Investigation of the Substrate Range of CYP199A4: Modification of the Partition between Hydroxylation and Desaturation Activities by Substrate and Protein Engineering

ARTICLE *in* CHEMISTRY - A EUROPEAN JOURNAL · DECEMBER 2012

Impact Factor: 5.73 · DOI: 10.1002/chem.201202776 · Source: PubMed

7 AUTHORS, INCLUDING:

Stephen Graham Bell
University of Adelaide
70 PUBLICATIONS **1,487** CITATIONS
SEE PROFILE

Wen Yang
University of Washington Seattle
25 PUBLICATIONS **327** CITATIONS
SEE PROFILE

Alexander S. Gentleman
University of Oxford
9 PUBLICATIONS **38** CITATIONS
SEE PROFILE

Luet-Lok Wong
University of Oxford
123 PUBLICATIONS **3,239** CITATIONS
SEE PROFILE

CHEMISTRY
A EUROPEAN JOURNAL

FULL PAPER

DOI: 10.1002/chem.201202776

Investigation of the Substrate Range of CYP199A4: Modification of the Partition between Hydroxylation and Desaturation Activities by Substrate and Protein Engineering

Stephen G. Bell,*[a, b] Ruimin Zhou,[c] Wen Yang,[c] Adrian B. H. Tan,[a] Alexander S. Gentleman,[b] Luet-Lok Wong,*[a] and Weihong Zhou*[c]

Abstract: The cytochrome P450 enzyme CYP199A4, from *Rhodopseudomonas palustris* HaA2, can efficiently demethylate 4-methoxybenzoic acid. It is also capable of oxidising a range of other related substrates. By investigating substrates with different substituents and ring systems we have been able to show that the carboxylate group and the nature of the ring system and the substituent are all important for optimal substrate binding and activity. The structures of the veratric acid, 2-naphthoic acid and indole-6-carboxylic acid substrate-bound CYP199A4 complexes reveal the substrate binding modes and the side-chain conformational changes of the active site residues to accommodate these larger substrates. They also provide a rationale for the selectivity of product oxidation. The oxidation of alkyl substituted benzoic acids by CYP199A4 is more complex, with desaturation reactions competing with hydroxylation activity. The structure of 4-ethylbenzoic acid-bound CYP199A4 revealed that the substrate is held in a similar position to 4-methoxybenzoic acid, and that the C^{β} C–H bonds of the ethyl group are closer to the heme iron than those of the C^{α} (3.5 vs. 4.8 Å). This observation, when coupled to the relative energies of the reaction intermediates, indicates that the positioning of the alkyl group relative to the heme iron may be critical in determining the amount of desaturation that is observed. By mutating a single residue in the active site of CYP199A4 (Phe185) we were able to convert the enzyme into a 4-ethylbenzoic acid desaturase.

Keywords: biocatalysis · cytochromes · desaturation · hydroxylation · X-ray crystal structure

Introduction

Cytochrome P450 (CYP) enzymes are heme-dependent monooxygenases with physiological roles that include hormone and secondary metabolite biosynthesis, as well as xenobiotic metabolism.[1] CYP enzymes primarily catalyze the insertion of an oxygen atom from dioxygen into the carbon–hydrogen bonds of organic molecules. Mechanistically, hydrogen abstraction by Compound I is followed by oxygen rebound to give the hydroxylated compound.[2] Alternatively, a second hydrogen atom abstraction at an adjacent atom can yield the dehydrogenation product.[2b] Other types of activity are also known and include C–C bond cleavage, oxidative dealkylation, C–C bond formation and various ring alterations.[1b,c,3] The stable, soluble nature of bacterial CYP enzymes and the wide range of compounds they oxidise are of interest for applications in synthesis involving C–H bond oxidation.[4] The ability to engineer these bacterial CYP enzymes by rational or random methods can further enhance their activities and broaden their substrate range.[4b,5] A detailed knowledge of the enzyme–substrate interactions is critical in these efforts in order to maximise the catalytic efficiency of the biocatalytic systems and to achieve the turnover numbers required for synthetic applications.

Rhodopseudomonas palustris is a family of purple photosynthetic bacteria that are isolated in temperate, aquatic sediments and possess extraordinary metabolic versatility. Bacteria classified as *Rhodopseudomonas* spp. can grow with or without light or oxygen, fix nitrogen and have a wide range of biodegradation abilities.[6] Studies have shown that different *Rhodopseudomonas* species share many characteristics but different species have unique sets of functional genes to take advantage of the microenvironment in which they are found.[7] We recently reported the properties of the complete CYP199A2 and CYP199A4 class I P450 systems[8] from the *R. palustris* strains CGA009 and HaA2.[9] Both CYP199A2 and CYP199A4 have a high affinity for 4-

[a] Dr. S. G. Bell,[+] A. B. H. Tan, Dr. L.-L. Wong
Department of Chemistry, University of Oxford
South Parks Road, Oxford, OX1 3QR (UK)
E-mail: luet.wong@chem.ox.ac.uk

[b] Dr. S. G. Bell,[+] A. S. Gentleman
School of Chemistry and Physics
University of Adelaide, SA 5005 (Australia)
E-mail: stephen.bell@adelaide.edu.au

[c] R. Zhou,[+] Dr. W. Yang, Dr. W. Zhou
State Key Laboratory of Medicinal Chemical Biology
College of Life Sciences, Nankai University
Tianjin 300071 (P.R. China)
E-mail: zhouwh@nankai.edu.cn

[+] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.201202776.

methoxybenzoic acid and use similar electron transfer chains consisting of a [2Fe-2S] ferredoxin (Pux and HaPux) and a flavin-dependent ferredoxin reductase (PuR and HaPuR) to mediate heme reduction by NADH.[8–10]

In addition to 4-methoxybenzoic acid CYP199A4 and CYP199A2 also bind and oxidise 4-ethylbenzoic acid and veratric acid. Both 4-methoxybenzoic acid and veratric acid are *O*-demethylated to 4-hydroxybenzoic acid and vanillic acid, respectively. This could proceed via oxidation of a methyl C–H bond followed by spontaneous decomposition of the resulting hemiacetal.[11] 4-Ethylbenzoic acid oxidation gives a more complex mixture with two majority and three minority metabolites being formed. The two majority products (>90%) were identified as 4-vinylbenzoic acid, formed by desaturation of the C^{α}–C^{β} bond (36–38%), and 4-(1-hydroxyethyl)benzoic acid from hydroxylation at the C^{α} benzylic carbon (51–59%).[9,10b] The alkene desaturation product could arise from initial C^{α} hydrogen atom abstraction followed by a second hydrogen atom abstraction from C^{β} of the radical intermediate. However, the substrate radical can also undergo one-electron oxidation by the Compound II species to a carbocation that loses a proton to form the alkene.[2b] Alternatively, initial hydrogen abstraction could occur at the terminal carbon (C^{β}) followed by electron transfer oxidation to form a primary carbocation that rearranges to the more stable, benzylic carbocation, which undergoes a C^{β} proton abstraction.[2b,10b,12]

CYP199A2 has also been reported to be a biocatalyst for the oxyfunctionalisation of various aromatic carboxylic acids, including 2-naphthoic acid and indole-6-carboxylic acid. Using a hybrid whole-cell oxidation system with non-physiological electron transfer proteins Furuya and Kino identified the products as 7- and 8-hydroxy-2-naphthoic acid and 2-indolinone-6-carboxylic acid, respectively.[13]

The crystal structure of substrate-free CYP199A2 has been determined at 2.0 Å resolution (PDB code: 2FR7).[10b] The B′ helix found in some P450 structures is missing in CYP199A2 and the G helix is broken into two (forming G and G′ helices). These G and G′ helices are bent back from the extended BC loop and the I helix to open up a clearly defined substrate access channel. The crystal structures of 4-methoxybenzoic acid-bound CYP199A2 and CYP199A4 have also been reported (1.8 and 2.0 Å; PDB codes: 4DNJ and 4DO1, respectively).[14] In the substrate-bound forms of the enzymes there are significant changes in the positions of residue side chains in the substrate access channel. These changes and the binding of an anion (chloride), which caps the entrance to the channel, result in a closed conformation.[14] In both these structures the substrate carboxylate group interacts with polar and basic residue side chains whereas hydrophobic interactions with the benzene ring help hold the substrate in position. In this arrangement the 4-methoxy substituent of the substrate is oriented towards the heme iron (3.9 Å), which is consistent with exclusive attack by CYP199A2 and CYP199A4 at the methoxy methyl group, leading to demethylation to form 4-hydroxybenzoic acid as the sole product.[10b,14]

Here we compare the binding and kinetic properties of CYP199A4 with selected substrates in order to more fully understand fundamental aspects of binding and catalysis and to better determine the biocatalytic scope of this enzyme. By modifying the substrate and the protein we have been able to alter the partition between hydroxylation and desaturation activity with alkyl substituted compounds. In addition, we report the crystal structures of CYP199A4 bound with 4-ethylbenzoic acid, veratric acid, indole-6-carboxylic acid and 2-naphthoic acid. All the structures have a closed conformation with a bound chloride ion in a similar position to that observed in the 4-methoxybenzoic acid-bound structures of CYP199A4 and CYP199A2. The substrates are orientated in the CYP199A4 active site in such a fashion that the reported regioselectivity of oxidation can be rationalised. These structures also provide a basis for understanding how substrates that contain a carboxylic acid functional group are recognised and bound by CYP199A4 and how the enzyme could be engineered for improved substrate binding and modified product profiles.

Results and Discussion

Substrate binding studies on the CYP199A4 system: A range of substrates was tested for binding with CYP199A4 (Figure 1). We had shown previously that both 4-methoxy- and 4-ethylbenzoic acid induced a ≥95% type I spin state shift (Table 1 and Figure S1 in the Supporting Information) whereas veratric acid shifted the spin state to 70% high-spin.[9] The binding of veratric acid was considerably weaker (29.5 μM) compared to 4-ethyl- and 4-methoxybenzoic acid (0.34 and 0.28 μM, respectively). The nature of the *para*-sub-

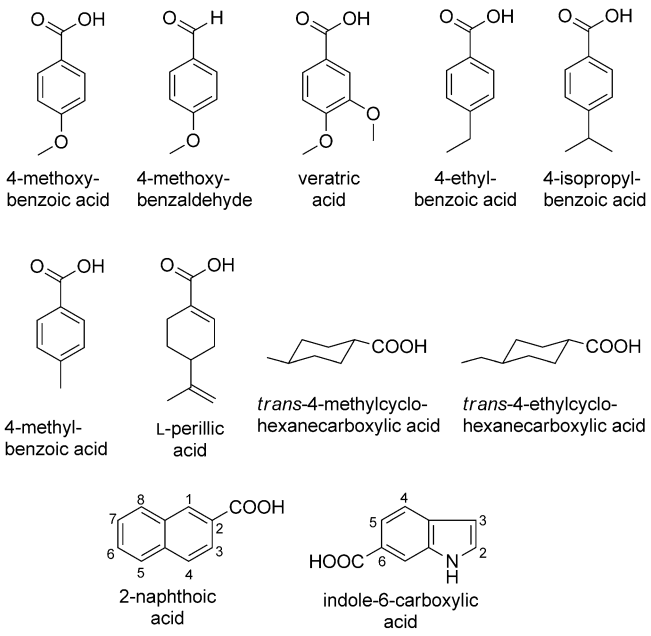


Figure 1. Substrates tested for binding with CYP199A4.

Table 1. Substrate binding parameters and catalytic turnover activity data for CYP199A4 with various substrates.[a]

Substrate	HS [%]	K_d [μM]	N[b]	PFR[c]	Coupling [%][d]
4-methoxybenzoic acid	≥95	0.28±0.01	2180±57	1970±47	91±2
4-ethylbenzoic acid	≥95	0.34±0.02	1180±34	1030±52	88±2
veratric acid	70	29.5±3.1	1100±59	1060±88	96±3
4-isopropylbenzoic acid	≥95	0.29±0.1	321±24	287±18	89±2
4-methylbenzoic acid	70	0.66±0.05	444±8.0	397±22	89±4
4-methoxybenzaldehyde	80	197±7	15±2.0	3±0.1	17±1
trans-4-methylcyclohexanecarboxylic acid	0	–[e]	19±3.0	2±0.1	13±4
trans-4-ethylcyclohexanecarboxylic acid	55	172±16	80±1.0	60±7.0	75±9
trans-4-isopropylcyclohexanecarboxylic acid	50	187±15	86±2.0	n.d.	n.d.
L-perillic acid	90	10.1±0.6	216±13	197±20	91±4
2-naphthoic acid	5	26.0±1.0	70±3.0	–[f]	–[f]
indole-6-carboxylic acid	50	44.4±2.2	696±21	–[f]	–[f]

[a] The data are given as mean ± S.D. with $n \geq 3$. The reaction mixtures (50 mM Tris, pH 7.4) contained P450 (0.5 μM), HaPux (5 μM) and HaPuR (0.5 μM). Rates are given as nmol (nmol CYP)$^{-1}$ min^{-1}. [b] NADH turnover rate. [c] Product formation rate. [d] The percentage of NADH consumed in the reaction that led to the formation of products. [e] Not determined due to absence of spin state shift. [f] Not determined due to further reactions of the products.

To appraise the importance of the planar benzene ring, *trans*-4-ethyl-, *trans*-4-isopropyl- and *trans*-4-methylcyclohexanecarboxylic acid were assayed with CYP199A4 (Table 1). All three compounds induced smaller spin state shifts (55, 50 and 0% respectively) and showed weaker binding (172-(±16) μM for 4-ethyl- and 187-(±15) μM for 4-isopropylcyclohexanecarboxylic acid) compared with their benzoic acid analogues (Table 1 and Figures S1 and S2). The lack of any discernible spin state shift with

stituent of the substrate is important for the binding affinity to CYP199A2, for example, benzoic acid induces a smaller spin state shift of 35%.[10a] In order to further investigate the effect of the size of the *para* substituent on substrate binding to CYP199A4 we studied the binding of 4-methyl- and 4-isopropylbenzoic acids. The spin state shift and binding of 4-isopropylbenzoic acid was comparable to that of 4-ethylbenzoic acid (≥95% and 0.29 μM, Figure S2 in the Supporting Information) whereas 4-methylbenzoic acid binding resulted in a smaller shift to the high-spin state (70%, Figure S1) and the binding was also weaker (0.66 μM, Table 1 and Figure S2). Based on the location of 4-methoxybenzoic acid in the structure of substrate-bound CYP199A4 (PDB code: 4DO1)[14] and of 4-ethylbenzoic acid (PDB code: 4EGM, vide infra) the smaller methyl substituent may not be close enough to the heme to induce complete dissociation of the water ligand from the heme iron centre.

In the 4-methyoxybenzoic acid-bound crystal structure of CYP199A4 (PDB code: 4DO1), the substrate carboxylate group forms hydrogen bonds to the side chains of Arg92, Ser95 and Ser244 (Figure 2). In addition, Arg243, which substantially changes its location between the substrate-bound and substrate-free forms, interacts with the substrate carboxylate group through a bridging water molecule via hydrogen bonds. All of these interactions help to position the *para*-substituent over the heme iron, and account for the substrate specificity and product selectivity. Mutation of these residues dramatically weakened substrate binding and reduced enzyme activity.[10a,14] We have also found that the substrate carboxylate group is important for substrate binding: whereas 4-methoxybenzaldehyde induced a large spin state shift (80%) the binding was three orders of magnitude weaker (197(±7) μM) compared to 4-methoxybenzoic acid (Table 1 and Figure 3). Similar results were obtained with 4-ethyl- and 4-isopropylbenzaldehyde (both induced 90% spin state shift) whereas the corresponding alcohols hardly induced any spin state shift (4-ethylbenzylalcohol, 10% and 4-isopropylbenzylalcohol 0%, data not shown).

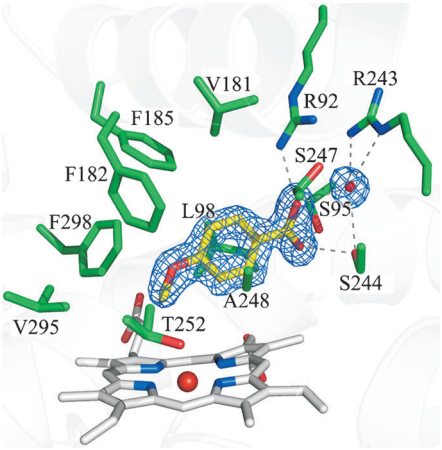


Figure 2. The active site of 4-methoxybenzoic acid-bound CYP199A4. The $2mF_o-DF_c$ density of the 4-methoxybenzoate (yellow) is shown contoured at approximately 0.35 e Å^3 (blue mesh). The heme and the active site residues are shown in grey and green, respectively. Wat170 is shown as a red sphere and hydrogen bonding interactions as dashed lines.

4-methylcyclohexanecarboxylic acid reinforces the tenet that the *para*-substituent requires a minimum of two atoms in order for it to approach close enough to the heme iron to displace the sixth water ligand effectively. Increasing the rigidity of the six membered ring by introducing a double bond was tested by evaluating L-perillic acid binding. This induced a larger spin state shift (90%) than *trans*-4-ethyl- and *trans*-4-isopropylcyclohexanecarboxylic acid and the binding was also significantly tighter (K_d=10.1 μM, Figure S2) though still 30-fold weaker than for 4-ethyl- and 4-isopropylbenzoic acid (Table 1). The data suggest that the planar benzene ring is important for tight binding.[14]

CYP199A2 has been reported to oxidise larger aromatic carboxylic acids including 2-naphthoic acid and indole-6-carboxylic acid.[13] 2-Naphthoic acid elicited a small spin state shift with CYP199A4 (<10%, Figure S1) and the binding (K_d=26(±1) μM) was also weaker compared with the benzo-

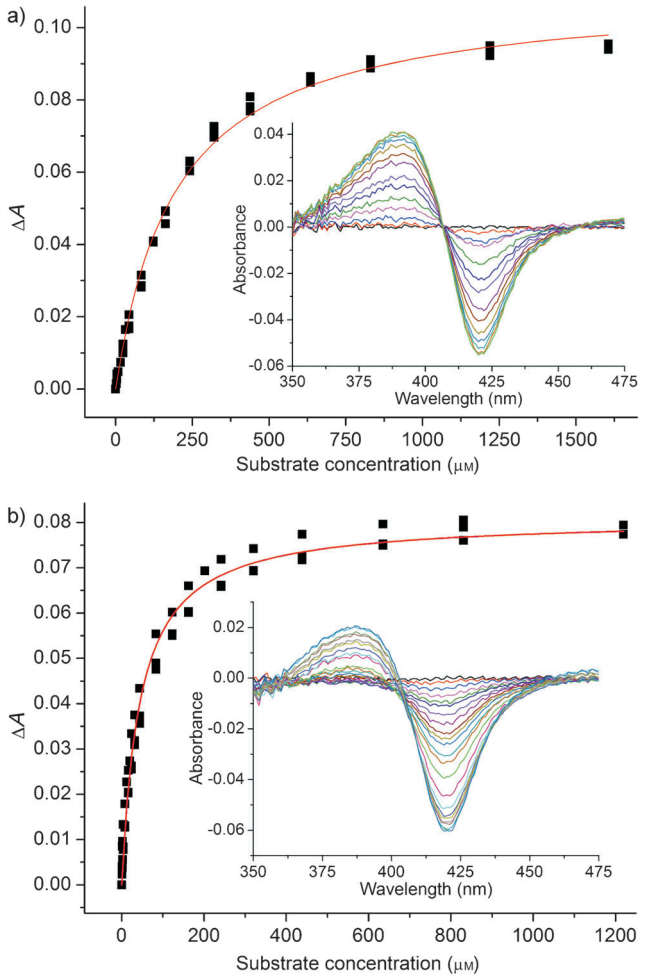


Figure 3. Substrate binding constant analysis of: a) 4-methoxybenzaldehyde (K_d = 197(±7) μM), and b) indole-6-carboxylic acid (K_d = 44.4(±2.2) μM) with CYP199A4.

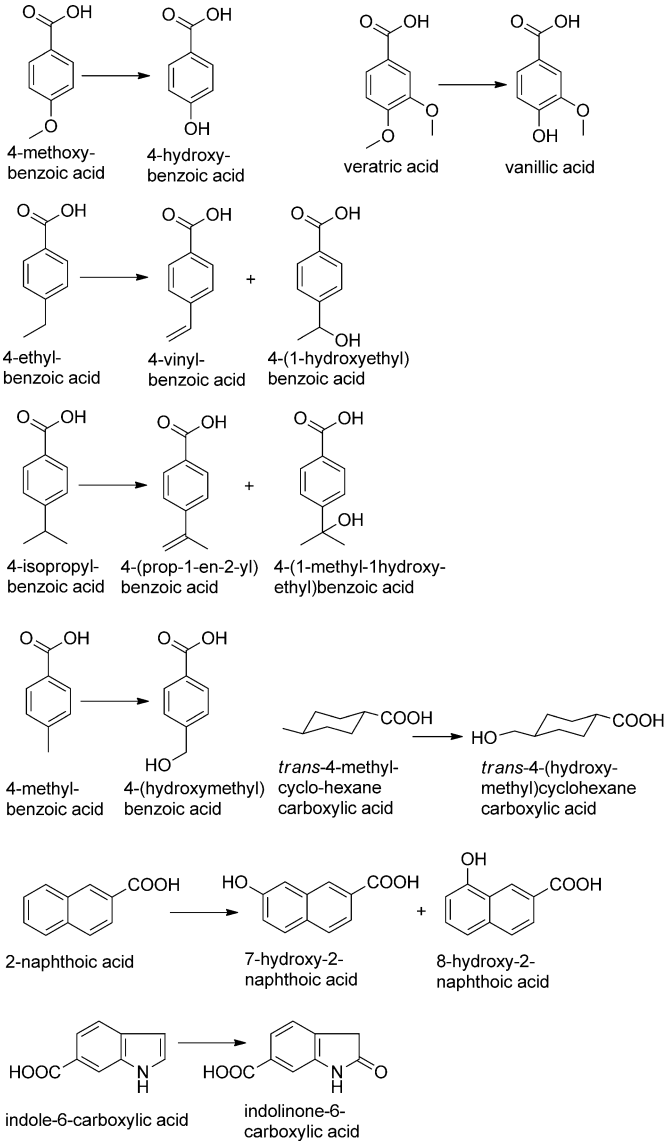


Figure 4. Substrates tested for product formation with CYP199A4. All the products shown have been confirmed by GC co-elution experiments with authentic standards and GC-MS, with the exception of the 4-isopropylbenzoic acid products, which were rationalised on the basis of GC-MS experiments and by analogy with the products of 4-ethylbenzoic acid. The products of 2-naphthoic acid and indole-6-carboxylic acid are from the work of others.[13]

ic acids. Indole-6-carboxylic acid produced a larger spin state shift (20%, Figure S1) but the binding affinity was lower (K_d = 44.4(±2.2) μM, Figure 3).

Activity and product formation assays: The oxidation of 4-methoxybenzoic acid by CYP199A4 results in a single product (4-hydroxybenzoic acid, Figure 4) with high NADH consumption activity and tight coupling of the reducing equivalents to product formation.[9] The NADH consumption rates with veratric and 4-ethylbenzoic acids were lower but the reactions were still tightly coupled (Table 1). Veratric acid also produced a single product (vanillic acid) from selective demethylation of the *para*-methoxy group whereas 4-ethylbenzoic acid yielded two major products, 4-vinylbenzoic acid (38%, TMS-derivatised product: *m*/*z* 221.0984 vs. 223.1064 for the starting material) formed by C^{α}–C^{β} bond desaturation, and 4-(1-hydroxyethyl)benzoic acid (51%, TMS-derivatised product: *m*/*z* 311.1473, Figure 4 and Figure S3 in the Supporting Information).[9,10b]

The turnover activities with 4-methyl- and 4-isopropylbenzoic acid were lower than 4-ethylbenzoic acid, predominantly due to a decrease in the NADH consumption rate with coupling remaining high for both (Table 1 and Figure S4 in the Supporting Information). Oxidation of 4-methylbenzoic acid gave a single product, which was identified as 4-(hydroxymethyl)benzoic acid by GC co-elution with the TMS-derivatised authentic compound and by GC-MS (TMS-derivatised product: *m*/*z* 297.1221, Figure 4 and Figure S5 in the Supporting Information). 4-Isopropylbenzoic acid oxidation resulted in two major products (Figures 4 and 5). These were assigned by GC-MS as the desaturation product 4-(prop-1-en-2-yl)benzoic acid (54%, TMS-derivatised product: *m*/*z* 235.1107 vs. 237.131083 for the starting material)

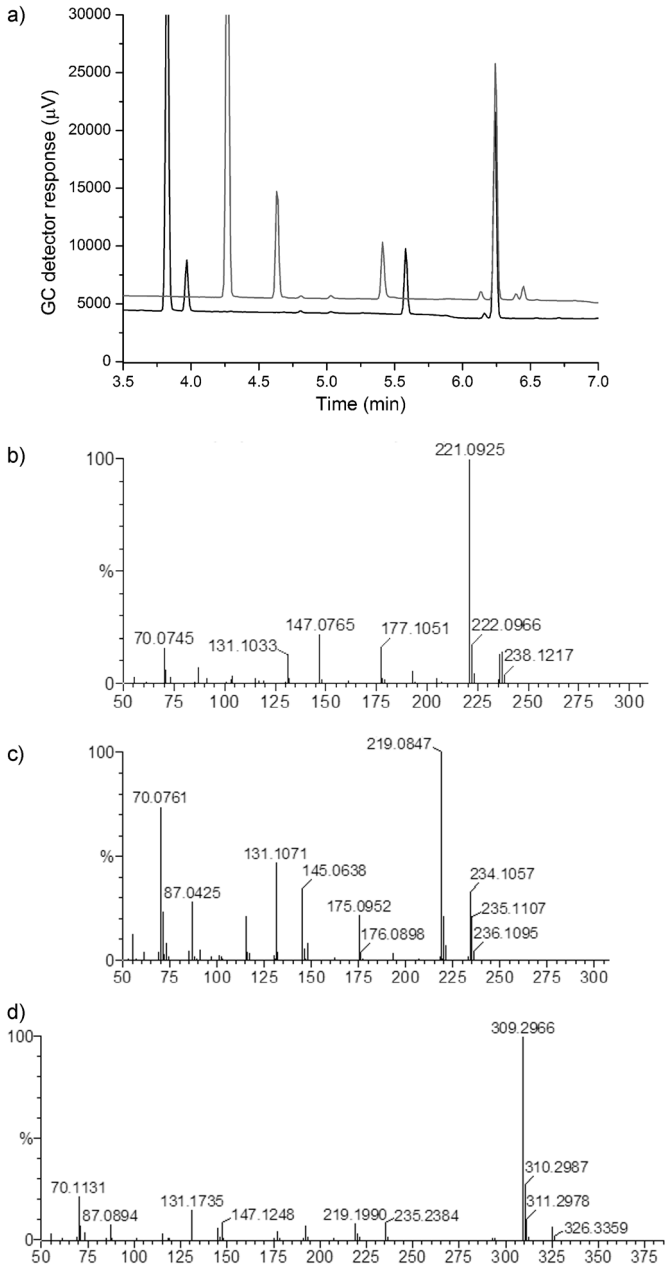


Figure 5. a) Gas chromatography analysis of CYP199A4 turnovers with 4-ethylbenzoic acid (black, $t_R = 3.6$ min) and 4-isopropylbenzoic acid (grey, substrate $t_R = 4.3$ min). The 9-fluorenol internal standard ($t_R = 6.3$ min) is also shown. For clarity the chromatograms have been offset along the *y* axis. MS analysis of: b) the TMS derivative of 4-isopropyl benzoic acid, and the peaks assigned to: c) the TMS derivatives of 4-(prop-1-en-2-yl)benzoic acid, and d) the hydroxylated product assigned as 4-(1-methyl-1-hydroxyethyl)benzoic acid. The GC-MS analysis of the 4-ethylbenzoic acid turnover is provided in the Supporting Information (Figure S2).

and 4-(1-methyl-1-hydroxyethyl)benzoic acid (30%, TMS-derivatised product: m/z ~325.3, calcd: 325.165526 for $[M+H]^+$; Figure 5). The remaining unidentified products (total 16%) had longer GC retention times and may be other hydroxylation products, for example, 4-(1-methyl-2-hydroxyethyl)benzoic acid or further oxidation products.[9,10b] The NADH turnover activities with all the *para*-substituted benzaldehydes and benzyl alcohols tested were very low and little or no products were formed.

The turnover activity with the cyclohexanecarboxylic acids was much lower than the benzoic acids. For example, the product formation rate for *trans*-4-ethylcyclohexanecarboxylic acid was 60(±7) versus 1030(±52) nmol (nmol CYP)$^{-1}$min^{-1} with 4-ethylbenzoic acid (Table 1 and Figure S4). This reduction in activity predominantly arose from a drop in the NADH consumption rate but the coupling of the reducing equivalents to product formation was also lower. The single product arising from *trans*-4-methylcyclohexanecarboxylic acid oxidation was identified as *trans*-4-(hydroxymethyl)cyclohexanecarboxylic acid by co-elution with one of the peaks in a derivatised standard containing *cis*- and *trans*-4-(hydroxymethyl)cyclohexanecarboxylic acid. The oxidation of *trans*-4-ethylcyclohexanecarboxylic acid resulted in two products. The TMS derivatives of both had GC retention times that were far removed from the substrate ($t_R = 5.4$ and 5.6 min vs. 3.2 min, Figure S6a in the Supporting Information). Comparing the retention times of these to those of the derivatised products of 4-ethylbenzoic and *trans*-4-methylcyclohexanecarboxylic acid indicates that both products are likely to be hydroxylation products rather than desaturation products (Figure S6b). The GC-MS data for both products indicated that this was indeed the case (observed masses: 317.1673 and 317.1743, calcd: 317.1968 for $[M+H]^+$, Figure S6c). It is also worth noting that the oxidation of *trans*-4-isopropylcyclohexanecarboxylic acid resulted in a single majority product at an extended retention time with mass of approximately 331.2, indicating that this is also a hydroxylation product (calcd: 331.212476 for $[M+H]^+$, Figure S7 in the Supporting Information). For these compounds hydroxylation could potentially occur at C^{α}, C^{β} or even the C^4 and C^3 carbons of the cyclohexyl ring and further investigation is required in order to fully characterise them.

The addition of a single double bond to the cyclohexane ring in L-perillic acid resulted in an increase in the activity versus *trans*-4-ethyl- and *trans*-4-isopropylcyclohexanecarboxylic acids (Table 1 and Figure S4). Four product peaks were observed in the GC analysis—potentially due to the more complex product profile that may arise from the prop-1-en-2-yl substituent or the double bond on the cyclohexene ring (Figure S8 in the Supporting Information). Due to the multiple products we have as yet been unable to identify the exact nature of each of the metabolites but the retention times and GC-MS data indicate that the products are derived from hydroxylation products (m/z of trimethylsilyl derivatives 327.1778, calcd: 327.181176 for $[M+H]^+$). We were also able to estimate that the coupling of the reducing electrons to product formation was high (91%).

The activity of 2-naphthoic acid oxidation was low (Table 1 and Figure S4) and it yielded two products, both of which fluoresced under UV light and are presumably the 7- and 8-hydroxynaphthoic acid products identified previously (Figure 4 and Figure S9 in the Supporting Information).[13b]

Table 2. Data collection and structure refinement statistics.

Data set	4-Ethylbenzoic acid PDB: 4EGM	Veratric acid PDB: 4EGN	Indole-6-carboxylic acid PDB: 4EGO	2-Naphthoic acid PDB: 4EGP
wavelength [Å]	1.5418	1.0000	0.9796	1.5418
space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$C222_1$
cell dimensions *a*/*b*/*c* [Å]	106.9/143.3/172.5	106.5/143.4/172.8	106.9/143.7/172.9	143.6/172.2/106.3
resolution [Å]	50–2.90 (2.95–2.90)	50–2.00 (2.07–2.00)	50.0–1.76 (1.82–1.76)	50.0–3.00 (3.11–3.00)
average $I/\sigma(I)$[a]	14.9 (3.6)	16.6 (3.1)	15.4 (2.4)	15.5 (2.6)
completeness [%][a]	99.7 (99.6)	98.5 (98.5)	90.0 (66.5)	99.9 (100)
redundancy[a]	4.4 (4.3)	6.0 (5.7)	4.6 (3.7)	5.3 (5.3)
R_{merge} [%][a,b]	14.1 (56.6)	10.0 (59.5)	6.9 (41.7)	13.8/74.9
structure refinement statistics				
average *B*-factor [Å^2]	39.5	24.3	23.2	62.6
R_{work}/R_{free} [%][c]	18.1/24.9	17.0/20.7	18.3/21.3	23.3/28.7
r.m.s.d. bond lengths [Å]	0.017	0.028	0.030	0.014
r.m.s.d. bond angles [°]	1.719	2.094	2.22	1.798
Ramachandran favoured [%]	95.91	98.34	98.4	85.8
Ramachandran outlier [%]	0.32	0.06	0	1.3
molprobity score	2.01	1.26	1.19	2.36
poor rotamers [%]	2.35	1.25	0.47	1.3
all atom clash scores	8.22	3.96	4.1	13.72

[a] Values in parentheses correspond to the highest-resolution shell. [b] $R_{merge}=\Sigma_i|I_i-\langle I\rangle|/\Sigma\langle I\rangle$, where I_i is an individual intensity measurement and $\langle I\rangle$ is the average intensity for all the reflections. [c] $R_{work}/R_{free}=\Sigma||F_o|-|F_c||/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively.

The monooxygenase activity with indole-6-carboxylic acid was higher, with a NADH consumption rate of 696-($\pm$21) nmol (nmol CYP)$^{-1}$ min^{-1} (Figure S4). No product could be detected by GC after derivatisation but the turnover mixtures had a distinct blue tinge. The product 2-indolinone-6-carboxylic acid has been reported previously (Figure 4).[13c] In an attempt to identify the products of indole-6-carboxylic acid and 2-naphthoic acid we used an in vivo HaPuR/HaPux/CYP199A4 expression system.[9] Addition of the indole-6-carboxylic acid or 2-naphthoic acid to this whole-cell oxidation system resulted in a change in the colour of the media to purple and red, respectively. These could be pigments formed from the coupling together of two molecules of monooxygenase product as is observed with indole and other aromatic compounds.[12c,15] Preparative scale reactions will be required for purification of the oxidation products of 4-isopropylbenzoic acid, *trans*-4-ethyl- and *trans*-4-isopropylcyclohexanecarboxylic acids, L-perillic acid, indole-6-carboxylic acid and 2-naphthoic acid for full product characterisation.

Structure of CYP199A4 with bound aromatic carboxylic acid substrates: In previous studies we solved the crystal structures of substrate-free CYP199A2 and the 4-methoxybenzoic acid-bound forms of both CYP199A2 and CYP199A4.[10b,14] Significant conformational changes and residue side-chain movements in the substrate access channel were found in the transition from the open structure of the substrate-free CYP199A2 to a closed conformation for the substrate-bound form, and a chloride ion is bound near the top of the substrate pocket and closes off the active site to the external solvent. The substrate carboxylate group interacts with polar and basic residue side chains whereas hydrophobic interactions hold the benzene ring and position the substrate methyl group closest to the iron, consistent with the exclusive formation of the demethylation product. We have now solved the structures of CYP199A4 complexes with four other substrates, 4-ethylbenzoic acid (PDB code: 4EGM), 3,4-dimethoxybenzoic (veratric) acid (PDB code: 4EGN), 2-naphthoic acid (PDB code: 4EGP) and indole-6-carboxylic acid (PDB code: 4EGO) to provide further insight into substrate binding, specificity and product selectivity of these enzymes. Data collection and refinement statistics of the complexes are given in Table 2. The electron densities of individual substrates in the active sites, and the residues around, are shown in Figure 6.

In the crystal structures of the new CYP199A4–substrate complexes, there are two (2-naphthoic acid) or four molecules (the others) in each asymmetric unit, which can be traced from residues 17 to 409 for each molecule. The refinement of the second molecule of 2-naphthoic acid in the asymmetric unit was poor although it could be traced from residue 17 to the C-terminal residue. This substrate yielded a thin, plate-like crystal form, which may not have been a perfect single crystal. In addition radiation damage from the longer data collection time from the home source X-ray generator potentially led to lower quality diffraction data. Nevertheless, geometry and chemical restraints evaluations indicated that the structural parameters were in acceptable regions (Table S1 in the Supporting Information).

The overall structure of these substrate complexes of CYP199A4 is very similar to that of the 4-methoxybenzoic acid-bound form. All show the closed conformation and a chloride ion is bound close to the surface, which shields the active site from the external solvent (Figure S10 in the Supporting Information). The tertiary structures of the enzyme

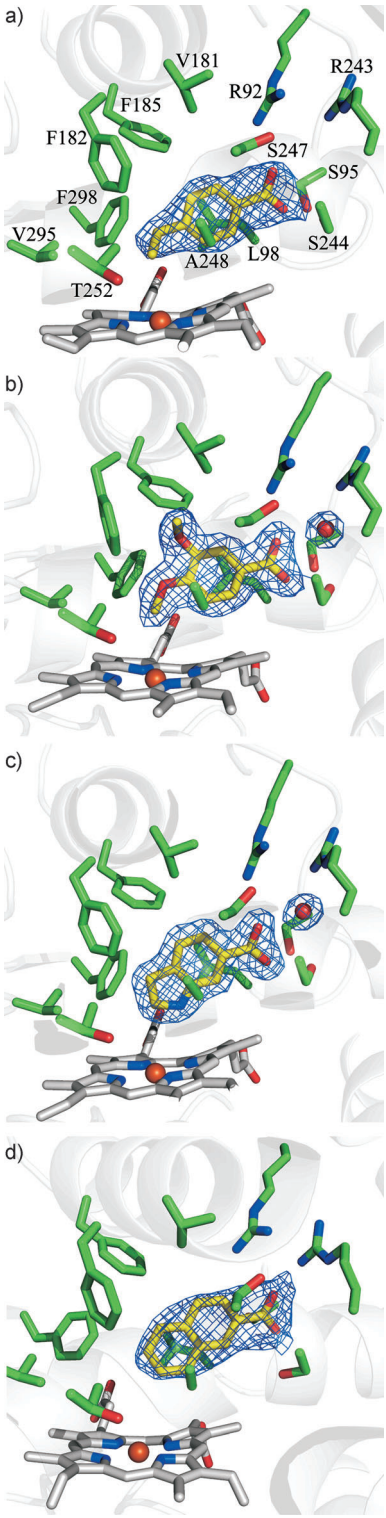


Figure 6. The active site of: a) 4-ethylbenzoic acid-bound, b) veratric acid-bound, c) indole-6-carboxylic acid-bound, and d) 2-naphthoic acid-bound CYP199A4. The $2mF_o-DF_c$ density of the substrates are shown contoured at approximately 0.35 e Å^3 (blue mesh). The substrates, the heme and the active site residues are shown in yellow, grey and green, respectively. The active site water (where present) equivalent to Wat170 in 4-methoxybenzoic acid-bound CYP199A4 is shown as a red sphere. In the 4-ethylbenzoic acid-bound structure, the density is too weak to assign a water molecule in this region. For clarity the residues are only labelled in (a).

in the different complexes are superimposable on one another and the r.m.s.d. for all resolved C^α between all molecules is <0.34 Å. In each instance the crystal structure can be used to rationalise the product selectivity of CYP199 family members with the different substrates. Chain A of each complex was used for further comparison and analysis.

There are two closely related substrate binding modes, one is found with 4-methoxy- and 4-ethylbenzoic acid, veratric acid and indole-6-carboxylic acid for which the benzene rings and the carboxylate groups are at virtually identical locations. The active site water molecule (Wat170 in PDB code: 4DO1), which bridges the substrate carboxylate group and the Arg243 found in the 4-methoxybenzoic acid-bound structure, is retained in the veratric acid and indole-6-carboxylic acid-bound structures. There is not a strong enough region of electron density to assign to a water molecule at this position in the 4-ethylbenzoic acid structure; this is possibly due the lower resolution and higher *B*-factor of the structure (Figure 6a–c). However, the distance between the substrate carboxylate group and Arg243 is similar to that found in the 4-methoxybenzoic acid-bound structure. The second binding mode is observed with 2-naphthoic in which the carboxylate group is displaced and as a result there is no water molecule in the active site pocket (Figure 6d and Figure S11 in the Supporting Information).

4-Ethylbenzoic acid is structurally the most closely related to 4-methoxybenzoic acid; the enzyme–substrate contacts are very similar and for both structures the terminal carbon (C^β) of the *para*-substituent is closest to the heme iron (Figure 6a and Figure S11). There is a small rotation about the C^α–C^{ipso} bond in the ethyl substituted compound, leading to the C^β of the ethyl group being closer to and more directly over the heme iron than the CH_3 of the methoxy group (3.5 vs. 4.1 Å and C–Fe–S(Cys358) angle = 164° vs. 155°, Table S2 in the Supporting Information). Veratric acid is bound with its benzene ring, the carboxylate group and 4-methoxy substituent being superimposable with 4-methoxybenzoic acid (Figure 6b and Figure S11). The 4-methoxy carbon is 3.9 Å from the heme, consistent with vanillic acid being the only oxidation product. The 3-methoxy substituent points away from the heme where it makes contact with the benzene ring of Phe182 and Phe185. The Phe182 side chain is displaced slightly from its position in the 4-methoxybenzoic acid complex to accommodate the additional methoxy group. This movement also affects the position of the side chains of Leu396 and Asp251 and these steric clashes presumably contribute to the weaker binding of this substrate (Figure 6b and Figure S11). Remarkably indole-6-carboxylic acid is bound in such an orientation that its carboxylic acid group and the benzene ring are superimposable on their counterparts in 4-methoxybenzoic acid. The indole NH is in close contact with the *meso* carbon and the adjacent carbons of the C- and D-pyrrole rings of the porphyrin, with strong C–H–π-like interactions (Figure 6c and Figure S11). The indole C^2 is closest to the heme iron (4.2 Å) with the indole nitrogen being the next nearest atom (4.4 Å); this is consistent with attack at this part of the molecule to form 2-indoli-

none-6-carboxylic acid.[13c] The close proximity of the indole nitrogen may contribute to the lower type I spin state shift observed with this substrate.

Although 2-naphthoic acid is similar in size to indole-6-carboxylic acid its binding mode is significantly different. Instead of the benzene ring bearing the carboxylate group it is the C^5–C^8 ring that is superimposable on the benzene ring of 4-methoxybenzoic acid (Figure 6d and Figure S11). The carboxylate group is thus located about 1.9 Å further towards the N terminus of the I helix, and the ordered water molecule (Wat170) found in the 4-methoxybenzoic acid-bound form is displaced. As a result one of the carboxylate oxygen atoms of 2-naphthoic acid forms a hydrogen bond directly with $N^{\eta 1}$ (3.0 Å) of Arg243 rather than through a water molecule (Wat170 in the 4-methoxybenzoic acid-bound complex). However, the hydrogen bonds between the other carboxylate oxygen atom and the side chains of Ser95 and Ser244 are maintained (Figure 6d). In order to maintain the hydrogen bond interaction between Ser95 and the substrate carboxylate the alcohol group of the serine side chain moves by 1.4–2.3 Å. In this binding mode the substrate C^5–C^8 ring is further away from the heme iron; the closest carbon atoms are C^7 (4.7 Å) and C^8 and C^6 (both 5.4 Å), in agreement with the observation of lower spin state shift on substrate binding and the products reported by others (though C^8, 62% was the majority product over C^7, 38%).[13a]

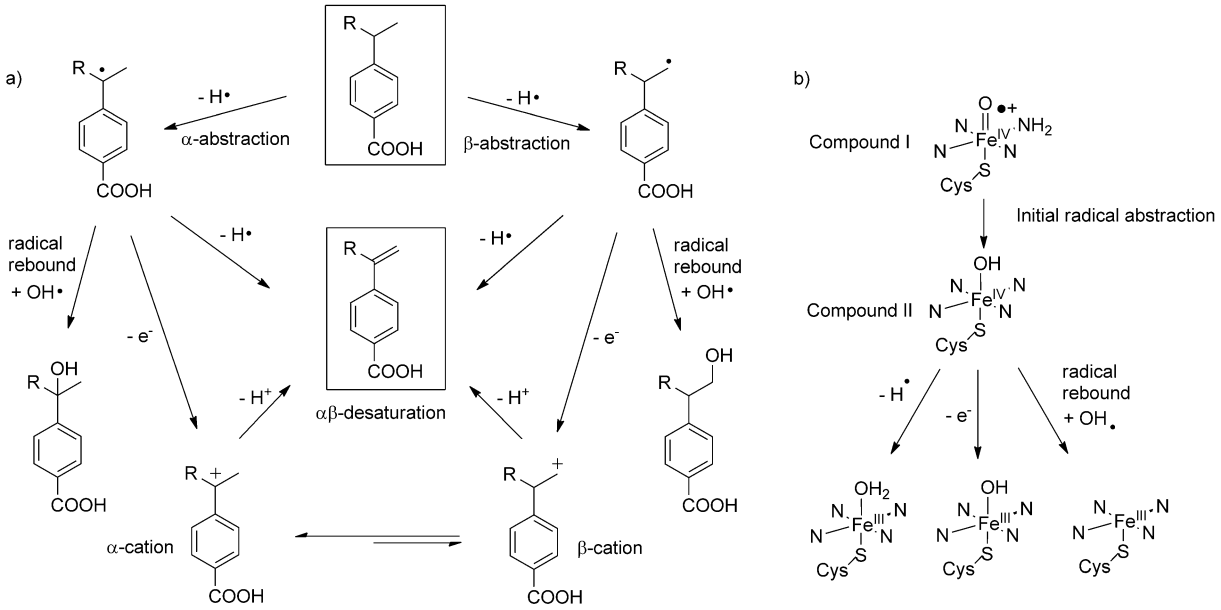
The crystal structures of the CYP199A4–substrate complexes account for the substrate specificity of the enzyme. The requirement of a carboxylate group is explained by the binding pocket for this group comprising the side chains of Arg92, Arg243, Ser95 and Ser244; neutral alcohol and aldehyde groups, for example, 4-methoxybenzaldehyde and 4-ethylbenzyl alcohol, do not form sufficiently strong hydrogen bonds to induce the conformational changes that occur on substrate binding, for example, relocation of Arg243 and cleavage of the Arg92–Glu99 ion pair. The side chains of Leu98 and Ala248 appear to function as a vice clamp to hold the substrate aromatic ring in position via strong van der Waals interactions. The alicyclic rings of cyclohexanecarboxylic acids and L-perillic acid would not fit well between these two key residue side chains, hence the low activity for their oxidation. It can be envisaged that mutations at these residues will relax the specificity for aromatic acids and potentially alter the product selectivity. Moreover, mutation of the cluster of three aromatic residues, Phe182, Phe185 and Phe298, will enable the binding of larger substrates and compounds with a larger substituent on the aromatic rings. During the preparation of this manuscript the importance of the Phe182 residue in substrate binding was demonstrated by mutations at the equivalent residue in CYP199A2 (Phe185). The mutation of this residue to smaller non-aromatic residues modified the substrate binding mode of 2-naphthoic acid and improved the activity with cinnamic acid derivatives.[14,16]

The indole-6-carboxylic acid and 2-naphthoic acid structures all contain a second molecule of the substrate located in a largely hydrophobic groove on the enzyme surface between the F and I helices and the long loop linking the two β4 strands (Figure S12a and b in the Supporting Information). The location of this second molecule of the substrate is equivalent to the groove on the surface where a camphor molecule has been reported to bind to CYP101A1 and CYP101D2.[17] These two substrates adopt slightly different orientations. Taking indole-6-carboxylic acid as an example, it contacts Glu153, Leu175, Pro176, and Gly179 in the F helix, Asp251 in the I helix, and Arg391 in the β4 loop (Figure S12a). The CYP199A4 complex with indole-6-carboxylic acid has a third substrate molecule bound close to the entrance to the substrate access channel, contacting the BC loop residues but also Lys89. This molecule is at the interface between two protein molecules in the asymmetric unit, and contacts with Arg142 and Gln141 of another CYP199A4 molecule (Figure S12c). The locations of the substrate molecules in relation to other molecules in the asymmetric suggest that these binding sites may not be significant in the mechanism and function of the enzyme.

Analysis of the desaturation versus hydroxylation partition: Desaturation between C^α and C^β of the *para*-substituent in substituted benzoic acids could be initiated by hydrogen atom abstraction at either carbon centre (Scheme 1).[10b,12a,b,18] Hydrogen abstraction, from the less activated C^β, followed by abstraction from C^α could occur. Alternatively electron transfer from the β-radical to Compound II yields a β-carbocation with a tendency to rearrange to the more highly substituted benzylic α-cation. Loss of a proton from the C^β of this cation forms the alkene desaturation product. The C–H bond of C^α, which is more activated for hydrogen atom abstraction, could be the initial site of attack followed by a second abstraction of a hydrogen atom from C^β (or a proton if the C^α radical transfers an electron to the Compound II intermediate, Scheme 1).

We observe higher proportions of the desaturation product with CYP199A4 and CYP199A2 with 4-ethylbenzoic acid compared to P450BM3 (CYP102A1) with ethylbenzene where C^α hydroxylation dominates.[12c] DFT calculations were performed to explore why the turnover of 4-ethylbenzoic acid by CYP199A4 is more prone to desaturation than the equivalent reaction of ethylbenzene with CYP102A1 (Table S3 and Figure S13 in the Supporting Information). The relative stabilities of the different intermediates in the desaturation and hydroxylation pathways of these substrates were assessed. We found that the energy differences between the intermediates for 4-ethylbenzoic acid and 4-ethylbenzene oxidation were not large enough to infer that the electronic effects of the carboxylic acid group is a major factor in the different reactivity profiles of these two compounds with CYP199A4 and CYP102A1 (Table S3 and Figure S13).

The additional branching in 4-isopropylbenzoic acid results in an increase in the proportion of the desaturation pathway compared to 4-ethylbenzoic acid. Increased desaturation percentages were also observed for the oxidation of



Scheme 1. a) The desaturation and hydroxylation pathways in the oxidation *para*-substituted benzoic acids. B) The heme species that are generated from Compound I during the various desaturation and hydroxylation pathways.

substituted benzenes by wild-type (WT) CYP102A1, for example, 8.9% for isopropylbenzene versus 0% for ethylbenzene.[12c] It is tempting to conclude from these observations and the absence of desaturation products in the oxidation of *trans*-4-ethylcyclohexanecarboxylic acid by CYP199A4, that the aromatic ring stabilises the intermediates of the desaturation pathway and is required for this activity. However, the first observations of C–C bond desaturation were in the oxidation of the aliphatic compounds valproic and lauric acid.[12b,19] We propose that the major factors for the observation of desaturation is slower radical rebound to form the normal, hydroxylation product and the accessibility of more than one carbon centre to the ferryl, Compound I intermediate. The radical could then undergo rearrangement or electron transfer oxidation to form the carbocation that in turn can rearrange. The inductive and resonance stabilisation effects of other parts of the substrate, saturated versus aromatic rings, tertiary versus secondary carbons, etc., will play a role in these rearrangements and the relative populations of intermediates that collapse to form the products. Radical rebound would be slowed down if the substrate and/or the group bearing the radical centre is mobile and can rapidly adopt a different location away from the heme after initial hydrogen atom abstraction. The partition between desaturation and hydroxylation will depend on the accessibility and reactivity of C–H bonds compared to simple rebound.

In the crystal structure of 4-ethylbenzoic acid-bound CYP199A4 the C^{β} of the ethyl substituent is located over the heme iron (3.5–3.7 Å) with C^{α} being 1.3–1.5 Å further removed. Therefore, with the caveat that the position of the substrate might be altered by oxygen binding and that the C^{α} C–H bond is more reactive than those of C^{β}, we would expect a significant proportion of C–H bond extraction to occur from C^{β}. Radical rebound could be slowed down by rotation about the C^{α}–C^{ipso} bond, a radical rearrangement or oxidation to the carbocation which undergoes rearrangement followed by a second abstraction could ensue and lead to desaturation. Alternatively, once the C^{β} has rotated away from the heme a second abstraction can occur at C^{α}. By comparison, CYP102A1 has a preference for hydrogen abstraction at subterminal locations for fatty acids, alkanes and alkyl benzenes.[5b,20] Therefore, the proximity of the C^{β} of 4-ethylbenzoic acid to the heme iron in the CYP199A4 structure may provide a more complete explanation of the higher levels of desaturation observed in this system. By contrast, the absence of desaturation products in 4-isopropyl- and *trans*-4-ethylcyclohexanecarboxylic acid oxidation may arise from these substrates being bound with their isopropyl and ethyl substituents located differently relative to the heme iron. As noted in the analysis of the binding interactions of the aromatic ring it would be of great interest to examine the effect of mutations at the pair of vice clamp residues Leu98 and Ala248 on substrate and product specificity. Overall the structural data suggest that the positions of C^{α} and C^{β} relative to the heme iron are significant in determining the relative amount of desaturation that is likely to be observed for 4-ethylbenzoic acid and 4-ethylbenzene.

The conversion of CYP199A4 into a 4-ethylbenzoic acid desaturase: We have previously shown that replacement of Phe185 with valine or leucine residues results in a lowering in the spin state shift, the affinity of binding and the activity with 4-methoxybenzoic acid. These mutations of the rigid phenyl side chain of phenylalanine with smaller and more flexible alkyl side chains, such as isoleucine or valine, could modify the position of the benzene ring or the *para*-substitu-

ent in the active site or allow increased substrate mobility both of which could potentially alter the product selectivity. If the location of C^{α} and C^{β} is altered relative to the heme iron the level of desaturation may change. Therefore, we analysed 4-ethylbenzoic acid binding and turnover with these mutants.

As was observed with 4-methoxybenzoic acid the spin shift and substrate binding affinities were lower with 4-ethylbenzoic acid and the Phe185Ile and Phe185Val mutants compared to WT CYP199A4. This is consistent with the substrate being positioned further away from the heme resulting in weaker binding and allowing increased water access to the ferric ion (Table 3). Both mutations also had a

Table 3. Substrate binding and catalytic parameters for CYP199A4 variants with 4-ethyl- and 4-isopropylbenzoic acid.[a]

CYP199A4	HS [%]	K_d [μM]	N[b]	PFR[c]	Coupling [%][d]
4-ethylbenzoic acid					
WT	≥95	0.34±0.02	1180±34	1030±52	88±2
Phe185Ile	70	10.9±0.4	19±2.0	3±0.1	16±2
Phe185Val	80	7.0±0.3	421±28	393±30	93±1
4-isopropylbenzoic acid					
WT	≥95	0.29±0.1	321±24	287±18	89±2
Phe185Ile	90	1.2±0.1	20±1.0	3±0.5	13±2

[a] The reaction mixtures (50 mM Tris-HCl, pH 7.4) contained P450 (0.5 μM), HaPux (5 μM) and HaPuR (0.5 μM). Rates are given as nmol (nmol CYP)$^{-1}$ min^{-1}. The data are given as mean ±S.D. with $n \geq 3$. [b] NADH turnover rate; [c] product formation rate. [d] The percentage of NADH consumed in the reaction that led to the formation of products.

deleterious effect on the activity of CYP199A4. The extra carbon unit in the side chain of isoleucine over valine reduced the NADH consumption activity and the coupling and hence product formation rate to a much greater degree (Table 3).

Product analysis of the Phe185Val turnover demonstrated an increased level of desaturation (68 vs. 54%). More dramatic was the effect of the Phe185Ile mutation, which shifted the product selectivity and resulted in the generation of 100% of the desaturation product, 4-vinylbenzoic acid, with no evidence of hydroxylated product (Figure 7). When the same variant was used to turnover 4-isopropylbenzoic acid, 4-(prop-1-en-2-yl)benzoic acid was the only product observed. The NADH consumption rate and coupling are also reduced for 4-isopropylbenzoic acid (Table 3).

The reduced activity and coupling of these Phe185 mutants likely arose from increased substrate mobility and modification of the substrate binding position relative to the heme iron. These alterations also promoted the desaturation pathway. It is likely that the C–H bond of C^{α} and C^{β} are further away from the heme iron in the mutants so the β-radical is at a far enough distance away from the heme to slow down radical rebound and enable the desaturation pathway to compete more effectively and indeed to dominate in the Phe185Ile mutant.

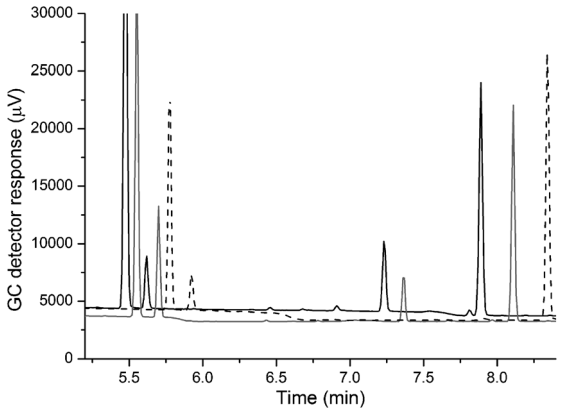


Figure 7. Gas chromatography analysis of the turnover of 4-ethylbenzoic acid (t_R = 5.6 min) with WT (black), Phe185Val (grey) and Phe185Ile (dashed line) CYP199A4. TMS derivatised 4-vinylbenzoic acid is at t_R = 5.7 min with the hydroxylated product at 7.3 min. For clarity the chromatograms have been offset along the *x* and *y* axes.

In conclusion, the crystal structures of substrate-bound forms of CYP199A4 have provided insight into the specificity of the CYP199 family for substituted benzoic acids and rationalisation for observed product distributions. Studies on the occurrence of C–C bond desaturation led to the proposal that this unusual activity of P450 enzymes requires not only for adjacent carbon centres to be close to the heme iron but depends mainly on retardation of the radical rebound step. Although the factors involved are far from fully understood it might be envisaged that the binding of a substrate and the relative rates of the radical rebound and a second hydrogen atom/proton abstraction might be manipulated to increasingly favour C–C desaturation, leading to novel strategies in functionalisation of chemically inert parts of complex molecules.

Experimental Section

General: General reagents, organic substrates and *N*,*O*-bis(trimethylsilyl)trifluoroacetamide with trimethylchlorosilane (BSTFA/TMCS, 99:1) were from Sigma–Aldrich, TCI or Merck (UK). Buffer components were from Anachem (UK) or the Beijing Chemical Company (China). NADH was from Roche Diagnostics (UK), and growth media and isopropyl-β-D-thiogalactopyranoside (IPTG) were from Melford Laboratories (UK) or Invitrogen (USA).

UV/Vis spectra and spectroscopic activity assays were recorded at 30-(±0.5)°C on a Varian CARY-50 or CARY-1E spectrophotometer. Gas chromatography analysis was performed on a ThermoFinnegan TRACE instrument equipped with a flame ionisation detector (FID) and either a DB-1 fused silica column (7 m × 0.32 mm) or a CP-SIL 8CB fused silica column (15 m × 0.32 mm) by using helium as the carrier gas. Both the injector and the FID were held at 250°C. GC-MS was performed on an Agilent-6890 GC coupled to a Micromass GCT (Waters) by using a HP6890 GC 15 m column with helium as the carrier gas and Cl^+ as the ionisation mode for the mass spectrometer.

Enzymes and molecular biology: General DNA and microbiological experiments were carried out by standard methods.[21] The expression and purification of CYP199A4, HaPux, and HaPuR have been described elsewhere.[10a,b] Proteins were stored at −20°C in Tris (pH 7.4, 50 mM) con-

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

The production of CYP199A4 for protein crystallography was carried out by using *Escherichia coli* BL21(DE3) as described elsewhere.[14] Briefly the cells were grown at 37°C in LB medium (1 L) containing kanamycin (50 $\mu g\,mL^{-1}$) to an OD_{600} of 0.6–0.8. Protein production was induced with IPTG (0.5 mM) and the cells were grown for a further 20 h at 25°C. The cell pellet obtained by centrifugation (4000 *g*) was resuspended in buffer A (20 mM HEPES, pH 7.4, 10 mM β-mercaptoethanol) and lysed by sonication at 4°C. The crude extracts were centrifuged at 27000 *g* for 30 min at 4°C and the supernatant was loaded onto a Q Fastflow Sepharose column (GE Healthcare). The target protein was eluted by using a linear salt gradient of KCl (0–1 M) in buffer A, and the pooled fractions were concentrated and then buffer-exchanged into buffer A. Further purification was carried out on a Resource Q column (GE Healthcare) by using a linear gradient of KCl (0–1 M) in buffer A. Gel filtration chromatography on a Superdex-75 column (GE Healthcare) was used for further purification, eluting with HEPES (pH 7.4, 20 mM), KCl (150 mM), β-mercaptoethanol (10 mM). The purity of the protein was estimated to be greater than 95% by SDS-PAGE analysis. The CYP199A4 protein concentration was calculated by using $\varepsilon_{419} = 119\ mM^{-1}\,cm^{-1}$.[14]

Crystallisation: For co-crystallisation of CYP199A4 with 4-ethylbenzoic acid, veratric acid, 2-naphthoic acid and indole-6-carboxylic acid, the purified protein was concentrated to approximately 40 $mg\,mL^{-1}$ in HEPES (pH 7.4, 20 mM), KCl (150 mM), β-mercaptoethanol (10 mM) and a saturated concentration of the substrate. The mixtures were kept on ice before setting up the crystallisation trials. Crystals were obtained by using the hanging-drop vapour-diffusion method at 20°C with protein solution (1 μL) mixed with reservoir solution (1 μL) and equilibrated with reservoir solution (200 μL). Crystal screening was carried out with Hampton Research Crystal Screen I, II and Index kits. Good quality crystals of the complexes were obtained after 2–4 weeks from optimisation around the conditions of Bis-Tris (pH 5.5, 0.1 M), ammonium sulfate (1.45 M), sodium chloride (0.1 M) by using different additives, such as NaF, NaI, $CoCl_2$ and 1,5-diaminopentane·2HCl. The final buffer conditions for each set of crystals are detailed in Table S4 in the Supporting Information.

Data collection and structure determination: X-ray diffraction data for CYP199A4 in complex with veratric acid were collected at −173°C on beamline BL17U1 of the Shanghai Synchrotron Radiation Facility (SSRF). The data for the CYP199A4 indole-6-carboxylic acid complex were collected on beamline BL-17A of the Photon Factory (KEK) in Japan. Diffraction data of CYP199A4 in complex with 4-ethylbenzoic acid and 2-naphthoic acid were collected in-house on a Rigaku R-AXIS HTC image plate by using $Cu_{K\alpha}$ radiation ($\lambda = 1.5418$ Å) from a Rigaku MicroMax-007 rotating anode X-ray generator operating at 40 kV and 30 mA. All crystals were cryoprotected by the addition of glycerol (20% v/v) to the crystallisation solutions. All diffraction data were indexed, integrated and scaled with the HKL2000 package.[22] Complete data collection statistics are summarised in Table 2.

The structures of the complexes were solved by using the molecular replacement method with the program Phaser[23] in the CCP4 suite[24] with the native structure of CYP199A4 (PDB code: 4DNZ) as the initial search model. The substrates were manually built and adjusted under the guidance of $F_o - F_c$ difference maps by using the program COOT.[25] Structural refinements were carried out with Refmac5.[26] The stereochemical quality of the refined structures were checked with the program MolProbity.[27] A detailed summary of the refinement statistics is provided in Table 2. The coordinates for the crystal structures of the complexes have been deposited in the Protein Data Bank with the accession codes 4EGM, 4EGN, 4EGO and 4EGP for the 4-ethylbenzoic acid, veratric acid, indole-6-carboxylic acid and 2-naphthoic acid-bound forms of CYP199A4, respectively.

Activity assays: NADH turnover rate assays were performed with mixtures (1.2 mL) containing Tris (pH 7.4, 50 mM), CYP199A4 (0.5 μM), HaPux (5 μM), HaPuR (0.5 μM) and bovine liver catalase (100 $\mu g\,mL^{-1}$). The buffer solution was oxygenated before use and the mixtures were equilibrated at 30°C for 2 min. Substrates were added as a stock solution (100 mM) in ethanol to a final concentration of 1 mM. NADH was added to about 320 μM (final $A_{340} = 2.00$) and the absorbance at 340 nm was monitored. The rate of NADH consumption was calculated by using $\varepsilon_{340} = 6.22\ mM^{-1}\,cm^{-1}$.

Substrate binding; spin state determination and binding titrations: The high-spin heme content was estimated (to approximately ±5%) by comparison with a set of spectra generated from the sum of the appropriate percentages of the spectra of the substrate-free (>95% low-spin, Soret maximum at 418 nm) and camphor-bound (>95% high-spin, Soret maximum at 392 nm) forms of WT CYP101A1.

For binding constant determination the CYP199A4 mutants were diluted to 0.5–2.0 μM by using Tris (pH 7.4, 50 mM) in a total volume of 2.5 mL, and aliquots (0.5–2 μL) of the substrate were added with a Hamilton syringe from 1, 10 or 100 mM stock solutions in ethanol. The peak-to-trough difference (ΔA) in absorbance between 700 and 250 nm was recorded. Further aliquots of substrate were added until the peak-to-trough difference did not change. The dissociation constants, K_d, were obtained by fitting ΔA against total substrate concentration [S] to a hyperbolic function:

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

where ΔA_{max} is the maximum absorbance difference. Several substrates exhibited tight binding, with $K_d < 1$ μM. In these instances the data were fitted to the tight binding quadratic equation:[28]

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

where ΔA_{max} is the maximum absorbance difference and [E] is the enzyme concentration.

Analysis of metabolites: After the NADH had been consumed in substrate oxidation incubations, 990 μL of the reaction mixture was mixed with 10 μL of an internal standard solution (25 mM 9-hydroxyfluorene in ethanol) and 2 μL of concentrated HCl. The mixture was extracted three times with ethyl acetate (400 μL) and the organic extracts were combined and dried over $MgSO_4$. Solvent was evaporated under a stream of dinitrogen and the sample was dissolved in acetonitrile (200 μL). Excess (25 μL) BSTFA/TMCS (99:1) was added and the mixture was left for at least 120 min to produce the trimethylsilyl ester of the carboxylic acid group and trimethylsilyl ether of the alcohol, if formed. The reaction mixtures were used directly for GC analysis. The oven temperature was held at 100°C for 1 min and then increased at 15°C min^{-1} up to 220°C. The retention times for the trimethylsilyl (TMS) derivatives are given in the Supporting Information. Products were calibrated against derivatised product samples of 4-vinylbenzoic acid, 4-hydroxybenzoic acid, vanillic acid and 4-(hydroxymethyl)benzoic acid.[10b] When authentic samples were not available the coupling was estimated based on the closest available compound from above. The same extractions of reaction mixtures were also used for GC-MS analysis.

Computational studies on desaturation: All geometry optimisations were performed under gas-phase conditions at 298 K by using the B3LYP[29] density functional and the 6-31+G(d,p)[30] basis set on all atoms. Upon attaining all optimised structures, harmonic vibrational frequency calculations at the same level of theory were subsequently performed to ascertain whether they were local minima (0 imaginary frequencies) or local transition state maxima (one imaginary frequency). Wave-function stability tests (Stable = Opt) were also performed in order to ensure that there were no instabilities in the optimised electronic wave-functions. All relative gas-phase energy values discussed and displayed are Gibbs free energy corrected and are in $kJ\,mol^{-1}$ unless otherwise specified. All calculations were carried out by using the Gaussian09 suite of programs.[31]

Acknowledgements

This work was supported by the Ministry of Science and Technology of China Project 973 grant 2010CB530100 (to W.Z.), and the Natural Science Foundation of China grant 31170684 (to W.Z.).

[1] a) F. P. Guengerich, *Chem. Res. Toxicol.* **2008**, *21*, 70–83; b) P. R. Ortiz de Montellano, in *Cytochrome P450: Structure, Mechanism, and Biochemistry,* Kluwer Academic/Plenum, New York, **2005**; c) A. Sigel, H. Sigel, R. Sigel, in *The Ubiquitous Roles of Cytochrome P450 Proteins, Vol. 3*, Wiley-VCH, Weinheim, **2007**.

[2] a) J. Rittle, M. T. Green, *Science* **2010**, *330*, 933–937; b) P. R. Ortiz de Montellano, *Chem. Rev.* **2010**, *110*, 932–948.

[3] a) E. M. Isin, F. P. Guengerich, *Biochim. Biophys. Acta Gen. Subj.* **2007**, *1770*, 314–329; b) M. J. Cryle, J. E. Stok, J. J. De Voss, *Aust. J. Chem.* **2003**, *56*, 749–762.

[4] a) G. Grogan, *Curr. Opin. Chem. Biol.* **2011**, *15*, 241–248; b) D. Monti, G. Ottolina, G. Carrea, S. Riva, *Chem. Rev.* **2011**, *111*, 4111–4140; c) V. B. Urlacher, M. Girhard, *Trends Biotechnol.* **2012**, *30*, 26–36.

[5] a) S. G. Bell, N. Hoskins, C. J. C. Whitehouse, L. L. Wong, *Design and Engineering of Cytochrome P450 Systems*, Wiley, New York, **2007**, p. 652; b) C. J. Whitehouse, S. G. Bell, L. L. Wong, *Chem. Soc. Rev.* **2012**, *41*, 1218–1260.

[6] F. W. Larimer, P. Chain, L. Hauser, J. Lamerdin, S. Malfatti, L. Do, M. L. Land, D. A. Pelletier, J. T. Beatty, A. S. Lang, F. R. Tabita, J. L. Gibson, T. E. Hanson, C. Bobst, J. L. Torres y Torres, C. Peres, F. H. Harrison, J. Gibson, C. S. Harwood, *Nat. Biotechnol.* **2004**, *22*, 55–61.

[7] Y. Oda, F. W. Larimer, P. S. Chain, S. Malfatti, M. V. Shin, L. M. Vergez, L. Hauser, M. L. Land, S. Braatsch, J. T. Beatty, D. A. Pelletier, A. L. Schaefer, C. S. Harwood, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 18543–18548.

[8] F. Hannemann, A. Bichet, K. M. Ewen, R. Bernhardt, *Biochim. Biophys. Acta Gen. Subj.* **2007**, *1770*, 330–344.

[9] S. G. Bell, A. B. Tan, E. O. Johnson, L. L. Wong, *Mol. Biosyst.* **2010**, *6*, 206–214.

[10] a) S. G. Bell, N. Hoskins, F. Xu, D. Caprotti, Z. Rao, L. L. Wong, *Biochem. Biophys. Res. Commun.* **2006**, *342*, 191–196; b) S. G. Bell, F. Xu, I. Forward, M. Bartlam, Z. Rao, L.-L. Wong, *J. Mol. Biol.* **2008**, *383*, 561–574; c) S. G. Bell, F. Xu, E. O. Johnson, I. M. Forward, M. Bartlam, Z. Rao, L. L. Wong, *J. Biol. Inorg. Chem.* **2010**, *15*, 315–328; d) F. Xu, S. G. Bell, Y. Peng, E. O. Johnson, M. Bartlam, Z. Rao, L. L. Wong, *Proteins Struct. Funct. Bioinf.* **2009**, *77*, 867–880.

[11] F. P. Guengerich, *Chem. Res. Toxicol.* **2001**, *14*, 611–650.

[12] a) C. A. Reilly, G. S. Yost, *Drug Metab. Dispos.* **2005**, *33*, 530–536; b) A. E. Rettie, M. Boberg, A. W. Rettenmeier, T. A. Baillie, *J. Biol. Chem.* **1988**, *263*, 13733–13738; c) C. J. Whitehouse, S. G. Bell, L. L. Wong, *Chem. Eur. J.* **2008**, *14*, 10905–10908.

[13] a) T. Furuya, K. Kino, *Biosci. Biotechnol. Biochem.* **2009**, *73*, 2796–2799; b) T. Furuya, K. Kino, *ChemSusChem* **2009**, *2*, 645–649; c) T. Furuya, K. Kino, *Appl. Microbiol. Biotechnol.* **2010**, *85*, 1861–1868.

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

[15] a) S. G. Bell, C. F. Harford-Cross, L. L. Wong, *Protein Eng.* **2001**, *14*, 797–802; b) E. M. Gillam, A. M. Aguinaldo, L. M. Notley, D. Kim, R. G. Mundkowski, A. A. Volkov, F. H. Arnold, P. Soucek, J. J. De Voss, F. P. Guengerich, *Biochem. Biophys. Res. Commun.* **1999**, *265*, 469–472; c) E. M. Gillam, L. M. Notley, H. Cai, J. J. De Voss, F. P. Guengerich, *Biochemistry* **2000**, *39*, 13817–13824; d) O. Shoji, C. Wiese, T. Fujishiro, C. Shirataki, B. Wunsch, Y. Watanabe, *J. Biol. Inorg. Chem.* **2010**, *15*, 1109–1115.

[16] T. Furuya, Y. Arai, K. Kino, *Appl. Environ. Microbiol.* **2012**, *78*, 6087–6094.

[17] a) W. Yang, S. G. Bell, H. Wang, W. Zhou, M. Bartlam, L. L. Wong, Z. Rao, *Biochem. J.* **2011**, *433*, 85–93; b) H. Yao, C. R. McCullough, A. D. Costache, P. K. Pullela, D. S. Sem, *Proteins Struct. Funct. Bioinf.* **2007**, *69*, 125–138.

[18] R. S. Obach, *Drug Metab. Dispos.* **2001**, *29*, 1599–1607.

[19] X. Guan, M. B. Fisher, D. H. Lang, Y. M. Zheng, D. R. Koop, A. E. Rettie, *Chem.-Biol. Interact.* **1998**, *110*, 103–121.

[20] a) M. J. Cryle, R. D. Espinoza, S. J. Smith, N. J. Matovic, J. J. De Voss, *Chem. Commun.* **2006**, 2353–2355; b) M. J. Cryle, J. M. Stuthe, P. R. Ortiz de Montellano, J. J. De Voss, *Chem. Commun.* **2004**, 512–513; c) C. J. Whitehouse, S. G. Bell, H. G. Tufton, R. J. Kenny, L. C. Ogilvie, L. L. Wong, *Chem. Commun.* **2008**, 966–968; d) Y. Miura, A. J. Fulco, *J. Biol. Chem.* **1974**, *249*, 1880–1888; e) Y. Miura, A. J. Fulco, *Biochim. Biophys. Acta Lipids Lipid Metab.* **1975**, *388*, 305–317.

[21] J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, **1989**.

[22] Z. Otwinowski, W. Minor, *Methods Enzymol.* **1997**, *276*, 307–326.

[23] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Crystallogr.* **2007**, *40*, 658–674.

[24] C. C. Project, *Acta Crystallogr. Sect. D* **1994**, *50*, 760–763.

[25] P. Emsley, K. Cowtan, *Acta Crystallogr. Sect. D* **2004**, *60*, 2126–2132.

[26] G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta Crystallogr. Sect. D* **1997**, *53*, 240–255.

[27] 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. Sect. D* **2010**, *66*, 12–21.

[28] J. W. Williams, J. F. Morrison, *Methods Enzymol.* **1979**, *63*, 437–467.

[29] a) A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098–3100; b) A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652; c) C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B Condens. Matter.* **1988**, *37*, 785–789.

[30] a) T. Clark, J. Chandrasekhar, G. W. Spitznagel, P. V. Schleyer, *J. Comput. Chem.* **1983**, *4*, 294–301; b) P. C. Hariharan, J. A. Pople, *Theor. Chim. Acta* **1973**, *28*, 213–222; c) W. J. Hehre, R. Ditchfie, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257; d) G. W. Spitznagel, T. Clark, J. Chandrasekhar, P. V. Schleyer, *J. Comput. Chem.* **1982**, *3*, 363–371.

[31] Gaussian 09, Revision A.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian, Inc. Wallingford CT, **2009**.

Received: August 2, 2012
Published online: November 7, 2012

