

# Molecular BioSystems

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

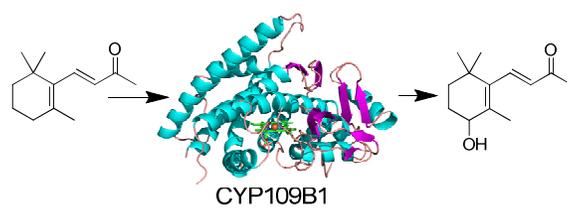
You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



[www.rsc.org/molecularbiosystems](http://www.rsc.org/molecularbiosystems)

## TOC



The structure of CYP109B1 from *Bacillus subtilis*, which catalyses the oxidation of ionones, has been determined. This will allow the future design of more efficient biocatalytic monooxygenase systems.

The crystal structure of the versatile cytochrome P450 enzyme CYP109B1 from *Bacillus subtilis*

Aili Zhang,<sup>1</sup> Ting Zhang,<sup>1</sup> Emma A. Hall,<sup>2</sup> Sean Hutchinson,<sup>2</sup> Max J. Cryle,<sup>3</sup> Luet-Lok Wong<sup>4</sup> and Weihong Zhou<sup>1,\*</sup> and Stephen G. Bell<sup>2,4,\*</sup>

<sup>1</sup> College of Life Sciences, Nankai University, Tianjin 300071, China

<sup>2</sup> School of Chemistry and Physics, University of Adelaide, SA 5005, Australia

<sup>3</sup> Department of Biomolecular Mechanisms, Max Planck Institute for Medical Research, Jahnstrasse 29, 69120 Heidelberg, Germany

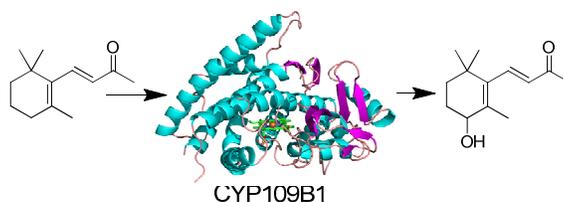
<sup>4</sup> Department of Chemistry, University of Oxford, Inorganic Chemistry Laboratory, South Parks Road, Oxford, OX1 3QR, UK

\* To whom correspondence should be addressed.

Stephen G. Bell ([stephen.bell@adelaide.edu.au](mailto:stephen.bell@adelaide.edu.au)) Tel: +61 8 83134822; Fax: +61 8 83034380

Weihong Zhou ([zhouwh@nankai.edu.cn](mailto:zhouwh@nankai.edu.cn)) Tel: +86 22 23507118; Fax: +86 22 23507118

## TOC



The structure of CYP109B1 from *Bacillus subtilis*, which catalyses the oxidation of ionones, has been determined. This will allow the future design of more efficient biocatalytic monooxygenase systems.

**Abstract**

The crystal structure of the versatile CYP109B1 enzyme from *Bacillus subtilis* has been solved at 1.8 Å resolution. This is the first structure of an enzyme from this CYP family, whose members are prevalent across diverse species of bacteria. In the crystal structure the enzyme has an open conformation with an access channel leading from the heme to the surface. The substrate-free structure reveals the location of the key residues in the active site that are responsible for binding the substrate in the correct orientation for regioselective oxidation. Importantly, there are significant differences among these residues in members of the CYP109 and closely related CYP106 families and these likely account for the variations in substrate binding and oxidation profiles observed with these enzymes. A whole-cell oxidation biosystem was developed, which contains CYP109B1 and a phthalate family oxygenase reductase (PFOR), from *Pseudomonas putida* KT24440, as the electron transfer partner. This electron transfer system is able to support CYP109B1 activity resulting in the regioselective hydroxylation of both  $\alpha$ - and  $\beta$ -ionone *in vivo* and *in vitro*. The PFOR is therefore a versatile electron transfer partner that is able to support the activity of CYP enzymes from other bacterium. The crystal structure of CYP109B1 has a positively charged proximal face and this explains why it can interact with the PFOR and adrenodoxin which are predominantly negatively charged around their [2Fe-2S] clusters.

## Introduction

The cytochromes P450 are a group of heme thiolate–ligated monooxygenase enzymes found in almost all forms of life. They catalyse reactions in important metabolic pathways, including a wide variety of oxidative reactions, the most common being the insertion an oxygen atom from dioxygen into chemically inert carbon-hydrogen bonds with high regio- and stereoselectivity.<sup>1,2</sup> As such they are involved in the synthesis of diverse organic molecules (such as the biosynthesis of lipids, steroid hormones) and complex antibiotics (such as vancomycin and erythromycin) as well as in the detoxification of xenobiotics including the metabolism of drugs.<sup>1-3</sup> The ability of bacterial monooxygenase enzymes to catalyse regioselective and stereoselective hydroxylation reactions under mild conditions has resulted in extensive efforts to isolate and engineer robust CYP enzymes for chemical applications.<sup>4-8</sup>

Dioxygen activation by CYP enzymes requires two electrons that are usually derived from NAD(P)H and delivered by electron transfer proteins.<sup>1,9</sup> Bacterial P450 enzymes use a variety of electron transfer partner systems including different types of NAD(P)H dependent ferredoxin reductases, ferredoxins, flavodoxin and fusion proteins.<sup>8-12</sup> The identification of a suitable electron transfer partner is often critical for efficient turnover.<sup>13-15</sup>

The CYP109 family enzyme, CYP109A1, was identified in *Bacillus subtilis* strain W23<sup>16,17</sup> and other members of this family are prevalent in many strains of *Bacillus*.<sup>15,16</sup> CYP109B1, from *Bacillus subtilis* strain 168, is the most extensively studied enzyme from this family and has been reported to hydroxylate compactin to pravastatin and valencene to nootkatol and nootkatone. The oxidation of fatty acids along with testosterone and other steroids has also been described.<sup>18-21</sup> Additionally, the CYP109C1, CYP109C2, and CYP109D1 enzymes from *Sorangium cellulosum* So ce56 have been investigated:<sup>22</sup> these enzymes have all been shown to bind and turnover fatty acids with oxidation

occurring at the  $\omega_{-1}$  to  $\omega_{-6}$  positions. The natural electron transfer systems of these CYP109 family enzymes are as yet unknown and mitochondrial adrenodoxin and adrenodoxin reductase have been used as surrogate electron transfer partners to support their activity.<sup>20, 23</sup>

Both CYP109B1 and CYP109D1 have been tested for their potential to oxidise the norisoprenoids  $\alpha$ - and  $\beta$ -ionone (Scheme 1).<sup>20, 23, 24</sup> Norisoprenoids are a family of natural products, related to terpenes, which serve a range of important physiological functions and are fragrance compounds with floral scents.<sup>25</sup> The oxygenated products of ionones are desirable compounds and can be used as synthetic building blocks, for example 4-hydroxy- $\beta$ -ionone is an important intermediate for the synthesis of carotenoids and deoxyabscisic acid.<sup>15, 26, 27</sup>

Here we report the crystal structure of CYP109B1, the first to be reported for a member of this CYP family. This allows detailed comparison of different CYP109 enzymes and members of other related families such as CYP106A2 and CYP278A1. We also have developed a whole-cell oxidation system for CYP109B1 exploiting a phthalate family oxygenase reductase (PFOR) from *Pseudomonas putida* KT24440 as the electron transfer partner.<sup>10</sup> Using this whole-cell system we were successfully able to regioselectively hydroxylate both  $\alpha$ - and  $\beta$ -ionone, which together with the structure of CYP109B1 will enable further development of this system towards the production of high value compounds.

## Results

### Phylogenetic analysis of the CYP109 family

In order to assess how common the CYP109 family is across the bacterial kingdom we utilised the Basic Local Alignment Search Tool (BLAST) to search for similar enzymes to CYP109B1 from *B. subtilis* str. 168.<sup>27</sup> Eleven P450s with > 90% sequence identity and a further nineteen with > 60% sequence identity were discovered in different strains of *Bacillus* bacteria. A much larger number, with > 40% sequence identity, were found (~ 200) in strains of *Bacillus* and related bacteria such as *Paenibacillus*,<sup>28</sup> *Brevibacillus*<sup>29</sup> and other bacterial strains such as *Arthrobacter*,<sup>30</sup> *Ktedonobacter*,<sup>31</sup> *Herpetosiphon*,<sup>32</sup> *Haloferax*<sup>33</sup> and *S. cellulosum* (CYP109C1, CYP109C2 and CYP109D1). The CYP109B1 enzyme also shares > 40% sequence identity with the CYP106A1 and CYP106A2 enzymes, both from strains of *B. megaterium*. The CYP106 enzymes oxidise steroids and di- and tri-terpenoids.<sup>34-36</sup> CYP109C1 and CYP109C2 shared > 40% sequence identity with other CYP enzymes from strains of *Paenibacillus*, *Bacillus* and *Arthrobacter* while CYP109D1 from *S. cellulosum* had a lower sequence homology with the *Bacillus* CYP enzymes. However CYP109D1 does share > 40% sequence identity with several CYP enzymes from *Mycobacterium* strains including CYP278A1 and CYP123B1 from *M. marinum* (Table 1).<sup>37</sup>

In order to better understand the relationship of these CYP enzymes from different bacterial strains we compared the sequence of CYP109B1 with other CYP109 and CYP106 members from selected *Bacillus* strains, and the other members of the CYPome of *B. subtilis* str. 168 (CYP102A2, CYP102A3, CYP107H1, CYP107J1, CYP107K1, CYP134A1 and CYP152A1) (Fig. 1).<sup>38-49</sup> A phylogenetic tree of these *Bacillus* CYP enzymes, and the CYP109 enzymes from *S. cellulosum*, CYP278A1 from *M. marinum* and other well characterised bacterial CYP enzymes highlights that, as expected, the CYP109 enzymes cluster together but that there are different levels of sequence

homology depending on the strain of *Bacillus* (Fig. 1 and Table 1). It is of note that CYP106A1 and CYP106A2 are more similar to the CYP109B subfamily enzymes than are CYP109 members from the A, C and D subfamilies suggesting these systems are related and would be expected to share functional similarities. The CYP109C1 and CYP109C2 enzymes from *S. cellulosum* have greater similarity to CYP109B1 than CYP109D1 from the same bacterium. However the CYP109C and CYP109D subfamily members all have a lower homology with CYP109B1 than the CYP109A1 enzyme from *B. subtilis* strain W23. CYP278A1 from *M. marinum* is most closely related to CYP109D1. CYP123B1 from *M. marinum* also has > 40% similarity with CYP109D1, whilst CYP123 from *M. tuberculosis* strain H37Rv clusters with the CYP107 family enzymes. The CYP107 family congregates together with the CYP109 family while the CYP102 family and the peroxygenase enzyme, CYP152A1, form a distinctive grouping (Fig. 1).

To analyse the local sequence similarities between the CYP109 family members we aligned the sequences of CYP109B1, CYP109C1, CYP109C2 and CYP109D1 with CYP109 enzymes from *B. mojavensis*, *B. amyloliquefaciens* and *B. cereus*, CYP106A2 from *B. megaterium* ATCC13368 and CYP278A1 from *M. marinum* (Fig. 2); we also included CYP101A1 (P450cam) for comparative purposes. CYP109B1 is broadly similar to the other CYP109 family members and there are numerous regions of high sequence identity and similarity between all the enzymes. The EXXR motif of the K-helix, which has been linked to heme incorporation, is strictly conserved in all the CYP109 P450s analysed in this work (EEXLR).<sup>50, 51</sup> Other areas that appear particularly well conserved include the I-helix and the proximal heme binding loop (Fig. 2 and S1). In the I-helix the acid-alcohol pair is conserved with glutamate and threonine residues being strictly maintained (Glu242 and Thr243 in CYP109B1) with the exception of CYP109D1, where the glutamate residue is replaced with an aspartate (Fig. 2 and S1).<sup>52</sup> Analysis of the residues between Leu236 and Thr245

of the I-helix (Fig. S1) reveals that the only residue that shows significant variation between the CYP109 family members and CYP278A1 is Asn241, in CYP109B1: this corresponds to a Thr residue in CYP109C1, a His residue in CYP109C2 and a Ser residue in CYP278A1. The only other changes in this region occur at Leu237 (Val residue in CYP278A1), Val238 (Ala residue in CYP109C1) and Thr244 (Ala residue in CYP109C1). Thus, the <sup>236</sup>LLVAGNETTT<sup>245</sup> sequence is conserved across all the members of the CYP109B and CYP109A subfamilies aligned in this work (Fig. 2). The CYP106A2 enzyme showed greater variations in this region compared to the CYP109 and CYP278A1 enzymes having the equivalent sequence **ILGAGVETTS** (with differences highlighted in bold).

The heme binding domain is also highly conserved (Fig. 2 and S1) with the residues highlighted in bold in the following motif **HXXFGXGXHFCXGAXL** being strictly maintained across the CYP109 and CYP106A2 enzymes. The phenylalanine, seven residues before the iron binding cysteine residue, is replaced by a leucine in CYP278A1 (Fig. S1).<sup>53</sup> The first and third residues after the cysteine, which are involved in the proximal hydrogen bonding interactions with the sulfur of this residue and stabilise the heme and regulate its redox potential, are well conserved. The first is a leucine or isoleucine whilst the latter is an alanine in all the CYPs aligned in Fig. 2.<sup>54-56</sup> CYP106A2, CYP109D1 and CYP278A1 have a different arrangement of proline residues, prior to the heme binding loop, when compared to Pro336 and Pro338 in CYP109B1, which may lead to subtle changes in this part of the substrate binding pocket in these enzymes.

Obvious areas of difference between the enzymes analysed include the insertion of additional amino acids between the A- and B-helices in CYP109D1 and CYP278A1 and a further insertion between the D- and E-helices in CYP278A1. CYP109D1 also has a deletion in the G-helix compared to the other CYP enzymes (Fig. 2) while CYP109 from *B. cereus* has a two amino acid

insertion at the end of this helix. CYP109B1 and the CYP109 enzymes from *B. mojavensis* and *B. amyloliquefaciens* possess fewer residues at the C-terminus of the L-helix. CYP106A2 has a ten amino acid insertion prior to the first substrate recognition site (SRS1, before Met70 in CYP109B1) and a truncated G-helix compared to CYP109B1.

There are differences in the six regions assigned as substrate recognition sites (SRS1-6) as defined by Gotoh.<sup>57</sup> These substrate recognition sites are the residues that comprise the active site of the CYP109 enzymes. For example the CYP109D1 interaction with ionone has been modelled and the His94 residue in SRS1 was postulated to be important in H-bonding to the carbonyl oxygen of the ionone substrate. This residue is not well conserved across the CYP109 enzymes being replaced by an asparagine residue, Asn83, in CYP109B1, *vide infra* (Fig. 2).

### **Structure of CYP109B1**

#### ***The overall structure of CYP109B1***

The crystal structure of CYP109B1 was solved at 1.8 Å resolution using molecular replacement and the PDB file 2WHW as a search model. The molecule was traced from residues Asp22 to Met396 with gaps due to unresolved electron density between Met70 to Gly78 and Glu216 to Gly221; these correspond to the flexible parts of cytochrome P450 enzymes between B/B' and H/I loops. Despite the absence of electron density in these regions it is apparent that CYP109B1 displays the classic P450 fold with eleven helices and nine β-sheets observed (Fig. 3a, Table S1).

CYP109B1 was crystallised in an open conformation with an access channel leading from the heme to the enzyme surface (Fig. S2). The F-helix and G-helix extend out forming the access channel and these helices along with the F/G loop (Val166 to Ala176) and B/B' loop, which is at the top of the channel are likely to play a role in substrate recognition and entry (Fig. 3 and Fig. S2). The tertiary structure of CYP109B1 matches closely those of other CYP enzymes which have been

crystallised with an open conformation including CYP101A1 (PDB: 3L61) and CYP101D2 (PDB: 3NV5) (Fig. S3).<sup>58, 59</sup>

### ***The heme environment***

The heme-iron is bound to the proximal cysteine ligand, Cys349 (2.3 Å) and Wat28 is the sixth iron ligand (Fe–O, 2.4 Å, Table S2). In this structure the iron is out of the plane of the heme by 0.5 Å toward the proximal cysteine sulfur. The sulfur of this cysteine forms hydrogen bonds with the main chain atoms of Phe342, His347, Phe348 and Gly351 (Fig. 3b). These residues are conserved across all members of the different CYP109 family (and CYP106A2) and their interactions with the atoms in the peptide chain backbone are conserved across other P450 enzymes even though the identity of the residues may differ.

Like other CYP enzymes, the B- and C-rings of the heme in CYP109B1 interact with the polypeptide chain via hydrophobic interactions with non-polar residues from the D-, I- and L-helices. The A-ring propionate forms a salt bridge with the side-chain of Arg291 and a hydrogen bond with three water molecules (Wat9, Wat65 and Wat78) (Fig. S4). The D-ring propionate forms salt bridges with N atoms from the side chains of His89, His347 and Arg93 and is in van der Waals contact with Ile82. The residues His89, Arg93, Arg291 and His347 are all conserved across the CYP109 family members and the CYP106A2 and CYP278A1 enzymes (Fig. S4). They are also found in other CYP enzymes, for example in the CYP101 subfamily of enzymes these Arg and His residues are conserved, with the exception of CYP101A1 where a Gln residue replaces His89, which suggests the heme to protein interactions have been preserved (Fig. 2).<sup>14, 58, 60, 61</sup> The Ile82 residue aligns structurally with Thr101 in CYP101A1, which extends into the active site and presumably plays a role in substrate binding (Fig. 3c). This residue varies among the CYP109 enzymes but is always an aliphatic hydrophobic residue (Ile, Leu or Val) while it is also a Thr

residue in CYP106A2.

### *The active site*

The first tier of the active site consists of predominantly hydrophobic residues including Ile82, Ala239, Pro285 and Leu289 (Fig. 3c). Ala239 and Pro285 are well conserved across the CYP109 family while Ile82 (*vide supra*) and Leu289 show more diversity. Leu289 varies from Val, Leu, Ile and Met residues in the CYP109 enzymes. Thr243, of the acid-alcohol pair, is also found in this tier and all of these residues are within 4 Å of the heme group (Fig. 3c). As part of the acid-alcohol pair, Thr243 is hypothesised to be involved in facilitating proton delivery to the bound dioxygen and stabilising the ferrous-oxy form, both of which are important in the cleavage of the O–O bond to generate the reactive intermediate.<sup>52, 62</sup> The acidic residue Glu242 points away from the active site in a similar fashion to that observed in the open conformations of other P450 systems. Glu242 hydrogen bonds directly to two water molecules and there is no obvious salt-bridge network as observed in the CYP101 family of enzymes (Fig. 3d).<sup>14, 63, 64</sup>

Asn83, which corresponds to His94 in CYP109D1 - the residue proposed to be important in ionone binding, sits in the tier of residues above those in the immediate vicinity of the heme, with the side chain pointing into the active site (Fig. 3d). While this residue is also conserved in the CYP109B subfamily enzymes from *B. mojavensis* and *B. amyloliquefaciens* and CYP278A1 it varies in more distant CYP109 relatives. The corresponding amino acid is a Gly residue in the CYP109B subfamily member for *B. cereus*, a Thr residue in CYP109A and a Phe residue in both CYP109C1 and CYP109C2 (Fig. 2).

Leu235 of the I-helix forms part of the second tier of residues which point into the active site (Fig. 3d). This corresponds to Leu244 of CYP101A1 and this leucine is well conserved across the CYP109B subfamily (Fig. 4a); however, it varies in the other CYP109 subfamilies; CYP109C1

(Phe), CYP109D1 (Gly) and in CYP278A1 (Thr) (Fig. 2). Val238 of CYP109B1, equivalent to Val247 in CYP101A1, would also be expected to be part of this second tier of active site residues. However, in the open structure it points away from the active site towards the access channel in a similar position to Val247 and Leu253 in the open structures of CYP101A1 and CYP101D2, respectively.<sup>58,59</sup> This residue is well conserved across the P450 enzymes discussed here with only CYP109C1 showing any variation (Ala).

The residues either side of Ala286, in SRS5, are both proline residues. One, Pro285 as noted earlier, is conserved across the CYP109 family while the other, Pro287, is only conserved in CYP109B1 and the closely related subfamily members from *B. mojavensis* and *B. amyloliquefaciens*. In the other enzymes the corresponding residue is a Gln (Ile and residues in CYP106A2 and CYP101A1, respectively, Fig. 2 and 3d). Proline residues are known to cause structural changes in the surrounding region and the orientation of the loop comprising these residues is different to the equivalent loops in the open conformations of CYP101A1 and CYP101D2 (Fig. 4b). A similar arrangement of two proline residues is found in P450Bm3 (CYP102A1), CYP102A2 and CYP102A3. In CYP102A1 the residues Pro326 and Pro329 are located in this region and the introduction of a third proline through the mutation Ala330Pro alters substrate selectivity and activity, virtually abolishing fatty acid binding;<sup>65</sup> therefore this region has been shown to be important in substrate binding. One effect of these proline residues in CYP109B1 is that the subsequent  $\beta$ -sheet ( $\beta$ 3-1) is longer than those of CYP101A1 and CYP101D2 and there are differences in the orientation of side chains of subsequent residues. Ala286 of CYP109B1 aligns structurally with Val295 of CYP101A1 and both point into the active site (Fig. 4b). Val288, two residues to the C-terminus of Ala286, points away from the active site while the acidic residue in the same position in the amino acid sequence of CYP101A1, Asp297 (and in CYP101D2, Glu303)

points towards the heme and interacts with the heme A-ring propionate. Val288 of CYP109B1 therefore aligns structurally with Ala296 of CYP101A1. Consequently Leu289 of CYP109B1 is structurally the equivalent of Asp297 of CYP101A1 (and Glu303 of CYP101D2) and forms part of the active site (Fig. 4b and Table S3). From this point onwards the secondary structure forms a  $\beta$ -sheet. Arg299 of CYP101A1 and the equivalent residue in CYP101D2 (Arg305), points toward and forms a salt bridge with the A-ring propionate,<sup>58</sup> while Arg290 in CYP109B1, which structurally aligns with Gly298 in CYP101A1 (Ala304 in CYP101D2), points away from the heme (Fig. 4b). The following residue in CYP109B1 is Arg291 and its side chain points toward the A-ring propionate with which it forms a salt bridge (Fig. 3d). The alignment shows that this double Arg arrangement (Arg290 and Arg291) is only conserved in the three CYP109 subfamily members that contain the second proline residue (Pro287) (Fig. 2). These modifications in primary structure result in substantial changes in the secondary structure. These may not be picked up by sequence alignment alone and as a result the crystal structure is a very useful tool to understand substrate binding and to facilitate the design of more efficient catalysts through protein engineering.

Val385 and Ile386, of SRS6, are located high in the active site in the loop region in between the last  $\beta$ -sheet (Fig. 3d). These residues are in the final substrate recognition site of P450 enzymes and correspond to Ile395 and Val396 of CYP101A1; their location in the open structure of CYP109B1 is similar to that in CYP101A1 (and the Ile401 and Val402 equivalents in the open structure of CYP101D2). In the closed form of the enzyme these would be expected to take up similar positions to those in the CYP101 family in the second tier of residues in the active site with Val385 being closer to the heme. These two residues are poorly conserved across the CYP109 family members and CYP278A1, with residues including Phe, Thr, Met, Pro and His being found in these positions, as well as Val and Ile.

Due to the open structure and the different length of some of the helices and loops it is more difficult to assess which residues of the F/G loop might be important in substrate binding in the closed form of the enzyme. Val166, which is conserved in all the CYP109B subfamily members, and aligns structurally with Thr185 in the open form of CYP101A1, is located at the end of the F-helix (Fig. 3d and 4a). The Val166 residue may interact with the substrate in a closed form of the enzyme in a similar fashion to Thr185. This residue is not conserved in CYP109D1 (Met), CYP109C2 (Leu) or CYP278A1 (Met).

It is unfortunate, albeit not uncommon, that one of the missing regions of the CYP109B1 structure corresponds to the region of the B/C loops where residues which have key interactions with the substrate are often located. For example this is the region where Phe87, Tyr96 and Phe98 of CYP101A1 reside. All of these residues of CYP101A1, in particular Tyr96, are important in substrate binding and recognition.<sup>61, 66, 67</sup> This region, corresponding to the residues between Met70 and Asn79, is extremely variable across the CYP109 family members and is likely to have an impact on the substrate preferences of these enzymes. Asn79 of CYP109B1, which corresponds to Phe98 of CYP101A1, and Met98 of CYP101D2, is the only one of the residues to be resolved in the structure and is located in a similar position to Phe98 and Met98 in the open forms of CYP101A1 and CYP101D2, respectively. This residue is very poorly conserved across the CYP109 family. The high mobility associated with this region in the absence of bound substrate and the resulting absence of resolvable electron density is not unusual in crystal structures of CYP enzymes. Attempts to crystallise the CYP109B1 enzyme in a closed conformation with a substrate to assist with resolving the electron density of all of the polypeptide chain have so far been unsuccessful. Both co-crystallisation and soaking experiments were undertaken with  $\beta$ -ionone and oleic acid and

while crystals were obtained they were unsuitable for X-ray diffraction analysis due to crystal twinning and low resolution.

Finally despite the relatively high overall sequence similarity of CYP106A2 with CYP109B1 the residues which constitute the active site, based on sequence alignment, show significant variation. For example Ile82, Asn83, Val166, Val238, Val385 and Ile386 in CYP109B1 align with Thr89, Glu90, Phe173, Gly242, Thr396 and Gly397 in CYP106A2 (Table S3). These differences in the substrate binding pocket of CYP106A2 are likely accentuated by the insertion of amino acids before SRS1 (Fig. 2) and the absence of proline residues in SRS5 (Fig. 4b), *vide supra*. However it is clear that the modified primary structure in these regions results in the different substrate preferences of the CYP106 family compared the CYP109 enzymes. The variability across the residues in the active site and access channel of the CYP109 enzymes provides a rationale for the different substrate binding properties of the various members of this family and the structure reported here provides information on the areas of the protein that could be targeted to alter the substrate selectivity. Overall, the results indicate that the subfamilies of the CYP109 family may have diverse roles in the bacterial strains in which they are found.

### **Turnover of $\alpha$ - and $\beta$ -ionone**

CYP109B1 has been reported to oxidise  $\alpha$ -ionone and  $\beta$ -ionone regioselectively at the allylic C3 and C4 positions, respectively.<sup>20</sup> Saturated fatty acids bind with larger shifts in the spin state to high spin than the norisoprenoids, however binding is weaker. Unsaturated fatty acids bind more tightly but with lower spin state shifts. The oxidation of fatty acids by CYP109B1 was not selective with five products being observed ( $\omega$ -1 to  $\omega$ -6) with the unsaturated fatty acids reported to produce greater than 20 products.<sup>20</sup> The crystal structure of CYP109B1 provides a few clues to the nature of fatty acid binding. The entrance to the substrate access channel has a positively charged patch due

to Arg290 and Pro287 which may facilitate substrate entry and binding (Fig. S2). There are other negatively charged residues present but analysis is hampered by missing residue electron density especially in regions important for substrate binding and recognition, for example Met70 to Gly78 of the B/B' loop.

As yet, the natural CYP electron transfer partners from *B. subtilis* are not known. We recently identified a PFOR from *Pseudomonas putida* KT24440 which was able to act as an electron transfer carrier to CYP238A1 from the same bacterium.<sup>10</sup> The PFOR protein was similar to the electron transfer system of Class IV self-sufficient fusion systems from *Rhodococcus* sp. (P450Rhf), *Cupriavidus metallidurans* and *Rhodococcus ruber*.<sup>68-70</sup> The PFOR domain of the *Rhodococcus* system has been used to reconstitute the turnover of many P450s by creating artificial fusion proteins.<sup>71-73</sup> We wanted to assess if the PFOR protein from *P. putida* was capable of acting as a promiscuous electron transfer partner to other bacterial P450 enzymes without the need to fuse the two together. Therefore we created a whole-cell oxidation system by cloning the genes for the PFOR protein and CYP109B1 into the pETDuet vector so each gene is downstream of its own promoter. By producing both proteins in *E. coli* and adding  $\alpha$ - and  $\beta$ -ionone as substrates we observed products for each norisoprenoid by GC-MS or HPLC (Fig. 5).

$\beta$ -Ionone oxidation resulted in a single product (Fig. 5) but  $\alpha$ -ionone resulted in a mixture of products (data not shown). Scaling up the reaction we were able to purify sufficient product by silica gel chromatography for characterisation by NMR. The products of oxidation were isolated (2-10 mg of each) and characterised as 4-hydroxy- $\beta$ -ionone and a mixture of *cis*- and *trans*-3-hydroxy- $\alpha$ -ionone (See experimental and Fig. S5).

We were also able to generate products from  $\beta$ -ionone using an *in vitro* turnover system consisting of CYP109B1, PFOR, yeast alcohol dehydrogenase and NADH to show that the PFOR

was responsible for supporting the catalytic cycle of CYP109B1 (Fig. 5): this shows that the PFOR electron transfer partner can drive the activity of CYP109B1. In addition this is the first example of this PFOR electron transfer partner being capable of supporting electron transfer to a non-physiological CYP enzyme. Combining the use of a simple whole-cell oxidation system for generating metabolites from P450 oxidations and the ability of the PFOR to transfer electrons to support other bacterial CYP enzymes will be generally applicable to other P450 systems where the electron transfer partners are unknown.

CYP109B1 has previously been shown to be functional when using adrenodoxin as an electron transfer partner: Adrenodoxin is predominantly negatively charged around the iron-sulphur cluster.<sup>14, 74</sup> The electrostatic potential of the CYP109B1 proximal face is mainly positive in the vicinity of the heme. This is due to the side-chains of Lys57, Lys63, Pro87, Arg103, Lys106, Lys223, Arg295 and Lys323 (Fig. 6). While some of these residues are conserved among the closely related CYP109 family members (Fig. 2) there is no obvious pattern. This is perhaps not surprising given that the electron transfer pathways in bacterial species may vary considerably.

The structure of the PFOR is not available with the only similar structurally characterised enzyme being that from *Pseudomonas cepacia* (now *Burkholderia cepacia*, PDB 2PIA).<sup>75</sup> These proteins share 42% sequence identity (53% similarity) and the key residues around the iron-sulfur cluster are conserved though there are modifications in the residues around the coordinating cysteines (Fig. S6). The electrostatic potential of the surface of this protein around the [2Fe-2S] iron-sulfur cluster is neutral though there are three major regions of negative charge which surround the cluster (Fig. 6b). The residues that contribute to these three negatively charged regions are Ser58, Glu104 and Glu223; Asp320 and the C-terminus carboxylate group and Asp298, Asp299 and Glu300. There are also regions of positive charge on the surface one due to Arg267 and another which is further

away from the cluster from Lys109 and Arg110. A comparison of the sequences of PP1957 with the PFOR from *P. cepacia* (2PIA) highlights that the majority of the negatively charged residues are conserved (Fig. S6 and Table S4). The exceptions are Glu104 which is modified to a His residue and Asp299 which is replaced by a Ser residue. The positively charged residues are not conserved with Lys106, Arg107 and Arg267 being replaced with Asp, Ala and Glu residues, respectively. Arg297 and Arg244 which are close to two of the negatively charged patches (next to Asp298 and Asp320, respectively) are both modified to Ser residues in PP1957 (Table S4). Overall this suggest that the negatively charged surface around the iron-sulfur cluster is maintained in PP1957, in agreement with its lower pI compared to 2PIA (4.9 vs. 5.6), which may in part explain why it is an effective electron transfer partner for CYP109B1.

## Experimental

### General

General DNA and microbiological experiments were carried out using standard methods.<sup>76</sup> The KOD polymerase, used for the PCR steps, and pET28a and pETDuet expression vectors were from Merck-Millipore. T4 DNA ligase and restriction enzymes for molecular biology were from New England Biolabs. General reagents and organic compounds were from Sigma-Aldrich or VWR, Australia. Isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG, Scimar, Australia), growth media and buffer components were from Melford Laboratories, UK, or VWR, Australia and NADH was from Roche Diagnostics, UK or Astral Scientific, Australia.

### Cloning, expression and purification

The gene encoding full-length *CYP109B1* was cloned into the vector pET28a(+) using the following primers which contain the restriction sites BamHI and Sall; CYP109B1 5', 5'-ggagtgaggatccatgaatgtgttaaaccgccg-3' and CYP109B1 3', 5'-agtctcctgcgacttacattttcacacggaagc. This construct adds the sequence MGSSHHHHHSSGLVPRGSHMASMTGGQQMGRGS to the N-terminus of target protein which contains a His tag for Ni<sup>2+</sup>-chelating affinity chromatography. The sequences of all genes were confirmed by DNA sequencing (Australian Genome Research Facility, Adelaide).

The recombinant plasmid was transformed into *Escherichia coli* strain BL21(DE3) and the transformed cells were cultured in 2xYT medium at 37 °C with 30  $\mu\text{g mL}^{-1}$  kanamycin. When the OD<sub>600</sub> of the culture reached 0.6 – 0.8, 0.5 mM IPTG was added in order to induce the production of the recombinant protein. After further growth for 18 hours at 28 °C, cells were harvested by centrifugation, resuspended in buffer P (1xPhosphate Buffered Saline, 10 mM  $\beta$ -mercaptoethanol, pH 7.4) and lysed by sonication on ice. The crude extracts were then centrifuged at 27000 g for 45

min at 4 °C to remove the cell debris.

The supernatant containing the protein was loaded onto a Ni<sup>2+</sup>-chelating affinity column (1.0 mL of Ni<sup>2+</sup>-NTA agarose) which had been equilibrated with buffer P, and washed twice with the same buffer and then subsequently with buffer P containing 10 mM imidazole (5 column volumes). CYP109B1 was eluted with buffer P containing 300 mM imidazole. The eluent was buffer-exchanged into buffer A (20 mM HEPES, 10 mM β-mercaptoethanol, pH 7.4) by ultrafiltration. The protein solution was concentrated to 1 mL in buffer A and injected onto a Resource Q column (GE Healthcare, USA) and eluted with a gradient of 0 to 1 M KCl in buffer A. CYP109B1 eluted at *ca.* 100 mM KCl as a single peak, which was collected and concentrated by ultrafiltration and then applied to a Superdex-75 size-exclusion chromatography column (GE Healthcare, USA) equilibrated in buffer B (buffer A with 150 mM KCl) for further purification. The purity of CYP109B1 was estimated to be greater than 95% by SDS-PAGE Gel analysis (Fig. S7).

### Crystallisation

The concentration of CYP109B1 was calculated using the extinction coefficient at 418 nm ( $\epsilon_{418}$  128 mM<sup>-1</sup> cm<sup>-1</sup>).<sup>20</sup> Purified CYP109B1 was concentrated in crystallisation buffer (200 mM KH<sub>2</sub>PO<sub>4</sub>, 200 mM KCl, 10 mM β-mercaptoethanol, pH 7.4) to 50 mg mL<sup>-1</sup>. The sitting-drop vapour-diffusion method was used with Crystal Screen, Crystal Screen 2 and Index (Hampton Research, USA) to screen for CYP109B1 crystals at 16 °C in 48-well plates (Tianjin Xiangyushun Macromolecule Technology Ltd, China). Two drops (1 μL each) of the protein solution, with 30 or 50 mg mL<sup>-1</sup> protein, were 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 Crystal Screen kit 2 condition No.7 (10% w/v polyethylene glycol 1000, 10% w/v polyethylene glycol 8000). However, none of the crystals were suitable for X-ray diffraction analysis. Further optimisation was

undertaken by varying the concentration of the two polyethylene glycol (PEG) precipitants, the protein concentration, volume ratio of protein to reservoir solution and additive screening. Finally, the single crystals (0.01×0.05×0.4 mm) which gave high resolution data were grown under optimised conditions consisting of 16% w/v PEG 1000, 10% w/v PEG 8000, 0.625% w/v PEG 3350, with 30 mg mL<sup>-1</sup> protein concentration and the ratio of the volume of protein to reservoir solution was 2:1.

### Crystallographic data collection and processing

X-ray diffraction data were collected to 1.8 Å resolution from CYP109B1 crystals flash-cooled at 100 K in a nitrogen gas stream. Data were collected on a Rigaku R-Axis HTC image plate using Cu K $\alpha$  radiation from an in-house Rigaku MicroMax-007 rotating-anode X-ray generator operating at 40 kV and 30 mA. No cryoprotectant was used. The intensity set was indexed, integrated and scaled with the HKL-2000 package.<sup>77</sup> The crystals belonged to space group P2<sub>1</sub>, with unit-cell parameters  $a = 54.00$ ,  $b = 67.57$ ,  $c = 56.41$  Å and  $\beta = 113.04$  °. There was one molecule in each asymmetric unit and the Matthews' coefficient was 2.10 with 41.6% solvent content. The complete data collection results and statistics are summarised in Table 2.

### Structure determination and refinement

For the structure determination of CYP109B1, the molecular-replacement (MR) program Phaser in the CCP4 suite<sup>78, 79</sup> was used to find the phases. The structure of CYP107L1 (PDB:2WHW), PikC, a macrolide hydroxylating P450, which has 38% sequence identity, was used as a search model.<sup>73</sup> The model was rebuilt with Coot<sup>80</sup> and refined with REFMAC5 and Phenix.refine.<sup>81, 82</sup> The stereochemical quality of the refined structure was checked with the program MolProbity.<sup>83</sup> A summary of the structure refinement statistics is provided in Table 2. The coordinates of the crystal structure have been deposited in the PDB (<http://www.pdb.org>; with accession code PDB: 4RM4).<sup>84</sup>

### Growth and expression of the *in vivo* system

The *PP1957* gene encoding the PFOR was previously cloned into the pETDuet vector using the *NcoI* (**ccatgg**) and *EcoRI* (**gaattc**) restriction sites.<sup>10</sup> A whole-cell oxidation system was generated by using the following primers to amplify the CYP109B1 gene. CYP109B1 5', 5'-ttaattcatatggctagcatgactggtggacag-3' CYP109B1 3', 5'-ttaattgtaccctattacattttcacacggaagctctttaatc-3'. The PCR product was cut using the *NdeI* and *KpnI* restriction enzymes and ligated into the pETDuetPP1957 vector cut with the same two enzymes to generate the pETDuetPP1957/CYP109B1 vector. The incorporation of the whole gene was confirmed by restriction enzyme digest and DNA sequencing.

The plasmid pETDuetPP1957/CYP109B1 was transformed into competent BL21(DE3) cells and grown on LB plates containing ampicillin, 100  $\mu\text{g mL}^{-1}$  ( $\text{LB}_{\text{amp}}$ ). A single colony was inoculated into 500 mL broth ( $2\times\text{YT}_{\text{amp}}$ ) in a 2 L flask and grown at 37 °C overnight. Protein expression was induced by the addition of 100  $\mu\text{M}$  IPTG (from a 0.4 M stock in  $\text{H}_2\text{O}$ ) and the temperature was reduced to 25 °C and the shaker speed to 140 rpm. The growths were allowed to continue for another 24 hours before the cell pellet was harvested by centrifugation and washed in *E. coli* minimal media (EMM).<sup>85</sup> The cell pellet was resuspended in an equal volume of antibiotic containing media ( $\text{EMM}_{\text{amp}}$ ). The ionone substrates were added to the resuspended cells to a concentration of 2 mM (from a 0.5 M stock in EtOH) and the reactions were then shaken at 220 rpm and 30 °C. The samples were analysed by HPLC or GC-MS. HPLC was performed using an Agilent 1260 Infinity pump equipped with an Agilent Eclipse Plus C18 column (250 mm x 4.6 mm, 5  $\mu\text{m}$ ) and an autoinjector. A gradient, 20 - 95%, of acetonitrile (with trifluoroacetic acid, 0.1%) in water (TFA, 0.1%) was developed over 30 min using a mobile phase of water/acetonitrile (0.1% trifluoroacetic acid) at a flow rate of 1 ml  $\text{min}^{-1}$ . GC-MS was performed on a Shimadzu GC-17A

with a DB-5 MS fused silica column (30 m x 0.25 mm, 0.25  $\mu$ m) and a QP5050A GC-MS detector. The injector was held at 250 °C and the interface at 280 °C. For the ionone substrates, the initial oven temperature was 120 °C which was held for 3 min, before increasing to 220 °C at 10 °C per minute, where it was held for 7 min. The GC (HPLC) retention times were as follows:  $\beta$ -ionone, 7.5 min (25 min); 4-hydroxy- $\beta$ -ionone, 9.9 min (14.7 min);  $\alpha$ -ionone, 6.7 min; 3-hydroxy- $\alpha$ -ionone, 9.4 min.

For product isolation the cells were removed by centrifugation and supernatant was extracted using EtOAc. The organic layer was separated and the extraction of the aqueous phase was repeated with EtOAc (2 x 100 mL). The combined organic layers were washed with brine (3 x 100 mL), then dried over MgSO<sub>4</sub>, filtered and concentrated in *vacuo*. The resultant oil, a mixture of substrate and product, was purified using silica gel flash chromatography (hexane/EtOAc, 2:1). The products were then isolated and characterised by NMR as described below.

#### $\alpha$ -Ionone and $\beta$ -ionone oxidation products

After extraction and silica gel chromatography the hydroxylated products were isolated (2-10 mg of each) and characterised using <sup>1</sup>H, <sup>13</sup>C, COSY, HMBC and HSQC NMR techniques using a Varian Inova-600 spectrometer (Fig. S6). The oxidation of  $\alpha$ -ionone resulted in a mixture of products. The major product was assigned by NMR as *trans*-3-hydroxy-ionone. The minor product was not isolated but the minor signals in the product NMR spectra matched those of *cis*-3-hydroxy-ionone.

Data for **4-hydroxy- $\beta$ -ionone**: <sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>);<sup>60, 86</sup>  $\delta$  7.19 (d,  $J$  = 16.4 Hz, 1H, H8), 6.12 (d,  $J$  = 16.4 Hz, 1H, H7), 4.02 (t,  $J$  = 4.9 Hz, 1H, H4), 2.31 (s, 3H, H10), 1.93 (ddd,  $J$  = 14.0, 7.4, 3.2 Hz, 1H, H3), 1.85 (s, 3H, H13), 1.73 (ddd,  $J$  = 13.0, 7.6, 3.9 Hz, 1H, H3), 1.69–1.63 (m, 1H, H2), 1.46 (ddd,  $J$  = 11.3, 7.3, 2.7 Hz, 1H, H2), 1.07 (s, 3H, H12), 1.04 (s, 3H, H11); <sup>13</sup>C NMR (150 MHz, CDCl<sub>3</sub>)  $\delta$  198.38 (C9), 142.63 (C8), 139.44 (C6), 133.88 (C5), 133.11 (C7), 69.94 (C4),

34.65 (C2), 28.85 (C12), 28.31 (C3), 27.49 (C11), 27.38 (C10), 18.48 (C13).

Data for ***trans*-3-hydroxy- $\alpha$ -ionone**:  $^1\text{H}$  NMR (600 MHz,  $\text{CDCl}_3$ );  $\delta$  6.54 (dd,  $J = 15.8, 10.1$  Hz, 1H, H7), 6.10 (d,  $J = 15.5$  Hz, 1H, H8), 5.63 (s, 1H, H4), 4.27 (s, 1H, H3), 2.50 (d,  $J = 10.1$  Hz, 1H, H6), 2.26 (s, 3H, H10), 1.84 (dd,  $J = 13.7, 5.6$  Hz, 1H, H2), 1.62 (s, 3H, H13), 1.41 (dd,  $J = 13.9, 5.8$  Hz, 1H, H2), 1.03 (s, 3H, H11), 0.89 (s, 3H, H12);  $^{13}\text{C}$  NMR (150 MHz,  $\text{CDCl}_3$ )  $\delta$  198.03 (C9), 147.10 (C7), 135.44(C5), 133.61 (C8), 125.82 (C4), 65.47 (C3), 54.30 (C6), 43.83 (C2), 33.86 (C1), 29.31 (C11), 27.20 (C10), 24.70 (C12), 22.66 (C13).

### ***Phylogenetic Analysis***

The data was taken from the CYP450 engineering database (<http://www.cyped.uni-stuttgart.de/>) or the databases at the National Center for Biotechnology Information (NCBI). The phenogram (Fig. 1) was generated at PhyloDendron (<http://iubio.bio.indiana.edu/treeapp/treeprint-form.html>). BLAST searches were performed using the databases at the NCBI. Sequence alignments were performed using ClustalW and ESPript.<sup>87, 88</sup> The substrate recognition sites of CYP109B1 were defined by alignment with those of other P450s, including CYP101A1, by Gotoh as follows; SRS1, 77-84; SRS2 and 3, 168-181; SRS4, 228-245; SRS5, 282-294 and SRS6 380-387.

## Conclusion

The structure of CYP109B1, a member of a bacterial family of cytochrome P450 enzymes with potential roles in biocatalysis, has been solved and this provides important information concerning residues that are involved in the enzyme-substrate interactions in the CYP109 family. The number of CYP109 enzymes identified which have unknown function and the high variability in the B/B' region may indicate varied substrate specificity across this diverse family. In addition, we have shown that CYP109B1 can accept electrons from a phthalate family oxygenase reductase both *in vitro* and *in vivo*. This will allow the facile identification of the oxidation products generated by CYP109 family members. Together, these results will facilitate the engineering this family of cytochrome P450 enzymes for more efficient and selective substrate turnover, which will in turn enable the development of biocatalytic routes to as yet unobtainable fine chemicals using CYP109 family members.

## Acknowledgements

E.A.H is grateful to the University of Adelaide for a M. Phil Scholarship. M.J.C. is grateful for the support of the Deutsche Forschungsgemeinschaft (Emmy–Noether Program, CR 392/1-1). S.G.B and M.J.C acknowledge the Group of Eight Australia and the Deutsche Akademischer Austausch Dienst for Group of Eight, Australia – Germany Joint Research Co-operation Scheme grants; (grant 56265933 to M.J.C.).

## References

1. P. R. Ortiz de Montellano, ed., *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 3rd edn., Kluwer Academic/Plenum Publishers, New York, 2005.
2. A. Sigel, H. Sigel and R. Sigel, *The Ubiquitous Roles of Cytochrome P450 Proteins*, 1st edn., John Wiley & Sons, 2007.
3. M. J. Cryle, J. E. Stok and J. J. De Voss, *Aust. J. Chem.*, 2003, **56**, 749-762.
4. R. Agudo, G. D. Roiban and M. T. Reetz, *Chembiochem*, 2012, **13**, 1465-1473.
5. S. G. Bell and L. L. Wong, *Biochem. Biophys. Res. Commun.*, 2007, **360**, 666-672.
6. S. Q. Pham, P. Gao and Z. Li, *Biotechnol. Bioeng.*, 2013, **110**, 363-373.
7. H. Venkataraman, S. B. A. de Beer, D. P. Geerke, N. P. E. Vermeulen and J. N. M. Commandeur, *Adv. Synth. Catal.*, 2012, **354**, 2172-2184.
8. C. J. Whitehouse, S. G. Bell and L. L. Wong, *Chem. Soc. Rev.*, 2012, **41**, 1218-1260.
9. F. Hannemann, A. Bichet, K. M. Ewen and R. Bernhardt, *Biochim. Biophys. Acta*, 2007, **1770**, 330-344.
10. S. G. Bell, L. French, N. H. Rees, S. S. Cheng, G. Preston and L.-L. Wong, *Biotechnol. Appl. Biochem.*, 2013, **60**, 9-17.
11. D. B. Hawkes, K. E. Slessor, P. V. Bernhardt and J. J. De Voss, *Chembiochem*, 2010, **11**, 1107-1114.
12. F. Xu, S. G. Bell, Y. Peng, E. O. Johnson, M. Bartlam, Z. Rao and L. L. Wong, *Proteins*, 2009, **77**, 867-880.
13. S. G. Bell, J. H. McMillan, J. A. Yorke, E. Kavanagh, E. O. Johnson and L. L. Wong, *Chem. Commun.*, 2012, **48**, 11692-11694.
14. W. Yang, S. G. Bell, H. Wang, W. Zhou, N. Hoskins, A. Dale, M. Bartlam, L. L. Wong and Z. Rao, *Journal of Biological Chemistry*, 2010, **285**, 27372-27384.
15. F. Hannemann, C. Virus and R. Bernhardt, *J. Biotechnol.*, 2006, **124**, 172-181.
16. D. R. Nelson, *Arch. Biochem. Biophys.*, 1999, **369**, 1-10.
17. D. R. Zeigler, *Microbiology (Reading, England)*, 2011, **157**, 2033-2041.
18. T. Furuya, T. Nishi, D. Shibata, H. Suzuki, D. Ohta and K. Kino, *Chem. Biol.*, 2008, **15**, 563-572.
19. T. Furuya, D. Shibata and K. Kino, *Steroids*, 2009, **74**, 906-912.
20. M. Girhard, T. Klaus, Y. Khatri, R. Bernhardt and V. B. Urlacher, *Appl. Microbiol. Biotechnol.*, 2010, **87**, 595-607.
21. M. Girhard, K. Machida, M. Itoh, R. D. Schmid, A. Arisawa and V. B. Urlacher, *Microb. Cell. Fact.*, 2009, **8**, 36.
22. Y. Khatri, F. Hannemann, M. Girhard, R. Kappl, A. Meme, M. Ringle, S. Janocha, E. Leize-Wagner, V. B. Urlacher and R. Bernhardt, *Biotechnol. Appl. Biochem.*, 2013, **60**, 18-29.
23. Y. Khatri, M. Girhard, A. Romankiewicz, M. Ringle, F. Hannemann, V. B. Urlacher, M. C. Hutter and R. Bernhardt, *Appl. Microbiol. Biotechnol.*, 2010, **88**, 485-495.
24. Y. Khatri, F. Hannemann, K. M. Ewen, D. Pistorius, O. Perlova, N. Kagawa, A. O. Brachmann, R. Muller and R. Bernhardt, *Chem. Biol.*, 2010, **17**, 1295-1305.
25. P. Winterhalter and R. Rouseff, in *Carotenoid-Derived Aroma Compounds*, American Chemical Society, 2001, vol. 802, pp. 1-17.
26. M. Fischer, M. Knoll, D. Sirim, F. Wagner, S. Funke and J. Pleiss, *Bioinformatics (Oxford,*

- England), 2007, **23**, 2015-2017.
27. S. F. Altschul, W. Gish, W. Miller, E. W. Myers and D. J. Lipman, *J. Mol. Biol.*, 1990, **215**, 403-410.
  28. R. Ding, Y. Li, C. Qian and X. Wu, *J. Bacteriol.*, 2011, **193**, 4537.
  29. M. Djukic, A. Poehlein, A. Thurmer and R. Daniel, *J. Bacteriol.*, 2011, **193**, 5535-5536.
  30. A. Pal, U. K. Mondal, S. Mukhopadhyay and A. K. Bothra, *Bioinformation*, 2011, **5**, 446-454.
  31. Y. J. Chang, M. Land, L. Hauser, O. Chertkov, T. G. Del Rio, M. Nolan, A. Copeland, H. Tice, J. F. Cheng, S. Lucas, C. Han, L. Goodwin, S. Pitluck, N. Ivanova, G. Ovchinnikova, A. Pati, A. Chen, K. Palaniappan, K. Mavromatis, K. Liolios, T. Brettin, A. Fiebig, M. Rohde, B. Abt, M. Goker, J. C. Detter, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and A. Lapidus, *Stand. Genomic Sci.*, 2011, **5**, 97-111.
  32. H. Kiss, M. Nett, N. Domin, K. Martin, J. A. Maresca, A. Copeland, A. Lapidus, S. Lucas, K. W. Berry, T. Glavina Del Rio, E. Dalin, H. Tice, S. Pitluck, P. Richardson, D. Bruce, L. Goodwin, C. Han, J. C. Detter, J. Schmutz, T. Brettin, M. Land, L. Hauser, N. C. Kyrpides, N. Ivanova, M. Goker, T. Woyke, H. P. Klenk and D. A. Bryant, *Stand. Genomic Sci.*, 2011, **5**, 356-370.
  33. N. A. Paramonov, L. A. Parolis, H. Parolis, I. F. Boan, J. Anton and F. Rodriguez-Valera, *Carbohydr. Res.*, 1998, **309**, 89-94.
  34. S. Bleif, F. Hannemann, M. Lisurek, J. P. von Kries, J. Zapp, M. Dietzen, I. Antes and R. Bernhardt, *Chembiochem*, 2011, **12**, 576-582.
  35. E. Brill, F. Hannemann, J. Zapp, G. Bruning, J. Jauch and R. Bernhardt, *Appl. Microbiol. Biotechnol.*, 2014, **98**, 1701-1717.
  36. D. Zehentgruber, F. Hannemann, S. Bleif, R. Bernhardt and S. Lutz, *Chembiochem*, 2010, **11**, 713-721.
  37. T. P. Stinear, T. Seemann, P. F. Harrison, G. A. Jenkin, J. K. Davies, P. D. Johnson, Z. Abdallah, C. Arrowsmith, T. Chillingworth, C. Churcher, K. Clarke, A. Cronin, P. Davis, I. Goodhead, N. Holroyd, K. Jagels, A. Lord, S. Moule, K. Mungall, H. Norbertczak, M. A. Quail, E. Rabinowitsch, D. Walker, B. White, S. Whitehead, P. L. Small, R. Brosch, L. Ramakrishnan, M. A. Fischbach, J. Parkhill and S. T. Cole, *Genome Res.*, 2008, **18**, 729-741.
  38. M. J. Cryle, S. G. Bell and I. Schlichting, *Biochemistry*, 2010, **49**, 7282-7296.
  39. M. J. Cryle and I. Schlichting, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 15696-15701.
  40. M. C. Gustafsson, O. Roitel, K. R. Marshall, M. A. Noble, S. K. Chapman, A. Pessegueiro, A. J. Fulco, M. R. Cheesman, C. von Wachenfeldt and A. W. Munro, *Biochemistry*, 2004, **43**, 5474-5487.
  41. I. Matsunaga, A. Ueda, N. Fujiwara, T. Sumimoto and K. Ichihara, *Lipids*, 1999, **34**, 841-846.
  42. J. J. Reddick, S. A. Antolak and G. M. Raner, *Biochem. Biophys. Res. Commun.*, 2007, **358**, 363-367.
  43. O. Shoji, T. Fujishiro, S. Nagano, S. Tanaka, T. Hirose, Y. Shiro and Y. Watanabe, *J. Biol. Inorg. Chem.*, 2010, **15**, 1331-1339.
  44. J. E. Stok and J. De Voss, *Arch. Biochem. Biophys.*, 2000, **384**, 351-360.
  45. N. H. Youssef, N. Wofford and M. J. McInerney, *Int. J. Mol. Sci.*, 2011, **12**, 1767-1786.

46. A. M. Earl, M. Eppinger, W. F. Fricke, M. J. Rosovitz, D. A. Rasko, S. Daugherty, R. Losick, R. Kolter and J. Ravel, *J. Bacteriol.*, 2012, **194**, 2378-2379.
47. X. H. Chen, A. Koumoutsi, R. Scholz, A. Eisenreich, K. Schneider, I. Heinemeyer, B. Morgenstern, B. Voss, W. R. Hess, O. Reva, H. Junge, B. Voigt, P. R. Jungblut, J. Vater, R. Sussmuth, H. Liesegang, A. Strittmatter, G. Gottschalk and R. Borriss, *Nat. Biotechnol.*, 2007, **25**, 1007-1014.
48. X. H. Chen, A. Koumoutsi, R. Scholz, K. Schneider, J. Vater, R. Sussmuth, J. Piel and R. Borriss, *J. Biotechnol.*, 2009, **140**, 27-37.
49. D. A. Rasko, J. Ravel, O. A. Okstad, E. Helgason, R. Z. Cer, L. Jiang, K. A. Shores, D. E. Fouts, N. J. Tourasse, S. V. Angiuoli, J. Kolonay, W. C. Nelson, A. B. Kolsto, C. M. Fraser and T. D. Read, *Nucleic Acids Res.*, 2004, **32**, 977-988.
50. C. A. Hasemann, R. G. Kurumbail, S. S. Boddupalli, J. A. Peterson and J. Deisenhofer, *Structure (London, England : 1993)*, 1995, **3**, 41-62.
51. S. Rupasinghe, M. A. Schuler, N. Kagawa, H. Yuan, L. Lei, B. Zhao, S. L. Kelly, M. R. Waterman and D. C. Lamb, *FEBS Lett.*, 2006, **580**, 6338-6342.
52. T. L. Poulos, *Chem. Rev.*, 2014, **114**, 3919-3962.
53. T. W. B. Ost, C. S. Miles, A. W. Munro, J. Murdoch, G. A. Reid and S. K. Chapman, *Biochemistry*, 2001, **40**, 13421-13429.
54. T. Tosha, S. Yoshioka, K. Ishimori and I. Morishima, *Journal of Biological Chemistry*, 2004, **279**, 42836-42843.
55. S. Yoshioka, S. Takahashi, K. Ishimori and I. Morishima, *J. Inorg. Biochem.*, 2000, **81**, 141-151.
56. S. Yoshioka, T. Tosha, S. Takahashi, K. Ishimori, H. Hori and I. Morishima, *J. Am. Chem. Soc.*, 2002, **124**, 14571-14579.
57. O. Gotoh, *Journal of Biological Chemistry*, 1992, **267**, 83-90.
58. W. Yang, S. G. Bell, H. Wang, W. Zhou, M. Bartlam, L. L. Wong and Z. Rao, *Biochem. J.*, 2011, **433**, 85-93.
59. Y. T. Lee, R. F. Wilson, I. Rupniewski and D. B. Goodin, *Biochemistry*, 2010, **49**, 3412-3419.
60. M. Ma, S. G. Bell, W. Yang, Y. Hao, N. H. Rees, M. Bartlam, W. Zhou, L. L. Wong and Z. Rao, *ChemBioChem*, 2011, **12**, 88-99.
61. T. L. Poulos, B. C. Finzel and A. J. Howard, *J. Mol. Biol.*, 1987, **195**, 687-700.
62. J. Rittle and M. T. Green, *Science*, 2010, **330**, 933-937.
63. C. Di Primo, E. Deprez, S. G. Sligar and G. Hui Bon Hoa, *Biochemistry*, 1997, **36**, 112-118.
64. V. Lounnas and R. C. Wade, *Biochemistry*, 1997, **36**, 5402-5417.
65. C. J. Whitehouse, W. Yang, J. A. Yorke, B. C. Rowlatt, A. J. Strong, C. F. Blanford, S. G. Bell, M. Bartlam, L. L. Wong and Z. Rao, *ChemBioChem*, 2010, **11**, 2549-2556.
66. X. Chen, A. Christopher, J. P. Jones, S. G. Bell, Q. Guo, F. Xu, Z. Rao and L. L. Wong, *Journal of Biological Chemistry*, 2002, **277**, 37519-37526.
67. F. Xu, S. G. Bell, J. Lednik, A. Insley, Z. Rao and L.-L. Wong, *Angew. Chem., Int. Ed.*, 2005, **44**, 4029-4032.
68. R. De Mot and A. H. Parret, *Trends Microbiol.*, 2002, **10**, 502-508.
69. L. Liu, R. D. Schmid and V. B. Urlacher, *Biotechnol. Lett.*, 2010, **32**, 841-845.
70. A. J. Warman, J. W. Robinson, D. Luciakova, A. D. Lawrence, K. R. Marshall, M. J. Warren, M. R. Cheesman, S. E. Rigby, A. W. Munro and K. J. McLean, *FEBS J.*, 2012, **279**,

- 1675-1693.
71. A. Robin, G. A. Roberts, J. Kisch, F. Sabbadin, G. Grogan, N. Bruce, N. J. Turner and S. L. Flitsch, *Chem. Commun.*, 2009, 2478-2480.
  72. M. Nodate, M. Kubota and N. Misawa, *Appl. Microbiol. Biotechnol.*, 2006, **71**, 455-462.
  73. S. Li, M. R. Chaulagain, A. R. Knauff, L. M. Podust, J. Montgomery and D. H. Sherman, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 18463-18468.
  74. A. V. Grinberg, F. Hannemann, B. Schiffler, J. Muller, U. Heinemann and R. Bernhardt, *Proteins*, 2000, **40**, 590-612.
  75. C. C. Correll, C. J. Batie, D. P. Ballou and M. L. Ludwig, *Science*, 1992, **258**, 1604-1610.
  76. J. Sambrook, E. F. Fritsch and T. Maniatis, *Molecular Cloning: A Laboratory Manual*, 2<sup>nd</sup> edn., Cold Spring Harbor Laboratory Press, New York, 1989.
  77. Z. Otwinowski and W. Minor, in *International Tables for Crystallography Volume F: Macromolecular Crystallography*, M. G. Rossmann, Kluwer Academic Publishers, Dordrecht/Boston/London, 2001, pp. 226-235.
  78. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J. Appl. Cryst.*, 2007, **40**, 658-674.
  79. M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson, *Acta Crystallogr. D Biol. Crystallogr.*, 2011, **67**, 235-242.
  80. P. Emsley and K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.*, 2004, **60**, 2126-2132.
  81. G. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin, *Acta Crystallogr. D Biol. Crystallogr.*, 2011, **67**, 355-367.
  82. P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart and P. D. Adams, *Acta Crystallogr. D Biol. Crystallogr.*, 2012, **68**, 352-367.
  83. V. B. Chen, W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson, *Acta Crystallogr. D Biol. Crystallogr.*, 2010, **66**, 12-21.
  84. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235-242.
  85. S. G. Bell, C. F. Harford-Cross and L.-L. Wong, *Protein Eng.*, 2001, **14**, 797-802.
  86. S. Lutz-Wahl, P. Fischer, C. Schmidt-Dannert, W. Wohlleben, B. Hauer and R. D. Schmid, *Appl. Environ. Microbiol.*, 1998, **64**, 3878-3881.
  87. R. Chenna, H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins and J. D. Thompson, *Nucleic Acids Res.*, 2003, **31**, 3497-3500.
  88. X. Robert and P. Gouet, *Nucleic Acids Res.*, 2014, **42**, W320-324.

**Table 1** The sequence identities of CYP109B1 and selected other CYP109 family members with different CYP enzymes including CYP106 family members from other *Bacillus* strains (CYP106A1; *B. megaterium* DSM319 and CYP106A2; *B. megaterium* ATCC 13368 – P450meg) and CYP278A1 from *Mycobacterium marinum*.

<b>Protein; CYP109B1</b>	<b>Accession number</b>	<b>Identities</b>	<b>Positives</b>	<b>Gaps</b>	<b>Score</b>
CYP109 <i>B. mojavensis</i>	WP_010333858	360/396(91%)	382/396(96%)	0/396(0%)	758
CYP109 <i>B. amyloliquefaciens</i>	YP_001420823	253/386(66%)	309/386(80%)	0/386(0%)	545
CYP109 <i>B. cereus</i>	NP_979552	187/390(48%)	264/390(67%)	9/390(2%)	386
CYP109A1	YP_003866322	167/382(44%)	257/382(67%)	8/382(2%)	334
CYP109C1 <i>S. cellulosum</i>	YP_001610759	136/347(39%)	194/347(55%)	20/347(5%)	297
CYP109C2 <i>S. cellulosum</i>	YP_001619565	158/360(44%)	228/360(63%)	3/360(0%)	319
CYP109D1 <i>S. cellulosum</i>	WP_012237265	126/368(34%)	195/368(52%)	23/368(6%)	227
CYP106A1 <i>B. megaterium</i>	KGJ82431	151/372(41%)	220/372(59%)	20/372(5%)	274
CYP106A2 <i>B. megaterium</i>	Q06069	156/373(42%)	227/373(60%)	18/373(4%)	288
CYP278A1 <i>M. marinum</i>	WP_012394577	129/379(34%)	195/379(51%)	30/379(7%)	200
<b>CYP109C1</b>	YP_001610759	<b>Identities</b>	<b>Positives</b>	<b>Gaps</b>	<b>Score</b>
CYP109D1	WP_012237265	163/396(41%)	221/396(55%)	22/396(5%)	254
<b>CYP109D1</b>	WP_012237265	<b>Identities</b>	<b>Positives</b>	<b>Gaps</b>	<b>Score</b>
CYP278A1 <i>M. marinum</i>	WP_012394577	180/427(42%)	244/427(57%)	27/427(6%)	299

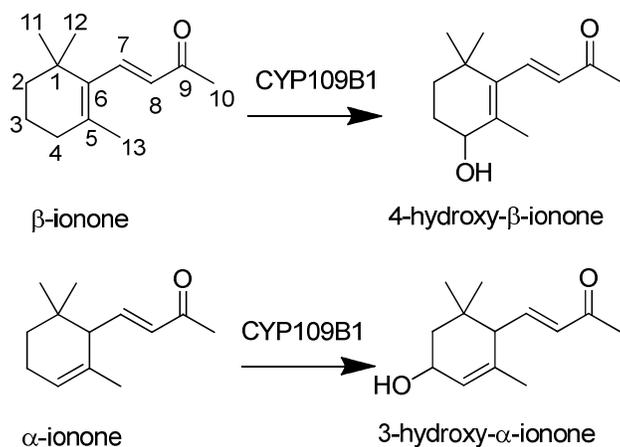
**Table 2** Data collection and refinement statistics of CYP109B1

<b>CYP109B1</b>	
<b>Data collection statistics</b>	
Space group	P2 <sub>1</sub>
Cell dimensions <i>a/b/c</i> (Å)	54.0/67.6/56.4
<i>α/β/γ</i> (°)	90.00/113.0/90.00
Wavelength (Å)	1.5418
Resolution (Å)	50-1.77(1.83-1.77)
Average <i>I</i> / <i>σ</i> ( <i>I</i> ) <sup>1</sup>	21.7(3.7)
Completeness (%) <sup>1</sup>	85.2(93.7)
Average Redundancy <sup>1</sup>	2.5 (2.3)
<i>R</i> <sub>merge</sub> (%) <sup>1,2</sup>	3.7 (22.9)
Molecules in one ASU	1
<b>Structure refinement statistics</b>	
Resolution (Å)	51.90-1.77
Average <i>B</i> -factor (Å <sup>2</sup> )	23.0
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%) <sup>3</sup>	20.2/25.2
r.m.s.d. bond lengths (Å)	0.007
r.m.s.d. bond angles (°)	1.07
Ramachandran favoured (%)	95.8
Ramachandran outliers (%)	0
MolProbity score	1.83
Poor rotamers (%)	0.6
Clash scores, all atoms	10.5

<sup>1</sup> Values in parentheses correspond to the highest-resolution shell.

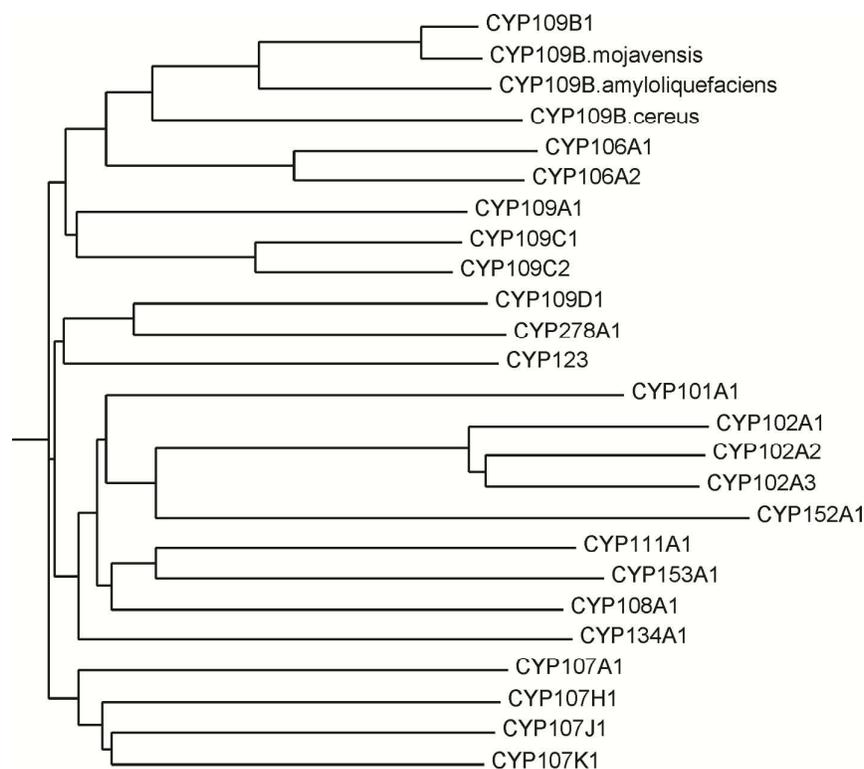
<sup>2</sup>  $R_{\text{merge}} = \frac{\sum_h \sum_i |I_i - \langle I \rangle|}{\sum_h \sum_i I_i}$ , where  $I_i$  is the *i*th intensity measurement of reflection *h*, and  $\langle I \rangle$  is the average intensity from multiple observations<sup>3</sup>  $R_{\text{work}}/R_{\text{free}} = \frac{\sum ||F_o| - |F_c||}{\sum |F_o|}$ , where  $F_o$  and  $F_c$  are the observed and calculated structure factors, respectively.

<sup>3</sup>  $R_{\text{work}}/R_{\text{free}} = \frac{\sum ||F_o| - |F_c||}{\sum |F_o|}$ , where  $F_o$  and  $F_c$  are the observed and calculated structure factors, respectively.

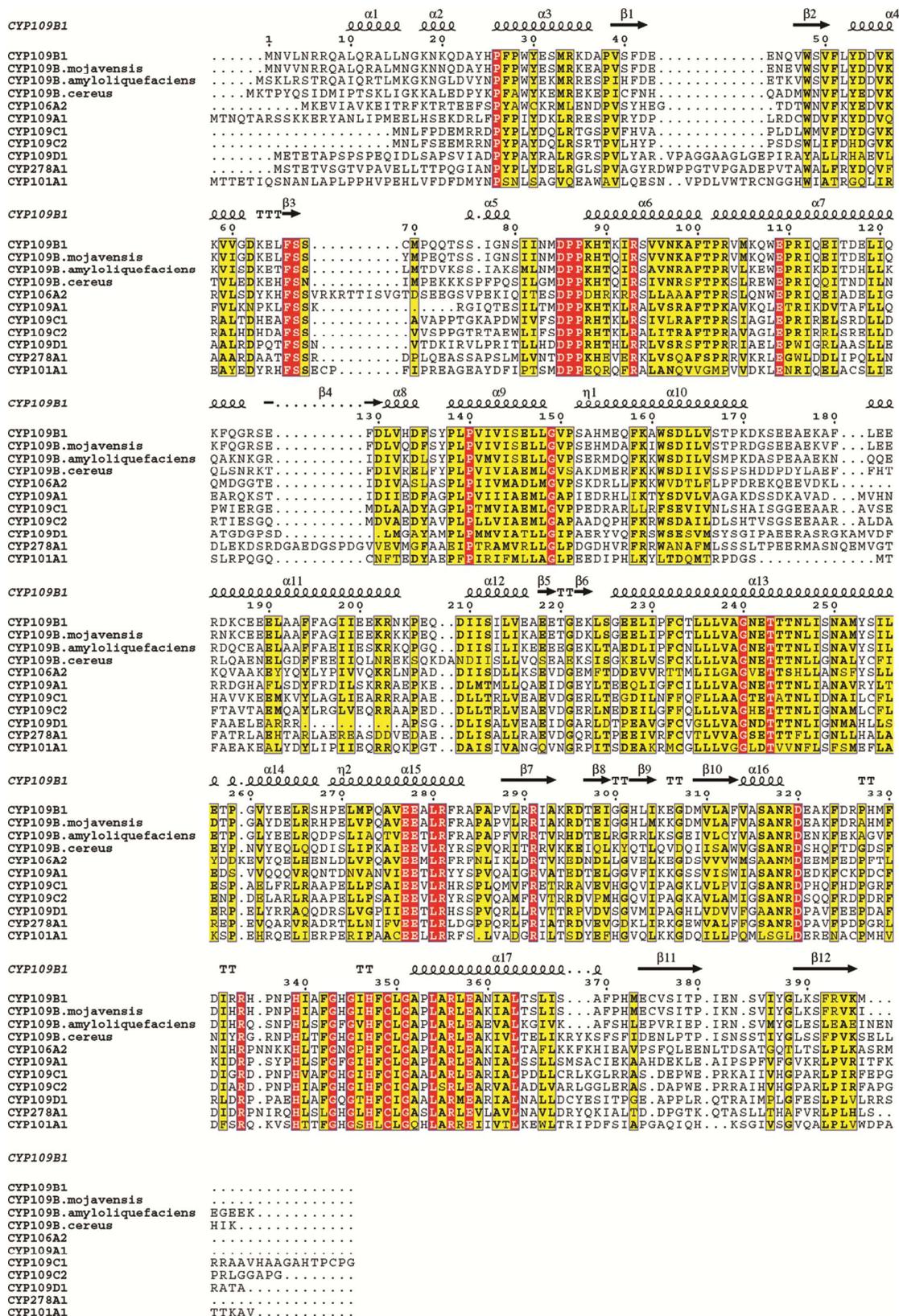


**Scheme 1.** Product isolated from the oxidation of  $\alpha$ - and  $\beta$ -ionone by CYP109B1 using the phthalate family oxygenase reductase from *Pseudomonas putida* KT24440 as an electron transfer partner.

## Figures



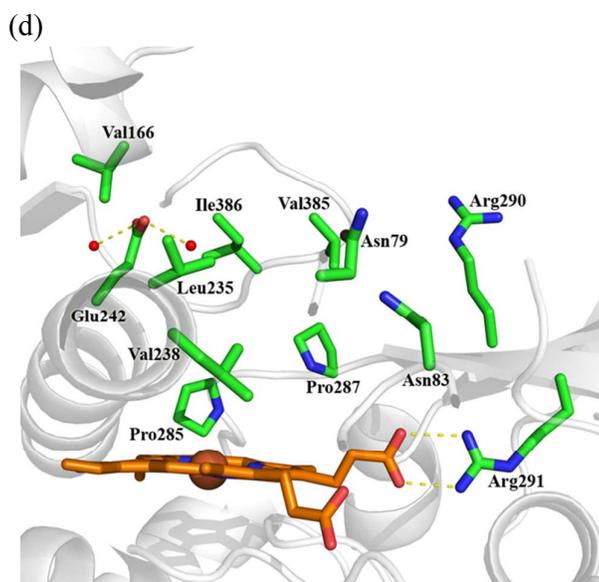
**Figure 1** Phylogenetic tree of selected CYP109 family enzymes. The CYP106A1 and CYP106A2 (P450Bm1) enzymes from strains of *Bacillus megaterium*, the other CYP enzymes from *B. subtilis* str. 168 (CYP102A2, CYP102A3, CYP107H1, CYP107J1, CYP107K1, CYP134A1 and CYP152A1) and other well characterised bacterial CYP enzymes are included. CYP101A1 (P450cam) from *Pseudomonas putida*, CYP102A1, P450Bm3, from *Bacillus megaterium*, CYP111A1, P450lin from a *Pseudomonas*, CYP153A1 from *Acinetobacter sp.* EB104, CYP108A1, P450terp from a *Pseudomonas* and CYP107A1, P450eryF from *Saccharopolyspora erythraea* are the CYP enzyme included for comparison. CYP123 from *M. tuberculosis* and CYP278A1 from *M. marinum* are also included. CYP109B1 is clustered with the other members of the CYP109 family from *Bacillus sp* (CYP109 is from *Bacillus amyloliquefaciens* FZB42). The CYP109D1 enzyme from *Sorangium cellulosum* So ce56 clusters more closely with CYP278A1 from *M. marinum* than with the other CYP109 enzymes. The data was taken from the CYPengineering database (<http://www.cyped.uni-stuttgart.de/>) which clusters CYP123 from *M. tuberculosis* with the CYP109 enzymes. However, our analysis found very little sequence overlap between these enzymes. The phenogram generated at PhyloDendron (<http://iubio.bio.indiana.edu/treeapp/treeprint-form.html>).



**Figure 2** Sequence alignment of CYP109B1 from *Bacillus* with other members of the CYP109 family, CYP106A2 from *Bacillus megaterium* (P450Bm1), CYP278A1 from *Mycobacterium marinum* and CYP101A1 (P450cam). Black arrows and cylinders indicate the  $\beta$ -sheets and

$\alpha$ -helices, respectively. Conserved and similar residues are highlighted with red and yellow highlighting, respectively. CYP109C1, CYP109C2 and CYP109D1 are from *Sorangium cellulosum* So ce56, CYP109A1 from *Bacillus subtilis* strain W23, CYP109amy from *Bacillus amyloliquefaciens* FZB42, CYP109moj from *B. mojavensis* and the CYP109 enzyme from *B. cereus*. Helices  $\alpha$ 13 and  $\alpha$ 15 are the I and K helices, respectively.

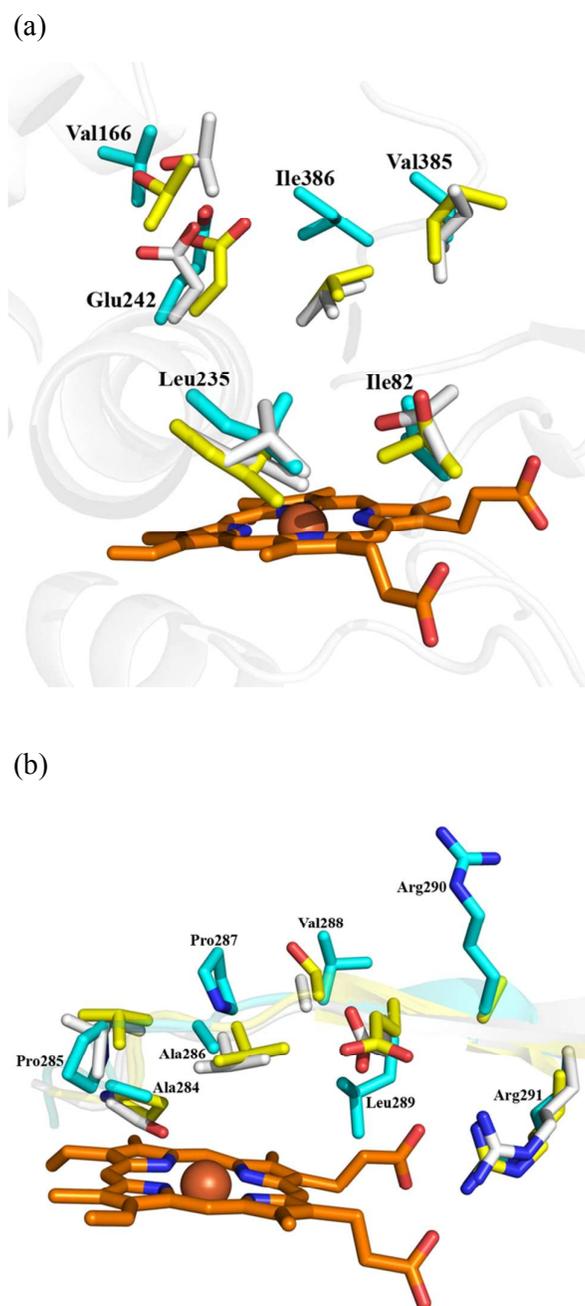




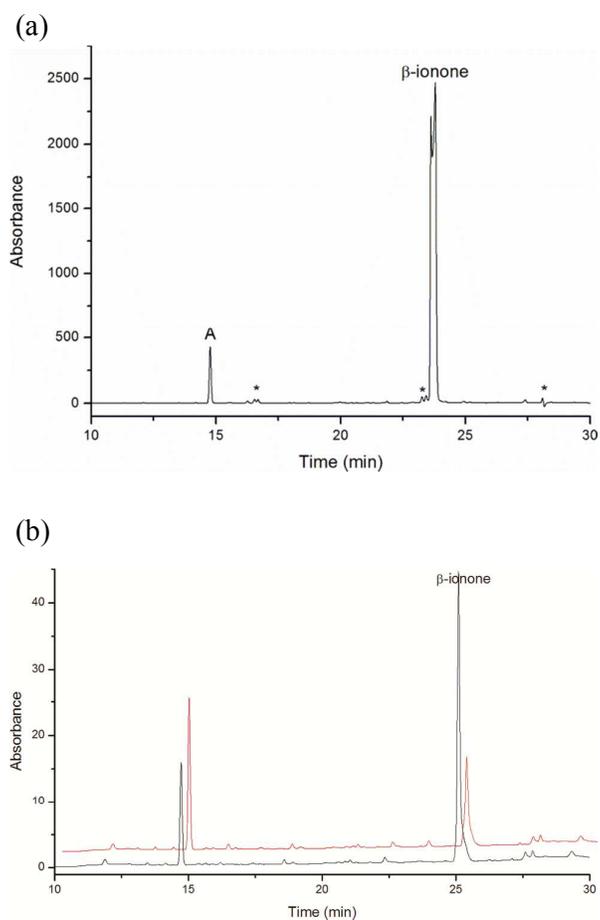
**Figure 3** (a) The overall structure of CYP109B1. The heme, the  $\alpha$ -helices,  $\beta$ -sheets and loops are shown in green, cyan, magenta and pink, respectively.

(b) The environment of the proximal side of the heme. The thiolate sulfur of the heme ligand, Cys349, is hydrogen-bonded to the NH of Gly351 and its carbonyl oxygen forms a hydrogen bond with the carbonyl oxygen of Phe348 (dashed line). The carbonyl groups of Phe342 and His347 interact with the backbone amide NH of Cys349 via hydrogen bonds.

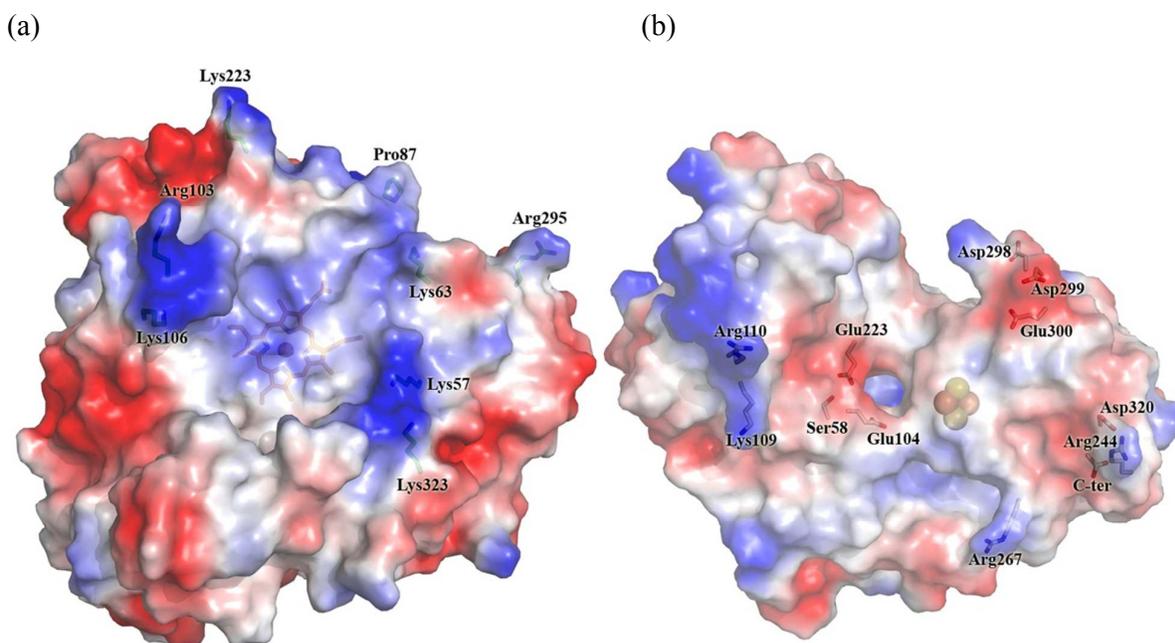
(c) and (d) The substrate binding pocket of CYP109B1 showing residues in tier 1 (c) and those higher up in the active site (d). Water molecules are shown in red, the heme in orange and hydrogen bonding interactions as dashed lines.



**Figure 4.** Overlay of (a) important active site residues in CYP109B1 (cyan) with their equivalents in CYP101A1 (PDB code: 3L62, grey) and CYP101D2 (PDB code: 3NV5, yellow) and (b) the loop region around SRS5 which contains two proline residues in CYP109B1 (cyan) with the equivalent regions in CYP101A1 (grey) and CYP101D2 (yellow).



**Figure 5** (a) HPLC analysis of the whole-cell oxidation of  $\beta$ -ionone with CYP109B1 and (b) HPLC analysis of the *in vitro* oxidation of the same substrate with CYP101B1 supported with PFOR after 30 minutes (black) and 90 minutes (red). 4-Hydroxy- $\beta$ -ionone is labeled (A) as are impurities (\*)



**Figure 6** (a) The electrostatic potential surface of the proximal face of CYP109B1. The area around the heme is dominated by positively charged residues. (b) The electrostatic potential surface of the phthalate family oxygenase reductase enzyme from *Pseudomonas cepacia* (PDB: 2PIA) around the iron-sulfur cluster. The area around the cluster is dominated by three negatively positively charged regions. These residues are conserved in PP1957 (Fig S6). Positively charged residues are shown in blue with negatively charged residues shown in red.