Supplementary Information

The crystal structures of 4-methoxybenzoate bound CYP199A2 and CYP199A4: closing of the substrate access channel and the identification of an anion binding site

Stephen G. Bell[*‡a], *Wen Yang*[‡2], *Adrian B. H. Tan* [a], *Ruimin Zhou*[b], *Eachan O.D. Johnson*[a], *Aili Zhang*[b], *Weihong Zhou*[*b], *Zihe Rao*[b,c] *and Luet-Lok Wong*[*a]

[a] Department of Chemistry, University of Oxford, Inorganic Chemistry Laboratory, South Parks Road, Oxford, OX1 3QR, UK

[b] State Key Laboratory of Medicinal Chemical Biology, College of Life Sciences, Nankai University, Tianjin 300071, China

[c] Laboratory of Structural Biology, Tsinghua University, Beijing, 100084, China.

Corresponding authors

[*] S.G. Bell; present address; The School of Chemistry and Physics, The University of Adelaide, North Terrace, Adelaide, SA 5005, Australia; (stephen.bell@adelaide.edu.au)

L.-L. Wong; (luet.wong@chem.ox.ac.uk)

W. Zhou; (zhouwh@nankai.edu.cn)

Table S1. Residues that make up the active site and substrate access channel of CYP199A2 and CYP199A4. All residues are conserved between these two enzymes. The residues highlighted in red are conserved in greater that 75% of the proteins identified as CYP199 family members by having greater than 40% sequence homology with CYP199A2. The residues highlighted in bold are conserved in all of these enzymes (for more detailed analysis see ESI, Fig. S8).

R. palustris CGA009 CYP199A2	*R. palustris* HaA2 CYP199A4
Val82	Val80
Glu90	Glu88
Lys91	Lys89
Pro92	Pro90
Trp93	**Trp91**
Arg94	**Arg92**
Pro95	Pro93
Pro96	Pro94
Ser97	**Ser95**
Leu98	Leu96
Ile99	Ile97
Leu100	Leu98
Glu101	**Glu99**
Tyr180	Tyr177
Leu183	Leu180
Val184	Val181
Phe185	**Phe182**
Ala187	Ala184
Phe188	Phe185
Ile199	Ile196
Arg201	Arg198
Ser202	Ser199
Ala203	Ala200
Gln206	Gln203
Leu243	Leu240
Arg246	**Arg243**
Ser247	Ser244
Ser250	Ser247
Ala251	Ala248
Asp254	**Asp251**
Thr255	**Thr252**
Val298	Val295
Thr300	Thr297
Phe301	Phe298
Cys361	**Cys358**
Thr398	Thr395
Leu399	Leu396

Fig. S1. The open structure of SF CYP199A2 in (a) is compared with the closed structure of the 4-methoxybenzoate bound (SB) form of CYP199A4 in (b). Negatively and positively charged surface areas are coloured in red and blue respectively. A close up of the surface at the entrance of the substrate access channel of SF CYP199A2 (c) which is closed in the SB form of CYP199A4 (d). The heme is shown in yellow and the F, G and I helices are in blue, orange and magenta, respectively.

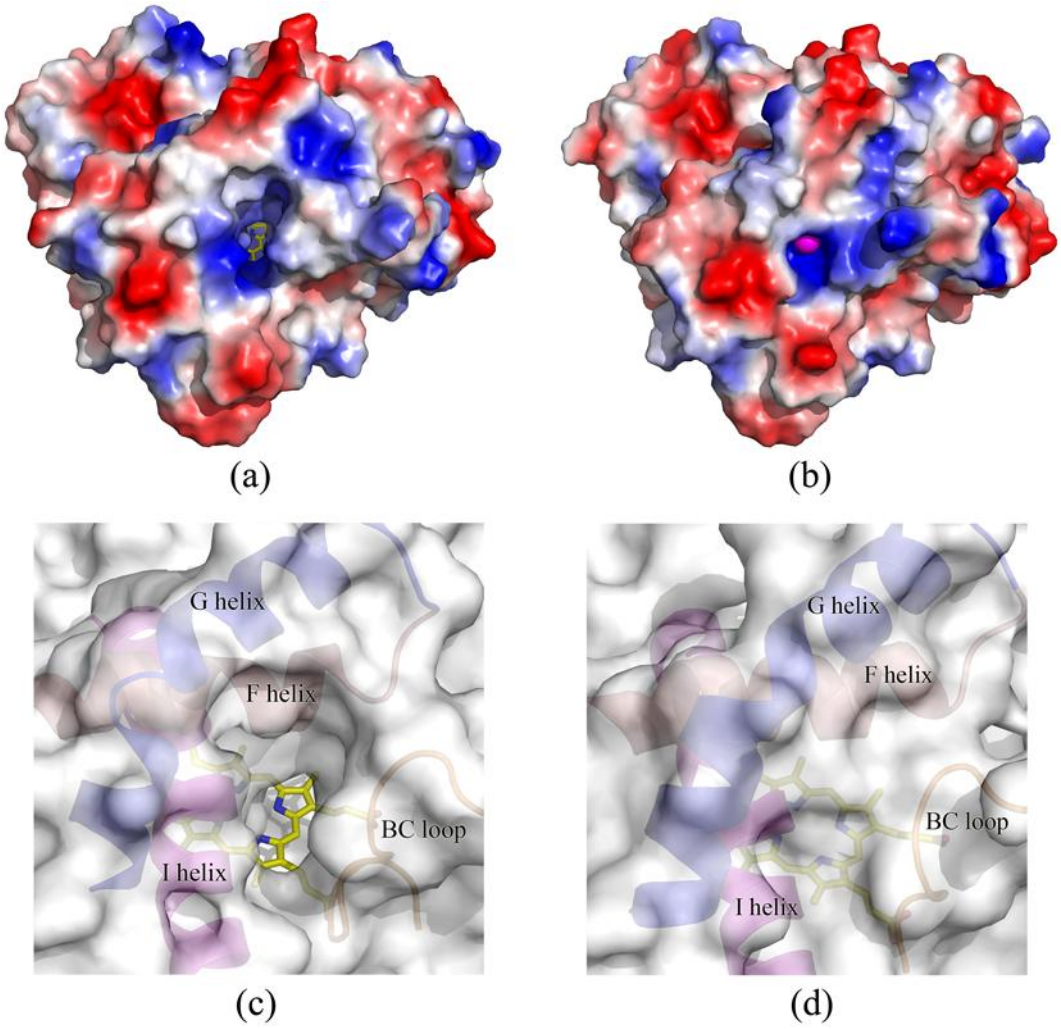


Fig. S2. The active site of 4-methoxybenzoate bound CYP199A4. The $2mF_o - DF_c$ density of the 4-methoxybenzoate (pale green) is shown contoured at ~0.35 e/Å^3 (blue mesh). The heme and the active site residues are shown in yellow and pink respectively. Wat170 is shown as a red sphere and hydrogen bonding interactions as dashed lines.

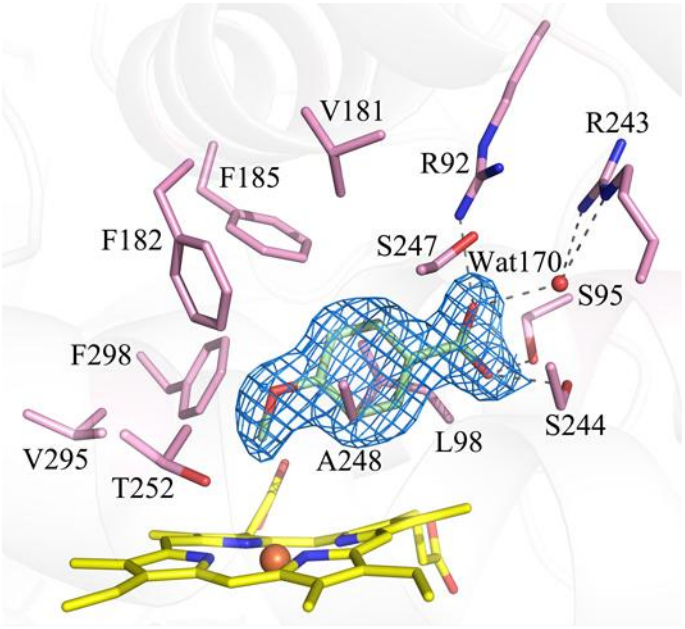


Fig. S3. The active site of native CYP199A4. The unresolved $2mF_o - DF_c$ density is contoured at ~0.35 e/Å^3 (blue mesh). The heme and the active site residues are shown in yellow and pink respectively.

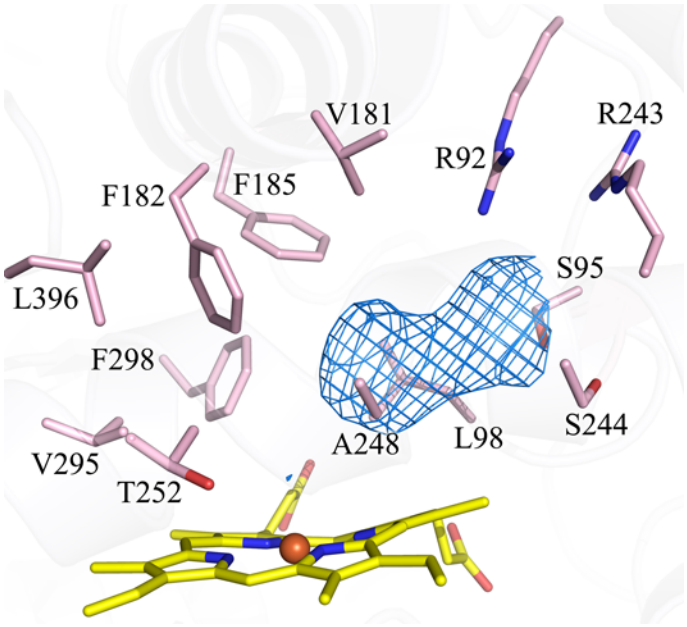


Fig. S4. Comparison of the access channel and active site of SF CYP199A2 (grey) and SB CYP199A4 (pink). The heme and 4-methoxybenzoate substrate are shown in yellow and pale green respectively. Wat170 in SB CYP199A4 is shown as a red sphere and hydrogen bonding interactions as dashed lines. The residue numbers of CYP199A2 are given in parentheses. Significant residue movements are highlighted using a green arrow.

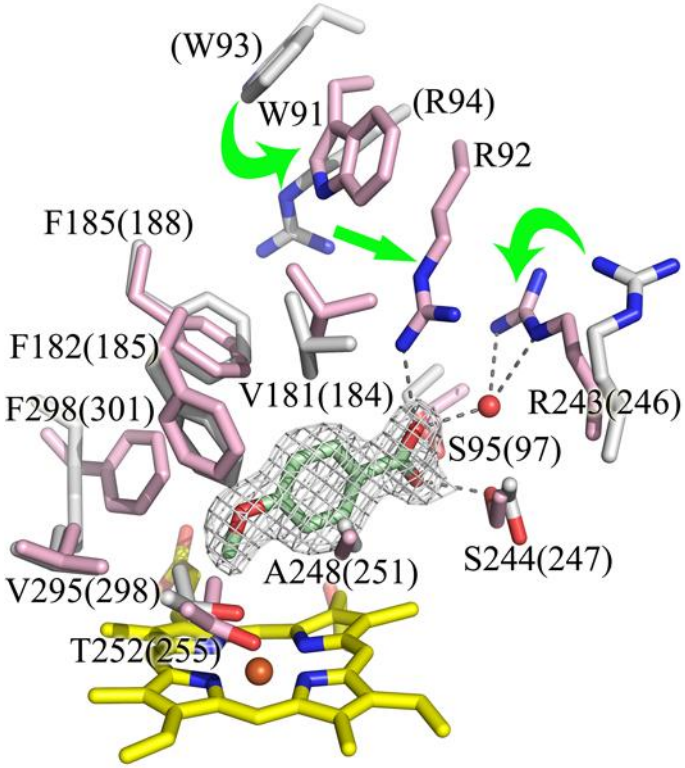


Fig. S5. The chloride binding site of CYP199A4. The $2mF_o - DF_c$ density of the chloride ion is shown contoured at ~0.35 e/Å^3 (blue mesh) The four-fold coordination environment of the bound chloride is highly irregular, comprising the phenolic OH of Tyr177 in the F helix, the side chain NH_2 of Gln203 in the G helix and two water molecules. The Nη1 of Arg92 in the BC loop, and Nη2 of Arg243 in the I helix are further distant. The chloride binding site residues are shown in green and and 4-methoxybenzoate is in yellow. Water molecules and the chloride ion are shown as red and magenta spheres and hydrogen bond and chloride-residue interactions as dashed lines.

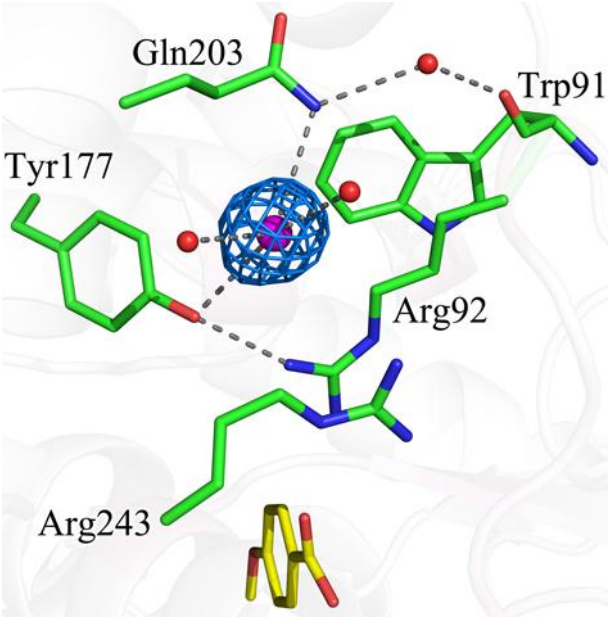
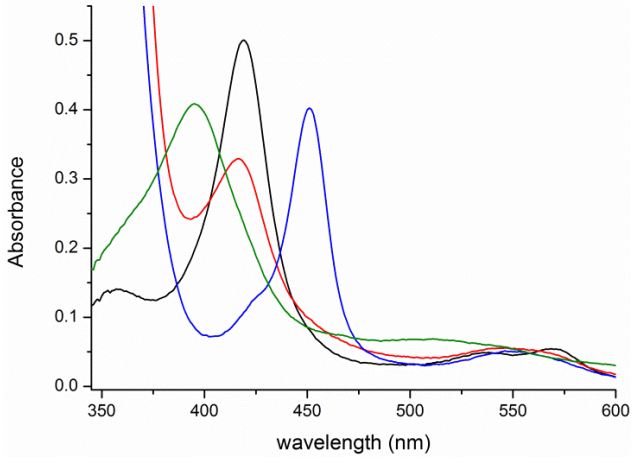
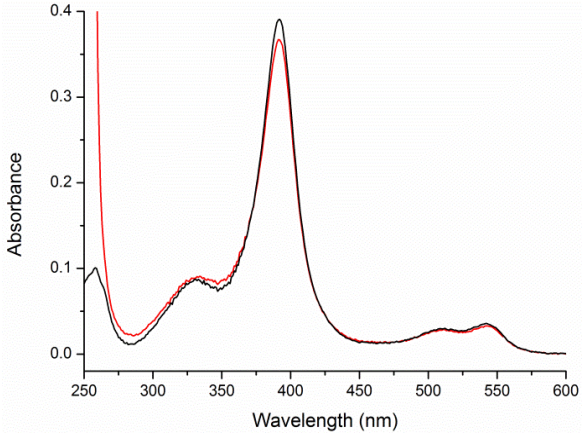


Fig. S6 (a) The UV/Vis spectra of substrate-free ferric (black), ferrous (red), ferrous-CO (blue) and 4-methoxybenzoic acid bound (green) forms of the CYP199A4 Arg92Glu mutant. Spin-state shift of the (b) Ser95Val, (c) Phe185Ile, (d) Phe185Val and (e) Arg243Thr variants of CYP199A4 with 4-methoxybenzoate. The substrate-free and substrate bound CYP199A4 spectra are shown in black and red respectively. Substrate binding titration analysis of 4-methoxybenzoate binding to the (f) WT (7 μM), (g) Phe185Val (0.25 μM), (h) Phe185Ile (0.25 μM) and (i) Arg243Thr (0.25 μM) variants of CYP199A4. The mutants were analysed with hyperbolic fitting and show a 20 to 100 fold weaker binding compared to WT.

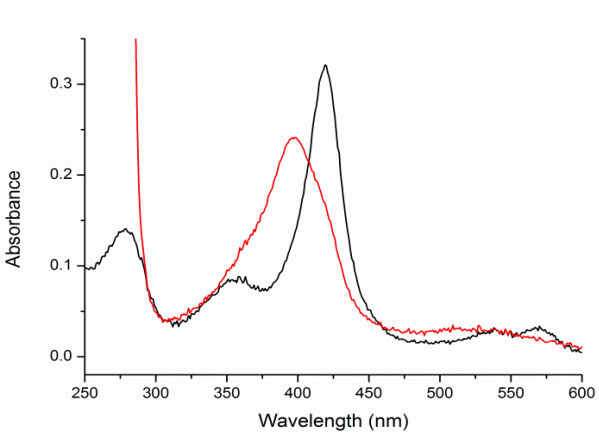
(a)



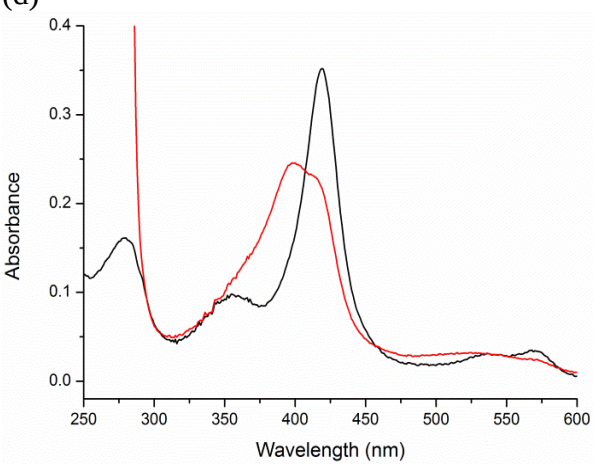
(b)



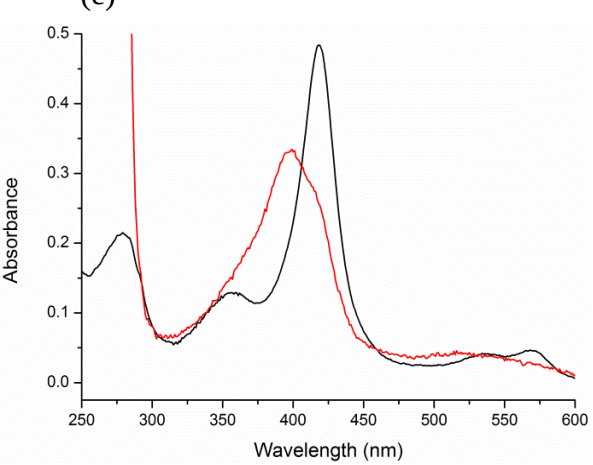
(c)



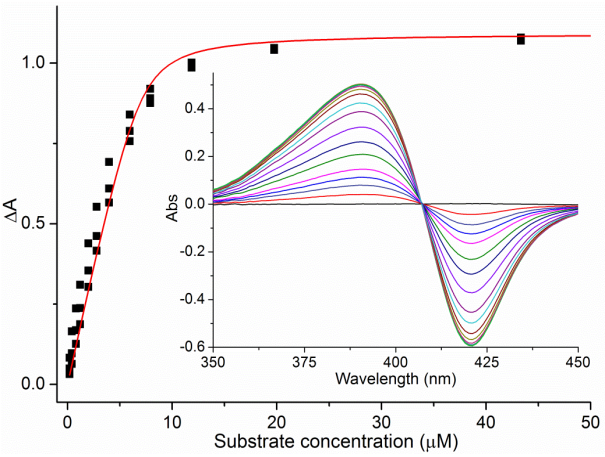
(d)



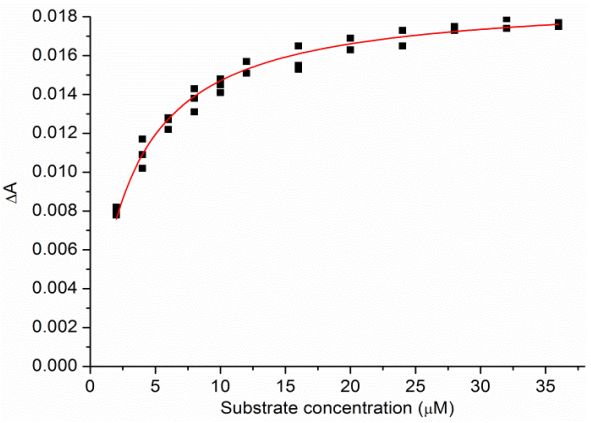
(e)



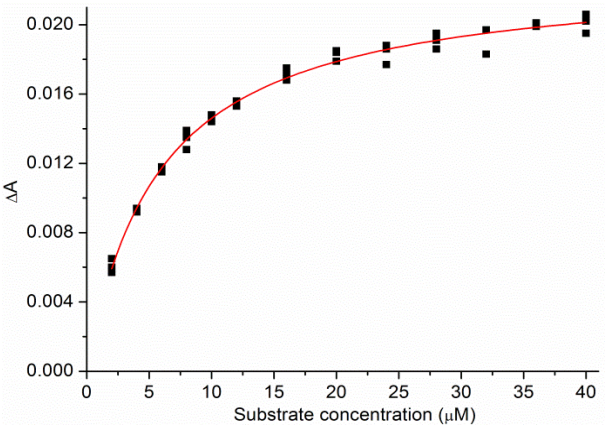
(f)



(g)



(g)



(h)

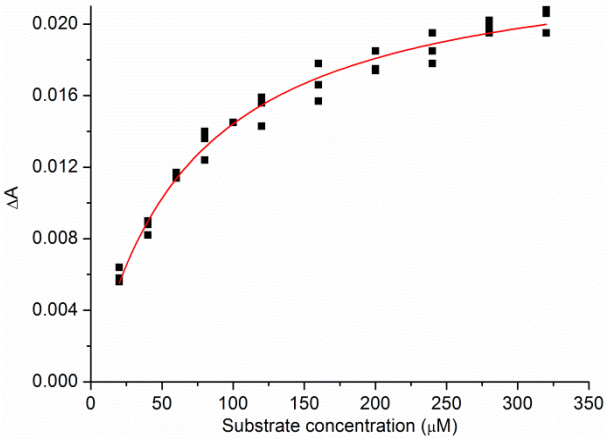


Fig. S7. (a) NADH consumption rate kinetic analysis of WT (black), Phe185Val (violet) and Arg243Glu (orange) variants of CYP199A4 with 4-methoxybenzoic acid. (b) NADH consumption rate kinetic analysis of WT (black), Phe185Ile (blue), Ser95Val (magenta) and Arg92Glu (red) variants of CYP199A4 with 4-methoxybenzoic acid.

(a) (b)

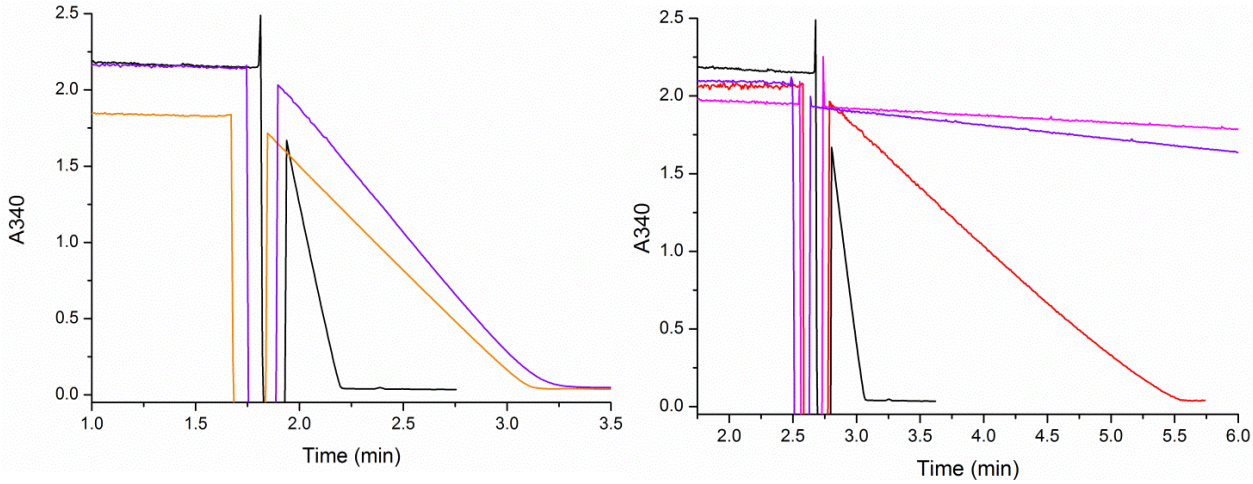
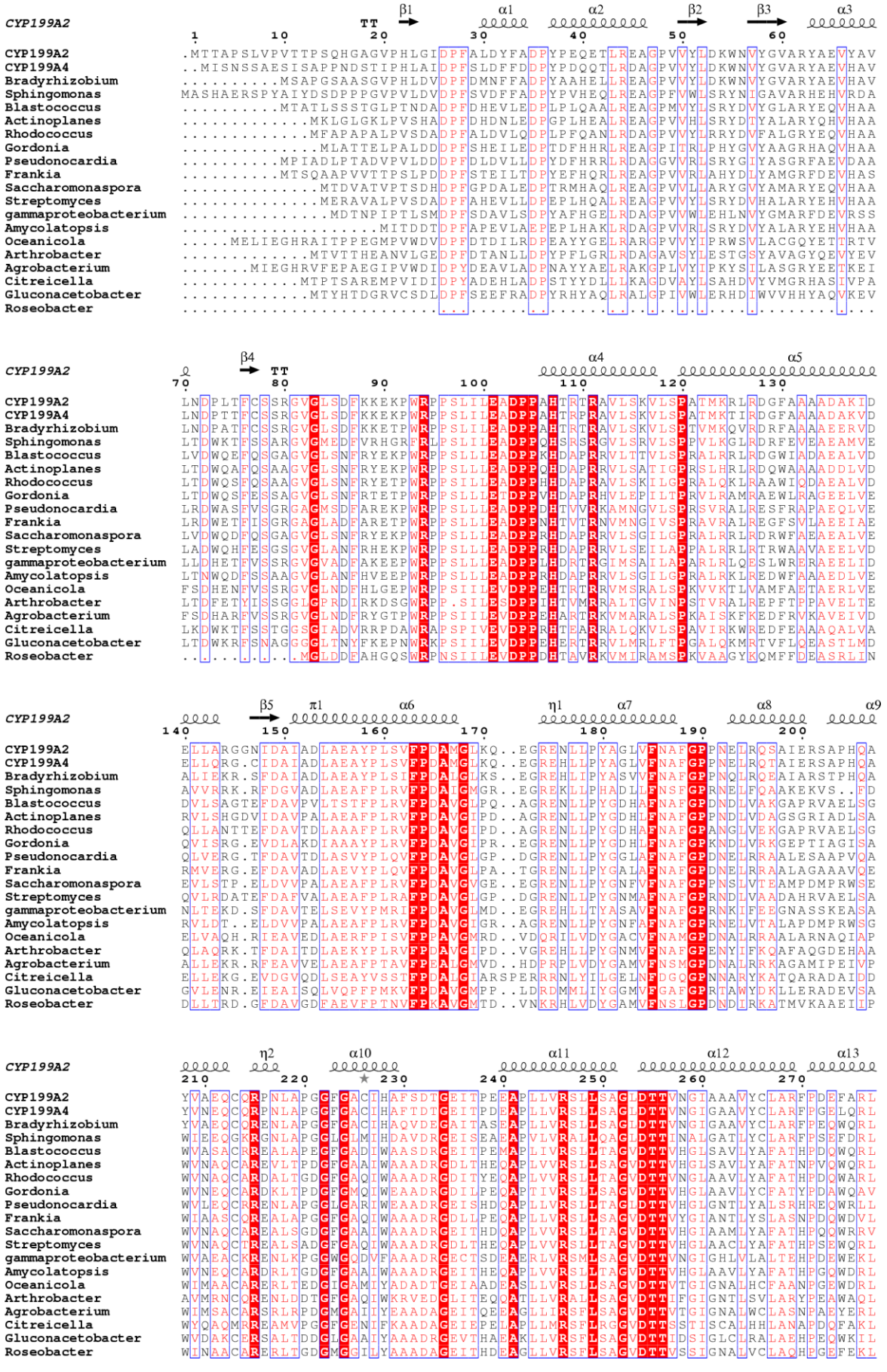
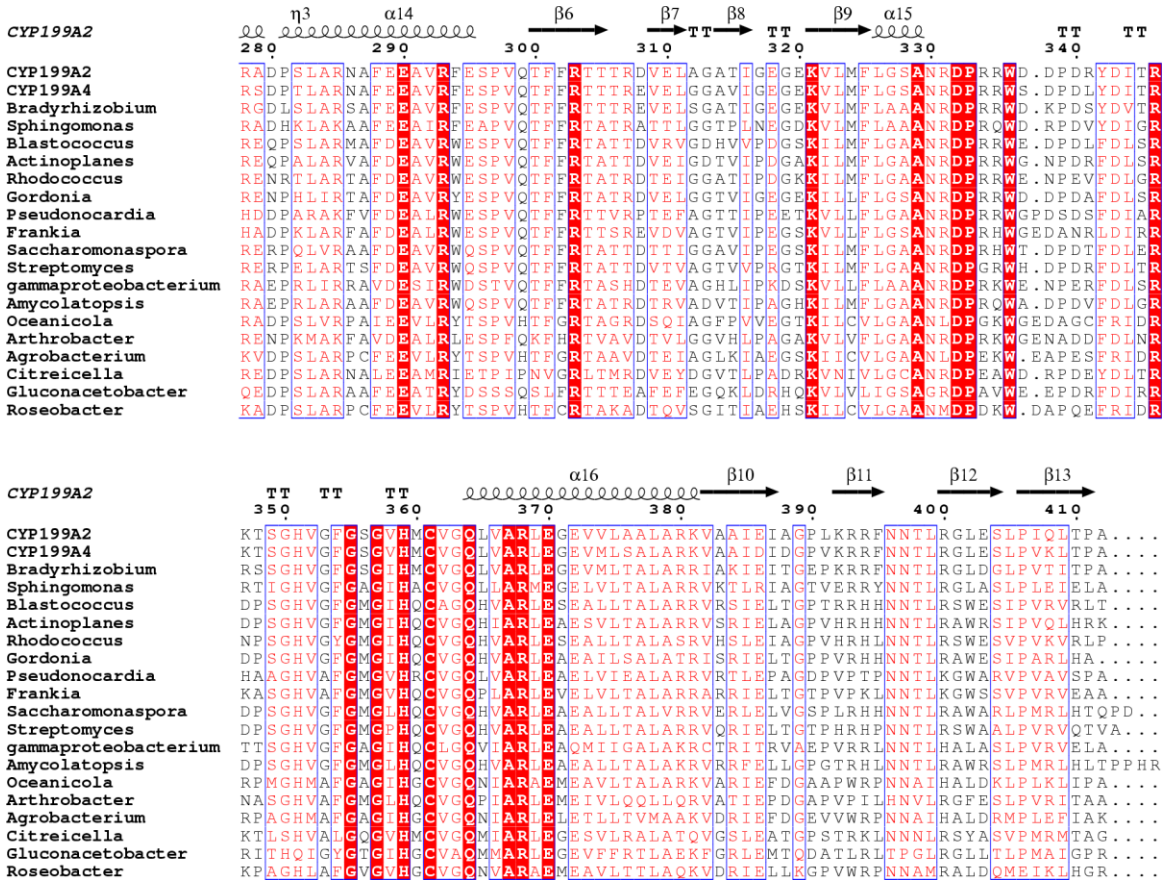


Fig. S8. Sequence alignment of CYP199A2 and CYP199A4 and other selected members of the CYP199 family.[a]





[a] The alignments are numbered according to the sequence of CYP199A2. One member of each bacteria was chosen as a representative example. Therefore other CYP199 family members from for example *Rhodopseudomonas*, *Bradyrhizobium*, *Gordonia* and *Rhodococcus* species are also known. The bacterial species, gene ID and the overall identity to CYP199A2 (in parentheses) are are listed below; *Bradyrhizobium japonicum* USDA 110, blr1048, (79%);
Sphingomonas wittichii RW1, Swit_3269, (59%);
Blastococcus saxobsidens DD2, BLASA_2994, (55%);
Actinoplanes missouriensis 431, AMIS_8190, (58%);
Rhodococcus jostii RHA1, RHA1_ro02948, (57%);
Gordonia amarae NBRC 15530, GOAMR_61_00040, (55%);
Pseudonocardia dioxanivorans CB1190, Psed_3265, (56%);
Frankia sp. EUN1f, FrEUN1fDRAFT_6833, (56%);
Saccharomonospora marina XMU15, SacmaDRAFT_1596 (55%)
Streptomyces hygroscopicus ATCC 53653, SSOG_08952, (55%);
gamma proteobacterium NOR5-3, NOR53_1231, (51%);
Amycolatopsis sp. ATCC 39116, AATC3_020100014369, (54%);
Oceanicola sp. S124, OS124_010100009522, (49%);
Arthrobacter sp. FB24, Arth_1727, (50%);
Agrobacterium vitis S4, Avi_7683 (44%);
Citreicella sp. SE45, CSE45_4276, (44%);
Gluconacetobacter oboediens 174Bp2, Gobo1_010100012098, (43%);
Roseobacter sp. GAI101, RGAI101_2590, (45%)

There was no sequence alignment of the important residues of the CYP199 family to those of the benzoic acid acid hydroxylase family from fungal species (CYP53 family).

Fig. S9. SDS-PAGE analysis of CYP199A4.