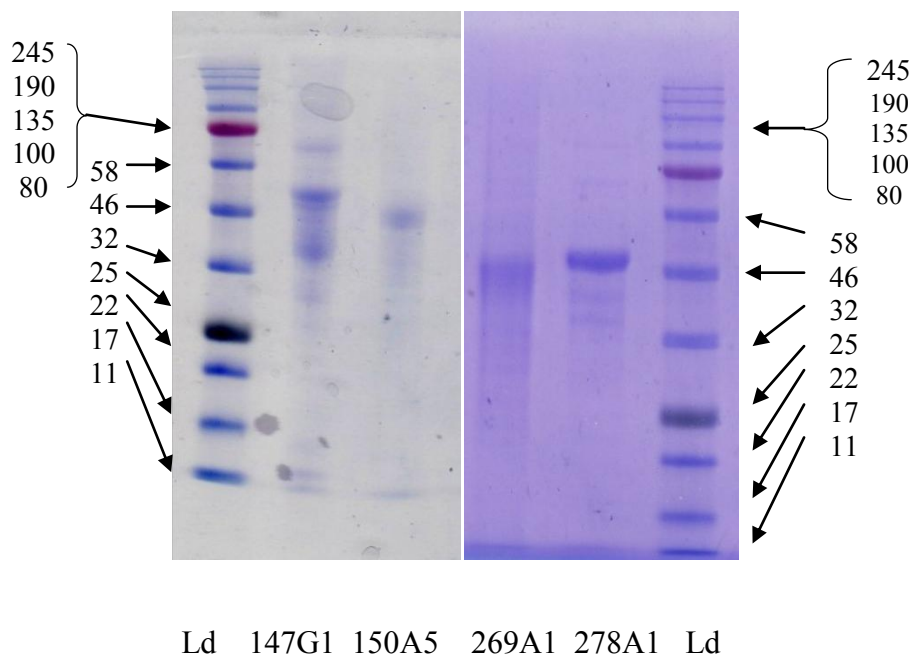


Figure S10: Absorbance (upper) and CD (lower) spectra of Fdx4 (panels A and D), Fdx5 (panels B and E) and Fdx9 (panels C and F) as isolated (red traces), incubated with ferricyanide (black traces), incubated with ascorbate (blue traces) and Eu^{2+} (dark cyan traces).

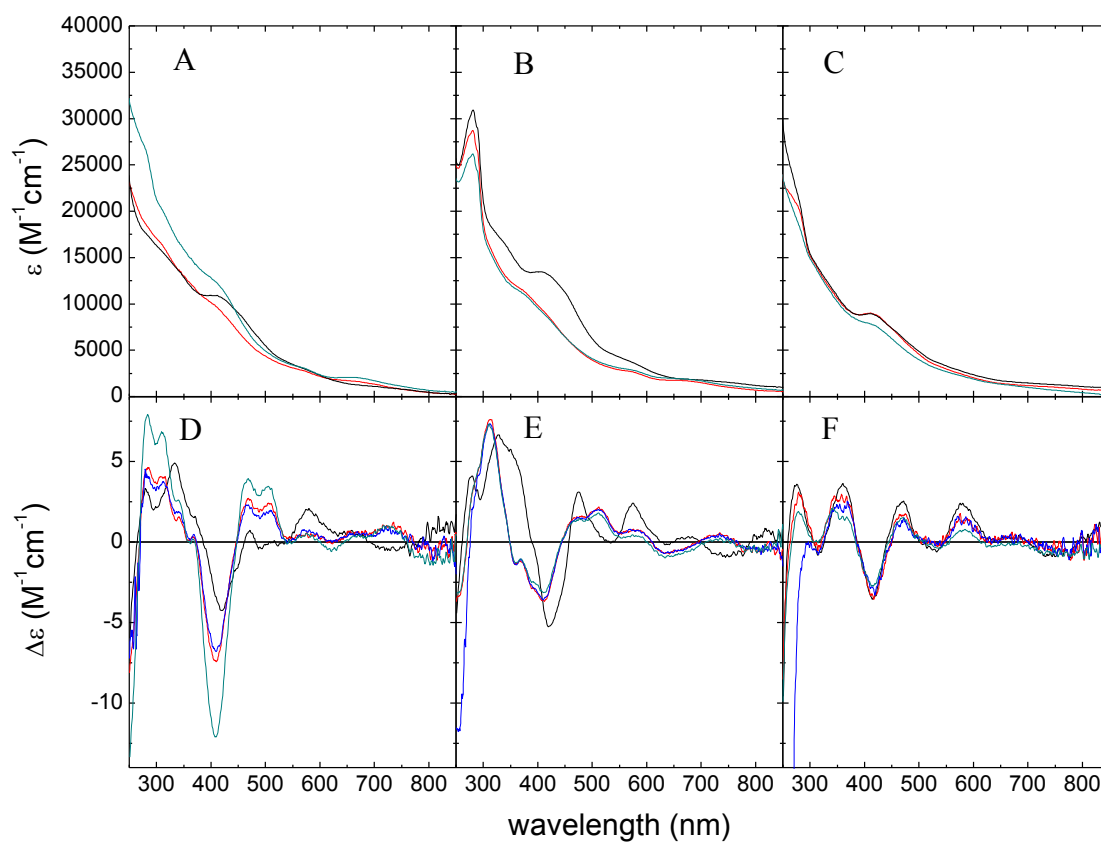


Figure S12: Cluster identification by mass spectrometry. Native (upper) and LC (lower) mass spectra of Fdx4 (panel A), Fdx5 (panel B) and Fdx9 (panel C). Each of the LC-MS spectra contain a peak at the predicted mass of the apo protein and an additional peak 131 Da lighter than this, consistent with cleavage of the N-terminal methionine. The native mass spectra contain peaks 296 Da higher in mass than each of these features (indicated by red arrows) consistent with binding of a [3Fe-4S] cluster by each of the ferredoxins.

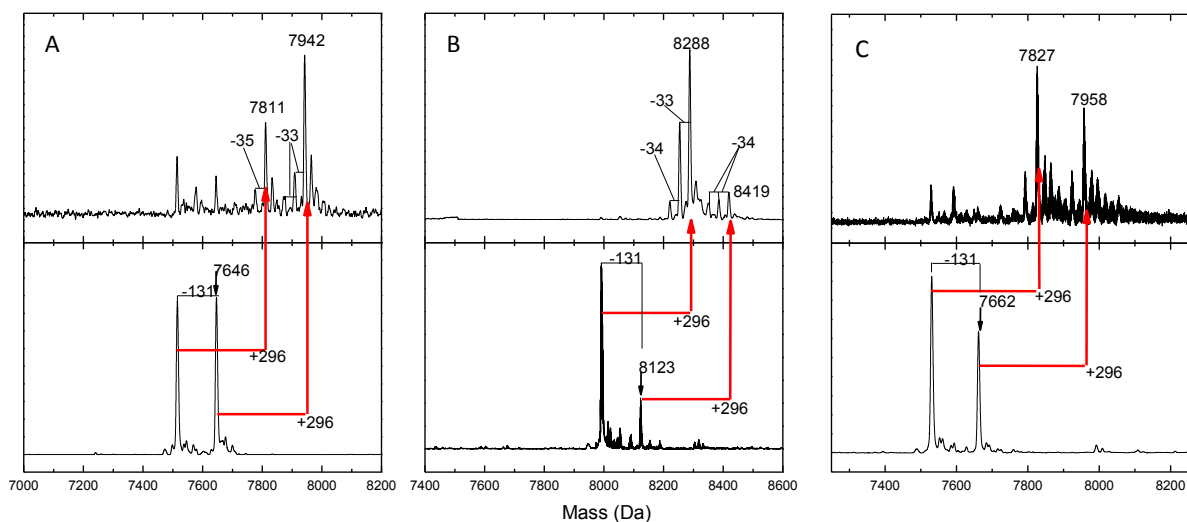


Figure S12 :EPR of spectroscopy of Fdx4 (Panel A), Fdx5 (Panel B) and Fdx9 (Panel C). The form of the spectra was consistent with the oxidised form of a [3Fe-4S] cluster in each case. The black traces are those recorded for the proteins as isolated, red traces those following incubation with ferricyanide and blue traces those following incubation with Eu^{2+} .

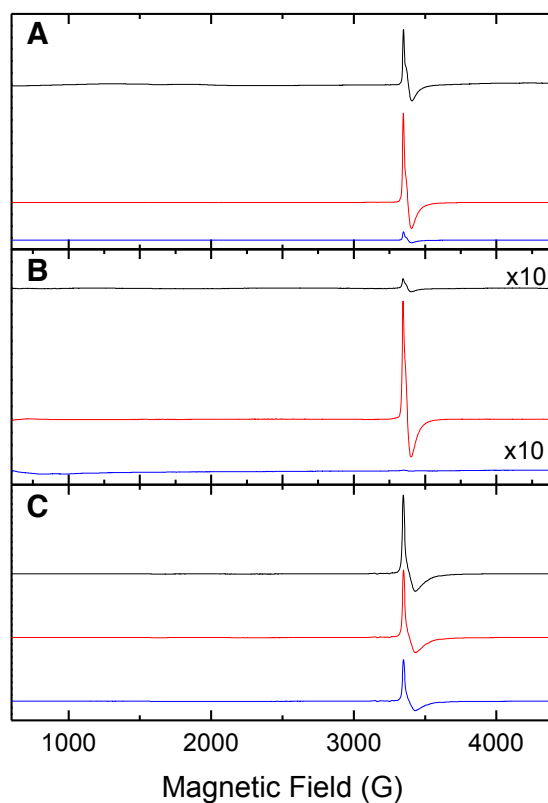


Figure S13: CD Spectra of Fdx4 isolated utilising a Ni²⁺ charged IMAC column (black trace) or a combination of anion exchange and size exclusion chromatography (red trace). Blue trace represents the latter sample following incubation with 1.0 stoichiometric equivalent of (NH₄)₂Fe(SO₄)₂

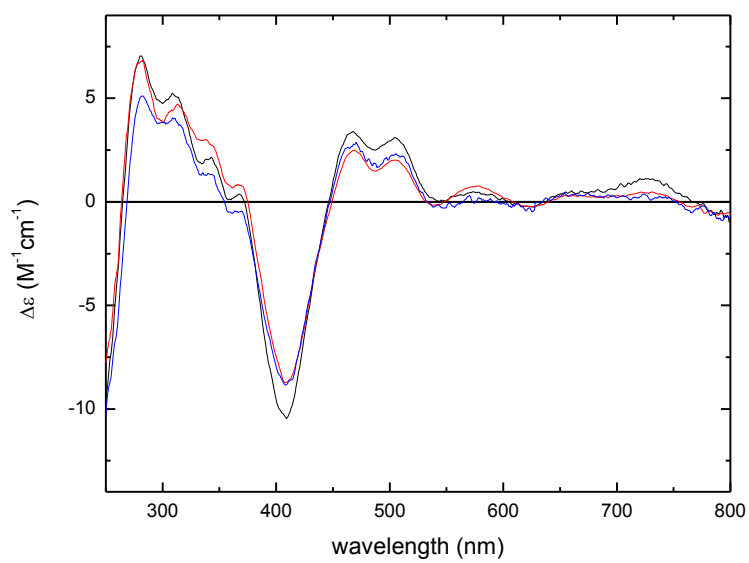
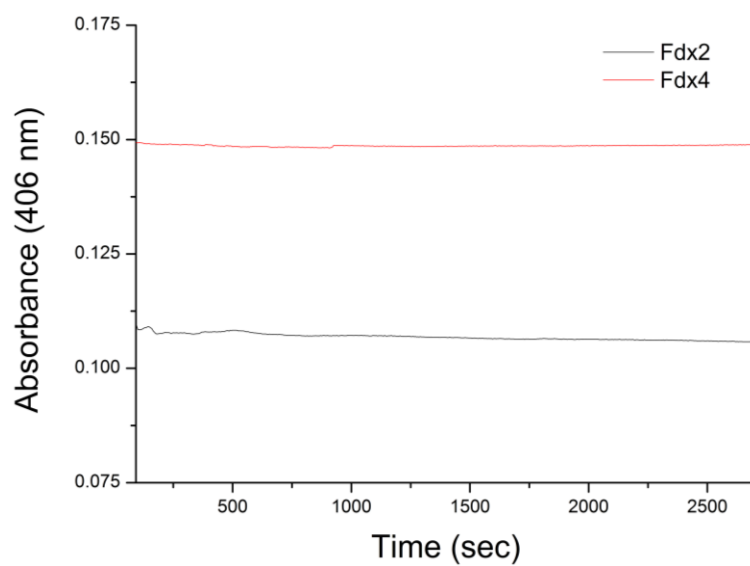


Figure S14: Measurement of the UV spectra (absorbance at 406 nm) of Fdx2 and Fdx4 which were purified under anaerobic conditions and then exposed to dioxygen.



References

1. Rupasinghe, S. et al. The cytochrome P450 gene family CYP157 does not contain EXXR in the K-helix reducing the absolute conserved P450 residues to a single cysteine. *FEBS Letters* 580, 6338-6342 (2006).
2. Poulos, T.L. Heme enzyme structure and function. *Chem Rev* 114, 3919-62 (2014).
3. Clark, J.P. et al. The role of Thr268 and Phe393 in cytochrome P450 BM3. *J. Inorg. Biochem.* 100, 1075-90 (2006).
4. Ouellet, H., Podust, L.M. & de Montellano, P.R. *Mycobacterium tuberculosis* CYP130: crystal structure, biophysical characterization, and interactions with antifungal azole drugs. *J. Biol. Chem.* 283, 5069-80 (2008).
5. Driscoll, M.D. et al. Expression and characterization of *Mycobacterium tuberculosis* CYP144: common themes and lessons learned in the *M. tuberculosis* P450 enzyme family. *Biochim. Biophys. Acta* 1814, 76-87 (2011).
6. Brezna, B. et al. Molecular characterization of cytochrome P450 genes in the polycyclic aromatic hydrocarbon degrading *Mycobacterium vanbaalenii* PYR-1. *Appl. Microbiol. Biotechnol.* 71, 522-32 (2006).
7. Ouellet, H., Johnston, J.B. & Ortiz de Montellano, P.R. The *Mycobacterium tuberculosis* cytochrome P450 system. *Arch. Biochem. Biophys.* 493, 82-95 (2010).
8. Garcia-Fernandez, E. et al. A highly conserved mycobacterial cholesterol catabolic pathway. *Environ. Microbiol.* 15, 2342-59 (2013).
9. Bell, S.G. et al. Cytochrome P450 enzymes from the metabolically diverse bacterium *Rhodospseudomonas palustris*. *Biochem. Biophys Res. Commun.* 342, 191-6 (2006).
10. Bell, S.G. & Wong, L.L. P450 enzymes from the bacterium *Novosphingobium aromaticivorans*. *Biochem. Biophys. Res. Commun.* 360, 666-72 (2007).
11. Bellamine, A., Mangla, A.T., Nes, W.D. & Waterman, M.R. Characterization and catalytic properties of the sterol 14alpha-demethylase from *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U. S. A.* 96, 8937-42 (1999).
12. Capyk, J.K. et al. Mycobacterial cytochrome p450 125 (cyp125) catalyzes the terminal hydroxylation of c27 steroids. *J. Biol. Chem.* 284, 35534-42 (2009).
13. Driscoll, M.D. et al. Structural and biochemical characterization of *Mycobacterium tuberculosis* CYP142: evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen. *J. Biol. Chem.* 285, 38270-82 (2010).
14. Johnston, J.B., Kells, P.M., Podust, L.M. & Ortiz de Montellano, P.R. Biochemical and structural characterization of CYP124: a methyl-branched lipid omega-hydroxylase from *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U. S. A.* 106, 20687-92 (2009).
15. Belin, P. et al. Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U. S. A.* 106, 7426-31 (2009).
16. Ouellet, H. et al. *Mycobacterium tuberculosis* CYP125A1, a steroid C27 monooxygenase that detoxifies intracellularly generated cholest-4-en-3-one. *Mol. Microbiol.* 77, 730-42 (2010).
17. Nelson, D.R. Cytochrome P450 nomenclature. *Methods Mol. Biol.* 107, 15-24 (1998).
18. Nelson, D.R. Cytochrome P450 nomenclature, *Methods Mol. Biol.* 320, 1-10 (2006).
19. Cole, S.T. et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537-44 (1998).
20. Stinear, T.P. et al. Reductive evolution and niche adaptation inferred from the genome of *Mycobacterium ulcerans*, the causative agent of Buruli ulcer. *Genome Res.* 17, 192-200 (2007).
21. Jones, D.T., Taylor, W.R. & Thornton, J.M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275-82 (1992).

22. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30, 2725-9 (2013).
23. Barr, I. & Guo, F. Pyridine Hemochromagen Assay for Determining the Concentration of Heme in Purified Protein Solutions. *Bio. Protoc.* 5(2015).
24. Williams, J.W. & Morrison, J.F. The kinetics of reversible tight-binding inhibition. *Methods Enzymol.* 63, 437-67 (1979).
25. Horak, T. et al. Analysis of free fatty acids in beer: comparison of solid-phase extraction, solid-phase microextraction, and stir bar sorptive extraction. *J. Agric. Food Chem.* 57, 11081-5 (2009).
26. Jenner, L.P. & Butt, J.N. Personal Communication.
27. Overmars, L., Kerkhoven, R., Siezen, R.J. & Francke, C. MGcV: the microbial genomic context viewer for comparative genome analysis. *BMC Genomics* 14, 209 (2013).
28. Williams, J.G. & Gratzer, W.B. Limitations of the detergent-polyacrylamide gel electrophoresis method for molecular weight determination of proteins. *J. Chromatogr.* 57, 121-5 (1971).

Chapter 3 Supplementary Information

Table S1. Sequence identities of CYP147G1, analogous proteins from *Mycobacterium* species, other named members of the CYP147 family and various other proteins for comparison.

Species	CYP name ^a	% identity	Query cover	NCBI accession number
<i>M. ulcerans</i> subsp. <i>shinshuense</i>	-	98	100	WP_096370728.1
<i>M. liflandii</i> 128FXT	-	97	91	AGC62648.1
<i>M. bohemicum</i>	-	85	100	WP_085181067.1
<i>M. saskatchewanense</i>	-	85	99	WP_085257246.1
<i>M. gastri</i>	-	84	99	WP_036416250.1
<i>M. kansasii</i> ATCC 12478	-	84	96	WP_042313145.1
<i>Nocardia vinacea</i>	-	70	100	WP_040690456.1
<i>M. rhodesiae</i>	-	69	100	WP_083121537.1
<i>M. aromaticivorans</i>	-	69	100	WP_036340410.1
<i>M. vanbaalenii</i> PYR-I	147G2	68	99	WP_011777722.1
<i>Nocardia fusca</i>	-	67	100	WP_063129094.1
<i>Streptomyces tubercidus</i>	147C1	50	99	AAT45277.1
<i>Myxococcus xanthus</i>	147A1	47	99	WP_011554130.1
<i>Streptomyces bingchenggensis</i>	147F2	45	100	WP_014172979.1
<i>Rhodococcus jostii</i> RHA1	147B2	45	96	WP_011595234.1
<i>Ktedonobacter racemifer</i>	-	44	100	WP_007923086.1
<i>Streptomyces bingchenggensis</i>	147F3	44	98	WP_014177798.1
<i>Streptomyces avermitilis</i>	147B1	44	95	WP_037650463.1
<i>Streptomyces peucetius</i>	147F1	43	98	WP_100110235.1
<i>Methanosarcina barkeri</i>	147E1	42	99	WP_048107618.1
<i>Magnetospirillum magnetotacticum</i>	147D1	42	98	^b
<i>M. smegmatis</i>	164A2	34	98	3R9C_A
<i>M. tuberculosis</i> H37Rv	124A1	34	76	WP_003411654.1
<i>Bacillus subtilis</i>	107H1	32	98	P53554
	(P450BioI)			

^a CYP name given in accordance with the NCBI database and Dr Nelson P450 homepage for bacterial P450s where listed[1]. ^b could not be found on NCBI so is given in accordance with the P450 homepage only[1].

NB: Additional analogues of CYP147G1 were found in *Mycobacterium* species (between 82% and 98% similarity, and a further group between 59% and 70% which includes CYP147G2), and a number of *Nocardia* (all between 65% and 70%) and *Streptomyces* (40% and 50%) species. There were no analogues of the enzyme found between 72% and 82%. Among the analogues revealed by the BLAST search, there were additional species containing similar enzymes including *Thermoactinospira rubra* (50%), *Saccharopolyspora shandongensis* (51%), *Frankia discariae* (44%), *Actinopolymorpha singaporensis* (46%), *Arthrobacter* sp. MA-N2 (43%) and *Rhodococcus koreensis* (44%).

Table S2. Sequence identities of Fdx3 and analogous proteins from other Mycobacterium species and elsewhere. Almost all the ferredoxins listed contain the tyrosine residue replacing the third cysteine of the iron-sulfur cluster motif; exceptions are underlined. All of the associated CYPs share >40% similarity to CYP147G1. Where multiple proteins are listed from the same species, the same strain is used for all.

Species	% identity	Query cover	NCBI accession number	Associated FdR	Associated CYP
<i>M. liflandii</i> 128FXT	99	100	AGC62647.1	-	WP_020787308.1
<i>M. ulcerans</i> subsp. <i>shinshuense</i>	99	100	WP_015355775.1	-	WP_096370728.1
<i>M. kansasii</i> ATCC 12478	94	100	AGZ49085.1	WP_023364433.1	WP_042313145.1
<i>M. bohemicum</i>	94	100	ORU99153.1	WP_085181066.1	WP_085181067.1
<i>M. gastri</i>	94	100	WP_036416246.1	WP_036416248.1	WP_036416250.1
<i>M. rhodesiae</i> DSM44223	94	98	WP_083121535.1	WP_083121536.1	WP_083121537.1
<i>M. aromaticivorans</i>	93	98	WP_036346035.1	WP_036340407.1	WP_036340410.1
<i>M. saskatchewanense</i>	93	95	WP_085257248.1	WP_085257247.1	WP_085257246.1
<i>M. vanbaalenii</i> PYR-1	84	98	ABM11247.1	WP_011777721.1 ^c	WP_011777722.1
<u><i>Mycobacterium</i> sp. 1274761.0^a</u>	80	85	WP_066982528.1	-	-
<i>Nocardia fusca</i>	78	91	WP_063129092.1	WP_063129093.1	WP_063129094.1
<i>Nocardia vinacea</i>	77	91	WP_040690458.1	WP_040690457.1	WP_040690456.1
<i>M. gastri</i>	72	81	WP_036416270.1	WP_036416261.1	WP_036416260.1
<i>Ktedonobacter racemifer</i>	70	85	WP_052569463.1	WP_007923087.1	WP_007923086.1
<i>M. kansasii</i> ATCC 12478	69	85	AGZ49073.1	AGZ49074.1	AGZ49075.1
<i>Streptomyces bingchenggensis</i>	69	83	ADI08329.1	WP_014177797.1	WP_014177798.1
<i>Frankia</i> sp. CN3	67	85	WP_084174339.1	- ^c	WP_007510563.1
<i>Streptomyces avermitilis</i>	66	98	WP_037650459.1	WP_037650461.1	WP_037650463.1
<i>Streptomyces bingchenggensis</i>	66	85	ADI03498.1	WP_014172978.1	WP_014172979.1
<u><i>Streptomyces</i> sp. WM6378^b</u>	64	85	WP_078966815.1	WP_014677360.1	-
<i>Streptomyces tubercidus</i>	62	85	AAT45278.1	AAT45279.1	AAT45277.1
<i>Methanosarcina barkeri</i>	62	84	WP_048110266.1	WP_048107620.1	WP_048107618.1
<i>Myxococcus xanthus</i> DK 1622	62	81	ABF90123.1	- ^c	ABF88234.1
<i>Rhodococcus jostii</i> RHA1	61	87	WP_011595232.1	WP_009475197.1	WP_011595234.1
<i>Streptomyces peucetius</i>	60	94	WP_100110237.1	WP_100110236.1	WP_100110235.1
<i>Rhodopseudomonas palustris</i>	54	37	4ID8_A		
<i>Clostridium thermoaceticum</i> ^d	28	81	WP_075515755.1		
<i>Thermococcus litoralis</i>	27	100	WP_004067745.1		

^a The tyrosine residue in the cluster binding motif is replaced with Cys (there is an associated peroxygenase, WP_066982530.1) ^b Tyr residue replaced with His ^c fusion protein (of Fdx and FdR domains) ^d species renamed *Moorella thermoacetica*. ^e has a second FdR protein co-located (WP_011777719.1) NB: There were a large number of Fdx analogues in *Mycobacterium* species (with sequence similarities from 100% down to 69%). There are two Tyr containing Fdx in several *Mycobacterium* species such as *M. kansasii* (94 and 69%) and *M. gastri* (94 and 72%). *Rhodococcus jostii* RHA1 also contains a second Tyr-containing Fdx (43%) while *S. avermitilis* contains three (66%, 59% and 56%, all with associated CYPs and the third being a fused Fdx/FdR, WP_037648248.1). Analogues of the Fdx found in *Nocardia* have similarities between at 77% and 80% while in *Streptomyces* were at 75% and below. In the first 250 results from the BLAST search, only two did not have the tyrosine residue indicating it is highly conserved. They are listed in the table.

Table S3: Sequence identities of FdR1 and analogous proteins from other *Mycobacterium* species and elsewhere.

Species	% identity	Query cover	NCBI accession number
<i>M. kansasii</i> ATCC 12478	91	99	AGZ49084.1
<i>M. gastri</i>	91	99	WP_036416248.1
<i>M. saskatchewanense</i>	90	99	WP_085257247.1
<i>M. bohemicum</i>	90	99	WP_085181066.1
<i>M. aromaticivorans</i>	83	97	WP_036340407.1
<i>M. rhodesiae</i> DSM44223	82	97	WP_083121536.1
<i>M. vanbaalenii</i> PYR-1	81	98	WP_011777721.1
<i>Nocardia vinacea</i>	79	98	WP_040690457.1
<i>Nocardia fusca</i>	75	99	WP_063129093.1
<i>Thermoactinospira rubra</i>	74	99	WP_084959902.1
<i>Streptomyces tubercidicus</i>	74	99	AAT45279.1
<i>Saccharopolyspora antimicrobica</i>	64	96	WP_093160449.1
<i>Microtetraspora niveoalba</i> ^a	63	96	WP_067168930.1
<i>Ktedonobacter racemifer</i>	63	95	WP_007923087.1
<i>Streptomyces avermitilis</i>	63	95	WP_037650461.1
<i>Streptomyces olivochromogenes</i> ^a	62	96	WP_067381331.1
<i>Arthrobacter</i> sp. ok909 ^a	61	98	WP_091252245.1
<i>Mycococcus xanthus</i> DK 1622 ^a	61	95	ABF90123.1
<i>M. kansasii</i> ATCC 12478	60	98	AGZ49074.1
<i>Streptomyces bingchenggensis</i>	60	97	WP_014177797.1
<i>M. kansasii</i> ATCC 12478	60	96	AGZ49074.1
<i>Frankia</i> sp. CN3 ^a	60	96	WP_084174339.1
<i>Streptomyces bingchenggensis</i>	59	97	WP_014172978.1
<i>Methanosarcina barkeri</i>	57	93	WP_048107620.1

^a The protein is a fusion of ferredoxin reductase and ferredoxin domains. Additional fusions were found in the following species: *Arthrobacter* sp. ov407, *Arthrobacter* sp. 49Tsu3.1M3, *Streptomyces* sp. NRRL S-1824 and *Streptomyces* sp. 93. In all cases the Tyr residue of the ferredoxin cluster binding motif was conserved.

NB: Similar ferredoxin reductases are found in many *Mycobacterium* species (down to 62%, although below 72% all are from *M. kansasii*), *Nocardia* species (from 75% to 78%) and *Streptomyces* (from 75% and below). There are two similar proteins in *M. kansasii* (91% and 60%) but only one in *M. gastri* and it is not present at all in *M. liflandii* (the highest similarity protein from that species is 37% according to the BLAST search, and the gene next to the ferredoxin in the genome is frameshifted) or any *M. ulcerans* strains (also 37% similarity). There are also two copies in *S. bingchenggensis* (at 60% and 59% similarities) and three in *S. avermitilis* although they have lower sequence similarities (58%, 43% and 37%).

Table S4: Conserved domains of the CYP147 family of enzymes. The conserved residues are in bold, alterations to those residues are in red and underlined.

CYP name	I-helix	K-helix	Heme binding motif
147A1	GH ETT V	ELLR	FGSGIH YC
147B1	GH ETT V	ELLR	FGSGIH LC
147B2	GH ETT V	ELLR	<u>L</u> GS GIH SC
147C1	GH ETT V	ELLR	FSQGIH FC
147D1	GH ETT V	ELLR	FGGGLH YC
147E1	GH ETT V	EFLR	FGSGIH YC
147F1	GH ETT V	ELLR	<u>L</u> GS GIH SC
147F2	GH ETT V	EVL R	FYTGIH YC
147F3	GH ETT V	EVL R	FGSGIH IC
147G1	GH <u>DS</u> TV	EVQR	<u>W</u> GS GIH TC
147G2	GH <u>DS</u> TV	EVQR	<u>W</u> GS GIH TC

Table S5: Additional spin state shifts of CYP147G1 with a variety of substrates.

Substrate	Spin state shift (% HS)
Linoleic	30
Behenic acid	25
<i>n</i> -Propylbenzoic acid	20
<i>n</i> - <i>t</i> -Butylbenzoic acid	20
Octadec-9-enoic acid	20
Eicosanoic acid	15
2-Naphthoic acid	10
4-Heptylbenzoic acid	≤5
2-Indanone	≤5
Indole	≤5

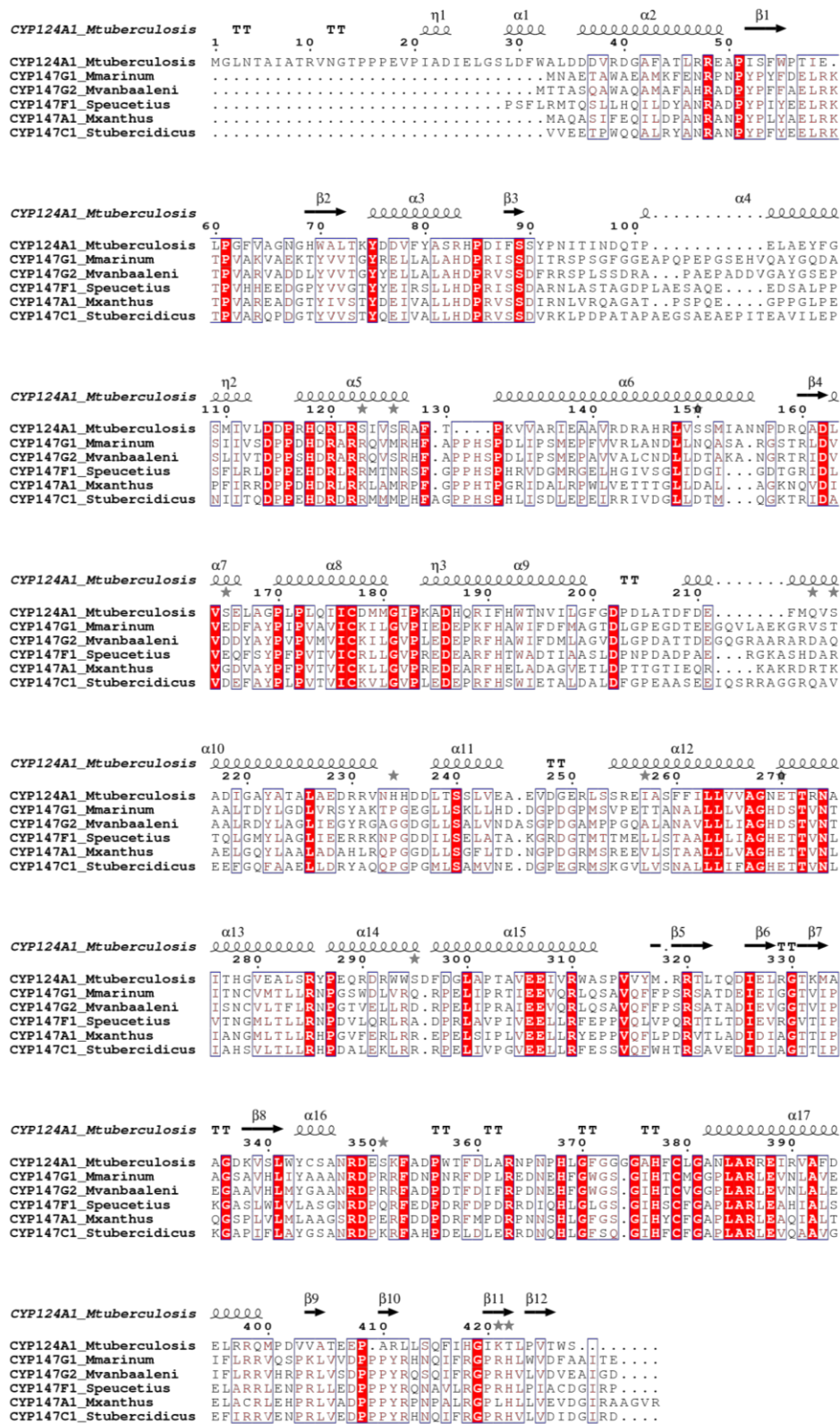


Figure S1: Protein sequences of CYP147G1, CYP147G2 enzyme from *Mycobacterium vanbaalenii* PYR1, CYP147A1 from *Myxococcus xanthus*, CYP147C1 from *Streptomyces tubercidicus* and CYP147F1 from *Streptomyces peucetius* aligned to the structural elements of CYP124A1 from *Mycobacterium tuberculosis* (PDB: 2WM5)[2] using ESPript. [3]

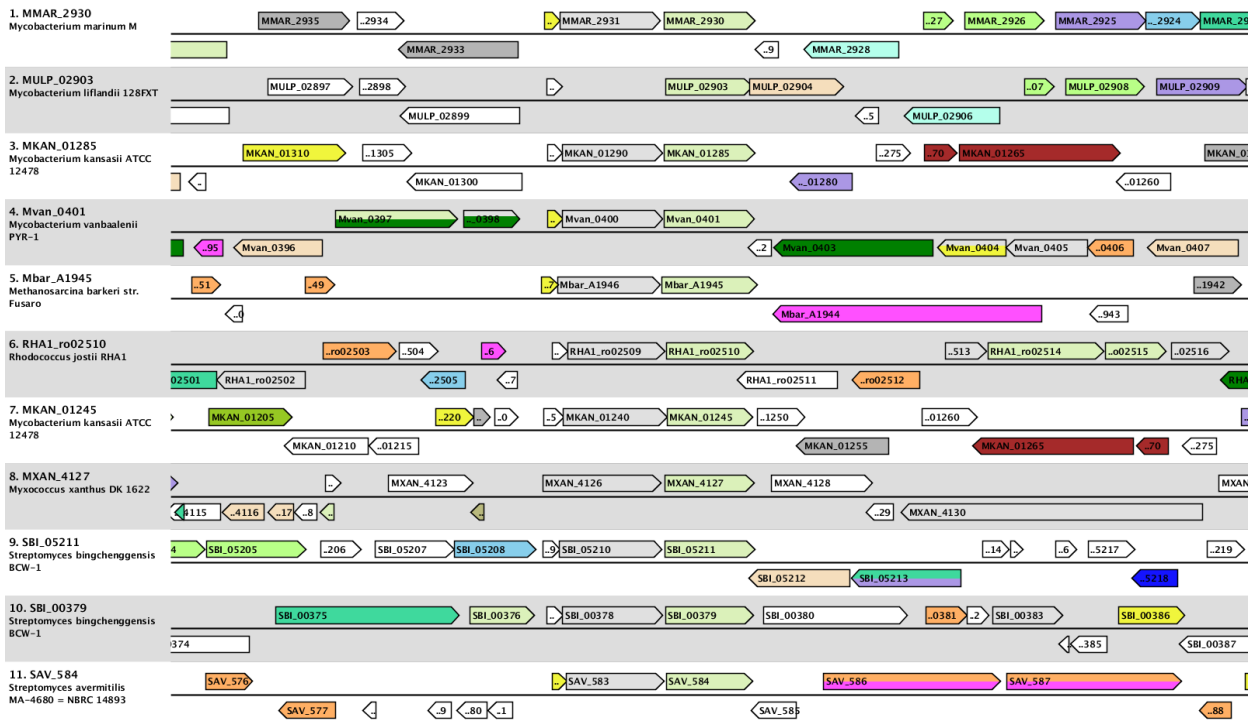


Figure S2: The Microbial Genomic Context Viewer (MGcV)[4] was used to compare the surrounding genome regions (40 kbp) of homologous CYPs. Genes are labelled by their locus tag and coloured by COG (Clusters of Orthologous Groups). The genes surrounding the *Mmar_2930* gene encoding CYP147G1 (green) from *M. marinum* M are compared to those from *M. liflandii* 128FXT (*Mulp_02093*), *M. kansasii* ATCC (*Mkan_01285* and *Mkan_01245*), *M. vanbaalenii* PYR-1 (*Mvan_0401*), *Methanosarcina bakeri* str. Fusaro (*Mbar_A1945*), *Rhodococcus jostii* RHA1 (*RHA1_ro02510*), *Myxococcus xanthus* DK 1622 (*Mxan_4127*), *Streptomyces bingchenggensis* BCW-1 (*Sbi_05211* and *Sbi_00379*) and *Streptomyces avermitilis* (*Sav_584*). While the surrounding gene regions show little similarity between species, the CYP147G1 is conserved in almost all instances alongside genes encoding an FdR (*Mmar_2931*) and Fdx (*Mmar_2932*). The exceptions to this are *M. liflandii* 128FXT which is missing the FdR gene, and *M. xanthus* which has a single, fused FdR/Fdx protein (*Mxan_4126*). Two proteins, a PE-PGRS family protein (*Mmar_2933*) and a cutinase (*Mmar_2934*) are conserved in the *Mycobacterium* species *M. kansasii* and *M. liflandii*.

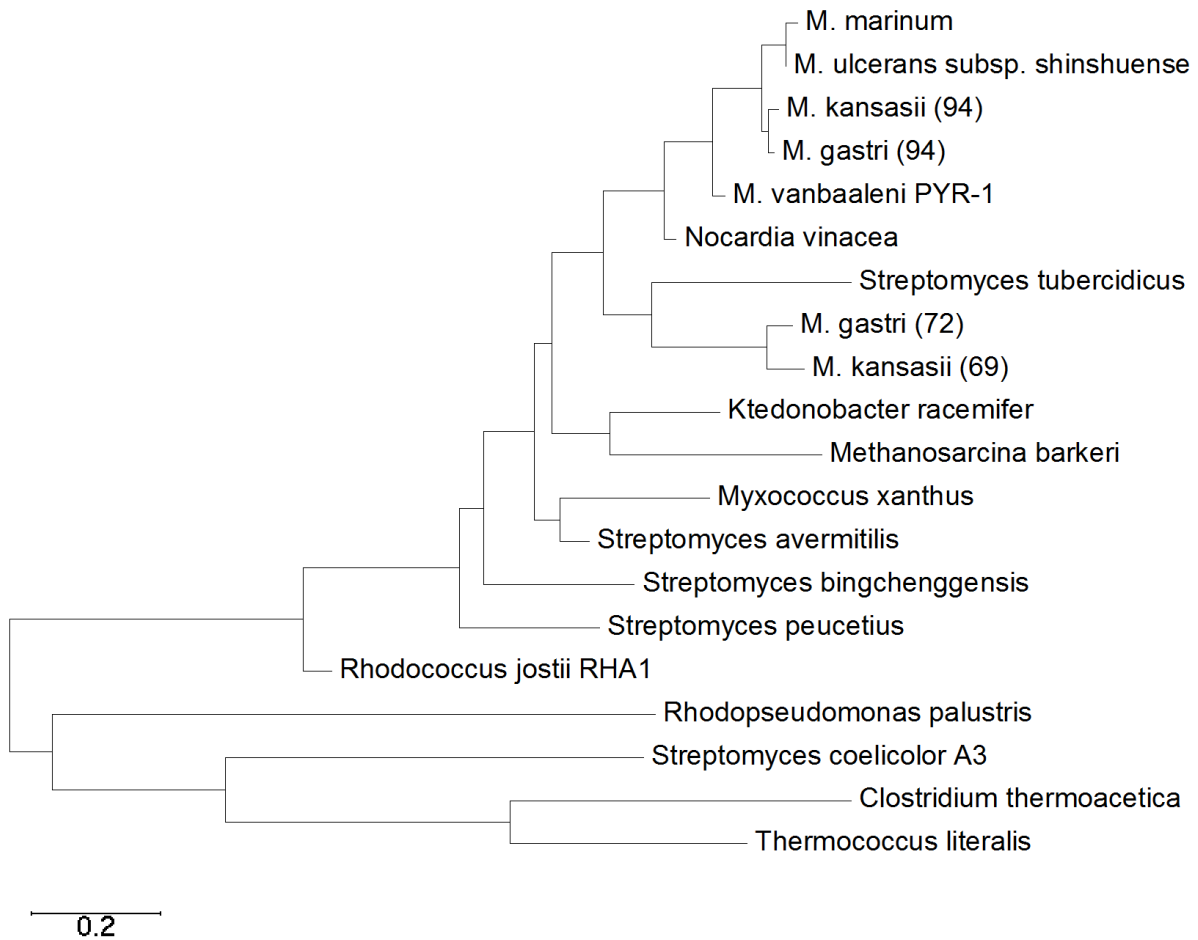
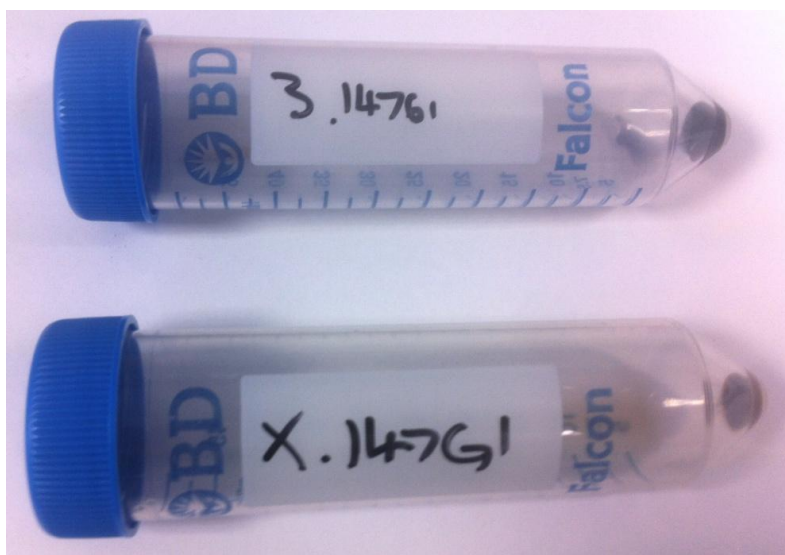
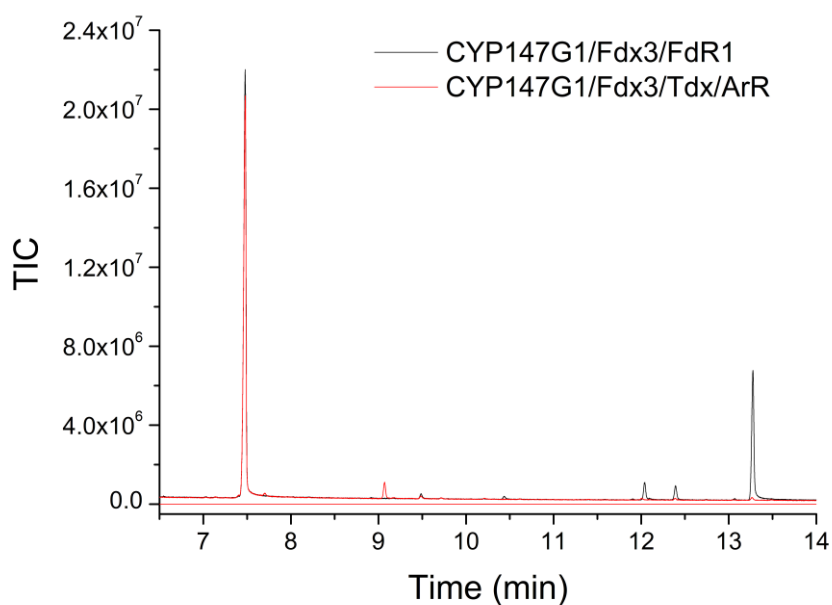


Figure S3: A phylogenetic tree (phenogram) of Fdx3 from *M. marinum*, and the similar proteins from *Mycobacterium* and other species (all listed by species name, for protein sequence similarities and accession numbers see Table S2). Where two similar ferredoxins exist in the same genome (eg. *M. kansasii*), the percentage similarity of each to Fdx3 is shown in parentheses. The ferredoxin in *Myxococcus xanthus* is a ferredoxin-ferredoxin reductase fusion protein. For comparison the [3Fe-4S] ferredoxins *Rhodopseudomonas palustris* (His), *Streptomyces coelicolor* (Ala), and the [4Fe-4S] ferredoxins from *Clostridium thermoacetica* and *Thermococcus litoralis* (Cys) have been included. Aside from those four, all the ferredoxins shown have a tyrosine residue in the iron-sulfur cluster binding motif (CXXYXXC(X)_nCP). The scale shows number of substitutions per site.



(a)



(b)

Figure S4: (a) CYP147G1-mediated indigo formation (via the hydroxylation of indole), as qualitatively measured by the blue colour in the cell pellet, is reduced when coupled with the Fdx/FdR fusion gene *pp_1957* from *Pseudomonas putida* [5] (bottom tube) compared to expression with the Fdx3/FdR1 native pair (top tube) and (b) the hydroxylation of undecanoic acid by CYP147G1 is not supported by the Tdx/ArR pair (from *Pseudomonas sp* [6] and *Novosphingobium aromaticivorans*[7] respectively) used to support CYP268A2 activity [8] (RT undecanoic acid 7.5 min, 10-hydroxyundecanoic acid 13.25 min). This highlights that CYP147G1 activity is reduced when coupled with non-native electron transfer partners and the selectivity of the enzyme for its physiological partners.

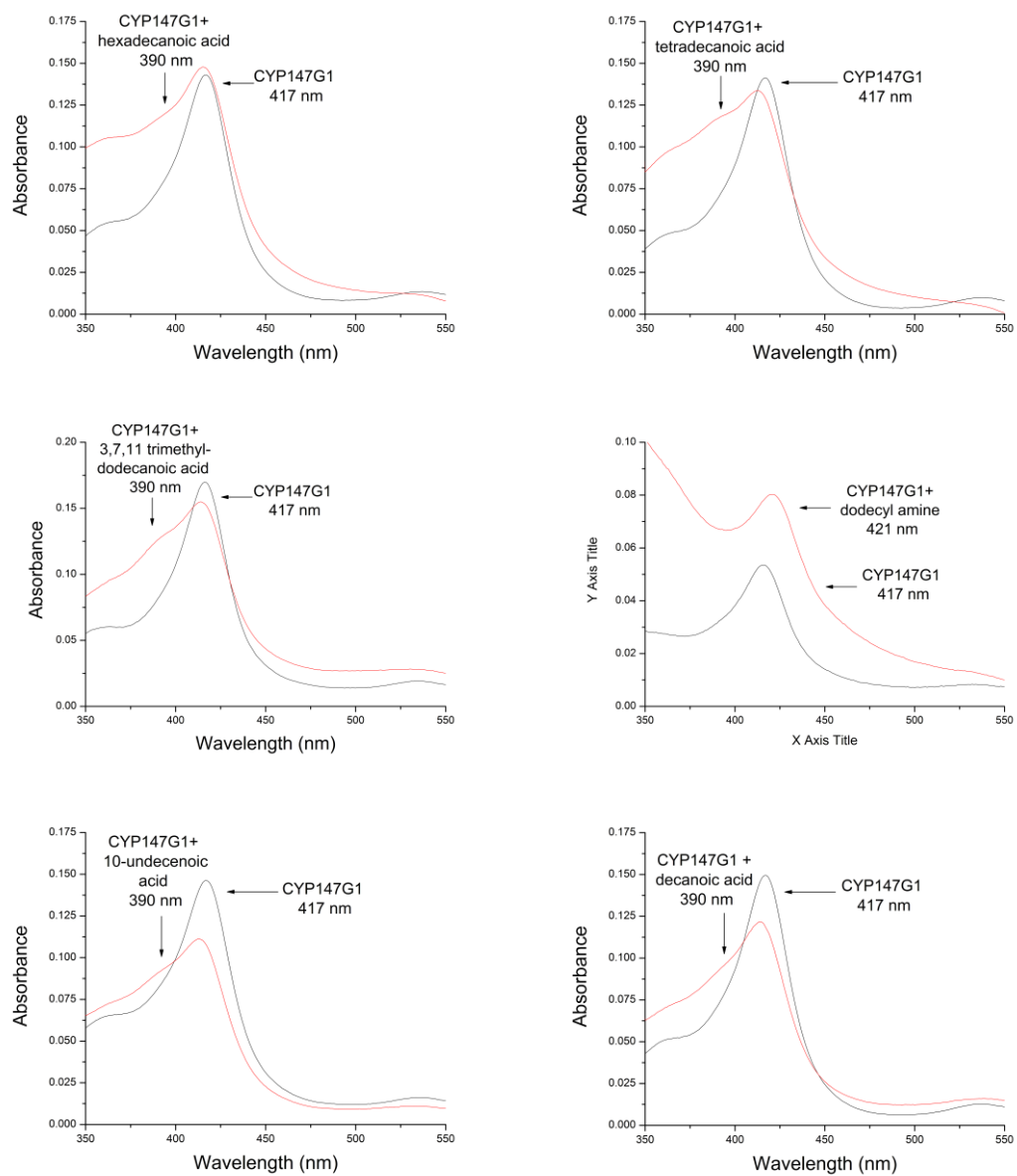


Figure S5: Selected spin state shifts of CYP147G1 for a variety of substrates.

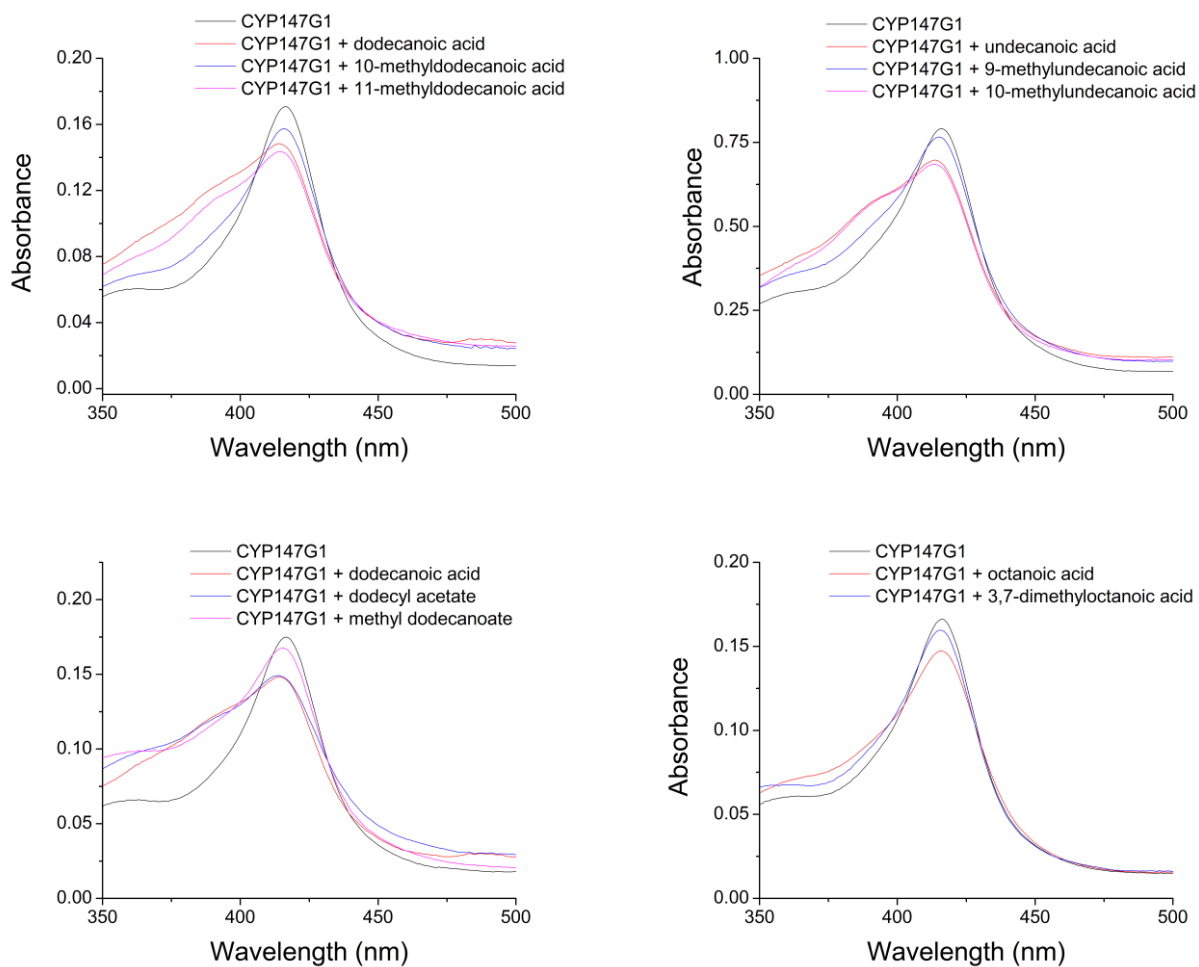
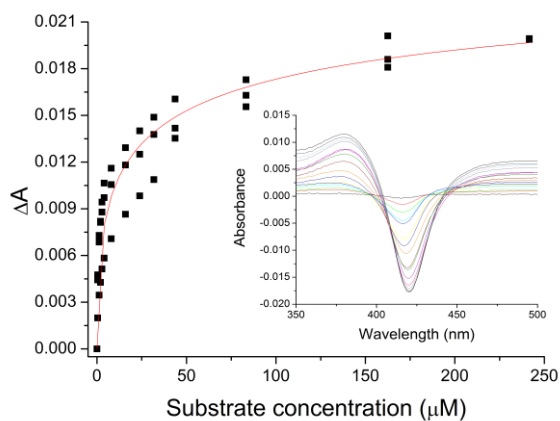
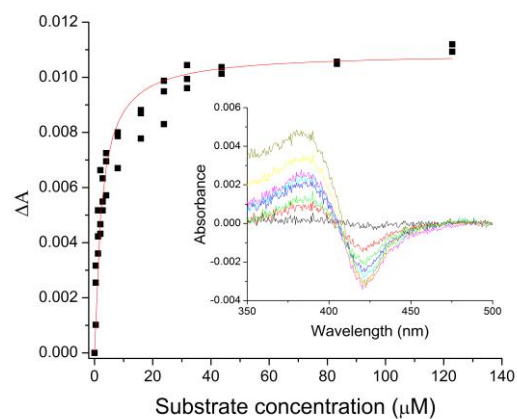


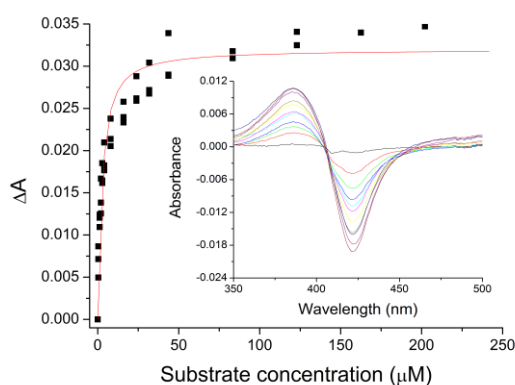
Figure S6: Spin state shifts of various substrates with CYP147G1 including (a) dodecanoic acid, 10-methyldodecanoic acid and 11-methyldodecanoic acid (b) undecanoic acid, 9-methylundecanoic acid and 10-methylundecanoic acid, (c) dodecanoic acid, dodecyl acetate and methyl dodecanoate and (d) octanoic acid and 3,7-dimethyloctanoic acid.



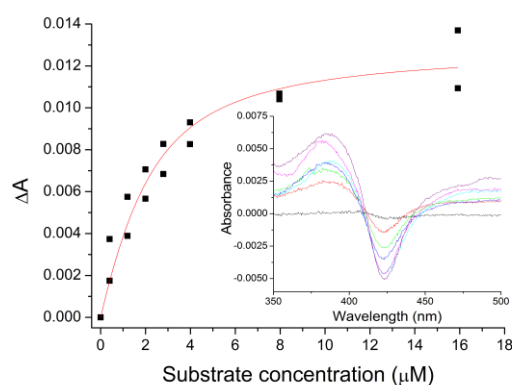
(a) 10-Undecenoic acid



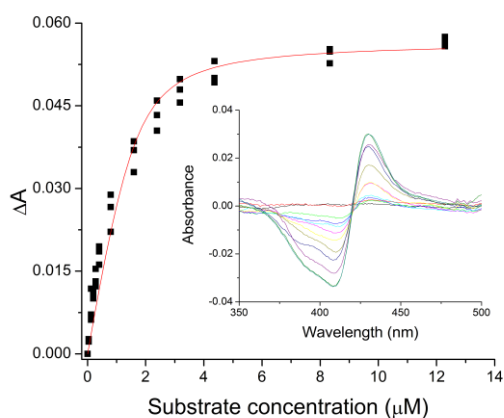
(b) Dodecyl acetate



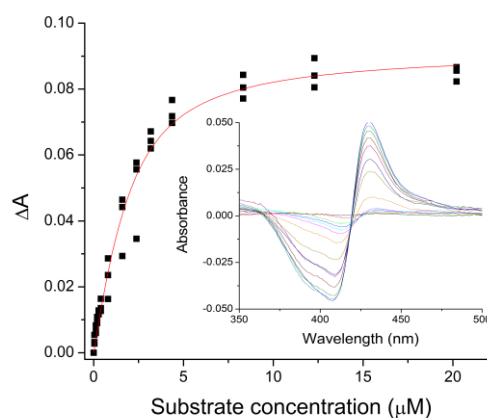
(c) Farnesol



(d) Farnesyl acetate



(e) Miconazole

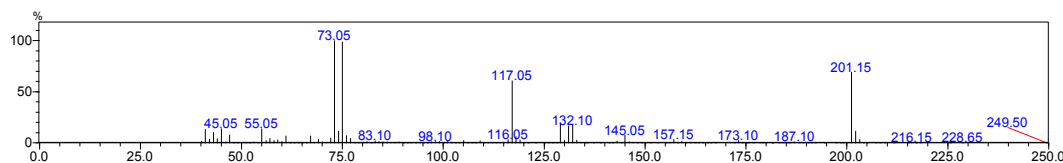


(f) Econazole

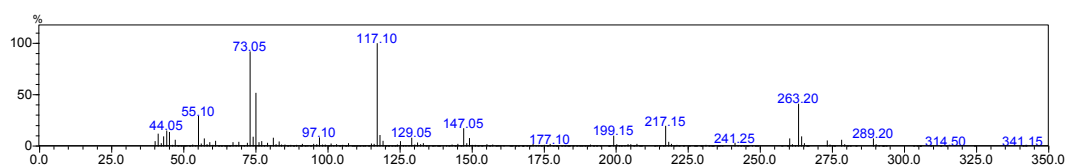
Figure S7: Additional binding constants with CYP147G1. The inset represents a typical substrate binding titration. The peak to trough difference in absorbance was measured from 420 nm to 390 nm in each, except with miconazole and econazole where a Type II shift was recorded (peak to trough 409 to 430 nm for both).

Figure S8: Mass spectra of TMSCl derivatised fatty acid substrates and their hydroxylated metabolites

Octanoic acid Fig 3(a)

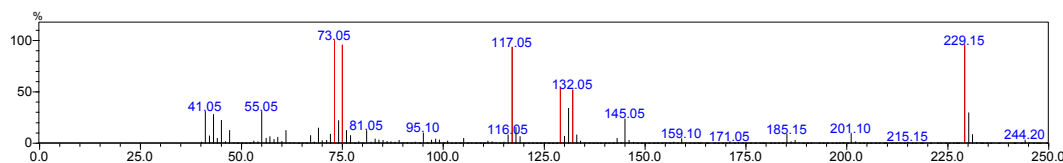


(a) MS of peak at 4.3 min – substrate octanoic acid (expected mass 216.15 m/z, -15 peak seen at 201.15 m/z)

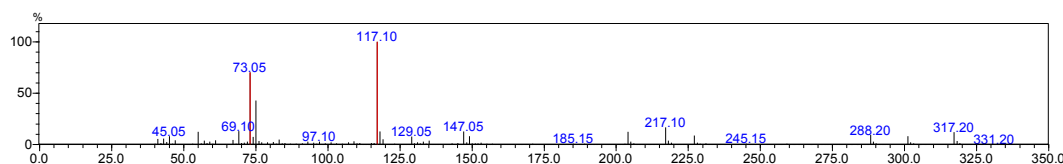


(b) MS of peak at 9.1 min – octanoic acid product ω -1 hydroxylation (expected mass 304.19 m/z, -15 peak seen at 289.20 as well as peak at 117.10 m/z)

Decanoic acid Fig 3(b)



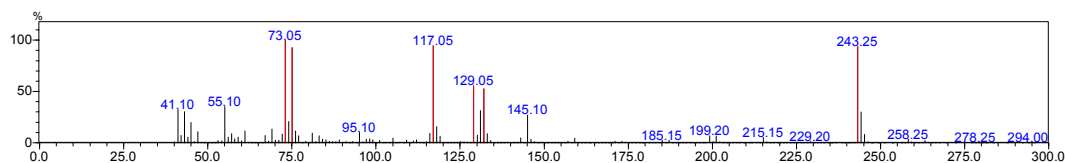
(c) MS of peak at 7.5 min – substrate decanoic acid (expected mass 244.19 m/z, -15 peak seen at 229.15 m/z)



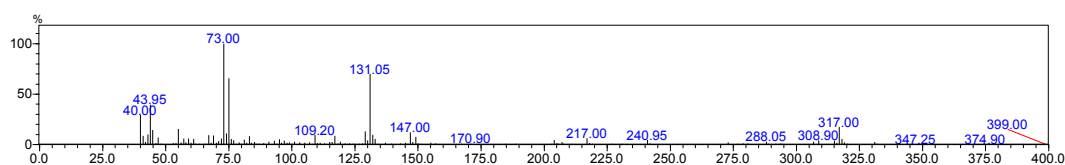
(d) MS of peak at 12.4 min – decanoic acid product ω -1 hydroxylation (expected mass 332.22 m/z, -15 peak seen at 317.20 as well as peak at 117.10 m/z)

Figure S8 (cont): Mass spectra of TMSCl derivatised fatty acid substrates and their hydroxylated metabolites

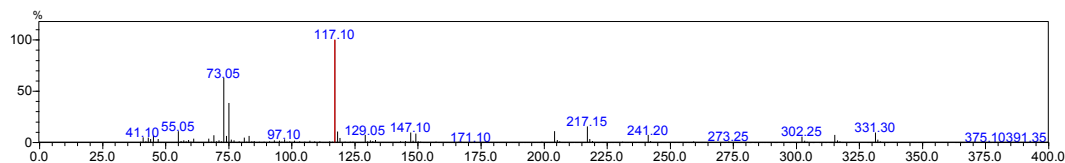
Undecanoic acid Fig 3(c)



(e) MS of peak at 9.25 min – substrate undecanoic acid (expected mass 258.20 m/z, -15 peak seen at 243.25 m/z)



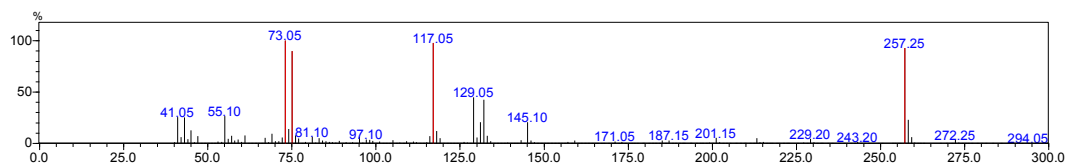
(f) MS of peak at 13.2 min – undecanoic acid minor product ω -2 hydroxylation (expected mass 346.24 m/z, peak seen at 347.25 as well as distinctive ω -2 peaks at 317.00 m/z, -29, and 131.05 m/z)



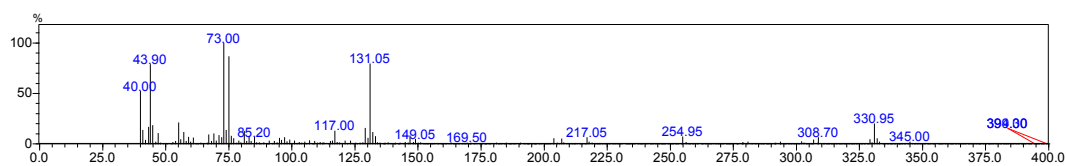
(g) MS of peak at 13.9 min – undecanoic acid product ω -1 hydroxylation (expected mass 346.24 m/z, -15 peak seen at 331.30 as well as peak at 117.10 m/z)

Figure S8 (continued): Mass spectra of TMSCl derivatised fatty acid substrates and their hydroxylated metabolites

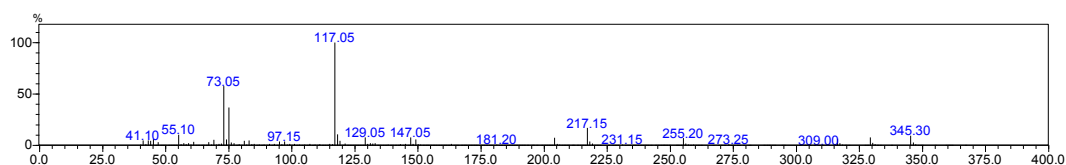
Dodecanoic acid Fig 3(d)



(h) MS of peak at 11.0 min – substrate dodecanoic acid (expected mass 272.22 m/z, -15 peak seen at 257.25 m/z)



(i) MS of peak at 14.5 min – dodecanoic acid minor product ω -2 hydroxylation (expected mass 360.25 m/z, peak seen at 345.05 m/z as well as peaks at 330.95 m/z and 131.05 m/z)



(j) MS of peak at 15.4 min – dodecanoic acid product ω -1 hydroxylation (expected mass 360.25 m/z, -15 peak seen at 345.30 as well as peak at 117.05 m/z)

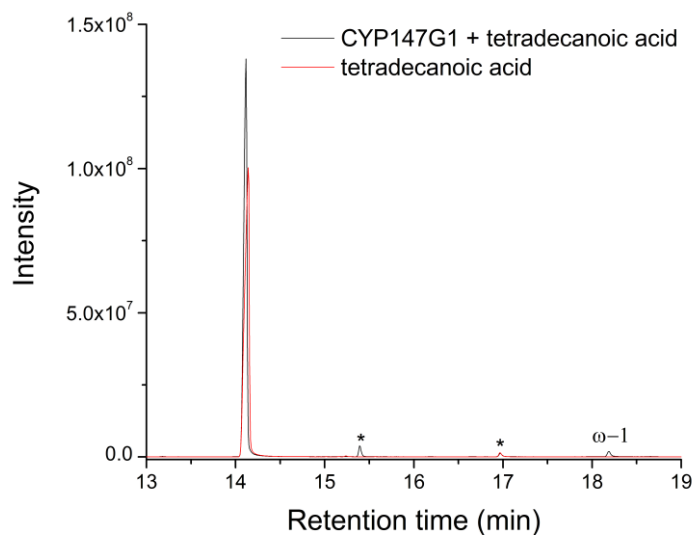
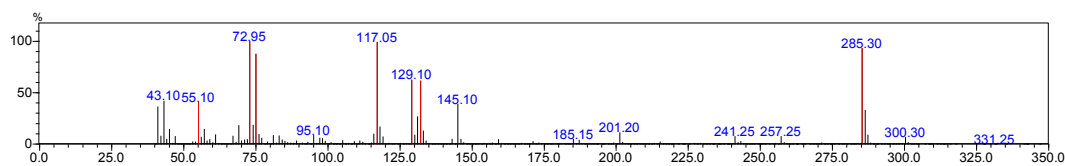
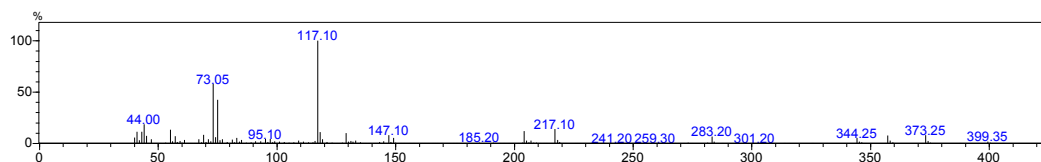


Figure S9: (a) GC trace of CYP147G1 with tetradecanoic acid. * indicates impurities identified from MS as the ω -1 hydroxylation product of dodecanoic acid (15.4 min) and hexadecanoic acid (17.0 min). Dodecanoic acid is also present as an impurity in both the turnover and the substrate control (11.0 min).



(b) MS of peak of (a) at 14.1 min – substrate tetradecanoic acid (expected mass peak at 300.25 m/z seen, -15 peak seen at 285.30 m/z)



(c) MS of peak at 18.2 min – tetradecanoic acid product ω -1 hydroxylation (expected mass 388.28 m/z, -15 peak seen at 373.25 as well as peak at 117.10 m/z)

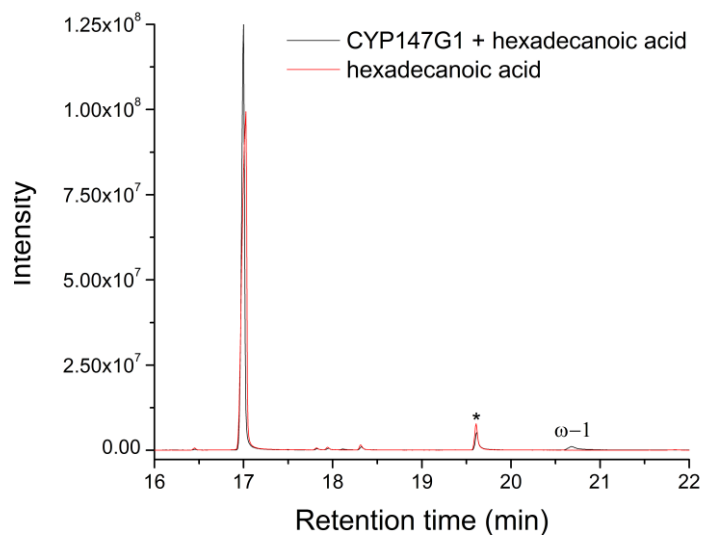
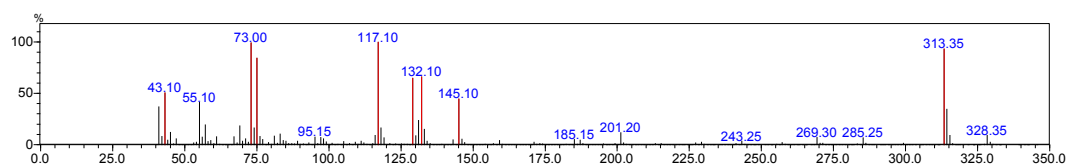
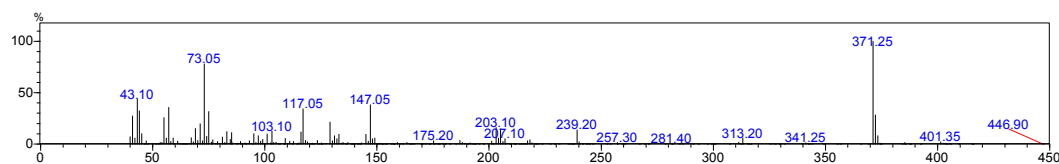


Figure S10: (a) GC trace of CYP147G1 with hexadecanoic acid. * indicates an impurity identified from MS as octadecanoic acid.



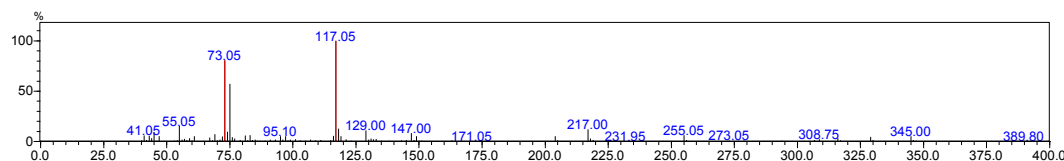
(b) MS of peak of (a) at 17.0 min – substrate hexadecanoic acid (expected mass peak at 328.28 m/z seen, -15 peak seen at 313.35 m/z)



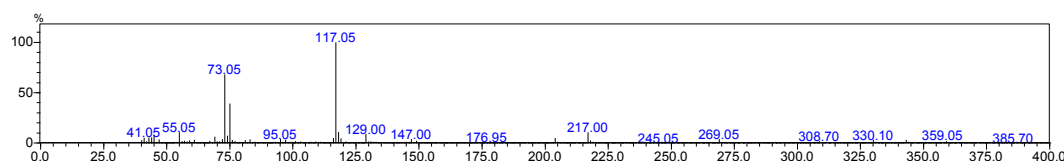
(c) MS of peak at 20.7 min – hexadecanoic acid product $\omega-1$ hydroxylation (expected mass 416.31 m/z , -15 peak seen at 401.35 as well as peak at 117.05 m/z .)

Figure S11: Mass spectra of TMSCl derivatised linear and branched dodecanoic acid substrates and their hydroxylated metabolites

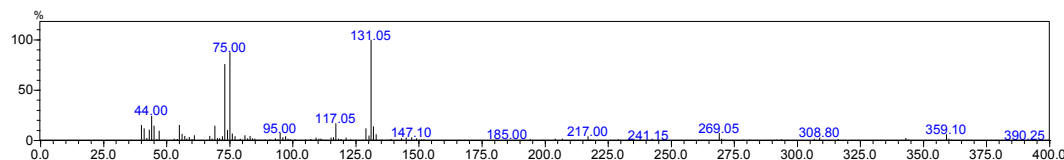
Dodecanoic acid, 10-methyldodecanoic acid, 11-methyldodecanoic acid, Fig. 4(a)



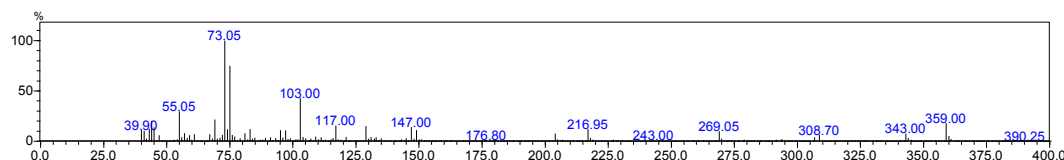
(a) MS of GC peak at 14.7 min – dodecanoic acid ω -1 hydroxylation product (expected mass 360.25 m/z, -15 peak seen at 345.00 as well as peak at 117.05 m/z)



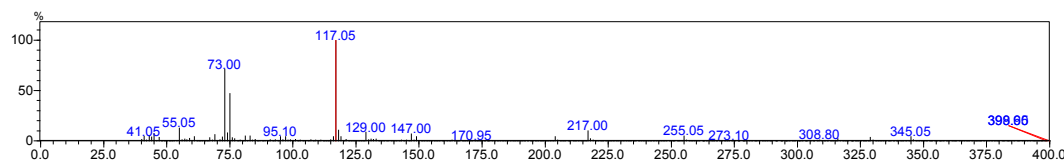
(b) MS of GC peak at 15.4 min –10-methyldodecanoic acid ω -1 hydroxylation product (expected mass 374.27 m/z, -15 peak seen at 359.05 as well as peak at 117.05 m/z)



(c) MS of GC peak at 15.35 min –11-methyldodecanoic acid ω -1 hydroxylation product (expected mass 374.27 m/z, -15 peak seen at 359.10 as well as peak at 131.05 m/z, indicating the ω -1 product in a methyl branched substrate as the fragment is now $\text{CH}_3\text{CCH}_3\text{OSiMe}_3^+$)



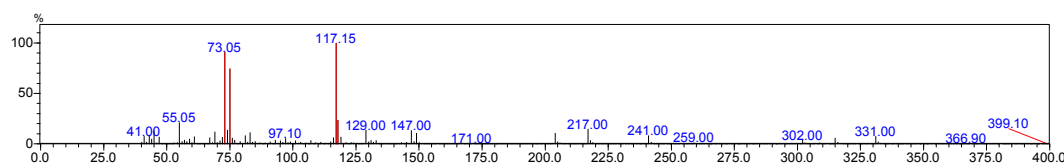
(d) MS of GC peak at 16.0 min –11-methyldodecanoic acid ω hydroxylation product (expected mass 374.27 m/z, -15 peak seen at 359.00 as well as peak at 103.00 m/z)



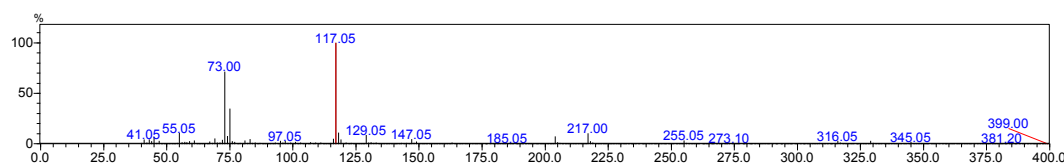
(e) MS of GC peak at 14.7 min – methyl dodecanoate acid ω -1 hydroxylation product (expected mass 360.25 m/z if the acetate group is cleaved in derivatisation, -15 peak seen at 345.05 as well as peak at 117.05 m/z)

Figure S12: Mass spectra of TMSCl derivatised linear and branched undecanoic acid substrates and their hydroxylated metabolites

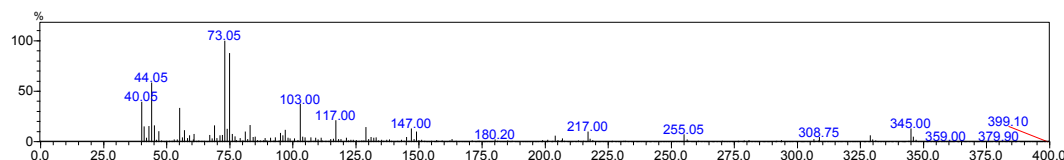
Undecanoic acid, 9-methylundecanoic acid, 10-methylundecanoic acid, Fig. 4(b)



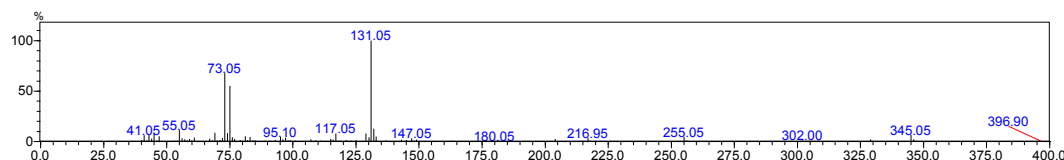
(a) MS of GC peak at 13.4 min – undecanoic acid ω -1 hydroxylation product (expected mass 346.24 m/z, -15 peak seen at 331.00 as well as peak at 117.15 m/z)



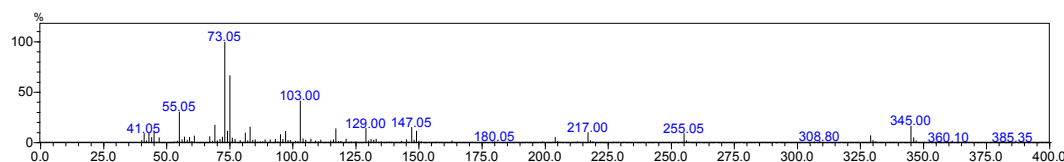
(b) MS of GC peak at 14.1 min – 9-methylundecanoic acid ω -1 hydroxylation product (expected mass 360.25 m/z, -15 peak seen at 345.05 as well as peak at 117.05 m/z)



(c) MS of GC peak at 14.4 min – 9-methylundecanoic acid ω hydroxylation product (expected mass 360.25 m/z, -15 peak seen at 345.00 as well as peak at 103.00 m/z)



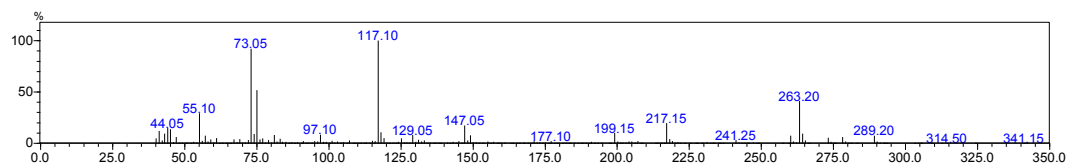
(d) MS of GC peak at 14.0 min – 10-methylundecanoic acid ω -1 hydroxylation product (expected mass 360.25 m/z, -15 peak seen at 345.05 as well as peak at 131.05 m/z, indicating the ω -1 product in a methyl branched substrate)



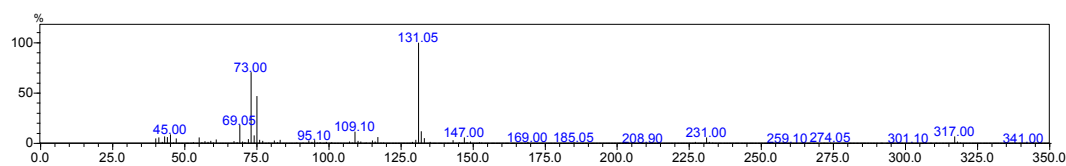
(e) MS of GC peak at 14.75 min – 10-methylundecanoic acid ω hydroxylation product (expected mass 360.25 m/z, -15 peak seen at 345.00 as well as peak at 103.00 m/z)

Figure S13: Mass spectra of TMSCl derivatised linear and branched octanoic acid substrates and their hydroxylated metabolites

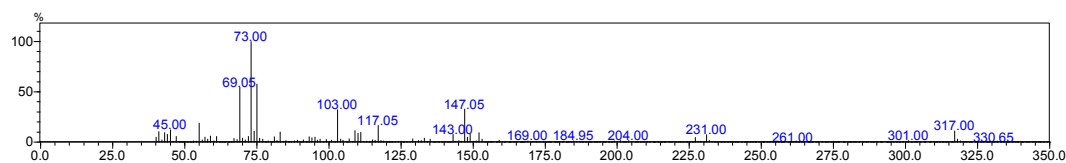
Octanoic acid, 3,7-dimethyloctanoic acid, Fig. 4(c)



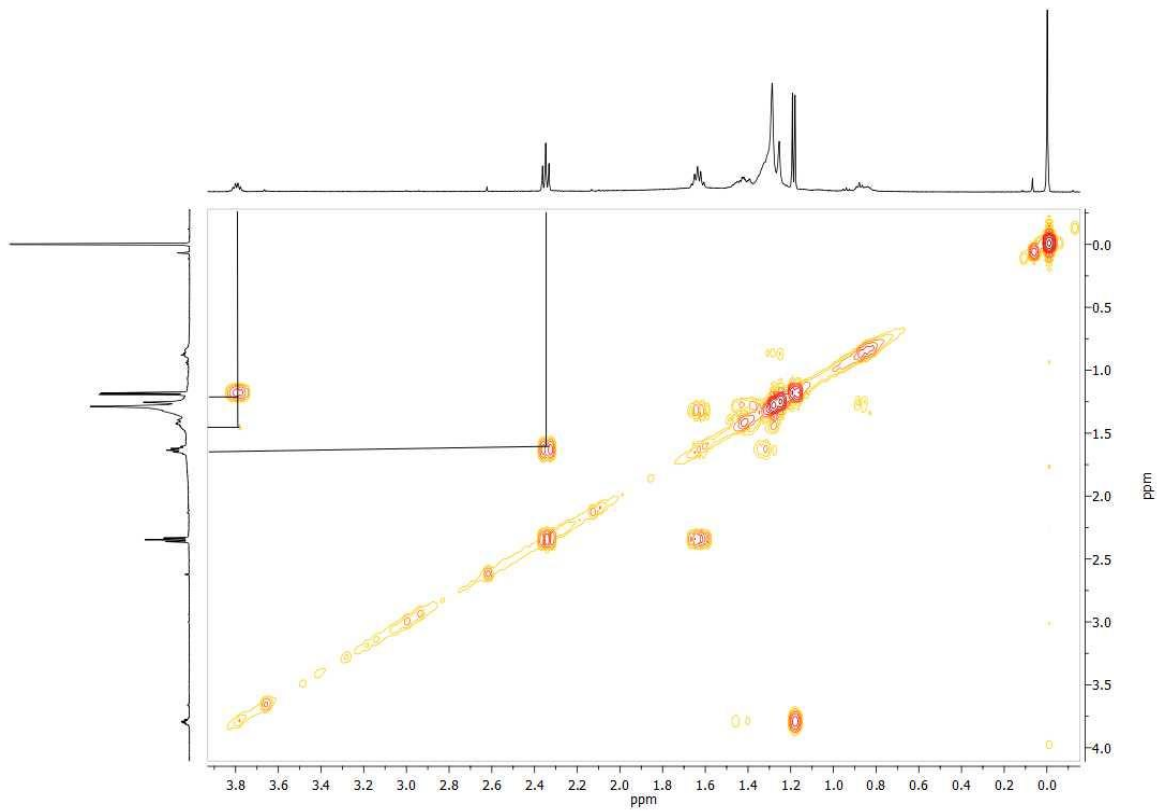
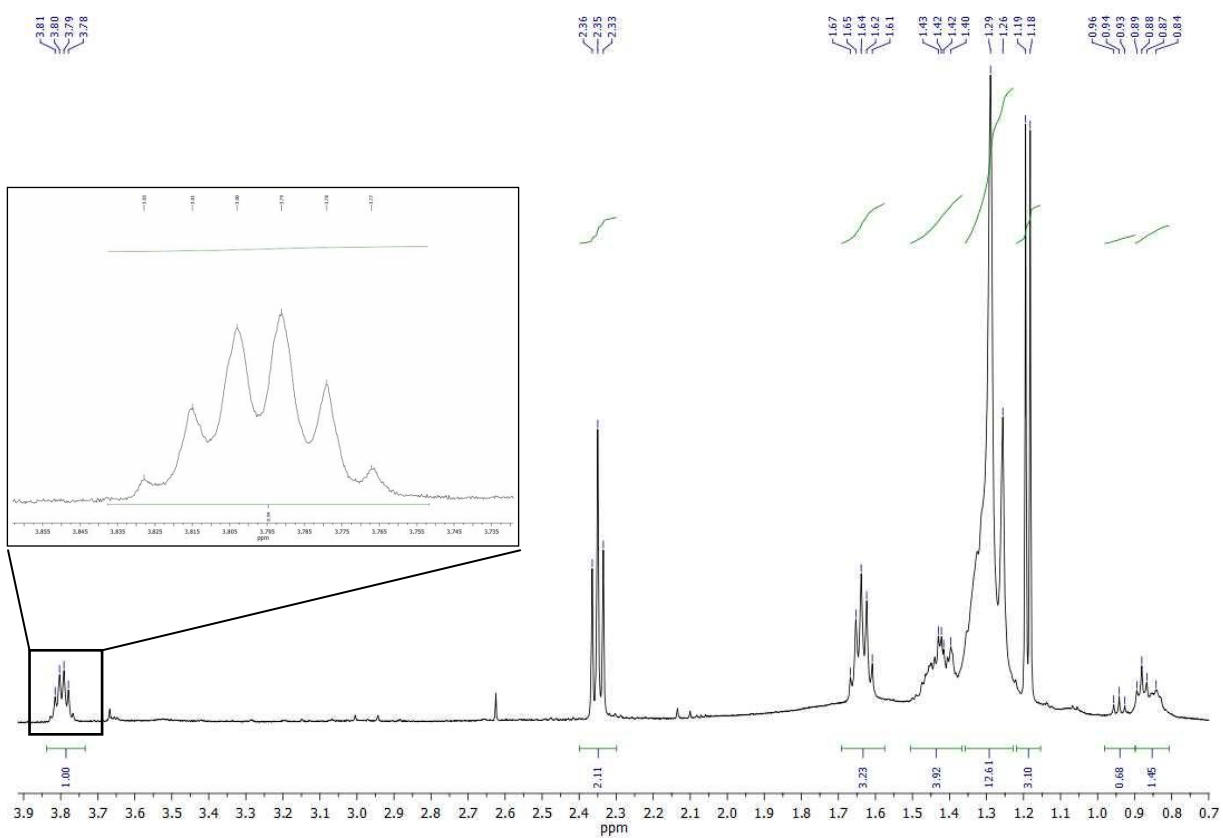
(a) MS of GC peak at 9.2 min – octanoic acid ω -1 hydroxylation product (expected mass 304.19 m/z, -15 peak seen at 289.20 as well as peak at 117.10 m/z)



(b) MS of GC peak at 10.1 min – 3,7-dimethyloctanoic acid ω -1 hydroxylation product (expected mass 332.22 m/z, -15 peak seen at 317.00 as well as peak at 131.05 m/z, indicating the ω -1 product in a methyl branched substrate)



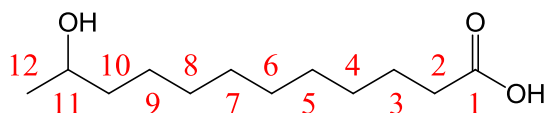
(c) MS of GC peak at 11.0 min – 3,7-dimethyloctanoic acid ω hydroxylation product (expected mass 332.22 m/z, -15 peak seen at 317.00 as well as peak at 103.00 m/z)



(a)

(b)

Figure S14: (a) ^1H NMR of the product of the CYP147G1 turnover with dodecanoic acid, showing an expansion of the diagnostic triplet of quartets at δ 3.8 ppm (b) COSY-NMR of the product.



Scheme S1: Dodecanoic acid hydroxylation product of CYP147G1. ^1H NMR (500 MHz, CDCl_3) δ 3.8 (tq, $J = 6.2$ Hz, 1H, **H11**), 2.35 (t, $J = 7.4$ Hz, 2H, **H2**), 1.64 (tt, $J = 7.4$ Hz, 2H, **H3**), 1.42 (m, 2H, **H10**), 1.27 (broad m, 12H, **H4-9**), 1.19 (d, $J = 6.2$ Hz, 3H, **H12**). ^{13}C NMR (126 MHz, CDCl_3) δ 68.23, 39.32, 33.55, 29.71, 29.52, 29.44, 29.26, 29.12, 28.99, 25.70, 24.67, 23.50.

The triplet of quartets at approximately δ 3.8 ppm is distinctive for ω -1 hydroxylation (Figure S13). The 3H doublet peak at δ 1.19 ppm confirms the hydroxylation of the substrate at the ω -1 position, as it can be assigned by COSY as the terminal CH_3 group, split by the single H on the CHOH group.

References

- [1] D.R. Nelson, The cytochrome p450 homepage, *Human genomics*, 4 (2009) 59-65.
- [2] J.B. Johnston, P.M. Kells, L.M. Podust, P.R. Ortiz de Montellano, Biochemical and structural characterization of CYP124: a methyl-branched lipid omega-hydroxylase from *Mycobacterium tuberculosis*, *Proc. Natl. Acad. Sci. U. S. A.*, 106 (2009) 20687-20692.
- [3] X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server, *Nucleic Acids Research*, 42 (2014) W320-W324.
- [4] L. Overmars, R. Kerkhoven, R.J. Siezen, C. Francke, MGcV: the microbial genomic context viewer for comparative genome analysis, *BMC genomics*, 14 (2013) 209.
- [5] S.G. Bell, L. French, N.H. Rees, S.S. Cheng, G. Preston, L.L. Wong, A phthalate family oxygenase reductase supports terpene alcohol oxidation by CYP238A1 from *Pseudomonas putida* KT2440, *Biotechnol. Appl. Biochem.*, 60 (2013) 9-17.
- [6] S.G. Bell, W. Yang, J.A. Yorke, W. Zhou, H. Wang, J. Harmer, R. Copley, A. Zhang, R. Zhou, M. Bartlam, Z. Rao, L.L. Wong, Structure and function of CYP108D1 from *Novosphingobium aromaticivorans* DSM12444: an aromatic hydrocarbon-binding P450 enzyme, *Acta Crystallogr. D*, 68 (2012) 277-291.
- [7] S.G. Bell, A. Dale, N.H. Rees, L.L. Wong, A cytochrome P450 class I electron transfer system from *Novosphingobium aromaticivorans*, *Appl. Microbiol. Biotechnol.*, 86 (2010) 163-175.
- [8] S.A. Child, E.F. Naumann, J.B. Bruning, S.G. Bell, Structural and functional characterisation of the cytochrome P450 enzyme CYP268A2 from *Mycobacterium marinum*, *Biochem. J.*, (2018).

Chapter 4 Supplementary information

Table S1: Alternative electron transfer partners paired with CYP150A5 for whole-cell product formation assays. The ability of all to support turnover was assessed with the substrate β -ionone.

pRSF vector	pET vector
CYP150A5	Adx/AdR (from bovine mitochondria)[1, 2]
CYP150A5	HaPux/HaPuR (<i>Rhodopseudomonas palustris</i>)[3]
CYP150A5	Arx/ArR (<i>Novosphingobium aromaticivorans</i>)[4]
CYP150A5	Tdx/ArR (<i>Pseudomonas sp</i> and <i>N. aromaticivorans</i>)[5, 6]
CYP150A5	pp1957 (<i>Pseudomonas putida</i> KT2440)[7]

Table S2: Comparison of the sequence identities of CYP150A5 and CYP150A6 and analogous proteins from other *Mycobacterium* species, other named members of the CYP150 family and various other proteins.

Species	CYP name ^a	% identity to A5	Query cover	% identity to A6	Query cover	NCBI accession number
<i>M. marinum</i> M	150A5	-	-	55	100	WP_012396275.1
<i>M. liflandii</i> 128FXT	-	99	99	55	100	WP_015357170.1
<i>M. ulcerans</i> subsp. <i>shinshuense</i>	-	99	99	54	100	WP_096369782.1
<i>M. kansasii</i> ATCC 12478	-	89	97	56	92	AGZ50574.1
<i>M. avium</i> subsp. <i>paratuberculosis</i>	-	85	99	56	99	WP_010948947.1
<i>M. gastri</i>	-	85	99	55	100	WP_085105105.1
<i>M. colombiense</i>	-	83	99	53	99	WP_064882548.1
<i>M. vanbaalenii</i> PYR-1	150A9	74	98	55	97	WP_011781335.1
<i>M. smegmatis</i> MC2 155	150A3	74	98	55	91	AFP41148.1
<i>Frankia</i> sp. EAN1pec	150A11	71	99	58	95	WP_020462178.1
<i>Frankia</i> sp. EAN1pec	150A13	70	99	58	99	WP_020461226.1
<i>Frankia</i> sp. EAN1pec	150A12	69	98	54	98	WP_020461462.1
<i>M. vanbaalenii</i> PYR-1	150A10	63	98	54	98	WP_011781472.1
<i>M. smegmatis</i> MC2 155	150A2	58	95	72	98	AFP41107.1
<i>M. kansasii</i> ATCC 12478	-	57	99	85	99	AGZ50528.1
<i>Norcadia vinacea</i>	-	57	98	74	99	WP_040701857.1
<i>M. colombiense</i>	-	56	96	81	99	WP_007777258.1
<i>Frankia</i> sp. EAN1pec	150A14	56	99	61	98	WP_020458249.1
<i>M. marinum</i> M	150A6	55	100	-	-	WP_020726567.1
<i>M. ulcerans</i> Agy99	150A6	55	99	98	99	WP_011738669.1
<i>M. liflandii</i> 128FXT	-	55	99	98	99	WP_015357135.1
<i>M. vanbaalenii</i> PYR-1	150A7	55	99	71	99	WP_011781296.1
<i>Mycobacterium</i> sp. FM10	150A1	55	99	70	100	AAC97519.1
<i>Frankia</i> sp. Eu11c	-	55	99	50	100	WP_013426440.1
<i>Frankia</i> sp. EAN1pec	150A15	53	83	59	83	ABW10161.1
<i>M. smegmatis</i> MC2 155	150A4	49	100	64	100	AFP42303.1
<i>M. vanbaalenii</i> PYR-1	150A8	48	99	61	99	WP_011782311.1
<i>Streptomyces himastatinicus</i>	HmtT	33	80	34	94	4GGV_A
<i>M. tuberculosis</i>	144A1	32	99	31	90	5HDI_A
<i>M. marinum</i>	144A4	32	95	31	97	WP_085979817.1
<i>Bacillus subtilis</i>	P450 _{Biol}	28	95	30	95	3EJB_B

^a CYP name given in accordance with the NCBI database and Dr Nelson P450 homepage for bacterial P450s where listed [8].

Table S3: Additional substrate spin state shifts for CYP150A5 and CYP150A6

Substrates	CYP150A5 Spin state shift (%)	CYP150A6 Spin state shift (%)
(S)-(-)-Camphor	60	<5
1-Naphthol	60	<5
Cyclohexyl butyrate	55	<5
Borneol	45	10
α -Ionone	45	<5
Nopol	45	5
Ambroxide	35	5
Fenchone	35	-
1,4-Cineole	30	20
α -Santonin	30	15
<i>cis</i> -Jasmone	20	-
Tetralin	15	10
β -Damascone	15	<5
Pseudoionone	15	5
<i>p</i> -Cymene	15	<5
Undecanoic acid	15	<5
4-Phenylcyclohexanone	10	<5
Lidocaine	10	<5
Progesterone	10	<5
Quinoline	10	-
4-Biphenylmethanol	10	-
2-Methylnaphthalene	10	-
Naphthalene	<5	<5
Nootkatone	<5	<5
4-Phenylphenol	<5	-
4- <i>n</i> -Heptylbenzoic acid	-	20
Testosterone	-	<5
1-Indanone	-	<5
Diclofenac	-	<5
4-Hydroxybenzoic acid	-	<5

(-) indicates the substrate was not tested with the enzyme.

Table S4: The active site residues of additional CYP150 family members, compared to CYP150A6 and CYP150A5 from *M. marinum*. Bold indicates the residue matches that of CYP150A6, underlined indicates it matches CYP150A5. For emphasis those that are conserved in all those listed in the main text (Table 4) are given in red.^a Abbreviations in the table are as follows: *M. marinum* (Mmar), *Mycobacterium* sp. (Msp), *M. smegmatis* (Msmeg), *Frankia* sp. (Fran).

Mmar 150A6	Mmar 150A5	Msp 150A1	Msmeg 150A2	Msmeg 150A4	Fran 150A11	Fran 150A12	Fran 150A13	Fran 150A14	FraE^b 150
A64	<u>S64</u>	A64	A76	A64	<u>S64</u>	<u>S64</u>	<u>S64</u>	<u>S64</u>	I64
L65	<u>V65</u>	<u>V65</u>	<u>V77</u>	I65	<u>V65</u>	<u>V65</u>	<u>V65</u>	<u>V65</u>	I65
M94	<u>F94</u>	M94	M106	I94	M94	M94	M94	M94	Q96
V99	<u>P99</u>	V99	V111	V99	<u>P99</u>	<u>P99</u>	<u>P102</u>	V99	V101
F248	<u>N248</u>	F250	F261	F254	<u>N250</u>	<u>N250</u>	<u>N253</u>	F249	<u>N248</u>
A251	<u>S251</u>	A253	A264	A257	<u>S253</u>	<u>S253</u>	<u>S257</u>	<u>S253</u>	F251
<u>A252</u>	<u>A252</u>	<u>A254</u>	<u>A265</u>	<u>A258</u>	<u>A254</u>	<u>A254</u>	<u>A258</u>	<u>A254</u>	<u>A252</u>
<u>V299</u>	<u>V299</u>	<u>V301</u>	<u>V312</u>	T305	<u>V301</u>	<u>V301</u>	<u>V305</u>	<u>V300</u>	I300
<u>I409</u>	<u>I409</u>	<u>I411</u>	<u>I422</u>	<u>I415</u>	<u>I414</u>	<u>I411</u>	<u>I415</u>	<u>I410</u>	L410

^a E255 and T256 (the acid alcohol pair) are also conserved in all but are not listed here.

^b Indicates the *Frankia* sp. EuI1c CYP150A enzyme (WP_013426440.1) listed in Table 1.

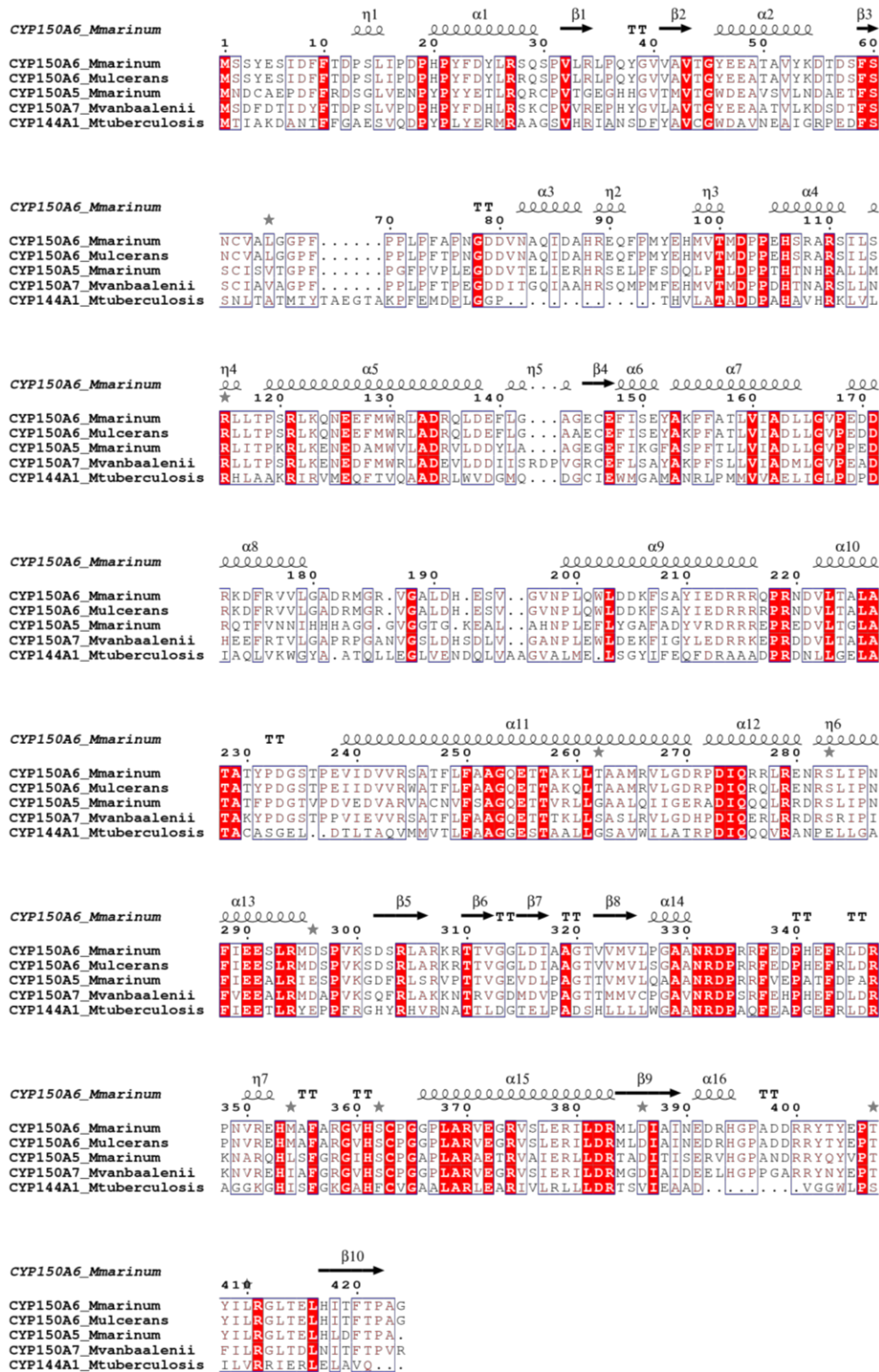


Figure S1: Sequence alignment of CYP150A6 from *M. marinum*, CYP150A6 (*M. ulcerans*), CYP150A5 (*M. marinum*), CYP150A7 (*M. vanbaalenii*) and CYP144A1 (*M. tuberculosis*) to the structural elements of CYP150A6 using ESPript [9].

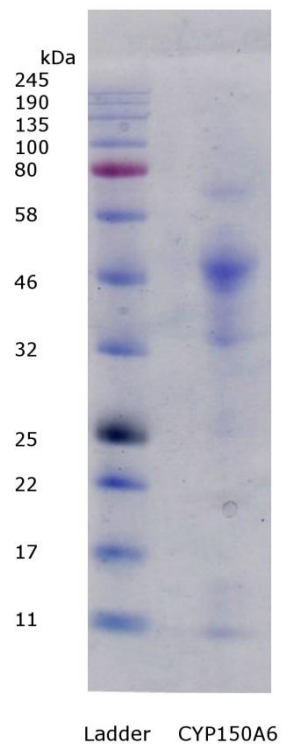


Figure S3: SDS-PAGE of CYP150A6 after ion exchange chromatography. Expected mass of CYP150A6 is 47.72 kDa.

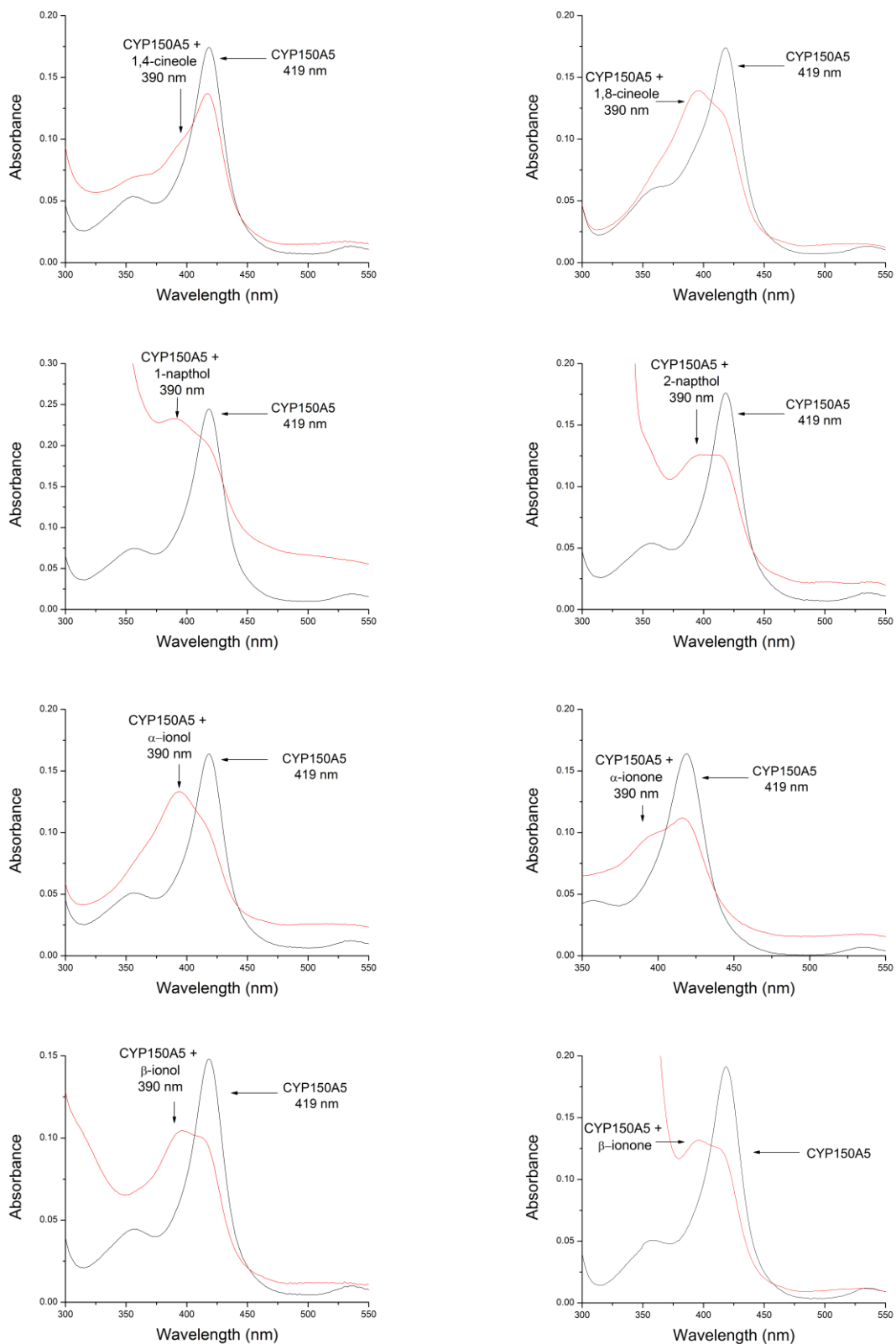


Figure S4: Selected spin state shifts of substrates and possible inhibitors with CYP150A5

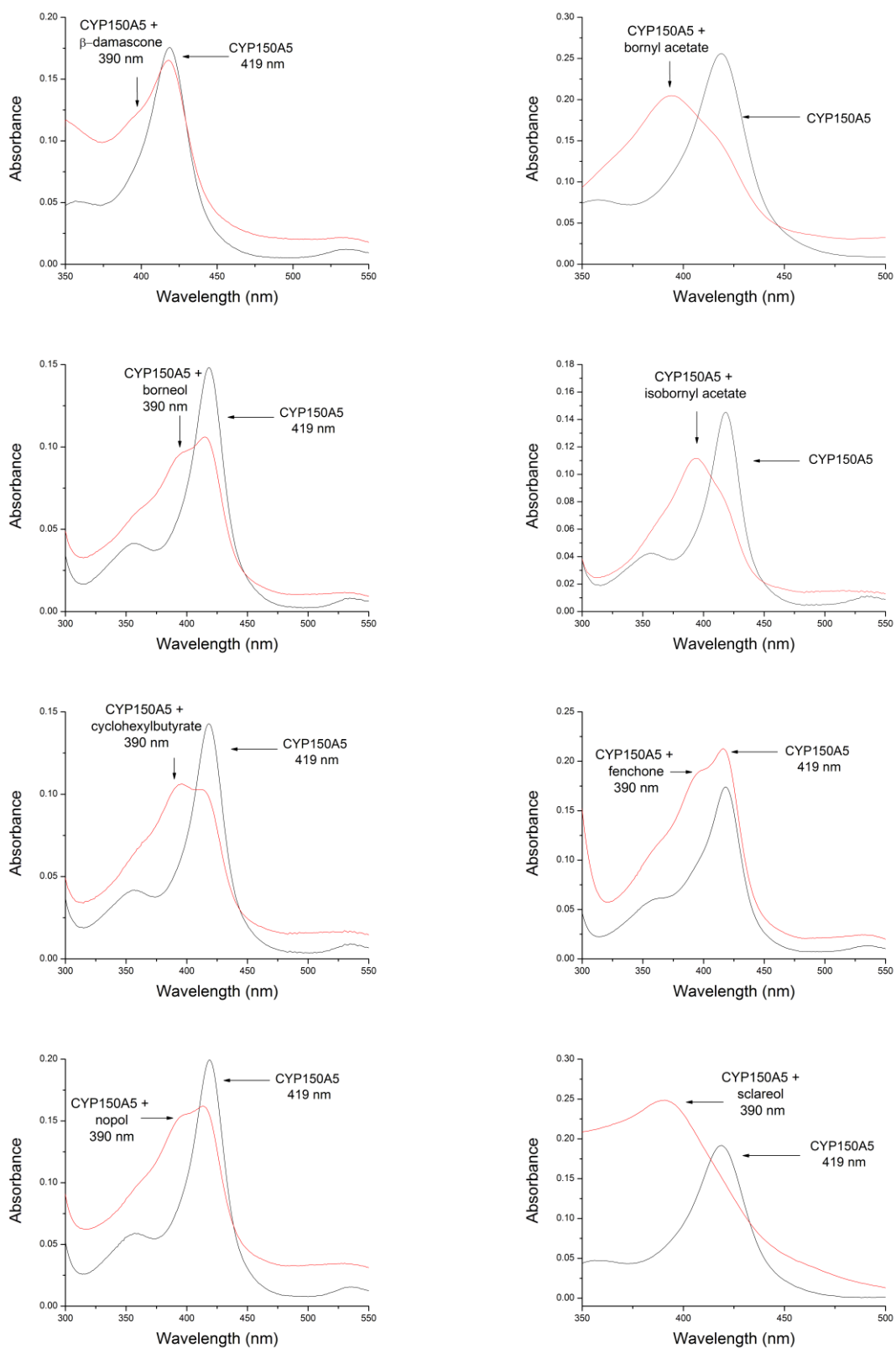


Figure S4 (continued): Selected spin state shifts of substrates and possible inhibitors with CYP150A5

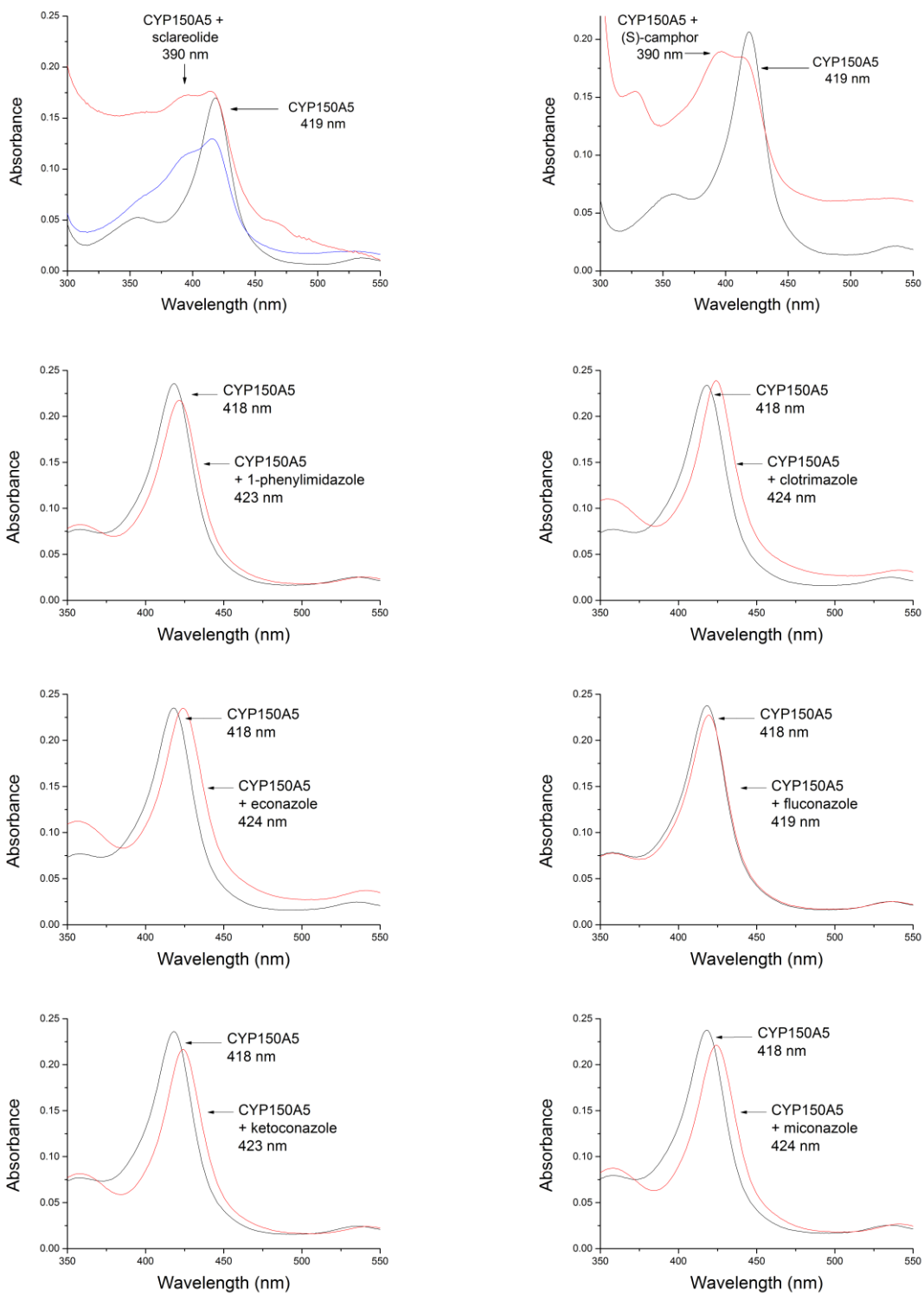
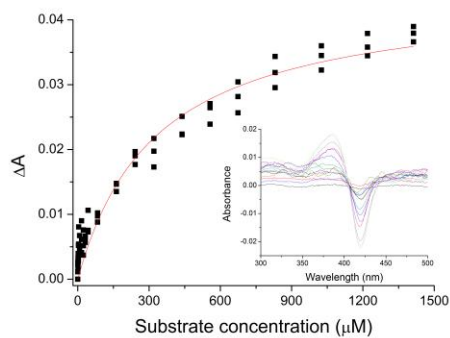
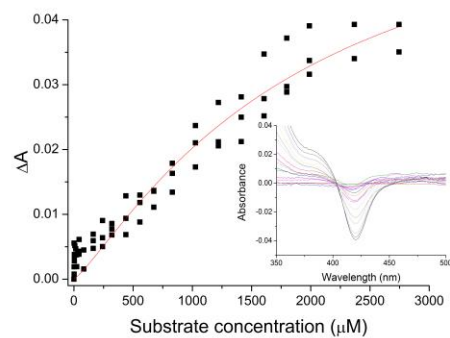


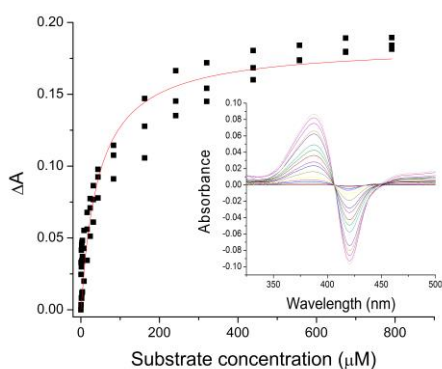
Figure S4 (continued): Selected spin state shifts of substrates and possible inhibitors with CYP150A5



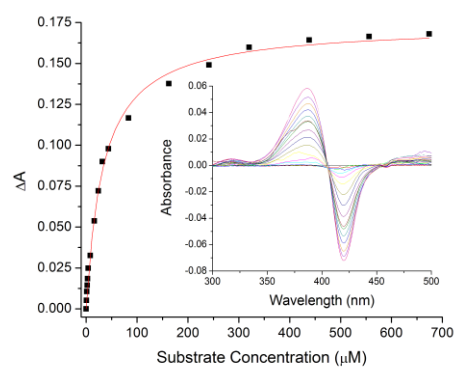
1,8-Cineole



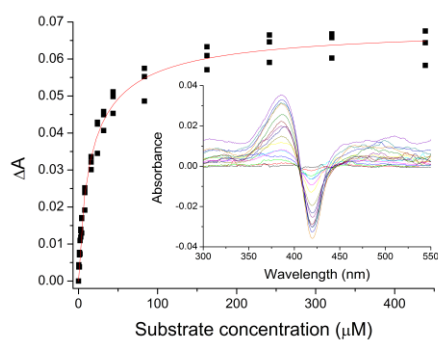
2-Naphthol



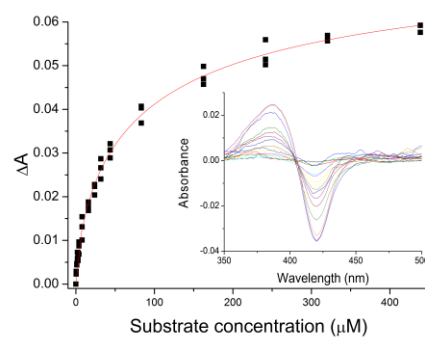
Bornyl acetate



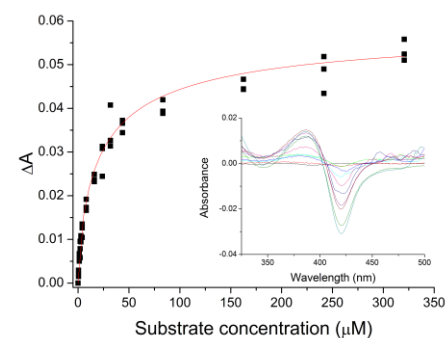
Isobornyl acetate



α-Ionol

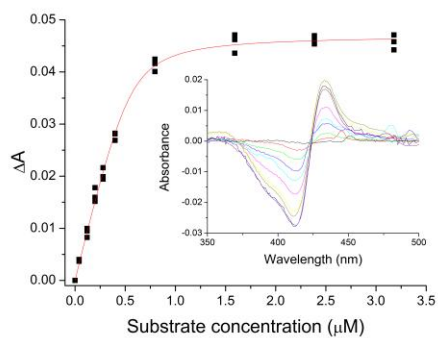


Fenchyl acetate

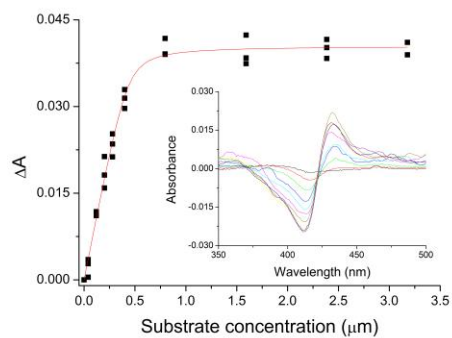


Methyl-α-ionone

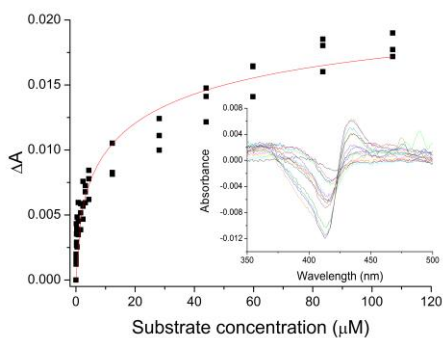
Figure S5: Dissociation constants of additional substrates with CYP150A5



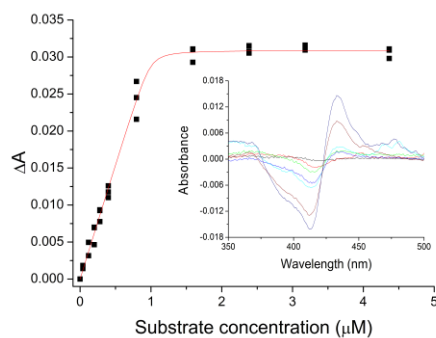
Miconazole



Clotrimazole



1-Phenylimidazole



Econazole

Figure S6: Dissociation constants of possible inhibitors with CYP150A5. Peak to trough difference in absorbance was measured at the following wavelengths: miconazole, 1-phenylimidazole and econazole, 433 to 413 nm; clotrimazole, 432 to 412 nm.

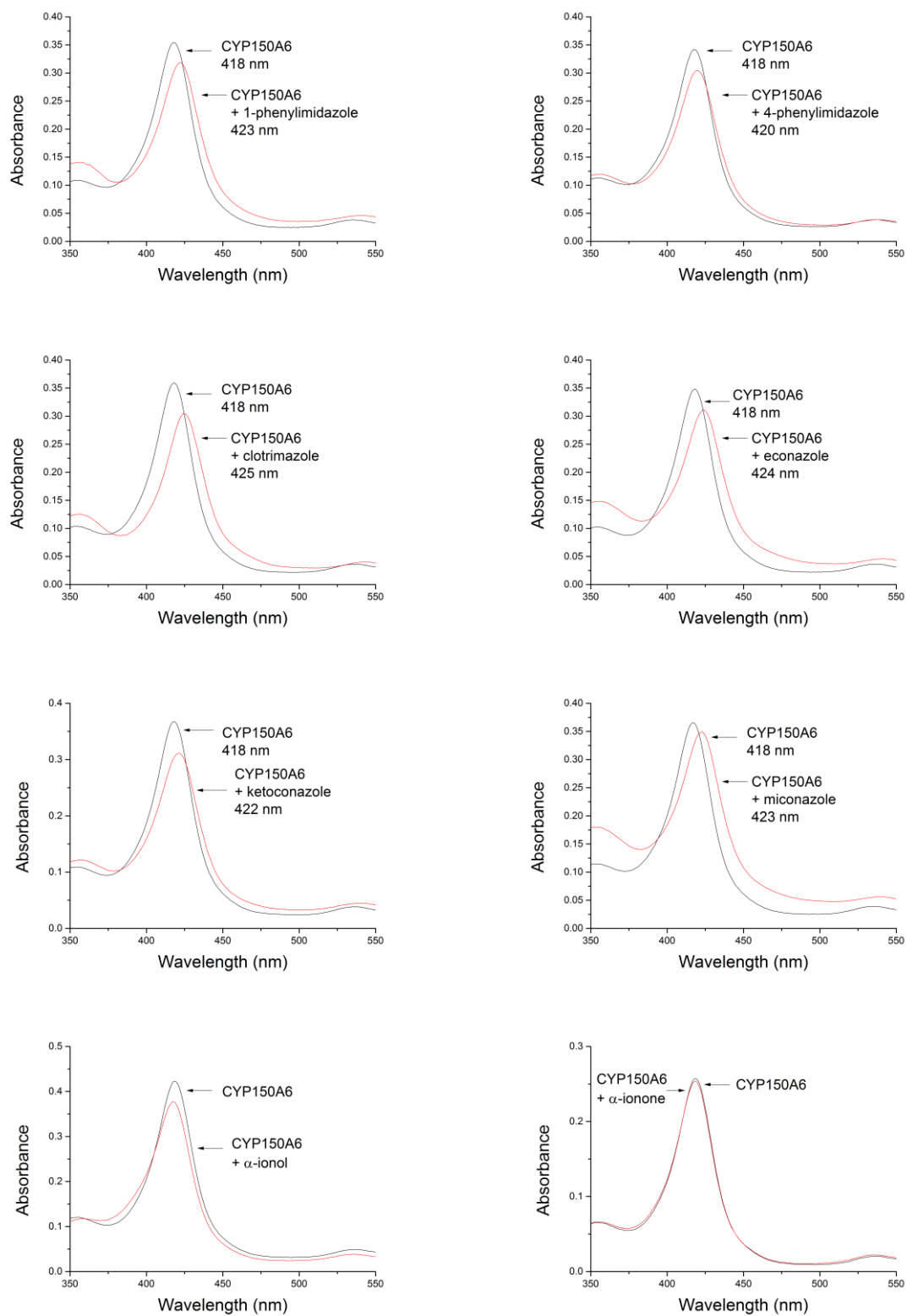


Figure S7: Selected spin state shifts of substrates with CYP150A6

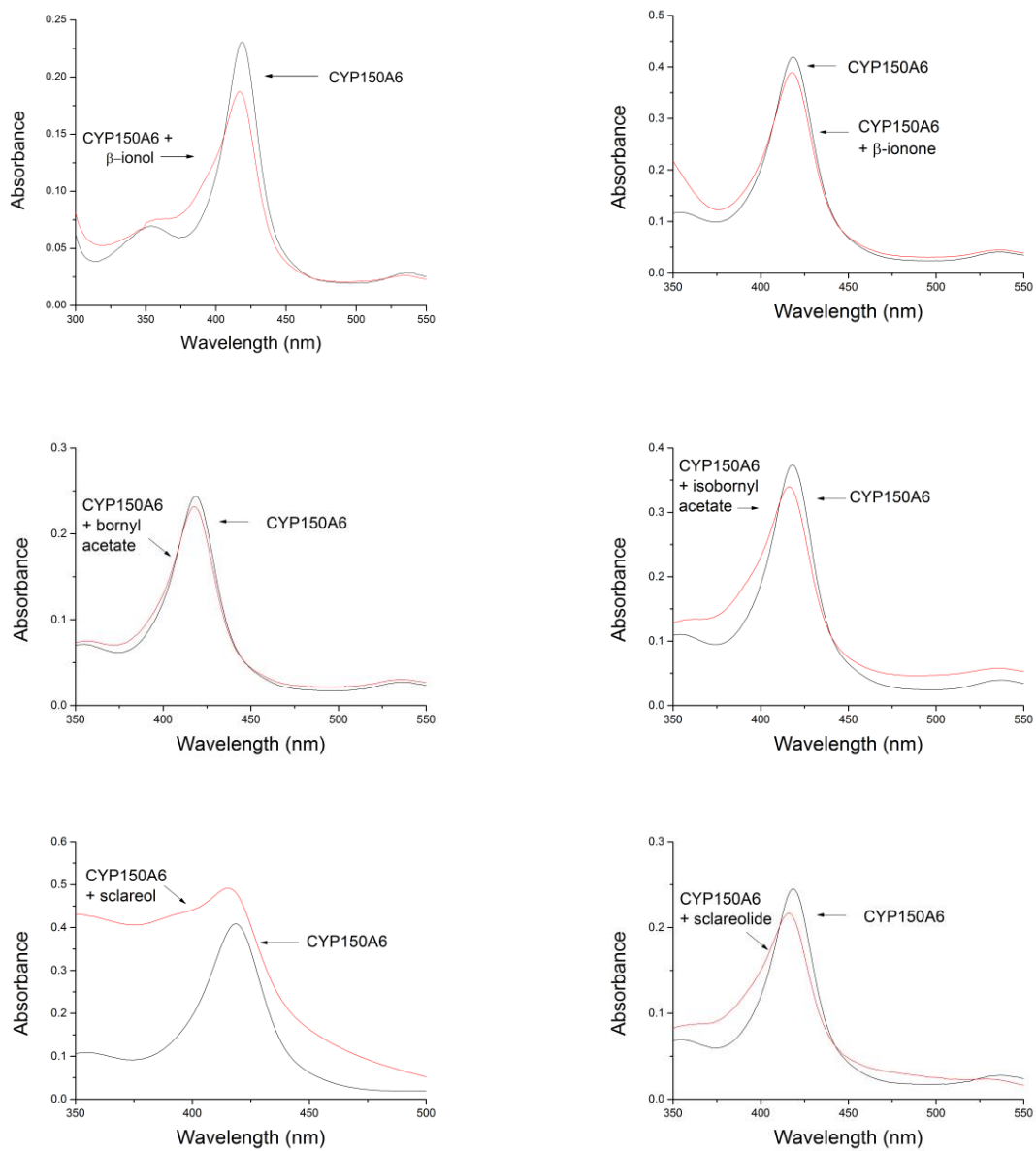
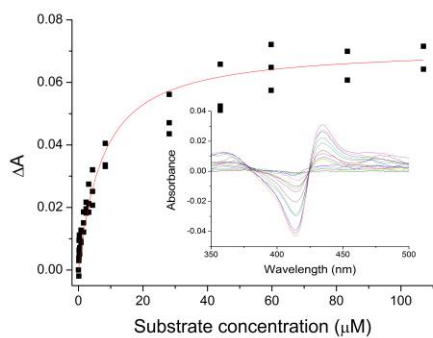
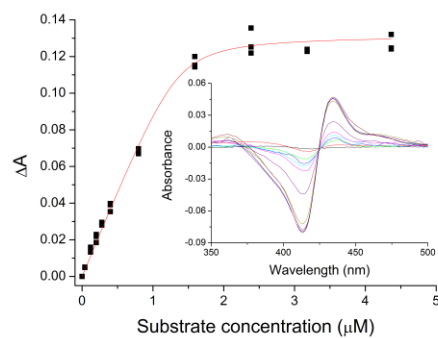


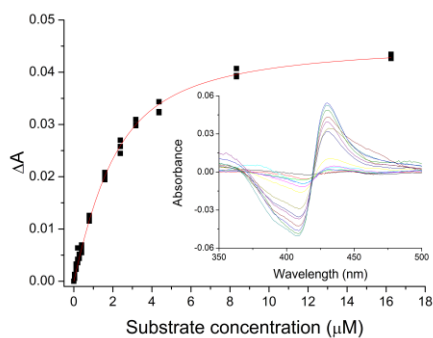
Figure S7 (continued): Selected spin state shifts of substrates and inhibitors with CYP150A6



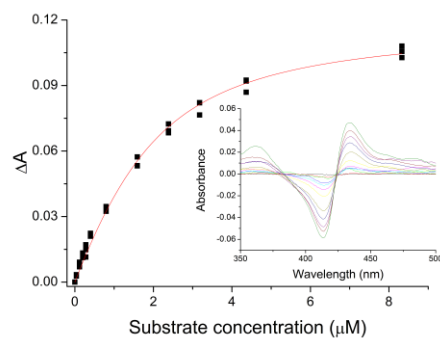
Ketoconazole



Clotrimazole



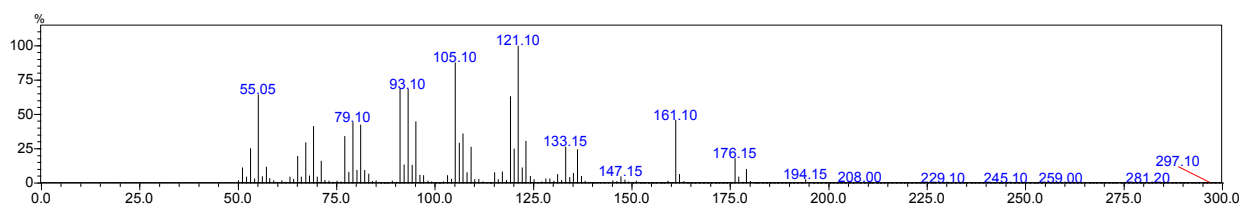
Econazole



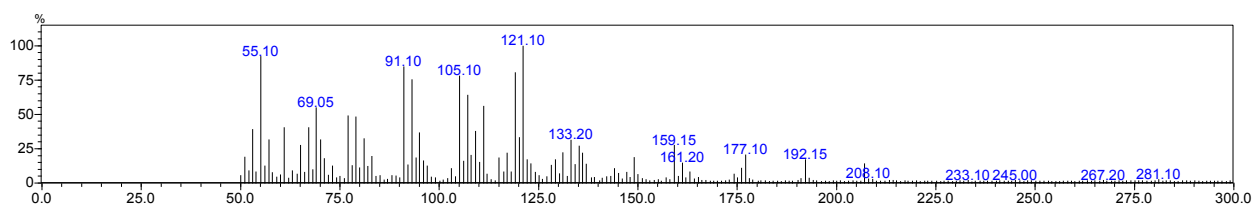
Miconazole

Figure S8: Additional binding constants of possible inhibitors with CYP150A6. Peak to trough difference in absorbance was measured at the following wavelengths: miconazole, clotrimazole and miconazole, 434 to 414 nm; econazole, 430 to 408 nm;

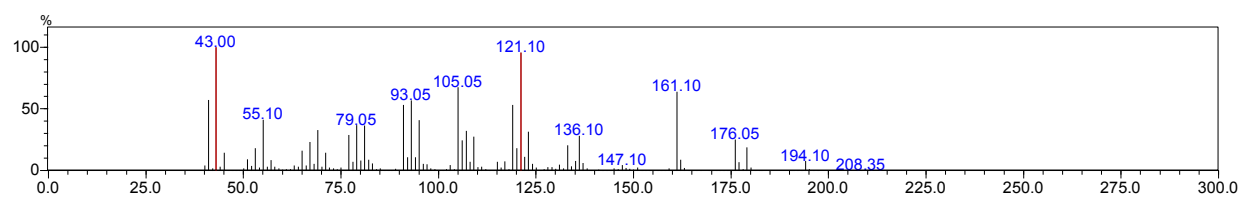
Figure S9: Mass spectra of the GC chromatogram of β -ionol with CYP150A5.



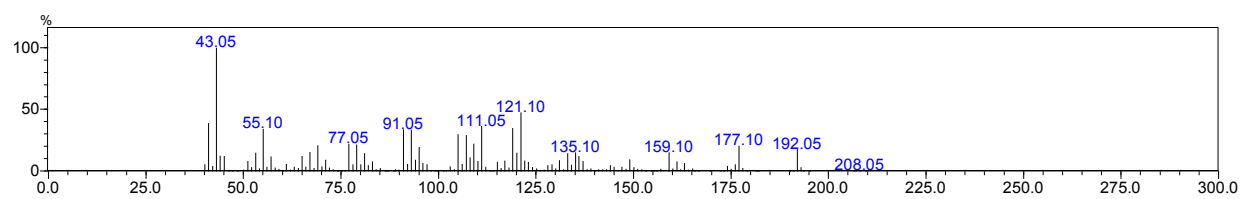
(a) MS of peak at 6.8 min – identified as substrate β -ionol (expected mass 194.3 m/z).



(b) MS of peak at 9.4 min – 4-hydroxy- β -ionol product (expected mass of single hydroxylated product 210.3 m/z).



(c) MS of substrate control for β -ionol (RT 6.8 min)



(d) MS of the CYP101B1 product 4-hydroxy- β -ionol [11]

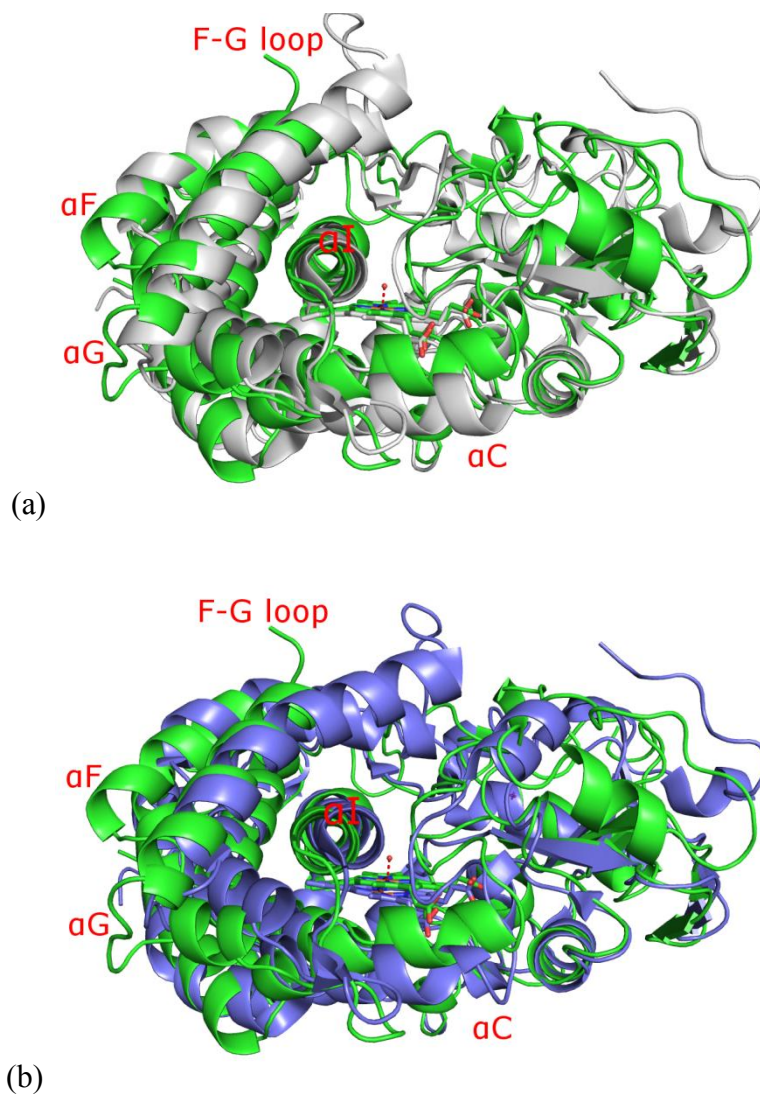


Figure S10: Superimposition of CYP150A6 (green) with the (a) open (white, 3L62) and (b) closed (blue, 3L63) form of P450_{cam} (CYP101A1) [12] demonstrating that the CYP150A6 enzyme is in a more open conformation. In both the view is centred on the I helix, showing the alteration to the F-G helix loop is the primary structural change between the open and closed structures. The C helix of CYP150A6 is closer to the haem than in either of the P450_{cam} structures. The change in B-C region is discussed further in the Fig S11.

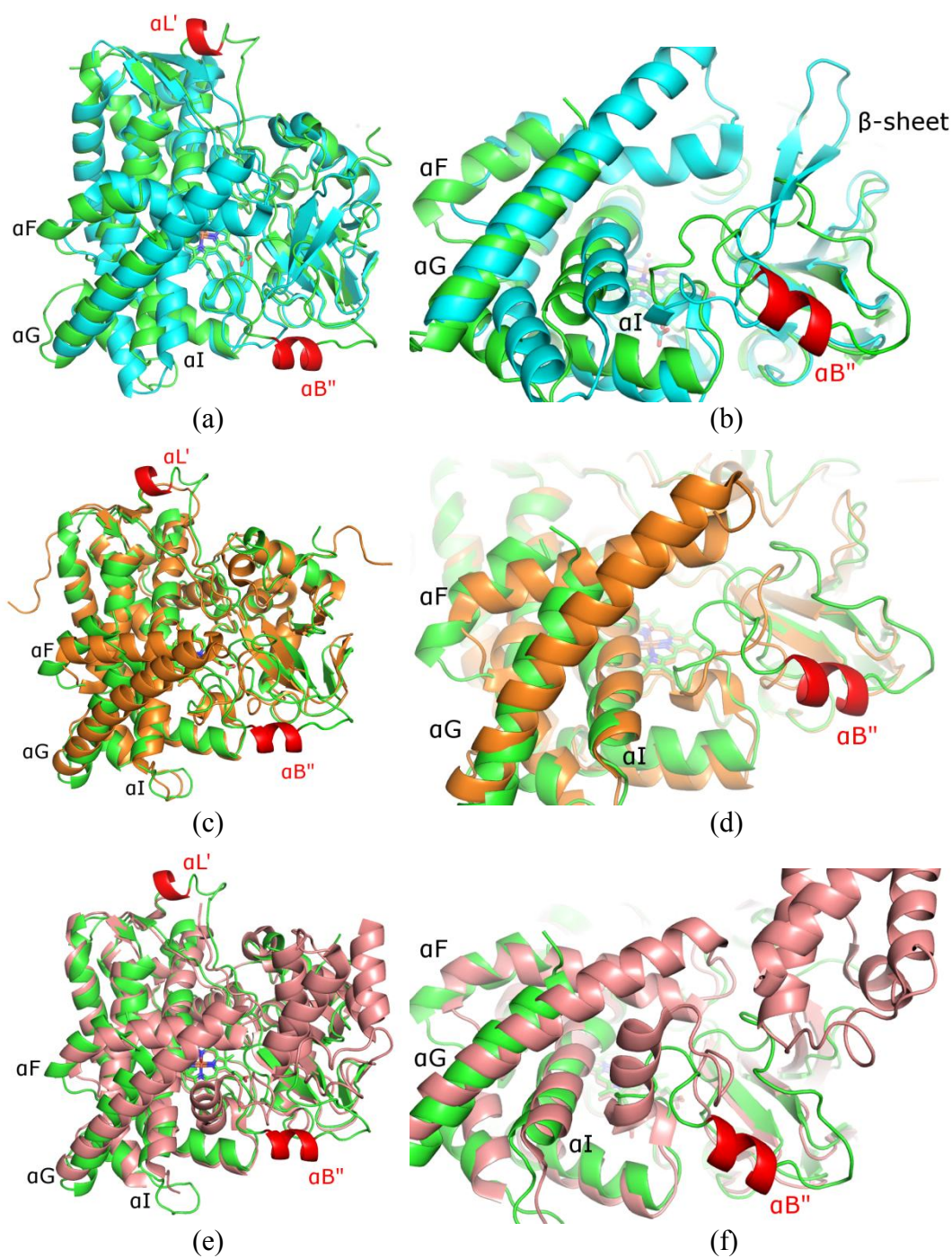


Figure S11: (a) CYP150A6 (green) overlay with *M. tuberculosis* CYP144A1 (blue, 5HDI) [13], (c) with *M. smegmatis* CYP142A2 (orange, 3ZBY)[14] and (e) with P450_{BioI} (pink, 3EJD)[15] all showing the two apparently non-conserved helices in CYP150A6 in red (L' and B''). Panels (c), (d) and (f) show the F-G region and B-C loop of the same enzymes with the B'' helix marked in red. The β -sheet region in the loop of CYP144A1 is not present in CYP150A6, which shows more similarity with CYP142A2.

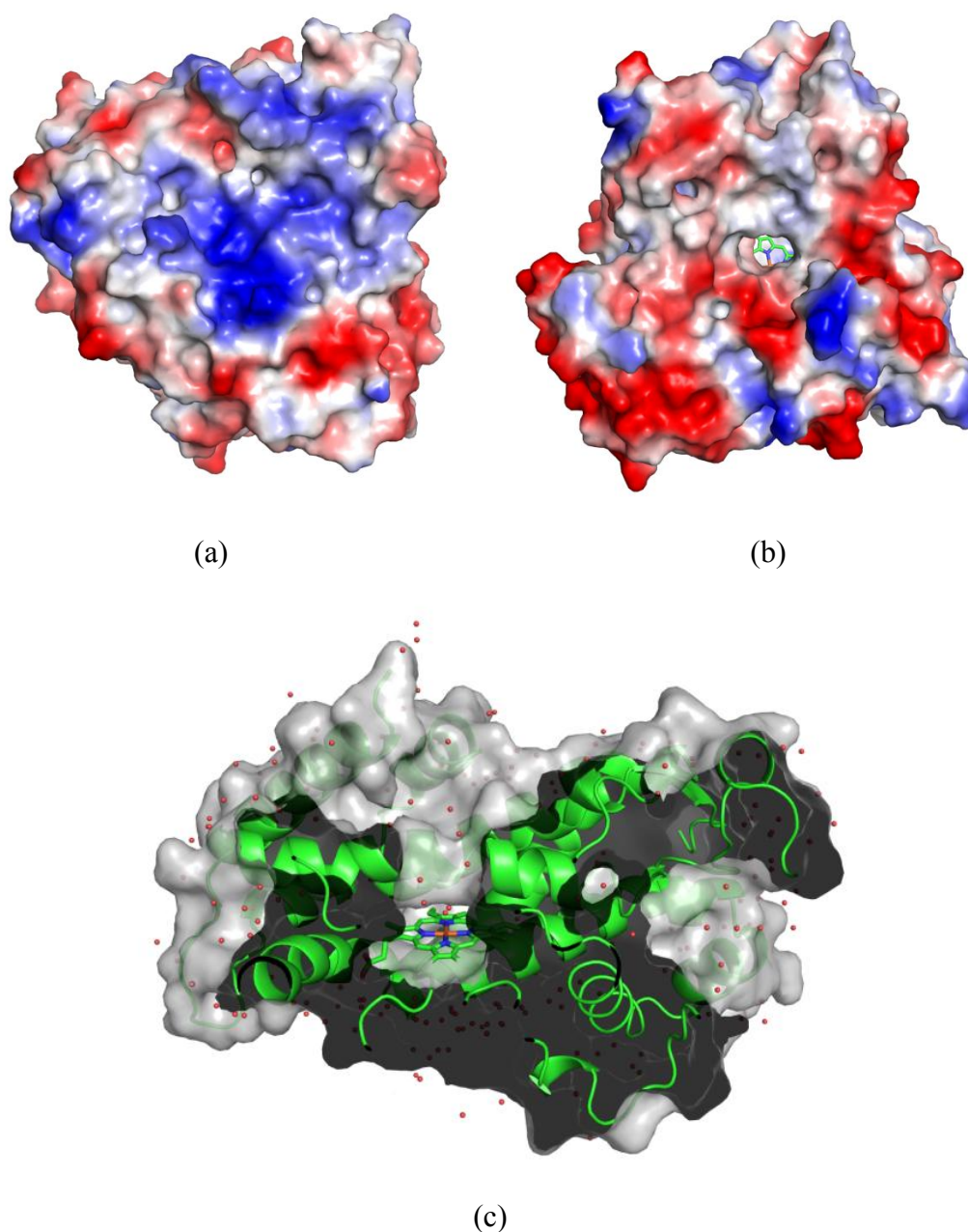


Figure S12: (a) Proximal and (b) distal surface model of CYP150A6 with applied protein contact potential (negatively charged shown in red, positively charged in blue). The substrate binding site is mostly non-polar (not shown). Note the missing residues of the F-G loop make the surface model of the distal side of the enzyme less reliable, particularly the degree to which the substrate binding site is exposed to the surface. The proximal surface (ordinarily the binding site for electron transfer partners) is strongly positively charged. (c) A surface model of CYP150A6 showing the active site cavity from (b) as a slice-through. The missing residues of the F-G loop make the cavity very large at the opening.

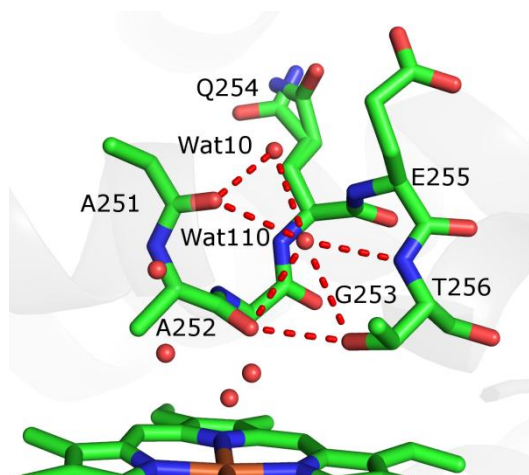


Figure S13: The water molecule (W110) bound in the kink in the I-helix and the network of stabilising hydrogen bonds (red).

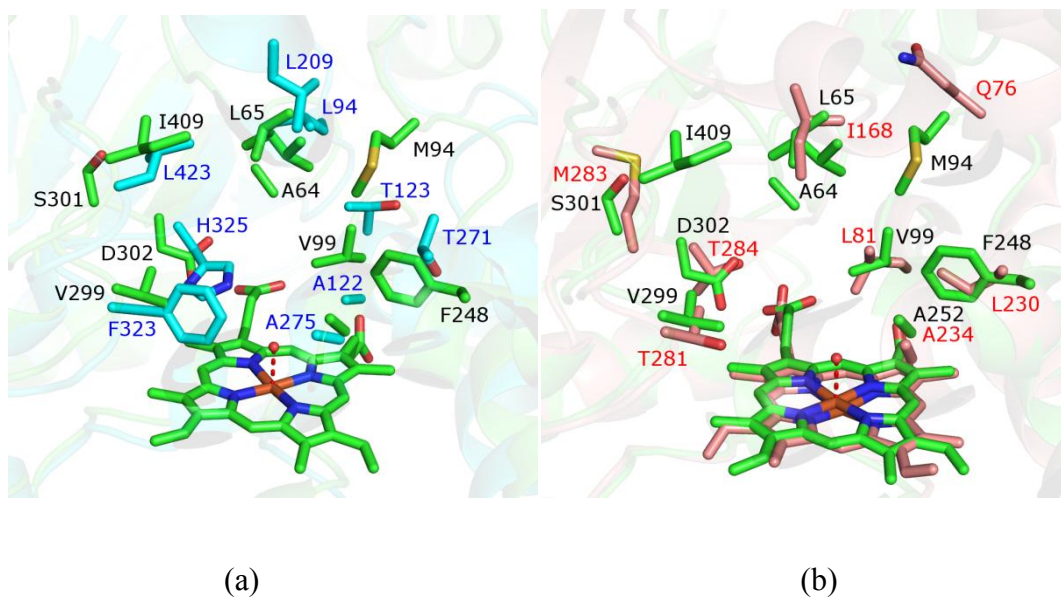


Figure S14: The active site overlay of CYP150A6 (green) and (a) CYP144A1 (blue, 5HDI) [13] and (b) P450_{BioI} (pink, 3EJD) [15] showing the active site residues in both. The region of CYP150A6 that contains M94 is not closely matched in either structure. The active site residues are not similar to those of either: for example the CYP144A1 F321 and H324 side chains identified as key active site residues are replaced in CYP150A6 with V299 and D30. In P450_{BioI} residues M283 and T281 are S301 and V299, respectively.

References

- [1] M. Tanaka, M. Haniu, K.T. Yasunobu, T. Kimura, The amino acid sequence of bovine adrenodoxin, *J. Biol. Chem.*, 248 (1973) 1141-1157.
- [2] K. Suhara, K. Nakayama, O. Takikawa, M. Katagiri, Two forms of adrenodoxin reductase from mitochondria of bovine adrenal cortex, *FEBS J.*, 125 (1982) 659-664.
- [3] S.G. Bell, A.B. Tan, E.O. Johnson, L.L. Wong, Selective oxidative demethylation of veratric acid to vanillic acid by CYP199A4 from *Rhodopseudomonas palustris* HaA2, *Molecular BioSystems*, 6 (2010) 206-214.
- [4] S.G. Bell, A. Dale, N.H. Rees, L.-L. Wong, A cytochrome P450 class I electron transfer system from *Novosphingobium aromaticivorans*, *Appl. Microbiol. Biotechnol.*, 86 (2010) 163-175.
- [5] S.G. Bell, W. Yang, J.A. Yorke, W. Zhou, H. Wang, J. Harmer, R. Copley, A. Zhang, R. Zhou, M. Bartlam, Z. Rao, L.L. Wong, Structure and function of CYP108D1 from *Novosphingobium aromaticivorans* DSM12444: an aromatic hydrocarbon-binding P450 enzyme, *Acta Crystallogr. D*, 68 (2012) 277-291.
- [6] S.A. Child, E.F. Naumann, J.B. Bruning, S.G. Bell, Structural and functional characterisation of the cytochrome P450 enzyme CYP268A2 from *Mycobacterium marinum*, *Biochemical J.*, 475 (2018) 705-722.
- [7] S.G. Bell, L. French, N.H. Rees, S.S. Cheng, G. Preston, L.L. Wong, A phthalate family oxygenase reductase supports terpene alcohol oxidation by CYP238A1 from *Pseudomonas putida* KT2440, *Biotechnol. Appl. Biochem.*, 60 (2013) 9-17.
- [8] D.R. Nelson, The cytochrome p450 homepage, *Human genomics*, 4 (2009) 59-65.
- [9] X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server, *Nucleic Acids Research*, 42 (2014) W320-W324.
- [10] L. Overmars, R. Kerkhoven, R.J. Siezen, C. Francke, MGcV: the microbial genomic context viewer for comparative genome analysis, *BMC genomics*, 14 (2013) 209.
- [11] E.A. Hall, M.R. Sarkar, J.H.Z. Lee, S.D. Munday, S.G. Bell, Improving the Monooxygenase Activity and the Regio- and Stereoselectivity of Terpenoid Hydroxylation Using Ester Directing Groups, *ACS Catalysis*, 6 (2016) 6306-6317.
- [12] Y.-T. Lee, R.F. Wilson, I. Rupniewski, D.B. Goodin, P450cam visits an open conformation in the absence of substrate, *Biochemistry*, 49 (2010) 3412-3419.
- [13] J. Cheng, M.E. Kavanagh, M.D. Driscoll, K.J. McLean, D.B. Young, T. Cortes, D. Matak-Vinkovic, C.W. Levy, S.E. Rigby, D. Leys, C. Abell, A.W. Munro, Structural characterization of CYP144A1 - a cytochrome P450 enzyme expressed from alternative transcripts in *Mycobacterium tuberculosis*, *Scientific Reports*, 6 (2016) 26628.
- [14] E. Garcia-Fernandez, D.J. Frank, B. Galan, P.M. Kells, L.M. Podust, J.L. Garcia, P.R. Ortiz de Montellano, A highly conserved mycobacterial cholesterol catabolic pathway, *Environ. Microbiol.*, 15 (2013) 2342-2359.
- [15] M.J. Cryle, I. Schlichting, 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 (2008) 15696-15701.

Chapter 5 Supplementary Information

Gene fragments and oligonucleotides used in this work

The CYP268A2 gene was cloned using the following primers:

CYP268A2 NdeI 5'
5' – TTAATTCATATGACGGCAATTTTCGGCGAGCG-3'

CYP268A2 KpnI 3'
5' – TTAATTAAGCTTGGTACCCTATTAGAAGGTGCAGGGCATGCTGCG-3'

HindIII KpnI

A NdeI site was incorporated at the 5' and HindIII and KpnI at the 3' end. The restriction sites are underlined, start and stop codons highlighted in bold. A double stop codon has been inserted at the C terminus.

Terpredoxin

The terpredoxin gene was obtained as a gblock:

TTATTATCCATGGCCCCACGCGTTGTGTTTCATCGACGAACAGTCCGGTGAATATG
CGGTCGATGCCAGGACGGTCAGAGTCTGATGGAAGTCGCAACCCAAAACGGTG
TGCCGGGCATTGTGGCCGAATGCGGTGGATCGTGCGTCTGTGCTACTTGTGCTAT
CGAAATCGAGGACGCTTGGGTTGAAATCGTAGGCGAAGCGAATCCGGATGAAAA
CGACTTACTGCAGAGTACGGGTGAGCCGATGACCGCCGGCACCCGTCTCAGTTGT
CAAGTGTTTCATTGATCCTTCGATGGATGGACTGATTGTACGGGTGCCTCTGCCGG
CGTAATAGAAGCTTTAATAA

NcoI sites and HindIII sites were incorporated at the 5' and 3' ends respectively. The restriction sites NcoI and HindIII are underlined, start and stop codons highlighted in bold. A double stop codon has been inserted at the C terminus. Note that an additional GGC codon encoding an alanine residue has been inserted to incorporate the NcoI site.

Modified Terpredoxin amino acid sequence:

MAPRVVFIDEQSGEYAVDAQDGQSLMEVATQNGVPGIVAECGGSCVCATCRIEIED
AWVEIVGEANPDENDLLQSTGEPMTAGTRLSCQVFIDPSMDGLIVRVPLPA

Table S1: Details including sequence identities of analogous and similar proteins to CYP268A2.

Species	CYP name ^a	% identity	Query cover	NCBI accession number
<i>Mycobacterium liflandii</i>	-	98	100	WP_015356535.1
<i>M. ulcerans</i> subsp. <i>shinshuense</i>	-	97	99	BAV42537.1
<i>M. ulcerans</i> Agy99 ^b	268A2P	98	100	NC_008611.1
<i>M. kansasii</i>	-	85	99	WP_063470951.1
<i>M. avium</i> subsp. <i>paratuberculosis</i>	268A1	79	98	AAS04578.1
<i>M. yongenense</i>	-	79	98	WP_065500185.1
<i>M. avium</i>	-	79	75	WP_038863172.1
<i>M. smegmatis</i>	268A3	78	98	WP_003895997.1
<i>M. tuberculosis</i> TKK-01-055	-	78	98	KBZ61308.1
<i>M. colombiense</i>	-	77	99	WP_077092567.1
<i>Streptomyces. bingchengensis</i>	268A4	54	99	WP_014180281.1
<i>M. smegmatis</i>	268B2	51	99	WP_011728434.1
<i>M. vanbaalenii</i> PYR-1	268C1	51	96	WP_011781149.1
<i>M. vanbaalenii</i> PYR-1	268B1	49	99	WP_011779461.1
<i>M. smegmatis</i>	124A1	44	97	AFP39933.1
<i>M. vanbaalenii</i> PYR-1	142B1	43	99	WP_011777337.1
<i>M. vanbaalenii</i> PYR-1	124A1	43	97	ABM13832.1
<i>S. griseus</i>	124G1	42	96	WP_042498075.1
<i>M. marinum</i> M	124A1	41	97	WP_012395007.1
<i>M. tuberculosis</i> H37Rv	124A1	41	97	WP_003411654.1
<i>M. ulcerans</i> Agy99	124A1	41	97	ABL03846.1
<i>M. sp</i> JLS	142B2	41	97	WP_011855908.1
<i>M. tuberculosis</i> H37Rv	125A1	40	89	WP_003419304.1
<i>M. ulcerans</i> Agy99	125A7	40	89	ABL06152.1
<i>M. marinum</i> M	125A6	40	87	WP_012394490.1
<i>M. marinum</i> M	125A6	40	86	WP_012396556.1
<i>M. smegmatis</i> MC2 155	125A3	38	90	WP_003897396.1
<i>M. tuberculosis</i> H37Rv	142A1	33	96	WP_003900082.1
<i>M. smegmatis</i>	142A2	33	96	AFP42192.1
<i>M. marinum</i> M	142A3	32	96	WP_012396527.1
<i>M. ulcerans</i> Agy99	142A3	32	96	ABL06126.1

NB: A large number of *Mycobacterium* analogues were found, of which a selected few are presented here. The first ~300 results were solely from *Mycobacterium* species, with sequence identities > 70%. There were no analogues found with identity between 72% to 65%. Many non-*Mycobacterial* species were found with a sequence identity lower than 65%. These include *Nocardia violaceofusca*, *Nocardia yamanashiensis* and other *Nocardia* strains, *Cryptosporangium aurantiacum*, *Cryptosporangium arvum* and others, *Rhodococcus erythropolis*, *Rhodococcus* sp. ACS1 and other strains, *Pseudonocardia spinosispora*, *Pseudonocardia ammonioxydans* and others, *Saccharomonospora* sp. CUA-673, *Blastococcus aggregatus*, *Corynebacteriales*, *Skermania piniformis*, *Sporichthya polymorpha*, *Frankia coriariae*, *Streptomyces* sp. NBS 14/10, *Streptomyces bingchengensis* (listed above), *Gordonia* sp. HS-NH1, *Gordonia westfalica* and other strains, *Haloethinotrix alba*, *Williamsia herbipolensis*, and *Patulibacter americanus* before sequence identity drops below 50%.

^a CYP name given in accordance with the NCBI database and Dr Nelson P450 homepage for bacterial P450s where listed [1].

^b full sequence used, despite truncation, with the large nucleotide insertion removed. As the protein is not expected to exist, the accession number given here is for the gene.

Table S2: Additional spin state shifts of CYP268A2 with a variety of substrates.

CYP268A2 substrates	Spin state shift (% HS)	CYP268A2 substrates	Spin state shift (% HS)
<i>cis</i> -Nerolidol	55	Bornyl acetate	10
3-Phenyltoluene	55	Biphenyl	10
Indane	55	Dodecyl acetate	10
1-Naphthol	50	Linoleic acid	10
Nopol	50	Menthyl acetate	10
Phenyl acetate	50	Myristic acid	10
Phenylcyclohexane	50	Octanoic acid	10
Linoleic acid	50	Quinoline	10
α -Bisabolol	40	α -Santonin	10
Palmitelaidic acid	35	4-Cholest-3-one	10
4-Phenylcyclohexanone	30	1-Dodecanol	5
<i>p</i> -Cymene	30	α -Ionone	<5
Palmitoleic acid	30	Cholesterol	<5
11-Methyl dodecanoic acid	30	Cholesterol acetate	<5
Indole	20	7-Dehydrocholesterol	<5
Methyl laurate	20	Palmitic acid	<5
Phytanic acid	20	Progesterone	<5
10-Methyl dodecanoic acid	20	Stigmasterol	<5
15-Methyl hexadecanoic acid	15	Estriol	<5
4-n-Heptyl benzoic acid	10	Estrone	<5
Arachidic acid	10	Calciferol	<5

Table S3: Binding of possible inhibitors to CYP268A2

CYP268A2 substrates	Type I or II	%HS	Shift (nm)	K_d (μM)
1-Phenylimidazole	II	-	421	0.9 ± 0.3
4-Phenylimidazole	II	-	423	4.5 ± 0.6
Econazole	I	20	-	
Miconazole	I	20	-	
2-Phenylimidazole	I	10	-	
Clotrimazole	I	10	-	
Fluconazole	I	10	-	
Ketoconazole	I	10	-	

Table S4: Potential electron transfer partners for CYP268A2, screened with various substrates.*

pRSF vector	pET vector
CYP268A2	Adx/AdR (from bovine mitochondria)[2, 3]
CYP268A2	HaPux/HaPuR (<i>Rhodopseudomonas palustris</i>)[4]
CYP268A2	Arx/ArR (<i>Novosphingobium aromaticivorans</i>)[5]
CYP268A2	Tdx/ArR (<i>Pseudomonas sp</i> and <i>N. aromaticivorans</i>)[6]
CYP268A2	pp1957 (<i>Pseudomonas putida</i> KT2440)[7]

*Tdx/ArR best supported CYP268A2 activity with geranyl acetate, followed by pp1957, with very little to no product formation when supported by Adx/AdR, Arx/ArR and HaPux/HaPuR.

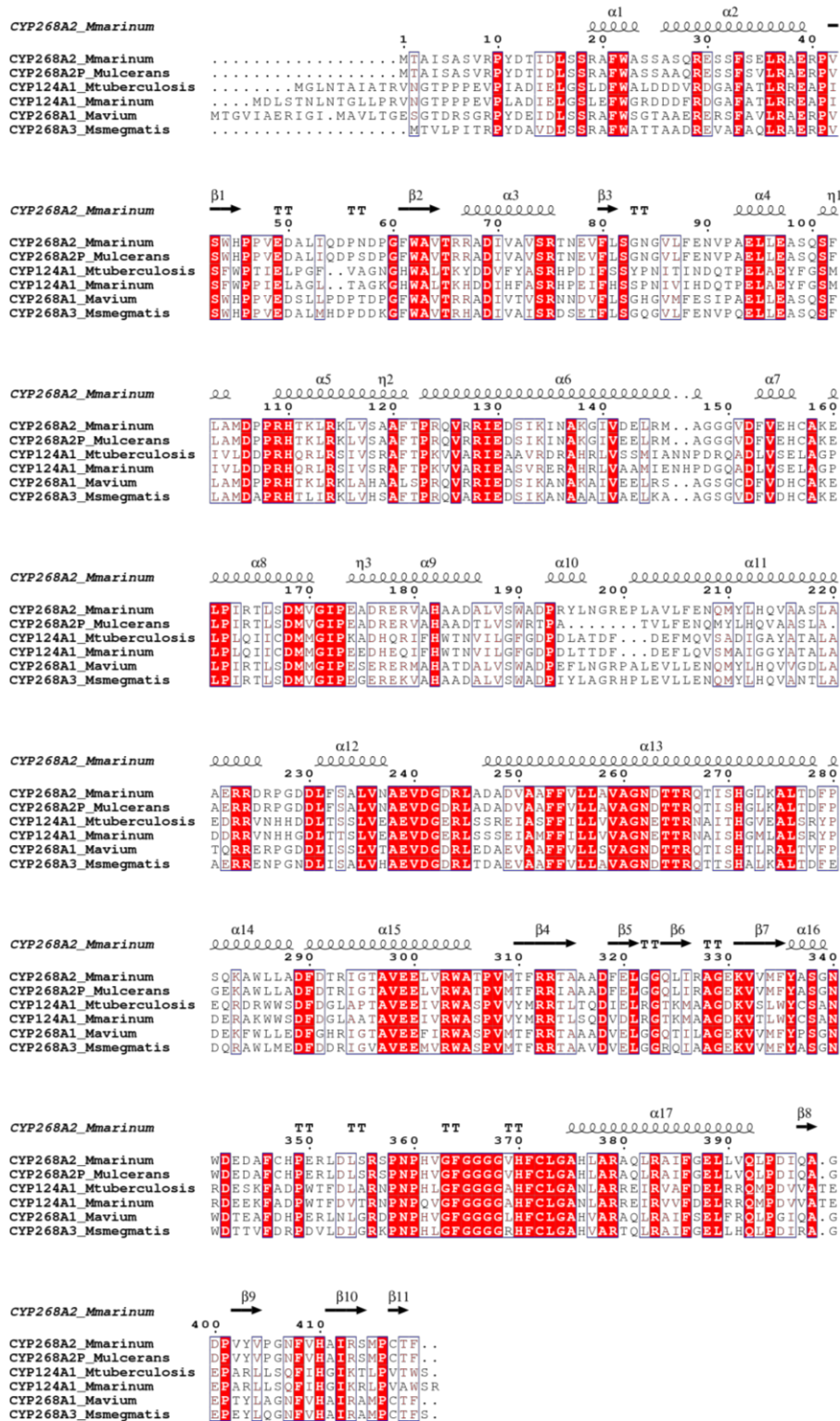


Figure S1: Alignment of CYP268A2 from *M. marinum*, CYP268A2P from *M. ulcerans*, CYP124A1 from *M. tuberculosis*, CYP124A1 from *M. marinum*, CYP268A1 from *M. avium* and CYP268A3 from *M. smegmatis* to the structural elements of CYP268A2 (PDB: 6BLD) using ESPrict [8].

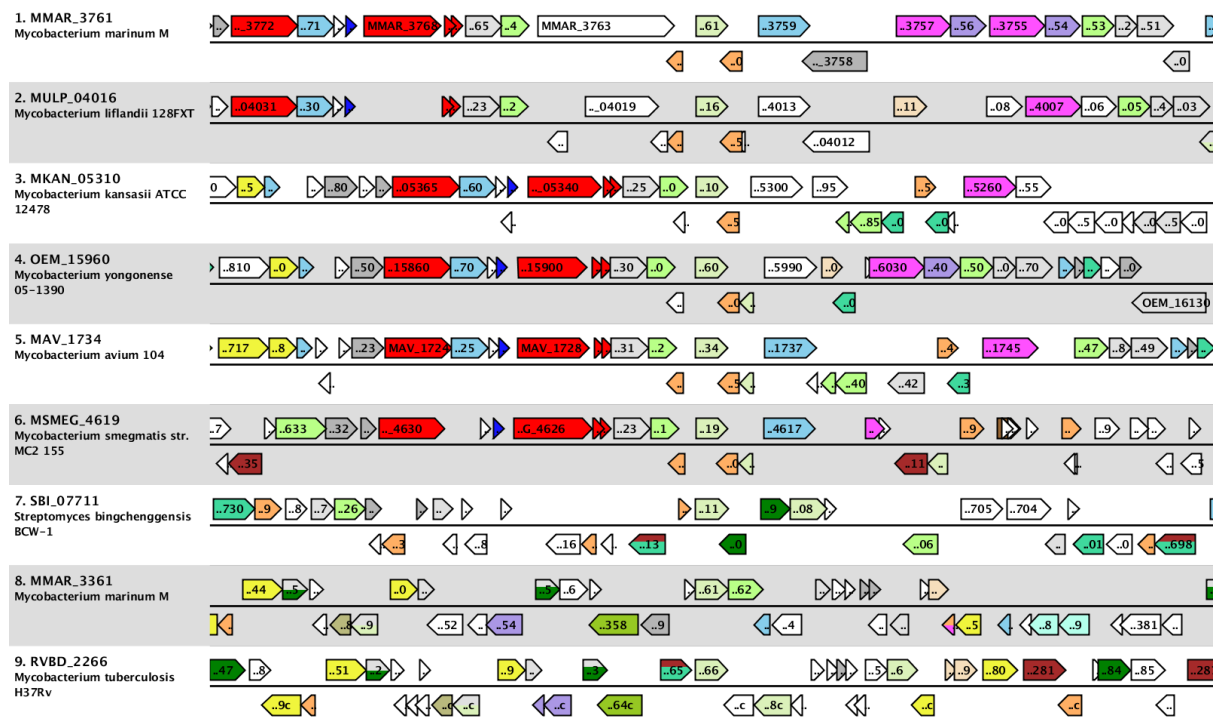


Figure S2: The Microbial Genomic Context Viewer (MGcV) [9] was used to compare the surrounding genome regions (40 kbp) of homologous CYPs. Genes are labelled by their locus tag and coloured by COG (Clusters of Orthologous Groups). The genes surrounding the *Mmar_3761* gene encoding CYP268A2 (green) from *Mycobacterium marinum* M are compared to those from *Mycobacterium liflandii* 128FXT (*Mulp_04016*), *Mycobacterium kansasii* ATCC (*Mkan_05310*), *Mycobacterium youngonense* (*OEM_15960*), *Mycobacterium avium* subsp. *paratuberculosis* (*Map2261c*), *Mycobacterium smegmatis* (*Msmeg_4619*) and *Streptomyces bingchengensis* (*SBI_07711*). For reference *Mmar_3361* from *M. marinum* and *RVBD_2266* from *Mycobacterium tuberculosis* H37Rv both encoding CYP124A1 are also included. The highly conserved operon present in *M. kansasii*, *M. youngonense*, *M. avium* and *M. smegmatis*, has a gamma-glutamyl kinase (COG0263, green), a GTPase (COG0536, grey), a large ribonuclease gene (COG1530, red) and two other ribosomal proteins (COG0261 and COG0211, red) upstream. This is present in *M. marinum* further upstream from the *cyp268* gene (within 20 kbp, *Mmar_3764* to *Mmar_3772*). In *M. marinum* the *cyp268A2* gene is flanked on both sides by a PE-PGRS gene (*Mmar_3758* and *Mmar_3763*, COG4907), and similar to the other strains, has an NAD synthetase (COG0171, blue) and two regulatory proteins (AcrR and Sir2-like, COG1309 and COG0263, orange) nearby. The *cyp268* gene in *S. bingchengensis* has none of those genes from the *Mycobacteria* nearby, the only similarity being the presence of a regulatory protein directly upstream.

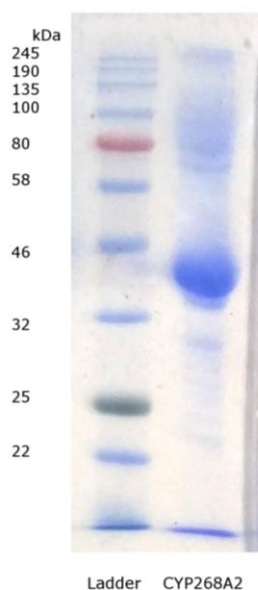


Figure S3: SDS page of CYP268A2 after ion exchange chromatography

Mass spectrometry was used to confirm the mass of CYP268A2. Protein mass measurements were carried out under denaturing conditions using an Agilent 6560 ion mobility quadrupole time-of-flight instrument with Dual AJS electrospray ionisation source, coupled to an Agilent 1290 Infinity II LC System. The protein was buffer exchanged into 250 mM ammonium acetate, concentrated to ~10 mM, then diluted 1:1 with acetonitrile. 3 μL of sample was injected and electrosprayed using 50% aqueous acetonitrile/0.01% formic acid at a flow rate of $0.1 \text{ mL}\cdot\text{min}^{-1}$, without chromatographic separation. ESI-MS conditions were: positive-ion mode; capillary voltage, 3500 V; nozzle voltage 1000 V; fragmentor, 400 V; gas 8 L/min; gas temperature, 300 $^{\circ}\text{C}$; sheath gas 11 L/min; and sheath gas temperature, 350 $^{\circ}\text{C}$. Spectra were deconvoluted using BioConfirm software (Agilent).

Mass determined 45871.21 (Expected mass 45975.25)

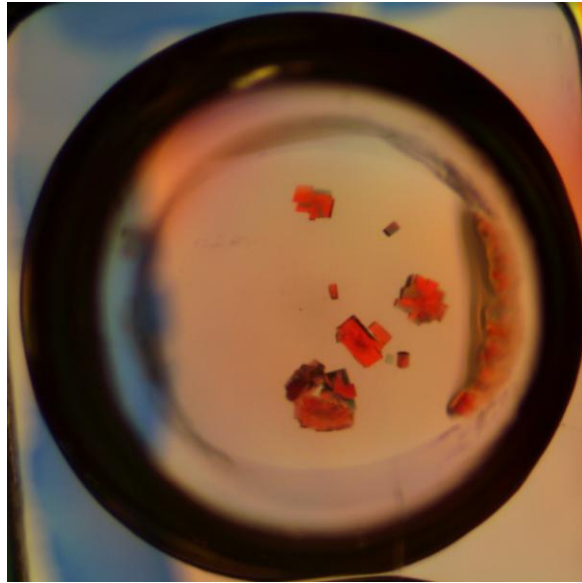


Figure S4: Crystals of CYP268A2 before optimization of crystal conditions

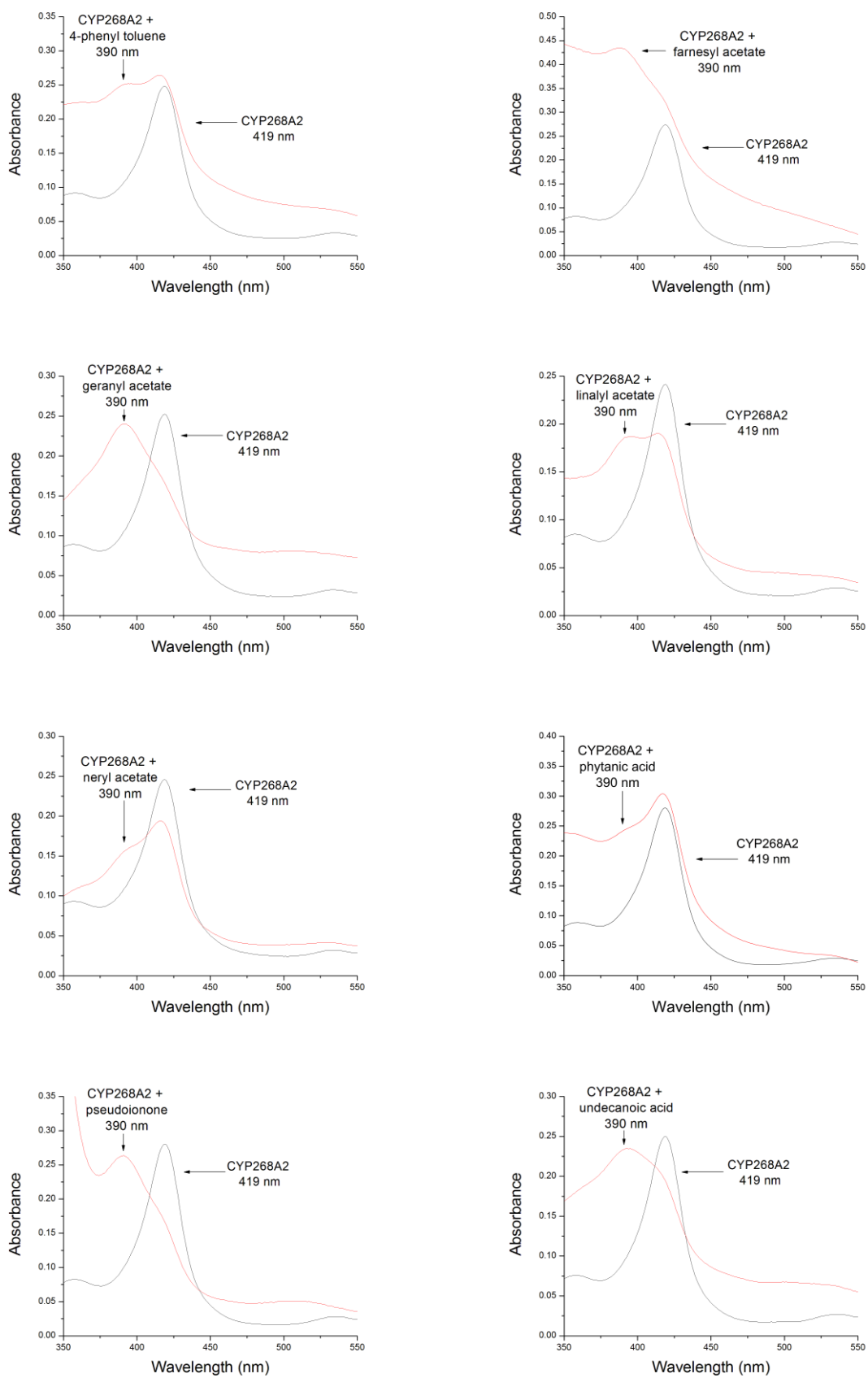


Figure S5: Spin state shifts of selected substrates with CYP268A2

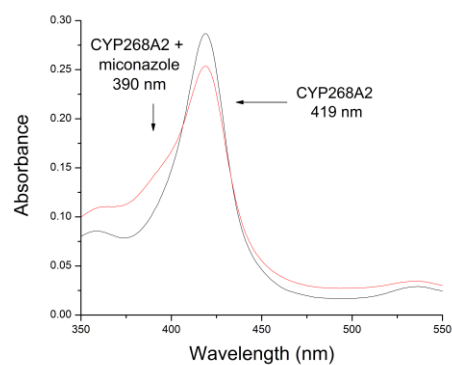
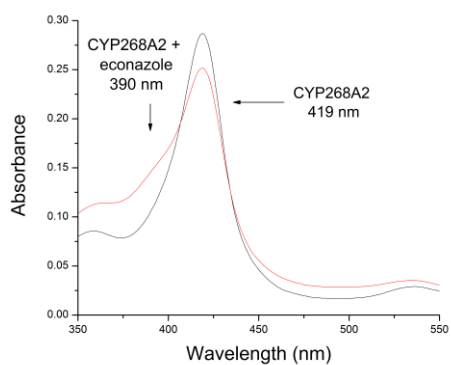
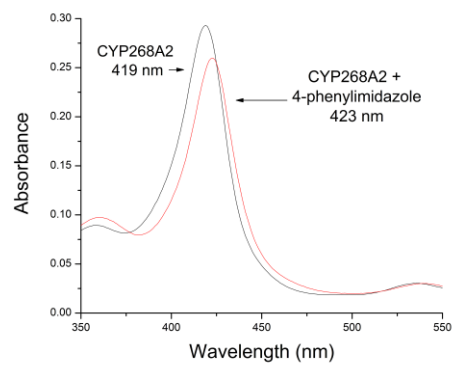
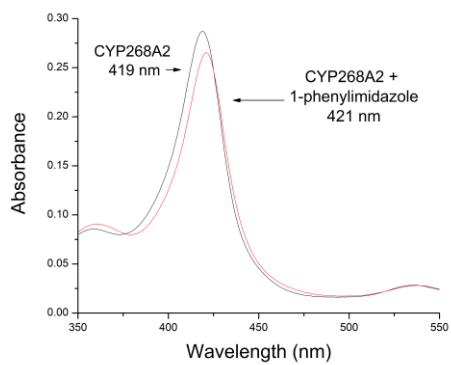
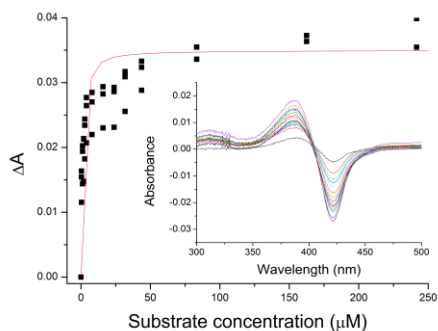
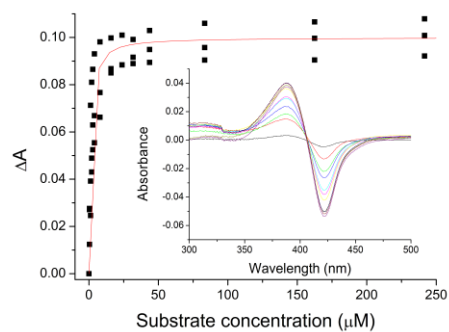


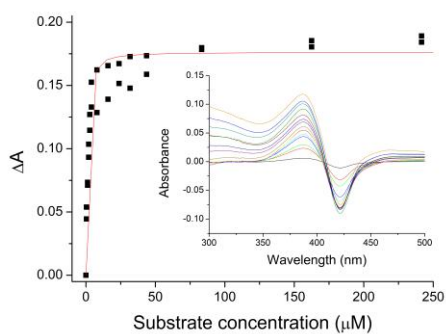
Figure S6: Spin state shifts of selected potential inhibitors with CYP268A2



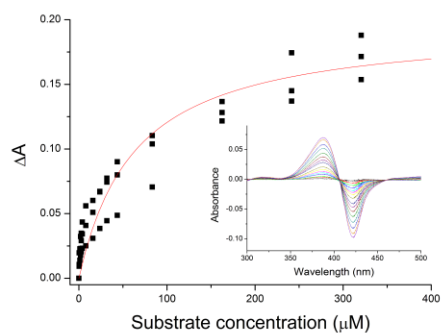
A) Capric acid



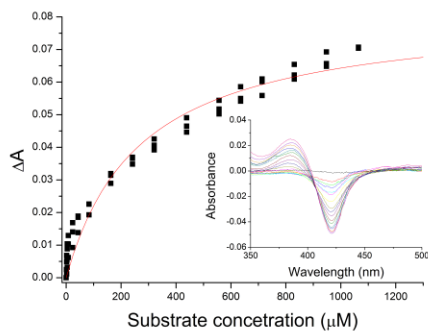
B) Farnesol



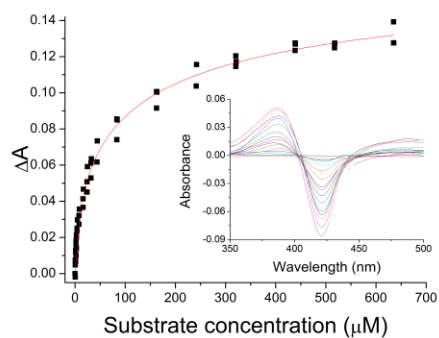
C) Farnesyl acetate



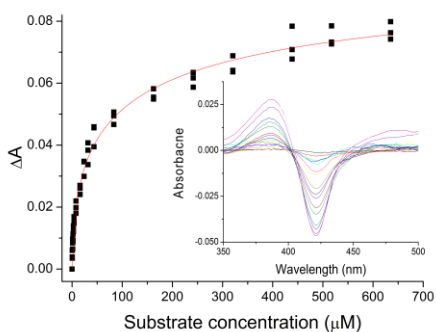
D) Geraniol



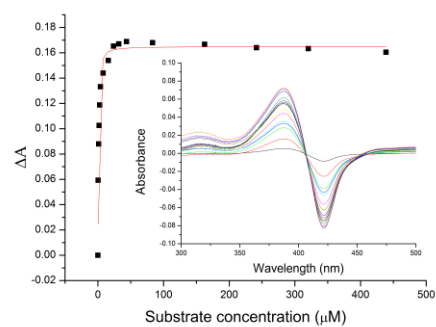
E) Lauric acid



F) Linalyl acetate



G) Neryl acetate



H) 10-Undecenoic acid

Figure S7: Additional binding constant analysis of CYP268A2 with various substrates

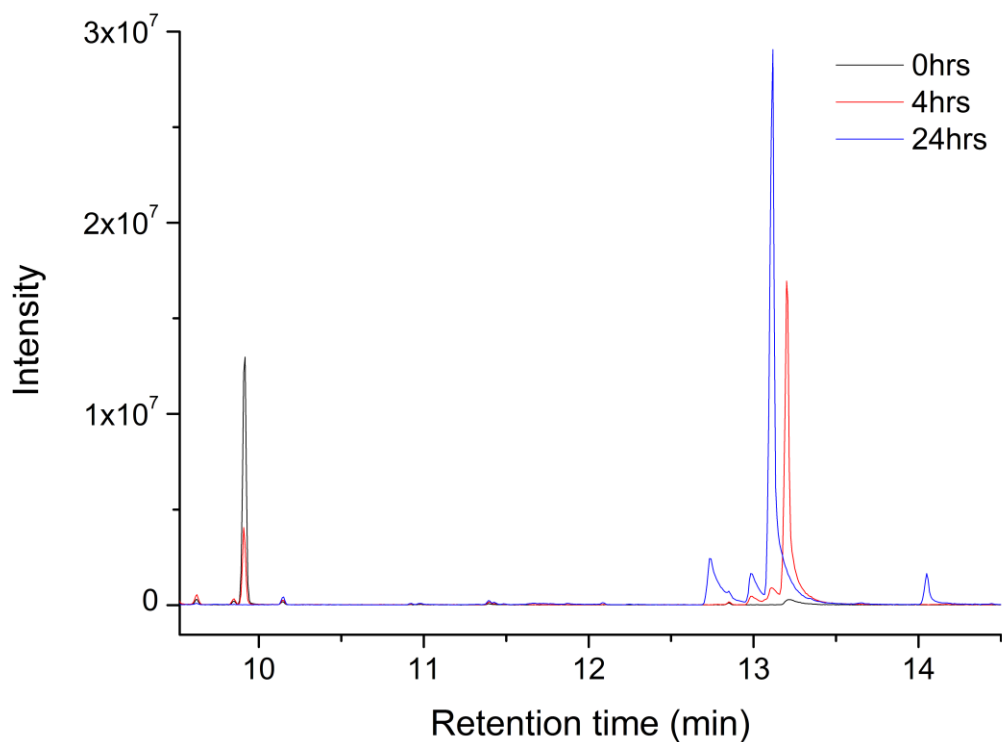
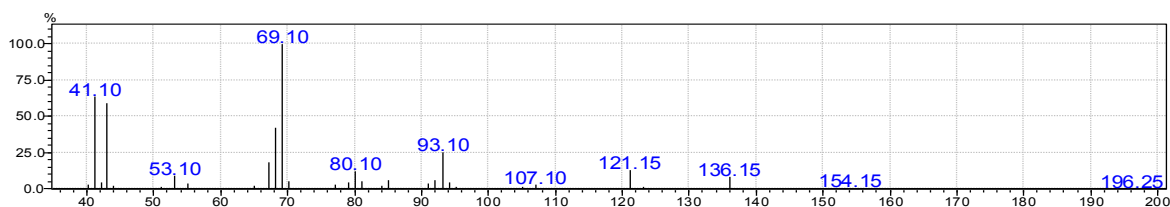
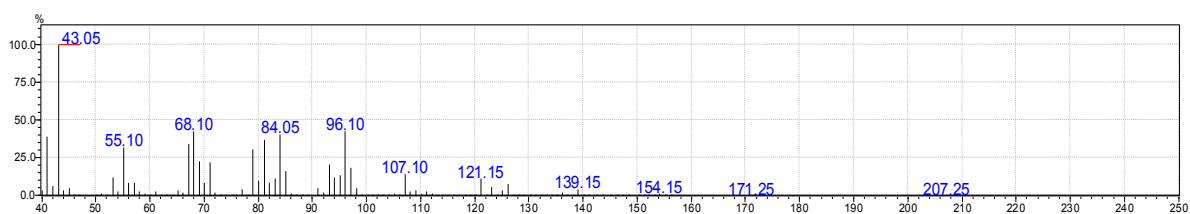


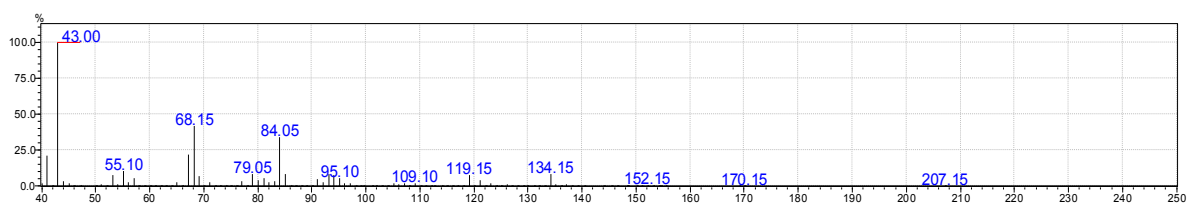
Figure S8: (a) GC trace of the extracted *in vivo* turnover of CYP268A2 with geranyl acetate. The major product at 4hrs is consumed to produce the 24hr major product.*



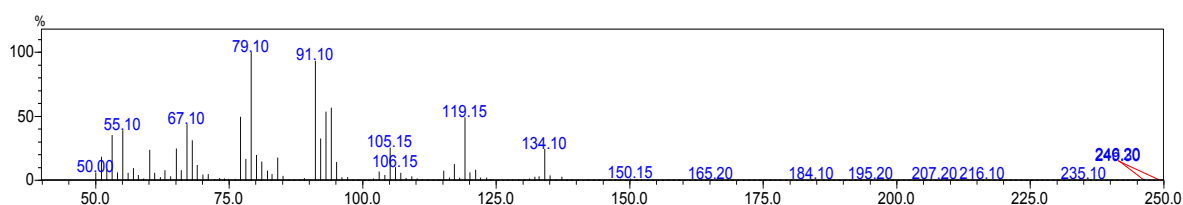
(b) MS of peak at 9.9 mins – substrate (expected mass 196.15)



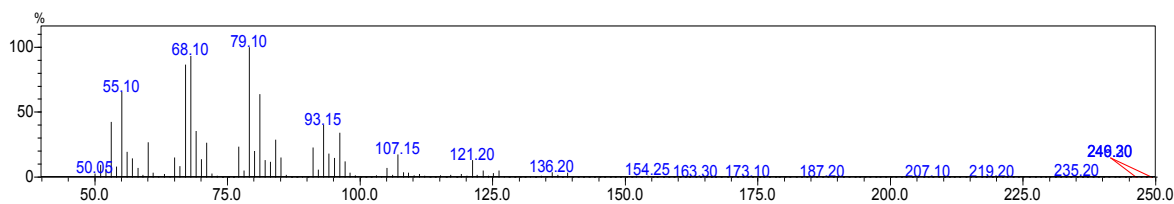
(c) MS of the peak in (a) at 13.25 mins – major product at 24 hrs



(d) MS of the peak in (a) at 13.5 mins – major product at 4 hrs



(e) MS of Figure 5 peak 9.5 mins – major product *in vitro* (expected mass of hydroxylation product is 214, peak at 152 (-acetate) is present but not labelled) *equivalent to peak at 13.5 mins in Figure S8



(f) MS of Figure 5 peak at 9.4 mins – major product *in vivo* *equivalent to peak at 13.25 mins in Figure S8

* At 4 hours the major product is at 13.5 mins, and has an m/z of 152.15. By 24 hours, this peak has disappeared in favour of the peak at 13.25 min, which has mass peaks two mass units higher (154.15), suggesting hydrogenation followed hydroxylation.

Additionally three minor *in vivo* products appear at 12.9, 13.0 and 14.1 mins, which combined form less than 15% of the total product. From m/z values they also appear to be further oxidation products but were not produced in sufficient yield to allow purification.

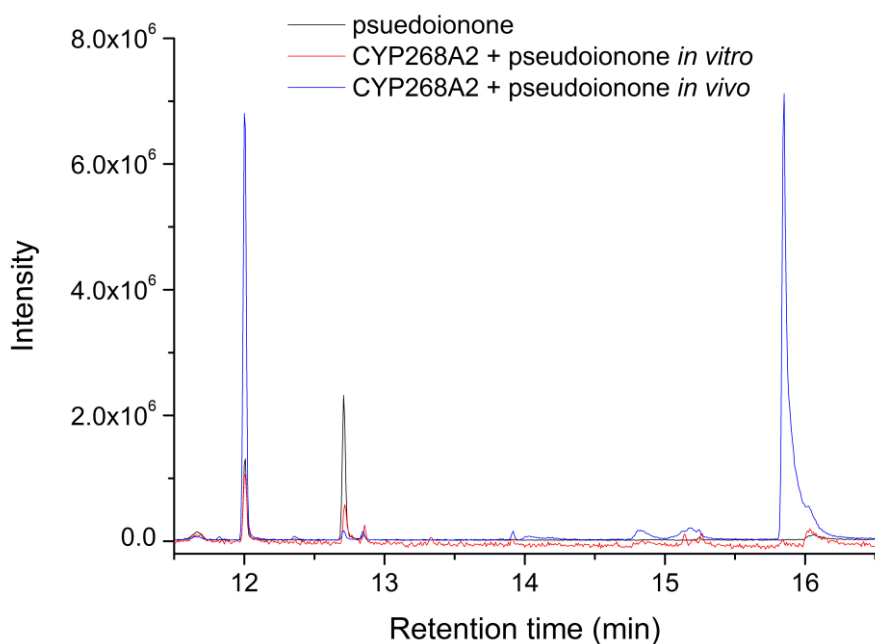
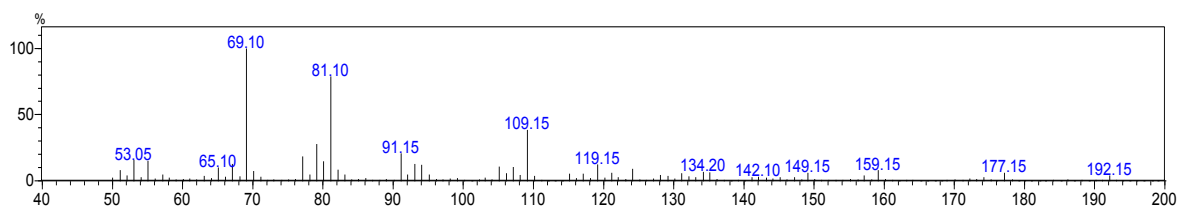
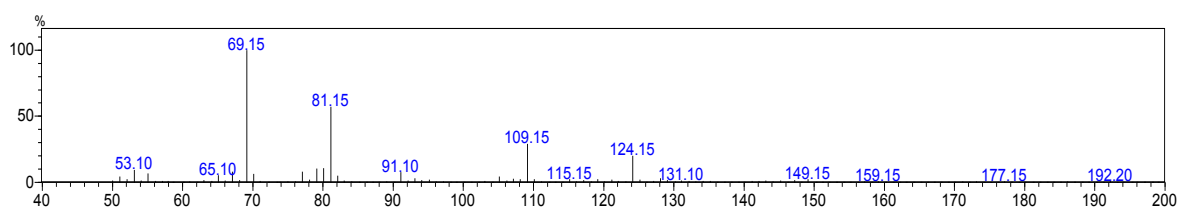


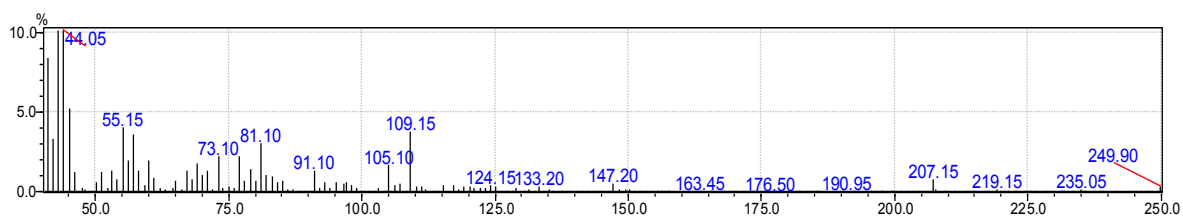
Figure S9: (a) GC trace of the *in vitro* and *in vivo* turnovers of CYP268A2 with pseudoionone showing the change in product distribution over time.



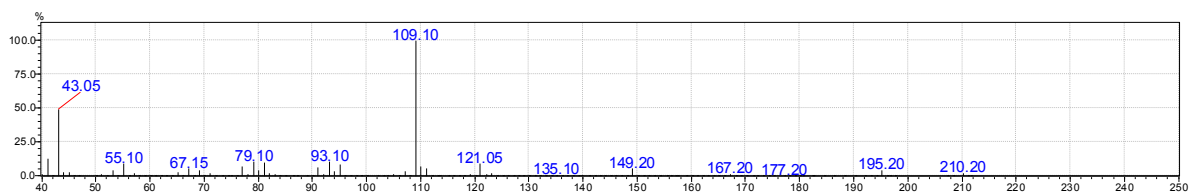
(b) MS of peak at 12.0 mins – *cis* substrate (expected mass 192.15)



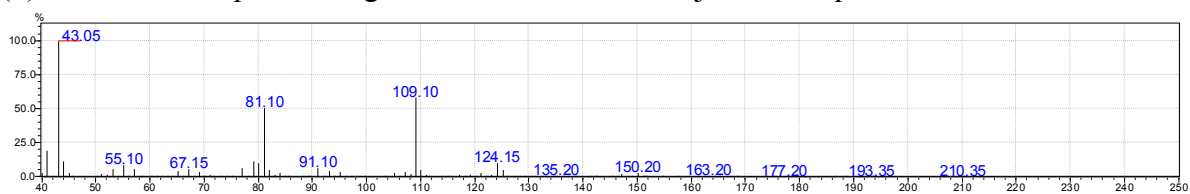
(b) MS of peak at 12.7 mins – *trans* substrate (expected mass 192.15)



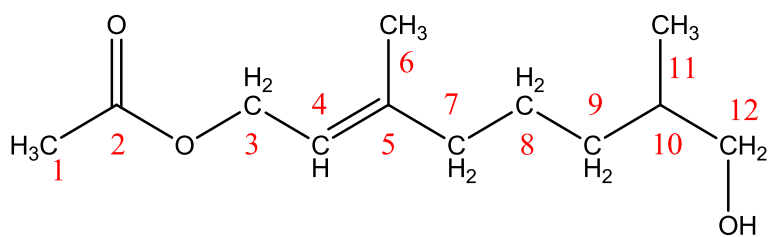
(c) MS of peak at 16.1 mins – major product *in vitro* (expected mass 208, peak at 208.15 present)



(d) MS of the GC peak of Figure 5b at 15.9 mins – major *in vivo* product at 24 hrs



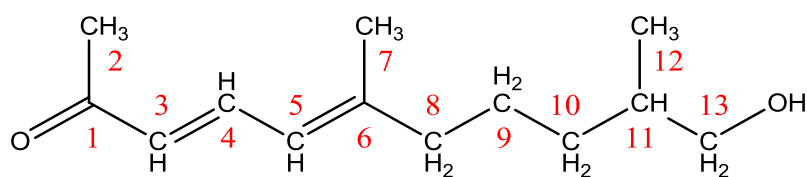
(e) MS of the GC peak of Figure 5b at 16.1 mins – major *in vivo* product at 4 hrs



Scheme S1: Geranyl acetate oxidation product

^1H NMR (500 MHz, CDCl_3) δ 5.41 – 5.28 (tq, $J = 7.25, 1.25$ Hz, 1H, **H4**), 4.59 (d, $J = 7.25$ Hz, 2H, **H3**), 3.55 – 3.38 (m, 2H, **H12**), 2.06 (s, 3H, **H1**), 2.13 – 1.98 (m, 2H, **H7**), 1.70 (s, 3H, **H6**), 1.68 – 1.59 (m, 1H, **H10**), 1.55 – 1.37 (m, 2H, **H8**), 1.45 – 1.35 (m, 2H, **H9**), 1.14 – 1.05 (m, 2H, **H9**), 0.93 (d, $J = 6.7$ Hz, 3H, **H11**).

^{13}C NMR (126 MHz, CDCl_3) δ 171.21 (**C2**), 142.43 (**C5**), 118.36 (**C4**), 68.37 (**C12**), 61.48 (**C3**), 39.77 (**C7**), 35.71 (**C10**), 32.75 (**C9**), 24.95 (**C8**), 21.14 (**C1**), 16.60 (**C11**), 16.42 (**C6**).



Scheme S2: Pseudoionone oxidation product

^1H NMR (600 MHz, CDCl_3) δ 7.42 (ddd, $J = 15.3, 11.4, 5.4$ Hz, 1H, **H4**), 6.08 (d, $J = 15.3$ Hz, 1H, **H3**), 6.00 (d, $J = 11.4$ Hz, 1H, **H5**), 3.60 – 3.32 (m, 2H, **H13**), 2.27 (d, $J = 5.2$ Hz, 3H, **H2**), 2.15 (m, 2H, **H8**), 1.89 (d, $J = 9.7$ Hz, 3H, **H7**), 1.69 – 1.59 (m, 1H, **H11**), 1.54 (s, 6H, water), 1.59 – 1.41 (m, 2H, **H9**), 1.46 – 1.37 (m, 2H, **H10**), 1.17 – 1.06 (m, 2H, **H10**), 0.93 (m, 3H, **H12**).

^{13}C NMR (126 MHz, CDCl_3) δ 201.46 (**C1**), 154.27 + 153.90 (**C6**), 142.20 + 141.70 (**C4**), 131.08 + 130.85 (**C3**), 127.18 + 126.36 (**C5**), 70.86 + 70.85 (**C13**), 43.29 (**C8**), 38.28 + 38.26 (**C11**), 35.64 (**C2**), 35.48 + 35.38 (**C10**), 30.43 + 30.17 (**C2**), 28.34 + 27.70 (**C9**), 27.15 + 20.06 (**C7**), 19.16 (**C12**).

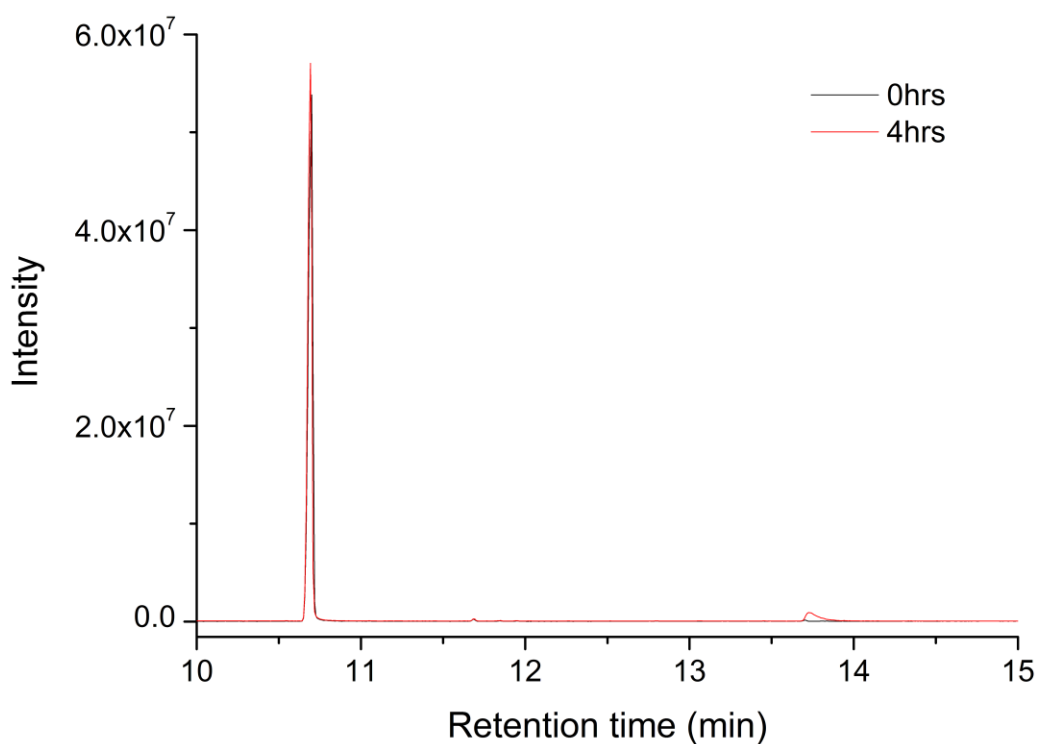
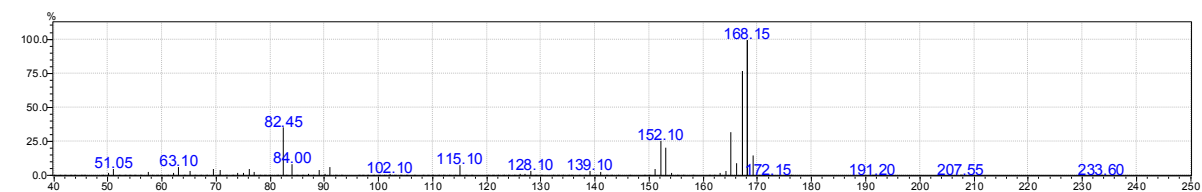
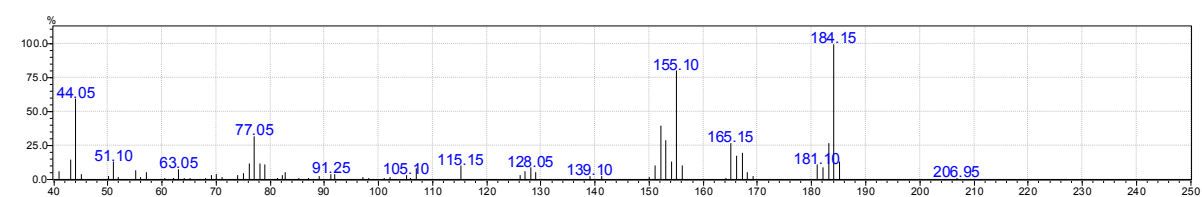


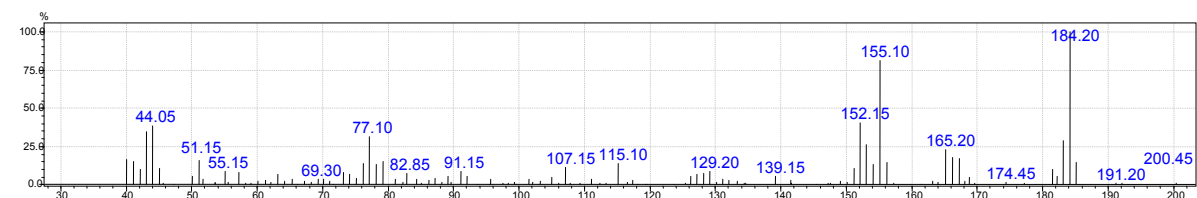
Figure S10: (a) GC trace of the extracted turnover of CYP268A2 with 4-phenyltoluene



(b) 10.8 mins – substrate 4-phenyltoluene



(c) 13.7 mins – product (4-biphenylmethanol)



(d) 4-biphenylmethyl product of 101B1 with 4-phenyltoluene [10]

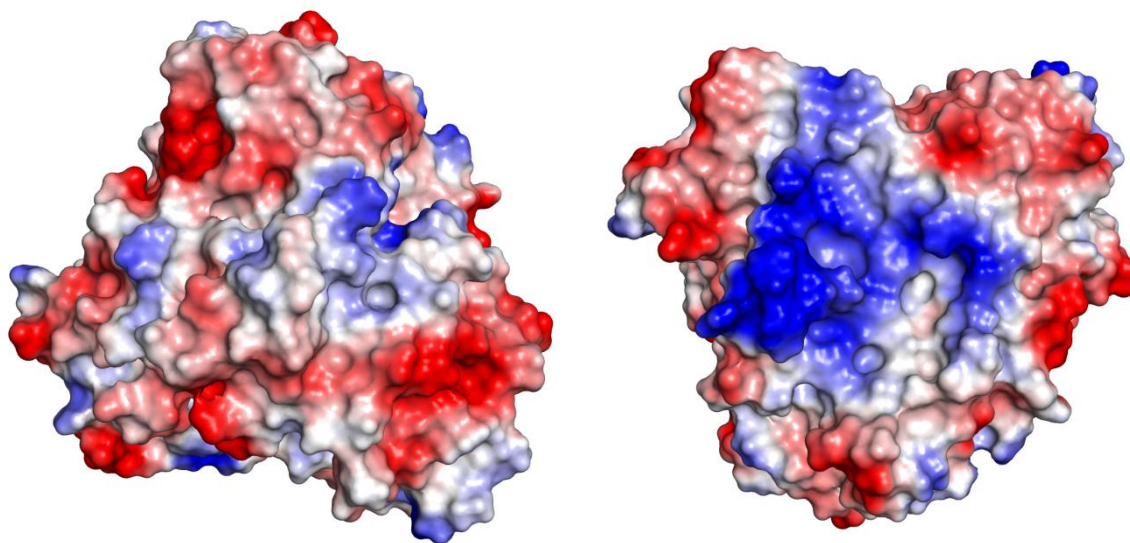


Figure S11: The (a) distal and (b) proximal surface of CYP268A2 (to the heme), with an applied protein contact potential (negatively charged in red, positively charged in blue). Note there is no visible access to the substrate binding site in the distal face.

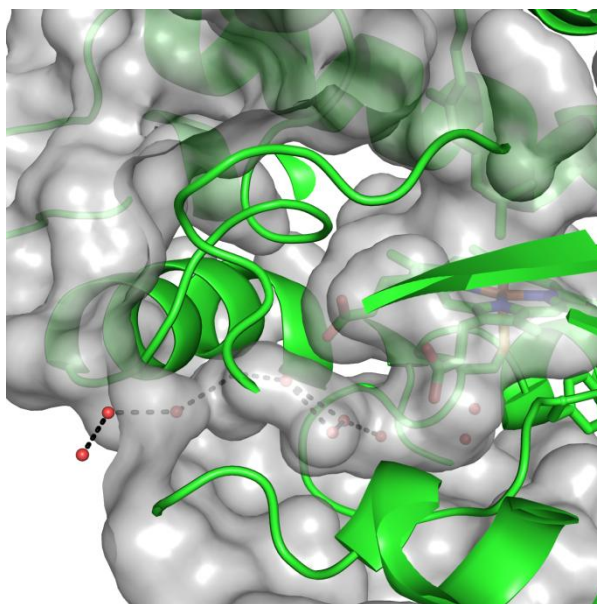


Figure S12: CYP268A2 (green) showing a proximal water channel (waters red, hydrogen bonds black), leading from the surface of the protein to the proximal face of the heme.

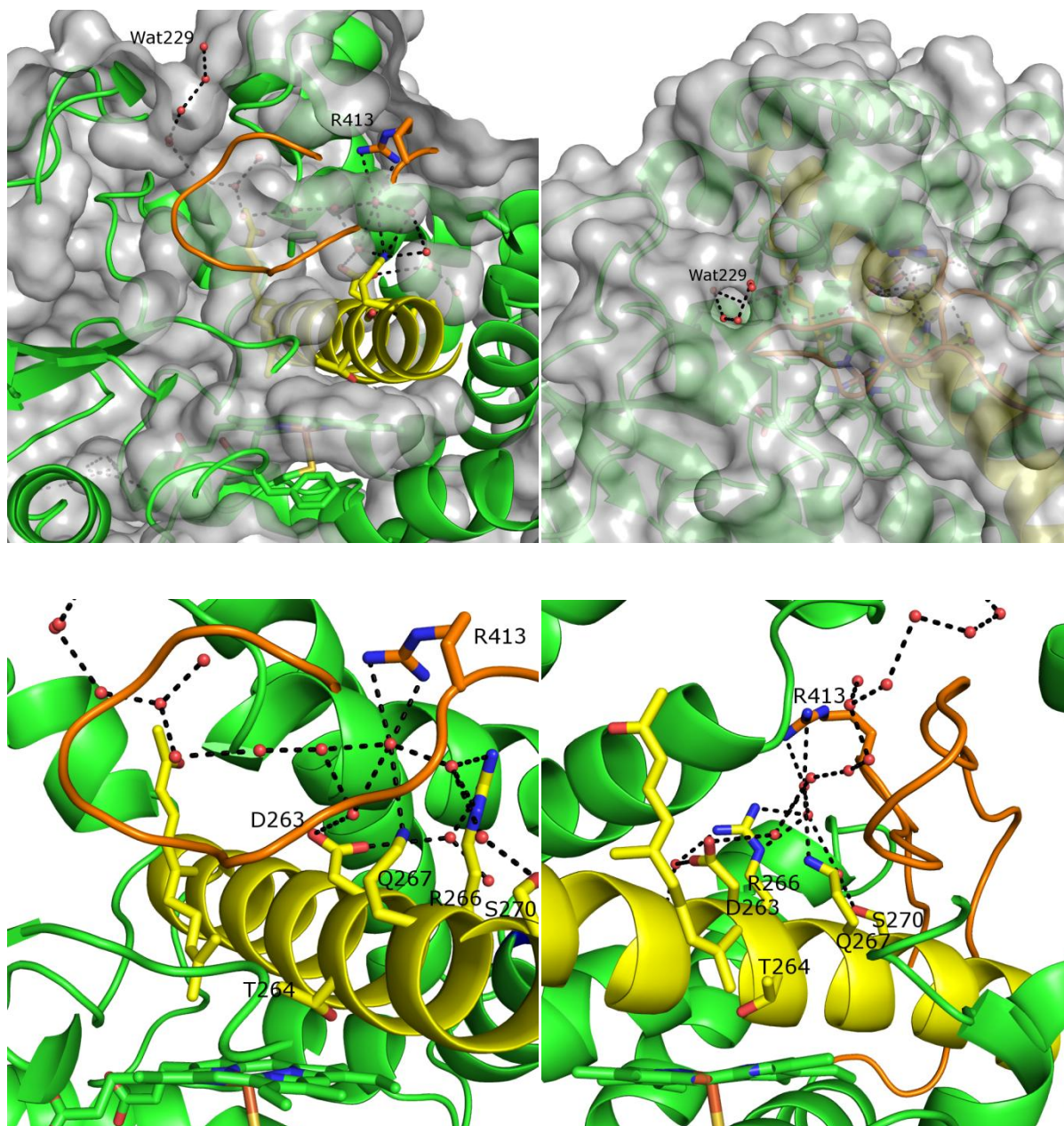


Figure S13: (a) CYP268A2 (green with I-helix (residues 247- 278) in yellow and the unstructured region of the C terminus in orange (residues 393 – 419), pseudoionone also in yellow) showing internal solvent channel (b) external surface of the protein showing solvent access channel (c and d) residues of the I helix and C-terminal end in a network of hydrogen bonds to solvent molecules, including D263 (the acid of the acid-alcohol pair).

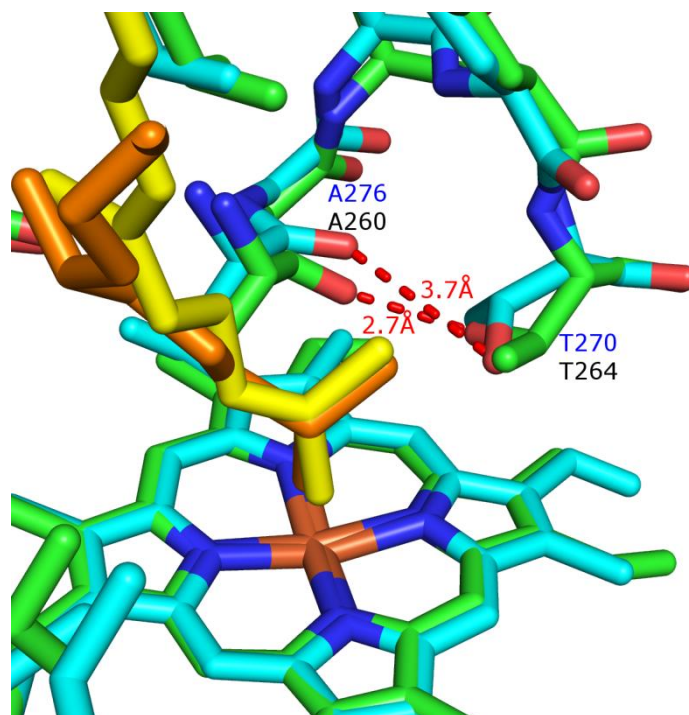


Figure S14: CYP268A2 (green) and CYP124A1 (blue) region of the I-helix loop containing the threonine alcohol residue showing hydrogen bond (red) to the alanine four residues earlier in both enzymes. The flipped threonine residue in CYP124A1 increases the putative hydrogen bond distance from 2.7 Å in CYP268A2 to 3.7 Å.

References

- 1 Nelson, D. R. (2009) The cytochrome p450 homepage. *Human genomics*. 4, 59-65
- 2 Tanaka, M., Haniu, M., Yasunobu, K. T. and Kimura, T. (1973) The amino acid sequence of bovine adrenodoxin. *J. Biol. Chem.*. 248, 1141-1157
- 3 Suhara, K., Nakayama, K., Takikawa, O. and Katagiri, M. (1982) Two forms of adrenodoxin reductase from mitochondria of bovine adrenal cortex. *FEBS J.* 125, 659-664
- 4 Bell, S. G., Tan, A. B., Johnson, E. O. and Wong, L. L. (2010) Selective oxidative demethylation of veratric acid to vanillic acid by CYP199A4 from *Rhodopseudomonas palustris* HaA2. *Molecular bioSystems*. 6, 206-214
- 5 Bell, S. G., Dale, A., Rees, N. H. and Wong, L.-L. (2010) A cytochrome P450 class I electron transfer system from *Novosphingobium aromaticivorans*. *Appl. Microbiol. Biotechnol.*. 86, 163-175
- 6 Bell, S. G., Yang, W., Yorke, J. A., Zhou, W., Wang, H., Harmer, J., Copley, R., Zhang, A., Zhou, R., Bartlam, M., Rao, Z. and Wong, L. L. (2012) Structure and function of CYP108D1 from *Novosphingobium aromaticivorans* DSM12444: an aromatic hydrocarbon-binding P450 enzyme. *Acta Crystallogr. D*. 68, 277-291
- 7 Bell, S. G., French, L., Rees, N. H., Cheng, S. S., Preston, G. and Wong, L. L. (2013) A phthalate family oxygenase reductase supports terpene alcohol oxidation by CYP238A1 from *Pseudomonas putida* KT2440. *Biotechnol. Appl. Biochem.*. 60, 9-17
- 8 Robert, X. and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*. 42, W320-W324
- 9 Overmars, L., Kerkhoven, R., Siezen, R. J. and Francke, C. (2013) MGcV: the microbial genomic context viewer for comparative genome analysis. *BMC genomics*. 14, 209
- 10 Hall, E. A., Sarkar, M. R. and Bell, S. G. (2017) The selective oxidation of substituted aromatic hydrocarbons and the observation of uncoupling via redox cycling during naphthalene oxidation by the CYP101B1 system. *Catal Sci Technol*. 7, 1537-1548

Chapter 6 Supplementary Information

Gene fragments and oligonucleotides used in this work.

Primers used to clone the cytochrome P450 genes of MmarCYP124A1, CYP125A6, CYP125A7, and CYP142A3. The restriction sites are underlined and the start and stop codons are highlighted in bold. Cloned into pEt26 NdeI HindIII, (CYP124A1), NdeI XhoI (CYP142A3 and CYP125A6) and NdeI EcoRI (CYP125A7).

CYP125A6 NdeI 5'

5' – TTAATTCAT**ATG**CCAGCTGCCGAGCCAACC- 3'

CYP125A6 XhoI 3'

5' – TTAATTCTCGAGCT**ATT**AGTGCGCGACAGGACATTTCCC – 3'

Cannot use KpnI or HindIII therefore ferredoxin must be cloned first for in vivo system

CYP124A1 NdeI 5'

5' – TTAATTCAT**ATG**GACCTCAGCACGAACCTCAAC- 3'

CYP124A1 KpnI 3'

5' – TTAATTAAGCTTGGTACCCT**ATTA**ACGGCTCCACGCGACTGGATC–3'

CYP125A7 NdeI 5'

5' – TTAATTCAT**ATG**CCCTTGCCCCAACCTTCCGC- 3'

CYP125A7 KpnIEcoRI 3'

5' – TTAATTGAATCCGGTACCCT**ATTA**ATGTGAAACCGGGCACTTG–3'

Has a HindIII site present and must go into the in vivo system second

CYP142A3 NdeI 5'

5' – TTAATTCAT**ATG**ACTAAGCCGTTGATCAAACC-3'

CYP142A3 XhoI 3'

5' – TTAATTCTCGAGCT**ATT**AGCTCAGTGGCCGGCTGGGAGTG-3'

MtbCYP124A1, CYP125A1 and CYP142A1 were each obtained as a gblock. NdeI and HindIII sites were incorporated at the 5' and 3' ends, respectively. A double stop codon has been inserted at the C terminus. Start and stop codons are underlined, and restriction enzyme sites of NdeI and HindIII are in bold.

M. tuberculosis CYP124A1 Mtb NdeI - HindIII

TCTGAT**CATATG**GGCCTGAACACGGCCATAGCTACACGTGTCAATGGCACTCCA
CCACCGGAAGTTCCAATCGCTGATATTGAGTTAGGCTCCCTTGATTTCTGGGCGT
TGGACGACGACGTTAGAGACGGTGCTTTCGCCACACTGCGCAGAGAGGCACCAA
TTTCATTTTGGCCTACAATAGAATTACCGGGTTTTGTAGCCGGGAACGGACATTG
GGCTCTGACCAAATACGACGATGTTTTCTACGCTAGTCGGCACCCCGATATTTTT
TCGAGCTATCCCAATATCACGATAAACGACCAGACACCTGAGCTTGCAGAATAC
TTTGGCAGTATGATTGTTCTGGATGACCCTCGCCACCAGAGACTTCGCTCCATTG
TGTCGAGAGCGTTTACGCCAAGGTAGTGGCACGTATAGAGGCCCGCAGTGAGAG
ACCGCGCTCATAGACTTGTAAGTTCAATGATCGCAAACAATCCGGACCGGCAGG
CGGACCTGGTAAGTGAATTAGCGGGTCCATTACCCCTGCAGATtATATGCGATAT
GATGGGTATTCCAAAGGCCGACCATCAACGGATTTTCCATTGGACAAATGTCATA
TTGGGGTTCGGGGATCCTGATCTGGCGACTGATTTTCGATGAGTTCATGCAAGTTT
CAGCTGACATCGGAGCCTATGCAACCGCGCTGGCAGAGGACAGAAGAGTAAATC
ACCATGACGACTTAACTTCCTCTCTTGTGTAAGCCGAGGTAGATGGTGAAGACT
GTCATCACGCGAAATTGCCTCCTTTTTTATATTGTTGGTTGTAGCGGGGAATGAA
ACTACTCGTAACGCCATCACACACGGAGTACTGGCTCTTAGTAGATACCCTGAGC
AGCGGGACCGCTGGTGGTCAGATTTTCGATGGCCTTGCACCGACGGCGGTGGAAG
AAATAGTAAGATGGGCCTCCCCCGTAGTCTATATGAGACGCACACTTACTCAAG
ACATTGAGCTTCGCGGTAATAAAATGGCTGCGGGTGACAAAGTTAGTCTTTGGTA
CTGCAGTGCAAACCGGGACGAGTCTAAGTTCGCTGATCCTTGGACCTTCGATTTA
GCCCACAACCCGAATCCGCACTTAGGGTTCGGTGGTGGTGGAGCTCACTTCTGCT
TGGGCGCCAATCTTTCGAGACGCGAGATACGTGTGGCTTTTGACGAATTACGTCG
TCAGATGCCCGATGTGGTCGCAACCGAGGAACCTGCGCGGCTTTTATCCCAGTTT
ATCCACGGTATTAAGACTTTGCCTGTGACGTGGTCCCATCATCACCACCATCATT
AATGAGGTACCAAGCTTTTATGC

M. tuberculosis CYP125A1 Mtb NdeI - HindIII

TCTGAT**CATATG**AGTTGGAACCACCAGTCGGTTGAAATAGCAGTTCGTCGTACC
ACGGTCCCATCCCCTAACTTGCCTCCGGGCTTCGATTTCACTGATCCTGCTATTTA
CGCTGAGCGTCTGCCGGTGGCAGAATTCGCAGAGTTGCGGTGAGCAGCCCCAT
CTGGTGGAAATGGACAGGACCCGGGCAAGGGTGGCGGCTTCCATGACGGAGGTTT
TTGGGCGATCACAAAACCTAACGATGTGAAGGAGATAAGTCGTCATTTCGGACGT
ATTTAGCTCCTATGAAAACGGCGTAATCCCACGCTTTAAGAATGATATTGCGCGC
GAAGACATTGAGGTCCAACGGTTTTGTGATGCTTAATATGGACGCGCCCCATCATA
CACGGCTTCGGAAGATTATATCAAGAGGGTTCACGCCACGCGCGGTCGGGAGAC
TTCACGATGAGCTTCAAGAACGCGCACAGAAAATAGCTGCAGAAGCAGCGGCTG
CTGGATCGGGTACTTTGTTGAGCAAGTCAGTTGTGAGCTTCCCCTTCAGGCAAT
AGCGGGTTTGCTGGGAGTACCTCAGGAAGACCGGGGGAAGTTGTTCCACTGGTC

GAACGAAATGACTGGGAACGAGGACCCGGAGTACGCTCATATAGACCCGAAGG
CTTCGTCAGCAGAATTAATTGGGTACGCAATGAAGATGGCGGAGGAGAAGGCCA
AGAACCCGGCAGATGATATAGTAACACAGTTAATACAAGCAGATATAGACGGGG
AGAAaCTTTCGGACGACGAATTCGGTTTCTTCGTTGTCATGCTTGCGGTAGCCGG
GAACGAGACCACACGTAACCTCGATCACCCAAGGGATGATGGCTTTTGCGGAGCA
CCCGGATCAGTGGGAGCTGTATAAAAAAGTCCGTCCGGAAACCGCAGCAGATGA
AATAGTGCCTGGGCGACCCCTGTAACAGCCTTTCAGCGGACAGCGCTTAGAGA
TTACGAATTGAGCGGCGTACAAATTA AAAAGGGACAACGCGTGGTTCATGTTTTAT
CGCTCGGCCAATTTTGACGAGGAGGTATTCCAGGACCCATTTACATTCAATATTC
TGCGGAATCCTAACCCACACGTGGGCTTCGGGGGGACGGGAGCCCATTATTGCA
TAGGCGCGAATCTGGCGCGTATGACCATCAATCTGATTTTCAACGCGGTGCGCGA
TCACATGCCCGATCTGAAACCTATATCGGCTCCTGAACGCCTGCGCTCTGGGTGG
CTGAACGGAATTAAGCATTGGCAGGTTGATTATACGGGTAGATGCCCGGTGGCT
CATCATCACCACCATCATTAATGAGGTACCAAGCTTTTATGC

M. tuberculosis CYP142A1 NdeI-HindIII

TCTGATCATATGACCGAAGCCCCGATGTAGATTTGGCAGATGGGAACTTCTAC
GCAAGCCGTGAAGCGAGAGCTGCGTATCGCTGGATGCGTGCGAACCAACCTGTA
TTCAGAGACCGTAATGGCTTAGCAGCAGCCTCCACATACCAAGCAGTCATCGAT
GCAGAGCGTCAGCCGGAGTTGTTCTCCAACGCTGGGGGCATCAGACCCGACCAA
CCGGCATTGCCCATGATGATAGACATGGACGACCCTGCACACCTGCTTCGCCGCA
AGTTGGTAAATGCAGGTTTCACTCGTAAAAGAGTTAAGGACAAAGAAGCCTCTA
TTGCCGCTCTTTGTGACACTCTGATTGATGCCGTGTGCGAGCGTGGCGAATGCGA
CTTCGTAAGAGATTTAGCCGCCCATTTGCCGATGGCCGTAATTGGGGATATGCTT
GGTGTGCGCCCAGAACACGCGATATGTTCTGAGATGGAGTGATGACCTTGTG
ACATTTTATCAAGCCACGTAAGCCAAGAAGACTTTCAAATCACtATGGACGCGT
TTGCCGCTTACAACGATTTTACTCGCGCTACCATAGCCGCCCGGCGCGCAGACCC
TACGGATGATCTTGTTAGCGTTCTTGTGTCAGCAGTGAGGTGGACGGCGAACGTCTT
AGTGATGATGAACTTGTGATGGAGACACTTTTAATCCTGATCGGCGGCGACGAA
ACGACTAGACACACTTTGAGTGGCGGGACCGAGCAGTTGCTGCGCAACCGGGAC
CAGTGGGACCTGCTTCAGCGGGATCCGAGCTTGCTGCCTGGGGCGATTGAAGAG
ATGTTGCGGTGGACAGCACCAGTAAAAAATATGTGTCGCGTGTTGACGGCAGAT
ACTGAGTTCCACGGgACCGCGCTTTGCGCGGGGGAGAAAATGATGTTGTTGTTTCG
AGAGCGCCAACCTTTGACGAGGCCGTCTTCTGCGAACCTGAAAAGTTCGACGTGC
AACGTAACCCTAACTCGCATTTGGCATTTCGGGTTTGGCACACATTTTGCCTGGG
AAACCAGCTGGCACGGCTTGAGCTTCTCTGATGACGGAGCGTGTCTGAGACGT
CTTCTGACCTGCGCCTTGTAGCGGATGACTCAGTATTACCACTTAGACCAGCTA
ACTTCGTTTCCGGTCTGGAATCCATGCCAGTTGTTTTACACCATCGCCTCCCCTG
GGCCATCATCACCACCATCATTAATGAGGTACCAAGCTTTTATGC

Table S1: Active site residues of CYP125 enzymes from *M. marinum*, *M. smegmatis* and *Rhodococcus* sp. RHA1 compared to the identified active site residues of *M. tuberculosis* [1].

Where the CYP125A1 residue is conserved, it is indicated in bold and underlined.

Mtb 125A1	Mmar 125A6	Mmar 125A7	Mulc 125A7	RHA 125	Msmeg 125A3	Msmeg 125A4
<u>I97</u>	<u>I</u>	<u>I</u>	<u>I</u>	<u>I</u>	<u>I</u>	L
<u>F100</u>	T	<u>F</u>	<u>F</u>	<u>F</u>	W	Y
<u>D108</u>	Q	<u>D</u>	<u>D</u>	N	A	Q
<u>V111</u>	R	<u>V</u>	<u>V</u>	M	L	Q
<u>Q112</u>	G	<u>Q</u>	<u>Q</u>	<u>Q</u>	<u>Q</u>	G
<u>V115</u>	<u>V</u>	<u>V</u>	<u>V</u>	I	<u>V</u>	<u>V</u>
<u>M200</u>	<u>M</u>	<u>M</u>	<u>M</u>	<u>M</u>	<u>M</u>	<u>M</u>
<u>G202</u>	<u>G</u>	<u>G</u>	<u>G</u>	<u>G</u>	A	<u>G</u>
<u>P213</u>	A	A	A	<u>P</u>	<u>P</u>	G
<u>K214</u>	L	<u>K</u>	<u>K</u>	Q	A	A
<u>S217</u>	<u>S</u>	<u>S</u>	<u>S</u>	<u>S</u>	<u>S</u>	A
<u>I221</u>	<u>I</u>	<u>I</u>	<u>I</u>	<u>I</u>	<u>I</u>	<u>I</u>
<u>F260</u>	<u>F</u>	<u>F</u>	<u>F</u>	<u>F</u>	<u>F</u>	<u>F</u>
<u>V263</u>	I	<u>V</u>	<u>V</u>	I	V	V
<u>V267</u>	<u>V</u>	<u>V</u>	<u>V</u>	<u>V</u>	<u>V</u>	<u>V</u>
<u>W414</u>	<u>W</u>	<u>W</u>	<u>W</u>	<u>W</u>	<u>W</u>	<u>W</u>

NB: residues that interact with androstenedione: V263, M200, G202, V111, V115, Q112, I97, F100, W414, P213, K214, S217. Residues that interact with econazole: V115 F260, V263, V267, I97, I221, M200, S217, V111, D108, Q112. The acid-alcohol pair are not included in this assessment but are conserved as a glutamate/threonine pair in all.

Table S2: Additional spin state shifts for CYP142A3 and CYP142A1 with selected steroids.

	<i>M.mar142A3</i> Spin state (HS%)	<i>M.tb142A1</i> Spin state (HS%)
Stigmasterol	>5	0
Estriol	>5	0
Lanosterol	50	90
Calciferol	>5	-
Pregnenolone	0	-
Progesterone	0	-
Testosterone	>5	-

Table S3: Active site residues of CYP142 enzymes from *M. marinum*, *M. smegmatis* and *Rhodococcus* sp. RHA1 compared to the identified active site residues of *M. tuberculosis* [2]. CYP125A1 and CYP124A1 from *M. tuberculosis* are included for comparison.

Mtb 142A1	Mmar 142A3	Mulc 142A3	Msmeg 142A2	RHA 142	Mtb 124	Mtb 125
<u>I65</u>	<u>I68</u>	<u>I70</u>	<u>I68</u>	<u>I69</u>	<u>I94</u>	V96
<u>I76</u>	<u>I79</u>	<u>I81</u>	<u>I79</u>	<u>I83</u>	<u>I111</u>	L117
<u>V160</u>	<u>V163</u>	<u>V165</u>	<u>V163</u>	M167	L198	T201
<u>L226</u>	<u>L229</u>	<u>L231</u>	<u>L229</u>	<u>L230</u>	<u>L263</u>	M264
<u>I229</u>	<u>I232</u>	<u>I234</u>	<u>I232</u>	<u>I236</u>	V266	V267
<u>G230</u>	<u>G233</u>	<u>G235</u>	<u>G233</u>	<u>G237</u>	A267	A268
<u>T234</u>	<u>T237</u>	<u>T239</u>	<u>T237</u>	<u>T241</u>	<u>T271</u>	T272
<u>V277</u>	<u>V280</u>	<u>V282</u>	<u>V280</u>	<u>I284</u>	<u>V315</u>	<u>V313</u>
<u>M280</u>	<u>M283</u>	<u>M285</u>	<u>M283</u>	<u>M287</u>	<u>M318</u>	F316
<u>F380</u>	<u>F383</u>	<u>F385</u>	<u>F382</u>	<u>F386</u>	<u>F416</u>	W414
<u>V381</u>	<u>V384</u>	<u>V386</u>	<u>V381</u>	<u>V387</u>	I417	L415

Table S4: Additional spin state shifts for CYP124 from *M. marinum* and *M. tuberculosis*. Data from elsewhere is noted, otherwise was performed as recorded in the experimental section.

	<i>M.mar124</i> Spin state (HS%)	<i>M.tb124</i> Spin state (HS%)	K_D (μM)
Lauric acid	65	75	$>100^a$ 26.57 ± 4.51^b
Palmitic acid	55	60	$>100^a$
Arachidic acid	50	55	-
Palmitoleic acid	55	55	-
Palmitic acid	55	60	$>100^a$
15-methyl hexadecanoic acid	70	90	1.01 ± 0.07^a
Farnesol	100	100	1.04 ± 0.05^a
Phytanic acid	100	100	0.22 ± 0.006^a
Geraniol	30	<5	25 ± 1.8^a
Clotrimazole	Type II, 426 nm	Type II	0.22 ± 0.05

^a Johnston *et al*, 2009 [3]

^b Vasilevskaya *et al*, 2017 [4]

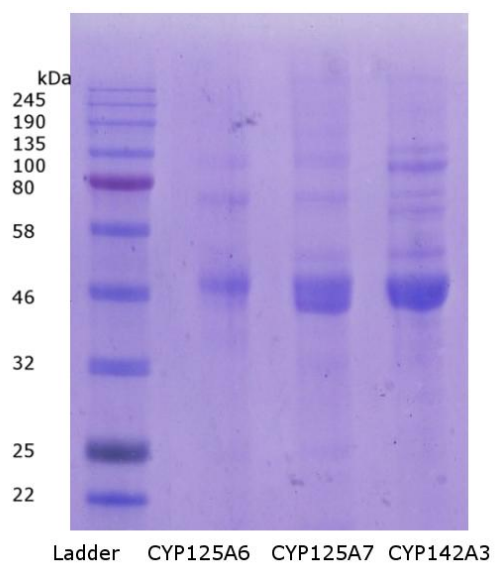


Figure S1: SDS-PAGE of purified *M. marinum* CYP125A6, CYP125A7 and CYP142A3

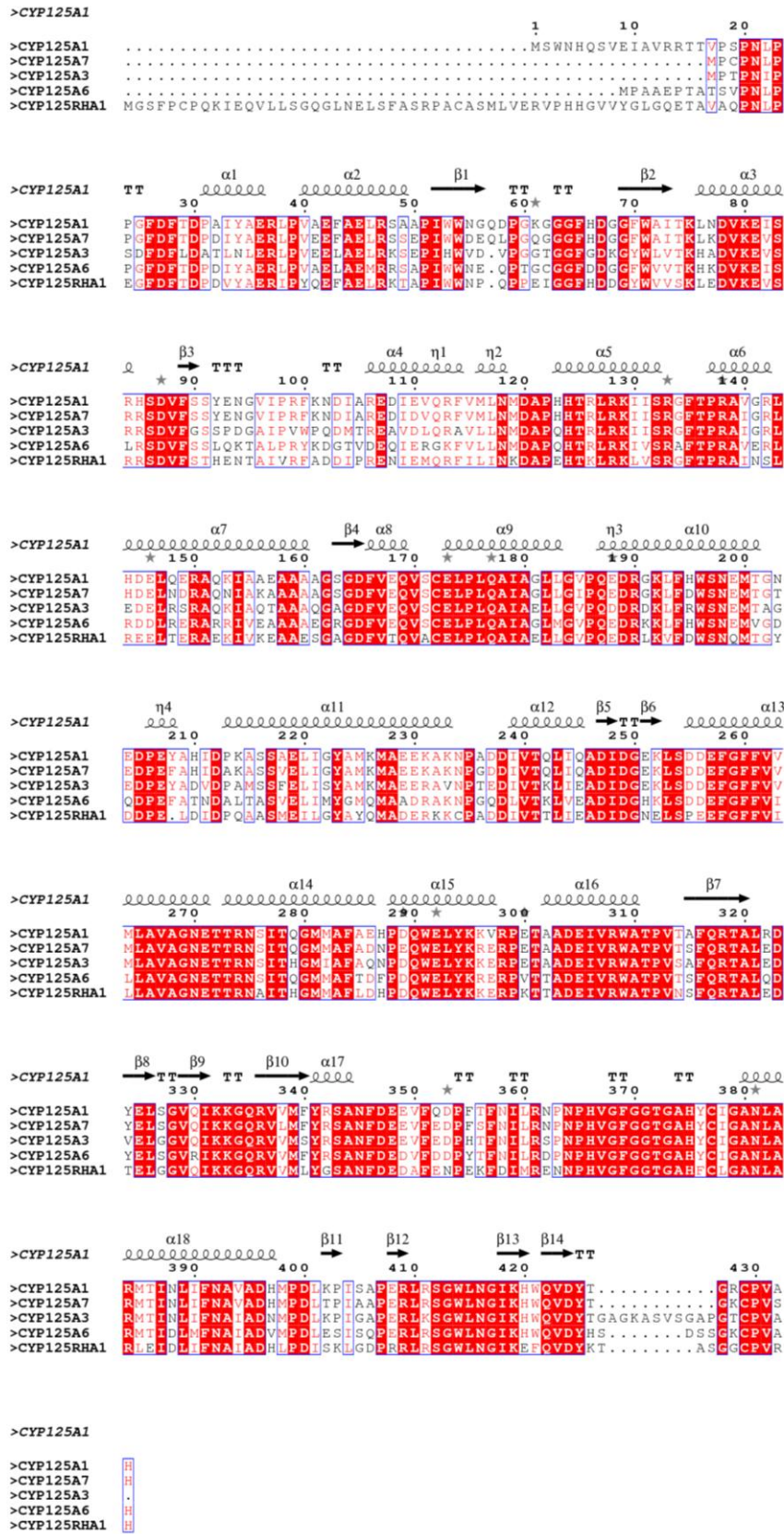


Figure S2: Sequence alignment of CYP125 family members from *M. tuberculosis*, *M. marinum*, *M. smegmatis* and *Rhodococcus* RHA

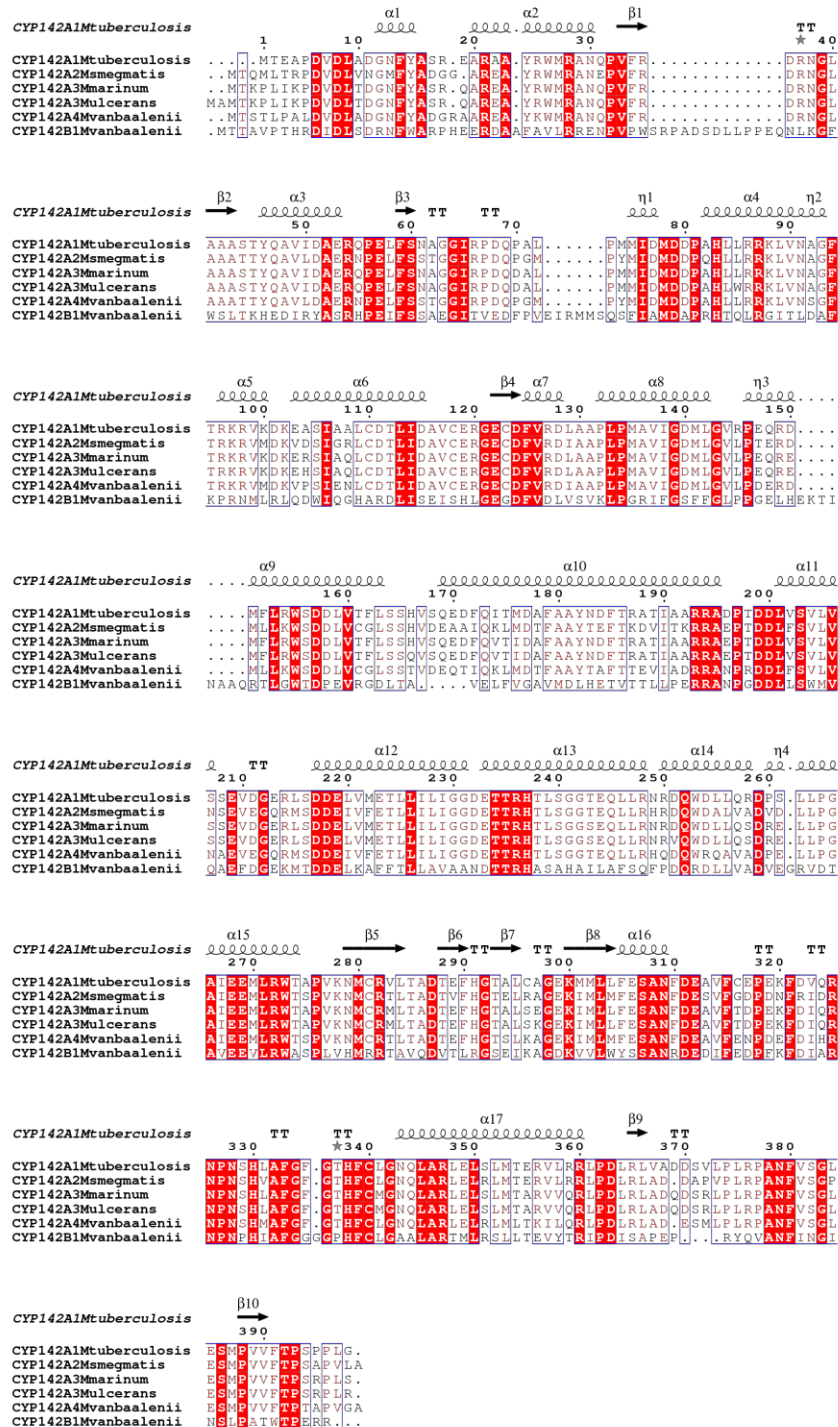
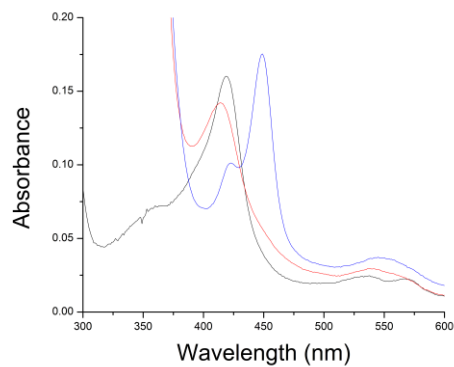


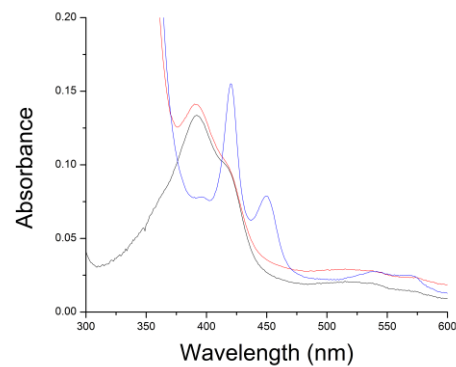
Figure S3: Sequence alignment of CYP142 members from *Mycobacterium* species, including *M. tuberculosis* CYP142A1, *M. smegmatis* CYP142A2, *M. marinum* CYP142A3 and others.



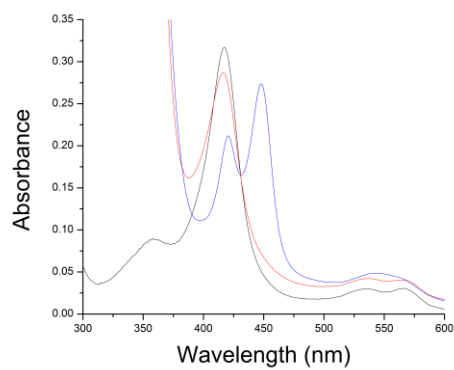
Figure S4: Sequence alignment of CYP124A family members from *M. tuberculosis*, *M. marinum*, *M. ulcerans*, *M. smegmatis* and *M. vanbaalenii*.



(a) Mtb124A1



(b) Mtb125A1



(c) Mtb142A1

Figure S5: CO binding spectra for the *M. tuberculosis* enzymes CYP124A1, CYP125A1 and CYP142A1. For each, the resting state (black), the reduced enzyme (red) and the reduced CO bound (blue) spectra are shown. The CYP125A1 enzyme had split A_{450}/A_{420} peak in the CO-bound state.

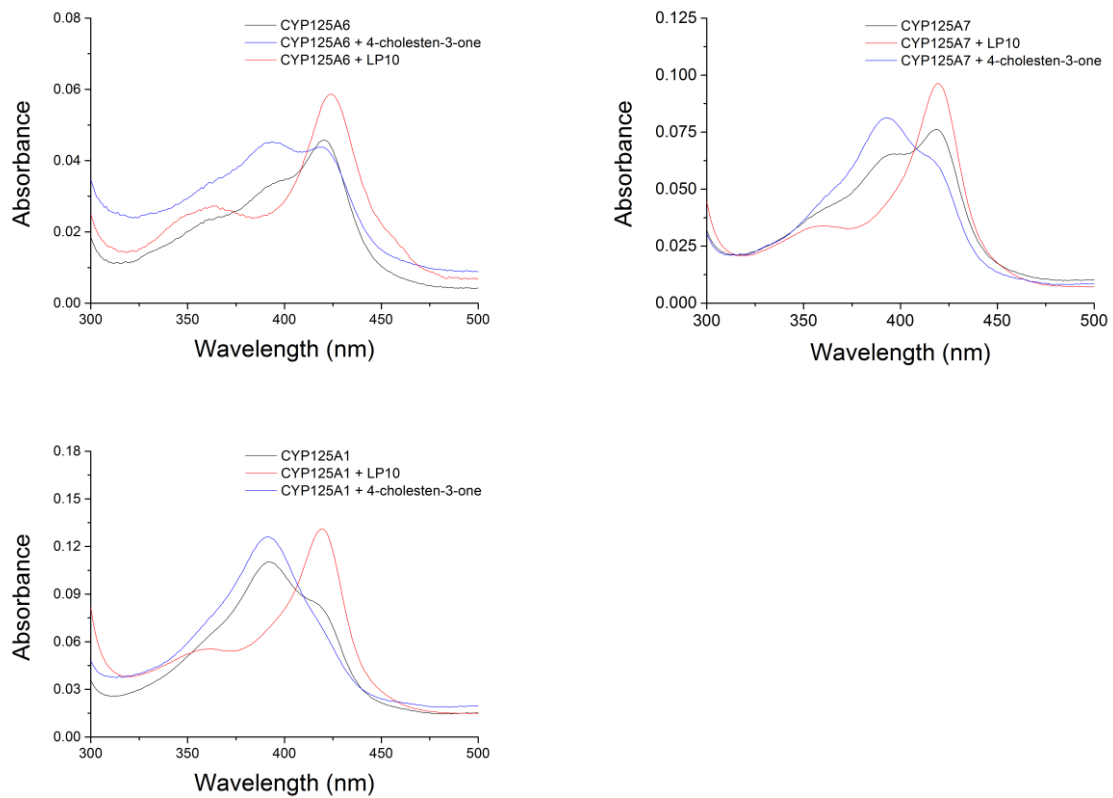
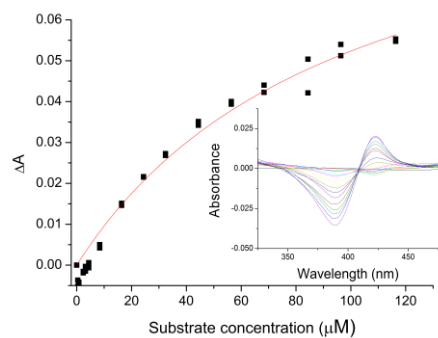
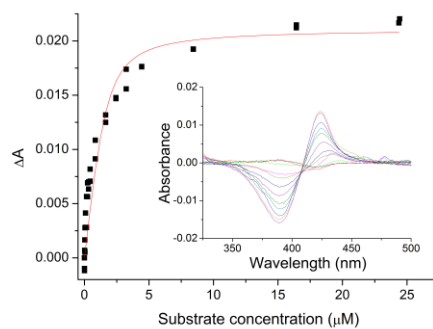


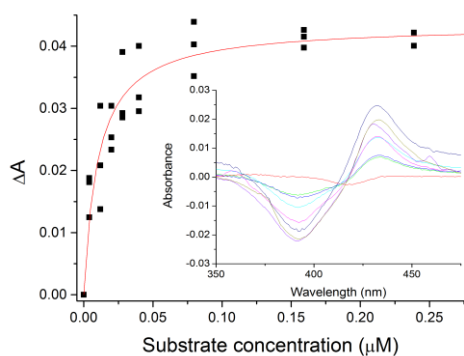
Figure S6: Selected spin state shifts with CYP125 enzymes



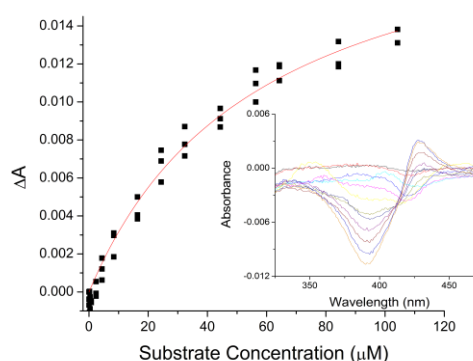
CYP125A1 with LP10



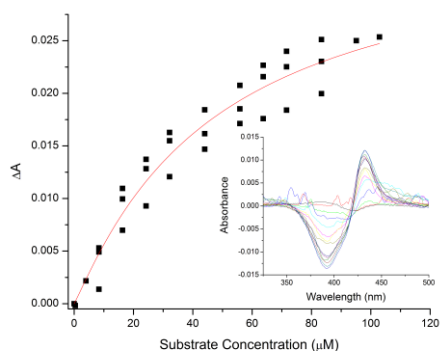
CYP125A6 with LP10



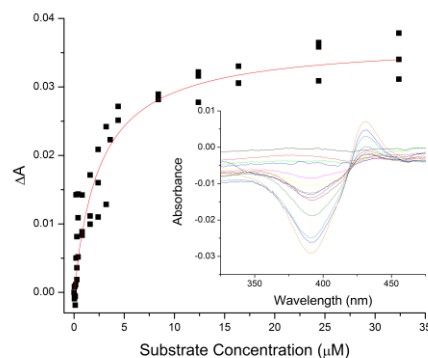
CYP125A7 with LP10



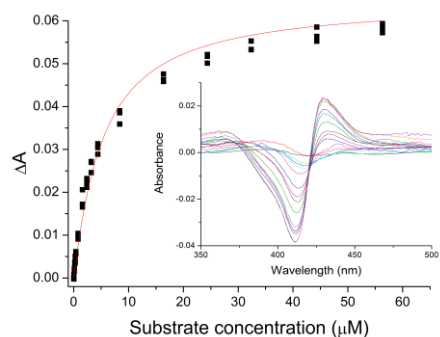
CYP125A6 with 1-phenylimidazole



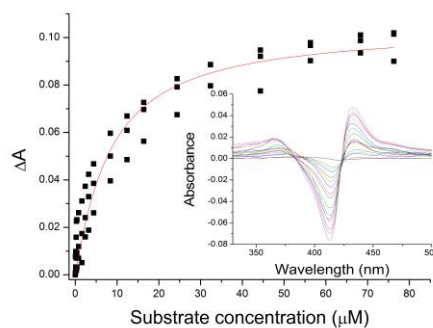
CYP125A6 with 4-phenylimidazole



CYP125A6 with clotrimazole

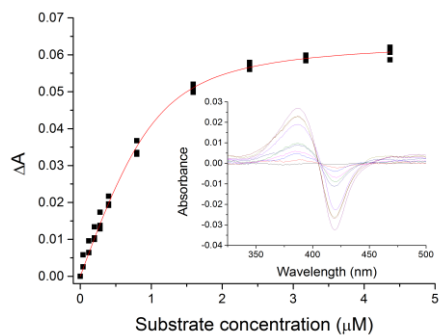


CYP142A3 with 1-phenylimidazole

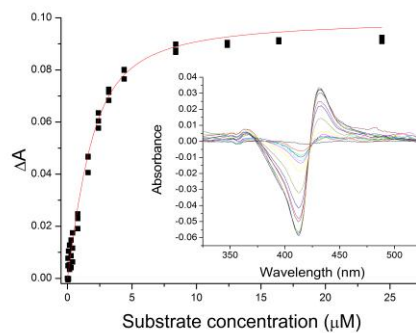


CYP142A3 with 4-phenylimidazole

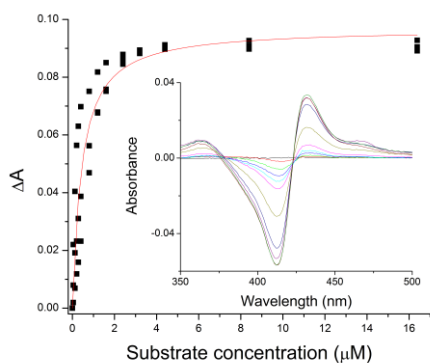
Figure S7: Additional dissociation constants with *M. marinum* and *M. tuberculosis* CYPs. The peak to trough for LP10 with CYP125A1 was recorded at ~422 to 388 nm; CYP125A6 was 423 to 389 nm; CYP125A7 was 432 to 391 nm. Peak to trough with azole substrates varied by enzyme: for CYP125A6, 428 to 390 nm; CYP125A7, 422 to 387 nm; most others 434 to 414 nm.



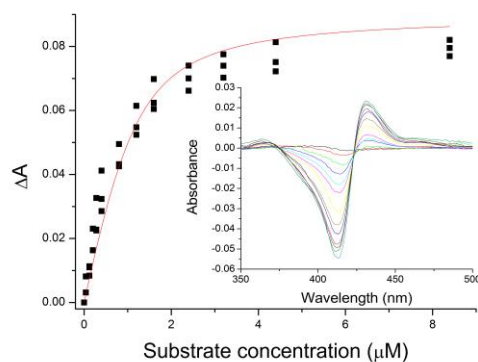
CYP142A3 with cholesterol



CYP142A3 with econazole



CYP142A3 with miconazole



CYP142A3 with clotrimazole

Figure S7: (continued) Additional dissociation constants with *M. marinum* and *M. tuberculosis* CYPs.

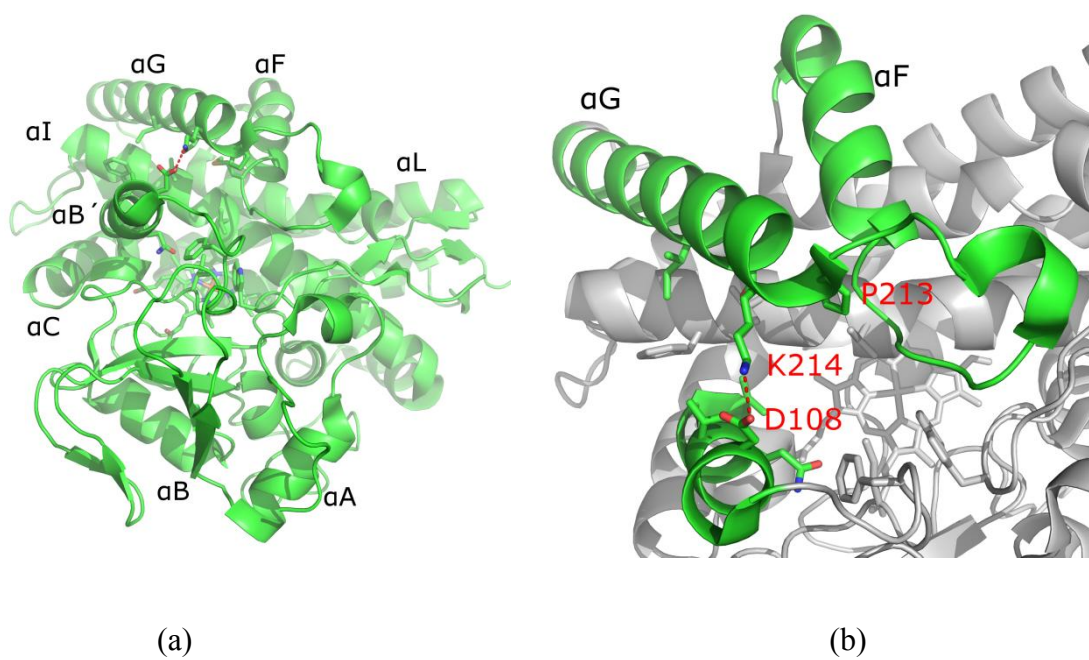
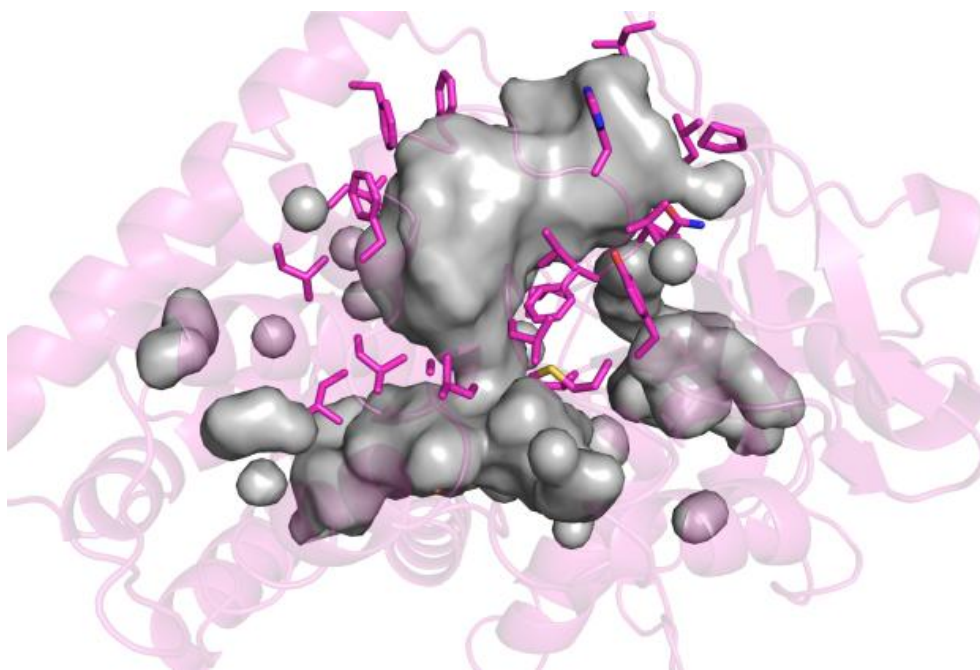
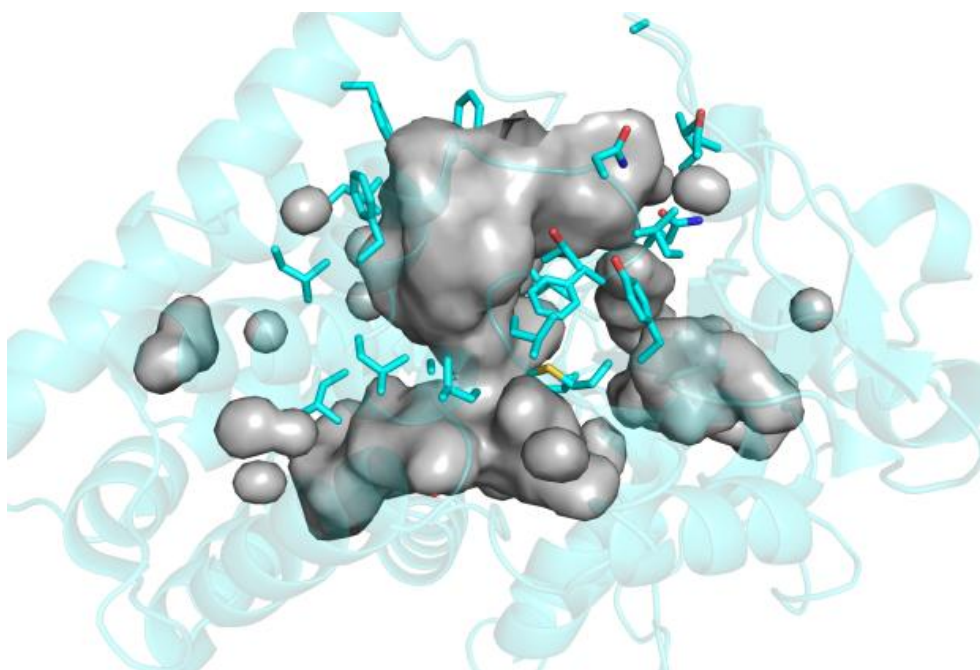


Figure S8: For reference, the CYP125A1 structure (PDB: 2X5L [1]) showing (a) the entire structure and (b) position of the key residues (the D108-K214 salt bridge and P213) discussed in the text.

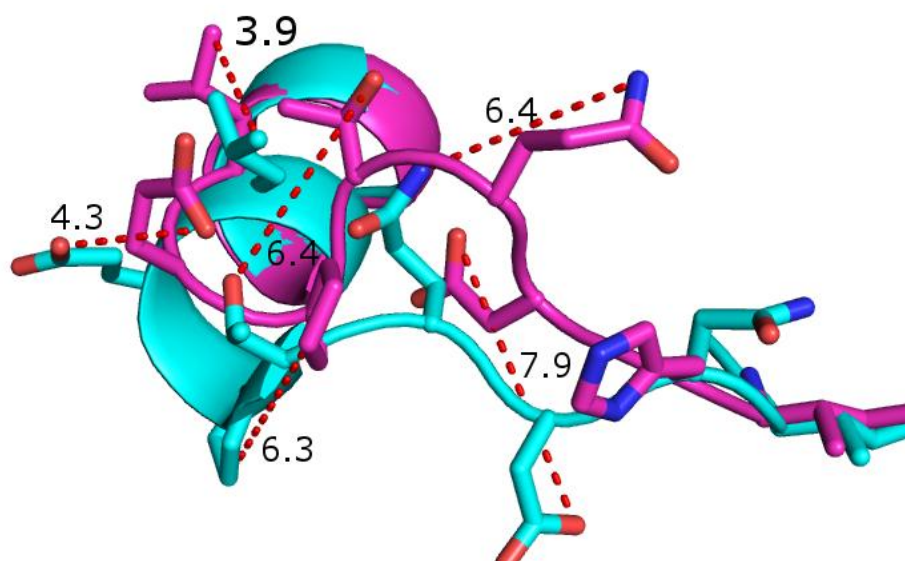


(a)

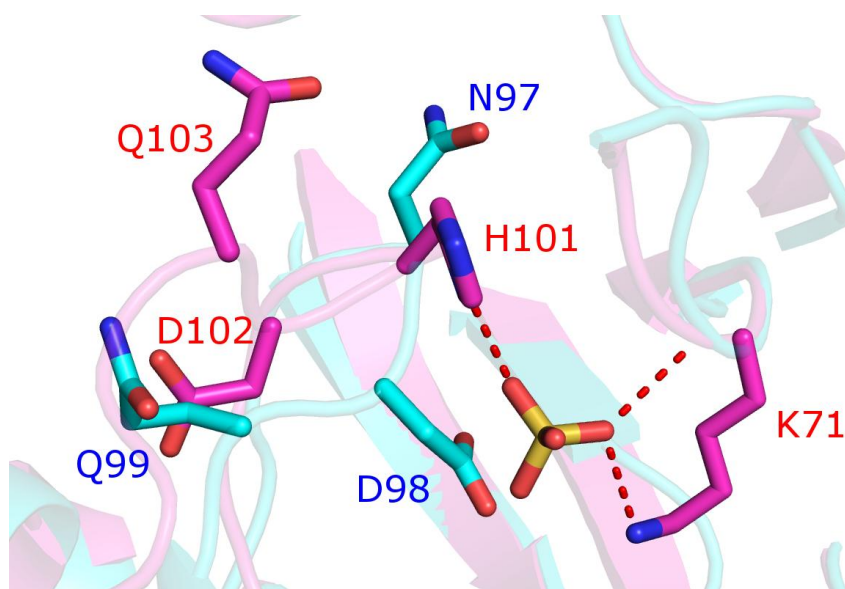


(b)

Figure S9: Visualisation of the active site of (a) substrate-free MmarCYP124 (purple) with the internal surface cavities showing, and for comparison (b) substrate-free MtbCYP124 (blue, PDB: 2WM5 [3]). The active site of MmarCYP124 appears to be more closed toward the exit due to the movement of the residues 101-104.



(a)



(b)

Figure S10: Overlay of MmarCYP124 (purple, with red labels) and MtbCYP124 (PDB: 2WM5 [3], blue cartoon and labels) shows (a) alterations in the position of the B-C loop (b) The interaction of the H101 residue with a bound SO₄ molecule, and the movement away by Q103 are highlighted. The equivalent residue pairs are N97/H101, D98/D102 and Q99/Q103.

References

- [1] K.J. McLean, P. Lafite, C. Levy, M.R. Cheesman, N. Mast, I.A. Pikuleva, D. Leys, A.W. Munro, The Structure of *Mycobacterium tuberculosis* CYP125: Molecular basis for cholesterol binding in a P450 needed for host infection, *J. Biol. Chem.*, 284 (2009) 35524-35533.
- [2] M.D. Driscoll, K.J. McLean, C. Levy, N. Mast, I.A. Pikuleva, P. Lafite, S.E. Rigby, D. Leys, A.W. Munro, Structural and biochemical characterization of *Mycobacterium tuberculosis* CYP142: evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen, *J. Biol. Chem.*, 285 (2010) 38270-38282.
- [3] J.B. Johnston, P.M. Kells, L.M. Podust, P.R. Ortiz de Montellano, Biochemical and structural characterization of CYP124: a methyl-branched lipid omega-hydroxylase from *Mycobacterium tuberculosis*, *Proc. Natl. Acad. Sci. U. S. A.*, 106 (2009) 20687-20692.
- [4] A.V. Vasilevskaya, A.V. Yantsevich, G.V. Sergeev, A.P. Lemish, S.A. Usanov, A.A. Gilep, Identification of *Mycobacterium tuberculosis* enzyme involved in vitamin D and 7-dehydrocholesterol metabolism, *J. Steroid Biochem. Mol. Biol.*, 169 (2017) 202-209.