

University of Groningen

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](https://doi.org/10.1038/s41594-019-0347-2)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Final author's version (accepted by publisher, after peer review)

Publication date:
2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Nicoll, C. R., Bailleul, G., Fiorentini, F., Mascotti, M. L., Fraaije, M. W., & Mattevi, A. (2019). Ancestral-sequence reconstruction unveils the structural basis of function in mammalian FMOs. *Nature Structural & Molecular Biology*, 27(1), 14-24. <https://doi.org/10.1038/s41594-019-0347-2>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

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

3
4 Callum R. Nicoll¹, Gautier Bailleul², Filippo Fiorentini¹,
5 María Laura Mascotti^{3,*}, Marco W. Fraaije^{2,*} and Andrea Mattevi^{1,*}
6

7 ¹ Department of Biology and Biotechnology “Lazzaro Spallanzani”, University of Pavia, via Ferrata
8 9, 27100 Pavia, Italy

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

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

13
14 *Correspondence to María Laura Mascotti, Marco Fraaije, Andrea Mattevi
15 E-mail: mlmascotti@unsl.edu.ar, m.w.fraaije@rug.nl, andrea.mattevi@unipv.it
16

17 Abstract

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

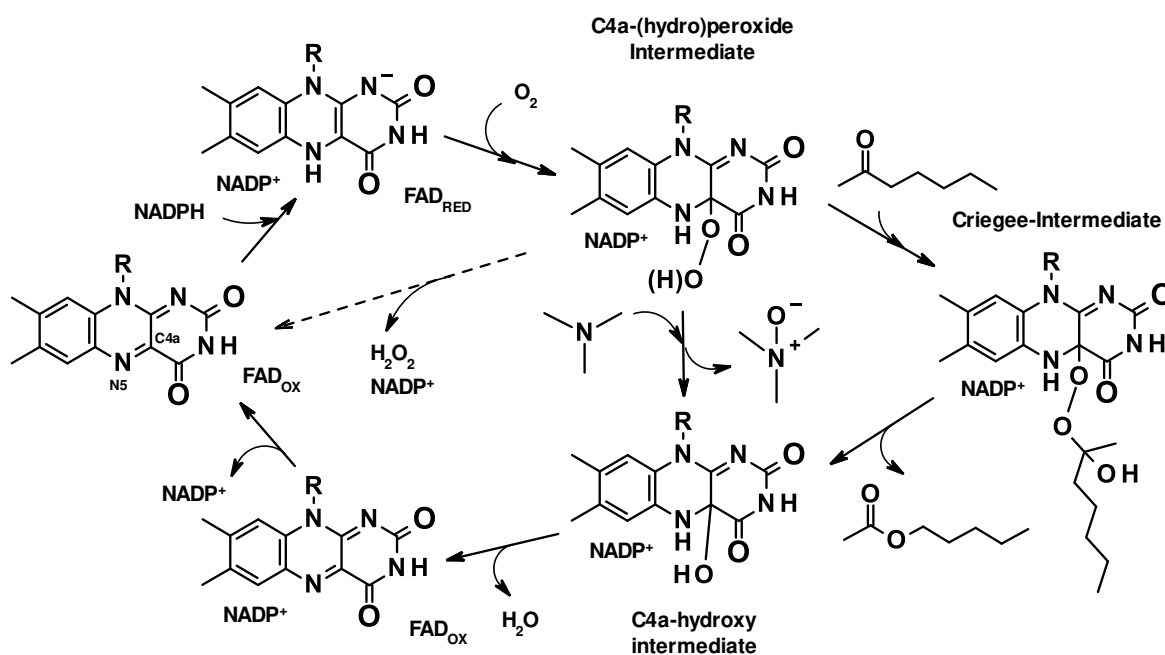
31

32 Introduction

33 Xenobiotic metabolism is an ancient and imperative process pursued by all organisms. With
34 evolution resulting in the production of, and thus exposure to, a vast number of noxious and toxic
35 natural products,¹ organisms have employed a multitude of intricate detoxification systems, to
36 tackle the sheer quantity of diverse chemicals.¹⁻⁶ Flavin-containing monooxygenases (FMOs; EC
37 1.14.13.8) represent one of these detoxifying protein families and are prevalent in all domains of
38 life.^{4,7} FMOs are members of the Class B flavin-dependent monooxygenases and utilize the
39 cofactors FAD and NADP(H), and dioxygen for activity.⁸⁻¹⁰ Typically, FMOs pursue catalysis as
40 illustrated in **Scheme 1**, whereby a soft nucleophile (here demonstrated with trimethylamine)
41 receives the distal oxygen atom from the C4a-(hydro)peroxyflavin intermediate.^{11,12} The more
42 water-soluble hydroxylated product is then released by the enzyme to be excreted from the host.

43 Humans possess five FMO **isoforms that** are differentially expressed in many different tissues such
44 as the kidney, lung, and liver.^{2,10,13,14} The **human FMO** genes are found on chromosome 1, with
45 *FMO1-4* clustering over 220 kb, and *FMO5* found on a separate chromosome region.^{15,16} The
46 **human FMO** family contains six non-expressed pseudogenes which are also located on
47 chromosome 1.¹⁶ FMOs are involved in phase I of xenobiotic detoxification.^{2,3} They oxidize an
48 array of compounds bearing soft nucleophilic centers such as nitrogen and sulfur atoms,^{2,17-19}
49 making them clinically important regarding drug metabolism.^{3,6,12,15,17,20,21} The most extensively
50 characterized FMO is **human FMO3**, renowned for its production of trimethylamine N-oxide.²²⁻²⁶

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



67
 68 **Scheme 1.** The catalytic mechanism of FMOs. FAD_{ox} is reduced by NADPH. FAD_{red} is
 69 consequently oxidized by a molecule of dioxygen to generate the C4α-(hydro)peroxide
 70 intermediate. The typical mode of action of FMOs with the distal oxygen atom from the
 71 intermediate being inserted onto a soft nucleophile through nucleophilic addition is shown with
 72 reference to trimethylamine. The Baeyer-Villiger monooxygenation activity conducted by human
 73 FMO5 is shown on the right with reference to heptan-2-one. The dotted arrow indicates the
 74 uncoupling reaction whereby the C4α-(hydro)peroxide intermediate decays with the release of
 75 NADP⁺ and hydrogen peroxide.

76

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

88

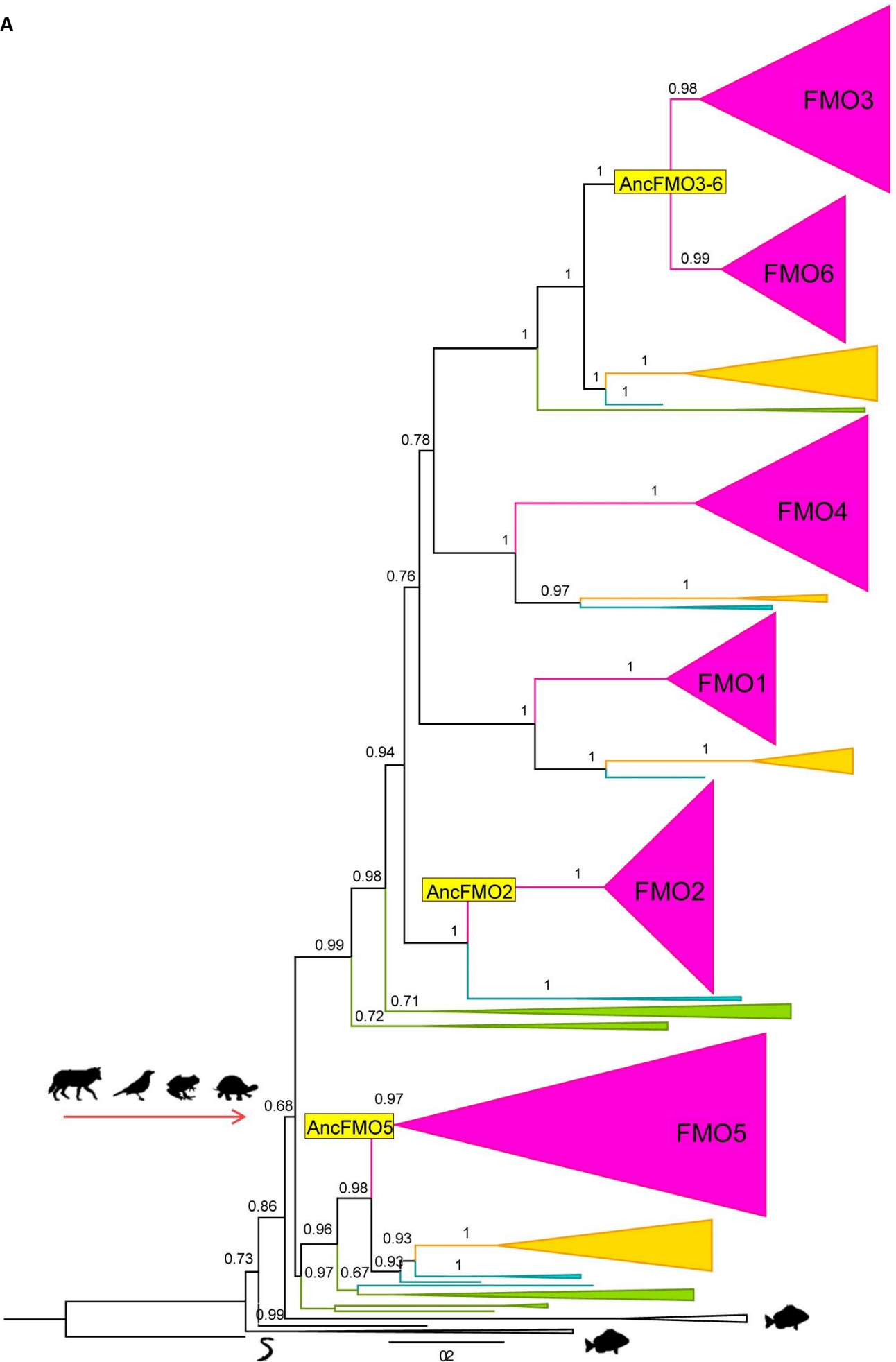
89 **Results**

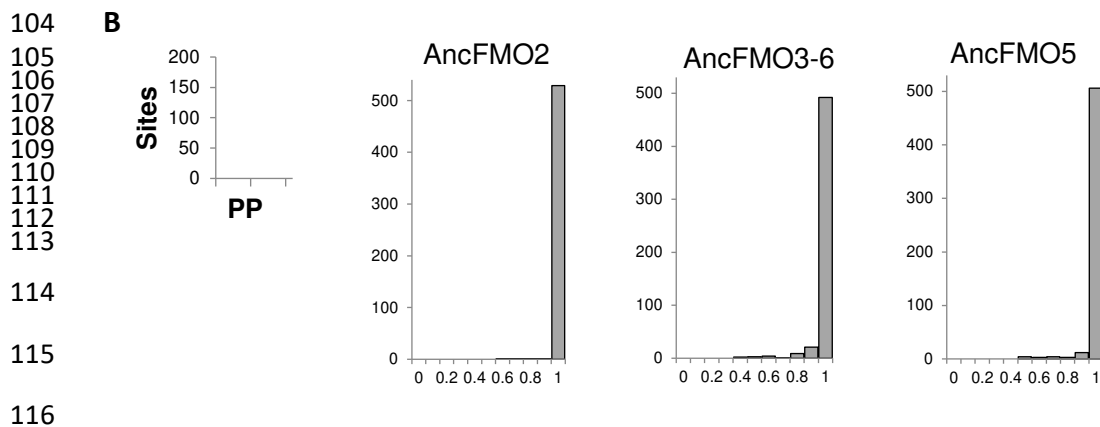
90 *Ancestral sequence reconstruction of mammalian flavin-containing monooxygenases*

91 We inferred the evolutionary history of FMOs from a full phylogeny constructed by including
92 experimentally-characterized enzymes from Bacteria and Eukarya, plus sequences found by
93 extensive sequence homology searching and HMM profiling (**Supplementary Figure 1 and Data 1**).
94 **Our work confirmed the findings of the previous studies by Hernandez et al.¹⁶ and Hao et al.⁴⁴: (i)**
95 **jawed vertebrate FMOs are monophyletic and derived from a single common ancestor (Figure 1A,**
96 **Supplementary Figure 2); (ii) several duplication events occurred in the terrestrial vertebrates; (iii)**
97 **the ancestor of mammals already encoded the five FMO paralogs resulting from four major gene-**
98 **duplication events (Figure 1A, Supplementary Figure S2 and Data 1); (iv) a sixth mammalian**
99 **paralog (FMO6) resulted from a late gene duplication event. FMO6 has been described as a**
100 **pseudogene in humans⁴⁵ but it might be functional in mouse¹⁶ and its nature is unknown in other**
101 **mammals.**

102

A





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

128
129

130 By performing ancestral sequence reconstruction, we obtained the protein sequences of AncFMOs
 131 from mammals with high posterior probabilities (ranging from 0.98-0.99) (**Figure 1B**,
 132 **Supplementary Data 2**). In the phylogeny, we observed that FMO5 diverged earlier in agreement
 133 with previous reports,⁴⁴ followed by FMO2, FMO1, FMO4 and the FMO3-6 hybrid. This topology
 134 suggests that the gene duplication events took place simultaneously rendering no clear paralog
 135 couples as it has been previously proposed (**Figure 1A**, **Supplementary Figure 2**).^{16,44} Among the
 136 whole clade of present-day FMO2s, 80% of sites are conserved, while the rest are likely responsible
 137 for functional differences among species. We observe that from AncFMO2 to **human FMO2**, 42
 138 substitutions have occurred of which 18 are conservative (as defined by Grantham⁴⁷). In the case of
 139 AncFMO3-6, the ancestor underwent an early duplication event originating the FMO3 and FMO6
 140 paralogs in mammals. As a general trend, comparing the pre-duplication ancestor to modern FMO3
 141 and FMO6, 70% of the sites are conserved. Along each branch to the **human FMO3** or **human**
 142 **FMO6** sequences, 94-98 substitutions occurred, 28-30 of them being conservative. The lower
 143 degree of conservation is not surprising considering the duplication scenario. Finally, FMO5 is the
 144 most enigmatic of all extant FMOs due to its Baeyer-Villiger oxidation activity.³⁷ AncFMO5 shows
 145 44 changes along the branch to **human FMO5**, with 19 conservative substitutions. In light of this

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

148

149 *AncFMOs portray catalytic rates similar to those of extant mammalian FMOs*

150 Critically for our project, the hitherto generated AncFMOs sequences proved to encode stable
151 proteins that can be effectively produced and purified as recombinant, FAD-loaded, and
152 catalytically competent enzymes in *E. coli*. Thus, the first relevant result was that a convenient
153 bacterial expression system for the study and biocatalytic exploitation of close homologs to human
154 FMOs was established (see materials and methods). We next verified whether these enzymes
155 retained enzymatic activities by performing steady-state kinetics experiments using a NADPH-
156 depletion spectrophotometric assay. The NADPH oxidase activity was initially tested (NADPH
157 consumption in absence of an organic substrate; NADPH_{uncoupling} in Table 1). This was followed by
158 the measurements of the kinetics of the reaction in the presence of known oxygen-accepting
159 substrates of FMO2 and FMO3 (methimazole, trimethylamine and thioanisole), and FMO5 (heptan-
160 2-one). The results were reassuring in that AncFMO2, AncFMO3-6, and AncFMO5 proved to be
161 enzymatically active with kinetic parameters very similar with those reported for their extant
162 human-derived enzymes.^{2,17,34,37,48-51} The k_{cat} , K_M^{NADPH} and uncoupling values ranged between
163 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
164 that the AncFMOs displayed a high affinity towards the coenzyme NADPH and a significantly
165 higher NADPH consumption rate when a suitable substrate was present. This result is in full
166 agreement with the canonical catalytic mechanism observed for FMOs and sequence related
167 flavoprotein monooxygenases (Scheme 1). These features were further demonstrated by stopped-
168 flow kinetic studies. NADPH-reduced AncFMO2 and AncFMO3-6 were found to react rapidly with
169 oxygen to form a stable and detectable C4α-(hydro)peroxyflavin intermediate with its well-
170 defined spectroscopic properties (Figure 2). Based on the steady-state kinetics data, AncFMO5 is
171 assumed to behave similarly. Collectively, these experiments convincingly demonstrated that our
172 AncFMO2/3-6/5 enzymes are enzymatically competent and exhibit the typical catalytic features of
173 Class B flavoprotein monooxygenases.

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

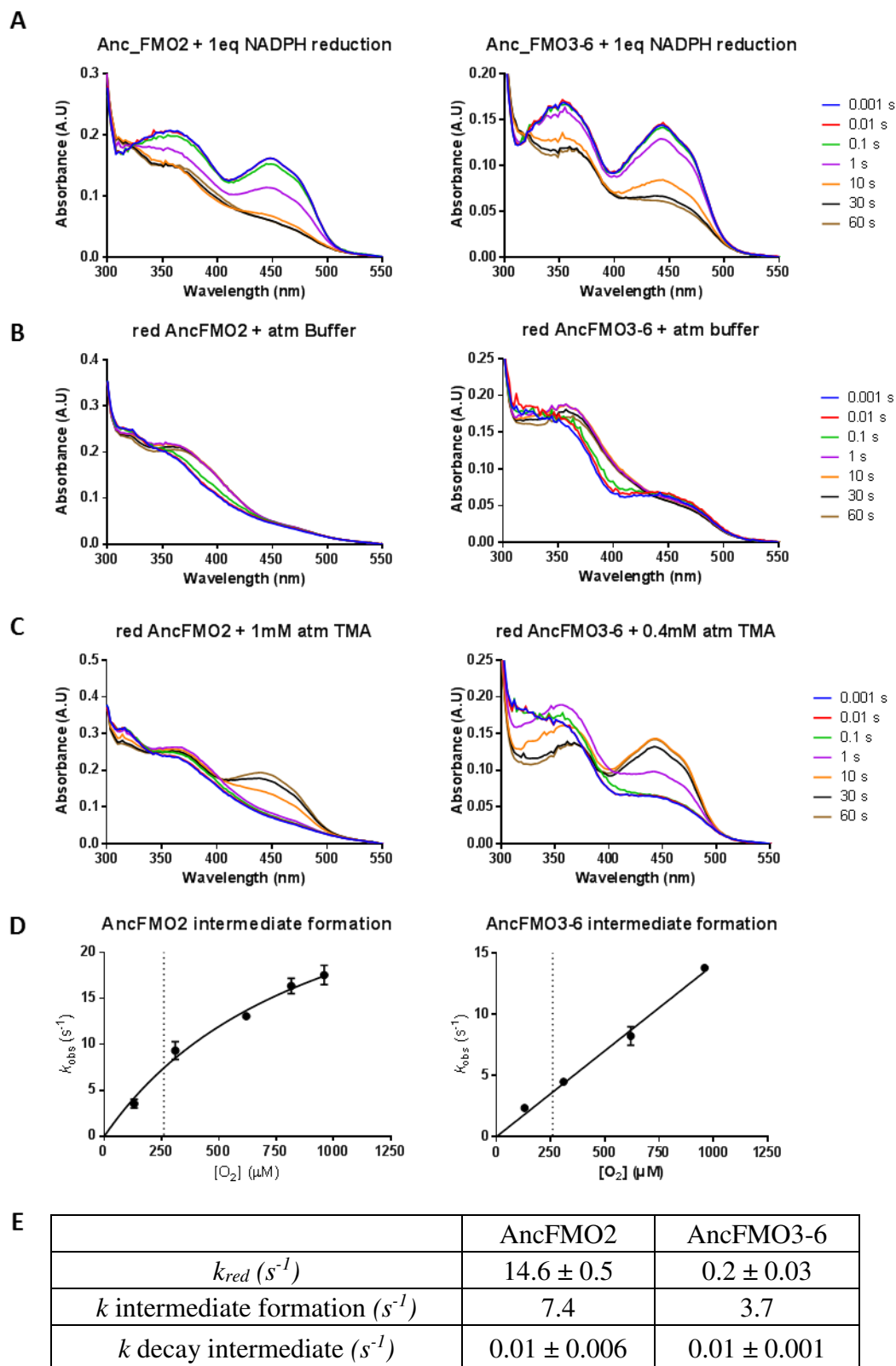


Figure 2. Stopped-flow kinetics studies on AncFMO2 and AncFMO3-6. **A:** Enzyme reduction upon the anaerobic addition of NADPH. **B:** Mixing reduced enzyme with dioxygen (0.13 mM) reveals the appearance of a peak at 360 nm which is characteristic for a C4a-(hydro)peroxyflavin intermediate (Scheme 1).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

223 (hydro)peroxyflavin intermediate which subsequently decays to form the reoxidized flavin species.
 224 **D:** Dependence of the rate of C4a-(hydro)peroxyflavin formation ($A_{360\text{nm}}$) on varying oxygen
 225 concentrations. The dotted lines correspond to the atmospheric concentration of dioxygen (0.26
 226 mM). For AncFMO2, the observed saturation behavior suggests a binding event taking place before
 227 dioxygen reacts with the reduced flavin. Interestingly, such a saturation behavior was also reported
 228 for pig liver FMO1.⁵² **E:** Rates of reduction, C4a-(hydro)peroxyflavin formation, and C4a-
 229 (hydro)peroxyflavin decay in the absence of substrate (0.26 mM dioxygen; dotted line on panels D).
 230

231 **Table 1:** Steady state-kinetics.

	Substrate ^a	k_{cat} (s ⁻¹)	K_M (μM)
<i>Ancient FMOs</i>			
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
<i>Extant FMOs</i>			
human FMO3	NADPH ^c	0.06 ± 0.16	46 ± 9
human FMO5	NADPH ^c	0.197 ± 0.009	59 ± 8

232 ^a Rates were determined by following NADPH consumption (absorbance decrease at 340 nm). The
 233 buffer was composed of 50 mM potassium phosphate (pH 7.5), 250 mM NaCl, 0.05% (v/v) triton
 234 X-100 reduced. The reactions were run at 37 °C. For the determination of the K_M of the substrates,
 235 100 μM and 50 μM NADPH was used for AncFMO2 and AncFMO3-6, and AncFMO5
 236 respectively. For the determination of the K_M for NADPH, 1 mM trimethylamine was used as
 237 oxygen-accepting substrate for AncFMO2 and AncFMO3-6, whilst 30 mM of heptan-2-one was
 238 used for AncFMO5. NADPH_{uncoupling} rates were determined in the absence of substrates. The
 239 increase (“burst”) in NADPH consumption rates upon addition of the substrates demonstrate that
 240 the AncFMOs are highly coupled and effectively oxygenate their substrates.

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

243 ^c The rates for NADPH consumption in the presence and absence of the substrate are the same
 244 because of a high degree of uncoupling in the extant human FMOs. The data for human FMO3 are
 245 shown in Supplementary Figure 8. The data for human FMO5 are taken from Fiorentini et al.³⁷
 246

247

248

249

250

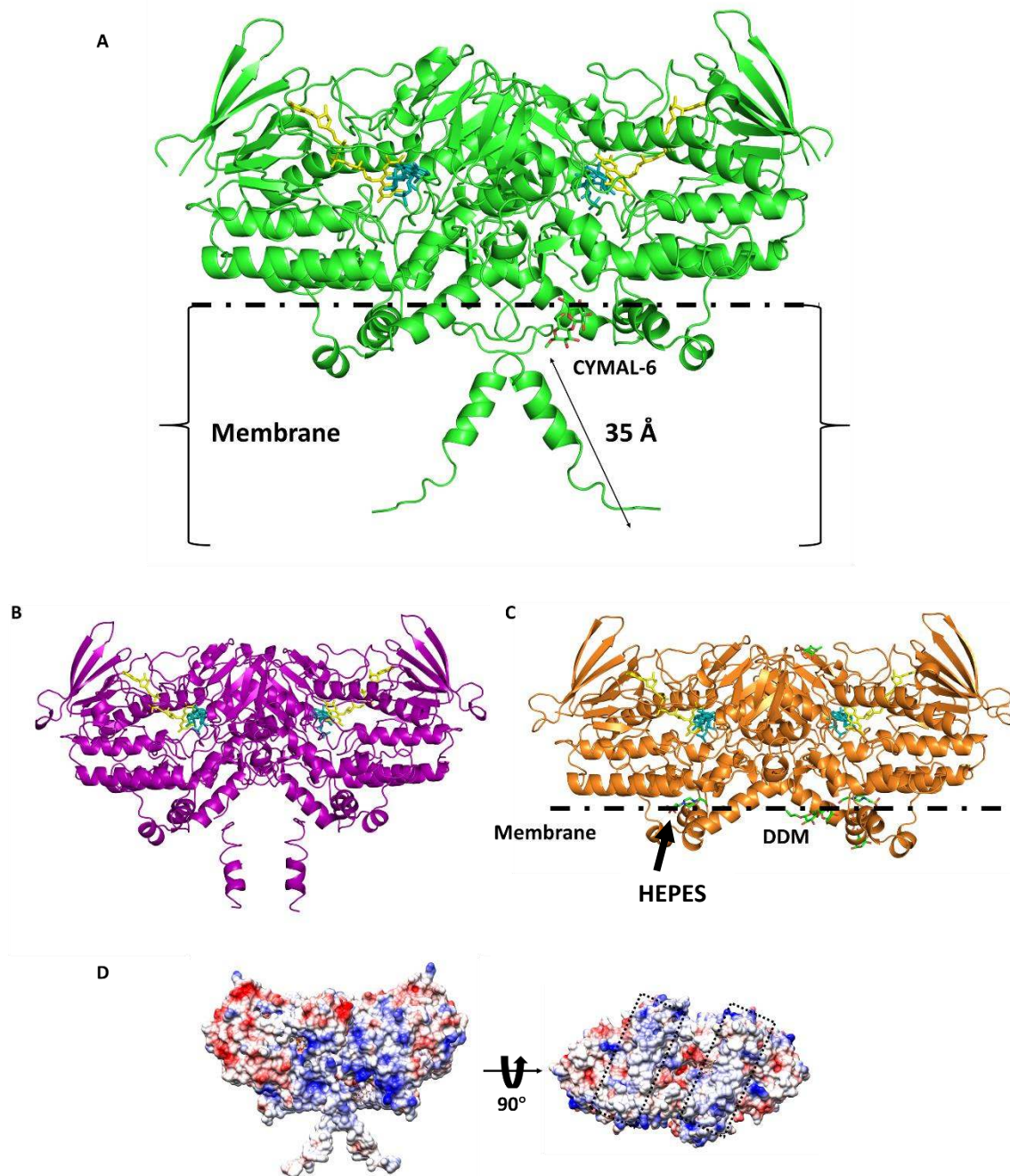
251 *AncFMOs crystallize as dimers with extensive membrane binding features*

252 To investigate the role of the AncFMOs in detail, the crystal structures of each AncFMO in the
253 presence of NADP⁺ were determined (Figure 3). AncFMO2 was also crystallized in the absence of
254 NADP⁺ but, akin to Class B flavin-dependent monooxygenases, no major conformational changes
255 were observed between the apo- or holo-enzyme crystal structures (Supplementary Figure 3). The
256 structures were solved at 2.7, 3.0, 2.8, and 2.7 Å resolution for AncFMO2 (without NADP⁺),
257 AncFMO2, AncFMO3-6, and AncFMO5 (all including NADP⁺), respectively (Supplementary
258 Table 1, Supplementary Figure 4). For the purpose of the structural analysis, it must be highlighted
259 that the AncFMOs display high sequence identities to their extant human FMO counterparts: 92%,
260 83%, and 92% for AncFMO2, AncFMO3-6, and AncFMO5 respectively, making them excellent
261 structural models of human FMOs (Supplementary Figure 5).

262 Our crystal structures depict the AncFMOs as dimers: they possess an extensive monomer-
263 monomer interface over approximately 2000 Å² (calculated by the PISA server).⁵³ Furthermore,
264 their well-conserved FAD and NADP(H) dinucleotide-binding domains are accompanied by two
265 large trans-membrane helices (one from each monomer) that project outwards, approximately
266 parallel to the twofold axis (Figure 3A-D). Pairwise structural superpositions of AncFMO2,
267 AncFMO3-6, and AncFMO5 show that their ordered ~480 Cα atoms overlap with root-mean-
268 square deviations of less than 1 Å. This result reveals a high degree of structural similarity among
269 the FMO structures. We additionally notice that the dimerization interface of the AncFMOs is
270 different from the dimer interfaces exhibited by soluble FMOs (e.g. FMO from *Roseovarius*
271 *nubinihibens*, PDB entry 5IPY; FMO from *Methylophaga aminisulfidivorans*, PDB entry 2VQ7).

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

283



284

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

297

298 It was reported that truncation of the C-terminal helices was insufficient for protein
299 solubilization.^{54,55} This indicated that the enzyme possessed additional membrane-binding features.
300 To understand what elements promote membrane association, the charge distribution on the protein
301 surface was scrutinized. Intriguingly on the underside of the dimer, two large hydrophobic strips,
302 about 30 Å in length, extend across the enzyme surface (Figure 3D). These strips are lined by an
303 extensive ring of positively charged residues. Collectively, these features equip the enzyme for
304 binding to the membrane surface. The array of hydrophobic residues penetrate, monotonically, into
305 the phospholipid bilayer and are held in place by the ionic-based interactions introduced between
306 the negatively charged, polar head group of the phospholipids and the positively charged amino
307 acids. Serendipitously, in the crystal structures of AncFMO2 and AncFMO5, we were able to
308 observe the polar head groups of the detergent molecules, CYMAL-6 and dodecyl-β-D-maltoside
309 (DDM) respectively, that were used for protein solubilisation and crystallisation (Figure 3A, 3C,
310 3D). These molecules delineate the membrane-enzyme interface and further validates that the
311 hydrophobic strips monotonically embed within the membrane. These findings rationalize the
312 extensive membrane binding nature employed for this class of enzymes and corroborates that
313 truncation of the C-terminal helix alone is not sufficient to facilitate protein solubilization.⁵⁴ FMOs
314 employ both bitopic and monotopic membrane-binding features to grapple onto the membrane
315 effectively and abstract lipophilic substrates from within the membrane.

316

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

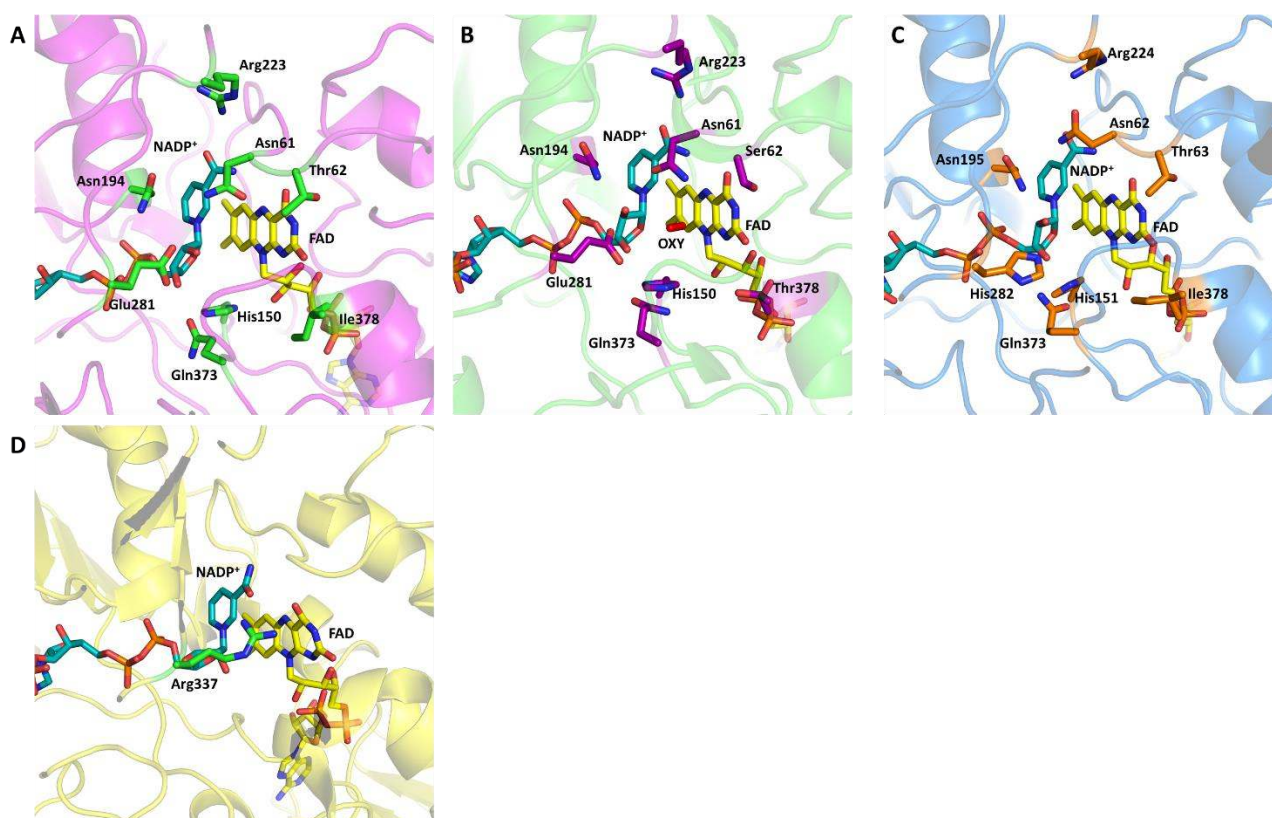
318 To comprehend the unique and distinct structural features associated with mammalian FMOs, we
319 compared them with structurally characterized soluble FMOs. Consistent with Class B flavin-
320 dependent monooxygenases, the AncFMOs have two well-conserved dinucleotide-binding domains
321 for cofactors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

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

335

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

337 With the AncFMOs active sites no longer being open clefts like their soluble homologs, we
338 scrutinized each closely to determine the functions of each residue and whether the mode of
339 NADP(H) binding is akin to Class B flavin-dependent monooxygenases. Notably, most residues in
340 the active and NADPH-binding sites are conserved with near-identical conformations (**Figure 4A-**
341 **C**). Thr62/Ser62/Thr63 for AncFMO2/3-6/5 respectively, are within hydrogen bonding distance to
342 the N3 atom of the isoalloxazine ring and orientate the FAD towards the substrate pocket.
343 Additionally, Asn61/61/62 is observed in all active sites and is strictly conserved among **human**
344 **FMOs** (**Supplementary Figure 5**). This residue situates close to the C4a of the isoalloxazine ring
345 (4.6 Å) and is likely fundamental for the stabilization of the flavin-(hydro)peroxide intermediate
346 (**Scheme 1**). Consistently, mutating this residue in **human FMO3** causes trimethylaminuria, further
347 verifying its integral role within the active site.^{12,28}



348

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

354 molecule of dioxygen (OXY). For the sake of comparison, panel **D** shows the binding of NADP⁺ to
355 the active site of phenylacetone monooxygenase, a prototypical class B monooxygenase (PDB code
356 2ylr). Arg337 is a conserved residue that is essential for the Baeyer-Villiger activity of this and
357 similar enzymes.
358

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

373

374 ***A Glu-His mutation in the mammalian specific insertion may promote Baeyer-Villiger oxidation*** 375 ***in FMO5***

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

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

397

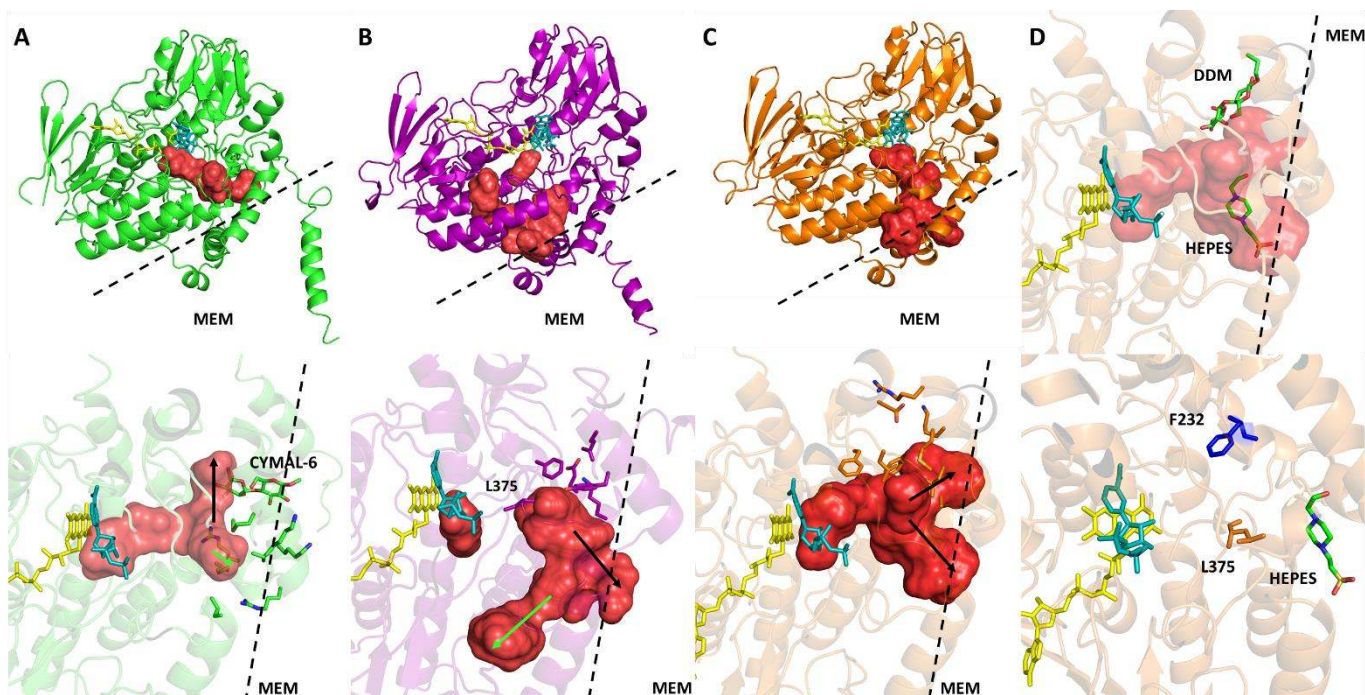
398 *AncFMOs possess a conserved substrate tunnel that branches out towards the membrane*

399 As the mammalian FMOs are notorious for their broad substrate profiles, we conducted extensive
400 research to elaborate how the substrates navigate through the enzyme using the *HOLLOW* server.⁶³

401 The conserved tunnel is roughly perpendicular to the face of the isoalloxazine ring and extends
402 outwards (approximately 16 Å) towards the membrane, before deviating in multiple directions
403 (Figure 5). In all three structures, the inner segment of the tunnel features a conserved leucine that
404 acts as a gate keeper (Leu375 in AncFMO3-6; Figure 5B, lower panel): in an upward position, it
405 creates a closed cavity at the active site (AncFMO3-6), and in the downward position, it opens the
406 tunnel to the protein-membrane interface (AncFMO2 and AncFMO5). This leucine is also
407 conserved in human FMO1-3 and 5, implying an integral role in gating the inner “catalytic” part of
408 the tunnel and affording a solvent-protected environment for catalysis (Supplementary Figure 5).

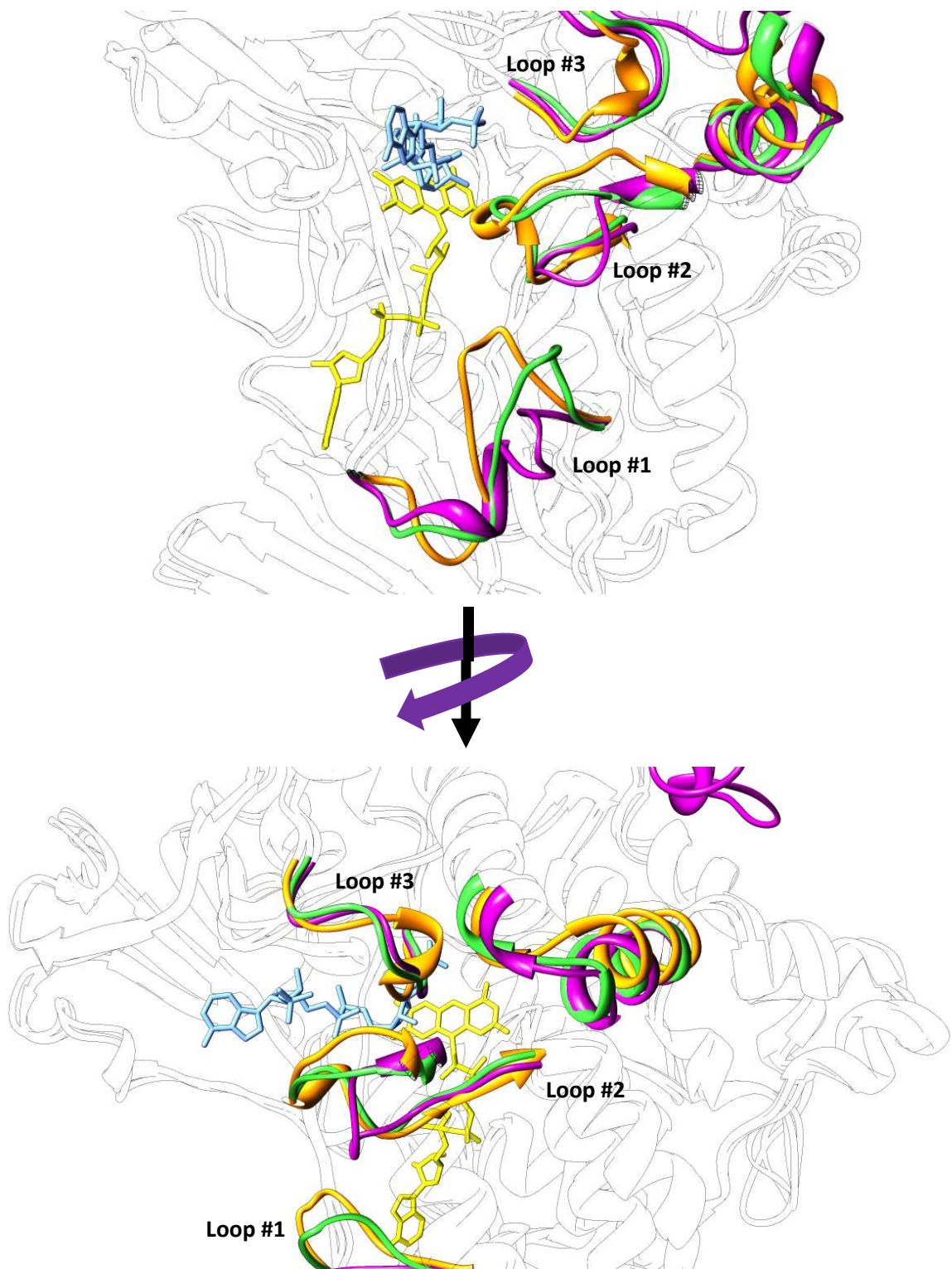
409 The substrates/products penetrate/exit the tunnels through the subdomain found in the 80-residue
410 insertion. Here, the paths are heavily dictated by the conformations of the residues in and around the
411 subdomain (Figure 5A-D). Specifically, a few noticeable changes were observed (Figure 6).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

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



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

443
 444
 445
 446
 447



448
 449
 450
 451
 452
 453
 454
 455
 456
 457

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 residues 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.

458 *AncFMOs are thermostable enzymes that are stabilized significantly in the presence of NADP⁺*
459 *and are reliable models for human FMOs*

460 Allegedly, highly thermostable enzymes are highly prone to crystallization.⁶⁴ Considering that all
461 three AncFMOs crystallized, melting temperature (T_m) assays were conducted using the
462 ThermoFAD technique,⁶⁵ to investigate the thermal stability of AncFMOs compared to human
463 FMOs. Remarkably, our AncFMOs in storage buffer conditions (see Materials & Methods) reached
464 T_m s of 60 °C. Comparing AncFMO3-6 and AncFMO5 with human FMO3 and human FMO5
465 directly, we observed increases of the T_m of up to +22 and +11 °C respectively (Supplementary
466 Table 3, Supplementary Figure 8).³⁷ Generally, the differences between the AncFMOs and their
467 respective human equivalents are found dispersed across the protein (Supplementary Figure 9).
468 These patterns of highly distributed and non-systematic amino acid replacements between ancestral
469 and extant enzymes validate the notion that AncFMOs are very reliable models for the human
470 FMOs. Noticeably, at the periphery of the active site, the small Ala232 in human FMO5 is mutated
471 to a bulky Phe232 in AncFMO5 (Figure 5D, lower panel). This substitution may well allow larger
472 substrates in human FMO5.-Intriguingly, Fiorentini et al. documented that NADP⁺ has no effect on
473 the T_m of human FMO5,³⁷ a result also observed for human FMO3 (Supplementary Figure 8).
474 However, the melting temperatures of all three AncFMOs increased in the presence of NADP⁺ by
475 +17, +7 and +4 °C for AncFMO2, AncFMO3-6 and AncFMO5 respectively (Supplementary Table
476 3). With AncFMOs exhibiting a low degree of uncoupling, it corroborates that tight binding of
477 NADP⁺ is necessary for highly coupled reactions (Table 1; Figure 1).

478

479 Discussion

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

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

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

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

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

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

541

542 **Materials & Methods**

543 **Phylogenetic Inference and Ancestral Sequence Reconstruction.** To obtain a robust and
544 representative phylogeny of FMOs, sequences from the Bacteria and Eukarya were collected by
545 homology searches using BLAST and HMM profiling. 310 sequences were collected and aligned
546 with MAFFT v7.⁷³ Best-fit model parameters were obtained by the Akaike information criterion in
547 ProtTest v3.4. Phylogenies were inferred employing the maximum likelihood method in PhyML
548 v3.0 or RAxML v0.6.0 with 500/1000 bootstraps and transfer bootstrap expectation (TBE)
549 subsequently applied.⁷⁴ As FMOs are not monophyletic, derived clades BVMOs and NMOs were
550 included in the phylogeny.^{7,9} Later, the Gnathostomata FMOs phylogeny was constructed for
551 ancestral sequence reconstruction. To do this, a dataset of 361 sequences was collected including
552 also a cephalochordate sequence to root the tree according to species tree (Supplementary Data 1).⁷⁵
553 Ancestral sequence reconstruction was performed using the maximum likelihood inference method
554 in PAMLX v.4.9.^{43,76} Sequences were analyzed using an empirical amino acid substitution model
555 (model = 3), 4 gamma categories and LG substitution matrix. The posterior probability distribution
556 of ancestral states at each site was analyzed at nodes AncFMO2, AncFMO3-6 and AncFMO5. Sites
557 were considered ambiguously reconstructed if alternative states displayed PP >0.2.⁷⁷ Alternative
558 sites were found to be 2 for AncFMO2, 15 for AncFMO3-6 and 11 for AncFMO5 (Supplementary
559 Table 4). These are mostly conservative amino acid substitutions, and after mapping in the crystal

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

562 **Cloning and Expression of the AncFMOs.** Genes were ordered from Genescript containing BsaI
563 sites at both the 5' and 3' ends of the insert. The insert contained overhangs TGGT and CAAG at
564 the 5' and 3' ends respectively to then be inserted into common pBAD-NK destination vectors with
565 the following modifications: three BsaI sites were eliminated and two were introduced to facilitate
566 the cloning that incorporated SUMO and 8xHis-tag regions to the N-terminus. Inserts were fused
567 into the destination vectors through Golder Gate cloning. The sample was prepared with the
568 following: 75 ng of golden gate entry vector, a molar ratio of 2:1 between insert and vector, BsaI(-
569 HF) (15 U), 30 WU T4 DNA ligase (15 U), T4 DNA ligase buffer (1x) and Nuclease free water
570 added to a final volume of 20 μ L. During the cloning procedure, a negative control was prepared
571 with the fragments/inserts omitted. The Golden Gate assembly was conducted in the following
572 manner, where maximum efficiency was desired: A cycle of 37 °C for 5 mins followed by 16 °C for
573 10 minutes was repeated 30 times; followed by 55 °C for 10 minutes, 65 °C for a further 20
574 minutes, finishing with 8 °C for 20 minutes. Once cloned, the plasmids were transformed by heat
575 shock into *E. coli*. BL21 cells (25 seconds, 42 °C). Cells from the resulting colonies were pre-
576 inoculated into 100 mL of LB broth containing 100 μ g/mL of ampicillin and grown overnight at 37
577 °C. The cultures were then transferred to 1L Terrific Broth cultures (15 mL) and grown at 24 °C,
578 rpm 180 for 5 – 6 hours until the OD reached 0.3. The cultures were then induced with a sterilized
579 arabinose solution (20% w/v), final concentration of 0.02% (w/v) and incubated at 24 °C, 180 rpm
580 for an additional 24 hours. Cells were then harvested by centrifugation (5000g, 15 mins, 10 °C)
581 flash frozen in liquid nitrogen and stored at -80 °C. For site-directed mutagenesis, a PCR-reaction
582 mixture was prepared with 10 μ M primer forward and reverse, 100 ng of template DNA, 1.6 %
583 DMSO, 0.8 mM MgCl₂ and 1x *Pfu* Ultra II Hotstart Master Mix (Agilent). The Quickchange PCR
584 cycle was performed using the following method: firstly a 5 minute incubation at 95°C, then cycles
585 (95 °C for 5 minutes - 60 °C for 30 seconds - 72 °C for 6 minutes) were repeated 25 times; followed
586 by 72 °C for 10 minutes and finishing with 8 °C on hold. The PCR mixture was digested with *DpnI*
587 overnight and transformed into *E. coli*.

588 **Cell Disruption, Extraction, and Purification of AncFMOs.** All procedures were carried out in
589 ice or at 4 °C. Cells (*ca.* 20 g) were resuspended (1:5) in buffer A (250 mM NaCl, 50 mM KH₂PO₄,
590 pH 7.8) and included additional protease inhibitors: phenylmethylsulfonyl fluoride (1 mM),
591 leupeptin (10 μ M), pepstatin (10 μ M) and DNase I (5 μ g/g of cell paste). The solution was stirred
592 and incubated at 4 °C for 45 minutes before cell lysis was conducted using sonication or a high-
593 pressure homogenizer. Sonication was conducted using the following conditions: 50 mL solution, 5

594 seconds on, 10 seconds off, 1-minute cycles with a total sonication time of 3 minutes using a
595 microtip (70% amplitude). Cells were passed through a high-pressure homogenizer twice. Lysed
596 cells were then spun down (1200g, 12 mins, 4 °C) to remove the cell debris. The resultant
597 supernatant was then centrifuged further (56,000g, 1 hr. 40 mins, 4 °C) to collect the membrane
598 pellet which was then re-homogenized in buffer A (15 mL,) and centrifuged again (56,000g, 1 hr.
599 20 mins, 4 °C) to further purify the insoluble material. The resulting pellet was re-homogenized in
600 buffer A (7 mL) and diluted to a final concentration of 13 mg/mL (assayed using Biuret reagent).
601 Triton X-100 Reduced (TRX-100-R) (Sigma-Aldrich) was then added to the solution (0.5% (v/v)
602 final concentration) and mixed overnight at 4 °C. The detergent-solubilized fraction containing the
603 AncFMOs was then abstracted by collecting the supernatant after centrifugation (56,000g, 1 hr. 20
604 mins, 4 °C). The supernatant was then transferred to a pre-equilibrated (with buffer A and 0.05%
605 (v/v) TRX-100-R) gravity column containing a Ni-resin (GE Healthcare). The supernatant was
606 washed with buffer A, containing 0.05% TRX-100-R, and then with increasing concentrations of
607 buffer B (50 mM KH₂PO₄, 500 mM NaCl, 300 mM imidazole, pH 7.8), also containing 0.05% (v/v)
608 TRX-100-R, in step-by-step fashion: 5 mM imidazole wash, 30 mM imidazole wash and finally a
609 300 mM imidazole wash, where the protein then eluted. The buffers were then exchanged using a
610 centrifugal filter unit (50 kDa cut-off) and multiple washes with buffer A with 0.05% (v/v) TRX-
611 100-R. This step was important to remove high concentrations of imidazole employed during the
612 elution. The protein sample was then concentrated down to a final volume between 500 and 1000
613 µL. The sample was then mixed with a 6xHis-tagged SUMO protease (1.2 mg/mL) to a volume
614 ratio of 10:1 respectively and incubated overnight at 4 °C. The sample was then loaded onto an
615 Akta purification system (GE Healthcare) endowed with a multiwavelength detector (set at
616 280/370/450 nm) and then onto a nickel-affinity His-trap column (GE Healthcare). The column was
617 pre-equilibrated with buffer A containing 0.05% (v/v) TRX-100-R, as stated before, with the
618 proteins eluting out in the presence of 6 mM imidazole derived from buffer B (2%) containing
619 0.05% (v/v) TRX-100-R. The SUMO-His-tag cleaved protein was then concentrated, and buffer
620 exchanged using a concentrating centrifugal filter unit (50 kDa cut-off) to a final volume between
621 250 and 500 µL. The sample was incubated for 1 h with 100 µM FAD at 4 °C and then loaded onto
622 a gel filtration column (Superdex 200 10/300, GE Healthcare) pre-equilibrated with a storage buffer
623 (50 mM Tris-HCl pH 8.5 at 4 °C, 10 mM NaCl) and a detergent of choice to obtain a higher degree
624 of purity (obtained from anatrace). Typically DDM was used (0.03% (w/v) DDM (analytical
625 grade)), but other detergents were used for crystallization screenings at 3x their respective Critical-
626 Micelle Concentration (CMC). The protein eluted with a very high purity and homogeneity
627 (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, $\epsilon_{340} = 6.22 \text{ mM}^{-1} \text{ cm}^{-1}$ 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 ($\epsilon_{340} = 6.22 \text{ mM}^{-1} \text{ cm}^{-1}$ 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

662 the anaerobic enzyme solution with anaerobic buffer containing increasing concentrations of
663 NAD(P)H (Oriental Yeast Co. LTD.). The rate of flavin reduction was determined by measuring the
664 decrease of absorbance at 447 nm or 442 nm for AncFMO2 and AncFMO3-6, respectively. In order
665 to reduce the flavin cofactor in the FMOs for the oxidative half reaction, NADPH was added to an
666 anaerobic solution containing an equivalent amount of AncFMO. The resulting solution was
667 incubated on ice until the bleaching of the FAD was complete, indicating complete reduction to
668 FADH₂. The anaerobically reduced FMOs were mixed with air-saturated buffer first and then with
669 air-saturated buffer containing 1.0 mM or 0.4 mM of trimethylamine, respectively. This allowed us
670 to follow the spectral changes during the oxidative half-reactions. The C4alpha-(hydro)peroxyflavin
671 intermediate formation and decay was specifically monitored using the PMT module at 360 nm. For
672 determining the rates of reoxidation, the reduced enzymes were mixed with buffers containing
673 different concentrations of dioxygen. The final concentrations of dioxygen (0.13, 0.31, 0.61, 0.96
674 mM after mixing) were achieved by mixing the anaerobic enzyme solution with (i) air-saturated
675 buffer; (ii) mixing equal volumes of 100% argon buffer and 100% O₂ buffer; (iii) 100% O₂ buffer;
676 (iv) 100% O₂ buffer on ice. All solutions were bubbled for ten minutes at room temperature except
677 the last one which was done on ice. In the case of AncFMO2, in order to confirm its saturating
678 behaviour, an additional measurement was performed at 0.816 mM of O₂ by mixing 100% O₂
679 buffer on ice with 100% argon buffer.

680 Observed rates (k_{obs}) were determined by fitting traces to exponential functions. All data were
681 analysed using the software Pro-Data (Applied Photophysics, Surrey, UK) and GraphPad Prism
682 6.05 (La Jolla, CA, USA).

683 **ThermoFAD Assays.**⁶⁵ A Bio-Rad MiniOpticon Real-Time PCR System was employed to perform
684 ThermoFAD screenings (temperature gradient 25–70 °C, fluorescence detection every 0.5 °C at 485
685 ± 30 nm excitation and 625 ± 30 nm emission for 5 s). Concentrations were determined using a
686 molar extinction coefficient of 12 mM⁻¹ cm⁻¹ for the FAD band at 442 nm. Experiments were
687 performed in triplicate using **human FMO5** and the AncFMOs, in the presence or absence of
688 NADP⁺. Each sample contained the protein of interest (4 μM), and with or without NADP⁺ (200
689 μM), made to a final volume of 20 μL using the storage buffer and incubated in ice for one hour.
690 Melting temperatures for **human FMO3** (0.05% (v/v) TRX-100-R) were performed in duplicates
691 and diluted to a final concentration of 5 μM in buffer (100 mM) with varying pH values (pH 6-6.5
692 MES, pH 7-7.5-8 HEPES, pH 8-8.5-9 Bicine, pH 9.5 CHES), KCl concentrations (0-500 mM in
693 HEPES pH 8), and NADP⁺ concentrations (5-500 μM in HEPES pH 8, 10 mM KCl, 0.05% (v/v)
694 TRX-100-R) in an attempt to generate optimal storage buffer conditions.

695 **Conversions.** Conversions were performed using AncFMO2 and AncFMO3-6 and their respective
696 E281H mutants ([Supplementary Table 2](#)). Reaction mixtures (1.0 mL) contained 5.0 mM substrate
697 (<1% EtOH), 0.10 mM NADPH, 2.0 μ M enzyme, 5.0 μ M phosphite dehydrogenase, 20 mM
698 phosphite, 50 mM KPi pH 7.5, 250 mM NaCl and 0.05% (v/v) TRX-100-R. The mixtures were
699 incubated for 18 hours at 30 °C and subsequently extracted with 1.0 mL ethyl acetate. The organic
700 phase was passed through anhydrous sulfate magnesium to remove residual water. Analysis was
701 carried out using a GCMS-QP2010 Ultra (Shimadzu) equipped with a HP-1 column, using electron
702 ionization MS detection.

703 **Crystallization and Structural Determination of the AncFMOs.** Each AncFMO crystallized in a
704 range of conditions with multiple detergents. Typically, PEG 4000 was optimal for crystallization.
705 The highest diffracting crystallization conditions for each AncFMO are described below. AncFMO2
706 (with and without NADP⁺): 12-15 mg/mL of AncFMO2 (in storage buffer and CYMAL-6 (0.09%
707 (w/v))) was incubated with crystallization conditions of HEPES buffer (0.1 M, pH 7.5) and PEG
708 4000 (10%) at 20 °C with a ratio of 1:1 on a sitting drop. Sitting drop was 2 μ L after mixing and
709 contained a reservoir of 1 mL. Prior to crystallization, NADP⁺ (1 mM final) was incubated with 12-
710 15 mg/mL of AncFMO2 for 1 hour at 4 °C. After two days, large yellow crystals formed.
711 AncFMO3-6: 12-15 mg/mL of AncFMO3-6 (in storage buffer and DDM (0.03% (w/v))) was
712 incubated with crystallization conditions of Sodium Acetate buffer (0.1 M, pH 5.5) and PEG 4000
713 (7.5%) at 20 °C with a ratio of 1:1 on a sitting drop. Sitting drop was 2 μ L after mixing and
714 contained a reservoir of 1 mL. Prior to crystallization, NADP⁺ (1 mM final) was incubated with 12-
715 15 mg/mL of AncFMO3-6 for 1 hour at 4 °C. After one day, large yellow crystals formed.
716 AncFMO5: 12 mg/mL of AncFMO5 (in storage buffer and DDM (0.03% (w/v))) was incubated with
717 crystallization conditions of HEPES buffer (0.1 M, pH 6.9) and PEG 4000 (9%) at 20 °C with a
718 ratio of 1:1 on a sitting drop. Sitting drop was 2 μ L after mixing and contained a reservoir of 1 mL.
719 Prior to crystallization, NADP⁺ (1 mM final) was incubated with 12-15 mg/mL of AncFMO5 for 1
720 hour at 4 °C. After one day, large yellow hexagon shaped crystals formed. During crystal fishing, a
721 cryo-protectant was prepared containing modified crystallizations conditions with 20% glycerol and
722 PEG 4000 (15%).

723 Data were collected at the European Synchrotron Radiation Facility (Grenoble, France) and the
724 Swiss Light Source, (Villigen, Switzerland) and processed with the XDS⁷⁸ and CCP4 packages.⁷⁹
725 Aimless was used to merge the observations into average densities ([Supplementary Table 1](#)).
726 STARANISO was additionally used for AncFMO2 which suffered greatly from anisotropy.^{80,81} The
727 phase problem was solved by Molecular Replacement using a recently solved insect FMO
728 (PDB:5nmw)⁸² as a search model, and then AncFMO2 for the proceeding AncFMOs, using Phaser

729 and Molrep.⁸³ The phases were greatly improved by density averaging with DM.^{79,84} Model
730 Building and Refinement were then conducted using COOT,⁸⁵ Buccaneer,⁷⁹ and Refmac5⁸⁶
731 respectively (**Supplementary Table 1**). Figures were then generated using UCSF Chimera,⁸⁷
732 PyMOL (DeLano Scientific; www.pymol.org) and CCP4mg.⁷⁹

733

734 **Acknowledgements**

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

740

741 **Author Contributions**

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

752

753 **Declaration of Interest**

754 The authors declare no conflict of interest.

755

756 **Data availability**

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

760

761 **References**

762 1. Jakoby, W. B. & Ziegler, D. M. The enzymes of detoxication. *J. Biol. Chem.* **265**, 20715–
763 20718 (1990).

- 764 2. Krueger, S. K. & Williams, D. E. Mammalian flavin-containing monooxygenases:
765 Structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacol. Ther.*
766 **106**, 357–387 (2005).
- 767 3. Cruciani, G. *et al.* Flavin monooxygenase metabolism: Why medicinal chemists should
768 matter. *J. Med. Chem.* **57**, 6183–6196 (2014).
- 769 4. Huijbers, M. M. E., Montersino, S., Westphal, A. H., Tischler, D. & Van Berkel, W. J. H.
770 Flavin dependent monooxygenases. *Arch. Biochem. Biophys.* **544**, 2–17 (2014).
- 771 5. Ziegler, D. M. Flavin-containing monooxygenases: enzymes adapted for multisubstrate
772 specificity. *Trends Pharmacol. Sci.* **11**, 321–324 (1990).
- 773 6. Cashman, J. R. Some distinctions between flavin-containing and cytochrome P450
774 monooxygenases. *Biochem. Biophys. Res. Commun.* **338**, 599–604 (2005).
- 775 7. Mascotti, M. L., Lapadula, W. J. & Ayub, M. J. The origin and evolution of Baeyer - Villiger
776 Monooxygenases (BVMOs): An ancestral family of flavin monooxygenases. *PLoS One* **10**,
777 1–16 (2015).
- 778 8. D. M. Ziegler & Pettit, F. H. Formation of an intermediate N-oxide in the oxidative
779 demethylation of N,N-dimethylaniline catalyzed by liver microsomes. *Biochem. Biophys.*
780 *Res. Commun.* **15**, 188–193 (1964).
- 781 9. Mascotti, M. L., Juri Ayub, M., Furnham, N., Thornton, J. M. & Laskowski, R. A. Chopping
782 and Changing: the Evolution of the Flavin-dependent Monooxygenases. *J. Mol. Biol.* **428**,
783 3131–3146 (2016).
- 784 10. Cashman, J. R. & Zhang, J. Human Flavin-Containing Monooxygenases. *Annu. Rev.*
785 *Pharmacol. Toxicol.* **46**, 65–100 (2006).
- 786 11. Romero, E., Castellanos, J. R. G., Gadda, G., Fraaije, M. W. & Mattevi, A. The same
787 substrate, many reactions: oxygen activation in flavoenzymes. *Chem. Rev.* **118**, 1742-1769
788 (2017).
- 789 12. Alfieri, A., Malito, E., Orru, R., Fraaije, M. W. & Mattevi, A. Revealing the moonlighting
790 role of NADP in the structure of a flavin-containing monooxygenase. *Proc. Natl. Acad. Sci.*
791 **105**, 6572–6577 (2008).
- 792 13. Zhang, J. QUANTITATIVE ANALYSIS OF FMO GENE mRNA LEVELS IN HUMAN
793 TISSUES. *Drug Metab. Dispos.* **34**, 19–26 (2005).
- 794 14. Dolphin, C. T., Cullingford, T. E., Shephard, E. A., Smith, R. L. & Phillips, I. R. Differential
795 developmental and tissue-specific regulation of expression of the genes encoding three
796 members of the flavin-containing monooxygenase family of man, FMO1, FMO3 and FMO4.
797 *Eur. J. Biochem.* **235**, 683–689 (1996).
- 798 15. McCombie, R. R., Dolphin, C. T., Povey, S., Phillips, I. R. & Shephard, E. A. Localization of
799 human flavin-containing monooxygenase genes FMO2 and FMO5 to chromosome 1q.
800 *Genomics* **34**, 426–429 (1996).
- 801 16. Hernandez, D., Janmohamed, A., Chandan, P., Phillips, I. R. & Shephard, E. A. Organization
802 and evolution of the flavin-containing monooxygenase genes of human and mouse.
803 *Pharmacogenetics* **14**, 117–130 (2004).
- 804 17. Fiorentini, F. *et al.* Baeyer-Villiger Monooxygenase FMO5 as Entry Point in Drug
805 Metabolism. *ACS Chem. Biol.* **12**, 2379–2387 (2017).
- 806 18. Poulsen, L. L. & Ziegler, D. M. Multisubstrate flavin-containing monooxygenases:
807 applications of mechanism to specificity. *Chem. Biol. Interact.* **96**, 57–73 (1995).
- 808 19. Henderson, M. C., Krueger, S. K., Siddens, L. K., Stevens, J. F. & Williams, D. E. S-
809 oxygenation of the thioether organophosphate insecticides phorate and disulfoton by human
810 lung flavin-containing monooxygenase 2. *Biochem. Pharmacol.* **68**, 959–967 (2004).
- 811 20. Cashman, J. R. Role of flavin-containing monooxygenase in drug development. *Expert Opin.*
812 *Drug Metab. Toxicol.* **4**, 1507–1521 (2008).
- 813 21. Siddens, L. K., Krueger, S. K., Henderson, M. C. & Williams, D. E. Mammalian flavin-
814 containing monooxygenase (FMO) as a source of hydrogen peroxide. *Biochem. Pharmacol.*

- 815 **89**, 141–147 (2014).
- 816 22. Li, C. Y. *et al.* Structural mechanism for bacterial oxidation of oceanic trimethylamine into
817 trimethylamine N-oxide. *Mol. Microbiol.* **103**, 992–1003 (2017).
- 818 23. Miao, J. *et al.* Flavin-containing monooxygenase 3 as a potential player in diabetes-
819 associated atherosclerosis. *Nat. Commun.* **6**, 1–10 (2015).
- 820 24. Krueger, S. K. *et al.* GENETIC POLYMORPHISMS OF FLAVIN- CONTAINING
821 MONOOXYGENASE (FMO). **34**, 523–532 (2002).
- 822 25. Dolphin, C. T., Janmohamed, A., Smith, R. L., Shephard, E. A. & Phillips, I. R. Compound
823 heterozygosity for missense mutations in the flavin-containing monooxygenase 3 (FMO3)
824 gene in patients with fish-odour syndrome. *Pharmacogenetics* **10**, 799–807 (2000).
- 825 26. Veeravalli, S. *et al.* Effect of flavin-containing monooxygenase genotype, mouse strain, and
826 gender on trimethylamine N-oxide production, plasma cholesterol concentration, and an
827 index of atherosclerosis. *Drug Metab. Dispos.* **46**, 20–25 (2018).
- 828 27. Phillips, I. R. & Shephard, E. A. Flavin-containing monooxygenases: mutations, disease and
829 drug response. *Trends Pharmacol. Sci.* **29**, 294–301 (2008).
- 830 28. Hernandez, D. *et al.* Trimethylaminuria and a human FMO3 mutation database. *Hum. Mutat.*
831 **22**, 209–213 (2003).
- 832 29. Dolphin, C. T., Janmohamed, A., Smith, R. L., Shephard, E. A. & Phillips, Ian R. Missense
833 mutation in flavin-containing mono-oxygenase 3 gene, FMO3, underlies fish-odour
834 syndrome. *Nat. Genet.* **17**, 491–494 (1997).
- 835 30. Lang, D. H. *et al.* Isoform specificity of trimethylamine N-oxygenation by human flavin-
836 containing monooxygenase (FMO) and P450 enzymes Selective catalysis by fmo3. *Biochem.*
837 *Pharmacol.* **56**, 1005–1012 (1998).
- 838 31. Hodgson, E., Rose, R. L., Cao, Y., Dehal, S. S. & Kupfer, D. Flavin-containing
839 monooxygenase isoform specificity for the N-oxidation of tamoxifen determined by product
840 measurement and NADPH oxidation. *J. Biochem. Mol. Toxicol.* **14**, 118–120 (2000).
- 841 32. Mushiroda, T., Douya, R., Takahara, E. & Nagata, O. The involvement of flavin-containing
842 monooxygenase but not CYP3A4 in metabolism of itopride hydrochloride, a gastroprokinetic
843 agent: Comparison with cisapride and mosapride citrate. *Drug Metab. Dispos.* **28**, 1231–
844 1237 (2000).
- 845 33. Shephard, E. A. & Phillips, I. R. The potential of knockout mouse lines in defining the role
846 of flavin-containing monooxygenases in drug metabolism. *Expert Opin. Drug Metab.*
847 *Toxicol.* **6**, 1083–1094 (2010).
- 848 34. Krueger, S. K., VanDyke, J. E., Williams, D. E. & Hines, R. N. The role of flavin-containing
849 monooxygenase (FMO) in the metabolism of tamoxifen and other tertiary amines. *Drug*
850 *Metab. Rev.* **38**, 139–147 (2006).
- 851 35. Veeramah, K. R. *et al.* The potentially deleterious functional variant flavin-containing
852 monooxygenase 2*1 is at high frequency throughout sub-Saharan Africa. *Pharmacogenet.*
853 *Genomics* **18**, 877–886 (2008).
- 854 36. Dolphin, C. T. *et al.* The flavin-containing monooxygenase 2 gene (FMO2) of humans, but
855 not of other primates, encodes a truncated, nonfunctional protein. *J. Biol. Chem.* **273**, 30599–
856 30607 (1998).
- 857 37. Fiorentini, F. *et al.* Biocatalytic Characterization of Human FMO5: Unearthing Baeyer-
858 Villiger Reactions in Humans. *ACS Chem. Biol.* **11**, 1039–1048 (2016).
- 859 38. Leiser, S. F. *et al.* Promotes Longevity and Health Span. *Science.* **350**, 1375–1378 (2015).
- 860 39. Warriar, M. *et al.* The TMAO-Generating Enzyme Flavin Monooxygenase 3 Is a Central
861 Regulator of Cholesterol Balance. *Cell Rep.* **10**, 326–338 (2015).
- 862 40. Gonzalez Malagon, S. G. *et al.* The phenotype of a knockout mouse identifies flavin-
863 containing monooxygenase 5 (FMO5) as a regulator of metabolic ageing. *Biochem.*
864 *Pharmacol.* **96**, 267–277 (2015).
- 865 41. Scott, F. *et al.* Identification of flavin-containing monooxygenase 5 (FMO5) as a regulator of

- 866 glucose homeostasis and a potential sensor of gut bacteria. *Drug Metab. Dispos.* **45**, 982–989
867 (2017).
- 868 42. Risso, V. A., Sanchez-Ruiz, J. M. & Ozkan, S. B. Biotechnological and protein-engineering
869 implications of ancestral protein resurrection. *Curr. Opin. Struct. Biol.* **51**, 106–115 (2018).
- 870 43. Hochberg, G. K. A. & Thornton, J. W. Reconstructing Ancient Proteins to Understand the
871 Causes of Structure and Function. *Annu. Rev. Biophys.* **46**, 247–269 (2017).
- 872 44. Hao, D. C., Chen, S. L., Mu, J. & Xiao, P. G. Molecular phylogeny, long-term evolution, and
873 functional divergence of flavin-containing monooxygenases. *Genetica* **137**, 173–187 (2009).
- 874 45. Hines, R. N. Alternative Processing of the Human FMO6 Gene Renders Transcripts
875 Incapable of Encoding a Functional Flavin-Containing Monooxygenase. *Mol. Pharmacol.*
876 **62**, 320–325 (2002).
- 877 46. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for Timelines,
878 Timetrees, and Divergence Times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
- 879 47. Grantham, R. Amino acid difference formula to help explain protein evolution. *Science.* **185**,
880 862–864 (1974).
- 881 48. Krueger, S. K. *et al.* Characterization of sulfoxxygenation and structural implications of
882 human flavin-containing monooxygenase isoform 2 (FMO2.1) variants S195L and N413K.
883 *Drug Metab. Dispos.* **37**, 1785–1791 (2009).
- 884 49. Overby, L. H., Carver, G. C. & Philpot, R. M. Quantitation and kinetic properties of hepatic
885 microsomal and recombinant flavin-containing monooxygenases 3 and 5 from humans.
886 *Chem. Biol. Interact.* **106**, 29–45 (1997).
- 887 50. Ripp, S. L., Itagaki, K., Philpot, R. M. & Elfarra, A. A. Methionine S-oxidation in human
888 and rabbit liver microsomes: Evidence for a high-affinity methionine S-oxidase activity that
889 is distinct from flavin-containing monooxygenase 3. *Arch. Biochem. Biophys.* **367**, 322–332
890 (1999).
- 891 51. Lin, J. & Cashman, J. R. N-oxygenation of phenethylamine to the trans-oxime by adult
892 human liver flavin-containing monooxygenase and retroreduction of phenethylamine
893 hydroxylamine by human liver microsomes. *J. Pharmacol. Exp. Ther.* **282**, 1269–1279
894 (1997).
- 895 52. Beaty, N. B. & Ballou, D. P. The reductive half-reaction of liver microsomal FAD-
896 containing monooxygenase. *J. Biol. Chem.* **256**, 4611–4618 (1981).
- 897 53. Krissinel, E. Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Res.*
898 **43**, W314–W319 (2015).
- 899 54. Geier, M. *et al.* Human FMO2-based microbial whole-cell catalysts for drug metabolite
900 synthesis. *Microb. Cell Fact.* **14**, 1–10 (2015).
- 901 55. Korsmeyer, K. K. *et al.* N-glycosylation of pig flavin-containing monooxygenase form 1:
902 Determination of the site of protein modification by mass spectrometry. *Chem. Res. Toxicol.*
903 **11**, 1145–1153 (1998).
- 904 56. Wierenga, R. K., De Maeyer, M. C. H. & Hoi, W. G. J. Interaction of Pyrophosphate
905 Moieties with α -Helixes in Dinucleotide Binding Proteins. *Biochemistry* **24**, 1346–1357
906 (1985).
- 907 57. Orru, R., Torres Pazmiño, D. E., Fraaije, M. W. & Mattevi, A. Joint functions of protein
908 residues and NADP(H) in oxygen activation by flavin-containing monooxygenase. *J. Biol.*
909 *Chem.* **285**, 35021–35028 (2010).
- 910 58. Lončar, N. *et al.* Characterization of a thermostable flavin-containing monooxygenase from
911 *Nitricola lacisaponensis* (NiFMO). *Appl. Microbiol. Biotechnol.* **103**, 1755–1764 (2019).
- 912 59. Fürst, M. J., Fiorentini, F. & Fraaije, M. W. Beyond active site residues: overall structural
913 dynamics control catalysis in flavin-containing and heme-containing monooxygenases. *Curr.*
914 *Opin. Struct. Biol.* **59**, 29–37 (2019).
- 915 60. Romero, E., Castellanos, J. R. G., Mattevi, A. & Fraaije, M. W. Characterization and Crystal
916 Structure of a Robust Cyclohexanone Monooxygenase. *Angew. Chemie - Int. Ed.* **55**, 15852–

- 917 15855 (2016).
- 918 61. Torres Pazmiño, D. E., Baas, B. J., Janssen, D. B. & Fraaije, M. W. Kinetic mechanism of
919 phenylacetone monooxygenase from *Thermobifida fusca*. *Biochemistry* **47**, 4082–4093
920 (2008).
- 921 62. Criegee, R. Die Umlagerung der Dekalin-peroxydester als Folge von kationischem
922 Sauerstoff. *Justus Liebigs Ann. Chem.* **560**, 127–135 (1948).
- 923 63. Ho, B. K. & Gruswitz, F. HOLLOW: Generating accurate representations of channel and
924 interior surfaces in molecular structures. *BMC Struct. Biol.* **8**, 1–6 (2008).
- 925 64. Gumulya, Y. *et al.* Engineering highly functional thermostable proteins using ancestral
926 sequence reconstruction. *Nat. Catal.* **1**, 878–888 (2018).
- 927 65. Forneris, F., Orru, R., Bonivento, D., Chiarelli, L. R. & Mattevi, A. ThermoFAD, a
928 Thermofluor®-adapted flavin ad hoc detection system for protein folding and ligand binding.
929 *FEBS J.* **276**, 2833–2840 (2009).
- 930 66. Wheeler, L. C., Lim, S. A., Marqusee, S. & Harms, M. J. The thermostability and specificity
931 of ancient proteins. *Curr. Opin. Struct. Biol.* **38**, 37–43 (2016).
- 932 67. Lawton, M. P. & Philpot, R. M. Functional characterization of flavin-containing
933 monooxygenase 1B1 expressed in *Saccharomyces cerevisiae* and *Escherichia coli* and
934 analysis of proposed FAD- and membrane-binding domains. *J. Biol. Chem.* **268**, 5728–5734
935 (1993).
- 936 68. Allen, K. N., Entova, S., Ray, L. C. & Imperiali, B. Monotopic Membrane Proteins Join the
937 Fold. *Trends Biochem. Sci.* **44**, 7–20 (2019).
- 938 69. Nagata, T., Williams, D. E. & Ziegler, D. M. Substrate Specificities of Rabbit Lung and
939 Porcine Liver Flavin-Containing Monooxygenases: Differences due to Substrate Size. *Chem.*
940 *Res. Toxicol.* **3**, 372–376 (1990).
- 941 70. Young Mi Kim & Ziegler, D. M. Size limits of thiocarbamides accepted as substrates by
942 human flavin-containing monooxygenase 1. *Drug Metab. Dispos.* **28**, 1003–1006 (2000).
- 943 71. Beaty, N. B. & Ballou, D. P. Transient Kinetic Study of Liver. *J. Biol. Chem.* **255**, 3817–
944 3819 (1980).
- 945 72. Beaty, N. B. & Ballou, D. P. The Oxidative Half-reaction of Liver Microsomal FAD-
946 containing Monooxygenase. *J. Biol. Chem.* **256**, 4619–4625 (1981).
- 947 73. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence
948 alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 1–7 (2017).
- 949 74. Lemoine, F. *et al.* Renewing Felsenstein’s phylogenetic bootstrap in the era of big data.
950 *Nature* **556**, 452–456 (2018).
- 951 75. Hug, L. A. *et al.* A new view of the tree of life. *Nat. Microbiol.* **1**, 1–6 (2016).
- 952 76. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**,
953 1586–1591 (2007).
- 954 77. Siddiq, M. A., Loehlin, D. W., Montooth, K. L. & Thornton, J. W. Experimental test and
955 refutation of a classic case of molecular adaptation in *Drosophila melanogaster*. *Nat. Ecol.*
956 *Evol.* **1**, 1–6 (2017).
- 957 78. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially
958 unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
- 959 79. Project, C. C. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. Sect.*
960 *D Biol. Crystallogr.* **50**, 760–763 (1994).
- 961 80. Tickle, I. J. *et al.* Staraniso. *Glob. Phasing Ltd* (2018).
- 962 81. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta*
963 *Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
- 964 82. Kubitzka, C. *et al.* Crystal structure of pyrrolizidine alkaloid N-oxygenase from the
965 grasshopper *Zonocerus variegatus*. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 422–432
966 (2018).
- 967 83. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674

- 968 (2007).
- 969 84. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein
970 structure alignment in three dimensions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**,
971 2256–2268 (2004).
- 972 85. Jung, W. S., Singh, R. K., Lee, J. K. & Pan, C. H. Crystal Structure and Substrate Specificity
973 of D-Galactose-6-Phosphate Isomerase Complexed with Substrates. *PLoS One* **8**, 2–11
974 (2013).
- 975 86. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures
976 by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53**, 240–
977 255 (1997).
- 978 87. Pettersen, E. F. *et al.* UCSF Chimera - a visualization system for exploratory research and
979 analysis. *J. Comput. Chem.* **25**, 1605–12 (2004).
- 980
- 981
- 982

983 **Supplementary Information**

984

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

987

988 Callum R. Nicoll¹, Gautier Bailleul², Filippo Fiorentini¹,
989 María Laura Mascotti^{3,*}, Marco W. Fraaije^{2,*} and Andrea Mattevi^{1,*}

990

991 ¹ Department of Biology and Biotechnology “Lazzaro Spallanzani”, University of Pavia, via Ferrata
992 9, 27100 Pavia, Italy

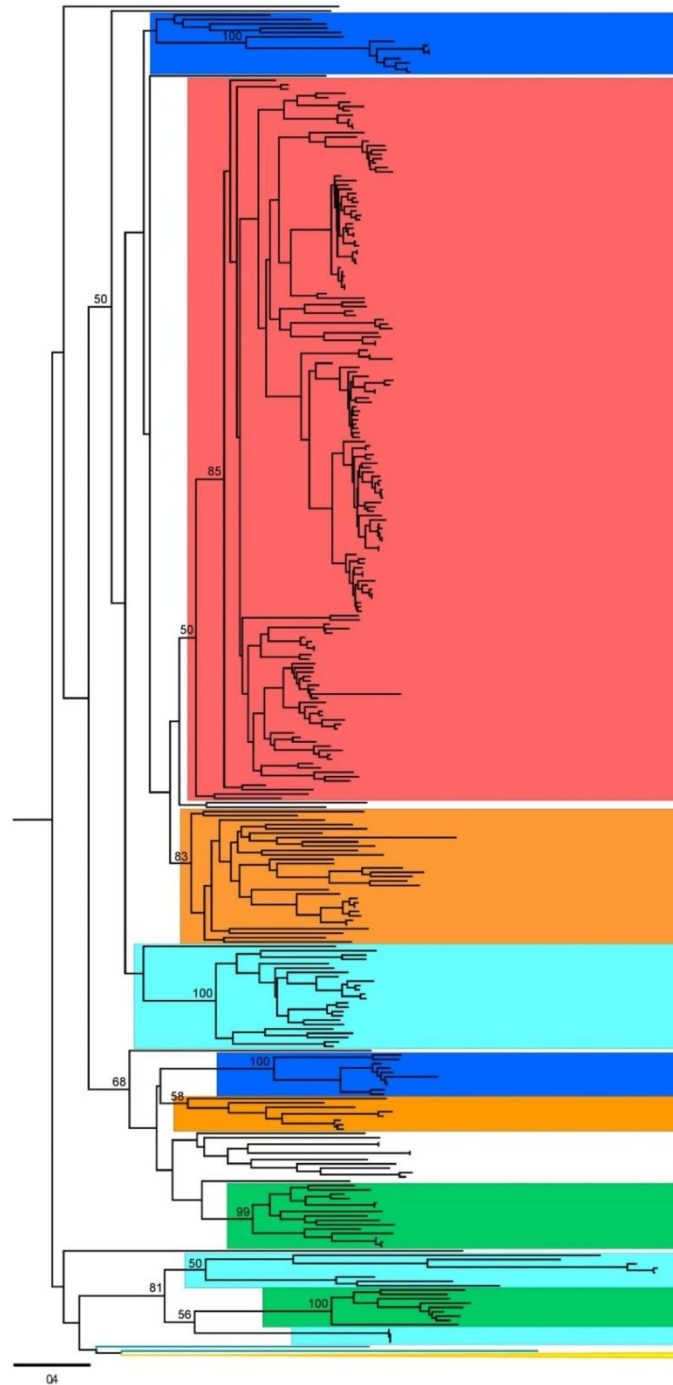
993 ² Laboratory of Biochemistry, Groningen Biomolecular Sciences and Biotechnology Institute,
994 University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

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

997

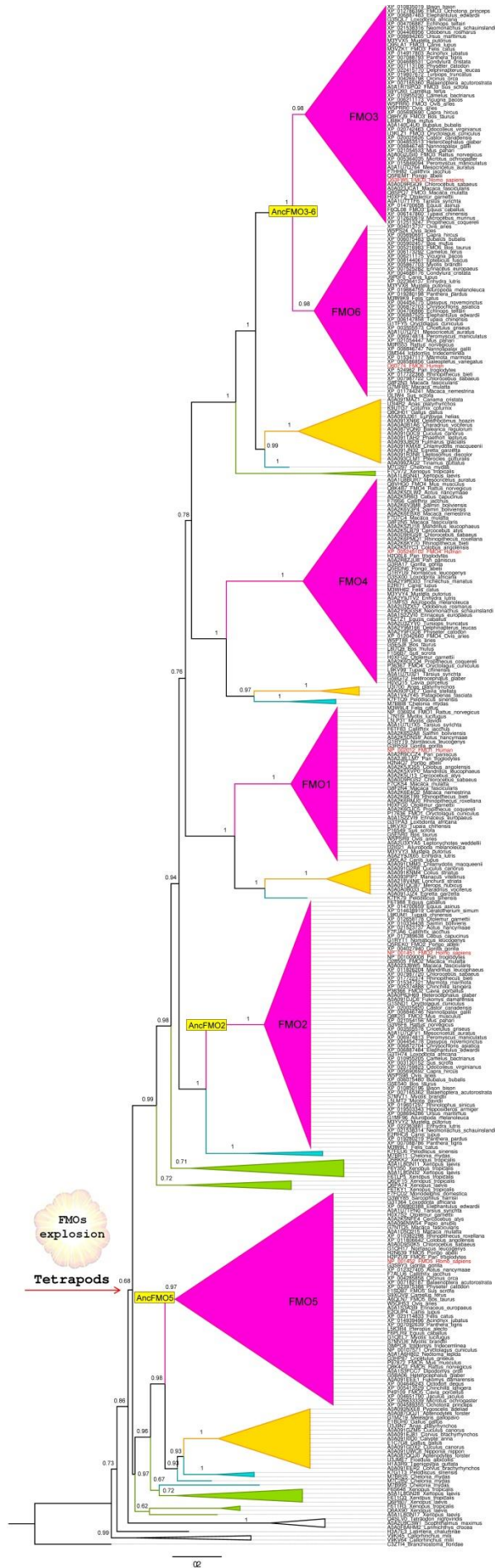
998 *Correspondence to María Laura Mascotti, Marco Fraaije, Andrea Mattevi
999 E-mail: mlmascotti@unsl.edu.ar, m.w.fraaije@rug.nl, andrea.mattevi@unipv.it

1000



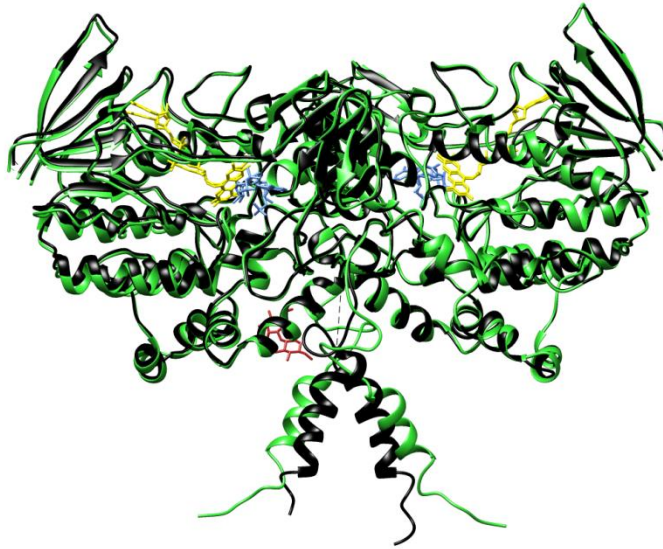
1002

1003 **Supplementary Figure 1:** Full phylogeny of FMOs. Tree was constructed in PhyML v3.0, 500
 1004 bootstraps were run, and best-fit model parameters were obtained with ProtTest v3.4. The employed
 1005 multiple sequence alignment was trimmed in single sequence extensions and contained 341 taxa
 1006 and 501 sites. Tree topology was analyzed according to Class B flavoprotein monooxygenases
 1007 phylogeny. As FMOs are paraphyletic (*i.e.* they do not display a single origin), derived clades
 1008 are shown collapsed: Baeyer-Villiger monooxygenases (yellow) and N-hydroxylating monooxygenases
 1009 (deep teal). FMOs are shown in black branches and monophyletic clades according to organism
 1010 classes are highlighted as follows: Bacteria (cyan), plants (green), vertebrates (red), arthropods
 1011 (orange), and fungi (blue). Statistical support (BS >50) at major divergence points is labeled at the
 1012 nodes. The tree was prepared in Figtree v1.4.2.



1014 **Supplementary Figure 2: Vertebrates phylogeny of FMOs.** Tree was constructed in RAxML
1015 v0.6.0, 1000 bootstraps were run and best-fit model parameters were obtained with ProtTest v3.4.
1016 Bootstrap transfer was applied and TBE values are shown at the nodes. The employed MSA was
1017 trimmed in single sequence extensions and contained 361 taxa and 569 sites. Clades are collapsed
1018 and colored according to tetrapods classes: mammalia (magenta), aves (light orange), amphibia
1019 (green) and testudines (teal). Actinopterygii, coelacanthimorpha and chondrichthyes (all jawed
1020 vertebrates) and the cephalochordate sequence in the root are shown in black. Mammalian FMO
1021 clades are marked inside the clades. The FMOs explosion concordant with the emergence of
1022 terrestrial vertebrates (tetrapods, 352 mya) on the left. The three ancestral nodes that were
1023 experimentally characterized are plotted at the nodes and labelled with yellow squares. The tree was
1024 prepared in Figtree v1.4.2. Accession codes to sequences can be found in [Supplementary Data 1](#).

1025



1026

1027 **Supplementary Figure 3:** Crystal Structure of AncFMO2 in the absence of NADP⁺ (green)
1028 superposed to the structure bound to NADP⁺ (dark green). AncFMO2 crystallizes in an identical
1029 manner with or without NADP⁺. The root-mean-square deviation between the native AncFMO2 and
1030 its NADP⁺ complex is 0.23 Å over 530 Cα atom pairs. The orientation of the dimer depicts the
1031 structure sitting on top of the phospholipid bilayer as shown in the other structures (see **Figure 3**).
1032

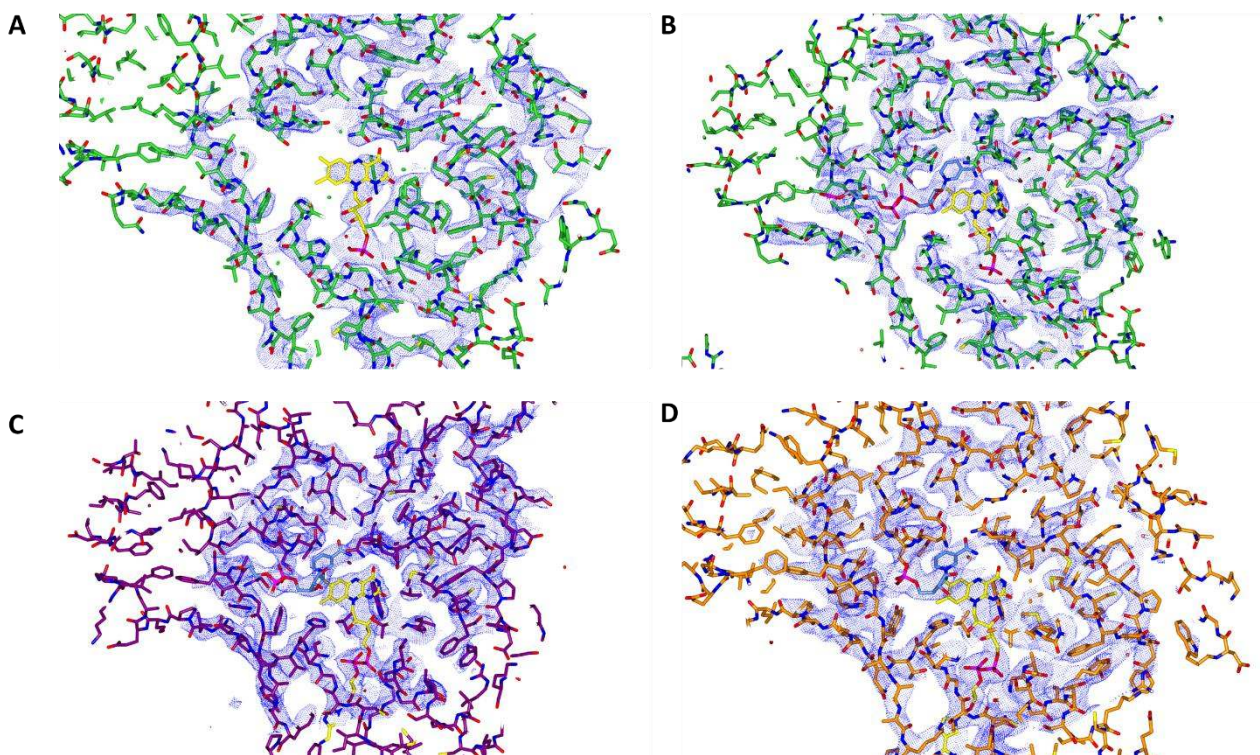
1033

1034

1035

1036

1037



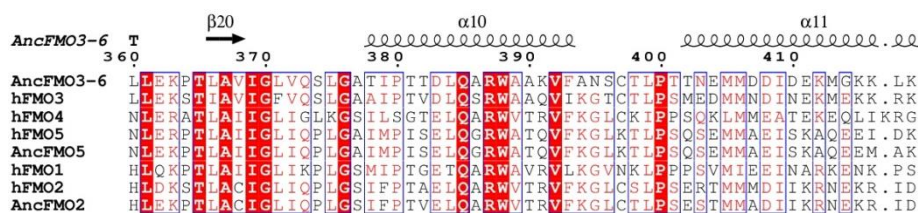
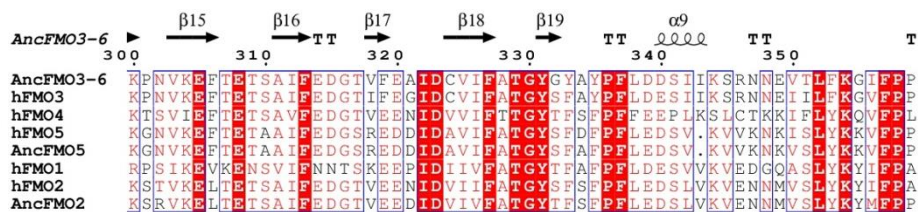
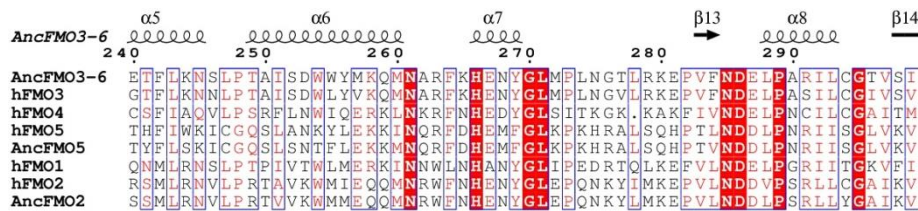
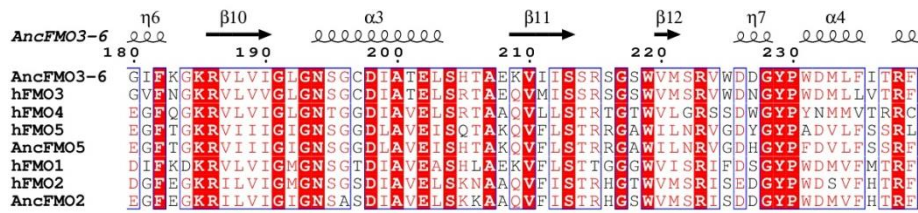
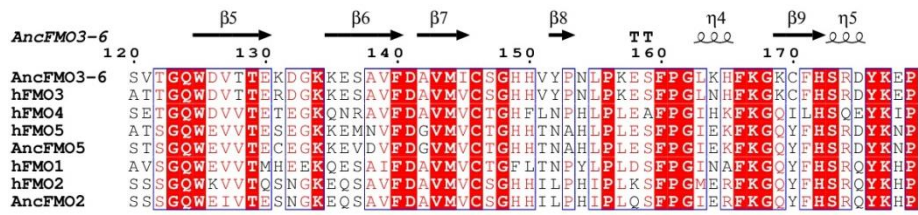
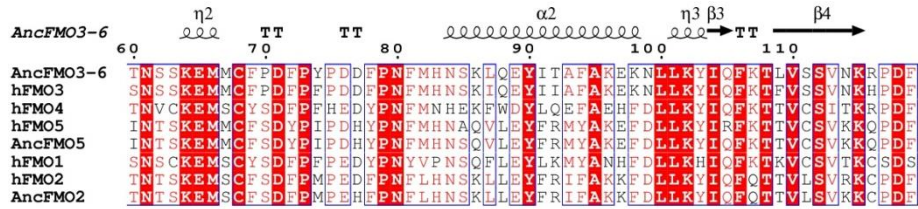
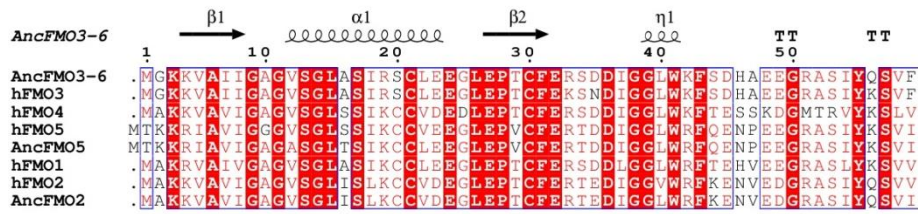
1038 **Supplementary Figure 4:** Electron density maps of the AncFMOs. Structure determination was
 1039 greatly facilitated by density averaging because the asymmetric unit of the crystals of AncFMO2,
 1040 AncFMO3-6, AncFMO5 contain four, six, and two protein molecules. The depicted 2Fo-Fc maps
 1041 were calculated by averaging the electron density maps obtained after molecular replacement
 1042 (shown in blue). **A:** AncFMO2 shown in lime green, without NADP⁺. **B:** AncFMO2 shown in lime
 1043 green, with NADP⁺. **C:** AncFMO3-6 shown in dark magenta, with NADP⁺. **D:** AncFMO5 shown in
 1044 dark orange, with NADP⁺. Cofactors FAD and NADP⁺ are shown in yellow and cornflower blue
 1045 respectively. The contour level is 1.4 σ .

1046

1047

1048

1049



```

AncFMO3-6  Q Q      α12      α13      α14      η8      T
            420      430      440      450      460      470
AncFMO3-6  W F G Q S . . Q T L Q T D Y I T Y M D E L G S F I G A K P N I P W L F L T D P Q L A I E V F F G P C S P Y Q F R L M G P
hFMO3      W F G K S . . E T I Q T D Y I V Y M D E L S S F I G A K P N I P W L F L T D P K L A M E V Y F G P C S P Y Q F R L V G P
hFMO4      V F K D . . T S K D K F D Y I A Y M D D I A A C I G T K P S I P L F L K D P R L A W E V F F G P C T P Y Q Y R L M G P
hFMO5      R Y V E S Q R H T I Q G D Y I D T M E E L A D L V G V R P N L L S L A F T D P K L A L H L L L G P C T P I H Y R V Q G P
AncFMO5    R Y V D S Q R H T I Q G D Y I D T M E E I A D L V G V R P N L L S L A F T D P K L A L K L F F G P C T P V Q Y R L Q G P
hFMO1      W F G L C Y C K A L Q S D Y I T Y I D E L L T Y I N A K P N L F S M L L T D P H L A L T V F F G P C S P Y Q F R L T G P
hFMO2      L F G E S Q S Q T L Q T N Y V D Y L D E L A L E I G A K P D F C S L L F K D P K L A V R L Y F G P C N S Y Q Y R L V G P
AncFMO2    L F G E S Q S Q I L Q T N Y I D Y L D E L A L E I G A K P D L S L L L K D P K L A M K L Y F G P C N S Y Q Y R L V G P

```

```

AncFMO3-6  T      α15      α16      η9      α17      T
            480      490      500      510      520      530
AncFMO3-6  G K W D G A R N A I L T Q W D R T L K P T R T R A V G E A R R P S L F Y N L L K I L L F P V L L L A V L L A F Y . . . .
hFMO3      G Q W P G A R N A I L T Q W D R S L K F M Q T R V V G R L Q K P C F F H W L K L F A I P I L L I A V F L V I T . . . .
hFMO4      G K W D G A R N A I L T Q W D R T L K P L K T R I V P D S S K P A S M S H Y L K A W G A P V L L A S L L L I C K S S L F
hFMO5      G K W D G A R K A I L T D D R I R K P L M T R V V E R S S S M T S T M T I G K F . M L A L A F F A I I I A Y F . . . .
AncFMO5    G K W D G A R K T I L T E D R I R K P L M T R V I E K S N S M T S T M T M G R F . M L A V V F F A I I M A Y F . . . .
hFMO1      G K W E G A R N A I M T Q W D R T R K V I K A R V V Q E S P S P F . . E S F L K V F S F L A L L V A I F L I F L . . . .
hFMO2      G O W E G A R N A I F T Q K Q R I L K P L K T R A L K D S S N F S V S F L . L K I L G . . . L L A V V V A F F C Q L Q
AncFMO2    G O W E G A R N A I F T Q K Q R I L K P L K T R A L K A S S N F P V S F L . L K I L G . . . L L A V V V A F F F Q L Q

```

```

AncFMO3-6
AncFMO3-6  . . . . .
hFMO3      . . . . .
hFMO4      L K L V R D K L Q D R M S P Y L V S L W R G
hFMO5      . . . . .
AncFMO5    . . . . .
hFMO1      . . . . .
hFMO2      W S . . . . .
AncFMO2    W F . . . . .

```

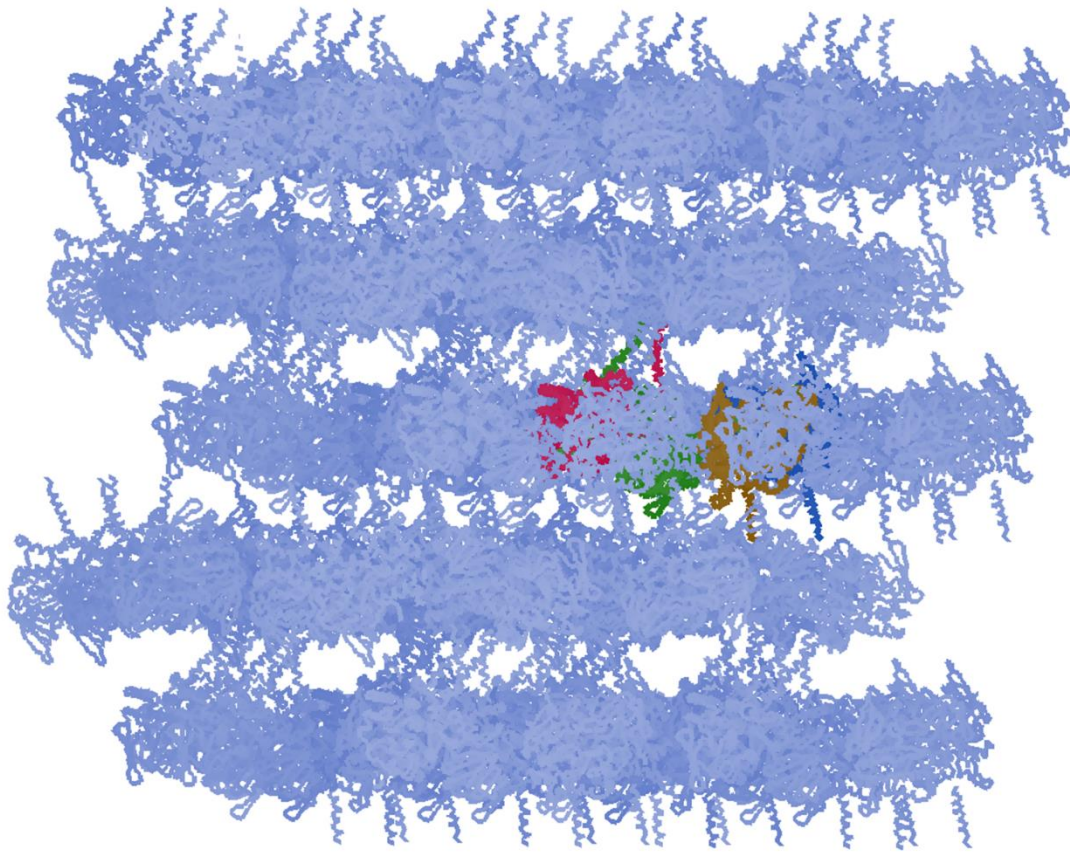
1051

1052 **Supplementary Figure 5:** Structure based sequence alignment using ESPrpt 3.0. The alignment
 1053 was generated by using AncFMO3-6 pdb structure as input file and inserting the Ancestral FMO
 1054 and hFMO sequence files.

1055

1056

1057



1058

1059 **Supplementary Figure 6:** The crystal packing of AncFMO2 forms multiple planes of soluble
1060 dimer-dimer interactions that extend across the lattice. The asymmetric unit is depicted by the four
1061 differently coloured monomer units of dark yellow, dark red, dark green and dark blue. In between
1062 each plane, we see multiple trans-membrane helices projecting upwards and downwards from each
1063 asymmetric unit. Each dimer projects its trans-membrane helices towards its reciprocal dimer.

1064

1065

1066

1067

1068

1069

1070

1071

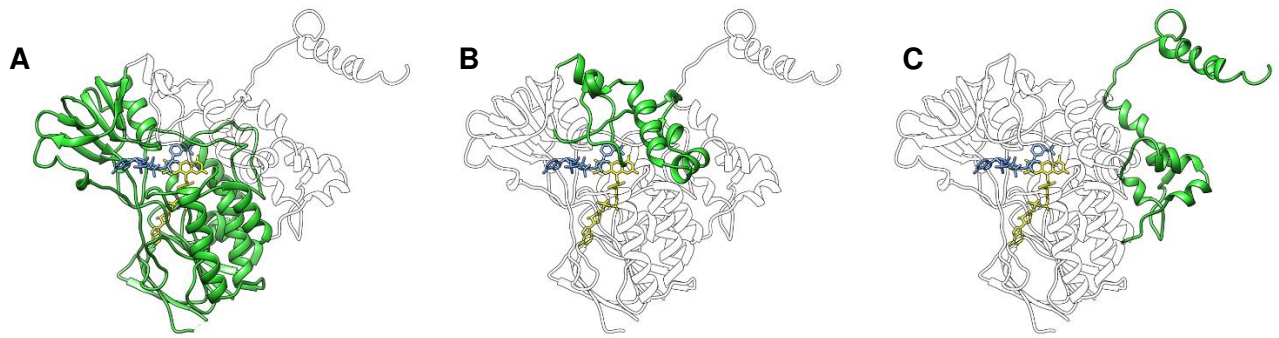
1072

1073

1074

1075

1076



1077

1078 **Supplementary Figure 7:** Topological features of the mammalian FMOs. **A:** Highly conserved
 1079 NADP(H) and FAD dinucleotide binding domains that are observed in all FMOs. **B:** The
 1080 characteristic 80-residue insertion (residues 214-295 in AncFMO3-6) that covers the FAD and
 1081 binds to the membrane monotopically through an alpha helical triad. **C:** The additional C-terminal
 1082 (residues 443-528) that orchestrates both monotopic and bitopic membrane binding features through
 1083 an α -helical triad and a C-terminal helix respectively.
 1084

1085

1086

1087

1088

1089

1090

1091

1092

1093

1094

1095

1096

1097

1098

1099

1100

1101

1102

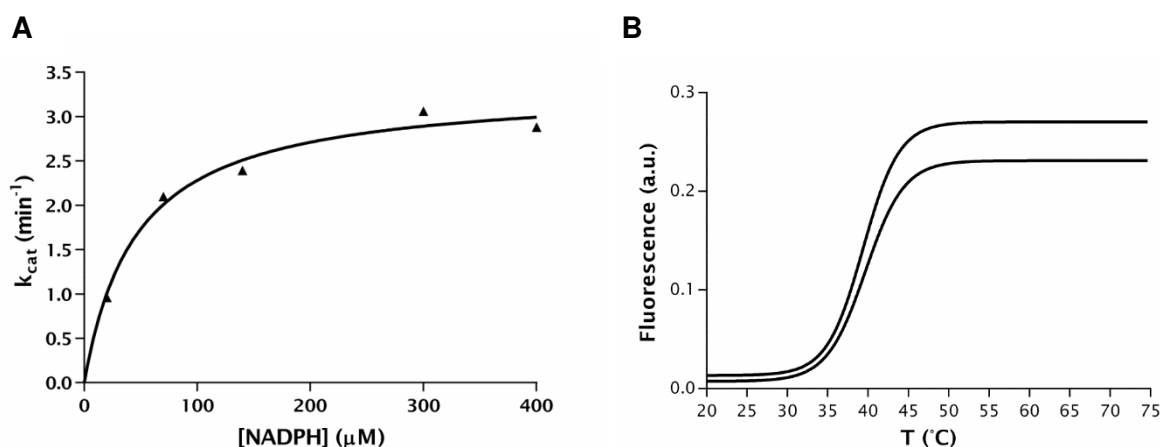
1103

1104

1105

1106

1107



1108

1109 **Supplementary Figure 8: NADPH oxidase activity and Melting temperature of hFMO3.** A:
1110 NADPH consumption was not altered by the presence of substrate and so a Michaelis-Menten curve
1111 was plotted at differing NADPH concentrations. The K_M and k_{cat} for NADPH were determined at $46 \pm 9 \mu\text{M}$
1112 $\pm 9 \mu\text{M}$ and $0.06 \pm 0.16 \text{ s}^{-1}$, respectively. B: Extensive buffer screenings for hFMO3 resulted in a
1113 maximum melting temperature of 44.5 $^{\circ}\text{C}$ (with and without 200 μM NADP^+ in the upper and
1114 lower curves, respectively) in buffer conditions of 100 mM HEPES pH 7.5, 10 mM KCl and 0.05%
1115 (v/v) TRX-100-R.

1116

1117

1118

1119

1120

1121

1122

1123

1124

1125

1126

1127

1128

1129

1130

1131

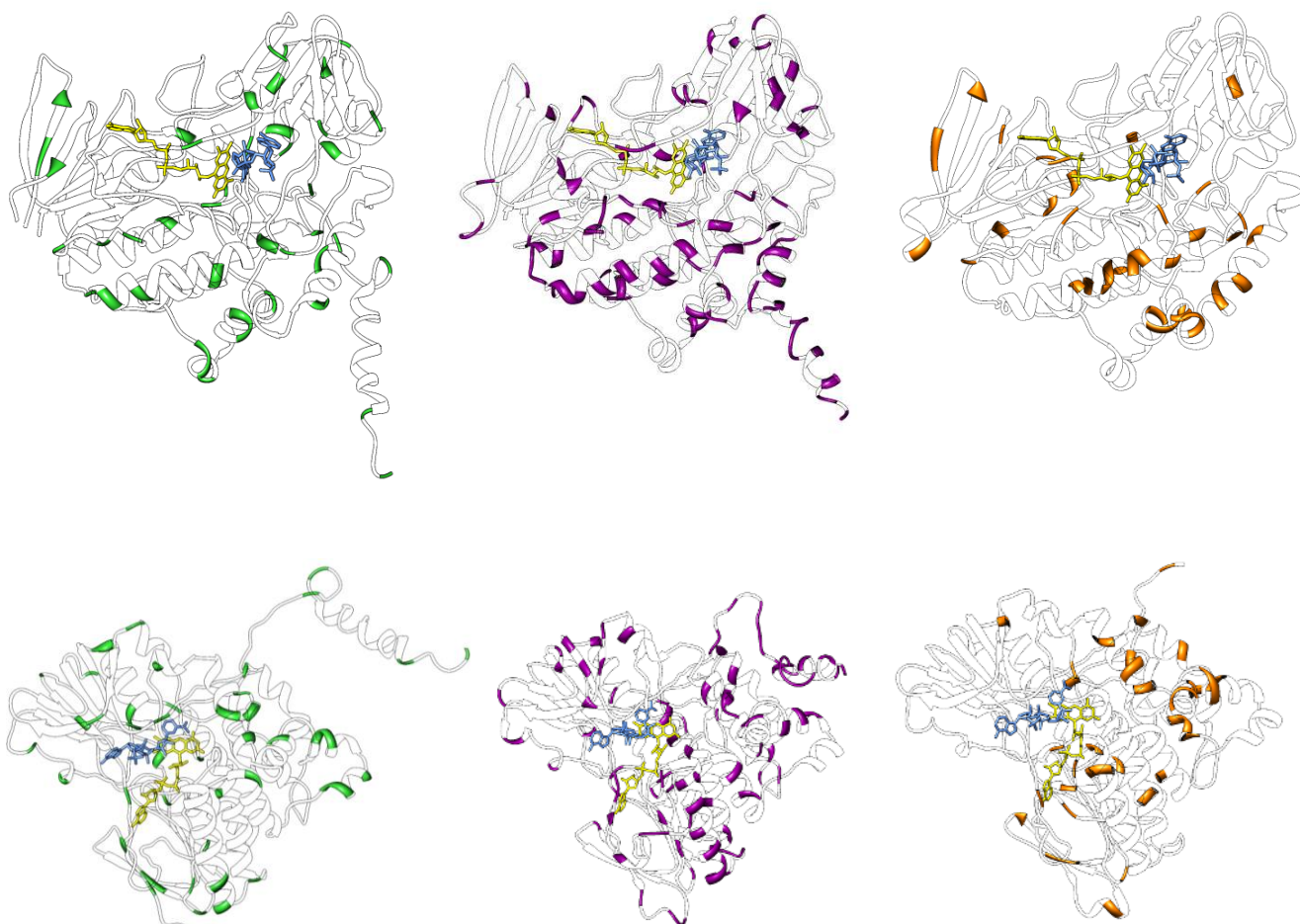
1132

1133

1134

1135

1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170



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.

1171 **Supplementary Table 1:** Crystallographic table for structures AncFMO2, AncFMO3-6 and
 1172 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 ₁ 21	P22 ₁ 2 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

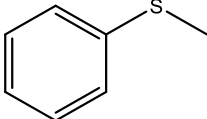
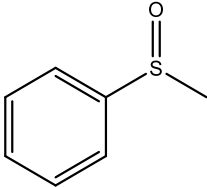
1173 Statistics for the highest-resolution shell are shown in parentheses. Data derived from Aimless and
 1174 Staraniso used for AncFMO2 are described.

1175

1176

1177

1178 **Supplementary Table 2:** Conversions by AncFMO2 and AncFMO3-6 variants.

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

1179 Cyclohexanone showed no conversion for all four enzymes.

1180

1181

1182

1183

1184

1185 **Supplementary Table 3:** Melting Temperatures of each AncFMO

Protein	NADP ⁺ (μM)	T _m (°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

1186

1187 Measured in triplicate by using the ThermoFAD technique (see materials and methods).

1188

1189 **Supplementary Table 4: Ambiguously reconstructed sites in AncFMOs**

AncFMO2					
Site	ML State	PP	Alt State	PP	
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	

1190
 1191 Ambiguously reconstructed sites defined by the Alt state highest posterior probability (PP)>0.2 are
 1192 presented for each AncFMO.
 1193

1194

1195

1196

1197

1198

1199

1200

1201

1202

1203 **Supplementary Data 1: FMOs sequences.** Separate spreadsheet.

1204

1205 **Supplementary Data 2: AncFMOs sequences.**

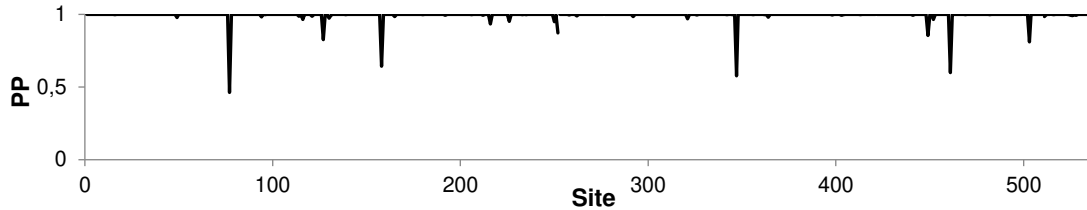
1206 >AncFMO2

1207 MAKKVAVIGAGVSGGLISLKCCVDEGLEPTCFERTEDIGGLWRFKENVEDGRASIYQSVITNTSKEMSCFSDFPMPEHFNP
1208 FLHNSKLLLEYFRIFAKKFDLLKYIQFQTTVLSVKKCPDFSSSGQWEIVTESNGKEQSAVFDAMVCSGHHILPHIPLQSF
1209 PGIERFKGQYFHSRQYKHPEGFEGKRILVIGIGNSASDIARELSKAAQVFI STRHGSWVMSRISDDGYPWDMVFHTRFS
1210 SMLRNVLPRTVVKWMEQQMNRWFNHENYGLEPQNKLYMKEPVLNDDLPSRLLYGAIKVKSRVKELTETS AIFEDGTVEE
1211 DIDVIVFATGYTFSFPFLEDSLKVENNMVSLYKYMFPHPLEKPTLACIGLIQPLGSI FPTVELQARWVTRVFKGLCTLP
1212 SESTMADI IKRNEKRIDLFGESQSQILQNTYIDYLDLELALEIGAKPDLLSLLKDPK LAMKLYFGPCNSYQYRLVGPQG
1213 WEGARNAIFTQQRILKPLKTRALKASSNFPV SFLKILGLLAVVVAFFFLQWFW

1214

1215

Average of reconstruction= 0.994



1216

1217

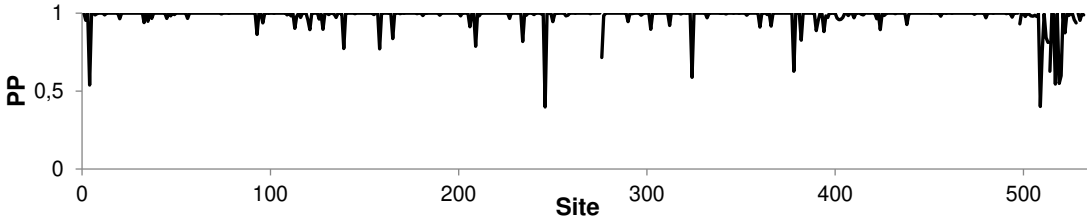
1218 >AncFMO3-6

1219 MGKKVAIIGAGVSGLASIRSCLEEGLEPTCFERSDDIGGLWKFSDHAEGRASIYQSVFTNSSKEMMCFDFPYPDDFPN
1220 FMHNSKLQEYITAFAKEKNLLKYIQFKTLVSSVNKRPDFSVTGQWDVTEKDGKESAVFDAMVICSGHHVYPNLPKESF
1221 PGLKHFKGKCFHSRDYKEPGIFKGRVVLVIGLNSGCDIATELSHTAEKVI I SSRSGSWVMSRVWDDGYPWDMFLITRFE
1222 TFLKNSLPTAISDWYMKQMNARFKHENYGLMPLNGTLRKEPVFNDELPARILCGTVS IKPNVKEFTETS AIFEDGTVEE
1223 AIDCVIFATGYGYAYPFLDDSI IKSRNNEVTLFKGI FPPLEKPTLAVIGLVQSLGATI PTTDLQARWAAKVFANSCTLP
1224 TTNEMDDIDEKMGKLLKWFQSQTLQTDYITYMDELGSFIGAKPNIPWLFLTD PQLALEVFFGPCSPYQFRLMGP GKWD
1225 GARNAILTQWDRTLKPTRTRAVGEAKRPSLFYNLLKILLFPVLLAVLLAFY

1226

1227

Average of reconstruction= 0.982



1228

1229

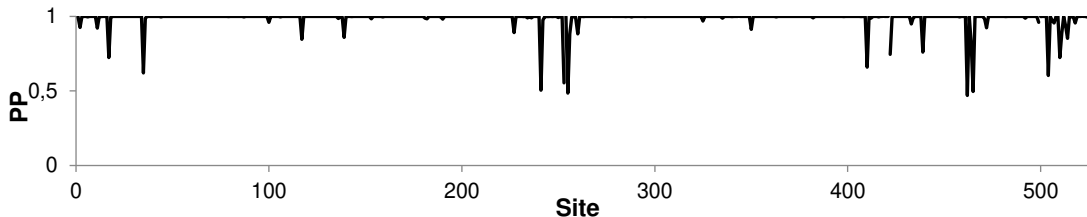
1230 >AncFMO5

1231 MTKKRIAVIGAGASGLT SIKCCLEEGLEPVC FERTDDIGGLWRFQENPEEGRASIYKSVIINTSKEMMCFSDYPIPDHYP
1232 NFMHNSQVLEYFRMYAKEFDLLKYIQFKTTVCSVKKQPDFSTSGQWEVVTECEGKKEVDVFDGVMVCTGHHTNAHLPLES
1233 FPGIEKFKGQYFHSRDYKNPEGFTGKRVI IIGIGNSGGDLAVEISHTAKQVFLSTRGAWILNRVGDHGYPFDFLFSRF
1234 TYFLSKICGQSLNFTLEKMNQRFDEHMFGLKPKHRALSQHPTVNDLDPNRIISGLVKVKGNVKEFTETA AIFEDGSRE
1235 DDIDAVIFATGYSFAFPFLEDSVKKVKNKVS LYKVFPPNLEKPTLAIIGLIQPLGAIMPISELOGRWATQVFKGLKTL P
1236 SQSEMMAEISKAQEEMAKRYVDSQRHTIQGDYIDTMEE IADLVGVRPNLLSLAFTDPKLALKLFFGPCTPVQYRLQGP GK
1237 WDGARKTILTTEDRIRKPLMTRVIEKSN SMTSTMTMGRFMLAVVFFAIIMAYF

1238

1239

Average of reconstruction= 0.9873



1240

1241

1242 For the three characterized ancestral proteins, sequences of the ML (maximum likelihood) ancestors
1243 are shown. Graphs display the posterior probabilities of each reconstructed site for AncFMO2,
1244 AncFMO3-6 and AncFMO5.

1245

1246