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

Stella A. Child,[1] Elise F. Naumann,[1] John B. Bruning[2] and Stephen G. Bell[1]*

[1]Department of Chemistry, School of Physical Sciences, University of Adelaide, SA 5005, Australia

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

* To whom correspondence should be addressed.

Stephen G. Bell (stephen.bell@adelaide.edu.au) Tel: +61 8 83134822

Abstract

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 *tran*s-pseudoionone at a terminal methyl group to yield (*2E,6E*)-8-hydroxy-3,7-dimethylocta-2,6-dien-1-yl acetate and (*3E,5E,9E*)-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.

Abbreviations

ArR, a ferredoxin reductase from *Novosphingobium aromaticivorans*; BSTFA-TMSCl, N,O-bis(trimethylsilyl) trifluoroacetamide and trimethylsilylchloride; CYP Cytochrome P450 enzyme; DTT, dithiolthreitol; EMM, *E. coli* minimal media; FAD, flavin adenine dinucleotide; FFFAS, Fold and Function Assignment System; IPTG, Isopropyl β-D-1-thiogalactopyranoside; LB, Lysogeny broth (also known as Luria or Lennox Broth); MAC, *Mycobacterium avium* complex; MRSAD, Molecular Replacement Single Wavelength Anomalous Diffraction; MTBC, *Mycobacterium tuberculosis* complex; NAD(P)H, reduced nicotinamide adenine dinucleotide (phosphate); SOC, Super Optimal broth with Catabolite repression; Tdx, terpredoxin (a ferredoxin from *Pseudomonas sp*); PCR, polymerase chain reaction; PDIM, phthiocerol dimycocerosates; PE-PGRS, proline glutamate (PE) polymorphic GC-rich sequence proteins.

Introduction

Cytochrome P450 (CYP) enzymes are a family of heme 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 [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 heme 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-sulfur ferredoxin, together known as a Class 1 electron transfer system [8].

Upon sequencing of the *Mycobacterium tuberculosis* genome [9], 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 [10]. Due 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 [11, 12]. Approximately one third of the global population has the latent form of tuberculosis according to WHO estimates, with over 10 million new cases of the active form each year [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 *Mycobacterium tuberculosis* complex (MTBC) has been taken as evidence that they perform general or housekeeping roles, while conservation only in *M. 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 biosynthetic 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, a number of 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 and IPTG were from Astral Scientific. Restriction enzymes used for cloning were purchased from New England Biolabs. KOD polymerase, used for the PCR 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; 2xYT 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), $MgCl_2$ (1 g) and 5 mL of 40 % glucose per litre; EMM (*E. coli* minimal media); K_2HPO_4 (7 g), KH_2PO_4 (3 g), Na_3citrate (0.5 g), $(NH_4)_2SO_4$ (1 g), $MgSO_4$ (0.1 g), 20 % glucose (20 mL) and glycerol (1 % v/v) in one litre; Trace elements; 0.74 g $CaCl_2{\cdot}H_2O$, 0.18 g $ZnSO_4{\cdot}7H_2O$, 0.132 g $MnSO_4{\cdot}4H_2O$, 20.1 g Na_2EDTA, 16.7 g $FeCl_3{\cdot}6H_2O$, 0.10 g $CuSO_4{\cdot}5H_2O$, 0.25 g $CoCl_2{\cdot}6H_2O$. Antibiotics were added to the following working concentrations; ampicillin, 100 μg mL^{-1} and kanamycin, 30 μg 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 GC-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). The injector was held at 250 °C and the interface at 280 °C. Column flow was set at 1.5 mL/min 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 per min 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 per min 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 x LC-20AR pumps, SIL-20AC HT autosampler, SPD-M20A photodiode array detector and a CT0-20AC column oven. Separation was performed using an Ascentis Si HPLC column (25 cm × 10 mm x 5 μm; Sigma-Aldrich). NMR was performed using an Agilent DD2 spectrometer at 500 MHz for ^{1}H and 126 MHz for ^{13}C.

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 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 ºC for 45 s followed by annealing at 55 ºC for 30 s and extension at 68 ºC 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 *Escherichia 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 6 flasks) at 37°C for 5 hours 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 mins, IPTG (0.1 mM) were added to induce protein expression [29]. The growths were continued for a further 16 hours before harvesting of the cell pellet by centrifugation (5000 g, 15 mins). 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 secs 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 mins) and then loaded onto a DEAE Sepharose column, (XK50, 200 mm x 40 mm, GE Healthcare) and eluted with a linear gradient of 100 mM to 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 x 40 mm) with elution using Buffer T. The protein sample was then loaded onto a further Source-Q anion exchange column (XK26, 80 mm x 30 mm, GE Healthcare) and eluted with a gradient of 0 to 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 $\varepsilon_{450} = 91\ mM^{-1}\ cm^{-1}$ for the reduced CO bound form [30]. The CYP268A2 concentration was determined using $\varepsilon_{419} = 108 \pm 7\ mM^{-1}\ cm^{-1}$.

Spin state shifts and substrate binding titrations

To determine the extent of substrate binding, CYP268A2 was diluted to ~1 µM 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 nm and 250 nm was recorded on the UV spectrophotometer until no further spectral change was observed. The high spin percentage was estimated (± 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 nm 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 µM, less than five times the concentration of the enzyme), the data were instead fitted to the tight-binding quadratic equation:

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

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 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 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 hours 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 mins 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 hours the turnover was then centrifuged (15 min, 5000 g) and the supernatant 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 $MgSO_4$ 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 $\varepsilon_{340} = 6.22$ µM. 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 x 300 mm) before being concentrated to 30-35 mg mL^{-1} in Buffer T. The substrate pseudoionone (100 mM EtOH, mix 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 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 x 140 x 20 µM, Figure S4) were obtained after 2 weeks at 16 °C from the condition containing 0.96 M ammonium phosphate, 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 N_2. Data was 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]. Initial molecular replacement phasing was attempted using the modified search model (residues 1-23, 145-152, 322-337, 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 protocol of Auto-Rickshaw [37, 38] using the native dataset. Within the pipeline various programs 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 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 *Mycobacterium 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 *Mycobacterium colombiense* [50]. *M. colombiense* is a member of the *Mycobacterium 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 (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 (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, 7 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 (Figure S1).

The area surrounding putative *cyp268* genes is marked by the presence of a highly conserved operon (Figure S2), containing a GTPase, a large ribonuclease, two regulatory proteins (AcrR and Sir2-like) and a downstream NAD synthetase. *Cyp268A2* is flanked on both sides by a PE-PGRS 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 (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 nm to 450 nm, with only a small shoulder (< 5 %) at 420 nm (Figure 2). The extinction coefficient for the Soret absorbance [58], ε_{450} = 91 mM cm^{-1}, was used to determine the extinction coefficient for CYP268A2 ε_{419} = 108 ± 7 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). A number 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 (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 a range of aromatic compounds could bind 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 ± 0.2 µM. Undecanoic acid gave K_d, 1.1 ± 0.5 µM. Pseudoionone, a linear ionone precursor, gave 80 % HS and K_d, 3.6 ± 0.6 µM (as a mixture of isomers). 4-Phenyltoluene also bound tightly, K_d, 13 ± 2.4 µM. Notably the longer substrates preferred by CYP124A1 such as phytanic acid failed to induce a significant spin state shift in CYP268A2 (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-methyllauric acid (30 % HS) and 10-methyllauric acid (20 % HS), although 15-methylhexadecanoic acid (15 %) was more effective than palmitic acid (0 %). The binding data suggested 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 ± 1.9 µM), over the *cis* form, neryl acetate (30 % HS, K_d, 106 ± 29 µM, 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 demonstrates 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.

Many azole drugs have been reported to bind CYPs as competitive inhibitors, coordinating directly to the Fe atom of the heme 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]. A number of possible azole 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-Phenylimidzole bound more tightly (K_d, 0.9 ± 0.3 µM) than 4-phenylimidazole (K_d, 4.5 ± 0.6 µM). The addition of 2-phenylimidazole and other larger azole 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 (Table S3). The pair terpredoxin (2Fe-2S ferredoxin, from a *Pseudomonas* sp.) and an FAD-containing ferredoxin reductase ArR (from *Novosphingobium aromaticivorans*) were selected (based on levels of product formation with geranyl acetate, supplementary material) to enable

characterisation of the products of the enzyme. I*n 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 ~60 nmol.nmolP450^{-1} min^{-1} and when extracted and analysed by GC-MS, showed a major product (9.5 mins, 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 hours. 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 mins 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 the major product *in vivo* was also hydrogenated in addition to the CYP-mediated hydroxylation (Supplementary material, 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 material Figure S15). The *in vitro* turnover experiment with geranyl acetate confirms the hydroxylation is CYP-mediated while it is likely 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 a generated a single major product at 16.1 mins with a mass peak at 208.1 (expected hydroxylation product at 208.15). Addition of pseudoionone to the *in vivo* CYP/Tdx/ArR system generated a single major product after 24 hours with a retention time of 15.9 mins (Figure 5b). Mass peaks of the *in vivo* product (210.2 and 195.20 and others) were again two mass units higher than the *in vitro* product (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 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 to expected molecular ion peak at 192.15) which are the *cis* (3*E*, 5*Z)* and *trans* (3*E*, 5*E)* isomer, 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 (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 enzyme prefers the straight chain isomer over the bent form. The absence of product for geraniol suggests 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 with 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 two weeks and diffraction data was 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*, *5E*) 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 unstructured 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 the pseudoionone is arranged with one of the ω methyl groups held directly over the heme (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 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 (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 has shown 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 heme, where 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 heme 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 (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 comprises the solvent channel (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 (2WM4) [21]. This structure shows the substrate bound in a similar manner to that of pseudoionone close to the heme, 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 carbonyl (Ala260 in CYP268A2, Ala267 in CYP124A1), with the bond distance in CYP268A2 2.7 Å compared to 3.7 Å in CYP124A1 (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 insight 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.

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 activity can be efficiently reconstituted with non-native electron transfer proteins including a [2Fe-2S] ferredoxin. This 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. Further, 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 heme 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.

Acknowledgements

The authors wish to 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*. This work was supported by ARC through a Future Fellowship (FT140100355 to SGB). The authors also acknowledge the award of University of Adelaide Faculty of Sciences Divisional Scholarship (PhD to SAC). X-ray diffraction data collection was undertaken on the MX1 beamline at the Australian Synchrotron, part of ANSTO. We acknowledge financial support from the Australian Synchrotron in the form of MXCAP9674.

Authorship statement

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

References

1 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

2 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

3 Nelson, D. R., Zeldin, D. C., Hoffman, S. M., 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

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

5 Guengerich, F. P. (2001) Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. Chemical research in toxicology. **14**, 611-650

6 Khatri, Y., Girhard, M., Romankiewicz, A., Ringle, M., Hannemann, F., Urlacher, V. B., Hutter, M. C. and Bernhardt, R. (2010) Regioselective hydroxylation of norisoprenoids by CYP109D1 from Sorangium cellulosum So ce56. Applied microbiology and biotechnology. **88**, 485-495

7 Scheps, D., Malca, S. H., Hoffmann, H., Nestl, B. M. and Hauer, B. (2011) Regioselective omega-hydroxylation of medium-chain n-alkanes and primary alcohols by CYP153 enzymes from Mycobacterium marinum and Polaromonas sp. strain JS666. Organic & biomolecular chemistry. **9**, 6727-6733

8 Hannemann, F., Bichet, A., Ewen, K. M. and Bernhardt, R. (2007) Cytochrome P450 systems--biological variations of electron transport chains. Biochimica et biophysica acta. **1770**, 330-344

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

10 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. Biochemical Society transactions. **40**, 573-579

11 Young, D. B. and Cole, S. T. (1993) Leprosy, tuberculosis, and the new genetics. Journal of bacteriology. **175**, 1-6

12 Bloom, B. R. and Murray, C. J. (1992) Tuberculosis: commentary on a reemergent killer. Science. **257**, 1055-1064

13 (2017) World Health Organisation: Global Tuberculosis Report. ed.)^eds.), http://www.who.int/tb/publications/global_report/en/

14 Stinear, T. P., Seemann, T., Harrison, P. F., Jenkin, G. A., Davies, J. K., Johnson, P. D., Abdellah, Z., Arrowsmith, C., Chillingworth, T., Churcher, C., Clarke, K., Cronin, A., Davis, P., Goodhead, I., Holroyd, N., Jagels, K., Lord, A., Moule, S., Mungall, K., Norbertczak, H., Quail, M. A., Rabbinowitsch, E., Walker, D., White, B., Whitehead, S., Small, P. L.,

Brosch, R., Ramakrishnan, L., Fischbach, M. A., Parkhill, J. and Cole, S. T. (2008) Insights from the complete genome sequence of Mycobacterium marinum on the evolution of Mycobacterium tuberculosis. Genome research. **18**, 729-741

15 Demangel, C., Stinear, T. P. and Cole, S. T. (2009) Buruli ulcer: reductive evolution enhances pathogenicity of Mycobacterium ulcerans. Nature reviews. Microbiology. **7**, 50-60

16 Weerdenburg, E. M., Abdallah, A. M., Rangkuti, F., Abd El Ghany, M., Otto, T. D., Adroub, S. A., Molenaar, D., Ummels, R., Ter Veen, K., van Stempvoort, G., van der Sar, A. M., Ali, S., Langridge, G. C., Thomson, N. R., Pain, A. and Bitter, W. (2015) Genome-wide transposon mutagenesis indicates that Mycobacterium marinum customizes its virulence mechanisms for survival and replication in different hosts. Infection and immunity. **83**, 1778-1788

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

18 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. The Journal of biological chemistry. **291**, 7325-7333

19 Ouellet, H., Chow, E. D., Guan, S., Cox, J. S., Burlingame, A. L. and de Montellano, P. R. (2013) Genetic and mass spectrometric tools for elucidating the physiological function(s) of cytochrome P450 enzymes from Mycobacterium tuberculosis. Methods in molecular biology (Clifton, N.J.). **987**, 79-94

20 Barry, C. E., 3rd, Lee, R. E., Mdluli, K., Sampson, A. E., Schroeder, B. G., Slayden, R. A. and Yuan, Y. (1998) Mycolic acids: structure, biosynthesis and physiological functions. Progress in lipid research. **37**, 143-179

21 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 omega-hydroxylase from Mycobacterium tuberculosis. Proceedings of the National Academy of Sciences of the United States of America. **106**, 20687-20692

22 McLean, K. J. and Munro, A. W. (2008) Structural biology and biochemistry of cytochrome P450 systems in Mycobacterium tuberculosis. Drug metabolism reviews. **40**, 427-446

23 Forrellad, M. A., Klepp, L. I., Gioffre, A., Sabio y Garcia, J., Morbidoni, H. R., de la Paz Santangelo, M., Cataldi, A. A. and Bigi, F. (2013) Virulence factors of the Mycobacterium tuberculosis complex. Virulence. **4**, 3-66

24 Jackson, C. J., Lamb, D. C., Marczylo, T. H., Parker, J. E., Manning, N. L., Kelly, D. E. and Kelly, S. L. (2003) Conservation and cloning of CYP51: a sterol 14 alpha-demethylase from Mycobacterium smegmatis. Biochemical and biophysical research communications. **301**, 558-563

25 Ouellet, H., Johnston, J. B. and Ortiz de Montellano, P. R. (2010) The Mycobacterium tuberculosis cytochrome P450 system. Archives of biochemistry and biophysics. **493**, 82-95

26 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. Journal of bacteriology. **181**, 3419-3426

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. Research in microbiology. **155**, 1-9
28 Brezna, B., Kweon, O., Stingley, R. L., Freeman, J. P., Khan, A. A., Polek, B., Jones, R. C. and Cerniglia, C. E. (2006) Molecular characterization of cytochrome P450 genes in the polycyclic aromatic hydrocarbon degrading Mycobacterium vanbaalenii PYR-1. Applied microbiology and biotechnology. **71**, 522-532
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
30 Omura, T. and Sato, R. (1964) The Carbon Monoxide-Binding Pigment of Liver Microsomes. I. Evidence for Its Hemoprotein Nature. The Journal of biological chemistry. **239**, 2370-2378
31 Dogne, J. M., de Leval, X., Benoit, P., Rolin, S., Pirotte, B. and Masereel, B. (2002) Therapeutic potential of thromboxane inhibitors in asthma. Expert opinion on investigational drugs. **11**, 275-281
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. Applied microbiology and biotechnology. **86**, 163-175
33 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 crystallographica. Section D, Biological crystallography. **68**, 277-291
34 McPhillips, T. M., McPhillips, S. E., Chiu, H. J., Cohen, A. E., Deacon, A. M., Ellis, P. J., Garman, E., Gonzalez, A., Sauter, N. K., Phizackerley, R. P., Soltis, S. M. and Kuhn, P. (2002) Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. Journal of synchrotron radiation. **9**, 401-406
35 Kabsch, W. (2010) Xds. Acta crystallographica. Section D, Biological crystallography. **66**, 125-132
36 Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement. Acta crystallographica. Section D, Biological crystallography. **66**, 133-144
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 crystallographica. Section D, Biological crystallography. **61**, 449-457
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 crystallographica. Section D, Biological crystallography. **65**, 1089-1097
39 Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. Acta crystallographica. Section D, Biological crystallography. **67**, 235-242
40 Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L.

M., Simonson, T. and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D. **54**, 905-921
41 Brunger, A. T. (2007) Version 1.2 of the Crystallography and NMR system. Nat Protoc. **2**, 2728-2733
42 Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D. **67**, 355-367
43 Cowtan, K. (2000) General quadratic functions in real and reciprocal space and their application to likelihood phasing. Acta crystallographica. Section D, Biological crystallography. **56**, 1612-1621
44 Sheldrick, G. M. (2002) Macromolecular phasing with SHELXE. Z Kristallogr. **217**, 644-650
45 Terwilliger, T. C. (2000) Maximum-likelihood density modification. Acta crystallographica. Section D, Biological crystallography. **56**, 965-972
46 Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. Acta crystallographica. Section D, Biological crystallography. **62**, 1002-1011
47 Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica. Section D, Biological crystallography. **66**, 213-221
48 Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (2010) Features and development of Coot. Acta crystallographica. Section D, Biological crystallography. **66**, 486-501
49 Yoshida, M., Nakanaga, K., Ogura, Y., Toyoda, A., Ooka, T., Kazumi, Y., Mitarai, S., Ishii, N., Hayashi, T. and Hoshino, Y. (2016) Complete Genome Sequence of Mycobacterium ulcerans subsp. shinshuense. Genome Announcements. **4**
50 Riojas, M. A. M., K.; Hazbon, M. H. NGS-Based Phylogenomic Analysis Supports Reclassification of All Species Within the *Mycobacterium tuberculosis* Complex as *Mycobacterium tuberculosis*. ed.)^eds.), Bei Resources, https://www.beiresources.org/Portals/2/PDFS/NGS-Based%20Phylogenomic%20Analysis%20of%20MTB%20Complex.pdf.
51 Maya-Hoyos, M., Leguizamon, J., Marino-Ramirez, L. and Soto, C. Y. (2015) Sliding Motility, Biofilm Formation, and Glycopeptidolipid Production in Mycobacterium colombiense Strains. BioMed research international. **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. The Journal of steroid biochemistry and molecular biology. **169**, 202-209
53 Frank, D. J., Madrona, Y. and Ortiz de Montellano, P. R. (2014) Cholesterol ester oxidation by mycobacterial cytochrome P450. The Journal of biological chemistry. **289**, 30417-30425
54 Capyk, J. K., Kalscheuer, R., Stewart, G. R., Liu, J., Kwon, H., Zhao, R., Okamoto, S., Jacobs, W. R., Jr., Eltis, L. D. and Mohn, W. W. (2009) Mycobacterial cytochrome p450 125 (cyp125) catalyzes the terminal hydroxylation of c27 steroids. The Journal of biological chemistry. **284**, 35534-35542
55 Poulos, T. L., Finzel, B. C., Gunsalus, I. C., Wagner, G. C. and Kraut, J. (1985) The 2.6-A crystal structure of Pseudomonas putida cytochrome P-450. The Journal of biological chemistry. **260**, 16122-16130

56 Banu, S., Honore, N., Saint-Joanis, B., Philpott, D., Prevost, M. C. and Cole, S. T. (2002) Are the PE-PGRS proteins of Mycobacterium tuberculosis variable surface antigens? Molecular microbiology. **44**, 9-19
57 Espitia, C., Laclette, J. P., Mondragon-Palomino, M., Amador, A., Campuzano, J., Martens, A., Singh, M., Cicero, R., Zhang, Y. and Moreno, C. (1999) The PE-PGRS glycine-rich proteins of Mycobacterium tuberculosis: a new family of fibronectin-binding proteins? Microbiology (Reading, England). **145 (Pt 12)**, 3487-3495
58 Omura, T. and Sato, R. (1964) The Carbon Monoxide-Binding Pigment of Liver Microsomes. Ii. Solubilization, Purification, and Properties. The Journal of biological chemistry. **239**, 2379-2385
59 Driscoll, M. D., McLean, K. J., Levy, C., Mast, N., Pikuleva, I. A., Lafite, P., Rigby, S. E., Leys, D. and Munro, A. W. (2010) Structural and biochemical characterization of Mycobacterium tuberculosis CYP142: evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen. The Journal of biological chemistry. **285**, 38270-38282
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
61 Ouellet, H., Podust, L. M. and de Montellano, P. R. (2008) Mycobacterium tuberculosis CYP130: crystal structure, biophysical characterization, and interactions with antifungal azole drugs. The Journal of biological chemistry. **283**, 5069-5080
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. Archives of biochemistry and biophysics. **464**, 228-240
63 Mueller, N. J., Stueckler, C., Hauer, B., Baudendistel, N., Housden, H., Bruce, N. C. and Faber, K. (2010) The Substrate Spectra of Pentaerythritol Tetranitrate Reductase, Morphinone Reductase, N-Ethylmaleimide Reductase and Estrogen-Binding Protein in the Asymmetric Bioreduction of Activated Alkenes. Advanced Synthesis & Catalysis. **352**, 387-394
64 Grau, M. M., van der Toorn, J. C., Otten, L. G., Macheroux, P., Taglieber, A., Zilly, F. E., Arends, I. W. C. E. and Hollmann, F. (2009) Photoenzymatic Reduction of C□C Double Bonds. Advanced Synthesis & Catalysis. **351**, 3279-3286
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
66 Yang, W., Bell, S. G., Wang, H., Zhou, W., Hoskins, N., Dale, A., Bartlam, M., Wong, L. L. and Rao, Z. (2010) Molecular characterization of a class I P450 electron transfer system from Novosphingobium aromaticivorans DSM12444. The Journal of biological chemistry. **285**, 27372-27384
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
68 Cojocaru, V., Winn, P. J. and Wade, R. C. (2007) The ins and outs of cytochrome P450s. Biochimica et biophysica acta. **1770**, 390-401
69 Poulos, T. L., Finzel, B. C. and Howard, A. J. (1987) High-resolution crystal structure of cytochrome P450cam. Journal of molecular biology. **195**, 687-700
70 Poulos, T. L. (2014) Heme Enzyme Structure and Function. Chemical reviews. **114**, 3919-3962

Tables and Figures

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 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

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

Data collection statistics[a]	
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)[b]
Number of unique reflections	26452 (1743)
Completeness	99.2 (89.4)
Redundancy	7.4 (7.0)
(I)/[σ(I)]	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%

[a] Data collected from one crystal

[b] Values in parenthesis are for highest resolution shell.

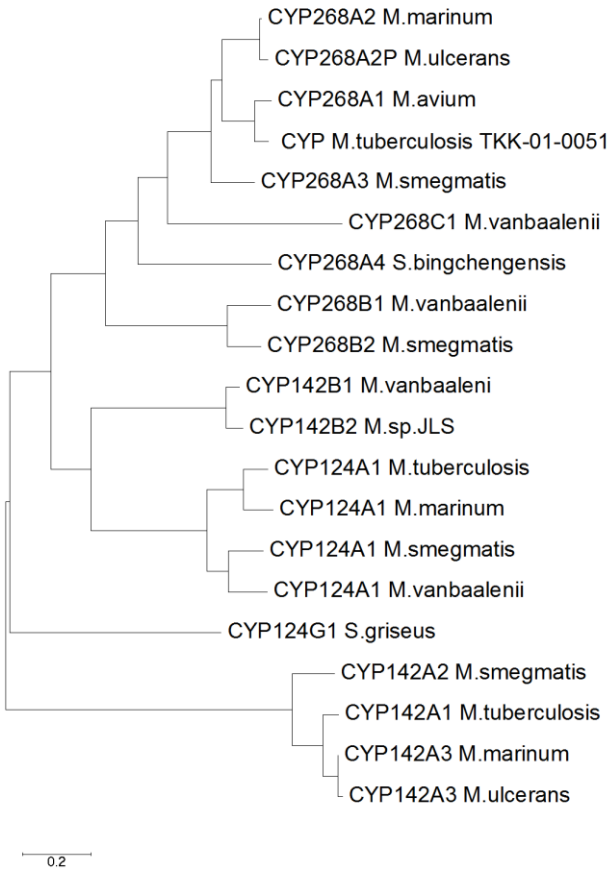

Figure 1: Phylogenetic tree of the CYP268 family. Included are some members of the CYP124 and CYP142 families, both of which have been predominately found in *Mycobacterium* species. The grouping shows that CYP268A2 is closely related to members of the 268 family, with significant sequence similarity 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 Table S1. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. Scale measures number of substitutions per site.

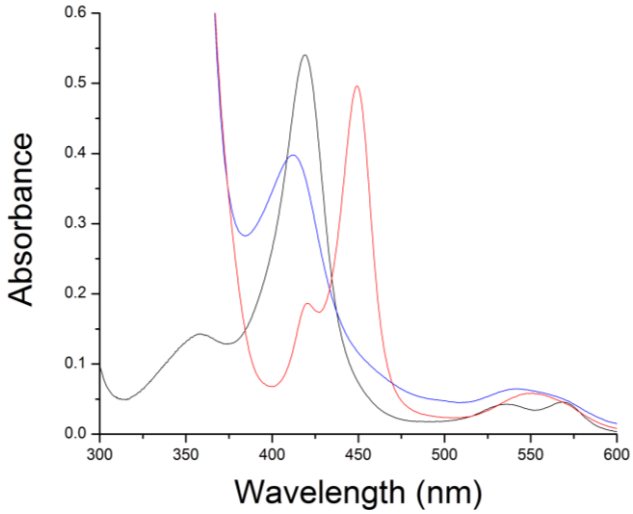

Figure 2: CYP268A2 (black, A_{419}) in the reduced form (blue, A_{412}), bound to CO (red, A_{449}), showing the characteristic absorbance at ~450 nm. The shoulder at 420 nm in the ferrous-CO form comprises ~ 5% of the total area.

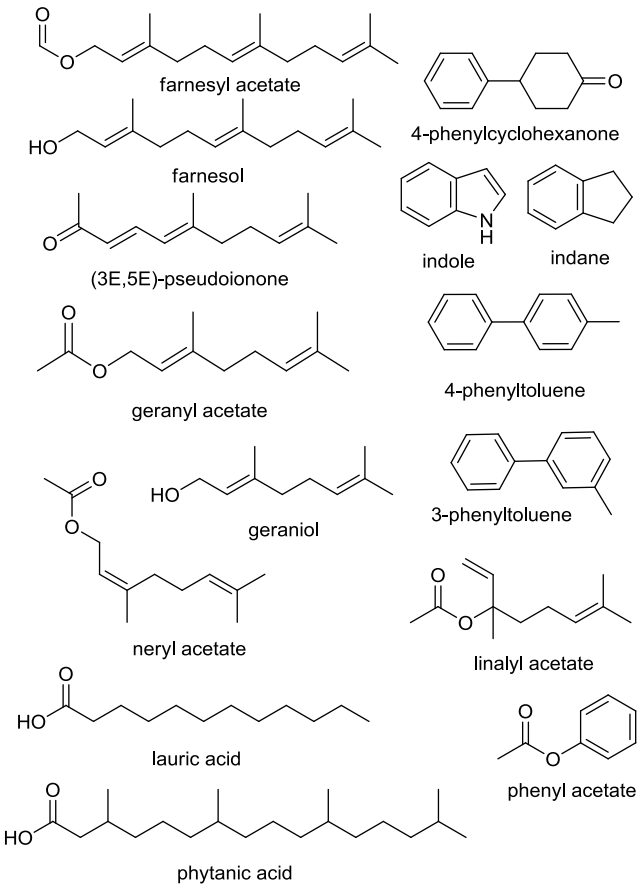

Figure 3: Structures of selected compounds tested as substrates for CYP268A2

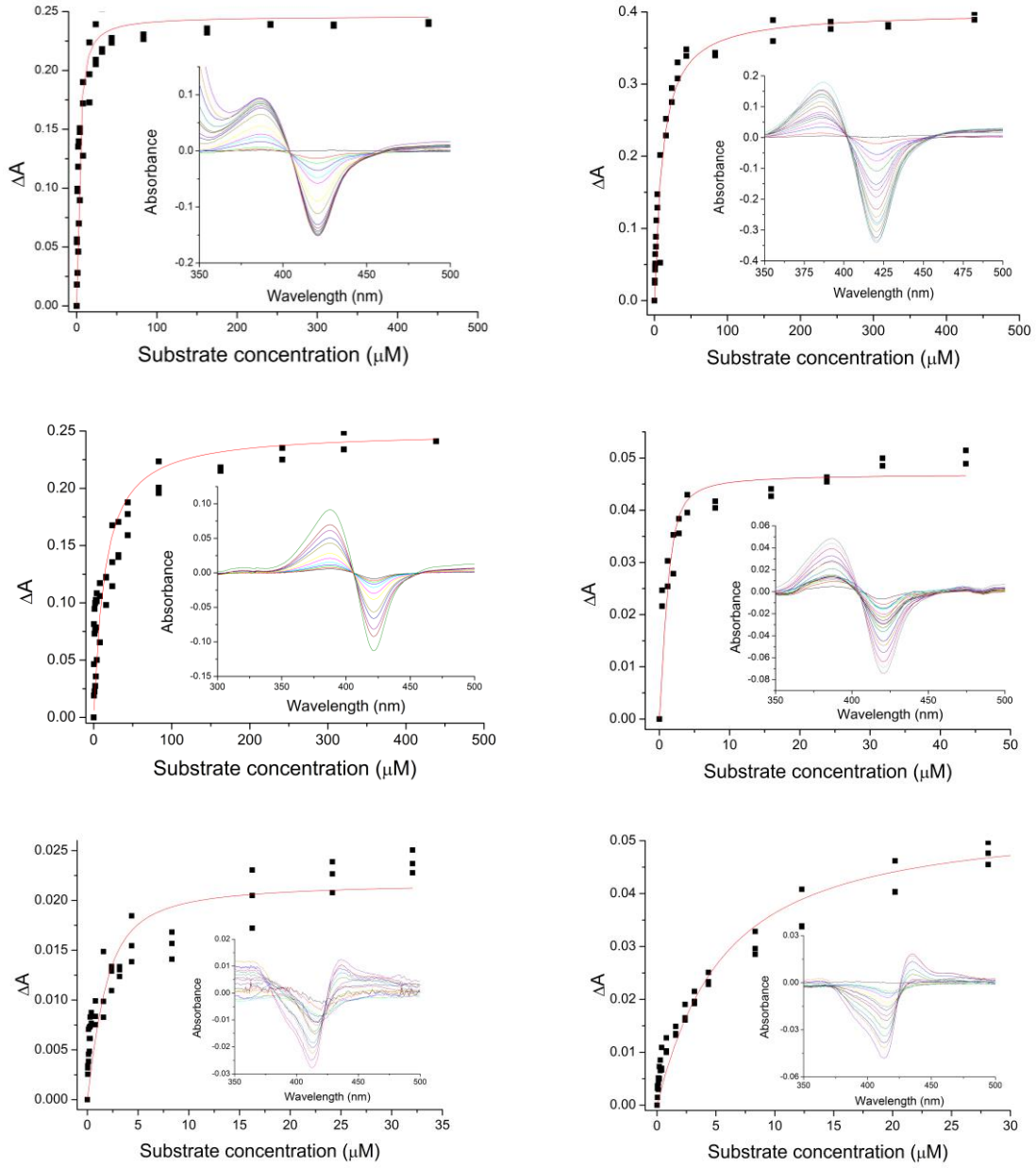

Figure 4: CYP268A2 binding constant analysis with the substrates A) pseudoionone B) geranyl acetate C) 4-phenyltoluene D) undecanoic acid E) 1-phenylimidazole (peak to trough 413 to 435 nm) F) 4-phenylimidazole (414 to 434 nm)

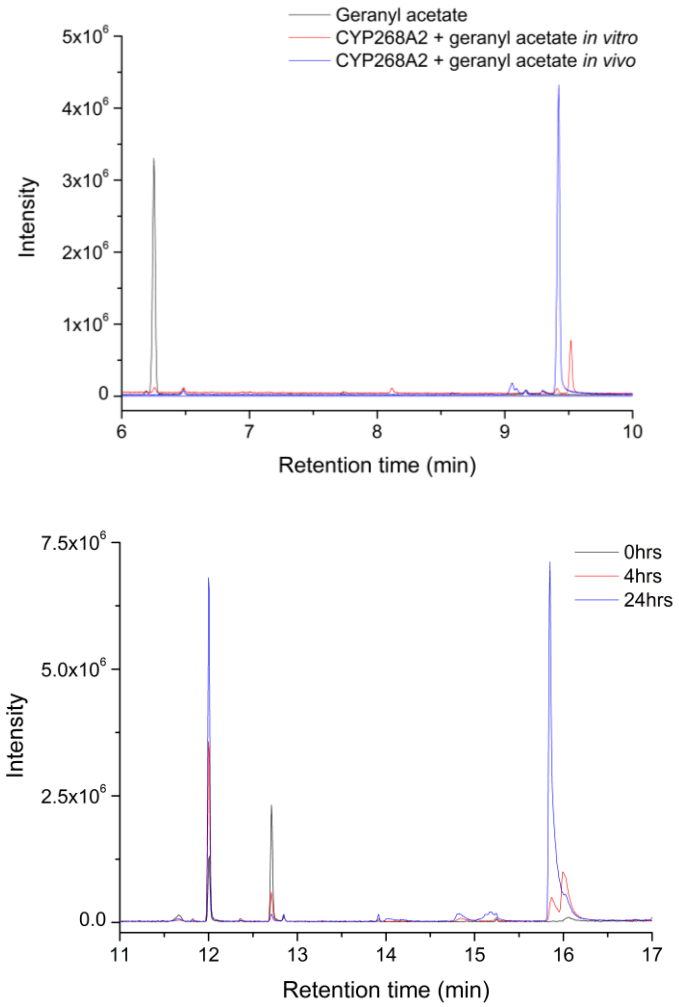

Figure 5: (a) GC trace of the *in vivo* and *in vitro* turnovers of CYP268A2 with geranyl acetate (b) *in vivo* turnover with pseudoionone monitored over 24 hours. Identification of isomer peaks was done by retention time comparison with neryl acetate (*cis*) and geranyl acetate (*trans*), where the *cis* form had the shorter retention time.

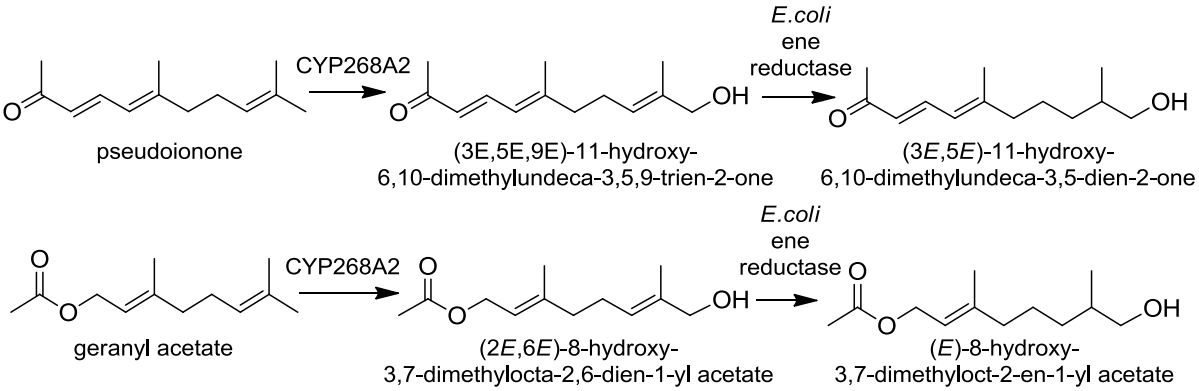

Figure 6: Hydroxylation products of CYP268A2. *In each, the hydroxylation at the terminal end would have generated a stereo-centre at the ω-1 carbon. However *in vivo* the hydrogenation occurred after the hydroxylation, and as a result the stereo-selectivity of the isolated product would be dependent on the *E.coli* ene reductase enzyme.

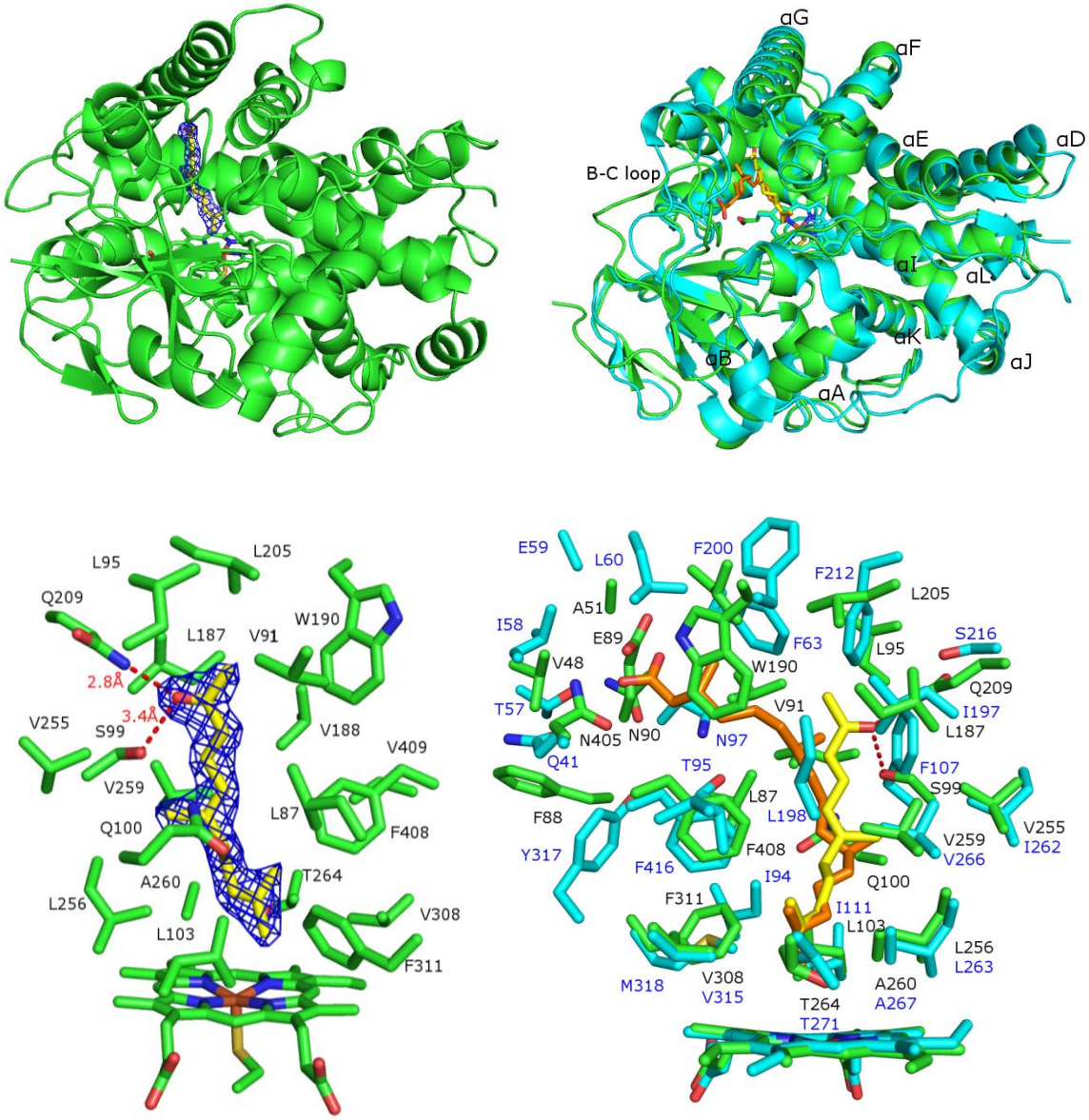

Figure 7: (a) *M. marinum* CYP268A2 (green) with pseudoionone bound (yellow), showing the composite omit $2F_o - F_c$ electron density at $\sigma = 1.5$ (blue mesh) of the pseudoionone in the active site (b) CYP268A2 (green) overlaid with *M. tb* CYP124A1 (blue) with phytanic acid (orange) bound, showing the conserved P450 helices[55] (c) CYP268A2 active site region showing side chains of amino acids within 5Å 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) the active site region showing side chains of amino acids from both enzymes within 5Å of the substrate molecules (CYP268A2 labels in black, CYP124A1 blue).

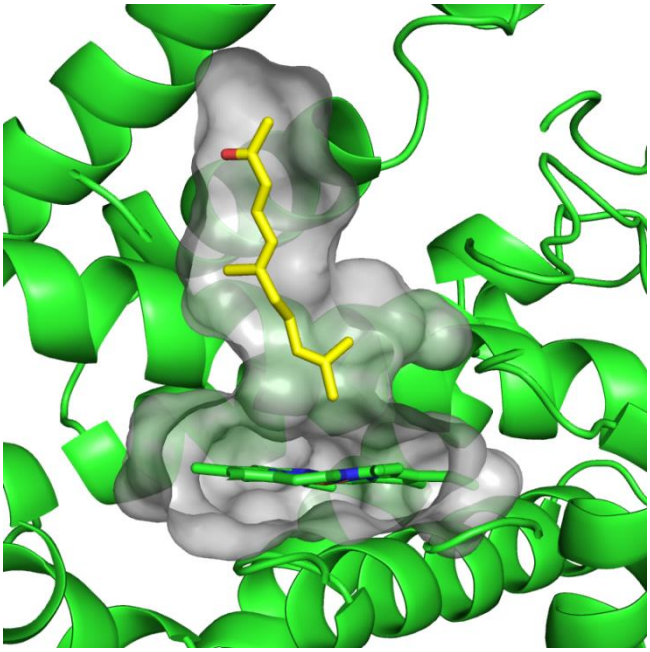

Figure 8: CYP268A2 (green) with the inner surface cavity of the peptide enclosing the pseudoionone (yellow). There is a small empty cavity near the heme.