ACCEPTED VERSION

Ian C.K. Lau, René Feyereisen, David R. Nelson, Stephen G. Bell
Analysis and preliminary characterisation of the cytochrome P450 monooxygenases from Frankia sp. EuI1c (Frankia inefficax sp.)
Archives of Biochemistry and Biophysics, 2019; 669:11-21

Final publication at http://dx.doi.org/10.1016/j.abb.2019.05.007

Analysis and Preliminary Characterisation of the Cytochrome P450 Monooxygenases from *Frankia* sp. EuI1c (*Frankia inefficax* sp.)

Ian C.-K. Lau,[a], René Feyereisen,[b], David R. Nelson,[c] and Stephen G. Bell*[a]

[a] Department of Chemistry, University Adelaide, Adelaide, SA, 5005, Australia
E-mail: stephen.bell@adelaide.edu.au
[b] Department of Plant and Environmental Sciences, University of Copenhagen, Copenhagen, Denmark.
[c] University of Tennessee Health Science Center, Dept. of Microbiology, Immunology and Biochemistry, 858 Madison Ave. Suite G01, Memphis, TN 38163, USA.

Highlights

Frankia sp. EuI1c (*Frankia inefficax* sp.), contains 68 CYP encoding genes.

The CYP189 family (21 members) has undergone gene duplication events classified as a "bloom".

Genes for eight [3Fe-4S] and one [2Fe-2S] cluster containing ferredoxins were identified

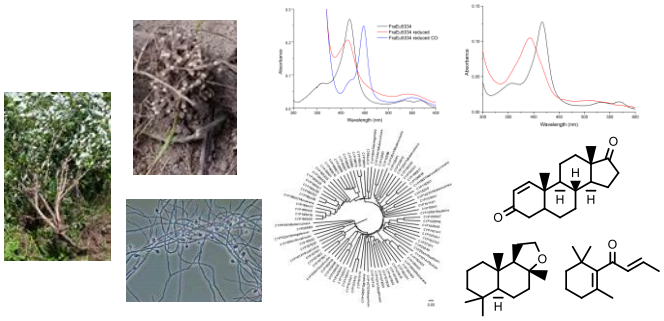
Three CYP family from *Frankia* sp. EuI1c, were produced and purified using *Escherichia coli*

Substrates including steroids and terpenoids which bind to all three CYPs were identified.

Abstract
Frankia bacteria are nitrogen fixing species from the *Actinobacterium* phylum which live on the root nodules of plants. They have been hypothesised to have significant potential for natural product biosynthesis. The cytochrome P450 monooxygenase complement of *Frankia* sp. EuI1c (*Frankia inefficax* sp.), which comprises 68 members, was analysed. Several members belonged to previously uncharacterised bacterial P450 families. There was an unusually high number of CYP189 family members (21) suggesting that this family has undergone gene duplication events which are classified as "blooms". The likely electron transfer partners for the P450 enzymes were also identified and analysed. These consisted of predominantly [3Fe-4S] cluster containing ferredoxins (eight), a single [2Fe-2S] ferredoxin and a couple of ferredoxin reductases. Three of these CYP family members were produced and purified, using *Escherichia coli* as a host, and their substrate range was characterised. CYP1027H1 and CYP150A20 bound a broad range of norisoprenoids and terpenoids. CYP1074A2 was highly specific for certain steroids including testosterone, progesterone, stanolone and 4-androstene-3,17-dione. It is likely that steroids are the physiological substrates of CYP1074A2. These results also give an indication that terpenoids are the likely substrates of CYP1027H1 and CYP150A2. The large number of P450s belonging to distinct families as well as the associated electron transfer partners found in different *Frankia* strains highlights the importance of this family of enzymes has in the secondary metabolism of these bacteria.

Keywords: Cytochrome P450 Monooxygenases; Frankia; ferredoxins; terpenoids; bacterial metabolism

Graphical Abstract



1. Introduction

The *Frankia* genus is comprised of nitrogen fixing bacteria known to live on the root nodules of actinorhizal plants.[1] These are a group of flowering plants (angiosperms) characterised by their symbiotic relationship with the *Frankia* bacteria. This relationship allows the plants to live in harsh environments and diverse ecological conditions including volcanic ash fields, glacial till, sand dunes, and both water-saturated and dry soils.[2] The first named species in the genus was *Frankia alni* which lives in symbiosis with the Alder genus of flowering plants. Other strains of the bacteria are specific to different plant species. Phylogenetic analysis has revealed four distinct clusters among *Frankia* strains.[3] Cluster 1 is a broad group that contains strains such as CcI3 (*F. casuarinae*), ACN14a (*F. alni*), and QA3, and are also known to infect the *Alnus* (Alder) genus of flowering plants. Cluster 2 and 3 are known to be microsymbiots of a multitude of other flowering plants and include the species BMG5.1 (*F. coriariae*), BMG5.12 (*F. elaeagni*) and EAN1pec.[4-8] The *Frankia* sp. EuI1c strain is classified in Cluster 4 phylogeny, which are atypical *Frankia* strains.[9, 10] These strains were isolated from actinorhizal plant nodules, but were unable to re-infect the host plant, or were ineffective strains with other examples being the strains CN3 (*F. saprophytica*), M16386 (*F. asymbiotica*) and G2 (*F. irregularis*).[11] *Frankia* sp. EuI1c, isolated from the root nodules of *Elaeagnus umbellate* (Japanese silverberry) is abnormal within the Cluster 4 *Frankia* strains, as it is capable of reinfecting its host-plant after isolation. However, after re-infection the EuI1c strain forms ineffectual root nodules.[12] Recent studies have attributed its inability to form root nodules to the lack of nitrogen fixation genes within its genome.[13] Recent chemotaxonomic, phenotypic, phylogenetic and genomic studies on the strain, have led to the proposal of the name "*Frankia inefficax* sp.".[10]

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 found to contain gene clusters related to secondary metabolite and natural product metabolism, xenobiotic degradation and heavy metal resistance.[11, 14-18] 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 S1).[14] This makes the metabolic and biosynthetic pathways of *Frankia* sp. strains attractive targets for understanding the synthesis of novel natural products.

The genome of *Frankia* sp. EuI1c strain has been sequenced, revealing a large genome size of 8.8 million base pairs (Mb).[9] Typical of *Frankia* bacteria, the EuI1c strain contains a high percentage of Guanine-Cytosine (GC) DNA base-pair content.[13] The entire EuI1c strain contains ≥ 7,000 protein encoding genes, including 68 cytochrome P450 monooxygenase (P450) genes, which is an unusually high number making up close to 1% of the total gene count. These P450 enzymes carry out oxidative transformations, including hydroxylations, as well as other more complex reactions.[19-22] They are heme monooxygenases and utilise dioxygen which is activated by the addition of two electrons from NAD(P)H. In bacteria these reducing equivalents are most often supplied by ferredoxin reductase and ferredoxin proteins.[23, 24] P450 enzymes are capable of catalysing their reactions with high selectivity and under ambient conditions.[21] The P450s within *Frankia* sp. EuI1c are hypothesised to play roles in facilitating the biosynthesis and/or metabolism of the various secondary metabolites and natural products the *Frankia* species are known to generate or encounter.[1, 14] The combination of these factors result in P450s, such as those found in *Frankia* bacteria, being attractive enzymes for biocatalytic applications.

Here we present the phylogenetic analysis of the P450 enzyme genes from the bacterium *Frankia* sp. EuI1c along with any associated ferredoxin and ferredoxin reductase genes. We also report the production and purification of selected P450 enzymes including the identification of their likely substrates through enzyme binding assays. These results presented here provide the first insight into the potential functions of the P450s from *Frankia* sp. EuI1c and other related species of this genus.

2 Experimental

2.1 General

Reagents, solvents and organic substrates were purchased from Sigma-Aldrich, VWR, Fluorochem, Acros Organics and Tokyo Chemical Industry. 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 hot start polymerase, used for the polymerase chain reaction (PCR) steps, and the pET26a vector were from Merck Millipore. The genomic DNA of *Frankia* sp. EuI1c was generously provided by Prof. Louis Tisa, University of New Hampshire. General DNA and microbiological experiments were carried out by standard methods.

UV/Vis spectra and spectroscopic activity assays were recorded at 30 ± 0.5 ^{o}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.

2.2 Bioinformatics and gene cloning

CYP family and subfamily assignments were made by Dr. David Nelson according to the P450 nomenclature[25] and are used as given in the National Centre for Biotechnology Information (NCBI) database.

2.2.1 Rare Codon Analysis

The %MinMax algorithm was used to analyse rare codon clusters (RCC) in P450 genes. 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 RCC analyses.[26]

2.2.2 Gene cloning

The CYP genes chosen for study were amplified by PCR using oligonucleotide primers (Table S1). They were amplified by 30 cycles of strand separation at 95 ºC for 45 s followed by annealing at 55 ºC for 30 s and extension at 68 ºC for 80 s. incorporated into the pET26 vector (Merck-Millipore) between NdeI and HindIII (or XhoI see Table S1) restriction sites. Successful incorporation of the genes and mutants into each vector was confirmed by restriction enzyme digest followed by DNA sequencing (AGRF, Adelaide node) using the primers; T7 promoter, T7 terminator appropriate for the pET26 parent vector (Merck-Millipore).

2.3 Protein Expression and Purification

Growth media 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)

2.3.1 Recombinant P450 Production

Plasmids (pET26a) containing the P450 genes were transformed in BL21(DE3) competent *E. coli* cells (New England Biolabs). These were held in ice for 45 min, heat shocked (42 °C for 1 min) and subsequently grown in SOC media (1 hour, 37 °C, 200 rpm) before incubation on a LB plate containing 30 μg L^{-1} kanamycin (37 °C, 16 hours).

A small scale growth was first performed for expression testing. For these a single colony inoculated in 50 mL 2x YT containing 30 μg mL^{-1} kanamycin and 150 μL of trace elements. The cultures were grown at 37 ºC for ~ 4 hours (150 rpm). 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 18 hours at 100 rpm. The cells were harvested from the culture by centrifugation at 5,000 *g* for 20 min. An indication of the levels of folded P450 was assessed by visual inspection of the cell pellet, with red colouration indicating P450 formation. In order to determine whether the red colour observed was from 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 min. Cell debris was removed by centrifugation at 4,000 rpm for 10 min. 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).

For larger scaler scale growths 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 ºC 14 hours at 100 rpm. The incubator temperature was lowered to 18 ºC 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 hours. The cells were harvested from the culture by centrifugation at 5,000 *g* for 10 min 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 min. 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, pH 7.4, 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, pH 7.4. 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 min 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 ºC. Glycerol was removed immediately before use by gel filtration on a 5 mL PD-10 column (GE Healthcare) by eluting with 50 mM Tris, pH 7.4.

To quantify the purified P450 enzymes, a carbon monoxide difference spectrum was measured. The enzyme was diluted in 50 mM Tris, pH 7.4, in 0.5 mL. Sodium dithionite (< 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 $\varepsilon_{446\text{-}490}$ = 91 $mM^{-1}cm^{-1}$ for the CO spectrum minus the reduced ferrous spectrum.[27] Using this value, the extinction coefficients for the ferric resting state of the CYP enzymes were calculated to be as follows; CYP1027H1 ε_{418} = 130 $mM^{-1}cm^{-1}$ and CYP150A20 ε418 = 151 $mM^{-1}cm^{-1}$.

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) (≈ 418 nm) to high-spin (HS) (≈ 390 nm) shift did not proceed further. The weighted averages of the spectra of camphor-free and camphor-bound P450cam (≈ 0% to ≈ 100% HS) were used to estimate the % HS of the P450 enzymes.[28]

2.3.2 Dissociation 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. The cumulative volume of each solution added was no more than 10 µL. The dissociation constants, Kd, were obtained by fitting the peak-to-trough difference against substrate concentration to a hyperbolic function (Equation 2)

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

Where ΔA is the peak-to-trough absorbance difference, ΔAmax is the maximum absorbance difference and [S] is the substrate concentration.

3 Results and Discussion

3.1 Analysis of the P450 content of *Frankia* sp. EuI1c

Frankia sp. EuI1c has 68 different P450 encoding genes within its genome (Table S2). These were named using the standard P450 nomenclature (Table S2).[25, 29] A phylogenetic tree of P450s from *Frankia* sp. EuI1c was created along with related P450s from a range of bacteria (Figure 1). While many of these from *Frankia* sp. EuI1c were primarily related to P450s from other *Frankia* species, a few are related to known species from other organisms. For example, CYP147F19 was related to CYP147G1 (45%, sequence identity) and CYP150A20 was related to CYP150A6 (55%), both of which are from *Mycobacterium marinum* M.[30-32]

Of the 68 P450 superfamily members 21 belonged to the CYP189 family (11 members of A, 6 members of B and 2 members of the C subfamilies and one each from those of E and F). This type of gene duplication and lineage-specific expansions in *CYP* subfamilies have been named 'blooms'.[33] While blooms have been reported in eukaryotes and in fungi they have yet to be extensively observed in prokaryotes. Analysis of other members of the *Frankia* genus revealed a contrasting picture. Frankia species such as ACN14a (*F. alni*), ARgP5 (*F. canadensis*) and CcI3 (*F. casuarinae*) (all cluster 1) contain no genes encoding CYP189 family members (they have between 14 and 51 P450 encoding genes). However, the *Frankia* sp. EAN1pec (cluster 3, 15 out of 46), *F. irregularis* (cluster 4, 18 out of 64), sp. EUN1f (cluster 1, 15 out of 69) and *F. asymbiotica* (cluster 4, 17 out of 76) all show a similar "bloom" of CYP189 family members. *F. elaeagni* (strain BMG5.12, cluster 3) contained eight CYP189 family members (out of 31 P450 genes, See Table S3 for more details). The data highlights that the number of P450 encoding genes is not consistent across the *Frankia* species (Table S3). This infers that different species have evolved to adapt to their symbiotic environments. It is commonly held that eukaryotes have a much smaller effective population size (N*e*) than prokaryotes, a key determinant of the longer half-life of duplicated genes which favours the generation of blooms.[34] Duplicated genes, when retained, have a greater probability of preservation by neofunctionalization in prokaryotes, while in eukaryotes the more common duplicates are more likely to be maintained in the genome by subfunctionalization. The CYP189 blooms onserved in some Frankia species may therefore constitute a set of P450s with new functions. The symbiotic nature of different *Frankia* species may also be a factor in determining the relative CYPomes sizes, in that N*e* for *Frankia* species may be significantly smaller than for free-living actinobacteria, thus relaxing selection and allowing the generation of blooms. At the same time, the variety of CYP genes may be beneficial to the peculiar ecology of each species of *Frankia*.

The role of the CYP189 family of enzymes is currently unknown hampering an understanding into why it might be a favourable candidate for gene duplication. CYP189A1 from *Mycobacterium smegmatis* was hypothesised to be involved in cholesterol metabolism as its gene was upregulated in response to this steroid.[35] However, after production, purification and biochemical analysis no substrate binding response or activity was observed for this and several other substrates.[36]

The EuI1c genome has 6 genes encoding members of the CYP107 family and 4 of the CYP150A subfamily (Table S2). There were also many P450 family members which are present in other actinobacteria such as *Mycobacteria*, *Rhodococcus* and *Streptomyces* including CYP108 (x2), CYP124 (x2), CYP125 (x3), CYP142 (x1), CYP147 (x1), CYP155 (x1), CYP157 (x4) and CYP190 (x1).[37] There were nine families where the P450 enzymes from this bacterium are the first named member (CYP1242A1, CYP1162A1, CYP1074A1, CYP1316A1, CYP1323A1, CYP1073A1, CYP1163A1 and CYP1161A1, Table S2).

The EXXR sequence motif is conserved in all the P450s in *Frankia* sp. EuI1c with the exception of each of the four CYP157 family members where the arginine was replaced by a tryptophan residue (Table S2).[38] The P450 enzymes encoded by these four genes also lacked the acid alcohol pair which facilitates oxygen activation in the P450s.[20] In these CYPs the acid group was replaced with a glutamine or leucine and the alcohol with a proline or alanine (Table S2). The acid residue was replaced in eight of the other CYPs including CYP1073A1 (Ser), CYP254A9 (His), CYP107BG3 (Val), CYP109K1 (Leu), CYP155D2, CYP150A23, CYP1034A3 and CYP1161A1 (all Gly). Based on this sequence data several of the P450s from *Frankia* sp. EuI1c would appear to have evolved to have different substrates or functions to those previously characterised.

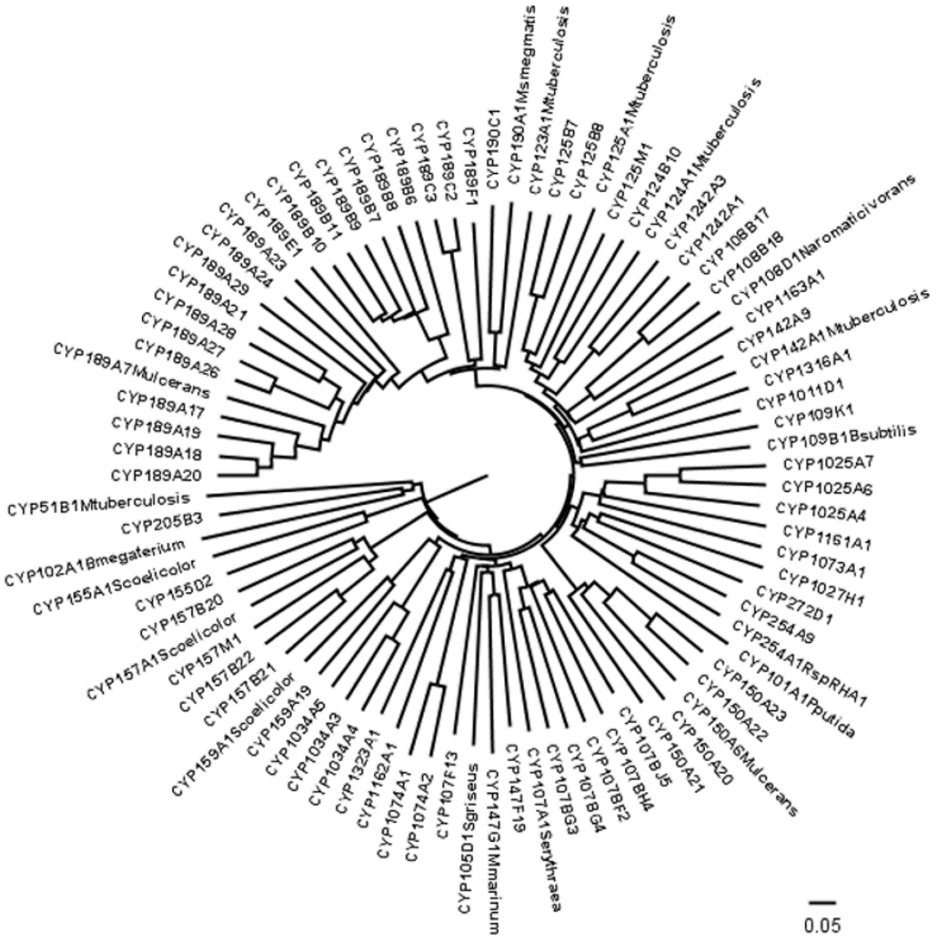


Figure 1 A phylogenetic tree of the 68 P450s from *Frankia* sp. EuI1c and related bacterial P450s. These include CYP189A7 (*Mycobacterium ulcerans*), CYP190A1 (*Mycobacterium smegmatis*), CYP123A1, CYP125A1, CYP125A1, CYP142A1, CYP51B1 (all *Mycobaterium tuberculosis*), CYP109B1 (*Bacillus subtilis*), CYP108D1 (*Novosphingobium aromaticivorans*), CYP101A1 (*Pseudomonas putida*), CYP159A1, CYP157A1 (both *Streptomyces coelicolor*), CYP254A1 (*Rhodococcus* sp. RHA1), CYP102A1 (*Bacillus megaterium*), CYP105D1 (*Streptomyces griseus*), CYP107A1 (*Saccharopolyspora erythraea*), CYP147G1 and CYP150A6 (both *Mycobacterium marinum*). The tree was computed by the CLUSTAL Omega server at EBI (http://www.ebi.ac.uk/Tools/msa/clustalo/).[39] The tree was midpoint rooted and drawn in Figtree v1.4.3.

The genome sequence of the bacterium was analysed in more detail in order to select the P450s to be investigated further. The P450 catalytic cycle requires that electrons and protons to be

transferred to the heme.[20] 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.[24] Identification of the ferredoxins and ferredoxin reductases for the P450s from *Frankia* sp. EuI1c would enable the P450 systems to be analysed in greater detail. Of the 68 P450 genes of this bacterium, nine of them are coupled with a ferredoxin gene. Furthermore only two out of these nine are clustered with a ferredoxin reductase gene (Table 1). This close arrangement of these types of genes infers that they are the likely electron transfer partners to their clustered P450s.[40, 41]

Table 1 The P450 genes of *Frankia* sp. EuI1c which are clustered with ferredoxins and ferredoxin reductase electron transfer partners. The gene size molecular weight, and non-standard residue of ferredoxins or cofactor of associated ferredoxin reductases if known are provided. More details on the nine P450 genes associated with these ferredoxins are also given in Table 2 (other P450s are described in Table S2). [a] *FraEuI1c5082* is a ferredoxin reductase/ferredoxin fusion protein.

Ferredoxin gene	P450 gene	Gene size (bp)	Protein weight (kDa)	cluster type	Residue ?
FraEuI1c_5370	*FraEuI1c_5369*	189	6.75	[3/4Fe-4S]	His
FraEuI1c_5333	*FraEuI1c_5334*	195	6.67	[3/4Fe-4S]	Ser
FraEuI1c_2495	*FraEuI1c_2494*	216	7.32	[3/4Fe-4S]	His
FraEuI1c_1414	*FraEuI1c_1415*	213	7.67	[3/4Fe-4S]	Thr
FraEuI1c_4132	*FraEuI1c_4131*	219	7.09	[3/4Fe-4S]	Ser
FraEuI1c_5082[a]	*FraEuI1c_5083*	1632	57.9	[3/4Fe-4S]	Tyr
FraEuI1c_5464	*FraEuI1c_5466*	189	6.71	[3/4Fe-4S]	His
FraEuI1c_4178	*FraEuI1c_4179*	204	6.81	[3/4Fe-4S]	Ser
FraEuI1c_3227	*FraEuI1c_3226*	319	10.8	[2Fe-2S]	-

Reductase gene	P450 gene	Gene size (bp)	Protein weight (kDa)	Related enzyme	cofactor
FraEuI1c_3228	*FraEuI1c_3226*	1251	43.2	PdR	FAD
FraEuI1c_5082[a]	*FraEuI1c_5083*	1632	57.9	-	-

One of these *FraEu1c_3227* encodes a [2Fe-2S] cluster protein and was also located next to a ferredoxin reductase gene (Table 1). This ferredoxin reductase gene, *FraEuI1c_3228*, is therefore clustered with the [2Fe-2S] ferredoxin (*FraEuI1c_3227*) and CYP1163A1 (*FraEu1c_3227*) forming a complete electron transfer system. Another of the potential electron transfer partner genes *FraEu1c_5082* is a fusion of a ferredoxin reductase and ferredoxin. This gene is next to that of CYP147F19 (*FraEuI1c_5082*). The ferredoxin component of this fused gene is a single [3Fe-4S] cluster system.[30] The remaining ferredoxin genes encode proteins whose sequences indicate that they would encode [3Fe-4S] or [4Fe-4S] cluster containing ferredoxins (Table 1).[30, 42] These types of iron-sulfur ferredoxins generally contain sequence motifs of CXX**?**XXC(X)nCP.[43] The cysteine residues in this motif bind to the iron sulfur cluster and if the variable residue at position ? is a cysteine the cluster type is normally a [4Fe-4S]. If it is an alanine or glycine it would be a [3Fe-4S] iron-sulfur cluster ferredoxin. All the potential [3Fe-4S] ferredoxins in *Frankia* sp. EuI1c have non-standard residues at the "?" position in the motif. Previous studies indicate that these ferredoxins contain [3Fe-4S] clusters but that the identity of the variable residue affects the redox properties of the ferredoxin.[30] The ferredoxin genes clustered next to the P450s *Frankia* sp. EuI1c contain histidine, serine, threonine and tyrosine residues at this variable position (Table 1).

Given that one of the associated ferredoxin reductase genes is fused to its own [3Fe-4S] cluster domain and the other is linked to a [2Fe-2S] containing ferredoxin gene, it seems likely that this bacterium could contain additional ferredoxin reductases that can support the other [3Fe-4S] ferredoxins linked with these P450s. These would facilitate the electron transfer between NAD(P)H and the ferredoxin and would allow detailed turnover assays to be undertaken. It is of note that, like other actinobacteria (for example, *Mycobacterium marinum*) there are significantly more ferredoxins of this [3Fe-4S] type in the genome.[30]

3.2 Production and Purification of selected P450s from *Frankia* sp. EuI1c

Strains of *Frankia* bacteria are known to have a high guanine-cytosine (GC) content. The *Frankia* sp. EuI1c genome contains 72.31% GC content compared to ~50% in *E. coli*.[9] The GC content for the 9 selected P450s ranged from 69-75% (Table 2). Given such a large difference in base content the *Frankia* CYP genes may possess codons whose tRNA are present in low amounts in *E. coli* (rare codons).[44, 45] Analysis of these 9 P450 genes for rare codon clusters (RCC) revealed multiple RCCs which could impair protein production using *E. coli* (Table 2, Figure S2).[46] The rare codon percentage was significant, ranging from 8-16% (Table 2).

Small scale expression tests of the nine P450s in Table 2 were undertaken to observe and identify which of these might be able to be produced at a level appropriate for further investigation. The level of P450 expression was ranked according to the intensity of the red colour, which is characteristic of the heme-containing enzyme, within the cell pellets: CYP1027H1 > CYP150A20 > CYP1074A2 > CYP1034A4 > CYP147F19 ~ CYP150A21 ~ CYP1161A1 ~ CYP1034A5 ~ CYP1163A1. While the cell pellet containing CYP1027H1 was dark red in colour that of CYP150A20 was pink. The pellets of CYP1074A2 and CYP1034A4 were a lighter shade of pink indicative of a low level of protein. Little to no P450 holoenzyme was produced for CYP147F19 and the other *Frankia* P450s tested here, as determined by the beige coloured cell pellets.

Table 2 Gene size, protein molecular weight and GC content of the P450s from *Frankia* sp. EuI1c. The codons encoding for arginine (AGG, AGA, CGG, CGA), isoleucine (ATA), and proline (CCC) were considered rare codons in *E. coli* (< 0.4% codon usage).[45] [26, 44]

P450 Gene	P450 name	Gene size (bp)	Protein weight (kDa)	GC content (%)	Rare codon (%)
FraEuI1c_2494	CYP1027H1	1248	46.47	72	11
FraEuI1c_5334	CYP150A20	1275	47.30	70	7
FraEuI1c_1415	CYP1074A2	1272	47.23	68	10
FraEuI1c_4131	CYP1034A4	1248	45.87	70	10
FraEuI1c_5369	CYP150A21	1380	49.83	70	8
FraEuI1c_5083	CYP147F19	1272	45.91	75	16
FraEuI1c_5466	CYP1161A1	1215	44.17	72	9
FraEuI1c_4179	CYP1034A5	1260	46.33	72	8
FraEuI1c_3226	CYP1163A1	1263	48.20	69	11

The four P450s that showed any evidence of fully folded enzyme production were grown on a scale viable for protein purification. SDS-PAGE gel analysis showed that the *E. coli* cells

generated a protein of between 44-48 kDa about the size expected for the P450 enzymes above (Table 2, data not shown). Large scale growth of *E. coli* producing CYP1027H1 resulted in a significant quantity of a red coloured soluble protein after cell lysis and centrifugation to remove cell debris. The levels of purified protein obtained with CYP150A20 and CYP1074A2 was lower and correlated with the colour of the cell pellet described above (Figure S3). While the pellet of CYP103A4 was similar in colour to CYP1074A2, little to no soluble P450 protein was isolated (Figure S3).

The purified P450s were analysed using UV/Vis spectroscopy (Table 3, Figure S3). In order to determine the concentration of the P450s, the spectra of the ferrous carbon monoxide (CO) bound form was determined. CYP1027H1 and CYP150A20 each showed an almost complete shift (≥95%) from to the CO-bound ferrous form (Figure S4). A small shoulder was observed at 420 nm for both the CO bound reduced P450s (Figure S4). The ferrous-CO assay for CYP1074A2 produced a different spectrum. After reduction, the P450 spectrum peak shifted to 426 nm (Figure S4). Once CO was bubbled into the mixture, a large peak at 422 nm peak was observed (Table 3). This indicated formation of the inactive species "P420" form which is thought to arise from protonation of the cysteine thiolate, resulting in thiol ligation to the heme iron.[47, 48] This suggested that either all of the CYP1074A2 protein is present in an inactive or denatured form, or the enzyme is sensitive to the method of reduction or addition of CO. The inactive form of P450s can be induced by treatment with acetone or exposing the P450 to a high pressure environment.[49, 50] It was subsequently found that CYP1074A2 could bind steroids with high affinity, inducing a full type I spin state shift (≥95% high-spin, see Section 3.3.3). Additional CO bound spectra for CYP1074A2 were obtained in the presence of the substrate 4-androsterene-3,17-dione (Figure S4). After the addition of CO, a proportion of the spectra displayed the characteristic 450 nm peak (Table 3, Figure S4).

Table 3 The UV/Vis wavelengths of the absorbance peak maxima for the ferric, ferrous and ferrous-CO spectra of CYP1027H1, CYP150A20 and CYP1074A2.

P450	Ferric (nm)	Ferrous (nm)	Ferrous-CO (nm)
CYP1027H1	418	410	447
CYP150A20	418	414	448
CYP1074A2	417	426	422
CYP1074A2+substrate	393	393	449 and 421

Through the CO difference assays, the extinction coefficients of CYP1027H1 and CYP150A20 in their resting states were calculated (CYP1027H1 ε_{418} = 130 $mM^{-1}cm^{-1}$ and CYP150A20 ε_{418} = 151 $mM^{-1}cm^{-1}$.The amount of CYP1027H1 protein expressed per litre of cell culture was approximately 60 mg L^{-1}. That of CYP150A20 being 14.8 mg of protein per litre of cell culture. The level of CYP1074A2 generated was estimated to be ≤ 2 mg L^{-1}.

The level of protein production using *E. coli* as a host for many of the tested P450s from *Frankia* sp. EuI1c was 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. However, the RCC content of the genes did not directly correlate to the levels of production of the enzymes. While the P450 genes from *Frankia* sp. EuI1c showed high GC content they did not appear to have an unsustainable number of rare codons but several had extensive RCCs. In order to improve the poor expression caused by these clusters, codon optimised genes for *E. coli* or an alternative host could be used. However, we cannot rule out that poor heme

incorporation or protein folding is the cause for the low level of formation of P450 holoenzyme. Further optimisation of the production of the majority of the P450s is required in order to enable more detailed analysis of their function.

3.3 Assessment of the potential substrate range of CYP1027H1, CYP150A20 and CYP1074A2

The first step of the P450 catalytic cycle often involves the displacement of the distal water ligand by an appropriate substrate.[20] When this occurs a low spin to high spin (HS) state shift occurs and this change can be monitored by UV/Vis spectroscopy as it results in a blue-shift in absorbance of the main Soret peak from approximately 418 nm to 390 nm. In order to help determine which substrates 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 identities.[51] As many of the *Frankia* P450s assessed here did not share significant sequence identity to functionally characterised enzymes other classes of molecules were also tested. These included cyclic mono- and sesqui-terpenoids, short and medium chain (C6-C12) linear alkanes, benzene and polycyclic aromatic hydrocarbons, fatty acids, substituted phenols, substituted benzoic acids and steroid compounds.

3.3.1 Investigation of Substrate Binding to CYP1027H1

CYP1027H1 is most closely related to other uncharacterised P450s from *Frankia*, sharing up to 53% amino acid sequence identity with these P450 enzymes and others from *Streptomyces*, *Gordonia*, *Mycobacterium*, *Thermomonospora*, *Nocardia* and *Rhodococcus* bacteria (Table 4). It showed low amino acid sequence identity (28-34%) to the terpenoid oxidising enzymes CYP101A1 (P450cam), CYP176A1 (P450cin) and CYP108A1 (P450terp, Table 4).[52-54] CYP1027H1 was also 25% identical to the amino acid sequence of CYP109B1 from *Bacillus subtilis* (Table 4).[55]

A range of compounds including norisoprenoids, monoterpenoids, sesquiterpenoids, steroids and fatty acids were screened with CYP1027H1. While certain mono- and sesqui-terpenoids, induced spin state shifts, no shift was observed for any of the fatty acids and steroids (Figure 2 and Figure S5 and S6). The norisoprenoids, α-ionone, β-ionol and β-damascone all induced commensurate spin state shifts of between 40-50% in CYP1027H1 (Figure 2). They also exhibited similar binding affinities to the enzyme (Kd = 7.6, 12 and 12 μM, respectively, Figure 2 and Figure S7). Unexpectedly, β-ionone did not induce any spin state shift in CYP1027H1, despite having a very similar structure to the other norisoprenoids. Bornyl acetate, fenchyl acetate and 4-*tert*-butylcyclohexyl acetate, which all possess a cyclohexane group with oxygen-containing carbon chain substituents similar to the norisoprenoids, induced a spin state shift in CYP1027H1 (35-40%, Figure S5 and S6). The binding affinities were similar to those of the norisoprenoids (Kd = 8.4, 8.6 and 11 μM, respectively, Figure 2 and Figure S7).

Decalin (*trans*-decahydronapthalene) induced a HS state shift (60%) when added to CYP1027H1 and also had a high binding affinity (Kd = 4.9 μM, Figure S6 and S7). However, the related compound tetralin, which contains an aromatic ring, induced only a 30% spin shift and the binding affinity was significantly lower (Kd = 52 μM, Figure 2, Figure S5). Naphthalene, which is comprised of two interconnected unsaturated benzene rings, induced a negligible spin state shift (Figure S6). This suggests that saturated cyclic rings are a core component in the natural substrates of CYP1027H1.

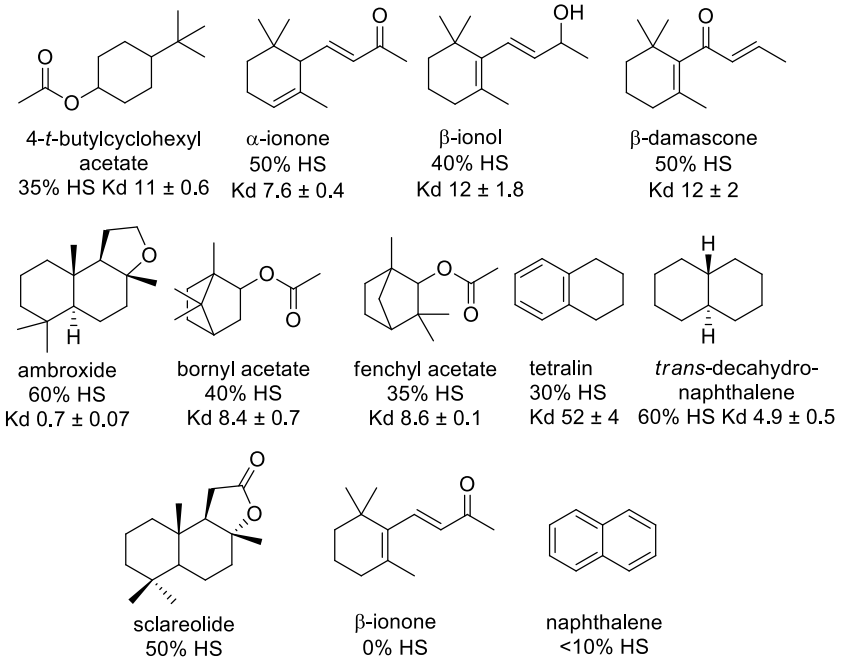
Larger compounds such as ambroxide, which has a structure consisting of two fused cyclohexane rings, similar to *trans*-decahydronaphthalene, but with an additional 5 membered ring carbon ring containing an ether oxygen (Figure 2), induced a 60% spin state shift of the heme iron upon addition to CYP1027H1. This substrate bound with the highest affinity to the enzyme (Kd = 0.7 μM, Figure 2 and Figure S5). A 50% spin state shift in CYP1027H1 was also observed when sclareolide was added. Sclareolide is structurally related to ambroxide but possesses an extra carbonyl group (Figure 2 and Figure S6). The dissociation constant for sclareolide was unable to be determined as larger amounts of substrate were required to saturate the spectrum and this was beyond the aqueous solubility limit. However, it had a lower affinity than ambroxide as judged by the significantly reduced response of the difference spectrum to substrate addition. These saturated cyclic ring containing compounds are promising potential substrates for CYP1027H1.

Table 4 The amino acid sequence identity of CYP1027H1 with other P450s. These were determined using a BLAST sequence search on the NCBI database. Three additional similar P450 enzymes were found in *Frankia* sp. Iso899 (47-49% sequence identity, not shown in the Table) along with others in other *Frankia* species including *Franki*a sp. QA3 (47%). Other strains of *Mycobacteria* and *Nocardia* also contained enzymes which shared 40-45% sequence identity. Note; CYP1027A1 and CYP1027B1 are from *Frankia alni* ACN14a (sequence identity 41 and 47%, respectively). The P450 name is given, where it has been assigned; if not the Genbank protein ID or NCBI accession number is provided. [a] Max Score 835.

P450	Source organism	Total score[a]	Sequence identity	Known substrates	PDB
WP_022912145.1	*Frankia* sp. Iso899	429	53%	-	-
WP_022914489.1	*Frankia* sp. Iso899	393	49%	-	-
KOV68049.1	*Streptomyces* sp. NRRL WC-3618	384	49%	-	-
GAC70340.1	*Gordonia soli* NBRC 108243	339	46%	-	-
CQD18807.1	*Mycobacterium europaeum*	335	45%	-	-
ACY98928.1	*Thermomonospora curvata*	323	47%	-	-
PPJ16295.1	*Nocardia nova*	323	45%	-	-
SMG27105.1	*Rhodococcus rhodochrous* J3	321	44%	-	-
CYP254A1	*Rhodococcus* sp. RhA1	197	35%	-	-
CYP101D1[41]	*Novosphingobium aromaticivorans*	173	34%	camphor	3LXH
CYP127A1	*Sinorhizobium sp. NGR234*	178	31%	-	-
CYP143A1	*Mycobacterium tuberculosis*	198	35%	-	-
CYP101A1[56]	*Pseudomonas putida*	164	29%	camphor	2CPD
CYP176A1[52]	*Citrobacter braakii*	157	29%	1,8-cineole	1T2B
CYP108A1[53]	*Pseudomonas sp.*	156	28%	α-terpineol	1CPT
CYP109B1[55]	*Bacillus subtilis*	134	25%	fatty acids	4RM4

Overall CYP1027H1 had a broad substrate range. The majority of these substrates contain oxygen containing substituents. These include ketone, ester, lactone, ether and alcohol groups, suggesting that the oxygen atom may interact with hydrophilic active site residues of CYP1027H1. These substituents, are hydrogen-bond acceptors/donors and are likely to interact with amino acids that contain hydrogen-bond donating side chains. However, these functionalities appear to be less important for CYP1027H1 than in other P450 systems for example P450cam (CYP101A1) and P450cin (CYP176A1).[57, 58] For example, *trans*-decahydronaphthalene, which is a hydrocarbon, exhibited similar binding parameters to the norisoprenoids.

Figure 2 Selected compounds assessed with CYP1027H1. Percentage high spin (HS) is denoted as % HS. The dissociation (Kd) is given as the mean ± S.D. with n ≥ 3 and given in μM. The dissociation constant of sclareolide and naphthalene were not determined due to insolubility of substrate preventing full analysis. β-Ionone did not induce an observable spin state shift and is shown for reference.



3.3.2 Investigation of Substrate Binding to CYP150A20

CYP150A20 has a high sequence similarity (83-88%) to other uncharacterised *Frankia* P450s (Table 5). It is also has a 55-64% amino acid sequence identity with CYP150 family members from *Mycobacterium* and *Blastococcus* species. For example, CYP150A6 and CYP150A5, both from *Mycobacterium marinum* species share 50% and 55% amino acid sequence identity, respectively with CYP150A20 (Table 5).[32] The three other CYP150A subfamily members of *Franki*a sp. EuI1c shared 50-51% (A21, A22 and A23) amino acid sequence identity to CYP150A20 (Table 5). The closest other structurally characterised P450s are CYP108A1 (P450terp) and CYP109B1, which have a low amino acid sequence identity at 32% (Table 5).

Members of the CYP150 family has been reported to oxidise polycyclic aromatic hydrocarbons and CYP150A5 from *M. marinum* has been demonstrated to bind a range of compounds including norisoprenoids and terpenoid based substrates including the aromatic sesquiterpene guaiazulene (1,4-dimethyl-7-isopropylazulene).[32, 59] Therefore CYP150A20 was screened with a range of aromatic hydrocarbons, fatty acids, monoterpenoids and steroids (Figure 3). Steroids, such as testosterone, and fatty acids did not induce any spin state shift. CYP150A20 shares 27% amino acid sequence identity with CYP1027H1 and some of the compounds screened with this P450 displayed an affinity towards CYP150A20 (Figure 2 and 3 and Figure S8-S10).

A number of norisoprenoids were able to induce spin state shifts in CYP150A20. β-Ionone, β-damascone and β-ionol, all induced shifts to 40% HS, and displayed moderate binding affinity (Kd = 107, 74 and 107 μM, respectively; Figure 3 and Figure S8 and S9) These three compounds are very similar in structure and only differ in the position of their functional groups, such as the ketone and alkene in the butenone moiety. Although these norisoprenoids

induced the highest spin state shifts compared to the all other substrates tested here, the binding affinities were relatively low. This suggested that while they bind they are not an optimal fit for the substrate binding pocket.

Table 5 The amino acid sequence identity of CYP150A20 with other P450 enzymes. Similar enzymes were found across species of *Frankia* (55-88% sequence identity), *Mycobacterium* (55-56% sequence identity), *Blastococcus* (55-64% sequence identity) and *Streptomyces* (55-56% sequence identity). CYP150A1 was originally found in a *Mycobacterium* species.[60] Several *Frankia* species had P450s with a high similarity to CYP150A20 (sp. EUN1h, *F. asymbiotica*, *F. saprophytica* and sp. BMG5.36) which are all members of cluster 4 atypical *Frankia* (not shown). The P450 name is given when it has been assigned if not the Genbank protein ID or NCBI accession number is provided. [a] Max Score 835.

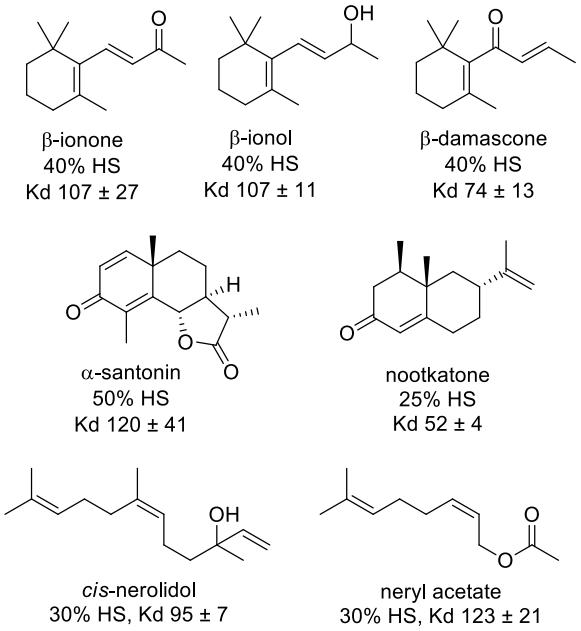
P450	Source organism	Total score[a]	Sequence identity	Known substrates	PDB
OHV39124.1	*Frankia* sp. EUN1h	739	83%	-	-
CYP150A23	*Frankia* sp. EuI1c	435	51%	-	-
CYP150A22	*Frankia* sp. EuI1c	429	50%	-	-
CYP150A21	*Frankia* sp. EuI1c	425	50%	-	-
CYP150A5[32]	*Mycobacterium marinum*	469	55%	sclareol	-
CYP150A6[32]	*Mycobacterium marinum*	442	50%	-	6DCD
SEP05279.1	*Blastococcus endophyticus*	553	64%	-	-
KKD07836.1	*Streptomyces* sp. WM6386	468	56%	-	-
CYP109B1[55]	*Bacillus subtilis*	155	32%	fatty acids	4RM4
CYP108A1[53]	*Pseudomonas sp.*	152	32%	α-terpineol	1CPT
CYP1027H1	*Frankia* sp. EuI1c	105	27%	ambroxide	-

Larger substrates with additional cyclic rings such as nootkatone induced a smaller shift of the Soret band (25%). Despite producing a smaller shift compared to several other substrates nootkatone exhibited the tightest binding to the P450 (Kd = 52 μM). α-Santonin shares structural features with nootkatone and induced a shift of 50% (Figure 3 and Figure S8 and S9) which was higher than any other substrate reported. However, its binding affinity was lower (Kd = 120 μM). Other large multiple ring structure compounds such as sclareolide and ambroxide were also tested with CYP150A20, but did not generate any spin state shift.

Acyclic compounds with hydroxyl or ketone groups were also screened and the sesquiterpene *cis*-nerolidol bound to CYP150A20 with an affinity similar to those of the norisoprenoids (Kd = 111 μM, Figure 3). However, it induced a lower spin state shift at around 30%. Neryl acetate, which shares structural features to *cis*-nerolidol, induced a similar spin state shift (30%) and binding affinity (Kd = 123 μM, Figure S9 and S10). The *cis*-conformation of the double bond appears to be important for binding as the *trans* isomer of neryl acetate (geranyl acetate), did not result in a shift in the Soret band.

CYP150A20 was able to bind a wide range of compounds but with relatively low affinity. While norisoprenoids bound to both CYP1027H1 and CYP150A20 the binding affinities for all substrates screened with CYP150A20 were lower compared to those found for CYP1027H1. In contrast to CYP1027H1. Compounds that do not have an oxygen containing functional group did not appear to bind to CYP150A20 with any significant affinity. Therefore while potential substrates have been determined, further screening and investigation is required to obtain the full range for this enzyme.

Figure 3 Compounds tested with CYP150A20. The percentage high spin and dissociation constants are presented as in Figure 2.



3.3.3 Investigation of Substrate Binding to CYP1074A2

CYP1074A2 only shares high amino acid sequence identity to uncharacterised P450s in the NCBI database; the highest similarity was to CYP1074A1 also from *Frankia* sp. EuI1c at 71%. It is also closely related to other uncharacterised P450s from species of *Mycobacteria* and *Blastococcus*, with amino acid sequence identities ranging from 62% to 65%. However, CYP1074A2 was not closely related to any characterised P450, with the majority of the closest related P450s sharing ≤ 30% amino acid sequence identity (Table 6). For instance, CYP106A2, which has been reported to regioselectively hydroxylate steroids at the 15β position, shares only 26% of its sequence with CYP1074A2.[61]

A range of compounds, including norisoprenoids, sesquiterpenoids, fatty acids, and polycylic aromatics, were screened with CYP1074A2 to assess if they could bind. While the majority of these classes of chemicals did not induce any noticeable spin state shifts (Supplementary Material) the steroid testosterone did. Testosterone bound tightly to CYP1074A2 and induced a complete spin state shift to the HS form (Kd = 7.6 µM, ≥ 95%, Figure 4 and Figure S11). The essentially complete type I spin state transition from 420 nm to 390 nm indicates that testosterone is positioned close enough to the heme centre to completely displace the distal water ligand. The low dissociation constant (Kd) demonstrates that the steroid molecule is held tightly within the active site of CYP1074A2. As a result of this, other steroids and steroid-like compounds were also tested for binding affinity towards CYP1074A2.

Steroids with structural similarities to testosterone were also able to induce an almost complete shift (≥ 95%). These include progesterone (Kd = 2.9 µM), 4-androstene-3,17-dione (Kd = 8.3 µM) and stanolone (Kd = 24 µM, Figure 4 and Figure S11). *cis*-Androsterone, a steroid similar to 4-androstene-3,17-dione, also induced a ≥ 95% spin state shift when added to CYP1074A2. However, the binding affinity (Kd = 204 µM, Figure S11) was almost two orders of magnitudes lower than with 4-androstene-3,17-dione. These compounds have the same general steroid ring structure consisting of three cyclohexane rings and one cyclopentane ring (Figure 4). Estrone and pregnenolone exhibited high binding affinity to CYP1074A2 but, unlike androsterone,

both induced a lower spin state shift (Figure 4). This infers that they are not positioned close enough to the iron-heme centre to completely displace the distal water ligand or that their mode of binding interferes with this step.

Table 6 The amino acid sequence identity of CYP1074A2 with other P450s. These were determined using a BLAST sequence search on the NCBI database. Similar P450 enzymes were found in other strains of *Blastococcus* and *Soliminas* bacteria (with similar sequence identities as those presented below) as well as *Algiphilus aromaticivorans* (45% identity) and *Nevskia soli* (42% sequence identity). The P450 name is given when it has been assigned if not the Genbank protein ID or NCBI accession number is provided. [a] Max Score 863.

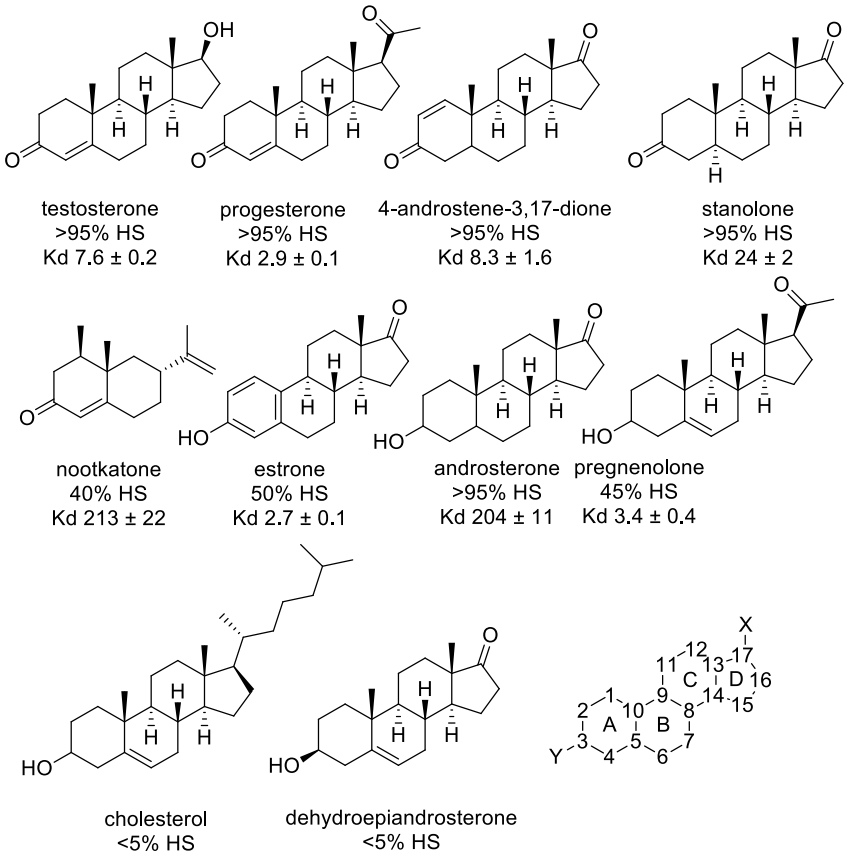
P450	Source organism	Total score[a]	Sequence identity	Known substrates	PDB
CYP1074A1	*Frankia* sp. EuI1c	630	71%	-	-
SEP19814.1	*Blastococcus endophyticus*	543	65%	-	-
ORW83025.1	*Mycobacterium riyadhense*	540	62%	-	-
SFF51085.1	*Fontimonas thermophila*	398	49%	-	-
PPE73593.1	*Solimonas* sp. HR-BB	378	44%	-	-
EIT67795.1	*Hydrocarboniphaga effusa*	369	45%	-	-
PTU30351.1	*Sinobacteraceae* sp. GT1R17	363	43%	-	-
CYP167A1[62]	*Sorangium cellulosum*	190	30%	Epothilone	1PKF
CYP105A1[63]	*Streptomyces griseolus*	126	28%	Vitamin D3	2ZBX
CYP107H1[64]	*Bacillus subtilis*	152	27%	fatty acids	3EJD
CYP106A2[61]	*Bacillus megaterium*	152	26%	steroids	4YT3

Androsterone, estrone and pregnenolone are all sterols, characterised by a hydroxyl group at the C3 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 arrangement of the alkene double bond, the binding affinity to CYP1074A2 and/or the induced spin state shift were significantly decreased. For example, both progesterone and pregnenolone have very similar structures and while their binding affinity was similar, progesterone induced a significantly higher shift than pregnenolone (≥ 95% compared to 45%). Overall steroids with a hydroxyl substituent at the C3 position resulted in a lower spin state shift or binding affinity compared to those containing a carbonyl group at the same position. Thus the C3 substituent not only affects the ability of the steroid to displace the distal water ligand (or driving water bound heme to the HS form),[65-67] but also has a role in the affinity of the substrate binding.

The steroid substrates with the highest affinity for CYP1074A2 contain an oxygen containing moiety at the C17 position. Each of the steroids which bound to CYP1074A2 contain an oxygen containing functional group on C3 and C17. Testosterone, which is structurally very similar to progesterone differing only at the C17 substituent, bound with an almost three-fold lower affinity highlighting the importance of this position. Dehydroepiandrosterone failed to induce any spin state shift in CYP1074A2 (Figure 4). It contains a C3 hydroxyl group and a C17 ketone, but the location of the double bond has moved to the B ring of the steroid (Figure 4). The reason for the absence of a spectral change on addition of dehydroepiandrosterone is not clear as it is structurally very similar to other steroids tested but the combination of a C3 hydroxy group, a C5-C6 alkene and a C17 ketone may not favour binding. The position of the cyclic double bond therefore appears to be important with the steroids. Those containing a conjugated double bond with the C3 ketone displayed the best binding parameters compared

to those which had saturated or aromatic A rings or the alkene in the B ring. For example, stanolone and 4-androstene-3,17-dione only differ by the addition of this double bond with the former having a decreased binding affinity. Cholesterol also did not induce any spin state shift in CYP1074A2. It does contain a C3 hydroxyl group but has a long branched carbon chain at the C17 position (Figure 4).

Figure 4 Compounds identified which bind to CYP1074A2. The steroid ring labels and the carbon numbering scheme are provided. The percentage high spin and dissociation constants are presented as in Figure 2.



The only non-steroid compound tested that induced a significant spin state shift was nootkatone. This sesquiterpenoid shares some structural similarities with steroids and induced a lower spin state shift and binding affinity compared to the best steroids (40%, Kd = 213 µM). It contains the same A and B cyclohexane ring skeleton and has a carbonyl group at C3 which is conjugated to an alkene (Figure 4). This provides additional evidence that these structural features are important for binding to CYP1074A2. However it lacks the cyclohexane (C) and cyclopentane (D) rings which make up the core structure of a steroid. Other substrates with structural similarities to nootkatone but which were larger, such as ambroxide and sclareolide, failed to induce any spin state shift.

Overall although the structures of some of these steroids are very similar, the variation in the binding affinities, suggest that small structural differences play a role in binding. The cumulative effects of the substituents at C3 and C17, as well as the position of the double bond within the steroids contribute significantly to the binding affinity with CYP1074A2. Both the binding affinities and induced spin states of these substrates were higher than those screened with CYP1027H1 and CYP150A20 inferring that the likely physiological substrate of CYP1074A2 is based on a steroid structure.

4 Conclusion

Soil-dwelling actinomycetes, such as *Frankia*, are known for being producers of many bioactive compounds and natural products. Recent bioinformatic analysis has revealed an abundance of secondary metabolic biosynthesis gene clusters in the other *Frankia* strains ACN14a, CcI3, and EAN1pec. These secondary metabolites and natural products include cyclic peptides, siderophores, aromatic compounds, signalling molecules and lipids. These gene clusters are also unique to individual strains of *Frankia*, suggesting species-specific biosynthetic diversity. The large number of P450s belonging to distinct families as well as the associated electron transfer partners found in different *Frankia* strains highlights the importance of these enzyme for forming novel natural products and secondary metabolites in these bacteria. Here we have identified and analysed the 68 cytochrome P450 encoding genes of *Frankia* sp. EuI1c along with associated ferredoxin (nine) and ferredoxin reductase (two) electron transfer partners. A selection of these P450s were produced, purified and screened for substrate binding with a range of different compounds. These results give an indication as to what their substrates may be (terpenoids and steroids). It is likely that steroids or related terpenoids are the physiological substrates of CYP1074A2 confirming the important role these enzymes have in *Frankia* sp. EuI1c secondary metabolism.

Acknowledgements

This work was supported by the Australian Research Council through a Future Fellowship (FT140100355 to S.G.B.). The authors also acknowledge the award of a University of Adelaide M. Phil. Scholarship (I.C.-K.L.). We thank Prof. Lou Tisa (University of New Hampshire) for the genomic DNA of *Frankia* sp. Eu1Ic and Prof Ray Rodgers (University of Adelaide) for providing some of the steroid samples used in this work. We also acknowledge and thank Dr Alex Taylor (https://www.alexandertaylor.me/) for the pictures of *Elaeagnus umbellate* and its root nodules used in the graphical abstract.

References

[1] D.R. Benson, J.M. Brooks, Y. Huang, D.M. Bickhart, J.E. Mastronunzio, The Biology of Frankia sp. Strains in the Post-Genome Era, Mol. Plant-Microbe Interact., 24 (2011) 1310-1316.
[2] D.R. Benson, W.B. Silvester, Biology of Frankia strains, actinomycete symbionts of actinorhizal plants, Microbiol. Rev., 57 (1993) 293-319.
[3] I. Nouioui, F. Ghodhbane-Gtari, N.J. Beauchemin, L.S. Tisa, M. Gtari, Phylogeny of members of the Frankia genus based on gyrB, nifH and glnII sequences, Antonie van Leeuwenhoek, 100 (2011) 579-579.
[4] T.V. Nguyen, D. Wibberg, K. Battenberg, J. Blom, B. Vanden Heuvel, A.M. Berry, J. Kalinowski, K. Pawlowski, An assemblage of Frankia Cluster II strains from California contains the canonical nod genes and also the sulfotransferase gene nodH, BMC Genomics, 17 (2016) 796.
[5] I. Nouioui, M. Del Carmen Montero-Calasanz, F. Ghodhbane-Gtari, M. Rohde, L.S. Tisa, H.P. Klenk, M. Gtari, Frankia discariae sp. nov.: an infective and effective microsymbiont isolated from the root nodule of Discaria trinervis, Arch. Microbiol., 199 (2017) 641-647.
[6] P. Normand, I. Nouioui, P. Pujic, P. Fournier, A. Dubost, G. Schwob, H.P. Klenk, A. Nguyen, D. Abrouk, A. Herrera-Belaroussi, J.F. Pothier, V. Pfluger, M.P. Fernandez, Frankia canadensis sp. nov., isolated from root nodules of Alnus incana subspecies rugosa, Int. J. Syst. Evol. Microbiol., 68 (2018) 3001-3011.
[7] I. Nouioui, F. Ghodhbane-Gtari, M. Rohde, H.P. Klenk, M. Gtari, Frankia coriariae sp. nov., an infective and effective microsymbiont isolated from Coriaria japonica, Int. J. Syst. Evol. Microbiol., 67 (2017) 1266-1270.
[8] I. Nouioui, A. Gueddou, F. Ghodhbane-Gtari, M. Rhode, M. Gtari, H.P. Klenk, Frankia asymbiotica sp. nov., a non-infective actinobacterium isolated from Morella californica root nodule, Int. J. Syst. Evol. Microbiol., 67 (2017) 4897-4901.
[9] L.S. Tisa, N. Beauchemin, M. Gtari, A. Sen, L. Wall, What stories can the Frankia genomes start to tell us?, J. Biosci., 38 (2013) 719-726.
[10] I. Nouioui, F. Ghodhbane-Gtari, M. Del Carmen Montero-Calasanz, M. Rohde, L.S. Tisa, M. Gtari, H.P. Klenk, Frankia inefficax sp. nov., an actinobacterial endophyte inducing ineffective, non nitrogen-fixing, root nodules on its actinorhizal host plants, Antonie Van Leeuwenhoek, 110 (2017) 313-320.
[11] P. Normand, P. Lapierre, L.S. Tisa, J.P. Gogarten, N. Alloisio, E. Bagnarol, C.A. Bassi, A.M. Berry, D.M. Bickhart, N. Choisne, A. Couloux, B. Cournoyer, S. Cruveiller, V. Daubin, N. Demange, M.P. Francino, E. Goltsman, Y. Huang, O.R. Kopp, L. Labarre, A. Lapidus, C. Lavire, J. Marechal, M. Martinez, J.E. Mastronunzio, B.C. Mullin, J. Niemann, P. Pujic, T. Rawnsley, Z. Rouy, C. Schenowitz, A. Sellstedt, F. Tavares, J.P. Tomkins, D. Vallenet, C. Valverde, L.G. Wall, Y. Wang, C. Medigue, D.R. Benson, Genome characteristics of facultatively symbiotic Frankia sp. strains reflect host range and host plant biogeography, Genome Res., 17 (2007) 7-15.
[12] D. Baker, W. Newcomb, J.G. Torrey, Characterization of an ineffective actinorhizal microsymbiont, Frankia sp. EuI1 (Actinomycetales), Can. J. Microbiol., 26 (1980) 1072-1089.
[13] L.S. Tisa, R. Oshone, I. Sarkar, A. Ktari, A. Sen, M. Gtari, Genomic approaches toward understanding the actinorhizal symbiosis: an update on the status of the Frankia genomes, Symbiosis, 70 (2016) 5-16.
[14] D.W. Udwary, E.A. Gontang, A.C. Jones, C.S. Jones, A.W. Schultz, J.M. Winter, J.Y. Yang, N. Beauchemin, T.L. Capson, B.R. Clark, E. Esquenazi, A.S. Eustaquio, K. Freel, L. Gerwick, W.H. Gerwick, D. Gonzalez, W.T. Liu, K.L. Malloy, K.N. Maloney, M. Nett, J.K.

Nunnery, K. Penn, A. Prieto-Davo, T.L. Simmons, S. Weitz, M.C. Wilson, L.S. Tisa, P.C. Dorrestein, B.S. Moore, Significant natural product biosynthetic potential of actinorhizal symbionts of the genus frankia, as revealed by comparative genomic and proteomic analyses, Appl. Environ. Microbiol., 77 (2011) 3617-3625.
[15] M. Rehan, T. Furnholm, R.H. Finethy, F. Chu, G. El-Fadly, L.S. Tisa, Copper tolerance in Frankia sp. strain EuI1c involves surface binding and copper transport, Appl. Microbiol. Biotechnol., 98 (2014) 8005-8015.
[16] E. Baker, Y. Tang, F. Chu, L.S. Tisa, Molecular responses of Frankia sp. strain QA3 to naphthalene, Can. J. Microbiol., 61 (2015) 281-292.
[17] M. Rehan, M. Kluge, S. Fränzle, H. Kellner, R. Ullrich, M. Hofrichter, Degradation of atrazine by Frankia alni ACN14a: gene regulation, dealkylation, and dechlorination, Appl. Microbiol. Biotechnol., 98 (2014) 6125-6135.
[18] J.W. Richards, G.D. Krumholz, M.S. Chval, L.S. Tisa, Heavy Metal Resistance Patterns of Frankia Strains, Appl. Environ. Microbiol., 68 (2002) 923-927.
[19] I.G. Denisov, S.G. Sligar, Cytochrome P450, in: P.R.O. de Montellano (Ed.), Springer International Publishing, Place Published, 2015, pp. 69-109.
[20] T.L. Poulos, Heme enzyme structure and function, Chem. Rev., 114 (2014) 3919-3962.
[21] A. Greule, J.E. Stok, J.J. De Voss, M.J. Cryle, Unrivalled diversity: the many roles and reactions of bacterial cytochromes P450 in secondary metabolism, Nat. Prod. Rep., 35 (2018) 757-791.
[22] F.P. Guengerich, Common and Uncommon Cytochrome P450 Reactions Related to Metabolism and Chemical Toxicity, Chem. Res. Toxicol., 14 (2001) 611-650.
[23] C.M. Krest, E.L. Onderko, T.H. Yosca, J.C. Calixto, R.F. Karp, J. Livada, J. Rittle, M.T. Green, Reactive intermediates in cytochrome p450 catalysis, J. Biol. Chem., 288 (2013) 17074-17081.
[24] F. Hannemann, A. Bichet, K.M. Ewen, R. Bernhardt, Cytochrome P450 systems - biological variations of electron transport chains, Biochim. Biophys. Acta (BBA) - Gen. Subj., 1770 (2007) 330-344.
[25] D.R. Nelson, Cytochrome P450 Nomenclature, 2004, in: I. Phillips, E. Shephard (Eds.) Cytochrome P450 Protocols, Humana Press, Place Published, 2006, pp. 1-10.
[26] A. Theodosiou, V.J. Promponas, LaTcOm: a web server for visualizing rare codon clusters in coding sequences, Bioinformatics, 28 (2012) 591-592.
[27] T. Omura, R. Sato, The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature, J. Biol. Chem., 239 (1964) 2370-2378.
[28] N.K. Maddigan, S.G. Bell, The self-sufficient CYP102 family enzyme, Krac9955, from Ktedonobacter racemifer DSM44963 acts as an alkyl- and alkyloxy-benzoic acid hydroxylase, Arch. Biochem. Biophys., 615 (2017) 15-21.
[29] D.R. Nelson, T. Kamataki, D.J. Waxman, F.P. Guengerich, R.W. Estabrook, R. Feyereisen, F.J. Gonzalez, M.J. Coon, I.C. Gunsalus, O. Gotoh, The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature, DNA Cell Biol., 12 (1993) 1-51.
[30] S.A. Child, J.M. Bradley, T.L. Pukala, D.A. Svistunenko, N.E. Le Brun, S.G. Bell, Electron transfer ferredoxins with unusual cluster binding motifs support secondary metabolism in *Mycobacteria* and are prevalent in many other bacteria, Chem. Sci., 9 (2018) 7948 - 7957.
[31] S.A. Child, V.P. Rossi, S.G. Bell, Selective omega-1 oxidation of fatty acids by CYP147G1 from Mycobacterium marinum, Biochim. Biophys. Acta Gen. Subj., 1863 (2019) 408-417.
[32] S.A. Child, K.L. Flint, J.B. Bruning, S.G. Bell, The characterisation of two members of the cytochrome P450 CYP150 family: CYP150A5 and CYP150A6 from Mycobacterium

marinum, Biochim. Biophys. Acta Gen. Subj., (2019).Accepted article DOI: 10.1016/j.bbagen.2019.02.016.
[33] R. Feyereisen, Arthropod CYPomes illustrate the tempo and mode in P450 evolution, Biochim. Biophys. Acta Proteins Proteom., 1814 (2011) 19-28.
[34] M. Lynch, J.S. Conery, The origins of genome complexity, Science, 302 (2003) 1401-1404.
[35] E. García-Fernández, D.J. Frank, B. Galán, P.M. Kells, L.M. Podust, J.L. García, P.R. Ortiz de Montellano, A highly conserved mycobacterial cholesterol catabolic pathway, Environ. Microbiol., 15 (2013) 2342-2359.
[36] D. Frank, C.A. Waddling, M. La, P.R. Ortiz de Montellano, Cytochrome P450 125A4, the Third Cholesterol C-26 Hydroxylase from Mycobacterium smegmatis, Biochemistry, 54 (2015) 6909-6916.
[37] M. Parvez, L.B. Qhanya, N.T. Mthakathi, I.K.R. Kgosiemang, H.D. Bamal, N.S. Pagadala, T. Xie, H. Yang, H. Chen, C.W. Theron, R. Monyaki, S.C. Raselemane, V. Salewe, B.L. Mongale, R.G. Matowane, S.M.H. Abdalla, W.I. Booi, M. van Wyk, D. Olivier, C.E. Boucher, D.R. Nelson, J.A. Tuszynski, J.M. Blackburn, J.H. Yu, S.S. Mashele, W. Chen, K. Syed, Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: Special focus on mycobacterial P450s, Sci. Rep., 6 (2016).
[38] S. Rupasinghe, M.A. Schuler, N. Kagawa, H. Yuan, L. Lei, B. Zhao, S.L. Kelly, M.R. Waterman, D.C. Lamb, The cytochrome P450 gene family CYP157 does not contain EXXR in the K-helix reducing the absolute conserved P450 residues to a single cysteine, FEBS Lett., 580 (2006) 6338-6342.
[39] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol Syst Biol, 2011, pp. 539.
[40] S.G. Bell, A.B.H. Tan, E.O.D. Johnson, L.-L. Wong, Selective oxidative demethylation of veratric acid to vanillic acid by CYP199A4 from Rhodopseudomonas palustris HaA2, Mol. Biosyst., 6 (2009) 206-214.
[41] W. Yang, S.G. Bell, H. Wang, W. Zhou, N. Hoskins, A. Dale, M. Bartlam, L.-L. Wong, Z. Rao, Molecular Characterization of a Class I P450 Electron Transfer System from Novosphingobium aromaticivorans DSM12444, J. Biol. Chem., 285 (2010) 27372-27384.
[42] T. Zhang, A. Zhang, S.G. Bell, L.-L. Wong, W. Zhou, The structure of a novel electron-transfer ferredoxin from Acta Crystallogr. Sect. D, 70 (2014) 1453-1464.
[43] M. Bruschi, F. Guerlesquin, Structure, function and evolution of bacterial ferredoxins, FEMS Microbiol. Lett., 54 (1988) 155-175.
[44] I.V.T.F. Clarke, P.L. Clark, Rare Codons Cluster, PLoS ONE, 3 (2008) 1-5.
[45] Y. Nakamura, T. Gojobori, T. Ikemura, Codon usage tabulated from the international DNA sequence databases: status for the year 2000, Nucleic Acids Res., 28 (2000) 292-292.
[46] J.F. Kane, Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli, Curr. Opin. Biotechnol., 6 (1995) 494-500.
[47] Y. Sun, W. Zeng, A. Benabbas, X. Ye, I. Denisov, S.G. Sligar, J. Du, J.H. Dawson, P.M. Champion, Investigations of Heme Ligation and Ligand Switching in Cytochromes P450 and P420, Biochemistry, 52 (2013) 5941-5951.
[48] R. Perera, M. Sono, J.A. Sigman, T.D. Pfister, Y. Lu, J.H. Dawson, Neutral thiol as a proximal ligand to ferrous heme iron: implications for heme proteins that lose cysteine thiolate ligation on reduction, Proc. Natl. Acad. Sci. U. S. A., 100 (2003) 3641-3646.
[49] P.M. Champion, I.C. Gunsalus, G.C. Wagner, Resonance Raman investigations of cytochrome P450CAM from Pseudomonas putida, J. Am. Chem. Soc., 100 (1978) 3743-3751.

[50] G. Hui Bon Hoa, C. Di Primo, I. Dondaine, S.G. Sligar, I.C. Gunsalus, P. Douzou, Conformational changes of cytochromes P-450cam and P-450lin induced by high pressure, Biochemistry, 28 (1989) 651-656.
[51] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol., 215 (1990) 403-410.
[52] D.B. Hawkes, G.W. Adams, A.L. Burlingame, P.R. Ortiz de Montellano, J.J. De Voss, Cytochrome P450cin (CYP176A), Isolation, Expression, and Characterization, J. Biol. Chem., 277 (2002) 27725-27732.
[53] S.S. Boddupalli, C.A. Hasemann, K.G. Ravichandran, J.-Y. Lu, E.J. Goldsmith, J. Deisenhofer, J.A. Peterson, 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, Proc. Nat. Acad. Sci. U. S. A., 89 (1992) 5567-5571.
[54] M. Katagiri, B.N. Ganguli, I.C. Gunsalus, A Soluble Cytochrome P-450 Functional in Methylene Hydroxylation, J. Biol. Chem., 243 (1968) 3543-3546.
[55] M. Girhard, T. Klaus, Y. Khatri, R. Bernhardt, V.B. Urlacher, Characterization of the versatile monooxygenase CYP109B1 from Bacillus subtilis, Appl. Microbiol. Biotechnol., 87 (2010) 595-607.
[56] T.L. Poulos, B.C. Finzel, A.J. Howard, High-resolution crystal structure of cytochrome P450cam, J. Mol. Biol., 195 (1987) 687-700.
[57] C. Di Primo, G. Hui Bon Hoa, P. Douzou, S. Sligar, Mutagenesis of a single hydrogen bond in cytochrome P-450 alters cation binding and heme solvation, J. Biol. Chem., 265 (1990) 5361-5363.
[58] Y.T. Meharenna, K.E. Slessor, S.M. Cavaignac, T.L. Poulos, J.J. De Voss, The critical role of substrate-protein hydrogen bonding in the control of regioselective hydroxylation in p450cin, J. Biol. Chem., 283 (2008) 10804-10812.
[59] B. Brezna, O. Kweon, R. Stingley, J. Freeman, A. Khan, B. Polek, R. Jones, C. Cerniglia, Molecular characterization of cytochrome P450 genes in the polycyclic aromatic hydrocarbon degrading Mycobacterium vanbaalenii PYR-1, Appl. Microbiol. Biotechnol., 71 (2006) 522-532.
[60] D.R. Nelson, The Cytochrome P450 Homepage, Hum. Genomics, 4 (2009) 59-65.
[61] C. Virus, M. Lisurek, B. Simgen, F. Hannemann, R. Bernhardt, Function and engineering of the 15beta-hydroxylase CYP106A2, Biochem. Soc. Trans., 34 (2006) 1215-1218.
[62] F. Kern, T.K. Dier, Y. Khatri, K.M. Ewen, J.P. Jacquot, D.A. Volmer, R. Bernhardt, Highly Efficient CYP167A1 (EpoK) dependent Epothilone B Formation and Production of 7-Ketone Epothilone D as a New Epothilone Derivative, Sci. Rep., 5 (2015) 14881.
[63] D.P. O'Keefe, J.M. Tepperman, C. Dean, K.J. Leto, D.L. Erbes, J.T. Odell, Plant Expression of a Bacterial Cytochrome P450 That Catalyzes Activation of a Sulfonylurea Pro-Herbicide, Plant Physiol., 105 (1994) 473-482.
[64] M.J. Cryle, N.J. Matovic, J.J. De Voss, Products of cytochrome P450(BioI) (CYP107H1)-catalyzed oxidation of fatty acids, Org. Lett., 5 (2003) 3341-3344.
[65] K.P. Conner, A.M. Schimpf, A.A. Cruce, K.J. McLean, A.W. Munro, D.J. Frank, M.D. Krzyaniak, P. Ortiz de Montellano, M.K. Bowman, W.M. Atkins, Strength of axial water ligation in substrate-free cytochrome P450s is isoform dependent, Biochemistry, 53 (2014) 1428-1434.
[66] K.P. Conner, C.M. Woods, W.M. Atkins, Interactions of cytochrome P450s with their ligands, Arch. Biochem. Biophys., 507 (2011) 56-65.
[67] R. Raag, T.L. Poulos, The structural basis for substrate-induced changes in redox potential and spin equilibrium in cytochrome P-450CAM, Biochemistry, 28 (1989) 917-922.