Investigation of the requirements for efficient and selective cytochrome P450 monooxygenase catalysis across different reactions

Matthew N. Podgorski,[a] Tom Coleman,[a] Rebecca R. Chao,[a] James J. De Voss,[b] John B. Bruning,[c] and Stephen G. Bell*[a]

[a] Department of Chemistry, University of Adelaide, Adelaide, SA 5005, Australia

E-mail: stephen.bell@adelaide.edu.au

[b] School of Chemistry and Molecular Bioscience, University of Queensland, St Lucia, Qld 4072, Australia

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

* Corresponding Author; Stephen G. Bell; E-mail: stephen.bell@adelaide.edu.au; Fax, +61 8 8313 4380

Abstract

The cytochrome P450 metalloenzyme CYP199A4 from *Rhodopseudomonas palustris* HaA2 catalyzes the highly efficient oxidation of *para*-substituted benzoic acids. Here we determined crystal structures of CYP199A4, and the binding and turnover parameters, with different *meta*-substituted benzoic acids in order to establish which criteria are important for efficient catalysis. When compared to the *para* isomers, the *meta*-substituted benzoic acids were less efficiently oxidized. For example, 3-formylbenzoic acid was oxidized with lower activity than the equivalent *para* isomer and 3-methoxybenzoic acid did not undergo O-demethylation by CYP199A4. The structural data highlighted that the *meta*-substituted benzoic acids bound in the enzyme active site in a modified position with incomplete loss of the distal water ligand of the heme moiety. However, for both sets of isomers the *meta*- or *para*-substituent pointed towards, and was in close proximity, to the heme iron. The absence of oxidation activity with 3-methoxybenzoic acid was assigned to the observation that the C-H bonds of this molecule point away from the heme iron. In contrast in the *para* isomer they are in an ideal location for abstraction. These findings were confirmed by using the bulkier 3-ethoxybenzoic acid as a substrate which removed the water ligand and reoriented the *meta*-substituent so that the methylene hydrogens pointed towards the heme, enabling more efficient oxidation. Overall we show relatively small changes in substrate structure and position in the active site can have a dramatic effect on the activity.

Keywords Metalloenzyme; C-H bond abstraction; X-Ray crystal structures; Heme monooxygenase; Enzyme catalysis; Molecular docking

1. Introduction

The heme-dependent cytochrome P450 (CYP) monooxygenase family of metalloenzymes catalyzes the oxidation of biological molecules often with high selectivity.[1-3] Moreover, they can catalyze C-H bond abstractions resulting in carbon hydroxylation at ambient temperatures using a reactive ferryl Compound I (Cpd I) intermediate.[1, 4-8] The insertion of the oxygen atom into the C-H bond is hypothesized to occur via a radical rebound mechanism after the Cpd I intermediate abstracts a hydrogen atom from the alkyl substrate.[5, 9] Importantly, members of this enzyme superfamily can catalyze a diverse range of other oxidative transformations including epoxidations, heteroatom dealkylations, sulfoxidations and other more complex reactions.[10-16]

Investigations into the substrate range of the soluble bacterial P450 CYP199A4 revealed that the enzyme is predisposed to bind and rapidly oxidize *para*-substituted benzoic acid substrates.[17-24] The fastest reaction catalyzed by CYP199A4 is *O*-demethylation of 4-methoxybenzoic acid, performed at a rate of 1220 min^{-1} with a coupling efficiency of 91%.[17, 19] CYP199A4 is also able to bind and oxidize the bulkier substrate 3,4-dimethoxybenzoic acid (veratric acid). In this instance, demethylation occurs only at the *para* position, affording 4-hydroxy-3-methoxybenzoic acid; the *meta*-substituent is not attacked.[21] CYP199A4 does not hydroxylate the aromatic ring of any of these substrates and also fails to oxidize benzoic acid.[19]

Crystal structures of substrate-bound CYP199A4 have been solved to elucidate the orientation of substrates in the binding pocket and examine the tendency of CYP199A4 to attack only the *para*-substituent (Figure 1).[18, 19, 22] These crystal structures revealed that the benzoate moiety, held almost perpendicular to the heme, forms both hydrophilic and hydrophobic interactions with residues in the binding pocket.[22] The substrate carboxylate hydrogen bonds to the hydroxyl groups of S95 and S244 and forms a salt bridge with the guanidinium of R92.[22]

Additionally, it interacts with a molecule of water which is in turn hydrogen-bonded to S244 and R243.[22]

If the substrate carboxylate is replaced with alternate functional groups, it abolishes the ability of the enzyme to tightly bind the substrate.[23, 24] Van der Waals interactions exist between the benzene ring and L98, A248, V181, and F185 and also between the *para*-substituent and F182 and F298.[18] Consistent with the ≥95% shift to high-spin observed when 4-methoxybenzoic acid binds, the crystal structure (PDB: 4DO1) revealed that the iron-bound water is completely displaced by this substrate. The *para*-methoxy group is held in close proximity to the heme iron (the carbon is 4.1 Å from the iron) and thus C-H bond abstraction occurs solely at this position, resulting in *O*-demethylation.[22] When veratric acid binds (PDB: 4EGN), its benzoate moiety and *para*-methoxy group occupy equivalent positions to those of 4-methoxybenzoic acid (Figure 1).[18] As a result, the *para*-methoxy group is close to the heme, but the *meta*-methoxy group of veratric acid is oriented away from the heme and is too far (8.1 Å) from the iron to react.[18]

The activity of CYP199A4 towards multiple other substituted benzoic acid substrates has been investigated (Figure S1).[17-21, 23, 24] 2,4-Dimethoxybenzoic acid was also solely oxidized at the *para* position.[19] On the other hand, 2- and 3-methoxybenzoic acid were not oxidized at all.[19] 3-Methoxybenzoic acid binds substantially less tightly than 4-methoxybenzoic acid (K_d = 69 μM vs 0.22 μM) and only shifts the spin-state to 40% high-spin, implying that the substrate is not bound in a suitable orientation relative to the heme to efficiently displace the iron-bound water.[19] This is consistent with previous studies which found that other *meta*-substituted benzoic acids bind weakly to CYP199A2, which shares 86% sequence identity with CYP199A4.[22] For example, 3-chlorobenzoic acid was found to induce a substantially smaller spin-state shift than the *para* isomer in CYP199A2 (10% vs 80%).[26] In the absence of a CYP199A4 crystal structure

with a substrate containing only a *meta*-arrangement of substituents, it was proposed that 3-methoxybenzoic acid could bind to the active site in a similar orientation to that of veratric acid (Figure 1). This would position the *meta*-substituent to point away from the heme iron, rather than towards it.[19] If this was the case the *meta*-substituent would be held too far from the iron to react. Of note is that 3,5-dimethoxybenzoic acid, which must bind with one *meta*-methoxy group pointing away from the heme and one methoxy group pointing towards it, is oxidized with low efficiency (at a product formation rate of 7 μM $(\mu M\text{-}P450)^{-1}$ min^{-1}).[19]

CYP199A4 can catalyze reactions other than *O*-dealkylation such as the sulfoxidation of 4-methythiobenzoic acid, *N*-dealkylation of 4-methylaminobenzoic acid and the hydroxylation of 4-methylbenzoic (4-toluic) acid.[17, 18] It also oxidizes bulkier substrates such as 4-ethoxy-, and 4-isopropyl-benzoic acid.[17, 18] Crystal structures have been obtained of CYP199A4 in complex with several of these substrates (PDB codes: 4-methylthio-, 5KT1; 4-methylamino-, 5U6W and 4-ethoxy-benzoic acid, 5U6T).

In contrast to 3-methoxybenzoic acid, both 3-methylamino- and 3-methylthio-benzoic acids were oxidized by CYP199A4[17]; however, the activities were significantly lower than those of the *para* isomers. We therefore set out to more fully investigate the ability of CYP199A4 to bind and oxidize benzoic acid substrates with alternative substituents at the *meta*-position. Docking and structural studies were performed to assess whether 3-substituted benzoic acids bind with the *meta*-substituent pointing away from the heme. The substrates were chosen to enable exploration of a range of different substituents which would alter the steric bulk and reactivity at this position (Figure 2).

The results were compared with those of the *para*-substituted isomers and used to clarify the binding mode of these substrates in order to rationalize the preference of CYP199A4 for *para*-

compared to *meta*-substituted benzoic acids. This provided a greater insight into how the coordination environment of the heme iron and the substrate orientation relative to this cofactor affects the catalytic activity in this important and ubiquitous superfamily of heme metalloenzymes.

2. Experimental methods

2.1 General methods

Production and purification of CYP199A4, crystal growth and structure determination, substrate binding analysis, *in vitro* NADH activity assays, chromatographic analysis of metabolites, hydrogen peroxide determination and catalytic activity as well as the use of a whole-cell oxidation system to generate metabolites for characterization were all performed using methods which have been reported previously (full details also provided in the Supplementary information).[17, 19, 21-23]

2.2 Docking of ligands into CYP199A4

To identify the likely binding mode of the substrates in the binding pocket of CYP199A4, the ligands were docked into a crystal structure of CYP199A4 (PDB ID: 5U6W) using ICM-Pro software, version 3.8-6a (Molsoft LLC, San Diego CA).[27] First, hydrogens were added to the crystal structure and optimized. Charges were also automatically assigned to residues which should be protonated/deprotonated at neutral pH and the orientation of histidine, proline, asparagine, glutamine and cysteine residues was also optimized.[27, 28] Waters were deleted from the structure with the exception of W618 which forms a hydrogen bond with the carboxylate of bound benzoic acid substrates, as well as active site residues (Figure 1).[17, 18, 22] The co-crystal ligand (4-methylaminobenzoic acid) was then removed from the active site of the enzyme and this region was used to define the binding site. The 4-methylaminobenzoic acid ligand was modified to 3-methylaminobenzoic acid and this ligand was re-docked into the binding pocket. The top scoring (lowest energy) pose had the 3-methylamino substituent pointing down towards the heme. A positional restraint was imposed on the nitrogen atom in order to dock the ligand in an alternative

configuration with the 3-methylamino moiety pointing away from the heme. The docking ‘score’ was then used to evaluate the docked poses, with a value of < -32 indicating a good fit.[27, 28] The other ligands were docked into CYP199A4 using the same method.[29]

3. Results

3.1 Substrate binding and catalytic activity of CYP199A4 towards *meta*-substituted benzoic acids

The ability of CYP199A4 to bind and oxidize 3-methyl- and 3-formyl-benzoic acid was explored. Compared to 3-methoxybenzoic acid, which induces a 40% spin-state shift in the heme iron to the high-spin form upon binding to CYP199A4 (Figure S2),[19] both 3-methyl- and 3-formyl-benzoic acid induced smaller type I shifts: 30% and 10%, respectively (Table 1, Figure S2, Table S2). These spin-state shifts are also lower than those reported for the corresponding *para* isomers and similar to those previously reported for 3-methylthio- and 3-methylamino-benzoic acid.[17] 4-Methylamino-, 4-methylthio- and 4-methyl-benzoic acid all induce 70% shifts to high-spin,[17] consistent with the *para*-substituent being held in close proximity to the heme and resulting in loss of the iron-bound water. By way of contrast, 4-formylbenzoic acid induced a smaller spin-state shift (25%), though this is larger than that induced by 3-formylbenzoic acid (Table 1, Figure S2, Table S2).

Based on the weak binding reported for 3-methoxy-, 3-methylamino- and 3-methylthio-benzoic acid, the substrates investigated here were predicted to bind weakly to CYP199A4.[17] 3-Methylbenzoic acid binds more than 100-fold less tightly than 4-methylbenzoic acid, and 3-methylamino- and 3-methylthio-benzoic acids (K_d = 89 μM, compared to 31 and 33 μM, respectively, Table 1, Figure S3). In line with its lower spin-state shift, 4-formylbenzoic acid bound with a lower affinity than the other *para*-substituted benzoic acids. 3-Formylbenzoic acid binds even more weakly than the other *meta*-substituted benzoic acids studied (K_d = 416 μM, Figure S3) and ~9-fold more less tightly than the corresponding *para* isomer (Table 1). This data indicates that all the *meta*-substituted benzoic acids fit less well in the CYP199A4 active site compared to

the corresponding *para* isomers. It has been hypothesized that this weaker binding is due to lower complementarity between the enzyme active site and the *meta*-substituted benzoic acids.[19]

The smaller spin-state shifts induced by the *meta*-substituted benzoic acids implied that the substrate is not bound close enough to the heme to efficiently displace the iron-bound water. Both 3-formyl- and 3-methylamino-benzoic acid slightly red shifted the Soret maximum (by 0.5 nm) despite giving type I difference spectra, implying that the heteroatoms of these substituents may interact with a heme-bound water (Supporting information, Figure S2, Table S2). The spin state experiments also suggest that NADH oxidation by CYP199A4 in the presence of these substrates will be slow because electron transfer is gated by displacement of the water and the related changes in spin state and redox potential.[17, 19, 30-33] If these substrates were to bind exclusively with the *meta*-substituent on the side of the benzene ring which is further away from the heme, we would predict little or no product to be formed. Therefore, *in vitro* NADH activity assays were performed to evaluate whether these substrates are metabolized by CYP199A4 (Figure S4). The rates of NADH oxidation and product formation by CYP199A4 are substantially lower with the *meta*-substituted benzoic acids than the equivalent *para* isomers (Table 1). Whereas 3-methoxybenzoic acid is not oxidized by CYP199A4, HPLC and GC-MS analysis revealed the presence of oxidation metabolites arising from CYP199A4-catalyzed oxidation of all the other *meta*-substituted substrates. We have previously reported some of these including 3-aminobenzoic acid and 3-methylsulfinylbenzoic acid, which arise from *N*-dealkylation and sulfoxidation of the nitrogen- and sulfur-containing substrates, respectively.[17]

In vitro turnovers with 3-methylbenzoic acid oxidized NADH at a rate of 78 μM $(\mu M\text{-}P450)^{-1}$ min^{-1} (henceforth abbreviated to min^{-1}), compared to a rate of 444 min^{-1} with the *para* isomer (Table 1). CYP199A4 has previously been reported to convert the *para* isomer exclusively into 4-

hydroxymethylbenzoic acid.[18] CYP199A4 similarly converted 3-methylbenzoic acid into a single product (Figure 3). The mass of the TMS-derivatized product (*m/z* = 296.10) is consistent with the calculated mass of doubly derivatized 3-hydroxymethylbenzoic acid (*m/z* = 296.1264). The product could also conceivably be 4-hydroxy-3-methylbenzoic acid, arising from hydroxylation of the benzene ring. However, the retention time of authentic 4-hydroxy-3-methylbenzoic acid was different to that of the P450 product, discounting the possibility that the substrate was hydroxylated at the *para* aromatic C-H bond of the benzoic acid moiety (Figure S5 and S6). To verify that the product was 3-hydroxymethylbenzoic acid, this compound was chemically synthesized via reduction of 3-formylbenzoic acid with $NaBH_4$. The CYP199A4 oxidation product and synthesized 3-hydroxymethylbenzoic acid co-eluted via HPLC (Figure S6). The coupling efficiency for this reaction was 48%, and the product formation rate 38 min^{-1}. CYP199A4 therefore oxidizes 3-methylbenzoic acid at a ~10-fold slower rate than it hydroxylates 4-methylbenzoic acid.

The CYP199A4 system oxidized NADH in the presence of 3-formylbenzoic acid at a rate (10.6 min^{-1}) roughly equal to the leak rate (the NADH oxidation rate in the absence of substrate; ~9 min^{-1}). However, low levels of the carboxylic acid product (isophthalic acid) were detected (Figure S7). The coupling efficiency and product formation rate were only 8% and 0.9 min^{-1}, respectively. CYP199A4 converts 4-formylbenzoic acid into terephthalic acid at a rate of 110 min^{-1} with a coupling efficiency of ~83% (Figure 4). This demonstrates that CYP199A4 is capable of catalyzing efficient aldehyde oxidation. Control reactions with these aldehyde-containing substrates and the others showed that no product was detected when the P450 was omitted from the reaction mixture (Figure 4 and S7).

Finally, we assessed the binding and turnover of the *ortho*-substituted substrate 2-methylbenzoic acid. CYP199A4 displayed weak affinity for 2-methylbenzoic acid (K_d = 611 μM) and only a small spin-state shift (~5%) was observed (Table 1, Figure S8). In line with the small spin-state shift induced by this substrate, the NADH oxidation rate was only 13 min^{-1}, barely faster than the leak rate (Table 1). No product was detected by GC-MS or HPLC analysis of the turnover mixture (Figure S9 and S10). This result reaffirms that CYP199A4 does not oxidize *ortho*-substituted benzoic acids.[19]

Minimal levels of H_2O_2 were detected in all CYP199A4-catalyzed oxidations. However, it is important to note here that when the rate of NADH oxidation is slow, as it is with the *meta*-benzoic acid isomers, H_2O_2 generated via uncoupling may be lost before the reaction is complete. Indeed, in control experiments we found that when ~300 μM H_2O_2 was incubated with a mixture containing HaPux, HaPuR and CYP199A4 in Tris-HCl buffer (but no NADH, substrate or catalase) or with just the HaPux [2Fe-2S] ferredoxin, almost all H_2O_2 was lost over a 60-minute period (Supporting information, Figure S11). Therefore, this method is unlikely to provide an accurate measurement of the amount of H_2O_2 formed at the end of a slow reaction (>15 min) and the values obtained are likely to be an underestimate of the actual amount generated (Figure S11). This is an important observation when considering the quantitation of the contribution of different uncoupling pathways in P450 reactions.

3.2 Docking of substrates into the active site of CYP199A4

The ability of CYP199A4 to oxidize all the *meta*-substituted benzoic acids with the exception of 3-methoxybenzoic acid led us to speculate that the orientation of these substrates in the binding pocket of CYP199A4 may differ. We therefore decided to dock the *meta*-substituted benzoic acids into the active site of CYP199A4 to predict their likely binding mode. ICM-Pro software was used

to dock all the substrates (except 3-formylbenzoic acid due to the lower binding affinity of the aldehyde isomers) into the CYP199A4 structure (PDB: 5U6W).[27]

In all instances the top-scoring pose had the *meta*-moiety pointing towards the heme (Figure 5, Supporting information and Table S3). For all the substrates, except 3-methylthiobenzoic acid, an alternative (lower-scoring) pose having this substituent pointing away from the heme was found. For 3-methylthiobenzoic acid the next lowest scoring pose also had the *meta*-substituent pointing towards the heme.

In contrast, when veratric acid was docked into the active site, the top-scoring poses had the *meta*-substituent pointing away from the heme, consistent with the crystal structure (Figure S12). Docking of 2-methoxybenzoic acid was conducted as an example of an *ortho*-substituted substrate and revealed that it preferred to bind with the *ortho*-methoxy group pointing away from the heme (Figure S12).

For 3-methylaminobenzoic acid the methyl carbon was only 4.1 Å from the heme iron and the sulfur of 3-methylthiobenzoic acid and methyl group of 3-methylbenzoic acid were at distances of 4.2 Å and 4.0 Å, respectively (Figure 5 and Table S3). This could rationalize why these substrates were oxidized at these positions. However, the top-scoring pose of 3-methoxybenzoic acid, which is not oxidized by CYP199A4, also had the *meta*-substituent pointing towards the heme, with the methyl group 4.0 Å from the iron. Another important observation from these docking experiments was that the position of the benzene moiety of the *meta*-substituted substrates was slightly different to that observed in the X-ray crystal structures of the *para*-substituted benzoic acids (Figure S13). To more satisfactorily answer the question of how these different *meta*-substituted benzoic acids bind, we decided to solve crystal structures of these substrates bound to CYP199A4.

3.3 X-ray crystal structures of CYP199A4 with *meta*-substituted benzoic acids

CYP199A4 was co-crystallized with 3-methoxy-, 3-methylthio-, 3-methylamino-, 3-methyl- and 4-methyl-benzoic acid (Figure S14) and X-ray diffraction data were collected at the MX1 or MX2 beamlines at the Australian Synchrotron at 100 K.[34, 35] Crystal structures of these protein-ligand complexes were solved at resolutions of 1.60-1.89 Å (PDB codes 6PQ6, 6PQD, 6PRR, 6PQW and 6PQS). Data collection and refinement statistics are presented in Table S4. The RMSD between Cα atoms of the superimposed structures is ≤0.178 Å, indicating that the overall protein fold is virtually identical for all five complexes. These crystal structures also had a similar fold to previously reported structures of CYP199A4 (Figure S15). Clear electron density corresponding to the substrate permitted modelling of each ligand into each structure (Figure 6). These substrates all bind with the *meta*-substituent pointing towards the heme in agreement with the docking studies. Overall this rationalizes why, with the exception of 3-methoxybenzoic acid, oxidation occurred at this position.

The crystal structures of the *meta*-substituted benzoic acids also demonstrated that the water ligand was maintained with a degree of occupancy at the sixth coordination position (21-90%; Table 1, Figure 6). The crystal structures also provide evidence that the water ligand and the N, S, or O atoms of the *meta*-substituent of 3-methylamino-, 3-methylthio- and 3-methoxy-benzoic acid, were interacting with this water molecule. The small red shift observed in the Soret maximum with 3-methylaminobenzoic acid may result from such an interaction. 3-Methylbenzoic acid, on the other hand, which has a more hydrophobic substituent, largely displaced the water in the 3-methylbenzoate-CYP199A4 crystal structure (water occupancy = 21%).

The occupancy of the water ligand in the other *meta*-substituted benzoic acid crystal structures closely correlated with the measured fraction of high/low-spin heme (Table 1). For

example, the 3-methylaminobenzoic acid-CYP199A4 complex is 90% low-spin and the occupancy of the water ligand in the crystal structure is also 90%. The 3-methylthiobenzoic acid-CYP199A4 complex is 70% low-spin and the refined water occupancy is 77%, while the 3-methoxybenzoic acid-CYP199A4 complex is 60% low-spin and the water ligand occupancy is 50%. 4-Methylbenzoic acid, on the other hand, induces a larger spin-state shift to the high-spin form (70%) and a water ligand is not observed in the crystal structure. The *para*-substituted benzoic acids displace the water molecule and induce larger shifts to the high-spin ferric state.[17] It should be noted that partial spin-state shifts and water occupancy in the crystal structure have also been reported with different substrates and $P450_{cam}$ (CYP101A1).[36, 37]

To ascertain why we observe such an acute difference in the binding and activity of *meta*- versus *para*-substituted benzoic acids, we next compared the structures of the different substrate-bound analogues. When the structures of the *meta*-substituted benzoate-bound CYP199A4 were superimposed with the equivalent *para* substrate-bound structures, only subtle differences in the positions of active-site residues were observed (Figures 7 and 8). However, the benzene ring of the *meta* isomers are all shifted up away from the heme relative to same moiety of the equivalent *para*-substituted benzoic acids. This was in agreement with what was observed in the docking experiments and the positions of 3-methoxy-, 3-methylthio-, 3-methylamino-, and 3-methyl-benzoic acid in the crystal structure agreed with the top-scoring docked poses. When the position of the substrates was compared in the X-ray and the modelled structures all the atoms, including those of the substituent, were found to align (Figure S16). This suggests that energy minimization modelling could be a valuable tool in understanding the binding of different benzoic acids and related substrates in this family of P450 enzymes. This movement of the substrate, when the

position of substitution changes, seems to occur to avoid a steric clash between the *meta*-substituent and the heme (Figure S17).

A common feature of all of the structures with *meta*-substituted isomers is that there is more space directly above the heme center as the substituent is more offset from the area directly above the iron (Figures 7 and 8). This is consistent with the partial occupancy of the 6th ligand in these structures compared to the *para*-substituted benzoic acid counterparts, in which this water was not present. Based on the productive catalytic activity and NADH oxidation rates observed (Table 1) it appears that this water molecule can be displaced enabling the enzyme to bind dioxygen and enabling turnover of the catalytic cycle albeit at a lower rate than for the equivalent *para* isomer.

While the heme iron in 4-methylthiobenzoate-bound CYP199A4 is five-coordinate, when 3-methylthiobenzoate is bound a high-occupancy water ligand is present (Fe–O distance 2.3 Å; Table S5). The sulfur of 3-methylthiobenzoic acid is 3.3 Å from this water. The methylthio moiety of 3-methylthiobenzoic acid is rotated out of the plane of the benzene ring by 13˚ and that of 4-methylthiobenzoic acid is rotated by 34˚. In both structures, the methylthio moiety is held close to the heme, though the sulfur of 3-methylthiobenzoic acid is held nearer to the iron than that of 4-methylthiobenzoic acid (4.3 vs 5.0 Å; Table S5). The methyl group of 3- and 4-methylthiobenzoic acid is also held in close proximity to the heme iron (4.4 vs 4.5 Å). When comparing the structures, we wanted to ascertain how close the substrate would be to the oxygen of the reactive Cpd I intermediate (as well as the heme iron). It has been calculated that the Fe=O bond length of Cpd I is 1.62 Å.[38, 39] This is in good agreement with the measured Cpd I Fe=O bond length using EXAFS, which was found to be 1.67 Å.[40] Therefore, to estimate the likely distances between the ferryl oxygen of Cpd I and the relevant atoms of the bound substrate, an oxygen atom was modelled 1.62 Å above the heme iron directly opposite the Fe-S bond and we assumed that the

position of the substrate is not significantly changed after oxygen binding and activation (Table S5). The measured distance between the Cpd I oxygen and the sulfur of 3-methylthiobenzoic acid is only 3.2 Å. This is slightly shorter than the distance between the Cpd I oxygen and the sulfur of 4-methylthiobenzoic acid (3.4 Å). For the *meta*-methylthio isomer the Fe^{IV}=O–S angle is 126.2˚, whereas for the *para* isomer the angle is larger (162.5˚).

In the crystal structure of 3-methylaminobenzoate-bound CYP199A4, a water ligand to the heme is present at 90% occupancy (Fe–O distance 2.2 Å; Table S5), and this water is hydrogen-bonded to the substrate nitrogen (3.1 Å). The nitrogen and methyl group of 3-methylaminobenzoic acid are 4.2 and 4.4 Å from the heme iron, respectively. The distance between the Cpd I oxygen and methyl carbon of 3-methylaminobenzoic acid is 3.1 Å, which is similar to that of the methyl carbon of 4-methylaminobenzoic acid (3.0 Å). The Fe^{IV}=O–C_{Me} angle was similar for both the *meta* and *para* isomers (129.9 vs 124.6˚). The nitrogen heteroatom of the *para* isomer had a larger Fe^{IV}=O–N angle than 3-methylaminobenzoic acid (148.6 vs 126.8˚; Table S5).

In both the structures of 3-methoxybenzoic acid- and 4-methoxybenzoic acid-bound CYP199A4, the methoxy groups are held above the heme. The O and C_{Me} of 3-methoxybenzoic acid are equidistant from the heme iron (4.4 Å), and the methoxy group is rotated 20˚ out of the plane of the benzene ring. In contrast, the methoxy group of 4-methoxybenzoic acid lies almost in the plane of the benzene ring, and the C_{Me} is closer to the iron (4.1 Å) with the oxygen being further away (5.2 Å). When the ferryl oxygen of Cpd I was considered, the C_{Me} and O of 3-methoxybenzoic acid are both 3.1 Å from the Cpd I oxygen, whereas the C_{Me} and O of 4-methoxybenzoic acid are 2.7 and 3.6 Å from the ferryl oxygen. The Fe^{IV}=O–C_{Me} angle is similar for both 3-methoxy- and 4-methoxy-benzoic acid (136.3 vs 140.7˚).

The methyl group of both 3-methyl- and 4-methyl-benzoic acid is held close to the heme iron (4.4 vs 4.5 Å). A low-occupancy water ligand is present in the 3-methylbenzoate-CYP199A4 structure with a Fe–O distance of 2.5 Å, but no water ligand was detected in the 4-methylbenzoic acid structure. The benzene ring of 3-methylbenzoic acid is shifted away from the heme relative to that of 4-methylbenzoic acid, though the carboxylate groups are held in similar positions. The methyl of 3-methylbenzoic acid is held slightly further from the ferryl oxygen than that of 4-methylbenzoic acid (3.2 vs 3.0 Å). The Fe^{IV}=O–C_{Me} angle is larger for 4-methylbenzoic acid than for 3-methylbenzoic acid (154.4 vs 130.4˚).

3.4 Substrate binding and catalytic activity of bulkier *meta*-substituted benzoic acids

The crystal structures of 3-methoxy-, 3-methylamino-, and 3-methylthio-benzoate-bound CYP199A4 revealed that although the *meta*-substituent was positioned close to the heme, the water ligand was not fully displaced. We hypothesized that substrates with bulkier *meta*-substituents would more efficiently displace the water ligand and that CYP199A4 would display higher oxidation activity towards them. We selected 3-ethoxybenzoic acid and 3-*tert*-butylbenzoic acid and docked these into the active site of CYP199A4 (Figure S18). The top-scoring poses have the *meta*-substituent positioned on the heme side of the substrate (Table S6). However, docking 3-*tert*-butylbenzoic acid into CYP199A4 resulted in a steric clash between the *tert*-butyl group and a heme nitrogen. This implies that this substrate may be too bulky to fit into the active site.

We subsequently measured the spin-state shifts induced by these substrates. The minimal spin-state shift (<5%) induced by 3-*tert*-butylbenzoic acid implies that the bulky *tert*-butyl group hinders binding (Table 3, Figure S19). The *para* isomer, 4-*tert*-butylbenzoic acid, bound to CYP199A4 with a 90% shift to the high spin form, but had lower binding affinity (39 ± 2 μM,

Figure S20) than the other *para*-substituted benzoic acid substrates, such as 4-methoxybenzoic acid (Table 1). In contrast to the other *meta*-substituted benzoic acids we tested, 3-ethoxybenzoic acid induced an almost complete (85%) shift of the ferric iron to the HS state, indicating efficient removal of the 6th water ligand (Figure S19). This spin-state shift is larger than that induced by certain *para*-substituted benzoic acids (Table 1). However, 3-ethoxybenzoic acid was found to bind substantially less tightly than the *para* isomer (K_d = 82 ± 2 vs 0.2 ± 0.1 μM, Figure S21).

There was no increase in the NADH oxidation rate above the leak rate (as defined in Table 1) in the presence of 3-*tert*-butylbenzoic acid. No product was detected by HPLC analysis of the reaction mixture (Figure S22). These observations are evidence that 3-*tert*-butylbenzoic acid does not bind to CYP199A4. The corresponding *para* isomer is hydroxylated by CYP199A4 at a rate of 227 min^{-1} with 100% coupling efficiency and chromatographic analysis indicated it was oxidized to a single metabolite (Table 3, Figure S23 and S24). The GC-MS of the TMS-derivatized turnover confirmed the formation of a single alcohol product (*m/z* = 338.25 vs expected 338.17, Figure S24). The product was prepared using a whole-cell oxidation system[21] and, after conversion to the methyl ester, purified and characterized by NMR as 4-(1,1-dimethyl-2-hydroxyethyl)benzoic acid (Supporting information, Figure S25). This arose from methyl hydroxylation due to the absence of any benzylic hydrogens within the substrate.

On the other hand, the CYP199A4 system oxidized NADH upon addition of 3-ethoxybenzoic acid at a moderately fast rate (208 min^{-1}). Whereas CYP199A4 is unable to oxidize 3-methoxybenzoic acid, the enzyme did catalyze the *O*-deethylation of 3-ethoxybenzoic acid. The coupling efficiency and product formation rate were calculated to be 66% and 138 min^{-1}, and only 7% of the NADH consumed was measured as being channeled into H_2O_2 production (Table 3). A single product, 3-hydroxybenzoic acid, was detected, indicating that no hydroxylation occurred at

the less reactive terminal CH_3 group (Figure 9). CYP199A4 converts the *para* isomer, 4-ethoxybenzoic acid, into 4-hydroxybenzoic acid at a rate of 527 min^{-1} and with a coupling efficiency of 100%.[17] CYP199A4 therefore still exhibits substantially higher activity towards 4-ethoxybenzoic acid than 3-ethoxybenzoic acid, which again demonstrates the generality of the preference of this enzyme for *para*- over *meta*-substituted benzoic acids.

3.5 Crystal structure of 3-ethoxybenzoate-bound CYP199A4

We obtained a crystal structure of 3-ethoxybenzoate-bound CYP199A4 which was solved at 2.37 Å resolution (PDB code 6PRS, Figure 10, Table S7, Figure S26). It confirms that the *meta*-substituent points down towards the heme and that the 6^{th} iron water ligand is displaced, in agreement with the 85% shift to high spin and the docked structure (Figure S27). There is a change in the (CH_2-O-C_3-C_4, Table S8) dihedral angle of the larger ethoxy substituent compared to that of the methoxy substituent which results in the methylene group being held more over the heme iron than the methyl group of the methoxy substituent (Figure 10, Table S8).

Both the methylene (CH_2) and methyl carbons (CH_3) of the *meta*-ethoxy substituent are held in close proximity to the heme iron (3.6 and 4.2 Å, respectively, Table S8). When the distances were measured between the substrate and the hypothetical position of the Cpd I oxygen, the methylene carbon (CH_2) is also held closer than CH_3 (2.1 vs 2.7 Å). The Fe^{IV}=O–CH_2 and Fe^{IV}=O–CH_3 angles are 155.3 and 147.7˚, respectively. C-H bond abstraction occurs solely at CH_2 by CYP199A4, with no hydroxylation being observed at the methyl group (Table S8). This can be rationalized by CH_2 being the closest carbon to the iron and the weaker strength of the secondary methylene C-H bonds (especially those adjacent to oxygen)[1] compared to the primary C-H bonds of the terminal methyl group. Overall this structure highlights that certain *meta*-substituted benzoic

acid substrates can efficiently bind to the active site and thus remove the 6^{th} water ligand of CYP199A4 and that this enables more efficient catalytic oxidation.

4. Discussion

The cytochrome P450 family of enzymes catalyze many reactions which are important in xenobiotic catabolism and secondary metabolite formation. The majority of cytochrome P450s are thought to follow the consensus catalytic cycle observed with P450cam, as does CYP199A4. The reactive Cpd I intermediate has been calculated to have similar reactivity across different forms. We hypothesize that the observations reported here, with CYP199A4, would be applicable to many other P450 enzymes. However, it is important to note that an increasing number of P450 enzymes which catalyze varied reactions are being discovered, highlighting the complex roles they play in metabolism and their potential for mechanistic and function diversification.

CYP199A4 displays a strong preference for benzoic acid substrates with a *para*-substituent. Despite having no activity towards 3-methoxybenzoic acid and 3-*tert*-butylbenzoic acid, this enzyme nevertheless is capable of oxidizing certain *meta*-substituted benzoic acids. The binding mode of *meta*-substituted benzoic acids revealed by the crystal structures has the *meta*-substituent positioned close to the heme, not directed away as previously observed in the disubstituted veratric acid structure. All the *meta*-substituted benzoic acid substrates studied here bind with the benzene ring and carboxylate held in virtually identical positions (Figure S28). While the distance of the *meta*-substituent to the heme iron/Cpd I model oxygen is often comparable to that of the equivalent position in the *para*-substituted species, the altered position of the benzene ring and change in substitution pattern results in there being more space directly above the heme iron in the former set of substrates.

Within each pair of *meta* and *para* isomers, the extent of the spin-state shift correlated with the NADH oxidation rate, suggesting that this step is somewhat gated by displacement of the water. The observation of partial occupancy of the 6th heme iron water ligand in the crystal structures of

the *meta*-substituted benzoic acid substrates would result in a lower spin-state shift to the high-spin form and slower turnover of the catalytic cycle which is consistent with the first electron reduction being the rate determining step in the catalytic cycle. This is consistent with previous observations with this enzyme and other related class I bacterial systems such as $P450_{cam}$. However, the extent of the spin-state shift does not always directly correlate with NADH oxidation rate or product formation of CYP199A4 across all these substrates. Another observation from this work is that hydrogen peroxide is consumed by the P450 system (by the [2Fe-2S] ferredoxin) which can make its quantification in slow reactions difficult or unfeasible.

The binding orientations of 3-methylamino-, 3-methylthio- and 3-methyl-benzoic acid enable one to explain why these compounds can be oxidized by CYP199A4. Despite their lower activities, the coupling efficiency of these oxidation reactions were comparable to the equivalent *para* isomers, in line with their similar distances and orientations relative to the heme iron. The crystal structures and molecular docking experiments with the *meta*-substituted benzoic acids provide excellent insight into why *meta*-substituted and larger substrates are not as good substrates for the CYP199A4 enzyme. These would be beneficial in the future design of variant forms of CYP199A4 to improve the oxidation activity of these and similar substrates. If more space was generated at the top of the active site by mutating the hydrophobic residues in this region (F182, F185 and V181, Figure 7 and 8) the *meta* isomer could bind in an orientation which more efficiently displaces the 6^{th} water ligand and improve the activity. Such changes may also enhance the binding and activity of larger substrates such as substituted naphthoic acids. However, such amino acid substitutions could result in the substrate binding with the *meta*-substituent pointing into the space generated, which as with veratric acid, would place them further away from the heme.

The results with the methylthio and methylamino analogues are perhaps unsurprising given their higher reactivity. The sulfur of 3-methylthiobenzoic acid is in range for oxidation by the ferryl Cpd I reactive intermediate. For sulfur-containing substrates, Cpd I reacts with the heteroatom via direct oxygen transfer, or via single electron transfer followed by oxygen rebound (we note that other intermediates of the P450 catalytic cycle have been proposed to catalyze sulfoxidations).[8, 12] The observation that 3-methylaminobenzoic acid is oxidized by CYP199A4 but 3-methoxybenzoic acid is not requires more detailed analysis. These substrates have virtually identical orientations in the two crystal structures and the distances of the methyl group and heteroatom of each of the substituents to the heme iron are comparable. The reactivity of methylamines and the potential of *N*-dealkylation to occur via the alternative single electron transfer (SET) pathway - which would involve initial reaction at the closer nitrogen atom- rather than a hydrogen abstraction (HAT) mechanism, are likely to be important here.[12]

The CYP199A4 oxidations of 3-methoxy-, 3-ethoxy- and 3-methyl-benzoic acids (and their *para* isomer equivalents) would all proceed via an initial C-H bond abstraction. Factors that would affect whether Cpd I can abstract a hydrogen are the distance of the C-H bond from the ferryl oxygen, the C-H bond strength and the angle of approach of the substrate (Figure 11).[41] The C-H bond strength should not be problematic in any of these oxidations. 3-Methoxybenzoic acid is not oxidized by CYP199A4 which indicates that it must be inappropriately positioned in the active site for C-H bond activation despite turnover of the catalytic cycle (as judged by the NADH oxidation rate, Table 1). Whereas 3-methoxybenzoic acid is not oxidized, CYP199A4 is able to dealkylate the bulkier 3-ethoxybenzoic acid and hydroxylate the methyl C-H bonds of 3-methylbenzoic acid. There is a significant change in the conformation of the *meta*-substituent in

the ethoxy structure which pushes the methylene group closer to the heme iron and changes its orientation with respect to where the reactive oxygen atom would be located.

It has been proposed that a linear arrangement of the C-H-O atoms is ideal for hydrogen abstraction[42] and that it is optimal for the substrate to approach the Fe^{IV}=O species at an angle of ~130˚ (Fe-O-C angle).[41] In the 4-methoxybenzoic acid crystal structure, the methoxy group and its hydrogens point towards the heme. However, the methoxy group of 3-methoxybenzoic acid is oriented such that the hydrogens could not form a linear C-H-O arrangement with the iron-oxo atom of Cpd I (Figure 11). The change in the conformation in the 3-ethoxybenzoic acid side chain results in the C-H bonds on the methylene pointing towards the iron-oxo atom in contrast to those on the respective methyl groups of 3-methoxy and 3-ethoxybenzoic acid (Figure 11).

It is important to note that while the benzoic acid moiety of the substrates seems to be tightly confined within the active site, we cannot rule out that the orientation of the substituent observed in the crystal structures may not be the same as in solution and its position relative to the heme iron may be altered once oxygen binds or in the presence of the reactive Cpd I intermediate. However, the crystal structure and biochemical data reported are in agreement with theoretical work related to this family of enzymes (Figure 11).[41, 42] The presence of heteroatoms, which can potentially interact with ligands bound to the heme iron, could affect heme reduction or oxygen binding and activation thus altering different steps of the catalytic cycle. In the 4-methoxy- and 3-ethoxy-benzoic acid structures, the methyl or methylene groups are held significantly closer to the heme than the heteroatom. In contrast, in the 3-methoxybenzoic acid structure, the methyl group and oxygen are held at almost equal distances from the iron. There is a possibility that this close approach of the oxygen heteroatom could interfere with substrate oxidation.

5. Conclusions

Overall the crystal structures of CYP199A4 provide detailed information on the factors which promote efficient catalytic monooxygenase activity for a range of substrates. In summary, this study reveals that CYP199A4 does oxidize certain *meta*-substituted benzoic acids and provides an understanding of why the activity towards these substrates is substantially lower than those of the *para* isomers. The crystal structures of CYP199A4 with these substrates bound in the active site provide important insights into the structural parameters, in this system, which define whether C-H bond abstraction (or other P450-catalyzed activities) can occur efficiently. This provides experimental validation of theoretical studies on other forms of these metalloenzymes and, accordingly these observations could be extended to other P450s. This work highlights that relatively minor changes in the substrate and its position in the enzyme active site relative to the heme can have a dramatic effect on the activity of oxidation. Some P450 activities, for example the C-H abstraction that results in *O*-dealkylation, were more sensitive to these parameters than others which involve more reactive substituents (*N*-dealkylation and *S*-oxidation). These considerations are essential when designing or evolving enzyme systems for efficient biocatalytic reaction or understanding and interpreting the reactivity of computational docking or structural studies of these enzymes. The range of monooxygenase reactions catalyzed by CYP199A4 (Figure S1) infer that this system could be used to investigate other types of cytochrome P450-catalyzed activities, which are important in drug metabolism or natural product synthesis.

Abbreviations

AcCN, acetonitrile; BA, benzoic acid; Cpd I, iron(IV)−oxo porphyrin radical cation intermediate; CYP, cytochrome P450 enzyme; CYP199A4, cytochrome P450 enzyme from *Rhodopseudomonas palustris* strain HaA2; FAD, flavin adenine dinucleotide; FEM, feature enhanced maps; GC-MS, gas chromatography-mass spectrometry; HaPuR, FAD containing flavoprotein from *Rhodopseudomonas palustris* strain HaA2 (Uniprot: Q2ITV9); HaPux; [2Fe-2S] ferredoxin from *Rhodopseudomonas palustris* strain HaA2 (Uniprot: Q2IU01); HAT, hydrogen abstraction; HPLC, high performance liquid chromatography; HS, high spin; NADH, reduced nicotinamide adenine dinucleotide; PFR, product formation rate; SET, single electron transfer; TFA, trifluoracetic acid; TMS, trimethylsilyl; Tris, tris(hydroxymethyl)aminomethane.

Supplementary Information. Experimental methods, additional substrate binding parameters and catalytic turnover activity data for CYP199A4, details of the distances and angles of pertinent substrate atoms from the heme iron in the crystal structures of CYP199A4, spin state shift assays, dissociation constant analysis, examples of NADH oxidation assays, H_2O_2 assays, HPLC and GC-MS analysis of the *in vitro* turnovers, NMR spectra and details of isolated products, computational docking studies, crystallography details and structures of the Cα chains of different CYP199A4 crystal structures.

Acknowledgements

This work was in part supported by ARC grant DP140103229 (to JJDV and SGB). SGB acknowledges the ARC for a Future Fellowship (FT140100355). The authors also acknowledge the award of an Australian Government Research Training Program Scholarships (PhD to TC

and MPhil to MNP and RRC). We would like to thank the scientists at MX1 and MX2 beamline for help with data collection. We acknowledge ANSTO for financial support and in providing the facility used in this work.

References

[1] P.R. Ortiz de Montellano, Chem. Rev., 110 (2010) 932-948.

[2] T.L. Poulos, Chem. Rev., 114 (2014) 3919-3962.

[3] P.R. Ortiz de Montellano, in: P.R. Ortiz de Montellano (Ed.), Cytochrome P450: Structure, Mechanism, and Biochemistry, Springer International Publishing, Cham, 2015, pp. 111-176.

[4] F. Ogliaro, S.P. de Visser, S. Cohen, P.K. Sharma, S. Shaik, J. Am. Chem. Soc., 124 (2002) 2806-2817.

[5] J. Rittle, M.T. Green, Science, 330 (2010) 933-937.

[6] P.K. Sharma, S.P. De Visser, S. Shaik, J. Am. Chem. Soc., 125 (2003) 8698-8699.

[7] B. Meunier, S.P. de Visser, S. Shaik, Chem. Rev., 104 (2004) 3947-3980.

[8] S. Shaik, S. Cohen, Y. Wang, H. Chen, D. Kumar, W. Thiel, Chem. Rev., 110 (2010) 949-1017.

[9] X. Huang, J.T. Groves, J. Biol. Inorg. Chem., 22 (2017) 185-207.

[10] M.J. Cryle, J.E. Stok, J.J. De Voss, Aust. J. Chem., 56 (2003) 749-762.

[11] P.R. Ortiz de Montellano, J.J. De Voss, Nat. Prod. Rep., 19 (2002) 477-493.

[12] F.P. Guengerich, Chem. Res. Toxicol., 14 (2001) 611-650.

[13] E.M. Isin, F.P. Guengerich, Biochim. Biophys. Acta, 1770 (2007) 314-329.

[14] J.B. Behrendorff, W. Huang, E.M. Gillam, Biochem. J., 467 (2015) 1-15.

[15] M.T. Reetz, J. Am. Chem. Soc., 135 (2013) 12480-12496.

[16] V.B. Urlacher, M. Girhard, Trends Biotechnol., 30 (2012) 26-36.

[17] T. Coleman, S. Wong, M. Podgorski, J. Bruning, J. De Voss, S. Bell, ACS Catal., 8 (2018) 5915-5927.

[18] S.G. Bell, R. Zhou, W. Yang, A.B. Tan, A.S. Gentleman, L.L. Wong, W. Zhou, Chemistry, 18 (2012) 16677-16688.

[19] T. Coleman, R.R. Chao, J.B. Bruning, J. De Voss, S.G. Bell, RSC Adv., 5 (2015) 52007 - 52018.

[20] R.R. Chao, I.C.K. Lau, J.J. De Voss, S.G. Bell, ChemCatChem, 8 (2016) 3626-3635.

[21] S.G. Bell, A.B. Tan, E.O. Johnson, L.L. Wong, Mol. Biosyst., 6 (2010) 206-214.

[22] S.G. Bell, W. Yang, A.B. Tan, R. Zhou, E.O. Johnson, A. Zhang, W. Zhou, Z. Rao, L.L. Wong, Dalton Trans., 41 (2012) 8703-8714.

[23] T. Coleman, R.R. Chao, J. De Voss, S.G. Bell, Biochim. Biophys. Acta Proteins Proteomics, 1864 (2016) 667-675.

[24] R.R. Chao, J.J. De Voss, S.G. Bell, RSC Adv., 6 (2016) 55286-55297.

[25] S.G. Bell, F. Xu, I. Forward, M. Bartlam, Z. Rao, L.-L. Wong, J. Mol. Biol., 383 (2008) 561-574.

[26] S.G. Bell, N. Hoskins, F. Xu, D. Caprotti, Z. Rao, L.L. Wong, Biochem. Biophys. Res. Commun., 342 (2006) 191-196.

[27] R. Abagyan, M. Totrov, D. Kuznetsov, J. Comput. Chem., 15 (1994) 488-506.

[28] A.J.W. Orry, R. Abagyan, in: A.J.W. Orry, R. Abagyan (Eds.), Homology Modeling: Methods and Protocols, Humana Press, Totowa, NJ, 2012, pp. 351-373.

[29] R. Abagyan, A.J.W. Orry, E. Raush, M. Totrov.

[30] M.T. Fisher, S.G. Sligar, J. Am. Chem. Soc., 107 (1985) 5018-5019.

[31] S.G. Sligar, Biochemistry, 15 (1976) 5399-5406.

[32] M.J. Honeychurch, A.O. Hill, L.L. Wong, FEBS Lett., 451 (1999) 351-353.

[33] S.G. Bell, J.H. McMillan, J.A. Yorke, E. Kavanagh, E.O. Johnson, L.L. Wong, Chem. Commun. , 48 (2012) 11692-11694.

[34] D. Aragao, J. Aishima, H. Cherukuvada, R. Clarken, M. Clift, N.P. Cowieson, D.J. Ericsson, C.L. Gee, S. Macedo, N. Mudie, S. Panjikar, J.R. Price, A. Riboldi-Tunnicliffe, R. Rostan, R. Williamson, T.T. Caradoc-Davies, J. Synchrotron Radiat., 25 (2018) 885-891.

[35] N.P. Cowieson, D. Aragao, M. Clift, D.J. Ericsson, C. Gee, S.J. Harrop, N. Mudie, S. Panjikar, J.R. Price, A. Riboldi-Tunnicliffe, R. Williamson, T. Caradoc-Davies, J. Synchrotron Radiat., 22 (2015) 187-190.

[36] R. Raag, T.L. Poulos, Biochemistry, 30 (1991) 2674-2684.

[37] R. Raag, T.L. Poulos, Biochemistry, 28 (1989) 917-922.

[38] R. Lonsdale, J. Olah, A.J. Mulholland, J.N. Harvey, J. Am. Chem. Soc., 133 (2011) 15464-15474.

[39] J.C. Schoneboom, H. Lin, N. Reuter, W. Thiel, S. Cohen, F. Ogliaro, S. Shaik, J. Am. Chem. Soc., 124 (2002) 8142-8151.

[40] C.M. Krest, A. Silakov, J. Rittle, T.H. Yosca, E.L. Onderko, J.C. Calixto, M.T. Green, Nat. Chem., 7 (2015) 696-702.

[41] R. Lonsdale, K.T. Houghton, J. Żurek, C.M. Bathelt, N. Foloppe, M.J. de Groot, J.N. Harvey, A.J. Mulholland, J. Am. Chem. Soc., 135 (2013) 8001-8015.

[42] J.C. Schöneboom, S. Cohen, H. Lin, S. Shaik, W. Thiel, J. Am. Chem. Soc., 126 (2004) 4017-4034.

Investigation of the requirements for efficient and selective cytochrome P450 monooxygenase catalysis across different reactions

Matthew N. Podgorski,[a] Tom Coleman,[a] Rebecca R. Chao,[a] James J. De Voss,[b] John B. Bruning,[c] and Stephen G. Bell*[a]

[a] Department of Chemistry, University of Adelaide, Adelaide, SA 5005, Australia

E-mail: stephen.bell@adelaide.edu.au

[b] School of Chemistry and Molecular Bioscience, University of Queensland, St Lucia, Qld 4072, Australia

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

* Corresponding Author; Stephen G. Bell; E-mail: E-mail: stephen.bell@adelaide.edu.au; Fax, +61 8 8313 4380

Supplementary Information

Contents

Experimental

Materials

Substrates, reagents and derivatization agents were purchased from Sigma-Aldrich, Tokyo Chemical Industry Co, Ltd, Fluorochem Ltd or Enamine Ltd. HPLC grade solvents, TFA and salts were from Sigma-Aldrich, Ajax Finechem and Chem-Supply. Enzymes, NADH and molecular biology reagents were supplied by Sigma-Aldrich and Astral Scientific. Crystallization reagents and Paratone-N oil were from Hampton Research. MicroMounts, MicroLoops and MicroTools were purchased from MiTeGen, LCC. *E. coli* BL21 (DE3) competent cells were from New England Biolabs. The composition of the *E. coli* growth media and buffers is described in Table S1. Ferredoxin reductase (HaPuR) and ferredoxin (HaPux) were prepared and quantified as described previously.[1]

Instrumentation

Cells were lysed using a Vibra-Cell VC 505 ultrasonic processor (Sonics & Materials, USA) with a CV334 probe at 70% amplitude. A Cary 60 UV-Vis spectrophotometer (Agilent Technologies) coupled to a Peltier unit was used to record UV-Vis spectra and kinetic traces at 30 ± 0.5 ˚C; a quartz cuvette with a path length of 10 mm was used. Analytical high-performance liquid chromatography was performed using either a Shimadzu Prominence HPLC system or an Agilent 1260 Infinity system equipped with a Kinetex XB-C18 reversed-phase liquid chromatography column (100 Å pore size, 250 × 4.6 mm, 5 μm; Phenomenex). A gradient of 0-50% or 20-95% AcCN in water (with 0.1% TFA) over 20 minutes was used to elute the samples at a rate of 1 mL min^{-1} and the eluate was monitored at 254 nm. Preparative-scale HPLC was performed using a Kinetex EVO C18 reversed-phase column (100 Å pore size, 150 × 21.2 mm, 5 μm; Phenomenex) using a flow rate of 5 mL min^{-1}. Gas chromatography-mass spectrometry (GC-MS) was performed using a Shimadzu GC-17A gas chromatograph with a GCMS-QP5050A mass spectrometer detector in electron ionization (EI) mode; the column used was a DB-5MS column (30 m × 0.25 mm, 0.25 μm; Agilent). The interface and injection port temperatures were set to 280 and 250 ˚C. For the first 3 min the column temperature was held at 120 ˚C and then increased to 200 ˚C at a rate of 5 ˚C min^{-1}. The column was held at this temperature for 2 min, and the temperature was then increased to 250 ˚C at a rate of 20 ˚C min^{-1}. Analysis was also performed using a Shimadzu GC-2010 gas chromatograph equipped with an autoinjector and a GCMS-QP2010S detector; the column used was a DB-5MS UI column. The interface and injection port temperatures were held at 280 and 250 ˚C. The column was held at 120 ˚C for 3 min, and the temperature was then increased to 140 ˚C at a rate of 5 ˚C min^{-1} and held at 140 ˚C for 9 min. The temperature was then raised to 240 ˚C at 5 ˚C min^{-1}.

Table S1 Composition of *E. coli* growth media and buffers

Medium or buffer	Constituents (L^{-1})
LB broth (Lennox)	Sodium chloride (5 g), Bacto yeast extract (5 g), Bacto tryptone (10 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)
Super optimal broth with catabolite repression (SOC)[2]	Tryptone (20 g), yeast extract (5 g), NaCl (0.584 g), KCl (0.186 g), $MgCl_2$ (0.952 g), $MgSO_4$ (1.204 g) and 20 mM glucose
LB agar	Sodium chloride (5 g), Bacto yeast extract (5 g), Bacto tryptone (10 g), Agar No. 1 (15 g)
Buffer T	50 mM Tris-HCl buffer, pH 7.4, 1 mM DTT (dithiothreitol)

Production and purification of CYP199A4

The pET28a plasmid DNA containing the CYP199A4 gene (100 ng) was transformed into *E. coli* BL21 (DE3) competent cells. The bacteria were then plated on an LB agar plate containing kanamycin (30 mg L^{-1}) and incubated at 37 ˚C overnight (for ~20 h). Six 750 mL volumes of LB broth containing kanamycin (LB_{kan}) were each inoculated with a single colony of *E. coli* and the bacteria were grown for ~16 h at 37 ℃ and 95 rpm. The bacteria were then incubated for 30 min at a reduced temperature (18 ˚C) before 2% v/v EtOH and 0.02% v/v benzyl alcohol were added to induce the expression of chaperones which can improve protein folding.[3, 4] The medium was also supplemented with 2 mL of trace elements solution (Table S1). The culture was incubated for a further 30 minutes at 18 ˚C before protein expression was induced by addition of ~30 μM IPTG (isopropyl β-D-1-thiogalactopyranoside). After incubating the culture for ~72 hours at 18 ˚C and 95 rpm, the red cell pellet was harvested by centrifugation (5000 g, 10 min, 4 ˚C) and stored frozen at -20 ˚C.

Before cell lysis, the frozen cell pellet was thawed and resuspended in 200 mL of buffer T (Table S1). The cells were lysed by sonication (sixty 20 s pulses) on ice with manual stirring, with a 40 s interval between each 20 s pulse to prevent overheating. The cell debris was separated from the protein-containing supernatant by centrifugation (35,000 g, 30 min, 4 ˚C). The P450 was then isolated by precipitation with ammonium sulfate. Precipitated protein was collected by centrifugation (20,400 g, 15 min, 4 ˚C). The 0-30% fraction was discarded and the 30-60% ammonium sulfate fraction, which contained the P450, was redissolved in buffer T. A Sephadex G-25 coarse grain column (250 mm × 40 mm; GE Healthcare) was used to desalt the protein, using

buffer T as the eluent. The protein solution was then purified on a DEAE Fast Flow Sepharose ion-exchange column (XK 50, 200 × 40 mm; GE Healthcare). A gradient of 100-400 mM KCl in buffer T was used to elute the P450 at a rate of 6 mL min^{-1}. Red fractions were pooled and concentrated by ultrafiltration using a Vivacell 100 centrifugal concentrator with a 10 kDa molecular weight cut-off membrane (Sartorius) and the protein was then desalted using a Sephadex G-25 medium grain column (250 mm × 40 mm; GE Healthcare) using buffer T as the eluent. The protein solution was concentrated to ~10 mL and further purified on a Source Q column (XK26, 80 x 30 mm; GE Healthcare). A gradient of 0-160 mM KCl in buffer T was used to elute the protein at a rate of 6 mL min^{-1}. The purity of the fractions was assessed by measuring the Reinheitszahl, RZ = $Abs_{420\ nm}/Abs_{280\ nm}$, and fractions with a RZ value >1.8 were combined and concentrated to >40 μM. Before crystal growth experiments the CYP199A4 enzyme was passed down a HiPrep Sephacryl S-200 HR size-exclusion column (60 cm x 16 mm; GE Healthcare). The protein was eluted using buffer T as the eluent at a rate of 1 mL min^{-1}. The concentrated protein was centrifuged at 7000 rpm for 10 min to remove particulate matter, and autoclaved glycerol was added to a final concentration of 40-50% v/v before the solution was filter-sterilised through a 0.2 μm filter. The purified protein was stored at -20 ˚C. Before use, glycerol was first removed from the purified proteins using a PD-10 desalting column (GE Healthcare).

Spin-state shift analysis

Binding of a substrate typically displaces the heme-bound water shifting the heme iron from low spin (LS; λ_{max} = ~418 nm) to high spin (HS; λ_{max} = ~390 nm).[5-7] To measure the spin-state shift, 1 μL aliquots of a 100 mM substrate stock solution in EtOH or DMSO (see Table S2 for details) were successively added to 500 μL of ~1 μM P450 in Tris-HCl buffer (50 mM, pH 7.4), recording the UV-Vis spectrum after each addition, until there was no further shift. The fraction of the P450 enzyme in the high-spin state (% HS) was estimated by visually comparing the spectrum of the substrate-bound enzyme to a series of spectra of $P450_{cam}$ which show varying percentages of HS and LS enzyme.[8],[9]

Binding constant analysis

Titrations were performed to determine substrate binding affinity (K_d).[10] The UV-Vis spectrophotometer was first referenced against substrate-free 2.5 mL of ~2 μM P450 in Tris-HCl buffer (50 mM, pH 7.4). Aliquots (0.5-5 μL) of 1, 10, or 100 mM substrate stock solutions in EtOH/DMSO (Table S2) were then added using a 5 μL Hamilton syringe. After each addition of substrate, the difference spectrum was recorded from 300 to 600 nm. The type I difference spectra all contained a peak at ~390 nm and trough at ~420 nm.[11] Aliquots of substrate were added until

no further spectral shift occurred and no more than 10 μL of each stock solution was added to avoid diluting the enzyme. Titrations were performed in triplicate. The peak-to-trough absorbance difference, ΔA ($A_{peak} - A_{trough}$), was then plotted against substrate concentration.

All substrates investigated here bind weakly so the data were fitted to the hyperbolic equation (Equation 1).

$$\Delta A = \frac{\Delta A_{max} \times [S]}{K_d + [S]} \quad (1)$$

In this equation, ΔA is the peak-to-trough absorbance difference, ΔA_{max} is the ΔA at infinite substrate concentration, and [S] is the concentration of substrate added to the P450. K_d is the dissociation constant of the P450-ligand complex.

In vitro NADH activity assays

In vitro NADH turnovers[12] were performed at 30 ˚C and contained 0.5 μM CYP199A4 (WT or mutant), 5 μM HaPux, 0.25 μM HaPuR and 100 ng μL^{-1} bovine liver catalase in oxygenated Tris-HCl buffer (50 mM, pH 7.4) in a total volume of 1.2 mL. Catalase was added to remove H_2O_2 generated via uncoupling in assays where this uncoupling pathway was not being monitored.[13] The absorbance at 340 nm was set to zero and the mixture was incubated at 30 ˚C for 2 min before NADH was added to a concentration of ~320 μM, corresponding to an absorbance of ~2.0. The rate of NADH background oxidation (the 'leak' rate) was measured before initiating the reaction. To start the reaction, 1 mM substrate was added from a 100 mM stock in EtOH/DMSO (Table S2) and NADH oxidation was monitored at 340 nm.[14] The rate of NADH consumption (N) by the P450 enzyme in units of (μM-NADH) $(\mu M\text{-P450})^{-1}$ min^{-1} (abbreviated to min^{-1}) was calculated from the slope ($\Delta A_{340}/\Delta t$) of the graph of A_{340} versus time using an extinction coefficient of ε_{340} = 6.22 mM^{-1} cm^{-1} (Equation 2).

$$N = \frac{\frac{\Delta A_{340}}{\Delta t} \times 1000 \times 2}{6.22} \quad (2)$$

All experiments were performed at least three times and the mean and standard deviation reported. Control reactions were also performed in which either the P450 or NADH was omitted from the turnover mixture (replaced with the same volume of buffer).

Analysis of metabolites

Products were identified and quantitated by HPLC or GC-MS. To prepare turnover mixtures for HPLC, 132 μL of the *in vitro* turnover mixture was mixed with 66 μL of AcCN and 2 μL of internal standard (10 mM 9-hydroxyfluorene (9-OHFlu) in EtOH) and centrifuged to remove particulate

matter. HPLC analysis was performed as described above. Products were identified by co-elution with authentic samples of the expected oxidation products.

Samples were prepared for GC-MS analysis by first adding 10 μL of internal standard (10 mM 9-OHFlu) and 3 μL of 3 M HCl to 990 μL of the turnover mixture. After extraction with EtOAc (3 × 400 μL), the organic extracts were pooled, dried over anhydrous $MgSO_4$ and solvent was evaporated under N_2. The residue was redissolved in anhydrous AcCN (150 μL) and the substrate, product(s) and internal standard were converted into TMS derivatives by treatment with excess derivatization reagent (>35 μL BSTFA + TMCS, 99:1). The mixtures were left for 2 hours at 37˚C prior to analysis by GC-MS. GC-MS analysis was performed using the instruments and methods described above.

For the 3-methylaminobenzoic acid turnovers, a modified method was used. After extraction, the dry samples were redissolved in anhydrous EtOAc (300 μL) and 45 μL of derivatization reagent (BSTFA + TMCS, 99:1) was added. The samples were left for 36 h to fully convert 3-aminobenzoic acid into the doubly derivatized form. Samples were then analyzed by GC-MS.

Calibration curves were constructed to quantify the product. Solutions of metabolites with concentrations of 20, 50, 100, 200 and 500 μM were prepared for HPLC analysis in the same way as the turnovers. A plot of product peak area/IS peak area *vs.* product concentration was then constructed. The calibration factor was the slope of this graph and had units of μM^{-1}. 3-Hydroxymethylbenzoic acid was quantified using the substrate, 3-methylbenzoic acid.

Analysis of catalytic activity

Coupling efficiency (C) is the percentage of NADH consumed that was channelled into substrate oxidation (Equation 3).

$$\mathrm{C} = \frac{[product]}{[NADH\ consumed]} \times 100\% \qquad (3)$$

The product formation rate (PFR) was calculated from the NADH consumption rate (*N*) and the coupling efficiency (C) (Equation 4).[15] The PFR has units of (μM-product) $(\mu M\text{-P450})^{-1}$ min^{-1} (abbreviated to min^{-1}).

$$\mathrm{PFR} = N \times \frac{\mathrm{C}}{100} \qquad (4)$$

Solvents used to prepare stock solutions of substrates and products

Compound	**Solvent**	**Concentration**
3-EthoxyBA	DMSO	100 mM
3-HydroxyBA	EtOH	100 mM
3-*tert*-ButylBA	DMSO	100 mM
4-*tert*-ButylBA	EtOH	100 mM
3-MethylaminoBA	EtOH	100 mM
3-AminoBA	DMSO/EtOH	100 mM
3-MethylthioBA	DMSO/EtOH	100 mM
3-MethylsulfinylBA	DMSO/EtOH	10 mM
3-MethylBA	DMSO/EtOH	100 mM
4-Hydroxy-3-methylBA	DMSO/EtOH	100 mM
4-MethylBA	DMSO	100 mM
2-MethylBA	DMSO/EtOH	100 mM
3-FormylBA	DMSO/EtOH	100 mM
Isophthalic acid	DMSO/EtOH	100 mM
4-FormylBA	DMSO	100 mM
Terephthalic acid	DMSO	100 mM
3-MethoxyBA	EtOH	100 mM

Quantification of H_2O_2

Hydrogen peroxide formation during *in vitro* NADH turnovers was quantitated using the HRP/phenol/4-AP assay.[12, 16] This assay detects H_2O_2 at levels as low as 1 μM.[16] *In vitro* NADH activity assays were performed without catalase and H_2O_2 was quantified immediately after all NADH had been consumed. The turnover mixture (400 μL) was mixed with 200 μL of 50 mM phenol and 200 μL of 5 mM 4-AP in Tris-HCl buffer (50 mM, pH 7.4). After the absorbance of the mixture at 510 nm had been set to zero, 1 μL of 20 mg mL^{-1} HRP in Milli-Q water was added. The concentration of H_2O_2 was calculated from the absorbance of the resulting mixture at 510 nm.

If H_2O_2 is present, it reacts rapidly with phenol and 4-aminoantipyrine to generate pink-coloured quinoneimine dye, catalyzed by HRP.[17] Quinoneimine was quantified using ε_{510} = 6580 M^{-1} cm^{-1} (Equation 5).[12]

$$[Quinoneimine\ dye] = \frac{A_{510} \times 10^6}{6580}\mu M \qquad (5)$$

The formation of one quinoneimine molecule requires two molecules of H_2O_2. The concentration of H_2O_2 in the original (undiluted) turnover mixture is therefore given by four times the calculated quinonimine concentration (Equation 6).

$$[H_2O_2] = 4 \times [Quinoneimine\ dye]_{\text{diluted mixture}} = \frac{A_{510} \times 10^3}{1.645}\mu M \qquad (6)$$

Detection of H_2O_2 after incubation with turnover components

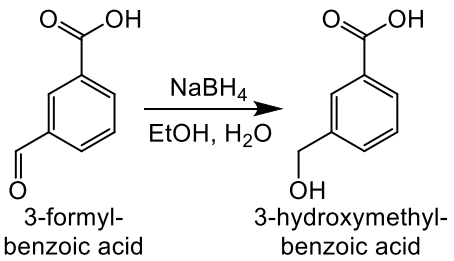
To assess if H_2O_2 generated via uncoupling could be lost before slower *in vitro* turnover reactions were complete, ~300 μM H_2O_2 was added to a control turnover mixture (1.2 mL) containing 5 μM HaPux, 0.25 μM HaPuR and 0.5 μM CYP199A4 in Tris-HCl buffer (but no NADH, substrate or catalase) and the mixture was incubated at 30 ˚C for 60 min. H_2O_2 levels were measured (as already described) at time points of 2, 10, 20, 40 and 60 minutes. Measurements were made in triplicate. H_2O_2 was also incubated with the individual components of the turnover mixture in Tris-HCl buffer.

Characterization of 4-*tert*-butylbenzoic acid metabolite

For characterization of the turnovers of 4-*tert*-butyl-benzoic acid, the product was derivatized into the equivalent methyl ester using excess N,N-dimethylformamide dimethyl acetyl (DMF-DMA). To isolate enough product for characterization a whole-cell turnover was used to scale up the reaction. pETDuet-HaPux-HaPuR and pRSFDuet-HaPux-CYP199A4 plasmids were transformed into BL21(DE3) and grown in 200 mL LB broth, as described above. The culture was grown for a further 24 hours, and then harvested by centrifugation (5,000 *g*, 10 min). The harvested cells were resuspended in 200 mL *E. coli* minimal media (EMM; K_2HPO_4 7 g, KH_2PO_4 3 g, Na_3citrate 0.5 g, $(NH_4)_2SO_4$ 1 g, $MgSO_4$ 0.1 g, 20 % glucose (20 mL) and glycerol (1 % v/v) per litre in a 2 L baffled flasks. Substrate (2 mM) was added to the whole-cell reaction and the mixture was shaken at 160 rpm at 30 ^oC for 20 h. The supernatant (200 mL) was then acidified, extracted in ethyl acetate (3 x 100 mL), washed with brine (100 mL) and dried with $MgSO_4$. The organic extracts were pooled and the solvent was removed under reduced pressure before derivatization with DMF-DMA as above. The resulting product was purified using silica column chromatography which was then characterised by NMR and GC-MS. The purified product was also evaluated by GC-MS to assign it to the observed metabolite peak.

NMR spectra were acquired on a Varian Inova spectrometer spectrometer operating at 600 MHz for ^1H and 151 MHz for ^{13}C.

Synthesis of 3-hydroxymethylbenzoic acid



To a stirred solution of 3-formylbenzoic acid (9 mg, 0.06 mmol) in ethanol (1 mL) was added dropwise a solution of sodium borohydride (1.7 mg, 0.045 mmol) in water (1 mL) and the resulting mixture was stirred for 40 min. Water (3 mL) and 50 μL conc. HCl were then added and the reaction was analyzed by HPLC.

X-ray crystallography

A HiPrep Sephacryl S-200 HR size-exclusion column (60 cm x 16 mm; GE Healthcare) was used to purify the CYP199A4 protein for crystallization. The protein was eluted using Tris-HCl buffer (50 mM, pH 7.4) at a flow rate of 1 mL min^{-1}. The purified protein (RZ >2) was concentrated to 40 mg mL^{-1} (~900 μM) in 50 mM Tris-HCl buffer pH 7.4 by ultrafiltration using a Microsep Advance centrifugal device (10 kDa MWCO, Pall Corporation). Substrate was then added to a concentration of 1-4 mM from a 100 mM stock in DMSO or EtOH (Table S2).

Diffraction-quality crystals were obtained using optimised conditions for the WT enzyme reported previously: 200 mM magnesium acetate, 100 mM Bis-Tris buffer (adjusted with acetic acid to pH 5.0-5.75) and 20-32% w/v polyethylene glycol (PEG) 3350.[12, 18] Crystallization of the protein was accomplished using the hanging-drop vapour-diffusion method using a 24-well crystallization tray. Hanging drops of 1.2-2 μL of protein mixed with an equal volume of crystallization reagent were equilibrated with a 500 μL reservoir of the crystallization reagent at a temperature of 16 ˚C. Under these conditions, clusters of red plate-like crystals appeared within one week.

Single crystals were mounted onto MicroMounts or MicroLoops (MiTeGen LLC, New York, USA) and dragged through Parabar 10312 oil (Paratone-N, Hampton Research, California, USA) before they were flash-cooled in liquid N_2. High-resolution X-ray diffraction data (360 images per crystal) were obtained at 100 K at the Australian Synchrotron, beamlines MX1 or MX2.[19, 20] The exposure time was 1 s, oscillation angle 1˚ and wavelength 0.9537 Å. iMosflm[21] was used to index and integrate the data, and the data were then scaled and merged using Aimless[22] (part of the CCP4 suite of programs).[23] PhaserMR in the CCP4 suite of programs was used to solve the structure by the molecular replacement method.[24] A high-resolution (1.54-Å) structure of CYP199A4 was used as the search model (PDB: 5UVB, with the heme, chloride, 4-cyclopropylbenzoate substrate and waters deleted).[18] Coordinates and restraints for the ligand were generated using Phenix eLBOW[25] and the protein model was manually rebuilt in Coot.[26] Refinement was performed using phenix.refine,[27, 28] and this was followed by multiple rounds of manual rebuilding in Coot and refinement to improve the model. Solvent was added automatically using phenix.refine and positive density in the anion binding site of CYP199A4 was modelled as a chloride ion.[29] The Fe-S(Cys) and Fe-OH_2 bond lengths were

not restrained during refinement. The occupancy of the substrate and water ligand, if present, was refined using phenix.refine. In the final stages of refinement, waters were manually deleted if they were in low electron density or unrealistically close to other waters or residues.[30, 31] Composite omit maps[32] and feature-enhanced maps[33] were generated using Phenix to verify the presence of the ligand in the enzyme's active site. The validation tool MolProbity was used to assess the quality of the model before the structure was deposited into the Protein Data Bank (www.rcsb.org).[34] Structures were considered fully refined for deposition when they satisfied the following quality criteria: R_{work} <25%, R_{free} <25%, Ramachandran favoured >98%, poor rotamers <1%, RMSD bond lengths <0.02 Å, RMSD bond angles <2˚.[34-36] Images of the refined crystal structures were generated using PyMOL.[37]

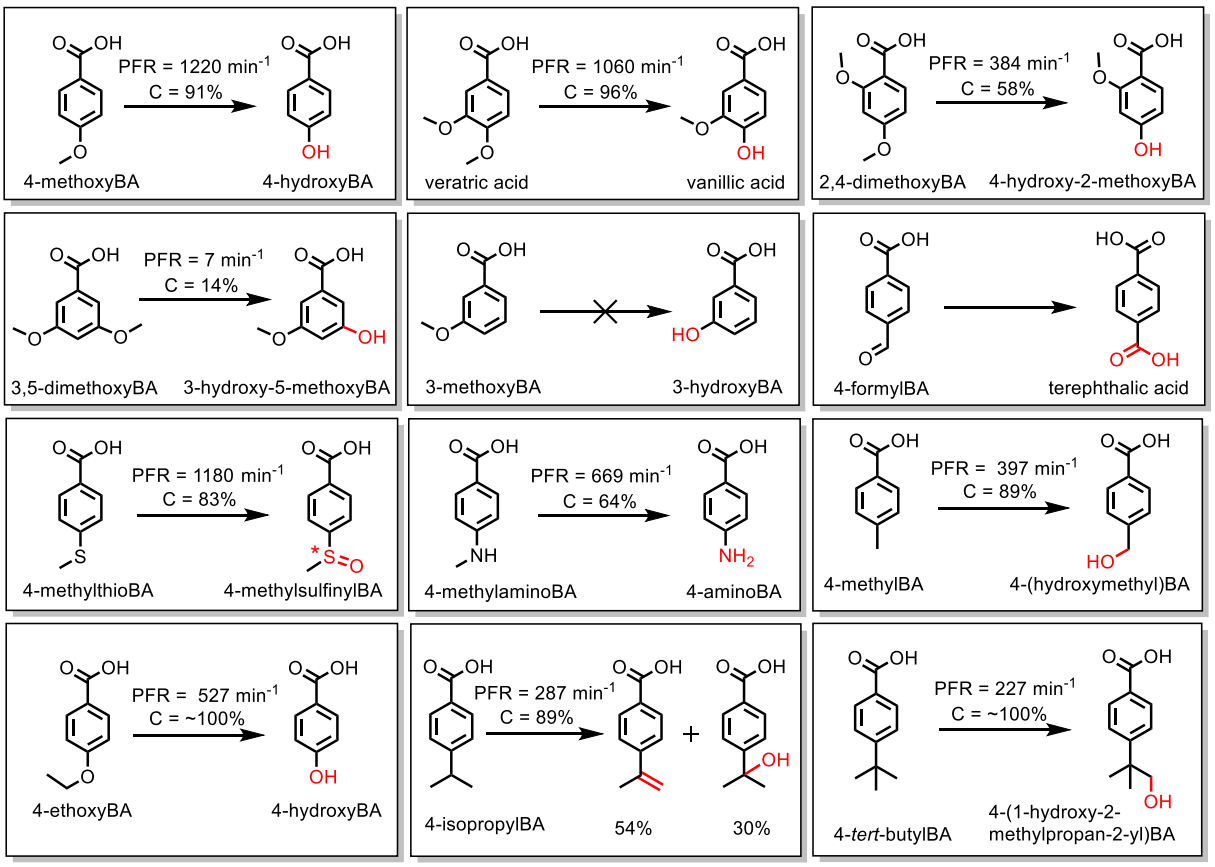
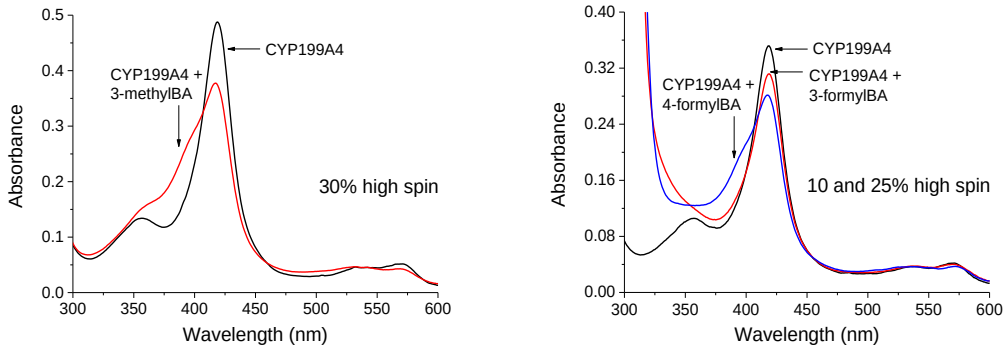


Figure S1 Selected reactions catalyzed by CYP199A4 from this and previous work.[1, 8, 12, 18, 38-40] Product formation rates (PFR) and coupling efficiencies (C) are given. Note that details of the characterization of these reactions and those of 3-methoxy, 3-methylthio and 3-methylamino-benzoic acid have been reported previously.[12, 38]

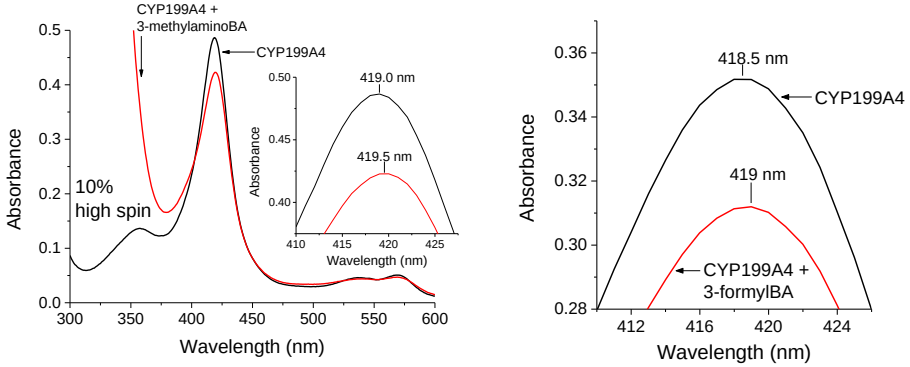


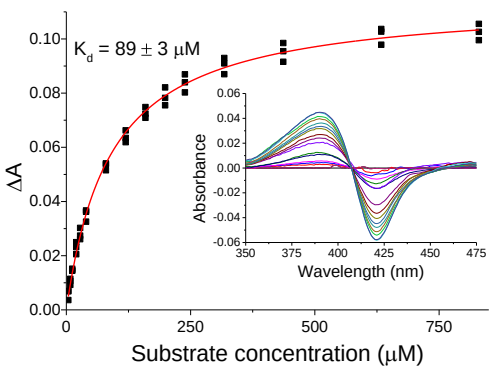
(a) **(b)**

Figure S2 Spin-state shifts of CYP199A4 induced by **(a)** 3-methylbenzoic acid and **(b)** 3-formylbenzoic acid. In black is the substrate-free form and in red is the substrate-bound enzyme. In blue is 4-formylbenzoic acid-bound CYP199A4 for comparison. The spin-state shifts of 3-methoxybenzoic acid,[38] 3-methylaminobenzoic acid and 3-methylthiobenzoic acid and the *para*-isomers have been reported previously.[12]

Table S2 Shift in Soret band position upon binding of substrates to CYP199A4

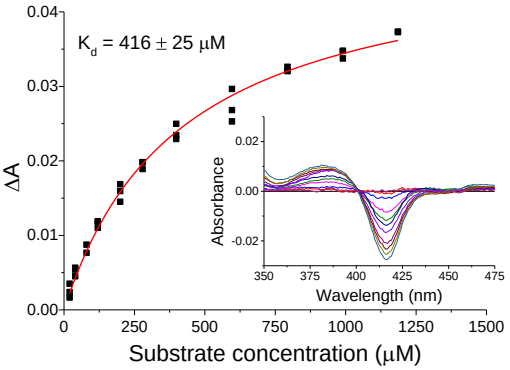
Substrate	**Shift in Soret peak position**
3-methoxyBA	2.5 nm blue shift
3-methylaminoBA	**0.5 nm red shift**
3-methylthioBA	0.5 nm blue shift
3-methylBA	2 nm blue shift
3-formylBA	**0.5 nm red shift**
4-formylBA	0.5 nm blue shift
3-ethoxyBA	22 nm blue shift
3-*tert*-butylBA	no shift
3-isopropylBA	3 nm blue shift





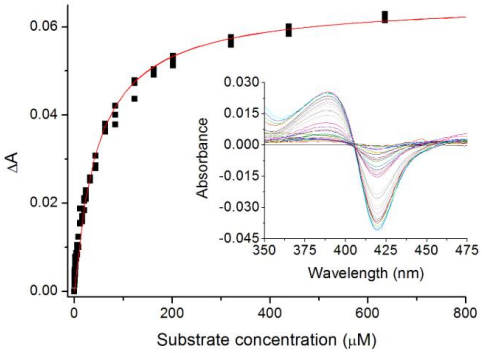
(a) 3-methylbenzoic acid

(2.2 μM, peak: 390 nm, trough: 421 nm)



(b) 3-formylbenzoic acid

(3.7 μM, peak: 386 nm, trough: 416 nm)



(c) 4-formylbenzoic acid

(1.5 μM, peak: 389 nm, trough: 419 nm)

Figure S3 UV-Vis titrations to determine the dissociation constant of CYP199A4 with benzoic acid substrates. The concentration of enzyme used and the peak and trough wavelengths are given under each graph. The dissociation constants of 3-methoxybenzoic acid,[38] 3-methylaminobenzoic acid and 3-methylthiobenzoic acid and the *para*-isomers have been reported previously.[12]

***In vitro* turnover data**

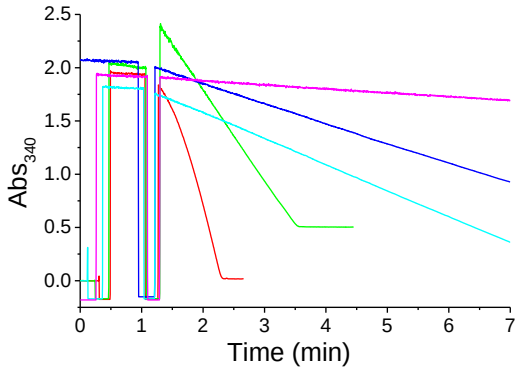


Figure S4 Rates of NADH consumption by CYP199A4 with 3-methoxybenzoic acid (**red**), 3-methylaminobenzoic acid (**green**), 3-methylthiobenzoic acid (**dark blue**), 3-methylbenzoic acid (cyan), and 3-formylbenzoic acid (magenta).

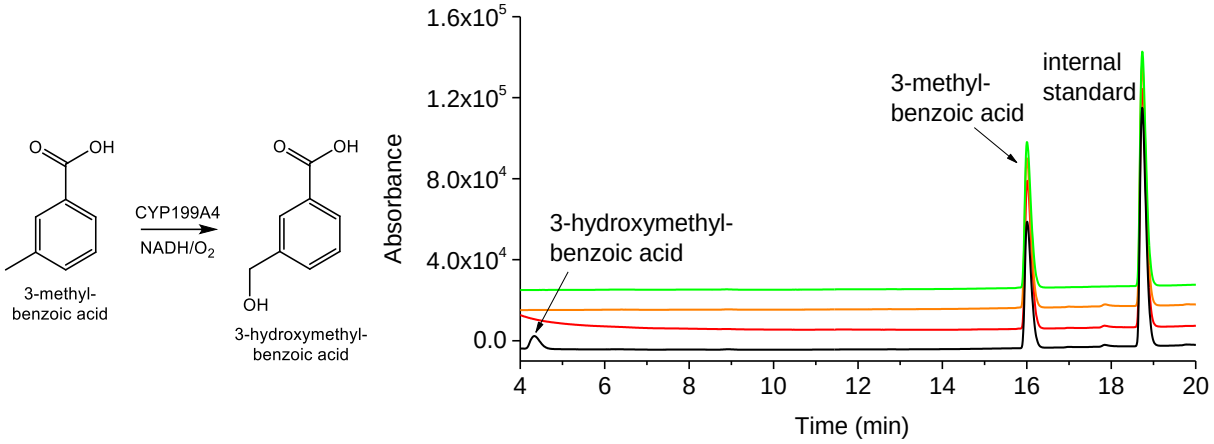
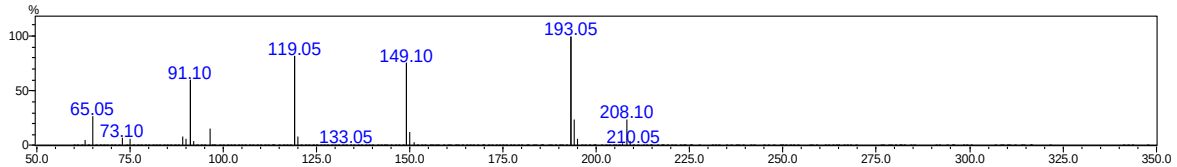
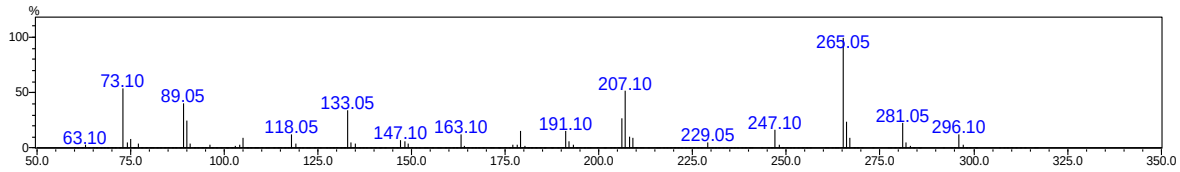


Figure S5 HPLC analysis of the reaction with 3-methylbenzoic acid. In **black** is the *in vitro* turnover, in **red** and **orange** are control reactions omitting the P450 and NADH, and in **green** is the substrate control (RT = 16 min). Product appears at RT = 4.3 min. Gradient: 20-95% AcCN in H_2O with 0.1% TFA. Detection wavelength: 254 nm (Shimadzu Prominence HPLC).

GC-MS data Mass spectral data for the CYP199A4 *in vitro* reaction with 3-methylbenzoic acid.

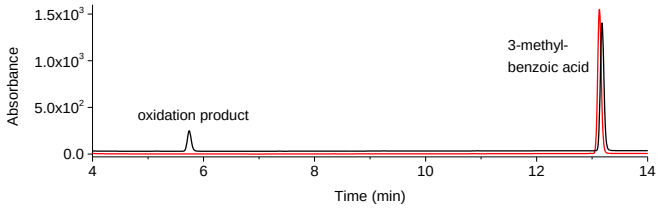


3-Methylbenzoic acid substrate (singly derivatized): *m/z* = 208.10 (exp. *m/z* = 208.0920), RT = 5.1 min.

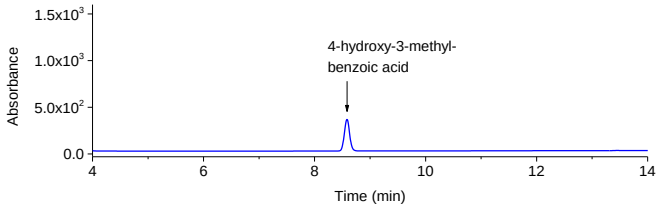


Oxidation product; 3-hydroxymethylbenzoic acid (doubly derivatized): *m/z* = 296.10 (exp. *m/z* = 296.1264), RT = 12.3 min.

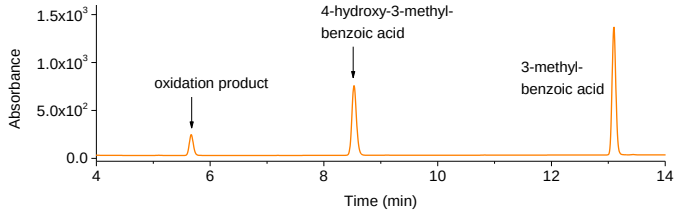
Analysis of the product of the 3-methylbenzoic acid turnover



(a) *In vitro* turnover (**black**) and control reaction omitting the P450 (**red**)



(b) Authentic 4-hydroxy-3-methylbenzoic acid



(c) *In vitro* turnover spiked with authentic 4-hydroxy-3-methylbenzoic acid

Figure S6 (a) HPLC analysis of the *in vitro* CYP199A4 reaction with 3-methylbenzoic acid. The oxidation product had a retention time of 5.7 min. **(b)** Authentic 4-hydroxy-3-methylbenzoic acid had a retention time of 8.5 min and **(c)** did not co-elute with the oxidation product. Gradient: 20-95% AcCN in H_2O with 0.1% TFA. Detection wavelength: 240 nm (Agilent 1260 Infinity). CYP199A4 could conceivably convert 3-methylbenzoic acid into 4-hydroxy-3-methylbenzoic acid via hydroxylation of the aromatic ring but these results show that the product generated is different and most likely 3-hydroxymethylbenzoic acid.

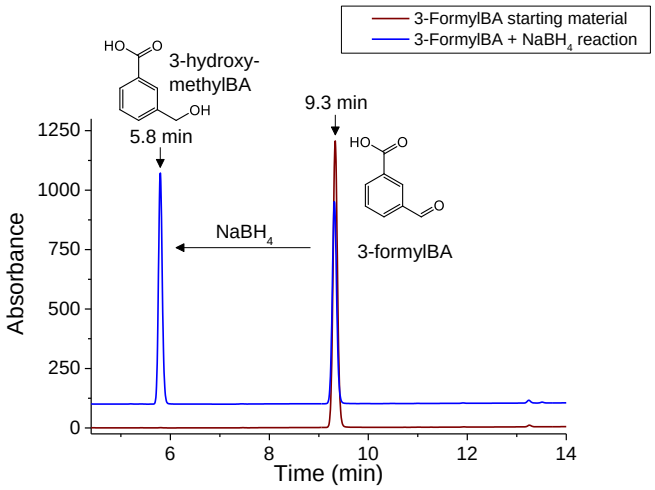


Figure S6 (d) HPLC analysis of the 3-formylbenzoic acid starting material (**wine**, RT = 9.3 min) and the product, 3-hydroxymethylbenzoic acid, arising from reduction of the aldehyde by sodium borohydride (**blue**, RT = 5.8 min). Gradient: 20 → 95% AcCN in H_2O with 0.1% TFA. Detector wavelength: 240 nm.

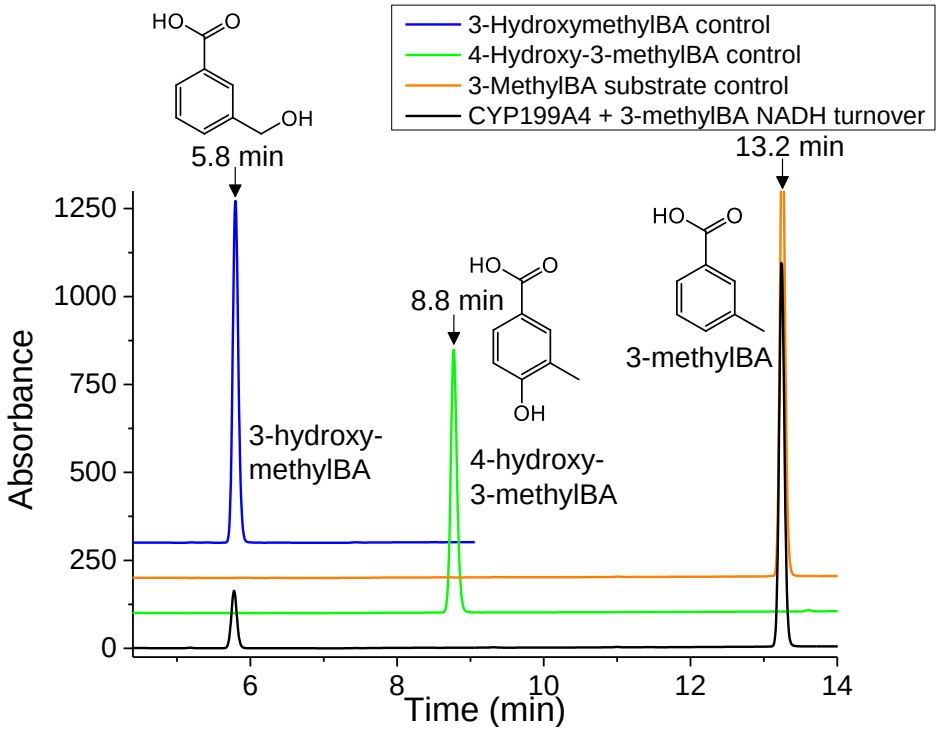


Figure S6 (e) HPLC analysis of the *in vitro* NADH-supported reaction of CYP199A4 with 3-methylbenzoic acid (**black**). In **orange** is the 3-methylbenzoic acid substrate control (RT = 13.2 min), in **blue** is chemically synthesized 3-hydroxymethylbenzoic acid (RT = 5.8 min) and in **green** is authentic 4-hydroxy-3-methylbenzoic acid (RT = 8.8 min). The P450 oxidation product (RT = 5.8 min) co-elutes with 3-hydroxymethylbenzoic acid. Gradient: 20 → 95% AcCN in H_2O with 0.1% TFA. Detector wavelength: 240 nm.

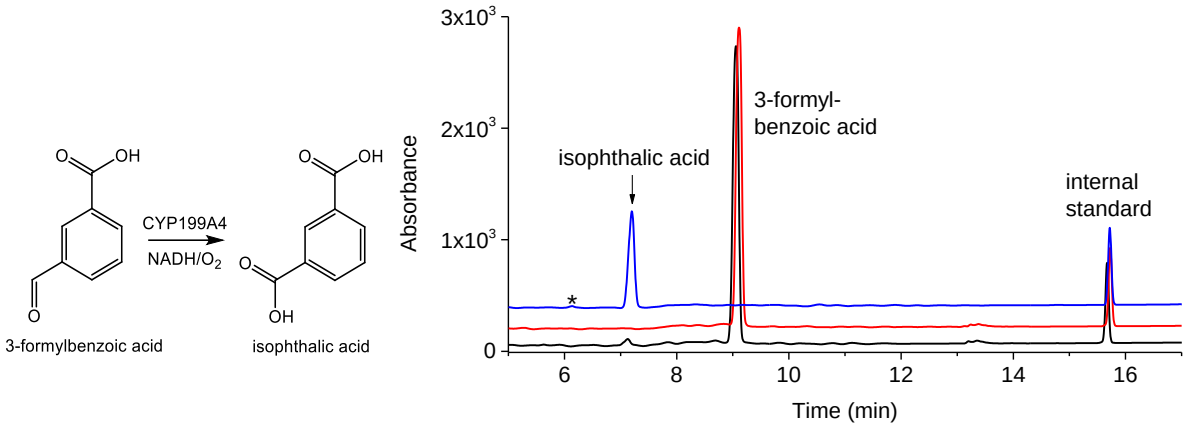
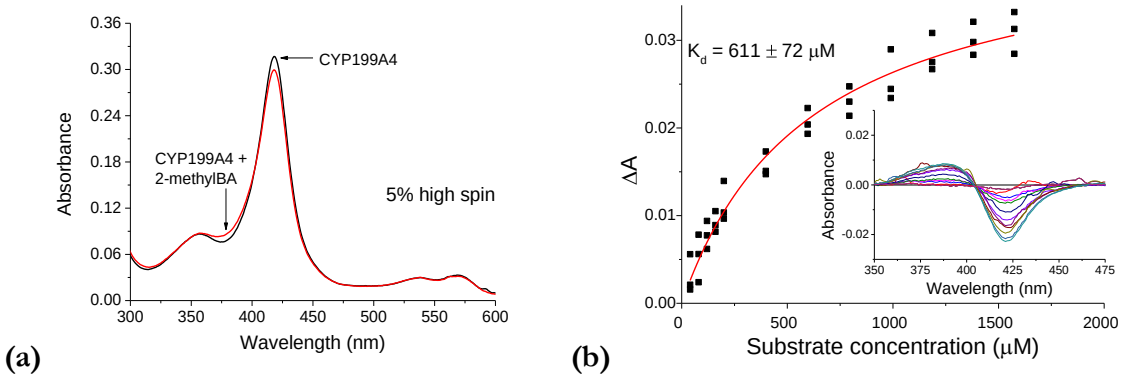


Figure S7 HPLC analysis of the CYP199A4 *in vitro* reaction with 3-formylbenzoic acid. In **black** is the *in vitro* turnover, in **red** is a control reaction performed without P450, and in **blue** is authentic isophthalic acid (RT = 7.1 min). The substrate appears at RT = 9.1 min. Gradient: 20-95% AcCN in H_2O with 0.1% TFA. Detection wavelength: 220 nm. * The sample of authentic isophthalic acid was contaminated with ~6% terephthalic acid (RT = 6.1 min).

Binding and turnover data for 2-methylbenzoic acid



(3.7 µM, peak: 388, trough: 421 nm)

Figure S8 Spin-state shift induced by binding of 2-methylbenzoic acid to CYP199A4 **(a)**, and **(b)** determination of the dissociation constant.

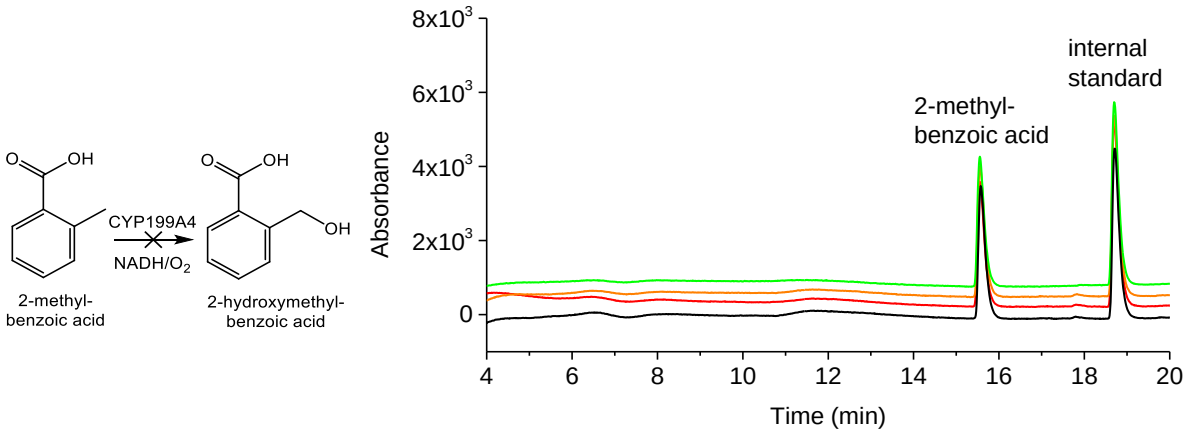


Figure S9 HPLC chromatogram of the CYP199A4 *in vitro* reaction with 2-methylbenzoic acid. No product was detected. In **black** is the *in vitro* turnover, in **red** is a control reaction omitting the P450, in **orange** is a control reaction performed without NADH, and in **green** is the 2-methylbenzoic acid substrate control (RT = 15.6 min). Gradient: 20-95% AcCN in H_2O with 0.1% TFA. Detection wavelength: 254 nm.

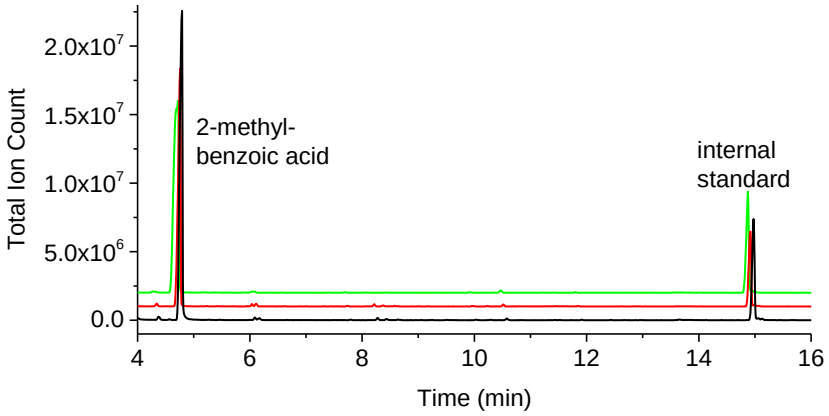
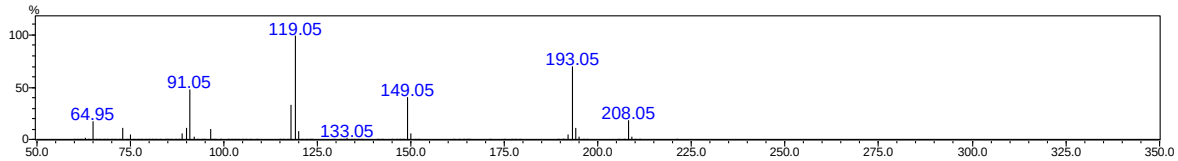


Figure S10 GC-MS analysis of the CYP199A4 *in vitro* reaction with 2-methylbenzoic acid. In **black** is the *in vitro* turnover, in **red** is a control reaction performed without P450 and in **green** is the 2-methylbenzoic acid substrate control (RT = 4.7 min). No oxidation product was detected.



2-Methylbenzoic acid substrate (singly derivatized): *m/z* = 208.05 (exp. *m/z* = 208.0920), RT = 4.7 min.

Rate of disappearance of H_2O_2 added to a control turnover mixture

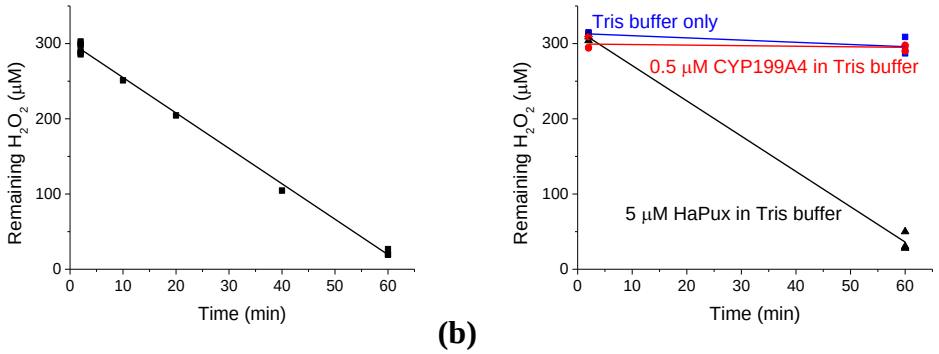


Figure S11 (a) Rate of disappearance of 300 μM H_2O_2 incubated in a control turnover mixture containing 5 μM HaPux, 0.25 μM HaPuR and 0.5 μM CYP199A4 in Tris-HCl buffer (but no NADH, substrate or catalase) at 30 ˚C. **(b)** Rate of disappearance of H_2O_2 incubated with the individual components of the turnover mixture in Tris-HCl buffer (**blue** = Tris-HCl buffer only, **red** = 0.5 μM CYP199A4 in Tris buffer, **black** = 5 μM HaPux in Tris buffer).

When the rate of NADH consumption by the P450 is slow, a portion of the H_2O_2 generated via uncoupling may be lost before the H_2O_2 concentration is measured at the end of the reaction. We would therefore be underestimating the amount of H_2O_2 generated.

To assess whether any loss of H_2O_2 occurs over time, 300 μM H_2O_2 was incubated in a control turnover mixture containing 5 μM HaPux, 0.25 μM HaPuR and 0.5 μM CYP199A4 in Tris-HCl buffer (but no NADH, substrate or catalase) at 30 ˚C for 1 h. The H_2O_2 concentration was measured at intervals using the horseradish peroxidase/phenol/4-aminoantipyrine assay. Substantial decreases in H_2O_2 concentration were observed over the sixty-minute period. No significant loss of H_2O_2 occurred, however, when H_2O_2 was incubated in Tris buffer, or Tris buffer containing only 0.5 μM CYP199A4 (**blue**, **red**). The ferredoxin appeared to be responsible for the loss of H_2O_2 (**black**).

ICM-Pro docking scores

ICM-Pro scores for **3-methylaminobenzoic acid** docked into CYP199A4 in the 'down' and 'up' orientations.[41]

Pose	Score	Vls Score	Strain kcal/mol	Steric	Torsion	Electro	Hbond kcal/mol	Hydroph kcal/mol	Surface
'Down'	-40.19	-43.22	3.028	-13.41	1	-9.149	-10.66	-3.202	6.331
'Up'	-26.73	-36.42	9.698	-6.917	1	-11.1	-10.08	-3.049	6.341

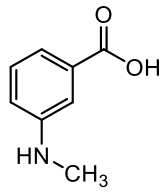


Table S3 Selected distances in Å between 3-methylaminobenzoic acid and active-site residues/heme

Distance (Å)	'Down'	'Up'
Fe - CH_3	4.1	7.4
Fe - N	4.6	7.2
N - nearest heme N	4.0	-
CH_3 - nearest heme N	3.6	-
CH_3 - nearest heme C	3.7	-
CH_3 - closest C of F182	3.8	2.8
CH_3 - closest C of F298	3.8	6.4
CH_3 - closest C of F185	5.5	4.2

ICM-Pro scores for **3-methylthiobenzoic acid** docked into CYP199A4

Pose	Score	Vls Score	Strain kcal/mol	Steric	Torsion	Electro	Hbond kcal/mol	Hydroph kcal/mol	Surface
1	-37.52	-39.79	2.272	-17.83	1	-3.519	-9.179	-3.085	6.74
2	-35.45	-37.55	2.101	-14.09	1	-4.815	-9.431	-3.076	6.728

ICM-Pro scores for **3-methylbenzoic acid** docked into CYP199A4 in the 'up' and 'down' orientations

Pose	Score	Vls Score	Strain kcal/mol	Steric	Torsion	Electro	Hbond kcal/mol	Hydroph kcal/mol	Surface
'Down'	-38.05	-39.48	1.425	-16.03	1	-4.867	-8.859	-3.182	4.942
'Up'	-36.88	-39.42	2.535	-8.641	1	-11.47	-10.04	-3.136	4.939

ICM-Pro scores for **3-methoxybenzoic acid** docked into CYP199A4 in the 'up' and 'down' orientations

Pose	Score	Vls Score	Strain kcal/mol	Steric	Torsion	Electro	Hbond kcal/mol	Hydroph kcal/mol	Surface
'Down'	-37.67	-40.63	2.963	-16.55	1	-5.051	-9.507	-3.175	6.446
'Up'	-30.02	-38.16	8.143	-8.934	1	-10.83	-10.07	-3.07	6.483

ICM-Pro scores for **veratric acid** docked into CYP199A4

Pose	**Score**	**Vls Score**	**Strain** kcal/mol	**Steric**	**Torsion**	**Electro**	**Hbond** kcal/mol	**Hydroph** kcal/mol	**Surface**
1	-34.31	-42.1	7.793	-13.86	1	-9.939	-10.17	-3.571	7.777
2	-29.19	-37.54	8.348	-7.206	1	-11.54	-10.48	-3.55	7.439

ICM-Pro scores for **3-ethoxybenzoic acid** docked into CYP199A4

Pose	**Score**	**Vls Score**	**Strain** kcal/mol	**Steric**	**Torsion**	**Electro**	**Hbond** kcal/mol	**Hydroph** kcal/mol	**Surface**
1	-38.79	-42.52	3.727	-16.68	2	-7.03	-9.792	-3.507	6.299
2	-37.55	-41.55	3.997	-15.39	2	-7.129	-9.885	-3.536	6.301

ICM-Pro scores for **3-*tert*-butylbenzoic acid** docked into CYP199A4

Pose	**Score**	**Vls Score**	**Strain** kcal/mol	**Steric**	**Torsion**	**Electro**	**Hbond** kcal/mol	**Hydroph** kcal/mol	**Surface**
1	2.3	8.14	5.84	167	2	464	3.536	3.983	4.94

ICM-Pro scores for **2-methoxybenzoic acid** docked into CYP199A4

Pose	**Score**	**Vls Score**	**Strain** kcal/mol	**Steric**	**Torsion**	**Electro**	**Hbond** kcal/mol	**Hydroph** kcal/mol	**Surface**
'Up'	-22.63	-29.79	7.163	-9.916	1	-4.492	-7.886	-3.275	6.177
'Down'	-5.564	-10.39	4.831	-7.732	1	5.909	-3.796	-3.301	6.205

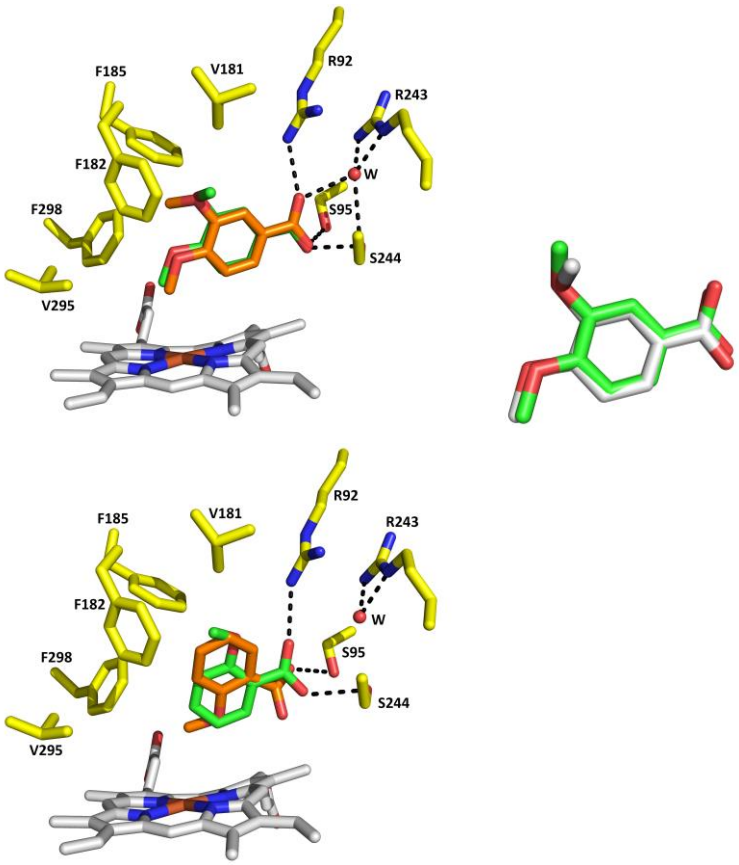


Figure S12 Veratric acid (top) and 2-methoxybenzoic acid (bottom) docked into CYP199A4 (PDB code: 5U6W). A comparison of the top-scoring docked pose (grey) with the crystal structure pose (green) of veratric acid is also shown. ICM-Pro correctly predicted the binding mode. The RMSD between the docked pose and crystallographic pose is 0.326 Å. For 2-methoxybenzoic acid the top-scoring pose **(green**) has the *ortho* methoxy group pointing away from the heme. The alternate pose (**orange**), which has the *ortho* group pointing towards the heme, scored poorly.

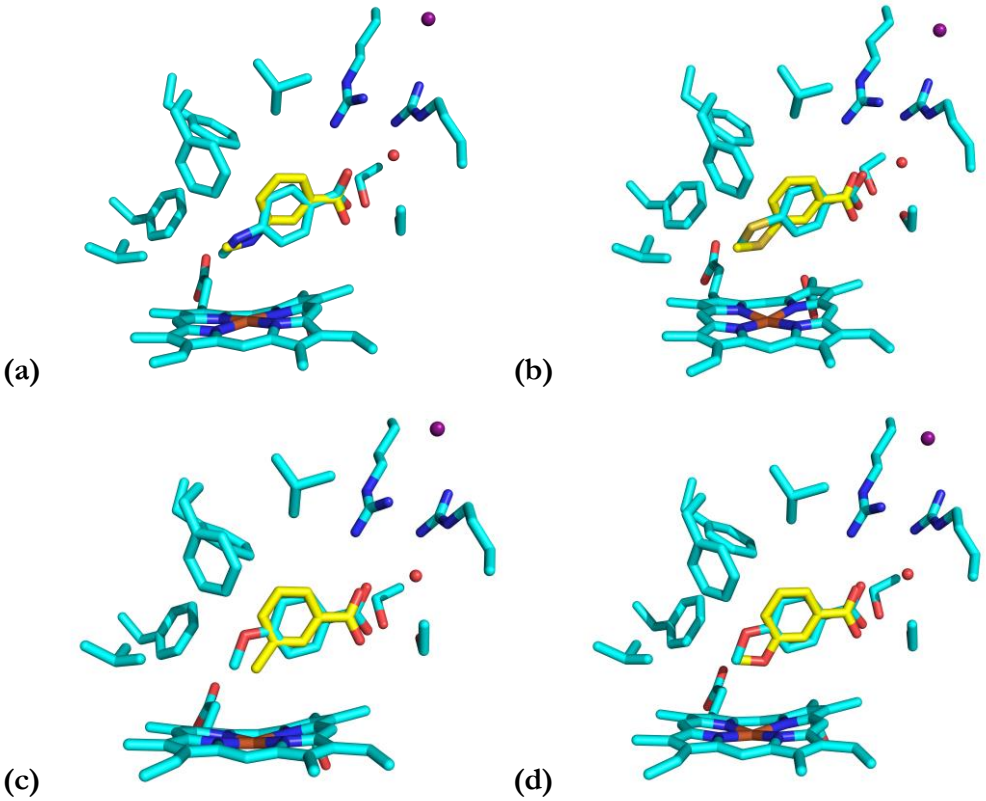


Figure S13 Docked poses of **(a)** 3-methylamino-, **(b)** 3-methylthio-, **(c)** 3-methyl- and **(d)** 3-methoxy-benzoic acid, shown in yellow, are overlaid with the crystal structures of the *para* isomers (cyan). The benzene ring of the *meta* isomers is shifted away from the heme relative to that of the *para* isomers. Note in **(c)** that the docked pose of 3-methylbenzoic acid is overlaid with the crystal structure of 4-methoxybenzoic acid as that of 4-methylbenzoic acid was not available at the time.

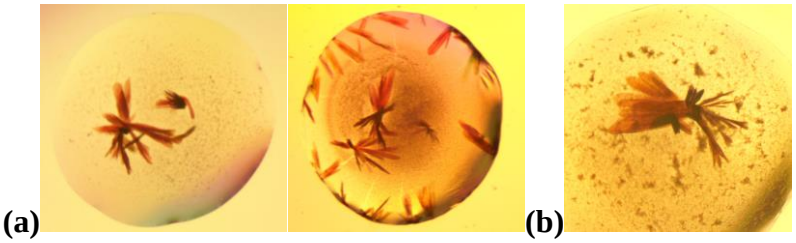


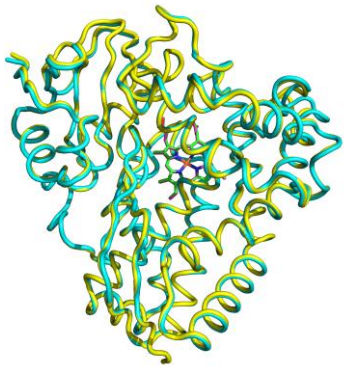
Figure S14 (a) 3-Methylthio- and **(b)** 3-methoxy-benzoate-bound CYP199A4 crystals under magnification.

Data collection and crystal structure refinement statistics

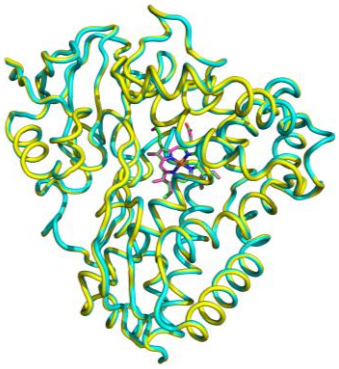
Table S4 Statistics for data collection and refinement of crystal structures of CYP199A4 with 3-methoxy-, 3-methylthio-, 3-methylamino-, 3-methyl- and 4-methyl-benzoic acid bound. Values in parentheses correspond to the highest resolution (outer) shell.

Statistic	3-methoxy	3-methylthio	3-methylamino	3-methyl	4-methyl
PDB code	6PQ6	6PQD	6PRR	6PQW	6PQS
X-ray wavelength	0.9537	0.9537	0.9537	0.9537	0.9537
Unit cell parameters	*a* = 44.5 *b* = 51.4 *c* = 78.9 α = 90.0 β = 92.5 γ = 90.0	*a* = 44.3 *b* = 51.3 *c* = 78.7 α = 90.0 β = 92.6 γ = 90.0	*a* = 44.4 *b* = 51.4 *c* = 79.5 α = 90.0 β = 92.1 γ = 90.0	*a* = 44.5 *b* = 51.2 *c* = 79.0 α = 90.0 β = 92.2 γ = 90.0	*a* = 44.2 *b* = 51.3 *c* = 78.9 α = 90.0 β = 92.1 γ = 90.0
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Molecules per asymmetric unit	1	1	1	1	1
Resolution range	43.05 – 1.66 (1.69 – 1.66)	44.25 – 1.89 (1.93 – 1.89)	44.32 – 1.67 (1.70 – 1.67)	44.42 – 1.68 (1.71 – 1.68)	44.19 – 1.60 (1.63 – 1.60)
<I/σ(I)>	12.3 (2.4)	13.6 (4.0)	16.7 (4.1)	14.3 (2.6)	10.3 (1.2)
Unique reflections	41800	28336	40980	40638	45536
Completeness of data	98.9 (94.3)	99.5 (95.1)	98.0 (62.9)	99.8 (97.3)	97.2 (91.8)
Multiplicity	6.3 (5.1)	6.7 (6.1)	6.7 (6.3)	6.9 (6.9)	6.9 (6.8)
R_{merge} (all I+ and I−)	9.5 (66.2)	10.2 (46.2)	6.2 (34.9)	7.4 (56.8)	15.9 (149.2)
R_{pim} (all I+ and I−)	4.0 (30.8)	4.2 (20.2)	2.6 (15.0)	3.0 (23.1)	6.4 (59.7)
$CC_{1/2}$	99.7 (83.2)	99.7 (90.2)	99.9 (91.7)	99.9 (91.5)	99.6 (46.0)
R_{work}	0.1555	0.1540	0.1547	0.1451	0.1598
R_{free} (5% held)	0.1847	0.1999	0.1836	0.1874	0.1939
Ramachandran favoured (%)	98.2	97.95	97.95	98.47	98.21
Ramachandran outliers (%)	0	0	0	0	0
Unusual rotamers (%)	0	0	0	0	0
RMSD bond angles (°)	0.708	0.845	0.903	1.036	0.834
RMSD bond lengths (Å)	0.003	0.006	0.007	0.009	0.004
Omit map type, contour level	composite omit map, 1.0 σ	composite omit map, 1.5 σ	composite omit map, 1.5 σ	polder omit map, 2.5 σ	composite omit map, 1.5 σ

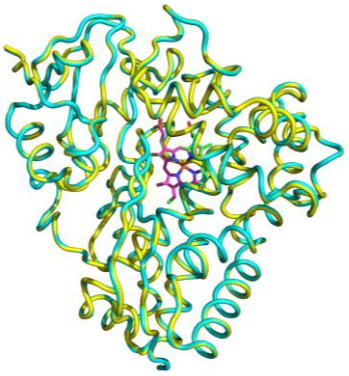
Three-dimensional fold of CYP199A4 in complex with 3-methoxy-, 3-methylamino-, 3-methylthio-, and 3-methyl-benzoic acid



(a) 3- and 4-methoxybenzoic acid (RMSD = 0.588 Å)



(b) 3- and 4-methylaminobenzoic acid (RMSD = 0.222 Å)



(c) 3- and 4-methylthiobenzoic acid (RMSD = 0.140 Å)

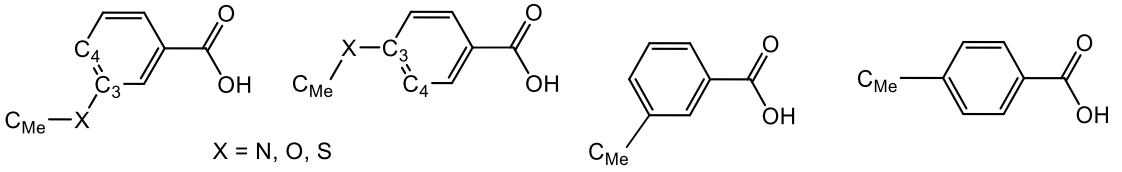
(d) 3- and 4-methylbenzoic acid (RMSD = 0.131 Å)

Figure S15 Overall fold of CYP199A4. In yellow are the crystal structures of CYP199A4 with *meta*-substituted benzoic acids bound. Superimposed with these are the structures of CYP199A4 with the corresponding *para* isomers bound (blue cartoon, green heme). The root-mean-square deviation between the superimposed structures (measured over all 393 pairs of Cα atoms) is given in parentheses under each figure.

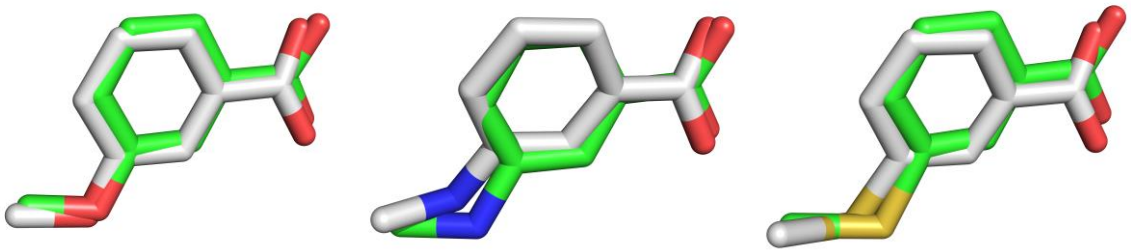
4-Methoxybenzoate-bound CYP199A4 (PDB: 4DO1) was crystallized under different conditions than the other crystal structures, accounting for the fact that the RMSD between the 4- and 3-methoxybenzoic acid structures is higher. If the 3-methoxybenzoic acid structure is superimposed with the 4-methylthiobenzoic acid structure, the RMSD is only 0.155 Å.

Table S5 Distances between the heme iron and *meta*-substituted benzoic acid substrates. Values are also given for the corresponding *para* isomers,[12] which were previously reported, for comparison. Angles and distances were measured between the substrate and the hypothetical position of the Cpd I oxygen (inserted 1.62 Å above the heme iron). Substrate names are abbreviated: **3-SMe** = 3-methylthioBA; **4-SMe** = 4-methylthioBA; **3-NMe** = 3-methylaminoBA; **4-NMe** = 4-methylaminoBA; **3-OMe** = 3-methoxyBA; **4-OMe** = 4-methoxyBA.

Distance (Å) (X = S, N, or O)	**3-SMe**	**4-SMe**	**3-NMe**	**4-NMe**	**3-OMe**	**4-OMe**	**3-Me**	**4-Me**
C_{Me} - Fe	4.4	4.5	4.4	4.1	4.4	4.1	4.4	4.5
C_{Me} - O=Fe^{IV}	3.1	3.3	3.1	3.0	3.1	2.7	3.2	3.0
X - Fe	4.3	5.0	4.2	4.2	4.4	5.2		
X - O=Fe^{IV}	3.2	3.4	3.0	2.8	3.1	3.6		
Fe - OH_2 ligand	2.3	-	2.2	-	2.4	-	2.5	-
X - OH_2 ligand	3.3	-	3.1	-	3.2	-	3.1 (C_{Me})	-
Angle (˚)								
Fe^{IV}=O-C_{Me}	136.8	131.9	129.9	124.6	136.3	140.7	130.4	154.4
Fe^{IV}=O-X	126.2	162.5	126.8	148.6	130.0	158.3	-	-
Dihedral angle C_{Me}-X-C_3-C_4	13.1	33.9	32.7	50.0	20.0	2.1	-	-



Comparison of docked ligand poses and crystallographic poses



(a) 3-methoxybenzoic acid

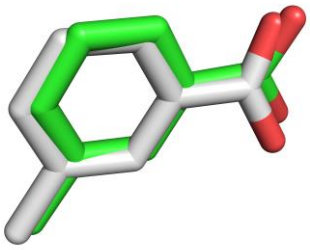
RMSD = 0.357 Å

(b) 3-methylaminobenzoic acid acid

RMSD = 0.461 Å

(c) 3-methylthiobenzoic

RMSD = 0.538 Å



(d) 3-methylbenzoic acid

RMSD = 0.423 Å

Figure S16 Comparison of docked ligand poses (grey) and crystallographic poses (green). For each substrate, the top-scoring docked pose agrees well with the crystallographic pose. The RMSD between the predicted pose and crystal structure pose is given beneath each figure.

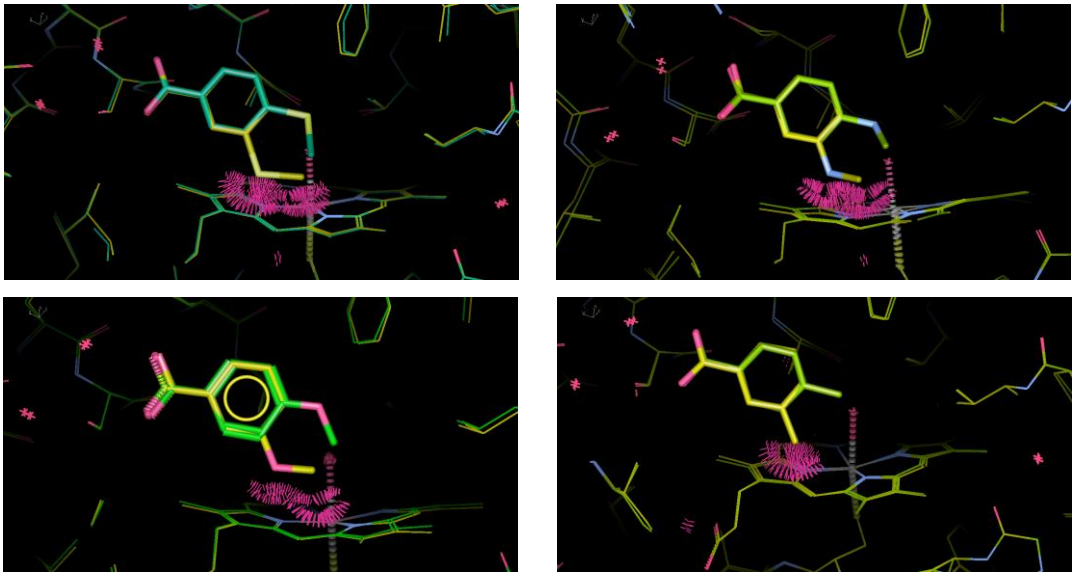


Figure S17 The benzene ring of 3-substituted benzoic acids is held further away from the heme than the benzene ring of 4-substituted benzoic acids. If the *meta*-substituted benzoic acids are repositioned so that their benzene ring is superimposable with that of 4-substituted benzoic acids, as shown above, it results in a steric clash between the *meta* substituent and the heme. MolProbity[34] uses hot-pink spikes to represent steric clashes.

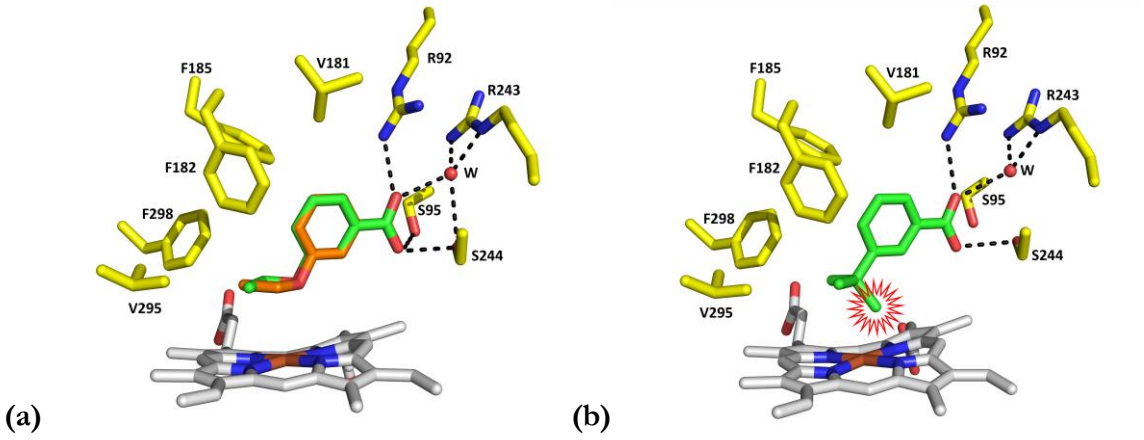
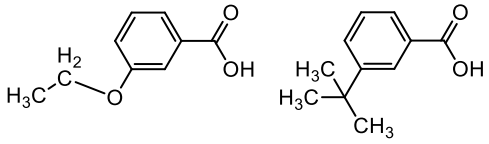


Figure S18 Bulky *meta*-substituted benzoic acids docked into CYP199A4 (PDB ID: 5U6W). The docked substrates are **(a)** 3-ethoxybenzoic acid and **(b)** 3-*tert*-butylbenzoic acid. The top scoring pose is shown in green, and lower scoring poses in orange. A significant steric clash is indicated by a red star.

Table S6 Selected distances (in Å) between the docked ligands and active-site residues/heme (Figure S18).

Docked ligand	Distance (Å)	
3-ethoxyBA	CH_2 - Fe	4.2
	CH_3 - Fe	3.6
3-*tert*-butylBA	CH_3 - Fe	3.0, 3.4, 4.7
	CH_3 - closest heme N	2.1



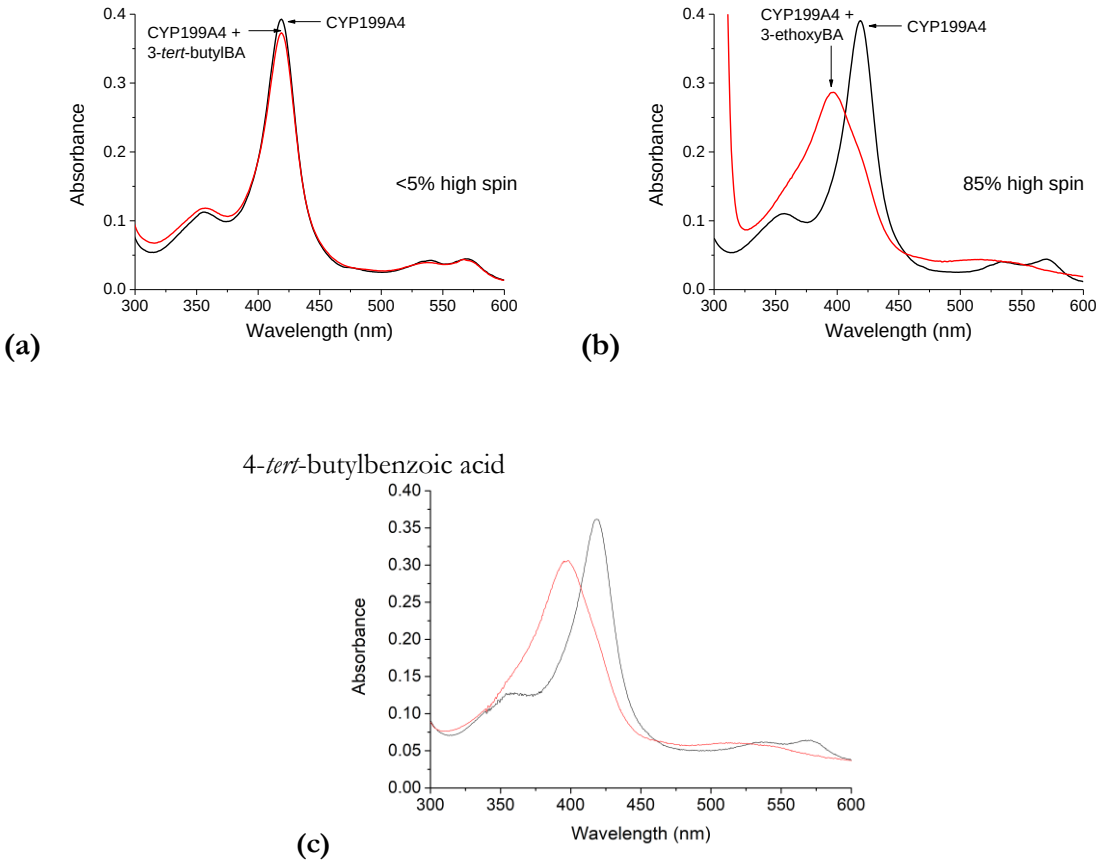
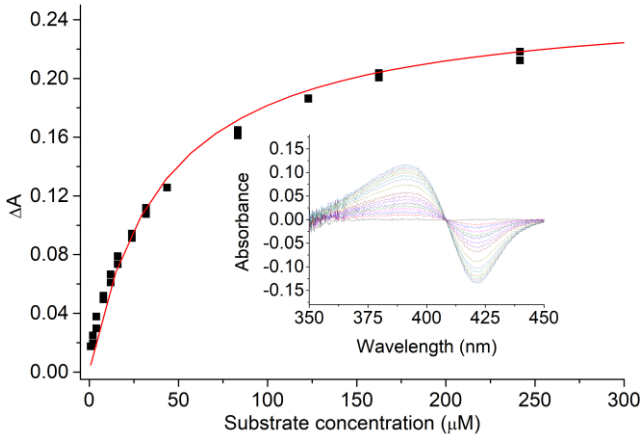
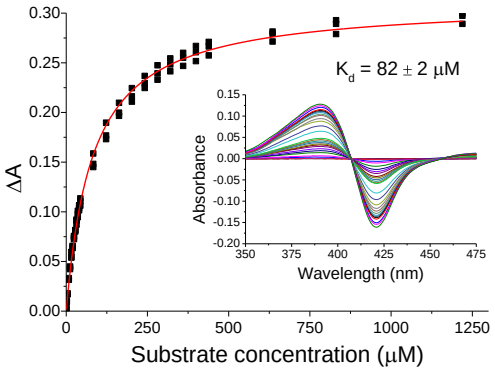


Figure S19 Spin-state shifts of CYP199A4 induced by **(a)** 3-*tert*-butylbenzoic acid, **(b)** 3-ethoxybenzoic acid and (c) 4-*tert*-butylbenzoic acid.



4-*tert*-butylbenzoic acid
(1.8 μM, peak: 421nm, trough: 391nm)

Figure S20 UV-Vis titrations to determine the dissociation constant of CYP199A4 with 4-*tert*-butylbenzoic acid.



3-ethoxybenzoic acid

(2.9 µM, peak: 391 nm, trough: 421 nm)

Figure S21 UV-Vis titrations to determine the dissociation constant of CYP199A4 with 3-ethoxybenzoic acid.

HPLC analysis of the *in vitro* CYP199A4 reaction with 3-*tert*-butylbenzoic acid.

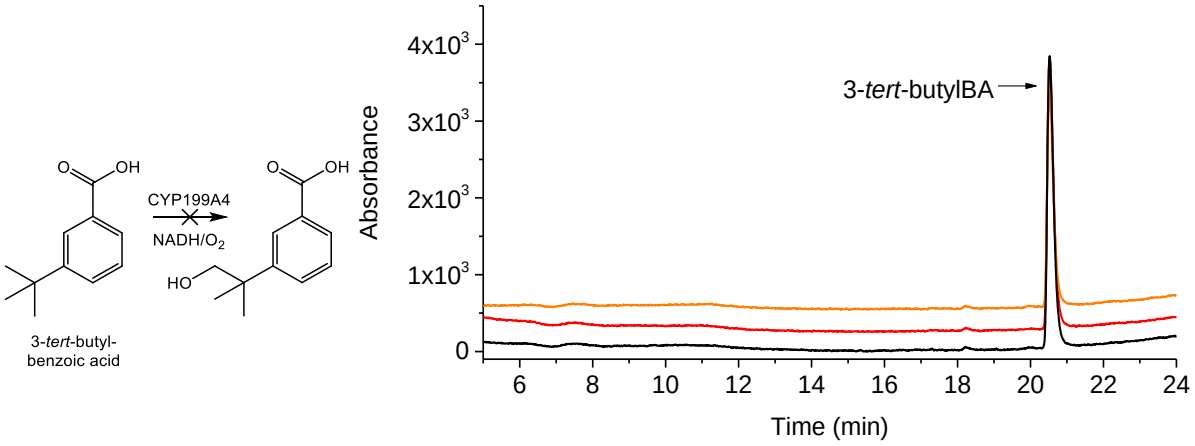


Figure S22 HPLC analysis of the CYP199A4 *in vitro* reaction with 3-*tert*-butylbenzoic acid, which yielded no product. In **black** is the turnover andin **red** and **orange** are control reactions without P450 or NADH. The substrate appears at RT = 20.5 min. Gradient: 20-9% AcCN in H_2O with 0.1% TFA. Detection wavelength: 254 nm.

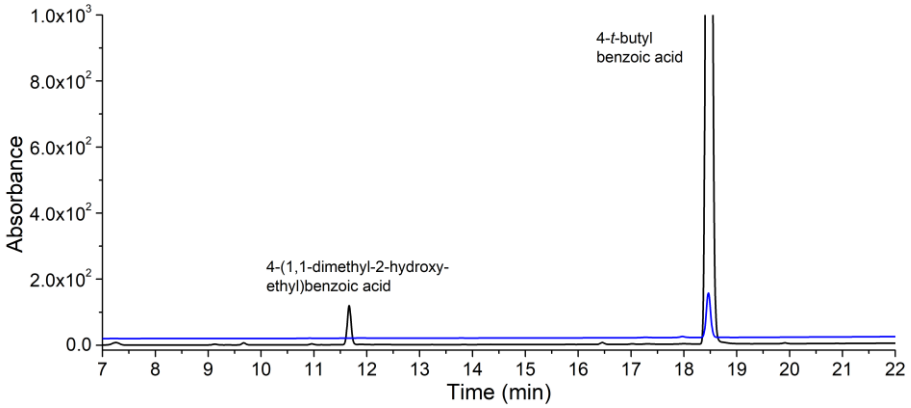


Figure S23 HPLC analysis of the *in vitro* CYP199A4 turnover of 4-*tert*-butylbenzoic acid. For clarity the chromatograms have been offset along the y axis. The chromatogram was monitored at 254 nm and the method used was a 20-95% gradient of acetonitrile to H_2O (both containing 0.1% trifluoroacetic acid, TFA). Black = CYP199A4 *in vitro* turnover and blue = substrate control. The substrate retention time was 18.5 min, and the hydroxylation product RT = 11.7 min.

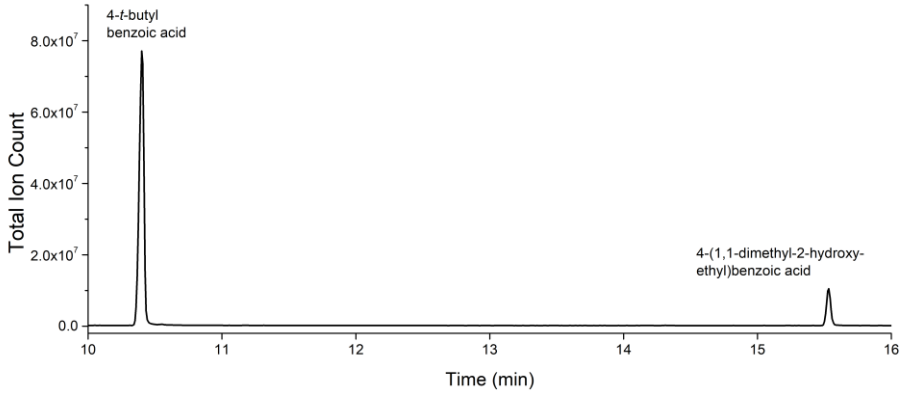
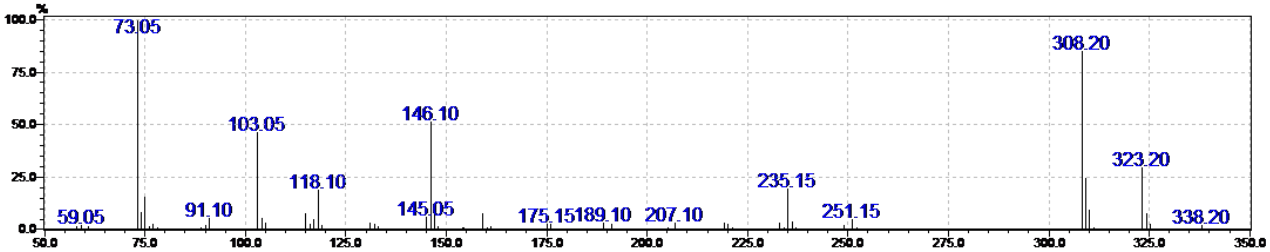


Figure S24 GC-MS analysis of the *in vitro* CYP199A4 turnover of 4-*tert*-butylbenzoic acid. The product and substrate are labelled. The substrate retention time was 10.4 min, the product RT was 15.5 min.



4-(1,1-dimethyl-2-hydroxyethyl)benzoic acid. RT = 15.5 min. Observed m/z = 338.20 versus expected 338.17.

NMR data for metabolite 4-*tert*-Butylbenzoic acid:

4-(1,1-dimethyl-2-hydroxyethyl) methylbenzoate.

^{1}H NMR (600 MHz, $CDCl_3$): δ 1.36 (s, 6H), 3.65 (s, 2H), 3.91 (s, 3H), 7.46 (d, 2H, J = 7.5 Hz), 8.00 (d, 2H, J = 7.5 Hz).[42]

^{13}C NMR (600 MHz, $CDCl_3$): δ 25.2, 40.5, 52.0, 72.7, 126.3, 128.1, 129.7, 152.0, 167.0

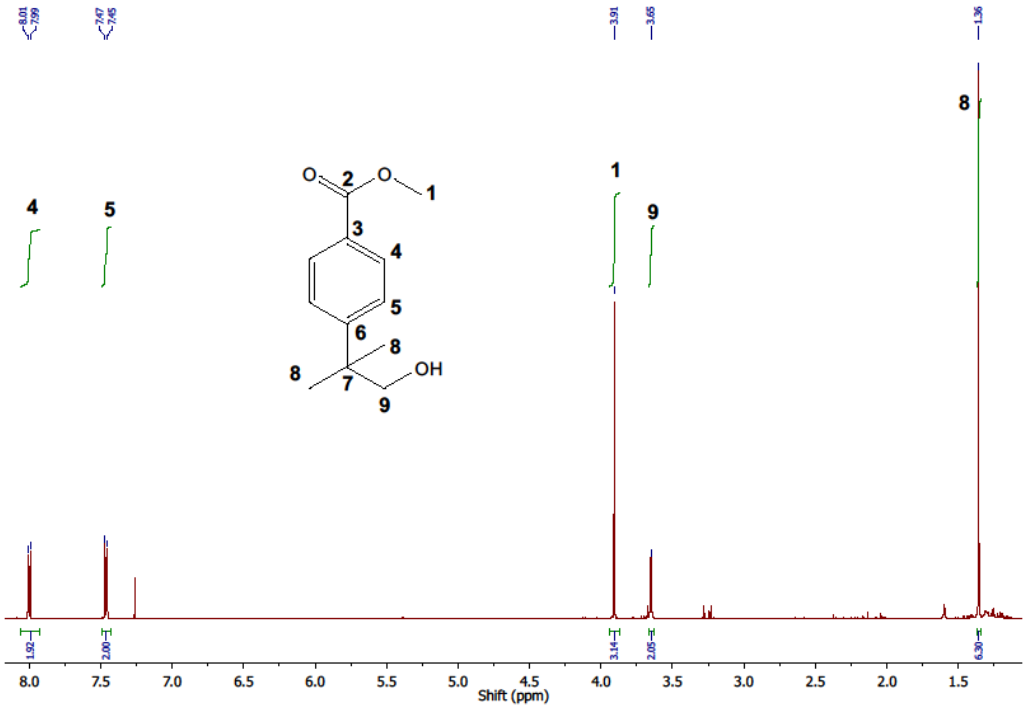


Figure S25 NMR analysis of the methyl ester derivatized metabolite of 4-*tert*-butylbenzoic acid.

Data collection and refinement statistics for the crystal structure of 3-ethoxybenzoic acid-bound CYP199A4 and overall fold of the enzyme

Table S7 Statistics for data collection and refinement of the crystal structure of CYP199A4 with 3-ethoxybenzoic acid bound. Values in parentheses correspond to the highest resolution (outer) shell.

Statistic	
PDB code	6PRS
X-ray wavelength	0.9537
Unit cell parameters	*a* = 44.6 *b* = 51.4 *c* = 79.4 α = 90.0 β = 92.2 γ = 90.0
Space group	$P12_11$
Molecules per asymmetric unit	1
Resolution range	44.56 – 2.37 (2.45 – 2.37)
<I/σ(I)>	5.8 (1.5)
Unique reflections	14637
Completeness of data	98.3 (84.1)
Multiplicity	6.2 (5.2)
R_{merge} (all I+ and I−)	19.2 (75.0)
R_{pim} (all I+ and I−)	8.3 (35.7)
$CC_{1/2}$	98.5 (70.3)
R_{work}	0.1825
R_{free} (5% held)	0.2221
Ramachandran favoured (%)	97.4
Ramachandran outliers (%)	0
Unusual rotamers (%)	0.31
RMSD bond angles (°)	0.612
RMSD bond lengths (Å)	0.002
Omit map type, contour level	feature-enhanced map, 1.0 σ

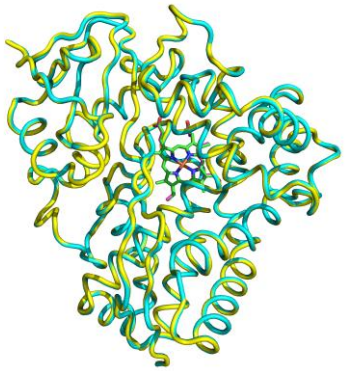
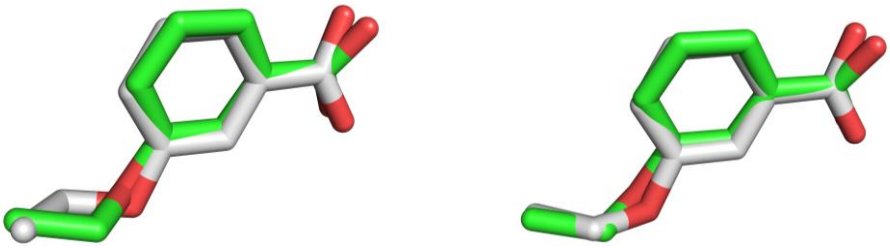


Figure S26 Overlaid structures of 3-ethoxybenzoic acid- (yellow cartoon, magenta heme) and 4-ethoxybenzoic acid-bound (blue cartoon, green heme) CYP199A4. The RMSD between the Cα atoms is 0.182 Å (over all 393 pairs).

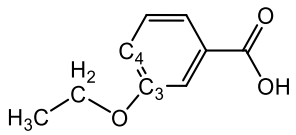


(a) RMSD = 0.536 Å **(b)** RMSD = 0.367 Å

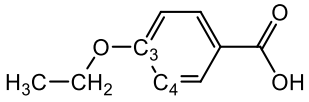
Figure S27 Comparison of the docked ligand poses (grey) with the crystallographic pose of 3-ethoxybenzoic acid (green). In **(a)** the top-scoring docked pose is compared to the crystallographic pose. In **(b)** the second-ranked docked pose is compared to the crystallographic pose. The crystallographic pose is more similar to the second-ranked docked pose shown in **(b)** than the top-ranked docked pose shown in **(a)**.

Table S8 Angles and distances between the heme iron and 3-ethoxybenzoic acid. Distances between the substrate and the hypothetical position of the Cpd I oxygen are also given.

Distance (Å)	**3-OEt**	**4-OEt**
CH_2 - Fe	3.6	4.2
CH_2 - O=Fe^{IV}	2.1	2.8
CH_3 - Fe	4.2	4.3
CH_3 - O=Fe^{IV}	2.7	3.0
O - Fe	4.6	5.1
O - O=Fe^{IV}	3.3	3.6
Angle (°)		
Fe^{IV}=O-CH_2	155.3	144.8
Fe^{IV}=O-CH_3	147.7	137.3
Fe^{IV}=O-O	139.2	163.4
Dihedral angle CH_2-O-C_3-C_4	-52.4	9.2



3-OEt



4-OEt

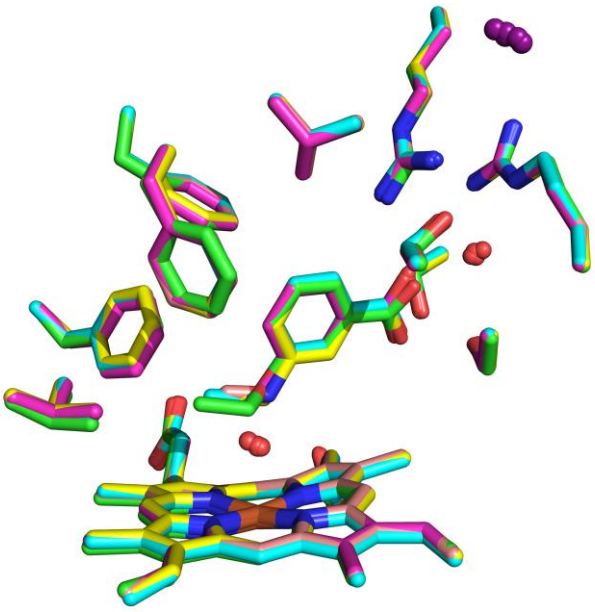


Figure S28 Superimposed crystal structures of CYP199A4 in complex with 3-methoxy- (salmon), 3-methylamino- (yellow), 3-methylthio- (cyan), 3-methyl- (magenta), and 3-ethoxy-benzoic acid (green). In all of these structures, the active-site residues and benzoic acid moiety are held in virtually identical positions.

References

[1] S.G. Bell, A.B. Tan, E.O. Johnson, L.L. Wong, Mol. Biosyst., 6, (2010), 206-214.
[2] J.C. Lessard, in: J. Lorsch (Ed.), Methods Enzymol., vol. 533, Academic Press, 2013, pp. 181-189.
[3] A. de Marco, Microb. Cell Fact., 12, (2013), 101.
[4] A. de Marco, L. Vigh, S. Diamant, P. Goloubinoff, Cell Stress Chaperon., 10, (2005), 329-339.
[5] T.C. Pochapsky, N. Wong, Y. Zhuang, J. Futcher, M.-E. Pandelia, D.R. Teitz, A.M. Colthart, Biochim. Biophys. Acta (BBA) - Proteins and Proteomics, 1866, (2018), 126-133.
[6] A. Luthra, I.G. Denisov, S.G. Sligar, Arch. Biochem. Biophys., 507, (2011), 26-35.
[7] K.P. Conner, C.M. Woods, W.M. Atkins, Arch. Biochem. Biophys., 507, (2011), 56-65.
[8] S.G. Bell, R. Zhou, W. Yang, A.B. Tan, A.S. Gentleman, L.L. Wong, W. Zhou, Chemistry, 18, (2012), 16677-16688.
[9] N.K. Maddigan, S.G. Bell, Arch. Biochem. Biophys., 615, (2017), 15-21.
[10] L.N. Jeffreys, H.M. Girvan, K.J. McLean, A.W. Munro, in: N. Scrutton (Ed.), Methods Enzymol., vol. 608, Academic Press, 2018, pp. 189-261.
[11] J.S. Harbort, J.J. De Voss, J.E. Stok, S.G. Bell, J.R. Harmer, in: G. Hanson, L. Berliner (Eds.), Future Directions in Metalloprotein and Metalloenzyme Research, Springer International Publishing, Cham, 2017, pp. 103-142.
[12] T. Coleman, S.H. Wong, M.N. Podgorski, J.B. Bruning, J.J. De Voss, S.G. Bell, ACS Catal., 8, (2018), 5915-5927.
[13] J.A. Fruetel, R.L. Mackman, J.A. Peterson, P.R. Ortiz de Montellano, J. Biol. Chem., 269, (1994), 28815-28821.
[14] A. Glieder, P. Meinhold, in: F.H. Arnold, G. Georgiou (Eds.), Directed Enzyme Evolution: Screening and Selection Methods, Humana Press, Totowa, NJ, 2003, pp. 157-170.
[15] I. Hanukoglu, Drug Metabol. Rev., 38, (2006), 171-196.
[16] F. Xu, S.G. Bell, Z. Rao, L.L. Wong, Protein Eng. Des. Sel., 20, (2007), 473-480.
[17] M.R. Pincus, J.L. Bock, R. Rossi, D. Cai, in: R.A. McPherson, M.R. Pincus (Eds.), Henry's Clinical Diagnosis and Management by Laboratory Methods, 2017, pp. 428-440.e421.
[18] T. Coleman, Department of Chemistry, vol. PhD, The University of Adelaide, Adelaide, March 2018.
[19] D. Aragão, J. Aishima, H. Cherukuvada, R. Clarken, M. Clift, N.P. Cowieson, D.J. Ericsson, C.L. Gee, S. Macedo, N. Mudie, S. Panjikar, J.R. Price, A. Riboldi-Tunnicliffe, R. Rostan, R. Williamson, T.T. Caradoc-Davies, J. Synchrotron Rad., 25, (2018), 885-891.
[20] N.P. Cowieson, D. Aragao, M. Clift, D.J. Ericsson, C. Gee, S.J. Harrop, N. Mudie, S. Panjikar, J.R. Price, A. Riboldi-Tunnicliffe, R. Williamson, T. Caradoc-Davies, J. Synchrotron Rad., 22, (2015), 187-190.
[21] T.G. Battye, L. Kontogiannis, O. Johnson, H.R. Powell, A.G. Leslie, Acta Crystallogr. D, 67, (2011), 271-281.
[22] P.R. Evans, G.N. Murshudov, Acta Crystallogr. D, 67, (2013), 1204-1214.
[23] M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R.M. Keegan, E.B. Krissinel, A.G. Leslie, A. McCoy, S.J. McNicholas, G.N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin, K.S. Wilson, Acta Crystallogr. D, 67, (2011), 235-242.
[24] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, J. Appl. Crystallogr., 40, (2007), 658-674.
[25] N.W. Moriarty, R.W. Grosse-Kunstleve, P.D. Adams, Acta Crystallogr. D, 65, (2009), 1074-1080.
[26] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Acta Crystallogr. D, 66, (2010), 486-501.

[27] 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, P.D. Adams, Acta Crystallogr. D, 68, (2012), 352-367.
[28] P.D. Adams, P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.-W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, Acta Crystallogr. D, 66, (2010), 213-221.
[29] S.G. Bell, W. Yang, A.B. Tan, R. Zhou, E.O. Johnson, A. Zhang, W. Zhou, Z. Rao, L.L. Wong, Dalton Trans., 41, (2012), 8703-8714.
[30] A. Wlodawer, W. Minor, Z. Dauter, M. Jaskolski, FEBS J., 280, (2013), 5705-5736.
[31] G. Rhodes Journal (2014) Pages.
[32] T.C. Terwilliger, R.W. Grosse-Kunstleve, P.V. Afonine, N.W. Moriarty, P.D. Adams, R.J. Read, P.H. Zwart, L.-W. Hung, Acta Crystallogr. D, 64, (2008), 515-524.
[33] P.V. Afonine, N.W. Moriarty, M. Mustyakimov, O.V. Sobolev, T.C. Terwilliger, D. Turk, A. Urzhumtsev, P.D. Adams, Acta Crystallogr. D, (2015), 646-666.
[34] 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, D.C. Richardson, Acta Crystallogr. D 66, (2010), 12-21.
[35] J.-M. Rondeau, H. Schreuder, in: C.G. Wermuth, D. Aldous, P. Raboisson, D. Rognan (Eds.), The Practice of Medicinal Chemistry (Fourth Edition), Academic Press, San Diego, 2015, pp. 511-537.
[36] K.R. Acharya, M.D. Lloyd, Trends Pharmacol. Sci., 26, (2005), 10-14.
[37] Schrodinger, LLC, 2015.
[38] T. Coleman, R.R. Chao, J.B. Bruning, J. De Voss, S.G. Bell, RSC Adv., 5, (2015), 52007 - 52018.
[39] R.R. Chao, I.C.K. Lau, J.J. De Voss, S.G. Bell, ChemCatChem, 8, (2016), 3626-3635.
[40] T. Coleman, R.R. Chao, J. De Voss, S.G. Bell, Biochim. Biophys. Acta Proteins Proteomics, 1864, (2016), 667-675.
[41] R. Abagyan, A.J.W. Orry, E. Raush, M. Totrov.
[42] H.K. Reimschuessel, B.T. DeBona, Macromolecules, 13, (1980), 1582-1587.

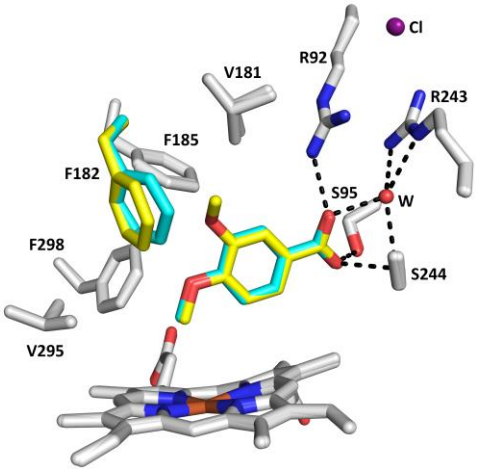


Figure 1. Binding modes of 4-methoxybenzoic acid (cyan; 4DO1) and veratric acid (yellow; 4EGN) in the binding pocket of CYP199A4. F182 rotates to accommodate the *meta*-methoxy group of veratric acid, which points away from the heme. A chloride ion (purple sphere) is also bound at the entrance to the active site. This chloride is thought to stabilize the closed conformation of the enzyme which prevents excess water entering the active site.[22, 25] Hydrophilic interactions are represented by dashed lines and the distances are all between 2.5-3.1 Å.

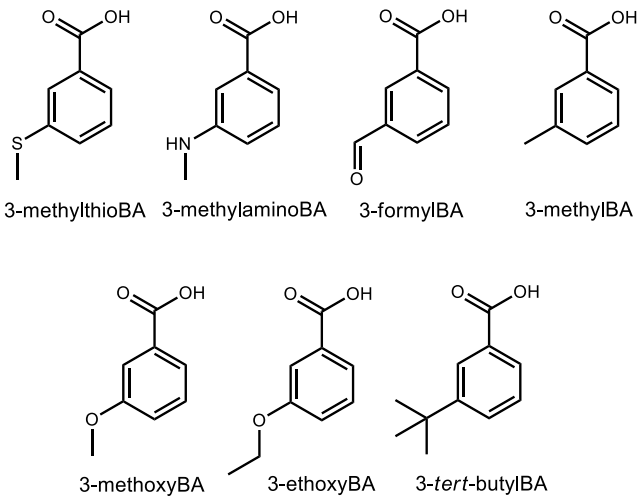


Figure 2. The *meta*-substituted benzoic acid substrates investigated with CYP199A4. Several of the *para*-substituted isomers have been studied previously while others were investigated in this work.[17-19]

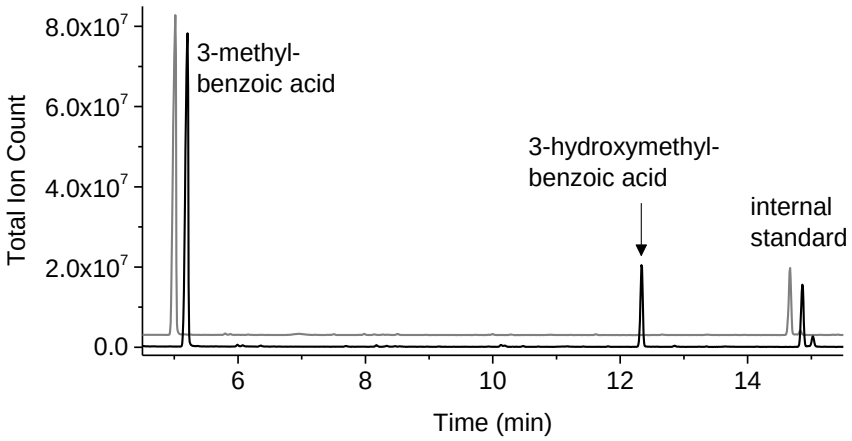


Figure 3. GC-MS analysis of the CYP199A4 *in vitro* reaction with 3-methylbenzoic acid. In black is the *in vitro* turnover, and in gray is a control reaction omitting the P450. Derivatized 3-methylbenzoic acid appears at 5.1 min and the product, 3-hydroxymethylbenzoic acid, appears at 12.3 min. The chromatograms have been offset along both the *x*- and *y*-axes for clarity.

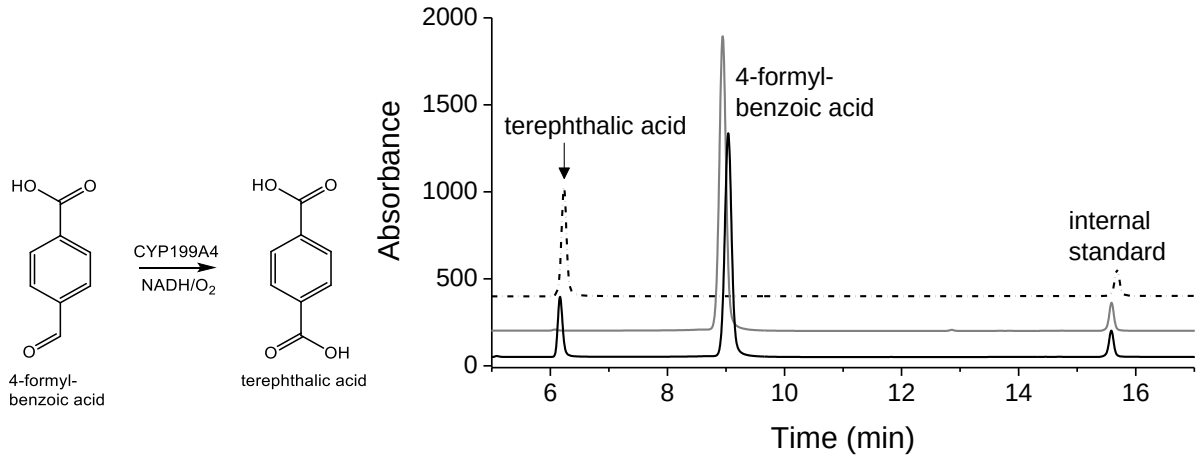


Figure 4. HPLC analysis of the CYP199A4 *in vitro* reaction with 4-formylbenzoic acid. In black is the *in vitro* turnover and in gray is a control reaction performed without P450. Authentic terephthalic acid (dashed chromatogram) co-eluted with the enzyme product at 6.1 min. The substrate appears at 8.9 min. Gradient: 20-95% AcCN in H_2O with 0.1% TFA. Detection wavelength: 254 nm.

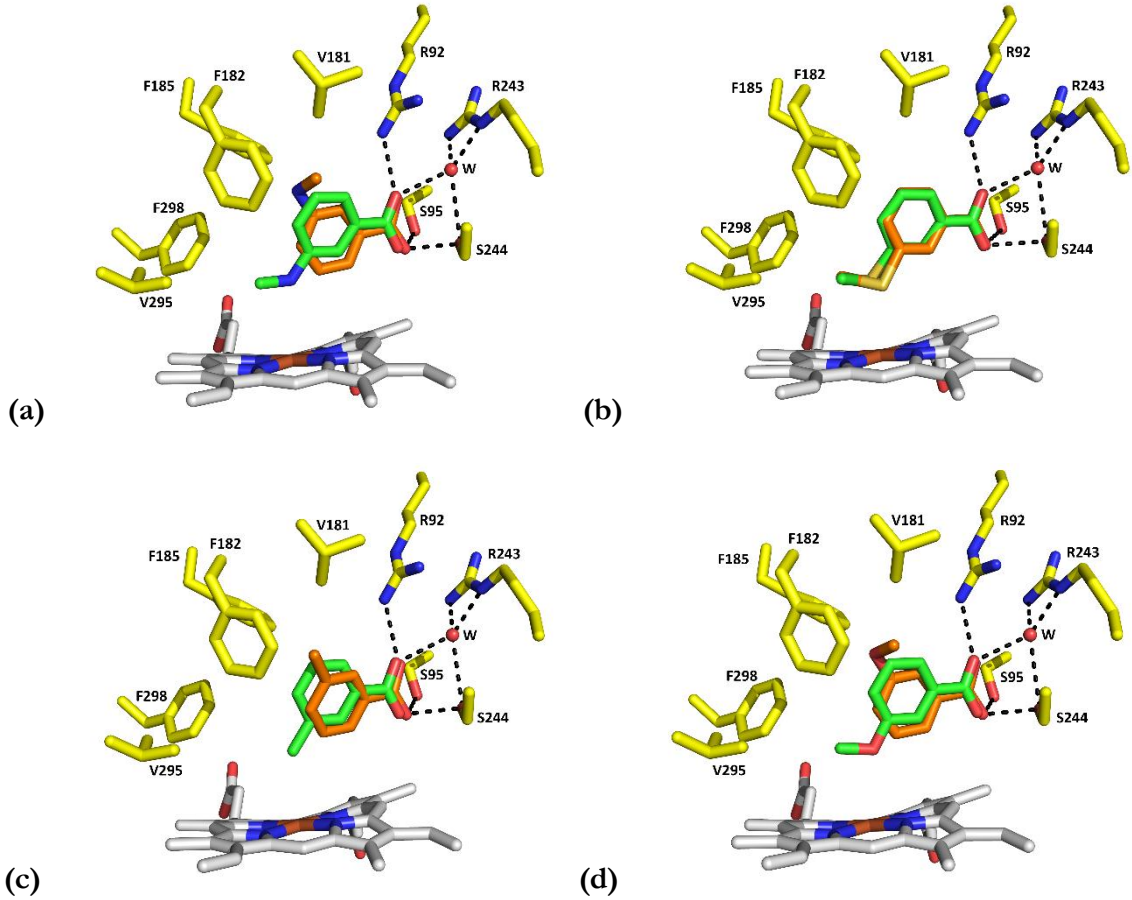


Figure 5. Benzoic acids docked into CYP199A4 (PDB ID: 5U6W). The docked substrates are **(a)** 3-methylaminobenzoic acid, **(b)** 3-methylthiobenzoic acid, **(c)** 3-methylbenzoic acid and **(d)** 3-methoxybenzoic acid. The top-scoring pose is shown in **green**, and lower-scoring poses in **orange**.

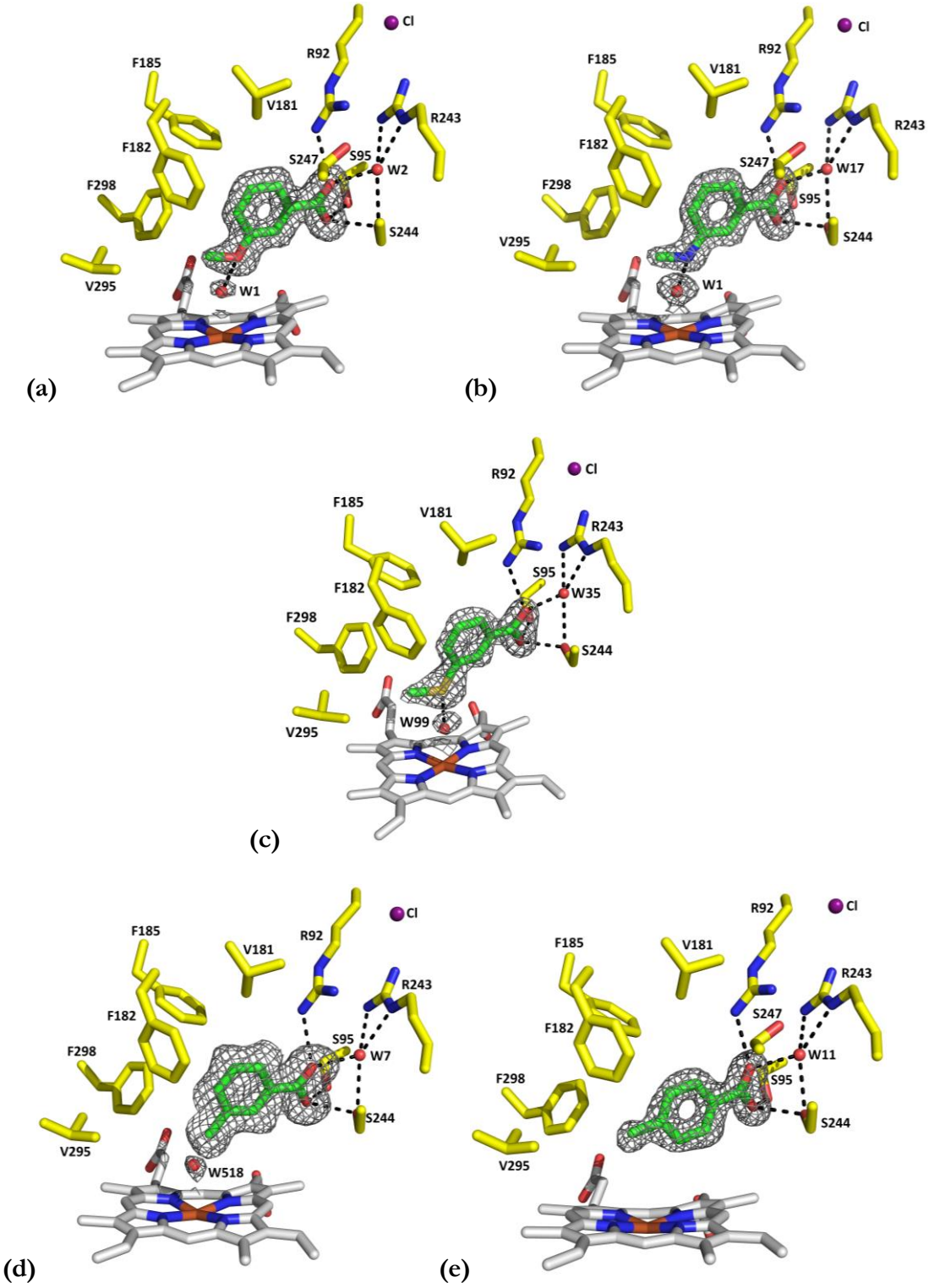


Figure 6. Crystal structures of **(a)** 3-methoxy-, **(b)** 3-methylamino-, **(c)** 3-methylthio-, **(d)** 3-methyl- and **(e)** 4-methyl-benzoate-bound CYP199A4. The 2mFo-DFc electron density of the substrate and iron-bound water ligand is shown as grey mesh. The electron density map is either a composite omit map or a polder omit map. The details of the contour levels are given

in Table S4. Hydrogen bonds are represented by black dashed lines. In grey is the heme, active-site residues are in yellow, the substrate is in green, waters are represented by red spheres, and the chloride ion is in purple.

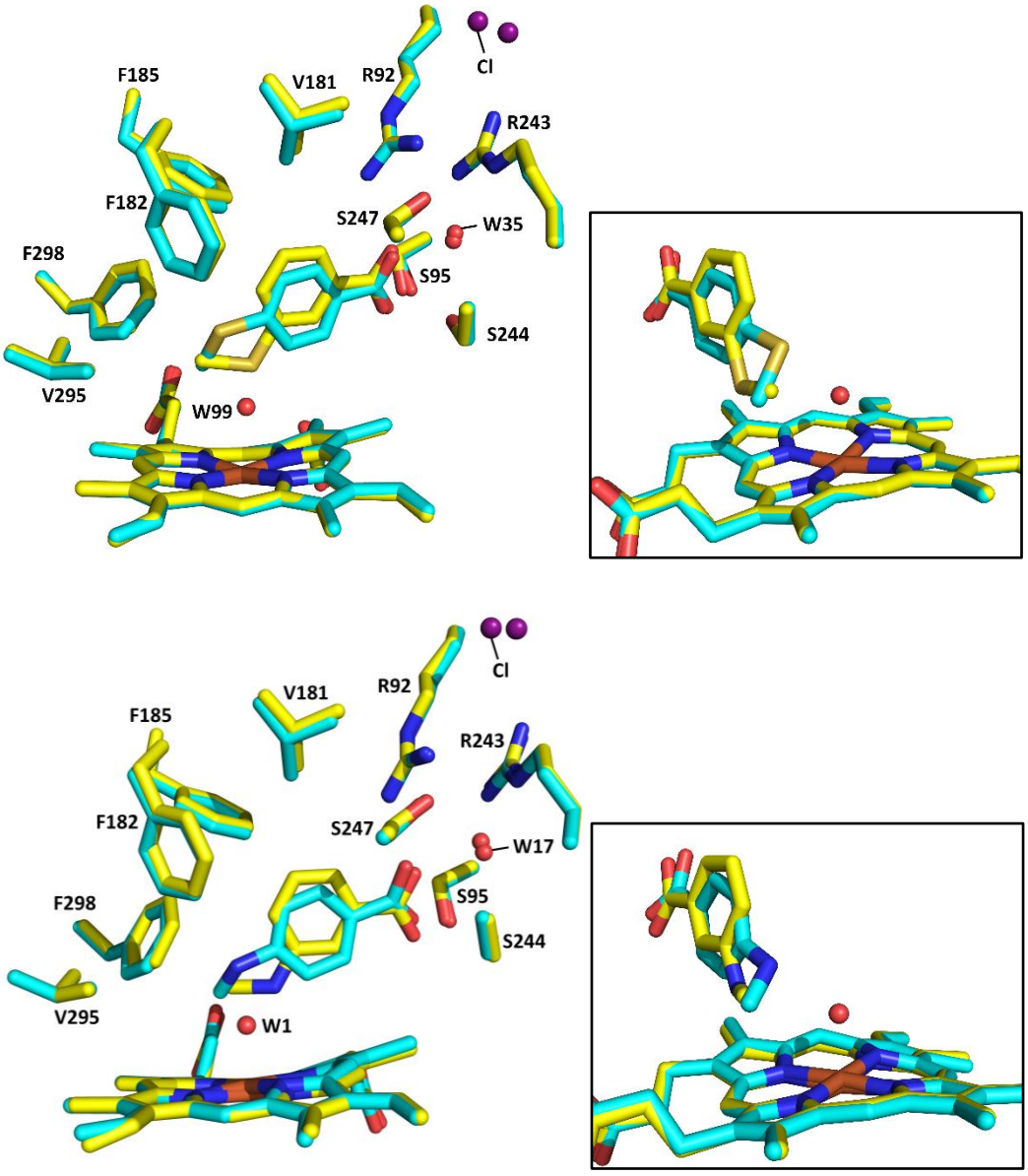


Figure 7. (top) Overlaid crystal structure of 4-methylthio- (cyan) and 3-methylthio-benzoate-bound CYP199A4 (yellow). (bottom) Overlaid crystal structure of 4-methylamino- (cyan) and 3-methylamino-benzoate-bound CYP199A4 (yellow). The waters and chloride in the *meta*-substituted structures are labelled.

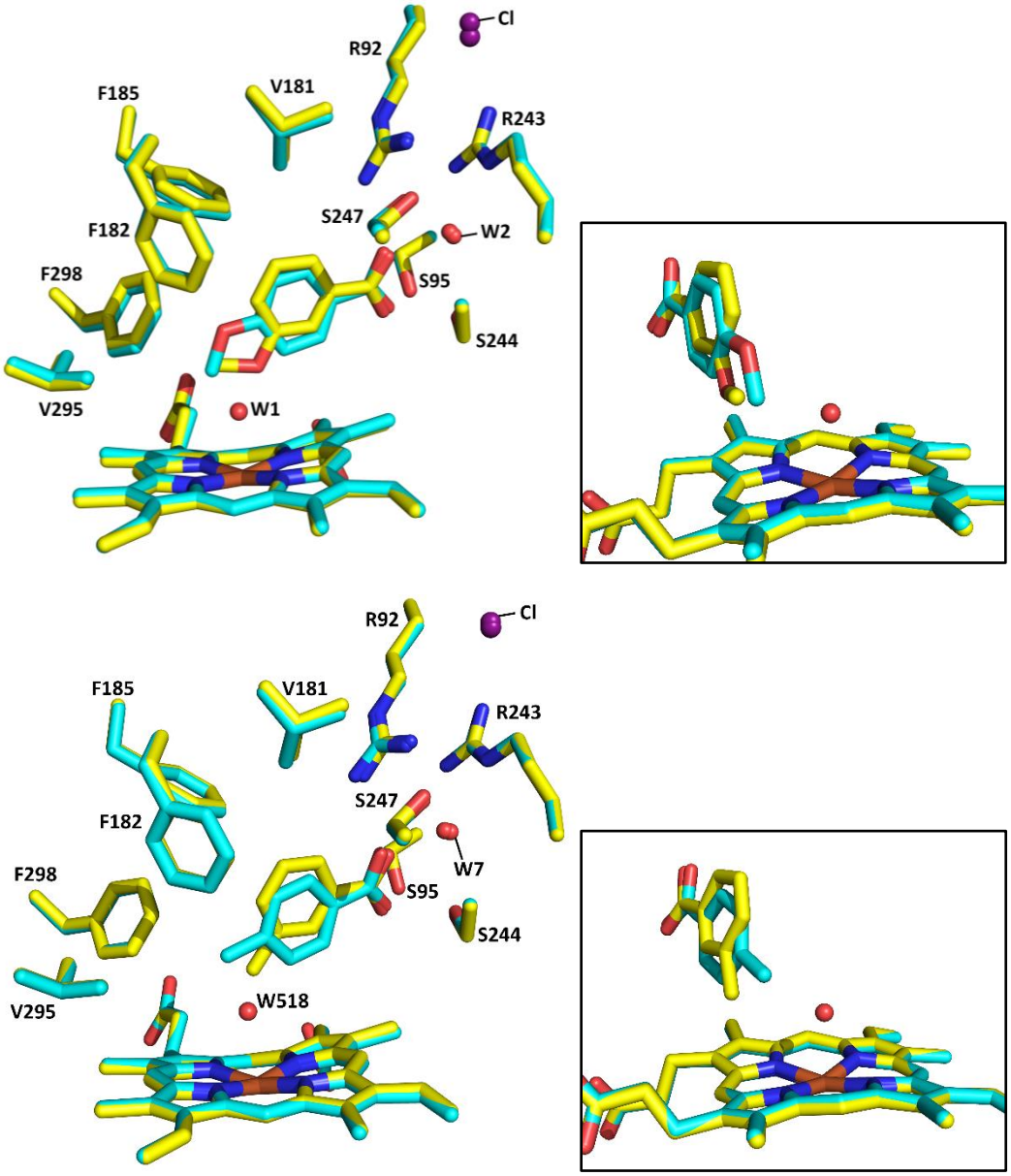


Figure 8. (top) Overlaid crystal structure of 4-methoxy- (cyan) and 3-methoxy-benzoate-bound CYP199A4 (yellow). (bottom) Overlaid crystal structures of 4-methyl- (cyan) and 3-methyl-benzoate-bound CYP199A4 (yellow). The waters and chloride in the *meta*-substituted benzoic acid structure are labelled.

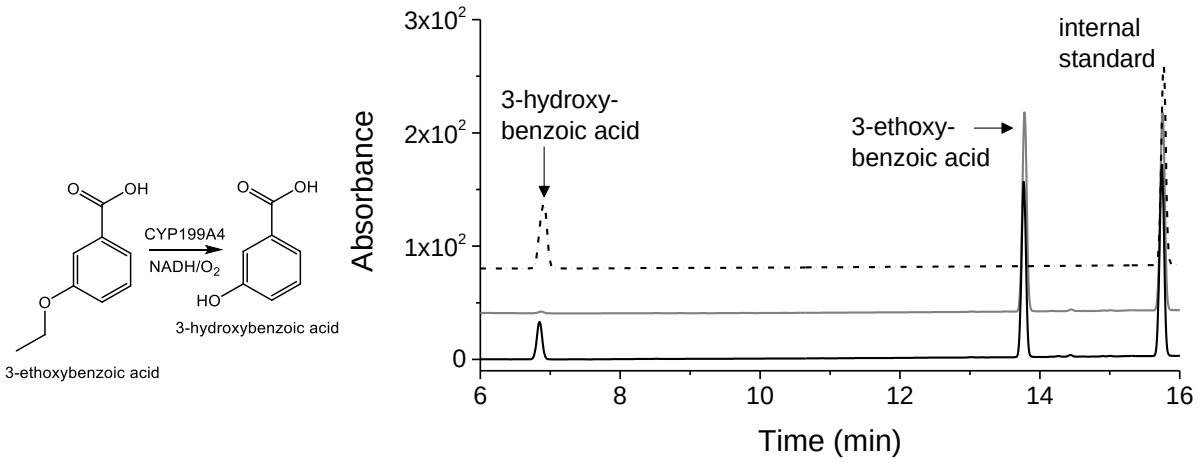


Figure 9. HPLC analysis of the CYP199A4 *in vitro* reaction with 3-ethoxybenzoic acid. In black is the *in vitro* reaction mixture and in gray is a control reaction without P450. Authentic 3-hydroxybenzoic acid (dashed line) co-eluted with the enzyme product at 6.9 min. The substrate, 3-ethoxybenzoic acid, appears at 13.8 min. Gradient: 20-95% AcCN in H_2O with 0.1% TFA. Detection wavelength: 254 nm.

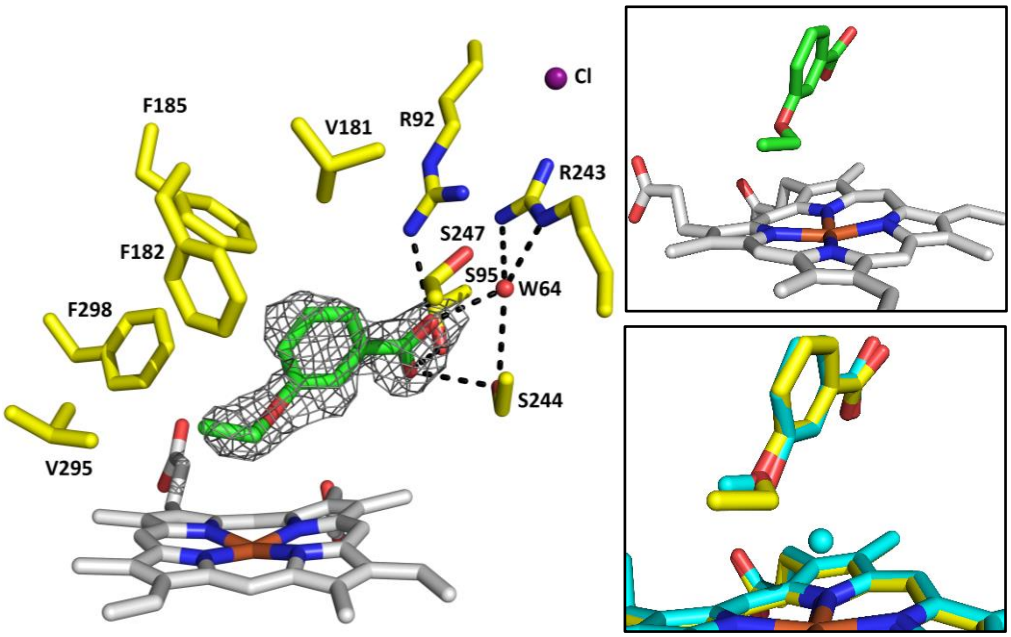


Figure 10. Crystal structure of 3-ethoxybenzoate-bound CYP199A4. A feature-enhanced map of the substrate is shown as grey mesh contoured at 1.0 σ. The insets show the orientation of the 3-ethoxybenzoic acid substrate relative to the heme and the overlaid structures of 3-methoxy- (cyan) and 3-ethoxy-benzoate-bound CYP199A4 (yellow).

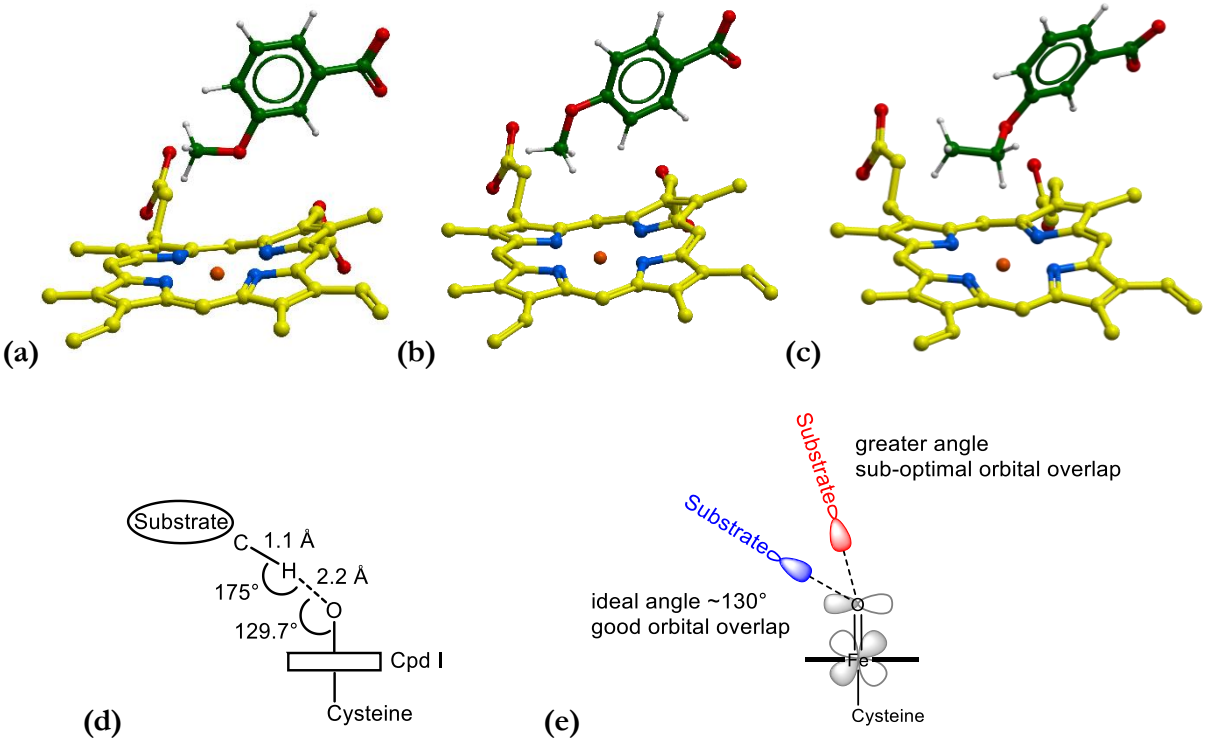


Figure 11. Crystal structures of **(a)** 3-methoxy-, **(b)** 4-methoxy- and **(c)** 3-ethoxy-benzoate-bound CYP199A4 with modelled hydrogens. Hydrogens were added to the crystal structure using ICM-Pro. **(d)** A linear C-H-O angle is calculated to be ideal for hydrogen abstraction[42] and **(e)** the optimal approach angle of the substrate to the Fe^{IV}=O species is postulated to be ~130˚.[41] It has been reported that it is not ideal for the C-H bond of the substrate to approach at angles >130˚ because this results in larger barriers to abstraction.[41]

Table 1. Binding data and catalytic activity of CYP199A4 towards *meta*-substituted benzoic acids. Turnover data for *para*-substituted benzoic acid substrates is included for comparison. Values given are the mean ± SD, with $n \geq 3$. The NADH oxidation rate (N) and product formation rate (PFR) are given in units of μM (μM-P450)$^{-1}$ min^{-1} and were determined in the presence of 1 mM substrate (see Supporting Information for experimental details). C is the coupling efficiency. The NADH leak rate of the system in the absence of a substrate was ~9 min^{-1}.

Substrate	% HS (% H_2O)[a]	K_d (μM)	N (min^{-1})	PFR (min^{-1})	C (%)[d]
3-methoxyBA[17, 19]	40% (50%)	69 ± 2	498 ± 5	–[b]	–[b]
4-methoxyBA[19, 22]	≥95% (0%)	0.22 ± 0.02	1340 ± 28	1220 ± 120	91 ± 2
3-methylaminoBA[17]	10%[e] (90%)	31 ± 1	255 ± 2	175 ± 1	69 ± 1
4-methylaminoBA[17]	70% (0%)	1.6 ± 0.07	923 ± 200	669 ± 15	64 ± 2
3-methylthioBA[17]	30% (77%)	33 ± 0.5	56 ± 0.2	37 ± 1	66 ± 2
4-methylthioBA[17]	70% (0%)	2.3 ± 0.3	1430 ± 180	1180 ± 130	83 ± 3
3-methylBA	30% (21%)	89 ± 3	78 ± 1	37.8 ± 0.6	48 ± 1
4-methylBA[18]	70% (0%)	0.66 ± 0.05	444 ± 8	397 ± 22	89 ± 4
3-formylBA	10%[e] (-)	416 ± 25	10.6 ± 0.2[c]	0.9 ± 0.1	8 ± 1
4-formylBA	25% (-)	48 ± 2	134 ± 1	110 ± 2	83 ± 1

a The % occupancy of the 6th ligand water molecule in the X-ray crystal structure is given in brackets (where known). *b* No product was formed. *c* The NADH oxidation rate reported is the average rate over the first 15 minutes. *d* The coupling efficiency of the reaction is the proportion of the NADH reducing equivalents which are converted into oxidized metabolites. The level of H_2O_2 formed during each turnover was found to be ≤4% in all instances. *e* 3-Methylamino- and 3-formyl-benzoic acid, which induce 10% type I switches to the HS form, both cause the Soret maximum to red shift (0.5 nm, Table S2, Figure S2).

The data for 2-methylbenzoic acid was also measured but no product formation was observed (%HS, ~5%; K_d, 611 μM; and NADH, 13 μM (μM-P450)$^{-1}$ min^{-1}).

Table 2. *In vitro* turnover data for WT CYP199A4 with bulky *meta*-substituted benzoic acid substrates. The details are as defined in Table 1. Turnover data for CYP199A4 with the corresponding *para* isomers, 4-ethoxybenzoic acid which was previously reported,[17] is included for comparison.

Substrate	% HS	K_d (μM)	N (min^{-1})	PFR (min^{-1})	C (%)
3-ethoxyBA	85	82 ± 2	208 ± 5	138 ± 6	66 ± 1[d]
4-ethoxyBA**[17]**	95	0.2 ± 0.1	527 ± 10	527 ± 10	100 ± 8
3-*tert*-butylBA	<5%	_[a]	8.3[b]	_[c]	_[c]
4-*tert*-butylBA	90	39 ± 2	227 ± 4	227 ± 32	100 ± 13

a The lack of any appreciable spin-state shift prevented measurement of binding affinity. *b* The NADH oxidation rate reported is the average rate over the first 15 minutes. *c* No product detected. *d* The highest levels of H_2O_2 production were observed with 3-ethoxybenzoic acid and was 7 ± 0.4% compared to ≤1% for the 4-ethoxy isomer.[17]