Ancestral-sequence reconstruction unveils the structural basis of function in mammalian FMOs

Nicoll, Callum R; Bailleul, Gautier; Fiorentini, Filippo; Mascotti, María Laura; Fraaije, Marco W; Mattevi, Andrea

Published in:
Nature Structural & Molecular Biology

DOI:
10.1038/s41594-019-0347-2

Publication date:
2019

Ancestral Sequence Reconstruction Unveils the Structural Basis of Catalysis and Membrane Binding in Mammalian Flavin-Containing Monooxygenases

Callum R. Nicoll[1], Gautier Bailleul[2], Filippo Fiorentini[1],
María Laura Mascotti[3,*], Marco W. Fraaije[2,*] and Andrea Mattevi[1,*]

[1] Department of Biology and Biotechnology "Lazzaro Spallanzani", University of Pavia, via Ferrata 9, 27100 Pavia, Italy

[2] Molecular Enzymology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

[3] IMIBIO-SL CONICET, Facultad de Química Bioquímica y Farmacia, Universidad Nacional de San Luis, Ejército de los Andes 950, San Luis D5700HHW, Argentina

[*]Correspondence to María Laura Mascotti, Marco Fraaije, Andrea Mattevi
E-mail: mlmascotti@unsl.edu.ar, m.w.fraaije@rug.nl, andrea.mattevi@unipv.it

Abstract

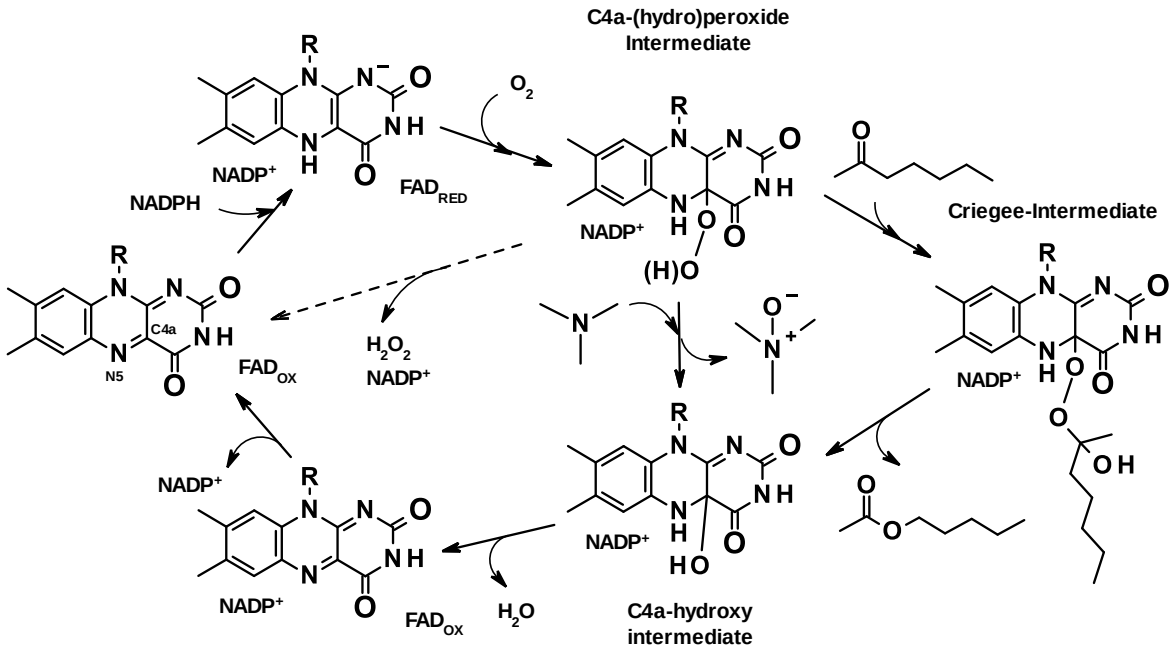
Flavin-containing monooxygenases (FMOs) are ubiquitous in all domains of life and metabolize a myriad of xenobiotics including toxins, pesticides and drugs. However, despite their pharmacological significance, structural information remains bereft. To further our understanding behind their biochemistry and diversity, we scrutinized three ancient mammalian FMOs: AncFMO2, AncFMO3-6 and AncFMO5, using Ancestral Sequence Reconstruction, kinetic and crystallographic techniques. Remarkably, all AncFMOs could be crystallized, and were structurally resolved between 2.7 and 3.2 Å. These crystal structures depict the unprecedented topology of mammalian FMOs. Each employs extensive membrane-binding features and intricate substrate-profiling tunnel networks through a conspicuous membrane-adhering insertion. Furthermore, a glutamate–histidine switch is speculated to induce the distinctive Baeyer-Villiger oxidation activity of FMO5. The AncFMOs exhibited catalysis akin to human FMOs and, with sequence identities between 82 and 92%, represent excellent models. Our study demonstrates the power of ancestral sequence reconstruction as a strategy for the crystallization of proteins.

Introduction

Xenobiotic metabolism is an ancient and imperative process pursued by all organisms. With evolution resulting in the production of, and thus exposure to, a vast number of noxious and toxic natural products,[1] organisms have employed a multitude of intricate detoxification systems, to tackle the sheer quantity of diverse chemicals.[1–6] Flavin-containing monooxygenases (FMOs; EC 1.14.13.8) represent one of these detoxifying protein families and are prevalent in all domains of life.[4,7] FMOs are members of the Class B flavin-dependent monooxygenases and utilize the cofactors FAD and NADP(H), and dioxygen for activity.[8–10] Typically, FMOs pursue catalysis as illustrated in Scheme 1, whereby a soft nucleophile (here demonstrated with trimethylamine) receives the distal oxygen atom from the C4a-(hydro)peroxyflavin intermediate.[11,12] The more water-soluble hydroxylated product is then released by the enzyme to be excreted from the host.

Humans possess five FMO isoforms that are differentially expressed in many different tissues such as the kidney, lung, and liver.[2,10,13,14] The human *FMO* genes are found on chromosome 1, with *FMO1-4* clustering over 220 kb, and *FMO5* found on a separate chromosome region.[15,16] The human FMO family contains six non-expressed pseudogenes which are also located on chromosome 1.[16] FMOs are involved in phase I of xenobiotic detoxification.[2,3] They oxidize an array of compounds bearing soft nucleophilic centers such as nitrogen and sulfur atoms,[2,17–19] making them clinically important regarding drug metabolism.[3,6,12,15,17,20,21] The most extensively characterized FMO is human FMO3, renowned for its production of trimethylamine N-oxide.[22–26]

FMO3 deactivation upon mutation induces trimethylaminuria ("fish odor syndrome"), whereby the body has an unpleasant smell due to the accumulation of trimethylamine.[27–30] Whilst FMO4 has not been extensively characterized, FMO1 and FMO3 were shown to have broad substrate ranges, metabolizing substrates as diverse as itopride (acetylcholine esterase inhibitor), and tamoxifen (anti breast-cancer drug).[2,31–34] Also FMO2 features a rather broad substrate profile, acting on pesticides such as napthylthiourea,[19] although its role in human metabolism remains partly unknown because FMO2 is not expressed in the majority of humans due to a mutation.[35,36] FMO5 is distinct from the other FMOs because it is able to perform Baeyer-Villiger oxidations (Scheme 1),[37] metabolizing ketone-containing drugs such as pentoxifylline (a muscle-pain killer).[17] Recent literature documents that FMOs are associated to diseases such as atherosclerosis and diabetes,[23,26] promote longevity,[38] and regulate cholesterol and glucose levels.[26,39–41] Despite their discovery over 30 years ago, the determinants underlying the existence of five isoforms remain unexplored and even more strikingly, no mammalian FMO has been structurally elucidated. This gap in our knowledge on these key enzymes of human drug metabolism likely reflects their distinctive feature: unlike bacterial, fungal, and insect FMOs that are soluble, mammalian FMOs are insoluble and reside in the membranes of the endoplasmic reticulum.[2,5]



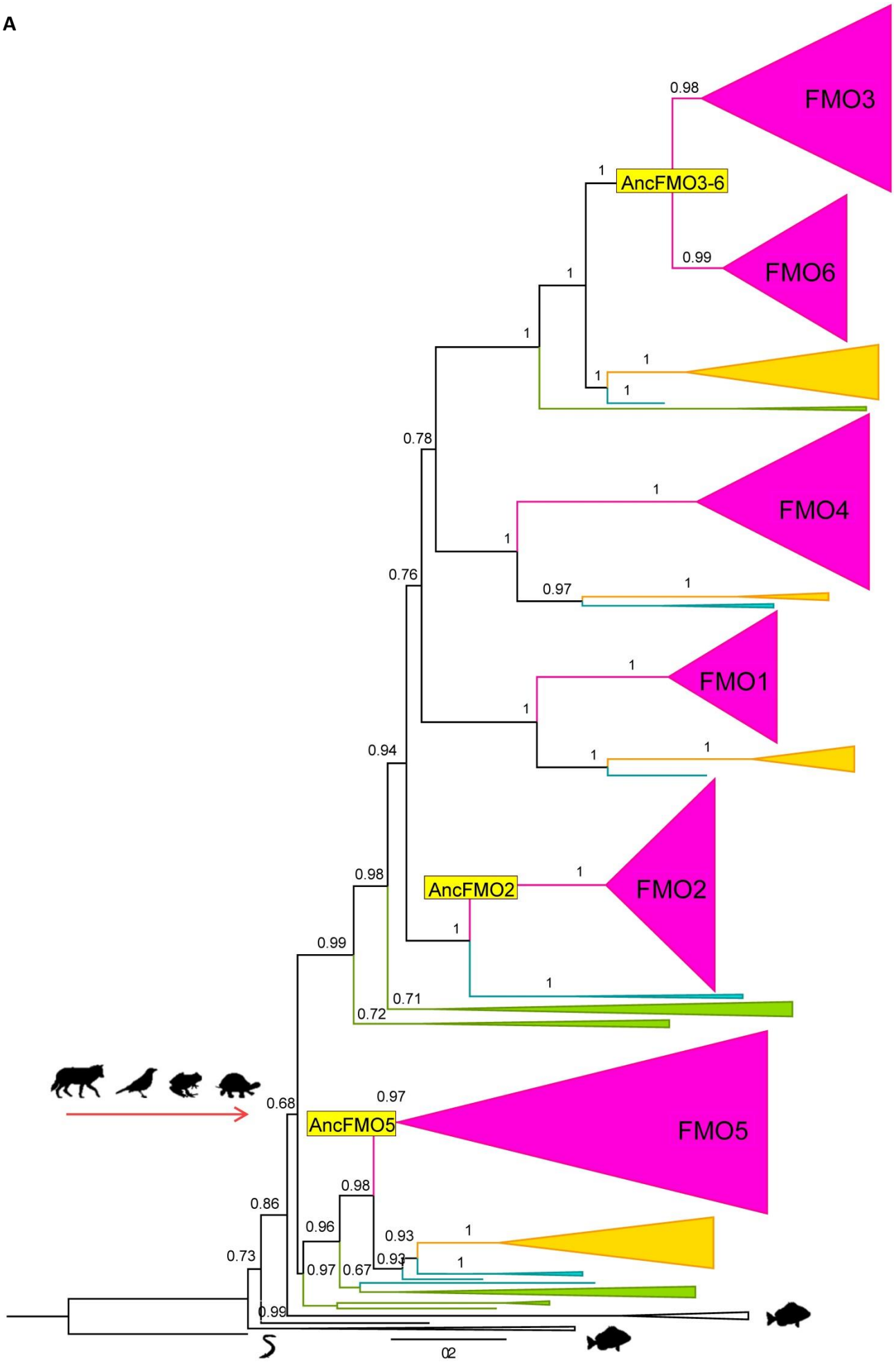
Scheme 1. The catalytic mechanism of FMOs. FAD_{ox} is reduced by NADPH. FAD_{red} is consequently oxidized by a molecule of dioxygen to generate the C4alpha-(hydro)peroxide intermediate. The typical mode of action of FMOs with the distal oxygen atom from the intermediate being inserted onto a soft nucleophile through nucleophilic addition is shown with reference to trimethylamine. The Baeyer-Villiger monooxygenation activity conducted by human FMO5 is shown on the right with reference to heptan-2-one. The dotted arrow indicates the uncoupling reaction whereby the C4alpha-(hydro)peroxide intermediate decays with the release of $NADP^+$ and hydrogen peroxide.

To gain insight into the historical events leading to the paralogs divergence in mammals, we generated three ancestral FMOs (*i.e.* the last common ancestors of extant mammalian FMO2s, FMO3s/FMO6s and FMO5s; herein referred to as AncFMOs) using Ancestral Sequence Reconstruction:[42,43] These enzymes were successfully expressed in *E. coli* and purified as holo (FAD-containing) and active enzymes. Despite countless failed crystallization attempts of human FMO3 and human FMO5, we were able to crystallize and structurally elucidate each AncFMO. In this article, we describe the unprecedented membrane-binding features associated with the mammalian FMO and we illustrate that substrate specificity is controlled by tunnel design rather than catalytic-site architecture. Furthermore, we demonstrate that the biochemistry of FMOs has been strictly conserved and that ancestral sequence reconstruction is a powerful tool to facilitate crystallization.

Results

Ancestral sequence reconstruction of mammalian flavin-containing monooxygenases

We inferred the evolutionary history of FMOs from a full phylogeny constructed by including experimentally-characterized enzymes from Bacteria and Eukarya, plus sequences found by extensive sequence homology searching and HMM profiling (Supplementary Figure 1 and Data 1). Our work confirmed the findings of the previous studies by Hernandez et al.[16] and Hao et al.[44]: (i) jawed vertebrate FMOs are monophyletic and derived from a single common ancestor (Figure 1A, Supplementary Figure 2); (ii) several duplication events occurred in the terrestrial vertebrates; (iii) the ancestor of mammals already encoded the five FMO paralogs resulting from four major gene-duplication events (Figure 1A, Supplementary Figure S2 and Data 1); (iv) a sixth mammalian paralog (FMO6) resulted from a late gene duplication event. FMO6 has been described as a pseudogene in humans[45] but it might be functional in mouse[16] and its nature is unknown in other mammals.


A
FMO3
AncFMO3-6
FMO6
FMO4
FMO1
AncFMO2
FMO2
AncFMO5
FMO5
0.98
1
0.99
1
1
1
1
1
0.78
1
1
0.97
1
0.76
1
1
1
0.94
0.98
1
0.99
1
0.71
0.72
0.97
0.68
0.98
0.86
0.96
0.93
1
0.73
0.93
1
0.97
0.67
0.99

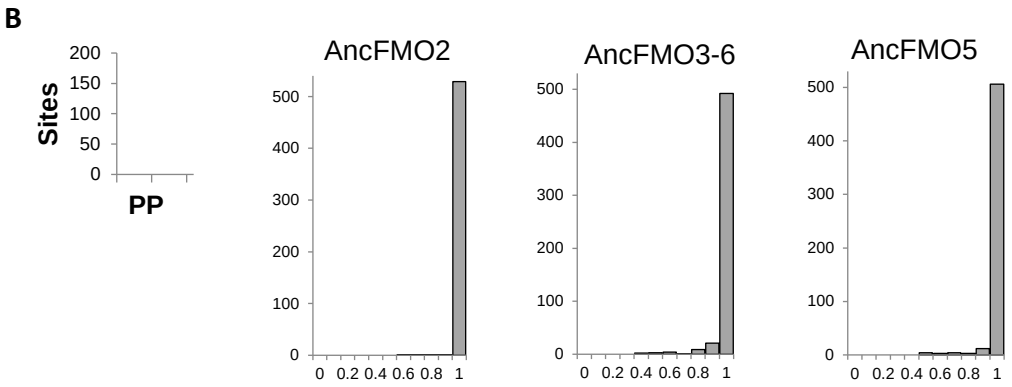


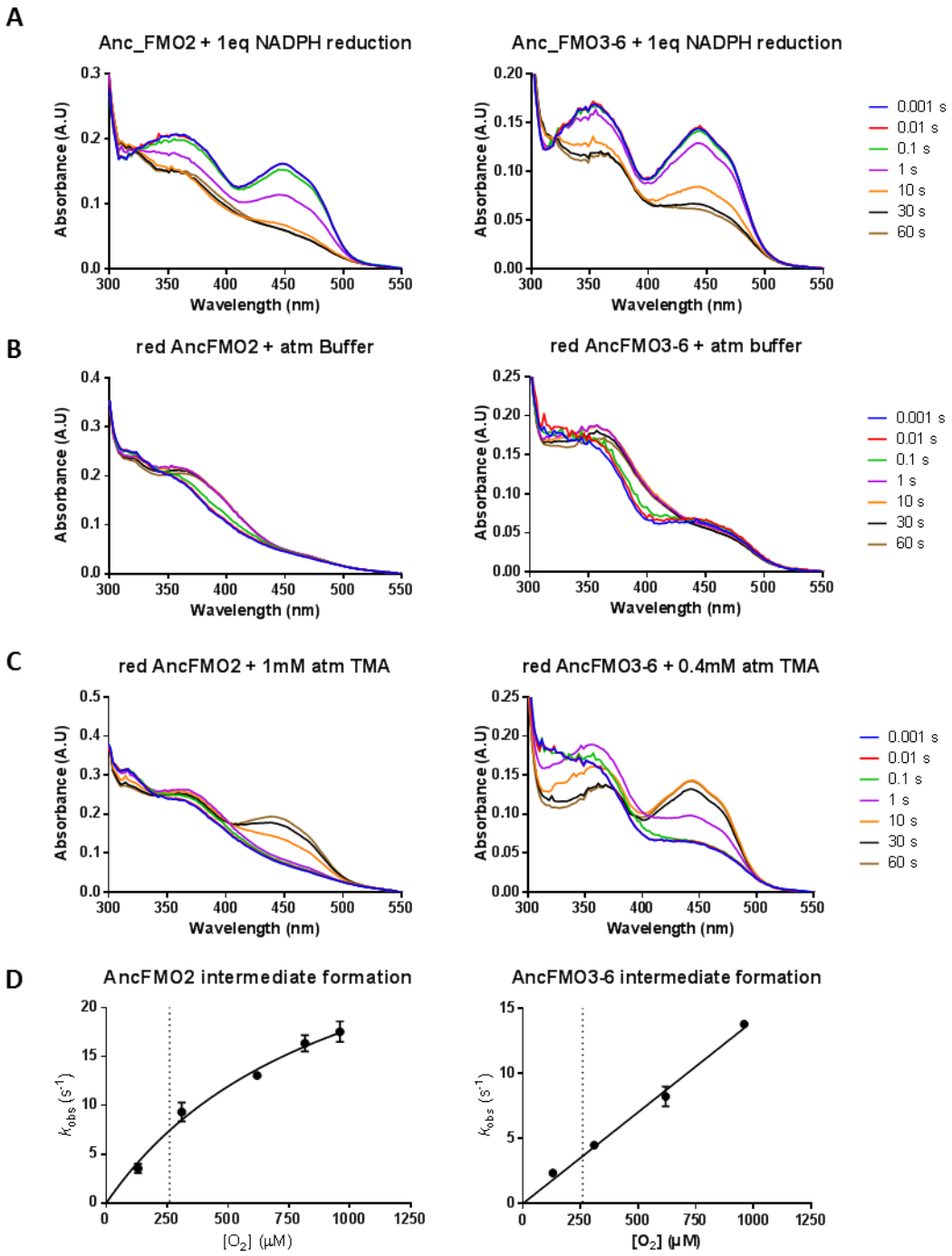
Figure 1. Ancestral Sequence Reconstruction of FMOs. **A**: Condensed Maximum Likelihood phylogeny of FMOs from jawed vertebrates. Clades are colored according to tetrapod classes: mammalia (magenta), aves (light orange), amphibia (green) and testudines (teal). Clades on the base are from other non-terrestrial gnathostomes (black). Rooting was performed according to the species tree. Above the branches transfer bootstrap expectation values are shown. The emergence of terrestrial vertebrates (tetrapods, 352 mya)[46] is marked with an arrow and cartoons on the left. The three ancestral nodes that were experimentally characterized are labeled with yellow squares. Fully annotated phylogeny is presented in Supplementary Figure 2. **B**: Statistical confidence of ancestral amino acid states. The highest posterior probability (PP) for each of the inferred ancestral states (sites) in AncFMOs is shown. Average accuracy for AncFMO2= 0.994, AncFMO3-6: 0.982 and AncFMO5: 0.987.

By performing ancestral sequence reconstruction, we obtained the protein sequences of AncFMOs from mammals with high posterior probabilities (ranging from 0.98-0.99) (Figure 1B, Supplementary Data 2). In the phylogeny, we observed that FMO5 diverged earlier in agreement with previous reports,[44] followed by FMO2, FMO1, FMO4 and the FMO3-6 hybrid. This topology suggests that the gene duplication events took place simultaneously rendering no clear paralog couples as it has been previously proposed (Figure 1A, Supplementary Figure 2).[16,44] Among the whole clade of present-day FMO2s, 80% of sites are conserved, while the rest are likely responsible for functional differences among species. We observe that from AncFMO2 to human FMO2, 42 substitutions have occurred of which 18 are conservative (as defined by Grantham[47]). In the case of AncFMO3-6, the ancestor underwent an early duplication event originating the FMO3 and FMO6 paralogs in mammals. As a general trend, comparing the pre-duplication ancestor to modern FMO3 and FMO6, 70% of the sites are conserved. Along each branch to the human FMO3 or human FMO6 sequences, 94-98 substitutions occurred, 28-30 of them being conservative. The lower degree of conservation is not surprising considering the duplication scenario. Finally, FMO5 is the most enigmatic of all extant FMOs due to its Baeyer-Villiger oxidation activity.[37] AncFMO5 shows 44 changes along the branch to human FMO5, with 19 conservative substitutions. In light of this

historical scenario, we selected AncFMO2, AncFMO3-6 and AncFMO5 for experimental characterization.

AncFMOs portray catalytic rates similar to those of extant mammalian FMOs

Critically for our project, the hitherto generated AncFMOs sequences proved to encode stable proteins that can be effectively produced and purified as recombinant, FAD-loaded, and catalytically competent enzymes in *E. coli*. Thus, the first relevant result was that a convenient bacterial expression system for the study and biocatalytic exploitation of close homologs to human FMOs was established (see materials and methods). We next verified whether these enzymes retained enzymatic activities by performing steady-state kinetics experiments using a NADPH-depletion spectrophotometric assay. The NADPH oxidase activity was initially tested (NADPH consumption in absence of an organic substrate; $NADPH_{uncoupling}$ in Table 1). This was followed by the measurements of the kinetics of the reaction in the presence of known oxygen-accepting substrates of FMO2 and FMO3 (methimazole, trimethylamine and thioanisole), and FMO5 (heptan-2-one). The results were reassuring in that AncFMO2, AncFMO3-6, and AncFMO5 proved to be enzymatically active with kinetic parameters very similar with those reported for their extant human-derived enzymes.[2,17,34,37,48–51] The k_{cat}, K_M^{NADPH} and uncoupling values ranged between 0.03-0.32 s^{-1}, 3.5-7.8 µM, and 0.016-0.03 s^{-1}, respectively (Table 1). It was especially noticeable that the AncFMOs displayed a high affinity towards the coenzyme NADPH and a significantly higher NADPH consumption rate when a suitable substrate was present. This result is in full agreement with the canonical catalytic mechanism observed for FMOs and sequence related flavoprotein monooxygenases (Scheme 1). These features were further demonstrated by stopped-flow kinetic studies. NADPH-reduced AncFMO2 and AncFMO3-6 were found to react rapidly with oxygen to form a stable and detectable C4alpha-(hydro)peroxyflavin intermediate with its well-defined spectroscopic properties (Figure 2). Based on the steady-state kinetics data, AncFMO5 is assumed to behave similarly. Collectively, these experiments convincingly demonstrated that our AncFMO2/3-6/5 enzymes are enzymatically competent and exhibit the typical catalytic features of Class B flavoprotein monooxygenases.



E

	AncFMO2	AncFMO3-6
k_{red} (s^{-1})	14.6 ± 0.5	0.2 ± 0.03
k intermediate formation (s^{-1})	7.4	3.7
k decay intermediate (s^{-1})	0.01 ± 0.006	0.01 ± 0.001

Figure 2. Stopped-flow kinetics studies on AncFMO2 and AncFMO3-6. **A**: Enzyme reduction upon the anaerobic addition of NADPH. **B**: Mixing reduced enzyme with dioxygen (0.13 mM) reveals the appearance of a peak at 360 nm which is characteristic for a C4a-(hydro)peroxyflavin intermediate (Scheme 1). **C**: Mixing reduced enzyme with dioxygen (0.13 mM) and trimethylamine (1 mM or 0.4 mM for AncFMO2 and 3-6, respectively) reveals again a rapid formation of the C4a-

(hydro)peroxyflavin intermediate which subsequently decays to form the reoxidized flavin species. **D**: Dependence of the rate of C4a-(hydro)peroxyflavin formation (A_{360nm}) on varying oxygen concentrations. The dotted lines correspond to the atmospheric concentration of dioxygen (0.26 mM). For AncFMO2, the observed saturation behavior suggests a binding event taking place before dioxygen reacts with the reduced flavin. Interestingly, such a saturation behavior was also reported for pig liver FMO1.[52] **E**: Rates of reduction, C4a-(hydro)peroxyflavin formation, and C4a-(hydro)peroxyflavin decay in the absence of substrate (0.26 mM dioxygen; dotted line on panels D).

Table 1: Steady state-kinetics.

	Substrate[a]	k_{cat} (s^{-1})	K_M (μM)
Ancient FMOs			
AncFMO2	Methimazole	0.19 ± 0.01	106 ± 22
	Thioanisole	0.3 ± 0.02	6.9 ± 1.6
	Trimethylamine	0.16 ± 0.008	445 ± 74
	NADPH	0.32 ± 0.05	7.8 ± 1.4
	$NADPH_{uncoupling}$	0.02 ± 0.001	20 ± 5.4
AncFMO3-6	Methimazole	0.19 ± 0.005	21 ± 2.3
	Thioanisole	0.1 ± 0.008	128 ± 38
	Trimethylamine	0.24 ± 0.01	41 ± 6.3
	NADPH	0.13 ± 0.008	3.5 ± 0.86
	$NADPH_{uncoupling}$	0.022 ± 0.002	16 ± 5.4
AncFMO5	Heptan-2-one[b]	0.07 ± 0.003	6.36 ± 1.2
	NADPH	0.06 ± 0.001	6.48 ± 0.38
	$NADPH_{uncoupling}$	0.03 ± 0.001	2.1 ± 0.5
Extant FMOs			
human FMO3	NADPH[c]	0.06 ± 0.16	46 ± 9
human FMO5	NADPH[c]	0.197 ± 0.009	59 ± 8

[a] Rates were determined by following NADPH consumption (absorbance decrease at 340 nm). The buffer was composed of 50 mM potassium phosphate (pH 7.5), 250 mM NaCl, 0.05% (v/v) triton X-100 reduced. The reactions were run at 37 °C. For the determination of the K_M of the substrates, 100 μM and 50 μM NADPH was used for AncFMO2 and AncFMO3-6, and AncFMO5 respectively. For the determination of the K_M for NADPH, 1 mM trimethylamine was used as oxygen-accepting substrate for AncFMO2 and AncFMO3-6, whilst 30 mM of heptan-2-one was used for AncFMO5. $NADPH_{uncoupling}$ rates were determined in the absence of substrates. The increase ("burst") in NADPH consumption rates upon addition of the substrates demonstrate that the AncFMOs are highly coupled and effectively oxygenate their substrates.

[b] Heptan-2-one is a typical ketone substrate for the Baeyer-Villiger oxidation catalyzed by FMO5 (Scheme 1).

[c] The rates for NADPH consumption in the presence and absence of the substrate are the same because of a high degree of uncoupling in the extant human FMOs. The data for human FMO3 are shown in Supplementary Figure 8. The data for human FMO5 are taken from Fiorentini et al.[37]

AncFMOs crystallize as dimers with extensive membrane binding features

To investigate the role of the AncFMOs in detail, the crystal structures of each AncFMO in the presence of $NADP^+$ were determined (Figure 3). AncFMO2 was also crystallized in the absence of $NADP^+$ but, akin to Class B flavin-dependent monooxygenases, no major conformational changes were observed between the apo- or holo-enzyme crystal structures (Supplementary Figure 3). The structures were solved at 2.7, 3.0, 2.8, and 2.7 Å resolution for AncFMO2 (without $NADP^+$), AncFMO2, AncFMO3-6, and AncFMO5 (all including $NADP^+$), respectively (Supplementary Table 1, Supplementary Figure 4). For the purpose of the structural analysis, it must be highlighted that the AncFMOs display high sequence identities to their extant human FMO counterparts: 92%, 83%, and 92% for AncFMO2, AncFMO3-6, and AncFMO5 respectively, making them excellent structural models of human FMOs (Supplementary Figure 5).

Our crystal structures depict the AncFMOs as dimers: they possess an extensive monomer-monomer interface over approximately 2000 $Å^2$ (calculated by the PISA server).[53] Furthermore, their well-conserved FAD and NADP(H) dinucleotide-binding domains are accompanied by two large trans-membrane helices (one from each monomer) that project outwards, approximately parallel to the twofold axis (Figure 3A-D). Pairwise structural superpositions of AncFMO2, AncFMO3-6, and AncFMO5 show that their ordered ~480 Cα atoms overlap with root-mean-square deviations of less than 1 Å. This result reveals a high degree of structural similarity among the FMO structures. We additionally notice that the dimerization interface of the AncFMOs is different from the dimer interfaces exhibited by soluble FMOs (e.g. FMO from *Roseovarius nubinhibens*, PDB entry 5IPY; FMO from *Methylophaga aminisulfidivorans*, PDB entry 2VQ7).

The mammalian FMOs were predicted to contain a highly hydrophobic C-terminal transmembrane helix (residue 510-532 in AncFMO3-6; Supplementary Figure 5). The crystal structures of AncFMO2 and AncFMO3-6 perfectly confirmed this prediction as both enzymes possess C-terminal trans-membrane helices that span 30 Å in length and are decorated with many hydrophobic residues (Figure 3A, 3B and 3D). Of notice, these α-helical scaffolds represent the key protein-protein interactions established within the crystal packing (Supplementary Figure 6). High disorder rendered the C-terminal residues untraceable in the crystal structure of AncFMO5. The transmembrane helices of AncFMO2 and AncFMO3-6 root themselves deep within the phospholipid bilayer through a bitopic membrane binding mode, whereby the final C-terminal residues exit the other side of the membrane. These two helices anchor the protein firmly into the membrane. Thus, Figure 3 depicts each enzyme as if it were sitting on the membrane.

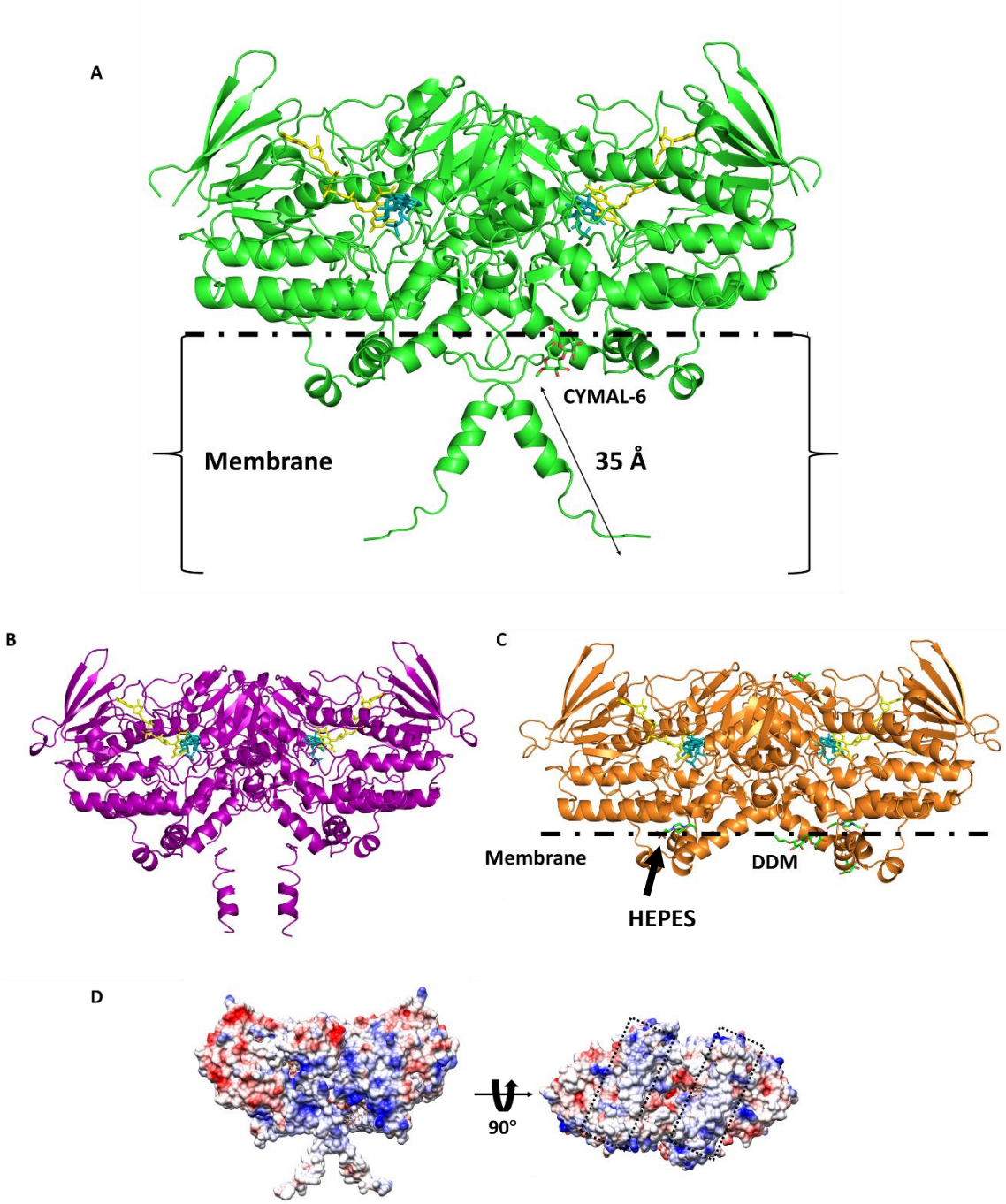


Figure 3: Crystal structures of the AncFMOs. Crystallographic dimers of AncFMO2, AncFMO3-6 and AncFMO5 are shown in lime green (**A**), dark magenta (**B**) and orange (**C**). FAD and $NADP^+$ are shown in yellow and cornflower blue, respectively. The orientations of the AncFMOs are identical, depicting their structures as if they were sitting on the phospholipid bilayer. In **A**, the lengths of the trans-membrane helices are portrayed at 35 Å, with the membrane cross-section indicated with brackets. In **C**, the membrane-protein interface is indicated by a horizontal dashed line, mapped with respect to the polar head group of the dodecyl-β-D-maltoside (DDM) detergent. Additionally, a molecule of HEPES buffer is observed entering the enzyme at the membrane-protein interface. **D**: Distribution of charge around the surface of AncFMO2, with red, white, and blue representing negative, neutral, and positive respectively. On rotation about 90°, we see the large parallel hydrophobic strips across the bottom of the dimer, lined by a ring of positively charged residues indicated by black dashed boxes.

It was reported that truncation of the C-terminal helices was insufficient for protein solubilization.[54,55] This indicated that the enzyme possessed additional membrane-binding features. To understand what elements promote membrane association, the charge distribution on the protein surface was scrutinized. Intriguingly on the underside of the dimer, two large hydrophobic strips, about 30 Å in length, extend across the enzyme surface (Figure 3D). These strips are lined by an extensive ring of positively charged residues. Collectively, these features equip the enzyme for binding to the membrane surface. The array of hydrophobic residues penetrate, monotopically, into the phospholipid bilayer and are held in place by the ionic-based interactions introduced between the negatively charged, polar head group of the phospholipids and the positively charged amino acids. Serendipitously, in the crystal structures of AncFMO2 and AncFMO5, we were able to observe the polar head groups of the detergent molecules, CYMAL-6 and dodecyl-β-D-maltoside (DDM) respectively, that were used for protein solubilisation and crystallisation (Figure 3A, 3C, 3D). These molecules delineate the membrane-enzyme interface and further validates that the hydrophobic strips monotopically embed within the membrane. These findings rationalize the extensive membrane binding nature employed for this class of enzymes and corroborates that truncation of the C-terminal helix alone is not sufficient to facilitate protein solubilization.[54] FMOs employ both bitopic and monotopic membrane-binding features to grapple onto the membrane effectively and abstract lipophilic substrates from within the membrane.

An eighty-residue insertion reshapes the active site and promotes membrane association

To comprehend the unique and distinct structural features associated with mammalian FMOs, we compared them with structurally characterized soluble FMOs. Consistent with Class B flavin-dependent monooxygenases, the AncFMOs have two well-conserved dinucleotide-binding domains for cofactors FAD (residues 2-154 and 331-442) and NADP(H) (residues 155-213 and 296-330) respectively, known as the paired Rossmann fold (Supplementary Figure 7A).[2,56] Superposition of a bacterial FMO (PDB: ID 2vq7, SEQ ID: 29%) from *Methylophaga* sp strain SK1,[12] shows a root-mean-square deviation of 1.1 Å over 205 Cα atom pairs, verifying a strict evolutionary conservation of the dinucleotide binding domains. However, close inspection of the structures reveals very substantial differences. In soluble FMOs, the FAD cofactor is exposed to the solvent and readily accessible by substrates. By contrast, an 80-residue insertion (214-295 in AncFMO3-6; Supplementary Figure 7B-C) shields the AncFMOs' active site from the cytosol and creates closed substrate-binding cavities. This insertion is comprised of a subdomain orchestrated by three small α-helical units that form a ridge-like, triangular fold. Additionally, this subdomain forms the first half of the hydrophobic strip mentioned above. Despite the FAD and the catalytic-center being

buried by the insertion, this subdomain provides a series of tunnels that branch out from the membrane towards the active site (see below). This finding implies that substrates navigate through tunnels manufactured by the insertion to access the closed catalytic cavity.

AncFMO consists of a buried active site and a well conserved NADP(H) binding mode

With the AncFMOs active sites no longer being open clefts like their soluble homologs, we scrutinized each closely to determine the functions of each residue and whether the mode of NADP(H) binding is akin to Class B flavin-dependent monooxygenases. Notably, most residues in the active and NADPH-binding sites are conserved with near-identical conformations (Figure 4A-C). Thr62/Ser62/Thr63 for AncFMO2/3-6/5 respectively, are within hydrogen bonding distance to the N3 atom of the isoalloxazine ring and orientate the FAD towards the substrate pocket. Additionally, Asn61/61/62 is observed in all active sites and is strictly conserved among human FMOs (Supplementary Figure 5). This residue situates close to the C4a of the isoalloxazine ring (4.6 Å) and is likely fundamental for the stabilization of the flavin-(hydro)peroxide intermediate (Scheme 1). Consistently, mutating this residue in human FMO3 causes trimethylaminuria, further verifying its integral role within the active site.[12,28]

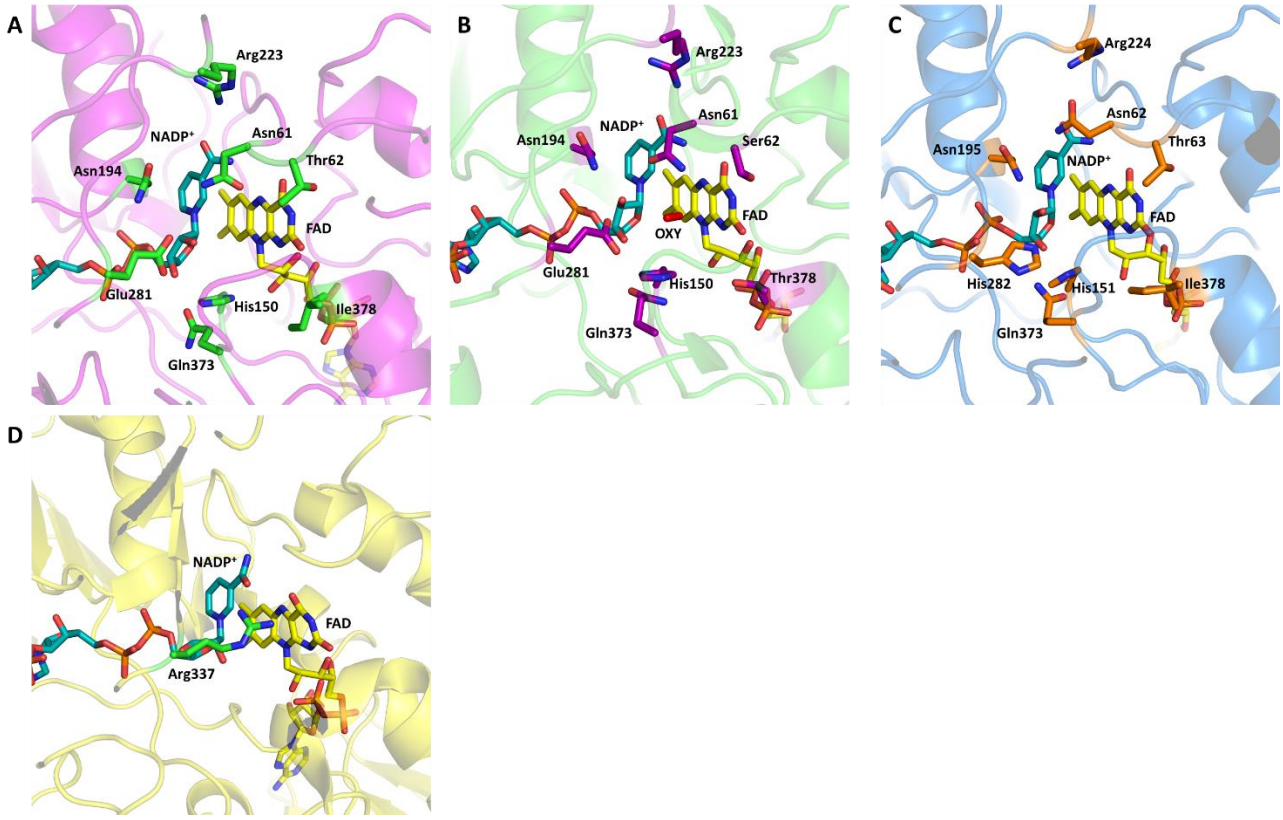


Figure 4: Active sites of the AncFMOs. The active sites for AncFMO2, AncFMO3-6 and AncFMO5 are shown in **A**, **B** and **C**, respectively. All three bear a high degree of similarity with most amino acids being strictly conserved and displaying identical conformations. The differing residues are as follows: Thr62/Ser62/Thr63; E281/E281/H282; Ile378/Thr378/Ile378 for AncFMO2/AncFMO3-6/AncFMO5 respectively. AncFMO3-6 also contains a tentatively assigned

molecule of dioxygen (OXY). For the sake of comparison, panel **D** shows the binding of $NADP^+$ to the active site of phenylacetone monooxygenase, a prototypical class B monooxygenase (PDB code 2ylr). Arg337 is a conserved residue that is essential for the Baeyer-Villiger activity of this and similar enzymes.

The binding mode of $NADP^+$ observed in the crystal structures is iconic to Class B flavin-dependent monooxygenases (Figure 4D).[4] The overhanging Arg223/223/224 is within hydrogen bonding distance of the carbonyl derived from the carbamide of $NADP^+$. Additionally, the amino group of the same carbamide forms a hydrogen bond with the N5 atom of the isoalloxazine ring. More so, the nicotinamide ring is sterically held in place by a well conserved Asn194/194/195 which acts like a *back door* for the cofactor (Figure 4A-C). This feature is not uncommon and portrayed in some soluble FMOs by a protruding tyrosine.[12,57,58] The hydroxyl groups of the ribose forms part of an intricate hydrogen bonding network. The 2'-OH group is within hydrogen bonding distance of the back-door residue Asn194/194/195 (3.0 Å) in AncFMO3-6 and AncFMO5, and Glu281 (2.9 Å) in AncFMO2 and AncFMO3-6. Additionally, the conserved Gln373 among the AncFMOs is within hydrogen-bonding distance of the 3'-OH group. Collectively, these hydrogen bonds and the steric interactions orientate the nicotinamide and the ribose in a manner characteristic to this class of enzymes and reiterate a significant role of $NADP^+$ in catalysis, most likely in C4a-(hydro)peroxyflavin formation/stabilization and substrate oxygenation (Scheme 1).[12,52,57,59]

A Glu–His mutation in the mammalian specific insertion may promote Baeyer-Villiger oxidation in FMO5

With AncFMO5 being structurally very similar to the AncFMO2 and AncFMO3-6, but at the same time functionally divergent, we sought out to clarify what features gave rise to its Baeyer Villiger oxidation activity. Inspecting the active site alone, the differing mode of action is likely derived from a Glu-to-His substitution. In AncFMO2 and AncFMO3-6, Glu281, derived from the above-described mammalian FMO-specific 80-residue insertion, points towards the flavin ring. With positively charged substrates being preferred by FMOs,[2] Glu281 is probably deprotonated and negatively charged within the cavity. In AncFMO5 however, this residue is substituted for His282 which optimally positions the N_ε-H of its imidazole ring towards the substrate pocket (Figure 4C) and likely serves as a hydrogen bond donor. This function is commensurate with Baeyer-Villiger monooxygenases, whereby hydrogen bond-donating residues (*i.e.* Arg; Figure 4D) are prevalent in the vicinity of the FAD ring to activate the carbonyl functional group of the substrate for electrophilic attack by the flavin-peroxy intermediate and stabilize the Criegee intermediate formed during Baeyer-Villiger oxidation catalysis (Scheme 1).[60–62] These observations rationalize the functional convergence observed among the FMO5 clade and Baeyer-Villiger monooxygenases. To

probe the importance of His282 in AncFMO5, the H282E mutant was prepared and analyzed. This revealed that the H282E AncFMO5 mutant fully lost its activity. Analogously, the E281H AncFMO2 and E281H AncFMO3-6 mutant enzymes were prepared which were found to retain FMO activity toward thioanisole. Yet, they were not able to perform Baeyer-Villiger oxidations (Supplementary Table 2). This could be due to the fact that the fine structural and geometric features for formation and stabilization of the Criegee intermediate (Scheme 1) needs further mutations, e.g. in the second-shell of active site residues.

AncFMOs possess a conserved substrate tunnel that branches out towards the membrane

As the mammalian FMOs are notorious for their broad substrate profiles, we conducted extensive research to elaborate how the substrates navigate through the enzyme using the *HOLLOW* server.[63] The conserved tunnel is roughly perpendicular to the face of the isoalloxazine ring and extends outwards (approximately 16 Å) towards the membrane, before deviating in multiple directions (Figure 5). In all three structures, the inner segment of the tunnel features a conserved leucine that acts as a gate keeper (Leu375 in AncFMO3-6; Figure 5B, lower panel): in an upward position, it creates a closed cavity at the active site (AncFMO3-6), and in the downward position, it opens the tunnel to the protein-membrane interface (AncFMO2 and AncFMO5). This leucine is also conserved in human FMO1-3 and 5, implying an integral role in gating the inner "catalytic" part of the tunnel and affording a solvent-protected environment for catalysis (Supplementary Figure 5).

The substrates/products penetrate/exit the tunnels through the subdomain found in the 80-residue insertion. Here, the paths are heavily dictated by the conformations of the residues in and around the subdomain (Figure 5A-D). Specifically, a few noticeable changes were observed (Figure 6). The largest conformational difference is seen at residues 337-352 and 338-352 for AncFMO2/3-6 and AncFMO5 respectively (herein referred to as loop 1). In AncFMO2 and AncFMO5, loop 1 forms a large arched fold that sits underneath the NADP(H) binding pocket. In AncFMO3-6, loop 1 instead forms a tightly coiled α-helix creating an open cavity below the $NADP^+$ binding pocket. This new cavity leads to the cytosolic tunnel observed in AncFMO3-6 (Figure 5B). The second difference observed comprises residues 419-431 for AncFMO2 and AncFMO5 and residues 419-429 for AncFMO3-6 (loop 2) in the neighborhood of the tunnel entrances. The final differences detected concern residues 273-282 of AncFMO2 and AncFMO3-6, and 274-283 of AncFMO5 (loop 3). In AncFMO5, loop 3 features a α-helical turn that blocks the cytosolic tunnel observed in AncFMO2 and AncFMO3-6. Moreover, AncFMO5 possesses a shorter α-helix in the subdomain which widens the cavity entrance site. These features have critical implications for the mechanisms of substrate binding and selectivity in FMOs. On the one hand, these structural variations on surface elements at

the tunnel entrances are likely to govern the similar but not identical substrate acceptance of the FMOs. On the other hand, despite these differences, all three AncFMO structures show that the tunnels can be accessed by both hydrophilic substrates that predictably diffuse from the cytosol, and by hydrophobic substrates that likely diffuse from the membrane. Likewise, hydrophilic and hydrophobic products can diffuse from the active site to the cytosol and to the membrane, respectively.

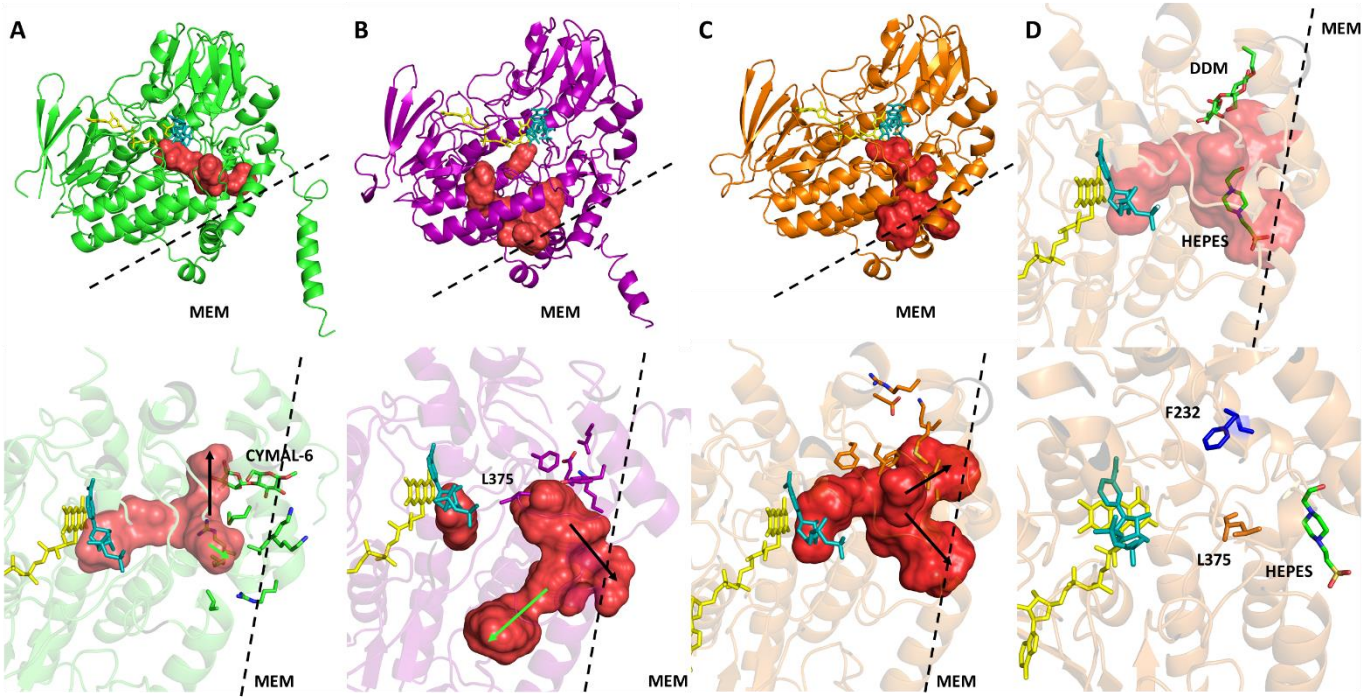


Figure 5: Substrate tunnels of the AncFMOs. Upper panels **A**, **B** and **C** portray the tunnels of AncFMO2, AncFMO3-6 and AncFMO5 respectively, with the protein-membrane interface labeled as MEM. Lower panels **A**, **B** and **C**, illustrate the directions of the tunnels for AncFMO2, AncFMO3-6 and AncFMO5 respectively with their directions towards the membrane or the cytosol depicted by black and green arrows respectively. The residues that block tunnel routes based on their conformations are shown. AncFMO2 and AncFMO3-6 contains two tunnel exits: one leading towards the membrane (black arrow) and one to the aqueous environment (green arrow). AncFMO5 contains two tunnels which both lead to membrane (black arrows). Upper panel **D** displays a molecule of DDM found above the α-helical triad to emphasize the protein-membrane interface in AncFMO5. Additionally, a molecule of HEPES is present in the tunnel passing below the helix, demonstrating a substrate accessible pathway. Lower panel **D** highlights residue Phe232 in AncFMO5 with respect to gatekeeper Leu375, inferring its vicinity to the FAD and how the change to alanine in human FMO5 is predicted to open the cavity.

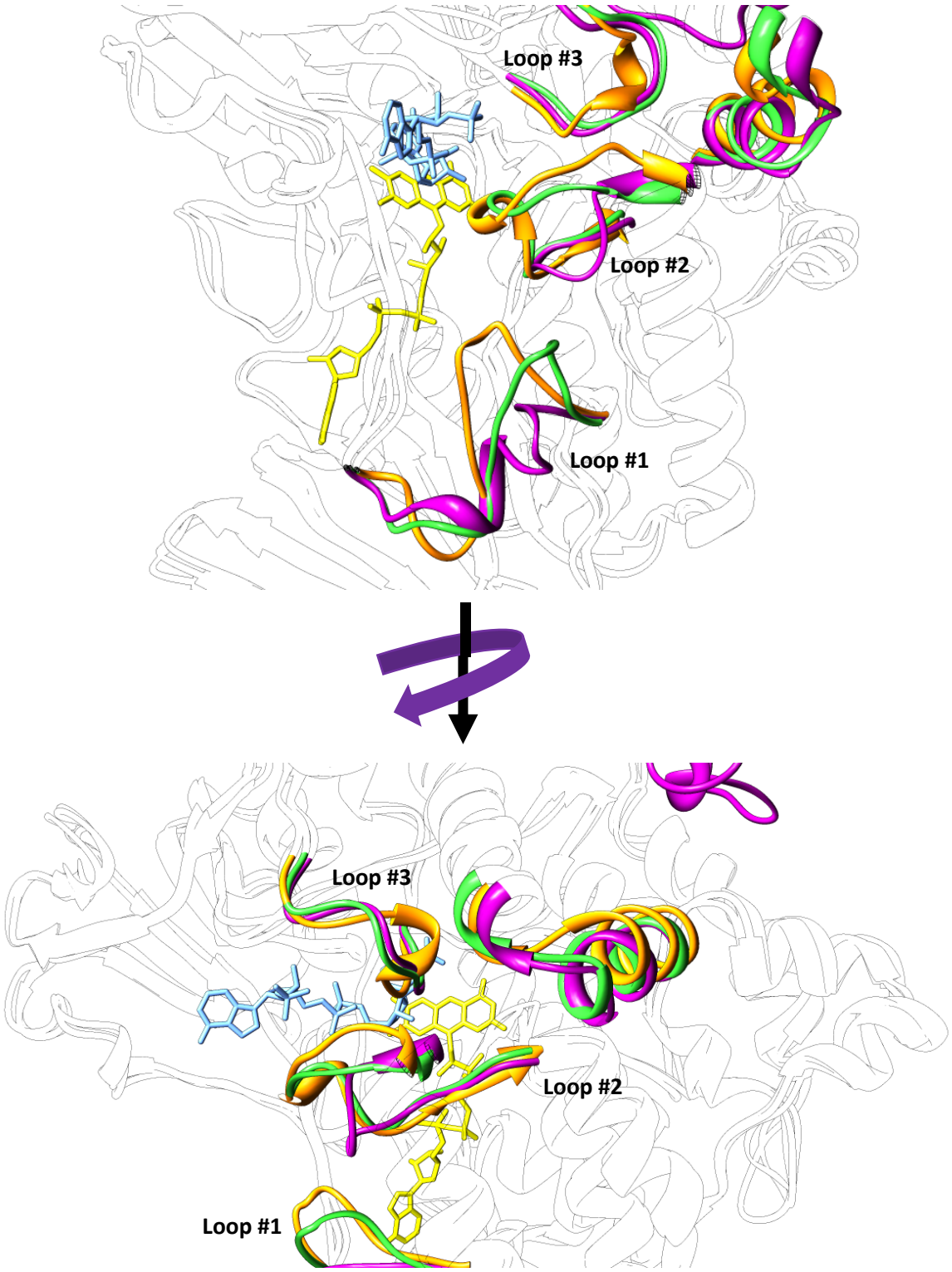


Figure 6: Structural differences among the AncFMOs. Upper and lower panels describe the conformational differences observed among the AncFMOs with AncFMO2, AncFMO3-6 and AncFMO5 depicted in lime green, dark magenta and orange respectively. Loop #1 contains residues 337-352 for AncFMO2 and AncFMO3-6, with resides 338-352 for AncFMO5. Loop #2 contains residues 419-431 for AncFMO2 and AncFMO5, with residues 419-429 for AncFMO3-6; Loop #3 contains residues 273-282 and 274-283 for AncFMO2 and AncFMO3-6, and AncFMO5 respectively. In the lower panel, a rotation of approximately 45° was imposed to portray the difference in the opening towards the FAD site.

AncFMOs are thermostable enzymes that are stabilized significantly in the presence of NADP$^+$ and are reliable models for human FMOs

Allegedly, highly thermostable enzymes are highly prone to crystallization.[64] Considering that all three AncFMOs crystallized, melting temperature (T_m) assays were conducted using the ThermoFAD technique,[65] to investigate the thermal stability of AncFMOs compared to human FMOs. Remarkably, our AncFMOs in storage buffer conditions (see Materials & Methods) reached T_ms of 60 °C. Comparing AncFMO3-6 and AncFMO5 with human FMO3 and human FMO5 directly, we observed increases of the T_m of up to +22 and +11 °C respectively (Supplementary Table 3, Supplementary Figure 8).[37] Generally, the differences between the AncFMOs and their respective human equivalents are found dispersed across the protein (Supplementary Figure 9). These patterns of highly distributed and non-systematic amino acid replacements between ancestral and extant enzymes validate the notion that AncFMOs are very reliable models for the human FMOs. Noticeably, at the periphery of the active site, the small Ala232 in human FMO5 is mutated to a bulky Phe232 in AncFMO5 (Figure 5D, lower panel). This substitution may well allow larger substrates in human FMO5.-Intriguingly, Fiorentini et al. documented that NADP$^+$ has no effect on the T_m of human FMO5,[37] a result also observed for human FMO3 (Supplementary Figure 8). However, the melting temperatures of all three AncFMOs increased in the presence of NADP$^+$ by +17, +7 and +4 °C for AncFMO2, AncFMO3-6 and AncFMO5 respectively (Supplementary Table 3). With AncFMOs exhibiting a low degree of uncoupling, it corroborates that tight binding of NADP$^+$ is necessary for highly coupled reactions (Table 1; Figure 1).

Discussion

Our work supports the notion that the number of FMOs in vertebrates significantly increased by successive gene duplication events, leading to the multiple paralogs observed in mammals today.[44] Tetrapods encode for four (amphibians, testudines and birds) or six (mammals) different FMOs, suggesting defined roles for each of these variants. Analyzing the different paralogs, we observed that FMO3 and FMO6 followed a common evolutionary path preceded by the diversification of FMO4, FMO1 and FMO2. FMO5 originated from the earliest gene duplication event and, intriguingly, is encoded by all the aforementioned terrestrial vertebrates' classes. With AncFMOs exhibiting substrate profiles and catalytic rates as their FMO successors, we propose that this class of enzymes have an evolutionary conserved mode of action. Moreover, two new features are derived from ancestral sequence reconstruction: (i) increased melting temperatures and (ii) the stabilizing effect induced by NADP$^+$ (see Supplementary Table 3). With the mutations scattered across the protein, it is unlikely that individual mutations stabilize the enzyme greatly. Their

summation however, enhances stability tremendously. Whether this higher thermal stability of the AncFMOs has a biological meaning remains unclear.[66]

Our research has resulted in the unveiling of the first structures hitherto of mammalian FMOs. Together, they demonstrate the extensive membrane-binding features employed by this enzyme class. Literature had always speculated that the C-terminus was involved in membrane association,[37,54,55,67] but the roles of the large insertions present in human FMOs were ostensibly more enigmatic. The dimerization observed in the crystal structure is not uncommon to membrane proteins and is now attributed to mammalian FMOs.[58,68] Specifically, the oligomerization state aids membrane insertion as the protein occupies a larger membrane-surface area.[68] The inserted residues together form a large monotopic binding feature, which constitutively holds the enzyme in the membrane, ensuring constant uptake and release of substrates and products from and to the membrane. These molecules are then siphoned through the enzyme via a series of tunnels implemented by this subdomain. These routes also open to the cytosolic side of the enzyme structures. Presumably, all FMOs are thereby capable of accepting and expelling soluble compounds from and into the cytosolic solvent as well as lipophilic compounds from and into the membrane bilayer.

With the AncFMOs all accommodating very similar active sites, substrate profiles are likely differentiated by the tunnels penetrating the scaffold. FMO2 is generally known to be slightly more restrictive in terms of substrate size, mostly metabolizing molecules possessing amino groups attached to large aliphatic tails.[2,69] Whilst, FMO3 and FMO1 are understood to be more promiscuous, occupying a breadth of substrate sizes.[2,18,70] The tunnels hereby depicted do not allow us to confidently rationalize these phenomena specifically. For example, the high activity of FMO3 towards trimethylamine likely arises from the combination of subtle factors including the distribution of charged residues, the partition of hydrophobic versus hydrophilic residues at the entrance and inside the FMO3 tunnel, and the flexibility of the residues that gate tunnel access and substrate diffusion. Nevertheless, the overall architecture of the catalytic site and of the access tunnels fully explains the broad substrate scopes of mammalian FMOs. Above all, the catalytic site promotes the flavin-mediated activation of oxygen through the formation of the flavin-(hydro)peroxide intermediate as observed in soluble FMOs as well as in Baeyer-Villiger monooxygenases.[71,72] After flavin reduction, the nicotinamide-ribose moiety of $NADP^+$ relocates to give access to the oxygen-reacting C4a atom of the flavin and thereby promote formation of the flavin-(hydro)peroxide that awaits a substrate to be monooxygenated (Scheme 1). Along these lines, no elements for the specific recognition of the substrate can be identified. The tunnel and the inner catalytic cavity of FMOs are rather designed to allow a “controlled” access to the flavin-

(hydro)peroxide without any strict or rigorous binding selectivity. It can easily be envisioned that the tunnels can adapt themselves depending on the bulkiness of the ligands. The gating elements (e.g. Leu375) may well seal the active-site cavity when small substrates are bound (Figure 5). Likewise, the same elements could enable the binding of bulky molecules whose non-reactive groups extend along the tunnel. Thus, FMOs exhibit typical features of enzymes that broadly function in xenobiotic detoxification. Their preference for nitrogen- and sulfur-containing substrates primarily reflects the pronounced reactivity of the flavin-(hydro)peroxide towards these soft-nucleophiles.

In conclusion, we have unveiled the first mammalian flavin-containing monooxygenase structures through the approach of ancestral sequence reconstruction. Additionally, our work adds to our understanding of the evolutionary history leading to the expansion of FMOs in terrestrial vertebrates. The elucidation of three ancient flavin-containing monooxygenases has allowed us to map the differences between FMOs, providing excellent templates for structure-based drug design. Furthermore, the thermostable but functionally and structurally conserved proteins delivered by this method should be seriously and routinely considered as a tool for protein crystallization.

Materials & Methods

Phylogenetic Inference and Ancestral Sequence Reconstruction. To obtain a robust and representative phylogeny of FMOs, sequences from the Bacteria and Eukarya were collected by homology searches using BLAST and HMM profiling. 310 sequences were collected and aligned with MAFFT v7.[73] Best-fit model parameters were obtained by the Akaike information criterion in ProtTest v3.4. Phylogenies were inferred employing the maximum likelihood method in PhyML v3.0 or RAxML v0.6.0 with 500/1000 bootstraps and transfer bootstrap expectation (TBE) subsequently applied.[74] As FMOs are not monophyletic, derived clades BVMOs and NMOs were included in the phylogeny.[7,9] Later, the Gnathostomata FMOs phylogeny was constructed for ancestral sequence reconstruction. To do this, a dataset of 361 sequences was collected including also a cephalochordate sequence to root the tree according to species tree (Supplementary Data 1).[75] Ancestral sequence reconstruction was performed using the maximum likelihood inference method in PAMLX v.4.9.[43,76] Sequences were analyzed using an empirical amino acid substitution model (model = 3), 4 gamma categories and LG substitution matrix. The posterior probability distribution of ancestral states at each site was analyzed at nodes AncFMO2, AncFMO3-6 and AncFMO5. Sites were considered ambiguously reconstructed if alternative states displayed PP >0.2.[77] Alternative sites were found to be 2 for AncFMO2, 15 for AncFMO3-6 and 11 for AncFMO5 (Supplementary Table 4). These are mostly conservative amino acid substitutions, and after mapping in the crystal

structures, it was evident that they all lay in the periphery of the protein, not affecting the catalytic core.

Cloning and Expression of the AncFMOs. Genes were ordered from Genescript containing BsaI sites at both the 5' and 3' ends of the insert. The insert contained overhangs TGGT and CAAG at the 5' and 3' ends respectively to then be inserted into common pBAD-NK destination vectors with the following modifications: three BsaI sites were eliminated and two were introduced to facilitate the cloning that incorporated SUMO and 8xHis-tag regions to the N-terminus. Inserts were fused into the destination vectors through Golder Gate cloning. The sample was prepared with the following: 75 ng of golden gate entry vector, a molar ratio of 2:1 between insert and vector, BsaI(-HF) (15 U), 30 WU T4 DNA ligase (15 U), T4 DNA ligase buffer (1x) and Nuclease free water added to a final volume of 20 µL. During the cloning procedure, a negative control was prepared with the fragments/inserts omitted. The Golden Gate assembly was conducted in the following manner, where maximum efficiency was desired: A cycle of 37 °C for 5 mins followed by 16 °C for 10 minutes was repeated 30 times; followed by 55 °C for 10 minutes, 65 °C for a further 20 minutes, finishing with 8 °C for 20 minutes. Once cloned, the plasmids were transformed by heat shock into *E. coli*. BL21 cells (25 seconds, 42 °C). Cells from the resulting colonies were pre-inoculated into 100 mL of LB broth containing 100 µg/mL of ampicillin and grown overnight at 37 °C. The cultures were then transferred to 1L Terrific Broth cultures (15 mL) and grown at 24 °C, rpm 180 for 5 – 6 hours until the OD reached 0.3. The cultures were then induced with a sterilized arabinose solution (20% w/v), final concentration of 0.02% (w/v) and incubated at 24 °C, 180 rpm for an additional 24 hours. Cells were then harvested by centrifugation (5000g, 15 mins, 10 °C) flash frozen in liquid nitrogen and stored at -80 °C. For site-directed mutagenesis, a PCR-reaction mixture was prepared with 10 µM primer forward and reverse, 100 ng of template DNA, 1.6 % DMSO, 0.8 mM $MgCl_2$ and 1x *Pfu* Ultra II Hotstart Master Mix (Agilent). The Quickchange PCR cycle was performed using the following method: firstly a 5 minute incubation at 95°C, then cycles (95 °C for 5 minutes - 60 °C for 30 seconds - 72 °C for 6 minutes) were repeated 25 times; followed by 72 °C for 10 minutes and finishing with 8 °C on hold. The PCR mixture was digested with *DpnI* overnight and transformed into *E. coli*.

Cell Disruption, Extraction, and Purification of AncFMOs. All procedures were carried out in ice or at 4 °C. Cells (*ca.* 20 g) were resuspended (1:5) in buffer A (250 mM NaCl, 50 mM KH_2PO_4, pH 7.8) and included additional protease inhibitors: phenylmethylsulfonyl fluoride (1 mM), leupeptin (10 µM), pepstatin (10 µM) and DNase I (5 µg/g of cell paste). The solution was stirred and incubated at 4 °C for 45 minutes before cell lysis was conducted using sonication or a high-pressure homogenizer. Sonication was conducted using the following conditions: 50 mL solution, 5

seconds on, 10 seconds off, 1-minute cycles with a total sonication time of 3 minutes using a microtip (70% amplitude). Cells were passed through a high-pressure homogenizer twice. Lysed cells were then spun down (1200g, 12 mins, 4 °C) to remove the cell debris. The resultant supernatant was then centrifuged further (56,000g, 1 hr. 40 mins, 4 °C) to collect the membrane pellet which was then re-homogenized in buffer A (15 mL,) and centrifuged again (56,000g, 1 hr. 20 mins, 4 °C) to further purify the insoluble material. The resulting pellet was re-homogenized in buffer A (7 mL) and diluted to a final concentration of 13 mg/mL (assayed using Biuret reagent). Triton X-100 Reduced (TRX-100-R) (Sigma-Aldrich) was then added to the solution (0.5% (v/v) final concentration) and mixed overnight at 4 °C. The detergent-solubilized fraction containing the AncFMOs was then abstracted by collecting the supernatant after centrifugation (56,000g, 1 hr. 20 mins, 4 °C). The supernatant was then transferred to a pre-equilibrated (with buffer A and 0.05% (v/v) TRX-100-R) gravity column containing a Ni-resin (GE Healthcare). The supernatant was washed with buffer A, containing 0.05% TRX-100-R, and then with increasing concentrations of buffer B (50 mM KH_2PO_4, 500 mM NaCl, 300 mM imidazole, pH 7.8), also containing 0.05% (v/v) TRX-100-R, in step-by-step fashion: 5 mM imidazole wash, 30 mM imidazole wash and finally a 300 mM imidazole wash, where the protein then eluted. The buffers were then exchanged using a centrifugal filter unit (50 kDa cut-off) and multiple washes with buffer A with 0.05% (v/v) TRX-100-R. This step was important to remove high concentrations of imidazole employed during the elution. The protein sample was then concentrated down to a final volume between 500 and 1000 µL. The sample was then mixed with a 6xHis-tagged SUMO protease (1.2 mg/mL) to a volume ratio of 10:1 respectively and incubated overnight at 4 °C. The sample was then loaded onto an Akta purification system (GE Healthcare) endowed with a multiwavelength detector (set at 280/370/450 nm) and then onto a nickel-affinity His-trap column (GE Healthcare). The column was pre-equilibrated with buffer A containing 0.05% (v/v) TRX-100-R, as stated before, with the proteins eluting out in the presence of 6 mM imidazole derived from buffer B (2%) containing 0.05% (v/v) TRX-100-R. The SUMO-His-tag cleaved protein was then concentrated, and buffer exchanged using a concentrating centrifugal filter unit (50 kDa cut-off) to a final volume between 250 and 500 µL. The sample was incubated for 1 h with 100 µM FAD at 4 °C and then loaded onto a gel filtration column (Superdex 200 10/300, GE Healthcare) pre-equilibrated with a storage buffer (50 mM Tris-HCl pH 8.5 at 4 °C, 10 mM NaCl) and a detergent of choice to obtain a higher degree of purity (obtained from anatrace). Typically DDM was used (0.03% (w/v) DDM (analytical grade)), but other detergents were used for crystallization screenings at 3x their respective Critical-Micelle Concentration (CMC). The protein eluted with a very high purity and homogeneity (evaluated by SDS-page electrophoresis and the shape of the peak in the chromatogram

628 respectively) with an elution volume of 10.5 – 11 mL. The sample was concentrated down to 100
629 µL using a centrifugal filter unit (50 kDa cut-off) with a final concentration ranging from 5 to 30
630 mg/mL.
631 **Preparation of human FMO3 and human FMO5.** Full-length cDNA encoding for *Homo sapiens*
632 FMO3 (UniProt P31513) and FMO5 (NCBI accession number: Z47553) were cloned into a
633 modified pET-SUMO vector (Invitrogen) to allow insertion of a cleavable N-terminal 8xHis-
634 SUMO tag. Expression, cell disruption, extraction, and purification were performed according to
635 the methods previously described for human FMO5.[37]
636 **Kinetic Assays of the AncFMOs**. Steady-state kinetics assays were performed in duplicates on a
637 Jasco V-660 spectrophotometer. Enzyme activity of the ancestral proteins was measured by
638 monitoring NADPH consumption (absorbance at 340 nm, ε_{340} = 6.22 mM^{-1} cm^{-1} for NADPH). The
639 buffer used for kinetic analyses was 50 mM potassium phosphate, 250 mM NaCl, 0.05% TRX-100-
640 R (Sigma-Aldrich), pH 7.5. The reaction volumes were 100 µL and contained variable NADPH and
641 substrate (methimazole, thioanisole, trimethylamine, heptanone) concentrations and 0.01-2.0 µM
642 enzyme. The spectrophotometer was set at 37 °C and the NADPH and substrate mix were also
643 incubated at 37 °C for 5 minutes before starting the reaction by adding enzyme. The pH and
644 temperature conditions were set based on literature studies for a fair comparison with previously
645 reported properties of mammalian FMOs.
646 **Kinetic Assays of human FMO3**. Kinetic assays were performed in duplicate at 25 °C on a Varian
647 spectrophotometer (Cary 100 Bio) equipped with a thermostatic cell compartment. The apparent K_M
648 for NADPH was measured by varying NADPH concentrations from 20 to 400 µM in aerated
649 reaction mixtures (150 µL) containing 2.5 µM human FMO3, 50 mM HEPES pH 7.5, 10 mM KCl,
650 5% (v/v) glycerol, 0.05% (v/v) TRX-100-R. Reactions were followed by monitoring the decrease of
651 NADPH concentration (decrease in absorbance) at 340 nm (ε_{340} = 6.22 mM^{-1} cm^{-1} for NADPH).
652 **Rapid Kinetics Analysis of the AncFMOs.** Stopped flow experiments were carried out using the
653 SX20 stopped flow spectrophotometer equipped with either the photodiode array detector or the
654 single channel photomultiplier (PMT) module (Applied Photophysics, Surrey, UK). Results were
655 obtained by mixing 50 µL of two solutions in single mixing mode. All solutions were prepared in
656 50 mM potassium phosphate, 10 mM NaCl and 0.05% TRX-100-R, pH 7.5 at 25 °C. For every
657 reaction a concentration of 10-15 µM enzyme was used and measurements were done in duplicate.
658 When needed, the solutions were supplemented with 5.0 mM glucose and the enzyme and/or
659 substrate solutions were made anaerobic by flushing solutions for 10 minutes with nitrogen,
660 followed by adding 0.3 µM glucose oxidase (*Aspergillus niger*, type VII, Sigma-Aldrich) to
661 consume the left-over oxygen. The monitoring of the reductive half reaction was done by mixing

the anaerobic enzyme solution with anaerobic buffer containing increasing concentrations of NAD(P)H (Oriental Yeast Co. LTD.). The rate of flavin reduction was determined by measuring the decrease of absorbance at 447 nm or 442 nm for AncFMO2 and AncFMO3-6, respectively. In order to reduce the flavin cofactor in the FMOs for the oxidative half reaction, NADPH was added to an anaerobic solution containing an equivalent amount of AncFMO. The resulting solution was incubated on ice until the bleaching of the FAD was complete, indicating complete reduction to $FADH_2$. The anaerobically reduced FMOs were mixed with air-saturated buffer first and then with air-saturated buffer containing 1.0 mM or 0.4 mM of trimethylamine, respectively. This allowed us to follow the spectral changes during the oxidative half-reactions. The C4alpha-(hydro)peroxyflavin intermediate formation and decay was specifically monitored using the PMT module at 360 nm. For determining the rates of reoxidation, the reduced enzymes were mixed with buffers containing different concentrations of dioxygen. The final concentrations of dioxygen (0.13, 0.31, 0.61, 0.96 mM after mixing) were achieved by mixing the anaerobic enzyme solution with (i) air-saturated buffer; (ii) mixing equal volumes of 100% argon buffer and 100% O_2 buffer; (iii) 100% O_2 buffer; (iv) 100% O_2 buffer on ice. All solutions were bubbled for ten minutes at room temperature except the last one which was done on ice. In the case of AncFMO2, in order to confirm its saturating behaviour, an additional measurement was performed at 0.816 mM of O_2 by mixing 100% O_2 buffer on ice with 100% argon buffer.

Observed rates (k_{obs}) were determined by fitting traces to exponential functions. All data were analysed using the software Pro-Data (Applied Photophysics, Surrey, UK) and GraphPad Prism 6.05 (La Jolla, CA, USA).

ThermoFAD Assays.[65] A Bio-Rad MiniOpticon Real-Time PCR System was employed to perform ThermoFAD screenings (temperature gradient 25−70 °C, fluorescence detection every 0.5 °C at 485 ± 30 nm excitation and 625 ± 30 nm emission for 5 s). Concentrations were determined using a molar extinction coefficient of 12 mM^{-1} cm^{-1} for the FAD band at 442 nm. Experiments were performed in triplicate using human FMO5 and the AncFMOs, in the presence or absence of $NADP^+$. Each sample contained the protein of interest (4 µM), and with or without $NADP^+$ (200 µM), made to a final volume of 20 µL using the storage buffer and incubated in ice for one hour. Melting temperatures for human FMO3 (0.05% (v/v) TRX-100-R) were performed in duplicates and diluted to a final concentration of 5 µM in buffer (100 mM) with varying pH values (pH 6-6.5 MES, pH 7-7.5-8 HEPES, pH 8-8.5-9 Bicine, pH 9.5 CHES), KCl concentrations (0-500 mM in HEPES pH 8), and $NADP^+$ concentrations (5-500 µM in HEPES pH 8, 10 mM KCl, 0.05% (v/v) TRX-100-R) in an attempt to generate optimal storage buffer conditions.

Conversions. Conversions were performed using AncFMO2 and AncFMO3-6 and their respective E281H mutants (Supplementary Table 2). Reaction mixtures (1.0 mL) contained 5.0 mM substrate (<1% EtOH), 0.10 mM NADPH, 2.0 µM enzyme, 5.0 µM phosphite dehydrogenase, 20 mM phosphite, 50 mM KPi pH 7.5, 250 mM NaCl and 0.05% (v/v) TRX-100-R. The mixtures were incubated for 18 hours at 30 °C and subsequently extracted with 1.0 mL ethyl acetate. The organic phase was passed through anhydrous sulfate magnesium to remove residual water. Analysis was carried out using a GCMS-QP2010 Ultra (Shimadzu) equipped with a HP-1 column, using electron ionization MS detection.

Crystallization and Structural Determination of the AncFMOs. Each AncFMO crystallized in a range of conditions with multiple detergents. Typically, PEG 4000 was optimal for crystallization. The highest diffracting crystallization conditions for each AncFMO are described below. AncFMO2 (with and without $NADP^+$): 12-15 mg/mL of AncFMO2 (in storage buffer and CYMAL-6 (0.09% (w/v)) was incubated with crystallization conditions of HEPES buffer (0.1 M, pH 7.5) and PEG 4000 (10%) at 20 °C with a ratio of 1:1 on a sitting drop. Sitting drop was 2 µL after mixing and contained a reservoir of 1 mL. Prior to crystallization, $NADP^+$ (1 mM final) was incubated with 12-15 mg/mL of AncFMO2 for 1 hour at 4 °C. After two days, large yellow crystals formed. AncFMO3-6: 12-15 mg/mL of AncFMO3-6 (in storage buffer and DDM (0.03% (w/v)) was incubated with crystallization conditions of Sodium Acetate buffer (0.1 M, pH 5.5) and PEG 4000 (7.5%) at 20 °C with a ratio of 1:1 on a sitting drop. Sitting drop was 2 µL after mixing and contained a reservoir of 1 mL. Prior to crystallization, $NADP^+$ (1 mM final) was incubated with 12-15 mg/mL of AncFMO3-6 for 1 hour at 4 °C. After one day, large yellow crystals formed. AncFMO5: 12 mg/mL of AncFMO5 (in storage buffer and DDM (0.03% (w/v)) was incubated with crystallization conditions of HEPES buffer (0.1 M, pH 6.9) and PEG 4000 (9%) at 20 °C with a ratio of 1:1 on a sitting drop. Sitting drop was 2 µL after mixing and contained a reservoir of 1 mL. Prior to crystallization, $NADP^+$ (1 mM final) was incubated with 12-15 mg/mL of AncFMO5 for 1 hour at 4 °C. After one day, large yellow hexagon shaped crystals formed. During crystal fishing, a cryo-protectant was prepared containing modified crystallizations conditions with 20% glycerol and PEG 4000 (15%).

Data were collected at the European Synchrotron Radiation Facility (Grenoble, France) and the Swiss Light Source, (Villigen, Switzerland) and processed with the XDS[78] and CCP4 packages.[79] Aimless was used to merge the observations into average densities (Supplementary Table 1). STARANISO was additionally used for AncFMO2 which suffered greatly from anisotropy.[80,81] The phase problem was solved by Molecular Replacement using a recently solved insect FMO (PDB:5nmw)[82] as a search model, and then AncFMO2 for the proceeding AncFMOs, using Phaser

and Molrep.[83] The phases were greatly improved by density averaging with DM.[79,84] Model Building and Refinement were then conducted using COOT,[85] Buccaneer,[79] and Refmac5[86] respectively (Supplementary Table 1). Figures were then generated using UCSF Chimera,[87] PyMOL (DeLano Scientific; www.pymol.org) and CCP4mg.[79]

Acknowledgements

The research for this work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 722390 and ANPCyT (Argentina) PICT 2016-2839 to MLM. MLM is a member of the Research Career of CONICET, Argentina. Authors thank Dr. Maximiliano Juri Ayub for his valuable comments and contributions on the manuscript

Author Contributions

All listed authors performed experiments and analyzed data: CRN generated purification protocols, crystallized the AncFMOs, collected the corresponding datasets at ESRF and SLS facilities, performed structural analysis and elucidated the AncFMO structures. GB and CRN performed golden gate cloning to insert the AncFMO genes into their respective vectors, designed by GB. GB carried out mutagenesis and extensive kinetic analysis and validated the substrate profiles using stopped-flow UV/Vis spectroscopy and GCMS for each AncFMO. MLM conducted thorough evolutionary analyses and performed ancestral sequence reconstruction obtaining AncFMO protein sequences. CRN, GB and MLM prepared the figures. CRN wrote the manuscript and edited by AM, MWF and MLM. All authors provided critical feedback and helped shape the research, analysis and manuscript. AM and MWF conceived the original idea.

Declaration of Interest

The authors declare no conflict of interest.

Data availability

Coordinates and structure factors have been deposited with the Protein Data Bank with accession codes 6SEM (AncFMO2), 6SF0 (AncFMO2 in complex with $NADP^+$), 6SE3 (AncFMO3-6), 6SEK (AncFMO5). All other data is available from the authors upon reasonable request.

References

1. Jakoby, W. B. & Ziegler, D. M. The enzymes of detoxication. *J. Biol. Chem.* **265**, 20715–20718 (1990).

2. Krueger, S. K. & Williams, D. E. Mammalian flavin-containing monooxygenases: Structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacol. Ther.* **106**, 357–387 (2005).
3. Cruciani, G. *et al.* Flavin monooxygenase metabolism: Why medicinal chemists should matter. *J. Med. Chem.* **57**, 6183–6196 (2014).
4. Huijbers, M. M. E., Montersino, S., Westphal, A. H., Tischler, D. & Van Berkel, W. J. H. Flavin dependent monooxygenases. *Arch. Biochem. Biophys.* **544**, 2–17 (2014).
5. Ziegler, D. M. Flavin-containing monooxygenases: enzymes adapted for multisubstrate specificity. *Trends Pharmacol. Sci.* **11**, 321–324 (1990).
6. Cashman, J. R. Some distinctions between flavin-containing and cytochrome P450 monooxygenases. *Biochem. Biophys. Res. Commun.* **338**, 599–604 (2005).
7. Mascotti, M. L., Lapadula, W. J. & Ayub, M. J. The origin and evolution of Baeyer - Villiger Monooxygenases (BVMOs): An ancestral family of flavin monooxygenases. *PLoS One* **10**, 1–16 (2015).
8. D. M. Ziegler & Pettit, F. H. Formation of an intermediate N-oxide in the oxidative demethylation of N,N-dimethylanaline catalyzed by liver microsomes. *Biochem. Biophys. Res. Commun*. **15**, 188–193 (1964).
9. Mascotti, M. L., Juri Ayub, M., Furnham, N., Thornton, J. M. & Laskowski, R. A. Chopping and Changing: the Evolution of the Flavin-dependent Monooxygenases. *J. Mol. Biol.* **428**, 3131–3146 (2016).
10. Cashman, J. R. & Zhang, J. Human Flavin-Containing Monooxygenases. *Annu. Rev. Pharmacol. Toxicol.* **46**, 65–100 (2006).
11. Romero, E., Castellanos, J. R. G., Gadda, G., Fraaije, M. W. & Mattevi, A. The same substrate, many reactions: oxygen activation in flavoenzymes. *Chem. Rev*. **118**, 1742-1769 (2017).
12. Alfieri, A., Malito, E., Orru, R., Fraaije, M. W. & Mattevi, A. Revealing the moonlighting role of NADP in the structure of a flavin-containing monooxygenase. *Proc. Natl. Acad. Sci.* **105**, 6572–6577 (2008).
13. Zhang, J. QUANTITATIVE ANALYSIS OF FMO GENE mRNA LEVELS IN HUMAN TISSUES. *Drug Metab. Dispos.* **34**, 19–26 (2005).
14. Dolphin, C. T., Cullingford, T. E., Shephard, E. A., Smith, R. L. & Phillips, I. R. Differential developmental and tissue-specific regulation of expression of the genes encoding three members of the flavin-containing monooxygenase family of man, FMO1, FMO3 and FMO4. *Eur. J. Biochem.* **235**, 683–689 (1996).
15. McCombie, R. R., Dolphin, C. T., Povey, S., Phillips, I. R. & Shephard, E. A. Localization of human flavin-containing monooxygenase genes FMO2 and FMO5 to chromosome 1q. *Genomics* **34**, 426–429 (1996).
16. Hernandez, D., Janmohamed, A., Chandan, P., Phillips, I. R. & Shephard, E. A. Organization and evolution of the flavin-containing monooxygenase genes of human and mouse. *Pharmacogenetics* **14**, 117–130 (2004).
17. Fiorentini, F. *et al.* Baeyer-Villiger Monooxygenase FMO5 as Entry Point in Drug Metabolism. *ACS Chem. Biol.* **12**, 2379–2387 (2017).
18. Poulsen, L. L. & Ziegler, D. M. Multisubstrate flavin-containing monooxygenases: applications of mechanism to specificity. *Chem. Biol. Interact.* **96**, 57–73 (1995).
19. Henderson, M. C., Krueger, S. K., Siddens, L. K., Stevens, J. F. & Williams, D. E. S-oxygenation of the thioether organophosphate insecticides phorate and disulfoton by human lung flavin-containing monooxygenase 2. *Biochem. Pharmacol.* **68**, 959–967 (2004).
20. Cashman, J. R. Role of flavin-containing monooxgenase in drug development. *Expert Opn. Drug Metab. Toxicol.* **4**, 1507–1521 (2008).
21. Siddens, L. K., Krueger, S. K., Henderson, M. C. & Williams, D. E. Mammalian flavin-containing monooxygenase (FMO) as a source of hydrogen peroxide. *Biochem. Pharmacol.*

89, 141–147 (2014).
22. Li, C. Y. *et al.* Structural mechanism for bacterial oxidation of oceanic trimethylamine into trimethylamine N-oxide. *Mol. Microbiol.* **103**, 992–1003 (2017).
23. Miao, J. *et al.* Flavin-containing monooxygenase 3 as a potential player in diabetes-associated atherosclerosis. *Nat. Commun.* **6**, 1–10 (2015).
24. Krueger, S. K. *et al.* GENETIC POLYMORPHISMS OF FLAVIN- CONTAINING MONOOXYGENASE (FMO). **34**, 523–532 (2002).
25. Dolphin, C. T., Janmohamed, A., Smith, R. L., Shephard, E. A. & Phillips, I. R. Compound heterozygosity for missense mutations in the flavin-containing monooxygenase 3 (FM03) gene in patients with fish-odour syndrome. *Pharmacogenetics* **10**, 799–807 (2000).
26. Veeravalli, S. *et al.* Effect of flavin-containing monooxygenase genotype, mouse strain, and gender on trimethylamine N-oxide production, plasma cholesterol concentration, and an index of atherosclerosis. *Drug Metab. Dispos.* **46**, 20–25 (2018).
27. Phillips, I. R. & Shephard, E. A. Flavin-containing monooxygenases: mutations, disease and drug response. *Trends Pharmacol. Sci.* **29**, 294–301 (2008).
28. Hernandez, D. *et al.* Trimethylaminuria and a human FMO3 mutation database. *Hum. Mutat.* **22**, 209–213 (2003).
29. Dolphin, C. T., Janmohamed, A., Smith, R. L., Shephard, E. A. & Phillips, lan R. Missense mutation in flavin-containing mono-oxygenase 3 gene, FMO3, underlies fish-odour syndrome. *Nat. Genet.* **17**, 491–494 (1997).
30. Lang, D. H. *et al.* Isoform specificity of trimethylamine N-oxygenation by human flavin-containing monooxygenase (FMO) and P450 enzymes Selective catalysis by fmo3. *Biochem. Pharmacol.* **56**, 1005–1012 (1998).
31. Hodgson, E., Rose, R. L., Cao, Y., Dehal, S. S. & Kupfer, D. Flavin-containing monooxygenase isoform specificity for the N-oxidation of tamoxifen determined by product measurement and NADPH oxidation. *J. Biochem. Mol. Toxicol.* **14**, 118–120 (2000).
32. Mushiroda, T., Douya, R., Takahara, E. & Nagata, O. The involvement of flavin-containing monooxygenase but not CYP3A4 in metabolism of itopride hydrochloride, a gastroprokinetic agent: Comparison with cisapride and mosapride citrate. *Drug Metab. Dispos.* **28**, 1231–1237 (2000).
33. Shephard, E. A. & Phillips, I. R. The potential of knockout mouse lines in defining the role of flavin-containing monooxygenases in drug metabolism. *Expert Opin. Drug Metab. Toxicol.* **6**, 1083–1094 (2010).
34. Krueger, S. K., VanDyke, J. E., Williams, D. E. & Hines, R. N. The role of flavin-containing monooxygenase (FMO) in the metabolism of tamoxifen and other tertiary amines. *Drug Metab. Rev.* **38**, 139–147 (2006).
35. Veeramah, K. R. *et al.* The potentially deleterious functional variant flavin-containing monooxygenase 2*1 is at high frequency throughout sub-Saharan Africa. *Pharmacogenet. Genomics* **18**, 877–886 (2008).
36. Dolphin, C. T. *et al.* The flavin-containing monooxygenase 2 gene (FMO2) of humans, but not of other primates, encodes a truncated, nonfunctional protein. *J. Biol. Chem.* **273**, 30599–30607 (1998).
37. Fiorentini, F. *et al.* Biocatalytic Characterization of Human FMO5: Unearthing Baeyer-Villiger Reactions in Humans. *ACS Chem. Biol.* **11**, 1039–1048 (2016).
38. Leiser, S. F. *et al.* Promotes Longevity and Health Span. *Science.* **350**, 1375–1378 (2015).
39. Warrier, M. *et al.* The TMAO-Generating Enzyme Flavin Monooxygenase 3 Is a Central Regulator of Cholesterol Balance. *Cell Rep.* **10**, 326–338 (2015).
40. Gonzalez Malagon, S. G. *et al.* The phenotype of a knockout mouse identifies flavin-containing monooxygenase 5 (FMO5) as a regulator of metabolic ageing. *Biochem. Pharmacol.* **96**, 267–277 (2015).
41. Scott, F. *et al.* Identification of flavin-containing monooxygenase 5 (FMO5) as a regulator of

glucose homeostasis and a potential sensor of gut bacteria. *Drug Metab. Dispos.* **45**, 982–989 (2017).
42. Risso, V. A., Sanchez-Ruiz, J. M. & Ozkan, S. B. Biotechnological and protein-engineering implications of ancestral protein resurrection. *Curr. Opin. Struct. Biol.* **51**, 106–115 (2018).
43. Hochberg, G. K. A. & Thornton, J. W. Reconstructing Ancient Proteins to Understand the Causes of Structure and Function. *Annu. Rev. Biophys.* **46**, 247–269 (2017).
44. Hao, D. C., Chen, S. L., Mu, J. & Xiao, P. G. Molecular phylogeny, long-term evolution, and functional divergence of flavin-containing monooxygenases. *Genetica* **137**, 173–187 (2009).
45. Hines, R. N. Alternative Processing of the Human FMO6 Gene Renders Transcripts Incapable of Encoding a Functional Flavin-Containing Monooxygenase. *Mol. Pharmacol.* **62**, 320–325 (2002).
46. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
47. Grantham, R. Amino acid difference formula to help explain protein evolution. *Science.* **185**, 862–864 (1974).
48. Krueger, S. K. *et al.* Characterization of sulfoxygenation and structural implications of human flavin-containing monooxygenase isoform 2 (FMO2.1) variants S195L and N413K. *Drug Metab. Dispos.* **37**, 1785–1791 (2009).
49. Overby, L. H., Carver, G. C. & Philpot, R. M. Quantitation and kinetic properties of hepatic microsomal and recombinant flavin-containing monooxygenases 3 and 5 from humans. *Chem. Biol. Interact.* **106**, 29–45 (1997).
50. Ripp, S. L., Itagaki, K., Philpot, R. M. & Elfarra, A. A. Methionine S-oxidation in human and rabbit liver microsomes: Evidence for a high-affinity methionine S-oxidase activity that is distinct from flavin-containing monooxygenase 3. *Arch. Biochem. Biophys.* **367**, 322–332 (1999).
51. Lin, J. & Cashman, J. R. N-oxygenation of phenethylamine to the trans-oxime by adult human liver flavin-containing monooxygenase and retroreduction of phenethylamine hydroxylamine by human liver microsomes. *J. Pharmacol. Exp. Ther.* **282**, 1269–1279 (1997).
52. Beaty, N. B. & Ballou, D. P. The reductive half-reaction of liver microsomal FAD-containing monooxygenase. *J. Biol. Chem.* **256**, 4611–4618 (1981).
53. Krissinel, E. Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Res.* **43**, W314–W319 (2015).
54. Geier, M. *et al.* Human FMO2-based microbial whole-cell catalysts for drug metabolite synthesis. *Microb. Cell Fact.* **14**, 1–10 (2015).
55. Korsmeyer, K. K. *et al.* N-glycosylation of pig flavin-containing monooxygenase form 1: Determination of the site of protein modification by mass spectrometry. *Chem. Res. Toxicol.* **11**, 1145–1153 (1998).
56. Wierenga, R. K., De Maeyer, M. C. H. & Hoi, W. G. J. Interaction of Pyrophosphate Moieties with α-Helixes in Dinucleotide Binding Proteins. *Biochemistry* **24**, 1346–1357 (1985).
57. Orru, R., Torres Pazmiño, D. E., Fraaije, M. W. & Mattevi, A. Joint functions of protein residues and NADP(H) in oxygen activation by flavin-containing monooxygenase. *J. Biol. Chem.* **285**, 35021–35028 (2010).
58. Lončar, N. *et al.* Characterization of a thermostable flavin-containing monooxygenase from Nitrincola lacisaponensis (NiFMO). *Appl. Microbiol. Biotechnol.* **103**, 1755–1764 (2019).
59. Fürst, M. J., Fiorentini, F. & Fraaije, M. W. Beyond active site residues: overall structural dynamics control catalysis in flavin-containing and heme-containing monooxygenases. *Curr. Opin. Struct. Biol.* **59**, 29–37 (2019).
60. Romero, E., Castellanos, J. R. G., Mattevi, A. & Fraaije, M. W. Characterization and Crystal Structure of a Robust Cyclohexanone Monooxygenase. *Angew. Chemie - Int. Ed.* **55**, 15852–

15855 (2016).
61. Torres Pazmiño, D. E., Baas, B. J., Janssen, D. B. & Fraaije, M. W. Kinetic mechanism of phenylacetone monooxygenase from *Thermobifida fusca*. *Biochemistry* **47**, 4082–4093 (2008).
62. Criegee, R. Die Umlagerung der Dekalin-peroxydester als Folge von kationischem Sauerstoff. *Justus Liebigs Ann. Chem.* **560**, 127–135 (1948).
63. Ho, B. K. & Gruswitz, F. HOLLOW: Generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* **8**, 1–6 (2008).
64. Gumulya, Y. *et al.* Engineering highly functional thermostable proteins using ancestral sequence reconstruction. *Nat. Catal.* **1**, 878–888 (2018).
65. Forneris, F., Orru, R., Bonivento, D., Chiarelli, L. R. & Mattevi, A. ThermoFAD, a Thermofluor®-adapted flavin ad hoc detection system for protein folding and ligand binding. *FEBS J.* **276**, 2833–2840 (2009).
66. Wheeler, L. C., Lim, S. A., Marqusee, S. & Harms, M. J. The thermostability and specificity of ancient proteins. *Curr. Opin. Struct. Biol.* **38**, 37–43 (2016).
67. Lawton, M. P. & Philpot, R. M. Functional characterization of flavin-containing monooxygenase 1B1 expressed in Saccharomyces cerevisiae and Escherichia coli and analysis of proposed FAD- and membrane-binding domains. *J. Biol. Chem.* **268**, 5728–5734 (1993).
68. Allen, K. N., Entova, S., Ray, L. C. & Imperiali, B. Monotopic Membrane Proteins Join the Fold. *Trends Biochem. Sci.* **44**, 7–20 (2019).
69. Nagata, T., Williams, D. E. & Ziegler, D. M. Substrate Specificities of Rabbit Lung and Porcine Liver Flavin-Containing Monooxygenases: Differences due to Substrate Size. *Chem. Res. Toxicol.* **3**, 372–376 (1990).
70. Young Mi Kim & Ziegler, D. M. Size limits of thiocarbamibes accepted as substrates by human flavin-containing monooxygenase 1. *Drug Metab. Dispos.* **28**, 1003–1006 (2000).
71. Beaty, N. B. & Ballou, D. P. Transient Kinetic Study of Liver. *J. Biol. Chem.* **255**, 3817–3819 (1980).
72. Beaty, N. B. & Ballou, D. P. The Oxidative Half-reaction of Liver Microsomal FAD-containing Monooxygenase. *J. Biol. Chem.* **256**, 4619–4625 (1981).
73. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 1–7 (2017).
74. Lemoine, F. *et al.* Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* **556**, 452–456 (2018).
75. Hug, L. A. *et al.* A new view of the tree of life. *Nat. Microbiol.* **1**, 1–6 (2016).
76. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
77. Siddiq, M. A., Loehlin, D. W., Montooth, K. L. & Thornton, J. W. Experimental test and refutation of a classic case of molecular adaptation in Drosophila melanogaster. *Nat. Ecol. Evol.* **1**, 1–6 (2017).
78. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry land cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
79. Project, C. C. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50**, 760–763 (1994).
80. Tickle, I. J. *et al.* Staraniso. *Glob. Phasing Ltd* (2018).
81. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
82. Kubitza, C. *et al.* Crystal structure of pyrrolizidine alkaloid N-oxygenase from the grasshopper *Zonocerus variegatus*. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 422–432 (2018).
83. McCoy, A. J. *et al.* Phaser crystallographic software . *J. Appl. Crystallogr.* **40**, 658–674

(2007).
84. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2256–2268 (2004).
85. Jung, W. S., Singh, R. K., Lee, J. K. & Pan, C. H. Crystal Structure and Substrate Specificity of D-Galactose-6-Phosphate Isomerase Complexed with Substrates. *PLoS One* **8**, 2–11 (2013).
86. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53**, 240–255 (1997).
87. Pettersen, E. F. *et al.* UCSF Chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–12 (2004).

Supplementary Information

Ancestral Sequence Reconstruction Unveils the Structural Basis of Catalysis and Membrane Binding in Mammalian Flavin-Containing Monooxygenases

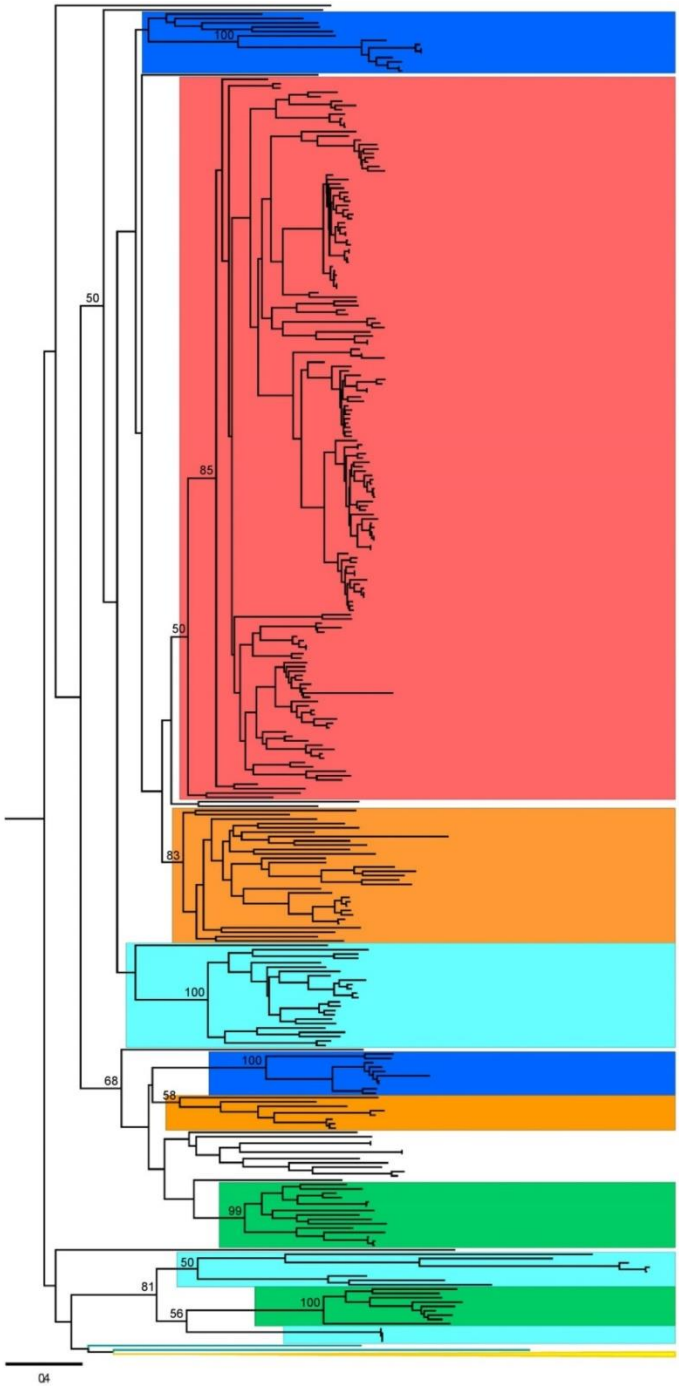
Callum R. Nicoll[1], Gautier Bailleul[2], Filippo Fiorentini[1],
María Laura Mascotti[3,*], Marco W. Fraaije[2,*] and Andrea Mattevi[1,*]

[1] Department of Biology and Biotechnology "Lazzaro Spallanzani", University of Pavia, via Ferrata 9, 27100 Pavia, Italy

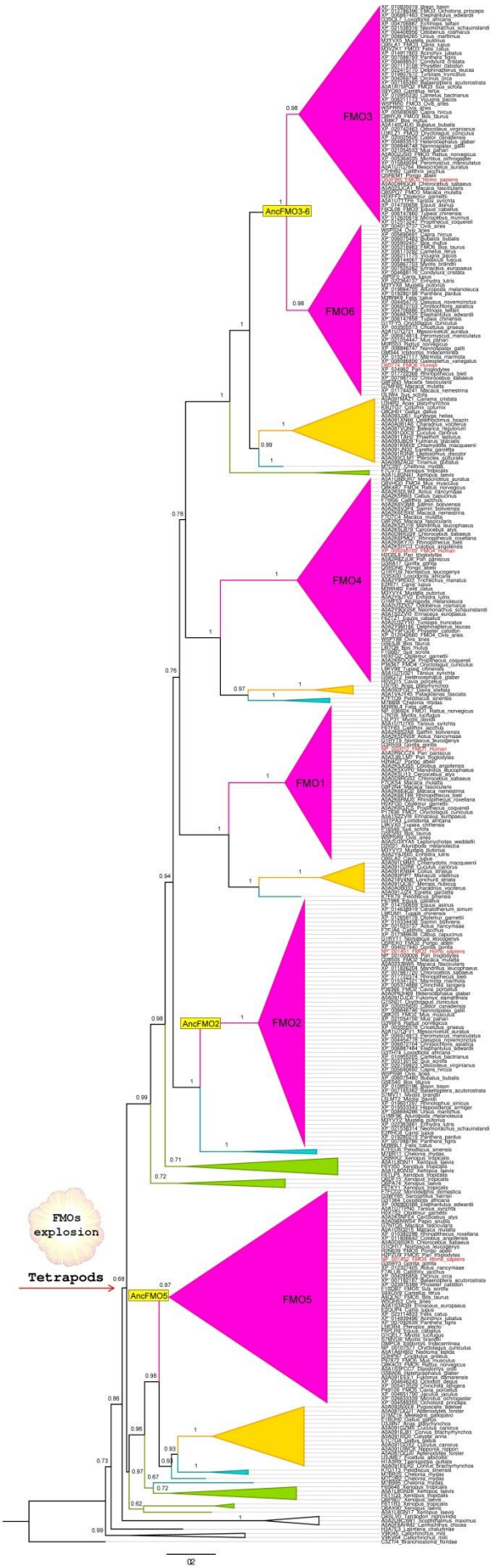
[2] Laboratory of Biochemistry, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

[3] IMIBIO-SL CONICET, Facultad de Química Bioquímica y Farmacia, Universidad Nacional de San Luis, Ejército de los Andes 950, San Luis D5700HHW, Argentina

[*]Correspondence to María Laura Mascotti, Marco Fraaije, Andrea Mattevi
E-mail: mlmascotti@unsl.edu.ar, m.w.fraaije@rug.nl, andrea.mattevi@unipv.it

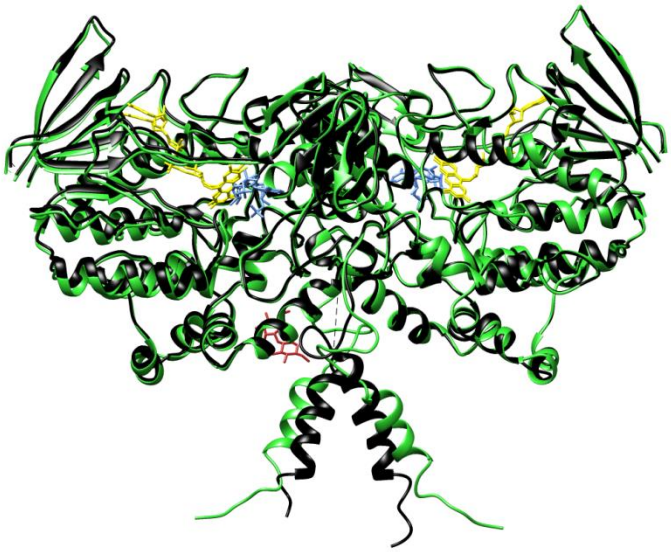


Supplementary Figure 1: Full phylogeny of FMOs. Tree was constructed in PhyML v3.0, 500 bootstraps were run, and best-fit model parameters were obtained with ProtTest v3.4. The employed multiple sequence alignement was trimmed in single sequence extensions and contained 341 taxa and 501 sites. Tree topology was analyzed according to Class B flavoprotein monooxygenases phylogeny. As FMOs are paraphyletic (*i.e*. they do not display a single origin), derived clades are shown collapsed: Baeyer-Villiger monooxygenases (yellow) and N-hydroxylating monooxygenases (deep teal). FMOs are shown in black branches and monophyletic clades according to organism classes are highlighted as follows: Bacteria (cyan), plants (green), vertebrates (red), arthropods (orange), and fungi (blue). Statistical support (BS >50) at major divergence points is labeled at the nodes. The tree was prepared in Figtree v1.4.2.

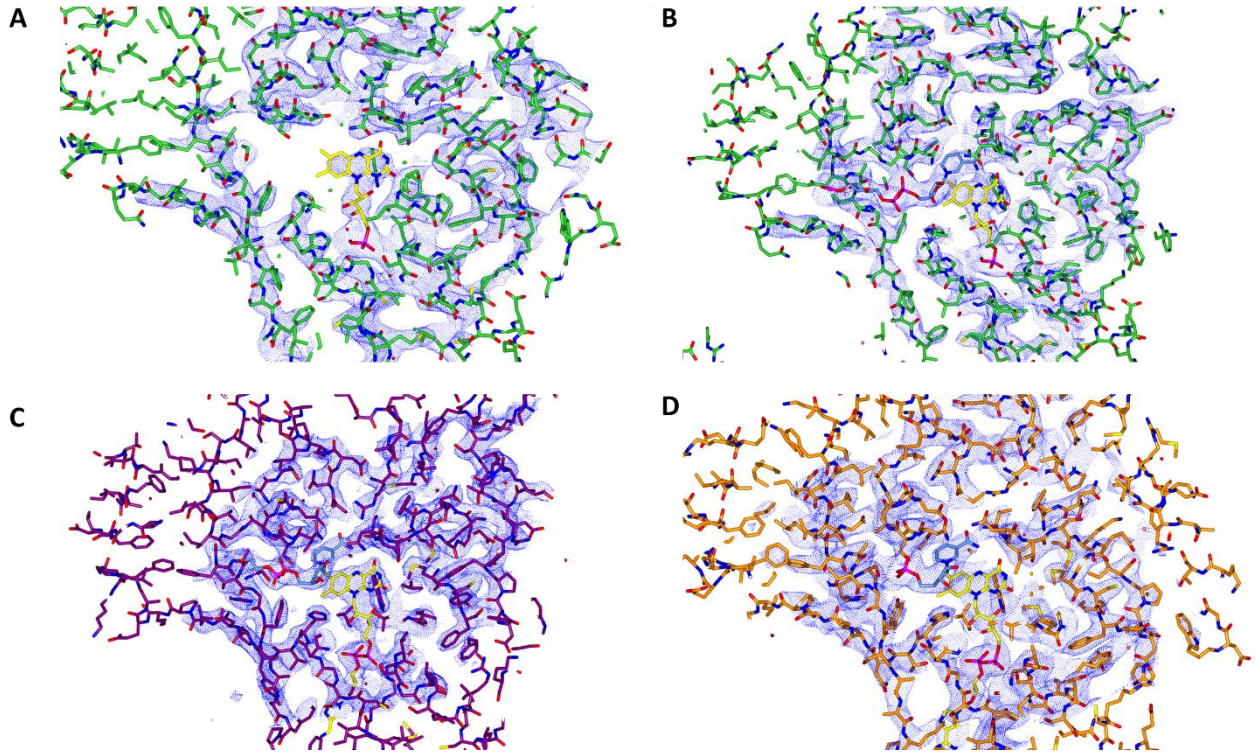

FMO3
AncFMO3-6
FMO6
FMO4
FMO1
AncFMO2
FMO2
FMOs
explosion
Tetrapods
AncFMO5
FMO5
0.98
1
0.98
1
1
0.99
1
0.78
1
1
0.76
0.97
1
1
1
0.94
1
0.98
1
0.99
1
0.71
0.72
0.68
0.97
0.98
0.86
0.96
1
0.93
0.91
1
0.73
0.97
0.67
0.72
0.62
1
0.99
02

1013

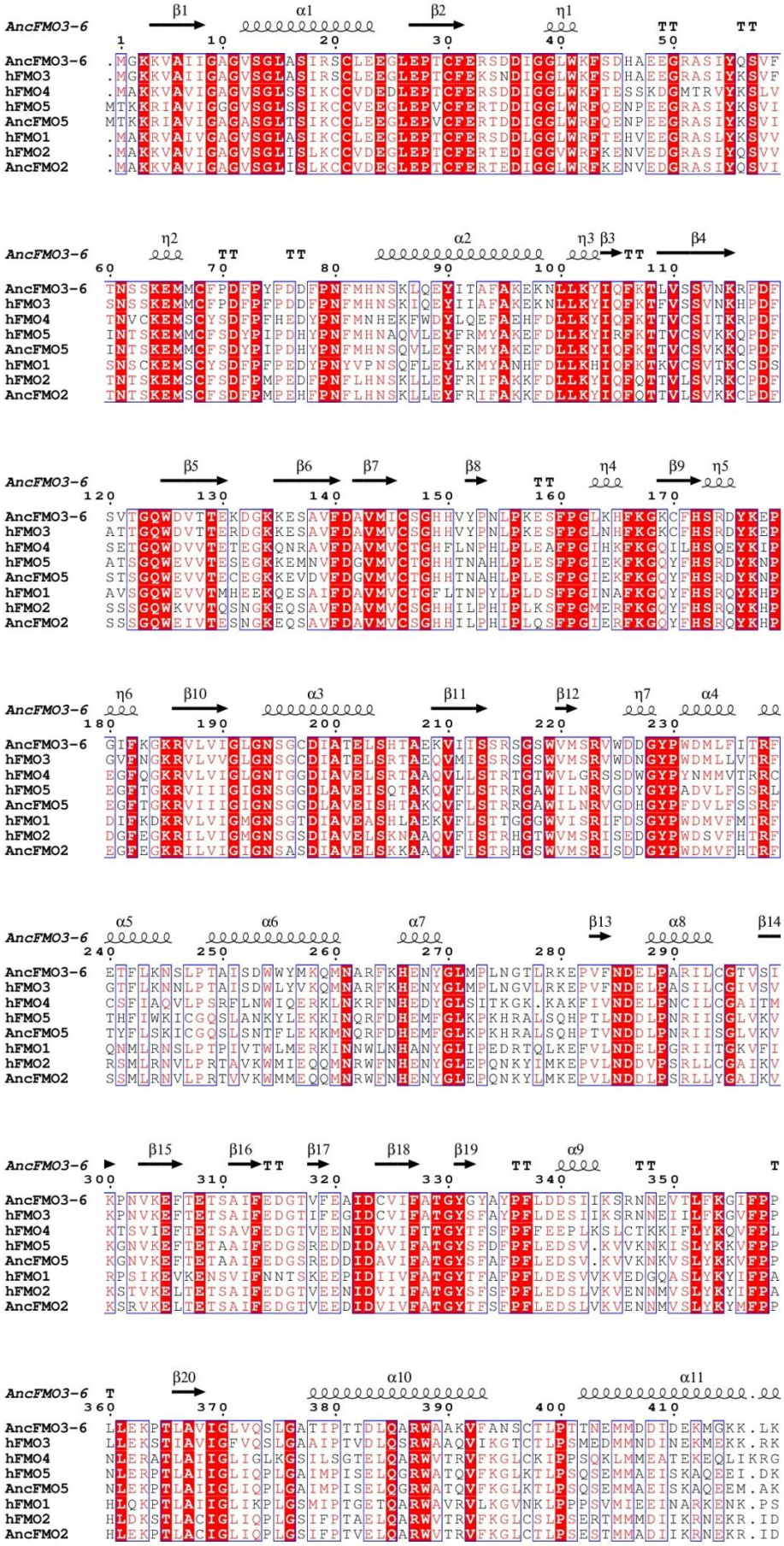
Supplementary Figure 2: Vertebrates phylogeny of FMOs. Tree was constructed in RAxML v0.6.0, 1000 bootstraps were run and best-fit model parameters were obtained with ProtTest v3.4. Bootstrap transfer was applied and TBE values are shown at the nodes. The employed MSA was trimmed in single sequence extensions and contained 361 taxa and 569 sites. Clades are collapsed and colored according to tetrapods classes: mammalia (magenta), aves (light orange), amphibia (green) and testudines (teal). Actinopterygii, coelacanthimorpha and chondrichthyes (all jawed vertebrates) and the cephalochordate sequence in the root are shown in black. Mammalian FMO clades are marked inside the clades. The FMOs explosion concordant with the emergence of terrestrial vertebrates (tetrapods, 352 mya) on the left. The three ancestral nodes that were experimentally characterized are ploted at the nodes and labelled with yellow squares. The tree was prepared in Figtree v1.4.2. Accession codes to sequences can be found in Supplementary Data 1.



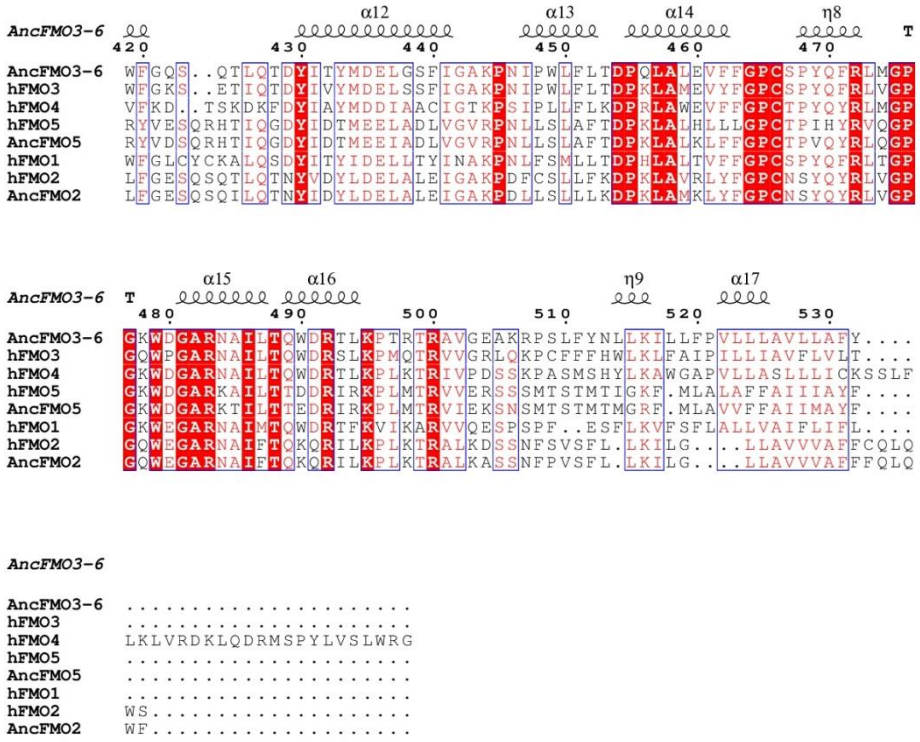
Supplementary Figure 3: Crystal Structure of AncFMO2 in the absence of $NADP^+$ (green) superposed to the structure bound to $NADP^+$ (dark green). AncFMO2 crystallizes in an identical manner with or without $NADP^+$. The root-mean-square deviation between the native AncFMO2 and its $NADP^+$ complex is 0.23 Å over 530 Cα atom pairs. The orientation of the dimer depicts the structure sitting on top of the phospholipid bilayer as shown in the other structures (see Figure 3).



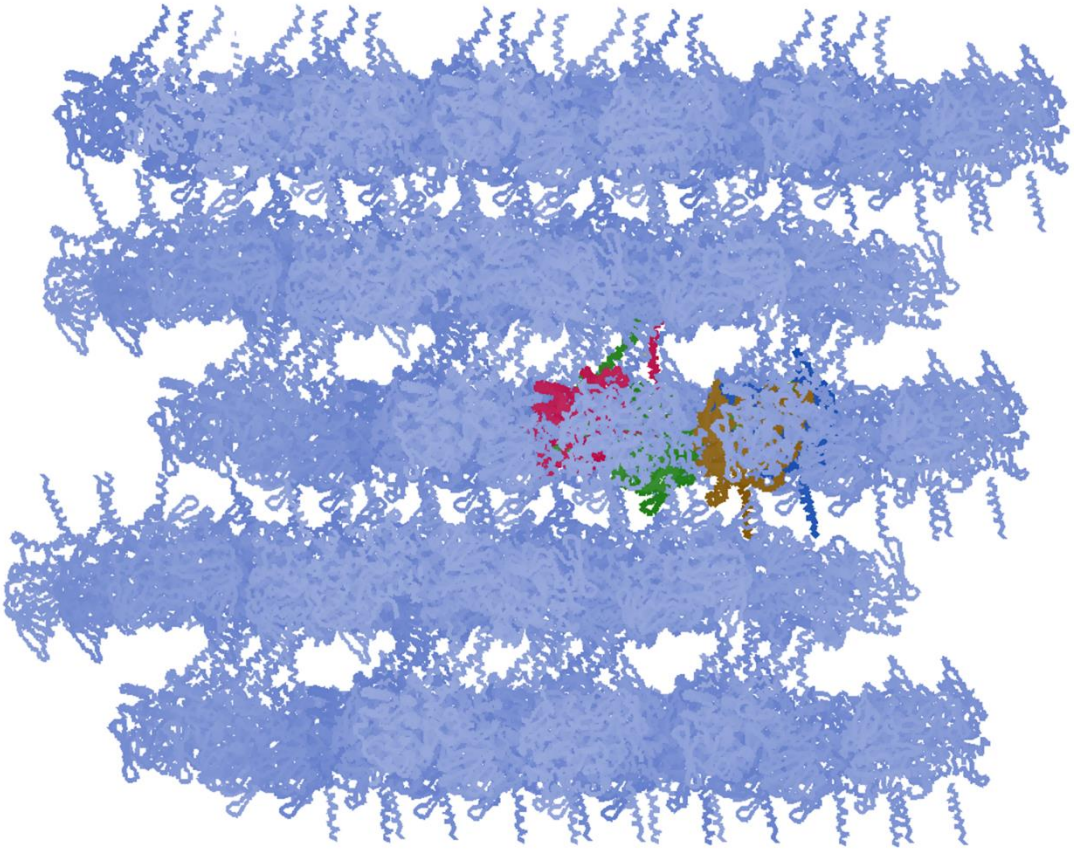
Supplementary Figure 4: Electron density maps of the AncFMOs. Structure determination was greatly facilitated by density averaging because the asymmetric unit of the crystals of AncFMO2, AncFMO3-6, AncFMO5 contain four, six, and two protein molecules. The depicted 2Fo-Fc maps were calculated by averaging the electron density maps obtained after molecular replacement (shown in blue). **A**: AncFMO2 shown in lime green, without $NADP^+$. **B**: AncFMO2 shown in lime green, with $NADP^+$. **C**: AncFMO3-6 shown in dark magenta, with $NADP^+$. **D**: AncFMO5 shown in dark orange, with $NADP^+$. Cofactors FAD and $NADP^+$ are shown in yellow and cornflower blue respectively. The contour level is 1.4 σ.


AncFMO3-6
AncFMO3-6
hFMO3
hFMO4
hFMO5
AncFMO5
hFMO1
hFMO2
AncFMO2

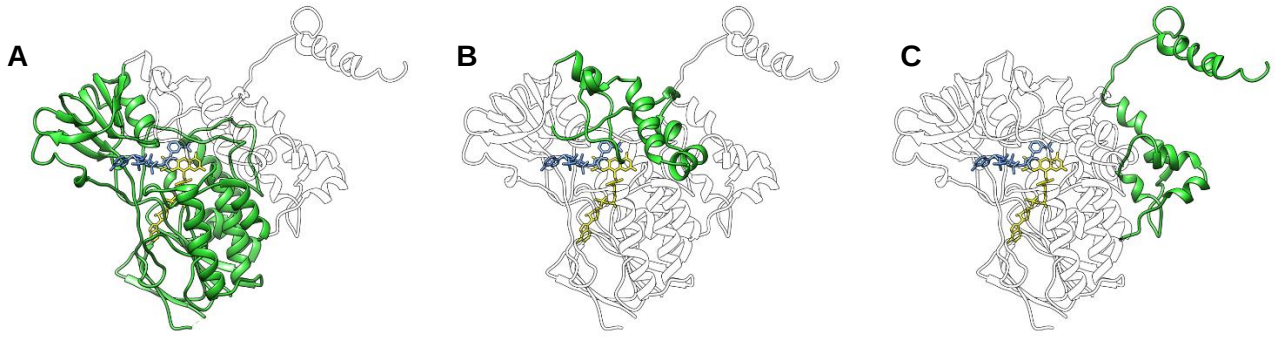
1050



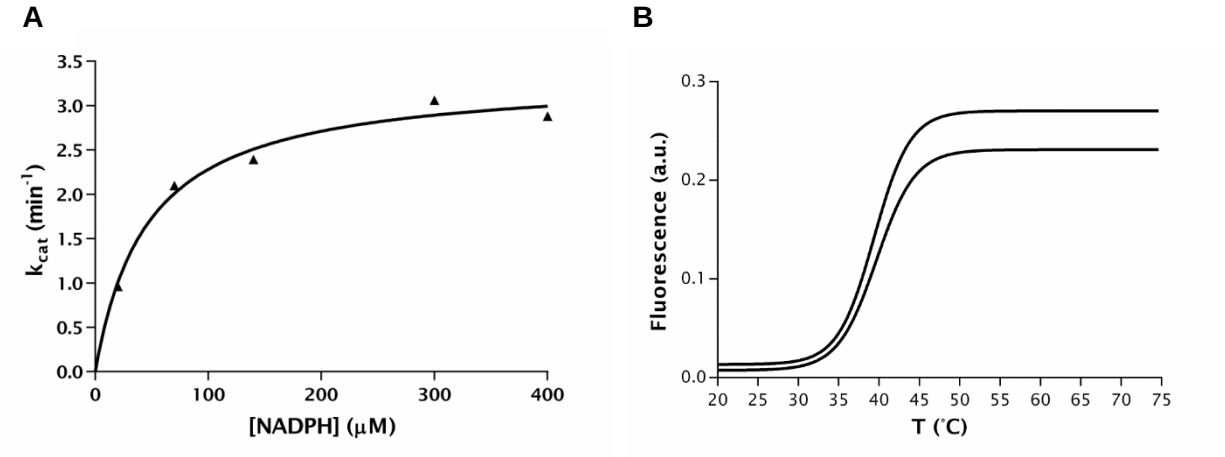
Supplementary Figure 5: Structure based sequence alignment using ESPript 3.0. The alignment was generated by using AncFMO3-6 pdb structure as input file and inserting the Ancestral FMO and hFMO sequence files.



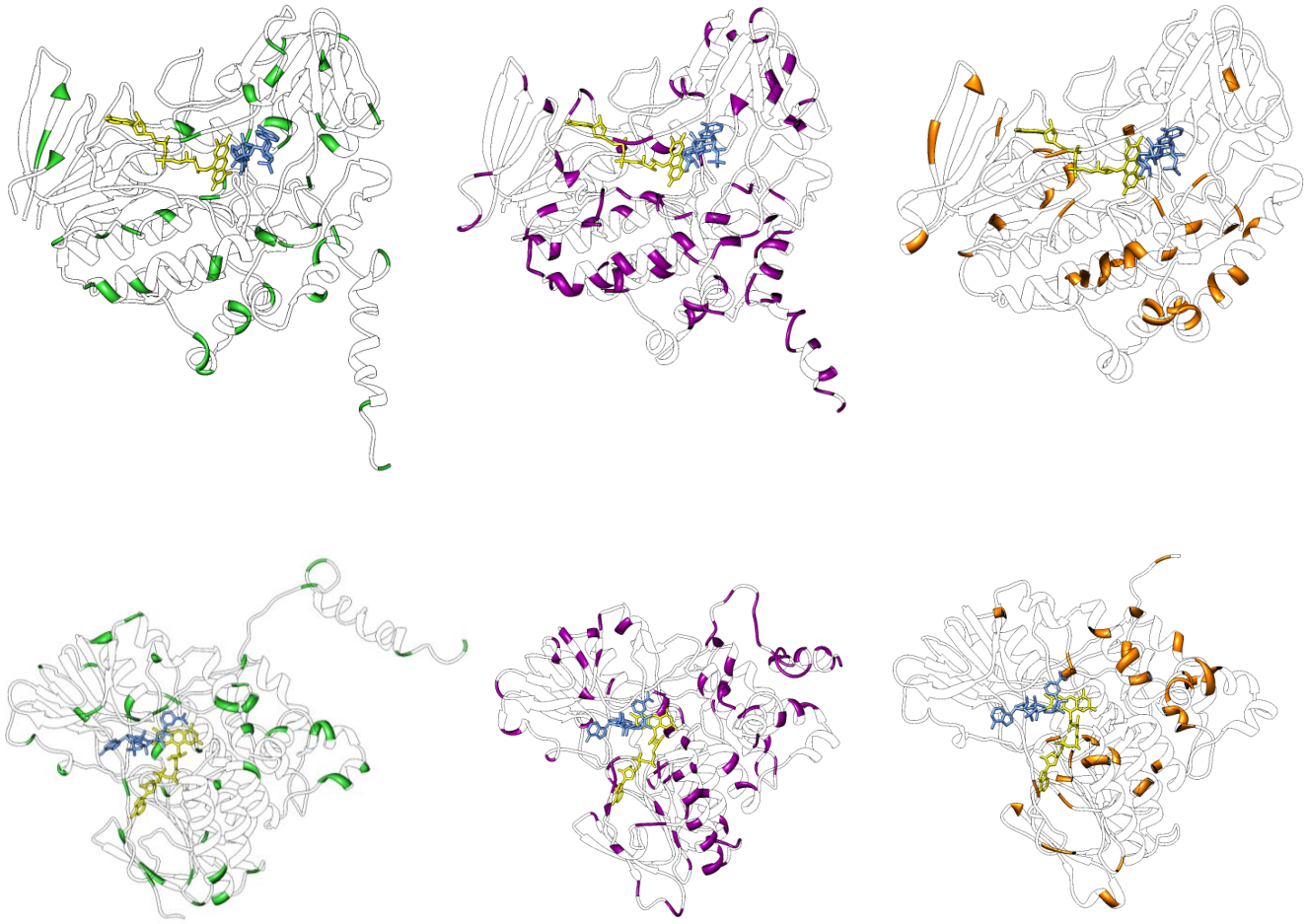
Supplementary Figure 6: The crystal packing of AncFMO2 forms multiple planes of soluble dimer-dimer interactions that extend across the lattice. The asymmetric unit is depicted by the four differently coloured monomer units of dark yellow, dark red, dark green and dark blue. In between each plane, we see multiple trans-membrane helices projecting upwards and downwards from each asymmetric unit. Each dimer projects its trans-membrane helices towards its reciprocal dimer.



Supplementary Figure 7: Topological features of the mammalian FMOs. **A**: Highly conserved NADP(H) and FAD dinucleotide binding domains that are observed in all FMOs. **B**: The characteristic 80-residue insertion (residues 214-295 in AncFMO3-6) that covers the FAD and binds to the membrane monotopically through an alpha helical triad. **C**: The additional C-terminal (residues 443-528) that orchestrates both monotopic and bitopic membrane binding features through an α-helical triad and a C-terminal helix respectively.



Supplementary Figure 8: NADPH oxidase activity and Melting temperature of hFMO3. A: NADPH consumption was not altered by the presence of substrate and so a Michaelis-Menten curve was plotted at differing NADPH concentrations. The K_M and k_{cat} for NADPH were determined at 46 ± 9 μM and 0.06 ± 0.16 s^{-1}, respectively. **B**: Extensive buffer screenings for hFMO3 resulted in a maximum melting temperature of 44.5 °C (with and without 200 μM $NADP^+$ in the upper and lower curves, respectively) in buffer conditions of 100 mM HEPES pH 7.5, 10 mM KCl and 0.05% (v/v) TRX-100-R.



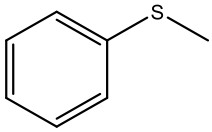
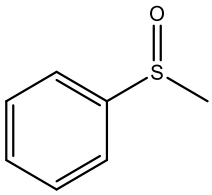
Supplementary Figure 9: Differing residues between AncFMOs and hFMOs. The upper and low panels display the structures in two orientations. The changes exhibited by AncFMO2, AncFMO3-6 and AncFMO5 compared to hFMOs are shown in lime green, dark magenta and orange, respectively. A close systematic analysis of the changes does not reveal any clear pattern of amino acid substitutions. Most are localised in the membrane-binding regions, implying that the enzyme can undergo multiple mutations in these parts of the protein as long as the hydrophobic nature of the side chains is conserved. This finding is further corroborated by the sequence alignment of the hFMOs and the AncFMOs sequences, with the sequences at the subdomains and C-termini, varying substantially (Supplementary Figure 5). The mutations however, are bereft in the FAD and NADP(H) binding domains, describing well-conserved sequence motifs among FMOs. Furthermore, the residues inside the enzyme and more importantly, lining the tunnels, are also well conserved. Only one overwhelming change in the core of the enzyme is observed in AncFMO5 as shown in Figure 5D (lower panel) of the main text.

Supplementary Table 1: Crystallographic table for structures AncFMO2, AncFMO3-6 and AncFMO5 complexed with or without $NADP^+$.

	AncFMO2 without $NADP^+$		**AncFMO2 with $NADP^+$**		**AncFMO3-6**	**AncFMO5**
	Aimless	Staraniso	Aimless	Staraniso	Aimless	Aimless
PDB codes	6SEM		6SF0		6SE3	6SEK
Resolution Range	48.395 - 2.749 (2.937 - 2.749)		138.06 - 3.01 (3.34 - 3.01)		49.30 - 2.80 (2.85 - 2.80)	49.89 - 2.70 (2.82 - 2.70)
Space Group	$C2_1$		$C2_1$		$P3_121$	$P22_12_1$
Unit Cell (Å, °)	152.96 147.78 144.93 β=96.91		153.75 148.68 139.12 β=97.08		156.09 156.09 370.61	96.78 100.07 143.15
Total Reflections	232952 (12654)	145667 (6653)	178736 (15850)	109279 (7605)	2621815 (133377)	258133 (31923)
Unique Reflections	77919 (4332)	48490 (2425)	51884 (4527)	32052 (2290)	129389 (6372)	38832 (4655)
Multiplicity	3.0 (2.9)	3.0 (2.7)	3.4 (3.5)	3.4 (3.3)	20.3 (20.9)	6.6 (6.9)
Completeness (%)	99.1 (98.0)	58.3 (16.2)	99.4 (99.9)	52.1 (13.9)	100.0 (99.9)	99.8 (99.9)
Completeness ellipsoidal (%)	-	93.6 (76.9)	-	92.4 (64.1)	-	-
Mean I/sigma (I)	8.8 (0.3)	13.7 (1.3)	3.3 (0.4)	5.3 (1.5)	14.3 (1.3)	9.9 (0.8)
R-merge	0.080 (3.85)	0.054 (0.85)	0.169 (2.37)	0.145 (0.74)	0.235 (2.88)	0.11 (2.19)
CC1/2	0.99 (0.19)	0.99 (0.47)	0.987 (0.36)	0.986 (0.63)	0.998 (0.41)	0.99 (0.42)
R-work (%)		21.0		21.1	21.7	22.2
R-free (%)		25.6		28.5	26.6	29.5
Number of non-hydrogen atoms		16905		17542	26169	8329
RMS (bonds) Å		0.008		0.008	0.010	0.007
RMS (angles) °		1.311		1.310	1.530	1.553
Ramachandran						
favoured (%)		86.3		81.7	90.1	94.5
allowed (%)		9.3		11.7	7.6	5.1
outliers (%)		4.4		6.6	2.3	0.4
Average B-factor						
Protein		86.5		99.5	73.6	90.6
FAD		78.1		98.2	56.8	98.1
$NADP^+$		-		85.0	67.6	75.1
Other ligands		-		102.3	57.0	126.6
Solvent		54.0		48.9	56.5	62.4

Statistics for the highest-resolution shell are shown in parentheses. Data derived from Aimless and Staraniso used for AncFMO2 are described.

Supplementary Table 2: Conversions by AncFMO2 and AncFMO3-6 variants.

		CONVERSION (%)			
SUBSTRATE	**PRODUCT**	AncFMO2 WT	AncFMO2 E281H	AncFMO3-6 WT	AncFMO3-6 E281H
		>99	43	>99	>99

Cyclohexanone showed no conversion for all four enzymes.

Supplementary Table 3: Melting Temperatures of each AncFMO

Protein	**$NADP^+$ (µM)**	**Tm (°C)**
AncFMO2	0	52.3
	200	70.0
AncFMO3-6	0	60.5
	200	66.5
AncFMO5	0	55.0
	200	59.0
Human FMO3	0	44.5
	200	44.5
Human FMO5	0	49.0
	200	48.5

Measured in triplicate by using the ThermoFAD technique (see materials and methods).

Supplementary Table 4: Ambiguously reconstructed sites in AncFMOs

Site	ML State	PP	Alt State	PP
AncFMO2				
158	Q	0.64	E	0.33
347	N	0.58	D	0.37
AncFMO3-6				
4	K	0.59	R	0.31
139	V	0.77	I	0.21
158	E	0.77	D	0.20
209	K	0.78	Q	0.20
246	S	0.45	T	0.40
276	G	0.67	S	0.20
324	C	0.59	Y	0.40
378	T	0.63	A	0.31
509	S	0.40	Q	0.36
510	L	0.75	P	0.22
514	L	0.63	I	0.28
517	I	0.54	L	0.28
519	L	0.55	F	0.45
520	F	0.60	C	0.37
AncFMO5				
17	T	0.72	A	0.24
35	T	0.62	S	0.21
241	T	0.51	K	0.43
253	S	0.56	I	0.39
255	T	0.48	H	0.44
410	S	0.66	T	0.20
422	D	0.75	K	0.30
439	I	0.76	L	0.21
462	K	0.47	E	0.40
504	I	0.60	V	0.49
510	M	0.72	S	0.24

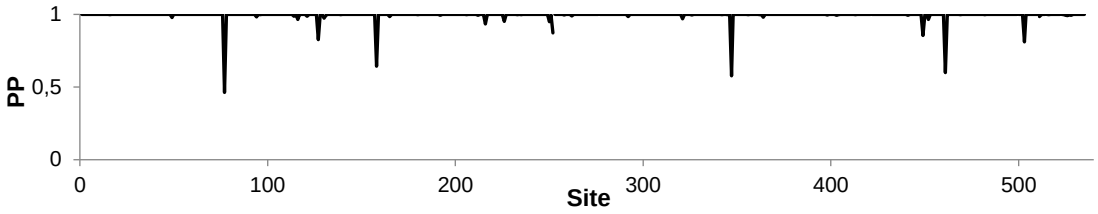
Ambiguously reconstructed sites defined by the Alt state highest posterior probability (PP)>0.2 are presented for each AncFMO.

Suplementary Data 1: FMOs sequences. Separate spreadsheet.

Suplementary Data 2: AncFMOs sequences.

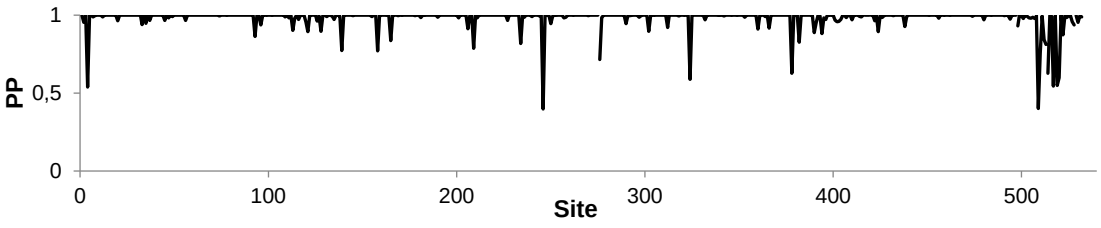
>AncFMO2
MAKKVAVIGAGVSGLISLKCCVDEGLEPTCFERTEDIGGLWRFKENVEDGRASIYQSVITNTSKEMSCFSDFPMPEHFPN
FLHNSKLLEYFRIFAKKFDLLKYIQFQTTVLSVKKCPDFSSSGQWEIVTESNGKEQSAVFDAVMVCSGHHILPHIPLQSF
PGIERFKGQYFHSRQYKHPEGFEGKRILVIGIGNSASDIAVELSKKAAQVFISTRHGSWVMSRISDDGYPWDMVFHTRFS
SMLRNVLPRTVVKWMMEQQMNRWFNHENYGLEPQNKYLMKEPVLNDDLPSRLLYGAIKVKSRVKELTETSAIFEDGTVEE
DIDVIVFATGYTFSFPFLEDSLVKVENNMVSLYKYMFPPHLEKPTLACIGLIQPLGSIFPTVELQARWVTRVFKGLCTLP
SESTMMADIIKRNEKRIDLFGESQSQILQTNYIDYLDELALEIGAKPDLLSLLLKDPKLAMKLYFGPCNSYQYRLVGPGQ
WEGARNAIFTQKQRILKPLKTRALKASSNFPVSFLLKILGLLAVVVAFFFQLQWF

Average of reconstruction= 0.994



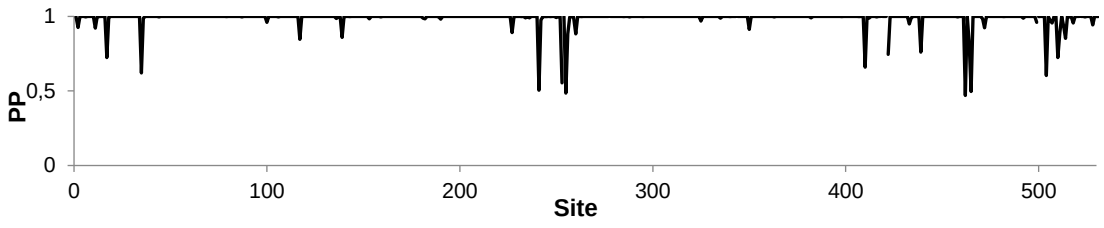
>AncFMO3-6
MGKKVAIIGAGVSGLASIRSCLEEGLEPTCFERSDDIGGLWKFSDHAEEGRASIYQSVFTNSSKEMMCFPDFPYPDDFPN
FMHNSKLQEYITAFAKEKNLLKYIQFKTLVSSVNKRPDFSVTGQWDVTTEKDGKKESAVFDAVMICSGHHVYPNLPKESF
PGLKHFKGKCFHSRDYKEPGIFKGKRVLVIGLGNSGCDIATELSHTAEKVIISSRSGSWVMSRVWDDGYPWDMLFITRFE
TFLKNSLPTAISDWWYMKQMNARFKHENYGLMPLNGTLRKEPVFNDELPARILCGTVSIKPNVKEFTETSAIFEDGTVFE
AIDCVIFATGYGYAYPFLDDSIIKSRNNEVTLFKGIFPPLLEKPTLAVIGLVQSLGATIPTTDLQARWAAKVFANSCTLP
TTNEMMDDIDEKMGKKLKWFGQSQTLQTDYITYMDELGSFIGAKPNIPWLFLTDPQLALEVFFGPCSPYQFRLMGPGKWD
GARNAILTQWDRTLKPTRTRAVGEAKRPSLFYNLLKILLFPVLLLAVLLAFY

Average of reconstruction= 0.982



>AncFMO5
MTKKRIAVIGAGASGLTSIKCCLEEGLEPVCFERTDDIGGLWRFQENPEEGRASIYKSVIINTSKEMMCFSDYPIPDHYP
NFMHNSQVLEYFRMYAKEFDLLKYIQFKTTVCSVKKQPDFSTSGQWEVVTECEGKKEVDVFDGVMVCTGHHTNAHLPLES
FPGIEKFKGQYFHSRDYKNPEGFTGKRVIIIGIGNSGGDLAVEISHTAKQVFLSTRRGAWILNRVGDHGYPFDVLFSSRF
TYFLSKICGQSLSNTFLEKKMNQRFDHEMFGLKPKHRALSQHPTVNDDLPNRIISGLVKVKGNVKEFTETAAIFEDGSRE
DDIDAVIFATGYSFAFPFLEDSVKVVKNKVSLYKKVFPPNLEKPTLAIIGLIQPLGAIMPISELQGRWATQVFKGLKTLP
SQSEMMAEISKAQEEMAKRYVDSQRHTIQGDYIDTMEEIADLVGVRPNLLSLAFTDPKLALKLFFGPCTPVQYRLQGPGK
WDGARKTILTTEDRIRKPLMTRVIEKSNSMTSTMTMGRFMLAVVFFAIIMAYF

Average of reconstruction= 0.9873



For the three characterized ancestral proteins, sequences of the ML (maximum likelihood) ancestors are shown. Graphs display the posterior probabilities of each reconstructed site for AncFMO2, AncFMO3-6 and AncFMO5.