Catalysis Science & Technology

Thermostable fatty acid hydroxylases from ancestral reconstruction of cytochrome P450 family 4 enzymes

Journal:	*Catalysis Science & Technology*
Manuscript ID	CY-ART-01-2024-000090.R1
Article Type:	Paper
Date Submitted by the Author:	25-Jun-2024
Complete List of Authors:	Harris, Kurt; University of Queensland Faculty of Science, School of Chemistry and Molecular Biosciences Zhang, Yichi; University of Queensland Faculty of Science, School of Chemistry and Molecular Biosciences Yang, Jade; University of Washington Seattle Campus, Department of Pharmaceutics Zeigler, Maxwell; University of Washington, Department of Medicinal Chemistry, School of Pharmacy Thomson, Raine; University of Queensland Faculty of Science, School of Chemistry and Molecular Biosciences Carrera-Pacheco, Saskya; University of Queensland Faculty of Science, School of Chemistry and Molecular Biosicences Russell, Drake; University of Washington, Department of Medicinal Chemistry, School of Pharmacy Okada, Shoko; CSIRO Synthetic Biology Future Science Platform, Black Mountain Science and Innovation Park Strohmaier, Silja; University of Queensland Faculty of Science, School of Chemistry and Molecular Bioscience Gumulya, Yosephine; University of Queensland Faculty of Science, School of Chemistry and Molecular Biosciences Scott, Colin; CSIRO Ecosystem Science, Biocatalysis and Synthetic Biology Team Totah, Rheem; University of Washington, Department of Medicinal Chemistry, School of Pharmacy Gillam, Elizabeth; University of Queensland Faculty of Science,

Data availability

The data for this paper are provided in the Supplementary information or at the following URL: https://www.scidb.cn/en/s/mMB7Rr.

Thermostable fatty acid hydroxylases from ancestral reconstruction of cytochrome P450 family 4 enzymes

Kurt L. Harris[1], Yichi Zhang[1], Jade Yang[2], Maxwell B. Zeigler[2], Raine E.S. Thomson [1], Saskya E. Carrera-Pacheco [3], Drake Russell [2], Shoko Okada[4], Silja J. Strohmaier[1], Yosephine Gumulya[1], Colin Scott[4], Rheem A. Totah[2], and Elizabeth M.J. Gillam[1]*

[1] School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, Brisbane, 4072 Australia

[2] Department of Medicinal Chemistry, School of Pharmacy, University of Washington, Seattle, WA, 98195-7631, USA

[3] Centro de Investigación Biomédica (CENBIO), Facultad de Ciencias de la Salud Eugenio Espejo, Universidad UTE, Quito 170527, Ecuador

[4] CSIRO Synthetic Biology Future Science Platform, Black Mountain Science and Innovation Park, Clunies Ross St, Canberra, 2601, Australia

*Corresponding author: e.gillam@uq.edu.au

Running title: Thermostable fatty acid hydroxylases from ancestral CYP4 enzymes

Keywords: cytochrome P450, fatty acids, CYP4, ancestral sequence reconstruction, thermostability, biopolymers, heme-thiolate proteins

Abstract

Biopolymers produced from plant sources offer sustainable alternatives to plastics derived from petrochemicals. Hydroxylated fatty acids (OHFAs) can be readily polymerized to form polyesters. The CYP4 family of cytochrome P450 monooxygenases are potential biocatalysts for fatty acid (FA) ω-hydroxylation, but existing CYP4 enzymes are limited by their inherently low thermostability. Here we resurrected ancestral FA hydroxylases belonging to the CYP4ABTXZ clade, that were significantly more thermostable than the corresponding extant forms with $^{10}T_{50}$ values up to 68 °C. All ancestors were active towards C_{12} - C_{20} FAs, but differed in chain length preference and hydroxylation regioselectivity. While extant CYP4A and CYP4B forms preferentially hydroxylate the terminal (ω) carbon, the ancestors showed less specificity for the ω-position for C_{12} – C_{18} FAs overall. All ancestors were more active towards arachidonic acid (C_{20}) than the extant forms tested, preferring hydroxylation at the ω-position. Reactions could be made more cost effective by using O_2 surrogates, obviating the need for O_2, NADPH or a co-expressed redox partner. These enzymes will serve as robust templates for the further engineering of stable FA hydroxlases with activity towards different FAs and alkyl chain positions to produce precursors for synthetic fibres with desired properties.

Introduction

With increasing awareness of the need to find sustainable alternatives to petrochemicals, there is growing interest in plastics produced from renewable resources. Biopolymers are more environmentally friendly, potentially biodegradable alternatives to plastics derived from petrochemicals [1]. Conventional production methods are reliant on the fermentative pathways of certain bacteria to produce polymers in the form of polyhydroxyalkanoates, or monomers such as lactic acid, which are subsequently chemically polymerized to form biopolymers, e.g. polylactic acid [2, 3]. While these processes are well established, their commercial applications and ability to compete with traditional petrochemical plastics, are constrained by limited chemical diversity, expensive growth conditions and purification processes, and a general lack of control over the biosynthetic process [2].

A novel approach to biopolymer synthesis lies in the use of ω-hydroxylated fatty acids (ω-OHFAs), which can be polymerized to form polyesters [4-6]. However, ω-hydroxylation of unactivated alkyl chains is difficult to achieve by traditional chemosynthetic approaches [7]. Biocatalysis offers an alternative route to generate ω-OHFAs, and enzymes exist in nature that carry out these reactions with a high degree of regioselectivity. Cytochrome P450 enzymes (P450s) are a diverse superfamily of heme-thiolate proteins, best known for their ability to monooxygenate unreactive C-H bonds [8]. In particular, animal P450 family 4 forms (CYP4s) and bacterial family 153 forms (CYP153) are mostly known as fatty acid (FA) hydroxylases, many of which exhibit a preference for the terminal ω-carbon in the alkyl chain [9-16].

In vertebrates, CYP4 enzymes are classified into seven subfamilies: CYP4A, CYP4B, CYP4F, CYP4T, CYP4V, CYP4X and CYP4Z [9, 17]. Each subfamily and isoform demonstrates discrete preferences in terms of FA chain length and regioselectivity (**Supplementary Table 1**). Based on the well-characterized chain-length and site preferences

of the few well-studied mammalian forms, CYP4A and CYP4B enzymes appear to preferentially catalyze ω-hydroxylation of FAs, with CYP4B subfamily members primarily metabolizing FAs of 7-12 carbons (C_{7-12}) [13, 18-20], and CYP4As metabolizing C_{12-16} FAs, arachidonic acid (AA, C_{20}) and its derivatives [13, 21-29]. The CYP4F subfamily acts mainly on AA and its derivatives (C_{18-26}), with different isoforms oxidizing at the ω, ω-1 and ω-2 positions [13, 19, 23, 30-36]. CYP4Ts have only been identified in some fish and amphibia, and have not been characterized in terms of FA metabolism. However, they are induced by FA-like marine pollutants such as perfluorooctanoic acid [37, 38]. Until recently, little was known about the native substrates or hydroxylation properties of the other three "orphan" CYP4 subfamilies. CYP4V2 has recently been shown to ω-hydroxylate myristic acid (C_{14}) [39]; CYP4X1 has been shown to form epoxides of the neurotransmitter anandamide (C_{22}), and to a lesser extent, AA [40]; and CYP4Z1 has been found to catalyze the in-chain hydroxylation of lauric and myristic acid, as well as 14,15-epoxidation of AA (C_{12}, C_{14}, C_{20}) [41, 42]. Little is known about CYP4 forms from non-mammalian vertebrates.

The ability of CYP4 enzymes to selectively hydroxylate FA chains of different lengths represents an opportunity to develop a highly controlled biopolymer production process, for which enzymes can be selected and tailored to meet the specific needs of the end application. However, there are several drawbacks to the industrial use of these enzymes. Chiefly, the native forms show poor thermal and solvent tolerance, weaknesses that have previously been the focus of protein engineering studies, most commonly concerning the P450 from *Bacillus megaterium,* $P450_{BM3}$ [43-45]. Another limitation lies in their reliance on the expensive cofactor NADPH and a redox partner, cytochrome P450 reductase (CPR), which also shows poor thermal stability. One way to overcome the requirement for both NADPH and a redox partner is to shunt the P450 catalytic cycle by directly supplying an activated oxygen to the haem via oxygen surrogates such as cumene hydroperoxide (CuOOH) or hydrogen peroxide (H_2O_2).

However, peroxides can damage the protein, and many P450s show reduced or no activity with these surrogates [46, 47].

Recent studies have shown that it is possible to obtain thermostable P450 enzymes using ancestral sequence reconstruction [48-50](ASR), a bioinformatic approach which uses the amino acid sequences of extant enzymes to infer the sequence of a progenitor [51-53], which can then be resurrected by gene synthesis and recombinant expression in a heterologous host, e.g. *Escherichia coli*. ASR presents a means to both investigate the evolution of the CYP4 family in terms of substrate specificity and chemo- and regio-selectivity, and to address the inherent lack of thermostability of CYP4 enzymes for potential industrial applications. The aim of this study was to infer, resurrect and characterize ancestral CYP4 enzymes, to explore their potential for use in biocatalytic FA hydroxylation.

Results and Discussion

Ancestral sequence reconstruction

The present study reports the characterization of thermostable FA hydroxylases based on the ancestral sequence reconstruction of a set of extant, vertebrate CYP4 enzymes. A maximum likelihood, joint reconstruction approach was employed to infer the ancestor of the clade containing the CYP4A, CYP4B, CYP4T, CYP4X and CYP4Z enzymes (CYP4ABTXZ, node (N)2), along with three intermediate ancestors (CYP4BT, CYP4A, CYP4XZ, i.e., nodes 4, 122 and 198 respectively of the tree shown in **Figure 1**). The nomenclature used to describe each ancestor consists of the prefix CYP for cytochrome P450, followed by the subfamily(ies) from which it was derived (e.g., CYP4XZ, includes CYP4X and CYP4Z subfamilies). A total of 248 sequences were used for the alignment, phylogenetic tree and ancestral inference. The tree obtained generally conformed to the expected tree of life to the level of vertebrate classes; however there were discrepancies in the clustering of CYP4B and

CYP4T sequences as noted previously for the CYP4ABTXZ clade [54-57]. The arrangement of subfamilies was generally similar to that observed by Kirischian and Wilson [54], with clustering of CYP4B/T and CYP4X/Z subfamilies; but in this case the CYP4A and CYP4B/T subfamilies shared a common ancestor as reported earlier [55, 57]. Reptilian and avian CYP4B sequences were split into two branches, one of which clustered with the amphibian/actinopterygian CYP4T clade, while the other grouped within the larger CYP4B group, potentially the result of a gene duplication in the avian/reptilian ancestor. In addition, a single reptilian sequence (K7F4P4) appeared to segregate with the mammalian CYP4B sequences rather than the reptilian/avian CYP4B clade.

Uncertainties and discrepancies in subfamily arrangement, as well as uncertainties found within the CYP4B/T clade compared to the tree of life, may be explained by possible gene duplication and/or loss, or incomplete lineage sorting in closely related evolutionary branches; indeed the CYP4 gene tree is not expected to match the vertebrate species tree [58]. However the relative paucity of sequence information available for CYP4 forms from non-mammalian species made it impossible to ascertain the basis to the differences seen. The ancestors inferred and resurrected here represent only the most probable ancestors at each position in the tree, based on the available sequence data. There is no certainty that these forms represent the historically correct ancestral forms characterization of these enzymes, but they can still provide some insight into the possible evolutionary history of the CYP4s, in addition to their potential industrial applications.

Since the extant forms of well-characterized subfamilies CYP4A and CYP4B generally show a strong preference for ω-hydroxylation of FAs, CYP4A (N122) was selected for reconstruction. Since it was not possible to select a node that included all CYP4Bs without including the CYP4Ts, CYP4BT (N4) was selected for reconstruction. While CYP4Ts have not been characterized towards FAs, they are phylogenetically similar to CYP4Bs and are

induced by the presence of FA-like molecules [37]. CYP4XZ (N198) was chosen over the ancestors of the individual subfamilies, since CYP4X and CYP4Z extant forms are less well characterized, and there was limited sequence information available for each subfamily individually. The amino acid sequence identity of the ancestors to selected extant forms ranged from 53-74 % (**Table 1**) whereas the ancestors were at least 80% identical to each other. By contrast the sequence identity among the extant forms ranged from 45 to 55%. The best-characterized representative extant forms from each subfamily were included for comparison, except for the CYP4Z subfamily for which no extant form has been successfully expressed in *E. coli* to date [59] and the CYP4T subfamily for which only limited characterisation has been undertaken so far.

Expression and thermostability analysis

All ancestors could be expressed in *E. coli* and displayed a characteristic peak at 450 nm upon binding of CO to the reduced heme, which indicates an intact P450 fold and thiolate linkage to the heme prosthetic group (**Table 2**). Expression yields were similar to those seen with the corresponding extant form(s).

The $^{10}T_{50}$ or $^{60}T_{50}$ values, i.e., the temperatures at which 50% of the protein retains the heme thiolate linkage indicating a functional P450 after heating for 10 or 60 minutes, were determined by quantifying the residual folded P450 using Fe(II).CO *vs.* Fe(II) difference spectroscopy. Each ancestor showed a significant increase in thermostability over the corresponding extant form(s), with increases in $^{60}T_{50}$ and $^{10}T_{50}$ values of at least 19 ºC and 21 ºC respectively (**Table 2**, **Figure 2**). CYP4BT and CYP4ABTXZ showed the largest increases over the corresponding extant forms (CYP4BT ~22 °C ($^{10}T_{50}$) and 24 °C ($^{60}T_{50}$) greater than rabbit CYP4B1 (rCYP4B1); CYP4ABTXZ ~29-34 °C ($^{10}T_{50}$) and ~23-30 °C ($^{60}T_{50}$) higher than the extant forms) The oldest ancestor, CYP4ABTXZ had a $^{60}T_{50}$ ~29 ºC

greater than any of the extant forms tested (**Table 2**). Each of the ancestral forms showed a significant increase in stability over the extant forms, . The $^{60}T_{50}$ values decreased with increasing branch length, a measure of evolutionary distance (**Figure 2B**).

The correlation observed between thermostability and evolutionary age (**Figure 2B**) was consistent with the reconstruction of ancestral P450s from several other P450 families [48, 50, 60]. As the vertebrate ancestors are not thought to have been present in an environment requiring such elevated thermal tolerance, we hypothesise that the elevated thermostability of these forms reflects the stability of more ancient enzymes, present in primordial thermophiles. Such thermostability may have been lost by neutral drift, in the absence of purifying selection pressure, resulting in a gradual decrease in the stability of these enzymes leading up to the modern-day forms. There is some suggestion in the literature of inherent bias in maximum likelihood methods towards increased thermal tolerance [61-64], so it is possible that the enhanced thermal stability of these ancestors is an artefact of the methods used in their creation. However, such a bias would be expected to be in the order of a few degrees, not tens of degrees as observed here [65, 66].

Transformation of the common marker substrate, Luciferin-MultiCYP

Luciferin-MultiCYP, a common substrate of many P450 forms from multiple families, was metabolised by all ancestral forms showing that they were catalytically active (**Figure 3**). The Luciferin MultiCYP demethylase activity of the ancestors was generally comparable to that of the extant forms, but overall, hCYP4A11 showed the highest activity (**Figure 3**). Amongst the ancestors, CYP4ABTXZ showed the highest activity while CYP4A and CYP4XZ showed the lowest activity towards this substrate.

Metabolism of C_8-C_{18} FAs and AA

However, the interest in CYP4 enzymes lies in their ability to hydroxylate unreactive FAs. Initially, the activity of the ancestors towards saturated C_{8-18} FAs (caprylic acid (C_8), capric acid (C_{10}), lauric acid (C_{12}), myristic acid (C_{14}), palmitic acid (C_{16}) and stearic acid (C_{18})), typically metabolized by hCYP4A11 and rCYP4B1 was assessed.

Of the C_{8-18} FAs metabolised by ancestral CYP4 forms, lauric acid was oxidized to the least extent by all forms (**Figure 4**), despite this FA being identified as a preferred substrate for both rCYP4B1 and hCYP4A11 [18, 25, 67, 68]. However lauric acid was also the only substrate from this set (C_{12-18}) for which any ancestor showed a preference for the ω- over the ω-1 position. CYP4ABTXZ and CYP4A showed ω:ω-1 ratios of 2.1:1 and 1.6:1 respectively, but these ratios are markedly lower than those observed for many extant CYP4A and CYP4B forms (**Supplementary Table 1**).

In general, the ancestors tended to prefer the ω-1 position over the ω position for C_{8-18} FAs (**Figure 4**). However CYP4ABTXZ showed activity towards the ω-position comparable to, or higher than, that towards the ω-1 position for all FAs except palmitic acid (C_{16}). By contrast, CYP4XZ showed the most activity towards ω-2 and ω-3 hydroxylation. In terms of chain length, CYP4ABTXZ and CYP4XZ preferred myristic acid (C_{14}), while CYP4A preferred stearic acid (C_{18}). CYP4BT showed the lowest activity overall compared to the other ancestors.

Interestingly, no activity was detected with any ancestor towards caprylic (C_8) or capric (C_{10}) acid (results not shown), FAs that are preferentially metabolised by rCYP4B1 at the ω-position [18]. However, capric acid was previously found to be ω-hydroxylated by rat, mouse, beef and pigeon liver microsomes, but not preparations from bullfrog or carp, suggesting that this activity may be limited to CYP4B forms in mammals and birds, and not a property of the corresponding CYP4T forms from fish and amphibians [69]. Since the CYP4BT ancestor,

which gave rise to both subfamilies, showed no activity towards this substrate, this activity may have evolved in the CYP4B subfamily rather than having been lost in the CYP4T subfamily lineage.

In contrast to the results with C_{8-18} FAs, when activity towards AA (C_{20}) was investigated, all ancestors exhibited a strong preference for ω-hydroxylation (**Figure 5**). Of the extant forms, only hCYP4A11 showed significant activity towards AA (**Figures 5A, 5F**). Characterised CYP4A and CYP4B forms primarily hydroxylate shorter FAs; however extant CYP4A forms also metabolize oleic acid and AA to 20- and 19-HETE to a lesser extent [25]. Human CYP4X1 is only reported to have low-level internal epoxidase activity towards AA [21, 24-26, 40, 67].

All ancestors produced 20-hydroxyeicosatetraenoic acid (20-HETE) as the major hydroxylation product, followed by 19-HETE (**Figure 5; Supplementary Figure 2)**. All ancestral forms produced 14,15-epoxyeicosatetraenoic acid (14,15-EET) in significant quantities, with trace amounts of 11,12-EET in incubations with CYP4XZ and CYP4ABTXZ (**Figure 5; Supplementary Figure 3**). hCYP4A11 produced 20- and 19-HETE, while neither rCYP4B1 nor hCYP4X1 showed significant amounts of any product. Amongst all forms, CYP4BT, CYP4ABTXZ and CYP4A produced the most 20-HETE. CYP4BT also produced the most 19-HETE, followed by CYP4A and CYP4ABTXZ.

In general, CYP4BT was less active than the other ancestors towards C_{12-18} substrates (**Figure 4**). It showed a slight preference for the ω-1- over the ω-position for these substrates, however for AA (C_{20}) it favored the ω-position (**Figure 5**). The apparent selectivity of this enzyme for longer chains, with reduced regioselectivity for shorter chains, contrasts with the characterized mammalian CYP4B forms, which hydroxylate the ω-position of C_7-C_{12} FAs [18, 20, 67, 70]. No activity towards AA was seen with CYP4B1 consistent with what is known of the characterized mammalian CYP4B forms [18, 20, 67, 70].

Human CYP4B1 is reportedly inactive; however when a single amino acid in the meander region was mutated (Pro427Ser), it demonstrated similar substrate- and regioselectivity towards C_{8-12} compared to the rabbit ortholog [67]. Serine was found at this position in the CYP4BT ancestor, so this is unlikely to have affected its activity compared to other forms. Limited studies have been reported using CYP4T enzymes so it is unclear whether they can hydroxylate FAs and how that impacts the activity of the ancestor.

The oldest ancestor preferentially hydroxylated myristic acid (C_{14}) (**Figure 4**) with an approximate 1:1 ratio of ω:ω-1 hydroxylation, and was also highly active towards AA (C_{20}), producing predominantly 20-HETE (~5:1 ω:ω-1 ratio, **Figure 5**), suggesting that this may be an ancestral activity of the 4ABTXZ clade. Given the important role that 20-HETE plays in blood pressure homeostasis in modern vertebrates, it would make sense for this activity to have been developed early on in CYP4 family evolution.

Like CYP4ABTXZ, CYP4XZ preferred myristic acid as a substrate, but showed a high degree of regioselectivity for the ω-1 position (ω:ω-1:ω-2 ratio of 1:13:4). In general, CYP4XZ showed a preference for in-chain hydroxylation (**Figure 4**), consistent with the limited studies of hCYP4Z1, which carries out in-chain hydroxylations at the ω-(1-5) positions for lauric and myristic acid [41]. The exception to this was AA, for which it was selective for the ω-position (**Figure 5**). However amongst the ancestors, this form produced the most 14,15-EET, a product reportedly formed by hCYP4Z1 and to a lesser extent, hCYP4X1 [40, 41]. However, in the present study, hCYP4X1 did not produce significant quantities of 14,15-EET or other epoxidation products compared to the negative control. The ability of CYP4XZ to hydroxylate subterminal positions of most FAs may reflect an intermediary shift in regioselectivity from the greater extent of terminal hydroxylation catalyzed by the CYP4ABTXZ ancestor, towards the reported in-chain hydroxylation

preferences (ω-4-OH for lauric and ω-3-OH for myristic acid for CYP4Z1; 14,15-epoxide for AA for CYP4Z1 and CYP4X1) of the extant forms.

CYP4A showed a preference for longer chains, with its activity increasing consistently with chain length (**Figure 4**). It selectively hydroxylated the ω-position of lauric acid, consistent with most characterized CYP4As, e.g., from human, rabbit, koala and rodents, being identified as lauric acid ω-hydroxylases [21, 24-26, 71]. However, for intermediate chain lengths (C_{14-18}) it became more selective for the ω-1 position, up until AA (C_{20}), for which it also demonstrated a preference for ω-hydroxylation (~3:1 ω:ω-1 ratio, **Figure 5**). In the present study, hCYP4A11 produced both 20-HETE and 19-HETE consistent with the literature[21, 24, 25, 72]. The increased activity of CYP4A for longer chain lengths, and its reduced regioselectivity for terminal hydroxylation of intermediate chains (C_{14-16}) suggests that the preference of modern forms for lauric acid may be recently evolved.

Homology modelling of CYP4 ancestors

To rationalize the differences in activity between the ancestors and their extant descendants, we undertook homology modelling of the ancestors based on the structure of rabbit CYP4B1 with octane bound (PDB: 5T6Q) and also predicted the structures of the ancestors using AlphaFold. The active site residues of the ancestors were compared to those of the extant forms (**Figures 6 and 7**). So far, the only CYP4 crystal structures available are of rabbit CYP4B1, in complex with octane (PDB: 5T6Q) and the inhibitor HET0016 (PDB:6C94) [70, 73]. **Figure 6** displays residues of interest mapped to homology models of the ancestors, prepared using rCYP4B1 (PDB: 5T6Q) as the template structure. Key active site residues identified in these studies as well as homology models with the other CYP4s are highlighted in **Figure 7**, along with the corresponding residues in the ancestral and characterized extant forms.

The heme-linked residue Glu310 in the I-helix, which is responsible for forming a covalent, ester linkage to the 5-methyl of the heme, has been proposed to be critical in positioning substrates for ω-hydroxylation[70]. (All residue numbering used herein is based on wild type rCYP4B1, UniProt: P15128, for consistency with Hsu et al. [70]). CYP4A and CYP4B extant forms contain Glu at this position, while CYP4X1 and CYP4Z1 are predicted not to have a covalently linked heme [42, 74-77]. All CYP4 ancestors except CYP4XZ contained Glu310 (**Figures 6 and 7**), suggesting that CYP4ABTXZ, CYP4BT and CYP4A were likely to form a covalent linkage with the heme, but CYP4XZ was not. In rCYP4B1, mutation of this residue to Ala, Gln or Asp significantly decreased ω-hydroxylase activity [70]. CYP4XZ contains Ala at this position, which may be responsible, at least in part, for its reduced terminal hydroxylase activity towards C_{12-18} FAs. However, the general preference for in-chain hydroxylation of these FAs by the other ancestors suggests that additional factors also influence selectivity for the ω-position.

Most other residues contributing to the shape of the active site hydrophobic pocket for CYP4B1 are conserved for all ancestors (His312, Phe309, Asp313, Thr314, Val486, Val375, Pro376, Tyr110, Leu122), in particular those closest to the heme iron. However, Leu485, which is highly conserved in all extant human CYP4 ω-hydroxylases but not in non-ω-hydroxylases, was replaced with Ile in the ancestors, analogous to the sub-terminal hydroxylase CYP4F8 (**Figure 7**) [32]. Mutation of this residue to Leu in the ancestors may lead to better ω-hydroxylation activity.

In CYP4B1, Val109, Phe113, Ile221, Met217, Gln218, Val378, Tyr379 and Gln377 are all implicated in shaping the binding pocket. The more distal, hydrophobic residues, Val109, Phe113, Ile221, and Met217 wrap around one side of the binding pocket along with polar Gln218, while Val378 and Tyr379 and polar Gln377 are near the top of the binding cavity in rCYP4B1. The ancestors contain nonconservative substitutions at Val109 (to Ser), Gln218 (to

Phe or His), Gln377 (to Ser or Gly) and Tyr379 (to Ser or Gly) (**Figures 6 and 7**), changing many of these positions to more polar residues, similar to the CYP4A, CYP4X and CYP4Z extant forms. Such changes could alter the hydrophobicity and size of the substrate binding pocket to accommodate larger substrates, potentially explaining the increased activity of the ancestors towards AA.

Val122 in CYP4Z1 is proposed to help with substrate positioning [42]. All ancestors contained leucine at this position, mirroring the majority of extant forms. In CYP4Z1 docking models, Val122 is in close proximity to the ω-carbon of lauric acid, with hydrophobic interactions playing a potential role in substrate positioning for internal hydroxylations [42]. The change to Leu122 in CYP4XZ, as seen in all ω-hydroxylases (**Figure 7**), may position the ω-end of medium chain FAs closer to the heme and cause a shift in selectivity towards the ω-carbon [42], resulting in the more terminal regioselectivity of CYP4XZ than hCYP4Z1 or hCYP4X1. Residues in positions identified as potentially important for CYP4Z1 substrate preference (113, 122, 375, 377, 379) [42] were conserved between CYP4XZ and CYP4ABTXZ, suggesting that these amino acids might confer a substrate preference for myristic acid. Key residues identified by this analysis which could make potential targets for mutagenesis, are Ile485 for all ancestors, Ala310 for CYP4XZ and Leu221 for CYP4ABTXZ and CYP4XZ. A full sequence alignment of the ancestors and extant forms, with structural information from rCYP4B1, can be found in **Supplementary Figure 1**.

Activity supported by O_2 surrogates

The improved thermostability of these ancestors makes them suitable for use in industrial biocatalysis, as thermostable enzymes typically show improved shelf-lives, and can carry out reactions at elevated temperatures or for longer periods[78, 79]. However the cost of cofactor supply presents an additional hurdle to the implementation of P450s as biocatalysts.

Therefore we investigated the ability of CYP4 forms to be supported by oxygen surrogates via the peroxide shunt [80-84]. Luciferin-MultiCYP dealkylation by ancestral and extant CYP4s was found to be supported at least one of the oxygen surrogates, CuOOH and BAIB, and, in the cases of hCYP4A11 and CYP4XZ at levels comparable to hCPR-supported catalysis. All ancestral forms and extant hCYP4A11 and rCYP4B1 were supported effectively by cumene hydroperoxide (CuOOH), although the relative activity compared to hCPR/NADPH-supported activity varied (**Figure 8**). By contrast, only hCYP4A11, CYP4XZ and CYP4ABTXZ showed significant activity with bis(acetoxy)iodobenzene (BAIB). At the single substrate and P450 concentration tested, hCYP4A11 and CYP4XZ showed comparable activity with 10-50 μM M BAIB and 0.5 mM CuOOH as did rCYP4B1 with 50-500 μM CuOOH. However, in all other cases O_2 surrogate-supported activity was markedly lower than that supported by the CPR and NADPH. Likewise, activity of hCYP4A11, CYP4XZ and CYP4ABTXZ towards AA supported by CuOOH was markedly lower than that supported by NADPH and hCPR (Supplementary Figures 5 and 6).

Oxygen surrogates are known to damage the P450s heme and apoprotein; therefore the stability of each P450 was assessed at the highest concentration used for supporting activity. Interestingly, the stability of the extant and ancestral P450s appeared to differ between lineages (Supplementary Figure 7). Whereas the CYP4ABTXZ ancestor and both ancestral and extant CYP4A and CYP4BT lineage forms were relatively stable over two-hour incubation with CuOOH, hemoprotein levels declined rapidly in incubations with CYP4XZ and CYP4X1. Interestingly, the CYP4XZ lineage lacks the conserved glutamate residue needed for formation of the covalent linkage between the heme and protein, whereas this glutamate is present in the oldest ancestor and CYP4A and CYP4BT ancestral and extant forms.

The use of an oxygen surrogate in place of hCPR would allow reactions to be carried out at elevated temperatures, while eliminating the need for the expensive cofactor, NADPH, oxygenation, or coexpression of a redox partner. While activity is reduced in incubations supported by CuOOH, the relative stability of the CYP4A and CYP4ABT lineages to oxidative damage suggests that

Conclusions

In summary, the present study has identified four thermostable CYP4 enzymes capable of hydroxylating FAs and demonstrated that they can be supported by inexpensive oxygen surrogates. While it was hypothesized that these ancestral CYP4 forms would favor ω-hydroxylase activity similar to the extant forms, for C_{12-18} FAs, hydroxylation was generally higher at the more thermodynamically favourable ω-1 position, suggesting reduced regioselectivity towards C_{12-18} substrates. The ancestors were considerably more active towards AA, while only hCYP4A11 showed any significant activity amongst the extant forms. Thus, metabolism of AA may represent an ancestral activity of the CYP4 family which has been lost in all but the CYP4A subfamily of those branches analysed here. Importantly, however, most of what is known of the preferences of CYP4 enzymes towards FA hydroxylation has been gleaned from studies on mammalian CYP4 forms, so it is possible that the substrate and reaction specificity of non-mammalian vertebrate CYP4 forms is quite different, and possibly more congruent with the ancestral FA hydroxylation specificities found here.

Hydroxylation at the terminal position of FAs is required for the synthesis of straight-chain polyesters, making the relatively high ω-hydroxylase activity and selectivity of CYP4ABTXZ and CYP4A towards AA highly desirable. Additionally, (ω-1)-OHFAs may

also be suited to biopolymer synthesis in some situations, as these compounds have greater relative reactivity compared to other OHFAs CYPs, and can be polymerized to yield synthetic fibres with excellent characteristics [85-87]. These forms will be good starting points for further protein engineering or directed evolution studies to optimize the regioselectivity of FA hydroxylation and activity towards desired FAs.

Investigation of the ability of engineered CYP4 forms to use cheaper feedstocks such as household vegetable oils (primarily composed of unsaturated C_{18} FAs [88]) and nutmeg butter (mainly composed of myristic acid [89]) will further enhance the potential commercial applications of these ancestors.

Methods and Materials

Materials

pCW plasmids containing human CYP4X1 (hCYP4X1) and human NADPH-cytochrome P450 reductase (hCPR) were obtained from Prof. P. F. Guengerich (Vanderbilt University, Nashville, USA) [90, 91]. The expression vector for human CYP4A11 (hCYP4A11) was obtained from Prof. J. Capdevila (Vanderbilt University, Nashville, USA) and the rCYP4B1 expression vector was as previously published [92]. The chaperone co-expression plasmid, pGro7, was donated by Prof. K. Nishihara (HSP Research Institute, Kyoto, Japan) [93]. The *E. coli* strain, DH5αF'IQ was purchased from Thermo Fisher Scientific (Scoresby, Australia). Gene strings and oligonucleotides were synthesized by GeneArt Thermo Fisher Scientific (Scoresby, Australia). Restriction enzymes, Gibson assembly kit and buffers were procured from NEB Biolabs (Ipswich, MA, USA).

Media ingredients were obtained from Becton Dickinson Pty Ltd (North Ryde, Australia). Luciferin-MultiCYP was purchased from Promega (Sydney, Australia). Emulgen 913 and 911 were supplied by Kao Chemicals (Kao Corporation, Japan). 1,2-Didodecanoyl-sn-

glycero-3-phosphocholine (DLPC) was obtained from Avanti Polar Lipids (Alabaster, AL, USA). All other materials including FA substrates and standards (sodium octanoate, sodium decanoate, sodium laurate, sodium myristate, sodium palmitate, sodium stearate, 8-hydroxyoctanoic acid, 9-hydoxydecanoic acid, 10-hydroxydecanoic acid and 15-hydroxypentadecanoic acid) and oxygen surrogates (BAIB and CuOOH) were purchased at the highest available grade from Sigma-Aldrich-Merck (Castle Hill, Australia) where possible, or at the highest grade available from a local supplier.

Ancestral sequence reconstruction

Amino acid sequences belonging to the CYP4A, CYP4B, CYP4T, CYP4X and CYP4Z subfamilies (denoted herein as the CYP4ABTXZ clade) were collected from the UniProt and NCBI databases via BLAST (Basic Local Alignment Search Tool) [94] using representative sequences selected from diverse phylogenetic classes and CYP4 subfamilies (Uniprot IDs: E7EVL8, W5UGX4, Q5XG47, H0ZDW2, P13584-2, Q8N118, Q86W10, H9GBN8, U3IPW5). CD-HIT [95] was used to remove sequences sharing >95 % identity to other forms in the collection in order to meet the cut-off of 300 sequences required by the ASR tool FastML [96]. Sequences were aligned using MAFFT [97] and curated manually to remove those that: were inconsistent with the subfamily characteristics (<55% AA identity globally or locally for over 20 residues); showed internal gaps longer than 10 residues; or contained likely sequencing errors (strings of undefined residues, recurring repeats, etc.) or incomplete open reading frames (missing exons). Where inserts were not found in the majority of subfamilies, the relevant sequences were removed from the alignment. Sequences were realigned with human CYP4V2 as an outgroup, and then a maximum likelihood phylogenetic tree was inferred using MEGA [98]. The resulting tree and MSA were submitted to the FastML web server [96], and ancestors were selected from the maximum likelihood, joint reconstruction of the CYP4ABTXZ clade representing the CYP4A subfamily ancestor, and common ancestors of

the CYP4B/CYP4T and CYP4X/CYP4Z clades, denoted herein as CYP4BT and CYP4XZ (**Figure 1**).

Resurrection of inferred ancestral forms

Amino acid sequences of the CYP4ABTXZ, CYP4BT, CYP4A and CYP4XZ ancestors were reverse translated to DNA using the GeneArt codon optimisation algorithm, optimising for *E. coli* expression and avoiding restriction sites of interest for cloning (i.e., NdeI, XbaI, HindIII, SalI). N-terminal modifications were considered for all forms and drawn from the modifications made for successful expression of extant CYP4 forms (**Supplementary Table 2**). The hCYP4A11-like modification was applied to CYP4ABTXZ, as the hCYP4A11 sequence showed the greatest sequence similarity (74%) to the ancestor (**Table 1**). The CYP4B1-like modification was applied to all intermediate nodes, as this sequence retained the greatest length of native sequence, allowing for subsequent shortening to apply alternative N-terminal modifications if required for expression (**Supplementary Table 2**) [40, 99-103]. Silent mutations were introduced into the DNA sequence from -21 to +96 nucleotides with reference to the start codon to minimize mRNA secondary structure formation, using mRNA Optimiser [104] followed by manual adjustment. Secondary structure formation was visualized in NUPACK [105]. A hexa-His tag was added to the C-terminus of each sequence to facilitate subsequent purification of the recombinant protein using immobilized metal affinity chromatography. Gene strings were synthesized with a 60 nucleotide overlap with the target plasmid at each end. The Gibson assembly method [106] was used to insert ancestral sequences into the pCW vector in bicistronic format with hCPR. Each ancestor was then subcloned into the pCW vector lacking hCPR (monocistronic format).

P450s and CPR were co-expressed and bacterial membranes were isolated as previously described [107] with the exception of hCYP4X1, which was expressed using a slight modification of the established method in which cultures were grown for 2 h at 37 °C to an

OD600 of ~0.5 before induction and then harvested after 20 h shaking at 28 °C. CPR was expressed for purification using media and additives described by Notley et al. [107], but using the protocol of Park et al. [108]. Cytochrome P450 content was quantified in whole cells and membranes using CO-difference spectroscopy as previously described [109] and CPR content was quantified using cytochrome *c* as a surrogate electron acceptor [110]. Thermostability of P450s was measured in whole cells as described .

Purification of enzymes

Membranes containing recombinant P450s were diluted six-fold in solubilization buffer (100 mM potassium phosphate buffer, pH 7.4, with 20% (w/v) glycerol, 0.1 mM DTT, 0.1 mM EDTA, and 1% (w/v) Emulgen 913) containing 1 mM phenylmethylsulfonyl fluoride and Protease Inhibitor Cocktail (Sigma Aldrich, Sydney, Australia) at 1000-fold dilution. Membranes were solubilized overnight with gentle mixing at 4 °C. Solubilized protein was isolated by ultracentrifugation at 250 000 x g for 60 minutes, and the supernatant was loaded onto a 1 ml HisTrap HP column (GE Health Care, Parramatta, NSW, AU) pre-equilibrated with solubilization buffer. The column was washed with solubilization buffer lacking detergent but containing 40 mM imidazole and then the protein was eluted with 500 mM imidazole. CYP4 proteins were further purified using a HiLoad Supadex 200 pg preparative size exclusion column (GE Health Care, Parramatta, NSW, AU) with solubilisation buffer minus detergent and imidazole. Fractions showing the highest A_{280} were collected and pooled before concentration using an Amicon Ultra-15 centrifugal filter unit (Merck, French's Forest, Australia). Aliquots (100 µl) were stored at -80 °C until use. For purification of CYP4XZ, the phosphate concentration of all buffers was increased to 300 mM, based on the previous work with hCYP4X1 [40], and the observation of protein aggregation at lower buffer strength.

CPR was purified as described [108], with modifications. Specifically, cells were lysed using a One Shot cell disrupter (Constant Systems Ltd., Low March, Daventry, UK) and membranes prepared as described [111]. Emulgen 911 1% (v/v), was used for solubilization instead of CHAPS and Tergitol. Imidazole concentrations of 40 mM and 300 mM were used for the wash and elution respectively, and salt and imidazole were removed using a PD-10 desalting column (GE Health Care, Parramatta, NSW, AU). Samples were stored in 100 µl aliquots at -80 °C.

Luciferin-MultiCYP demethylation assays

Demethylation of Luciferin-MultiCYP was measured according to the manufacturer's instructions, using microsomal membranes with a P450 concentration of 10 nM and a substrate concentration of 50 µM [112]. Since a low level of apparent activity was observed with hCPR in the absence of any P450, total hCPR concentrations were standardised by addition of membranes from cells expressing hCPR alone. Two different hCPR-only controls were used to match the hCPR concentrations in P450-containing incubations; hCPR#2 for hCYP4X1 (5.6 nM hCPR) and hCPR#1 (10.2 nM hCPR) for all other forms. Reactions were initiated by the addition of an NADPH-generating system (NGS; final concentrations: 250 µM $NADP^+$, 10 mM glucose-6-phosphate, 0.05 U/mL glucose-6-phosphate dehydrogenase) and carried out for 60 minutes at 37 °C.

For the assessment of O_2-surrogate supported activity, experiments were carried out as above, but with a substrate (Luciferin MultiCYP) concentration of 50 µM and a P450 concentration of 10 nM for a period of 60 minutes at 37 °C. Reactions were initiated either by the addition of an NGS, or by the addition of CuOOH, or BAIB at a final concentration of 2, 10, 50 or 500 µM.

C_8-C_{18} FA hydroxylation assays

Hydroxylation of C_8-C_{18} FAs was analysed with purified P450 (final concentration 0.2 μM) and reductase (0.4 μM) reconstituted in DLPC (150 μM). Enzymes were initially combined in the smallest possible volume (10-20 μl) and incubated for an hour at room temperature. Then, substrate was added to a final concentration of 100 μM, samples were made up to 450 μl with 100 mM potassium phosphate buffer, pH 7.4, and reactions were initiated by the addition of 50 μl NGS. Reactions were allowed to proceed for 30 minutes at 37 °C before addition of the internal standard, 15-hydroxypentanoic acid, to a final concentration of 10 μg/ml and quenching with 500 μl of ice cold 0.1 M HCl. Samples were extracted twice with 700 μl hexane:chloroform (8:1), and the organic phase was collected and dried down under N_2.

Lipids were methylated by the addition of 1 ml 10:1:1 CH_3OH:HCl:$CHCl_3$ at 100 °C for 1 hr, and the resulting methyl esters were extracted with hexane. The hexane phase was collected and samples were dried down under N_2 and the hydroxyl groups were derivatized with an excess of N,O-bis(trimethylsilyl)trifluoroacetamide at 90 °C for 1 hr. Samples were then analysed directly by gas chromatography mass spectrometry.

TMS-derivatized FA methyl esters (FAMEs) were analyzed on an Agilent 7890A gas chromatograph coupled to a 5975C mass spectrometer and a BPX70 column (SGE Analytical Science, 30m x 0.25 mm ID x 0.25 μm film thickness). Helium was used as the carrier gas at a constant flow rate of 2 ml/min, and 1 μl was injected into the column in split-less mode. For C_{12}-C_{18} substrates the column oven was initially held at 100 °C for 1 min, then heated at 20 °C/min to 150 °C, followed by 2.5 °C/min to 175 °C, then 40 °C/min to 250 °C, at which point it was held for 5 min. The inlet and detector were both set to 250 °C. The source and quadrupole of the mass spectrometer were heated to 230 °C and 150 °C, respectively, and a mass range of 40 to 400 a.m.u. was collected after the first 3 min of the protocol. For C_8 and

C_{10} substrates a slight modification of the above method was used: the oven was initially held at 100 °C for 3 min, then heated at 7 °C/min to 240 °C, and held for 5 min.

Analysis of AA metabolism

AA (C_{20}) incubations with P450s were conducted using a modified method [113]. Both ancestral and extant P450s were incubated at a total protein concentration of 0.4 mg/mL in a total reaction volume of 200 µL. The incubation buffer was nitrogen purged 100 mM potassium phosphate buffer, pH 7.4, 3 mM $MgCl_2$, and 4 µM pyruvate +/- 5 µM AA. P450 membranes were preincubated at 37°C for 5 minutes before the addition of NADPH (20 µL, 1 mM final incubation concentration) to initiate the reaction. After a 30-minute incubation at 37 °C, the reaction was quenched with 400 µL ice-cold ethyl acetate with antioxidants butylated hydroxytoluene (5 mM) and triphenylphosphine (0.5 mM). An internal standard mixture of 20 µL of 25 ng/mL 14,15 EET-d_{11}, 8,9 EET-d_{11}, 5,6 EET-d_{11}, and 12 HETE- d_8 was added before the quenched reaction mixture was centrifuged at 4,000 xg for 10 minutes. Samples were extracted with ethyl acetate twice. The organic phase was combined and evaporated to dryness with nitrogen gas.

Samples from the P450 AA extraction were derivatized with 1-ethyl-3-(-dimethylaminopropyl) carbodiimide (EDC), 1-hydroxybenzotriazole (HOBT), and 1-(4-(aminomethyl) phyneyl) pyridine-1-ium chloride (AMPP) as described in Zeigler et.al. [114]. The dried reaction mix was reconstituted with 30 µL of acetonitrile, 10 µL of 640 mM EDC, 20 µL of 5 mM HOBT, and 20 µL of 15 mM AMPP added sequentially before incubation for 30 minutes at 60°C. After incubation, the derivatized samples were cooled on ice for 5 minutes and the final volume was brought up to 160 µL with optima water. Samples were then centrifuged for 10 minutes at 4,000 xg and the supernatant was collected for analysis via LC-MS.

HETE and EET quantification was performed on a Waters Xevo-TQs-ESI mass spectrometer in positive mode using an Acquity UPLC BEH Shield RP18 column (1.7 μM, 2.1 x 100 mm) and Coretecs C18 guard column (1.6 μM). Samples were analyzed using an injection volume of 10 μL, 0.45 μL/min flow rate and with the LC column heated to 50°C. Mobile phases A (water) and B (acetonitrile) both contained 10 mM formic acid. Chromatographic separation was achieved with a mobile phase gradient of 0-11.6 minutes 65% A, 11.6-12.5 minutes 10% A, and 12.5-15 minutes 65% A. A representative LC chromatogram is shown in **Supplementary figure 4**. The mass spectrometer was set to positive ionization, capillary voltage 1 kV, cone voltage 10 V, source temperature 150°C, and desolvation temperature of 350°C. The mass transitions used to detect the fatty acids and the internal standards can be found in **Supplementary Table 3**. An 8-point calibration curve and three QC samples were used for method validation and fatty acid quantitation; all of which passed with an r^2 value of 0.99 or greater. The limit of detection and limit of quantitation for each HETE and EET was estimated from the calibration curves and shown in **Supplementary Table 4**. Data collection and peak area integration were performed using MassLynx and peak TargetLynx, respectively.

Homology modelling of CYP4 ancestors

Homology models were generated based on the crystal structure of CYP4B1 complexed with octane (PDB: 5T6Q) [70] using Modeller [115] in UCSF Chimera [116], and were visualised using Chimera. In the case of CYP4ABTXZ, CYP4A and CYP4BT, the predicted covalent heme linkage was simulated by substitution of the residue corresponding to Glu-310 in CYP4B1. The structures of the ancestors were also preduced using AlphaFold 2 as described previously [117]. AlphaFold 2 was used via the ColabFold v1.5.2-patch: AlphaFold2 web server, using MMseqs. Amino acid sequences of the ancestors were input as the query sequences and the

template_mode was set as pdb 100. All the other settings were kept as default. The top 5 ranked models were examined using Chimera.

Conflicts of interest statement

There are no conflicts of interest to declare.

Acknowledgments

This work was supported by Australian Research Council Discovery Project Grants DP120101772 and DP160100865 and by a CSIRO Future Science Platform Scholarship to KLH. Thanks are extended to Professors F. P. Guengerich and J. Capdevila (Vanderbilt University, Nashville, TN, USA) for donating some of the plasmids used in this study.

Author contributions

EMJG and KLH conceived the project with input from YG, CS and RAT. KLH, MBZ,YZ, SJS, SO, REST, SECP, JY, and DR performed the experiments and/or metabolite analysis. KLH, MBZ, YZ, SJS, SO and EMJG analysed the results. YG, SO, CS, RAT and MBZ contributed specialist expertise and resources. KLH, MBZ, RAT and EMJG wrote the paper with input from all other authors.

Data availability

The data for this paper are provided in the Supplementary information or at the following URL: https://www.scidb.cn/en/s/mMB7Rr.

References

1. M. Flieger, M. Kantorova, A. Prell, T. Rezanka and J. Votruba, *Folia Microbiol.*, 2003, **48**, 27-44.
2. S. Khanna and A. K. Srivastava, *Process Biochem.*, 2005, **40**, 607-619.
3. H.-J. Endres and A. Siebert-Raths, in *Engineering Biopolymers - Markets, Manufacturing, Properties and Applications*, eds. H.-J. Endres and A. Siebert-Raths, Hanser Publishers, Munich, 2011, ch. 4, pp. 71-113.
4. C. Liu, F. Liu, J. Cai, W. Xie, T. E. Long, S. R. Turner, A. Lyons and R. A. Gross, *Biomacromolecules*, 2011, **12**, 3291-3298.
5. M. Sokolsky-Papkov, A. Shikanov, A. Ezra, B. Vaisman and A. J. Domb, in *Polymer Degradation and Performance*, eds. M.C. Celina, J.S. Wiggins and N. C. Billingham, ACS Symposium Series, Vol. 1004, Washington, DC, 2009, ch.6, pp. 60-69.
6. J. P. Jain, M. Sokolsky, N. Kumar and A. J. Domb, *Polymer Rev.*, 2008, **48**, 156-191.
7. J. B. Johnston, H. Ouellet, L. M. Podust and P. R. Ortiz de Montellano, *Arch. Biochem. Biophys.*, 2011, **507**, 86-94.
8. F. P. Guengerich, *Chem. Res. Toxicol.*, 2001, **14**, 611-650.
9. A. E. C. M. Simpson, *Gen. Pharmacol.*, 1997, **28**, 351-359.
10. T. Aoyama, J. P. Hardwick, S. Imaoka, Y. Funae, H. V. Gelboin and F. J. Gonzalez, *J. Lipid Res.*, 1990, **31**, 1477-1482.
11. J. P. Hardwick, B. J. Song, E. Huberman and F. J. Gonzalez, *J. Biol. Chem.*, 1987, **262**, 801-810.
12. D. E. Williams, S. E. Hale, R. T. Okita and B. S. Masters, *J. Biol. Chem.*, 1984, **259**, 14600-14608.
13. J. P. Hardwick, *Biochem. Pharmacol.*, 2008, **75**, 2263-2275.

14. A. Kalsotra and H. W. Strobel, *Pharmacol. Therap.*, 2006, **112**, 589-611.

15. X. Wang, R. Ullrich, M. Hofrichter and J. T. Groves, *Proc. Natl. Acad. Sci. U.S.A.*, 2015, **112**, 3686-3691.

16. F. Fiorentini, A.-M. Hatzl, S. Schmidt, S. Savino, A. Glieder and A. Mattevi, *Biochemistry*, 2018, **57**, 6701-6714.

17. M.-H. Hsu, Ü. Savas, K. J. Griffin and E. F. Johnson, *Drug Metab. Rev.*, 2007, **39**, 515-538.

18. M. B. Fisher, Y. M. Zheng and A. E. Rettie, *Biochem. Biophys. Res. Comm.*, 1998, **248**, 352-355.

19. A. Rettie and E. Kelly, in *Cytochromes P450: Role in the Metabolism and Toxicity of Drugs and Other Xenobiotics*, ed. C. Ioannides, Royal Society of Chemistry, Cambridge, UK, 2008, ch. 12, pp. 348-414.

20. B. R. Baer and A. E. Rettie, *Drug Metab. Rev.*, 2006, **38**, 451-476.

21. L. J. Roman, C. N. A. Palmer, J. E. Clark, A. S. Muerhoff, K. J. Griffin, E. F. Johnson and B. S. S. Masters, *Arch. Biochem. Biophys.*, 1993, **307**, 57-65.

22. C. N. Palmer, T. H. Richardson, K. J. Griffin, M. H. Hsu, A. S. Muerhoff, J. E. Clark and E. F. Johnson, *Biochim. Biophys. Acta*, 1993, **1172**, 161-166.

23. P. K. Powell, I. Wolf, R. Jin and J. M. Lasker, *J. Pharmacol. Exp. Ther.*, 1998, **285**, 1327-1336.

24. Dominik N. Muller, C. Schmidt, E. Barbosa-Sicard, M. Wellner, V. Gross, H. Hercule, M. Markovic, H. Honeck, Friedrich C. Luft and W.-H. Schunck, *Biochem. J.*, 2007, **403**, 109-118.

25. U. Hoch, Z. Zhang, D. L. Kroetz and P. R. Ortiz de Montellano, *Arch. Biochem. Biophys.*, 2000, **373**, 63-71.

26. C. Helvig, E. Dishman and J. H. Capdevila, *Biochemistry*, 1998, **37**, 12546-12558.

27. K. M. Lukaszewicz and J. H. Lombard, *Clin. Sci. (Lond.)*, 2013, **124**, 695-700.

28. R. J. Roman, *Physiol. Rev.*, 2002, **82**, 131-185.

29. K. Z. Edson and A. E. Rettie, *Curr. Top. Med. Chem.*, 2013, **13**, 1429-1440.

30. F. Xu, J. R. Falck, P. R. Ortiz de Montellano and D. L. Kroetz, *J. Pharmacol. Exp. Ther.*, 2004, **308**, 887-895.

31. M. Fer, L. Corcos, Y. Dreano, E. Plee-Gautier, J. P. Salaun, F. Berthou and Y. Amet, *J. Lipid Res.*, 2008, **49**, 2379-2389.

32. J. Bylund, M. Hidestrand, M. Ingelman-Sundberg and E. H. Oliw, *J. Biol. Chem.*, 2000, **275**, 21844-21849.

33. J. Bylund, M. Bylund and E. H. Oliw, *Biochem. Biophys. Res. Comm.*, 2001, **280**, 892-897.

34. J. Bylund, A. G. Harder, K. G. Maier, R. J. Roman and D. R. Harder, *Arch. Biochem. Biophys.*, 2003, **412**, 34-41.

35. Y. Kikuta, E. Kusunose and M. Kusunose, *J. Biochem.*, 2000, **127**, 1047-1052.

36. Y. Kikuta, E. Kusunose, M. Ito and M. Kusunose, *Arch. Biochem. Biophys.*, 1999, **369**, 193-196.

37. Y. Liu, J. Wang, Y. Liu, H. Zhang, M. Xu and J. Dai, *Comp. Biochem. Physiol. C Toxicol. Pharmacol.*, 2009, **150**, 57-64.

38. C. Sabourault, J. B. Bergé, M. Lafaurie, J. P. Girard and M. Amichot, *Biochem. Biophys. Res. Comm.*, 1998, **251**, 213-219.

39. M. Nakano, E. J. Kelly and A. E. Rettie, *Drug Metab. Dispos.*, 2009, **37**, 2119-2122.

40. K. Stark, M. Dostalek and F. P. Guengerich, *FEBS J.*, 2008, **275**, 3706-3717.

41. A. Zöllner, C.-A. Dragan, D. Pistorius, R. Müller, B. Bode Helge, T. Peters Frank, H. Maurer Hans and M. Bureik, *Biol. Chem.*, 2009, **390**, 313.

42. M. G. McDonald, S. Ray, C. J. Amorosi, K. A. Sitko, J. P. Kowalski, L. Paco, A. Nath, B. M. Gallis, R. A. Totah, M. J. Dunham, D. M. Fowler and A. E. Rettie, *Drug Metab. Dispos.*, 2017, **45**, 1364-1371.

43. S. Kumar, L. Sun, H. Liu, B. K. Muralidhara and J. R. Halpert, *Protein Eng. Des. Sel.*, 2006, **19**, 547-554.

44. O. Salazar, P. C. Cirino and F. H. Arnold, *ChemBioChem*, 2003, **4**, 891-893.

45. P. A. Romero, A. Krause and F. H. Arnold, *Proc. Natl. Acad. Sci. U.S.A.*, 2013, **110**, E193-E201.

46. I. I. Karuzina and A. I. Archakov, *Free Rad. Biol. Med.*, 1994, **17**, 557-567.

47. S. Kumar, *Expert Opin. Drug Metab. Toxicol.*, 2010, **6**, 115-131.

48. Y. Gumulya, J.-M. Baek, S.-J. Wun, R. E. S. Thomson, K. L. Harris, D. J. B. Hunter, J. B. Y. H. Behrendorff, J. Kulig, S. Zheng, X. Wu, B. Wu, J. E. Stok, J. J. De Voss, G. Schenk, U. Jurva, S. Andersson, E. M. Isin, M. Bodén, L. Guddat and E. M. J. Gillam, *Nat. Catal.*, 2018, **1**, 878-888.

49. Y. Gumulya, W. Huang, S. A. D'Cunha, K. E. Richards, R. E. S. Thomson, D. J. B. Hunter, J.-M. Baek, K. L. Harris, M. Boden, J. J. De Voss, M. A. Hayes, E. M. Isin, S. Andersson, U. Jurva and E. M. J. Gillam, *ChemCatChem*, 2019, **11**, 841-850.

50. R. E. S. Thomson, *Structural and functional characterisation of ancestral cytochromes P450 from family 2 in tetrapods*, Ph.D. Thesis, The Univ. of Qld, 2021.

51. D. L. Swofford and W. Maddison, in *Systematics, Historical Ecology, and North American Freshwater Fishes*, ed. R. L. Mayden, Stanford University Press, Stanford, CA, 1992, ch. 5, pp. 186-223.

52. J. Ma, Encyclopedia of Life Sciences, 2008, Wiley, DOI: 10.1002/9780470015902.a0020736, 1-6.

53. V. A. Risso, J. A. Gavira and J. M. Sanchez-Ruiz, *Environ.Microbiol.*, 2014, **16**, 1485-1489.

54. N. L. Kirischian and J. Y. Wilson, *Mol. Phylogenet. Evol.*, 2012, **62**, 458-471.

55. Y. Fujita, H. Ohi, N. Murayama, K. Saguchi and S. Higuchi, *Comp. Biochem. Physiol. B*, 2004, **138**, 129-136.

56. J. H. Thomas, *PloS Genetics*, 2007, **3**, 720-728.

57. D. R. Nelson, *Arch. Biochem. Biophys.*, 2003, **409**, 18-24.

58. M. W. Hahn and L. Nakhleh, *Evolution*, 2016, **70**, 7-17.

59. Y. Xiao, *Functional annotation of orphan human P450 enzymes: heterologous expression and substrate searches by metabolomic approaches*, Ph.D. Thesis, Vanderbilt University, 2014.

60. K. L. Harris, R. E. S. Thomson, Y. Gumulya, G. Foley, S. E. Carrera-Pacheco, P. Syed, T. Janosik, A.-S. Sandinge, S. Andersson, U. Jurva, M. Bodén and E. M. J. Gillam, *Mol. Biol. Evol.*, 2022, **39**, msac116.

61. D. L. Trudeau, M. Kaltenbach and D. S. Tawfik, *Mol. Biol. Evol.*, 2016, **33**, 2633-2641.

62. J. K. Hobbs, C. Shepherd, D. J. Saul, N. J. Demetras, S. Haaning, C. R. Monk, R. M. Daniel and V. L. Arcus, *Mol. Biol. Evol.*, 2011, **29**, 825-835.

63. M. D. Giulio, *J. Theor. Biol.*, 2003, **221**, 425-436.

64. V. A. Risso, F. Manssour-Triedo, A. Delgado-Delgado, R. Arco, A. Barroso-delJesus, A. Ingles-Prieto, R. Godoy-Ruiz, J. A. Gavira, E. A. Gaucher, B. Ibarra-Molero and J. M. Sanchez-Ruiz, *Mol. Biol. Evol.*, 2014, **32**, 440-455.

65. P. D. Williams, D. D. Pollock, B. P. Blackburne and R. A. Goldstein, *PLoS Comput. Biol.*, 2006, **2**, e69.

66. V. A. Risso, J. M. Sanchez-Ruiz and S. B. Ozkan, *Curr. Opin. Struct. Biol.*, 2018, **51**, 106-115.

67. Y. M. Zheng, M. B. Fisher, N. Yokotani, Y. Fujii-Kuriyama and A. E. Rettie, *Biochemistry*, 1998, **37**, 12847-12851.

68. P. K. Powell, I. Wolf and J. M. Lasker, *Arch. Biochem. Biophys.*, 1996, **335**, 219-226.

69. I. Kosuke, K. Emi and K. Masamichi, *BBA - Lipid. Lipid Met.*, 1969, **176**, 704-712.

70. M. H. Hsu, B. R. Baer, A. E. Rettie and E. F. Johnson, *J. Biol. Chem.*, 2017, **292**, 5610-5621.

71. S. N. Ngo, R. A. McKinnon and I. Stupans, *Gene*, 2006, **376**, 123-132.

72. C. N. A. Palmer, T. H. Richardson, K. J. Griffin, M. H. Hsu, A. S. Muerhoff, J. E. Clark and E. F. Johnson, *Biochim.Biophys.Acta*, 1993, **1172**, 161-166.

73. G. K. Jennings, M. H. Hsu, L. S. Shock, E. F. Johnson and J. C. Hackett, *J. Biol. Chem.*, 2018, **293**, 11433-11446.

74. P. R. Ortiz de Montellano, *Drug Metab. Rev.*, 2008, **40**, 405-426.

75. U. Hoch and P. R. Ortiz de Montellano, *J. Biol. Chem.*, 2001, **276**, 11339-11346.

76. B. R. Baer, J. T. Schuman, A. P. Campbell, M. J. Cheesman, M. Nakano, N. Moguilevsky, K. L. Kunze and A. E. Rettie, *Biochemistry*, 2005, **44**, 13914-13920.

77. K. R. Henne, K. L. Kunze, Y.-M. Zheng, P. Christmas, R. J. Soberman and A. E. Rettie, *Biochemistry*, 2001, **40**, 12925-12931.

78. P. Grunwald, in *Biocatalysis: Biochemical Fundamentals and Applications*, ed. P. Grunwald, Imperial College Press, United Kingdom 2nd edn., 2017, ch. 15, pp. 695-1036.

79. S. Jemli, D. Ayadi-Zouari, H. B. Hlima and S. Bejar, *Crit. Rev. Biotechnol.*, 2016, **36**, 246-258.

80. E. G. Hrycay and P. J. O'Brien, *Arch. Biochem. Biophys.*, 1971, **147**, 14-27.

81. E. G. Hrycay, P. J. O'Brien, J. E. Van Lier and G. Kan, *Arch. Biochem. Biophys.*, 1972, **153**, 495-501.

82. A. D. Rahimtula and P. J. O'Brien, *Biochem. Biophys. Res. Comm.*, 1975, **62**, 268-275.

83. F. F. Kadlubar, K. C. Morton and D. M. Ziegler, *Biochem. Biophys. Res. Comm.*, 1973, **54**, 1255-1261.

84. J.-Å. Gustafsson and J. Bergman, *FEBS Lett.*, 1976, **70**, 276-280.

85. Z.-F. Wu, R. K. Bajpai and W. Yan, *Biocat. Biotrans.*, 2008, **26**, 444-449.

86. R. Herrington, K. F. Hock, F. Moore and W. L. Casati, in *Flexible Polyurethane Foams*, eds. R. Herrington and K. F. Hock, Dow Chemical Company, Midland, 2nd edn., 1997, ch. 2, pp. 2.1-2.35.

87. T. M. Kuo, T. Kaneshiro and C. T. Hou, in *Lipid Biotechnology*, eds. T. M. Kuo and H. Gardner, CRC Press, Boca Raton, FL, 1st edn., 2002, ch. 31, pp. 605-627.

88. A. Nuur Annisa and W. Widayat, A review of bio-lubricant production from vegetable oils using esterification transesterification process, The 24th Regional Symposium on Chemical Engineering (RSCE 2017), Semarang, Indonesia, 2017.

89. United States Department of Agriculture, Agricultural Research Service, Nutrient Data Laboratory, Basic Report: 04572, Oil, nutmeg butter, Beltsville MD 2018.

90. D. Kim, G.-S. Cha, L. D. Nagy, C.-H. Yun and F. P. Guengerich, *Biochemistry*, 2014, **53**, 6161-6172.

91. A. Parikh, E. M. J. Gillam and F. P. Guengerich, *Nat. Biotech.*, 1997, **15**, 784-788.

92. M. J. Cheesman, B. R. Baer, Y. M. Zheng, E. M. J. Gillam and A. E. Rettie, *Arch. Biochem. Biophys.*, 2003, **416**, 17-24.

93. K. Nishihara, M. Kanemori, M. Kitagawa, H. Yanagi and T. Yura, *Appl. Environ. Microbiol.*, 1998, **64**, 1694-1699.

94. S. F. Altschul, W. Gish, W. Miller, E. W. Myers and D. J. Lipman, *J. Mol. Biol.*, 1990, **215**, 403-410.

95. W. Z. Li and A. Godzik, *Bioinformatics*, 2006, **22**, 1658-1659.

96. H. Ashkenazy, O. Penn, A. Doron-Faigenboim, O. Cohen, G. Cannarozzi, O. Zomer and T. Pupko, *Nucl. Acids Res.*, 2012, **40**, W580-W584.

97. K. Katoh and D. M. Standley, *Mol. Biol. Evol.*, 2013, **30**, 772-780.

98. K. Tamura, G. Stecher, D. Peterson, A. Filipski and S. Kumar, *Mol. Biol. Evol.*, 2013, **30**, 2725-2729.

99. A. Bellamine, Y. Wang, M. R. Waterman, J. V. Gainer Iii, E. P. Dawson, N. J. Brown and J. H. Capdevila, *Arch. Biochem. Biophys.*, 2003, **409**, 221-227.

100. H. A. N. Songhee, E. U. N. Chang-Yong, H. A. N. Jung-Soo, C. Young-Jin, K. I. M. Dong-Hyun, Y. U. N. Chul-Ho and K. I. M. Donghak, *Biomol. Ther.*, 2009, **17**, 156-161.

101. R. Gasser and R. M. Philpot, *Mol.Pharmacol.*, 1989, **35**, 617-625.

102. M. J. Cheesman, *Arch. Biochem. Biophys.*, 2003, **416**, 17-24.

103. R. L. Strausberg, E. A. Feingold, L. H. Grouse, J. G. Derge, R. D. Klausner, F. S. Collins, L. Wagner, C. M. Shenmen, G. D. Schuler, S. F. Altschul, B. Zeeberg, K. H. Buetow, C. F. Schaefer, N. K. Bhat, R. F. Hopkins, H. Jordan, T. Moore, S. I. Max, J. Wang, F. Hsieh, L. Diatchenko, K. Marusina, A. A. Farmer, G. M. Rubin, L. Hong, M. Stapleton, M. B. Soares, M. F. Bonaldo, T. L. Casavant, T. E. Scheetz, M. J. Brownstein, T. B. Usdin, S. Toshiyuki, P. Carninci, C. Prange, S. S. Raha, N. A. Loquellano, G. J. Peters, R. D. Abramson, S. J. Mullahy, S. A. Bosak, P. J. McEwan, K. J. McKernan, J. A. Malek, P. H. Gunaratne, S. Richards, K. C. Worley, S. Hale, A. M. Garcia, L. J. Gay, S. W. Hulyk, D. K. Villalon, D. M. Muzny, E. J. Sodergren, X. Lu, R. A. Gibbs, J. Fahey, E. Helton, M. Ketteman, A. Madan, S. Rodrigues, A.

Sanchez, M. Whiting, A. Madan, A. C. Young, Y. Shevchenko, G. G. Bouffard, R. W. Blakesley, J. W. Touchman, E. D. Green, M. C. Dickson, A. C. Rodriguez, J. Grimwood, J. Schmutz, R. M. Myers, Y. S. N. Butterfield, M. I. Krzywinski, U. Skalska, D. E. Smailus, A. Schnerch, J. E. Schein, S. J. M. Jones, M. A. Marra and T. Mammalian Gene Collection Program, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 16899-16903.

104. P. Gaspar, G. Moura, M. A. Santos and J. L. Oliveira, *Nucl. Acids Res.*, 2013, **41**, e73.

105. J. N. Zadeh, C. D. Steenberg, J. S. Bois, B. R. Wolfe, M. B. Pierce, A. R. Khan, R. M. Dirks and N. A. Pierce, *J. Comput. Chem.*, 2011, **32**, 170-173.

106. D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison and H. O. Smith, *Nat. Methods*, 2009, **6**, 343-345.

107. L. M. Notley, K. H. Crewe, P. J. Taylor, M. S. Lennard and E. M. J. Gillam, *Chem. Res. Toxicol.*, 2005, **18**, 1611-1618.

108. H. G. Park, Y. R. Lim, S. Han, D. Jeong and D. Kim, *J. Microbiol. Biotechnol.*, 2017, **27**, 983-989.

109. W. A. Johnston, W. Huang, M. A. Hayes, J. J. De Voss and E. M. J. Gillam, *J. Biomol. Screen.*, 2008, **13**, 135-141.

110. F. P. Guengerich, in *Principles and Methods in Toxicology*, ed. A. W. Hayes, Raven Press, New York, 5th edn., 2007, ch. 40, pp. 1973-2040.

111. E. M. Gillam, T. Baba, B. R. Kim, S. Ohmori and F. P. Guengerich, *Arch. Biochem. Biophys.*, 1993, **305**, 123-131.

112. M. Sobol, D. Ma, J. Unch and J. J. Cali, 2008, *Promega Resources*, Promega Corporation, Madison WI.

113. T. Aliwarga, B. S. Raccor, R. N. Lemaitre, N. Sotoodehnia, S. A. Gharib, L. B. Xu and R. A. Totah, *Free Rad. Biol. Med.*, 2017, **112**, 131-140.

114. M. Zeigler, D. Whittington, N. Sotoodehnia, R. N. Lemaitre and R. A. Totah, *Chem. Phys. Lipids*, 2018, **216**, 162-170.

115. A. Sali and T. L. Blundell, *J. Mol. Biol.*, 1993, **234**, 779-815.

116. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605-1612.

117. Y. D. Ivanov, A. Taldaev, A. V. Lisitsa, E. A. Ponomarenko and A. I. Archakov, *Journal*, 2022, **27**, 1386.

118. B. Webb and A. Sali, *Curr. Protoc. Bioinform.*, 2016, **54**, 5.6.1-5.6.37.

119. P. Gouet and X. Robert, *Nucl. Acids Res.*, 2014, **42**, W320-W324.

Figures and Figure Legends

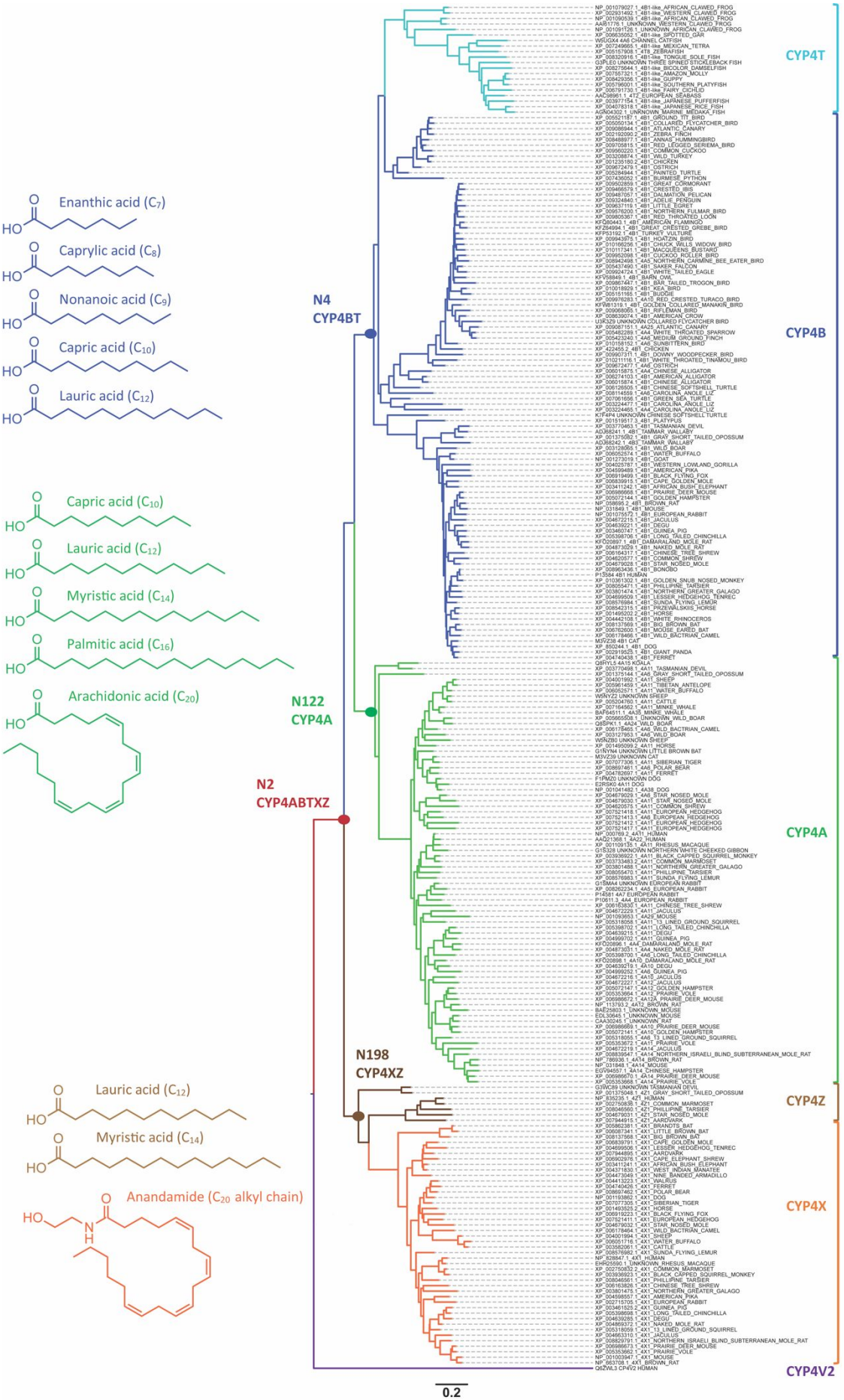

Enanthic acid (C7)
Caprylic acid (C8)
Nonanoic acid (C9)
Capric acid (C10)
Lauric acid (C12)
Capric acid (C10)
Lauric acid (C12)
Myristic acid (C14)
Palmitic acid (C16)
Arachidonic acid (C20)
Lauric acid (C12)
Myristic acid (C14)
Anandamide (C20 alkyl chain)
N4
CYP4BT
N122
CYP4A
N2
CYP4ABTXZ
N198
CYP4XZ
CYP4T
CYP4B
CYP4A
CYP4Z
CYP4X
CYP4V2
0.2

Figure 1. Phylogenetic and functional comparison of the CYP4 clade

A phylogenetic tree showing the evolutionary relationship between the sequences used to reconstruct CYP4ABTXZ and the intermediate ancestors. The tree is coloured to represent the individual subfamilies of the reconstruction, with each node selected for ancestral reconstruction labelled and indicated by a circle of the corresponding colour. The carbon chain length and structure of the known FA substrate preferences for the characterised extant forms from each subfamily are shown on the left. Structures are coloured to indicate which subfamilies have been identified to act on that substrate.

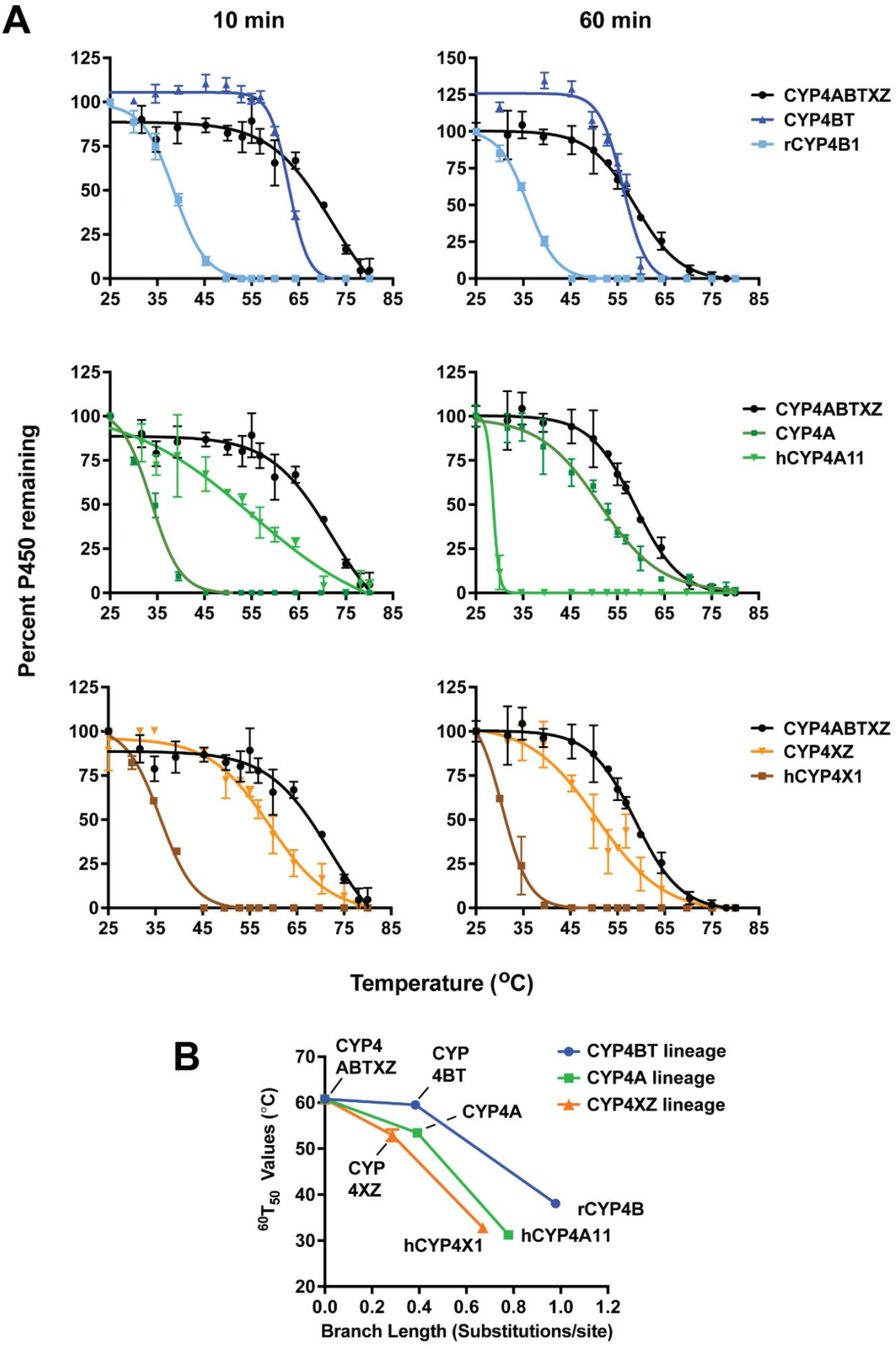

A
10 min
60 min
CYP4ABTXZ
CYP4BT
rCYP4B1
CYP4ABTXZ
CYP4A
hCYP4A11
CYP4ABTXZ
CYP4XZ
hCYP4X1
Percent P450 remaining
Temperature (°C)
B
CYP4 ABTXZ
CYP 4BT
CYP4A
CYP 4XZ
rCYP4B
hCYP4A11
hCYP4X1
CYP4BT lineage
CYP4A lineage
CYP4XZ lineage
60T50 Values (°C)
Branch Length (Substitutions/site)

Figure 2. Thermostability of ancestral and extant forms

The thermostability of each form was determined by quantifying the residual folded P450 using Fe(II).CO *vs.* Fe(II) difference spectroscopy, after heating for 10 (A, left panels) or 60 min (A, right panels) as indicated. The $^{10}T_{50}$ or $^{60}T_{50}$ was determined by fitting the data to a sigmoidal melting equation, which was used to interpolate the temperature at which 50% of the protein remained folding after heating. A shows the thermostability assays for the CYP4BT (top panels), CYP4A (middle), and CYP4XZ (bottom) lineages. B shows the correlation between determined $^{60}T_{50}$ values and branch length, a measure of evolutionary distance. A trend was observed towards greater stability in more ancient (shorter branch length) ancestors.

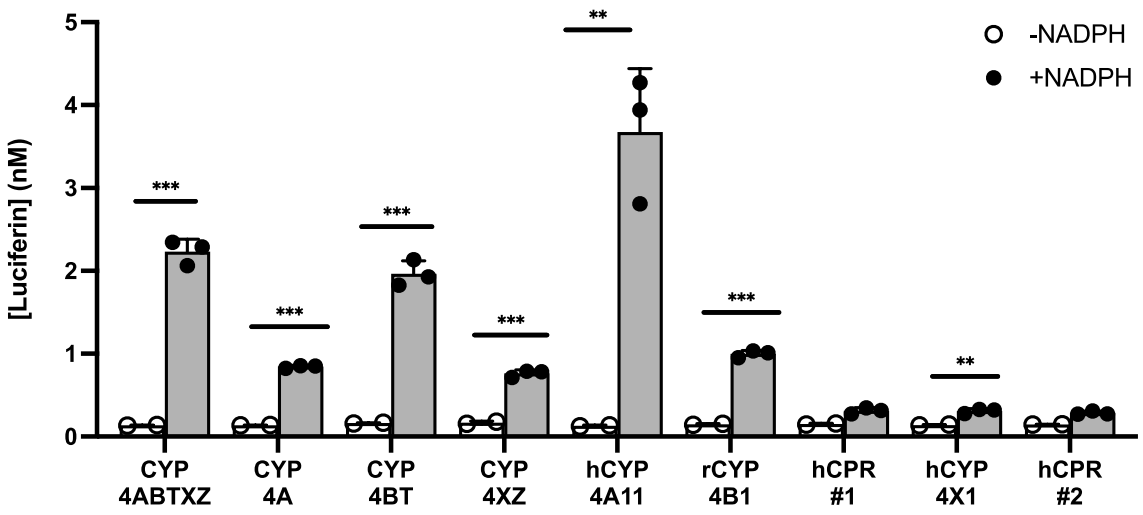


Figure 3: Activity of ancestral and extant CYP4 forms towards Luciferin-MultiCYP dealkylation

The activity of ancestral and extant CYP4 forms was tested towards the P450 marker substrate, Luciferin-MultiCYP. Reactions were carried out with 50 µM substrate and 10 nM P450 using membranes from bacteria coexpressing each form indicated with hCPR, for 60 minutes at 37 °C. Total hCPR concentrations were standardised by addition of membranes from cells expressing hCPR alone. Two different hCPR-only controls were used to match the hCPR concentrations in P450-containing incubations; hCPR#2 for hCYP4X1 (5.6 nM hCPR) and hCPR#1 (10.2 nM hCPR) for all other forms. Data represent the mean +/- SD of three replicates for + NADPH reactions and two replicates for -NADPH controls. Asterisks indicate that product formation was significantly higher than for the respective NADPH-deficient control (one-tailed, unpaired Student's *t*-test): *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$. All P450-containing, NADPH-replete reactions showed significantly higher product formation than the corresponding NADPH-replete hCPR controls ($p < 0.001$) except hCYP4X1 which was not significantly different to the hCPR control.

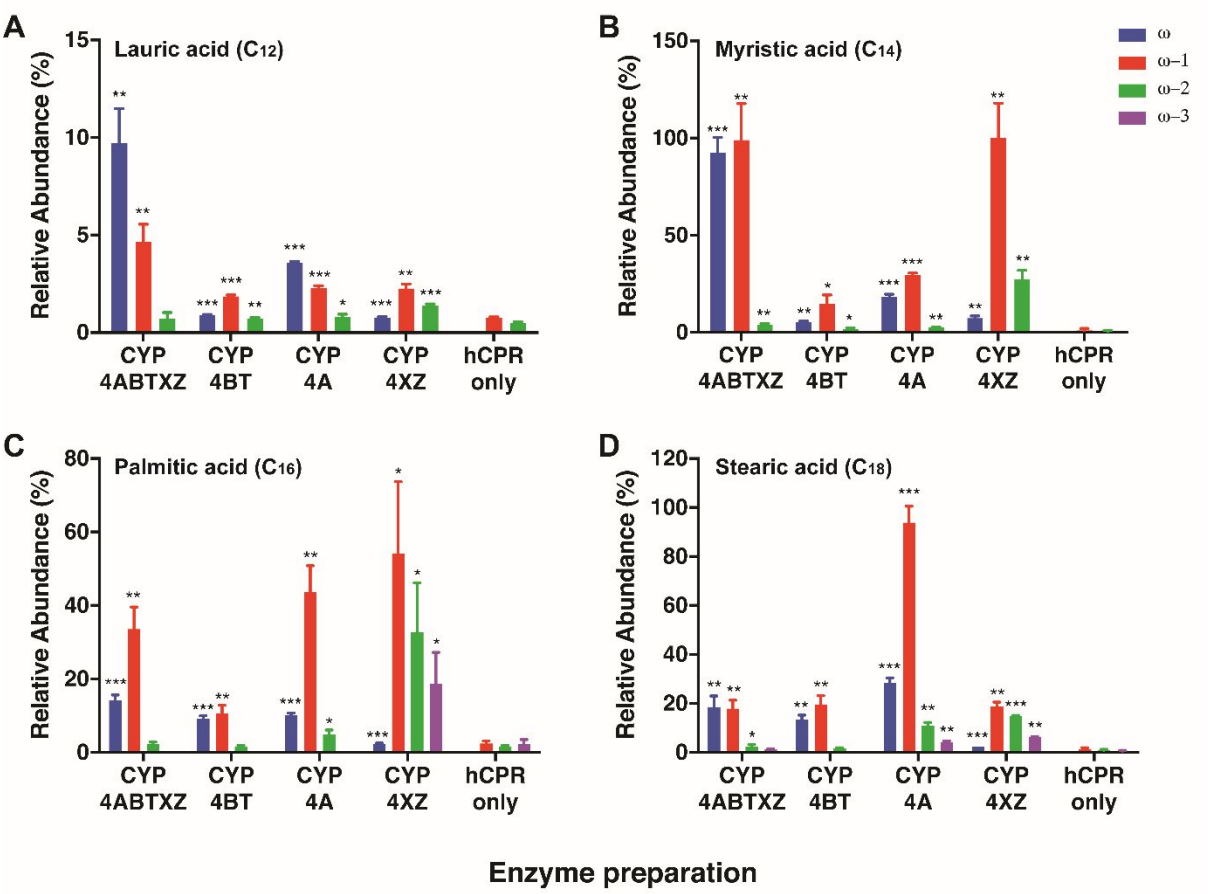
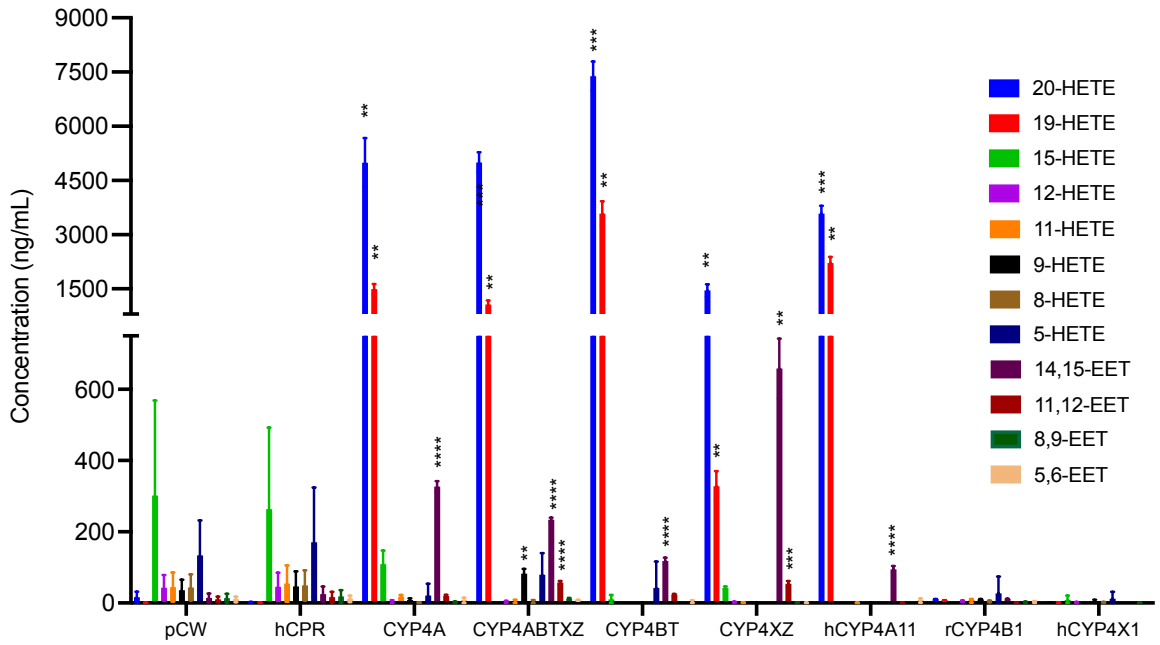


Figure 4. FA hydroxylase activity of the ancestral CYP4 forms

The relative activities and regioselectivities of the four ancestors towards lauric acid (**A**, C_{12}), myristic acid (**B**, C_{14}), palmitic acid (**C**, C_{16}) and stearic acid (**D**, C_{18}) were determined in reconstitutions of purified P450s with hCPR. Incubations contained 0.2 µM P450, 0.4 µM CPR, 150 µM DLPC and 100 µM substrate in a 500 µl volume, and were carried out for 30 mins at 37 ºC then analysed by GC-MS. The negative control (hCPR) contained only the redox partner and no P450. Metabolite production was normalised to the maximal production of the C14 w-1 hydroxylation product by CYP4XZ. Data represent the mean +/- SD of three replicates. Asterisks indicate statistical significance compared to the negative control (incubations containing hCPR but no P450) for each activity: *, $p < 0.05$; **, $p \leq 0.01$: ***, $p \leq 0.001$ (one-tailed, unpaired Student's *t*-test).

A


Concentration (ng/mL)
9000
7500
6000
4500
3000
1500
600
400
200
0
pCW
hCPR
CYP4A
CYP4ABTXZ
CYP4BT
CYP4XZ
hCYP4A11
rCYP4B1
hCYP4X1
20-HETE
19-HETE
15-HETE
12-HETE
11-HETE
9-HETE
8-HETE
5-HETE
14,15-EET
11,12-EET
8,9-EET
5,6-EET

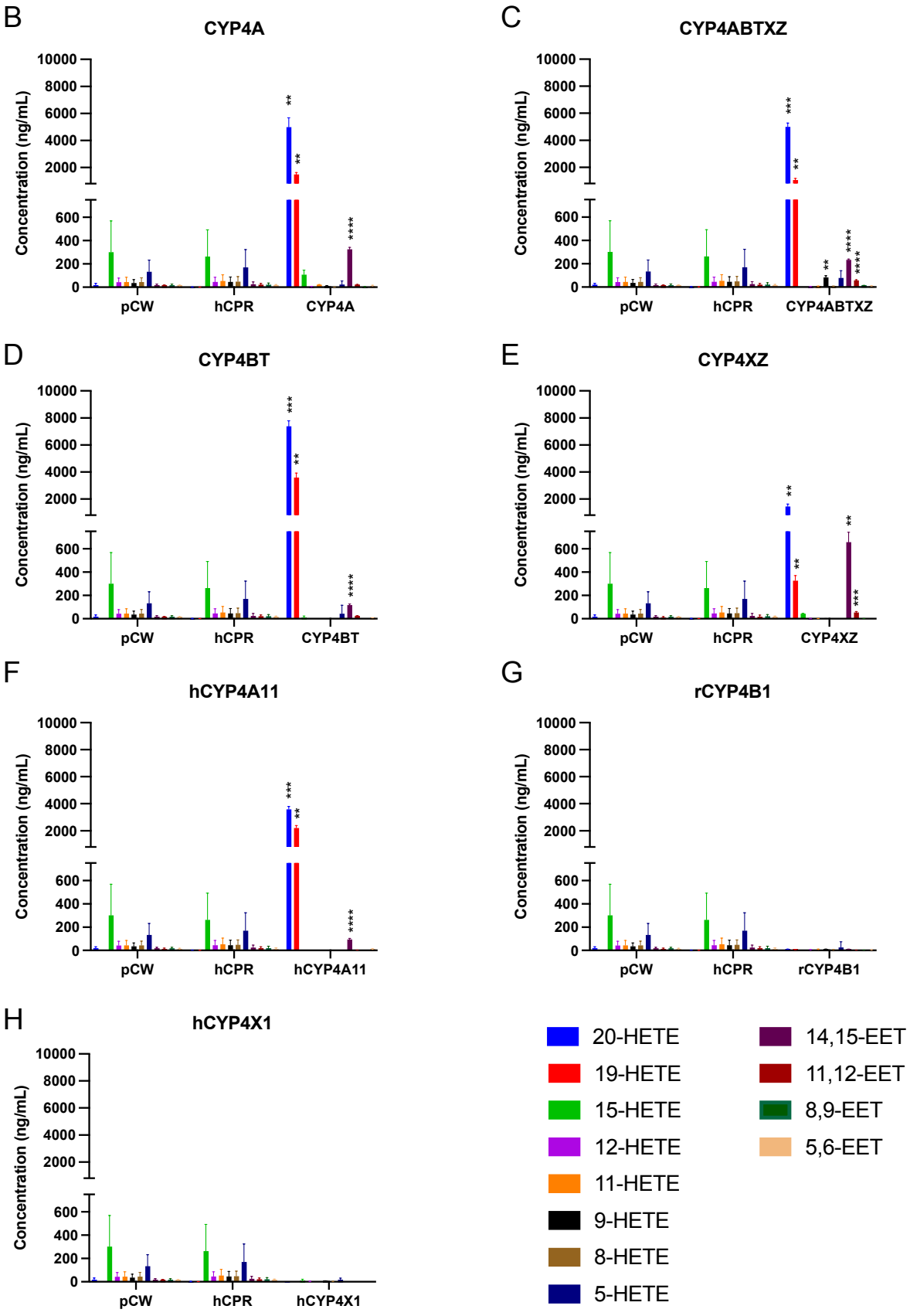

B
CYP4A
C
CYP4ABTXZ
D
CYP4BT
E
CYP4XZ
F
hCYP4A11
G
rCYP4B1
H
hCYP4X1
Concentration (ng/mL)
10000
8000
6000
4000
2000
600
400
200
0
pCW
hCPR
CYP4A
CYP4ABTXZ
CYP4BT
CYP4XZ
hCYP4A11
rCYP4B1
hCYP4X1
20-HETE
19-HETE
15-HETE
12-HETE
11-HETE
9-HETE
8-HETE
5-HETE
14,15-EET
11,12-EET
8,9-EET
5,6-EET

Figure 5. AA metabolism by CYP4 extant and ancestral forms

Incubations of the ancestral and extant CYP4s with AA were carried out and metabolites were analyzed with LC-MS (A). The data presented are the concentrations of EETs and HETEs for CYP4A (B), CYP4ABTXZ (C), CYP4BT (D), CYP4XZ (E), hCYP4A11 (F), rCYP4B1 (G), and hCYP4X1 (H). Asterisks indicate statistical significance compared to the pCW control bacteria lacking P450 expression for each metabolite formed using an one-tailed, unpaired Student's T-test (*, $p < 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$; ****, $p \leq 0.0001$).

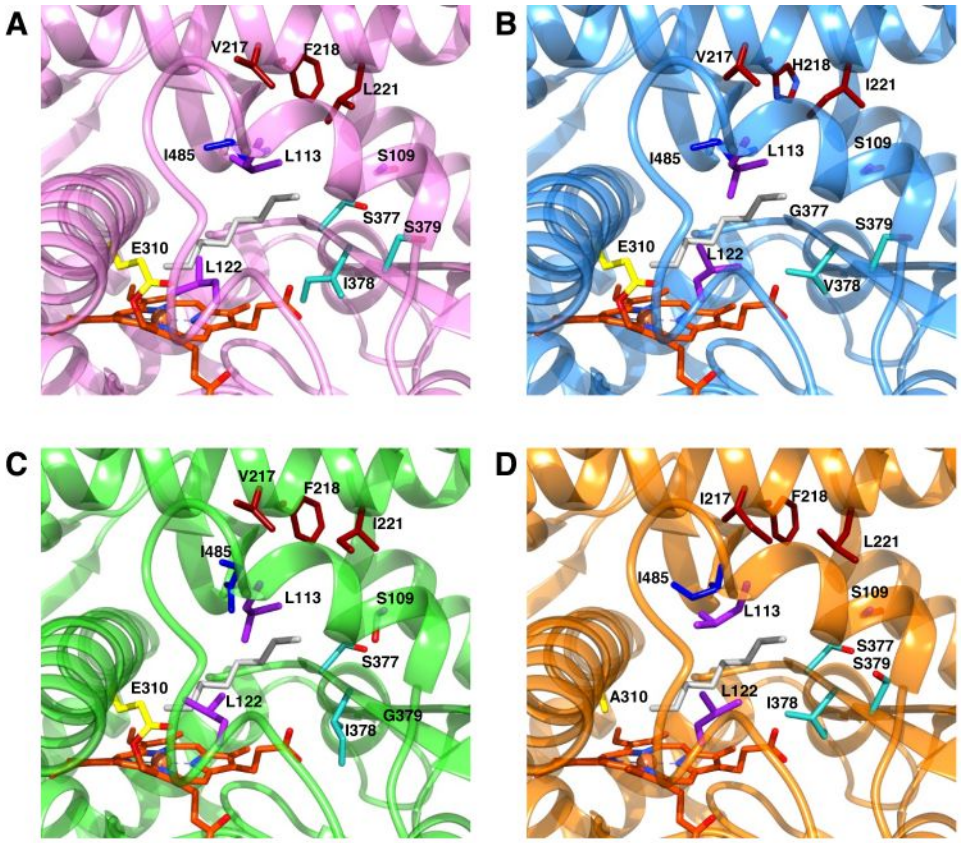


Figure 6 : Homology models of CYP4 ancestor substrate binding pockets

The active site cavities of CYP4ABTXZ (**A**) CYP4BT (**B**), CYP4A (**C**), and CYP4XZ (**D**) were modelled on the rCYP4B1 crystal structure in complex with octanoate (PDB: 5T6Q). Residues potentially responsible for changes in substrate and FA hydroxylation regioselectivity are shown as coloured sticks with positions 109, 113 and 122 in purple; 217, 218 and 221 in maroon; 310 in yellow; 377-379 in turquoise; and 485 in blue. The heme is shown as dark orange sticks and octanoate as light grey sticks. Homology models were created in Chimera[116] using Modeller[118].

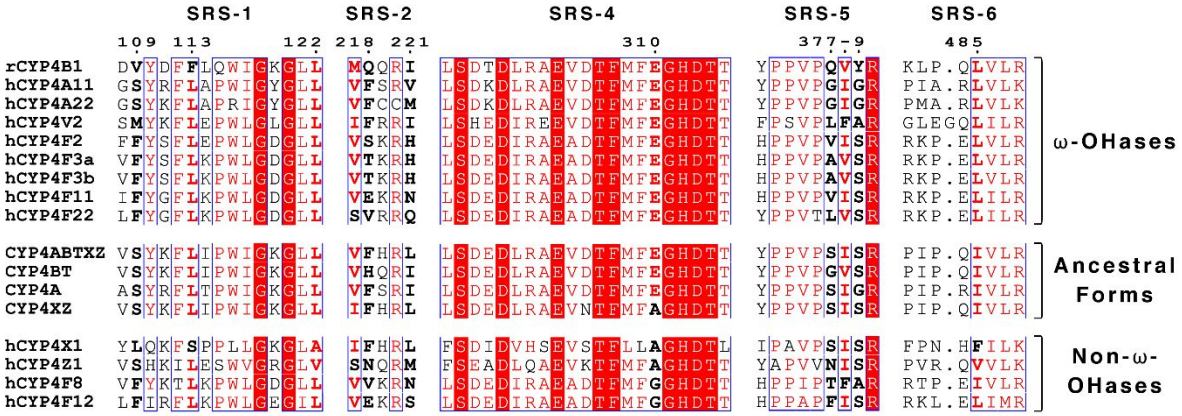


Figure 7. Key residues proposed to contribute to differences in FA hydroxylase activity between ancestral and extant forms

Positions that may contribute to the differences in activity observed between ancestral and extant forms are shown in bold [70]. Residues are coloured by conservation as calculated and visualised by ENDscript 3.0 [119]. Red boxes indicate identity, red characters indicate similarity, and blue frames surround regions of homology.

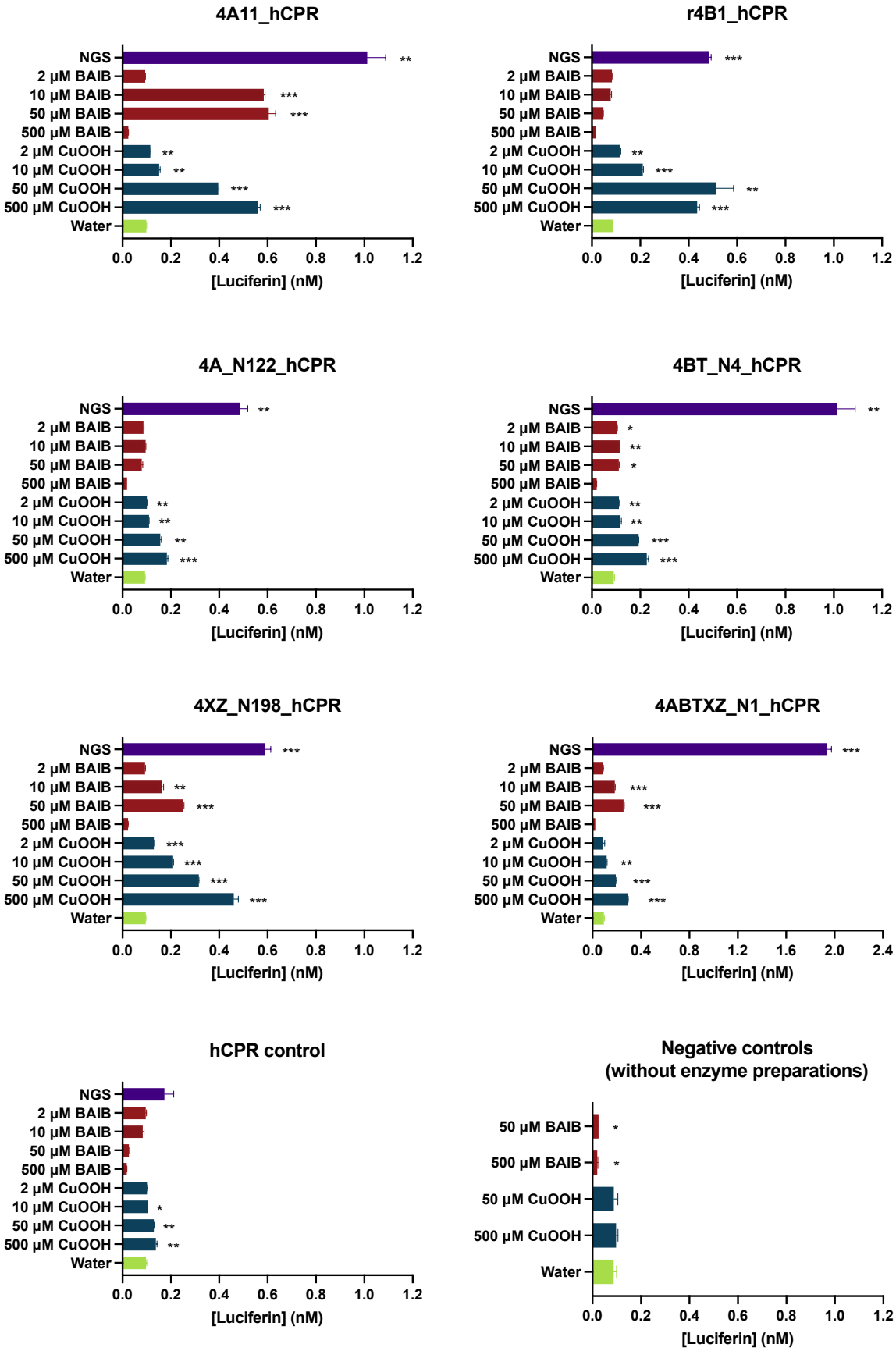

4A11_hCPR
NGS
2 μM BAIB
10 μM BAIB
50 μM BAIB
500 μM BAIB
2 μM CuOOH
10 μM CuOOH
50 μM CuOOH
500 μM CuOOH
Water
0.0 0.2 0.4 0.6 0.8 1.0 1.2
[Luciferin] (nM)
r4B1_hCPR
0.0 0.2 0.4 0.6 0.8 1.0 1.2
[Luciferin] (nM)
4A_N122_hCPR
0.0 0.2 0.4 0.6 0.8 1.0 1.2
[Luciferin] (nM)
4BT_N4_hCPR
0.0 0.2 0.4 0.6 0.8 1.0 1.2
[Luciferin] (nM)
4XZ_N198_hCPR
0.0 0.2 0.4 0.6 0.8 1.0 1.2
[Luciferin] (nM)
4ABTXZ_N1_hCPR
0.0 0.4 0.8 1.2 1.6 2.0 2.4
[Luciferin] (nM)
hCPR control
0.0 0.2 0.4 0.6 0.8 1.0 1.2
[Luciferin] (nM)
Negative controls
(without enzyme preparations)
0.0 0.2 0.4 0.6 0.8 1.0 1.2
[Luciferin] (nM)

Figure 8. Assessment of the ability of O_2 surrogates to support the activity of CYP4 forms

The relative activities of the four ancestors and two extant forms towards Luciferin MultiCYP *O*-demethylation supported by BAIB or CuOOH were determined using bacterial membranes containing P450s co-expressed with hCPR. Reactions were carried out with 50 μM substrate and 10 nM P450 using membranes from bacteria coexpressing each form indicated with hCPR. Total hCPR and protein concentrations in reactions were standardised at 5.6 nM and 60 μg/mL respectively by addition of membranes from bacteria expressing the hCPR or no recombinant proteins. Negative controls lacking enzyme preparations were analysed in parallel to detect product formation due to the oxygen surrogates alone. NGS denotes positive controls supported by an NADPH-generating system. Reactions were initiated by the addition of the NGS, oxygen surrogate or an equivalent volume of water as indicated and allowed to proceed for 60 minutes at 37 °C before being quenched by the addition of an equal volume of Luciferin Detection Reagent. Data represent the mean +/- SD of two replicates. Asterisks indicate statistical significance compared to the water control for each form of CYP4 if the means were higher than the water control: *, $p < 0.05$; **, $p \leq 0.01$: ***, $p \leq 0.001$ (one-tailed, unpaired Student's *t*-test). For the negative controls without membrane preps, asterisks indicate statistical significance compared to the water sample: *, $p < 0.05$; **, $p \leq 0.01$: ***, $p \leq 0.001$ (two-tailed, unpaired Student's *t*-test).

Tables

Table 1. Percentage amino acid sequence identity of ancestors and extant forms

	CYP 4ABTXZ	CYP 4BT	CYP 4A	CYP 4XZ	rCYP 4B1	sCYP 4T2[a]	hCYP 4A11	hCYP 4X1	hCYP 4Z1
CYP4ABTXZ	100								
CYP4BT	85	100							
CYP4A	85	80	100						
CYP4XZ	91	81	80	100					
rCYP4B1	61	69	63	61	100				
sCYP4T2[a]	57	60	53	54	51	100			
hCYP4A11	68	65	74	65	55	48	100		
hCYP4X1	59	55	53	63	47	44	48	100	
hCYP4Z1	61	55	55	64	45	45	50	52	100

[a] CYP4T2 from Sea Bass

Table 2. Expression and thermostability of ancestral and extant CYP4 enzymes

Protein	**Expression (nmol/L)**	**$^{10}T_{50}$ (°C)**	**$^{60}T_{50}$ (°C)**
CYP4A	267 ± 8	53 ± 2	51.3 ± 0.6
CYP4ABTXZ	320 ± 20	67.74 ± 0.05	58.6 ± 0.4
CYP4BT	1450 ± 60	62.9 ± 0.4	57.4 ± 0.4
CYP4XZ	440 ± 190	58 ± 3	50.8 ± 0.9
hCYP4A11	300 ± 50	33.8 ± 0.6	29.0 ± 2.0
rCYP4B1	451 ± 12	38.5 ± 0.5	35.9 ± 0.4
hCYP4X1	726 ± 18	36.0 ± 0.3	30.7 ± 0.1