ARTICLES
https://doi.org/10.1038/s41594-019-0347-2

nature structural & molecular biology

Corrected: Publisher Correction

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

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

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 importance, structural information remains bereft. To further our understanding behind their biochemistry and diversity, we used ancestral-sequence reconstruction, kinetic and crystallographic techniques to scrutinize three ancient mammalian FMOs: AncFMO2, AncFMO3-6 and AncFMO5. Remarkably, all AncFMOs could be crystallized and were structurally resolved between 2.7- and 3.2-Å resolution. 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.

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 FMOs and utilize the cofactors flavin adenine dinucleotide (FAD) and nicotinamide adenine dinucleotide phosphate (NADPH), and dioxygen for activity[8–10]. Typically, FMOs pursue catalysis as illustrated in Fig. 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 tissues such as the kidney, lung, and liver[2,10,13,14]. The human *FMO* genes are found on chromosome 1, with *FMO1*–*FMO4* clustering over 220 kilobases (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 (ref. [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 in regard to 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]. 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 active in the majority of humans due to a mutation that introduces a premature stop codon[35,36]. FMO5 is distinct from the other FMOs because it is able to perform Baeyer–Villiger oxidations (Fig. 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].

To gain insight into the historical events leading to the paralogues' divergence in mammals, we generated three ancestral FMOs (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 *Escherichia coli* and purified as holoenzymes (containing FAD) 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 that can facilitate crystallization.

Results

Ancestral-sequence reconstruction of mammalian FMOs. We inferred the evolutionary history of FMOs from a full phylogeny

[1]Department of Biology and Biotechnology "Lazzaro Spallanzani", University of Pavia, Pavia, Italy. [2]Molecular Enzymology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Groningen, The Netherlands. [3]IMIBIO-SL CONICET, Facultad de Química Bioquímica y Farmacia, Universidad Nacional de San Luis, San Luis, Argentina. *e-mail: mlmascotti@unsl.edu.ar; m.w.fraaije@rug.nl; andrea.mattevi@unipv.it

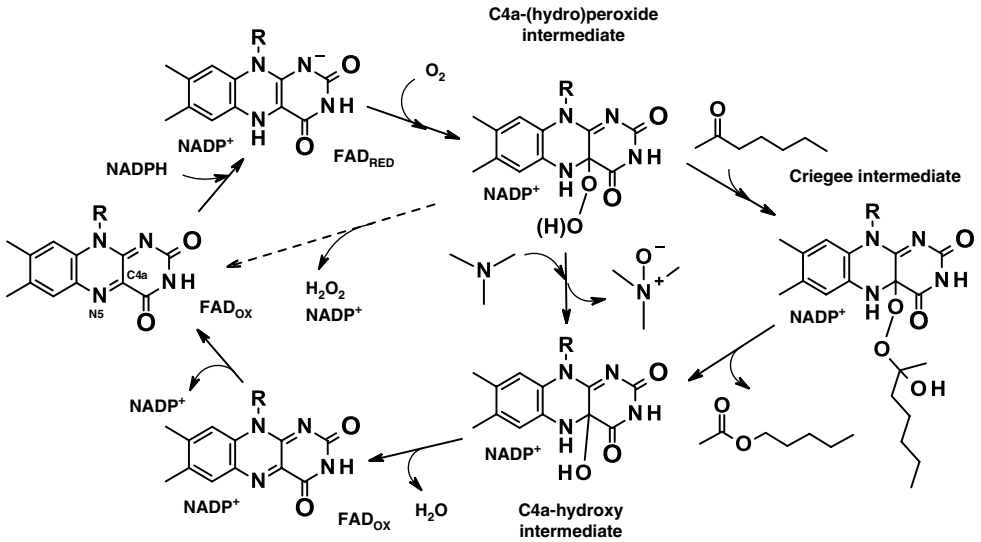


Fig. 1 | The catalytic mechanism of FMOs. Oxidized FAD (FAD_{OX}) is reduced (FAD_{RED}) by NADPH. FAD_{RED} is consequently oxidized by a molecule of dioxygen to generate the C4a-(hydro)peroxide intermediate. The typical mode of action of FMOs, with the distal oxygen atom from the intermediate being 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 C4a-(hydro)peroxide intermediate decays with the release of $NADP^+$ and hydrogen peroxide. R represents the ribityl-adenosine-diphosphate.

constructed by including experimentally characterized enzymes from Bacteria and Eukarya, plus sequences found by extensive sequence homology searching and hidden Markov model (HMM) profiling (Supplementary Fig. 1). Our work confirmed the findings of the previous studies by Hernandez et al.[16] and Hao et al.[44]: (1) jawed vertebrate FMOs are monophyletic and are derived from a single common ancestor (Fig. 2a and Supplementary Fig. 2); (2) several duplication events occurred in the terrestrial vertebrates; (3) the ancestor of mammals already encoded the five FMO paralogues resulting from four major gene-duplication events (Fig. 2a and Supplementary Fig. 2); and (4) a sixth mammalian paralogue (*FMO6*) resulted from a late gene-duplication event. *FMO6* has been described as a pseudogene in humans[45], but it might be functional in mice[16] and its nature is unknown in other mammals.

By performing ancestral-sequence reconstruction, we obtained the protein sequences of AncFMOs from mammals with high posterior probabilities (ranging from 0.98 to 0.99) (Fig. 2b and Supplementary Fig. 3). In the phylogeny, we observed that FMO5 diverged earlier than the other FMOs, in agreement with previous reports[44], and was followed by FMO2, FMO1, FMO4 and the FMO3-6 paralogue couple. This topology suggests that the gene-duplication events took place simultaneously, rendering no clear paralog couples as has been previously proposed (Fig. 2a and Supplementary Fig. 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[46]). In the case of AncFMO3-6, the ancestor underwent an early duplication event, originating the FMO3 and FMO6 paralogues in mammals. As a general trend, 70% of sites are conserved between the preduplication ancestors and modern FMO3 and FMO6. Along each branch to the human FMO3 or human FMO6 sequences, 94–98 substitutions occurred, with 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 owing 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.

Catalytic rates of AncFMOs resemble extant mammalian FMOs. Critically for our project, the hitherto-generated AncFMO 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 homologues to human FMOs was established (see 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 the absence of an organic substrate; $NADPH_{uncoupling}$ in Table 1). This was followed by 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[2,17,34,37,47–50]. 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 substantially 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 (Fig. 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 C4a-(hydro) peroxyflavin intermediate with its well-defined spectroscopic properties (Fig. 3). Based on the steady-state kinetics data, AncFMO5 is assumed to behave similarly. Collectively, these experiments convincingly demonstrated that our AncFMO2, AncFMO3-6 and AncFMO5 enzymes are enzymatically competent and exhibit the typical catalytic features of class B flavoprotein monooxygenases.

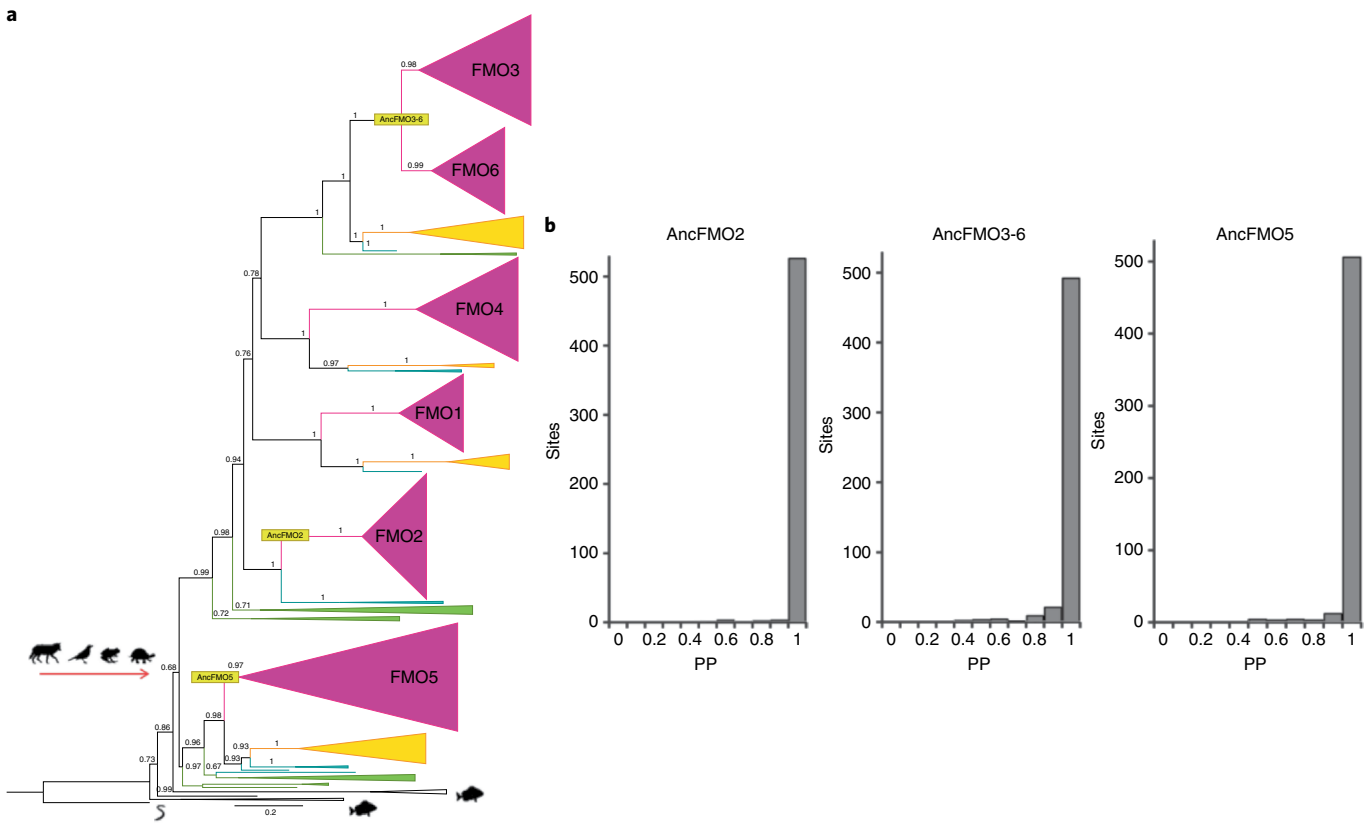


Fig. 2 | Ancestral-sequence reconstruction of FMOs. a, Condensed maximum-likelihood phylogeny of FMOs from jawed vertebrates. Clades are colored according to tetrapod classes: Mammalia (magenta), Aves (light orange), Amphibia (green) and Testudines (teal). Clades on the base are from other non-terrestrial gnathostomes (black). Rooting was performed according to the species tree. Above the branches, the transfer bootstrap expectation values are shown. The emergence of terrestrial vertebrates (tetrapods, 352 million years ago)[72] is marked with an arrow and cartoons on the left. The three ancestral nodes that were experimentally characterized are labeled with yellow squares. The fully annotated phylogeny is presented in Supplementary Fig. 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 was 0.994, for AncFMO3-6 was 0.982 and for AncFMO5 was 0.987. Accession codes to sequences can be found in the Source Data available with the paper online.

AncFMOs are 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 (Fig. 4). 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 holoenzyme crystal structures (Extended Data Fig. 1). 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 (Table 2 and Extended Data Fig. 2). 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 Fig. 4).

Our crystal structures depict the AncFMOs as dimers: they possess an extensive monomer–monomer interface over approximately 2,000 Å^2 (calculated by the PISA server)[51]. Furthermore, their well-conserved FAD and NADP(H) dinucleotide-binding domains are accompanied by two large transmembrane helices (one from each monomer) that project outwards, approximately parallel to the twofold axis (Fig. 4a–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 (for example, FMO from *Roseovarius nubinhibens*, PDB entry 5IPY; FMO from *Methylophaga aminisulfidivorans*, PDB entry 2VQ7).

The mammalian FMOs were predicted to contain a highly hydrophobic carboxy-terminal transmembrane helix (residues 510–532 in AncFMO3-6; Supplementary Fig. 4). The crystal structures of AncFMO2 and AncFMO3-6 perfectly confirmed this prediction, as both enzymes possess C-terminal transmembrane helices that span 30 Å in length and are decorated with many hydrophobic residues (Fig. 4a,b,d). Of notice, these α-helical scaffolds represent the key protein–protein interactions established within the crystal packing (Extended Data Fig. 3). 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, Fig. 4 depicts each enzyme as if it were sitting on the membrane.

It was reported that truncation of the C-terminal helices was insufficient for protein solubilization[52,53]. This indicated that the enzyme possessed additional membrane-binding features. To understand which elements promote membrane association,

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,d]	0.197 ± 0.009	59 ± 8

[a]Rates were determined by following NADPH consumption (absorbance decrease at 340 nm). 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 (Fig. 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. [d]The data for human FMO3 are shown in Extended Data Fig. 5. The data for human FMO5 are taken from Fiorentini et al.[37].

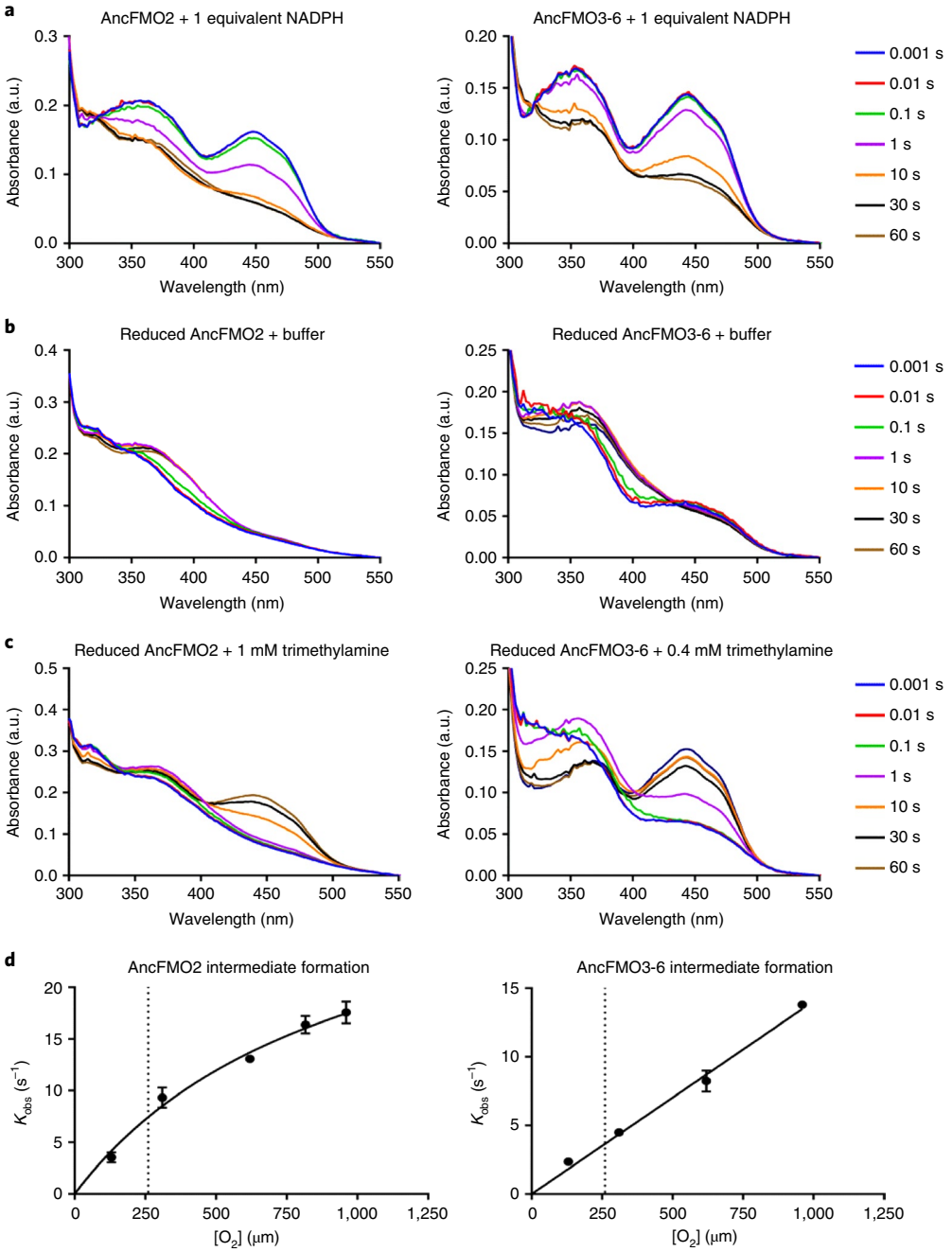
the charge distribution on the protein surface was scrutinized. Intriguingly on the underside of the dimer, two large hydrophobic strips, about 30 Å in length, extend across the enzyme surface (Fig. 4d). These strips are lined by an extensive ring of positively charged residues. Collectively, these features 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, respectively, that were used for protein solubilization and crystallisation (Fig. 4a,c,d). These molecules delineate the membrane–enzyme interface and further validate that the hydrophobic strips monotopically embed within the membrane. These findings rationalize the extensive membrane-binding nature employed for this class of enzymes and corroborate that the truncation of the C-terminal helix alone is not sufficient to facilitate protein solubilization[52]. FMOs employ both bitopic and monotopic membrane-binding features to grapple onto the membrane effectively and abstract lipophilic substrates from within the membrane.

An 80-residue insertion 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 (Extended Data Fig. 4a)[2,54]. Superposition of a bacterial FMO (PDB ID: 2VQ7, 29% identity) from *Methylophaga* sp. strain SK1 (ref. [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; Extended Data Fig. 4b,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 'AncFMOs show conserved membrane-accessible substrate tunnels'). This finding implies that substrates navigate through tunnels manufactured by the insertion to access the closed catalytic cavity.

AncFMOs possess a buried active site. With the AncFMOs' active sites no longer being open clefts like their soluble homologues, we scrutinized each closely to determine the functions of each residue and whether the mode of NADP(H) binding is akin to class B FMOs. Notably, most residues in the active and NADPH-binding sites are conserved with near-identical conformations (Fig. 5a–c). Thr62, Ser62 and Thr63 for AncFMO2, AncFMO3-6 and AncFMO5, respectively, are within hydrogen-bonding distance to the N3 atom of the isoalloxazine ring and orientate the FAD towards the substrate pocket. Additionally, Asn61, Asn61 and Asn62 are observed in all active sites, and an Asn residue at this position is strictly conserved among human FMOs (Supplementary Fig. 4). 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 (Fig. 1). Consistently, mutating this residue in human FMO3 causes trimethylaminuria, further verifying its integral role within the active site[12,28].

The binding mode of $NADP^+$ observed in the crystal structures is iconic to class B FMOs (Fig. 5d)[4]. The overhanging Arg223, Arg223 and Arg224 are 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. Moreover, the nicotinamide ring is sterically held in place by a well-conserved Asn194, Asn194 and Asn195, which acts like a backdoor for the cofactor (Fig. 5a–c). This feature is not uncommon and is portrayed in some soluble FMOs by a protruding tyrosine[12,55,56]. The hydroxyl groups of the ribose form part of an intricate hydrogen-bonding network. The 2′-OH group is within hydrogen-bonding distance of the backdoor residue Asn194, Asn194 and Asn195 (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 and substrate oxygenation (Fig. 1)[12,55,57,58].

A Glu-to-His mutation may promote Baeyer–Villiger oxidation. Considering that AncFMO5 is structurally very similar to AncFMO2 and AncFMO3-6, but at the same time is functionally divergent, we sought to clarify which 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, Glu 281, derived from the above-described mammalian FMO-specific 80-residue insertion, points towards the flavin ring. With positively charged substrates



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

Fig. 3 | 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 (Fig. 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[57]. **e**, Rates of reduction, C4a-(hydro)peroxyflavin formation and C4a-(hydro)peroxyflavin decay in the absence of substrate (0.26 mM dioxygen; dotted lines in **d**). Measurements were performed in technical duplicates. a.u., arbitrary units.

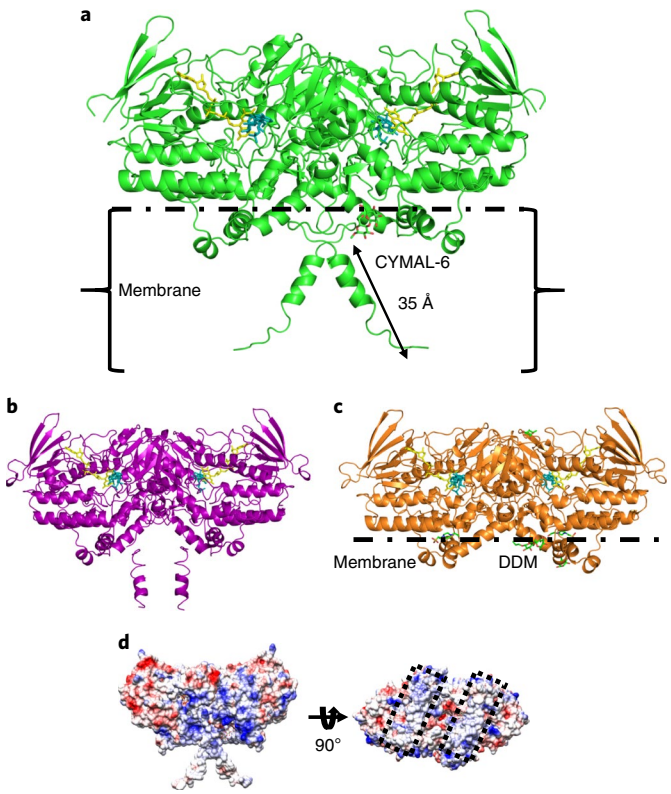


Fig. 4 | Crystal structures of the AncFMOs. a–c, Crystallographic dimers of AncFMO2, AncFMO3-6 and AncFMO5 are shown in lime green (**a**), dark magenta (**b**) and orange (**c**). FAD and NADP$^+$ are shown in yellow and cornflower blue, respectively. The orientations of the AncFMOs are identical, depicting their structures as if they were sitting on the phospholipid bilayer. In **a**, the lengths of the transmembrane 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 detergent molecule. 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°, the large parallel hydrophobic strips across the bottom of the dimer are visible, lined by a ring of positively charged residues indicated by black dashed boxes.

being preferred by FMOs[2], Glu 281 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 (Fig. 5c) and likely serves as a hydrogen-bond donor. This function is commensurate with Baeyer–Villiger monooxygenases, whereby hydrogen-bond-donating residues (that is, Arg; Fig. 5d) 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 (Fig. 1)[59–61]. 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 AncFMO5-H282E mutant fully lost its activity. Analogously, the AncFMO2-E281H and AncFMO3-6-E281H 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 1). This could be due to the fact that the fine structural and geometric features for formation and stabilization of the Criegee intermediate (Fig. 1) need further mutations, for example in the second shell of active-site residues.

AncFMOs show conserved membrane-accessible substrate tunnels. 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[62]. 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 (Fig. 6). In all three structures, the inner segment of the tunnel features a conserved leucine that acts as a gate keeper (Leu375 in AncFMO3-6; Fig. 6b, 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, 2, 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 Fig. 4).

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 (Fig. 6a–d). Specifically, a few noticeable changes were observed (Fig. 6e). The largest conformational difference is seen at residues 337–352, 337–352 and 338–352 for AncFMO2, AncFMO3-6 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 (Fig. 6b). 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 an 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.

AncFMOs are reliable and thermostable models for human FMOs. Allegedly, highly thermostable enzymes are highly prone to crystallization[63]. Considering that all three AncFMOs crystallized, melting temperature (T_m) assays were conducted using the ThermoFAD technique[64] to investigate the thermal stability of AncFMOs compared with that of human FMOs. Remarkably, our AncFMOs in storage buffer conditions (see 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 °C and +11 °C, respectively (Supplementary Table 2 and Extended Data Fig. 5)[37]. Generally, the differences between the AncFMOs and their respective human equivalents are found dispersed across the protein (Extended Data Fig. 6). These patterns of highly distributed and

Table 2 | Data collection and refinement statistics

	AncFMO2 without NADP+ (PDB 6SEM)		AncFMO2 with NADP+ (PDB 6SF0)		AncFMO3-6 (PDB 6SE3)	AncFMO5 (PDB 6SEK)
	Aimless	Staraniso	Aimless	Staraniso	Aimless	Aimless
Data collection[a]						
Space group	$C2_1$		$C2_1$		$P3_121$	$P22_12_1$
Cell dimensions						
a, *b*, *c* (Å)	152.96, 147.78, 144.93		153.75, 148.68, 139.12		156.09, 156.09, 370.61	96.78, 100.07, 143.15
α, β, γ (°)	90, 96.91, 90		90, 97.08, 90		90, 120, 120	90, 90, 90
Resolution (Å)[b]	48.39–2.75 (2.94–2.75)		138.06–3.01 (3.34–3.01)		49.30–2.80 (2.85–2.80)	49.89–2.70 (2.82–2.70)
R_{merge}	0.08 (3.85)	0.054 (0.85)	0.17 (2.37)	0.14 (0.74)	0.23 (2.88)	0.11 (2.19)
$I/\sigma(I)$	8.8 (0.3)	13.7 (1.3)	3.3 (0.4)	5.3 (1.5)	14.3 (1.3)	9.9 (0.8)
$CC_{1/2}$	0.99 (0.19)	0.99 (0.47)	0.987 (0.36)	0.986 (0.63)	0.998 (0.41)	0.99 (0.42)
Completeness (%)[b]	99.1 (98.0)	58.3 (16.2)	99.4 (99.9)	52.1 (13.9)	100.0 (99.9)	99.8 (99.9)
Ellipsoidal (%)[b]	-	93.6 (76.9)	-	92.4 (64.1)	-	-
Redundancy[b]	3.0 (2.9)	3.0 (2.7)	3.4 (3.5)	3.4 (3.3)	20.3 (20.9)	6.6 (6.9)
Refinement						
Resolution (Å)		2.75		3.01	2.80	2.70
No. reflections		48,490		32,052	129,389	38,832
R_{work}/R_{free}		0.21/0.256		0.211/0.285	0.217/0.266	0.222/0.266
No. of atoms[c]						
Protein		16639		17072	25188	8020
FAD		212		212	318	106
NADP+		-		192	288	96
Other ligands		-		24	12	87
Water		51		41	357	27
B-factors						
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
Water		54.0		48.9	56.5	62.4
R.m.s. deviations						
Bond lengths (Å)		0.008		0.008	0.010	0.007
Bond angles (°)		1.311		1.310	1.530	1.553

[a]Each dataset was measured from one crystal. [b]Values in parentheses are for highest-resolution shell. [c]The asymmetric unit contained 4, 4, 6 and 2 protein molecules for AncFMO2, AncFMO2 bound to NADP,+ AncFMO3-6 and AncFMO5, respectively.

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 (Fig. 6d, lower panel). This substitution may well allow larger substrates in human FMO5. Intriguingly, Fiorentini et al.[37] documented that NADP+ has no effect on the T_m of human FMO5, a result also observed for human FMO3 (Extended Data Fig. 5). However, the melting temperatures of all three AncFMOs increased in the presence of NADP+ by +17 °C, +7 °C and +4 °C for AncFMO2, AncFMO3-6 and AncFMO5, respectively (Supplementary Table 2). With AncFMOs exhibiting a low degree of uncoupling, it corroborates that the tight binding of NADP+ is necessary for highly coupled reactions (Table 1 and Fig. 2).

Discussion

Our work supports the notion that the number of FMOs in vertebrates substantially increased by successive gene-duplication events, leading to the multiple paralogues 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 paralogues, 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 vertebrate classes. With AncFMOs exhibiting similar substrate profiles and catalytic rates as their FMO successors, we propose that this class of enzymes has an evolutionarily conserved mode of action. Moreover, two new features are derived from ancestral-sequence reconstruction: (1) increased T_m and (2) the stabilizing effect induced by NADP+ (see Supplementary Table 2). 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[65].

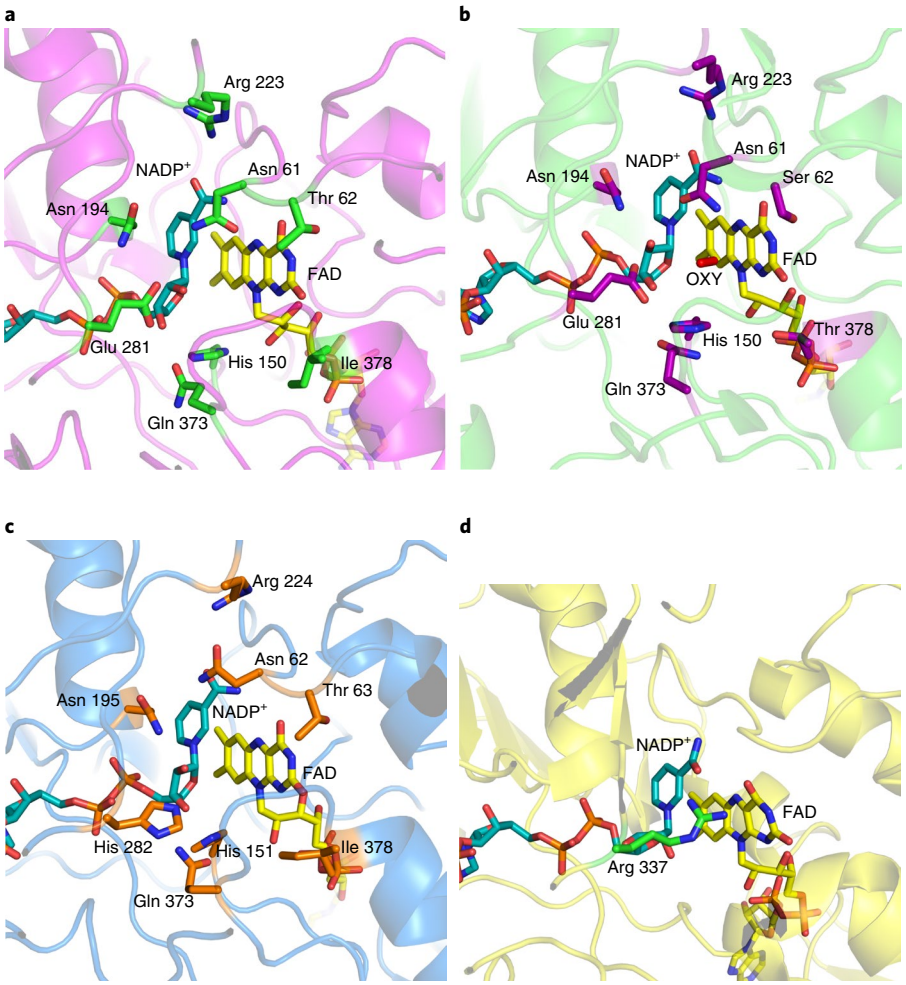


Fig. 5 | Active sites of the AncFMOs. a–c, The active sites for AncFMO2, AncFMO3-6 and AncFMO5 are shown in **a**, **b** and **c**, respectively. All three bear a high degree of similarity, with most amino acids being strictly conserved and displaying identical conformations. The differing residues are as follows: Thr62, Ser62, Thr63; Glu 281, Glu 281, His 282; Ile378, Thr378, Ile378. **d**, AncFMO3-6 also contains a tentatively assigned molecule of dioxygen (OXY). For the sake of comparison, the binding of NADP+ to the active site of phenylacetone monooxygenase, a prototypical class B monooxygenase (PDB 2YLR), is shown. Arg 337 is a conserved residue that is essential for the Baeyer–Villiger activity of this and similar enzymes.

Our research has resulted in the unveiling of the first structures until now of mammalian FMOs. Together, they demonstrate the extensive membrane-binding features employed by this enzyme class. The literature had always speculated that the C-terminus was involved in membrane association[37,52,53,66], 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[56,67]. Specifically, the oligomerization state aids membrane insertion, as the protein occupies a larger membrane-surface area[67]. 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[2,68]. FMO3 and FMO1 are understood to occupy a breadth of substrate sizes[2,18,69]. 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[70,71]. 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 (Fig. 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

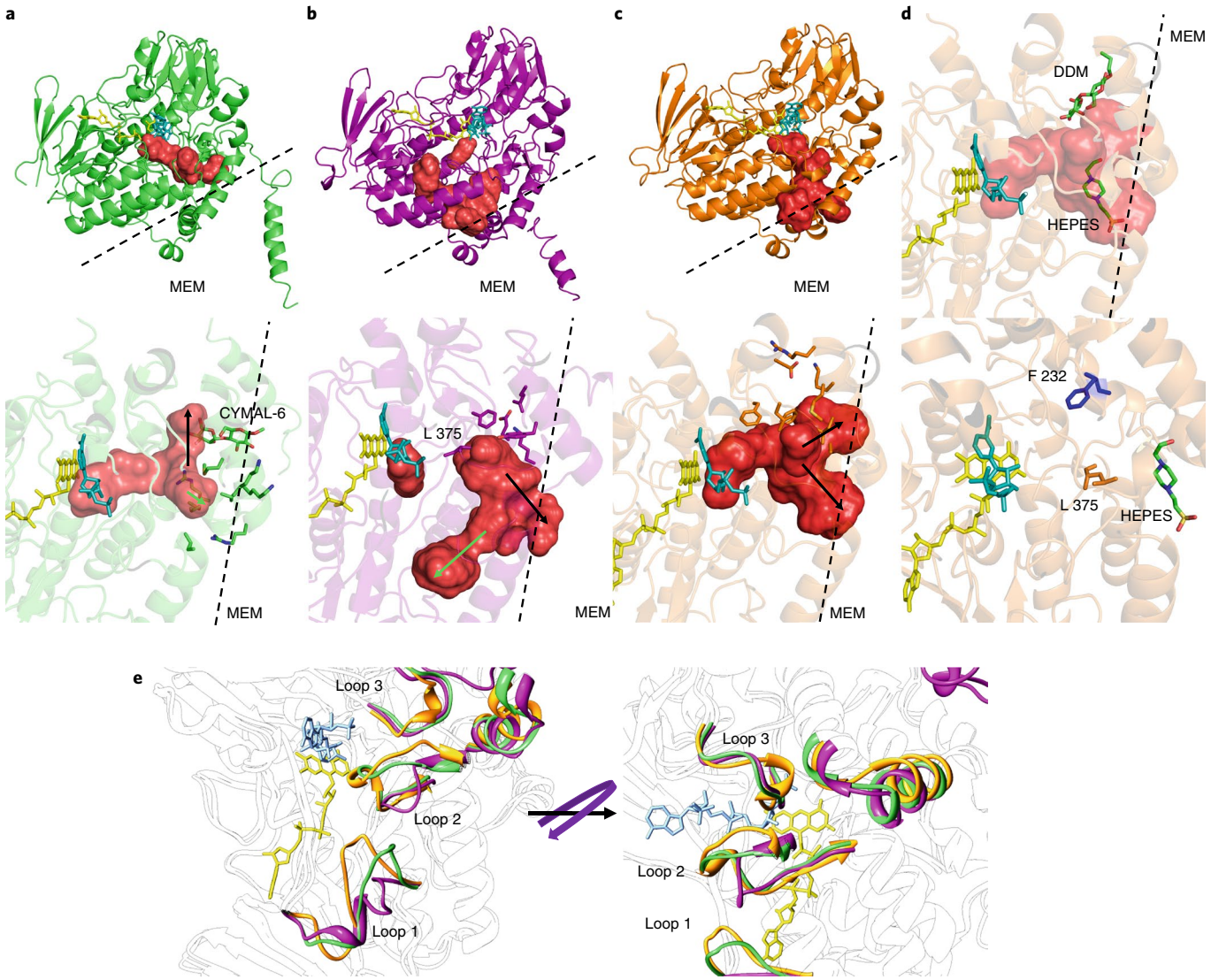


Fig. 6 | Substrate tunnels and structural differences in the AncFMOs. a–c, Top, the tunnels of AncFMO2 (**a**), AncFMO3-6 (**b**) and AncFMO5 (**c**) are shown, with the protein–membrane interface labeled as MEM. Bottom, The directions of the tunnels for AncFMO2 (**a**), AncFMO3-6 (**b**) and AncFMO5 (**c**) are shown, with their directions towards the membrane or the cytosol depicted by black or green arrows, respectively. The residues that block tunnel routes based on their conformations are shown. AncFMO2 and AncFMO3-6 contain 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). **d**, Top, a molecule of dodecyl-β-D-maltoside (DDM) found above the α-helical triad to emphasize the protein–membrane interface in AncFMO5 is shown. Additionally, a molecule of HEPES is present in the tunnel passing below the helix, demonstrating a substrate accessible pathway. Bottom, the Phe232 in AncFMO5 is shown with respect to gatekeeper Leu375, implying its vicinity to the FAD and how the change to alanine in human FMO5 is predicted to open the cavity. **e**, 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.

bulkiness of the ligands. The gating elements (for example Leu375) may well seal the active-site cavity when small substrates are bound (Fig. 6). 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 FMO 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 FMOs 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.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and

code availability are available at https://doi.org/10.1038/s41594-019-0347-2.

Received: 19 July 2019; Accepted: 1 November 2019;
Published online: 23 December 2019

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**, e0132689 (2015).
8. Ziegler, D. M. & 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. USA* **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**, 6498 (2015).
24. Krueger, S. K. et al. Genetic polymorphisms of flavin- containing monooxygenase (FMO). *Drug Metab. Rev.* **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, I. 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. Cell nonautonomous activation of flavin-containing monooxygenase 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. Grantham, R. Amino acid difference formula to help explain protein evolution. *Science.* **185**, 862–864 (1974).
47. 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).
48. 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).
49. 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).
50. 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).
51. Krissinel, E. Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Res.* **43**, W314–W319 (2015).
52. Geier, M. et al. Human FMO2-based microbial whole-cell catalysts for drug metabolite synthesis. *Microb. Cell Fact.* **14**, 1–10 (2015).

53. 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).
54. 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).
55. 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).
56. Lončar, N. et al. Characterization of a thermostable flavin-containing monooxygenase from *Nitrincola lacisaponensis* (NiFMO). *Appl. Microbiol. Biotechnol.* **103**, 1755–1764 (2019).
57. Beaty, N. B. & Ballou, D. P. The reductive half-reaction of liver microsomal FAD-containing monooxygenase. *J. Biol. Chem.* **256**, 4611–4618 (1981).
58. 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).
59. Romero, E., Castellanos, J. R. G., Mattevi, A. & Fraaije, M. W. Characterization and crystal structure of a robust cyclohexanone monooxygenase. *Angew. Chem. Int. Ed. Engl.* **55**, 15852–15855 (2016).
60. 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).
61. Criegee, R. Die umlagerung der dekalin-peroxydester als folge von kationischem sauerstoff. *Justus Liebigs Ann. Chem.* **560**, 127–135 (1948).
62. Ho, B. K. & Gruswitz, F. HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* **8**, 1–6 (2008).
63. Gumulya, Y. et al. Engineering highly functional thermostable proteins using ancestral sequence reconstruction. *Nat. Catal.* **1**, 878–888 (2018).
64. 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).
65. 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).
66. 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).
67. Allen, K. N., Entova, S., Ray, L. C. & Imperiali, B. Monotopic membrane proteins join the fold. *Trends Biochem. Sci.* **44**, 7–20 (2019).
68. 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).
69. Kim, Y. M. & Ziegler, D. M. Size limits of thiocarbamibes accepted as substrates by human flavin-containing monooxygenase 1. *Drug Metab. Dispos.* **28**, 1003–1006 (2000).
70. Beaty, N. B. & Ballou, D. P. Transient kinetic study of liver. *J. Biol. Chem.* **255**, 3817–3819 (1980).
71. Beaty, N. B. & Ballou, D. P. The oxidative half-reaction of liver microsomal FAD-containing monooxygenase. *J. Biol. Chem.* **256**, 4619–4625 (1981).
72. 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).

Methods

Phylogenetic inference and ancestral-sequence reconstruction. To obtain a robust and representative phylogeny of FMOs, sequences from Bacteria and Eukarya were collected by homology searches using BLAST and HMM profiling. We collected 310 sequences, which were aligned with MAFFT v7[73]. Best-fit model parameters were obtained by the Akaike information criterion in ProtTest v3.4. Phylogenies were inferred by employing the maximum-likelihood method in PhyML v3.0 or RAxML v0.6.0 with 500–1,000 bootstraps and transfer bootstrap expectation (TBE) subsequently[74]. As FMOs are not monophyletic, derived clades Baeyer–Villiger monooxygenases (BVMOs) and *N*-hydroxylating monooxygenases (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 a cephalochordate sequence, to root the tree according to species tree (Supplementary Fig. 2)[75].

Ancestral-sequence reconstruction was performed using the maximum-likelihood inference method in PAMLX v.4.9 (ref. [43,76]). Sequences were analyzed using an empirical amino acid substitution model (model = 3), four 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 posterior probability > 0.2 (ref. [77]). Alternative sites were found to be 2 for AncFMO2, 15 for AncFMO3-6 and 11 for AncFMO5 (Supplementary Table 3). 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. Complementary DNAs 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 6×His-tag regions to the amino-terminus. Inserts were fused into the destination vectors through Golden 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 (1×) and nuclease-free water added to a final volume of 20 µl. During the cloning procedure, a negative control was prepared with the fragments and inserts omitted. The Golden Gate assembly was conducted in the following manner, where maximum efficiency was desired: a cycle of 37 °C for 5 min followed by 16 °C for 10 min was repeated 30 times; followed by 55 °C for 10 min, 65 °C for a further 20 min; finishing with 8 °C for 20 min. Once cloned, the plasmids were transformed by heat shock into *E. coli*. BL21 cells (25 s, 42 °C). Cells from the resulting colonies were pre-inoculated into 100 ml of LB broth containing 100 µg ml^{-1} of ampicillin and grown overnight at 37 °C. The cultures were then transferred to 1 l Terrific Broth cultures (15 ml) and grown at 24 °C, 180 r.p.m. for 5–6 h until the optical density (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 r.p.m. for an additional 24 h. Cells were then harvested by centrifugation (5,000*g*, 15 min, 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 1 × *Pfu* Ultra II Hotstart Master Mix (Agilent). The Quickchange PCR cycle was performed using the following method: first a 5-min incubation at 95 °C, then cycles (95 °C for 5 min, 60 °C for 30 s, 72 °C for 6 min) were repeated 25 times; followed by 72 °C for 10 min 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 (approximately 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^{-1} of cell paste). The solution was stirred and incubated at 4 °C for 45 min before cell lysis was conducted using sonication or a high-pressure homogenizer. Sonication was conducted using the following conditions: 50 ml solution, 5 s on, 5 s off, with a total sonication time of 20 min using a microtip (70% amplitude). Cells were passed through a high-pressure homogenizer twice. Lysed cells were then spun down (1,200*g*, 12 min, 4 °C) to remove the cell debris. The resultant supernatant was then centrifuged further (56,000*g*, 1 h and 40 min, 4 °C) to collect the membrane pellet, which was then re-homogenized in buffer A (15 ml) and centrifuged again (56,000*g*, 1 h and 20 min, 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^{-1} (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 (56,000*g*, 1 h and 20 min, 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 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 for removing high concentrations of imidazole employed during the elution. The protein sample was then concentrated down to a final volume between 500 and 1,000 µl. The sample was then mixed with a 6×His-tagged SUMO protease (1.2 mg ml^{-1}) to a volume ratio of 10:1 and incubated overnight at 4 °C. The sample was then loaded onto an Äkta purification system (GE Healthcare) endowed with a multiwavelength detector (set at 280, 370 and 450 nm) and then onto a Ni-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 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, dodecyl-β-D-maltoside was used (0.03% (w/v), analytical grade), but other detergents were used for crystallization screenings at 3× their respective critical micelle concentration (CMC). The protein eluted with a very high purity and homogeneity (evaluated by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and the shape of the peak in the chromatogram, respectively) with an elution volume of 10.5–11 ml. The sample was concentrated to 100 µl using a centrifugal filter unit (50-kDa cut-off) with a final concentration ranging from 5 to 30 mg ml^{-1}.

Preparation of human FMO3 and human FMO5. Full-length cDNA encoding for *Homo sapiens* FMO3 (UniProt P31513) and FMO5 (Genbank Z47553) were cloned into a modified pET-SUMO vector (Invitrogen) to allow insertion of a cleavable N-terminal 8×His-SUMO tag. Expression, cell disruption, extraction and purification were performed according to the methods previously described for human FMO5 (ref. [37]).

Kinetic assays of the AncFMOs. Steady-state kinetics assays were performed in technical duplicates on a Jasco V-660 spectrophotometer. Enzyme activity of the ancestral proteins was measured by monitoring NADPH consumption (absorbance at 340 nm, $\varepsilon_{340} = 6.22\ mM^{-1}\ cm^{-1}$ for NADPH). The buffer used for kinetic analyses was 50 mM potassium phosphate, 250 mM NaCl, 0.05% TRX-100-R (Sigma-Aldrich), pH 7.5. For the determination of the K_M of the substrates, 100 µM, 100 µM and 50 µM NADPH were used for AncFMO2 (0.1 µM), AncFMO3-6 (0.1 µM) and AncFMO5 (2.0 µM), respectively. For the determination of the K_M for NADPH, 1 mM trimethylamine was used as the oxygen-accepting substrate for AncFMO2 and AncFMO3-6, and 30 µM of heptan-2-one was used for AncFMO5. The spectrophotometer was set at 37 °C and the NADPH and substrate mix were also incubated at 37 °C for 5 min before starting the reaction by adding the enzyme. The pH and temperature conditions were set based on literature studies for a fair comparison with previously reported properties of mammalian FMOs. $NADPH_{uncoupling}$ rates were determined in the absence of substrates.

Kinetic assays of human FMO3. Kinetic assays were performed in technical duplicates at 25 °C on a Varian spectrophotometer (Cary 100 Bio) equipped with a thermostatic cell compartment. The apparent K_M for NADPH was measured by varying NADPH concentrations from 20 to 400 µM in aerated reaction mixtures (150 µl) containing 2.5 µM human FMO3, 50 mM HEPES pH 7.5, 10 mM KCl, 5% (v/v) glycerol and 0.05% (v/v) TRX-100-R. Enzyme activity was measured by monitoring NADPH consumption (absorbance at 340 nm, $\varepsilon_{340} = 6.22\ mM^{-1}\ cm^{-1}$ for NADPH).

Rapid kinetics analysis of the AncFMOs. Stopped-flow experiments were carried out using the SX20 stopped-flow spectrophotometer equipped with either the photodiode array detector or the single-channel photomultiplier (PMT) module (Applied Photophysics, Surrey, UK). Results were obtained by mixing 50 µl of two solutions in single mixing mode. All solutions were prepared in 50 mM potassium phosphate, 10 mM NaCl and 0.05% TRX-100-R, pH 7.5 at 25 °C. For every reaction, a concentration of 10–15 µM enzyme was used, and measurements were done in technical duplicates. When needed, the solutions were supplemented with 5.0 mM glucose. Enzyme and substrate solutions were made anaerobic by flushing solutions for 10 min with nitrogen, followed by adding 0.3 µM glucose oxidase (*Aspergillus niger*, type VII, Sigma-Aldrich) to consume the leftover oxygen. The monitoring of the reductive half reaction was done by mixing the anaerobic enzyme solution with anaerobic buffer containing increasing concentrations of NADPH (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 trimethylamine, respectively. This allowed us to follow the spectral changes during the oxidative half-reactions. The C4a-(hydro) peroxyflavin intermediate formation and decay were 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 (1) air-saturated buffer; (2) equal volumes of 100% argon buffer and 100% O_2 buffer; (3) 100% O_2 buffer; (4) 100% O_2 buffer on ice. All solutions were bubbled for 10 min at room temperature, except the last one which was done on ice. In the case of AncFMO2, in order to confirm its saturating behavior, 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 analyzed using the software Pro-Data (Applied Photophysics, Surrey, UK) and GraphPad Prism 6.05 (La Jolla, CA, USA).

ThermoFAD Assays[64]**.** 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), with or without $NADP^+$ (200 µM), made to a final volume of 20 µl using the storage buffer and incubated in ice for 1 h. T_m for human FMO3 (0.05% (v/v) TRX-100-R) was determined in technical duplicates and with a final protein concentration of 5 µM in buffer (100 mM), with varying pH values (pH 6–6.5 MES, pH 7–8 HEPES, pH 8–9 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 1). Reaction mixtures (1.0 ml) contained 5.0 mM substrate (<1% ethanol), 0.1 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 h 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^{-1} 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 in a sitting drop. Sitting drop was 2 µl after mixing and the reservoir was 1 ml. Prior to crystallization, $NADP^+$ (1 mM final) was incubated with 12–15 mg ml^{-1} of AncFMO2 for 1 h at 4 °C. After 2 d, large yellow crystals formed. AncFMO3-6: 12–15 mg ml^{-1} of AncFMO3-6 (in storage buffer and CYMAL-6 (0.09% (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 in a sitting drop. Sitting drop was 2 µl after mixing and the reservoir was 1 ml. Prior to crystallization, $NADP^+$ (1 mM final) was incubated with 12–15 mg ml^{-1} of AncFMO3-6 for 1 h at 4 °C. After 1 d, large yellow crystals formed. AncFMO5: 12 mg ml^{-1} of AncFMO5 (in storage buffer and dodecyl-β-D-maltoside (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 in a sitting drop. The sitting drop was 2 µl after mixing and the reservoir was 1 ml. Prior to crystallization, $NADP^+$ (1 mM final) was incubated with 12–15 mg ml^{-1} of AncFMO5 for 1 h at 4 °C. After 1 d, large yellow hexagon-shaped crystals formed. During crystal fishing, a cryo-protectant was prepared containing modified crystallization 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 (Table 2). STARANISO was additionally used for AncFMO2, which suffered greatly from anisotropy[80]. The phase problem was solved by molecular replacement using a recently solved insect FMO (PDB 5NMW)[81] as a search model, and then AncFMO2 for the subsequent AncFMOs, using Phaser and Molrep[82]. The phases were greatly improved by density averaging with DM[79,83]. Model building and refinement were then conducted using COOT[84], Buccaneer[79] and Refmac5 (ref. [85]) (Table 2). The outliers in the Ramachandran plots were 4.4%, 6.6%, 2.3% and 0.4% of the residues for AncFMO2, AncFMO2 bound to $NADP^+$, AncFMO3 and AncFMO5, respectively. Figures were then generated using UCSF Chimera[86], PyMOL (DeLano Scientific; www.pymol.org) and CCP4mg[79].

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

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). Source data for Figs. 2 and 3, and Table 1, are available with the paper online.

References

73. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1166 (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. 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).
81. 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).
82. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
83. 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).
84. 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).
85. 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).
86. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Acknowledgements

The research for this work has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 722390; the Italian Ministry of Education, University and Research (MIUR) under the "Dipartimenti di Eccellenza (2018–2022)" program; and ANPCyT (Argentina) PICT 2016-2839 to M.L.M. M.L.M. is a member of the Research Career of CONICET, Argentina. The authors thank M. J. Ayub for his valuable comments and contributions on the manuscript.

Author contributions

All listed authors performed experiments and analyzed data. C.R.N. generated purification protocols, crystallized the AncFMOs, collected the corresponding datasets at the ESRF and SLS facilities, performed structural analysis and elucidated the AncFMO structures. G.B. and C.R.N. performed Golden Gate cloning to insert the AncFMO genes into their respective vectors, designed by G.B. G.B., C.R.N. and F.F. carried out mutagenesis and extensive kinetic analysis and validated the substrate profiles using stopped-flow ultraviolet–visible spectroscopy and GCMS for each AncFMO. M.L.M. conducted thorough evolutionary analyses and performed ancestral-sequence reconstruction to obtain AncFMO protein sequences. C.R.N., G.B. and M.L.M. prepared the figures. C.R.N. wrote the manuscript and A.M., M.W.F. and M.L.M. edited it. All authors provided critical feedback and helped shape the research, analysis and manuscript. A.M. and M.W.F. conceived the original idea.

Competing interests

The authors declare no competing interests.

Additional information

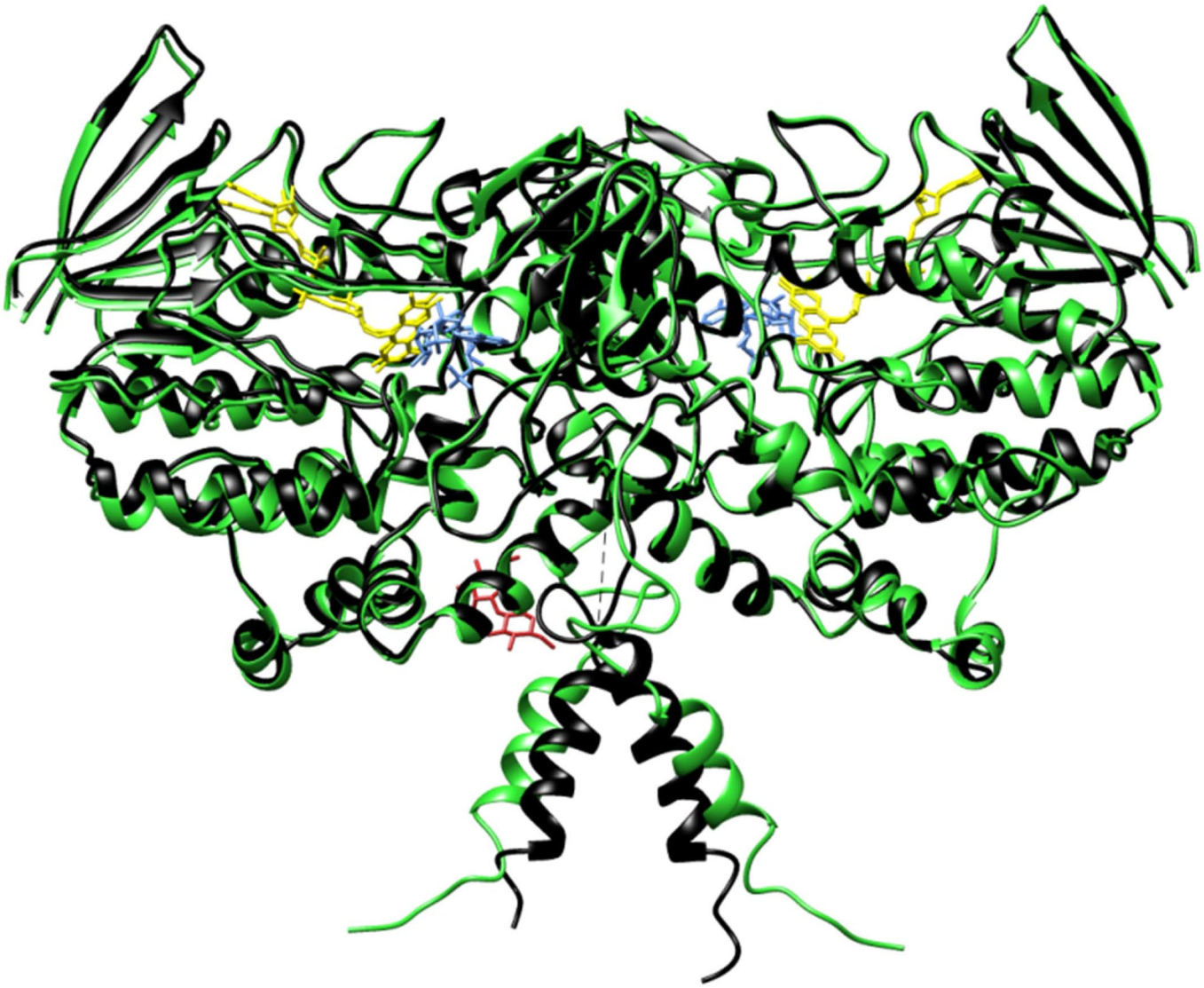
Extended data is available for this paper at https://doi.org/10.1038/s41594-019-0347-2.

Supplementary information is available for this paper at https://doi.org/10.1038/s41594-019-0347-2.

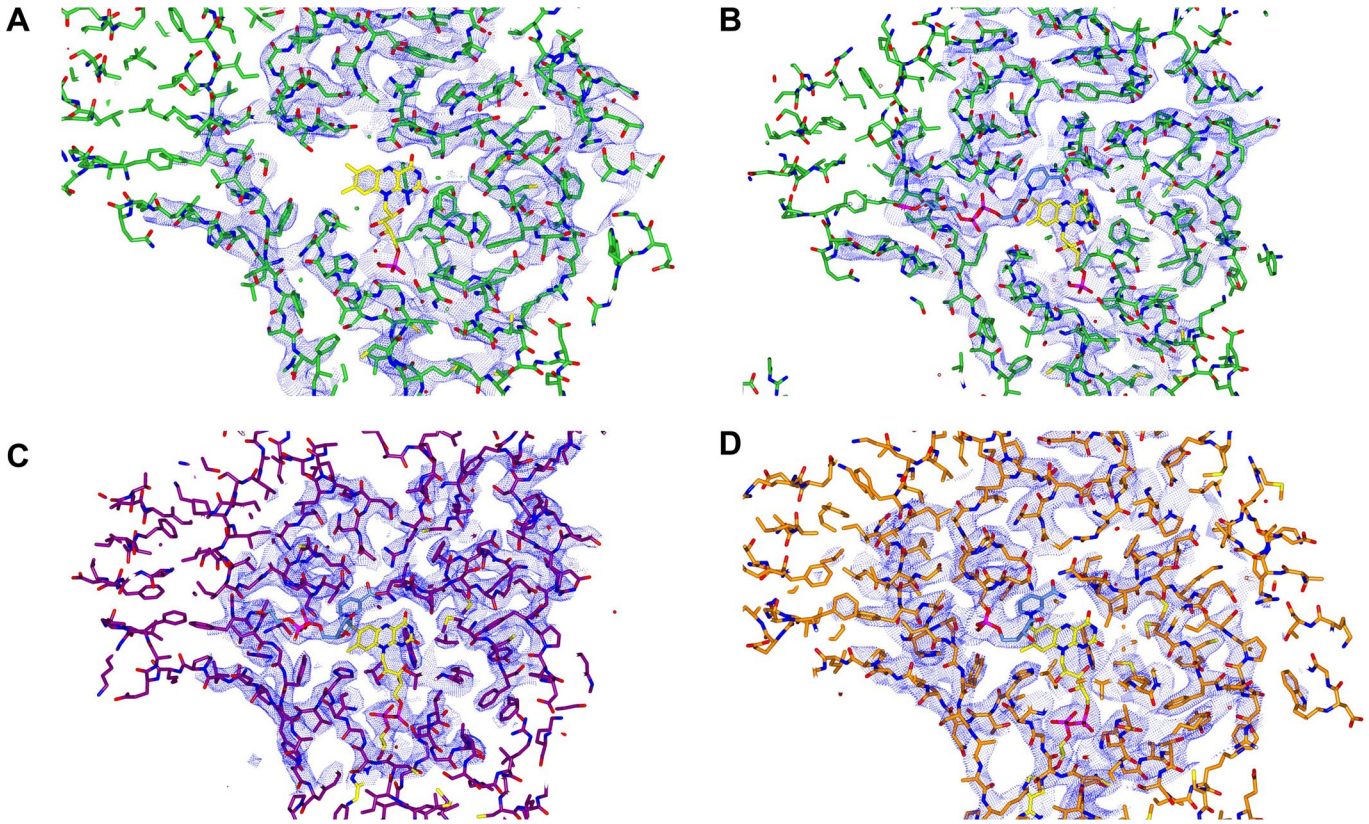
Correspondence and requests for materials should be addressed to M.L.M., M.W.F. or A.M.

Peer review information Katarzyna Marcinkiewicz was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

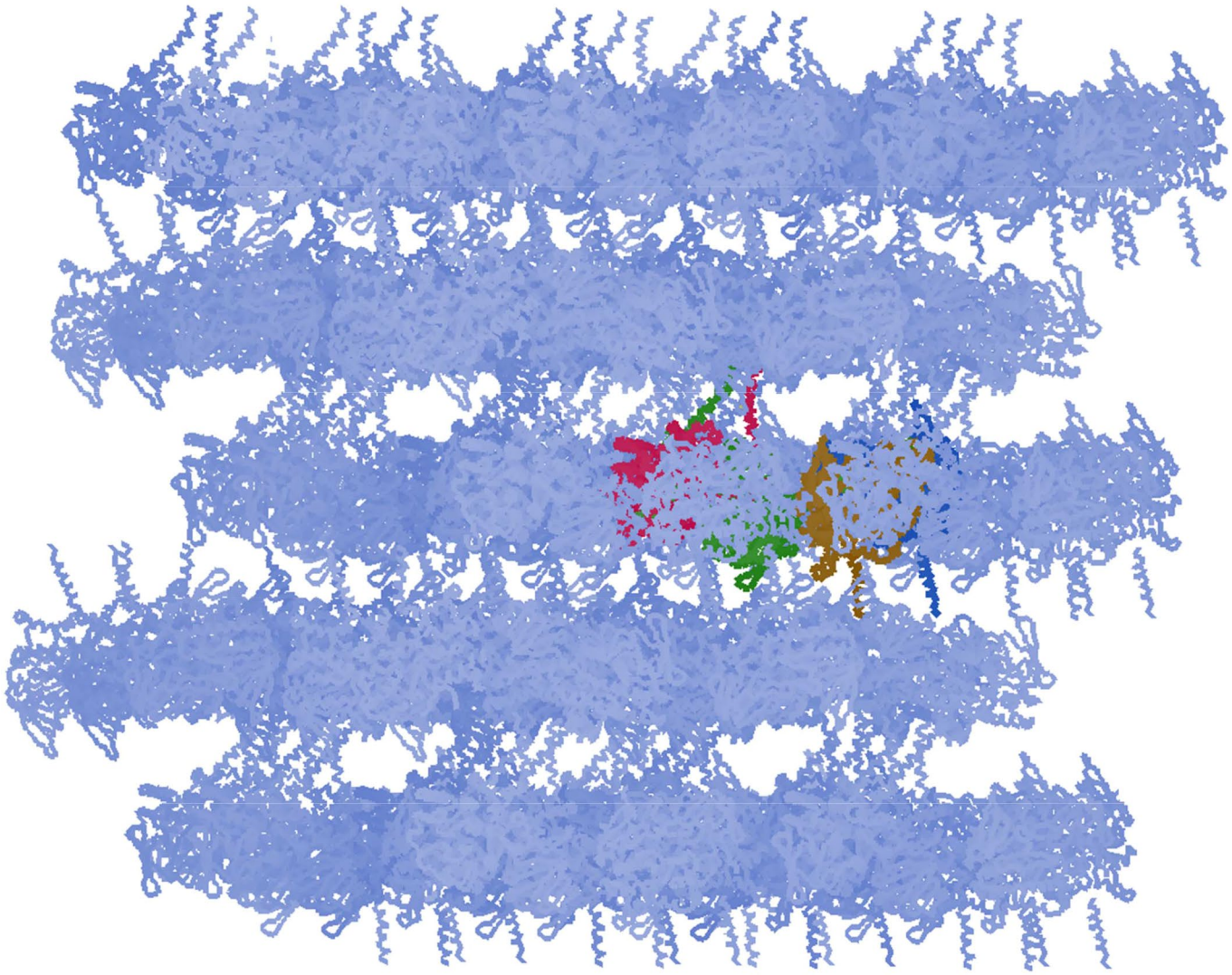
Reprints and permissions information is available at www.nature.com/reprints.



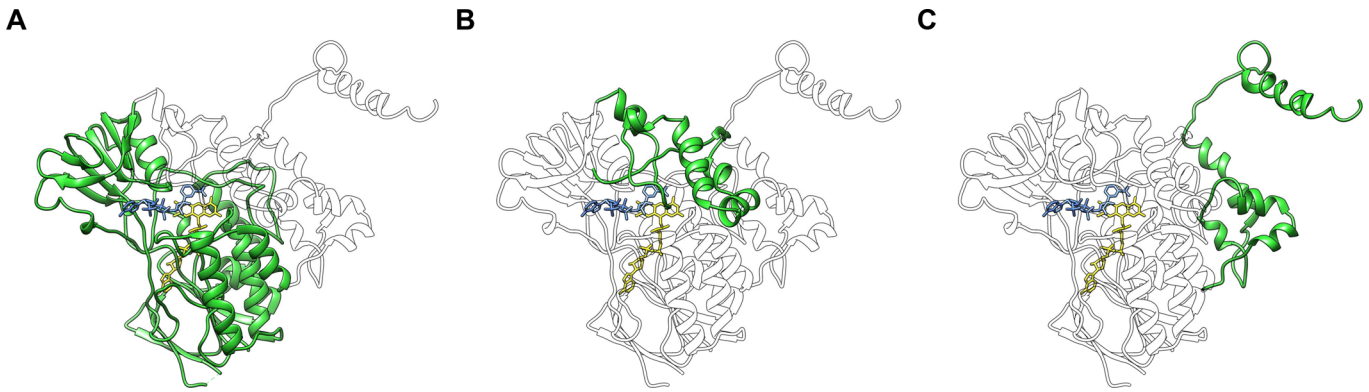
Extended Data Fig. 1 | Crystal Structure of AncFMO2 in the absence of $NADP^+$ (green) superposed to the structure bound to $NADP^+$ (dark green). AncFMO2 crystallizes in an identical manner with or without $NADP^+$. The root-mean-square deviation between the native AncFMO2 and its $NADP^+$ complex is 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 Fig. 4).



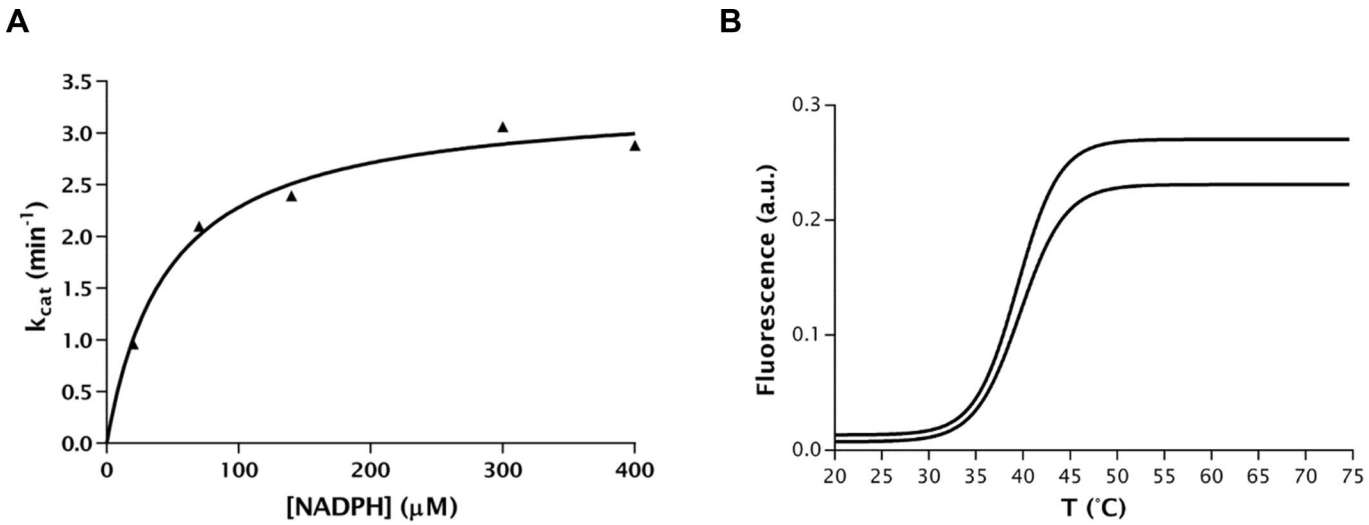
Extended Data Fig. 2 | Electron density maps of the AncFMOs. Structure determination was greatly facilitated by density averaging because the asymmetric unit of the crystals of AncFMO2, AncFMO3-6 and AncFMO5 contain four, six and two protein molecules, respectively. The depicted 2Fo-Fc maps were calculated 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.



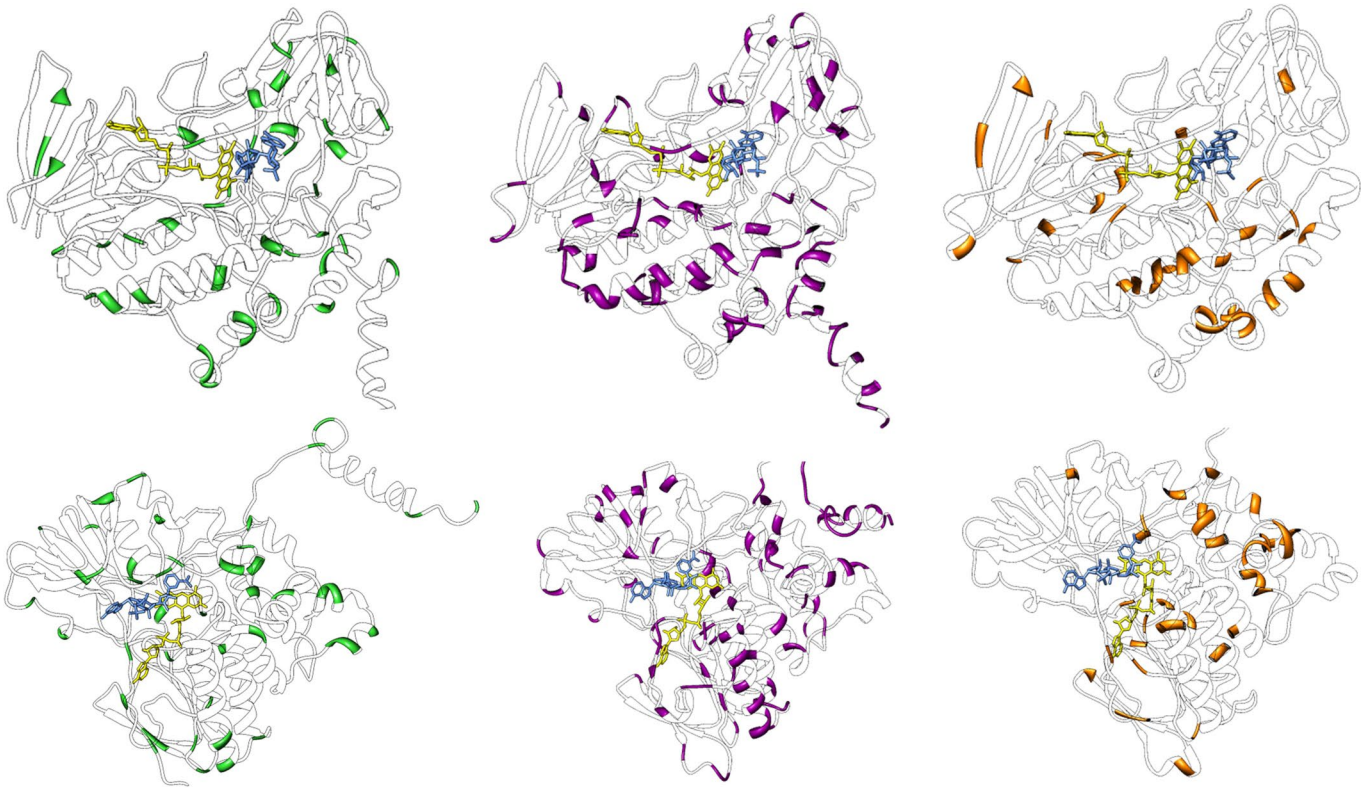
Extended Data Fig. 3 | The crystal packing of AncFMO2 forms multiple planes of soluble dimer–dimer interactions that extend across the lattice. The asymmetric unit is depicted by the four differently colored monomer units of dark yellow, dark red, dark green and dark blue. In between each plane, we see multiple transmembrane helices projecting upwards and downwards from each asymmetric unit. Each dimer projects its transmembrane helices towards its reciprocal dimer.



Extended Data Fig. 4 | 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 α-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.



Extended Data Fig. 5 | NADPH oxidase activity and melting temperature of human FMO3. 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 human FMO3 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. All measurements were performed in technical duplicates.



Extended Data Fig. 6 | Differing residues between AncFMOs and human FMOs. The upper and lower panels display the structures in two orientations. The changes exhibited by AncFMO2, AncFMO3-6 and AncFMO5 compared to human FMOs 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 localized 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 human FMOs and the AncFMOs sequences, with the sequences at the subdomains and C-termini varying substantially (Supplementary Fig. 4). 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 Fig. 6d (lower panel).

nature research

Corresponding author(s): Andrea Mattevi

Last updated by author(s): Oct 21, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Dataset were constructed using BLAST or HMMER tools and databases UniProtKB, Reference proteomes, Swiss Prot
Data analysis	Ancestral Sequence Reconstruction: MAFFT v7, ProtTest v3.4, PhyML v3.0, RAxML v0.6.0, PAMLX v.4.9. Enzyme kinetics: Pro-Data (Applied Photophysics, Surrey, UK), GraphPad Prism 6.05. Crystallography: STARANISO, CCP4 Program Suite v7.0.076 including COOT v0.8.9.2, UCSF Chimera, PyMOL v2.0, ccp4mg v2.10.0, HOLLOW v1.2.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequences, coordinates and structure factors will be made available through the PDB. Kinetic data will be available upon requests to the authors.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Not relevant
Data exclusions	Not relevant
Replication	Not relevant
Randomization	Not relevant
Blinding	Not relevant

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging