Isolation of new P450s and the modification of existing P450s for biocatalysis

Ian Lau

Supervisors:
Assoc. Prof. Stephen G. Bell
Prof. Hugh H. Harris

Thesis submitted for the degree of Master of Philosophy

May 2017

School of Physical Sciences

Contents

Abstract

Cytochrome P450s are a family of heme-containing monooxygenases that are ubiquitous in nature. Many P450s from bacterial sources, such as from *Frankia* sp. EuI1c which contain genes encoding 68 of these enzymes, have not previously been investigated. These P450s are potentially involved in the metabolism and biosynthesis of novel natural products including steroids, siderophores, fatty acids and antibiotics. Here four of these P450s were successfully expressed and purified. One of these, FraEu2494, was purified at high concentrations suitable for crystallisation. A selection of chemical compounds was screened with these enzymes to determine the substrate range of these P450s. In particular the P450 FraEu1415 exhibited high affinity towards steroid compounds such as testosterone, estrone and progesterone, highlighting the important compounds these P450s may metabolise.

The ferredoxin electron transfer partners of these P450s were also successfully expressed and purified. While the ferredoxin reductases were unable to be produced, a mutant library of the ferredoxin Fdx2495 was created. Non-standard amino acid residues within the iron-sulfur binding motif of the ferredoxins were investigated to ascertain if Fdx2495 could be used as a model for future study.

The cytochrome P450 CYP199A4 from *Rhodopseudomonas palustris* strain HaA2 is highly specific for the regioselective oxidation of *para*-substituted benzoic acids such as 4-methoxybenzoic acid. It has been reported that the activity of the CYP199A4 S244D mutant for the hydroxylation and demethylation of *para*-substituted non-benzoic acid derivatives is greater than with the wild-type enzyme. Here we report the potential scale up of these oxidation reactions by the S244D mutant with a system that contains an excess of NADH. A selection of similar *para*-substituted compounds, including styrenes, methylthio- and dimethyl-substituted benzene derivatives were tested with the enzyme to further investigate the mechanism and productivity of the mutant.

The sulfoxidation and epoxidation reactions by the S244D mutant of *para*-substituted benzene derivatives were investigated. The epoxidation reactions produced small amounts of aldehyde arising from a 1,2-rearrangement reaction, giving evidence of a non-concerted reaction pathway. Chiral analysis for the sulfoxidation and epoxidation reactions revealed a consistent bias for a single enantiomer, suggesting similar binding conformations for these *para*-substituted benzene substrates within the active site.

The crystal structure of 4-methoxybenzoic acid bound to the CYP199A4 S244D mutant (PDB: 5U5J) was solved. This revealed small differences between the mutant and the equivalent wild-type structure (PDB: 4DO1). Aside from the specific 244 amino acid mutation, the substrate binding site was largely unaffected by the S244D variant. A

small shift in the position of the substrate over the active site was observed. A large shift was discovered for the chloride ion which caps the active site from external solvent in the wild-type enzyme. The chloride ion in the S244D mutant coordinates to an asparagine residue, which plays no role in the coordination to this ion in the equivalent wild-type.

Declaration

I certify that 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 and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, 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.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also 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 an Australian Government Research Training Program Scholarship.

_____

Ian Lau
January 2017

Acknowledgements

Firstly I would like to thank my supervisor Associate Professor Stephen Bell for his guidance, effort and assistance throughout my candidature and drafting of my thesis, without whom I'll probably would have written a vastly inferior piece of text.

I'll like to say a big thank you to all the members of the Bell group, both past and present, for your continued help and guidance for me, and also for a lot of fun times that we had when we were supposed to be working. I'll also like to thank Professor Hugh Harris, and special thanks to Tom Coleman and Dr. John Bruning for helping me with the crystallography sections, especially Tom, who grew the CYP199A4 S244D crystals.

I would like to thank all the people who helped me proofread my thesis, including Stella Child, Tom Coleman, James Lau and Samuel Lau.

I'll also like to thank my family and friends who supported me through these past two years, especially those who made/bought me food. Food is, as any research student should know, is the best thing in the world.

Finally I would like to thank the readers, especially the ones who are reviewing this, as we all know you would rather be doing something more fun. Happy reading!

Abbreviations

AcCN	acetonitrile
ADH	alcohol dehydrogenase
AMU	atomic mass units
BA	benzoic acid
CO	carbon monoxide
DCM	dichloromethane
DFT	density functional theory
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
ee	enantiomeric excess
EtOH	ethanol
EMM	*E. coli* minimal media
GC content	guanine-cytosine content
GC-MS	gas chromatography-mass spectrometry
HCl	hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	high performance liquid chromatography
IPTG	isopropyl β-D-thiogalactopyranoside
LB	Luria-Bertani medium
NADH	reduced form of nicotinamide adenine dinucleotide
NMR	Nuclear magnetic resonance
QM/MM	quantum mechanics/molecular mechanics
S244D	CYP199A4 serine-244-aspartate mutant
SOC	super optimal broth with catabolite repression
TFA	trifluoroacetic acid
WT	wild-type

List of Figures

List of Tables

This page is intentionally left blank

1 Introduction

1.1 Cytochrome P450s

Cytochrome P450s are a superfamily of heme-containing monooxygenases that typically catalyse the insertion of one atom of atmospheric dioxygen into C−H bonds in a wide range of organic compounds.[1] These reactions require the transfer of two electrons, that are commonly derived from nicotinamide adenine dinucleotide (phosphate) (NAD(P)H), which is facilitated by electron transfer proteins (Equation 1).

$$\text{R-H} + \text{O}_2 + \text{H}^+ + \text{NAD(P)H} \rightarrow \text{R-OH} + \text{H}_2\text{O} + \text{NAD(P)}^+ \tag{1}$$

Cytochrome P450s are ubiquitous in nature, and are found in all domains of life. For example, in humans there are 57 genes encoding P450s which are known to have extensive activity in metabolising toxic xenobiotics and drugs.[2] An interesting feature of the cytochrome P450 family is the ability to catalyse regio- and stereospecific oxidation of non-functionalised hydrocarbons, which is highly important in the biosynthesis of bioactive compounds, such as steroids.[3] Currently there are over 21,000 members within the P450 superfamily.[4] Cytochrome "P450" are named for its signature red pigment colour and the 450 nm Soret absorption peak from its carbon monoxide reduced form, which is due to a $\pi \rightarrow \pi^*$ transition in the heme-centred porphyrin ring (**9** in Figure 1).[5] The family a P450 belongs to is denoted by the number following the CYP prefix. P450s that originate from the same kingdom of life (e.g. Animalia, Plantea) and share $\geq 40\%$ amino acid sequence similarity to each other belong to the same family.[6] Subfamily members share $\geq 55\%$ amino acid similarity, and are denoted by the letter following the family number. As an example, a P450 presented here for investigation is CYP199A4 from the bacterium *Rhodopseudomonas palustris* strain HaA2, which is a member of the P450 family 199, and is the fourth member of the A subfamily.[7]

The general P450 catalytic cycle is described in Figure 1. The iron within the heme group coordinates to the four nitrogen groups of the protoporphyrin IX ring. The heme group is anchored by a thiolate ligand from the conserved cysteine forming the fifth coordination site to the iron centre. The ferric resting state (**1**) has a H_2O ligand bounded to the sixth distal site. This water molecule is displaced by the substrate R-H (**2**), which results in the iron-heme shifting from a low spin-state with an absorbance peak at $\approx$ 419 nm to a high spin-state with a peak at $\approx$ 390 nm. This species is then reduced by an electron from NAD(P)H to its ferrous form (**3**). Molecular oxygen can then bind to the now unoccupied distal site creating a ferrous-dioxygen species (**4**), which is then further reduced by a second electron to reach a stable ferric-peroxy

anion intermediate (**5**). Protonation forms the ferric-hydroperoxo intermediate (**6**), which is followed by further protonation and heterolytic cleavage of the O-O bond, which produces a water molecule and the highly reactive Fe(IV)-oxo radical cation intermediate (**7**). This intermediate is responsible for the insertion of a single oxygen atom into the substrate via a radical rebound mechanism (Figure 2). Here, the highly reactive species (Cpd I) can "activate" the hydrocarbon substrate by abstracting a hydrogen from the C-H bond, generating a hydroxy bound to the heme group that could undergo radical recombination with the carbon radical in the substrate (**8**). The oxygen containing product then leaves the active site and a water molecule replaces it by binding at the distal site, thus returning the P450 to its resting state (**1**). This radical rebound mechanism by which the enzymatic reaction is believed to proceed by was first proposed by Groves and McClusky in 1976.[8,9]

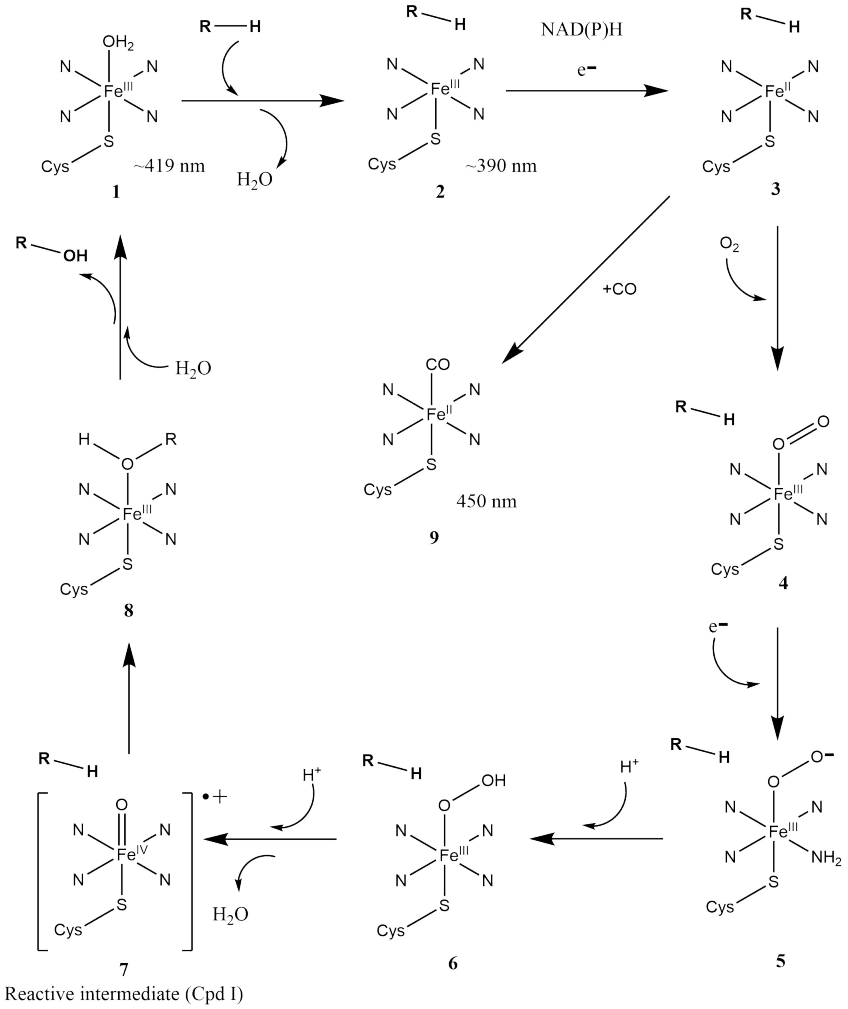


Figure 1: The general P450 catalytic cycle.

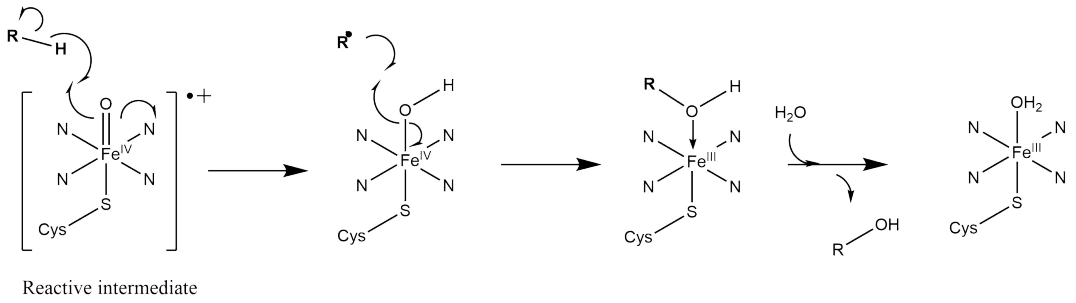


Figure 2: Radical rebound mechanism of the Fe(IV)-oxo porphyrin radical cation

P450s can selectively and efficiently oxidise non-activated C-H bonds, making them valuable for use within synthetic chemistry.[10,11] Because of the high efficiency and low environmental impact of cytochrome P450s, in that they produce little waste products, these enzymes are a highly attractive candidate for these oxidative reactions.

1.2 P450 Electron Transfer Systems

Most P450s require interactions with one or more redox partners in the electron transfer (ET) chain to obtain the electrons required for the catalytic cycle to progress. These partner proteins facilitate the transfer of two electrons from NAD(P)H (Figures 1 and 3). Bacterial and mammalian mitochondrial P450 Class I systems are comprised of a flavin adenine dinucleotide (FAD)-containing ferredoxin reductase, and iron-sulfur ferredoxin (most commonly 2Fe-2S) and the P450 itself.[12] Two electrons are passed from the NAD(P)H electron carrier to the FAD cofactor of the ferredoxin reductase, which then transfers one electron at a time to the iron-sulfur cluster ferredoxin. The electrons are then transferred to the P450 heme in two separate steps (Figure 1). In Class II systems which are found in eukaryotic systems, a single membrane-bound Cytochrome P450 Reductase (CPR) which has both FAD and flavin mononucleotide reductase (FMN) cofactors is responsible for the electron transfer. In terms of solubility, all three proteins in the class I systems of bacteria are soluble, while in class II eukaryotic systems both are membrane bound.[12–14] Due to the membrane bound nature of class II proteins, they are harder to produce and generally have lower solubility. Classes III and IV also feature in certain in bacterial systems and involve fusion of P450 and the ET partners (Figure 3).[15] These are not as common as the class I and II systems.

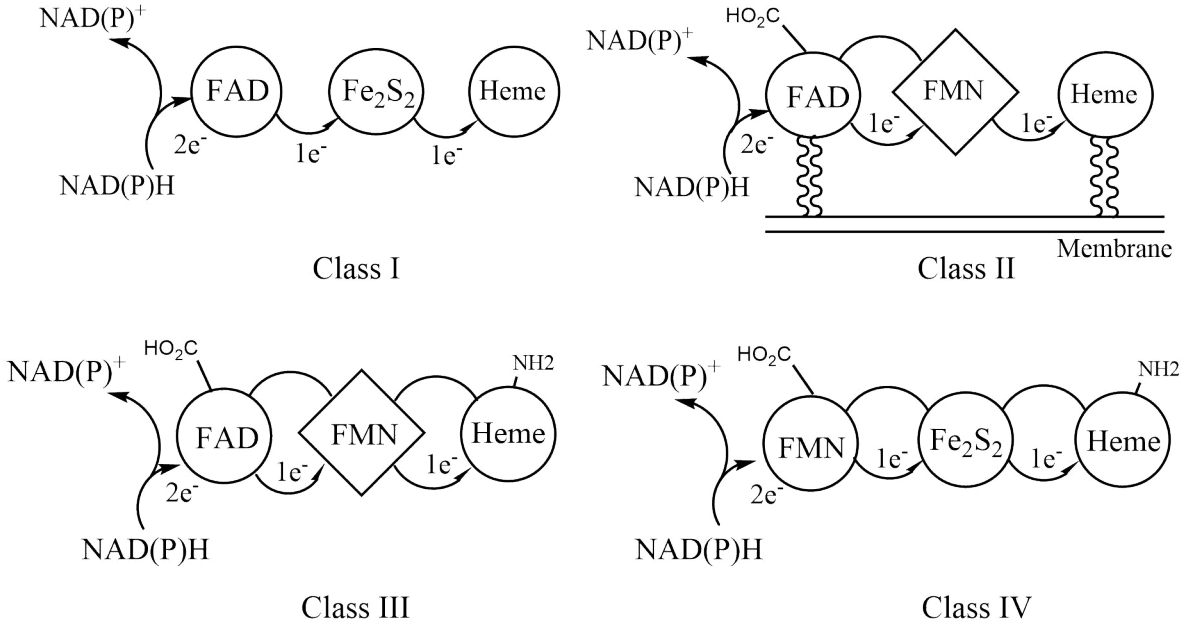


Figure 3: The different classes of electron transfer systems of P450 enzymes

While the iron-sulfur clusters of ferredoxins can vary, the majority of characterised bacterial ferredoxins are of the [2Fe-2S] cluster type.[1,16] Other types include [3Fe-4S] and [4Fe-4S] clusters, though these are less well understood.[12] The cluster is an essential

component for the ferredoxin to efficiently transfer reducing equivalents as they can delocalise the electron across the Fe and S atoms.[17] New insights into novel ferredoxins, such as characterisation of the 7Fe ferredoxin from *Streptomyces griseus* that contains both a [3Fe-4S] and a [4Fe-4S] cluster can further our understanding of these electron transfer partners.[18]

Bacterial P450s tend to have a high specificity towards their redox partners, i.e. the ferredoxin and ferredoxin reductase that transfers the electrons required for its catalytic cycle.[1] There are fewer ferredoxin and ferredoxin reductase encoding genes than those encoding P450 enzymes in most bacterial genomes. As a result these electron transfer partners are likely to support multiple P450 enzymes or other types of electron transport partners must exist. For example, *Streptomyces coelicolor* has 18 P450 encoding genes, and only six ferredoxins and four ferredoxin reductase encoding genes.[19] The bacterial strain *Frankia* sp. EuI1c has 68 P450 encoding genes, whereas genes encoding for ferredoxins are fewer. Using commercially available redox partners such as spinach ferredoxin/ferredoxin reductase will permit low level of bacterial P450 activity, but these approaches are expensive and not very effective.[20] Identification of the native electron transfer partners would allow for further insight on the mechanisms by which these enzymes interact and could result in increased activity. This is important as these electron transfer partners can then be used to facilitate the P450 catalytic cycle.

A proton relay is also required for oxygen activation in P450s (Figure 1). This is regulated by the highly conserved acid-alcohol pair of amino acids, commonly an aspartate and threonine/serine that sits in the I-helix above the heme active site (Asp251 and Thr252 for CYP199A4).[21] These residues have important roles in the transfer of protons to the peroxo and hydroperoxo intermediates (**5** and **6**) from the surrounding solvent.[22,23]

The rates of electron and proton transfer dictate the overall oxidation rate of substrate by the P450. In theory every molecule of NAD(P)H and oxygen that is consumed in a catalytic cycle can potentially produce an equivalent number of hydroxylated substrate. This only occurs when the substrate binds tightly within the active site of the P450 substrate. If the substrate does not fit well within the active site, this can often lead to unproductive uncoupling reactions. This can occur at different stages within the catalytic cycle (Figure 4).[22] The first possible uncoupling event can occur in the ferric-superperoxo species **4**, which can decompose to the ferric state by releasing superoxide. The second potential uncoupling event can occur from the ferric-hydroperoxide intermediate **6**. Hydrogen peroxide may be released if the second proton is transferred to the distal oxygen too slowly. This hydrogen peroxide uncoupling may be accelerated by excess water within the active site, which increases the competition of protonation of the proximal oxygen. The third and final uncoupling event involves an

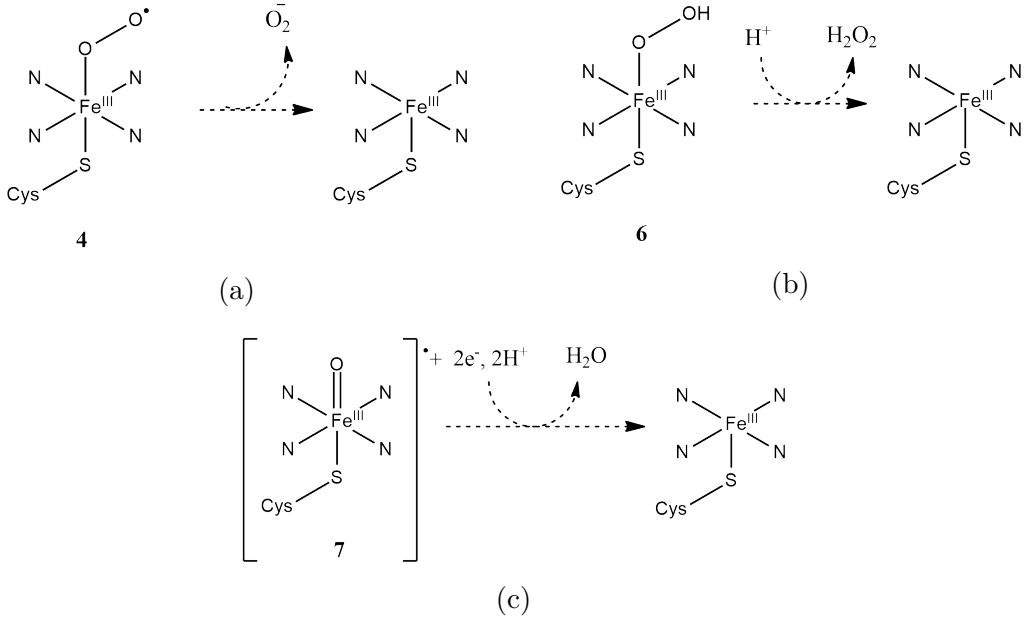


Figure 4: The potential uncoupling pathways of intermediate species **4** (a), **6** (Cpd 0) (b) and **7** (Cpd I) (c) within the P450 catalytic cycle.

unproductive two-electron reduction of Cpd I to water. This occurs when the substrate has no suitable C-H bonds or other target suitably positioned over the heme active centre that can be oxidised. The degree of uncoupling can potentially be limited by having an appropriate substrate within the active site.[24,25] If this substrate fits well within the binding pocket, this would expel the majority of water molecules, and place the substrate in the best position for efficient C-H bond abstraction.

1.3 Reactions by P450 systems

C-H bond oxidation by P450 enzyme is generally believed to be catalysed by the reactive iron-oxo porphyrin species compound I (Cpd I in Figure 1).[8,26,27] However, experimental studies have revealed that the electrophilic ferric-hydroperoxide species that proceeds Cpd I, commonly known as compound 0 (Cpd 0) may participate in some forms of oxidation activity, including sulfoxidation and alkene oxidation.[28–33] These two species are known to have different oxidative properties which dictate the preferences for certain oxidative reactions over others.[32,34] The most abundant species in this "two-oxidant" model will facilitate the oxygenase activity of the P450 and determine the resulting product profile.

Recently, it has been proposed that instead of a "two-oxidant" model of P450s facilitating different reactions, a "two-state" reactivity of Cpd I model can govern the P450 activity. Here Cpd I has two unpaired electrons in the iron orbital 3d π^* and one in the porphyrin ring a_{2u} π^* orbital. The name "two-state" comes from the high-spin quartet state of these orbital ($^4A_{2u}$), in which the three electrons spins are parallel, while the low-spin doublet state ($^2A_{2u}$) has the electron in the porphyrin orbital inverted (Figure 5).

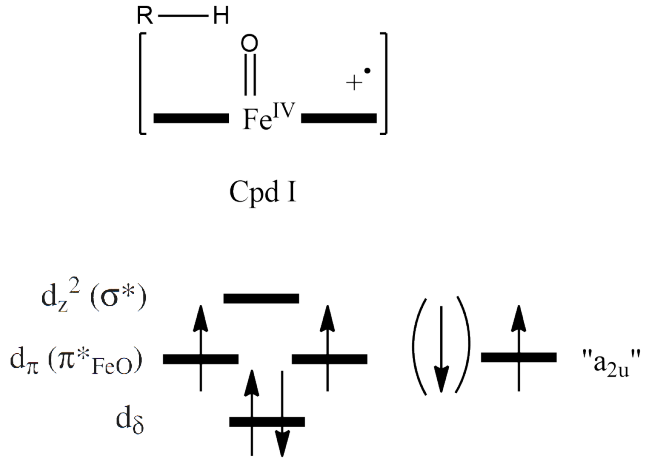


Figure 5: The "two-state" electronic structure of Cpd I

This "two-state" model has been supported by mechanical/molecular mechanical (QM/MM) and density functional theory (DFT) calculations.[26,35,36] These two spin-states have virtually the same energy, so the energy barriers leading from these two states for different oxidations determines the dominant spin-state present and thus the outcome of the substrate oxidation.

1.3.1 Hydroxylation

The insertion of molecular oxygen into an unactivated C-H bond is the most common reaction catalysed by P450s. The reaction is thought to proceed by a radical rebound mechanism from Cpd I (Figure 2). This "two-state" process involves the formation of a radical substrate. One of the first radical clock studies to measure the rate of oxygen rebound involved in ring-strained probes such nortricyclane, methylcyclopropane, and bicyclo[2,1,O]pentane.[37] These compounds can undergo rapid ring-opening via radical intermediates. The radical intermediates can give rise to both rearranged and unrearranged hydroxylation products, and the ratio of these products along with the known rate constant of ring-opening reactions allowed the lifetime of these radicals to be calculated. For instance, the lifetime of the radical intermediate of bicyclo[2,1,0]pentane by rat liver P450s was 50 ps (Figure 6).[37]

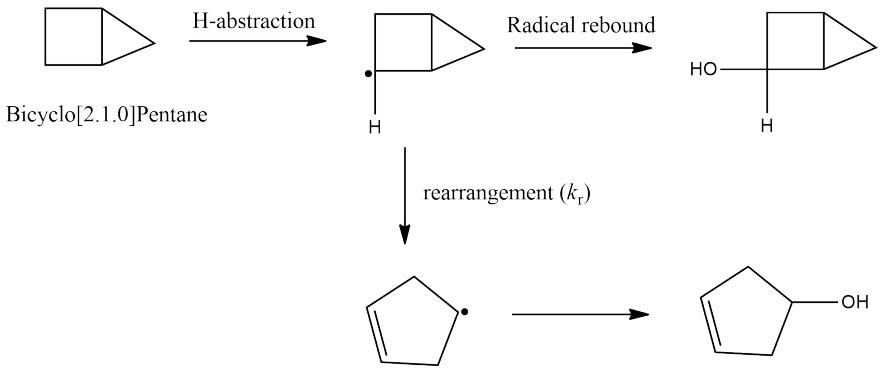


Figure 6: Radical clock study for bicyclo[2,1,0]pentane as a substrate for microsomal P450s. k_r is the rate of free-radical rearrangement used to calculate the lifetime of the radical intermediate.

The ferric hydroperoxide species proceeding Cpd I (Cpd 0) has also been proposed to be an alternate or even primary oxidising species.[32,38,39] Although evidence of a radical intermediate supported the "two-state" model, further ultra-fast radical clock substrates were found to have significantly shorter lifespans (70-200 fs) which were more akin to the lifetimes of transition states rather than radical intermediates.[40–42] There is some evidence that P450 oxidation reactions proceed by a cation intermediate through theoretical models such as DFT. Thus Cpd 0 has also been proposed to facilitate some hydroxylation activity. However, many theoretical studies have also ruled out the involvement of Cpd 0 for hydroxylation.[36,40,43–45]

Through Cpd I in the two-state reactivity model, the low-spin species proceeds in a "barrier-less" radical rebound mechanism after hydrogen abstraction (Figure 7). This pathway explains the rapid radical recombination outlined in the ultra-fast radical clock studies, as the "barrier-less" pathway is essentially concerted. Based on DFT calculations, the high-spin species must traverse a significant energy barrier to reach the radical intermediate before forming the hydroxylated product.[44] Thus in the context

of the two-state model, an abundance of the low spin-state of Cpd I should promote hydroxylation through this radical intermediate process.

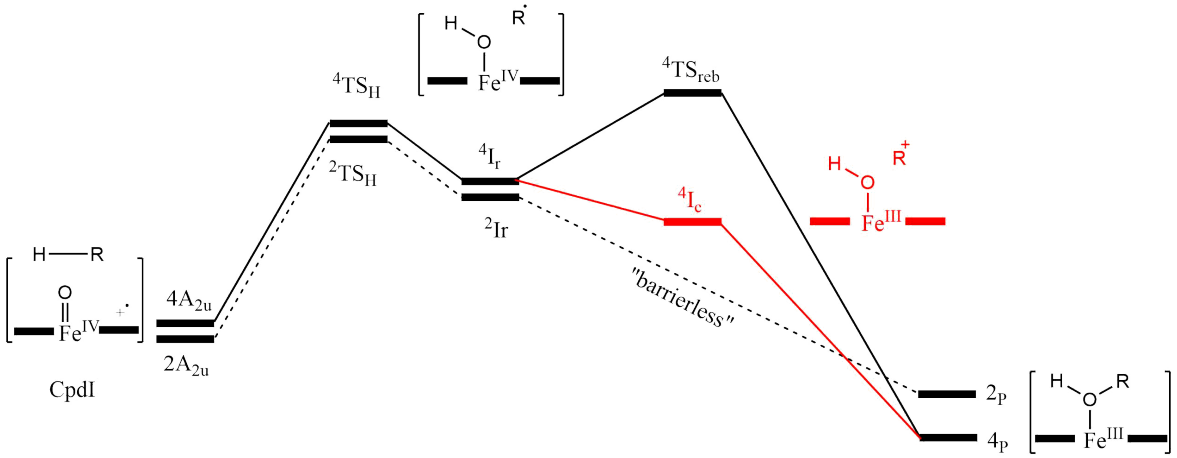


Figure 7: Energy profile for hydroxylation of an unactivated carbon-hydrogen bond for both high-spin ($^4A_{2u}$, solid line) and low-spin (($^2A_{2u}A$, dotted line) of Cpd I. TS_H and TS_{reb} represent the transition states of hydrogen abstraction and radical rebound respectively. I_r and I_c represent the radical and cation intermediates respectively. The high-spin pathway has been proposed to favour the cation intermediate (in red) over the radical intermediate pathway.[36,44]

DFT calculations have revealed significant positive charge buildup on the quartet pathway, implying that a cation intermediate pathway may bypass the large energy barrier required for a radical rebound pathway (Intermediate state ^4I_c in Figure 7).[40–42] As both doublet and quartet states of Cpd I are almost identical in energy, factors such as the identity of the bound substrate and the heme active site environment will determine the dominant spin-state, the radical/cation intermediates and ultimately the reaction pathway.[44,46,47]

1.3.2 Desaturation

P450 catalysed desaturation, which can occur alongside hydroxylation, is proposed to also be catalysed by Cpd I.[48] After hydrogen abstraction, instead of a radical recombination resulting in a hydroxylated product, an imperfect positioning of the carbon radical may facilitate the transfer of an electron, resulting in the formation of a carbocation which can be followed immediately by proton transfer to the ferryl oxygen.[49] Another potential pathway involves a second hydrogen abstraction by Cpd II of the carbon adjacent to the first hydrogen abstraction, directly forming the saturated product (Figure 8).

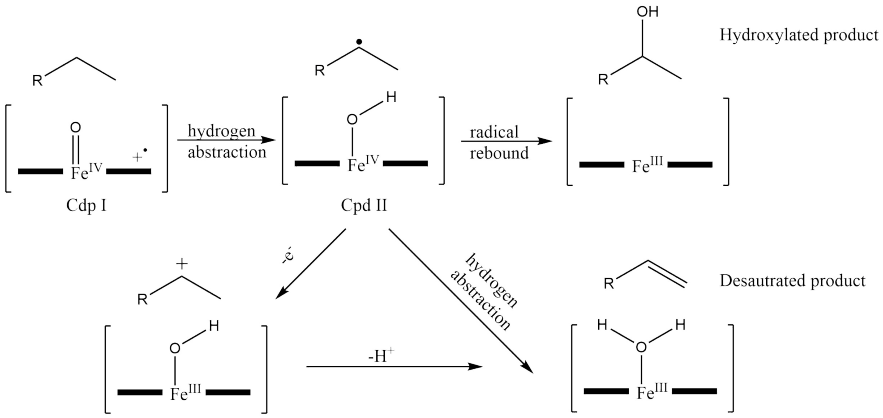


Figure 8: The mechanism of P450-catalysed desaturation reactions by Cpd I.

DFT and QM/MM studies have been used to determine the factors that partition the formation of hydroxylation and desaturation.[50,51] The doublet state of the two-state model is barrierless, and like hydroxylation the desaturation process may arise from this low-spin process. The quartet state also provides a significant barrier for both hydroxylation and desaturation, suggesting a cation intermediate. Factors that dictate the spin-state within the two-state model such as the type of substrate being oxidised and the heme active site environment also influences the distribution of hydroxylation and desaturation products.[50]

1.3.3 Alkene oxidation

The oxidation of alkenes (olefins) usually result in an epoxide product.[49] Experiments show a retention of stereochemistry, and so the interaction between the olefin and the P450 was proposed to proceed in a concerted fashion (Figure 9).

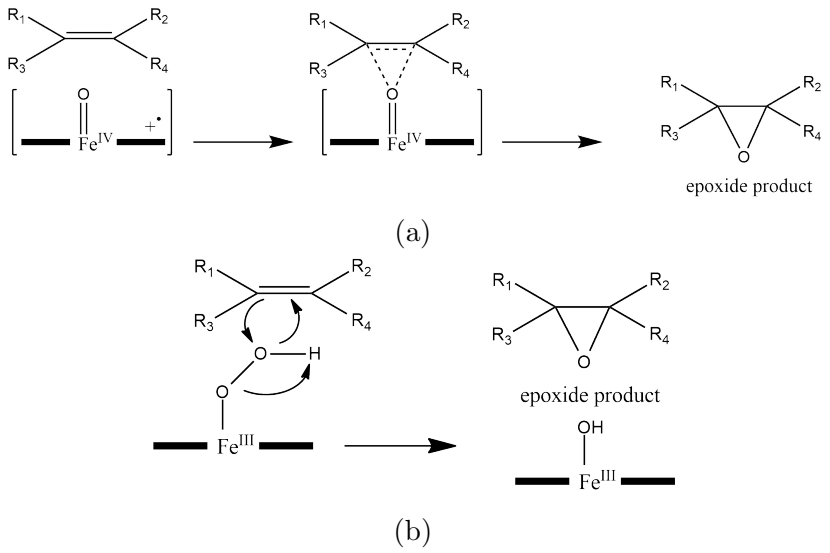


Figure 9: The mechanism of alkene epoxidation by Cpd I (a) and Cpd 0 (b)

While epoxides are usually the dominant product from alkene oxidation, other products such as aldehydes and ketones are formed from P450 catalysed olefin oxidation. In addition, the oxidation of terminal alkenes can result in an irreversible *N*-alkylation of the porphyrin ring, effectively rendering the P450 inactive. However, this process is uncommon with occurrences in only 1 in 200 attempted oxidations leading to inactivation of the enzyme.[29] In addition to these products, P450s that have been treated with their epoxide products do not produce these side-products, suggesting P450s are capable of oxidising alkenes in a non-concerted process.[49] This suggests that a species other than the radical-cation intermediate Cpd I must be responsible for these side reactions. While experimental evidence suggests that Cpd 0 may participate in olefin oxidation,[30,32,52] DFT and QM/MM theoretical studies determined that the hydroperoxide species would be a slow epoxidizing agent.[29,53–55] The quartet spin-state in the two-state model showed a large energy barrier that must be overcome, whereas with the doublet state the "barrier-less" reaction will only form the epoxide product (Figure 10).[44] Thus the quartet spin-state may also partition into a cation intermediate, which could be used to explain the 1,2-rearrangement which results in aldehydes and ketones, as well as *N*-alkylation of the porphyrin ring (Figure 10).

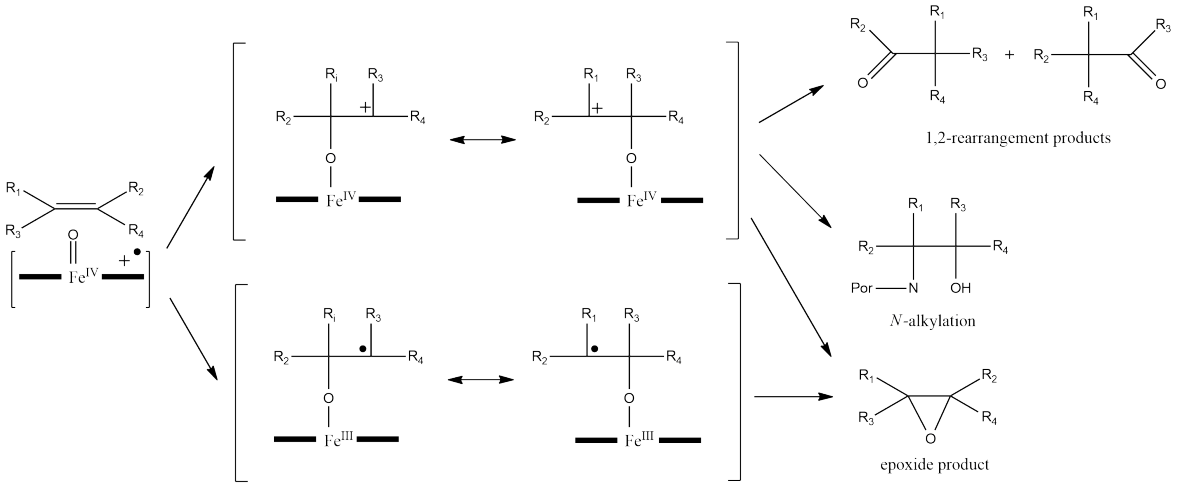


Figure 10: The mechanism of alkene epoxidation through radical intermediates of both doublet and quartet spin-states (bottom pathway) and cationic intermediate of the quartet spin-state (top pathway).

1.3.4 Heteroatom oxidation and dealkylation

There are two main mechanistic pathways proposed for P450 catalysed heteroatom dealkylation. Oxygen dealkylation is generally thought to proceed in a similar fashion as hydroxylation.[56] Here hydrogen abstraction occurs at the carbon alpha to the heteroatom, which initiates the radical rebound mechanism to yield a hydroxylation product. This sequence is known as hydrogen atom transfer (HAT). The unstable product either undergoes decomposition to yield the dealkylated product, or this cleaving process is facilitated by the P450 (Figure 11a).

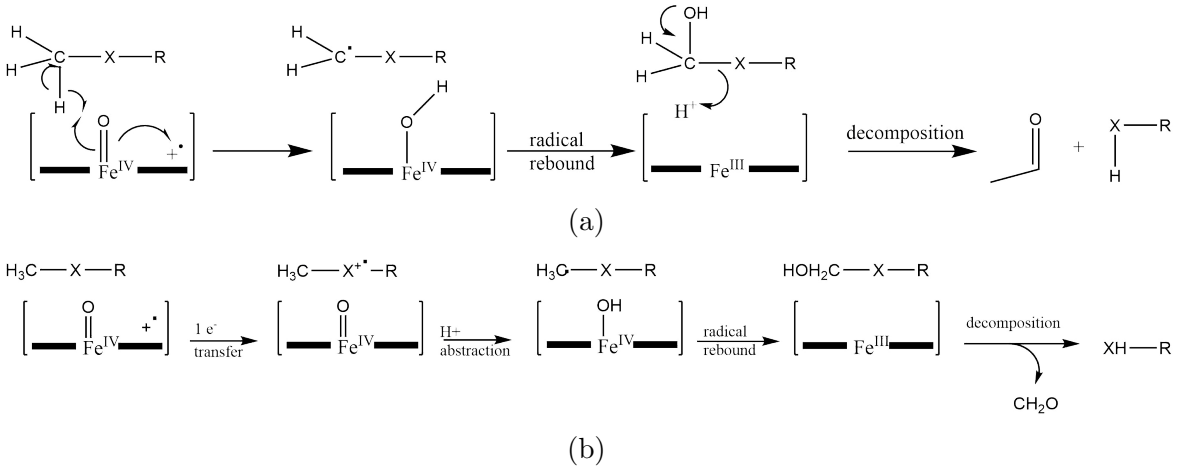


Figure 11: Dealkylation of an alkyl substrate via the HAT pathway (a) or via the SET pathway (b) (X = N, O or S).

Dealkylation of heteroatoms may also involve an initial single electron transfer (SET) from the heteroatom to the ferryl species Cpd I, generating a radical cation (Figure 11b). This species can then collapse with the iron-bound oxygen to give either an dealkylation or an oxide product. While the SET pathway suggests that these two product species would be in competition for each other, this is rarely the case for

nitrogen containing species.[49] A comparison study of the yields of *N*-dealkylation versus *N*-oxidation of *para*-substituted N,N-dimethylanilines. showed that when both processes were unhindered, the main product was the dealkylated nitrogen species.[57]

The oxidation of thioethers by P450s can produce both sulfur-dealkylated and sulfoxide products. Similar to the dealkylation of oxygen and nitrogen species, the reaction can occur either from hydroxylation of the carbon adjacent of the sulfur (HAT), or from a single electron transfer (SET) that generates a sulfur radical cation as an intermediate.[49] This stepwise electron transfer can also give rise to sulfur oxidation, which is more likely to occur than a direct attack of the oxygen atom (Figure 12).[56] The sulfur oxidation can proceed after a single electron transfer followed by sulfur-oxygen bond formation, which will rebound to form the sulfoxide (Figure 12).[29,49,58]

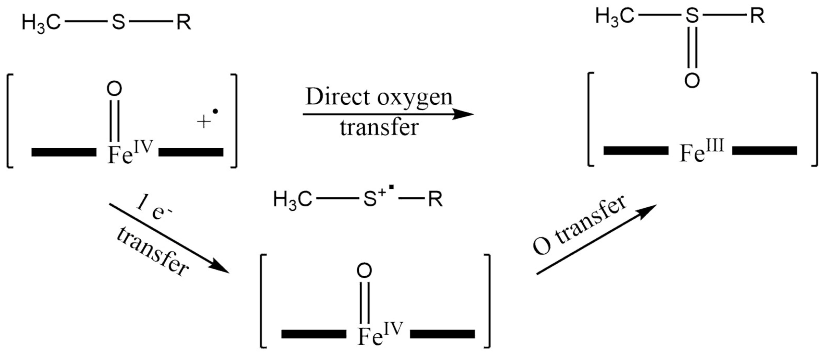


Figure 12: Sulfoxidation of a thioether via direct oxygen transfer and the SET pathway.

Computational studies have indicated that sulfoxidation is mediated by the high spin-state of Cpd I.[59,60] Studies involving the decrease of Cpd I activity through the mutation of the conserved threonine to alanine in P450cam had shown a decrease in hydroxylation activity but not sulfur oxidation.[61] The mutation affects the proton delivery that would convert Cpd 0 to Cpd I, hence a decrease of the Cpd I species does not appear to significantly decrease sulfur oxidation activity, suggesting a "two-state" reaction model. However, studies have also suggested that Cpd 0 is a slow oxidant for sulfoxidation in P450s.[59,60]

Protonation of the proximal oxygen atom of Cpd 0 can form a $Fe^{III}(H_2O_2)$ complex (Figure 13a). It has been suggested that this intermediate can facilitate sulfur oxidation. Recently computational calculations have shown that a concerted pathway for sulfur oxidation is energetically favoured by this species.[62] This pathway is described in Figure 13b, where the distal oxygen undergoes nucleophilic attack by the sulfur atom in the substrate, resulting in heterolytic O−O cleavage coupled to proton transfer.[62] The study found that the $Fe^{III}(H_2O_2)$ complex would facilitate oxidation of thioethers at a faster rate than Cpd I. Hence the "two-state" model may still be applicable for sulfur oxidation, as the $Fe^{III}(H_2O_2)$ species can function as a highly efficient oxidant that

bypasses Cpd I.

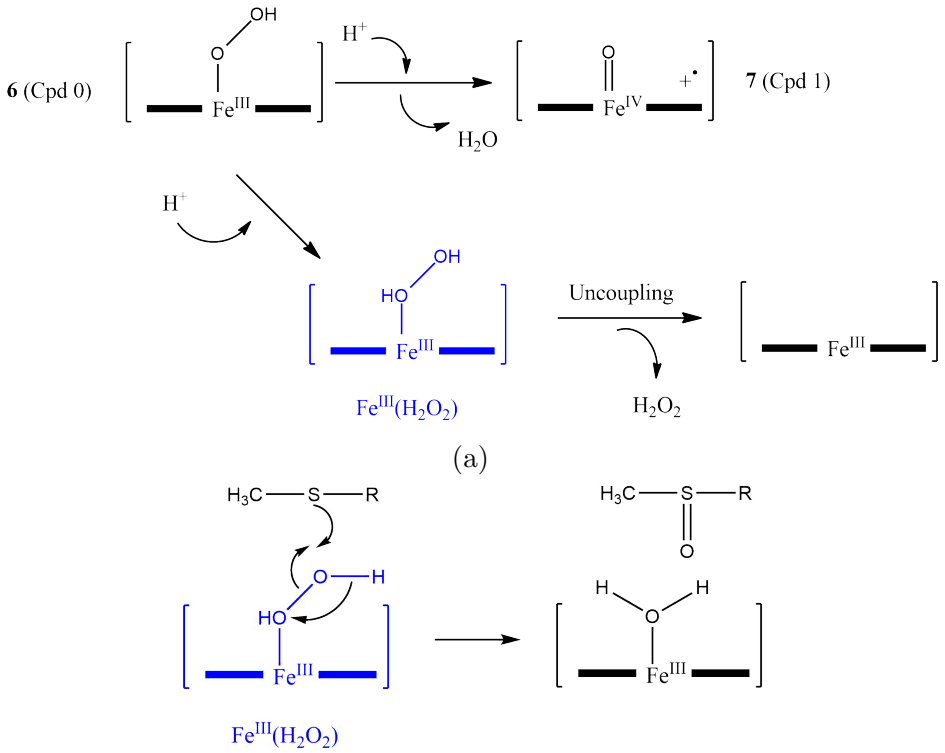


Figure 13: "Uncoupling" process of **6** (Cpd 0) forming the $Fe^{III}(H_2O_2)$ complex (in blue) (a), and the concerted sulfur oxidation by the $Fe^{III}(H_2O_2)$ complex (b).

1.4 CYP199A4 S244D mutant from *Rhodopseudomonas palustris* HaA2

Rhodopseudomonas palustris is a purple gram-negative bacterium which can grow without oxygen, fix nitrogen, produce hydrogen gas, convert carbon dioxide into biomass and biodegrade a wide range of substrates.[63–66] The gene sequence of different strains of *Rhodopseudomonas palustris* are known to differ to best suit their microenvironments.[67] The bacterial strain *Rhodopseudomonas palustris* HaA2 has been reported to contain nine P450 genes.[68,69] These P450s each belong to a different P450 family, suggesting that each of these enzymes oxidise a unique range of substrates.

The CYP199A4 enzyme from *Rhodopseudomonas palustris* HaA2 is a class I P450 with two electron transfer partner proteins; a [2Fe-2S] ferredoxin (HaPux) and a flavin-dependent ferredoxin reductase (HaPuR). Using NADH as an electron source, this system is able to oxidises a narrow range of *para*-substituted benzoic acids.[7,10,70–73] CYP199A4 is reported to bind to 4-methoxybenzoic acid with a high affinity ($K_d = 0.28$ µM) and inducing a $\geq 95\%$ spin-state shift.[7] CYP199A4 regioselectively demethylates 4-methoxybenzoic acid at the *para*-position to form 4-hydroxybenzoic acid (Figure 14).

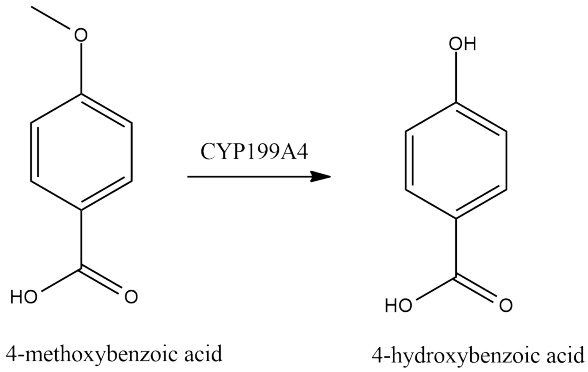


Figure 14: The demethylation of 4-methoxybenzoic acid to 4-hydroxybenzoic acid by CYP199A4.

The importance of the *para*-substituted framework for effective turnover activity is highlighted in recent studies on the oxidation range of CYP199A4 with substituted benzoic acid derivatives.[10,70,71] While the methoxy group in the *para*-substituted position is essential for optimal activity, closely related benzoic acid derivatives also show high affinity and induce different oxidative activities including hydroxylation, desaturation and heteroatom oxidation.[10,21,70,71,73] These reactions include hydroxylation of 4-methylbenzoic acid to 4-hydroxybenzoic acid, desaturation of 4-ethylbenzoic acid to 4-vinylbenzoic acid, and thioether oxidation of 4-(methylthio)benzoic acid to 4-(methylsulfinyl)benzoic acid.[10,73] The carboxy terminus of the *para*-substituted benzoic acids are important for tight binding to CYP199A4, as replacement of this functional group for other moieties, such as an aldehyde or alcohol has shown significant reduction

in binding affinity and turnover rate by the P450.[10,72,73]

Crystal structures of CYP199A4 with a range of benzoic acids bound to the active site have revealed the binding mechanisms that dictate selectivity and outcome of substrate oxidation (PDB: 4EGM, 4EGN and 4EGO).[10,21] The substrate is held tightly within the active site, with the carboxyl group or an equivalent substituent interacting with hydrophilic active site residues, while the benzene ring interacts with hydrophobic aromatic residues. The *para*-substituted moiety is held above the heme ring for oxidative enzymatic activity explaining the regioselectivity of the enzyme.

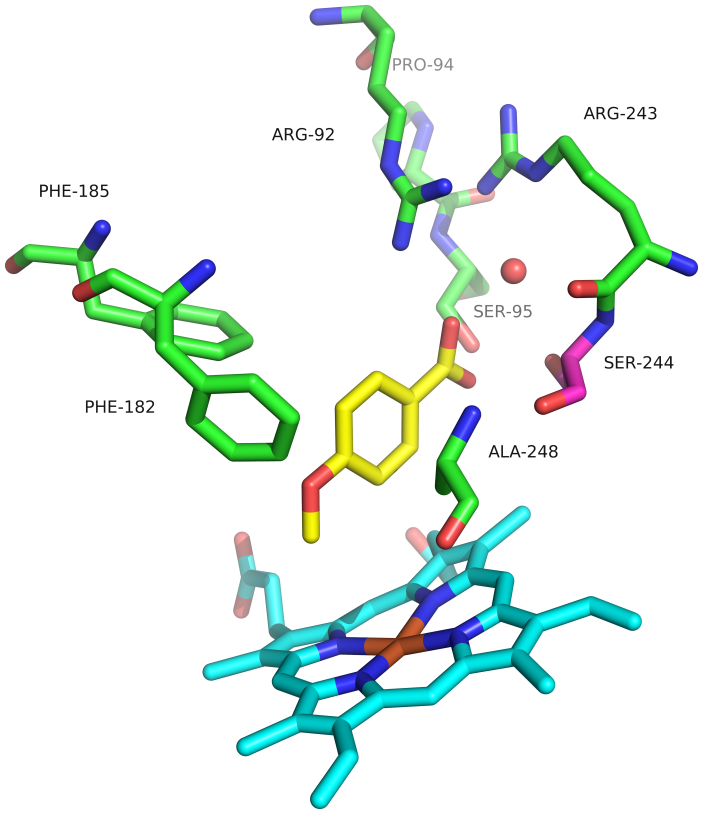


Figure 15: The active site of CYP199A4 with 4-methoxybenzoic acid. The amino acids that select for substrate specificity have been included. The heme active site is coloured cyan, while 4-methoxybenzoic acid is coloured yellow.

Mutating the active site residues have shown to have an effect on the substrate range of CYP199A4.[10,71,72] For instance, mutating a single residue phenylalanine 185 to an isoleucine increased the amount of the desaturated product 4-vinylbenzoic acid during 4-ethylbenzoic acid turnover when compared to the wild-type variant.[10] The serine 244 amino acid interacts with the carboxyl group in 4-methoxybenzoic acid. By mutating this polar amino acid to the acidic aspartate (S244D), the turnover rates for 4-methoxybenzoic acid is lowered to less than half of that from the wild-type (WT) enzyme.[72] However, this mutation increased the binding of other functional groups *para* to the methoxy moiety, including alcohols, aldehydes, amides and nitrile groups.[71,72] While the wild-type CYP199A4 exhibited low turnover activity for these

substrates, such as 4-methoxybenzaldehyde and 4-methoxyphenol, the S244D variant dramatically increased product formation and coupling.[71,72] In the case of demethylation of 4-methoxyphenol to 4-hydroxyphenol, a 900-fold increase of substrate oxidation rate was observed compared to the wild-type P450.[72] This increase of activity for substrates with modified carboxyl groups was also observed in *para*-methyl and *para*-ethyl substituted benzene derivatives. Therefore the increase in O-dealkylation and hydroxylation observed on a range of substituted benzenes for the S244D variation could potentially be applied to other P450 reactions such as sulfoxidation and epoxidation.

1.5 Bacterial Systems and P450s

Metagenomics is the study of how molecular biology affects the relationship between an organism and its environment.[74] These studies have proven to be rewarding due to the huge biodiversity of prokaryotes.[75] For example, sixty percent of drugs in the past 20 years have been sourced from natural compounds from biological sources which include bacteria, rather than from the chemical industry.[76] As such bacteria are considered an extremely promising source of novel chemical compounds. Metagenomic research can potentially lead to untapped resources of antibiotics,[77] anti-cancer compounds[78] and biologically active secondary metabolites. Metagenomics can also identify desirable resources outside of the medical world. Industrial enzymes that can help selectively synthesise chemicals, pharmaceuticals and industrial products such as high-performance detergents can be discovered with the help of metagenomics.[79] The synthesis of many of these bioactive natural products utilise cytochrome P450s, such as CYP154C3 from *Streptomyces griseus* which can oxidise various steroids like testosterone, and CYP107L1 from *Streptomyces venezuelae* which plays a major role in an antibiotic biosynthetic pathway, making the characterisation of bacterial P450 systems desirable.[80,81]

The number of different cytochrome P450 (CYP) genes in microbial genomes vary significantly, even within the same species. Genomic sequencing projects and metagenomics are continuing to reveal novel bacterial P450s and electron transfer enzymes for further study. The most extensively researched P450, P450cam is from a bacterial system *Pseudomonas putida*. P450cam is also the first P450 crystal structure to be solved, which still forms much of our understanding of P450 structures today.[16,24] Structural features and chemistry of oxygen activation is also largely influenced by the extensive characterisation of this prokaryotic P450.[16,82] The ease of which bacterial P450s can be purified, handled and manipulated compared to those derived from animals or plant cells allow these P450s to be easier to characterise and utilise for mechanistic studies and practical application.

Bacterial P450s tend to be more specialised than eukaryotic P450s and their substrate range is often narrower.[15] Bacterial P450s are thought to be involved in metabolism of compounds to be used as sources of nutrition, and have also shown extensive physiological roles, such as the production of antibiotics and other natural products that can have benefits towards human health and agriculture.[5] One example is the P450s from *Streptomyces himastatinicus* which are involved in creating the antibacterial drug himastatin.[83] Investigating P450s from novel bacteria will provide insight into the metabolic pathways of these organisms as well as potentially providing new and novel sources of secondary metabolites and natural products.

However, the industrial use of P450s remains a significant challenge at the current

moment.[3,84] Although bacterial P450s are far easier to express and produce in soluble form, many of these P450s have limited stability. High or low temperatures, and unfavoured solvents can destabilise the integrity of the P450s.[3,84] Expensive cofactors such as NAD(P)H and other cofactors required for P450 catalysis also limits the effectiveness for larger scale use. While this problem may be mitigated by using a whole-cell system that can regenerate these reducing equivalents, these systems are in turn limited by substrate solubility and toxicity of the substrates/products towards the cells.[3,84] Approaches such as directed evolution and rational design have been proposed in order to overcome these problems.[3,85,86] The effective engineering of libraries of P450 enzymes with desirable substrate specificity and activity suitable for application at a large scale is an active area of research.

1.6 *Frankia* actinobacteria

The bacterial phylum *actinobacteria* contains a diverse range of high guanine-cytosine (GC) gram-positive bacteria which are known for their ability to produce a diverse range of biologically active natural products.[87] Bacteria from the genus *Frankia* consists of mycelium-forming actinomycetes that are known for being nitrogen-fixing symbionts of actinorhizal plants. This symbiosis allows the actinorhizal plants to habitat harsh and diverse environmental conditions.[88] While actinomycetes are well-known for the ability to produce bioactive compounds, there has been few studies into the potential compounds from *Frankia*.[87] Little is understood about the interactions between the *Frankia* and its host plant, and virtually nothing is known about the infection process or the subsequent symbiotic development between the actinobacteria and its host.[89] Bioinformatic assays have shown an unusually large number of secondary metabolites biosynthetic gene clusters,[87] which many of which have not been observed or characterised. The identified cytochrome P450s must be integral enzymes to the biosynthesis of these compounds, and the characterisation of these holoenzymes could lead to an alternate biosynthetic route to novel natural products.

Figure 16: *Frankia alni* growth on root nodules. Photo by Gerhard Schuster.[90]

While there is only one named species of *Frankia* bacteria, *Frankia alni*, there are many strains which are specific to a particular plant host organism. Some of these strains, such as ACN14a, CcI3, and EAN1pec, encode a variety of natural products within their genome.[87] Not much is known about *Frankia* sp. EuI1c (also known as *Frankia inefficax* sp.),[91] aside from the strain having high tolerance for copper.[92,93] Genomic analysis of *Frankia* sp. EuI1c reveals this strain contains 68 P450 genes, which is an unusually high number for a bacterial system. This is also large compared to its overall gene size; with 7,264 genes in total, nearly 1% of its genes code for cytochrome P450s. Compared to *Mycobacterium tuberculosis*, another gram-positive bacteria that has 20 cytochrome P450s out of around 4000 genes[94] which is considered relatively high, the

greater number of P450 genes in this *Frankia* strain bacterium makes it an interesting target for cytochrome P450 characterisation.

Due to the abundance of cytochrome P450s in *Frankia* sp. EuI1c, their potential significance cannot be overlooked. The analysis of the structure, substrate range and mechanism of action can be beneficial towards characterisation of these enzymes. In this project the different cytochrome P450s of *Frankia* strain EuI1c and their electron transfer systems will be investigated.

1.7 Thesis objective

Bacterial systems such as *Frankia* are known to be a large source of natural products. Many of these bacterial sources, such as the bacterial family *Frankia*, are untapped for potentially desirable compounds. Although recent genomic studies have shown the potential for *Frankia* as a source of unique and novel natural products, little has been done to isolate these products. The P450s and their electron transfer partners from *Frankia* sp. EuI1c will be investigated for expression, purification and yields. The substrate range of the P450s will also be investigated through screening a variety of different chemical compounds. Turnover activity of these substrates with the various P450s will also be investigated.

The majority of the ferredoxin genes clustered next to P450 genes in the *Frankia* sp. EuI1c genome encode for the uncommon [3Fe-4S] iron-sulfur ferredoxin. The conserved sequence binding motif of this [3Fe-4S] will be investigated in terms of expression and stability by creating a mutant library of one of these purified ferredoxins.

The P450 mutant CYP199A4 S244D has demonstrated to have higher activity for many modified *para*-substituted aromatic compounds than its wild-type variant.[72] These include various modified benzene substrates with *para*-methyl, ethyl and methoxy groups. In order to investigate the maximum potential of the S244D mutant for turnover activity, these substrates will be tested in a regenerating NADH system.

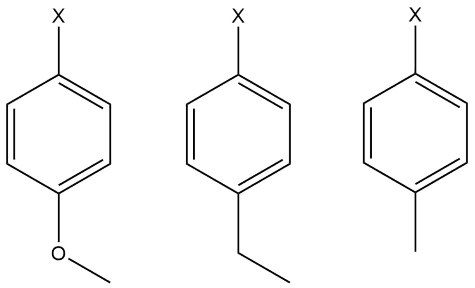


Figure 17: The different type of substrates to be tested for total turnovers with CYP199A4 S244D, where X represents various functional groups. (a) *para*-methoxy substituted aromatic substrates, (b) *para*-ethyl aromatic substrates and (c) *para*-methyl aromatic substrates.

Replacing this *para* moiety with other groups, such as thioethers (4-methylthio-) and alkenes (4-vinyl-/styrenes) have shown activity with the wild-type CYP199A4.[72,73] Exclusive sulfoxidation was observed for 4-(methylthio)benzoic acid, while epoxidation is the primary catalysed reaction for 4-vinylbenzoic acid. In order to further investigate the oxidative potential of the S244D variant compared to the wild-type CYP199A4, the mutant will be tested with a series of 4-methylthio and 4-vinyl-substituted aromatic compounds.

Both CYP199A4 and its S244D variant are reported to regioselectively oxidise substrates at the *para*-position. For example, both these enzymes can selectively oxidise 3,4-dimethoxybenzaldehyde exclusively at the *para*-position to 3-methoxy-4-hydroxybenzaldehyde.[72] A range of benzene derivatives with methyl groups at the *para* position and either the *meta* or *ortho* position will be investigated to determine their regioselective activity with the S244D mutant of CYP199A4.

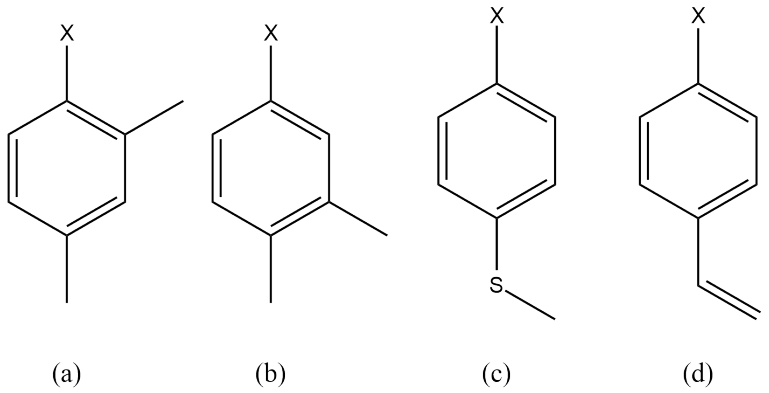


Figure 18: The different type of substrates to be tested with CYP199A4 S244D, where X represents various functional groups. (a) *para*- and *ortho*-substituted and (b) *para*- and *ortho*-substituted aromatic substrates, (c) *para*-methylthio substituted aromatic substrates and (d) *para*-vinyl aromatic substrates (styrenes).

Finally, the crystal structure of CYP199A4 has been solved with a variety of different substrates.[10,21] These structures have revealed the amino acids involved with the binding of *para*-substituted benzoic acid derivatives. In order to compare the changes the S244D variant has on the binding of these substrates, the crystal structure of a 4-methoxybenzoic acid bound S244D mutant will be investigated.

2 Experimental

2.1 General

Reagents, solvents and organic substrates were purchased from Sigma-Aldrich, VWR, Fluorochem, Acros Organics, Tokyo Chemical Industry and Matrix Scientific. Antibiotics, isopropyl β-D-thiogalactopyranoside (IPTG) and dithiothreitol (DTT) were purchased from Astral Scientific. T4 DNA ligase was from Lucigen and restriction enzymes for molecular biology were from New England Biolabs. KOD polymerase used for the polymerase chain reaction (PCR) steps, and the pET28, pET26a and pETDuet vectors were from Merck Millipore.

The genomic DNA of *Frankia* EuI1c was obtained from Prof. Louis Tisa, University of New Hampshire. General DNA and microbiological experiments were carried out by standard methods.[95]

UV/Vis spectra and spectroscopic activity assays were recorded at 30 ± 0.5 °C on an Agilent Cary 60 or Cary 5000 UV-vis spectrophotometer. Protein chromatography was performed on AKTA (GE healthcare) and GradiFrac (Amersham Pharmacia) systems. High performance liquid chromatography (HPLC) analysis was performed on an Agilent 1260 Infinity pump equipped with an autoinjector and Aligent Eclipse Plus C18 column (250 nm x 4.6 nm, 5 µm). For chiral analysis on the HPLC, a Lux 3u Cellulose-1 column (100 mm x 4.6 mm x 3µm; Phenomenex) was used. Gas chromatography/mass spectrometry (GC-MS) analysis was performed on a Shimadzu GC17A equipped with a QP5050A GC-MS detector and DB-5 MS fused silica column (30 m x 0.25 mm, 0.25 µm). For chiral analysis on the GC a β-DEX column (30 m x 0.25 mm, 0.25 µm) using a Shimadzu Tracera system equipped with a BID detector.

2.2 Protein Expression and Purification

2.2.1 Rare Codon Analysis

The %MinMax algorithm was used to analyse rare codon clusters (RCC) in P450 genes.[96] This algorithm compares the given mRNA sequence of the hypothetical protein with all the rarest codons (minimum) and the most common codons (maximum) to determine the frequency of codon usage. The online web tool LaTcOm was used for all rare codon cluster analyses.[97]

2.2.2 P450 Expression

The P450 encoding genes from *Frankia* sp. EuI1c (Table 2) were cloned into the vector pET26(+) between the restriction sites NdeI and HindIII. These recombinant CYP

Table 1: Growth media constituents.

Medium	**Constituents (L^{-1})**
LB	Tryptone (10 g), yeast extract (5 g), NaCl (10 g)
SOC	Tryptone (20 g), yeast extract (5 g), $MgCl_2$ (1 g), NaCl (0.5 g), KCl (0.2 g), glucose (0.2% w/v)
2x YT	Tryptone (16 g), yeast extract (10 g), NaCl (5 g)
Trace Elements	Na_2EDTA(20.1 g), $FeCl_3.6H_2O$ (16.7 g), $CaCl_2.H_2O$ (0.74 g), $CoCl_2.6H_2O$ (0.25 g), $ZnSO_4.7H_2O$ (0.18 g), $MnSO_4.4H_2O$ (0.132 g), $CuSO_4.5H_2O$ (0.10 g)
EMM	K_2HPO_4 (7 g), KH_2PO_4 (3 g), $(NH4)_2SO_4$ (1 g), Na_3Citrate (0.5 g), $MgSO_4$ (0.1 g), Glucose (4 g)

genes were transformed in BL21(DE3) competent *E.coli* cells (New England Biolabs) and subsequently grown on an LB plate containing 30 mg L^{-1} kanamycin. A small scale growth was first performed for expression testing (Section 2.2.4).

A single colony was inoculated into each of 5 x 500 mL 2x YT (Yeast Extract Tryptone) containing 30 µg mL^{-1} of kanamycin and 1.5 mL trace elements and grown at 37 °C overnight. The incubator temperature was lowered to 18 °C for 1 hour before the addition of 0.02% v/v benzyl alcohol and 2% v/v ethanol. After a further 1 hour, protein production was induced by 0.25 mM IPTG (from a stock of 0.5 M in water). The cultures were then grown for a further 48 to 72 hours. The cells were harvested from the culture by centrifugation at 5,000 *g* for 10 minutes and then resuspended in 200 mL Tris buffer (50 mM Tris, pH 7.4, 1 mM DTT). The resuspended cells were lysed by sonication using an Autotune CV334 Ultrasonic Processor equipped with a standard probe (136 mm x 13 mm; Sonics and Materials, US) using 30 x 20 s pulses with 40 s intervals. Cell debris was removed by centrifugation at 40,000 *g* for 30 minutes. The supernatant was loaded on to a DEAE Fastflow Sepharose column (XK50, 200 mm x 50 mm; GE Healthcare) and the protein was eluted using a linear salt gradient of KCl (100 - 400 mM) in Tris buffer at a flow rate of 6 mL min^{-1}. The red protein-containing fractions were collected and concentrated by ultrafiltration with a 10,000 kDa unit (Vivacell 100, Sartorius). The protein was desalted via gel filtration on a G-25 Sephadex medium column equilibrated with Tris buffer. After concentration by ultrafiltration, the protein was loaded onto an Source-Q ion-exchange column (80 mm x 30 mm; GE Healthcare) and eluted with solutions A (Tris buffer) and B (1 M KCl in Tris buffer) at a gradient of 10% - 40% solution B over 110 minutes at a flow rate of 6 mL min^{-1}. Purified fractions were combined and concentrated by ultrafiltration (10,000 kDa exclusion membrane, 50 mL). An equal volume of glycerol was added to the concentrated fractions, and the

solution was filtered through a 0.2 µm sterile syringe filter and stored at -20 °C.

Glycerol was removed immediately before use by gel filtration on a 5 mL PD-10 column (GE Healthcare, UK) by eluting with 50 mM Tris, pH 7.4.

2.2.3 Ferredoxin Expression

2.2.3.1 His-tagging and Purification of Ferredoxins The DNA for the ferredoxins and the mutant library for Fdx2495 were bought as gblock sequences (Figure 38, Appendix A1, Section 3.2.5) from Integrated DNA Technologies (IDT). The ferredoxin genes were cloned into a pETDuet vector between the restriction sites NcoI and HindIII. For the Fdx2495 mutant library, the codon of one of the amino acids in the iron-sulfur cluster binding motif was randomly changed from a histidine expressing codon CAC to NNC (where N is a random DNA base).

For the Fdx2495 mutants, forward and reverse primers were designed to insert a His-tagged tail flanking the end of the protein after PCR amplification (Figure 39, Section MutantLabel). The gene was amplified by 25 cycles of strand separation at 90°C for 30 s followed by annealing at 55°C for 20 s and extension at 70°C for 20 s. After amplification of the DNA by PCR, the gene was cloned into pET28 vectors using NcoI and HindIII restriction enzymes. The ligation reactions were then transformed into DH5α *E.coli* DNA competent cells (New England Biolabs) and subsequently grown on a LB plate containing 100 mg L^{-1} ampicillin. A single colony was inoculated into 15 mL LB broth and grown at 37 °C overnight. The cells were harvested by centrifugation at 4000 g for 10 minutes and the supernatant was discarded. DNA was extracted and purified with the EZ-10 spin column plasmid DNA miniprep kits (Bio Basic) following standard protocols. The purified DNA was sequenced at the Adelaide node of the Australian Genome Research Facility (AGRF).

2.2.3.2 Expression of Ferredoxins The purified ferredoxin encoding genes clustered next to P450 encoding genes (Table 2) were cloned into the pETDuet vector between the restriction sites NcoI and HindIII The plasmids containing recombinant ferredoxin genes were transformed in BL21 (DE3) competent *E.coli* cells (C2527 cells from New England Biolabs) and subsequently grown on an LB plate containing 100 mg L^{-1} ampicillin.

A single colony was inoculated into each of 5 x 500 mL 2x YT containing 100 µg mL^{-1} of ampicillin and 1.5 mL trace elements and grown at 37 °C overnight. The cells were grown and induced in the same way as previously outlined for P450 expression. The cells were harvested from the culture by centrifugation at 5000g for 10 minutes and

resuspended in 300 mL Tris buffer 50 mM, pH 7.4 containing 1 mM DTT, 20% v/v glycerol, 1% v/v β-mercaptoethanol, 1 mg mL^{-1} lysozyme and 2 mL Tween. The cells were stirred on ice for 1 hour and then lysed by sonication using 30 x 20 s pulses with 40 s intervals. The ferredoxins were then purified using the same method outlined in the P450 expression section after cell lysis, with minor changes to the salt gradients of KCl during elution from the DEAE sepharose column (150 - 400 mM).

For the His-tagged *FraEuI1c_2495* mutants (Figure 38, Section 3.2.5), the clarified supernatant after cell lysate centrifuging was loaded directly onto a HisTrap FF crude column (16 mm, 5 mL; GE Healthcare) using a syringe. The column was washed with 5 x column volume of binding buffer (50 mM Tris, 300 mM NaCl, 10 mM imizadole) and the protein was eluted in 5 x column volume of elution buffer (50 mM Tris, 300 mM NaCl, 300 mM imidazole). The protein solution were diluted in Tris solution (50 mM, pH 7.4) containing 1 mM DTT, and was concentrated by ultrafiltration (5 kDa, 50 mL Centricon Plus 70, Merck Millipore). This process was repeated 3 times for buffer exchange. An equal volume of glycerol was added to the final concentrated protein solution, and the solution was filtered through a 0.2 µm sterile syringe filter and stored at -20 °C.

2.2.4 Analysis of Recombinant Cytochrome P450 and Ferredoxins

For expression tests for P450s, a single colony of the recombinant P450 containing BL21 (DE3) competent *E. coli* cells was inoculated in 50 mL 2x YT containing 100 µg mL^{-1} ampicillin and 150 µL of trace elements. For the expression tests for ferredoxins 100 µg mL^{-1} ampicillin was added instead of kanamycin ampicillin. The cultures were grown at 37 °C for $\geq$ 4 hours. The incubator temperature was lowered to 18 °C for 1 hour before the recombinant protein production was induced by 0.25 mM IPTG (from a stock of 0.5 M in water). The cultures were grown for a further 12 - 24 hours. The cells were harvested from the culture by centrifugation at 5,000 g for 20 minutes. The expression levels of the P450 or ferredoxin were recorded by visual inspection of the cell pellet, with red and brown colouration indicating P450 and ferredoxins formation, respectively.

In order to determine whether the red/brown colour observed was from the soluble protein, the pellets were resuspended in 4 mL of Bacterial Protein Extraction Reagent (ThermoFisher Scientific) per 1 g of cell pellet and stirred for 20 minutes. Cell debris was removed by centrifugation at 4,000 rpm for 10 minutes. The protein supernatant was loaded onto a 12% SDS gel and analysed by gel electrophoresis in GTS buffer (2.5% glycerol, 20 mM Tris pH 8, 25 mM NaCl).

To determine the amount of P450 after purification, a carbon monoxide difference spectrum was measured.[5] The CO ferrous bound spectra were attained by diluting the

P450 to give a maximum UV/Vis peak absorbance 0.5 - 2.0 using 50 mM Tris, pH 7.4, in 0.5 mL. A small amount of sodium dithionite reducing agent (< 1 mg) was added and CO was gently bubbled through the sample. The spectrum was then recorded between 700 nm and 300 nm. The P450 protein concentrations were estimated using the extinction coefficient at $\epsilon_{446\text{-}490}$ = 91 $mM^{-1}cm^{-1}$ for the CO spectrum minus the reduced ferrous spectrum.[5,98] Using this value, the extinction coefficients for the ferric resting state of the CYP enzymes were calculated to be as follows; FraEu2494 ϵ_{418} = 130 $mM^{-1}cm^{-1}$ and FraEu5334 ϵ_{418} = 151 $mM^{-1}cm^{-1}$.

The expression and purification of CYP199A4, its S244D mutant and HaPuR was done by Coleman and Chao.[72,73] The ferredoxin HaPux was expressed and purified with the same methods as outlined in Section 2.2.3.

2.3 Substrate Binding Assays

2.3.1 Spin-state Shift Assays

P450 proteins were diluted to 0.5 - 1.0 µM using 50 mM, pH 7.4 Tris buffer, and the UV/Vis spectra was recorded. Aliquots of substrates (1 µL, 100 mM in ethanol/DMSO) were added to 500 µL of the protein until the low-spin (LS) (≈ 420 nm) to high-spin (HS) (≈ 390 nm) shift stopped. The weighted averages of the spectra of camphor-free and camphor-bound $P450_{cam}$ (≈ 0% to ≈ 100% HS) were used to estimate the % HS of the P450 enzymes (see Figure 29, section 3.2.2).

2.3.2 Binding Constant Assays

P450 proteins were diluted to 0.5 - 2.0 µM using 50 mM Tris, pH 7.4, in 2.5 mL. Increasing aliquots of substrate (0.5 - 3 µL) were added using a Hamilton syringe from a 1, 10 and then a 100 mM stock solution in ethanol and/or in DMSO. The samples were mixed and the peak-to-trough difference in absorbance was recorded on the UV spectrophotometer between 700 nm and 250 nm. Aliquots of substrate were added until the peak-to-trough difference did not shift any further (see Figure 29, section 3.2.2). The dissociation constants, K_d, were obtained by fitting the peak-to-trough difference against substrate concentration to a hyperbolic function (Equation 2)

$$\Delta A = \frac{\Delta A_{max} \times [S]}{K_d + [S]} \tag{2}$$

Where ΔA is the peak-to-trough absorbance difference, ΔA_{max} is the maximum absorbance difference and [S] is the substrate concentration.

For cases where the substrate bound tightly (K_d <5 [P450] µM), data was instead fit to the tight-binding (Morrison) equation[99] (Equation 3) to determine K_d.

$$\Delta A = \Delta A_{max} \frac{[E] + [S] + K_d - \sqrt{([E] + [S] + K_d)^2 - 4[E][S]}}{2} \quad (3)$$

2.4 *In Vitro* NADH Activity Assays

NADH activity assays were recorded by preparing a mixture (1.2 mL) containing oxygenated Tris (pH 7.4, 50 mM), CYP199A4 P450 enzyme (0.5 µM), HaPux ferredoxin (5 µM) and HaPuR ferredoxin reductase (1 µM). The mixture was equilibrated at 30 °C for 2 minutes. NADH (approximately 320 µM, $A_{340} \approx 2.0$) was then added to the mixture and the absorbance at 340 nm was recorded using a UV/Vis spectrophotometry. The substrate was then added to a final concentration of 1 mM. The rate of NADH oxidation was calculated by using $\epsilon = 6.22$ mM^{-1} cm^{-1}. When substrates such as 4-(methylthio)acetophenone and 4-(methylthio)benzaldehyde, which absorb strongly at 340 nm, were used, the NADH rate was calculated using the absorbance of NADH at 370 nm and calculated using $\epsilon = 2.60$ mM^{-1} cm^{-1}.

For total turnover assays, a mixture of ethanol (80 µL), P450 enzyme (0.1 µM), ferredoxin (1 µM), ferredoxin reductase (0.2 µM) in oxygenated Tris (2 mL, pH 7.4, 50 mM) was prepared. NADH (4 mM) was added to mixture, followed by the addition of substrate (2 mM). A replicate set of total turnovers was set up which also contained a NADH regenerating system through the addition of alcohol dehydrogenase (15 µL of 1 g in 34 mL solution, Roche). The mixture was then set aside at room temperature for > 12 hours before analysis by HPLC or GC-MS.

2.5 *In Vivo* Activity Assays

For CYP199A4 S244D *in vivo* turnovers, the pRSFDuet-*HaPux-HaPuR* and pRSFDuet-*HaPux-CYP199A4 S244D* plasmids were transformed into BL21(DE3) competent cells and grown on LB plates containing 100 mg L^{-1} ampicillin and 30 mg L^{-1} kanamycin.[7,71] For *Frankia* sp. EuI1c, the P450 containing pET28 plasmid and a pETDuet plasmid containing the ferredoxin and ferredoxin reductase genes were transformed into BL21(DE3) competent cells and grown in the same way stated above. A single colony was inoculated into 200 mL LB broth containing 30 µg mL^{-1} of kanamycin, 100 µg mL^{-1} of ampicillin and 1.5 mL trace elements and grown at 37 °C overnight. The incubator temperature was lowered to 18 °C for 1 hour before the addition of 0.02% v/v benzyl alcohol and 2% v/v ethanol. After a further 1 hour, protein production was induced by

0.25 mM IPTG (from a stock of 0.5 M in water). The culture was grown for a further 24 - 48 hours, and then harvested by centrifugation (5000 *g*, 10 min).

The harvested cells were resuspended in EMM and split into 40 mL aliquots in 250 mL flasks. Substrate (2 mM) was added to the whole-cell reaction and the mixture was shaken at 180 rpm at 30 °C. After $\geq$ 18 h, a 1 mL aliquot was removed for analysis.

2.6 Analysis of Metabolites

Turnover products were analysed via HPLC or GC-MS. For GC-MS analysis 990 µL of the reaction mixture was mixed with 10 µL of internal standard solution (20 mM 9-hydroxyfluorene in ethanol). The mixture was then extracted once with 400 µL of ethyl acetate, and then a further two times with 300 µL of ethyl acetate. For most substrates the extracts were used directly for GC-MS.

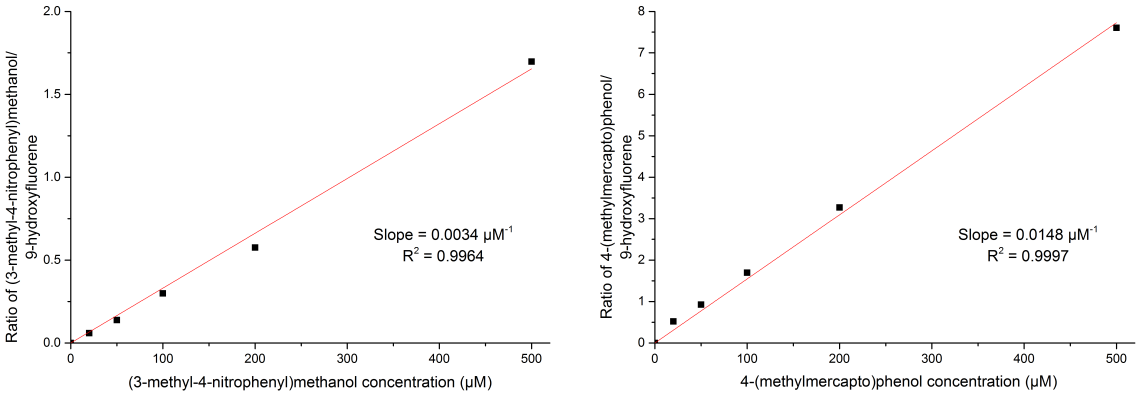
For acids and alcohols that required derivatisation, the extracts were dried over $MgSO_4$. Ethyl acetate was removed under a stream of nitrogen and the residue was dissolved in 200 µL of anhydrous acetonitrile. Excess (35 µL) BSTFA/TMCS (99:1) was added to the mixture and was incubated for 120 minutes to produce the trimethylsilyl ester of the carboxylic acid group and trimethylsilyl ether of the alcohol, if present. The derivatised samples were then analysed by GC-MS.

For the compounds that required derivatisation, the oven temperature was maintained at 120 °C for 3 minutes and then increased to 180 °C at 5 °C min^{-1} where it was held for 2 minutes. This was then increased to 220 °C at 20 °C min^{-1} before being left for a further 2 minutes. For all other substrates, the oven temperature was held at 80 °C for 3 minutes and then increased to 230 °C at 12 °C min^{-1} and maintained for 5 minutes. For chiral GC analysis, the oven temperature began at 80 °C for 3 minutes and then increased to 200 °C at 5 °C min^{-1} and then maintained for a further 3 minutes.

For HPLC analysis, 132 µL of the reaction mixture was mixed with 66 µL of acetonitrile and 2 µL of internal standard (20 mM 9-hydroxyfluorene in ethanol). Samples were eluted with a 20 - 95% or 20 - 50% gradient of water to acetonitrile with 0.1% trifluoroacetic acid (TFA). For chiral HPLC analysis, the samples were eluted with 5%, 10% or 15% isocratic acetonitrile with 0.1% TFA in water.

Authentic standards of the majority of the expected turnover products were not commercially available. As such, the products were calibrated against the closest available compound. In most cases, the this was done against the substrate. The identities of the products were confirmed via GC-MS by their mass fragmentation pattern and if available, coelution with the authentic product.

By using the calibration curve, the concentration of product for each turnover was determined. This along with the known amount of NADH added allowed the calculation of coupling efficiency, which is the percentage of NADH reducing equivalents used to generate product. The product formation rate was found by multiplying the NADH oxidation rate with the coupling efficiency divided by 100.



(a) (b)

Figure 19: An example of a calibration curve used to determine the product formed from CYP199A4 S244D mutant turnovers. The calibration curves shown are for (3-methyl-4-nitrophenyl)methanol (a) and 4-(methylmercapto)phenol (b).

2.7 Synthesis of Chiral Products

Oxidation of thioethers to sulfoxides: Hydrogen peroxide (0.5 mL, 30%) was added to a solution of the *para*-methylthio-modified benzene derivatives (2 mmol) in ethanol (1 mL). The mixture was stirred at room temperature for $\geq$ 48 hours.[100] The mixture diluted into Tris buffer and acetonitrile and then run by HPLC or further diluted into Tris buffer and further extracted with ethyl acetate for GC-MS analysis.

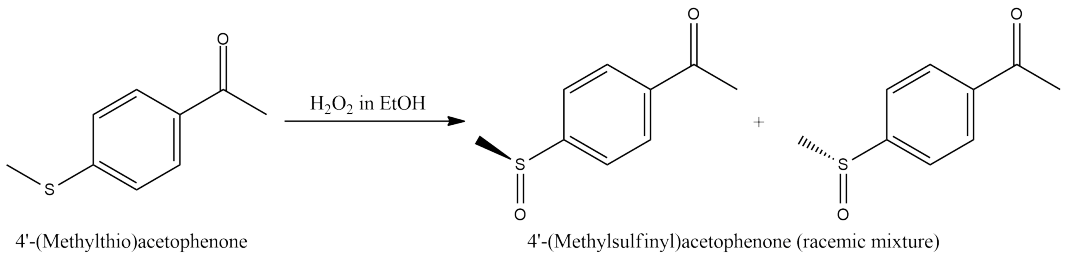


Figure 20: The synthesis of sulfoxides from thioethers using hydrogen peroxide

Oxidation of styrenes to epoxides: A solution of the *para*-substituted styrene (500 µL, 20 mmol in anhydrous acetonitrile) was added dropwise to a stirring solution of meta-chloroperoxybenzoic acid (mCPBA) (500 µL, 20 mmol) and excess sodium bicarbonate over five minutes. The reaction was stirred for 30 minutes over ice, and

then stirred for an additional 15 minutes at room temperature.[101] Solid impurities and undissolved sodium bicarbonate were removed by centrifugation at 10,000 *g* for 1 minute and the supernatant was extracted for GC-MS analysis as described above.

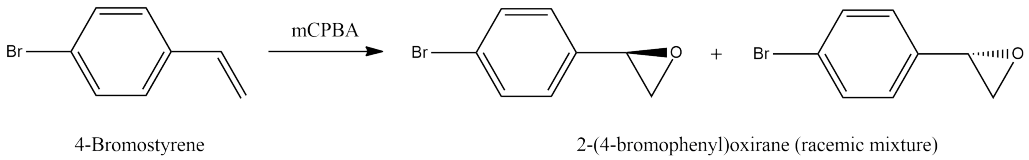


Figure 21: The epoxidation of styrene compounds using mCPBA

2.8 Crystallisation

The concentration of the P450 from *FraEu_2494* was calculated using the extinction coefficient at 418 nm (ϵ_{418} = 131 $mM^{-1}cm^{-1}$) (Section 2.2.4). Before crystallisation, the protein was further purified on a NGC^{TM} Medium-Pressure Liquid Chromatography System (Bio-Rad) equipped with an ENrich^{TM} SEC 650 size exclusion column (10 x 300 mm). Purified FraEu2494 cytochrome P450 was concentrated in crystallisation buffer (50 mM Tris, pH 7.4) to 40 mg mL^{-1}. The sitting-drop vapour-diffusion method was used with Hampton Research Crystal Screen Index and Hampton Research Crystal Screen PEG/ION for FraEu2494 crystals at 16 °C in 96-well Intelli plates (Art Robbins Instruments, USA). One drop (1 µL each) of the protein solution was mixed with 1 µL of reservoir solution and equilibrated against 100 µL reservoir solution. After a week several dark red thin-plate crystals were obtained from the Index crystal screen conditions No. 72 (0.2 M Sodium chloride, 0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3350).

For co-crystallisation of CYP199A4 S244D mutant with 4-methoxybenzoic acid, the purified protein was concentrated to approximately 40 mg mL^{-1} in crystallisation buffer Tris (50 mM, pH 7.4). Initial crystal screening was carried out with Hampton Research Crystal Screens HT and PEG/ION. These conditions were optimised, and the final conditions used for crystal growth were 0.1 M Bis-Tris, PEG 3350 (22%), magnesium acetate (MgAc) (0.2 M). These crystals were obtained using the hanging-drop vapour-diffusion method at 16^{o}C using a 1 µL protein solution mixed with 1 µL reservoir solution and equilibrated with 200 µL of the same solution.

All crystals were flash cooled with liquid nitrogen after briefly immersing into paratone-n as a cryoprotectant. The crystallisation process of the CYP199A4 S244D and flash cooling were performed by Tom Coleman.

2.9 Data Collection and Structure Determination

X-ray diffraction data for CYP199A4 S244D mutant in complex with 4-methoxybenzoate were collected on beamline MX1 of the Australian Synchrotron by Tom Coleman and beamline scientists. The data of 360 images were collected at 100 K. The diffraction data was indexed, integrated and scaled with the Mosflm package.[102] The space group of the mutant was found to be $P2_1$.

The structure of the complex was solved by using the molecular replacement method with the program Phaser[103] in the CCP4 suite[104] using the native structure of CYP199A4 (PDB code: 4DO1) as the initial search model.[21] 4-methoxybenzoate was manually built in eLBOW[105] and adjusted under the guidance of F_o-F_c difference maps by using the program COOT.[106] Structural refinements were carried out via multiple rounds of manual rebuilding in COOT followed by refinement in PHENIX.[105] The stereochemical quality of the refined structures were checked with the program MolProbity.[107]

3 Cytochrome P450s from *Frankia* sp. EuI1c

3.1 Introduction

The *Frankia* genus is comprised of nitrogen fixing bacteria known to live off the root nodules of actinorhizal plants.[108] Actinorhizal plants are a group of flowering plants (angiosperms) characterised by their symbiotic relationship with *Frankia*. This relationship allows these actinorhizal plants to live in harsh environments and diverse ecological conditions such as volcanic ashfields, glacial till, sand dunes, water-saturated soil and dry soils.[109] There is only one named species in the genus: *Frankia alni*. However, different strains of the bacteria are specific to different plant species. Phylogenetic analysis has revealed four distinct clusters among *Frankia* strains.[110] Cluster 1 is a broad group that contains strains such as CcI3, ACN14a, and QA3, and are known to infect the *Alnus* genus of flowering plants. Cluster 2 and 3 are known to be microsymbiots of a multitude of flowering plants.[111] The *Frankia* sp. EuI1c strain is classified in the Cluster 4 phylogeny which are comprised of atypical *Frankia* strains. These strains were isolated from actinorhizal plant nodules, but were unable to re-infect its host plant, or were ineffective strains.[88] *Frankia* sp. EuI1c is an atypical *Frankia* strain, yet unlike the other Cluster 4 *Frankia* strains, it is able to reinfect its host-plant after isolation. However, after re-infection the EuI1c strain forms ineffective root nodules.[108]

(a)

Figure 22: An example of an actinorhizal plant, *Allocasuarina luehmannii* (commonly known as Buloke) which is a native Australian plant. Photo by Mark Marathon.[112]

Frankia strains have a robust metabolism that includes nitrogen fixation, biosynthesis of secondary metabolites and energy-generating pathways. Recently the genomes of multiple *Frankia* sp. strains from each of the four clusters have been sequenced, and

further studies have revealed many of these genomes contain gene clusters related to secondary metabolite and natural product metabolism, xenobiotic degradation and heavy metal resistance.[87,92,111,113–115] A genomic and proteomic study on the *Frankia* strains ACN14a, CcI3 and EAN1pec revealed natural products ranging from lipids, siderophores, fatty acids, steroids, modified amino acids, spore pigments and potential antibiotics (Figure 23).[87] This makes the metabolic and biosynthetic pathways of *Frankia* sp. strains attractive targets for the synthesis of novel natural products.

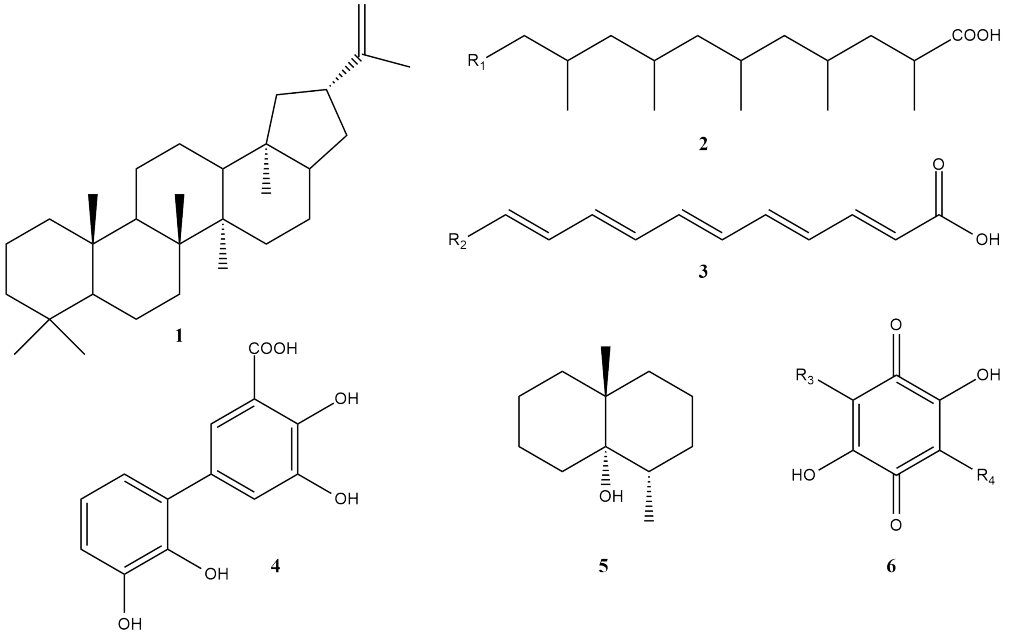


Figure 23: A selection of natural products from *Frankia* sp. strains ACN14a, CcI3, and EAN1pec. These include 1. Hopanoids, 2. Mycocerosate-like lipids, 3. Fatty acids, 4. Aromatic compounds, 5. Geosmin and 6. Quinones.[87] R_1 = extended methyl-branched carbon chain. R_2 = extended saturated carbon chain. R_3 and R_4 = OH or NH_2

Frankia sp. EuI1c is isolated from the root nodules of *Elaeagnus umbellata*. The genome of this strain has been recently sequenced, revealing a large genome size of 8.8 million base pairs (Mb).[88] Recent studies have attributed its inability to form root nodules to the lack of nitrogen fixation genes within its genome.[116] Coupling this with recent chemotaxonomic, phenotypic, phylogenetic and genomic studies on the strain, the name "*Frankia inefficax* sp." has been proposed.[91] Typical of *Frankia* bacteria, the EuI1c strain contains a high percentage of Guanine-Cytosine (GC) DNA base-pair content. The sequencing also revealed that the EuI1c strain contains 68 P450 genes, which is an unusually high number (Appendix B1). The entire EuI1c strain contains $\geq$ 7,000 protein encoding genes, where P450 genes make up close to 1% of the total gene count. The P450s within *Frankia* sp. EuI1c must play a role in facilitating the biosynthesis and/or metabolism of the various secondary metabolites and natural products the *Frankia* species are known to posses.[87,117]

The P450s from *Frankia* sp. EuI1c will be investigated. Initially the expression levels and protein production and purification will be assessed. After purification of a selection of these P450s, the substrate range will be investigated via spin-state shift and binding affinity assays in order to determine the potential substrate range of these enzymes. The electron transfer partners of these P450s will also be studied, such that *in vitro* and *in vivo* turnovers of the identified substrates can be performed.

Henceforth the P450 enzymes from any gene from *Frankia* sp. EuI1c will be written as FraEuXXXX, while the ferredoxins/ferredoxin reductase enzymes will be written as FdxXXXX/FdrXXXX, where the X represents the gene number the protein is from (e.g.. the P450 encoded by the gene *FrankiaEuI1c_2494* gene will be known as FraEu2494, while the ferredoxin from the *FraEuI1c_2495* gene will be known as Fdx2495). The *Frankia* genes will be in italics (e.g.. *FrankiaEuI1c_2494*).

3.2 Results

3.2.1 Expression and Purification of P450s from *Frankia* sp. EuI1c

The *Frankia* sp. EuI1c strain has 68 different P450 encoding genes within its genome (Appendix B1). A phylogenetic tree of P450s from *Frankia* sp. EuI1c was created to compare with each other and related bacterial P450s from a range of bacteria (Figure 24). While the majority of P450s from *Frankia* sp. EuI1c were primarily related to P450s from the *Frankia* species, a few P450s are related to known P450s from another organisms. For example, FraEu5083 was related to CYP147G1 from *Mycobacterium marinum* (45%), and FraEu5334 was related to CYP150A6, also from *Mycobacterium marinum* (55%). These P450s may provide insight on the potential substrates and functions of the P450s from *Frankia* sp. EuI1c.

The genome sequence of the bacterium was analysed in order to select the P450s to be investigated further. Only those P450s that had a ferredoxin gene clustered within 2 genes to the P450 genes were considered for initial screening (Table 2). These ferredoxins are likely to be electron transfer protein partners to their clustered P450s and should make reconstitution of the enzyme activity easier.

Table 2: The 9 P450 genes found in *Frankia* sp. EuI1c and their coupled ferredoxins. Note that *FraEuI1c_5082* is a ferredoxin/ferredoxin reductase fusion gene.

P450 gene	Ferredoxin gene	Ferredoxin reductase gene
FraEuI1c_5369	*FraEuI1c_5370*	N/A
FraEuI1c_5334	*FraEuI1c_5333*	N/A
FraEuI1c_2494	*FraEuI1c_2495*	N/A
FraEuI1c_1415	*FraEuI1c_1414*	N/A
FraEuI1c_4131	*FraEuI1c_4132*	N/A
FraEuI1c_5083	*FraEuI1c_5082*	*FraEuI1c_5082*
FraEuI1c_5466	*FraEuI1c_5464*	N/A
FraEuI1c_4179	*FraEuI1c_4178*	N/A
FraEuI1c_3226	*FraEuI1c_3227*	*FraEuI1c_3228*

There were 9 P450 genes found to be clustered with a ferredoxin. One of these P450s, *FraEuI1c_3226*, also had an associated ferredoxin reductase gene next to the ferredoxin gene (*FraEuI1c_3228* and *FraEuI1c_3227* respectively). Another P450 encoding gene, *FraEuI1c_5083*, was associated with *FraEu1c_5082* which encodes for a ferredoxin/ferredoxin reductase fusion protein. The genes encoding for the P450s were first analysed for rare codon clusters (RCC). Strains from *Frankia* bacteria are known

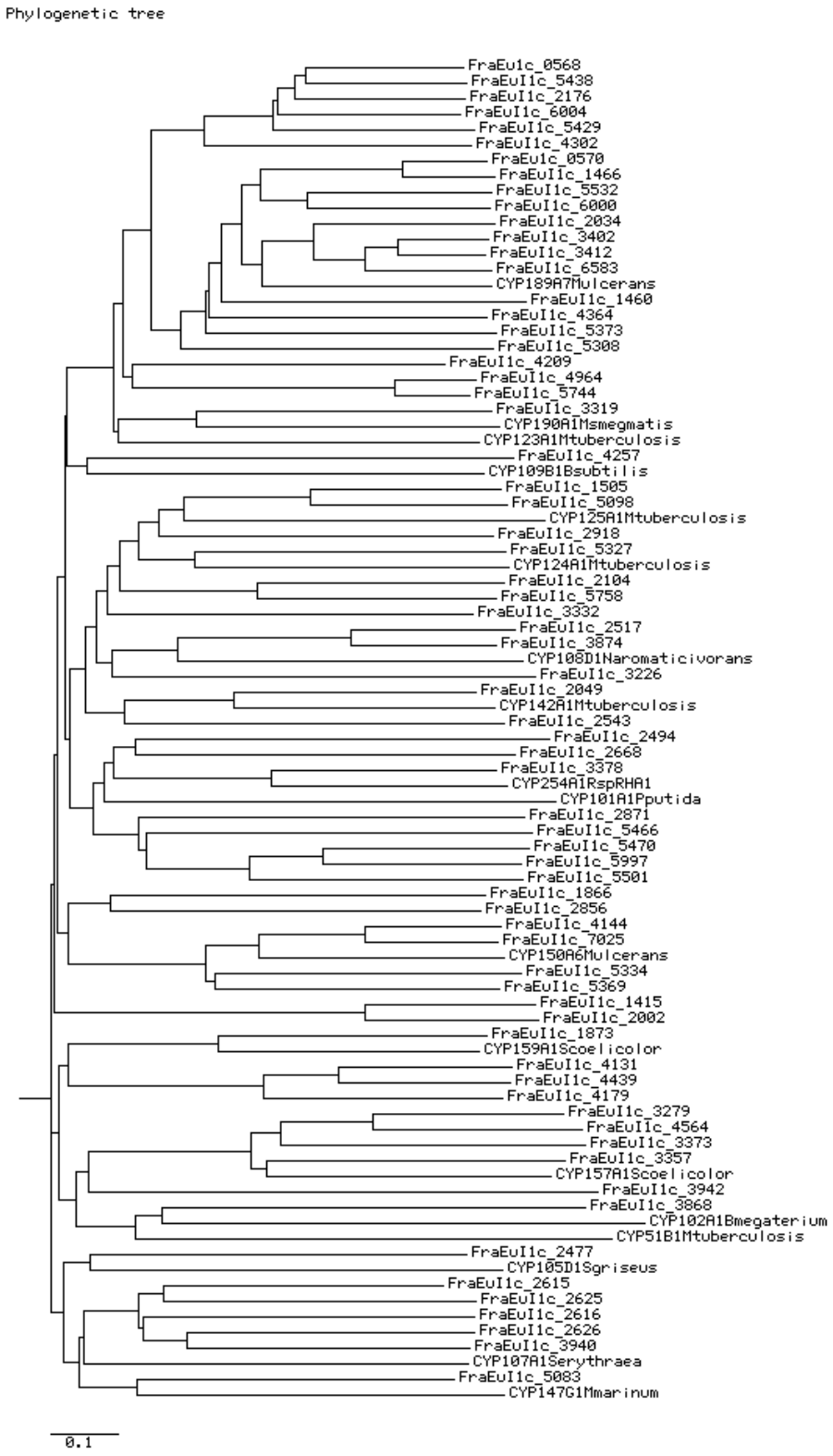


Figure 24: The phylogenetic tree of the 68 P450s from *Frankia* sp. EuI1c and related bacterial P450s. These include CYP1897, (*Mycobacterium sinense*), CYP190A1 (*Mycobacterium smegmatis*), CYP123A, CYP125A1, CYP125A1, CYP142A1, CYP51B1 (*Mycobaterium Tuberculosis*, CYP109B1 (*Bacillus subtilis*), CYP108D1 (*Novosphingobium aromaticivorans*), CYP101A1 (*Pseudomonas putida*), CYP159A1, CYP157A1 (*Streptomyces coelicolor*), CYP102A1 (*Bacillus megaterium*), CYP105D1 (*Streptomyces griseus*), CYP107A1 *Saccharopolyspora erythraea* and CYP147G1, CYP150A6 (*Mycobacterium marinum*).

to have a high guanine-cytosine (GC) content relative to *Escherichia coli* (*E. coli*). For example, the *Frankia* sp. EuI1c genome contains 72.31% GC content compared to ≈ 50% in *E. coli*.[88,118] A species with high GC content contains more G and C ending codons, which may increase the number of rare codons clusters within its genome. A rare codon for *E. coli* is considered here to make up less than 0.4% of the total codon usage (Appendix B2).[119] The percentage of rare codons that the P450 genes possess may reduce the protein production of an *E. coli* system as the tRNA that recognises rarely used codons are present in lower amounts. It is also thought that a gene with a low number of rare codons will be translated more quickly and with higher fidelity than a gene with a high number of rare codons.[120] Many of the P450s from *Frankia* sp. EuI1c had multiple rare codon clusters which may reduce protein expression levels (Figure 25, Table 3, Appendix B1).

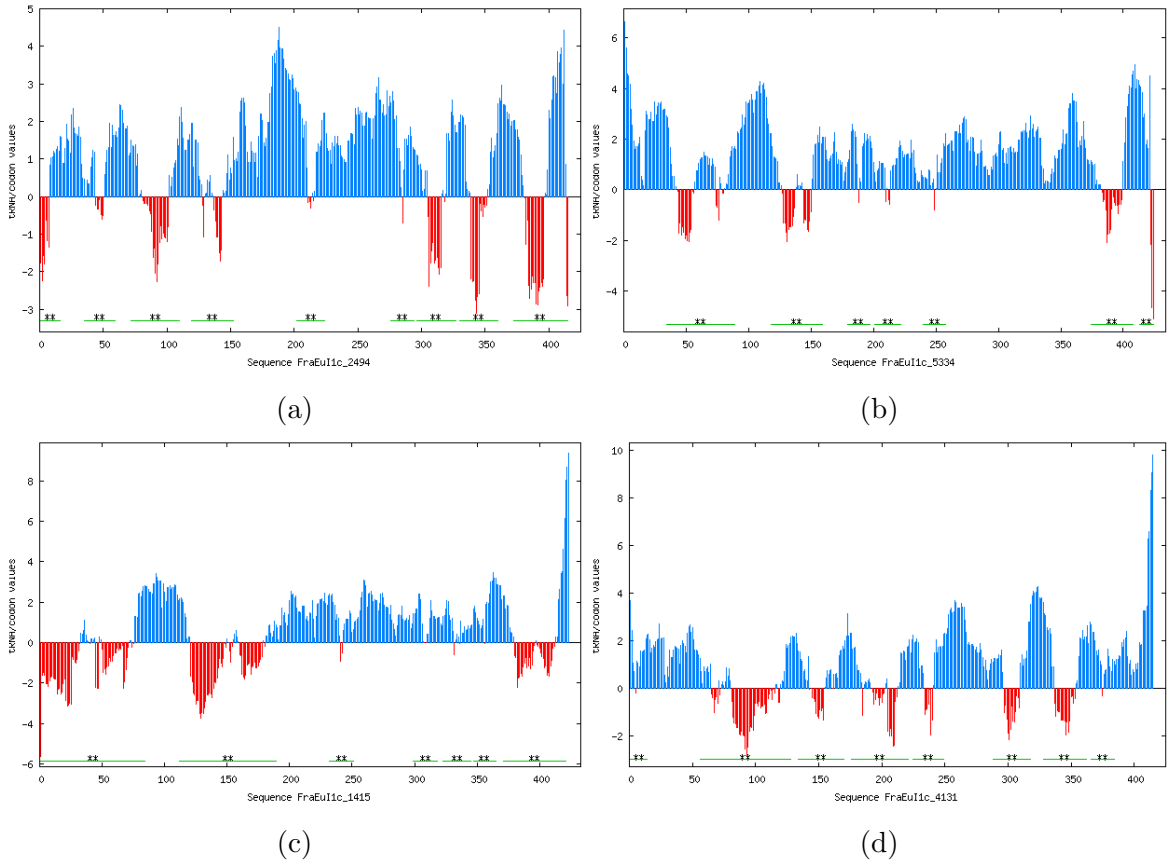


Figure 25: Rare codon clusters for *FraEuI1c_2494* (a), *FraEuI1c_5334* (b), *FraEuI1c_1415* (c) and *FraEuI1c_4131* (d). The intensity of the negative tRNA/codon values in red indicates the number of rare codon cluster centres, while positive values are non-rare codon cluster centres. The green lines indicate rare codon cluster ranges. The RCC identification method was %MinMax, with a window size of 19. The RCC calculation was done on LaTcOm.[97]

The GC content for the 9 selected P450s ranged from 69 - 75% (Table 3). The rare codon percentage was also fairly high, ranging from 8 - 16%, suggesting that these proteins may be difficult to prepare. A small scale expression test of all 9 P450s in

Table 3: Gene size, molecular weight and number of amino acids of P450s from *Frankia* sp. EuI1c. The codons encoding for arginine (AGG, AGA, CGG, CGA), isoleucine (ATA), and proline (CCC) are considered rare codon in *E. coli* (< 0.4% codon usage).

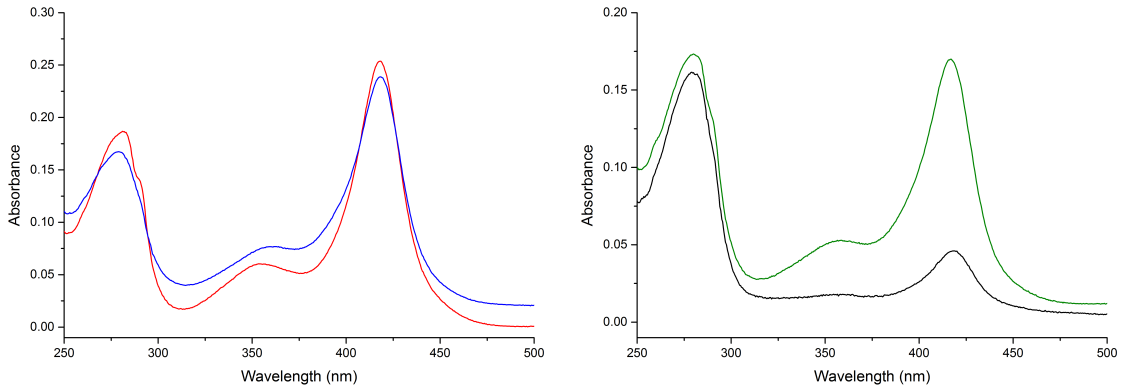
P450	Gene size (bp)	Protein weight (kDa)	GC content (%)	Rare codon (%)
FraEu2494	1248	46.47	72	11
FraEu5334	1275	47.30	70	7
FraEu1415	1272	47.23	68	10
FraEu4131	1248	45.87	70	10
FraEu5369	1380	49.83	70	8
FraEu5083	1272	45.91	75	16
FraEu5466	1215	44.17	72	9
FraEu4179	1260	46.33	72	8
FraEu3226	1263	48.20	69	11

Table 2 was undertaken to observe and identify P450s that were able to be produced at a level appropriate for large scale growth. This enables initial prioritisation of which P450s to scale up growth and purify.

The levels of P450 expression was ranked according to the intensity of the red colour characteristic of the heme-containing enzyme within the cell pellets: FraEu2494 > 5334 > 1415 > 4131 > 5083 ≈ 5369 ≈ 5466 ≈ 4179 ≈ 3226. A small amount of P450 was produced in the *E. coli* producing FraEu5083 and the equivalent *Frankia* P450s, as determined by the beige coloured cell pellets. While the pellet containing FraEu2494 was dark red, the pellet containing FraEu5334 exhibited a lighter shade of red. FraEu1415 and FraEu4131 were even lighter, forming a light pink coloured pellet. The four P450s that showed any red colouration in the respective cell pellets were produced in a scale considered viable for protein production and purification. A SDS gel was prepared to ensure that the protein pellet expressed the desired P450 between 44 - 48 kDa (Appendix B2).

Interestingly, the rare codon cluster analysis did not correlate to the trend of expression of the enzymes. For instance, while *FraEuI1c_5334* had the least number of rare codon clusters, it had a lower expression than *FraEuI1c_2494*. Some genes, such as *FraEuI1c_5083* and *FraEuI1c_4179* showed rare codon clusters similar in intensity and range to that of *FraEuI1c_2494*. However, the expression levels of these P450s were as poor as *FraEuI1c_5369*, which had the highest number of rare codon clusters. Therefore, it cannot be concluded that the appearance of rare codon clusters are indicative of poor expression.

Large scale growth of *E. coli* producing FraEu2494 resulted in a dark red pellet after centrifugation. A large quantity of the soluble protein was collected after purification. The colour of the cell pellets for FraEu5334 and FraEu1415 correlated with the purified



(a) (b)

Figure 26: UV/Vis spectra of (a) FraEu2494 (red) and FraEu5334 (blue) and (b) FraEu1415 (green) and FraEu4131 (black). The A_{420}/A_{280} absorption ratio were: 1.33:1 for FraEu2494, 1.43:1 for FraEu5334, 0.96:1 for FraEu1415 and 0.29:1 for FraEu4131

yields, and these P450s were used in further investigations (Section 3.2.2). While the pellet of FraEu4131 was similar in colour to FraEu1415, only a small amount of soluble P450 protein was purified.

The purified P450s were analysed using UV/Vis spectroscopy to find the ratio between the P450 Soret band at 418 nm and the aromatic amino acid absorption band at 280 nm. As nearly all proteins absorb at 280 nm, this ratio indicates the relative amount of non-P450 protein present after purification compared to P450 protein. As FraEu4131 had a very high 280 nm peak, the A_{418}/A_{280} ratio indicated low purity (Figure 26). Both FraEu2494 and FraEu5334 showed a higher A_{418}/A_{280} absorption ratio, while FraEu1415 had an almost 1:1 peak intensity ratio. The final yields for the P450s from *Frankia* sp. EuI1c followed the results from the expression tests, with FraEu2494 being purified in significantly greater amounts than FraEu4131 (Appendix B3). The preparation of FraEu5334 and FraEu1415 were attempted with the addition of 200 µM of likely substrates (Section 3.2.2) during the protein expression stage. The presence of a substrate may help stabilise the folding of the P450, but these tests did not appear to improve the production of the enzymes.

In order to determine the concentration of the P450s, carbon monoxide (CO) binding assays were carried out. The CO assay reveals the amount of active P450 protein generated compared to the inactive/denatured P420 form. The CO difference assay for FraEu4131 was not attempted due to the small amount of FraEu4131 available.

FraEu2494 and FraEu5334 each showed an almost complete shift ($\approx 95\%$) from the ferric form to its CO-bound ferrous form (Figure 27). The characteristic peaks of the reduced P450 bound to CO at 450 nm is almost at the same intensity as the unbound ferric form at 390 nm. A small shoulder is observed at 420 nm for both the CO bound

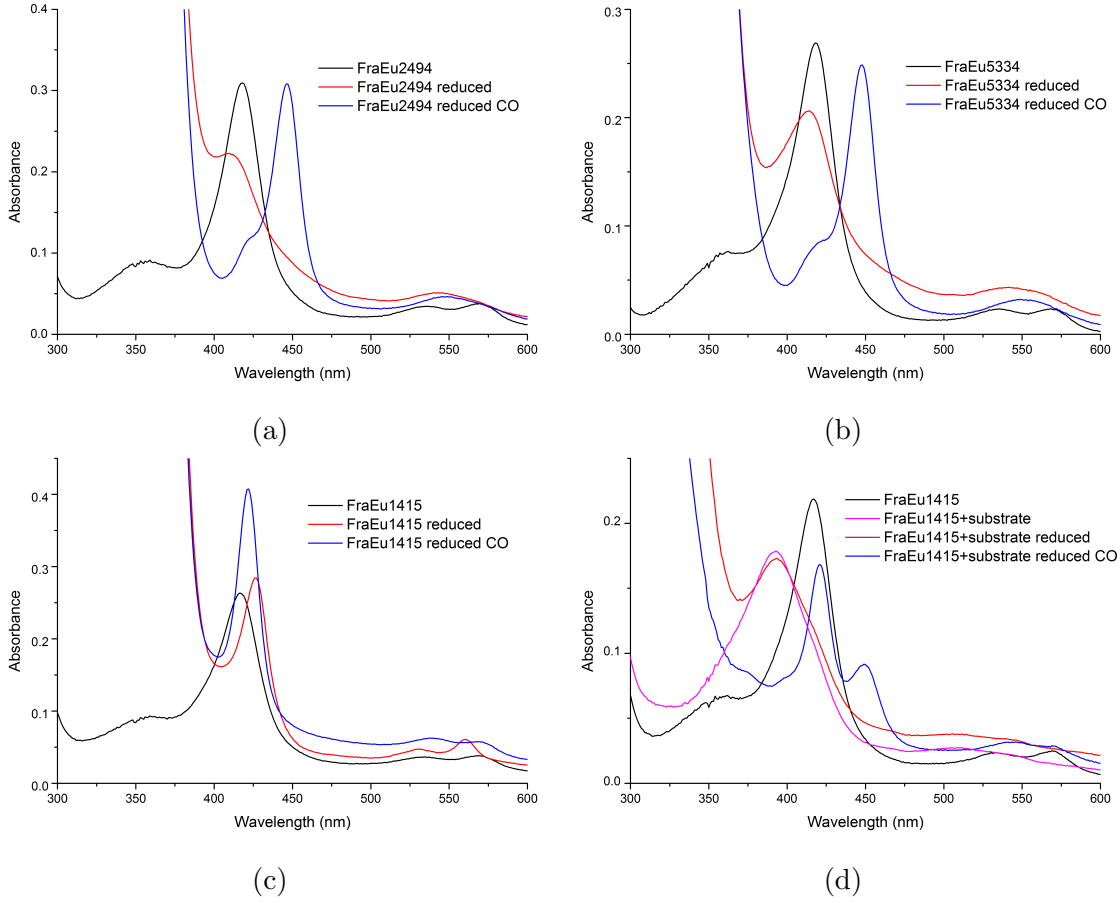


Figure 27: UV/Vis analysis showing the CO-ferrous difference spectra of FraEu2494 (a), FraEu5334 (b), FraEu1415 (c) and FraEu1415 with 4-androstene-3,17-dione substrate (d).

Table 4: The UV/Vis wavelengths absorbance peaks for the CO spectra of FraEu2494, FraEu5334 and FraEu1415.

P450	Wavelength (nm)	Absorbance
FraEu2494	418	0.309
FraEu2494 reduced	410	0.226
FraEu2494 reduced + CO	447	0.308
FraEu5334	418	0.269
FraEu5334 reduced	414	0.206
FraEu5334 reduced + CO	448	0.249
FraEu1415	417	0.256
FraEu1415 reduced	426	0.276
FraEu1415 reduced + CO	422	0.393
FraEu1415 (apo)	417	0.186
FraEu1415+substrate	393	0.179
FraEu1415+substrate reduced	393	0.173
FraEu1415+substrate reduced + CO	421 and 449	0.168 and 0.0917

reduced P450s, and either there was a small amount of soluble protein being denatured by CO, or alternatively it was inactivated from misfolding or from the purification process.

The CO difference assay for FraEu1415 produced a different spectra. After reduction, the P450 spectrum peak shifted to 426 nm instead of the expected 416 nm (Figure 27c). Once CO was bubbled into the mixture, a large $\approx$ 420 nm peak was observed. This inactive "P420" form is characterised by a CO-bound Soret peak near 420 nm, and the blue-shifted 420 nm peak is thought to arise from protonation of the cysteine thiolate, resulting in thio ligation to the heme iron.[121,122] Inactive forms of P450s can be formed by various methods such as treatment with acetone or exposing the P450 to a high pressure environment.[123,124] This suggests that either all of the FraEu1415 protein is in an inactive or denatured form, or the enzyme is sensitive to the method of reduction or addition of CO. It was later found that FraEu1415 could bind steroids very tightly, inducing a full spin-state shift ($\geq$ 95% high-spin, Section 3.2.2.3). An additional CO spectra for FraEu1415 was repeated with the substrate 4-androsterene-3,17-dione added to the enzyme mixture before reduction (Figure 27d). After the addition of CO, a proportion of the spectra had the characteristic 450 nm peak (Figure 27d). This suggests that FraEu1415 is very sensitive and easily denatured by the method used for analysing the reduced CO form.

Through the CO difference assays, the extinction coefficients of FraEu2494 and FraEu5334 in their resting states were calculated (FraEu2494 ϵ_{418} = 130 $mM^{-1}cm^{-1}$ and FraEu5334 ϵ_{418} = 151 $mM^{-1}cm^{-1}$, Section 2.2.4). The concentration of FraEu2494 was approximately 5 mg mL^{-1}, with an excess of 30 mL collected and purified. The amount of FraEu2494 protein expressed per litre of cell culture was approximately 60 mg L^{-1}. The concentration for purified FraEu5334 was found to be approximately 2.2 mg mL^{-1}, with 14.8 mg of protein produced per litre of cell culture. The concentration of FraEu1415 was not determined successfully due to the sensitivity of the protein, but it is expected to be lower than FraEu5334 due to its reduced intensity in colour (both visually and in UV/Vis scans, Appendix B3). While the concentration of FraEu4131 was also not determined, it is reasonable to assume it has significantly lower P450 levels than the FraEu1415. These results have shown successful expression and purification of the P450s FraEu2494, FraEu5334, FraEu1415 and, to a certain extent, FraEu4131.

3.2.2 Substrate Binding Assays for P450s from *Frankia* sp. EuI1c

The first step of the P450 catalytic cycle involves the displacement of the distal water ligand by an appropriate substrate. Therefore a substrate that can bind tightly within the active site of the enzyme and displace the water ligand will be able to start the P450 catalytic cycle. When the distal water ligand is displaced, a low spin-state to high spin-state shift occurs within the heme active site, with a blue-shifted absorbance peak from 418 nm to 390 nm. The proportion of the peak absorbance shift determines the percentage of P450 in the high spin-state (Experimental 2.3.1).

The binding affinity was determined by adding small aliquots of substrate in order to observe the gradual shift from 418 nm to 390 nm. The dissociation constant, K_{d} could be obtained by fitting the peak-to-trough difference against concentration to a hyperbolic function (See Experimental 2.3.2).

In order to determine which substrate to test with each of the P450s from *Frankia* sp. EuI1c, the amino acid sequence of the different P450s were initially compared to P450s of known function. A BLAST search was undertaken for each of the P450s to find others with close sequence similarities (Table 5, Table7 and Table 9).[125] The substrates of the P450s with the highest similarity to FraEu2494, FraEu5334 and FraEu1415 were then tested as a starting point for potential substrates. Other classes of molecules which were structurally similar or unrelated to these substrates were also tested. These included cyclic mono- and sesqui-terpinoids, short and medium chain (C_6-C_{12}) linear alkanes, benzene and polycyclic aromatic hydrocarbons, fatty acids, sesquiterpenes, substituted phenols, substituted benzoic acids and steroidal compounds.

3.2.2.1 Investigation of the Substrate Binding Range of FraEu2494

FraEu2494 is not closely related to any structurally characterised P450 (Figure 24). It is most closely related to uncharacterised P450s from *Frankia*, sharing up to 53% amino acid sequence identity. It showed low sequence similarity to CYP101A1 (P450cam), CYP176A1 (P450cin) and CYP108A1 (P450terp) (28 - 34% sequence similarity, Figure 5), which are documented to hydroxylate camphor,[126] 1,8-cineole[127] and α-terpineol[128] respectively. While FraEu2494 was not able to bind to any of these terpenoids, it provided a starting point for screening a range of similar substrates for binding affinity. FraEu2494 also has a low similarity to CYP109B1 from *Bacillus subtilis* (25%, Figure 5). The CYP109 family of P450s are known to regioselectively hydroxylate a range of compounds such as norisoprenoids, fatty acids and steroids, and these classes of substrates were tested with FraEu2494 (Figure 28).[129–133]

Table 5: The amino acid sequence similarity of FraEu2494 found through a BLAST sequence search.

P450	Source organism	Total score	Sequence identity	Known substrates	PDB
CYP101D1	*Novosphingobium aromaticivorans*	173	34%	camphor	3LXH
CYP127A1	*Sinorhizobium* sp. NGR234	178	31%	-	-
CYP143	*Mycobacterium tuberculosis*	198	35%	-	-
CYP101A1	*Pseudomonas putida*	164	29%	Camphor	2CPD
CYP176A1	*Citrobacter braakii*	157	29%	1,8-Cineole	1T2B
CYP108A1	*Pseudomonas* sp.	156	28%	α-Terpineol	1CPT
CYP109B1	*Bacillus subtilis*	134	25%	Fatty acids	4RM4

A range of compounds including norisoprenoids, monoterpenoids, sesquiterpenes, acetates, steroids and fatty acids were screened with FraEu2494 (Appendix B3). While a few of the monoterpenoids, acetates and sesquiterpenes induced spin-state shifts, none was observed for any of the fatty acids and steroids. The substrates that exhibited binding to FraEu2494, as measured with a shift of the spin-state, are outlined in Table 6.

The norisoprenoids, α-ionone, β-ionol and β-damascone all induced similar spin-state shifts of between 40 - 50% in FraEu2494 (Appendix B4, Table 6). They also exhibited similar binding affinities to the enzyme as measured by the dissociation constant (K_{d} = 7.6, 12 and 12 µM, respectively, Table 6, Appendix B7). The position of the oxygen containing functional group in α-ionone and β-ionol differs by two carbon atoms compared to β-damascone, however this did not appear to have a significant effect on the binding affinity to the P450 active site. However, β-ionone did not induce any spin-state shift in FraEu2494, despite being closely related to and having almost identical structures to β-ionol, α-ionone and β-damascone.

Table 6: Binding data for FraEu2494. Percentage high spin is denoted as % HS. The dissociation constant is denoted as K_d. The dissociation constant is given as mean ± S.D. with $n \geq 3$.

Substrate	% HS	K_d (µM)
4-*tert* butylcyclohexyl acetate	35	11 ± 0.6
α-ionone	50	7.6 ± 0.4
ambroxide	60	0.7 ± 0.07
β-ionol	40	12 ± 1.8
β-damascone	50	12 ± 2
bornyl acetate	40	8.4 ± 0.7
fenchyl acetate	35	8.6 ± 0.1
tetralin	30	52 ± 4
trans-decahydronapthalene	60	4.9 ± 0.5
sclareolide[a]	50	–

[a]Binding constant not determined due to insolubility of substrate preventing full analysis.

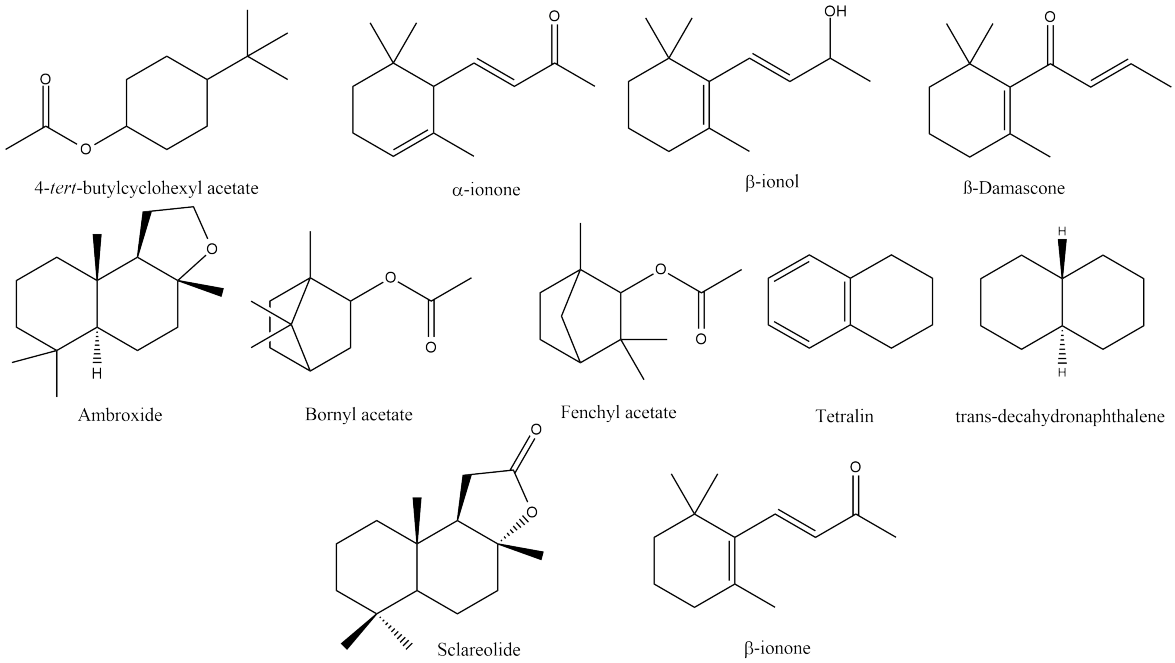


Figure 28: Substrates tested with FraEu2494. β-Ionone did not induce an observable spin-state shift and is shown as a reference.

A number of monoterpenoid acetates were screened with FraEu2494. Bornyl acetate, fenchyl acetate and 4-*tert*-butylcyclohexyl acetate all possess a cyclohexane group with oxygen-containing carbon chain substituents similar to the norisoprenoids (Figure 28). The monoterpenoid acetates each contain an ester group in contrast to the ketone or alcohol moieties found in norisoprenoids. These substrates induced a slightly lower

spin-state shift in FraEu2494 (35 - 40%) compared to α-ionone and β-damascone. The binding affinity was similar to the norisoprenoids (K_{d} = 8.4, 8.6 and 11 µM respectively, Appendix B7 and Figure 29).

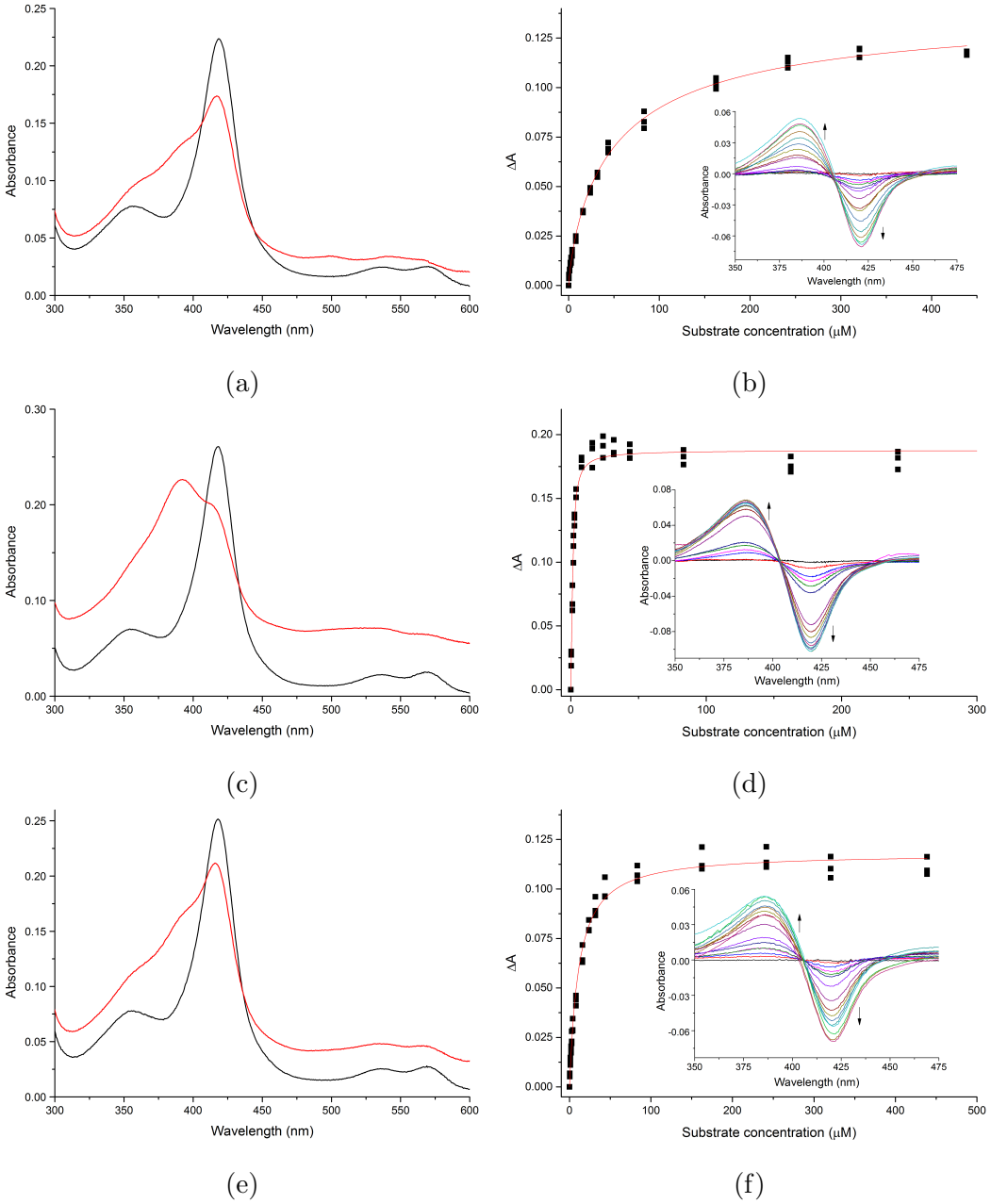


Figure 29: The respective spin-state shifts and dissociation constant analyses of FraEu2494 with tetralin (a) and (b), ambroxide (c) and (d), and 4-*tert*-butylcyclohexyl acetate (e) and (f).

The norisoprenoids and monoterpenoid acetates substrates which bind to FraEu2494 possess one or more cyclic rings. As such, *trans*-decahydronapthalene (also known as decalin), a simple unsaturated bicyclic compound, was also screened for binding. It induced a high spin-state shift (60%) when added to FraEu2494 and also had high binding affinity (K_{d} = 4.9 µM, Appendix B4). However the related compound tetralin, which contains an aromatic ring, induced only a 30% spin shift and the binding affinity was significantly lower (K_{d} = 52 µM, Figure 29a). Naphthalene, a chemical comprised

of two interconnected unsaturated benzene rings, induced a negligible spin-state shift (Appendix B5). This suggests that saturated cyclic rings are a core component in the natural substrates of FraEu2494.

Larger compounds with multiple cyclic rings such as ambroxide were also screened for binding affinity to FraEu2494 (Figure 28). Ambroxide contains a core ring structure consisting of two conjoining cyclohexane rings similar to *trans*-decahydronaphthalene, and also a 5 membered ring carbon ring containing oxygen (Figure 28). This polycyclic compound induced a 60% spin-state shift of the heme iron upon addition to the P450, and bound with the highest affinity to the enzyme ($K_{\mathrm{d}} = 0.7$ µM, Figure 29c). A 50% spin-state shift in FraEu2494 was also observed when sclareolide was added. Sclareolide is structurally related to ambroxide (Figure 28, Appendix B5), but possesses an extra ketone group. Therefore the presence of the carbonyl group did not appear to significantly effect the spin-state shift of FraEu2494. The binding constant for sclareolide was unable to be determined due to the low solubility of the substrate, but it is expected to have a lower affinity than ambroxide due to a lower response to substrate addition. Overall, these saturated cyclic ring containing compounds are promising substrates for FraEu2494.

3.2.2.2 Investigation of the Substrate Binding Range of FraEu5334

FraEu5334 has a high sequence similarity (83 - 88%) to other uncharacterised Frankia P450s. It is also has a 55 - 64% sequence similarity with many uncharacterised P450s from *Mycobacterium marinum* and *Blastococcus*, including members of the CYP150 P450 family. For example, CYP150A6 from *Mycobacterium marinum* species shares 50% sequence identity (Figure 24, Table 7). It is also shares 55% amino acid sequence similarity to CYP150A5 from *Mycobacterium marinum* (Table 7). The closest characterised P450s, CYP108A1 (P450terp) and CYP109B1, have a lower sequence similarity at 32%. The CYP150 family has been reported to oxidise polycyclic aromatic hydrocarbons. Therefore FraEu5334 was screened with a range of aromatic hydrocarbons, fatty acids, monoterpenoids and steroids for binding affinity (Figure 30, Appendix B4).[134]

Table 7: The amino acid sequence similarity of FraEu5334 found through a BLAST sequence search.

P450	Source organism	Total score	Sequence identity	Known substrates	PDB
CYP150A5	*Mycobacterium marinum*	469	55%	-	-
CYP150A6	*Mycobacterium marinum*	442	50%	Aromatic hydrocarbons	-
CYP109B1	*Bacilus subtilis*	155	32%	Fatty acids	4RM4
CYP108A1	*Pseudomas sp.*	152	32%	α-Terpineol	1CPT

Substrates such as the steroid testosterone and fatty acids also did not induce any

spin-state shift. FraEu5334 only shares 27% sequence identity with FraEu2494, and many of the compounds screened for FraEu2494 did not show any affinity towards FraEu5334 (Appendix B4). For instance, substrates that had high affinity for FraEu2494 such as ambroxide and nootkatone did not induce any spin-state shifts for FraEu5334.

A number of norisoprenoids were able to induce spin-state shifts in FraEu5334. Both β-damascone and β-ionol, which induced a spin-state shift of 40%, and showed moderate binding affinity ($K_{\mathrm{d}} = 74$ and 107 µM, respectively). β-Ionone generated a spin-state shift of 40% (Figure 31) and exhibited a similar binding affinity towards the active site of FraEu5334 ($K_{\mathrm{d}} = 107$ µM). These three related compounds are very similar in structure and differ only slightly in the position of their functional groups, such as the ketone moiety in β-ionone positioned two carbons away from that in β-damascone. Although these norisoprenoids induced the highest spin-state shifts compared to the all other substrates tested here, it was still relatively low. This, coupled with the relatively low affinity of these norisoprenoids suggests that they do not fit optimally inside the active site.

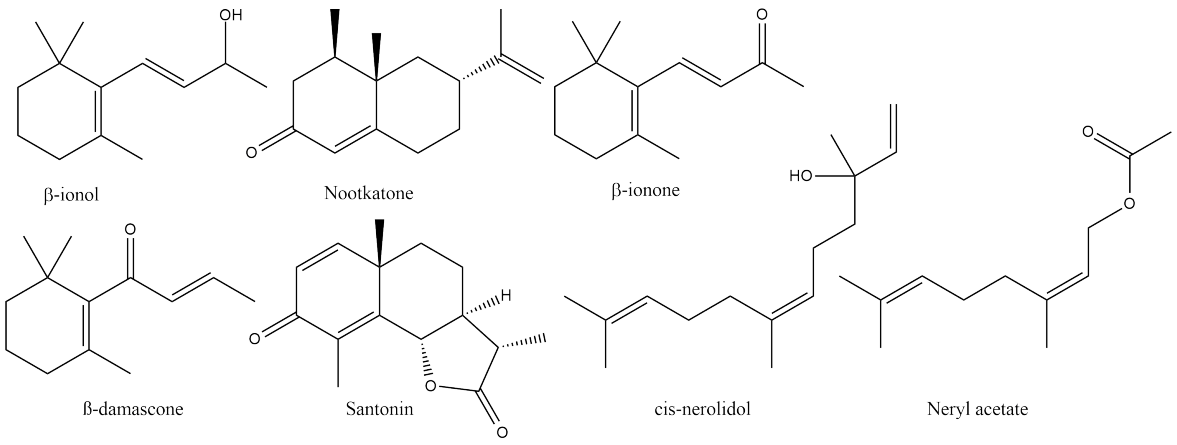


Figure 30: Substrates of FraEu5334

Table 8: Binding data for FraEu5334. Percentage high spin is denoted as % HS. The dissociation constant is denoted as K_{d}. The dissociation constant is given as mean ± S.D. with $n \geq 3$.

Substrate	% HS	K_{d} (µM)
α-santonin	50	120 ± 41
β-damascone	40	74 ± 13
β-ionol	40	107 ± 11
β-ionone	40	107 ± 27
cis-nerolidol	30	95 ± 7
neryl acetate	30	123 ± 21
nootkatone	25	52 ± 4

Larger substrates with additional cyclic rings such as nootkatone were screened for binding affinity against FraEu5334 as it showed structural similarities with the previously tested β-ionone and β-damascone. Nootkatone induced a smaller Soret band shift of approximately 25%. Despite producing the smallest shift out of the substrates reported in Table 8, it exhibited the tightest binding to the active site of the P450 ($K_{\mathrm{d}} = 52$ µM, Appendix B9). α-Santonin is similar to nootkatone and has multiple rings as well as a ketone moiety within its structure. This compound induced a 390 nm shift of 50% (Figure 31g) which was higher than any other substrate reported. However, its binding affinity to the active site to FraEu5334 was low ($K_{\mathrm{d}} = 120$ µM, Figure 31h). The low binding affinity may be due to its multiple ring structure being too bulky resulting in steric clashes with the active site amino acids. Other large multiple ring structure compounds such as sclareolide and ambroxide were also tested for affinity to FraEu5334, but did not produce any spin-state shift (Appendix B10).

Long carbon chain compounds with hydroxyl or ketone groups were also screened for binding affinity. The sesquiterpene *cis*-nerolidol binds to FraEu5334 with an affinity similar to the norisoprenoids ($K_{\mathrm{d}} = 111$ µM, Appendix B9). However it induced a lower spin-state shift at around 30% (Appendix B8). Neryl acetate, which shares structural features to *cis*-nerolidol, induced a similar spin-state shift (30%) and binding affinity ($K_{\mathrm{d}} = 123$ µM, Appendix B9).

As the active site of FraEu5334 was able to bind a wide range of compounds including norisoprenoids, bulky polycyclic substrates and long linear carbon chains, the substrate range of FraEu5334 appears to be fairly wide. In particular, norisoprenoids bound to both FraEu2494 and FraEu5334, although the binding affinities for all substrates screened with FraEu5334 were lower compared to those found for FraEu2494. These results show that while potential substrates have been determined, further screening and investigation is required to obtain the full substrate range of these enzymes.

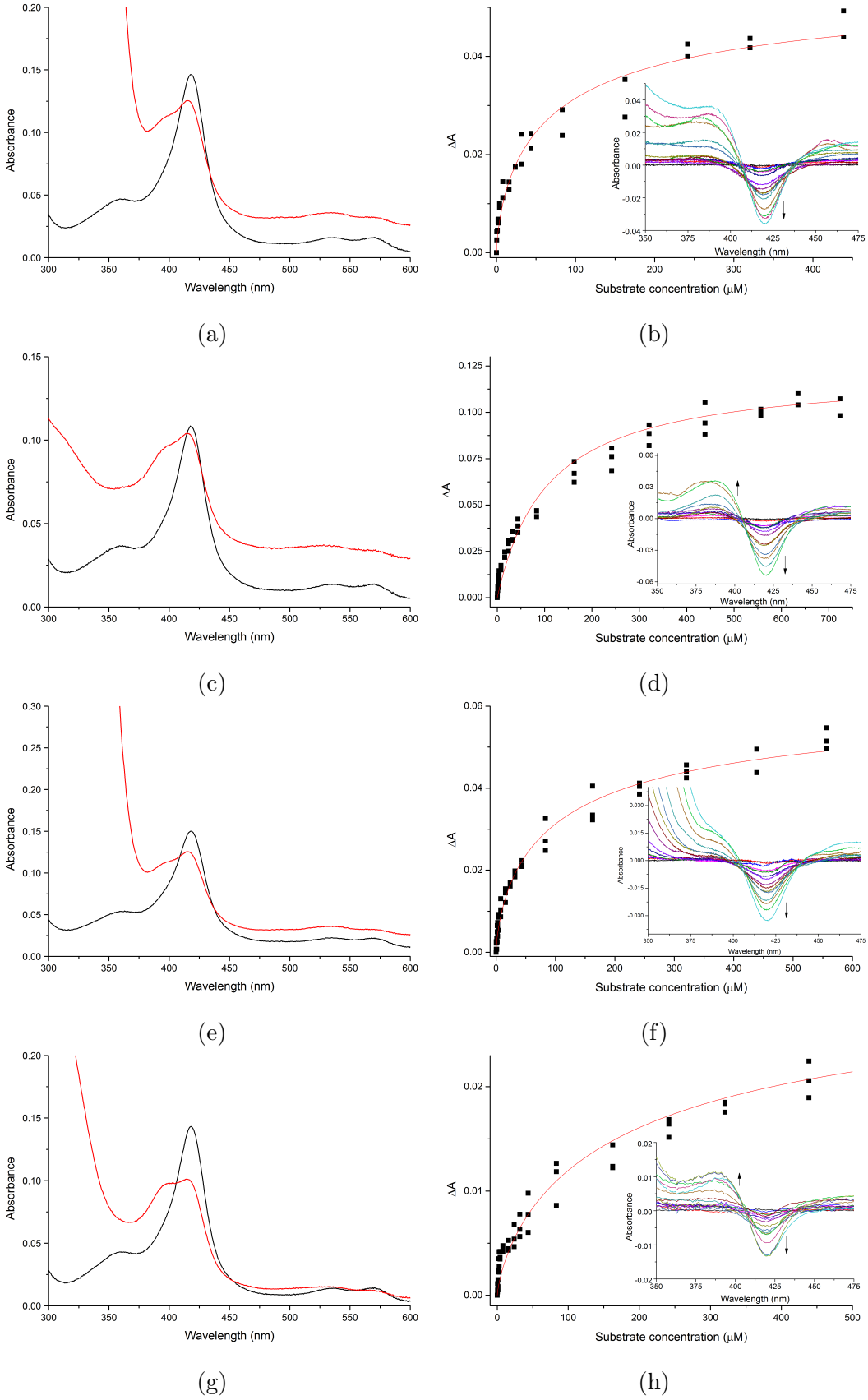


Figure 31: The respective spin-state shifts and dissociation constant analyses of FraEu5334 with β-damascone (a) and (b), β-ionol (c) and (d), β-ionone (e) and (f) and santonin (g) and (h). Due to substrate absorbance interference at lower wavelengths, the dissociation constants for β-damascone and β-ionone [(b) and (f)] were calculated with the difference between the trough at 420 nm and the wavelength at 600 nm (not shown here).

3.2.2.3 Investigation of the Substrate Binding Range of FraEu1415

FraEu1415 only shares high amino acid sequence similarity to uncharacterised P450s; the highest similarity was to another P450 from *Frankia* sp. EuI1c at 71%. It is also closely related to other uncharacterised P450s from *Blastococcus*, with sequence similarities ranging from 62% to 65%. However, FraEu1415 was not closely related to any characterised P450, with the majority of the closest related P450s sharing $\leq 30\%$ sequence similarity (Table 9). For instance, CYP106A2, which has been reported to regioselectively hydroxylate steroids at the 15β position, shares only 26% of its sequence with FraEu1415.[135]

Table 9: The amino acid sequence similarity of FraEu1415 found through a BLAST sequence search.

P450	Source organism	Total score	Sequence identity	Known substrates	PDB
CYP167A1	*Sorangium cellulosum*	190	30%	Epothilone	1PKF
CYP105A1	*Streptomyces griseolus*	126	28%	Vitamin D3	1PKF
CYP107H1	*Bacillus subtilis*	152	27%	Fatty acids	3EJD
CYP106A2	*Bacillus megaterium*	134	26%	Steroids	4YT3

A wide range of chemicals including norisoprenoids, sesquiterpenoids, fatty acids, and polycylic aromatics were screened for binding affinity with FraEu1415. While the majority of these classes of chemicals did not induce any noticeable spin-state shifts (Appendix B5), the steroid testosterone did. Testosterone bound tightly to FraEu1415 and induced a complete spin-state shift ($K_{\mathrm{d}} = 7.6$ µM, $\geq 95\%$, Table 10, Figure 33b). The complete class I spin-state shift from 420 nm to 390 nm means that testosterone is positioned close enough to the heme centre to completely displace the distal water ligand. The low dissociation constant (K_{d}) indicates that the substrate is held tightly and is a good fit within the active sight of FraEu1415. As a result, other steroids and steroid-like compounds were also tested for binding affinity towards FraEu1415.

Steroids with structural similarities to testosterone were also able to induce an almost complete shift ($\geq 95\%$), suggesting that they bind to the active site in a similar conformation. These include progesterone ($K_{\mathrm{d}} = 2.9$ µM, Figure 33d), 4-androstene-3,17-dione ($K_{\mathrm{d}} = 8.3$ µM, Appendix B11) and stanolone ($K_{\mathrm{d}} = 24$ µM, Appendix B11), which all exhibited high binding affinity to the enzyme. These compounds have the same general steroid ring structure consisting of three cyclohexane rings and one cyclopentane ring (Figure 32). In particular, each of these steroids contain a oxygen containing functional group on the carbon 3 (C-3) and carbon 17 (C-17) positions respectively. Although the structures of these steroids are very similar, the different binding affinities, such as the weaker affinity of stanolone towards FraEu1415, suggests that small structural differences play a large role in the mode of binding.

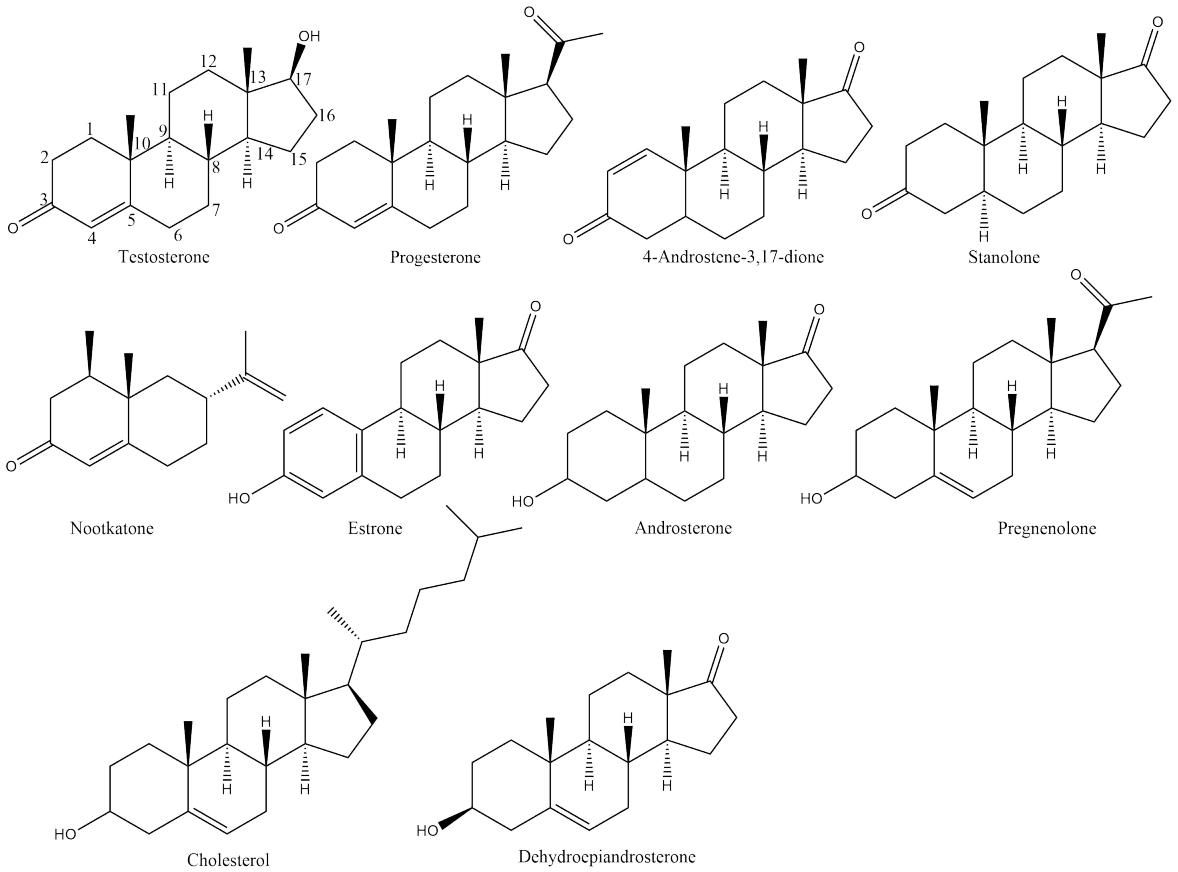


Figure 32: Substrates of FraEu1415. The parent ring structure of a steroid have been labelled on testosterone.

Table 10: Binding data for FraEu1415. Percentage high spin is denoted as % HS. The dissociation constant is denoted as K_{d}. The dissociation constant is given as mean ± S.D. with $n \geq 3$.

Substrate	% HS	K_{d} (µM)
testosterone	≥ 95	7.6 ± 0.2
progesterone	≥ 95	2.9 ± 0.1
4-androstene-3,17-dione	≥ 95	8.3 ± 1.6
stanolone	≥ 95	24 ± 2
androsterone	≥ 95	204 ± 11
estrone	50	2.7 ± 0.1
pregnenolone	45	3.7 ± 0.4
dehydroepiandrosterone	0	-
cholesterol	0	-
nootkatone	40	213 ± 22

Androsterone, a steroid similar to 4-androstene-3,17-dione, also induced a $\geq 95\%$ spin shift when added to FraEu1415. The binding affinity ($K_{\mathrm{d}} = 204$ µM, 10, Figure 33f) was several magnitudes lower than for 4-androstene-3,17-dione. However, the structurally

similar estrone and pregnenolone exhibited high binding affinity to FraEu1415. Unlike androsterone, both estrone (50% shift, K_d = 2.7 µM) and pregnenolone (45% K_d = 3.7 µM) exhibited tight binding but induced a lower spin-state shift (Appendix B12). These compounds bound tightly to the substrate binding pocket of FraEu1415, but the moderate spin-state shift suggests that they are not positioned close enough to the iron-heme centre to completely displace the distal water ligand. Androsterone, estrone and pregnenolone are all sterols, characterised by the hydroxyl group at the C-3 position. When compared to the steroids such as 4-androstene-3,17-dione, progesterone and testosterone which have a ketone at this position and a different double bond position, the binding affinity to FraEu1415 and/or the induced spin-state shift were significantly decreased. For example, both progesterone and pregnenolone have very similar structures, differing only in their double bond positions and their substituent at C-3. Although their binding affinity was fairly similar, pregnenolone induced a significantly lower shift than progesterone (45% compared to $\geq$ 95%). As such having a hydroxyl substituent at the C-3 position resulted in a lower spin-state shift or a lower binding affinity compared to having a carbonyl group at the same position.

Two other steroids, cholesterol and dehydroepiandrosterone, were also screened for binding affinity towards FraEu1415, however these steroids failed to induce any spin-state shift (Appendix B11). Both contain a C-3 hydroxyl group and an altered double bond position, although the low affinity for cholesterol could be due to its long branched carbon chain at the C-17 position (Figure 32).

The only non-steroid compound tested that induced a significant spin-state shift was nootkatone. Other substrates structurally similar to nootkatone such as ambroxide and sclareolide failed to induce any spin-state shift. Nootkatone shares structural similarities with steroids, as it contains the same A and B cyclohexane rings contained in the steroid structures (Figure 32 and 42). The spin-state shift and binding affinity was significantly lower than the steroids (40%, K_d = 213 µM, Table 10, Appendix B11), which may be due to the smaller size of nootkatone.

In summary, tight binding is observed for steroids, particularly those which possess a ketone at the C-3 position. The binding affinity of these substrates were greater than the substrates screened for both FraEu2494 and FraEu5334. Combined with the large spin-state shifts induced by these substrates in FraEu1415, it is inferred that the likely physiological substrate of FraEu1415 is based on a steroid structure.

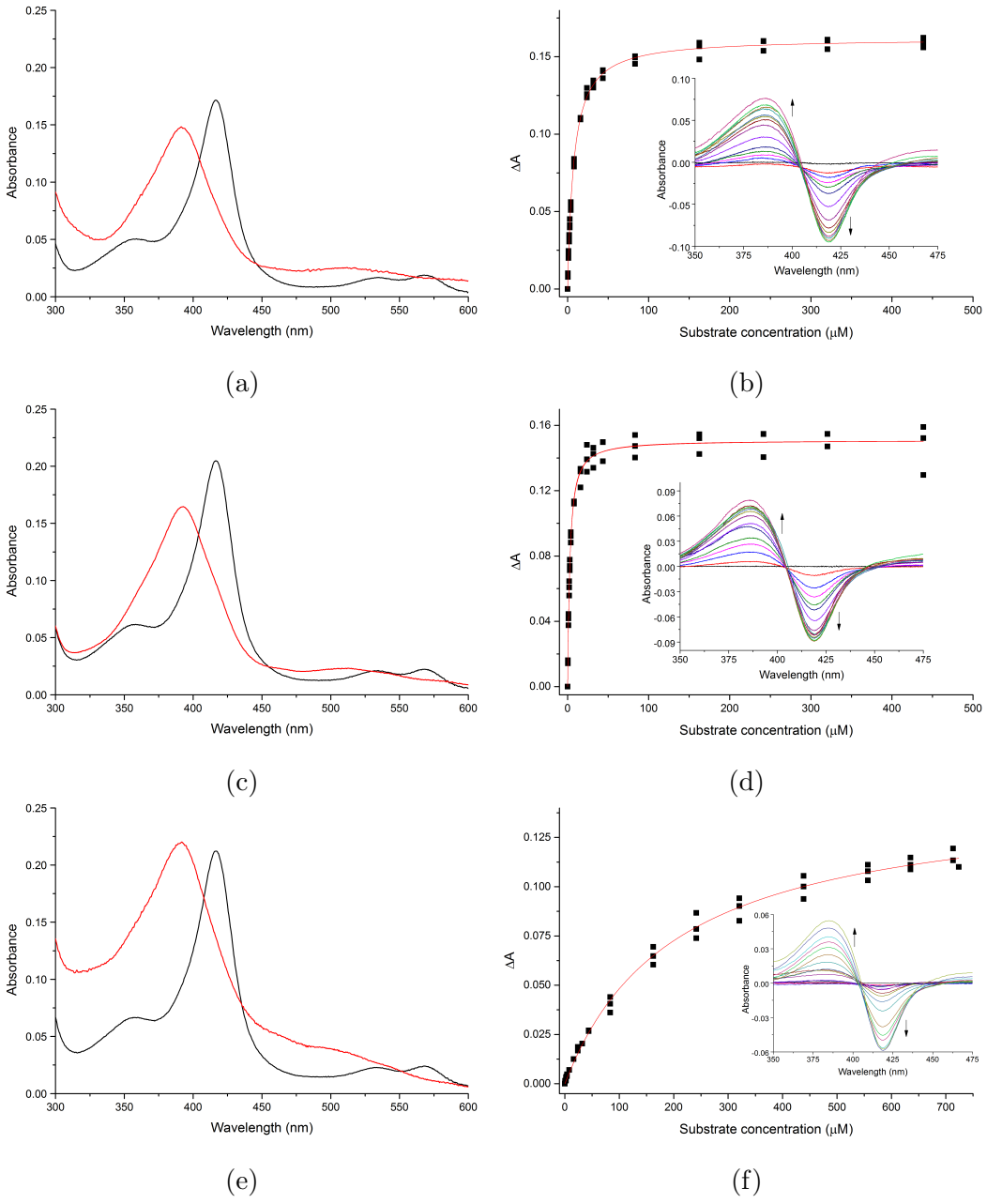


Figure 33: The respective spin-state shifts and dissociation constant analyses of FraEu1415 with testosterone (a) and (b), progesterone (c) and (d) and androsterone (e) and (f).

3.2.3 Crystallography of P450s from *Frankia* sp. EuI1c

In order to further characterise the P450s from *Frankia* sp. EuI1c, we attempted to crystallise FraEu2494 in order to solve its crystal structure. FraEu2494 was screened with Hampton Research Crystal Screen Index and PEG/ION in a sitting-drop vapour-diffusion method. The crystal conditions in the 72nd well of the INDEX crystal screen (0.2 M Sodium chloride, 0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3,350) facilitated the formation of large crystals after 1 week of growth (Figure 34). Attempts to optimise the crystal conditions by altering the concentrations of sodium chloride, HEPES and pH levels failed to reproduce similar sized crystals.

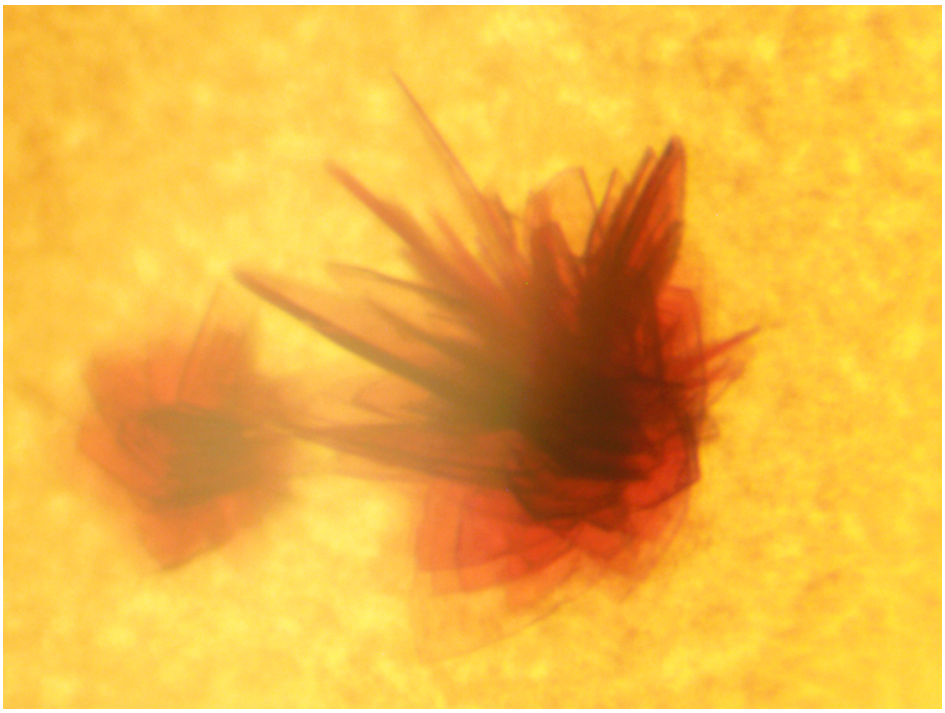


Figure 34: Crystal growth of FraEu2494 at 45x magnification

In order to create larger crystals, small amounts of the known substrate α-ionone and bornyl acetate (100 µM) were added to separate P450 solutions before crystal growth. A substrate bound P450 should be more stabilised compared to the apo protein, which hopefully allow further crystal growth. However, these substrate bound P450s generated similar or smaller and thinner crystals. This could be attributed to the majority of the substrates having relatively low solubility in the P450 buffer.

The majority of FraEu2494 crystals sheets were clustered together from a single nucleation point. The mounting of FraEu2494 crystal was difficult, and only a thin sheet of the crystal was obtained for X-ray diffraction. The X-ray source (0.9537 Å) from the Australian synchrotron MX1 beamline was unable to generate a diffraction pattern from this crystal. Although these crystals were too thin for X-ray diffraction, further optimisation of the crystal conditions should result in forming a crystal structure of this P450 protein. This could be used as a model to generate other crystal structures of the other P450s from *Frankia* sp. EuI1c, which would help help determine and understand the substrate specificity.

3.2.4 Expression and Purification of Electron Transfer Partners from *Frankia* sp. EuI1c

The P450 catalytic cycle requires electrons and protons to be transferred to the heme-active site. In the majority of bacterial P450 systems, both a ferredoxin and a ferredoxin reductase are required for the electron transfer from NAD(P)H to the P450 (Section 1.1). Purifying the ferredoxins and ferredoxin reductases for the P450s from *Frankia* sp. EuI1c will enable the P450 systems to be be analysed *in vitro*. Of the 68 P450 genes, 9 of them are coupled with a ferredoxin gene. Furthermore only 2 out of the 9 are clustered with a ferredoxin reductase gene. The two ferredoxin reductases are Fdr3228, which is co-located with the [2Fe-2S] ferredoxin Fdx3227, and Fdx/Fdr5082, which is a ferredoxin/ferredoxin reductase fusion protein. The remaining ferredoxin genes encode amino acids with sequence similarities to [3Fe-4S] and [4Fe-4S] clusters. These types of iron-sulfur ferredoxins generally contain sequence motifs of **CXXG/AXXC**$(X)_n$**CP** and **CXXCXXC**$(X)_n$**CP**, respectively.[136] The cysteine residues (C) in this sequence motif coordinates either the four or three irons in the respective [4Fe-4S] and [3Fe-4S] iron-sulfur clusters. All the potential [3Fe-4S] ferredoxins in *Frankia* sp. EuI1c have non-standard residues at the "?" position in **CXX?XXC**$(X)_n$**CP** (Figure 35). The non-standard residue is histidine for Fdx2495, serine for Fdx5333 and threonine for Fdx1414 (Table 11).

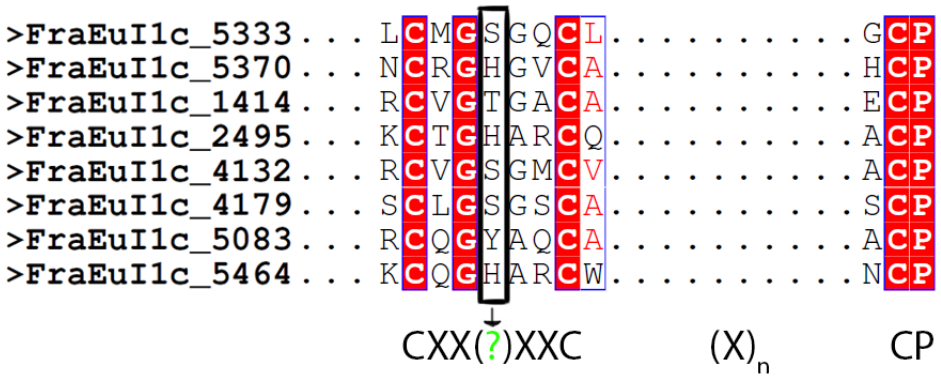


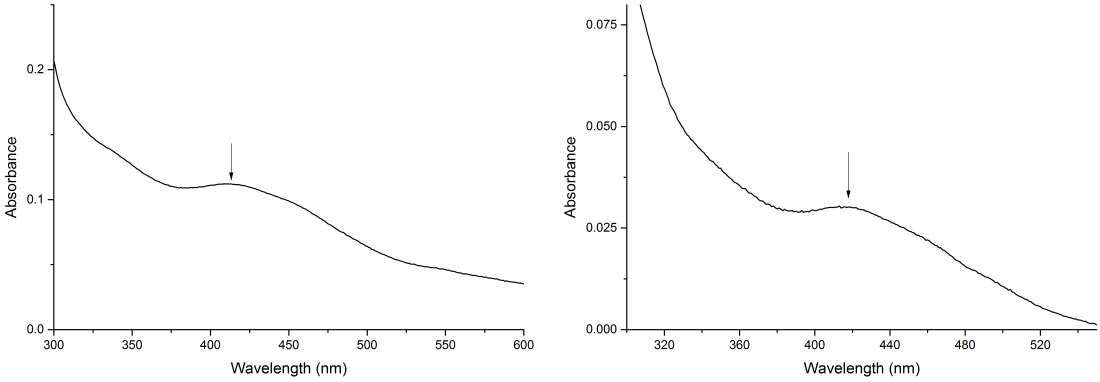
Figure 35: The sequences of the 8 suspected [3Fe-4S] ferredoxins of *Frankia* EuI1c at the cluster binding motif. The remaining coupled ferredoxin gene (*FraEuI1c_3227*) was not included as it has the sequence motif of a [2Fe-2S] cluster.

The ferredoxins genes of *FraEuI1c_2495*, *FraEuI1c_5333* and *FraEuI1c_1414* are clustered next to the P450s investigated previously. These electron transfer partner proteins were produced as outlined in the Experimental section 2.2.3. The codon-optimised ferredoxin genes were cloned in a pETDuet vector and subsequently transformed into BL21 (DE3) competent *E. coli* cells. After growing and inducing the cells, the brown

ferredoxin could be extracted and purified (Experimental section 2.2.3, Appendix B14). These ferredoxins exhibited broad characteristic peaks with a maximum absorbance at 414 - 418 nm in the UV/Vis spectra (Figure 36).

Table 11: Molecular weight, gene size, number amino acids and non-standard residue of ferredoxins from *Frankia* sp. EuI1c

Ferredoxin	Gene size (bp)	Protein weight (kDa)	Cluster motif	Non-standard residue
Fdx2495	216	7.32	[3/4Fe-4S]	Histidine
Fdx5333	195	6.67	[3Fe-4S]	Serine
Fdx1414	213	7.67	[3Fe-4S]	Threonine
Fdx4132	219	7.09	[3Fe-4S]	Serine
Fdx5370	189	6.75	[3Fe-4S]	Histidine
Fdx5464	189	6.71	[3Fe-4S]	Histidine
Fdx4179	204	6.81	[3Fe-4S]	Serine
Fdx5337	309	10.79	[2Fe-2S]	-



(a) (b)

Figure 36: UV/Vis spectra of ferredoxins Fdx2495 (a) and Fdx5333 (b). The characteristic peak absorbance at ≈ 414-418 nm is labelled.

The yields of Fdx2495 and Fdx5333 were relatively high (Appendix B14). However, the [3Fe-4S] ferredoxin Fdx1414 yields were low, although it still exhibited a small broad peak at 414-418 nm (Appendix B13). Due to the low expression of Fdx1414, it was unlikely that an *in vitro* turnover with FraEu1415 could be carried out without further optimisation of the Fdx1414 expression and purification.

The ferredoxin reductase genes were bought as a gblock and were cloned into the pET28 vector using the Nde-XhoI restriction site. Neither ferredoxin reductases showed any characteristic yellow protein in the initial expression tests. Because of this, it was unlikely that a large scale growth would produce any appreciable amount of this ferredoxin reductase protein, and as such large scale growth and purifications were not attempted.

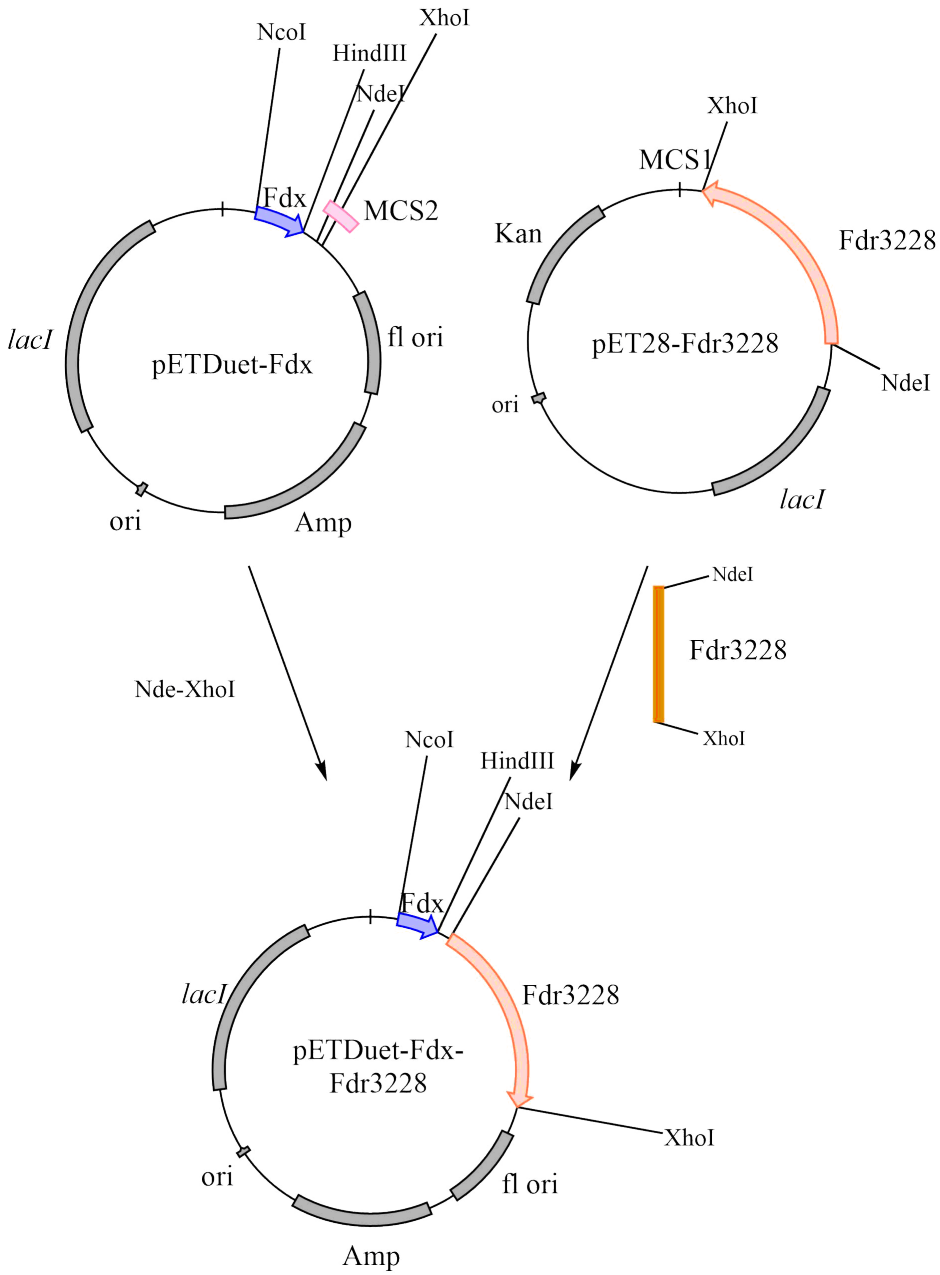


Figure 37: Construction of pETDuet vectors containing Fdx2495, Fdx5333 and Fdx1414 (represented as Fdx in blue) and ferredoxin reductase Fdr3228 (orange). The ferredoxin containing pETDuet vector is used as a backbone for ligation of the ferredoxin reductase gene. The multiple cloning sites 1 and 2 are highlighted in magenta and pink respectively.

While unable to purify the ferredoxin reductases outside the cell, turnovers within the cell (i.e. *in vivo* turnovers) may still be able to occur. As the substrate range for FraEu2494, FraEu5334 and FraEu1415 had been investigated in Chapter 3.2.2, these substrates can potentially be oxidised with an *in vivo* P450 system. An *in vivo* turnover requires the cells to have the necessary partner proteins and co-factors for the P450 reaction, and so the P450, ferredoxin and ferredoxin reductase must first be transformed into a single bacteria. The pET28 plasmid containing the ferredoxin reductase Fdr3228 was first digested with restriction enzymes NdeI-XhoI. Both pET28

and pETDuet plasmids have restriction sites for these enzymes in their respective multiple cloning sites. The ferredoxin reductase gene was then isolated and ligated into the pETDuet plasmid with the ferredoxin (either Fdx2495, Fdx5333 and Fdx1414). The ferredoxin was previously inserted with NcoI-HindIII into the multiple cloning site 1 (MCS1), allowing the ferredoxin reductase to be ligated into the other multiple cloning site (Figure 37). In a separate pRSFDuet vector, the P450 gene was inserted within the Nde-KpnI restriction sites. Finally the ferredoxin and ferredoxin reductase gene containing plasmid would be transformed along with the separate pRSFDuet-P450 containing plasmid into BL21(DE3) *E. coli* competent cells for large scale growth (Experimental 2.5). After this growth, the cells were then harvested and resuspended in *E. coli* minimum media (EMM), where substrate ($\approx$ 1 mM) was added to the culture (Section 3.3.2).

The substrates used for *in vivo* turnovers are outlined in Table 12. After 21 hours, these turnovers were extracted with ethyl acetate and analysed by GC-MS. However, none of the turnovers yielded in any detectable product. This suggests the ferredoxin reductase Fdr3228 may be unable to support *in vivo* turnovers of the ferredoxins Fdx2495, Fdx5333 and Fdx1414 with P450 systems FraEu2494, FraEu5334 or FraEu1415, or that it is not produced in the *E. coli* system.

Table 12: Substrates tested during *in vivo* turnovers for FraEu2494, FraEu5334 and FraEu1415 P450 systems. The ferredoxin reductase was either Fdr3228, or FdrMmar from *Mycobacterium marinum*

P450/Ferredoxin	Substrate
FraEu2494/Fdx2495	ambroxide
FraEu2494/Fdx2495	α-ionone
FraEu2494/Fdx2495	β-damascone
FraEu2494/Fdx2495	sclareolide
FraEu5334/Fdx5333	β-damascone
FraEu5334/Fdx5333	β-ionone
FraEu5334/Fdx5333	methyl ionone
FraEu5334/Fdx5333	santonin
FraEu1415/Fdx1414	nootkatone
FraEu1415/Fdx1414	progesterone

Non-native ferredoxin and ferredoxin reductases have been shown in a number of studies to be able to reconstitute native-like electron transfer activity.[131,132] A similar *in vivo* assay with a foreign ferredoxin reductase may be able to facilitate the initial electron transfer step from NADH. The ferredoxin from *Mmar_2932* (henceforth known as FdxMmar) from *Mycobacterium marinum* is phylogenetically related to the ferredoxins from *Frankia* sp. EuI1c, and the clustered ferredoxin reductase (henceforth known as

FdrMmar) may be able to facilitate electron transfer in the *Frankia* system (Appendix B15). The *in vivo* assays were repeated with the ferredoxin reductase FdrMmar replacing Fdr3228. However, no products could be detected by GC-MS, suggesting that this foreign ferredoxin reductase is also unable to support electron transfer.

As the ferredoxin reductases were not able to be produced at a large scale, *in vitro* turnovers could not analysed. However, yields of the novel [3Fe-4S] iron-sulfur cluster ferredoxins, Fdx2495 in particular, appeared to be at high enough levels that further investigation of the variable residue in the iron-sulfur cluster binding motif of the ferredoxin could be carried out (Section 3.2.5).

3.2.5 Fdx2495 Mutant Library

The majority of ferredoxin electron carriers identified in previous sections are [3Fe-4S] or [4Fe-4S] iron-sulfur complexes with non-standard residues in their conserved binding motif. How the identity of this particular residue can affect the expression, stability and function of these iron-sulfur complexes will be investigated. The ferredoxin Fdx2495 was expressed and purified at a large scale (Section 3.2.4). Fdx2495 has a histidine in its non-standard residue position at the "?" position in **C**XX**?**XXC(X)**CP** motif. It can be generated as a stable ferredoxin using *E. coli* cells (Section 3.2.4), and was chosen as the parent ferredoxin for this investigation.

The histidine residue is the thirteenth amino acid in the Fdx2495 peptide sequence, and is encoded by the codon CAC. To create a mutation library of this amino acid, a gblock sequence from Integrated DNA Technologies (IDT) was bought that had random DNA bases replacing the first two bases in this codon (i.e. NNC, where N is a random DNA base, Figure 38). This allowed for the possibility of generating potentially 14 of the 20 different amino acids at this position. Forward and reverse primers were also designed to insert a His-tagged tail flanking the end of the protein before the stop codon (Figure 39). This His-tagged gene will allow the use of a His-tagged affinity column after expression for purification of the protein. The His-tagged genes were ligated into a pET28 vector between the NcoI-HindIII restriction sites and the DNA was purified as outlined in Experimental Section 2.2.3. Only ≈ 36% of purified DNA showed a cut in the NcoI-HindIII DNA digest. This correlated to only 9 NNC mutants being successfully cloned and these were sequenced by the Australian Genome Research Facility (Figure 40). Four of the 9 NNC codons encoded for duplicate amino acids, and as such only 5 different amino acids were generated at the non-standard residue position in this first screen. The mutants were alanine (Fdx2495 H13A), tyrosine (Fdx2495 H13Y), asparagine (Fdx2495 H13N) and cysteine (Fdx2495 H13C). The final NNC mutant produced a histidine at H13, rendering it a His-tagged wild-type.

5`AATGTTCAGAATTACTTACCTGACCTTTCCATGCGTGCTCCATACCCCTCCACTCCCCCG
CTTATCCGTCCGAGGGGAGAGTTATTAT*CC**ATGG***CCAAGGTGGCTGTAGATGGTGCGAAA
TGCACAGGCNNCGCTCGCTGTCAGGCGACAGCACCAGAAGTTTTCGCGCTGGATGAGCT
CGGGTATGCTGTTCCGGGAGAACGCGAGATCGCGGCAGGTGCCGAGGCCGAAGCCCGTC
GTGGTGCATCGGCATGCCCTGAACGCGCTATTACCGTTTGGGGTGATCCGTTAGGTCGCC
CGGCCCGC**TAA***AAGCTT*TAATAA3`

Figure 38: The gblock for the codon optimised ferredoxin gene *FraEuI1c_2495* containing a random NNC mutant (in cyan). The forward and reverse primers annealing regions are in green and the expressed polypeptide chain is underlined in red. The start/end codons are in bold, while the restriction sites NcoI-HindIII are labelled in italics and/or highlighted yellow. The wild-type gene encoded "CAC" where it the NNC is labelled.

FraEuI1c2494 5` primer: TTTCATACCATGGCCAAGGTGGCTGTAG

FraEuI1c2495 3` primer:
TTAATTAAGCTTCTATTAATGATGGTGGTGATGATGGCGGGCCGGGGCGACCTAACGG

Figure 39: The primers for the codon optimised *FraEuI1c_2495* gene. The annealing regions are in green, and the His-tagged region is in purple. The start and end codons are in bold, while the restriction sites for NcoI (5′ primer) and HindIII (3′ primer) are underlined.

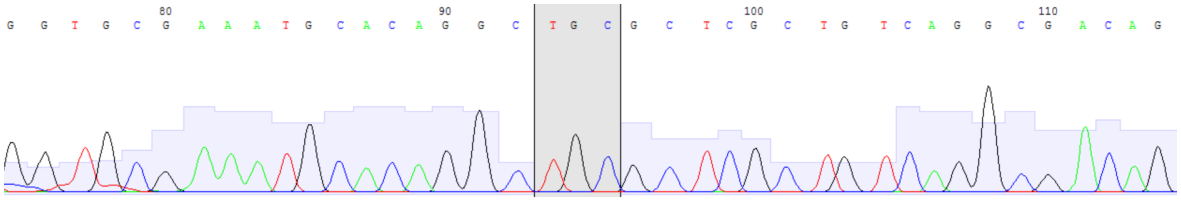
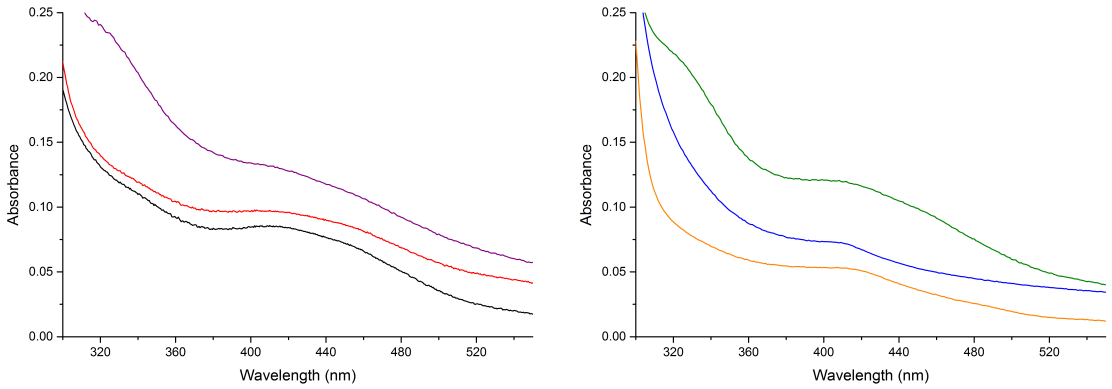


Figure 40: The DNA sequence of an *FraEuI1c_2495* NNC mutant, highlighting the NNC position (93-95), which encodes for cysteine (TGC).

After sequencing the Fdx2495 ferredoxin mutants, they were transformed into BL21 (DE3) *E. coli* cells. The Fdx2495 mutants were then prepared in a large scale outlined in the Experimental 2.2.3. The Fdx2495 mutants exhibited the characteristic ferredoxin peak centred around ≈ 414-418 nm, and the size was confirmed by SDS-PAGE gel analysis (not shown). Using an extinction coefficient of 11300 M^{-1} cm^{-1} at 412 nm for a similar [3Fe-4S] or [4Fe-4S] ferredoxin, the approximate concentration and protein yield for these mutants could be determined by recording the same absorbance peak of 412 nm using Beer-Lambert's law (Table 13).[137]



(a) (b)

Figure 41: UV/Vis spectra of mutant ferredoxins Fdx2495. (a) Spectra of the wild-type (in black), tyrosine mutant (in red) and the His-tagged wild-type (in purple). (b) Spectra of the asparagine mutant (in orange), cysteine mutant (in blue) and alanine mutant (in green). The His-tagged wildtype, alanine and tyrosine mutants have been diluted while the asparagine and cysteine mutants have been concentrated before UV/Vis analysis. For clarity the spectra have been offset along the y axis.

The order of protein yield/concentration of the ferredoxin library is as follows: His-tagged wild-type > alanine > tyrosine > asparagine > cysteine (Appendix B16, Table

Table 13: The yields of Fdx2495 and its his-tagged H13 mutants per litre of cell culture. The extinction coefficient used was 11300 M^{-1} cm^{-1} at 412 nm.[137]

His-tagged ferredoxin mutant	Product (mg)/Litre of cell culture
Fdx2495 WT	1.5
Fdx2495 Ala (H13A)	0.86
Fdx2495 Tyr (H13Y)	0.49
Fdx2495 Asn (H13N)	0.33
Fdx2495 Cys (H13C)	0.29

13).

Interestingly, mutating the non-standard residue to an alanine (H13A), which often is the standard residue for most [3Fe-4S] ferredoxins, decreased the amount of protein produced compared to the wild-type Fdx2495. The other mutants were produced at even lower levels, suggesting that this histidine residue is optimal for high yields or optimal protein folding. The cysteine mutant (H13C) creates a binding motif similar to [4Fe-4S] iron-sulfur ferredoxins. The side chains of the cysteine residues in the binding motif **C**XX**C**XXC(X)**CP** can each be coordinated to an iron metal, resulting in the possible formation of a [4Fe-4S] iron-sulfur complex. However as the yield of this ferredoxin mutant was the lowest out of the mutants presented, it may be unstable during the production and purification processes. [4Fe-4S] ferredoxins are known to be oxygen sensitive. As this mutation may have also changed the intensity of the absorbance at 412 nm and extinction coefficient, the relative yields of the ferredoxin mutants cannot be taken as definitive.

The small ferredoxin mutant library provided an initial insight into how the variable amino acids at the standard residue position can affect the stability and generation of these variants. It could potentially be used for activity studies and optimisation. Further mutations of this non-standard residue are required to generate a complete mutant library, allowing for an in depth analysis of the effect the variable residue has these types of ferredoxins.

3.3 Discussion

3.3.1 *Frankia* sp. EuI1c Protein Expression

The level of protein production using an *E. coli* host for many of the P450s from *Frankia* sp. EuI1c were poor. Analysis of the DNA sequence of the genes revealed a substantial number of rare codons which may slow down translation resulting in lower expression levels. As there are more total possible codon combinations than there are amino acids (64 compared to 20), most codons are redundant. However, while these amino acids have a number of redundant codons, not all redundant codons are equally favoured in different organisms.[96,138,139] Depending on the species that the encoded protein originates from, different codons are more or less common. While highly expressed genes are often enriched with common codons, rare codon clusters are abundant in prokaryotic and eukaryotic genomes, and are not just confined to rare or poorly expressed proteins.[96,140] The majority of P450s from *Frankia* sp. EuI1c showed low yields, and while they did not appear to have an unsustainable number of rare codons, many of the genes had extensive rare codon clusters. In order to improve the poor expression caused by these clusters, codon optimised genes for *E. coli* may improve yields. By changing and optimising the codons for the host cell, the same P450 protein can be expressed at higher rates and fidelity without changing the amino acid sequence. However we cannot rule out that low heme incorporation or poor protein folding is the cause for the low level of P450 holoenzyme generated.

The expression and purification of FraEu2494 was significantly greater than the other *Frankia* sp. EuI1c P450s. Optimising the expression of FraEu5334, FraEu1415 and FraEu4131 will further the understanding of these enzymes as potential novel biocatalysts. Adding known substrates to the P450s could help stabilise the active site of the P450s, and could help with increased production of the heme-containing protein. Initial expression tests of FraEu5334 and FraEu1415 with β-ionone and androstene-3,17-dione (200 µM) added during the protein growth phase revealed little benefit to the level of holoenzyme production for FraEu5334. The FraEu1415 expression test produced a pellet with lower colouration, although this could be due to the substrate having low solubility or potentially being toxic to the cells, reducing protein levels. The heme-prosthetic group is an important component within a cytochrome P450.[141] If the synthesis of heme can be increased, heme incorporation into P450 enzymes should also increase, thus increasing yields of the P450. For example, addition of the heme precursor σ-aminolevulinic acid (σ-ALA) usually improves yields of P450s, as it can help heme biosynthesis.[142] Other methods to increase the yields of the P450s involve using a different expression host that have a higher GC content, such as *Streptomyces venezuelae*, which has been used to produce a high GC% peptidase.[143] Attempting to

export the cytochrome P450 to the periplasm may also increase expression. CYP105D1 from *Streptomyces griseus* was engineered to be able to be heterologously prepared in *E. coli* and able to translocate across the bacterial inner membrane.[144] This process yielded a heme-incorporated P450 within the periplasmic space, which was also reported to be enzymatically active *in vivo* and *in vitro*. This process may be advantageous as it is in an easily accessible cellular location and has been shown to support protein production at high levels.[144] P450s from *Frankia* sp. EuI1c may also benefit from being exported into the periplasm to optimise the level of holoenzyme.

FraEu2494 and FraEu1415 do not belong to a pre-existing P450 family. FraEu5334, which has high sequence similarity to the CYP150 family, is one of the few P450s in *Frankia* sp. EuI1c to be related to a named P450 family. Thus the investigation on these novel P450s from *Frankia* sp. EuI1c may provide a reference for future P450s within new P450 families.

Crystallisation of FraEu2494 has been achieved. If the crystal structure was solved, this would provide insight on the size and shape of the enzyme active site, which would aid in the identification of substrates. Further investigation in optimising crystal growth for FraEu2494 is required to enable the generation of a suitable X-ray diffraction pattern.

Many of the ferredoxins from *Frankia* sp. EuI1c we tested were not generated as well as the P450s. While the P450 FraEu1415 was prepared, its clustered ferredoxin Fdx1414 was produced at low levels. Different methods of ferredoxin production may help stabilise the iron-sulfur cluster proteins. Iron–sulfur clusters are reactive towards a number of different redox-active species, such as reactive dioxygen/oxygen and nitrogen compounds. This makes these iron-sulfur cofactors susceptible to damage.[145] When these clusters come into contact with a strong coordinating species, the thiolate ligands of the iron-sulfur cluster may be displaced, or the iron/sulfide groups may be abstracted.[146–148] These interactions can lead to instability of the cluster, a change in the redox state and/or a complete loss of the iron-sulfur cluster. While the cellular environment usually protects the iron-sulfur protein from these highly oxidising species, there is little protection when the protein is outside the cell. Purification and reconstitution of these ferredoxins in anaerobic conditions can be employed to protect the delicate iron-sulfur clusters from oxidative damage.[145] This will help the preparation of these ferredoxins, and will allow their purification at higher levels. Other methods may involve adding iron or sulfur supplements to the cell culture during growth. Addition of ferric citrate and ferric ammonium have been shown to increase the expression of a plant-type reductase which contains a [2Fe-2S] iron-sulfur cluster.[149] Other iron additives such as ferrous sulfate have also been shown to increase the yields of iron-sulfur complexes.[150–153] Addition of cysteine, which is accepted to be a sulfur source for Fe-S cluster formation, may also increase the yields of the ferredoxins.[149,150]

3.3.2 Substrate Range of the *Frankia* P450s

The potential substrates for FraEu2494 all contain one or more saturated cyclic rings. The magnitude of the spin-state shifts and binding affinity can be correlated to the number of, or rather the lack of, unsaturated cyclic rings. For example, *trans*-decahydronaphthalene contains two saturated cyclic rings and induced a significant spin-state shift (60%). Tetralin contains one saturated and one unsaturated cyclic ring and induced a lower shift (30%). The related naphthalene, comprising two unsaturated cyclic rings, did not induce a noticeable spin-state shift. This suggests that planar unsaturated ring structures bind less favourably than their saturated equivalents. As fused saturated rings are contained in the best substrates of FraEu2494, this suggests that the natural substrates for FraEu2494 will also possess this structural feature. As substrates that bind to the active site of a P450 displaces the heme-bound water ligand and initiates the first electron transfer step, they will gate the commencement of the catalytic cycle and product formation.[154,155]

Many of these substrates also contain oxygen containing substituents that may be interacting with the residues of the substrate binding pocket of FraEu2494. These include ketone, ester and alcohol groups, suggesting that the oxygen atom may interact with hydrophilic active site residues. All these substituents, such as alcohol and ketones, are hydrogen-bond acceptors/donors and are likely to interact with amino acids that contain hydrogen-bond donating side chains. Norisoprenoid substrates such as β-ionol and β-damascone were shown to be potential substrates for both FraEu2494 and FraEu5334. While β-ionone bound to FraEu5334, it did not induce a spin-state shift for FraEu2494. β-Ionol and β-ionone only differ in a single functional group (alcohol for β-ionol and ketone for β-ionone) (Figure 28). α-Ionone and β-damascone are also very similar to β-ionone, with a small difference in the position of the double bond within the cyclic ring for α-ionone, or the position of the ketone for β-damascone. The subtle differences in the position of the hydrophilic ketone or alcohol groups for these norisoprenoids appear to have an effect on the substrates' interaction with the active site of FraEu2494 and FraEu5334. These groups may also be less important for FraEu2494 as *trans*-decahydronaphthalene has no oxygen atoms and exhibited a similar tight binding and spin-state shift as the norisoprenoids.

Th substrate range of FraEu5334 has not be determined as well as FraEu2494. However, a range of organic molecules of different shapes and sizes were able to bind within the active site of FraEu5334. Similar to FraEu2494, some of these compounds also contain one or more cyclic rings. For example, both santonin and nootkatone contain multiple carbon cyclic rings and could bind to FraEu5334. These substrates also contain one or more alkene groups, which may increase rigidity and shape of the molecule. All

of the compounds screened for FraEu5334 contained at least one oxygen containing functional group, such as an alcohol in the case of β-ionol and *cis*-nerolidol, a ketone in the case of β-damascone, or an ester group in neryl acetate. Organic compounds that did not have an oxygen group or were largely hydrophobic, such as naphthalene and *trans*-decahydronaphthalene, could not bind to FraEu5334. Thus these hydrophilic functional groups may create hydrogen bonds with the active site residues and increase their binding affinity to FraEu5334.

Neryl acetate and *cis*-nerolidol are linear chain substrates of FraEu5334. These substrates could interact with the active site in considerably different ways to the cyclic ring containing substrates. Neryl acetate and *cis*-nerolidol share structural features with each other, such as the position of the *cis* double bond at C-6. It is likely that both substrates bind in the active site in a similar conformation. *cis*-Nerolidol is slightly longer than neryl acetate, and this may facilitate further interactions with the active site amino acids as it exhibited a tighter binding to FraEu5334. The *cis*-conformation is important for binding as the *trans* isomer for neryl acetate (geranyl acetate), did not displace the distal water ligand (Appendix B8).

In general, the substrate range for FraEu5334, which includes bulky polycyclic ring structures and long carbon chain terpenes, is wider than than the substrate range of FraEu2494. However, while these substrates are able to displace the distal water ligand, their low binding affinities suggest that these substrates do not interact as favourably with the amino acids within the binding site. Additional substrates need to be screened with FraEu2494 and FraEu5334 to determine the likely natural substrates for these P450s.

Several steroids induced complete dissociation of the distal water ligand on addition to FraEu1415. These were bound with higher affinity than the substrates screened with FraEu2494 and FraEu5334. Due to their almost complete spin-state shifts and the high affinity of binding, it is likely that steroids or related terpenoids are the natural substrate targets for FraEu1415 within *Frankia* sp. EuI1c.

All of the steroid substrates which bind to FraEu1415 have either a carbonyl or an alcohol group at the C-3 position (Figure 42). An alcohol at this position causes the sterol to induce a lower spin-state shift and/or exhibit a lower binding affinity. If the C-3 substituent is positioned above the heme iron such that it displaces the water ligand, this could explain why there is a large difference in spin-state shifts between almost identical substrates, such as progesterone and pregnenolone. However, the C-3 substituent may also be interacting with a hydrophilic residue in the active site. Androsterone is the only sterol (i.e. a steroid with a C-3 hydroxyl group) added to FraEu1415 that completely displaced the water ligand, and is likely to bind in a similar

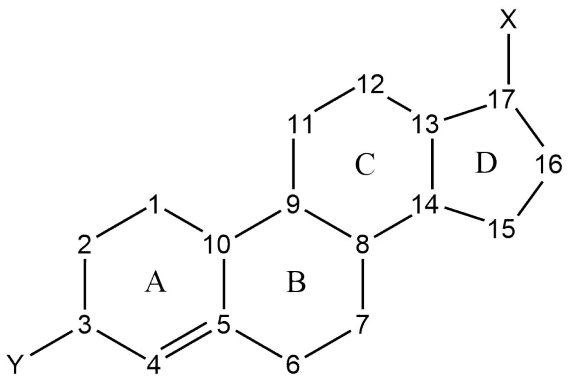


Figure 42: The parent ring system of a steroid, with rings ABCD labelled accordingly. Substituents X and Y are thought to help specificity for FraEu1415. A double bond is highlighted between C-4 and C-5, which is present in progesterone and testosterone, which bound to FraEu1415 very tightly.

conformation to 4-androstene-3,17-dione due to their structural similarities. However the affinity is much weaker for androsterone ($K_{\mathrm{d}} = 204$ µM compared to 8.3 µM, Figure 32). Thus the C-3 substituent could be interacting with a residue in the binding site that prefers a ketone over an alcohol. The substituent at the C-3 carbon not only affects the ability of the steroid to displace the distal water ligand, but also plays a major role in the affinity of the substrate.

The steroid substrates with the highest affinity to FraEu1415 also contain an oxygen containing moiety at the C-17 position, such as the alcohol in testosterone, and a acetyl group in pregnenolone and progesterone. A ketone at C-17 facilitates tight binding; for example progesterone and pregnenolone both bind with high affinity. Testosterone, which is structurally very similar to progesterone differing only at the C-17 position where it has an alcohol group, binds with an almost three-fold lower affinity (Figure 32). As such this ketone group must have a favourable interaction within the enzyme binding pocket. The steroids that have a carbonyl group at C-17 include 4-androstene-3,17-dione, androsterone, estrone, stanolone and dehydroepiandrosterone (Figure 32). However, dehydroepiandrosterone exhibited no spin-state shift. It has a C-3 hydroxyl group which is hypothesised to reduce the ability to displace the distal water ligand. However this is unlikely to be the sole reason why it does not induce any spin-state shift. Androsterone also has the same C-3 hydroxyl and C-17 carbonyl groups as well and was still able to induce a significant spin-state shift. In fact, androsterone and dehydroepiandrosterone only differ in one alkene bond at the C-5 position. The double bond in dehydroepiandrosterone may have increased the rigidity of the substrate altering its shape. This may prevent binding in the pocket or altering the position, resulting in no observable spin-state shift.

Like dehydroepiandrosterone, pregnenolone also has a double bond at the C-5 position, but was still able to induce a spin-state shift and bind very tightly to FraEu1415 (Figure

32). Thus the position of the double bond may just be one of many variables that determine whether the substrate can displace the distal water ligand. Stanolone and 4-androstene-3,17-dione also only differ by a double bond, where stanolone completely lacks a double bond in its structure. The lack of a double bond decreased the binding affinity of FraEu1415 towards stanolone. Androsterone also lacks any double bonds in its structure (Figure 32), and while it has a hydroxyl at C-3, it is still able to induce a full spin-state shift. The weak binding could be, as discussed previously, due to the C-3 moiety, but it could also be attributed to a lack of a double bond. Therefore the cumulative effects of the substituents at C-3 and C-17, as well as the position of the double bond within the compound must affect the binding affinity of FraEu1415 for the steroids.

When compared to the steroids, nootkatone has a lower affinity for FraEu1415. The decrease in affinity is thought to be because it lacks the additional cyclohexane ring (C) and cyclopentane ring (D) which make up the core structure of a steroid (Figure 42). While this difference is likely the reason why exhibits significantly weaker binding, nootkatone was still able to generate a respectable spin-state shift. The two rings in nootkatone is similar to the cyclic rings A and B in a steroid, and the carbonyl group at C-3 is in the same position as testosterone, progesterone, 4-androstene-3,17-dione and stanolone, all of which exhibit very tight binding. This further suggests that this carbonyl moiety is an important functional group for binding to FraEu1415.

Further investigation is required to determine the oxidation products of the substrates for FraEu1415. Optimisation of Fdx1414 expression and determining a ferredoxin reductase that could facilitate the electron transfer between NAD(P)H and the ferredoxin would allow for *in vitro* turnover assays. Identifying either a suitable native ferredoxin reductase for the FraEu1415 system or introducing foreign electron transfer partner proteins could enable *in vivo* turnovers of these steroids.

Use of non-native ferredoxin and ferredoxin reductase proteins to facilitate electron transfer in a native P450 system has been reported numerous times. For example, it has been previously reported that CYP109B1 from *Bacillus subtilis* hydroxylates valencene to nootkatol and nootkatone.[131] CYP109B1 was cloned into *E. coli* and was coexpressed with putidaredoxin reductase (PdR) and putidaredoxin (Pdx) from *Pseudomonas putida*. The PdR and Pdx electron transfer system supported successful hydroxylation from valence to nootkatol and nootkatone. Although the natural electron transfer system of CYP109B1 is currently unknown, it has been shown that multiple artificial redox proteins were able to facilitate enzymatic activity.[132] These include truncated bovine adrenodoxin and adrenodoxin reductase, flavodoxin and flavodoxin reductase from *E. coli*, and the flavodoxins YkuN and YkuP from *B. subtilis*.[132,133] These artificial redox partner proteins may also be able to facilitate turnover activity for FraEu5334. Further

investigation to find electron transfer partner proteins for the P450s from *Frankia* sp. EuI1c is required for study on their turnover activity.

Soil-dwelling actinomycetes, such as *Frankia*, are well known for being producers of many bioactive compounds and natural products. Recent bioinformatic analysis have revealed an abundance of secondary metabolic biosynthesis gene clusters in the *Frankia* strains ACN14a, CcI3, and EAN1pec.[87] These secondary metabolites and natural products include cyclic peptides, siderophores, aromatic compounds, signalling molecules and lipids. These gene clusters are unique to individual strains of *Frankia*, suggesting species-specific biosynthetic diversity.[87] Chemotaxonomic studies of *Frankia* sp. EuI1c has also revealed the metabolism of various phospholipids, sugars and fatty acids.[91] In combination with results from the three P450s tested here, this shows the potential of novel natural products and secondary metabolites from *Frankia* sp. EuI1c.

3.3.3 Ferredoxin Mutant Library

The variable non-standard residue "?" in the motif [**C**XX**?**XX**C**$(X)_n$**CP**] was investigated to ascertain how it affects the stability of the [3Fe-4S] ferredoxin Fdx2495. Fdx2495 has a non-standard histidine residue at H13, which is usually occupied by an alanine or glycine amino acid in most [3Fe-4S] iron-sulfur cluster ferredoxins. Similar histidine-containing ferredoxins in the non-standard residue position have been previously reported, including one from *Rhodopseudomonas palustris* (HaPuxC).[137,156,157] These [3Fe-4S] ferredoxins also possess a histidine residue replacing the standard glycine/alanine residue in the **C**XX**G/A**XX**C**$(X)_n$**CP** binding motif. The three-cysteine residue side chains act as ligands to the three irons in the cluster. In the crystal structure of HaPuxC the histidine side chain in the cluster-binding motif points away from the vacant site of the cluster (Figure 43).[156]

A structurally characterised [3Fe-4S]/[4Fe-4S] ferredoxin from *Pyrococcus furiosus* has been reported previously,[156] and this ferredoxin has an aspartate residue at the "?" position of the **C**XX**?**XX**C**$(X)_n$**CP** motif. The oxidised ferredoxin is a [3Fe-4S] iron-sulfur cluster, but when first isolated in anaerobic conditions, it is a [4Fe-4S] iron-sulfur complex.[146,158,159] The [4Fe-4S] cluster is retained and stabilised by the mutation of the aspartate D14 to a serine or a cysteine (D14S or D14C).[146] However, mutating the aspartate to a valine, histidine, tyrosine or asparagine isolated the ferredoxin with a [3Fe-4S] cluster.[158] It has also been reported in previous studies that a [3Fe-4S] to [4Fe-4S] iron-sulfur cluster conversion was possible via site-directed mutagenesis.[160] It is likely the Fdx2495 histidine wild-type, tyrosine and asparagine mutants will remain as a [3Fe-4S] iron-sulfur cluster. The cysteine mutant H13C had lower protein yields, and it may be possible that a [4Fe-4S] iron-sulfur cluster is not stable for Fdx2495.

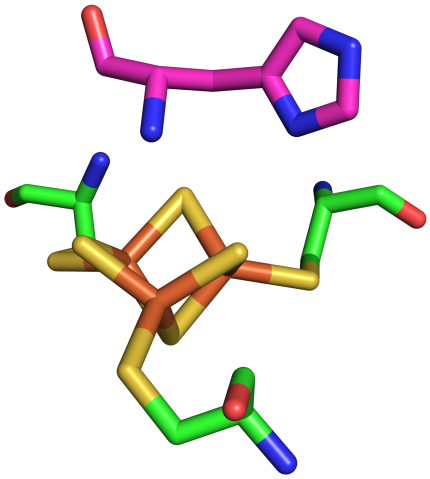


Figure 43: The binding site of the [3Fe-4S] iron-sulfur ferredoxin of HaPuxC from *Rhodopseudomonas palustris* (PDB: 4ID8).[156] The iron-sulfur complex (yellow and orange) is coordinated by the side chains of three cysteine residues (green). The histidine in the non-standard position is highlighted in purple.

However, further investigation into the cysteine mutant is needed to show whether or not the Fdx2495 mutant can be purified as a [4Fe-4S] iron-sulfur cluster, or if it remains a [3Fe-4S] cluster which is produced in lower yields. These investigations could include solving the crystal structure, or spectroscopic analysis such as variable-temperature magnetic circular dichroism (VTMCD), electron paramagnetic resonance (EPR) and resonance Raman (RR).[146,158,159]

Expression and purification of these ferredoxin mutants in anaerobic conditions and other conditions outlined earlier for the other *Frankia* sp. EuI1c ferredoxins may also help increase the yields of these mutants. Despite the low yields, Fdx2495 has been successfully mutated at the variable amino acid within the binding motif of the [3Fe-4S] ferredoxin. The successful purification of these mutants have shown that this system is able to produce additional mutants for further investigation.

4 CYP199A4 S244D Mutant Studies

4.1 Introduction

CYP199A4 is a cytochrome P450 enzyme from the *Rhodopseudomonas palustris* strain HaA2 which can effectively demethylate 4-methoxybenzoic acid at the *para* position.[1,7,68] This system comprises of the P450 CYP199A4 and its electron transfer partners; a [2Fe-2S] ferredoxin (HaPux) and a flavin adenine dinucleotide containing ferredoxin reductase (HaPuR) which mediate electron transfer via NADH. When combined with the electron transfer partners, CYP199A4 is known for having high hydroxylation, demethylation and desaturation activities of *para*-substituted benzoic acids.[7,11,161,162] In particular, CYP199A4 has high affinity for 4-methoxybenzoic acid (K_d = 0.28 µM), and demethylates at the *para*-position to form 4-hydroxybenzoic acid with a product formation rate of 1219 min^{-1}.[7,10,68] CYP199A4 also exhibits similarly high affinities for substrates that are closely related to 4-methoxybenzoic acid such as 4-ethylbenzoic acid.[7,21] It has been shown that moving the substituents at the *para* position to the *meta* or *ortho* positions, such as 3-methoxybenzoic acid and 2-methoxybenzoic acid, has a negative effect on substrate binding and product formation.[70]

Tight binding is observed for CYP199A4 with benzoic acid derivatives with a carboxy moiety at the *para* position. Replacing this carboxy group with an aldehyde or alcohol moiety results in a significant decrease in binding tightness and product turnover rates.[10,73] For example, 4-methoxybenzoic acid has a binding affinity of 0.28 µM, while 4-methoxybenzaldehyde has a binding affinity of 197 µM.[10,70] Although the wild-type CYP199A4 exhibits low oxidation activity for substrates in which the carboxy terminus of the benzoic acid moiety has been modified, the mutation of serine 244 to an aspartate residue (S244D) significantly improves the activity of the enzyme towards these substrates despite exhibiting generally weaker substrate binding.[71,72] The compounds previously screened for oxidation activity include those in which the carboxy group of 4-methoxybenzoic acid have been modified (henceforth these will be known as methoxy-modified substrates), e.g. 4-methoxybenzyl alcohol and 4-methoxybenzaldehyde. (Figure 44).[71,72] These previous studies showed an increase in oxidative demethylation of these methoxy-modified substrates with the S244D mutant. Similar methyl- and ethyl-substituted benzene derivatives (i.e. methyl- and ethyl-modified substrates) have also shown to be effectively oxidised by the S244D variant.[72]

While the CYP199A4 wildtype turns over 4-methoxybenzoic acid at a significantly higher rate than the S244D variant, the reverse is true when the carboxy moiety is changed to a different functional group, such as an aldehyde or alcohol.[10,71,72] For example, the product formation rate for 4-methoxybenzaldehyde is 3.0 min^{-1} for the

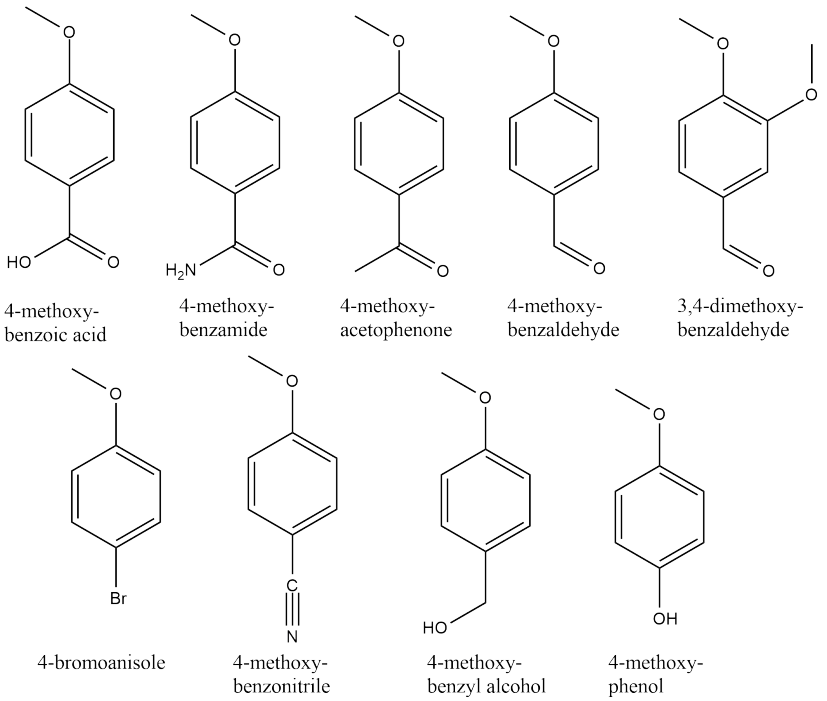


Figure 44: Methoxy-substituted benzenes tested with CYP199A4 wild-type its S244D variant

wild-type variant, while for the S244D it is 860 min^{-1}.[71,72] This relationship is also observed with methyl- and ethyl-modified benzoic acid derivatives (Figure 45). For instance, with 4-methylphenol the product formation rate is 65-fold greater with the S244D variant over the wild-type.[72] In order to assess the number of turnovers that can be catalysed by the S244D mutant, total turnover assays for methoxy-, methyl- and ethyl-modified substrates are investigated and presented here.

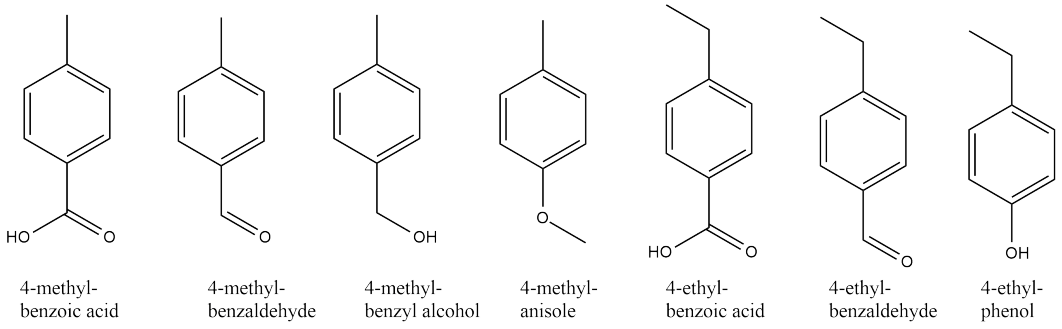


Figure 45: Methyl- and ethyl-modified substrates of CYP199A4 wild-type and S244D variant

Recently it was reported that little to no turnover activity was observed in CYP199A4 with benzoic acid derivatives that lacked a *para* substituent, such as 2-methoxybenzoic acid and 3-methoxybenzoic acid.[70] This highlights the importance for a methoxy substituent or an equivalent moiety at the *para*-position to facilitate the displacement of the distal water ligand bound to the heme iron which would initiate the catalytic cycle. Dimethyl-substituted benzoic acid derivatives will be analysed to investigate the

regioselectivity of hydroxylation by the S244D variant.

CYP199A4 has been shown to effectively oxidise 4-(methylthio)benzoic acid at the *para* position.[73] While demethylation is observed for 4-methoxybenzoic acid, sulfoxidation is the only reaction observed for the turnover of 4-(methylthio)benzoic acid. As the S244D mutant has been shown previously to oxidise *para*-substituted benzene derivatives at a higher product formation rate than the wild-type CYP199A4, the oxidation of *para*-methylthio-modified benzene derivatives by S244D will be investigated.

CYP199A4 has also been shown to effectively catalyse epoxidation of 4-vinylbenzoic acid (Figure 46b).[72] Epoxidation activity by the S244D variant will be investigated with a selection of *para*-substituted styrenes and compared to the wild-type CYP199A4 activity.

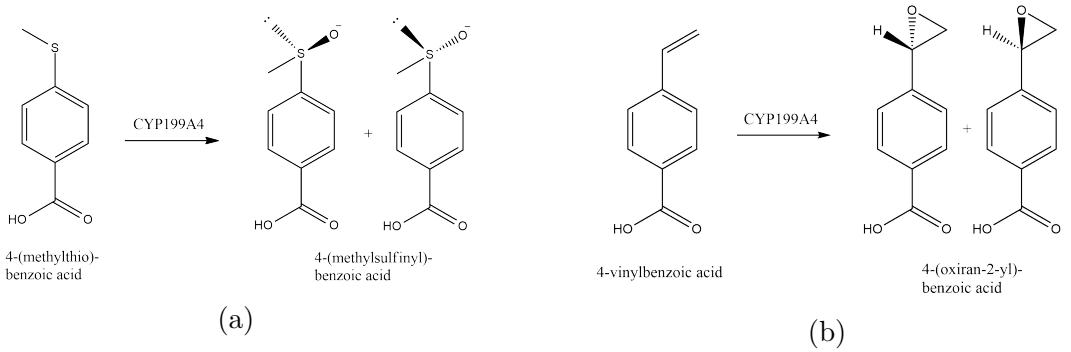


Figure 46: The oxidation of 4-(methylthio)benzoic acid to 4-(methylsulfinyl)benzoic acid (a) and 4-vinylbenzoic acid to 4-(oxiran-2-yl)benzoic acid (b).

The crystal structure of a 4-methoxybenzoic acid bound CYP199A4 has been solved, highlighting the active site residues that interact with the substrate (PDB:4DO1, Figure 15).[21] A selection of other substrates have also been co-crystallised with CYP199A4 which provide further insight into the substrate specificity and the outcomes of substrate oxidation (PDB: 4EGM, 4EGN and 4EGO).[10,21] For instance, the 4-ethylbenzoic acid bound CYP199A4 structure (PDB: 4EGM) reveals the C-H bonds of the C_β in the ethyl group are closer to the heme iron than those of the C_α. By mutating phenylalanine 185, one of the active site residues, to a valine variant (F185V), the enzyme was converted into a more effective 4-ethylbenzoic acid desaturase. This highlights the importance of the amino acids within the active site for selection towards a desired function.

The crystal structure of the S244D variant has not previously been solved. In order to understand the role of the serine 244 to aspartate mutation on the enzyme's substrates selectivity and product distributions, a crystal structure of 4-methoxybenzoic acid bound CYP199A4 S244D mutant will be produced and compared to the wild-type enzyme structure.

4.2 Results: Turnover Activity

4.2.1 Total Turnover Assays with Methoxy-modified Substrates

CYP199A4 and its S244D mutant have been reported to effectively and selectively demethylate 4-methoxybenzoic acid at the *para*-position.[68] Further investigation has shown that the S244D mutant was able to demethylate a range of carboxy-modified substrates at a higher efficiency than the wild-type.[72] For example, CYP199A4 was able to oxidise 4-methoxyphenol to 4-hydroxyphenol at a rate of 0.7 min^{-1} with a coupling efficiency of 9%, which was increased to 600 min^{-1} with 72% coupling efficiency for the S244D variant.[10,71] The coupling efficiencies of 4-methoxybenzoic acid and 4-methoxyacetophenone were also reported to be close to 100%.[71] The activity and coupling efficiencies of these turnovers by CYP199A4 S244D are high and well suited for larger scale biocatalysis. However, these studies are limited by the amount of NADH within the *in vitro* system, as product formation will cease once all of the NADH reducing equivalents are oxidised. By using excess NADH and a lower enzyme concentration, the number of turnovers the CYP199A4 S244D system is able to perform can be calculated. Alcohol dehydrogenase (ADH) can also be added to these systems to facilitate the regeneration of NADH via the oxidation of ethanol to ethanal (Figure 47).

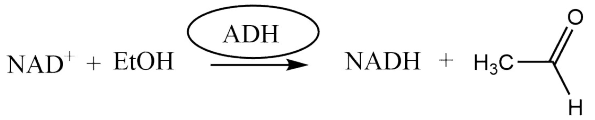


Figure 47: Regenerating NADH system with alcohol dehydrogenase and ethanol

In order to calculate the total number of turnovers, the product concentration is calculated with a calibration curve via HPLC or GC-MS analysis. Unless stated otherwise, the products for these turnovers were identified by coeluting authentic samples of the product with the turnovers using HPLC, GC-MS or both. The total turnover numbers (TTN) were calculated by dividing the final concentration of the product formed in the turnover reaction by the P450 concentration (0.1 µM). Here a turnover can have a theoretical maximum TTN of 20000. The NADH oxidation and product formation rates (PFR) were previously determined for each of the substrates in Figure 53 (Table 14).[72] This provided a measure of the catalytic activity of the S244D mutant. The coupling efficiency, which is the percentage of product formed compared to the amount of NADH reducing equivalents oxidised in the turnover, was also previously calculated.

The total turnover numbers of 4-methoxyacetophenone, 4-methoxybenzonitrile and 4-methoxyphenol with added ADH were lower than the systems without ADH. The other methoxy-modified substrates had similar total turnover numbers with both systems.

Table 14: *In vitro* total turnover data for CYP199A4 S244D with methoxy-modified substrates.[72] The coupling, NADH and PFR data are from a previous study and are given as mean ± S.D. with $n \geq 3$. Rates are given as nmol.nmol-P450^{-1}.min^{-1} (min^{-1}).

Substrate	NADH[72] (min^{-1})	PFR[72] (min^{-1})	Coupling[72] (%)	TTN with ADH	TTN without ADH
4-methoxybenzoic acid	566 ± 2	460 ± 27	82 ± 4	20000	20000
4-methoxybenzaldehyde	868 ± 32	860 ± 33	99 ± 2	20000	20000
4-methoxybenzamide	1050 ± 6	1050 ± 52	100 ± 5	20000	20000
4-methoxyacetophenone	160 ± 7	140 ± 3	85 ± 2	5900	8100
4-methoxybenzonitrile	89 ± 1	33 ± 2	37 ± 2	200	1100
4-methoxybenzyl alcohol	295 ± 2	260 ± 25	90 ± 2	8900	9100
4-methoxyphenol	841 ± 56	600 ± 35	72 ± 4	5600	14800
4-bromoanisole	773 ± 27	690 ± 34	90 ± 8	13500	13100
3,4-dimethoxybenzaldehyde	10 ± 1	8.4 ± 0.7	83 ± 16	820	880

As the results with ADH did not show higher total turnover numbers, turnover results with ADH will not be discussed further in this section.

The *in vitro* turnovers for 4-methoxybenzaldehyde, 4-methoxybenzoic acid and 4-methoxybenzamide with the S244D variant consumed all substrate after 12 hours. The product was also quantified and found to be ≈ 2 mM in each. This resulted in total conversion to a single product, giving the maximum total turnover number of 20000 (Table 14, Figure 48). These substrate total turnovers support previous *in vitro* turnover studies of the S244D mutant, where the coupling efficiencies of the NADH reducing equivalents to product formation were reported to be close to 100% (Table 14).[71,72]

The total turnover of 4-methoxyacetophenone showed little substrate remaining, although the TTN was 8100 (Figure 48b). The turnover for 4-methoxybenzyl alcohol had a slightly higher TTN (9100, Figure 49b), which correlated to the slightly higher PFR and NADH rates compared to 4-methoxyacetophenone (Table 14).[72] The TTN, PFR and NADH oxidation rates for 4-methoxyacetophenone and 4-methoxybenzyl alcohol were lower than that of 4-methoxybenzaldehyde and 4-methoxybenzamide despite having similar coupling efficiency, suggesting that an increase of PFR and NADH rates will also increase TTN. 4-Methoxyphenol and 4-bromoanisole were reported to have high PFR and NADH oxidation rates.[72] The total turnovers for these substrates also produced high total turnover numbers with little substrate remaining (14800 and 13100, respectively, Table 14, Appendix C1, Figure 49c).

The TTN for the 4-methoxybenzonitrile turnover was low (1100). Most of the substrate remained unconverted (Figure 48d), which correlates to the low coupling efficiency (37%) and PFR (33 min^{-1}) previously reported for 4-methoxybenzonitrile.[72] The S244D total turnover with 3,4-dimethoxybenzaldehyde also yielded a low TTN, especially compared

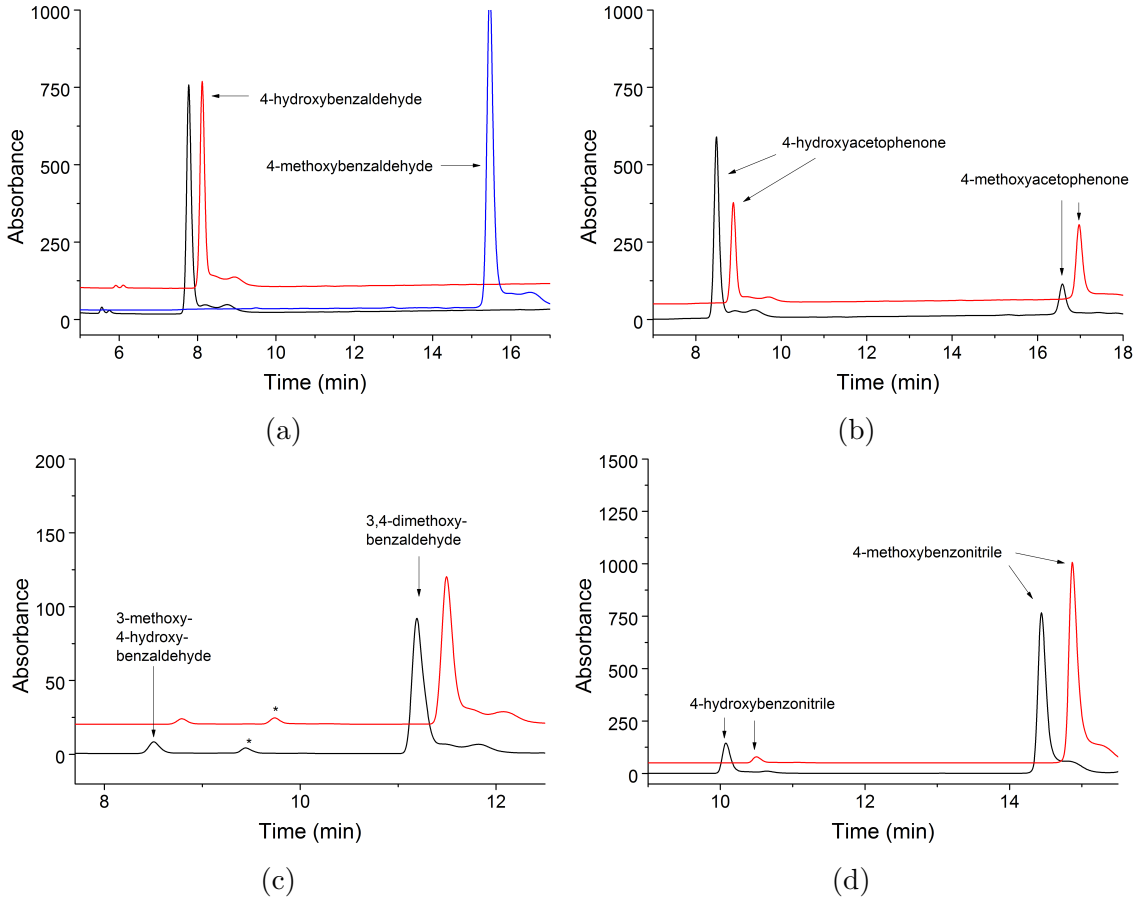


Figure 48: HPLC analysis of the total turnovers by S244D without ADH (black) and with ADH (red) of 4-methoxybenzaldehyde (t_R = 15.7 mins) to 4-hydroxybenzaldehyde (t_R = 7.9 mins) (a), 4-methoxyacetophenone (t_R = 16.6 mins) to 4-hydroxyacetophenone (t_R = 8.5 mins) (b), 3,4-dimethoxybenzaldehyde (t_R = 11.2 min) to 3-methoxy-4-hydroxybenzaldehyde (t_R = 8.5 mins) (c) and 4-methoxybenzonitrile (t_R = 14.3 mins) to 4-hydroxybenzonitrile (t_R = 10.1 mins) (d). Impurities have been marked as (*). The control for the substrate 4-methoxybenzaldehyde in (a) is shown in blue. For clarity the chromatograms have been offset along the x and y axes.

to the monosubstituted 4-methoxybenzaldehyde (880 compared to 20000).[72] This large decrease in TTN was expected as the S244D mutant and wild-type CYP199A4 oxidises 3,4-dimethoxybenzaldehyde to 3-methoxy-4-hydroxybenzaldehyde (vanillin) at a lower PFR rate compared to the monosubstituted 4-methoxybenzaldehyde (Table 14).[71,72] Despite this, both the S244D mutant and the wild-type enzyme have been shown to exclusively oxidise at the *para*-position, and here 3,4-dimethoxybenzaldehyde has been successfully hydroxylated to 3-methoxy-4-hydroxybenzaldehyde in a total turnover system.

The total turnover numbers for 4-methoxybenzoic acid derivatives were shown to correlate more strongly with the NADH oxidation and product formation rates than the coupling of the reaction. The high turnover numbers for the *para*-substituted methoxy-modified substrates show that the CYP199A4 S244D variant is an efficient catalyst for these substrates suitable for optimisation and for the scale-up of monooxygenase activity.

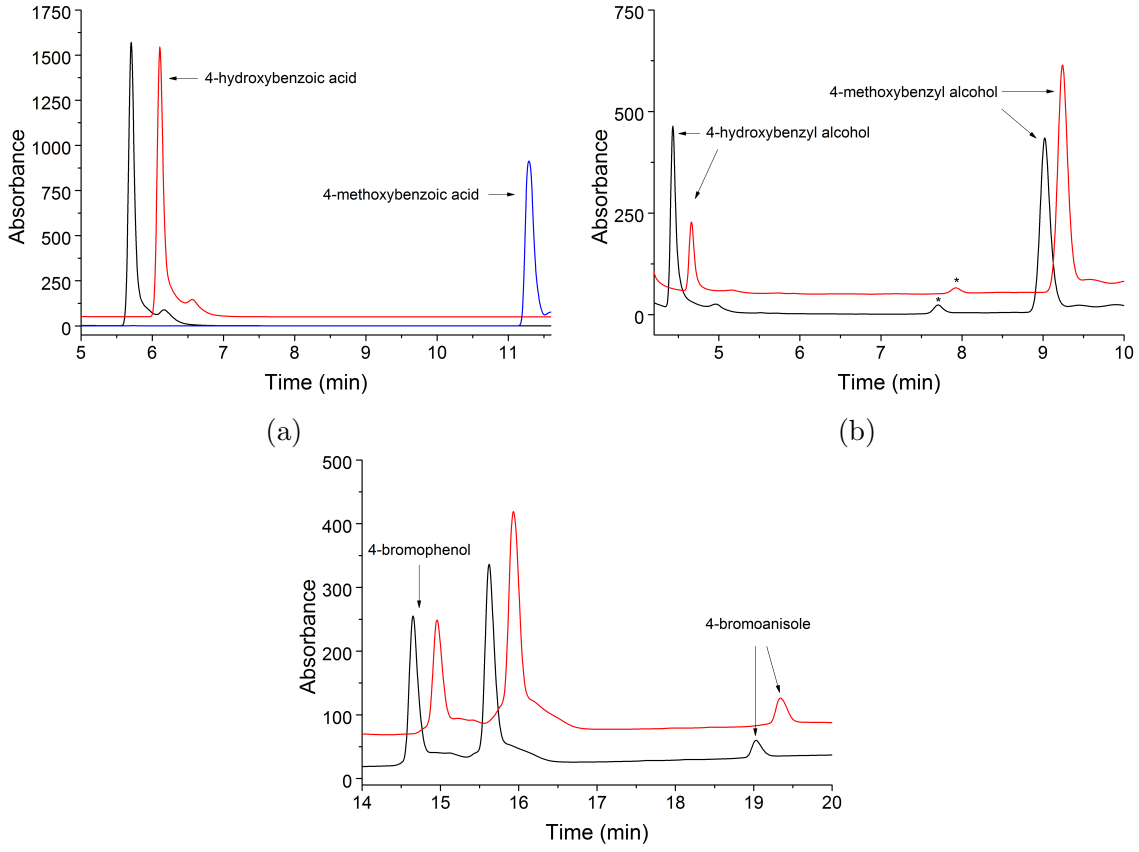


Figure 49: HPLC analysis of the total turnovers by S244D without ADH (black) and with ADH (red) of 4-methoxybenzoic acid to 4-hydroxybenzoic acid (t_R = 5.8 mins) (a), 4-methoxybenzyl alcohol (t_R = 9 mins) to 4-hydroxybenzyl alcohol (t_R = 4.4 mins) (b), and 4-bromoanisole (t_R = 19.1 mins) to 4-bromophenol (t_R = 14.6 mins) (c). The internal standard (t_R = 15.7 min) is also shown. Impurities have been marked as (*). The substrate control for 4-methoxybenzoic acid in (a) is shown in blue. For clarity the chromatograms have been offset along the x and y axes.

4.2.2 Total Turnover Assays with Methyl- and Ethyl-Modified Substrates

The CYP199A4 S244D mutant can bind and oxidise methyl and ethyl-modified benzoic acid derivatives. For example, the S244D mutant binds 4-ethylbenzaldehyde (K_d = 280 µM) which is hydroxylated and desaturated to form 4-(1-hydroxyethyl)benzaldehyde (79%) and 4-vinylbenzaldehyde (21%).[72] The oxidation occurs at high rates and coupling efficiency (564 min^{-1} and 57%). Hydroxylation occurs exclusively at the *para* position for methyl-modified substrates such as 4-methylbenzyl alcohol, forming 4-(hydroxymethyl)benzyl alcohol at a rate of 11 min^{-1} with high coupling efficiency (70%).[72] Although the coupling efficiency and the product formation rates were higher for the S244D mutant than for the wild-type CYP199A4, the binding affinity for 4-methoxybenzyl alcohol was low (K_d = 2800 µM).

These studies demonstrated increased oxidation activity for these methyl- and ethyl-modified substrates for the S244D variant. However these previous studies were limited by the amount of NADH within the system. Here, the total turnovers of methyl- and ethyl-modified substrates will be investigated and presented along the previously reported product formation rates, NADH oxidation rates and coupling efficiencies (Table 15).[72]

Table 15: *In vitro* total turnover data for CYP199A4 S244D with methyl-modified substrates. The coupling, NADH and PFR data are from a previous study[72] and are given as mean ± S.D. with $n \geq 3$. Rates are given as nmol.nmol-P450^{-1}.min^{-1} (min^{-1}).

Substrate	NADH (min^{-1})	PFR (min^{-1})	Coupling (%)	TTN with ADH	TTN without ADH
4-methylbenzoic acid	27 ± 2	8.5 ± 0.7	32 ± 0.2	470	880
4-methylbenzaldehyde	492 ± 13	360 ± 4	73 ± 2	170	1100
4-methylbenzyl alcohol	16 ± 0.4	11 ± 3	70 ± 17	620	1200
4-methylanisole	99 ± 3	42 ± 1	43 ± 2	1900	2000
4-ethylphenol	569 ± 42	500 ± 45	88 ± 3	20000	20000
4-ethylbenzaldehyde	564 ± 10	320 ± 13	57 ± 2	290	4800
4-ethylbenzoic acid	77 ± 2	58 ± 8	74 ± 8	2000	5400

All methyl- and ethyl-modified substrates showed greater total turnover numbers without alcohol dehydrogenase (ADH). In addition, 4-methylbenzaldehyde and 4-ethylbenzaldehyde reacted with ADH to form 4-methylbenzyl alcohol and 4-ethylbenzyl alcohol respectively. This was confirmed with a control experiment where no P450 enzyme and only ADH was added to the turnover system. Due to these results, the total turnover numbers for the system with ADH will not be discussed further and ADH was not used in future total turnover experiments.

The oxidation of 4-methylanisole by S244D formed a major product, 4-hydroxybenzyl

alcohol (71%), and a minor product, 4-methoxybenzyl alcohol (Figure 50c). 4-Hydroxybenzyl alcohol arises from double oxidation of the substrate. This is because the S244D mutant can oxidise 4-methoxybenzyl alcohol faster than 4-methylanisole, ultimately forming 4-hydroxybenzyl alcohol in greater amounts.[72] The total turnover number of 4-methylanisole was 2000. As the turnover had moderate NADH oxidation, product formation rates and coupling, a low TTN was not unexpected.[72] In addition, the methyl-modified substrates have been reported to have very low binding affinity compared to ethyl-modified substrates (e.g. K_d = 800 µM for 4-methylbenzaldehyde compared to K_d = 280 µM for 4-ethylbenzaldehyde). As such, lower TTN are expected for all methyl-modified substrates.[72]

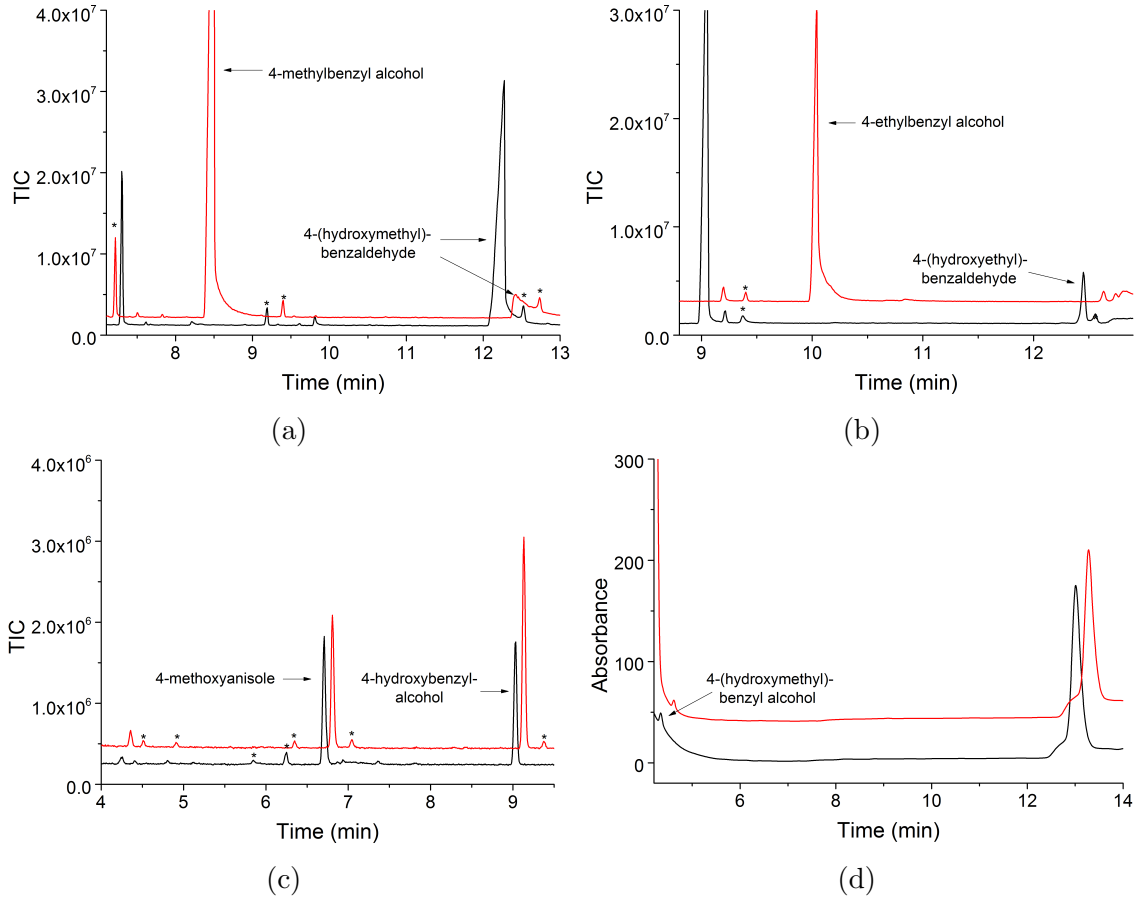


Figure 50: GC-MS analysis of the total turnovers by S244D without ADH (black) and with ADH (red) of 4-methylbenzaldehyde (t_R = 7.4 min) to 4-(hydroxymethyl)benzaldehyde (t_R = 12.3 mins) (a), 4-ethylbenzladehyde (t_R = 9 mins) to 4-vinylbenzaldehyde (t_R = 9.2 mins) and 4-(hydroxyethyl)benzaldehyde (t_R = 12.4 mins) (b), 4-methylanisole (t_R = 4.3 min) to 4-methoxybenzyl alcohol (t_R = 6.7 mins) and 4-hydroxybenzyl alcohol (t_R = 9 mins)(c), and HPLC analysis of 4-methylbenzyl alcohol (t_R = 12.9 min) to 4-(hydroxymethyl)benzyl alcohol (t_R = 4.4 mins) (d). Impurities have been marked as (*). 4-methylbenzyl alcohol and 4-ethylbenzyl alcohol were formed in control experiments containing ADH and NADH only. For clarity the chromatograms have been offset along the x and y axes.

The turnovers of the methyl-modified substrates 4-methylbenzoic acid, 4-methylbenzaldehyde and 4-methylbenzyl alcohol generated low total turnover numbers (880, 1100 and 1200, respectively, Table 15, Figure 50). Low TTN were expected for

4-methylbenzoic acid as previous studies report that *para*-substituted benzoic acid substrates have lower affinity and turnover activity for the S244D mutant compared to the wild-type CYP199A4.[71,72] When comparing the methyl-modified substrates, no obvious trend was observed between total turnover number, product formation rate and coupling.

The ethyl-modified compounds have greater product formation rates than the methyl-modified substrates (Table 15).[72] For example, the turnover of 4-ethylbenzaldehyde to 4-vinylbenzaldehyde and 4-(1-hydroxyethyl)benzaldehyde had a TTN of 4800, which is more than four times greater than 4-methylbenzaldehyde. (Table 16, Figure 50b). Similarly, the TTN of 4-ethylbenzoic acid (5400) was significantly higher than that for 4-methylbenzoic acid. The minor products 4-(1-hydroxyethyl)benzoic acid (4%) and 4-(2-hydroxyethyl)benzoic acid (20%) were observed along with the major product 4-vinylbenzoic acid (75%, Figure 52b). Other possible further oxidation products such as epoxides and ketones that were reported for 4-ethylbenzoic acid turnovers were not observed in this total turnover assay.[72] The TTN of 4-ethylbenzoic acid was greater than that of 4-ethylbenzaldehyde in line with its higher coupling efficiency (Table 15).

The TTN for 4-ethylphenol was 20000, which is the maximum potential turnover number (Table 15). The major product, 4-(hydroxyethyl)phenol, was observed in both systems with and without ADH. The minor product, 4-vinylphenol (10%), was only observed in the turnover system with ADH (Figure 51b). It is possible within the turnover system without ADH that further oxidation of 4-vinylphenol is responsible for the lack of this desaturated product. The high TTN is supported by high PFR and coupling efficiencies within *in vitro* turnovers of 4-ethylphenol (88% and 500 min^{-1} respectively, Table 15).[72] For the ethyl-modified substrates, larger total turnover numbers were observed for the turnovers with the highest product formation rates and coupling efficiencies (88% and 500 min^{-1} respectively, Table 15).[72]

In vivo turnovers with 4-methylbenzaldehyde, 4-ethylbenzaldehyde and 4-ethylphenol (2 mM) were undertaken to scale up the reaction system (outlined in the Experimental Section 2.5). The *in vivo* turnovers for 4-methylbenzaldehyde and 4-ethylbenzaldehyde were almost exclusively converted to 4-methylbenzyl alcohol and 4-ethylbenzyl alcohol respectively, producing similar results as the total turnovers with alcohol dehydrogenase (Appendix C2). However, this aldehyde to alcohol reduction may possibly be a product of the substrate interacting with other components within the cellular environment (such as ADH).

In vivo turnover analysis for 4-ethylphenol revealed high level (92%) of substrate conversion (Figure 51). 4-(1-Hydroxyethyl)phenol was the major product (89%, $\approx$ 1580 uM) with 4-vinylphenol making up the remainder (11% $\approx$ 220 uM). The high total turnover

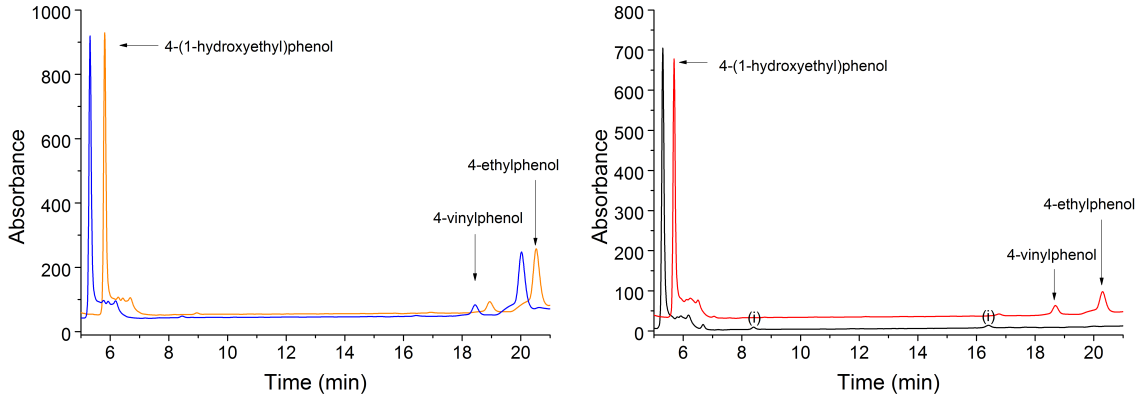


Figure 51: HPLC analysis of oxidation of 4-ethylphenol (t_R = 20) by S244D in (a) *in vivo* turnovers after 2 hours (blue) and 21 hours (orange) and (b) *in vitro* total turnovers without ADH (black) and with ADH (red) to 4-vinylphenol (t_R = 18.7 mins) and 4-(1-hydroxyethyl)phenol (t_R = 5.3 mins). Possible further oxidation products are denoted with (i). 4-Vinylphenol is is not observed in the total turnover without ADH in (b). For clarity the chromatograms have been offset along the x axis in (a) and the x and and y axes.

numbers and *in vivo* turnover results show the potential for upscale turnover activity of methyl- and ethyl-modified substrates with the S244D variant.

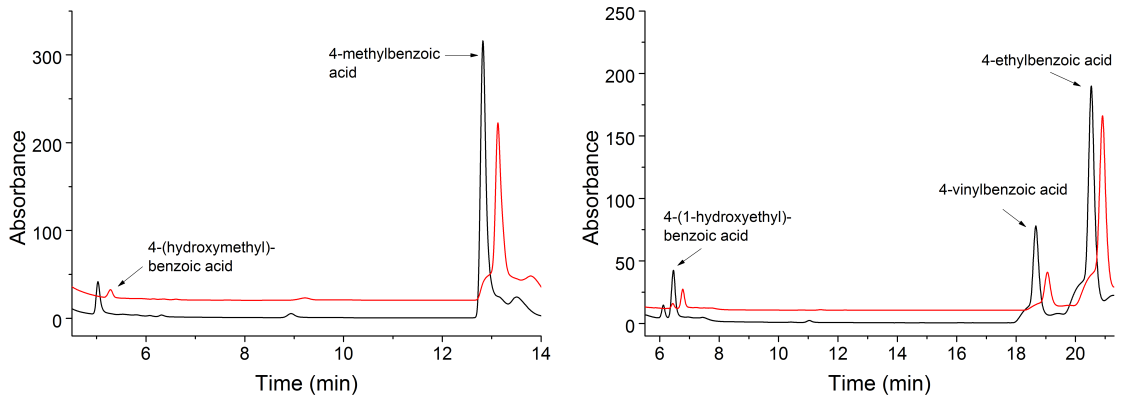


Figure 52: HPLC analysis of the total turnovers by S244D without ADH (black) and with ADH (red) of 4-methylbenzoic acid (t_R = 12.9 mins) to 4-(hydroxylmethyl)benzoic acid (t_R = 5.1 mins) (a), 4-ethylbenzoic acid (t_R = 20.5 min) to 4-vinylbenzoic acid (t_R = 18.5 mins), 4-(1-hydroxyethyl)benzoic acid (t_R = 6.5 mins) and 4-(2-hydroxyethyl)benzoic acid (t_R = 6.1 mins) (b). For clarity the chromatograms have been offset along the x and y axes.

4.2.3 Activity and Product Formation Assay with Methyl-modified Substrates

The S244D and wild-type CYP199A4 can oxidise a wide range of methoxy- and methyl-substituted benzoic acid derivatives exclusively at the *para*-position.[10,70,72] For instance, oxidative demethylation of the methoxy group *para* to the carboxyl moiety in 3,4-dimethoxybenzoic acid to form 3-methoxy-4-hydroxybenzoic acid is the only product observed with CYP199A4.[70] Both the wild-type and S244D variant have also been reported to selectively oxidise 3,4-dimethoxybenzaldehyde to 3-methoxy-4-hydroxybenzaldehyde exclusively at the *para*-position.[71] Here the S244D mutant will be investigated for activity with methyl-modified substrates containing additional *ortho* and *meta*-methyl substituents (Figure 53).

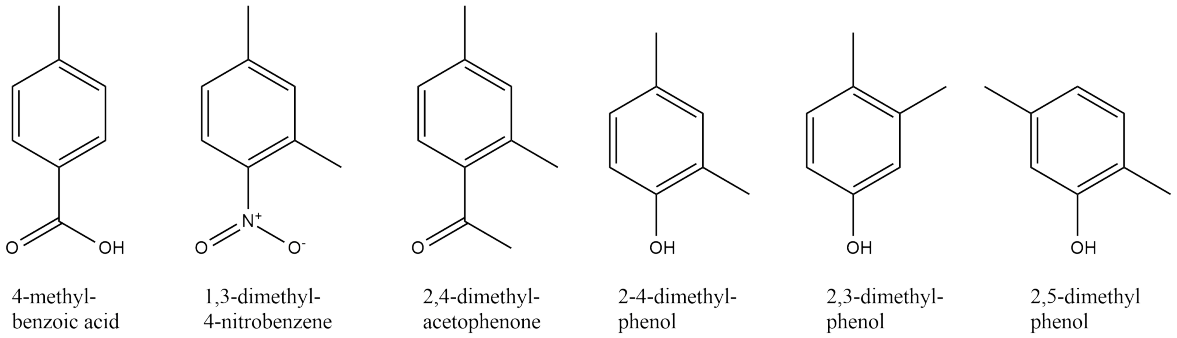


Figure 53: Methyl- and dimethyl-modified substrates of S244D CYP199A4

4-Methylbenzoic acid has been reported to be oxidised by CYP199A4 to 4-hydroxymethylbenzoic acid with high coupling efficiency ($\approx$ 90%).[10] The turnover with the S244D mutant had a lower coupling efficiency (31.2%, Table 16) and product formation rate (8.5 min^{-1}, Figure 54a). This lowered activity in the S244D variant compared to the wild-type was also observed in previous studies on the turnovers of 4-methoxybenzoic acid.[71,72]

The turnover activity of the S244D mutant for disubstituted methyl substrates was lower than that for monosubstituted methoxy-, ethyl- or methyl-substituted benzene derivatives (Table 16, Sections 4.2.1 and 4.2.2). The disubstituted substrate with the highest product formation rate was 1,3-dimethyl-4-nitrobenzene (99.7 min^{-1}). The product generated from the S244D turnover coeluted with an authentic standard of (3-methyl-4-nitrophenyl)methanol (Figure 54b). Here the S244D variant exclusively hydroxylated the substrate at the *para*-position with high coupling efficiency (51%). The wild-type turnover control showed a smaller product peak, suggesting it had a lower PFR (Figure 54b). No product was observed for the closely related 1,2-dimethyl-4-nitrobenzene turnover.

Table 16: *In vitro* turnover data for CYP199A4 S244D mutant with methyl-modified substrates. Unless otherwise stated, the turnovers is with the S244D mutant. The data are given as mean ± S.D. with $n \geq 3$. Rates are given as nmol.nmol-P450^{-1}.min^{-1} (min^{-1}).

Substrate	NADH (min^{-1})	PFR (min^{-1})	Coupling (%)
4-methylbenzoic acid	27 ± 2	8.5 ± 0.7	32 ± 0.2
4-methylbenzoic acid$_{WT}$ [10]	444 ± 8	397 ± 22	89 ± 4
2,4-dimethylphenol	28.7 ± 3	12.3 ± 1	43 ± 4
3,4-dimethylphenol	32.1 ± 3	26.6 ± 3	83 ± 5
3,4-dimethylphenol$_{WT}$	_[a]	_[b]	_[b]
2,5-dimethylphenol	11 ± 0.4	1.2 ± 0.1	11 ± 0.8
2,5-dimethylphenol$_{WT}$	_[a]	_[b]	_[b]
2,3-dimethylphenol	10 ± 1.2	_[b]	_[b]
2,6-dimethylphenol	6.9 ± 0.4	_[b]	_[b]
4-methylphenol [72]	295 ± 2	260 ± 25	90 ± 2
1,3-dimethyl-4-nitrobenzene	196 ± 11	99.7 ± 9	51 ± 3
1,2-dimethyl-4-nitrobenzene	16.4 ± 2	_[b]	_[a]
2,4-dimethylacetophenone	11 ± 0.4	1.2 ± 0.1	11 ± 0.8
4-methylacetophenone [72]	61 ± 2	60 ± 4	96 ± 4

[a]NADH rate did not differ from leak gradient of ≈ 6 min^{-1} [b]No product formation. The wildtype control turnovers had NADH consumption rates comparable to the average leak gradient, with the exception of 2,4-dimethylphenol, which had a NADH rate of 12 min^{-1}.

The product formation rate for the monosubstituted 4-methylacetophenone was 60 min^{-1} for the S244D mutant.[72] However, the turnover for the disubstituted 2,4-dimethylacetophenone was the lowest out of all disubstituted methyl-modified substrates (1.2 min^{-1}, Table 16). The turnover formed a small amount of product at low coupling efficiency (11%), in contrast to the high coupling efficiency observed for 4-methylacetophenone (96%).[72] In the absence of an authentic product standard, the presence of a hydroxylated product for 2,4-dimethylacetophenone was confirmed via MS analysis showing a product peak with an atomic mass unit (AMU) of 164 (Appendix C6). The hydroxylated product is likely to be 4-(hydroxymethyl)-2-methylacetophenone, as the S244D mutant has been reported to oxidise benzene derivatives exclusively at the *para*-position.[70–72]

All wildtype control turnovers for the dimethyl-modified substrates showed very little product formation or, in the case of 3,4-dimethylphenol and 2,4-dimethylphenol, no product formation at all (Figure 55). Previous studies reported high product formation rates for turnovers of 4-methylphenol (260 min^{-1}).[72] In contrast, turnovers for 2,4-dimethylphenol, 3,4-dimethylphenol and 2,5-dimethylphenol all had significantly lower product formation rates (12.3 min^{-1}, 26.6 min^{-1} and 1.2 min^{-1}, respectively, Ta-

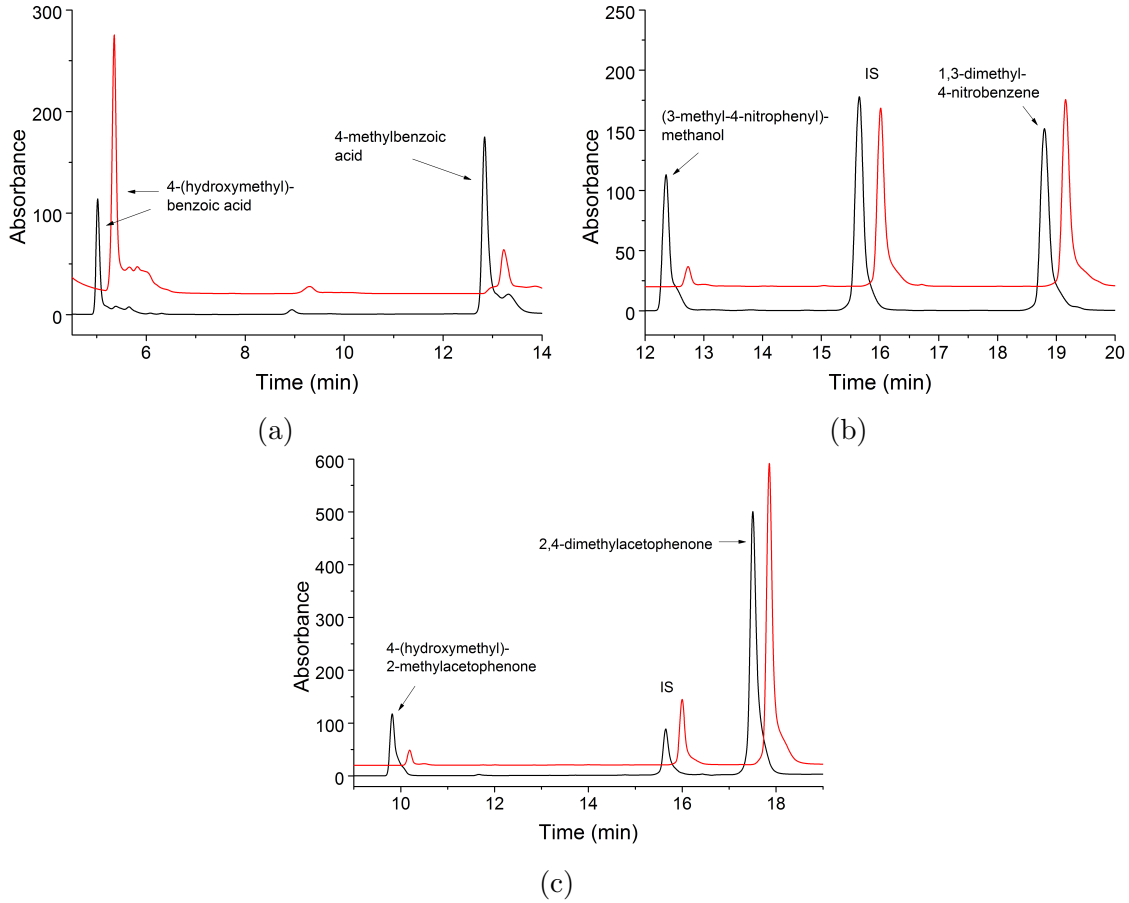


Figure 54: HPLC analysis of the *in vitro* turnovers by S244D (black) and wild-type (red) of 4-methylbenzoic acid (t_R = 12.9 min) to 4-(hydroxymethyl)benzoic acid (t_R = 5.1 mins) (a), 1,3-dimethyl-4-nitrobenzene (t_R = 18.7 mins) to (3-methyl-4-nitrophenyl)methanol (t_R = 12.3 mins) (b) and 2,4-dimethylacetophenone (t_R = 17.5 min) to 4-(hydroxymethyl)-2-methylacetophenone (t_R = 9.8 mins) (c). The internal standard (t_R = 15.8 min) is also shown. For clarity the chromatograms have been offset along the x and y axes.

ble 16). The turnover for 2,5-dimethylphenol had the lowest PFR out of these three dimethylphenol substrates, but this substrate lacks a substituent *para* to the phenol hydroxyl group (Figure 53). The substrates 2,3-dimethylphenol and 2,6-dimethylphenol also lack this *para*-methyl group and did not generate any detectable turnover product (Table 16). After derivatisation of the turnovers by BSTFA/TMSCl, a mass of 282.1 AMU was observed in the turnovers of 2,4-dimethylphenol, 3,4-dimethylphenol and 2,5-dimethylphenol, corresponding to a doubly TMS-derivatised product (Figure 55, Figure 56). This correlates to hydroxylation at one of the methyl positions. It is likely that the S244D catalysed hydroxylation process occurs at the *para*-position for 2,4-dimethylphenol and 3,4-dimethylphenol to form 4-(hydroxymethyl)-2-methylphenol and 4-(hydroxymethyl)-3-methylphenol, respectively. This is due to the highly efficient hydroxylation at the *para* position of 4-methylphenol.[72] Furthermore the lack of detectable turnover product in the turnovers of 2,6-dimethylphenol and 2,3-dimethylphenol suggests the *meta* and *ortho* substituted groups are not orientated in a suitable position for hydroxylation. As these products cannot be wholly confirmed by GC-MS analysis, a

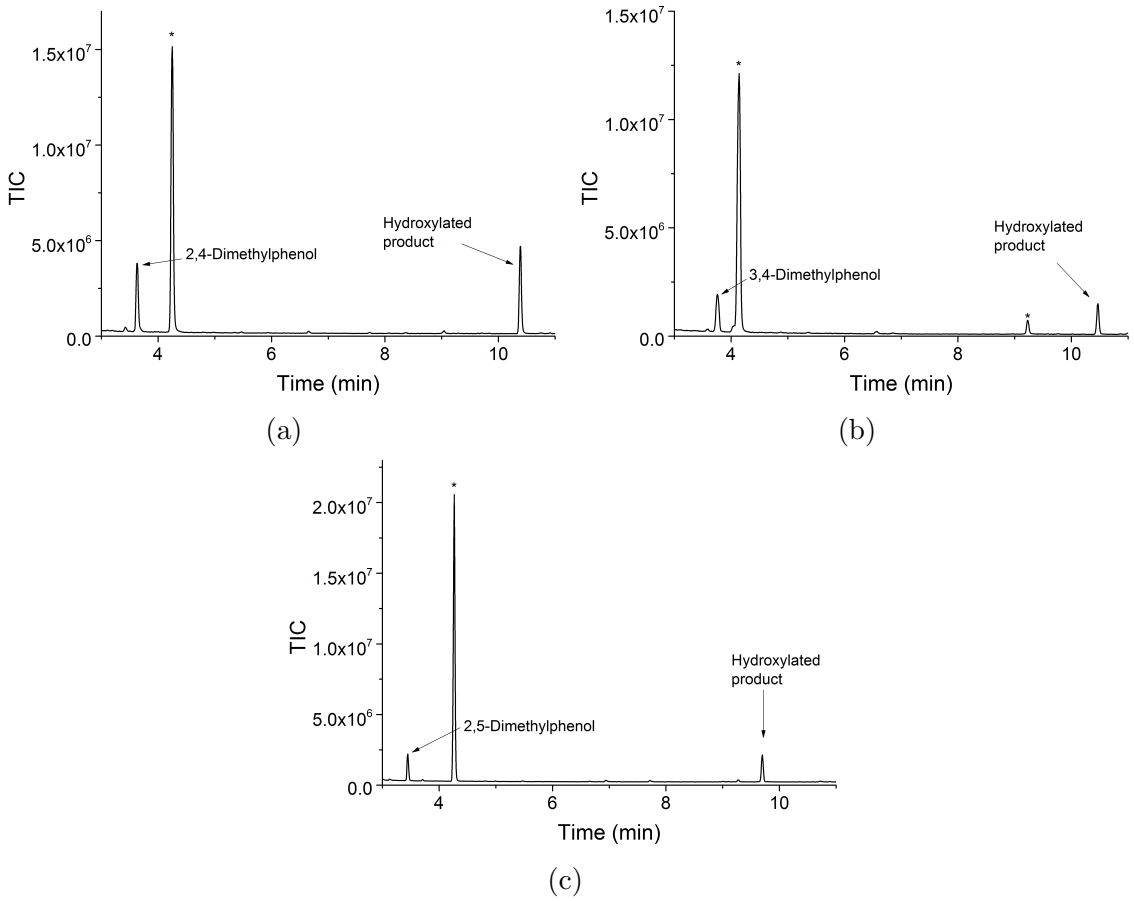
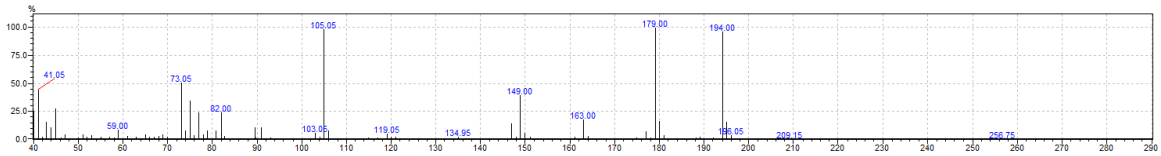
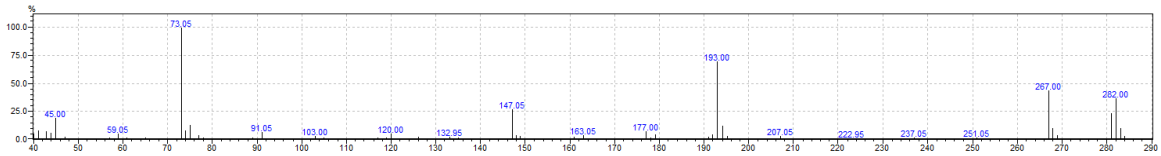


Figure 55: GC-MS analysis of the derivatised *in vitro* turnovers by S244D of 2,4-dimethylphenol (t_{R} = 3.6 min) to the assumed 4-(hydroxymethyl)-2-methylphenol (t_{R} = 10.4 mins) (a), 3,4-dimethylphenol (t_{R} = 3.75 mins) to the assumed 4-(hydroxymethyl)-3-methylphenol (t_{R} = 10.5 mins) (b) and 2,5-dimethylphenol (t_{R} = 3.4 min) to its hydroxylated product (t_{R} = 9.6 mins) (c). An impurity found in all three turnovers is observed at t_{R} = 4.1 mins. Impurities have been marked as (*).

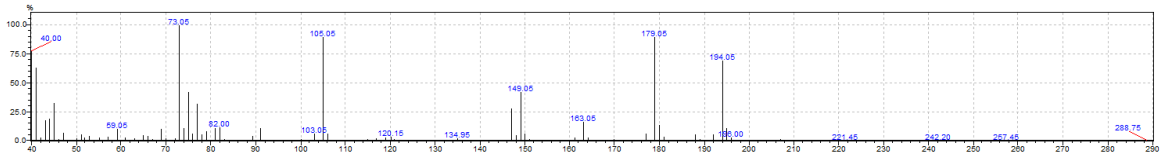
scale up turnover reaction would be required to generate enough of product for Nuclear Magnetic Resonance (NMR) analysis, which would allow confirmation of these products.



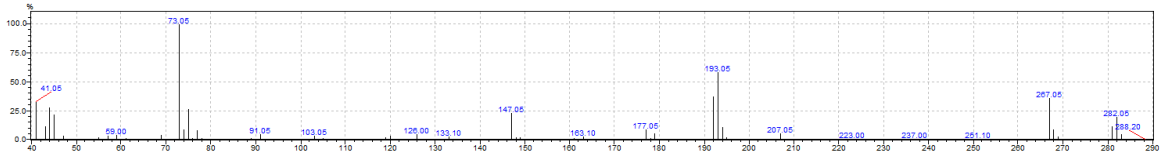
(a) 2,4-dimethylphenol (Predicted mass: 194 AMU)



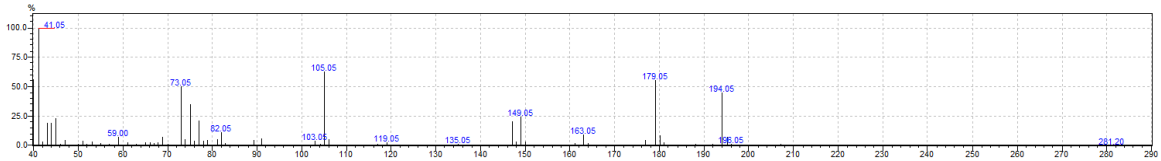
(b) Hydroxylated product of 2,4-dimethylphenol (Predicted mass: 282.1 AMU)



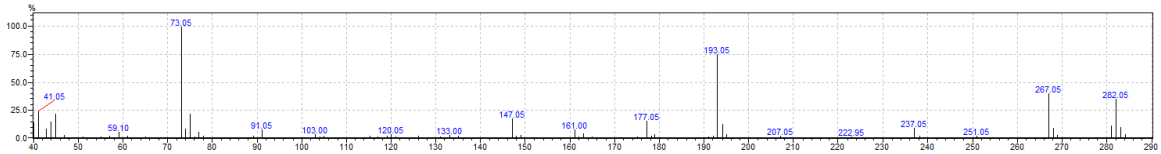
(c) 3,4-dimethylphenol (Predicted mass: 194 AMU)



(d) Hydroxylated product of 3,4-dimethylphenol (Predicted mass: 282.1 AMU)



(e) 2,5-dimethylphenol (Predicted mass: 194 AMU)



(f) Hydroxylated product of 2,5-dimethylphenol (Predicted mass: 282.1 AMU)

Figure 56: Mass spectra of 2,4-dimethylphenol (a) and its hydroxylated product (b), 2,3-dimethylphenol (c) and its hydroxylated product (d), and 2,5-dimethylphenol (e) and its hydroxylated product (f).

4.2.4 Activity and Product Formation Assays with Methylthio-modified Substrates

CYP199A4 binds and demethylates 4-methoxybenzoic acid derivatives at the *para*-position.[70] When the oxygen atom in the substituent is changed for a sulfur atom, the substrate (4-(methylthio)benzoic acid) is oxidised with high coupling efficiency.[73] Moreover, 4-(methylthio)benzoic acid is exclusively oxidised at the sulfur atom to form a sulfoxide product. To explore the efficiency of sulfoxidation activity by the S244D variant, *in vitro* turnovers of a range of methylthio-modified substrates were screened with the this mutant (Figure 57).

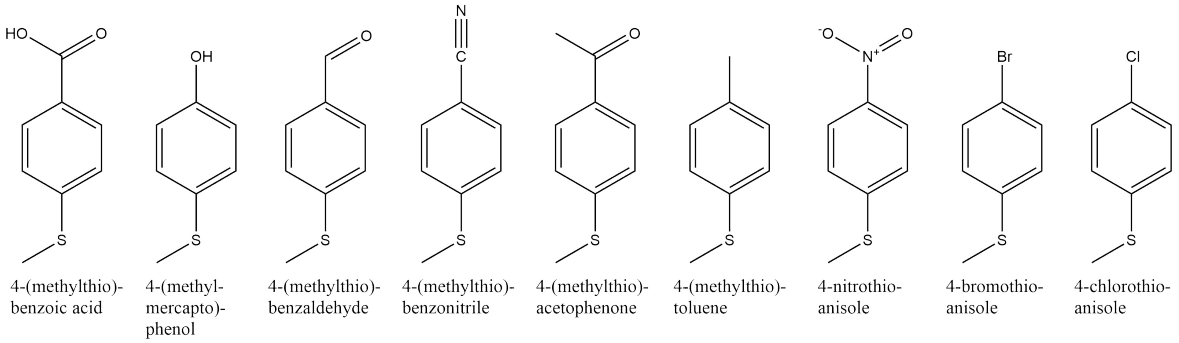


Figure 57: Methylthio-modified substrates of S244D CYP199A4

Products were identified by HPLC or GC-MS coelution of the turnovers with authentic product samples where possible. Where authentic products were not available, the sulfoxides were synthesised through a reaction of substrate (0.5 mM in EtOH) with hydrogen peroxide (3% v/v)(Figure 20, Section 2.7).

The product formation rate for the turnover of 4-(methylthio)benzoic acid was several orders of magnitude lower for the S244D variant than the wild-type enzyme (21 min^{-1} and 1175 min^{-1} respectively, Table 17), and the coupling efficiency also decreased (45% compared to 83%). This follows the trend of higher turnover rates for *para*-substituted benzoic acid substrates with wild-type CYP199A4 compared to the S244D variant.[71,73]

The turnover that generated the highest product formation rate and coupling was with 4-(methylthio)benzaldehyde (98.8 min^{-1} and 83% respectively). 4-(Methylmercapto)phenol also had a high product formation rate (66.6 min^{-1}), although at a much lower coupling efficiency (38%). The turnover of 4-(methylthio)benzaldehyde formed the sulfoxide product 4-(methylsulfinyl)benzaldehyde. A small amount of 4-(methylthio)benzyl alcohol was also present, though this may be an impurity (Figure 58c). The oxidation of 4-(methylmercapto)phenol forms only 4-(methylsulfinyl)phenol. These results are similar to that of a recent study on the hydroxylation of methyl-modified substrates with the CYP199A4 S244D mutant, where 4-methylbenzaldehyde and 4-methylphenol

Table 17: *In vitro* turnover data for CYP199A4 S244D with methylthio-modified substrates. The data are given as mean ± S.D. with $n \geq 3$. Rates are given as nmol.nmol-P450^{-1}.min^{-1} (min^{-1}).

Substrate	NADH (min^{-1})	PFR (min^{-1})	Coupling (%)
4-(methylthio)benzoic acid	46.4 ± 3.2	21 ± 3	45 ± 4.5
4-(methylthio)benzoic acid$_{WT}$ [163]	1430 ± 178	1180 ± 133	83 ± 3
4-(methylmercapto)phenol	176 ± 5.9	66.6 ± 3	38 ± 2
4-(methylthio)benzaldehyde	119 ± 8	98.8 ± 0.8	83 ± 6
4-(methylthio)benzonitrile	12.7 ± 1	7.2 ± 0.3	57 ± 6
4-(methylthio)acetophenone	11.5 ± 0.4	4.9 ± 0.5	43 ± 2.6
4-(methylthio)toluene	18.6 ± 0.3	4.9 ± 0.4	26 ± 2
1-bromo-4-(ethylthio)benzene	15 ± 1.4	0.72 ± 0.15	4.8 ± 0.9
4-fluorothioanisole	30.1 ± 2	10.7 ± 0.7	36 ± 3
4-chlorothioanisole	31.2 ± 1	2.6 ± 0.6	9.4 ± 2
4-bromothioanisole	288 ± 18	26.1 ± 3	9.1 ± 0.6
4-nitrothioanisole[a]	-	-	-

[a]Rates and coupling not determined due to interfering absorbance at 340 nm.

were reported to have the highest NADH oxidation rate, product formation rate and coupling.[72]

The oxidation of 4-(methylthio)benzonitrile to 4-(methylsulfinyl)benzonitrile occurred at a rate of 7.2 min^{-1} with 57% coupling efficiency (Figure 58b, Table 17). 4-(Methylthio)acetophenone and 4-(methylthio)toluene also shared a similar product formation rate (both 4.9 min^{-1}). The coupling efficiencies of these turnovers were lower than for 4-(methylthio)benzonitrile (43% and 26% respectively). While 4-(methylthio)benzonitrile and 4-(methylthio)acetophenone were oxidised to form the sulfoxide products 4-(methylsulfinyl)benzonitrile and 4-(methylsulfinyl)acetophenone respectively (Figure 58b and Appendix C3a), 4-(methylthio)toluene was not exclusively oxidised at the sulfur position. Very little of the sulfoxide product 4-(methylsulfinyl)toluene was observed; instead the major product was 4-(methylthio)benzyl alcohol (99%, Figure 58d).

4-Bromothioanisole and 1-bromo-4-(ethylthio)benzene differ in just one carbon atom, yet the product formation rate for the 4-bromothioanisole turnover is almost 40 times greater than the turnover for 1-bromo-4-(ethylthio)benzene (26.1 min^{-1} and 0.72 min^{-1}, respectively, Table 17). The turnover for 4-bromothioanisole with the S244D mutant oxidised NADH at a very fast rate (288 min^{-1}) but with low coupling efficiency (9.1%), suggesting significant uncoupling processes are occurring. The CYP199A4 control turnover of 4-bromothioanisole produced almost no detectable product (Figure 59a). The related substrates 4-fluorothioanisole and 4-chlorothioanisole were oxidised at the

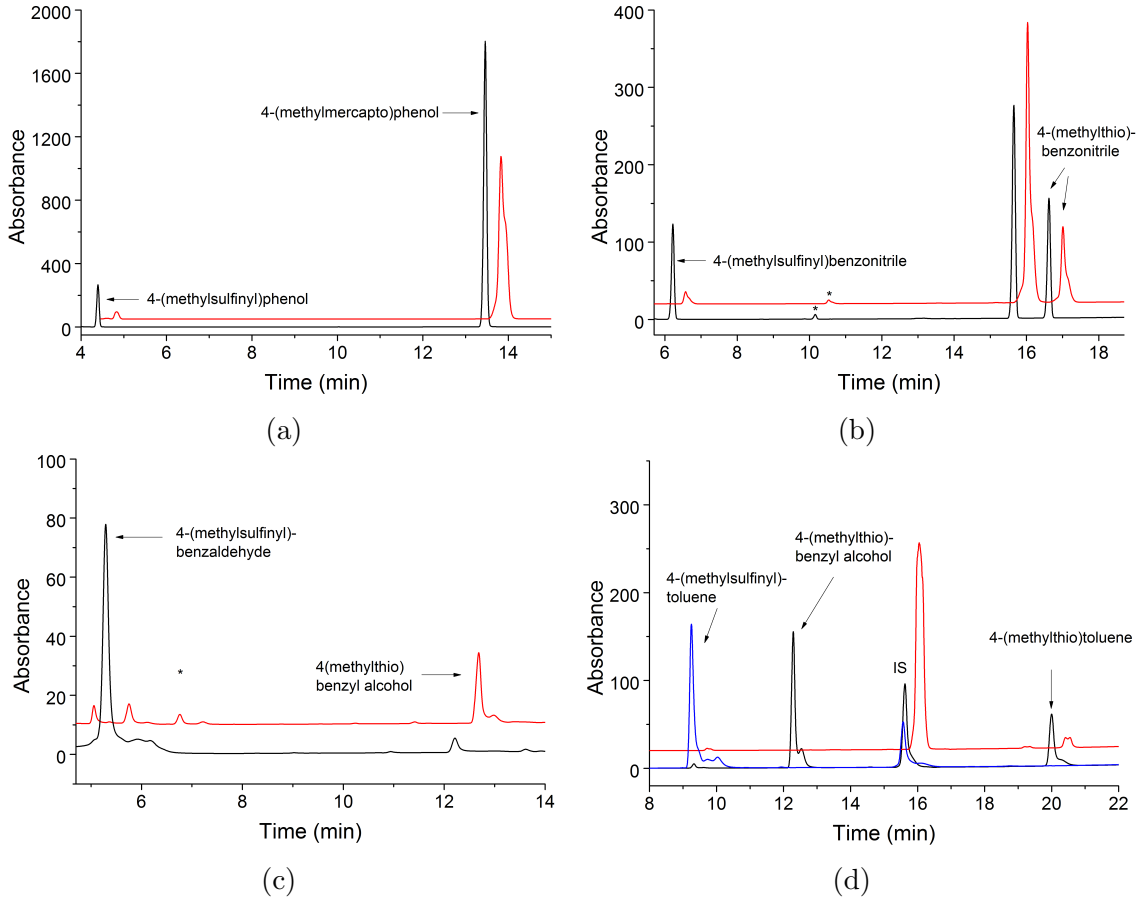


Figure 58: HPLC analysis of the *in vitro* turnovers by S244D (black) and wild-type (red) of 4-(methylmercapto)phenol (t_R = 13.5 min) to 4-(methylsulfinyl)phenol (t_R = 4.4 min) (a), 4-(methylthio)benzonitrile (t_R = 16.7 min) to 4-(methylsulfinyl)benzonitrile (t_R = 6.0 min) (b), 4-(methylthio)benzaldehyde (t_R = 15.5 min, not shown) to 4-(methylthio)benzyl alcohol (t_R = 12.2 min) and 4-(methylsulfinyl)benzaldehyde (t_R = 5.4 min) (c) and 4-(methylthio)toluene (t_R = 20 min) to 4-(methylthio)benzyl alcohol (t_R = 12.5 min) and 4-(methylsulfinyl)toluene (t_R = 9.4 min) (d). 4-(methylsulfinyl)toluene is shown as a control (in blue). Impurities have been marked as (*). The internal standard is seen at t_R = 15.6 min. The 4-(methylthio)benzyl alcohol observed in (c) may be an impurity. For clarity the chromatograms have been offset along the x and y axes.

sulfur at lower rates (PFR of 10.7 and 2.6 min^{-1}, respectively, Table 17). The coupling efficiency for 4-chlorothioanisole is similar to 4-bromothioanisole (9.4%), while for 4-fluorothioanisole it is higher (36%). Similar to 4-bromothioanisole, the sulfoxidation of 4-fluorothioanisole and 4-chlorothioanisole with the S244D mutant are much more efficient than with the wild-type, where virtually no product formation were observed in the wild-type control turnovers (Figure 59). The thioether in 4-nitrothioanisole was also oxidised by the S244D mutant to form 1-nitro-4-(methylsulfinyl)benzene (Figure 59d). However, no NADH oxidation and product formation rate was able to be calculated as the substrate absorbed UV/Vis radiation at the same wavelengths as NADH. No visible product was observed for the wild-type control turnover (Figure 59d). All product formation rates observed here were significantly lower than previously reported rates for *para*-methoxy derivatives.

Total turnovers for 4-(methylmercapto)phenol, 4-(methylthio)benzaldehyde and 4-

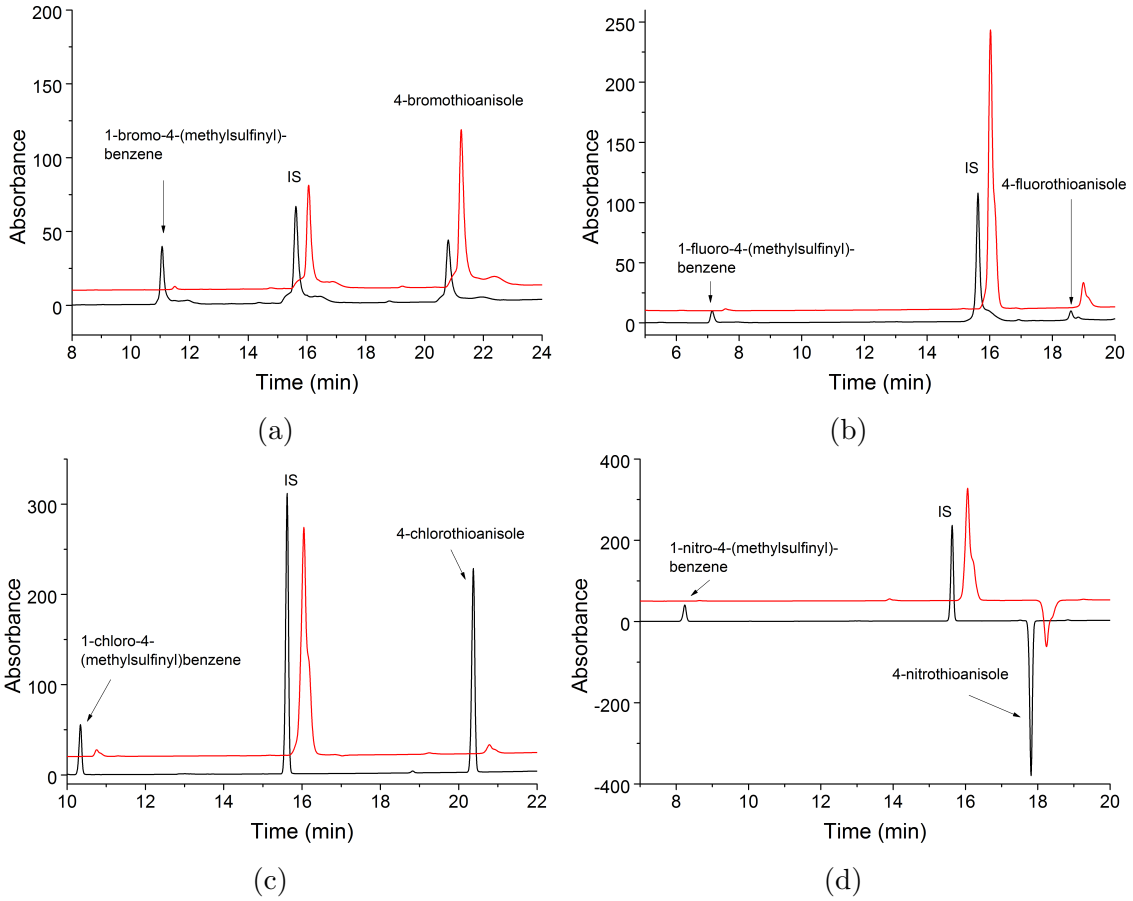


Figure 59: HPLC analysis of the *in vitro* turnovers by S244D (black) and WT (red) of 4-bromothioanisole (t_R = 20.8 min) to 1-bromo-4-(methylsulfinyl)benzene (t_R = 11 min) (a), 4-fluorothioanisole (t_R = 18.5 min) to 1-fluoro-4-(methylsulfinyl)benzene (t_R = 7.1 min) (b), 4-chlorothioanisole (t_R = 20.5 min) to 1-chloro-4-(methylsulfinyl)benzene (t_R = 10.4 min) (c) and 4-nitrothioanisole (t_R = 17.8 min) to 1-nitro-4-(methylsulfinyl)benzene (t_R = 8.1 min) (d). Impurities have been marked as (*). The internal standard is seen at t_R = 15.6 min. For clarity the chromatograms have been offset along the x and y axes.

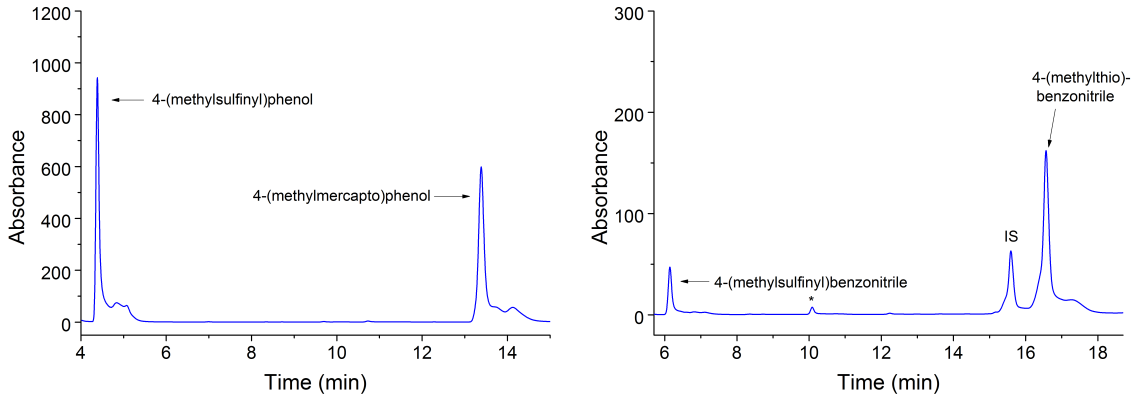
Table 18: *In vitro* total turnover data for CYP199A4 S244D with methylthio-modified substrates. These turnovers were run in the absence of ADH.

Substrate	TTN
4-(methylmercapto)phenol	910
4-(methylthio)benzaldehyde	3500
4-(methylthio)benzonitrile	450

(methylthio)benzonitrile were also analysed (Table 18, Appendix C4). Because the previous *in vitro* turnovers were limited by the amount of NADH within the system, these total turnovers gave an indication as to whether these substrates were suitable for up-scaling. The turnover for 4-(methylthio)benzaldehyde showed the greatest total turnover number (3500), which correlates its high levels of product formation. The lower total turnover numbers for 4-(methylmercapto)phenol (910) and 4-(methylthio)benzonitrile (450) also correlated with their lower product formation rates.

While the TTN of 4-(methylmercapto)phenol and 4-(methylthio)benzonitrile were low,

in vivo turnovers for 4-(methylmercapto)phenol and 4-(methylthio)benzonitrile with the CYP199A4 S244D mutant showed significant sulfur oxidation activity (Figure 60). Both generated the same products as the *in vitro* turnovers. Approximately 70% of the 4-(methylmercapto)phenol substrate (2 mM) was converted after 21 hours, while only $\approx$ 30% of the 4-(methylthio)benzonitrile turnover was converted after the same time. These *in vivo* assays show the potential for large scale sulfur oxidation by S244D.



(a) (b)

Figure 60: HPLC analysis of the *in vivo* S244D turnovers of 4-(methylmercapto)phenol (t_R = 13.5 min) to 4-(methylsulfinyl)phenol (t_R = 4.4 min) (a) and 4-(methylthio)benzonitrile (t_R = 16.7 min) to 4-(methylsulfinyl)benzonitrile (t_R = 6.0 min) (b). Impurities have been marked as (*). The internal standard is seen at t_R = 15.6 min.

In summary, exclusive sulfoxidation catalysed by the S244D mutant was reported in all the methylthio-modified substrates except for 4-(methylthio)toluene. In addition, these turnovers with S244D all formed significantly higher levels of product compared to the wild-type controls, many of which generated negligible product formation. However turnover activity was lower when compared to the equivalent methoxy substituted substrates (Section 4.2.1).[71]

4.2.5 Activity and Product Formation Assay with Styrenes

CYP199A4 effectively catalyses the oxidation of 4-vinylbenzoic acid (Figure 61).[72] The major epoxide product is 4-(oxiran-2-yl)benzoic acid and the minor product is thought to be an aldehyde rearrangement product. Alkene epoxidation activity of the S244D mutant was therefore investigated with a selection of *para*-substituted styrene compounds (Figure 62).

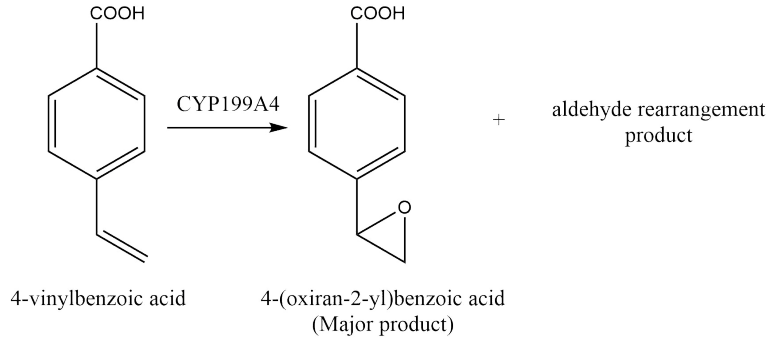


Figure 61: Oxidation of 4-vinylbenzoic acid by CYP199A4.

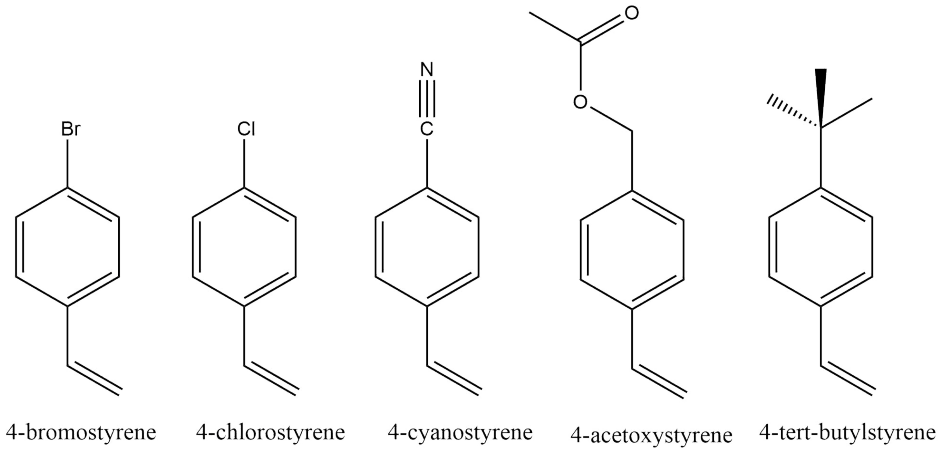


Figure 62: Styrene substrates of S244D CYP199A4

4-Bromostyrene, 4-chlorostyrene and 4-cyanostyrene were all oxidised by the S244D mutant but 4-acetoxystyrene and 4-*tert*-butylstyrene were not. This may be due to the the bulky substituents *para* to the alkene preventing binding to the active site (Figure 62). The product formation rates for the styrenes with the S244D mutant were higher than the equivalent wild-type control turnovers, but significantly lower than the wild-type turnover of 4-vinylbenzoic acid (Table 19).[72] The product formation activity of the three styrenes which were oxidised were similar ranging from 31.8 min^{-1} to 32.7 min^{-1} (Table 19). Their NADH oxidation rates were more varied (Table 19) and their coupling efficiencies ranged from 47% for 4-cyanostyrene, to 36% for 4-bromostyrene (Table 19). While no products were observed with the wild-type control turnovers of

4-chlorostyrene and 4-bromostyrene, a small amount of product was observed in the 4-cyanostyrene turnover (Figure 63a). To confirm the identity of the expected epoxide products, the turnovers were coeluted with authentic samples. Where no authentic product was available, such as 4-cyanostyrene, the epoxide was synthesised by addition of mCPBA to the substrate (Experimental 2.7). Small amounts of rearrangement products were observed in the CYP199A4 S244D turnovers of 4-bromostyrene and 4-chlorostyrene, and MS analysis suggests a aldehyde product as they posses the same mass with a fragmentation pattern of the characteristic loss of a aldehyde group (CHO, -29) (Appendix C8 and C9).

Table 19: *In vitro* turnover data for CYP199A4 S244D with *para*-substituted styrenes. The data are given as mean ± S.D. with $n \geq 3$. Rates are given as nmol.nmol-P450^{-1}.min^{-1} (min^{-1}).

Substrate	NADH (min^{-1})	PFR (min^{-1})	Coupling (%)
4-vinylbenzoic acid$_{WT}$ [72]	517 ± 3	220 ± 20	43 ± 4
4-cyanostyrene	68.2 ± 1	31.8 ± 3	47 ± 5
4-chlorostyrene	88.1 ± 7	36.4 ± 3	41 ± 0.6
4-bromostyrene	101 ± 5	36.6 ± 4	36 ± 4
4-*tert*-butylstyrene	7.9 ± 0.2	–[a]	–[a]
4-acetoxystyrene	37.3 ± 3.8	–[a]	–[a]

[a]No product formation.

Table 20: *In vitro* total turnover data for CYP199A4 S244D with styrenes. These turnovers were run in the absence of ADH.

Substrate	TTN
4-cyanostyrene	1160
4-bromostyrene	1300

The total turnover numbers of 4-bromostyrene and 4-cyanostyrene were determined to be 1300 and 1160, respectively (Table 20, Appendix C5). While these are lower than the potential maximum turnover number, *in vivo* turnovers of 4-bromostyrene and 4-chlorostyrene showed relatively high substrate conversion (Figure 64). 2 mM of substrate was added to the *in vivo* system, and 19% of 4-bromostyrene was calculated to be converted into product. The aldehyde rearrangement product was not observed, possibly due to further reactivity of this rearrangement product within the whole cell system. 4-Cyanostyrene also underwent epoxidation within the whole cell system. Less than 3% of the substrate remained after 21 hours, which was converted entirely into 2-(4-cyanophenyl)oxirane. This shows the potential for scaling up the epoxidation activity of S244D.

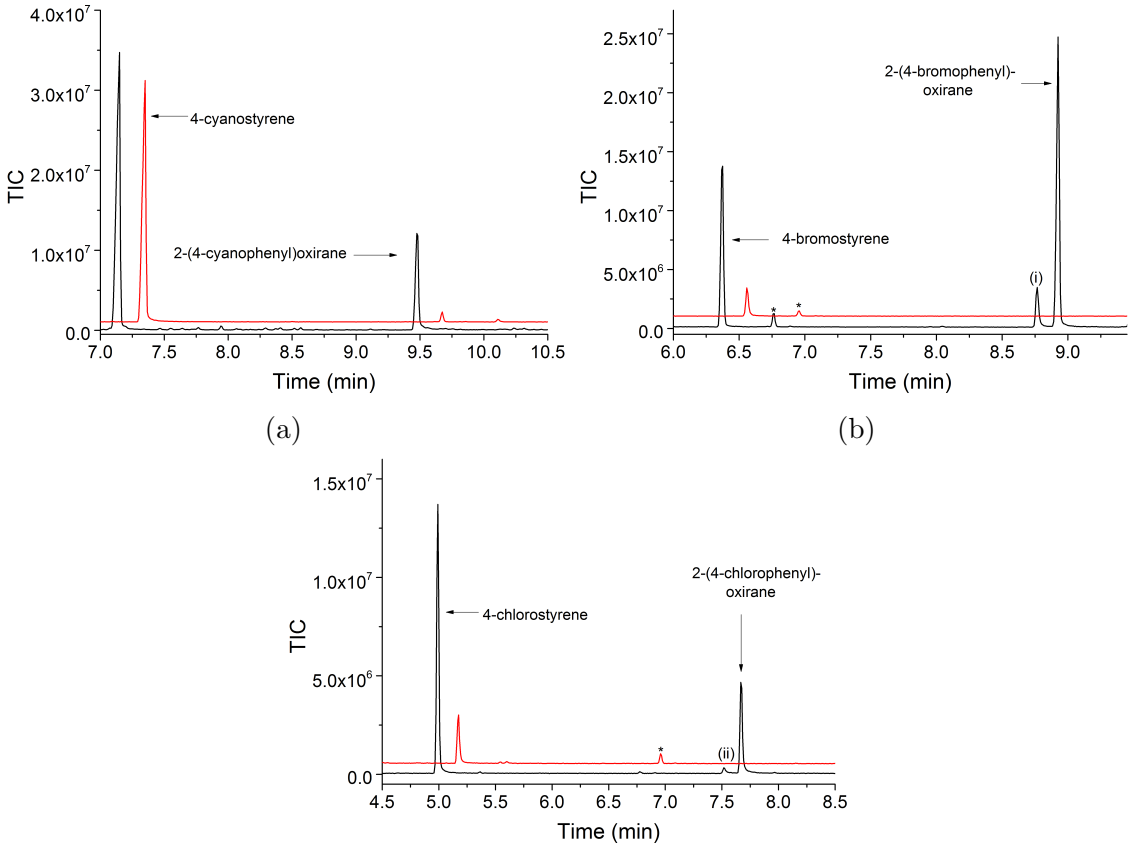


Figure 63: GC-MS analysis of the *in vitro* turnovers by S244D (black) and wild-type (red) of 4-cyanostyrene (t_R = 7.2 min) to 2-(4-cyanophenyl)oxirane (t_R = 9.5 mins) (a), 4-bromostyrene (t_R = 6.3 mins) to 2-(4-bromophenyl)oxirane (t_R = 8.9 mins) (b) and 4-chlorostyrene (t_R = 5.0 min) to 2-(4-chlorophenyl)oxirane (t_R = 7.7 mins) (c). Both the *m/z* (199.04) and mass spectrum of the peak at [i] and the *m/z* (154.59) and mass spectrum of the peak at [ii] suggests these are aldehyde rearrangement products (Figure 9) Impurities have been marked as (*). For clarity the chromatograms have been offset along the x and y axes.

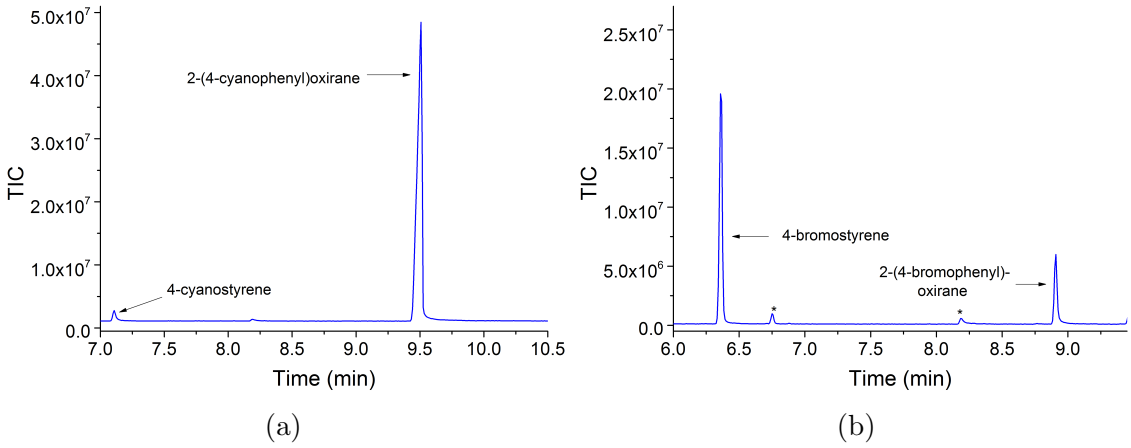


Figure 64: GC-MS analysis of *in vivo* S244D turnovers of 4-cyanostyrene (t_R = 7.2 min) to 2-(4-cyanophenyl)oxirane (t_R = 9.5 mins) (a) and 4-bromostyrene (t_R = 6.3 mins) to 2-(4-bromophenyl)oxirane (t_R = 8.9 mins) (b). Impurities have been marked as (*).

4.3 Discussion: Turnover Activity

The total turnover numbers for methoxy-modified substrates generally followed the trend in the PFR and NADH oxidation rates. The high TTN for methoxy- and ethyl-modified substrates show the potential for the oxidation of these compounds by the S244D mutant for large scale application. Styrenes, methyl- and methylthio- modified substrates showed significantly lower total turnover numbers. Differences in TTN could be due to the solubility of the substrate. A highly soluble substrate such as 4-methoxybenzoic acid would be expected to have a higher TTN compared to a substrate with low solubility which may precipitate. Both the substrate and the product formed in these turnovers may also be toxic to the P450, as such the total turnover numbers can vary between substrates with similar product formation rates and coupling efficiencies.

The crystal structure of the CYP199A4 wild-type with both 4-methoxybenzoic acid and 4-ethylbenzoic acid (PDB: 4DO1 and 4EGM) infers that a methyl group in 4-methylbenzoic acid will be further away from the heme iron.[21] Studies of methoxy, methyl and ethyl-modified substrates with CYP199A4 and its S244D variant have also revealed greater binding affinity, product formation rates and spin-state shifts for methoxy- and ethyl-modified substrates over methyl-modified substrates.[10,71,72] Here the lower total turnover numbers of CYP199A4 S244D for methyl substituted benzene derivatives compared to methoxy- and ethyl-modified compounds highlight the imperfect fit of these methyl-modified substrates within the binding pocket.

The S244D mutant preferentially hydroxylates dimethyl-substituted benzene derivatives at the *para* position. Dimethyl-modified substrates which lacked a *para*-methyl moiety showed little or no activity. 2,5-Dimethylphenol was the only substrate that was oxidised. CYP199A4 oxidatively demethylated 3,5-dimethoxybenzoic acid to 3-hydroxy-5-methoxybenzoic acid in low amounts.[70] This substrate also lacks a *para*-substituted moiety, yet the enzyme was able to demethylate at the *meta*-position regardless. Thus, while CYP199A4 almost always exclusively oxidises at the *para*-position, if a substrate lacks a *para*-substituent, small amounts of oxidation can still occur. Similarly, the S244D turnover of 2,5-dimethylphenol demonstrates the ability of the CYP199A4 variant to hydroxylate a substrate that lacks a *para*-substituted group. 2,5-Dimethylphenol is likely to be hydroxylated to form 5-(hydroxymethyl)-2-methylphenol, as the *meta* position may be closer to the heme iron than the *ortho* position. However, at this point we cannot rule out an alternative binding orientation of this substrate.

The turnovers of the *para*-substituted styrene substrates showed similar oxidation rates. The highly electron-withdrawing nitrile group in 4-cyanostyrene did not significantly decrease product formation compared to 4-chlorostyrene and 4-bromostyrene turnovers. Rearrangement products are also observed for the oxidation of these olefins (Figure 65).

The formation of the ketone/aldehyde rearrangement products of 4-bromostyrene and 4-chlorostyrene proceeds in a non-concerted fashion. The turnover for 4-cyanostyrene did not produce a ketone/aldehyde product. The electron withdrawing nitrile group would destabilise a cation intermediate, thus the concerted mechanism involving a radical intermediate may be the favoured pathway resulting in exclusive epoxide formation.

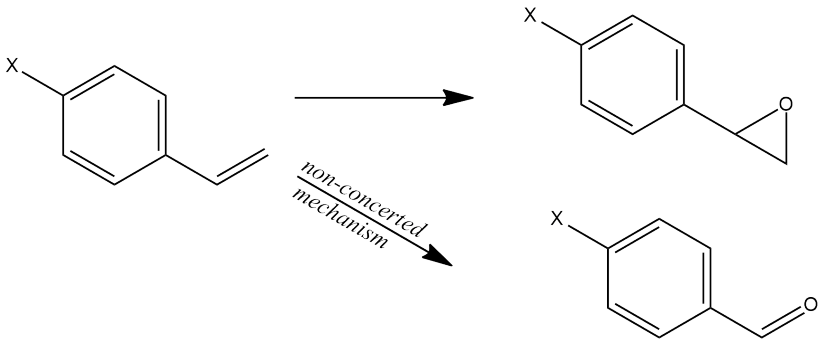


Figure 65: CYP199A4 S244D catalysed *para*-halogenated styrene oxidation.

Dealkylation of the methylthio-modified substrates were not observed. This contrasts with the methoxy-modified substrates which were exclusively demethylated. All the turnovers of the methylthio-modified substrates generated a sulfoxide, and had significantly slower product formation rates than the dealkylation of methoxy-modified compounds. The exception was 4-(methylthio)toluene, which was almost exclusively oxidised at the methyl group. This suggests that the majority of methylthio-modified substrates bind in the active site in a similar orientation as the methoxy- and methyl-modified substrates, with the thioether moiety closely positioned above the heme iron centre. Further oxidation to a sulfone was not observed for any of the methylthio-modified substrates, suggesting that further oxidation is not favoured, or that the sulfoxide product does not remain within the active site.

The binding pocket of the aspartate 244 mutation, which would interact with the carboxylate group of a 4-methoxybenzoic acid substrate is suggested to prefer hydrophilic moieties (Figure 15).[71] A recent study reported that 4-methoxytoluene was turned over by S244D to form 4-methoxybenzyl alcohol as a minor product and 4-hydroxybenzyl alcohol as the major product (Figure 66).[71] The hydrophilic pocket may prefer the more hydrophilic methoxy and methylthio groups, positioning the more hydrophobic methyl group above the heme iron-centre for hydroxylation. However, it was unknown if the major product 4-hydroxybenzyl alcohol was formed by further oxidation of 4-methoxybenzyl alcohol, or 4-methoxyphenol, which is not observed (Figure 66).[71] The preference of the binding pocket for hydrophilic moieties is supported by the turnover of 4-(methylthio)toluene being exclusively oxidised to 4-(methylthio)benzyl alcohol with no further oxidation (Figure 66).

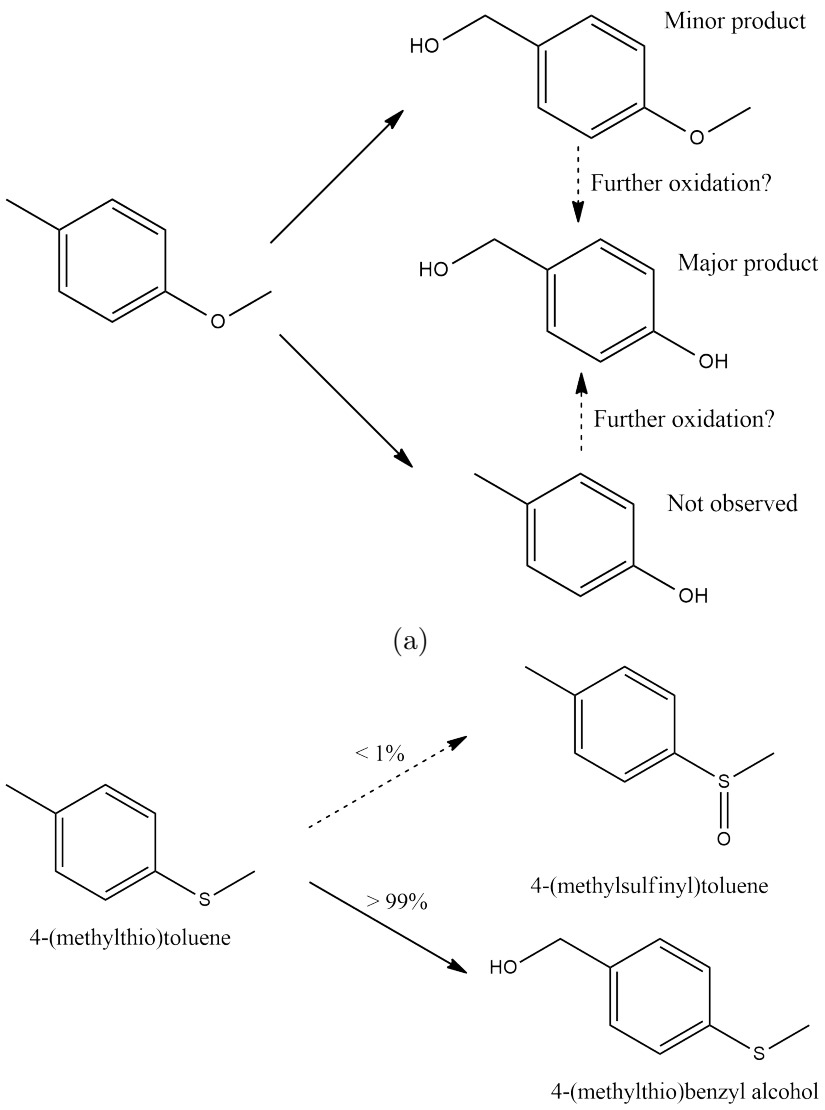


Figure 66: Oxidation of (a) 4-methoxytoluene to 4-hydroxybenzyl alcohol and 4-methoxybenzyl alcohol and (b) 4-(methylthio)toluene and 4-(methylthio)benzyl alcohol by CYP199A4 S244D.

The product formation rates for methylthio-modified substrates followed the trends for methyl-, ethyl- and methoxy-modified substrates. For example, the oxidation of 4-(methylthio)benzaldehyde with the S244D mutant was the highest of all the methylthio-modified compounds. This is also the case with the oxidation of *para*-methoxy-, methyl- and ethyl-benzaldehyde substrates with the S244D variant.[71,72] Therefore the aldehyde must have a highly favourable interaction with the aspartate 244 amino acid to position the sulfur group above the active iron-heme site for effective oxidation activity. The interactions between the negatively charged aspartate 244 in the S244D mutant and the carboxylic acid groups on *para*-substituted benzoic acids are thought to be the main reason why lower affinity and activity was observed for these substrates compared to the wild-type enzyme.[71] The H-bond-accepting carboxylate moiety in the substrate is replaced with a H-bond-donating phenol in 4-methoxyphenol. If the H-bond-donating

side chain of serine is replaced with an H-bond-accepting aspartate residue in the S244D variant, the acid-alcohol interaction is retained but reversed for a 4-methoxyphenol bound S244D mutant. As a result most *para*-substituted phenols also have a high affinity and activity with the S244D mutant. Here 4-(methylmercapto)phenol produced one of the highest product formation rates of all thioethers, further suggesting favourable interactions between the phenol and aspartate 244.

4.4 Results: Enantioselectivity of CYP199A4 S244D

Oxidation of sulfur containing compounds to sulfoxides generates a stereocenter on the sulfur atom. In order to determine the enantioselectivity of CYP199A4 S244D catalysed sulfoxidation reactions, the turnovers were analysed by chiral GC and HPLC. The turnovers were coeluted with available product standards (mixtures of enantiomers). Where authentic products were not available, a racemic mixture of the sulfoxides were synthesised by sulfoxidation using H_2O_2 as outlined in Section 2.7.

HPLC analysis enabled the separation of the enantiomers of 4-(methylsulfinyl)phenol. When the turnover of 4-(methylmercapto)phenol was coeluted with racemic 4-(methylsulfinyl)phenol, both enantiomers of the sulfoxide are observed, with a bias for one of the enantiomers present (Figure 68a). The precise enantioselectivity was not able to be accurately determined due to poor separation (≈ 75:25).

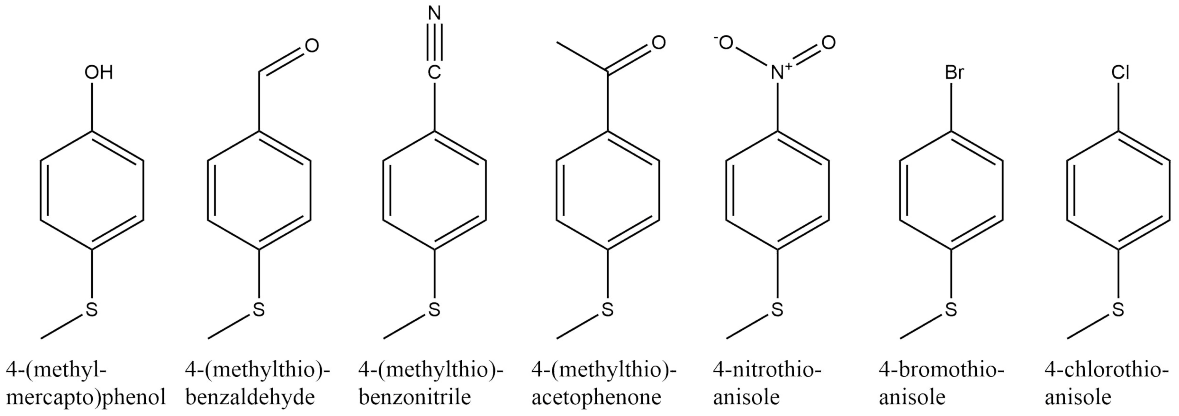
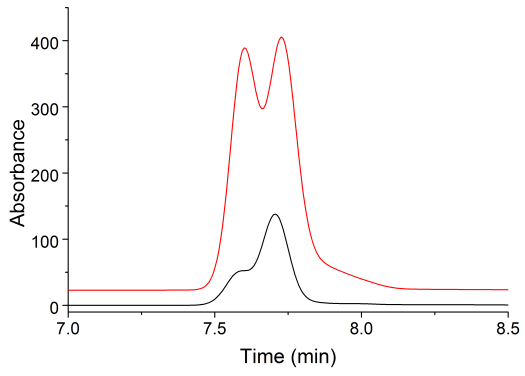


Figure 67: Methylthio-modified substrates of S244D CYP199A4 with chiral products



(a)

Figure 68: Chiral HPLC analysis of CYP199A4 S244D turnover (black) of 4-(methylmercapto)phenol (a). The authenticated products 4-(methylsulfinyl)phenol is shown in red. For clarity the chromatogram have been offset along the y axis.

The enantiomers arising from the turnovers for 4-(methylthio)benzaldehyde, 4-(methylthio)benzonitrile and 4-(methylthio)acetophenone could be partially separated

by chiral GC analysis. When these turnovers were coeluted with a racemic mixture of their potential products (4-(methylsulfinyl)benzaldehyde, 4-(methylsulfinyl)benzonitrile or 4-(methylsulfinyl)acetophenone), both enantiomers were observed as products (Figure 69). The turnover for 4-(methylthio)benzaldehyde, 4-(methylthio)benzonitrile and 4-(methylthio)acetophenone produced an enantiomeric excess of 40%, 58% and 46% respectively for the earlier eluting enantiomer. The turnovers for 4-chlorothioanisole, 4-bromothioanisole, 4-fluorothioanisole and 4-nitrothioanisole were also analysed on the GC chiral column. Like the other sulfoxidation reactions catalysed by the S244D mutant, these turnovers revealed a preference for one of the two enantiomeric products. This was most apparent in the turnover of 4-fluorothioanisole to 1-fluoro-4-(methylsulfinyl)benzene, which produced one enantiomer at an enantiomeric excess of 90% (Figure 70c). The turnovers for 4-nitrothioanisole, 4-bromothioanisole and 4-chlorothioanisole showed lower enantiomeric selectivity to 1-nitro-4-(methylsulfinyl)benzene (53% ee), 1-bromo-4-(methylsulfinyl)benzene (42% ee) and 1-chloro-4-(methylsulfinyl)benzene (41% ee) respectively (Figure 70).

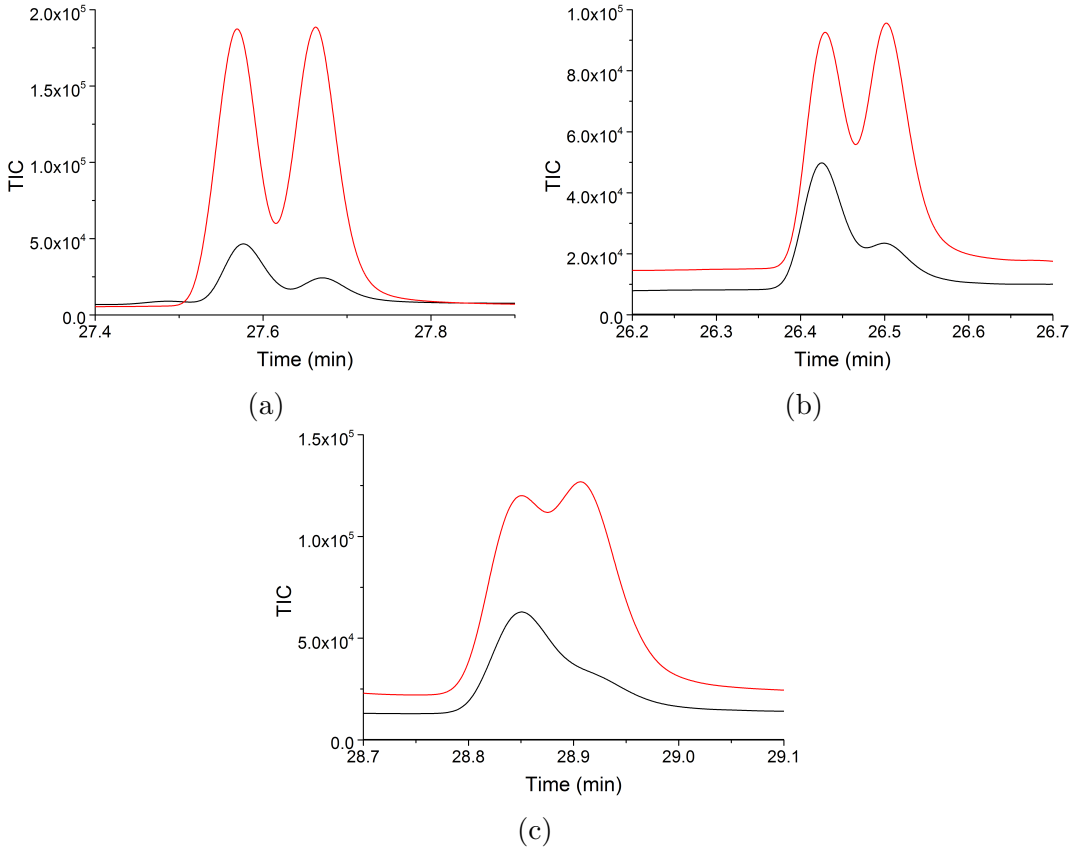


Figure 69: Chiral GC analysis of CYP199A4 S244D turnovers (black) of 4-(methylthio)benzonitrile (a), 4-(methylthio)benzaldehyde (b), and 4-(methylthio)acetophenone (c). The synthesised racemic sulfoxidation product standards are shown in red. For clarity the chromatogram have been offset along the y axis.

The most abundant stereoisomer was the earlier eluting enantiomer in the chiral GC analysis using a column which retains more polar compounds. The opposite is true

for the reverse phase HPLC chiral column, with the later eluted enantiomer found in excess. Therefore for all the turnovers presented here, it seems that the enantiomer in excess is always in same $(R)/(S)$ conformation. It was shown that wild-type CYP199A4 turnovers for 4-(ethylthio)benzoic acid and 4-ethylbenzoic acid were enantioselective for the S enantiomer.[163] A 91:9 and 86:14 S:R product ratios were observed in the turnovers of 4-(ethylthio)benzoic acid and 4-ethylbenzoic acid to 4-(ethylsulfinyl)benzoic acid and 4-(1-hydroxyethyl)benzoic acid respectively. The turnover products of CYP199A4 S244D catalysed sulfoxidation may also favour the S enantiomer if the substrate is bound in a similar orientation in the S244D and wild-type enzymes.[163]

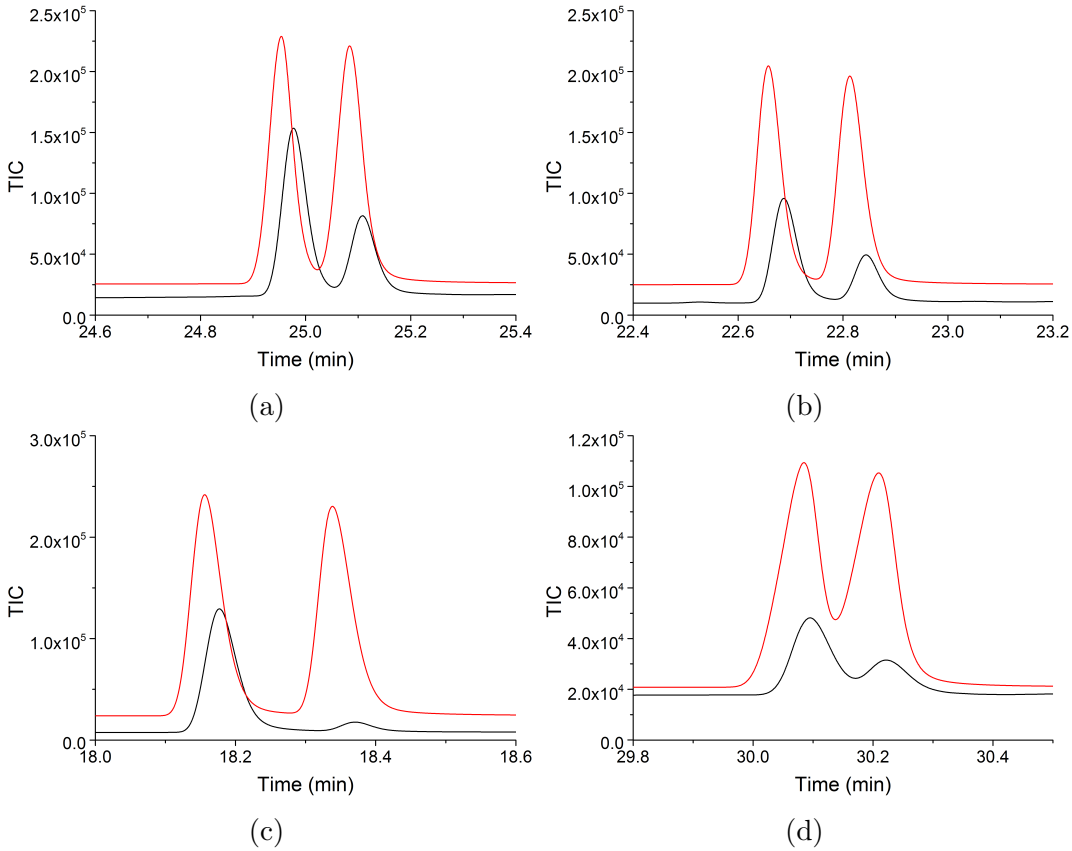


Figure 70: Chiral GC analysis of CYP199A4 S244D turnovers (black) of 4-bromothioanisole (a), 4-chlorothioanisole (b), 4-fluorothioanisole (c), and 4-nitrothioanisole (d). The synthesised racemic sulfoxidation product standards are shown in red. For clarity the chromatogram have been offset along the y axis.

Epoxidation of alkenes generates chiral products, meaning the S244D turnovers with styrenes could contain a mixture of enantiomers. When 4-cyanostyrene was coeluted with a racemic mixture of 2-(4-cyanophenyl)oxirane, a preference for one of the enantiomeric products was observed (72% ee, Figure 71a). Likewise, the turnover of 4-bromostyrene also showed a preference for one of the enantiomers of 2-(4-bromophenyl)oxirane (36% ee, Figure 71b). The turnover for 4-chlorostyrene revealed exclusive enantioselectivity for only one of the possible enantiomers from the epoxidation reaction when coeluted with a racemic mixture of the product 2-(4-chlorophenyl)oxirane (Figure 71c).

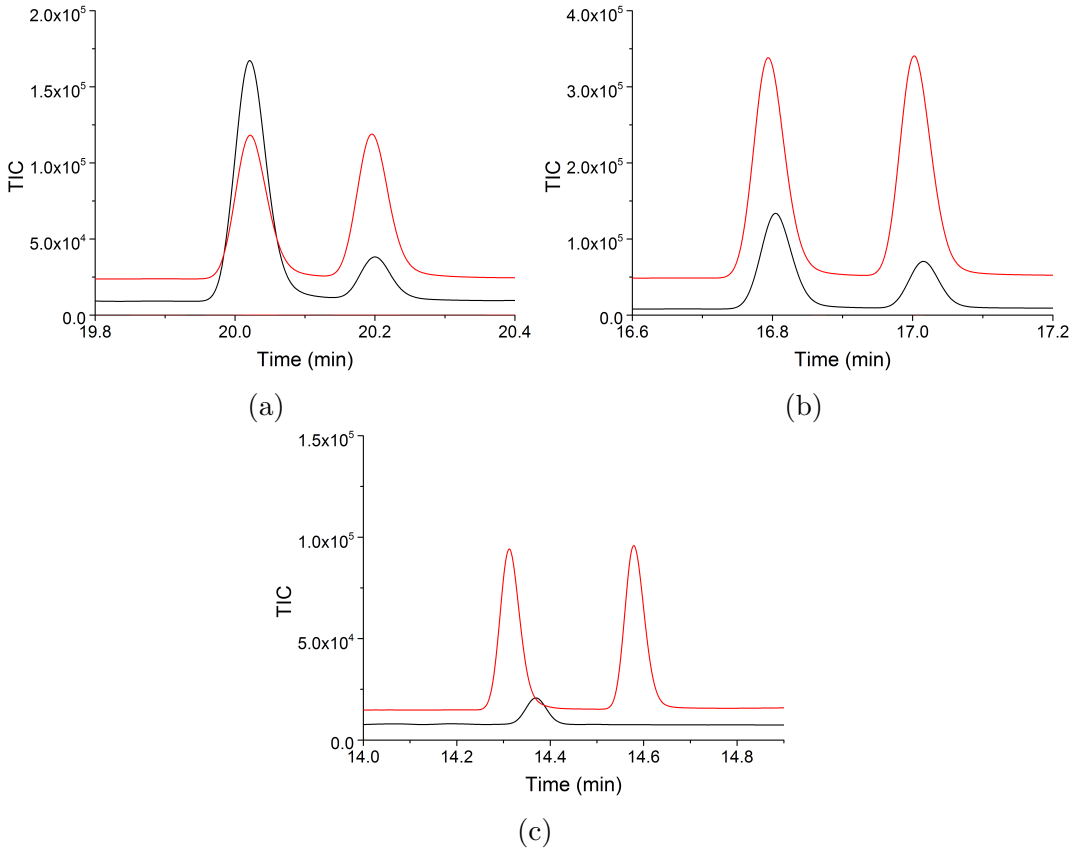


Figure 71: Chiral GC analysis of CYP199A4 S244D turnovers (black) of 4-cyanostyrene (a), 4-bromostyrene (b) and 4-chlorostyrene (c). The epoxidation products are shown in red. For clarity the chromatogram have been offset along the y axis.

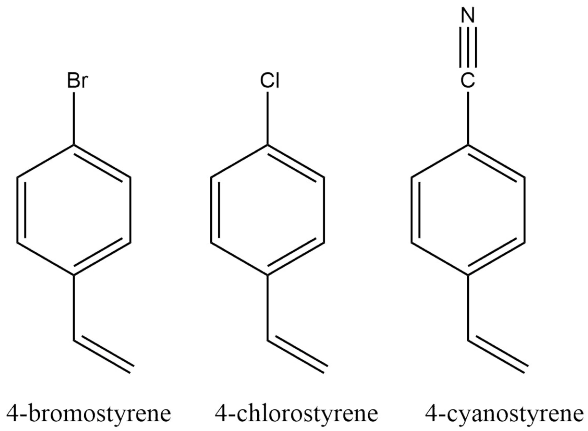


Figure 72: Styrene substrates of S244D CYP199A4 with chiral products

The CYP199A4 S244D mutant has a degree of enantioselectivity for sulfur oxidation and epoxidation. While it is not completely selective for a single enantiomer, the enantiomeric excess suggests the substrates are orientated above the heme-active site in such a way that stereoselective oxidation will occur.

4.5 Results: Crystal structure of CYP199A4 S244D

Several benzoic acid substrate-bound CYP199A4 crystal structures have been solved, including 4-methoxybenzoic acid, 4-ethylbenzoic acid, veratric acid and indole-6-carboxylic acid (PDB:4DO1, 4EGM, 4EGN and 4EGO respectively).[10,21] These studies revealed the carboxylic acid moiety of the substrate interacts through hydrogen bonding with the side chains of active site amino acids Arg92, Ser95 and Ser244 (Figure 73). The amino acid Arg243 also interacts with the carboxylate group through a bridging water molecule. The side chains of Leu98, Phe185 and Ala248 help hold the substrate within the active site via strong van der Waals interactions to the substrate benzene ring.[21] Phe182 and Phe298 both interact with the methoxy group (Figure 73).[10]

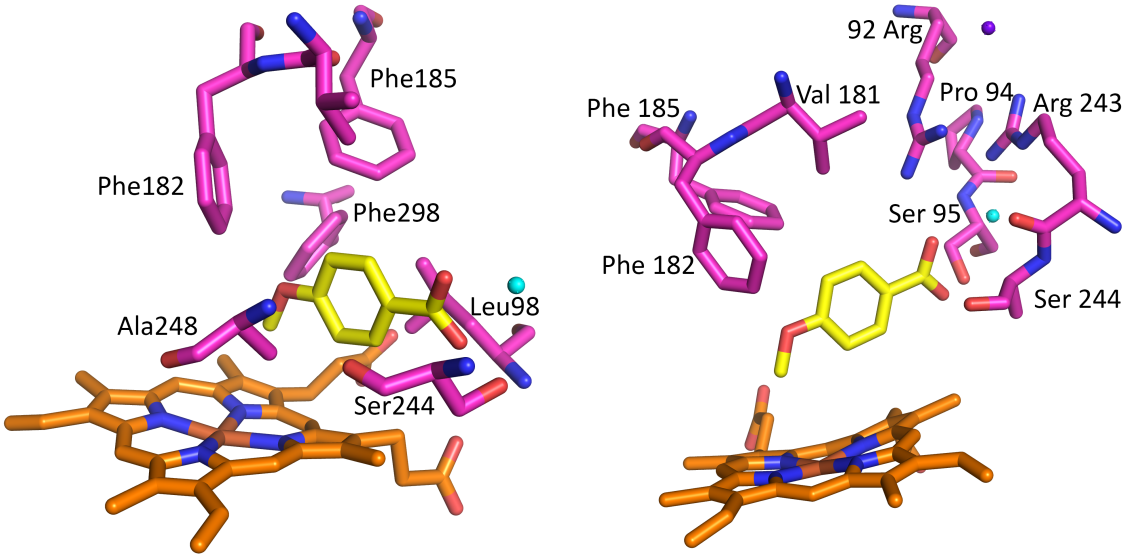


Figure 73: The crystal structure of the active site of 4-methoxybenzoic acid bound CYP199A4.[21] Selected active site residues are shown in magenta, while the substrate is highlighted in yellow. The heme ring is in orange. The bridging water molecule is shown as a cyan sphere, while the chloride ion is shown as a purple sphere.

Serine 244 forms a hydrogen bond with the carboxylate moiety of the substrate. By mutating this residue to the acidic aspartate residue, the substrate range of CYP199A4 expanded to include other *para*-substituted benzene derivatives including benzaldehydes and phenols. In order to determine the impact the S244D mutation has on the active site, the S244D mutant was co-crystallised with 4-methoxybenzoic acid. The crystal growth of the 4-methoxybenzoic acid bound CYP199A4 S244D was done by Tom Coleman.[163] The S244D protein was first concentrated to 40 mg mL^{-1} in crystallisation buffer Tris (50 mM, pH 7.4). Initial crystal screening was carried out with Hampton Research Crystal Screens HT and PEG/ION (Experimental 2.8). The conditions for crystal growth were optimised, and the final crystal conditions were 0.1 M Bis-Tris (pH 5.25), PEG 3350 (22%), MgAc (0.2 M).[163] The crystals were chosen in accordance to the size and ease of extraction for X-ray diffraction. Diffraction data were collected from the Australian

synchrotron MX1 beamline (0.9537 Å). The crystal structure of 4-methoxybenzoic acid bound CYP199A4 S244D mutant was solved at 1.89 Å resolution using molecular replacement with CYP199A4 wild-type co-crystallised with 4-methoxybenzoic acid as a search model (PDB: 4DO1).[21] The crystallography data and refinement statistics are reported in Table 21 (PDB: 5U5J).

Table 21: Data collection and refinement statistics of CYP199A4 S244D

	CYP199A4 S244D (PDB: 5U5J)
Data collection statistics[a]	
Wavelength (Å)	0.9537
Space group	$P2_1$
Cell dimensions *a*/*b*/*c* (Å)	41.29/51.22/79.18
Resolution (Å)	43 - 1.89 (1.93 - 1.89)
Average I/σ (I)	5.7 (1.1)
Completeness (%)	96.0456 (96.0419)
Redundancy	6.2 (4.8)
R_{merge}	30.4 (124.6)
R_{pim}	13 (63.8)
$CC_{1/2}$	97.5 (45.4)
Structure refinement statistics	
Average B-factor ($Å^{-1}$)	13.65
R_{work}/R_{free} (%)	17.36/26.29
r.m.s.d bond lengths (Å)	0.007
r.m.s.d bond angles (°)	1.028
Ramachandran favoured [%]	96.9
Ramachandran outliers [%]	0
MolProbability score	1.49
Poor rotamers [%]	0
Clash scores, all atoms	5.89

[a]Values in parentheses correspond to the highest resolution shell.

The root-mean-square deviation (r.m.s.d) between the 4-methoxybenzoic acid bound S244D and equivalent CYP199A4 wild-type crystal was 0.321 $Å^2$ (Figure 74). This low value was a result of a very similar overall structure with between the two models. Almost all of the active site residues appear to interact in a similar way to the 4-methoxybenzoic acid substrate (Table 22). For example, C^{ζ}-Phe182 and C^{ζ}-Phe298 both interact with the methoxy group of the substrate with the distances for both the wild-type and S244D variant being similar (Table 22). However, some important

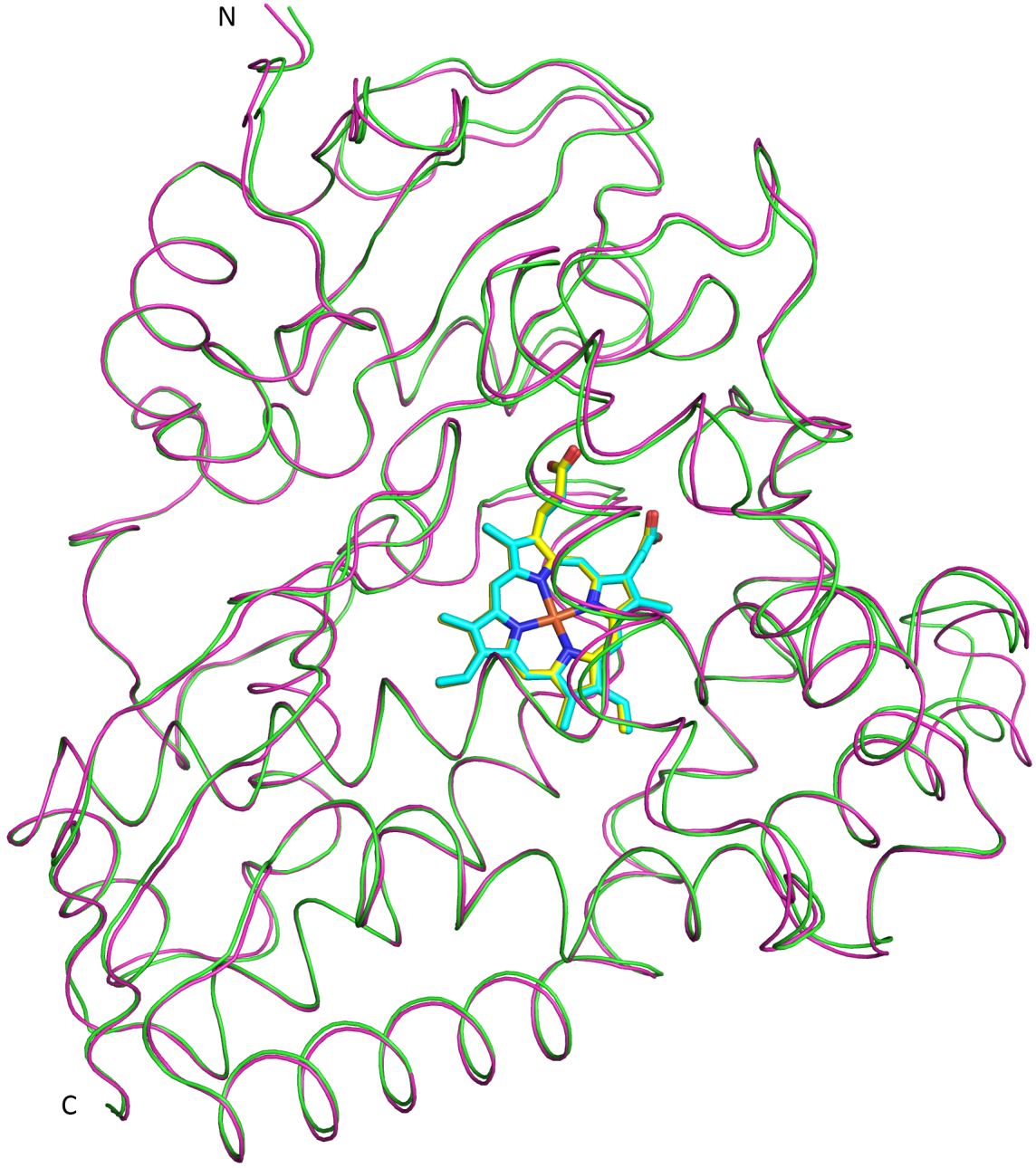


Figure 74: Cα trace of superimposed S244D mutant structure of CYP199A4. The CYP199A4 backbone is in magenta with the heme unit shown in cyan, with the overlaid S244D backbone in green with the heme unit shown in yellow. The N and C terminus is indicated in each. The structures displayed a r.m.s.d of 0.321 Å.

residues show altered positions between the S244D variant and its wild-type. The distance between the closest atom of the carboxylate group of 4-methoxybenzoic acid to $O^{\delta 1}$-Asp244 is shorter compared to the wildtype equivalent $O^{\gamma 1}$-Ser244 (2.4 Å compared to 2.6 Å) (Table 22). This can be partly attributed to the longer side chain of aspartate compared to serine. The S244D mutation also shifts the 4-methoxybenzoic acid substrate such that the methoxy-group is slightly further away from the heme-iron centre (4.2 Å compared to 4.1 Å, Figure 75a). A water molecule bridges the substrate with several

binding site amino acids. A minor shift in the position of the bridging water molecule (Wat31 in S244D) in relation to the substrate and amino acids Arg243 and Ser/Asp244 is observed (Table 22). The bond between $O^{\delta 1}$-Asp244 and the bridging water is shorter compared to the wildtype equivalent $O^{\gamma 1}$-Ser244 (2.7 Å compared to 3.0 Å). Interestingly in the wild-type, the carboxylate group in 4-methoxybenzoic acid is slightly closer to the bridging water molecule compared to the S244D mutant (2.6 Å compare to 2.8 Å, Table 22). These subtle differences may have an impact on the binding affinity of the substrate to the S244D variant, as the S244D mutant has a binding affinity several orders of magnitude lower for 4-methoxybenzoic acid than the wild-type.[70–72]

The structure of CYP199A4 show that the entrance of the active site for CYP199A4 is capped by a species which protects the the active site from external solvents while there is substrate bound (Figure 76).[21] Electron density at this position was modelled as an anion, with chloride being the closest fit. Similarly electron density near the entrance of the active site within the S244D variant was also modelled as a chloride ion. However, the position of this chloride ion is significantly different, and it coordinates to different amino acids. The chloride ion in S244D exhibits an irregular four-fold coordination with the phenolic OH of Tyr177 (2.8 Å), a water molecule (Wat232, 2.4 Å) and the N and C=O groups of the carboxamide of Asn207 (3.3 and 2.7 Å respectively, Figure 76b). Compared to the wild-type structure, the conformation of the asparagine 207 side chain has rotated to face towards the chloride ion in S244D. This facilitates the interaction of Asn207 with the chlorde (Figure 76). The chloride ion in the wild-type coordinates to two water molecules, but the different position of the chloride ion in the S244D structure and the movement of Asn207 may have caused the removal or displacement of one of the coordinating water molecules. Without this second water molecule, the chloride ion can coordinate with both the N and C=O groups in the Asn207 side chain.

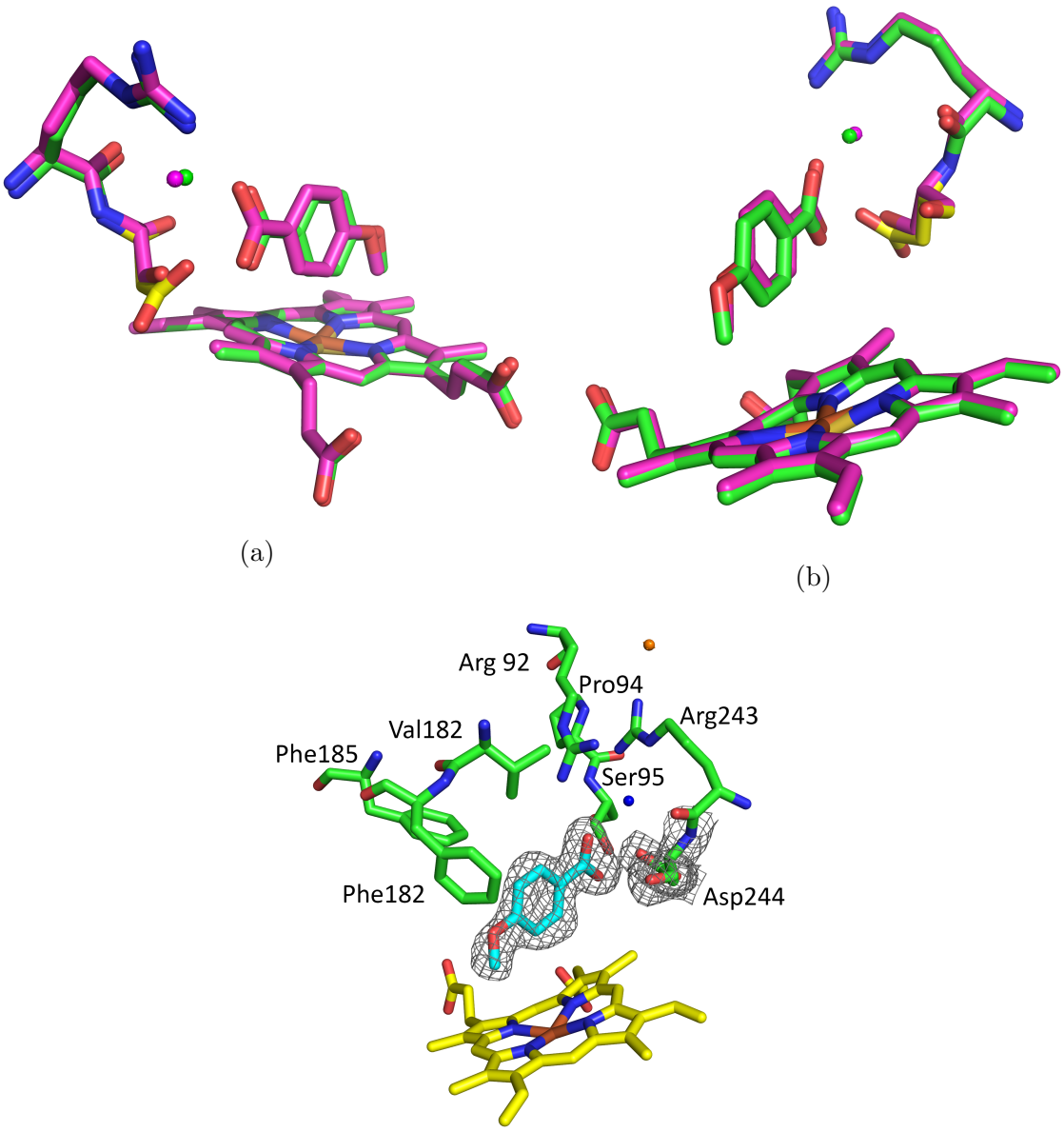


Figure 75: The crystal structures of the amino acids Arg243 and Ser244/Asp244 binding to the bridging water molecule of the wild-type (magenta) and S244D mutant (green) (a) and (b). The bridging water molecule is shown as spheres, and Asp244 in S244D is highlighted in yellow. (c) The active site of the CYP199A4 S244D mutant binding 4-methoxybenzoic acid. The $2mF_o$-DF_c density of the 4-methoxybenzoic acid (cyan) is shown contoured at ≈ 0.486 e/Å^3 (grey mesh). The heme and active site residues are shown in yellow and green respectively. The bridging water molecule and chloride ion are in blue and orange spheres respectively. Phe298 is not shown as it is obscured by the substrate.

Table 22: Key distances of S244D and CYP199A4[21] crystal structures. The distances from the heme group is measured from the iron centre. The atoms on the substrate 4-methoxybenzoic acid are specified in bold. Amino acid nomenclature follows the International Union of Pure and applied Chemistry/International Union of Biochemistry joint commission on biochemical nomenclature.[164] If the equivalent amino acid or molecule is different between the S244D and WT structures, they will be noted with a "/" (e.g. Asn/Ser244)

Species 1	Species 2	Distance: S244D (Å)	Distance: WT (Å)
Heme	S^3-Cys358	2.2	2.4
Heme	Substrate Me-**O**-	5.2	5.2
Heme	Substrate **Me**-O-	4.2	4.1
Heme	Substrate -**C**OOH	8.0	8.4
Cl	O^{η}-Tyr177	2.8	3.1
Cl	Wat232/Wat1129	2.4	3.1
Cl	$N^{\delta 2}$-Asn207/Wat679	2.7	3.1
Cl	$O^{\delta 2}$-Asn207/$N^{\epsilon 2}$-Gln203	3.3	3.3
Bridging H_2O	$O^{\delta 1}$-Asp244/$O^{\gamma 1}$-Ser244	2.7	3.0
Bridging H_2O	Substrate-CO**OH**	2.8	2.6
Bridging H_2O	N^{ϵ}-Arg243	2.9	2.9
Bridging H_2O	$N^{\eta 1}$-Arg243	2.9	2.9
Substrate Me-**O**-	C^{ζ}-Phe182	3.5	3.1
Substrate Me-**O**-	$C^{\epsilon 1}$-Phe182	3.1	3.2
Substrate Me-**O**-	C^{ζ}-Phe298	3.3	3.7
Substrate C^3-benzene	$C^{\epsilon 1}$-Phe185	3.8	3.9
Substrate -CO**O**H	$N^{\eta 1}$-Arg92	3.0	2.8
Substrate -CO**OH**	O^{γ}Ser95	3.5	2.5
Substrate -CO**OH**	$O^{\delta 1}$-Asp244/$O^{\gamma 1}$-Ser244	2.4	2.6
Substrate **C^2**-benzene	$C^{\delta 1}$Leu98	4.5	4.8
Substrate **C^5**-benzene	C^{β}Ala248	3.6	3.5

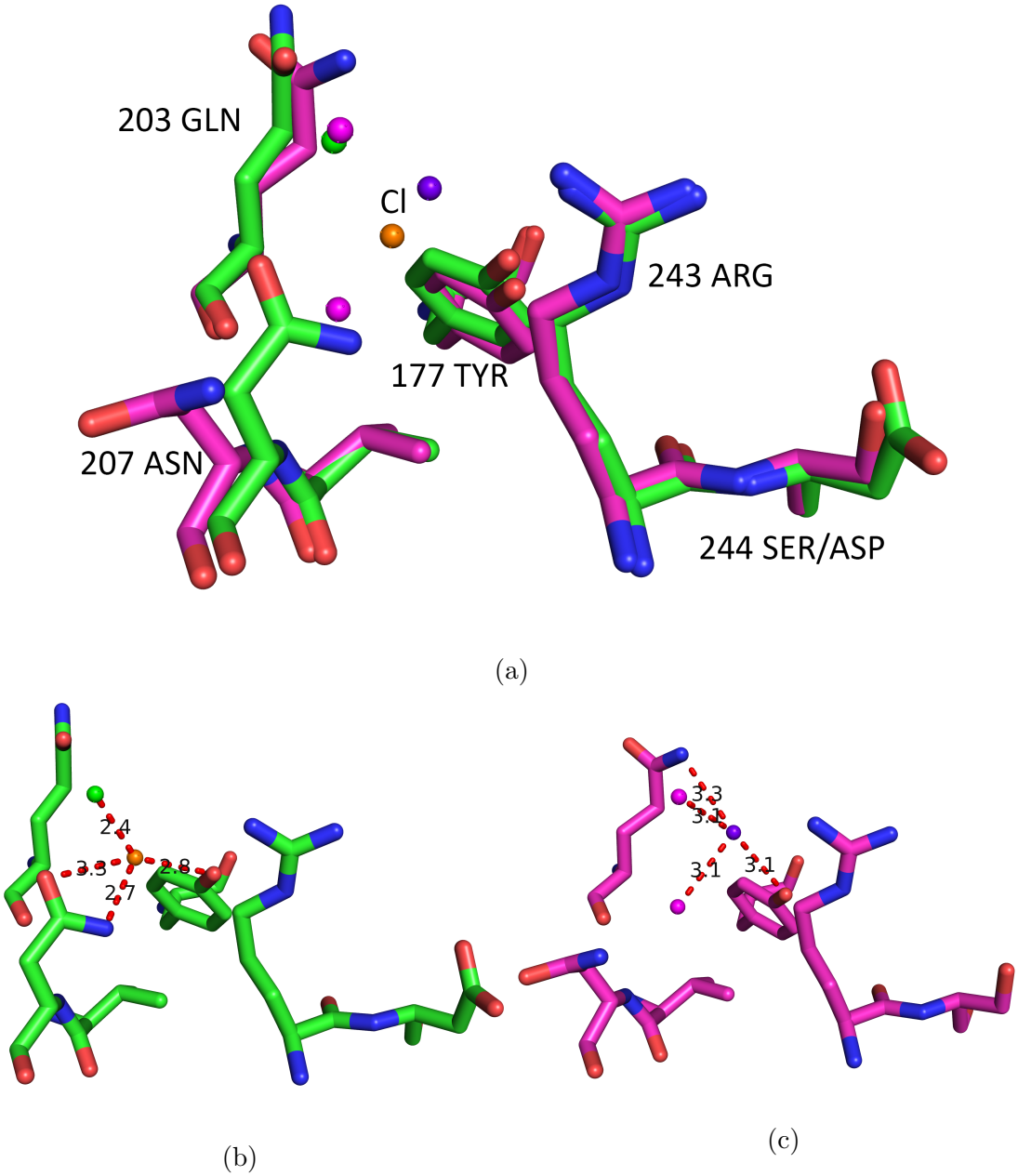


Figure 76: The chloride binding site of S244D (green) and wild-type (magenta).[21] The water and chloride molecules are shown as spheres, with the chlorides being orange for S244D and purple for the wild-type. The crystal structures of CYP199A4 S244D and its wild-type are overlaid in (a), and separated in (b) and (c) respectively with the distances of the chloride binding site highlighted.

4.6 Discussion: Crystal structure of CYP199A4 S244D

The S244D mutation increased the substrate range of *para*-substituted benzene derivatives. The mutant does not significantly change the type of reactions which is catalysed by the enzyme.[71] For instance, 4-methoxybenzaldehyde is still dealkylated to 4-hydroxymethylbenzaldehyde, and 4-(methylthio)benzoic acid is still oxidised to 4-(methylsulfinyl)benzoic acid.[71,73] However a change of the rate of oxidation was observed. The subtle shift in 4-methoxybenzoic acid over the heme active site supports this observation, as the methoxy-group is held over the heme-active site almost in the same position as the wild-type enzyme.[21]

The crystal structure of 4-methoxybenzoic acid bound S244D revealed a small shift in the substrate relative to the heme-iron centre compared to the CYP199A4 model. Aspartate 244, which was mutated from a serine, is closer to the benzoic acid carboxy terminus. Although this is probably due to the aspartate side chain being inherently longer, the interaction between the carboxy terminus with this acidic side chain must be less favourable compared to the wild-type enzyme, as the binding affinity and turnover rates for the S244D mutant with 4-methoxybenzoic acid are significantly lower than the wild-type P450 (Table 14).[71] It must be noted that the majority of other methoxy-modified substrates showed significantly higher turnover, but with weaker binding affinity to the S244D mutant. Further crystal structures of S244D with other substrates is required for future analysis on how methoxy-modified substrates interact within the active site of S244D.

The S244D mutant has been shown to enantioselectivity oxidise a single enantiomer for *para*-substituted ethylbenzene derivatives at the α-carbon position.[72] The α-hydroxylation reaction catalysed by S244D was found to be exclusively enantiomerically selective. However, both the sulfoxidation and epoxidation reactions catalysed by the CYP199A4 S244D mutant were less selective for the majority of sulfoxide and epoxide products. The degree of enantiomeric excess is affected by the different functional groups in methylthio-modified and styrene substrates. For instance, turnover product for 4-(methylthio)benzaldehyde produced enantiomeric excess of 40%. This was much lower than the turnover product for 4-fluorothioanisole, which showed an enantioselectivity of 90%. Different oxidation reactions catalysed by the wild-type CYP199A4 and the S244D mutant, such as S-oxidation and O-dealkylation, should also contribute to the degree of enantioselectivity. For example, 4-(methylthio)benzonitrile has a methylthio-moiety instead of an alkene group at the equivalent position in 4-cyanostyrene. The sulfoxide product of 4-(methylthio)benzonitrile had a lower enantiomeric excess than the epoxidation product of 4-cyanostyrene (58% compared to 72%). The *S* enantiomer is the preferred product for wild-type CYP199A4 turnovers with 4-(ethylthio)benzoic acid and

4-ethylbenzoic acid, while the enantioselectivity for styrenes and methylthio-modified substrates were less selective. The ethylthio/methylthio moiety must be positioned differently above the iron centre compared to the ethyl and alkene groups, allowing oxidative attack at predominately one side of the compound. This would explain why the turnovers of methylthio-modified substrates and styrenes are less biased for a single enantiomer. As a result both the carboxylic acid moiety or equivalent functional group and the *para*-substituted moiety have an affect on the enantioselectivity. Investigations on solving the structure of a 4-(methylthio)benzoic acid or styrene bound S244D mutant would help confirm the position of the substrate over the iron-centre. Finally a small shift the substrate position observed for 4-methoxybenzoic acid between the S244D and the wild-type enzyme in reference to the iron heme may also have an effect on the enantioselectivity. A difference of enantioselectivity may be observed between the wild-type CYP199A4 and its S244D variant for the same turnover due to this small shift.

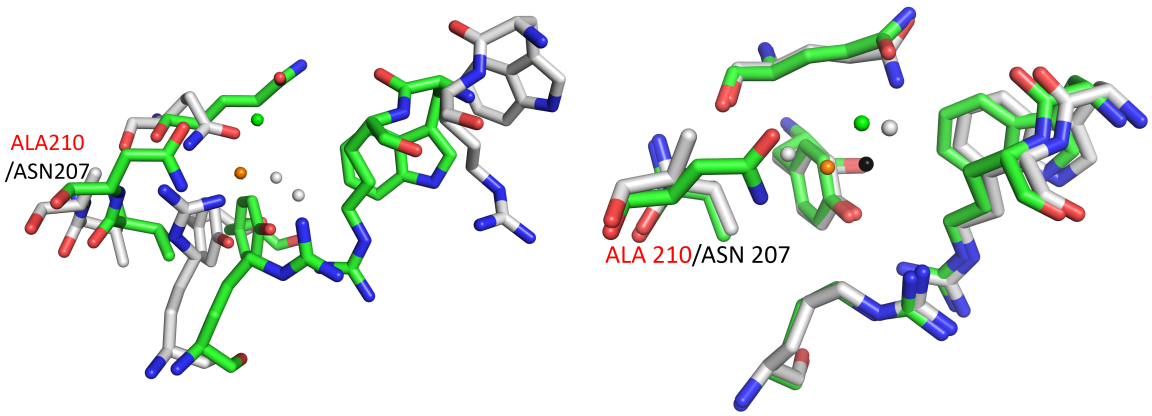
Turnover studies for CYP199A4 have shown that methyl-modified substrates typically bind with lower affinity and exhibit lower activity compared to ethyl-modified and methoxy-modified substrates.[10] Based on the position of 4-methoxybenzoic acid within the wild-type CYP199A4 and the S244D variant, this suggests that methyl-modified substrates are not held close enough to the iron-heme center within the active site of the S244D variant. Thus, it also cannot optimally displace the distal water ligand gating the initial step for P450 catalysis.

If dealkylation of the methoxy group is facilitated by the HAT mechanism, hydrogen abstraction at the carbon closest to the iron-heme group (i.e. α-carbon) is the first step in the radical rebound mechanism. Conversely, this α-carbon is not important in the SET pathway. Based on the location of the oxygen in 4-methoxybenzoic acid, the sulfur in methylthio-modified substrates can also be predicted to be held further away from the heme-active centre relative to the α-carbon. Despite the larger size of sulfur this may hinder direct oxygen transfer from the heme-active site to the sulfur atom. This could explain the lower product formation rates for methylthio-modified substrates compared to methoxy-modified substrates.[71]

The significantly different chloride ion position in the S244D compared to the wild-type may cause the chloride ion to be less effective in capping the active site from external solvent. Lower binding affinity for 4-methoxybenzoic acid may be due to a combination of the altered Asp207 side chain, the repulsive interactions between both carboxylate groups in Asp244 and 4-methoxybenzoic acid, and the altered chloride position. Shifting the chloride ion to a different position may allow easier access for a wider range of benzene derivatives to the active site. This suggests that the chloride may have control on the substrate entry to the active site, and altering the position of this chloride ion may also alter the types of substrates that can enter the active site.

The crystal structure of the closely related CYP199A2 has also been solved.[21] CYP199A2 hares 86% sequence identity and has a very similar structure as CYP199A4. The 4-methoxybenzoic acid bound crystal structure of CYP199A2 also reports a chloride capping the active site from external solvent. When the chloride binding site of the substrate bound CYP199A2 is compared to the CYP199A4 S244D variant, the positions of the amino acids in the binding site for CYP199A2 and the S244D equivalent are in similar positions and orientations (Figure 77b). The positions of Gln206 and Try180 in CYP199A2 are almost identical for the equivalent S244D amino acids Gln203 and Tyr177, even though Gln203 does not participate in the coordination of the chloride ion for the S244D mutant. The CYP199A2 chloride ion does not interact with alanine 210, which is the equivalent amino acid to asparagine 207 in the CYP199A4 structure. As alanine is a non-polar amino acid, it cannot interact with the chloride ion even if it were to move upon substrate binding. As such an equivalent S244D mutant of CYP199A2 may not show an altered chloride binding. The substrate-free CYP199A2 chloride binding site differs significantly from the positions of the amino acids of CYP199A4 S244D (Figure 77a), and Ala210 also does not point towards where the chloride ion would be. These comparisons show the unique position and coordination of the capping chloride ion within the S244D variant.

Crystal structure of the S244D mutant with other *para*-methoxy modified substrates and the equivalent CYP199A2 variant would determine if Asn207/Ala210 would move and coordinated to the chloride ion. Co-crystallisation of methylthio-modified compounds and styrenes will further the understanding on the oxidation and degree of enantioselectivity the S244D mutant has for many of these substrates.



(a) (b)

Figure 77: The chloride binding site of S244D (green) and the equivalent position for CYP199A2 (grey) for the substrate free (a) and substrate bound (b) forms.[21] The water and chloride molecules are shown as spheres, with the chlorides being orange for S244D and black for substrate bound CYP199A2. Asparagine 207 and its CYP199A2 equivalent alanine 210 are labelled in black and red respectively.

5 Conclusion and Future Directions

We have presented here the first analysis of the P450s and ferredoxins from *Frankia* sp. EuI1c. The successful expression and purification of FraEu2494, FraEu5334, FraEu1415 and FraEu4131 will enable the role of these P450s in this *Frankia* strain to be explored. Spin-state shift and binding affinity assays revealed potential substrates for FraEu2494, FraEu5334 and FraEu1415. The substrate range of FraEu2494 included norisoprenoids and other compounds containing cyclic rings. A range of norisoprenoids, linear carbon chain compounds and cyclic ring containing compounds bound weakly to FraEu5334. As a result the substrate range of FraEu5334 requires further investigation. FraEu1415 exhibits high binding affinity for specific steroids such as testosterone and progesterone. This confirms that the *Frankia* genus contains enzymes responsible for the synthesis or catabolism of various bioactive compounds. As such *Frankia* is a potential source of natural products. Further investigations on the turnover products of these steroids will be required to define the role of the P450s in metabolism.

FraEu2494 was screened for crystal growth for X-ray crystallography experiments. Although crystals were obtained, further optimisation experiments are required to generate a crystal which generates a diffraction pattern through X-ray crystallography. If the crystal structure of FraEu2494 can be solved, especially with a substrate bound, the interactions between the substrate and active site amino acids could be characterised. Future investigation on the crystallisation of FraEu5334 and FraEu1415 will also allow characterisation of the binding site of these P450s. This would be especially interesting for FraEu1415, as the tight binding of steroids suggests an optimal fit of these substrate within the active site of FraEu1415.

The ferredoxins Fdx2495, Fdx5333 and Fdx1414, whose genes are associated with the three P450s mentioned above, were also successfully generated and purified. These are [3Fe-4S] or [4Fe-4S] iron-sulfur cluster containing ferredoxins that have non-standard residues (histidine, serine and threonine respectively) in their binding motifs. Characterisation of the roles these variable residues have on the stability and function of the ferredoxin is required. None of the ferredoxin reductases from *Frankia* sp. EuI1c were able to be produced and purified. Further optimisation of the ferredoxin reductase expression and purification process is required to complete the electron transfer chain and enable P450 product formation. Using other electron transfer partner systems may facilitate transfer of the reducing equivalents from NAD(P)H to the P450, enabling product formation.

A mutant library of Fdx2495 was successfully generated and purified. The mutants studied were alanine (H13A), tyrosine (H13Y), asparagine (H13N) and cysteine (H13C). These mutant ferredoxins could provide insight into the properties of iron-sulfur cluster

ferredoxins with variable amino acid residues in their motif. The library created should contain further mutations and the investigation of these will provide additional information on how non-standard residues in the cluster binding motif can affect the integrity, structure and function of iron-sulfur ferredoxins.

Overall the results from *Frankia* sp. EuI1c demonstrate the potential of this bacterial strain as a source of novel P450s and [3Fe-4S] iron-sulfur ferredoxins. Future work will involve optimising the growth of these P450 enzymes and their electron transfer partners to facilitate turnover of substrates, and optimise crystallisation of these P450s for structural characterisation.

A range of *para*-substituted benzene substrates were tested with the S244D variant of the CYP199A4 enzyme from *Rhodopseudomonas palustris* strain HaA2. The activity of this variant was greater than the wild-type enzyme for non-benzoic acid substrates such as benzaldehydes and phenols. The activity was also influenced by the identity of the *para*-substituent. For example, ethyl oxidation and O-demethylation was more active than methyl oxidation, styrene epoxidation and S-oxidation. In the majority of cases the substrates were exclusively oxidised at the *para*-position to a single product.

To test the efficiency of the S244D variant, total turnover experiments were designed to investigate the maximum potential turnover numbers for the substrates. The high total turnovers numbers for methoxy-modified substrates matched with previously reported product formation rates. Similarly, total turnover numbers for ethyl-modified substrates followed the reported coupling efficiency. The comparatively low total turnover numbers for methyl-modified substrates affirmed previous results which suggested the methyl substituent is not held in an optimal location above the heme-iron site.

The regioselectivity of the S244D mutant towards dimethyl-modified substrates was also investigated. The turnovers were regioselective for the *para*-position. The hydroxylation of dimethyl-substituted benzene derivatives demonstrates the biocatalytic potential of CYP199A4 S244D for regioselective hydroxylation in the presence of additional functional groups.

Sulfoxidation was almost exclusively observed for methylthio-substituted benzene derivatives with the S244D mutant. No dealkylation or further oxidation products were detected. 4-(Methylthio)toluene was unusual in that it was predominately hydroxylated to form 4-(methylthio)benzyl alcohol. This shows that the substrate binds in a different orientation with the methylthio group in the hydrophilic binding pocket and the methyl group being held over the heme iron. The low product formation rates for these methylthio-modified substrates suggests they do not fit as well as methoxy-modified substrates and may be held further away from the heme iron.

The epoxidation of styrene substrates by the S244D variant was also studied. A selection

of styrenes were oxidised to epoxide products, and some exhibited 1,2-rearrangement to form an aldehyde product. The formation of the aldehyde implies the reaction pathway involves a cation intermediate. Therefore CYP199A4 and the S244D mutant could be used as model systems for the investigation of P450 reaction pathways and mechanisms. The epoxidation and sulfoxidation reactions both produced chiral products that were largely generated with enantiomeric excess of one stereoisomer. Based on previous results and GC and HPLC retention times, it appears the turnovers consistently produced an excess of the *S* enantiomer.

The crystal structure of 4-methoxybenzoic acid bound CYP199A4 S244D was solved. This showed the orientation of 4-methoxybenzoic acid was similar to that observed in the wild-type enzyme with the *para*-methoxy group being held over the heme-active site allowing for regioselective demethylation at this position. The binding pocket where the 244 aspartate mutation is positioned and the substrate positions are largely unchanged between the mutant and wild-type enzyme. The electronic interactions between the 244 amino acid and the substrate must largely dictate the differences in binding affinity between CYP199A4 and the S244D variant. Although large changes in backbone structure were not observed for the S244D mutant, the position and coordination of the capping chloride ion was altered. The change in the chloride ion position is accompanied by the flipping of the side chain of asparagine 207 in order to coordinate to the chloride. This may affect which substrates are allowed inside the active site of CYP199A4 S244D. Further crystallisation of the S244D mutant with other substrates will not only reveal how these compounds will bind within the active site, but also whether this chloride capping ion retains an altered coordination and position.

The results from the CYP199A4 S244D variant have further established this mutant as a potential biocatalyst for synthetically challenging regioselective functionalisation of a range of *para*-substituted benzene derivatives. Future work will involve crystallisation of CYP199A4 S244D with other substrates, as well as further investigation and mutagenesis of this P450 mutant for other oxidation activity.

References

[1] Bell, S. G.; Dale, A.; Rees, N.; Wong, L.-L. A cytochrome P450 class I electron transfer system from Novosphingobium aromaticivorans. *Applied Microbiology and Biotechnology* **2010**, *86*, 163–175.

[2] Guengerich, F. P. Cytochrome P450 and Chemical Toxicology. *Chemical Research in Toxicology* **2008**, *21*, 70–83, PMID: 18052394.

[3] O'Reilly, E.; Kohler, V.; Flitsch, S. L.; Turner, N. J. Cytochromes P450 as useful biocatalysts: addressing the limitations. *Chemical Communications* **2011**, *47*, 2490–2501.

[4] Nelson, D. The Cytochrome P450 Homepage. *Human Genomics* **2009**, *4*, 59–65.

[5] Omura, T.; Sato, R. The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. *Journal of Biological Chemistry* **1964**, *239*, 2370–2378.

[6] Nelson, D. R.; Kamataki, T.; Waxman, D. J.; Guengerich, F. P.; Estabrook, R. W.; Feyereisen, R.; Gonzalez, F. J.; Coon, M. J.; Gunsalus, I. C.; Gotoh, O. The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA and cell biology* **1993**, *12*, 1–51.

[7] Bell, S. G.; Tan, A. B. H.; Johnson, E. O. D.; Wong, L.-L. Selective oxidative demethylation of veratric acid to vanillic acid by CYP199A4 from Rhodopseudomonas palustris HaA2. *Molecular Biosystems* **2009**, *6*, 206–214.

[8] Groves, J. T.; McClusky, G. A. Aliphatic hydroxylation via oxygen rebound. Oxygen transfer catalyzed by iron. *Journal of the American Chemical Society* **1976**, *98*, 859–861.

[9] Groves, J. T.; McClusky, G. A.; White, R. E.; Coon, M. J. Aliphatic hydroxylation by highly purified liver microsomal cytochrome P-450. Evidence for a carbon radical intermediate. *Biochemical and Biophysical Research Communications* **1978**, *81*, 154 – 160.

[10] Bell, S. G.; Zhou, R.; Yang, W.; Tan, A. B. H.; Gentleman, A. S.; Wong, L.-L.; Zhou, W. Investigation of the Substrate Range of CYP199A4: Modification of the Partition between Hydroxylation and Desaturation Activities by Substrate and Protein Engineering. *Chemistry – A European Journal* **2012**, *18*, 16677–16688.

[11] Bell, S. G.; McMillan, J. H. C.; Yorke, J. A.; Kavanagh, E.; Johnson, E. O. D.; Wong, L.-L. Tailoring an alien ferredoxin to support native-like P450 monooxygenase activity. *Chemical Communications* **2012**, *48*, 11692–11694.

[12] Hannemann, F.; Bichet, A.; Ewen, K. M.; Bernhardt, R. Cytochrome P450 systems - biological variations of electron transport chains. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2007**, *1770*, 330 – 344, {P450}.

[13] Bernhardt, R. Cytochromes P450 as versatile biocatalysts. *Journal of Biotechnology* **2006**, *124*, 128 – 145.

[14] Gunsalus, I. C.; Sligar, S. G. *Advances in Enzymology and Related Areas of Molecular Biology*; John Wiley & Sons, Inc., 2006; pp 1–44.

[15] McLean, K. J.; Leys, D.; Munro, A. W. In *Cytochrome P450*, 4th ed.; de Montellano, P. R. O., Ed.; Springer International Publishing, 2015; Vol. 1; Chapter 6: Microbial Cytochromes P450, pp 261 – 407.

[16] Poulos, T. L.; Finzel, B. C.; Howard, A. J. High-resolution crystal structure of cytochrome P450cam. *Journal of Molecular Biology* **1987**, *195*, 687 – 700.

[17] Johnson, D. C.; Dean, D. R.; Smith, A. D.; Johnson, M. K. Structure, function, and formation of biological iron-sulfur clusters. *Annual Review of Biochemistry* **2005**, *74*, 247–281, PMID: 15952888.

[18] Trower, M. K.; Emptage, M. H.; Sariaslani, F. S. Purification and characterization of a 7Fe ferredoxin from Streptomyces griseus. *Biochimica et Biophysica Acta* **1990**, *1037*, 281–289.

[19] Chun, Y.-J.; Shimada, T.; Sanchez-Ponce, R.; Martin, M. V.; Lei, L.; Zhao, B.; Kelly, S. L.; Waterman, M. R.; Lamb, D. C.; Guengerich, F. P. Electron Transport Pathway for a Streptomyces Cytochrome P450: Cytochrome P450 105D5-Catalysed Fatty Acid Hydroxylation in Streptomycyes Coelicolor A3(2). *Journal of Biological Chemistry* **2007**, *282*, 17486–17500.

[20] Bhattarai, S.; Liou, K.; Oh, T.-J. Hydroxylation of long chain fatty acids by CYP147F1, a new cytochrome P450 subfamily protein from Streptomyces peucetius. *Archives of Biochemistry and Biophysics* **2013**, *539*, 63 – 69.

[21] Bell, S. G.; Yang, W.; Tan, A. B. H.; Zhou, R.; Johnson, E. O. D.; Zhang, A.; Zhou, W.; Rao, Z.; Wong, L.-L. The crystal structures of 4-methoxybenzoate bound CYP199A2 and CYP199A4: structural changes on substrate binding

and the identification of an anion binding site. *Dalton Transactions* **2012**, *41*, 8703–8714.

[22] Denisov, I. G.; Sligar, S. G. In *Cytochrome P450*, 4th ed.; de Montellano, P. R. O., Ed.; Springer International Publishing, 2015; Vol. 1; Chapter 3: Activation of Molecular Oxygen in Cytochromes P450, pp 69–109.

[23] Modi, A. R.; Dawson, J. H. In *Monooxygenase, Peroxidase and Peroxygenase Properties and Reaction Mechanisms of Cytochrome P450 Enzymes*; Hrycay, E., Bandiera, S., Eds.; Springer International Publishing,, 2015; Chapter 2, pp 63–81.

[24] Poulos, T. L.; Raag, R. Cytochrome P450cam: crystallography, oxygen activation, and electron transfer. *The FASEB Journal* **1992**, *6*, 674–9.

[25] Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I. Structure and Chemistry of Cytochrome P450. *Chemical Reviews* **2005**, *105*, 2253–2278.

[26] Shaik, S.; Kumar, D.; de Visser, S. l. P.; Altun, A.; Thiel, W. Theoretical Perspective on the Structure and Mechanism of Cytochrome P450 Enzymes. *Chemical Reviews* **2005**, *105*, 2279–2328, PMID: 15941215.

[27] Watanabe, Y. Alternatives to the oxoferryl porphyrin cation radical as the proposed reactive intermediate of cytochrome P450: two-electron oxidized Fe(III) porphyrin derivatives. *Journal of Biological Inorganic Chemistry* **2001**, *6*, 846–856.

[28] Davydov, R.; Perera, R.; Jin, S.; Yang, T.-C.; Bryson, T. A.; Sono, M.; Dawson, J. H.; Hoffman, B. M. Substrate Modulation of the Properties and Reactivity of the Oxy-Ferrous and Hydroperoxo-Ferric Intermediates of Cytochrome P450cam As Shown by Cryoreduction-EPR/ENDOR Spectroscopy. *Journal of the American Chemical Society* **2005**, *127*, 1403–1413, PMID: 15686372.

[29] Meunier, B.; de Visser, S.; Shaik, S. Mechanism of Oxidation Reactions Catalyzed by Cytochrome P450 Enzymes. *American Chemical Society* **2004**, *104*, 3947–3980.

[30] Jin, S.; Makris, T. M.; Bryson, T. A.; Sligar, S. G.; Dawson, J. H. Epoxidation of Olefins by Hydroperoxo-Ferric Cytochrome P450. *Journal of the American Chemical Society* **2003**, *125*, 3406–3407, PMID: 12643683.

[31] Cryle, M. J.; De Voss, J. J. Is the Ferric Hydroperoxy Species Responsible for Sulfur Oxidation in Cytochrome P450s? *Angewandte Chemie International Edition* **2006**, *45*, 8221–8223.

[32] Vaz, A. D. N.; McGinnity, D. F.; Coon, M. J. Epoxidation of olefins by cytochrome P450: Evidence from site-specific mutagenesis for hydroperoxo-iron as an electrophilic oxidant. *Proceedings of the National Academy of Sciences* **1998**, *95*, 3555–3560.

[33] Kerber, R. B., W.D; Goldberg, D. H2O2 Oxidations Catalyzed by an Iron(III) Corrolazine: Avoiding High-Valent Iron-Oxido Species? *Angewandte Chemie International Edition* **2007**, *46*, 3718–3721.

[34] Coon, M. J.; Vaz, A. D. N.; McGinnity, D. F.; Peng, H.-M. Multiple Activated Oxygen Species in P450 Catalysis. *Drug Metabolism and Disposition* **1998**, *26*, 1190–1193.

[35] Shaik, S.; Hirao, H.; Kumar, D. Reactivity patterns of cytochrome P450 enzymes: multifunctionality of the active species, and the two states-two oxidants conundrum. *Natural Product Reports* **2007**, *24*, 533–552.

[36] Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W. P450 Enzymes: Their Structure, Reactivity, and Selectivity—Modeled by QM/MM Calculations. *Chemical Reviews* **2010**, *110*, 949–1017, PMID: 19813749.

[37] Ortiz de Montellano, P. R.; Stearns, R. A. Timing of the radical recombination step in cytochrome P-450 catalysis with ring-strained probes. *Journal of the American Chemical Society* **1987**, *109*, 3415–3420.

[38] Newcomb, M.; Hollenberg, P. F.; Coon, M. J. Multiple mechanisms and multiple oxidants in P450-catalyzed hydroxylations. *Archives of Biochemistry and Biophysics* **2003**, *409*, 72 – 79.

[39] Newcomb, M.; Chandrasena, R. E. P. Highly reactive electrophilic oxidants in cytochrome P450 catalysis. *Biochemical and Biophysical Research Communications* **2005**, *338*, 394 – 403.

[40] Newcomb, M.; Shen, R.; Choi, S.-Y.; Toy, P. H.; Hollenberg, P. F.; Vaz, A. D. N.; Coon, M. J. Cytochrome P450-Catalyzed Hydroxylation of Mechanistic Probes that Distinguish between Radicals and Cations. Evidence for Cationic but Not for Radical Intermediates. *Journal of the American Chemical Society* **2000**, *122*, 2677–2686.

[41] Newcomb, M.; Le Tadic, M.-H.; Putt, D. A.; Hollenberg, P. F. An incredibly fast apparent oxygen rebound rate constant for hydrocarbon hydroxylation by cytochrome P-450 enzymes. *Journal of the American Chemical Society* **1995**, *117*, 3312–3313.

[42] Newcomb, M.; Le Tadic-Biadatti, M.-H.; Chestney, D. L.; Roberts, E. S.; Hollenberg, P. F. A nonsynchronous concerted mechanism for cytochrome P-450 catalyzed hydroxylation. *Journal of the American Chemical Society* **1995**, *117*, 12085–12091.

[43] Toy, P. H.; Newcomb, M.; Hollenberg, P. F. Hypersensitive Mechanistic Probe Studies of Cytochrome P450-Catalyzed Hydroxylation Reactions. Implications for the Cationic Pathway. *Journal of the American Chemical Society* **1998**, *120*, 7719–7729.

[44] Shaik, S.; Cohen, S.; de Visser, S. P.; Sharma, P. K.; Kumar, D.; Kozuch, S.; Ogliaro, F.; Danovich, D. The "Rebound Controversy": An Overview and Theoretical Modeling of the Rebound Step in C-H Hydroxylation by Cytochrome P450. *European Journal of Inorganic Chemistry* **2004**, *2004*, 207–226.

[45] Newcomb, M.; Toy, P. H. Hypersensitive Radical Probes and the Mechanisms of Cytochrome P450-Catalyzed Hydroxylation Reactions. *Accounts of Chemical Research* **2000**, *33*, 449–455, PMID: 10913233.

[46] Kamachi, T.; Yoshizawa, K. A Theoretical Study on the Mechanism of Camphor Hydroxylation by Compound I of Cytochrome P450. *Journal of the American Chemical Society* **2003**, *125*, 4652–4661, PMID: 12683838.

[47] Guallar, V.; Baik, M.-H.; Lippard, S. J.; Friesner, R. A. Peripheral heme substituents control the hydrogen-atom abstraction chemistry in cytochromes P450. *Proceedings of the National Academy of Sciences* **2003**, *100*, 6998–7002.

[48] Ortiz de Montellano, P. R. Hydrocarbon Hydroxylation by Cytochrome P450 Enzymes. *Chemical Reviews* **2010**, *110*, 932–948, PMID: 19769330.

[49] de Montellano, P. R. O. In *Cytochrome P450*, 4th ed.; de Montellano, P. R. O., Ed.; Springer International Publishing, 2015; Vol. 1; Chapter 4: Substrate Oxidation by Cytochrome P450 Enzymes, pp 111–176.

[50] Ji, L.; Faponle, A. S.; Quesne, M. G.; Sainna, M. A.; Zhang, J.; Franke, A.; Kumar, D.; van Eldik, R.; Liu, W.; de Visser, S. P. Drug Metabolism by Cytochrome P450 Enzymes: What Distinguishes the Pathways Leading to Substrate Hydroxylation Over Desaturation? *Chemistry - A European Journal* **2015**, *21*, 9083–9092.

[51] Lai, W.; Chen, H.; Cohen, S.; Shaik, S. Will P450cam Hydroxylate or Desaturate Alkanes? QM and QM/MM Studies. *The Journal of Physical Chemistry Letters* **2011**, *2*, 2229–2235.

[52] Vaz, A. D.; Pernecky, S. J.; Raner, G. M.; Coon, M. J. Peroxo-iron and oxenoid-iron species as alternative oxygenating agents in cytochrome P450-catalyzed reactions: switching by threonine-302 to alanine mutagenesis of cytochrome P450 2B4. *Proceedings of the National Academy of Sciences* **1996**, *93*, 4644–4648.

[53] Park, M. J.; Lee, J.; Suh, Y.; Kim, J.; Nam, W. Reactivities of Mononuclear Non-Heme Iron Intermediates Including Evidence that Iron(III)-Hydroperoxo Species Is a Sluggish Oxidant. *Journal of the American Chemical Society* **2006**, *128*, 2630–2634, PMID: 16492048.

[54] Altun, A.; Shaik, S.; Thiel, W. What is the Active Species of Cytochrome P450 during Camphor Hydroxylation? QM/MM Studies of Different Electronic States of Compound I and of Reduced and Oxidized Iron-Oxo Intermediates. *Journal of the American Chemical Society* **2007**, *129*, 8978–8987, PMID: 17595079.

[55] Ogliaro, F.; de Visser, S. P.; Cohen, S.; Sharma, P. K.; Shaik, S. Searching for the Second Oxidant in the Catalytic Cycle of Cytochrome P450: A Theoretical Investigation of the Iron(III)-Hydroperoxo Species and Its Epoxidation Pathways. *Journal of the American Chemical Society* **2002**, *124*, 2806–2817, PMID: 11890833.

[56] Guengerich, F. P. Common and Uncommon Cytochrome P450 Reactions Related to Metabolism and Chemical Toxicity. *Chemical Research in Toxicology* **2001**, *14*, 611–650.

[57] Seto, Y.; Guengerich, F. P. Partitioning between N-dealkylation and N-oxygenation in the oxidation of N,N-dialkylarylamines catalyzed by cytochrome P450 2B1. *Journal of Biological Chemistry* **1993**, *268*, 9986–9997.

[58] Hammons, G. J.; Guengerich, F. P.; Weis, C. C.; Beland, F. A.; Kadlubar, F. F. Metabolic Oxidation of Carcinogenic Arylamines by Rat, Dog, and Human Hepatic Microsomes and by Purified Flavin-containing and Cytochrome P-450 Monooxygenases. *Cancer Research* **1985**, *45*, 3578–3585.

[59] Sharma, P. K.; de Visser, S. P.; Shaik, S. Can a Single Oxidant with Two Spin States Masquerade as Two Different Oxidants? A Study of the Sulfoxidation Mechanism by Cytochrome P450. *Journal of the American Chemical Society* **2003**, *125*, 8698–8699.

[60] Li, C.; Zhang, L.; Zhang, C.; Hirao, H.; Wu, W.; Shaik, S. Which Oxidant Is Really Responsible for Sulfur Oxidation by Cytochrome P450? *Angewandte Chemie International Edition* **2007**, *46*, 8168–8170.

[61] Cryle, M. J.; De Voss, J. J. Is the Ferric Hydroperoxy Species Responsible for Sulfur Oxidation in Cytochrome P450s? *Angewandte Chemie International Edition* **2006**, *45*, 8221–8223.

[62] Wang, B.; Li, C.; Cho, K.-B.; Nam, W.; Shaik, S. The FeIII(H2O2) Complex as a Highly Efficient Oxidant in Sulfoxidation Reactions: Revival of an Underrated Oxidant in Cytochrome P450. *Journal of Chemical Theory and Computation* **2013**, *9*, 2519–2525, PMID: 26583848.

[63] Harwood, C. S.; Gibson, J. Anaerobic and aerobic metabolism of diverse aromatic compounds by the photosynthetic bacterium Rhodopseudomonas palustris. *Applied and Environmental Microbiology* **1988**, *54*, 712–717.

[64] Huang, J. J.; Heiniger, E. K.; McKinlay, J. B.; Harwood, C. S. Production of hydrogen gas from light and the inorganic electron donor thiosulfate by Rhodopseudomonas palustris. *Appl. Environ. Microbiol.* **2010**, *76*, 7717–7722.

[65] Pang, X.; Xu, F.; Bell, S. G.; Guo, D.; Wong, L.-L.; Rao, Z. Purification, crystallization and preliminary crystallographic analysis of cytochrome P450 203A1 from *Rhodopseudomonas palustris*. *Acta Crystallographica Section F* **2007**, *63*, 342–345.

[66] Peng, Y.; Xu, F.; Bell, S. G.; Wong, L.-L.; Rao, Z. Crystallization and preliminary X-ray diffraction studies of a ferredoxin reductase from *Rhodopseudomonas palustris* CGA009. *Acta Crystallographica Section F* **2007**, *63*, 422–425.

[67] Oda, Y.; Wanders, W.; Huisman, L. A.; Meijer, W. G.; Gottschal, J. C.; Forney, L. J. Genotypic and phenotypic diversity within species of purple nonsulfur bacteria isolated from aquatic sediments. *Applied and Environmental Microbiology* **2002**, *68*, 3467–3477.

[68] Bell, S. G.; Hoskins, N.; Xu, F.; Caprotti, D.; Rao, Z.; Wong, L.-L. Cytochrome P450 enzymes from the metabolically diverse bacterium Rhodopseudomonas palustris. *Biochemical and Biophysical Research Communications* **2006**, *342*, 191 – 196.

[69] Larimer, F. W. et al. Complete genome sequence of the metabolically versatile photosynthetic bacterium Rhodopseudomonas palustris. *Nature Biotechnology* **2004**, *22*, 55–61.

[70] Coleman, T.; Chao, R. R.; Bruning, J. B.; De Voss, J. J.; Bell, S. G. CYP199A4 catalyses the efficient demethylation and demethenylation of para-substituted benzoic acid derivatives. *RSC Advances* **2015**, *5*, 52007–52018.

[71] Chao, R. R.; Lau, I. C.-K.; De Voss, J. J.; Bell, S. G. Modification of an Enzyme Biocatalyst for the Efficient and Selective Oxidative Demethylation of para-Substituted Benzene Derivatives. *ChemCatChem* **2016**, *8*, 3626–3635.

[72] Chao, R. R. Utilising CYP199A4 from Rhodopseudomonas palustris HaA2 for Biocatalysis and Mechanistic Studies. M.Sc. thesis, University of Adelaide, 2016.

[73] Coleman, T. Investigation of the Substrate Range and Mechanism of CYP199A4 Catalysed Oxidations. Honours Thesis, The University of Adelaide, 2013.

[74] Kloet, R.; Buning, T. d. C.; Bunders, J. An introduction to and a reflection on the "ecogenomics promise". *Journal of Integrative Environmental Sciences* **2011**, *8*, 23–28.

[75] Curtis, T. P.; Sloan, W. T.; Scannell, J. W. Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences* **2002**, *99*, 10494–10499.

[76] Lefevre, F.; Robe, P.; Jarrin, C.; Ginolhac, A.; Zago, C.; Auriol, D.; Vogel, T. M.; Simonet, P.; Nalin, R. Drugs from hidden bugs: their discovery via untapped resources. *Research in Microbiology* **2008**, *159*, 153 – 161.

[77] Riesenfeld, C. S.; Goodman, R. M.; Handelsma&amp, J. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environmental Microbiology* **2004**, *6*, 981 – 989.

[78] Pettit, R. Soil DNA libraries for anticancer drug discovery. *Cancer Chemotherapy and Pharmacology* **2004**, *54*, 1–6.

[79] Maurer, K. Detergent proteases. *Current opinion in Biotechnology* **2004**, *15*, 330–334.

[80] Makino, T.; Katsuyama, Y.; Otomatsu, T.; Misawa, N.; Ohnishi, Y. Regio- and Stereospecific Hydroxylation of Various Steroids at the 16α Position of the D Ring by the Streptomyces griseus Cytochrome P450 CYP154C3. *Applied and Environmental Microbiology* **2014**, *80*, 1371–1379.

[81] Sherman, D. H.; Li, S.; Yermalitskaya, L. V.; Kim, Y.; Smith, J. A.; Waterman, M. R.; Podust, L. M. The Structural Basis for Substrate Anchoring, Active Site Selectivity, and Product Formation by P450 PikC from Streptomyces venezuelae. *Journal of Biological Chemistry* **2006**, *281*, 26289–26297.

[82] Rittle, J.; Green, M. T. Cytochrome P450 Compound I: Capture, Characterization, and C-H Bond Activation Kinetics. *Science* **2010**, *330*, 933–937.

[83] Ma, J.; Wang, Z.; Huang, H.; Luo, M.; Zuo, D.; Wang, B.; Sun, A.; Cheng, Y.-Q.; Zhang, C.; Ju, J. Biosynthesis of Himastatin: Assembly Line and Characterization of Three Cytochrome P450 Enzymes Involved in the Post-tailoring Oxidative Steps. *Angewandte Chemie International Edition* **2011**, *50*, 7797–7802.

[84] Julsing, M. K.; Cornelissen, S.; Buhler, B.; Schmid, A. Heme-iron oxygenases: powerful industrial biocatalysts? *Current Opinion in Chemical Biology* **2008**, *12*, 177 – 186.

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

[86] Urlacher, V. B.; Girhard, M. Cytochrome P450 monooxygenases: an update on perspectives for synthetic application. *Trends in Biotechnology* **2012**, *30*, 26 – 36.

[87] Udwary, D. W. et al. Significant natural product biosynthetic potential of actinorhizal symbionts of the genus frankia, as revealed by comparative genomic and proteomic analyses. *Applied and Environmental Microbiology* **2011**, *77*, 3617–3625.

[88] Tisa, L. S.; Beauchemin, N.; Gtari, M.; Sen, A.; Wall, L. What stories can the Frankia genomes start to tell us? *Journal of Biosciences* **2013**, *38*, 719–726.

[89] Rawnsley, T.; Tisa, L. S. Development of a physical map for three Frankia strains and a partial genetic map for Frankia EuI1c. *Physiologia Plantarum* **2007**, *130*, 427–439.

[90] Schuster, G. Frankia alni. 2011; `https://commons.wikimedia.org/wiki/File:Frankia_alni.jpg`.

[91] Nouioui, I.; Ghodhbane-Gtari, F.; del Carmen Montero-Calasanz, M.; Rohde, M.; Tisa, L. S.; Gtari, M.; Klenk, H.-P. Frankia inefficax sp. nov., an actinobacterial endophyte inducing ineffective, non nitrogen-fixing, root nodules on its actinorhizal host plants. *Antonie van Leeuwenhoek* **2016**, 1–8.

[92] Rehan, M.; Furnholm, T.; Finethy, R. H.; Chu, F.; El-Fadly, G.; Tisa, L. S. Copper tolerance in Frankia sp. strain EuI1c involves surface binding and copper transport. *Applied Microbiology and Biotechnology* **2014**, *98*, 8005–8015.

[93] Furnholm, T. R.; Tisa, L. S. The ins and outs of metal homeostasis by the root nodule actinobacterium Frankia. *BMC Genomics* **2014**, *15*, 1092.

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

[95] Sambrook, J.; Maniatis, T.; Fritsch, E. F. *Molecular cloning : a laboratory manual*, 2nd ed.; Cold Spring Harbor Laboratory Press, 1987.

[96] Clarke, T. F., IV; Clark, P. L. Rare Codons Cluster. *PLoS ONE* **2008**, *3*, 1–5.

[97] Theodosiou, A.; Promponas, V. J. LaTcOm: a web server for visualizing rare codon clusters in coding sequences. *Bioinformatics* **2012**, *28*, 591–592.

[98] Otey, C. In *Directed Enzyme Evolution*; Arnold, F., Georgiou, G., Eds.; Methods in Molecular Biology; Humana Press, 2003; Vol. 230; pp 137–139.

[99] Williams, J. W.; Morrison, J. F. *Enzyme Kinetics and Mechanism Part A Initial Rate and Inhibitor Methods*; Methods in Enzymology; Academic Press, 1979; Vol. 63; pp 437 – 467.

[100] Golchoubian, H.; Hosseinpoor, F. Effective Oxidation of Sulfides to Sulfoxides with Hydrogen Peroxide under Transition-Metal-Free Conditions. *Molecules* **2007**, *12*, 304–311.

[101] Crawley, J.; S., L. U.; Nagata, W. Collection Volume VI. *Organic Syntheses* **1988**, 948.

[102] Battye, T. G. G.; Kontogiannis, L.; Johnson, O.; Powell, H. R.; Leslie, A. G. W. *iMOSFLM*: a new graphical interface for diffraction-image processing with *MOSFLM*. *Acta Crystallographica Section D* **2011**, *67*, 271–281.

[103] McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *Journal of Applied Crystallography* **2007**, *40*, 658–674.

[104] Project, C. The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D* **1994**, *50*, 760–763.

[105] Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L.-W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C. *PHENIX*: building new software for automated crystallographic structure determination. *Acta Crystallographica Section D* **2002**, *58*, 1948–1954.

[106] Emsley, P.; Cowtan, K. *Coot*: model-building tools for molecular graphics. *Acta Crystallographica Section D* **2004**, *60*, 2126–2132.

[107] Chen, V. B.; Arendall, W. B., III; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *MolProbity*: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D* **2010**, *66*, 12–21.

[108] Baker, D.; Newcomb, W.; Torrey, J. G. Characterization of an ineffective actinorhizal microsymbiont, Frankia sp. EuI1 (Actinomycetales). *Canadian Journal of Microbiology* **1980**, *26*, 1072–1089, PMID: 7459722.

[109] Benson, D. R.; Silvester, W. B. Biology of Frankia strains, actinomycete symbionts of actinorhizal plants. *Microbiological Reviews* **1993**, *57*, 293–319.

[110] Nouioui, I.; Ghodhbane-Gtari, F.; Beauchemin, N. J.; Tisa, L. S.; Gtari, M. Phylogeny of members of the Frankia genus based on gyrB, nifH and glnII sequences. *Antonie van Leeuwenhoek* **2011**, *100*, 579.

[111] Normand, P. et al. Genome characteristics of facultatively symbiotic Frankia sp. strains reflect host range and host plant biogeography. *Genome Research* **2007**, *17*, 7–15.

[112] Marathon, M. Allocasuarina luehmannii foliage. 2015; `https://commons.wikimedia.org/wiki/File:Allocasuarina_luehmannii_foliage.jpg`.

[113] Rehan, M.; Kluge, M.; Fränzle, S.; Kellner, H.; Ullrich, R.; Hofrichter, M. Degradation of atrazine by Frankia alni ACN14a: gene regulation, dealkylation, and dechlorination. *Applied Microbiology and Biotechnology* **2014**, *98*, 6125–6135.

[114] Baker, E.; Tang, Y.; Chu, F.; Tisa, L. S. Molecular responses of Frankia sp. strain QA3 to naphthalene. *Canadian Journal of Microbiology* **2015**, *61*, 281–292, PMID: 25742598.

[115] Richards, J. W.; Krumholz, G. D.; Chval, M. S.; Tisa, L. S. Heavy Metal Resistance Patterns of Frankia Strains. *Applied and Environmental Microbiology* **2002**, *68*, 923–927.

[116] Tisa, L. S.; Oshone, R.; Sarkar, I.; Ktari, A.; Sen, A.; Gtari, M. Genomic approaches toward understanding the actinorhizal symbiosis: an update on the status of the Frankia genomes. *Symbiosis* **2016**, *70*, 5–16.

[117] Benson, D.; Brooks, J.; Huang, Y.; Bickhart, D.; Mastronunzio, J. The Biology of Frankia sp. Strains in the Post-Genome Era. *Molecular Plant-Microbe Interactions* **2011**, *24*, 1310–1316.

[118] Muto, A.; Osawa, S. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proceedings of the National Academy of Sciences of the United States of America* **1987**, *84*, 166–169.

[119] Nakamura, Y.; Gojobori, T.; Ikemura, T. Codon usage tabulated from the international DNA sequence databases: status for the year 2000. *Nucleic Acids Research* **2000**, *28*, 292.

[120] Kane, J. F. Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. *Current Opinion in Biotechnology* **1995**, *6*, 494 – 500.

[121] Sun, Y.; Zeng, W.; Benabbas, A.; Ye, X.; Denisov, I.; Sligar, S. G.; Du, J.; Dawson, J. H.; Champion, P. M. Investigations of Heme Ligation and Ligand Switching in Cytochromes P450 and P420. *Biochemistry* **2013**, *52*, 5941–5951, PMID: 23905516.

[122] Sabat, J.; Stuehr, D. J.; Yeh, S.-R.; Rousseau, D. L. Characterization of the Proximal Ligand in the P420 Form of Inducible Nitric Oxide Synthase. *Journal of the American Chemical Society* **2009**, *131*, 12186–12192, PMID: 19658411.

[123] Hui Bon Hoa, G.; Di Primo, C.; Dondaine, I.; Sligar, S. G.; Gunsalus, I. C.; Douzou, P. Conformational changes of cytochromes P-450cam and P-450lin induced by high pressure. *Biochemistry* **1989**, *28*, 651–656.

[124] Champion, P. M.; Gunsalus, I. C.; Wagner, G. C. Resonance Raman investigations of cytochrome P450CAM from Pseudomonas putida. *Journal of the American Chemical Society* **1978**, *100*, 3743–3751.

[125] Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **1990**, *215*, 403–410.

[126] Katagiri, M.; Ganguli, B. N.; Gunsalus, I. C. A Soluble Cytochrome P-450 Functional in Methylene Hydroxylation. *Journal of Biological Chemistry* **1968**, *243*, 3543–3546.

[127] Hawkes, D. B.; Adams, G. W.; Burlingame, A. L.; Ortiz de Montellano, P. R.; De Voss, J. J. Cytochrome P450cin (CYP176A), Isolation, Expression, and Characterization. *Journal of Biological Chemistry* **2002**, *277*, 27725–27732.

[128] Boddupalli, S. S.; Hasemann, C. A.; Ravichandran, K. G.; Lu, J.-Y.; Goldsmith, E. J.; Deisenhofer, J.; Peterson, J. A. Crystallization and Preliminary X-Ray Diffraction Analysis of P450Terp and the Hemoprotein Domain of P450BM-3, Enzymes Belonging to Two Distinct Classes of the Cytochrome P450 Superfamily.

Proceedings of the National Academy of Sciences of the United States of America **1992**, *89*, 5567–5571.

[129] Girhard, M.; Klaus, T.; Khatri, Y.; Bernhardt, R.; Urlacher, V. B. Characterization of the versatile monooxygenase CYP109B1 from Bacillus subtilis. *Applied Microbiology and Biotechnology* **2010**, *87*, 595–607.

[130] Khatri, Y.; Hannemann, F.; Girhard, M.; Kappl, R.; Meme, A.; Ringle, M.; Janocha, S.; Leize-Wagner, E.; Urlacher, V. B.; Bernhardt, R. Novel family members of CYP109 from Sorangium cellulosum So ce56 exhibit characteristic biochemical and biophysical properties. *Biotechnology and Applied Biochemistry* **2013**, *60*, 18–29.

[131] Girhard, M.; Machida, K.; Itoh, M.; Schmid, R. D.; Arisawa, A.; Urlacher, V. B. Regioselective biooxidation of (+)-valencene by recombinant E. coli expressing CYP109B1 from Bacillus subtilis in a two-liquid-phase system. *Microbial Cell Factories* **2009**, *8*, 36.

[132] Khatri, Y.; Girhard, M.; Romankiewicz, A.; Ringle, M.; Hannemann, F.; Urlacher, V. B.; Hutter, M. C.; Bernhardt, R. Regioselective hydroxylation of norisoprenoids by CYP109D1 from Sorangium cellulosum So ce56. *Applied Microbiology and Biotechnology* **2010**, *88*, 485–495.

[133] Zhang, A.; Zhang, T.; Hall, E. A.; Hutchinson, S.; Cryle, M. J.; Wong, L.-L.; Zhou, W.; Bell, S. G. The crystal structure of the versatile cytochrome P450 enzyme CYP109B1 from Bacillus subtilis. *Molecular Biosystems* **2015**, *11*, 869–881.

[134] Child, S. A. Unpublished PhD Thesis. Ph.D. thesis, Univeristy of Adelaide, 2017.

[135] Rauschenbach, R.; Isernhagen, M.; Noeske-Jungblut, C.; Boidol, W.; Siewert, G. Cloning sequencing and expression of the gene for cytochrome P450meg, the steroid-15β-monooxygenase from Bacillus megaterium ATCC 13368. *Molecular and General Genetics MGG* **1993**, *241*, 170–176.

[136] Bruschi, M.; Guerlesquin, F. Structure, function and evolution of bacterial ferredoxins. *FEMS Microbiology Letters* **1988**, *54*, 155–175.

[137] McLean, K.; Warman, A.; Seward, H.; Marshall, K.; Girvan, H.; Cheesman, M.; M.R, W.; A.W., M. Biophysical characterization of the sterol demethylase P450 from Mycobacterium tuberculosis, its cognate ferredoxin, and their interactions. *Biochemistry* **2006**, *45*, 8427–8443.

[138] Duret, L. Evolution of synonymous codon usage in metazoans. *Current Opinion in Genetics & Development* **2002**, *12*, 640 – 649.

[139] Erdmann, V. A. Collection of published 5S and 5.8S rRNA sequences and their precursors. *Nucleic Acids Research* **1980**, *8*, 197.

[140] Médigue, C.; Rouxel, T.; Vigier, P.; Hénaut, A.; Danchin, A. Evidence for horizontal gene transfer in Escherichia coli speciation. *Journal of Molecular Biology* **1991**, *222*, 851 – 856.

[141] Correia, M. A.; Sinclair, P. R.; De Matteis, F. Cytochrome P450 regulation: the interplay between its heme and apoprotein moieties in synthesis, assembly, repair, and disposal. *Drug Metabolism Reviews* **2011**, *43*, 1–26.

[142] Pritchard, M. P.; McLaughlin, L.; Friedberg, T. Establishment of functional human cytochrome P450 monooxygenase systems in Escherichia coli. *Methods in Molecular Biology* **2006**, *320*, 19–29.

[143] Binda, E.; Marcone, G. L.; Berini, F.; Pollegioni, L.; Marinelli, F. Streptomyces spp. as efficient expression system for a d,d-peptidase/d,d-carboxypeptidase involved in glycopeptide antibiotic resistance. *BMC Biotechnology* **2013**, *13*, 24.

[144] Kaderbhai, M. A.; Ugochukwu, C. C.; Kelly, S. L.; Lamb, D. C. Export of Cytochrome P450 105D1 to the Periplasmic Space of Escherichia coli. *Applied and Environmental Microbiology* **2001**, *67*, 2136–2138.

[145] Crack, J. C.; Green, J.; Thomson, A. J.; Le Brun, N. E. In *Metalloproteins: Methods and Protocols*; Fontecilla-Camps, J. C., Nicolet, Y., Eds.; Humana Press: Totowa, NJ, 2014; pp 33–48.

[146] Zhou, Z.; Adams, M. Site-directed mutations of the 4Fe-Ferredoxin from the hyperthermophilic archaeon Pyrococcus furiosus: Role of the cluster-coordinating aspartate in physiological electron transfer reactions. *Biochemistry* **1997**, *36*, 10892–10900.

[147] Averill, B. A.; Bale, J. R.; Orme-Johnson, W. H. Displacement of iron-sulfur clusters from ferredoxins and other iron-sulfur proteins. *Journal of the American Chemical Society* **1978**, *100*, 3034–3043.

[148] Macomber, L.; Imlay, J. A. The iron-sulfur clusters of dehydratases are primary intracellular targets of copper toxicity. *Proceedings of the National Academy of Sciences* **2009**, *106*, 8344–8349.

[149] Jaganaman, S.; Pinto, A.; Tarasev, M.; Ballou, D. P. High levels of expression of the iron-sulfur proteins phthalate dioxygenase and phthalate dioxygenase reductase in Escherichia coli. *Protein Expression and Purification* **2007**, *52*, 273 – 279.

[150] Nakamura, K., M. Saeki; Takahashi, Y. Hyperproduction of recombinant ferredoxins in Escherichia coli by coexpression of the ORF1- ORF2-iscS-iscU-iscA-hscB-hs cA-fdx-ORF3 gene cluster. *Journal of Biochemistry (Tokyo, Japan)* **1999**, 10–18.

[151] Chen, B.; Menon, N. K.; Dervertarnian, L.; Moura, J. J.; Przybyla, A. E. Cloning, sequencing and overexpression of the Desulfovibrio gigas ferredoxin gene in E. coli. *FEBS Letters* **1994**, *351*, 401 – 404.

[152] Metzler, D. E.; Metzler, C. M.; Scott, R. D.; Mollova, E. T.; Kagamiyama, H.; Yano, T.; Kuramitsu, S.; Hayashi, H.; Hirotsu, K.; Miyahara, I. NMR studies of 1H resonances in the 10-18-ppm range for aspartate aminotransferase from Escherichia coli. *Journal of Biological Chemistry* **1994**, *269*, 28027–33.

[153] Schrautemeier, B.; Cassing, A.; Bohme, H. Characterization of the genome region encoding an fdxH-type ferredoxin and a new 2[4Fe-4S] ferredoxin from the nonheterocystous, nitrogen-fixing cyanobacterium Plectonema boryanum PCC 73110. *Journal of Bacteriology* **1994**, *176*, 1037–1046.

[154] Daff, S. N.; Chapman, S. K.; ; Holt, R. A.; Govindaraj, S.; Poulos, T. L.; Munro, A. W. Redox Control of the Catalytic Cycle of Flavocytochrome P-450 BM3. *Biochemistry* **1997**, *36*, 13816–13823.

[155] Sligar, S. G. Coupling of spin, substrate, and redox equilibriums in cytochrome P450. *Biochemistry* **1976**, *15*, 5399–5406.

[156] Zhang, T.; Zhang, A.; Bell, S. G.; Wong, L.-L.; Zhou, W. The structure of a novel electron-transfer ferredoxin from *Rhodopseudomonas palustris* HaA2 which contains a histidine residue in its iron–sulfur cluster-binding motif. *Acta Crystallographica Section D* **2014**, *70*, 1453–1464.

[157] Jackson, C. J.; Lamb, D. C.; Marczylo, T. H.; Warrilow, A. G. S.; Manning, N. J.; Lowe, D. J.; Kelly, D. E.; Kelly, S. L. A Novel Sterol 14α-Demethylase/Ferredoxin Fusion Protein (MCCYP51FX) from Methylococcus capsulatusRepresents a New Class of the Cytochrome P450 Superfamily. *Journal of Biological Chemistry* **2002**, *277*, 46959–46965.

[158] Duderstadt, R.; Staples, C.; Brereton, P.; Adams, M.; Johnson, M. Effects of mutations in aspartate 14 on the spectroscopic properties of the [Fe3S4]+,0 clusters in Pyrococcus furiosus ferredoxin. *Biochemistry* **1999**, *38*, 10585–10593.

[159] Nielsen, M.; Harris, P.; Ooi, B.; Christensen, H. The 1.5 angstrom resolution crystal structure of [Fe3S4]-ferredoxin from the hyperthermophilic archaeon Pyrococcus furiosus. *Biochemistry* **2004**, *43*, 5188–5194.

[160] Rousset, M.; Montet, Y.; Guigliarelli, B.; Forget, N.; Asso, M.; Bertrand, P.; Fontecilla-Camps, J. C.; Hatchikian, E. C. [3Fe-4S] to [4Fe-4S] cluster conversion in Desulfovibrio fructosovorans [NiFe] hydrogenase by site-directed mutagenesis. *Proceedings of the National Academy of Sciences* **1998**, *95*, 11625–11630.

[161] Xu, F.; Bell, S. G.; Peng, Y.; Johnson, E. O.; Bartlam, M.; Rao, Z.; Wong, L. L. Crystal structure of a ferredoxin reductase for the CYP199A2 system from Rhodopseudomonas palustris. *Proteins* **2009**, *77*, 867–880.

[162] Coleman, T.; Chao, R. R.; Voss, J. J. D.; Bell, S. G. The importance of the benzoic acid carboxylate moiety for substrate recognition by CYP199A4 from Rhodopseudomonas palustris HaA2. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2016**, *1864*, 667 – 675.

[163] Coleman, T. Unpublished PhD Thesis. Ph.D. thesis, The University of Adelaide, 2017.

[164] Nomenclature and Symbolism for Amino Acids and Peptides. *European Journal of Biochemistry* **1984**, *138*, 9–37.

Appendix A Data for introduction and experimental

FraEuI1c_1414
5'TTATTATCCATGGCCACCGATGCCTGGCGTATTGAAGTGGATCGCCGTCGCTGT
GTGGGCACCGGTGCGTGTGCCTACGCGGCACCTGAAGTGTTTCGTCTGGATGATC
GCAGCTTAGCAACCGTCGTGGGCAAGGTCGACGGCGATGATGCCTTTCTCCGCG
ATGTGGTGGCCGAATGTCCTACGGAAGCTCTGCGTCTGTCTCCAATCGATGATGC
CCGTTAAAAGCTTTAATAA3'

FraEu1c_2495
5'TTATTATCCATGGCCAAGGTGGCTGTAGATGGCGCGAAATGCACAGGCCACGC
TCGCTGTCAGGCGACAGCACCAGAAGTTTTCGCGCTGGATGAGCTCGGGTATGCT
GTTCCGGGAGAACGCGAGATCGCGGCAGGCGCCGAGGCCGAAGCCCGTCGTGGC
GCATCGGCATGCCCTGAACGCGCTATTACCGTTTGGGGTGATCCGTTAGGTCGCC
CGGCCCGCTAAAAGCTTTAATAA3'

FraEuI1c_3227
5'TTATTATCCATGGCCTCTATTAAAGTCACGTTTCGCCTTGCGGATGGGTCCGCC
CGCGTGGTGGTTGCGGAACCGGGCACCTCGGTGATGCGCGTCGCAGTGGAAAAC
GGCGTTCCGGGCATTGTTGGTGAGTGTGGCGGTGAAATGTCCTGTGCGACTTGTC
ATGTGCATGTAGACCCAGATCTGGCATTCCGCAGTATGAGCGCGGATGAAGAGG
ACATGCTGGAAGCAGTGGACGATCGCTTACCGACTAGCCGTCTGGGCTGTCAGTT
GATTTTACGCGATGAACTGGGCGAAATTGCCGTGACAGTACCATCATAAAAGCTT
TAATAA3'

FraEuI1c_4132
5'TTATTATCCATGGCCGGCGGTGCGGCAACCATTGCGGTGGATCGCGCCCGCTGC
GTGGGTTCCGGCATGTGCGTCGTTTACGCACCAGATACCTTCACGCATGATGAGG
AAGCTAAAGTAGTTGTGCTCGATCCGTTAGGAGACCCAGCGGAAGTAGTGGAGA
GTGCCGTGGAAGCGTGTCCTACCGGCGCACTGACATTAGTGACGGCGGGAGAGG
CGGGCGGTTAAAAGCTTTAATAA3'

FraEuI1c_4178
5'TTATTATCCATGGAAAAATCAGGTTTGGCTTTTAGCGTAGATCGTGAAAGTTGC
TTAGGCAGCGGCAGCTGTGCGTTTCACGCGCCTGGCACCTTTGATCTGGACGATG
ATCTGAAGGTTGTTGTGCTTGCAGACGGCGATCCAGATGCGGCAGTAGCAGCCG
CAATTGATTCCTGTCCCTCCGGGGCGCTGCGCCGTAAAGAGACAGGCTAAAAGC
TTTAATAA3'

FraEuI1c_5333
5'TTATTATCCATGGTTGCACGTCACATCACGATTGATCGCGATCTGTGTATGGGT
AGCGGCCAATGCCTGATCTACGCACCGAACACTTTCGACTTGGATGATGATGCGA
TTGCGACTGTAGTGGATCCTGATGGTGACTCCGACGCAGAACTGGCTTCCGCGGT
AACCGGGTGTCCTACACAGGCCATTAGCGTGGCCCGCGATTAAAAGCTTTAATA
A3'

FraEuI1c_5370
5'CGTTATTATCCATGGTTAAAGTTTTCATTGATGACGATAACTGTCGTGGgCATG
GTGTCTGTGCGGCCGAATGTCCGCAAGTCTTCGTGGTCAATGATGACGGGTATGG
CGAAGTTATCGTTGATGAAGTTCCGGCGGAACTGGAAGATGCCGTTCGCGTTGTA
GCGACCCATTGTCCGGAACGTGCCATTACCGTCGAATAAAAGCTTTAATAA3'

FraEuI1c_5464
5'TTATTATCCATGGCCAAAGCCGAAGTTGATGCGGCGAAATGTCAAGGCCACGC
TCGTTGCTGGGAAATTTGTCCGGAAGTGTTTGCGCTGGATGAAGAAGGCTTCGGT
TCAGTCGTTGATCCAGAAGTCCCGGTCGAACTTGAAGCCAAAGCGCGTGAAGCC
GCGGATAACTGTCCTGAACGTGCAATTGTACTGGCCTAAAAGCTTTAATAA3'

FraEu1c_3755
5'TTATTATCCATGGTTACGGATGATCGTTCAGGGGTTGCGACGCCGGCAAACTCT
TCCGTGTCGGGCGGCTTACGTGTCCGCGTAGATCAGGCCGCCTGCACGGGAGAT
GGCCTGTGCGTCACGCGCGCTCCATCCGTCTTTGAACACGATATCGACGGACTGG
CGTATGTGAAAACCGACACCGGCGAACTGTTGACCCAACCGGGTGCCCGTGTTA
TTGTCCCGCTGCCTTTATTAGATGCTGTTATTGATGCGGCCGACGAATGCCCTGG
GGATTGCATTTATGTGGAACGCCCGGATGGAACGCTGGAAGCGGGCCCTGGGAG
TTAATGAAAGCTTTAATAA3'

FraEu1c_3991
5'TTATTATCCATGGCCAAGGTGGTGGTCGATTTTGATGTGTGCGAGTCTAATGCC
ATTTGTGTAGGCCTGGTGCCTTCAGTGTTCGAACTGGATGATAACGAGTATCTGA
GTGTGGAGGAAGAACCGGATGAGTCGCTGCGTCCGGATATCGAGAAGGCAGTCC
GGGCCTGCCCGAAACAGGCAATCAAAATTGTGGACTAAAAGCTTTAATTA3'

Figure A1: The codon optimised gblocks of ferredoxins from *Frankia* sp. EuI1c, purchased from Integrated DNA technology.

Appendix B Data for chapter 3

Table B1: The 68 P450 encoding genes of *Frankia* sp. EuI1c

FraEuI1c_0568	*FraEuI1c_0570*	*FraEuI1c_1415*	*FraEuI1c_1460*
FraEuI1c_1466	*FraEuI1c_1505*	*FraEuI1c_1866*	*FraEuI1c_1873*
FraEuI1c_2002	*FraEuI1c_2034*	*FraEuI1c_2049*	*FraEuI1c_2104*
FraEuI1c_2176	*FraEuI1c_2477*	*FraEuI1c_2494*	*FraEuI1c_2517*
FraEuI1c_2543	*FraEuI1c_2615*	*FraEuI1c_2616*	*FraEuI1c_2625*
FraEuI1c_2626	*FraEuI1c_2668*	*FraEuI1c_2856*	*FraEuI1c_2871*
FraEuI1c_2918	*FraEuI1c_3226*	*FraEuI1c_3279*	*FraEuI1c_3519*
FraEuI1c_3332	*FraEuI1c_3357*	*FraEuI1c_3373*	*FraEuI1c_3378*
FraEuI1c_3402	*FraEuI1c_3412*	*FraEuI1c_3868*	*FraEuI1c_3940*
FraEuI1c_3942	*FraEuI1c_4131*	*FraEuI1c_4144*	*FraEuI1c_4179*
FraEuI1c_4209	*FraEuI1c_4257*	*FraEuI1c_4302*	*FraEuI1c_4364*
FraEuI1c_4439	*FraEuI1c_4564*	*FraEuI1c_4964*	*FraEuI1c_5083*
FraEuI1c_1466	*FraEuI1c_5098*	*FraEuI1c_5308*	*FraEuI1c_5327*
FraEuI1c_5334	*FraEuI1c_5369*	*FraEuI1c_5373*	*FraEuI1c_5429*
FraEuI1c_5348	*FraEuI1c_5466*	*FraEuI1c_5470*	*FraEuI1c_5501*
FraEuI1c_5532	*FraEuI1c_5744*	*FraEuI1c_5758*	*FraEuI1c_5997*
FraEuI1c_6000	*FraEuI1c_6004*	*FraEuI1c_6583*	*FraEuI1c_7025*

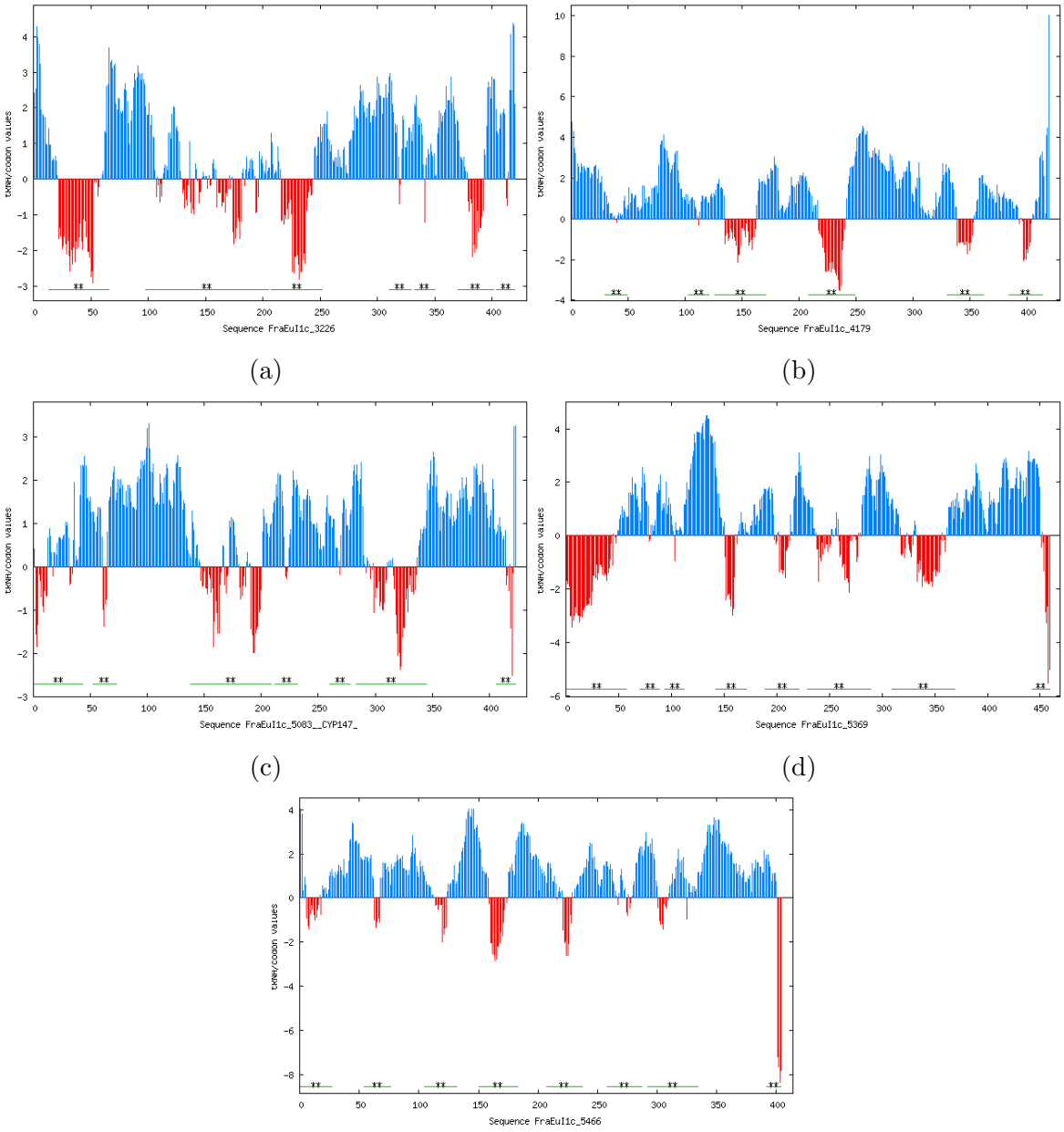


Figure B1: Rare codon clusters for P450s from *FraEuI1c_3226* (a), *FraEuI1c_4179* (b), *FraEuI1c_5083* (c), *FraEuI1c_5369* (d) and *FraEuI1c_5466* (e). The RCC identification method was %MinMax, with a window size of 19.

Table B2: Codon usage in *E. coli* K12

Codon	Percentage (%)	Codon	Percentage (%)	Codon	Percentage (%)	Codon	Percentage (%)
UUU	1.01	UCU	0.29	UAU	0.86	UGU	0.3
UUC	0.77	UCC	0.28	UAC	0.75	UGC	0.41
UUA	0.78	UCA	0.4	UAA	0.09	UGA	0.05
UUG	0.61	UCG	0.41	UAG	0.01	UGG	0.55
CUU	0.61	CCU	0.43	CAU	0.81	CGU	1.08
CUC	0.54	CCC	0.33	CAC	0.67	CGC	1.33
CUA	0.27	CCA	0.34	CAA	0.62	CGA	0.22
CUG	2.4	CCG	1.37	CAG	1.42	CGG	0.21
AUU	1.56	ACU	0.41	AAU	1.22	AGU	0.37
AUC	0.93	ACC	1.77	AAC	1.25	AGC	0.85
AUA	0.19	ACA	0.33	AAA	1.70	AGA	0.07
AUG	1.27	ACG	0.59	AAG	0.62	AGG	0.08
GUU	0.86	GCU	0.55	GAU	1.94	GGU	1.09
GUC	0.6	GCC	1.62	GAC	1.05	GGC	1.71
GUA	0.59	GCA	1.08	GAA	2.24	GGA	0.47
GUG	1.35	GCG	1.97	GAG	0.94	GGG	0.44

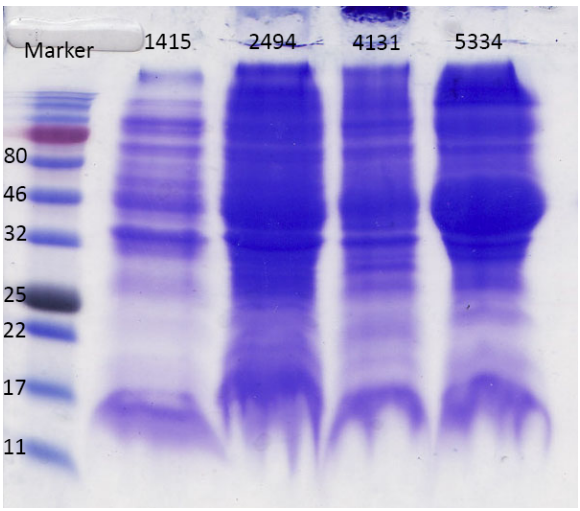


Figure B2: 12% SDS gels of (left to right) protein ladder marker, FraEu1415, FraEu2494, FraEu4131 and FraEu5334. Samples were taken from the cell pellet.

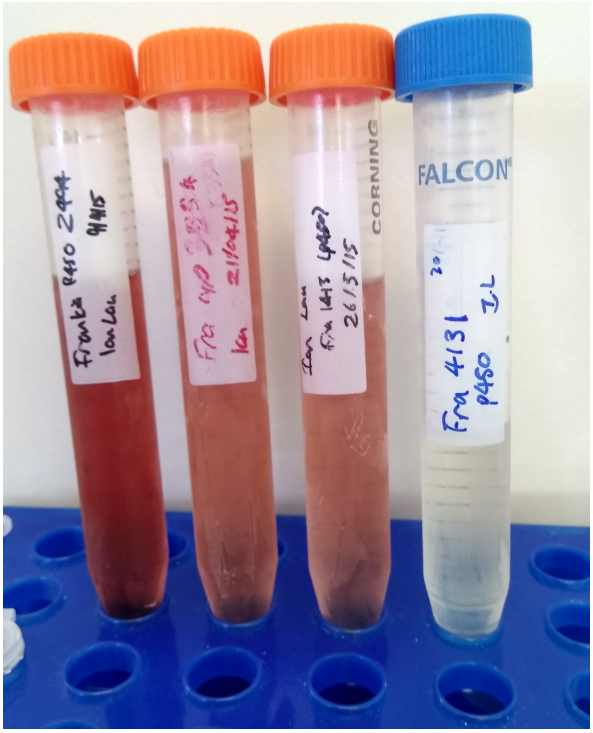


Figure B3: The purified P450s from *Frankia* sp. EuI1c: (from left to right) FraEu2494, FraEu5334, FraEu1415 and FraEu4131.

Substrates tested with FraEu2494 with no binding affinity

Table B3: Select substrates tested with no binding affinity to FraEu2494.

Substrate	Substrate class
Capric acid	Fatty acid
Farnesol	Acyclic sesquiterpene
Camphor	Monoterpenoid
Fenchone	Monoterpene
β-ionone	Norisoprenoid
Neryl acetate	Monoterpene acetate
4-Phenyl-2-butanone	Benzene ketone
Cyclohexyl acetate	6 member ring acetate
Nootkatone	Sesquiterpene
Testosterone	Steroid

Binding data for FraEu2494

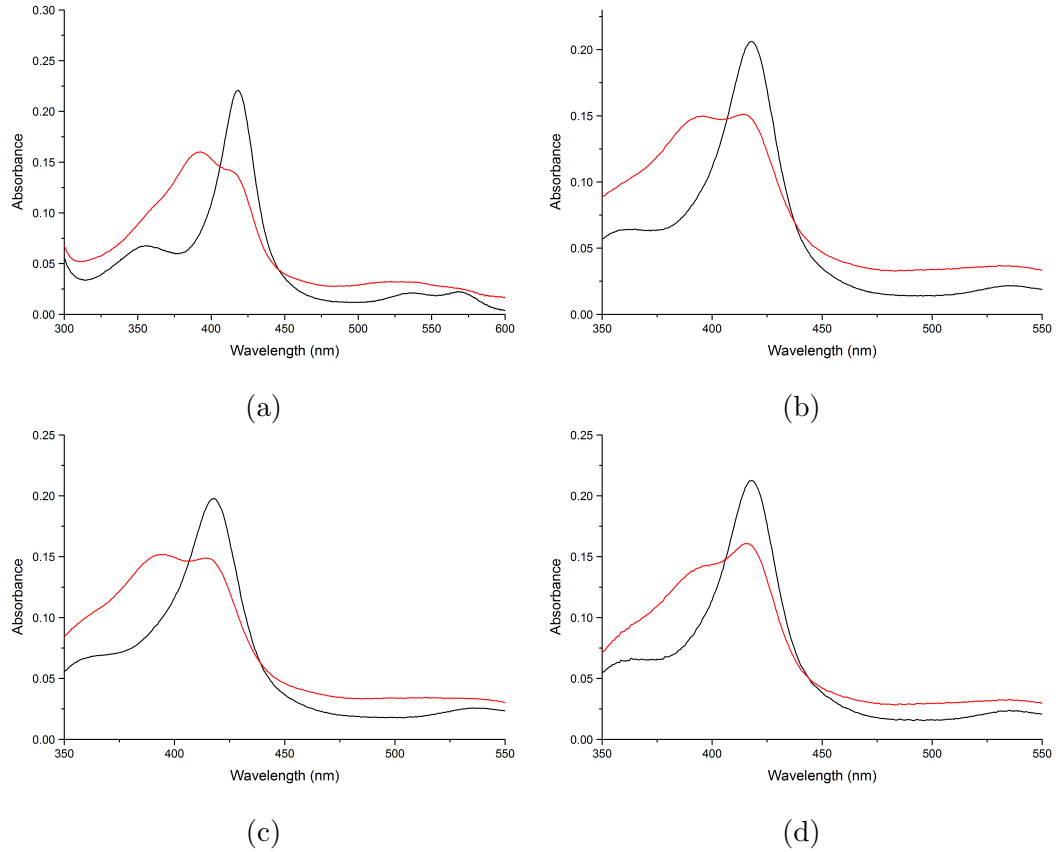


Figure B4: The spin-state shifts of FraEu2494 with *trans*-decahydronaphthalene (a), α-ionone (b), β-damascone (c) and β-ionol (d). For clarity the chromatograms have been offset along the y axes for (b), (c) and (d).

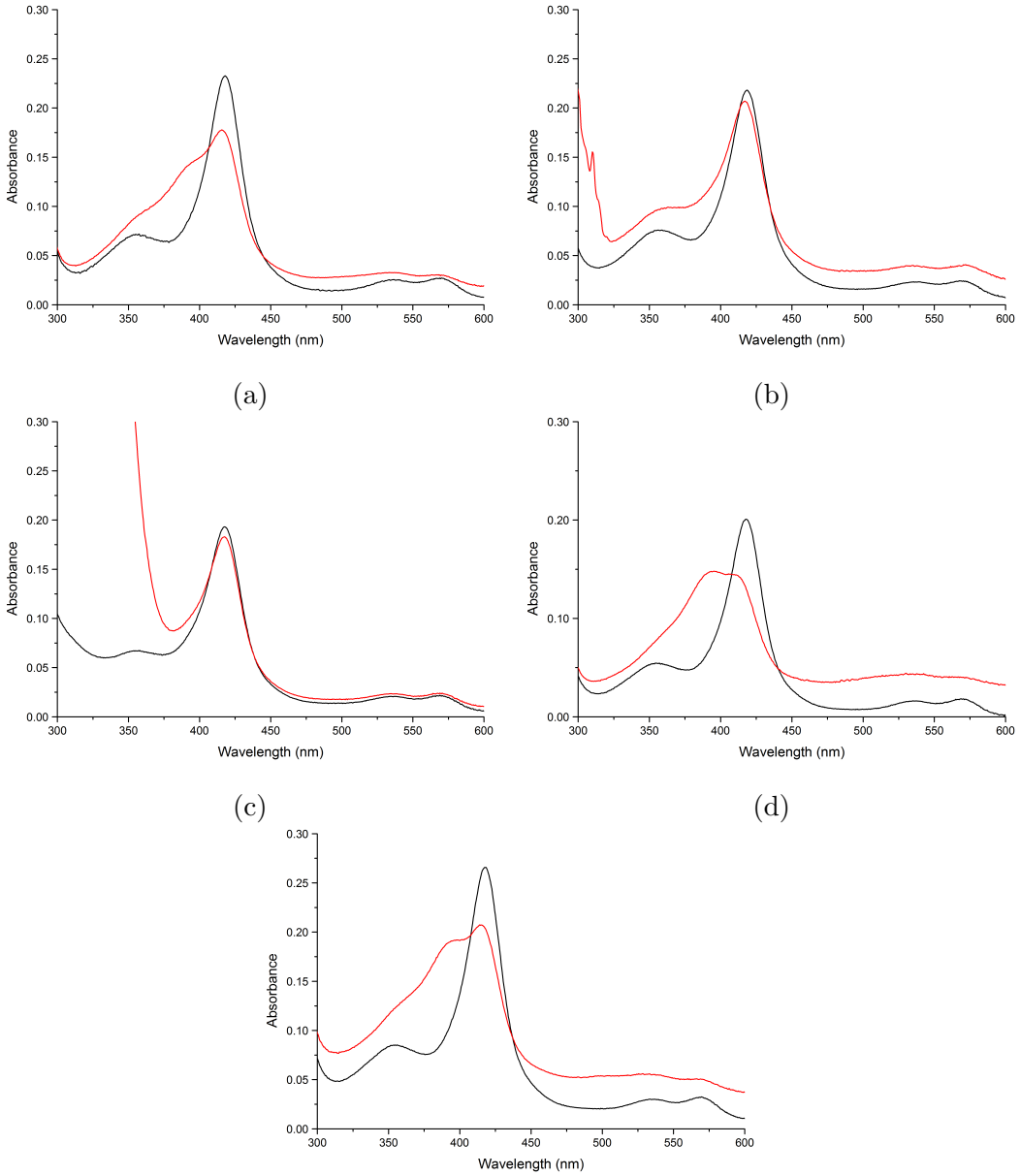


Figure B5: The spin-state shifts of FraEu2494 with fenchyl acetate (a), naphthalene (b), β-ionone (c), sclareolide (d) and bornyl acetate (e).

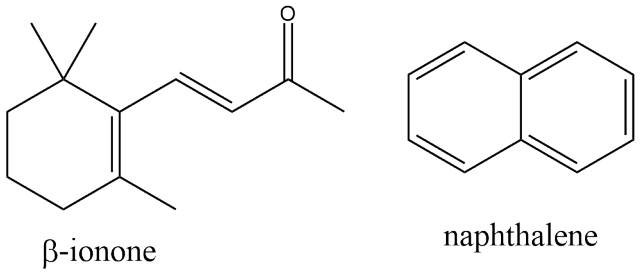


Figure B6: Structures of β-ionone and naphthalene

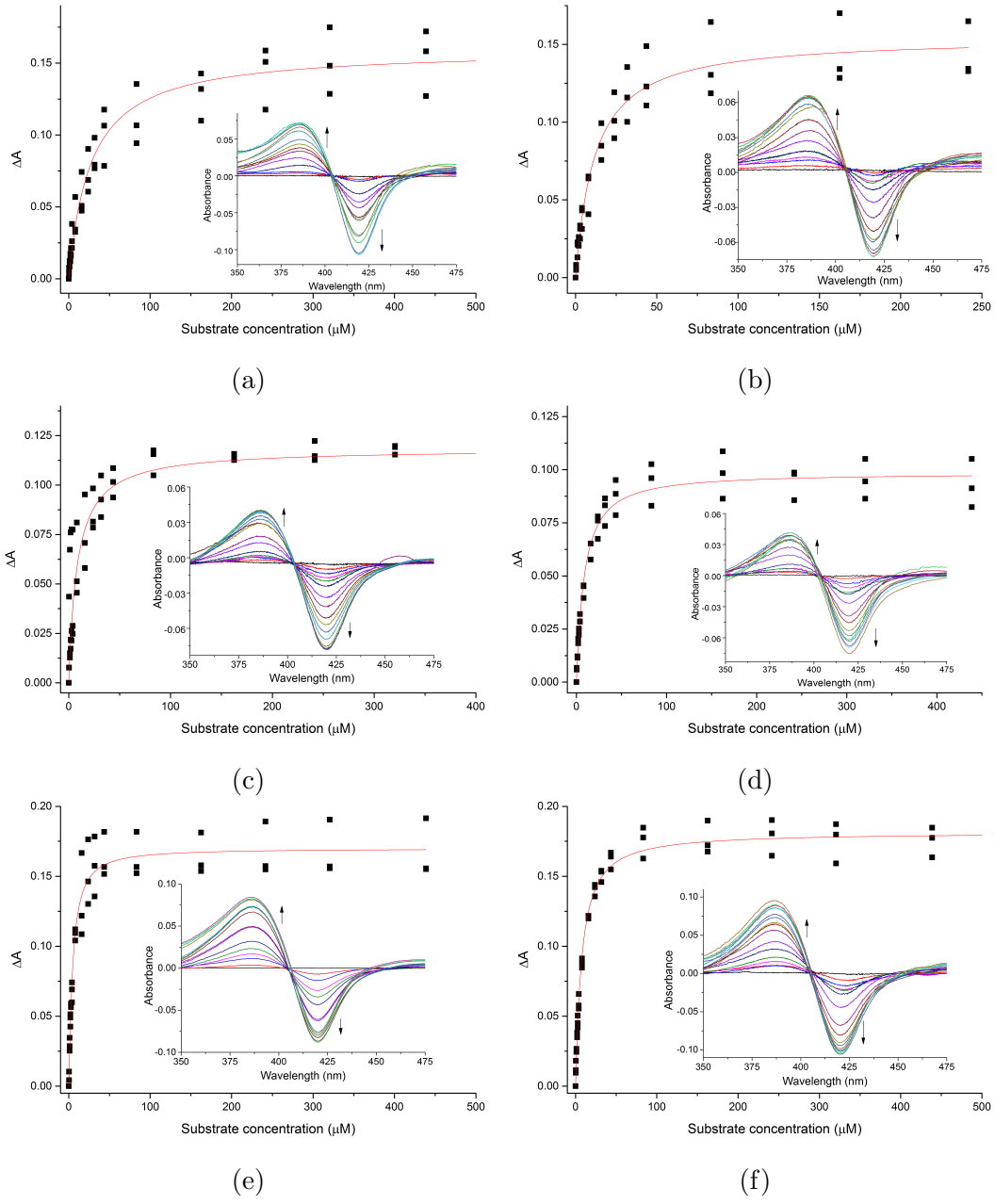


Figure B7: Dissociation analyses of FraEu2494 with β-damascone (a), β-ionol (b), bornyl acetate (c), fenchyl acetate (d), *trans*-decahydronaphthalene (e) and α-ionol (f).

Data for FraEu5334

Substrates tested with FraEu5334 with no binding affinity

Table B4: Select substrates tested with no binding affinity to FraEu5334.

Substrate	Substrate class
Farnesol	Acyclic sesquiterpene
Camphor	Monoterpenoid
Octanoic acid	Fatty acid
trans-decalin	Polycyclic carbon chain
Naphthalene	Polycyclic aromatics
Geranyl acetate	Monoterpene acetate
Ambroxide	Sesquiterpenoid
Nootkatone	Sesquiterpene
Testosterone	Steroid

Substrates tested with FraEu1415 with no binding affinity

Table B5: Select substrates tested with no binding affinity to FraEu1415.

Substrate	Substrate class
α-ionone	Norisoprenoid
Ambroxide	Sesquiterpenoid
Octanoic acid	Fatty acid
Tetralin	Bicyclic compounds
Naphthalene	Polycyclic aromatic
Camphor	Monoterpenoid
Cholesterol	Steroid
Fenchyl acetate	Bicyclic monoterpenoid
p-cymene	Alkylbenzene

Binding data for FraEu5334

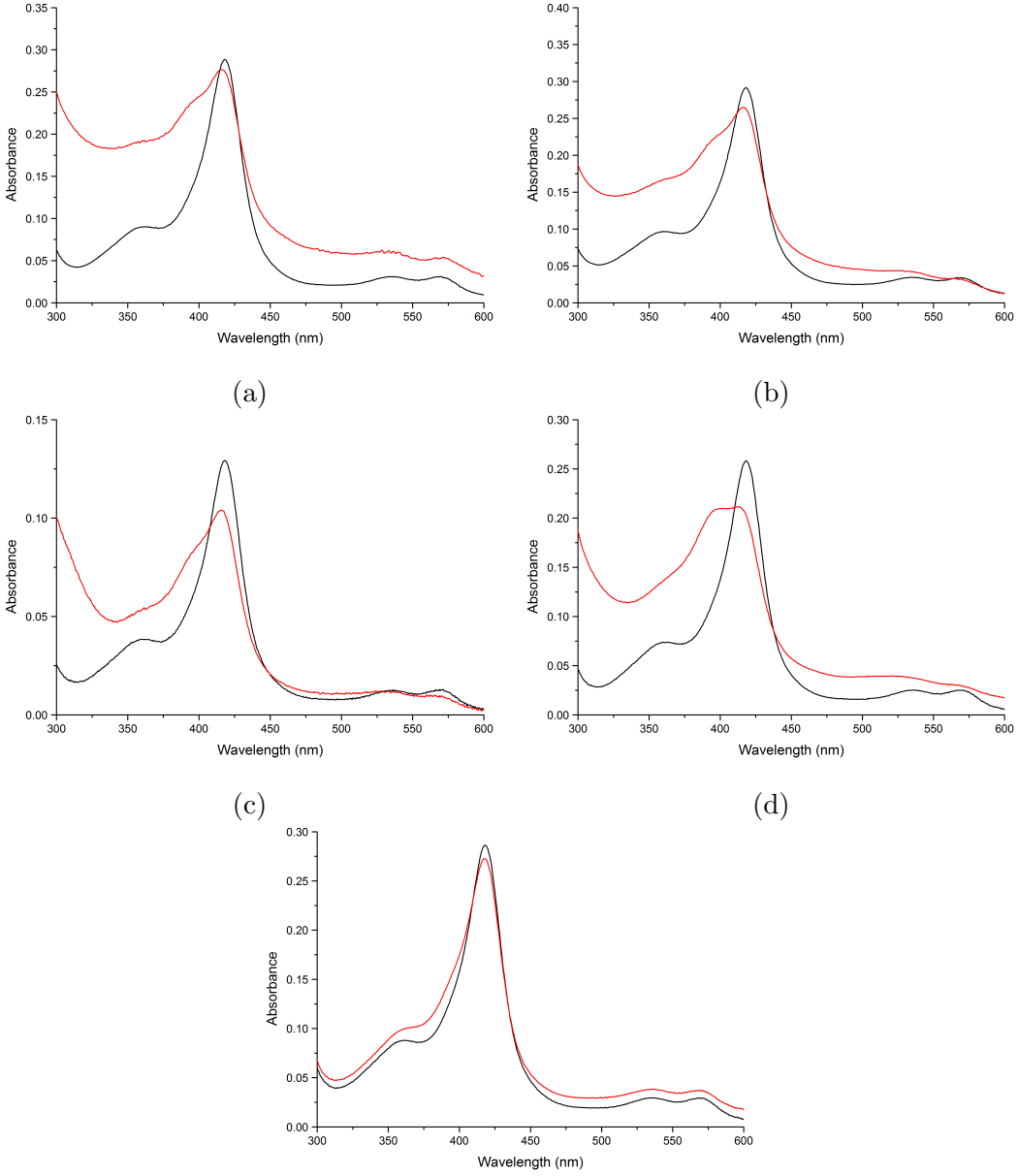


Figure B8: Spin-state shifts of FraEu5334 with *cis*-nerolidol (a), neryl acetate (b), nootkatone (c) methyl-ionone (d), and geranyl acetate (e).

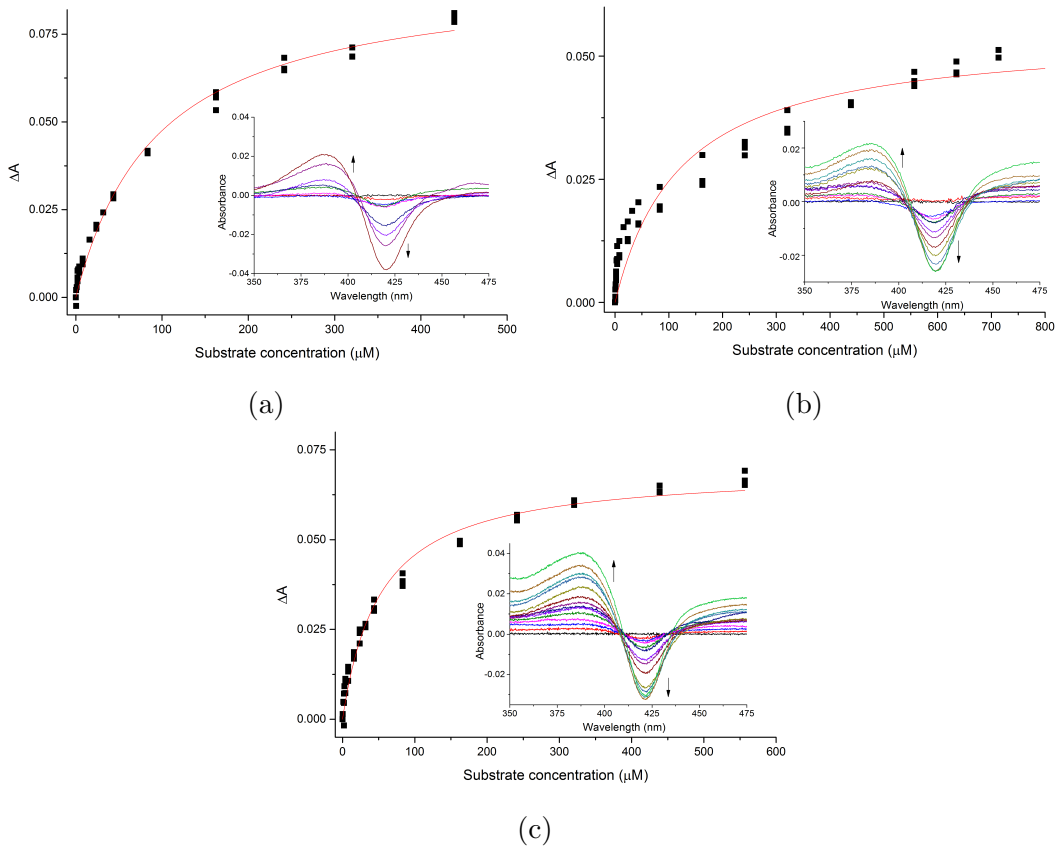


Figure B9: Dissociation analyses of FraEu5334 with *cis*-nerolidol (a), neryl acetate (b) and nootkatone (c).

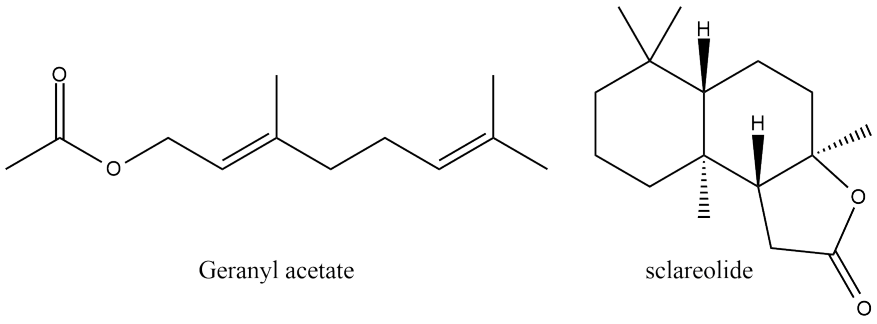


Figure B10: Structures of geranyl acetate and sclareolide

Data for FraEu1415

Binding data for FraEu1415

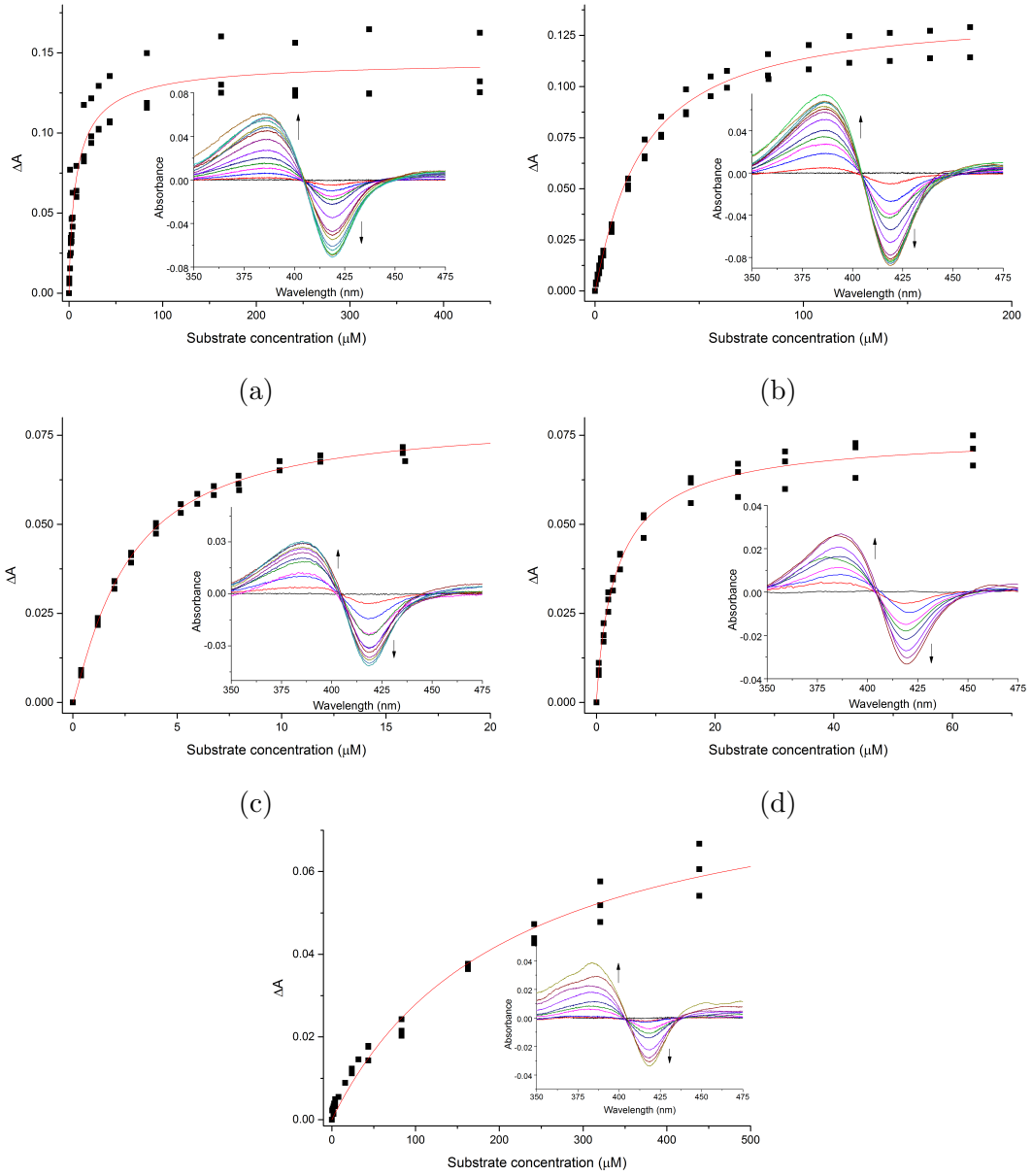


Figure B11: Dissociation analyses of FraEu1415 with 4-androstene-3,17-dione (a), stanolone (b), estrone (c), pregnenolone (d) and nootkatone (e).

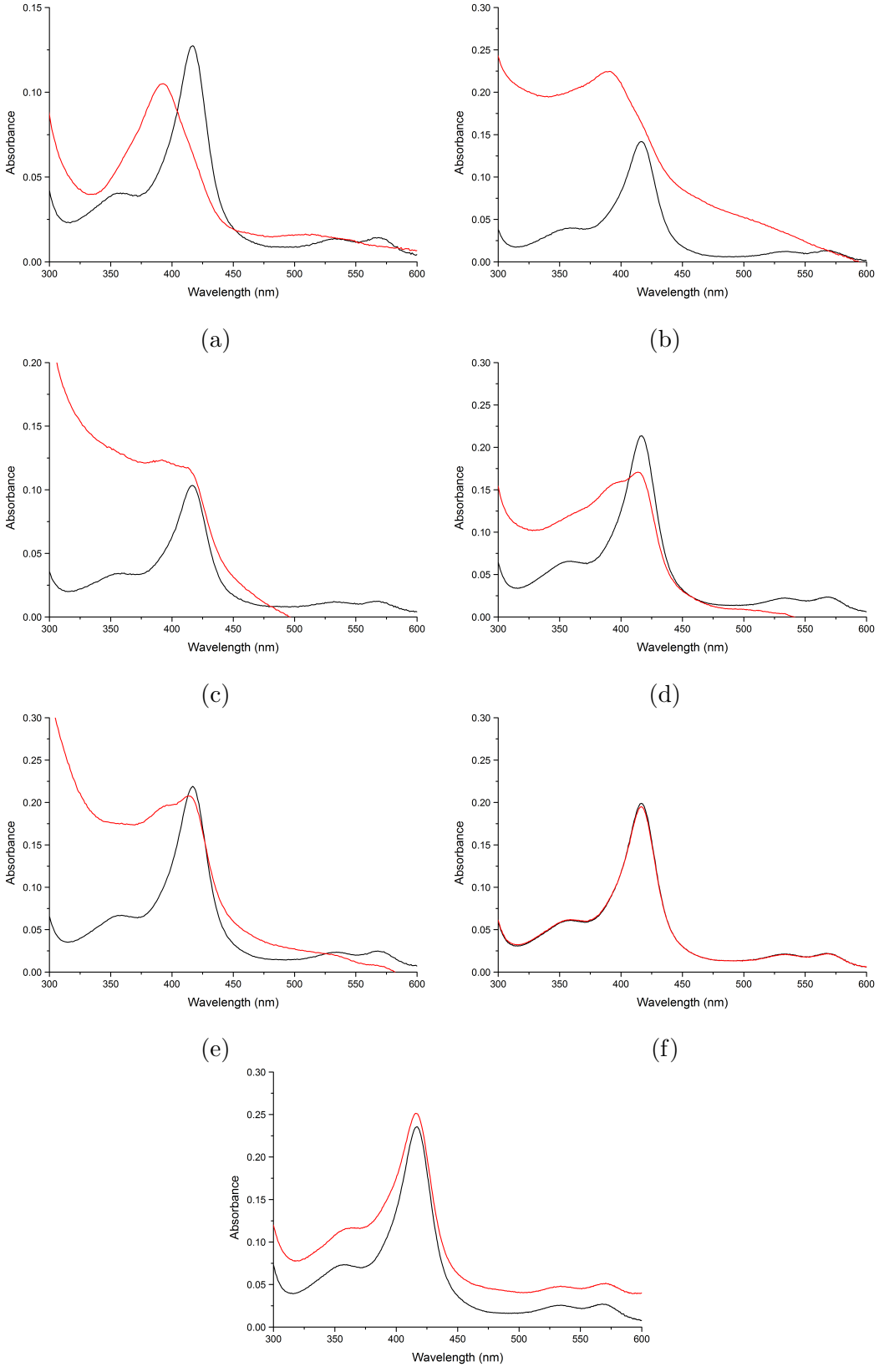
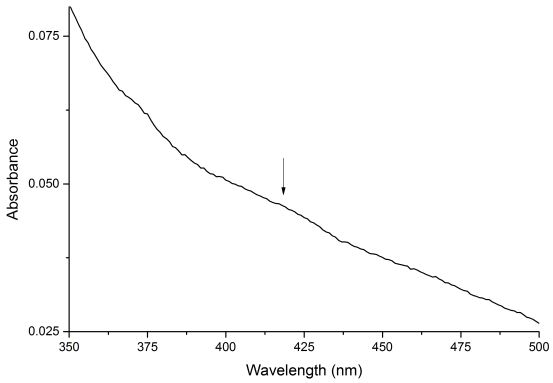


Figure B12: The spin-state shifts of FraEu1415 with 4-androstene-3,17-dione (a), stanolone (b), estrone (c), pregnenolone (d) nootkatone (e), cholesterol (f) and dehydroepiandrosterone (h).



(a)

Figure B13: UV/Vis spectra of ferredoxins Fdx1415. The characteristic peak absorbance at ≈ 416 nm is labelled.

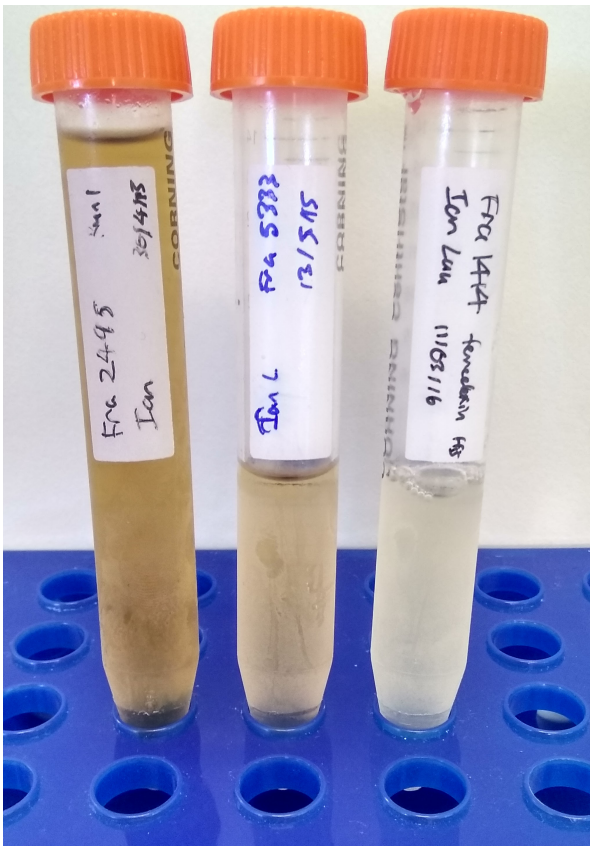


Figure B14: The purified ferredoxins from *Frankia* sp. EuI1c. (From left to right) Fdx2495, Fdx5333 and Fdx1414

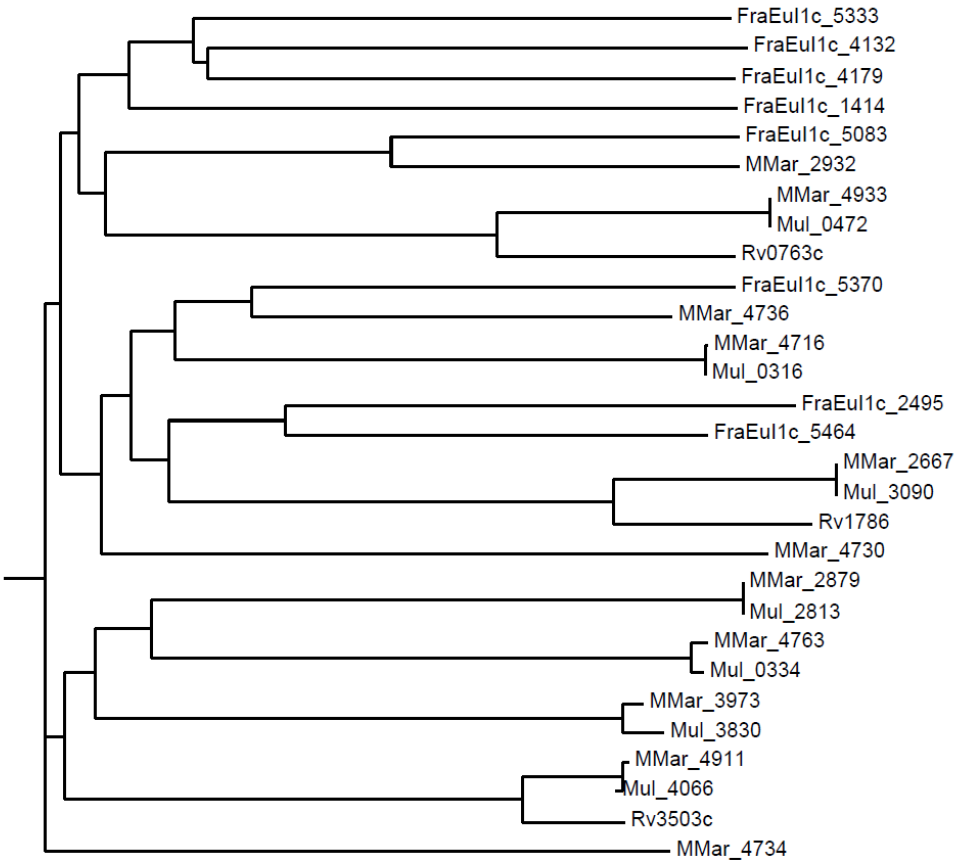


Figure B15: The phylogenetic tree of P450s from *Frankia* sp. EuI1c and related ferredoxins from *Mycobacterium marinum*, *Mycobacterium ulceran* and *Mycobacterium tuberculosis*.

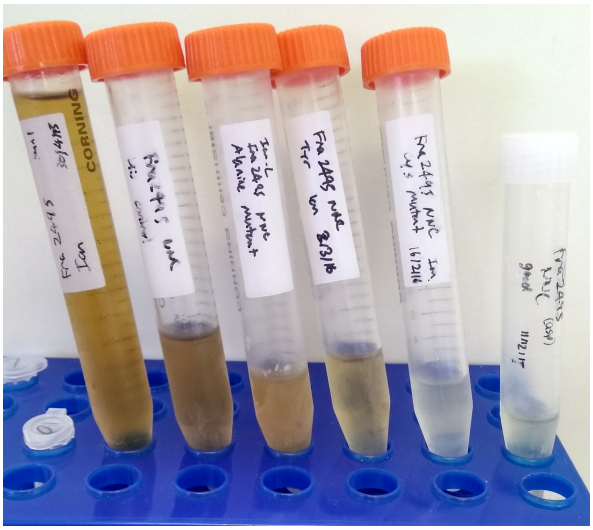
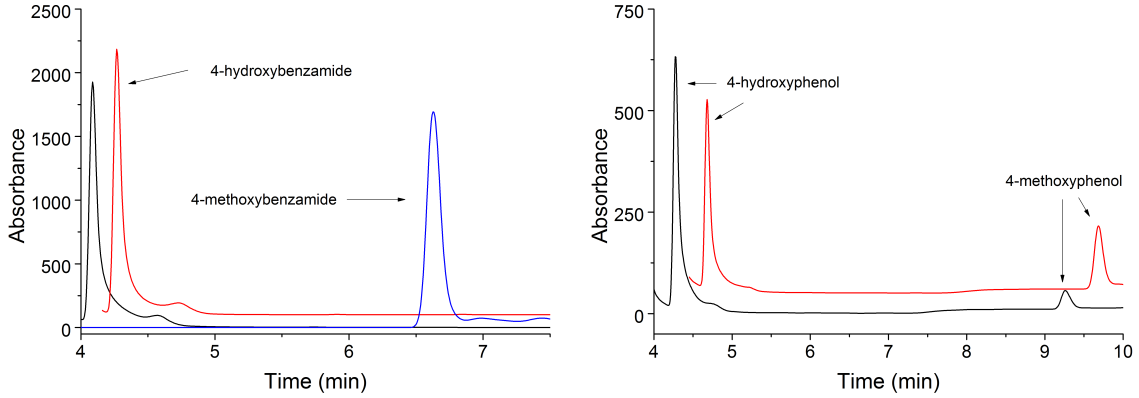


Figure B16: The purified ferredoxin Fdx2495 and its mutant library. (From left to right) wild-type, His-tagged wild-type, alanine mutant, tyrosine mutant, cysteine mutant and asparagine mutant

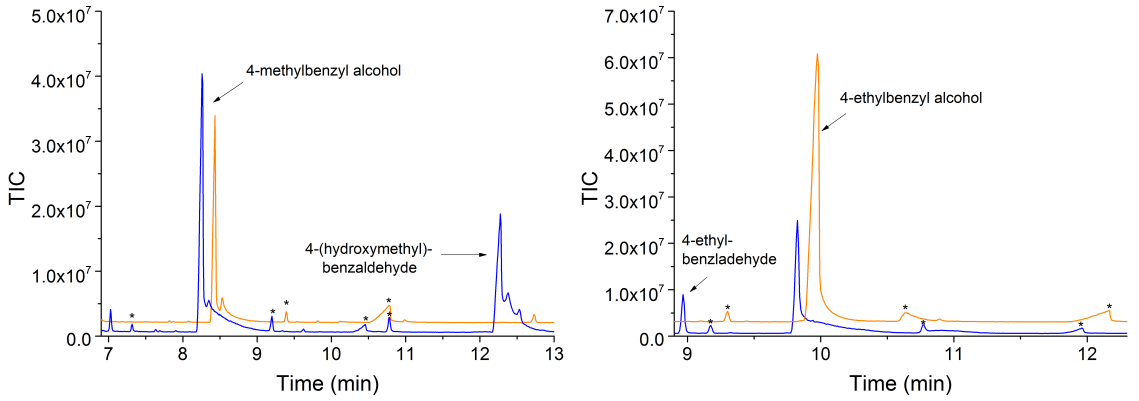
Appendix C Data for chapter 4

HPLC and GC analysis



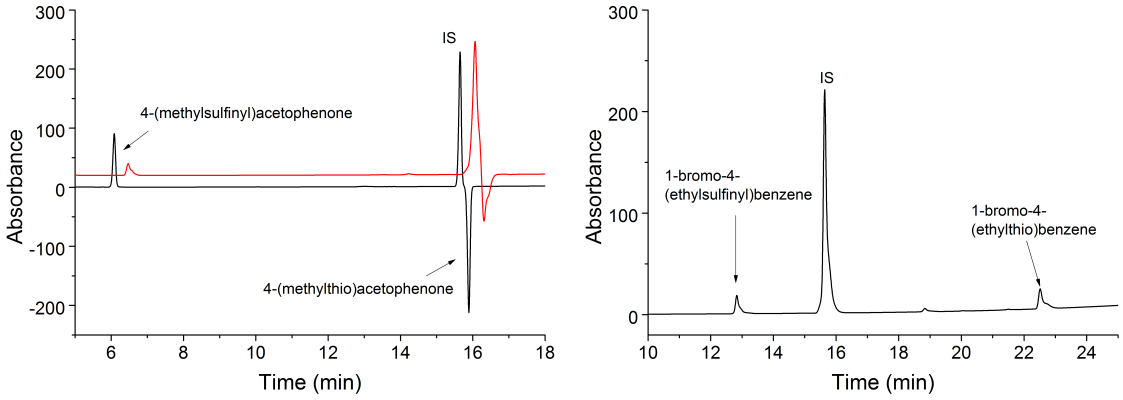
(a) (b)

Figure C1: HPLC analysis of the total turnovers by S244D without ADH (black) and with ADH (red) of 4-methoxybenzamide to 4-hydroxybenzamide (t_R = 4.1 mins) (a), 4-methoxyphenol (t_R = 9.2 mins) to 4-hydroxyphenol (t_R = 4.4 mins) (b). Impurities have been marked as (*). The substrate control for 4-methoxybenzamide in (a) are shown in blue. For clarity the chromatograms have been offset along the x and y axes.



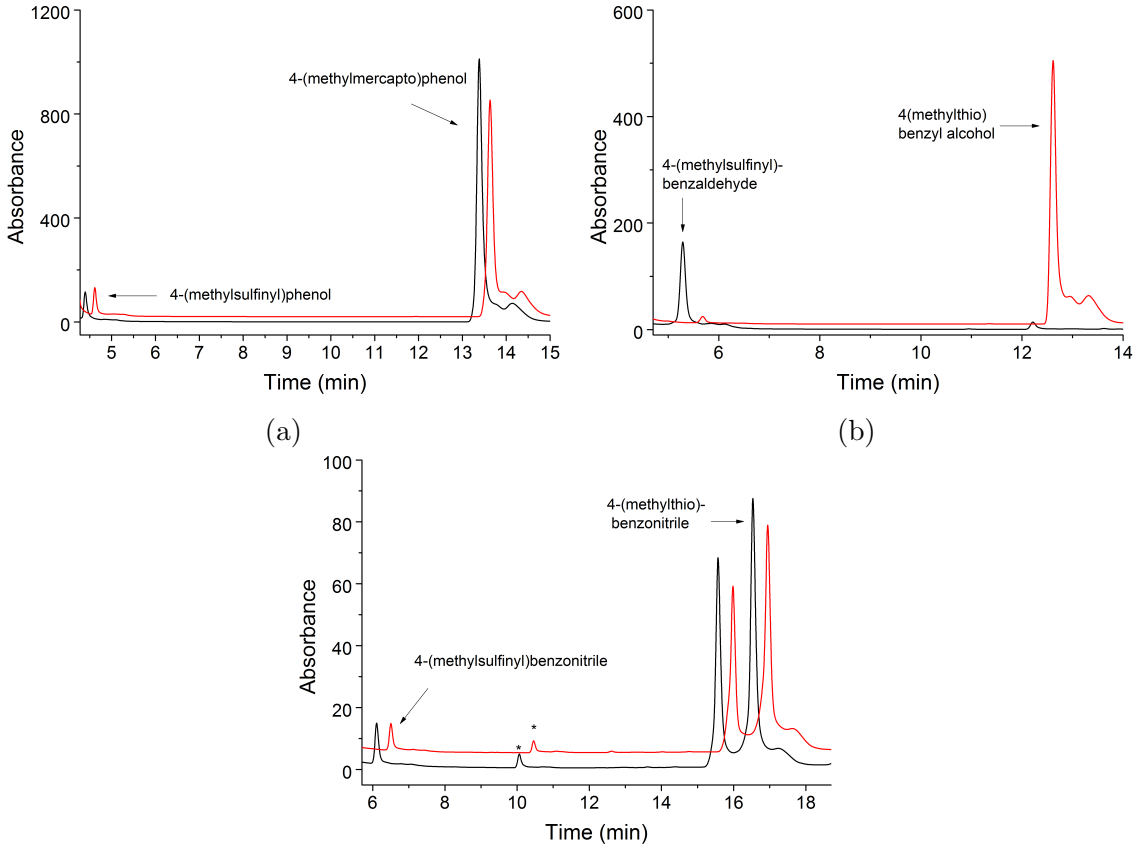
(a) (b)

Figure C2: GC-MS analysis of the *in vivo* turnovers by S244D after 4 hours (blue) and after 18 hours (orange) 4-methylbenzaldehyde (t_R = 7.4 min) to 4-(hydroxymethyl)benzaldehyde (t_R = 12.3 mins) and 4-(methylbenzyl) alcohol (t_R = 8.3 mins) (a), and 4-ethylbenzaldehyde (t_R = 9 mins) to 4-ethylbenzyl alcohol (t_R = 9.8 mins) (b). The benzaldehyde to alcohol conversion reactions are likely to be catalysed by other components of the whole cell environment and not by the P450. Impurities have been marked as (*). For clarity the chromatograms have been offset along the x and y axes.



(a) (b)

Figure C3: HPLC analysis of the *in vitro* turnovers by S244D (black) and WT (red) of 4-(methylthio)acetophenone (t_R = 16 min) to 4-(methylsulfinyl)acetophenone (t_R = 6.1 min) (a), and 1-bromo-4-(ethylthio)benzene (t_R = 23 min) to 1-bromo-4-(ethylsulfinyl)benzene (t_R = 12.75 min) (b). The internal standard (t_R = 15.7 min) is also shown. For clarity the chromatograms have been offset along the x and y axes.



(a) (b)

(c)

Figure C4: HPLC analysis of the total turnovers by S244D without ADH (black) and with ADH (red) of 4-(methylmercapto)phenol (t_R = 13.3 min) to 4-(methylsulfinyl)phenol (t_R = 4.4 min) (a), 4-(methylthio)benzaldehyde (t_R = 16 min, not shown) to 4-(methylthio)benzyl alcohol (t_R = 12.7 min) and 4-(methylsulfinyl)benzaldehyde (t_R = 5.4 min) (b), and 4-(methylthio)benzonitrile (t_R = 16.8 min) to 4-(methylsulfinyl)benzonitrile (t_R = 6.1 min) (c). The internal standard (t_R = 15.7 min) is also shown. For clarity the chromatograms have been offset along the x and y axes.

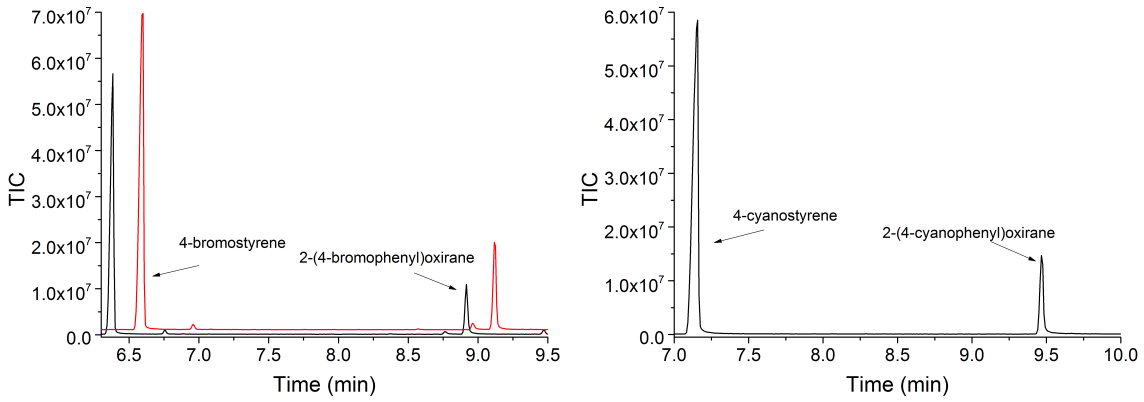
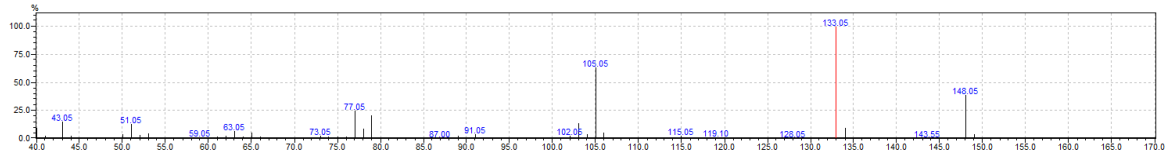
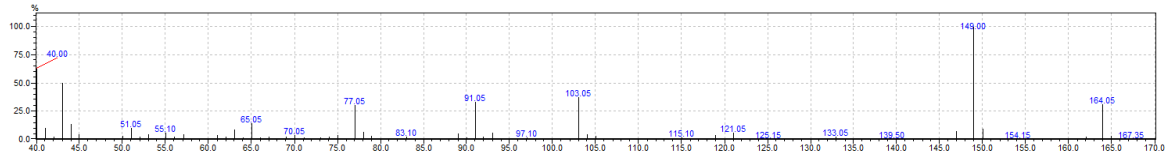


Figure C5: GC-MS analysis of the total turnovers by S244D without ADH (black) and with ADH (red) of 4-bromostyrene ($t_R = 6.4$ min) to 2-(4-bromophenyl)oxirane ($t_R = 8.9$ mins) (a), and 4-cyanostyrene ($t_R = 7.2$ mins) to 2-(4-cyanophenyl)oxirane ($t_R = 9.5$ mins) (b).

Mass spectra analysis

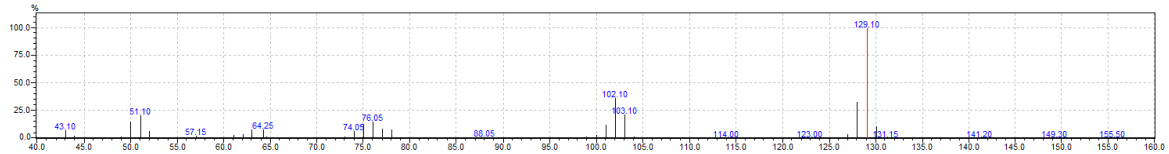


(a)

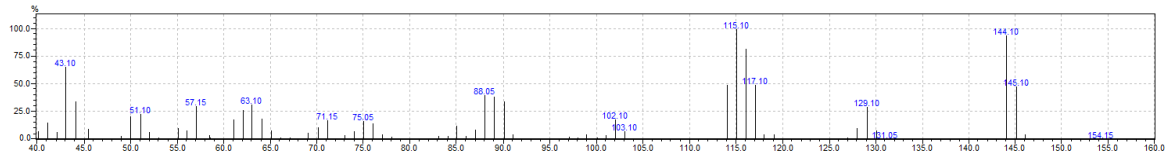


(b)

Figure C6: Mass spectra of (a)2,4-dimethylacetophenone and its oxidation product, presumed to be 4-(hydroxymethyl)-2-methylacetophenone (b).

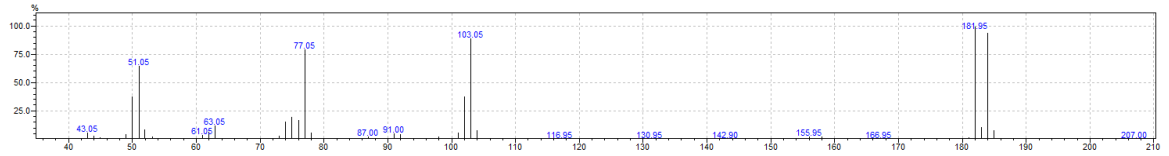


(a)

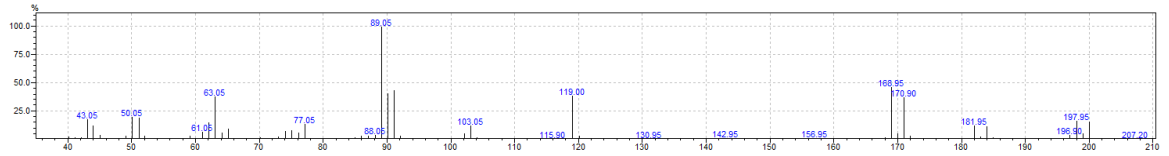


(b)

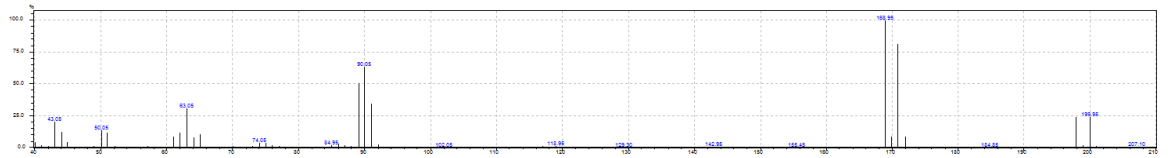
Figure C7: Mass spectra of (a) 4-cyanostyrene and its oxidation product (b) 2-(4-cyanophenyl)oxirane.



(a)

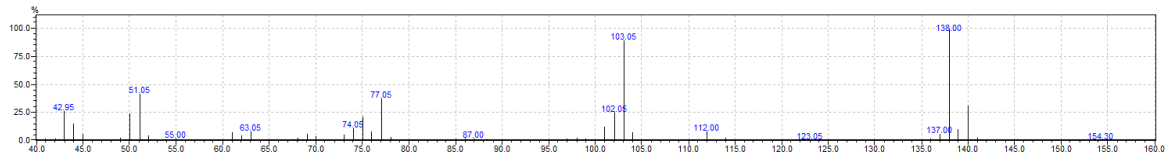


(b)

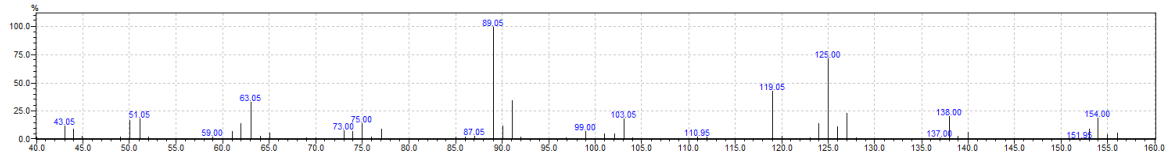


(c)

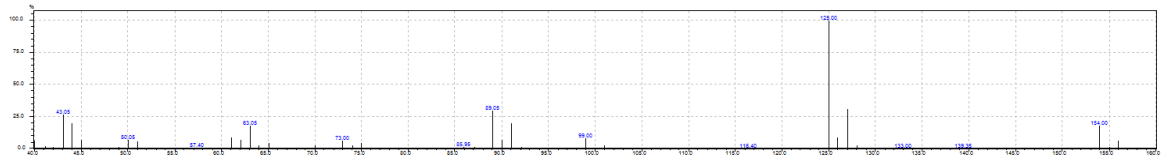
Figure C8: Mass spectra of (a) 4-bromostyrene and its oxidation product (b) 2-(4-Bromophenyl)oxirane and (c) the possible aldehyde rearrangement product of 4-bromostyrene.



(a)



(b)



(c)

Figure C9: Mass spectra of (a) 4-chlorostyrene and its oxidation product (b) 2-(4-chlorophenyl)oxirane and (c) the possible aldehyde rearrangement product of 4-chlorostyrene.