

## University of Groningen

### Snapshots of the past

Bailleul, Gautier

DOI:  
[10.33612/diss.180618224](https://doi.org/10.33612/diss.180618224)

**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*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2021

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

*Citation for published version (APA):*  
Bailleul, G. (2021). *Snapshots of the past: studies on extant and extinct flavin-containing monooxygenases*. University of Groningen. <https://doi.org/10.33612/diss.180618224>

#### 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.

# **Snapshots of the past**

Studies on extant and extinct  
flavin-containing monooxygenases

Gautier Bailleul

**The research described in this thesis was performed at**  
Molecular Enzymology Group, Groningen Biomolecular Sciences and Biotechnology  
Institute (GBB), University of the Groningen, The Netherlands.

**The research conducted in this thesis was financially supported by**  
The European Union's Horizon 2020 research and innovation programme under the  
Marie Skłodowska- Curie grant agreement No 722390.

**Printing of this thesis was financially supported by**  
University of Groningen

**Printed by** Ipskamp Printing  
Enschede, the Netherlands

**Design & layout** Bianca Pijl, [www.pijl.design.nl](http://www.pijl.design.nl)  
Groningen, the Netherlands



rijksuniversiteit  
 groningen

# Snapshots of the past

Studies on extant and extinct  
flavin-containing monooxygenases

PhD Thesis

to obtain the degree of PhD at  
the University of Groningen  
on the authority of the  
Rector Magnificus Prof. C. Wijmenga  
and in accordance with  
the decision by the College of Deans.

This thesis will be defended in public  
on Friday 8 October 2021 at 11.00 hours

by

**Gautier Patrick Quentin Bailleul**

born on 05 May 1993  
in Montivilliers, France

**Supervisors**

Prof. M. W. Fraaije

Prof. G. J. Poelarends

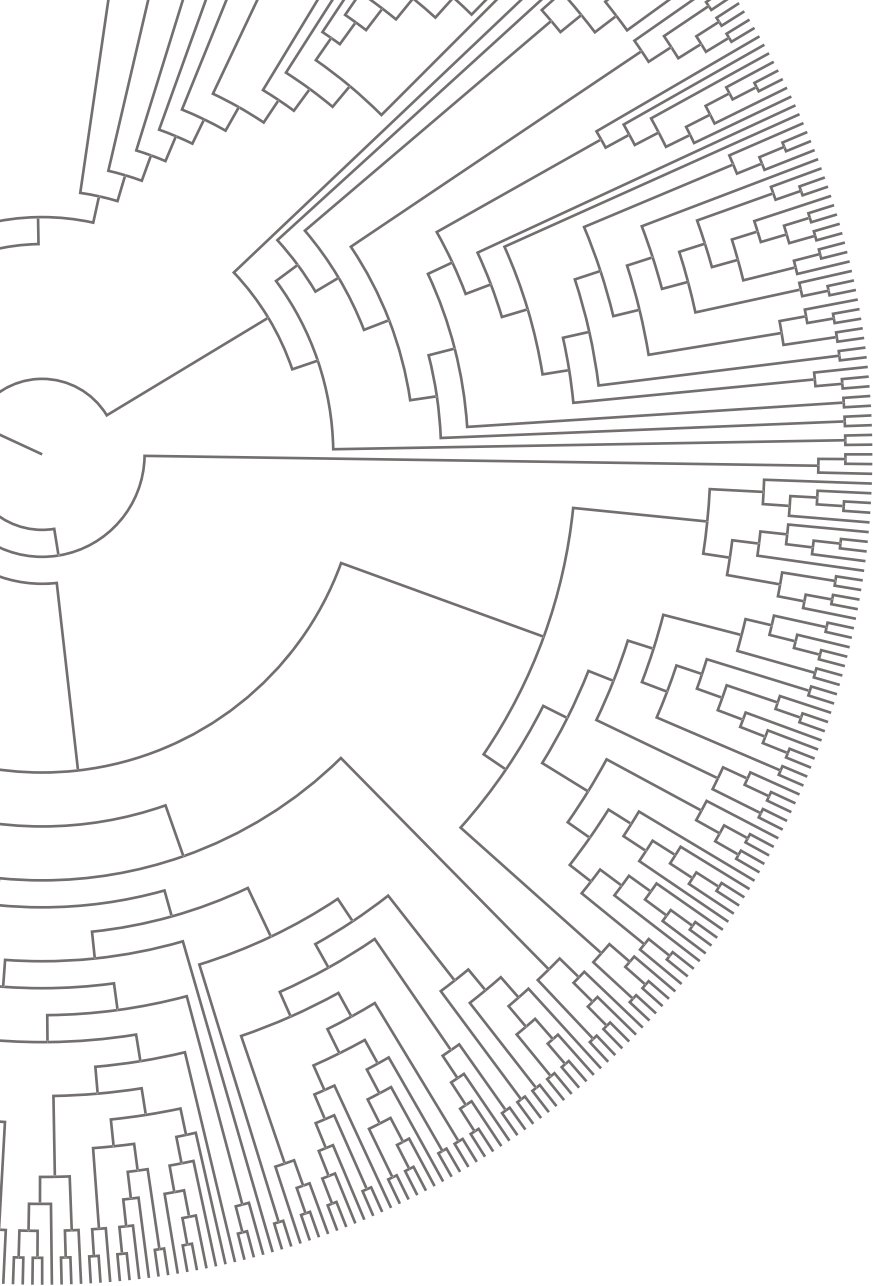
**Assessment Committee**

Prof. E. M. J. Gillam

Prof. D. B. Janssen

Prof. D. Tischler

*Dans ma vie j'avais deux ennemis, la peur de ne pas être assez bon et les poivrons.  
Maintenant j'ai un doctorat et je ne mange plus de poivrons. Merci, de rien, au  
revoir messieurs-dames.*



## Table of contents

Chapter I	General introduction to flavin-containing monooxygenases	9
Chapter II	Characterization of a thermostable flavin-containing monooxygenase from <i>Nitrocola lacisaponensis</i> (NiFMO)	35
Chapter III	Ancestral sequence reconstruction unveils the structural basis of catalysis and membrane binding in mammalian flavin-containing monooxygenases	55
Chapter IV	Ancestral reconstruction of mammalian FMO1 enables structural determination, revealing unique features that explain its catalytic properties	97
Chapter V	Reconstruction and characterization of ancestral tetrapod flavin-containing monooxygenases reveal the evolutionary path to the extant paralogs	147
Chapter VI	Summary and conclusions	183
	<b>Appendices</b>	191
	Nederlandse samenvatting	193
	Résumé en Français	197
	Curriculum Vitae	201
	List of Publications	203
	Acknowledgements	205





## CHAPTER I

# General introduction to flavin-containing monooxygenases



Gautier Bailleul  
Marco W. Fraaije

## **Abstract**

This chapter gives a comprehensive introduction to the classification of flavoprotein monooxygenases and the catalytic mechanisms of flavin-containing monooxygenases (FMOs). Bacterial, plant and mammalian FMOs are discussed in the context of their industrial prospects, role in metabolism and pharmaceutical relevance, respectively. Furthermore, the basal principles of ancestral sequence reconstruction (ASR) are explained as this has been used for a large part of the research on FMOs reported in this thesis.

## I. Flavoprotein monooxygenases

### a. Cofactors

Due to their high selectivity, enzymes receive a lot of attention as potential biocatalytic tools for industrial applications. They exhibit several attractive features for use in biotechnological processes and can be used in a broad range of fields, ranging from the production of biofuels to the synthesis of pharmaceuticals to food processing. The oxidoreductases, or redox enzymes, are particularly attractive for performing selective oxidation and reduction reactions. By their chemo-, regio- and enantioselectivity, they often outcompete chemical approaches. Flavoprotein monooxygenases (FPMOs), belonging to the class of monooxygenases (EC 1.14.x.x), are capable of selective oxygenations. With the help of an electron donor (typically NADH or NADPH), they insert one atom of oxygen, from molecular oxygen, in an organic substrate. Most monooxygenases require at least one type of cofactor for their biocatalytic activity. As their name suggests, FPMOs use a flavin cofactor as prosthetic group which is essential for catalysis. The existence of flavin cofactors, which name comes from the latin word *flavus* in reference to its typical yellow color, was first discovered by Warburg & Christian in 1932<sup>1</sup>. The most common flavin cofactor, flavin adenine dinucleotide (FAD), is formed by phosphorylation of riboflavin (vitamin B2) and coupling to an adenosine monophosphate moiety (Figure 1A).

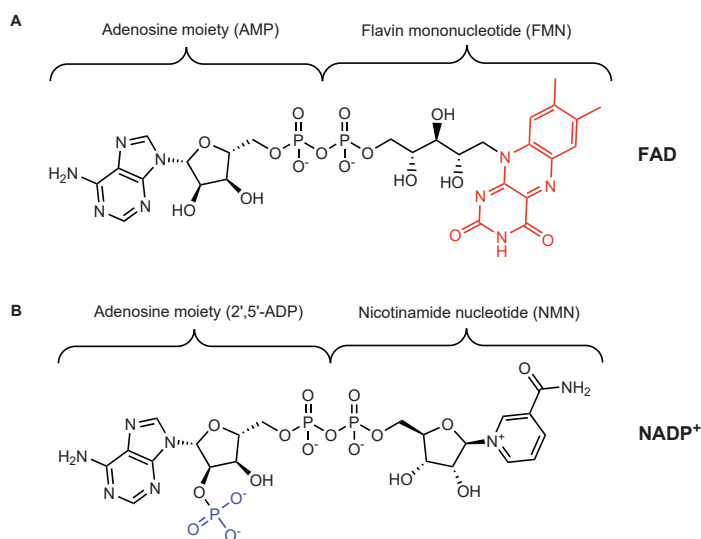


Figure 1. Cofactors structures

A, Flavin adenine dinucleotide (FAD). The isoalloxazine ring is shown in red.

B, Nicotinamide adenine dinucleotide phosphate (NADP<sup>+</sup>), a phosphorylated form of NAD<sup>+</sup>. The extra phosphate in NADP<sup>+</sup> is in blue.

The characteristic yellow color of flavin cofactors comes from the isoalloxazine moiety, a chromophore with a typical absorbance at around 450 nm. The isoalloxazine moiety is also the redox-active center of the flavin cofactor where the electrons are mainly exchanged via the central N<sub>5</sub> or C<sub>4a</sub> atom<sup>2</sup>. The majority of monooxygenases need an external electron donor such as a reduced coenzyme<sup>3</sup>. In this thesis, all the studied FPMOs employ nicotinamide cofactors as electron donors. Two common nicotinamide cofactors exist in nature: nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP). They differ only by the presence of an additional phosphate group on the ribose 2'-position (Figure 1B). In the reduced form, NAD(P)H can act as a hydride donor. Unlike the oxidized form, the reduced form absorbs at 340 nm ( $\epsilon = 6.22 \text{ mM}^{-1}\cdot\text{cm}^{-1}$ ) which makes it possible to follow enzyme activity of NAD(P)H-dependent enzymes by measuring NAD(P)H depletion on a spectrophotometer. This is also a commonly used method to determine FPMO activity. However, in the case of FPMOs, NAD(P)H consumption may also occur without performing substrate oxygenation by a process called uncoupling. Hence, monitoring NADPH depletion is not always the best assay for measuring oxygenase activity of FPMOs.

## b. Classification

The first comprehensive classification of FPMOs was proposed by Van Berkel *et al.* in 2006<sup>3</sup> based on catalytic and structural features. It identifies 6 classes of flavin-dependent monooxygenases (Figure 2). FPMOs can perform different types of oxidations which include hydroxylations, epoxidations, oxygenations of heteroatoms, halogenations and Baeyer-Villiger oxidations. FPMOs from class C to F are using a two-enzyme system expressed by two different genes, unlike class A and B FPMOs that rely on only one protomer. Class A FPMOs are mainly involved in hydroxylations and harbor one dinucleotide-binding domain easily identified by the Rossmann fold sequence motif GxGxxG, essential for FAD or NAD(P)H binding<sup>4</sup>. It reflects the fact that these FPMOs contain a tightly bound FAD in an N-terminal domain while they do not harbor a dedicated domain for binding NAD(P)H. In contrast, class B FPMOs harbor two dinucleotide binding domains for both redox cofactors (FAD and NAD[P]H) and this class includes of type I Baeyer-Villiger monooxygenases (BVMOs), N-hydroxylating monooxygenases (NMOs) and type I & II flavin-containing monooxygenases (FMOs). The BVMO- and FMO-specific sequence fingerprints are almost the same: FXGXXXHXXXW(P/D) and FXGXXXHXXX(Y/F), respectively<sup>5</sup>. As their name suggests, BVMOs catalyze the unusual oxidation reaction, discovered in 1899 by Baeyer and Villiger, by which ketones are converted into esters or lactones<sup>6</sup>. Although sequence-related to BVMOs, type I FMOs (EC 1.14.13.8) mainly catalyze N- or S-oxidations. Such enzymes were first identified in liver microsomes as 'mixed-function oxidases' but their name was later changed to FMOs<sup>7</sup> and the name persisted to describe this subclass even with a growing number of other 'flavin-containing monooxygenases' identified. Unfortunately, the name flavin-containing monooxygenase is also not well chosen as other FPMOs are also often named FMOs. The work described in this thesis mainly involves studies on class B type I FMOs and type I BVMOs.

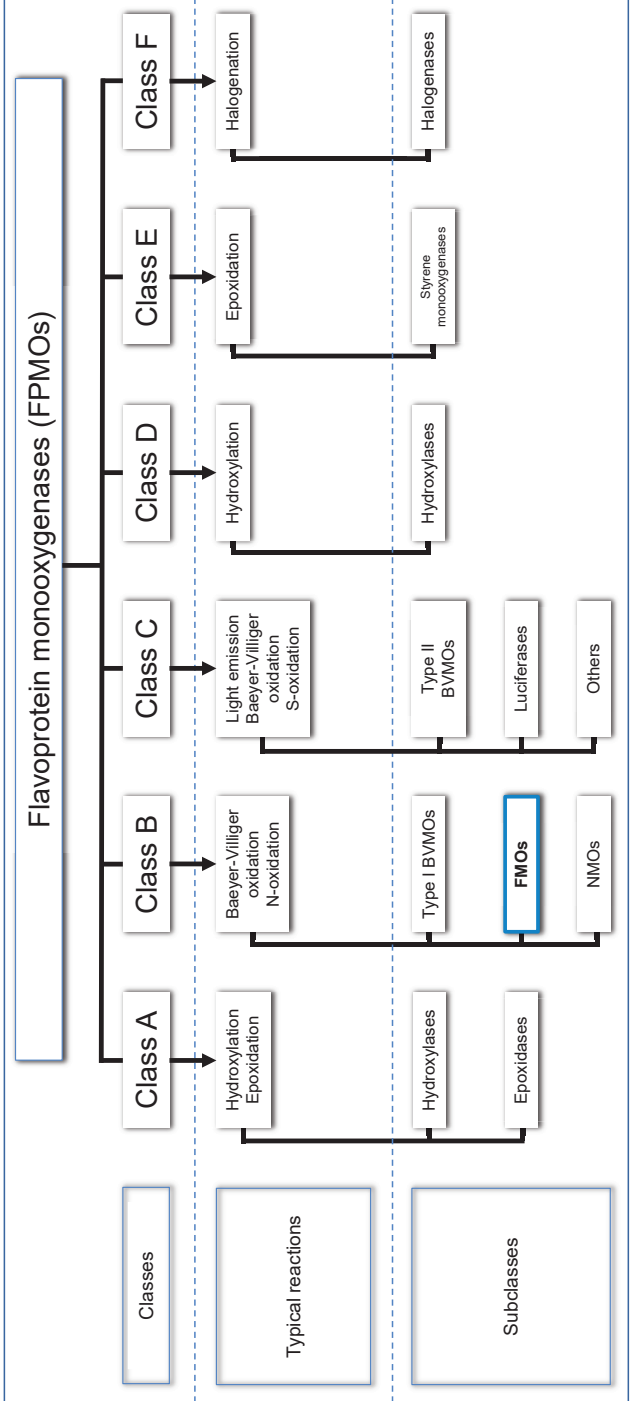


Figure 2. Classification of flavoprotein monooxygenases. Figure adapted from Riebel, A. (2013)<sup>115</sup>.

### c. Flavin-containing monooxygenases (FMOs)

As previously mentioned and as their name suggests, FMOs contain a tightly bound flavin cofactor (FAD). For activity, they also depend on NAD(P)H as cofactor. The catalytic cycle involves the following steps (Figure 3A):

1. The oxidized flavin is reduced by NAD(P)H via a hydride transfer to the N<sub>5</sub> of the isoalloxazine ring.
2. Dioxygen reacts with C<sub>4a</sub> resulting in formation of a reactive intermediate called the C<sub>4a</sub>- (hydro)peroxyflavin.
3. In absence of substrate, the intermediate will decay to the oxidized form of the flavin cofactor while forming hydrogen peroxide.
4. In presence of substrate, the peroxy-intermediate will transfer an oxygen atom to the substrate through a nucleophilic or electrophilic attack.
5. The resulting C<sub>4a</sub>-hydroxyflavin will then release water to regenerate oxidized FAD.
6. The final step corresponds to the release of NAD(P)<sup>+</sup>.

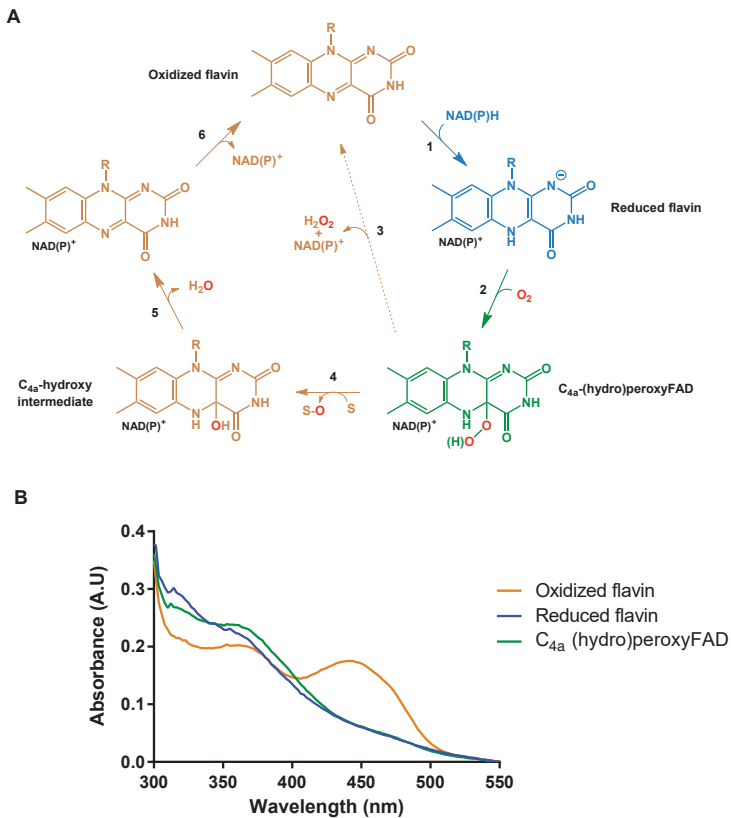


Figure 3. Flavin-containing monooxygenases catalytic cycle.

A, FAD reaction mechanisms on the isoalloxazine ring, adapted from chapter 4.

B, Absorption spectra taken from AncFMO2 in chapter 3 and adapted.

The type of oxygenation occurring in step 4 is dependent on both the type of enzyme and the substrate. An electrophilic attack will happen for N-oxidation, S-oxidation and hydroxylation reactions while a nucleophilic attack takes place when it concerns Baeyer-Villiger oxidations and epoxidations<sup>8</sup>. As a consequence of this mechanism, class B FPMO are sometimes described as a “loaded gun” as they spend most of their time as enzyme species with a reactive C4 $\alpha$ -(hydro)peroxyflavin, ready to transfer their oxygen atom to a substrate<sup>9,10</sup>. However, the relative stability of the peroxy-intermediate highly depends on the enzyme and the surrounding conditions. Formation of significant amounts of H<sub>2</sub>O<sub>2</sub>, known as uncoupling, can occur. As suggested in Figure 2B, it is possible to use UV-Vis absorbance features of the flavin cofactor to follow some of the kinetic steps in FMOs. Contrarywise to the oxidized form, the reduced flavin does not absorb significantly at 450 nm, it is therefore possible to measure the reduction rate in anaerobic conditions using stopped-flow equipment<sup>11</sup>. Additionally, upon mixing molecular oxygen with the reduced form of a FMO, the peroxy-intermediate can be observed which exhibits an absorbance maximum at 360 - 380 nm.

## II. Bacterial FMOs and their possible use in industrial applications

While FMOs were originally discovered and described based on enzymes isolated from mammalian tissue, research in the last few decades has revealed that FMOs are widespread in nature. Several bacterial FMOs have been now studied in detail. In fact, the first structure of a type I FMO was from a bacterial enzyme<sup>10</sup>. With the discovery of bacterial FMOs that are easily expressed as soluble proteins in recombinant hosts, these biocatalysts are nowadays also considered as valuable biocatalysts. The opportunities of bacterial FMOs for industrial applications are demonstrated below with two detailed examples: (1) use of FMOs in processing of food by-products, and (2) use of FMOs for indigo dye production.

*The use of FMOs to valorize fish processing waste* - The increase in human consumption of fish echoes on the amount of fish-related by-products. As it is not allowed to simply discharge unprocessed fish waste in the ocean, it is attractive to develop processes to utilize this protein-rich waste stream<sup>12</sup>. In the last decades, a lot of progress has been made in enzyme-based processing by which these by-products are converted into usable peptides and oils, mostly used for animal food. The use of these fish-based products is nonetheless limited due to the distinctive strong smell caused by trimethylamine. It severely limits applications in the human consumption market<sup>12</sup>. Previously, a FMO from *Methylocella silvestris* was found to be the key factor for this microorganism to grow on TMA as a sole carbon and/or nitrogen source<sup>13</sup>. In their recent work, Goris *et al.* screened several bacterial FMOs for their potential in industrial processing of trimethylamine in salmon protein hydrosylates<sup>14</sup>. The enzymes are catalyzing the oxygenation of TMA into the odorless trimethylamine N-oxide (TMAO) (Figure 4). The best three FMOs, which could be properly expressed, had activity toward TMA and displayed relatively good thermostability, were from the following three bacteria: *Methylophaga aminisulfivorans*, *Roseovarius nubinhibens*, *Nitricola lacisaponensis*. It is worth noting



that two of these enzymes were previously already identified for their potential use in indigo production (NiFMO assessed in chapter 2)<sup>15,16</sup>. Despite encouraging results, further research has to be performed to assess the true feasibility for an industrial use of FMOs in food product processing. Methods are already available to tune FMOs for their performance on TMA, such as directed evolution to improve the catalytic rate and stability, or the introduction of a NADPH recycling system.

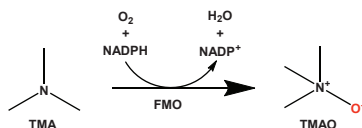
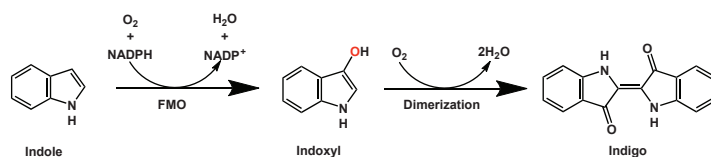


Figure 4. Oxygenation of trimethylamine by FMOs.

*The use of FMOs to produce indigo blue* - Indigo blue, or indigo, is one of the oldest dyes, already used during the late bronze age. Originally, it was extracted from plants like *Indigofera spp.* and *Polygonum tinctorum*<sup>17</sup>. The extraction and isolation process of the indole alkaloids from the plants are not cost-efficient enough to meet the current indigo production estimated to 80 000 tons per year<sup>18,19</sup>. Nowadays, to meet with the growing demand, indigo is almost entirely synthesized via a chemical process. However, the current production relies on cheap petrol-based feedstock, and harsh, not environment-friendly chemicals<sup>20</sup>. An alternative to avoid the creation of large waste streams and the use of non-renewable feedstock would be a fermentative process. There are several enzymatic routes known that start from indole and result in formation of indigo, typically via formation of indoxyl (Figure 5). In 2003, an FMO from the marine bacterium *Methylophaga aminisulfidivorans* (mFMO) was the first recombinant FMO discovered shown to be able to produce indigo in *E. coli*<sup>21</sup>. With some optimization, the use of recombinant *E. coli*, producing mFMO, resulted in 911 mg.L<sup>-1</sup> of indigo in a 3000 L fermenter<sup>22</sup>. The commercial viability of this biotechnological process greatly relies on the improvement made on the bacterial FMO. As mentioned earlier, directed evolution may be effective. In fact, recent engineering of mFMO improved the  $k_{\text{cat}}$  1.5 fold for indole<sup>23</sup>. Furthermore, it has been shown that mFMO can be produced fused to a NADPH-recycling enzyme<sup>23</sup>.



**Figure 5. Enzymatic pathway for indigo formation.** Indole is first transformed into indoxyl by FMOs, which spontaneously dimerizes into indigo in the presence of oxygen.

Recently, mFMO was also fused with tryptophanase, an enzyme naturally present in *E. coli* and that is able to convert tryptophan into indole<sup>24</sup>. Such fusion construct allows boosting the use of tryptophan as a cheap and less toxic precursor than indole. Other valuable indigoids can be produced through the same pathway. For example, indirubin, an often observed by-product of indigo production, may play a role in treatments of leukemia<sup>25</sup>. Production of indirubin was enhanced by supplementing cysteine in the culture, resulting in a switch of the ratio indigo:indirubin produced by mFMO in *E. coli*<sup>26</sup>. In a similar manner, Tyrian purple (6,6'-dibromoindigo) can be produced using an FMO. This dye was popular in previous centuries and was specifically obtained from Mediterranean sea snails<sup>27</sup>. Recently, it was shown that it can be produced in *E. coli* with the help of an FMO<sup>28</sup>. By setting up a two-cell reaction system with a tryptophan 6-halogenase, a flavin reductase, a tryptophanase and mFMO, it was possible to obtain 6,6'-dibromoindigo. Starting with tryptophan, a bromine is added by the halogenase and the subsequent bromotryptophan is then transformed into 6-bromoindole and 6,6'-dibromoindigo by the tryptophanase and mFMO, respectively. In another recent study, it was demonstrated that such biocatalytic cascade can also work by using the isolated enzymes<sup>29</sup>.

### III. Plants FMOs

Plant FMOs are highly more diversified compared to FMOs in other kingdoms<sup>30</sup>. While the human proteome contains 5 FMO paralogs, *Arabidopsis thaliana* and barley (*Hordeum vulgare*) have 29 and 41 FMOs, respectively<sup>31</sup>. Interestingly, based on the available genome sequence data, plants only seem to have class B FPMOs, recognizable by 3 conserved motifs: The FAD-binding motif (GXGXXG), the NADPH-binding motif (GXGXXG) and the FMO-identifying sequence motif (FXGXXXHXXXY/F)5. A recent review from Thodberg *et al.* gives a comprehensive insight on the diversity, functionality and application of plant FMOs<sup>31</sup>.

#### *S*- and *N*-oxidizing enzymes

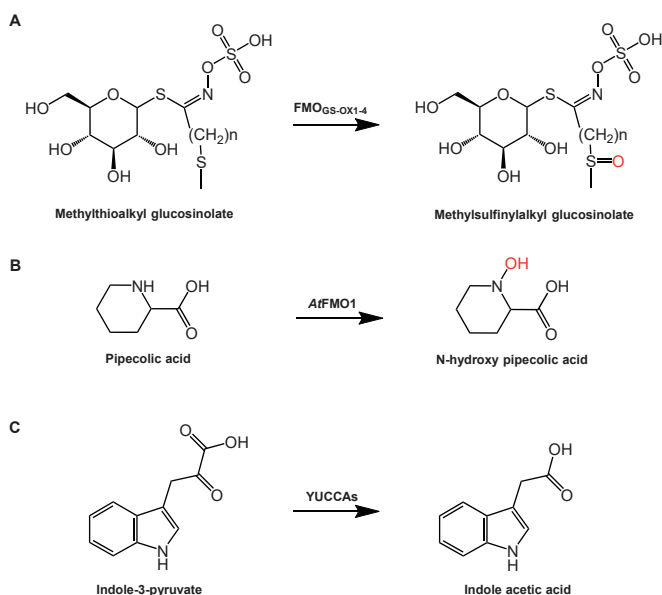
The first sulfur-oxidizing plant FMOs were identified from *Arabidopsis thaliana*. Their functional characterization revealed their involvement in the *S*-oxygenation of methionine-derived glucosinates. FMO<sub>GS-OX1-4</sub> catalyzes the oxygenation of

methylthioalkyl glucosinolate to the corresponding methyl-sulfinylalkyl (Figure 6A). Many orthologs of these FMOs have been found in other land plants<sup>31</sup>, highlighting their importance for the hydrolysis of glucosinolates which are known to possess many biological activities<sup>33</sup>.

Another FMO from *Arabidopsis thaliana*, AtFMO1, is able to induce the systemic acquired resistance (SAR) in the organism. AtFMO1 was shown to catalyze the oxidation of pipecolic acid into N-hydroxypipecolic acid (Figure 6B), the corresponding N-oxide being directly involved in the SAR mechanism<sup>34</sup>.

### YUCCAs

YUCCA from *Arabidopsis thaliana* was the first FMO to be functionally characterized in plants<sup>35</sup>. It was shown to be involved in the production of the plant hormone auxin, a hormone essential for the development of leaves, root and flower. The hormone is crucial in regulating plant growth and fruit development<sup>36,37</sup>. The auxin biosynthesis pathway is essential as the hormone is also involved in stress condition response like drought and shade tolerance<sup>38-40</sup>. YUCCAs catalyze the oxidative decarboxylation of indole-3-pyruvic acid into indole-3-acetic acid, the main auxine hormone (Figure 6C)<sup>41</sup>. YUCCAs are known to be promiscuous and can accept different  $\alpha$ -keto acids for decarboxylation<sup>42</sup> and perform N-hydroxylation of tryptamine<sup>35</sup>. Furthermore, YUC6 presents a second and distinct activity concerning auxine biosynthesis as it possesses a thiol-reductase functionality and its overexpression results in an improved drought tolerance<sup>43</sup>. Further research may elucidate other roles of YUCCAs.



**Figure 6.** Reaction catalyzed by plant FMOs, adapted from Thodberg et al. (2020)<sup>31</sup>. A, S-oxygenation of methionide-derived glucosinolate by FMO<sub>GS-OX1-432</sub>

B, Formation of the N-hydroxy picolic acid by AtFMO1<sup>34</sup>.

C, Decarboxylation of indole-3-pyruvate into indole-3-acetic acid catalyzed by YUCCAs<sup>41</sup>.

## IV. Mammalian FMOs

### a. FMO1-5

Mammalian FMOs play a significant role in the metabolism of nucleophilic heteroatom-containing chemicals and endogenous metabolites<sup>44</sup>. Their relevance is slowly getting more and more acknowledged as disease- and drug-related studies demonstrate their role in detoxification processes<sup>45-47</sup>. The human proteome contains five FMO paralogs, hFMO1-5 (human FMO is typically abbreviated into hFMO). An FMO-encoding pseudogene has been reported but does not encode a functional enzyme<sup>48</sup>. While studied purified FMOs from human or animals report some slight differences in activity<sup>47,49</sup>, the main *in vivo* variations of FMO activity can be largely explained by different expression levels, dependent on tissue type and/or developmental stage<sup>44</sup>.

#### *FMO1*

FMO1 was the first mammalian FMO investigated: it was purified from pig liver microsomes by Ziegler and Mitchell in 1972<sup>50</sup>. The enzyme, first described as a 'mixed function amine oxidase', was shown to be able to perform NADPH- and oxygen-dependent N- and S-oxidations<sup>51</sup>. The kinetic mechanism by which FMO1 performs catalysis was later described by Beaty and Ballou. Their detailed kinetic analyses described both the reductive and oxidative half-reactions of the enzyme and demonstrated the crucial role of a relatively stable peroxyflavin intermediate<sup>52,53</sup>. In rat and rabbit, FMO1 is the most abundant FMO in liver, which is not the case for humans where it is mostly found in the kidneys<sup>54,55</sup>.

FMO1 displays a rather relaxed substrate specificity. An example of a hFMO1 substrate is vandetanib, a multiple tyrosine kinase inhibitor applied for the treatment of medullary thyroid cancer. Radiolabeled vandetanib was found to be excreted in plasma, urine and feces, as different metabolites<sup>56</sup>. Recently, it was shown that, while a specific cytochrome P450 monooxygenase (CYP), CYP3A4, converts vandetanib into N-desmethylvandetanib, hFMO1 and hFMO3 generate the corresponding N-oxide (Figure 7C), with hFMO1 being 5 times more efficient than FMO3<sup>57,58</sup>. These studies exemplify the relevance of FMOs which should not be overlooked as they might produce different metabolites than their prominently-studied cousins, the CYPs.

Another recently recognized hFMO1-substrate is taurine. Taurine is considered to be an essential for mammals as it was found to be one of the most abundant amino acid in mammalian tissues and its deficiency has been involved in many pathologies<sup>59</sup>. Mammals obtain taurine either from their diet or generate it by *de novo* synthesis. Already in 1954, hypotaurine was found to be a precursor of taurine<sup>60</sup>. Nevertheless, the enzyme responsible for that oxidation reaction was never isolated nor identified for more than 65 years. Only in 2020, Veeravalli *et al.* proved that FMO1 is responsible for the conversion of hypotaurine into taurine<sup>61</sup>. They showed that knock-out mice for FMO1, FMO2 and FMO4 presented a deficit in taurine production and that *in vitro* human FMO1 is able to catalyze the oxygenation of hypotaurine into taurine, using NADH or NADPH as cofactor (Figure 7A).

## FMO2

Studies on hFMO2, also called the pulmonary FMO2, have shown that the major allele often encodes a truncated protein, which is catalytically inactive<sup>62</sup>. The early stop-codon at position 472 is a unique feature that emerged in the human lineage when compared with other mammals. Normally, the *fmo2* gene is encoding for a full-length protein of 535 amino acids<sup>62</sup>. In fact, hFMO2 polymorphisms consist of two different, frequently occurring, point mutations and result in either a premature stop codon (T1414C) or a frameshift by insertion of a thymidine<sup>63</sup>. Among the ethnicities investigated, the early-stop codon allele had a frequency of 13% in African-Americans. The overall occurrence of the truncated FMO2 allele is estimated to be between 2 and 7%<sup>64</sup>. The ethnic differences of hFMO2 can be explained by different evolutionary pathways. When expressed, hFMO2 is mostly present in lung tissue, an organ that is typically exposed to toxic chemicals through inhalation. Thus, hFMO2 may have a detoxifying role. Many S-containing chemicals, such as thioacetanilide and thiobenzylamine, have been identified as being able to undergo FMO2-dependent S-oxygenation<sup>65</sup>. These compounds can indeed enter the lung tissue and they are also known as lung toxicants<sup>66</sup>. The full-length hFMO2 was found to catalyze the oxygenation of two antitubercular drugs, thiacetazone and ethionamide (Figure 7B), with a catalytic efficiency 17 to 70 times higher than hFMO1 and hFMO3, respectively<sup>67</sup>. The interindividual variation of the hFMO2 allele can result in an inability to properly metabolize drugs and harmful compounds or activate prodrugs. More work is therefore needed to fully understand the role of hFMO2 and its importance to human metabolism.

## FMO3

hFMO3, mainly present in the liver, is the most studied human FMO. The enzyme receives considerable attention because of the hFMO3-related metabolic disorder trimethylaminuria (TMAU), also called fish-odour syndrome. TMAU is characterized by a strong and pungent odor from patients' urine, sweat and breath, often resulting in social problems, such as anxiety or chronic depression<sup>68</sup>. The smell is due to the accumulation of trimethylamine (TMA), a volatile and smelly aliphatic amine. Typically, the main way of clearance for TMA is its transformation into its non-odorous N-oxide (TMAO) (vide supra), which is secreted in the urine (Figure 4). It was found that TMAU was directly linked to an inactive hFMO3<sup>68</sup>. With its specific expression in the liver and its high activity toward TMA, when compared to other hFMOs, hFMO3 is indeed the main actor for TMA metabolism<sup>69</sup>. More than 40 genetic variants have been identified to cause TMAU. These mutations include missense and nonsense variants as well as deletions<sup>70</sup>. Similarly to hFMO2, some single nucleotide polymorphisms (SNPs) can be observed in different ethnic groups and result in modified activity of hFMO3 *in vitro*<sup>71</sup>. The interest in FMO3 is also linked to the high amount of drug candidates accepted as substrate by the enzyme. In their recent and extensive review, Phillips and Shephard identified in literature over 50 drugs as suitable substrates for hFMO3<sup>70</sup>. Related drugs are from various classes such as antifungal agents<sup>72</sup>, non-steroidal anti-inflammatory drugs<sup>73</sup>, antiestrogens<sup>74</sup>

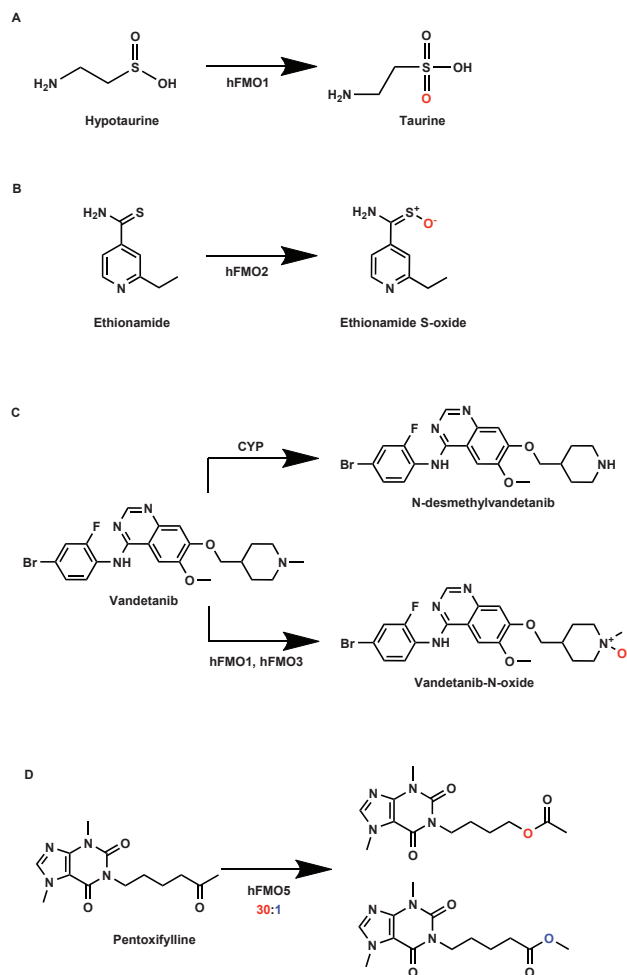
or even psychostimulants (*i.e.* amphetamine and methamphetamine)<sup>75</sup>. In spite of the importance for *in vitro* drug metabolism studies during the early stages of drug development, mammalian FMO3s have mostly been used and studied in microsomal preparations<sup>76,77</sup>. Similar to all mammalian FMOs, hFMO3 is a membrane-associated enzyme which makes it difficult to isolate. Yet, it has been shown that the enzyme can be purified when fused to a maltose binding protein and in presence of a detergent<sup>78,79</sup>. A group claimed in 2012 that by cleaving 17 C-terminal amino acids of hFMO3, they were able to obtain a fully soluble and active enzyme<sup>80</sup>. However, the lack of reproducibility by other groups and a rather cryptic methodology section about the purification raise concerns and brand these results as unreliable. The research on hFMO3 is still limited by the difficult isolation of the membrane protein and the absence of any structural information.

#### **FMO4**

There are no convincing experimental data that would confirm the presence of functional human FMO4. mRNA for expression of hFMO4 is found in several tissues, but in a relatively low yield compared to the other hFMOs<sup>55</sup>. The tissues with the most transcription are the adult liver and kidney<sup>44,77</sup>. The inability to express mammalian FMO4s in *E. coli*, yeast or insect cells<sup>81,82</sup> is most likely due to a problem occurring after transcription. Itagaki *et al.* showed that the lack of expression may be caused by a C-terminal elongation of around 20 amino acids. When relocating the stop codon 81 bases at the 5'-end, truncated-hFMO4 was found to be expressed as an active enzyme with characteristics close to its human paralogs<sup>83</sup>. These data suggest that mammalian FMO4 has lost its function during evolution.

#### **FMO5**

Although it is the most prevalent FMO in adult human liver<sup>44</sup>, hFMO5 is considered atypical due to its ability to perform Baeyer-Villiger oxidations<sup>84</sup>. In contrast to FMO1, FMO2 and FMO3, FMO5 is active on many carbonyl compounds but not so much on soft nucleophiles<sup>85</sup>. While hFMO5 was recently found to be functional as a Baeyer-Villiger monooxygenase, such reactivity could not be predicted based on its sequence features: it does not contain the specific protein sequence motifs characterizing all known bacterial and fungal BVMOs<sup>5,85</sup>. Following the unearthing of the BVMO activity in humans, two drugs, nabumetone (a non-steroidal anti-inflammatory) and pentoxifylline (a phosphodiesterase inhibitor) were exposed as being converted in the corresponding acetate esters by hFMO5 (Figure 7D)<sup>86</sup>. Except for an uncommon catalytic activity, hFMO5 has a substrate scope substantially different from the other hFMOs. Investigations on how hFMO5 is involved in metabolism of xenobiotics or how it can be exploited for drug development, will generate new insights into this atypical hFMO.



**Figure 7. Examples of reaction catalyzed by hFMOs**

*A, Hypotaurine S-oxygenation into taurine by hFMO1<sup>61</sup>.*

*B, Sulfoxidation of ethionamide by hFMO2<sup>67</sup>.*

*C, Vandetanib conversion to N-desmethylvandetanib by CYPs or to Vandetanib-N-oxide by hFMO1 and 3<sup>57</sup>.*

*D, Conversion of pentoxifylline into the corresponding acetate ester by hFMO5<sup>86</sup>.*

## b. Pharmaceutical relevance

An increasing number of studies suggest a relationship between FMOs and several diseases, including trimethylaminuria, atherosclerosis, cardiovascular disease, diabetes, neurodegeneration and iron dyshomeostasis<sup>45</sup>. Many drugs such as tamoxifen, benzydamine, ethionamide or voriconazole are recognized as substrates by hFMOs and are used as antiestrogen, anti-inflammatory, antitubercular or antifungal drugs, respectively<sup>87</sup>. While there are numerous studies showing the relevance of hFMOs in diseases and drug metabolisms, clinical implication studies remain scarce. Members of the cytochrome P450 monooxygenase family (CYPs) have always dominated the research on oxidative metabolism of xenobiotics. This can be explained by the fact that the human proteome contains about one order of magnitude more CYPs when compared with FMOs. This also translates in higher expression levels of CYP activities in the liver compared to FMOs<sup>58</sup>. Furthermore, CYPs have a different mode of action by using a heme cofactor, and are potent oxidizing catalysts, even able to hydroxylate an inactivated carbon. Yet, hFMOs also display unique catalytic properties (*i.e.* B-V oxidation by hFMO5) and may have been overlooked in some drug development or xenobiotic metabolism studies. Excluding hFMOs in such studies may partly be caused by the difficult isolation of the membrane-bound FMOs, leading many research groups to work on ill-defined microsomes<sup>58,88,89</sup>.

In his review about the role of FMOs in drug development, Cashman pointed out some problems related to the use of microsomes and the interest to develop FMOs-metabolized drugs<sup>90</sup>. The starting point of any *in vivo* study is to first investigate the *in vitro* characteristics of drug candidates *i.e.* the stability, identification of the metabolites formed and the enzymes responsible. It is therefore important to also work on isolated enzymes. While working on membrane extract and non-purified enzymes, the role of FMOs can be underestimated because of the presence of dominant CYP activity. According to the author, the limited interest in developing FMOs-metabolized drugs is related to their lower attractiveness in pharmacological studies. The contribution of hFMOs in drug metabolism might be advantageous as relying only on one enzyme system (like CYPs) increases the risk of adverse drug-drug interactions<sup>90</sup>. Secondly, FMOs are not subject to induction or inhibition (the apparent inhibition is only caused by substrate competition) and are producing polar and readily excreted metabolites which may result in fewer toxic consequences and more predictable pharmacokinetics. Finally, hFMOs' variability mainly relies on genetic polymorphism<sup>9</sup>, they can therefore also be advantageous for drug development based on genotype.



## V. Ancestral sequence reconstruction (ASR)

In pursuance of understanding how enzymes perform and their involvement in cellular metabolism, one must look into both their sequence and structure. Because an enzyme function is determined by a sufficient set of residues and their three-dimensional fold, minor changes in the amino acid sequence can result in complete enzyme inactivation or alteration of its function<sup>70,91,92</sup>. Computational methods are increasingly improving the rationalization behind enzymatic behaviors<sup>93</sup>, however, it remains arduous to fully explain variations in enzyme functionality. One method to analyze a specific enzyme is to compare it with a homologous enzyme, displaying a similar sequence and structure but a different function. The approach is called *horizontal*, as it compares different leaves of a phylogenetic tree. Despite being widely employed, this comparative analysis holds pitfalls as it ignores evolution. Extant proteins have evolved over time and not all the changes on their entire sequence are related to functionality<sup>94</sup>. Therefore, when comparing two homologs in the pursuit of understanding differences in enzyme functionality an immense number of variants will be experimentally assessed that will not provide an answer to our question<sup>95</sup>. In order to overcome this limitation, evolutionary biochemists are employing phylogenetics to study the evolutionary history of a protein family, commonly called *vertical* analysis. By analyzing how a protein family has changed over time, it is possible to better understand the effect of mutations and their relevance in the enzyme catalytic function. The first postulation about the possibility to reconstruct amino-acid sequences of ancestral polypeptide chains was made in 1963 by Pauling and Zuckerkandl<sup>96</sup>. Ancestral sequence reconstruction (ASR) aims to reconstruct the most likely sequence of an ancestral protein by inferring the sequences at internal nodes of a phylogenetic tree. Expression and experimental characterization of reconstructed enzymes can answer questions about functionality such as the determinants of substrate specificity, catalytic mechanisms, oligomerization states or specific physico-chemical properties<sup>97</sup>. The recent explosion of genomic sequences and gene synthesis techniques allowed the ASR approach to quickly bloom. In her recent book chapter (contribution to 'Enzyme Engineering' edition in *Methods in Molecular Biology*), Mascotti explains in detail the different steps for a proper ASR methodology. In ASR, once the starting point (extant sequence) has been determined, a homology search is conducted using the enzyme of interest as a query. The most important step is to build a phylogenetic tree that relies on a robust and representative dataset, hence the importance for the workflow to contain enough sequences along with removing unique sequences extensions and poorly aligned regions. A proper substitution model has to be chosen to offset the more frequent substitutions due to the several degrees of chemical likeness between amino acids<sup>98</sup>. Two different methods are used to infer the phylogeny and retrieve the ancestral protein, either Maximum Likelihood or Bayesian inference that relies on bootstrapping and Bayesian statistics, respectively<sup>99,100</sup>. ASR is a combined approach and experimental characterization of the reconstructed enzyme is, therefore, a fundamental prerequisite.

Two features of ancestral proteins that call the attention of protein engineers are their relatively high thermostability and/or their substrate promiscuity<sup>91,101,102</sup>. These features can be attractive when aiming at a new or improved biocatalyst. It has not yet been determined whether these are truly trends as there are no systematic studies of these. Also, in the case when ancestors behave like this the causes are usually unknown. The review by Wheeler *et al.* tackles both the thermostability and specificity of ancestral proteins<sup>103</sup>. They explain that while most of reconstructed proteins display a high stability, it should not always be correlated to a relatively warm environment in the past. The non-exhaustive list of reasons includes different requirements of enzymes for stability, ambiguity in the reconstruction and uncertainty in the relation between the melting temperature and the environmental temperature<sup>103</sup>. Anyway, in some cases ASR can provide a solution for obtaining proteins that display an increased stability by which they also may be easier to crystallize as is demonstrated in chapter 2 and 3 of this thesis. This is also reflected in the rapidly growing number of elucidated structures of reconstructed proteins. In 2017, more than 46 ancestral proteins structures were solved<sup>104</sup>. On a different angle, while being refuted in the phylogenetic field, researchers in the field of biocatalysis often describe ancestral proteins to frequently display a relaxed substrate acceptance profile. Acquiring new interactions or specific activities requires too many genetic changes compared to losing them and are therefore mainly obtained through gene duplication<sup>97</sup>. The loss of biochemical properties is consequently more likely to happen during evolution<sup>105,106</sup>. The most plausible scenario is the presence of many different activities already in the ancestral protein that later distributed those functions via gene duplication with potential fine-tuning<sup>106</sup>. Using ASR, several groups confirmed that extant proteins have a partitioned or enhanced specificity in comparison to their reconstructed ancestor<sup>107-109</sup>. However, there are also studies that reported the emergence of *de novo* functions, that ancestral proteins were lacking<sup>110,111</sup>. In a similar fashion as thermostability, the promiscuity of ancient enzyme is often observed but should not be seen as a fundamental truth. It is known that in some cases, *de novo* functionalities emerged following a gene duplication.

Ancestral reconstruction of protein sequences is a multi-aspect method often only pragmatically used to directly modify the biochemical characteristics of an enzyme. Conveniently, recent work on ancestral proteins unraveled how evolutionary processes are able to produce new molecular structures and functions<sup>112-114</sup>. Overall, ASR has become a powerful tool to expand the knowledge about evolutionary patterns. Work reported in this thesis demonstrates the value of ASR as it has resulted in detailed studies on so-called ancestral FMOs.

## VI. Thesis outline

Industrial biocatalysis and pharmaceuticals are fast-expanding fields that attract more and more attention due to the growing interest in sustainable alternatives. The bedrock for these areas of biotechnology is the seemingly endless power of enzymes. Currently, however, more research has to be done in order to be able to fully exploit the potential of these natural catalysts. Enzymes such as monooxygenases can catalyze the insertion of oxygen atoms into organic substrates with high selectivity. An important class of such selective oxidative biocatalysts are the flavin-containing monooxygenases (FMOs). These biocatalysts bear great promise as future industrial biocatalysts in various biotechnological applications. The research presented in this thesis was done as part of the OXYTRAIN-ITN training network (EU funded, Horizon2020) and aimed at expanding the knowledge about FMOs and to explore their possible applications.

In **Chapter 1**, the family of flavoprotein monooxygenases is described, including the biocatalytic mechanisms and different classes. The role of FMOs in different organisms is discussed together with their relevance in both industrial biocatalysis and pharmaceutical studies. Finally, the approach of ancestral sequence reconstruction (ASR) is discussed.

In **Chapter 2**, through a targeted genome mining approach, a bacterial FMO from *Nitriicola laciaponensis* is biochemically and structurally characterized. The enzyme, originated from an alkaliphilic extremophile organism, is expressed as active and soluble enzyme at high level in *Escherichia coli*. The thermostability, high organic solvent tolerance and broad range of accepted substrates highlight the potential of NiFMO in industrial indigoids production, such as indigo or Tyrian purple.

In **Chapter 3**, the aim was to close the gap in the lack of structural information of hFMOs. Despite their importance in the metabolism of xenobiotics, including toxins, pesticides and drugs, there was no 3D structure of any mammalian FMO available. By using ASR, three mammalian FMO ancestors were reconstructed: AncFMO2, AncFMO3-6 and AncFMO5. Remarkably, crystal structures of all AncFMOs could be elucidated, demonstrating the power of ASR as a strategy for the crystallization and structure determination of proteins. These ancestral structures can be used as accurate homology models for their extant forms as they share a high sequence identity and the same catalytic properties.

In **Chapter 4**, the collection of ancestral mammalian FMOs is completed with the reconstruction of AncFMO1 and AncFMO4. FMO4 is not expressed in mammals, most likely due to a C-terminal extension (see above). Also, the mammalian AncFMO4 could not be expressed. However, AncFMO1 could be expressed, purified and crystallized. The respective crystal structure revealed a dimeric protein structure. AncFMO1 displayed a typical FMO activity as demonstrated by the oxygenation of benzydamine, tamoxifen

and thioanisole. The activity on these drug-related compounds demonstrate that the ancestral form is a good mimic for the extant hFMO1. It was also found that AncFMO1 is able to use both NADH and NADPH as cofactor.

In **Chapter 5**, research is presented that aimed at identifying the changes in mammalian FMOs that results in the divergence into different monooxygenase (FMO vs. BVMO) activities. For this, the tetrapod ancestors of all FMOs were reconstructed. It was determined that the first tetrapod ancestor had the ability to perform both typical FMO and Baeyer-Villiger oxidations. Following gene duplication, this ability was split between FMO5, that retained the BVMO activity, and FMO1-4 which specialized in S- and N-oxidations. The main residues responsible for the switch were unveiled and reversing the activity by selective mutagenesis was shown to be possible. These findings shed more light on how the different FMOs evolved, and the obtained insights will help to redesign FMO and other sequence-related enzymes.

## References

- 1 Warburg, O. & Christian, W. Uber ein neues oxydationsferment und sein absorptionsspektrum. *Biochem Z* 254, 438–458 (1932).
- 2 Walsh, C. T. & Wenczewicz, T. A. Flavoenzymes: Versatile catalysts in biosynthetic pathways. *Nat. Prod. Rep.* 30, 175–200 (2013).
- 3 van Berkel, W. J. H., Kamerbeek, N. M. & Fraaije, M. W. Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *J. Biotechnol.* 124, 670–689 (2006).
- 4 Wierenga, R. K., Terpstra, P. & Hol, W. G. J. Prediction of the occurrence of the ADP-binding  $\beta\alpha\beta$ -fold in proteins, using an amino acid sequence fingerprint. *J. Mol. Biol.* 187, 101–107 (1986).
- 5 Fraaije, M. W., Kamerbeek, N. M., Van Berkel, W. J. H. & Janssen, D. B. Identification of a Baeyer-Villiger monooxygenase sequence motif. *FEBS Lett.* 518, 43–47 (2002).
- 6 Baeyer, A. & Villiger, V. Einwirkung des Caro'schen Reagens auf Ketone. *Ber. Dtsch. Chem. Ges.* 32, 3625–3633 (1899).
- 7 Ziegler, D. M. An overview of the mechanism, substrate specificities, and structure of FMOs. *Drug Metab. Rev.* 34, 503–511 (2002).
- 8 Walsh, C. T. & Chen, Y. J. Enzymic Baeyer-Villiger Oxidations by Flavin-Dependent Monooxygenases. *Angew. Chemie - Int. Ed.* 333–343 (1988).
- 9 Krueger, S. K. & Williams, D. E. Mammalian flavin-containing monooxygenases: Structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacol. Ther.* 106, 357–387 (2005).
- 10 Alfieri, A., Malito, E., Orru, R., Fraaije, M. W. & Mattevi, A. Revealing the moonlighting role of NADP in the structure of a flavin-containing monooxygenase. *Proc. Natl. Acad. Sci. U. S. A.* 105, 6572–6577 (2008).
- 11 Romero, E., Robinson, R. & Sobrado, P. Monitoring the reductive and oxidative half-reactions of a flavin-dependent monooxygenase using stopped-flow spectrophotometry. *J. Vis. Exp.* (2012) doi:10.3791/3803.
- 12 Kristinsson, H. G. & Rasco, B. A. Critical Reviews in Food Science and Nutrition Fish Protein Hydrolysates: Production, Biochemical, and Functional Properties. *Crit. Rev. Food Sci. Nutr.* 40, 43–81 (2000).
- 13 Chen, Y., Patel, N. A., Crombie, A., Scrivens, J. H. & Murrell, J. C. Bacterial flavin-containing monooxygenase is trimethylamine monooxygenase. *Proc. Natl. Acad. Sci. U. S. A.* 108, 17791–17796 (2011).
- 14 Goris, M. et al. Flavin containing monooxygenases for conversion of trimethylamine in salmon protein hydrolysates. *Appl. Environ. Microbiol.* (2020) doi:10.1128/aem.02105-20.
- 15 Han, G. H., Shin, H. J. & Kim, S. W. Optimization of bio-indigo production by recombinant *E. coli* harboring *fmo* gene. *Enzyme Microb. Technol.* 42, 617–623 (2008).
- 16 Lončar, N. et al. Characterization of a thermostable flavin-containing monooxygenase from *Nitrocola laciasaponensis* (NiFMO). *Appl. Microbiol. Biotechnol.* 103, 1755–1764 (2019).
- 17 Krizova, H. Natural dyes: their past, present, future and sustainability. *Recent Dev. Fibrous Mater. Sci.* 59–71 (2015).
- 18 Franssen, M.C.R.; Kircher, M.; Wohlgemuth, R. Industrial biotechnology in the chemical and pharmaceutical industry. *Ind. Biotechnol. Sustain. Growth Econ. Success* 323–350 (2010).
- 19 Brown, S., Clastre, M., Courdavault, V. & O'Connor, S. E. De novo production of the plant-derived alkaloid strictosidine in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 112, 3205–3210 (2015).
- 20 Fabara, A. N. & Fraaije, M. W. An overview of microbial indigo-forming enzymes. *Appl. Microbiol. Biotechnol.* 104, 925–933 (2020).

- 21 Choi, H. S. et al. A novel flavin-containing monooxygenase from *Methylophaga* sp. strain SK1 and its indigo synthesis in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 306, 930–936 (2003).
- 22 Han, G. H. et al. Bio-indigo production in two different fermentation systems using recombinant *Escherichia coli* cells harboring a flavin-containing monooxygenase gene (*fmo*). *Process Biochem.* 46, 788–791 (2011).
- 23 Lončar, N., van Beek, H. L. & Fraaije, M. W. Structure-Based Redesign of a Self-Sufficient Flavin-Containing Monooxygenase towards Indigo Production. *Int. J. Mol. Sci.* 20, 6148 (2019).
- 24 Fabara, A. N. & Fraaije, M. W. Production of indigo through the use of a dual-function substrate and a bifunctional fusion enzyme. *Enzyme Microb. Technol.* 142, 109692 (2020).
- 25 Wang, L. et al. Dissection of mechanisms of Chinese medicinal formula Realgar-Indigo naturalis as an effective treatment for promyelocytic leukemia. *Proc. Natl. Acad. Sci. U. S. A.* 105, 4826–4831 (2008).
- 26 Han, G. H., Gim, G. H., Kim, W., Seo, S. Il & Kim, S. W. Enhanced indirubin production in recombinant *Escherichia coli* harboring a flavin-containing monooxygenase gene by cysteine supplementation. *J. Biotechnol.* 164, 179–187 (2012).
- 27 Ziderman, I. I. Purple Dyes Made from Shellfish in Antiquity. *Rev. Prog. Color. Relat. Top.* 16, 46–52 (2008).
- 28 Lee, J. et al. Production of Tyrian purple indigoid dye from tryptophan in *Escherichia coli*. *Nat. Chem. Biol.* 17, (2020).
- 29 Schnepel, C., Doderer, V. I. & Sewald, N. Novel Arylindigoids by Late-Stage Derivatization of Biocatalytically Synthesized Dibromoindigo. *Chem. - A Eur. J.* 27, 5404–5411 (2021).
- 30 Nelson, D. & Werck-Reichhart, D. A P450-centric view of plant evolution. *Plant J.* 66, 194–211 (2011).
- 31 Thodberg, S. & Neilson, E. H. J. Application of Plant Flavoproteins. *Catalysts* 10, 329 (2020).
- 32 Li, J., Hansen, B. G., Ober, J. A., Kliebenstein, D. J. & Halkier, B. A. Subclade of flavin-monoxygenases involved in aliphatic glucosinolate biosynthesis. *Plant Physiol.* 148, 1721–1733 (2008).
- 33 Halkier, B. A. & Gershenzon, J. Biology and Biochemistry of Glucosinolates. (2006) doi:10.1146/annurev.arplant.57.032905.105228.
- 34 Hartmann, M. et al. Flavin Monooxygenase-Generated N-Hydroxypipecolic Acid Is a Critical Element of Plant Systemic Immunity. *Cell* 173, 456–469.e16 (2018).
- 35 Zhao, Y. et al. A role for flavin monooxygenase-like enzymes in auxin biosynthesis. *Science* (80-.). 291, 306–309 (2001).
- 36 Cheng, Y., Dai, X. & Zhao, Y. Auxin synthesized by the YUCCA flavin monooxygenases is essential for embryogenesis and leaf formation in *Arabidopsis*. *Plant Cell* 19, 2430–2439 (2007).
- 37 Cheng, Y., Dai, X. & Zhao, Y. Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in *Arabidopsis*. *Genes Dev.* 20, 1790–1799 (2006).
- 38 Lee, M. et al. Activation of a flavin monooxygenase gene YUCCA7 enhances drought resistance in *Arabidopsis*. *Planta* 235, 923–938 (2012).
- 39 Tivendale, N. D., Ross, J. J. & Cohen, J. D. The shifting paradigms of auxin biosynthesis. *Trends Plant Sci.* 19, 44–51 (2014).
- 40 Müller-Moullé, P. et al. YUCCA auxin biosynthetic genes are required for *Arabidopsis* shade avoidance. *PeerJ* 2016, (2016).
- 41 Mashiguchi, K. et al. The main auxin biosynthesis pathway in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 108, 18512–18517 (2011).

- 42 Dai, X. et al. The biochemical mechanism of auxin biosynthesis by an arabidopsis YUCCA flavin-containing monooxygenase. *J. Biol. Chem.* 288, 1448–1457 (2013).
- 43 Cha, J. Y. et al. A novel thiol-reductase activity of Arabidopsis YUC6 confers drought tolerance independently of auxin biosynthesis. *Nat. Commun.* 6, 1–13 (2015).
- 44 Cashman, J. R. & Zhang, J. Human flavin-containing monooxygenases. *Annu. Rev. Pharmacol. Toxicol.* 46, 65–100 (2006).
- 45 Rossner, R., Kaeberlein, M. & Leiser, S. F. Flavin-containing monooxygenases in aging and disease: Emerging roles for ancient enzymes. *J. Biol. Chem.* 292, 11138–11146 (2017).
- 46 Canyelles, M. et al. Trimethylamine N-oxide: A link among diet, gut microbiota, gene regulation of liver and intestine cholesterol homeostasis and HDL function. *Int. J. Mol. Sci.* 19, (2018).
- 47 Uno, Y., Shimizu, M. & Yamazaki, H. Molecular and functional characterization of flavin-containing monooxygenases in cynomolgus macaque. *Biochem. Pharmacol.* 85, 1837–1847 (2013).
- 48 Hines, R. N., Hopp, K. A., Franco, J., Saeian, K. & Begun, F. P. Alternative processing of the human FMO6 gene renders transcripts incapable of encoding a functional flavin-containing monooxygenase. *Mol. Pharmacol.* 62, 320–325 (2002).
- 49 Hu, S. X. Hepatic Flavin-containing Monooxygenase and Aldehyde Oxidase Activities in Male Domestic Pigs at Different Ages. *Drug Metab. Lett.* 12, 125–131 (2018).
- 50 Ziegler, D. M. & Mitchell, C. H. Microsomal oxidase IV: Properties of a mixed-function amine oxidase isolated from pig liver microsomes. *Arch. Biochem. Biophys.* 150, 116–125 (1972).
- 51 Ziegler, D. M. & Poulsen, L. L. Hepatic microsomal mixed-function amine oxidase. *Methods Enzymol.* 52, 142–151 (1978).
- 52 Beaty, N. B. & Ballou, D. P. The reductive half-reaction of liver microsomal FAD-containing monooxygenase. *J. Biol. Chem.* 256, 4611–4618 (1981).
- 53 Beaty, N. B. & Ballou, D. P. The oxidative half-reaction of liver microsomal FAD-containing monooxygenase. *J. Biol. Chem.* 256, 4619–4625 (1981).
- 54 Cashman, J. R. Structural and Catalytic Properties of the Mammalian Flavin-Containing Monooxygenase. *Chem. Res. Toxicol.* 8, 165–181 (1995).
- 55 Dolphin, C. T., Cullingford, T. E., Shephard, E. A., Smith, R. L. & Phillips, I. R. Differential developmental and tissue-specific regulation of expression of the genes encoding three members of the flavin-containing monooxygenase family of man, FMO1, FMO3 and FMO4. *Eur. J. Biochem.* 235, 683–689 (1996).
- 56 Martin, P. et al. Pharmacokinetic Drug Interactions with Vandetanib during Coadministration with Rifampicin or Itraconazole. *Drugs R D* 11, 37–51 (2011).
- 57 Indra, R. et al. Cytochrome P450 and flavin-containing monooxygenase enzymes are responsible for differential oxidation of the anti-thyroid-cancer drug vandetanib by human and rat hepatic microsomal systems. *Environ. Toxicol. Pharmacol.* 74, (2020).
- 58 Indra, R. et al. Identification of human enzymes oxidizing the anti-thyroid-cancer drug vandetanib and explanation of the high efficiency of cytochrome P450 3A4 in its oxidation. *Int. J. Mol. Sci.* 20, (2019).
- 59 Huxtable, R. J. Physiological Actions of Taurine. *Physiol. Rev.* 72, 101–163 (1992).
- 60 Cavallini, D., De Marco, C., Mondovì, B. & Stirpe, F. The biological oxidation of hypotaurine. *BBA - Biochim. Biophys. Acta* 15, 301–303 (1954).
- 61 Veeravalli, S. et al. Flavin-containing monooxygenase 1 catalyzes the production of taurine from hypotaurine. *Drug Metab. Dispos.* 48, 378–385 (2020).

- 62 Dolphin, C. T. et al. The flavin-containing monooxygenase 2 gene (FMO2) of humans, but not of other primates, encodes a truncated, nonfunctional protein. *J. Biol. Chem.* 273, 30599–30607 (1998).
- 63 Whetstine, J. R. et al. Ethnic differences in human flavin-containing monooxygenase 2 (FMO2) polymorphisms: Detection of expressed protein in African-Americans. *Toxicol. Appl. Pharmacol.* 168, 216–224 (2000).
- 64 Krueger, S. K. et al. Differences in FMO2\*1 allelic frequency between Hispanics of Puerto Rican and Mexican descent. *Drug Metab. Dispos.* 32, 1337–1340 (2004).
- 65 Krueger, S. K., Martin, S. R., Yueh, M. F., Pereira, C. B. & Williams, D. E. Identification of active flavin-containing monooxygenase isoform 2 in human lung and characterization of expressed protein. *Drug Metab. Dispos.* 30, 34–41 (2002).
- 66 Cashman, J. R., Traiger, G. J. & Hanzlik, R. P. Pneumotoxic effects of thiobenzamide derivatives. *Toxicology* 23, 85–93 (1982).
- 67 Francois, A. A., Nishida, C. R., Ortiz De Montellano, P. R., Phillips, I. R. & Shephard, E. A. Human flavin-containing monooxygenase 2.1 catalyzes oxygenation of the antitubercular drugs thiacetazone and ethionamide. *Drug Metab. Dispos.* 37, 178–186 (2009).
- 68 Mitchell, S. C. & Smith, R. L. Trimethylaminuria: The fish malodor syndrome. *Drug Metab. Dispos.* 29, 517–521 (2001).
- 69 Lang, D. H. et al. Isoform specificity of trimethylamine N-oxygenation by human flavin-containing monooxygenase (FMO) and P450 enzymes Selective catalysis by fmo3. *Biochem. Pharmacol.* 56, 1005–1012 (1998).
- 70 Phillips, I. R. & Shephard, E. A. Flavin-containing monooxygenase 3 (FMO3): genetic variants and their consequences for drug metabolism and disease. *Xenobiotica* 50, 19–33 (2020).
- 71 Cashman, J. R., Akerman, B. R., Forrest, S. M. & Treacy, E. P. Population-specific polymorphisms of the human FMO3 gene: Significance for detoxication. *Drug Metab. Dispos.* 28, 169–173 (2000).
- 72 El-Serafi, I. et al. Flavin-containing monooxygenase 3 (FMO3) role in busulphan metabolic pathway. *PLoS One* 12, (2017).
- 73 Mayatepek, E., Flock, B. & Zschocke, J. Benzydamine metabolism in vivo is impaired in patients with deficiency of flavin-containing monooxygenase 3. *Pharmacogenetics* 14, 775–777 (2004).
- 74 Parte, P. & Kupfer, D. Oxidation of tamoxifen by human flavin-containing monooxygenase (FMO) 1 and FMO3 to tamoxifen-N-oxide and its novel reduction back to tamoxifen by human cytochromes P450 and hemoglobin. *Drug Metab. Dispos.* 33, 1446–1452 (2005).
- 75 Cashman, J. R., Xiong, Y. N., Lifan, X. & Janowsky, A. N-oxygenation of amphetamine and methamphetamine by the human flavin-containing monooxygenase (form 3): Role in bioactivation and detoxication. *J. Pharmacol. Exp. Ther.* 288, 1251–1260 (1999).
- 76 Dolphin, C. T., Janmohamed, A., Smith, R. L., Shephard, E. A. & Phillips, I. R. Compound heterozygosity for missense mutations in the flavin-containing monooxygenase 3 (FMO3) gene in patients with fish-odour syndrome. *Pharmacogenetics* 10, 799–807 (2000).
- 77 Zhang, J. & Cashman, J. R. Quantitative analysis of FMO gene mRNA levels in human tissues. *Drug Metab. Dispos.* 34, 19–26 (2006).
- 78 Brunelle, A. et al. Characterization of two human flavin-containing monooxygenase (form 3) enzymes expressed in *Escherichia coli* as maltose binding protein fusions. *Drug Metab. Dispos.* 25, 1001–1007 (1997).



- 79 Reddy, R. R., Ralph, E. C., Motika, M. S., Zhang, J. & Cashman, J. R. Characterization of human flavin-containing monooxygenase (FMO) 3 and FMO5 expressed as maltose-binding protein fusions. *Drug Metab. Dispos.* 38, 2239–2245 (2010).
- 80 Catucci, G., Gilardi, G., Jeuken, L. & Sadeghi, S. J. In vitro drug metabolism by C-terminally truncated human flavin-containing monooxygenase 3. *Biochem. Pharmacol.* 83, 551–558 (2012).
- 81 Phillips, I. R. et al. The molecular biology of the flavin-containing monooxygenases of man. *Chem. Biol. Interact.* 96, 17–32 (1995).
- 82 Burnett, V. L., Lawton, M. P. & Philpot, R. M. Cloning and sequencing of flavin-containing monooxygenases FMO3 and FMO4 from rabbit and characterization of FMO3. *J. Biol. Chem.* 269, 14314–14322 (1994).
- 83 Itagaki, K., Carver, G. T. & Philpot, R. M. Expression and characterization of a modified flavin-containing monooxygenase 4 from humans. *J. Biol. Chem.* 271, 20102–20107 (1996).
- 84 Balke, K., Kadow, M., Mallin, H., Saß, S. & Bornscheuer, U. T. Discovery, application and protein engineering of Baeyer-Villiger monooxygenases for organic synthesis. *Org. Biomol. Chem.* 10, 6249–6265 (2012).
- 85 Fiorentini, F. et al. Biocatalytic Characterization of Human FMO5: Unearthing Baeyer-Villiger Reactions in Humans. *ACS Chem. Biol.* 11, 1039–1048 (2016).
- 86 Fiorentini, F. et al. Baeyer-Villiger Monooxygenase FMO5 as entry point in drug metabolism. *ACS Chem. Biol.* 12, 2379–2387 (2017).
- 87 Phillips, I. R. & Shephard, E. A. Drug metabolism by flavin-containing monooxygenases of human and mouse. *Expert Opin. Drug Metab. Toxicol.* 13, 167–181 (2017).
- 88 Yanni, S. B. et al. Role of flavin-containing monooxygenase in oxidative metabolism of voriconazole by human liver microsomes. *Drug Metab. Dispos.* 36, 1119–1125 (2008).
- 89 Henderson, M. C., Siddens, L. K., Morré, J. T., Krueger, S. K. & Williams, D. E. Metabolism of the anti-tuberculosis drug ethionamide by mouse and human FMO1, FMO2 and FMO3 and mouse and human lung microsomes. *Toxicol. Appl. Pharmacol.* 233, 420–427 (2008).
- 90 Cashman, J. R. Role of flavin-containing monooxygenase in drug development. *Expert Opin. Drug Metab. Toxicol.* 4, 1507–1521 (2008).
- 91 Chaloupkova, R. et al. Light-Emitting Dehalogenases: Reconstruction of Multifunctional Biocatalysts. *ACS Catal.* 9, 4810–4823 (2019).
- 92 Arabnejad, H. et al. Computational Design of Enantiocomplementary Epoxide Hydrolases for Asymmetric Synthesis of Aliphatic and Aromatic Diols. *ChemBioChem* 21, 1893–1904 (2020).
- 93 Risso, V. A. et al. Enhancing a: De novo enzyme activity by computationally-focused ultra-low-throughput screening. *Chem. Sci.* 11, 6134–6148 (2020).
- 94 Hochberg, G. K. A. & Thornton, J. W. Reconstructing Ancient Proteins to Understand the Causes of Structure and Function. *Annu. Rev. Biophys.* 46, 247–269 (2017).
- 95 Field, S. F. & Matz, M. V. Retracing Evolution of Red Fluorescence in GFP-Like Proteins from Faviina Corals. *Mol. Biol. Evol.* 27, 225–233 (2010).
- 96 Pauling, L. & Zuckerkandl, E. Chemical Paleogenetics. *Sciences (New York)*. 5, 9–11 (1965).
- 97 Voordeckers, K. et al. Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. *PLoS Biol.* 10, (2012).
- 98 Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* 27, 1164–1165 (2011).

- 99 Nascimento, F. F., Reis, M. Dos & Yang, Z. A biologist's guide to Bayesian phylogenetic analysis. *Nat. Ecol. Evol.* 1, 1446–1454 (2017).
- 100 Ehman, E. C. et al. Renewing Felsenstein's phylogenetic bootstrap in the era of Big Data. *Nature* 556, 452–456 (2018).
- 101 Gaucher, E. A., Govindarajan, S. & Ganesh, O. K. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* 451, 704–707 (2008).
- 102 Risso, V. A., Gavira, J. A., Mejia-Carmona, D. F., Gaucher, E. A. & Sanchez-Ruiz, J. M. Hyperstability and substrate promiscuity in laboratory resurrections of precambrian  $\beta$ -lactamases. *J. Am. Chem. Soc.* 135, 2899–2902 (2013).
- 103 Wheeler, L. C., Lim, S. A., Marqusee, S. & Harms, M. J. The thermostability and specificity of ancient proteins. *Curr. Opin. Struct. Biol.* 38, 37–43 (2016).
- 104 Gumulya, Y. & Gillam, E. M. J. Exploring the past and the future of protein evolution with ancestral sequence reconstruction: The 'retro' approach to protein engineering. *Biochem. J.* 474, 1–19 (2017).
- 105 Tokuriki, N. et al. Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nat. Commun.* 3, (2012).
- 106 Siddiq, M. A., Hochberg, G. K. & Thornton, J. W. Evolution of protein specificity: insights from ancestral protein reconstruction. *Curr. Opin. Struct. Biol.* 47, 113–122 (2017).
- 107 Baker, C. R., Hanson-Smith, V. & Johnson, A. D. Following gene duplication, paralog interference constrains transcriptional circuit evolution. *Science* (80-.). 342, 104–108 (2013).
- 108 Clifton, B. E. & Jackson, C. J. Ancestral Protein Reconstruction Yields Insights into Adaptive Evolution of Binding Specificity in Solute-Binding Proteins. *Cell Chem. Biol.* 23, 236–245 (2016).
- 109 Devamani, T. et al. Catalytic Promiscuity of Ancestral Esterases and Hydroxynitrile Lyases. *J. Am. Chem. Soc.* 138, 1046–1056 (2016).
- 110 Boucher, J. I., Jacobowitz, J. R., Beckett, B. C., Classen, S. & Theobald, D. L. An atomic-resolution view of neofunctionalization in the evolution of apicomplexan lactate dehydrogenases. *Elife* 2014, (2014).
- 111 Anderson, D. P. et al. Evolution of an ancient protein function involved in organized multicellularity in animals. *Elife* 5, (2016).
- 112 Clifton, B. E. et al. Evolution of cyclohexadienyl dehydratase from an ancestral solute-binding protein. *Nat. Chem. Biol.* 14, 542–547 (2018).
- 113 Pillai, A. S. et al. Origin of complexity in haemoglobin evolution. *Nature* 581, 480–485 (2020).
- 114 Hochberg, G. K. A. et al. A hydrophobic ratchet entrenches molecular complexes. *Nature* 1–6 (2020) doi:10.1038/s41586-020-3021-2.
- 115 Riebel, A. *Perspectives in Yellow: studies on Flavoprotein Monooxygenases.* (University of Groningen, 2013).





## CHAPTER II

# Characterization of a thermostable flavin-containing monooxygenase from *Nitriicola lakisaponensis* (NiFMO)

Nikola Lončar  
Filippo Fiorentini  
Gautier Bailleul  
Simone Savino  
Elvira Romero  
Andrea Mattevi  
Marco W. Fraaije

This chapter is based on a published article: *Appl. Microbiol. Biotechnol.* **103**, 1755–1764 (2019).

**Abstract**

The flavin-containing monooxygenases (FMOs) play an important role in drug metabolism but they also have a high potential in industrial biotransformations. Among the hitherto characterized FMOs there was no thermostable representative, while such biocatalyst would be valuable for FMO-based applications. Through a targeted genome mining approach, we have identified a gene encoding for a putative FMO from *Nitrocola lacisaponensis*, an alkaliphilic extremophile bacterium. Herein, we report the biochemical and structural characterization of this newly discovered bacterial FMO (NiFMO). NiFMO can be expressed as active and soluble enzyme at high level in *Escherichia coli* (90-100 mg/L of culture). NiFMO is relatively thermostable (melting temperature ( $T_m$ ) of 51 °C), displays high organic solvent tolerance, and accepts a broad range of substrates. The crystal structure of NiFMO was solved at 1.8 Å resolution, which allows future structure-based enzyme engineering. Altogether, NiFMO represents an interesting newly discovered enzyme with the appropriate features to develop into an industrially applied biocatalyst.

## Introduction

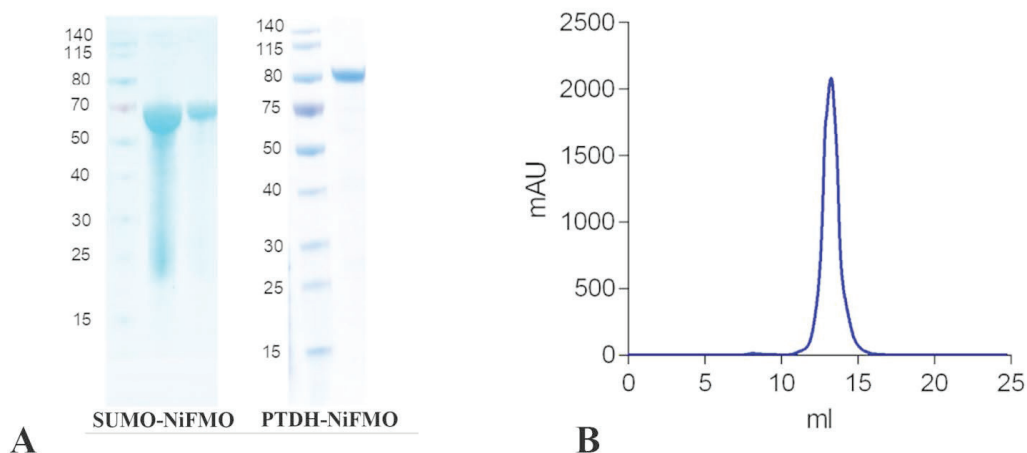
Mammalian FMOs are well-known monotopic membrane enzymes involved in the oxidation of xenobiotics and have been extensively studied for their contribution in phase-1 metabolic processes<sup>1-3</sup>. The original interest for the roles of these oxidative flavoenzymes in drug metabolism has been quickly matched by an equally strong appeal for their potential in biotransformations<sup>4,5</sup>. Besides the exploitation of soluble microbial homologs as mimics of mammalian FMOs for the enantioselective preparation of drug metabolites<sup>6</sup>, bacterial variants have also shown promise for the bioproduction of indigo and indigo derivatives, which can be very valuable both as dyes and as precursors of pharmaceuticals<sup>4,5,7,8</sup>. A designed metabolic pathway which involves a bacterial FMO for the hydroxylation of indole has been recently developed and allows a new dyeing strategy. Such FMO-based biotechnological process would offer a green alternative to the classic and highly-polluting indigo dyeing process<sup>8</sup>.

Although microbial FMOs are generally expressed in good amounts in convenient heterologous systems such as *Escherichia coli*, all characterized FMOs have not shown a very high thermal stability, a key requisite for industrial applications<sup>9,10</sup>. For this reason, we have been seeking for more stable FMOs through screening of the sequenced microbial genomes. Our search led us to identify a gene encoding for a putative FMO present in *Nitricola lacisaponensis*, an alkaliphilic bacterium isolated from the Soap Lake, an alkaline (pH ~ 9.8) and saline (NaCl ~ 10% w/v) lake located in Grant County, USA<sup>11</sup>. Here we report the biochemical and structural characterization of this novel bacterial FMO (NiFMO), which was found to display high thermostability, organic solvent tolerance, and a broad substrate profile. The robustness of this newly discovered FMO makes it an attractive biocatalyst for selective oxygenation reactions.

## Results

### *Substrate acceptance profile and thermostability of NiFMO*

NiFMO could be expressed in excellent amounts in *E. coli*. Using affinity chromatography, the enzymes could be isolated in high yields: 90-100 mg of purified protein per L of culture broth (Figure 1A). The His6-PTDH-NiFMO was used for biochemical and biocatalytic exploration while the His6-SUMO-NiFMO was used for crystallization purposes after cleavage of the tag. From gel permeation analysis the wild-type NiFMO proved to behave as a functional dimer (2 x 53 kDa), containing one molecule of non-covalently tightly bound FAD per monomer (Figure 1). This is similar to FMO from *Methylophaga* sp.<sup>9</sup>.




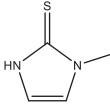

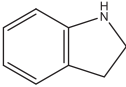
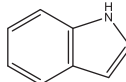
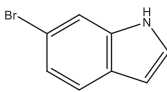
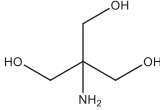
**Figure 1. Purification of NiFMO.** On the left, a SDS-PAGE gel shows purified samples of His<sub>6</sub>-SUMO-tagged and His<sub>6</sub>-PTDH-tagged NiFMO. On the right, gel-filtration analysis of purified tag-free NiFMO is shown. The wild-type enzyme elutes as a monodispersed peak corresponding to a MW of ~ 100 kDa, suggesting that NiFMO exists as a dimer in solution.

To establish whether NiFMO can be used for selective oxygenation, several potential substrates were tested. In the first experiments, it was found that already without adding a potential substrate an appreciable rate of NADPH oxidation was observed ( $\sim 0.1 \text{ s}^{-1}$ ). The rate for uncoupling is normally lower for FMOs or related class B flavoprotein monooxygenases<sup>21</sup>. Soon it was realized that NiFMO also accepts the used buffer, tris(hydroxymethyl)aminomethane, as a substrate. By changing to a phosphate buffer, the rate of uncoupling (NADPH oxidation in the absence of any substrate) could be determined:  $0.015 \text{ s}^{-1}$ .

Next, the steady-state kinetic parameters were determined for several substrates (Table 2, Figure S1). While tris(hydroxymethyl)aminomethane was a poor substrate in terms of kinetic parameters when compared with the other compounds, the  $K_M$  value of 21.5 mM turned out to fall in the range of the concentrations initially used in the experiments. This fully explains the interference of tris(hydroxymethyl)aminomethane when measuring activity of other test compounds.

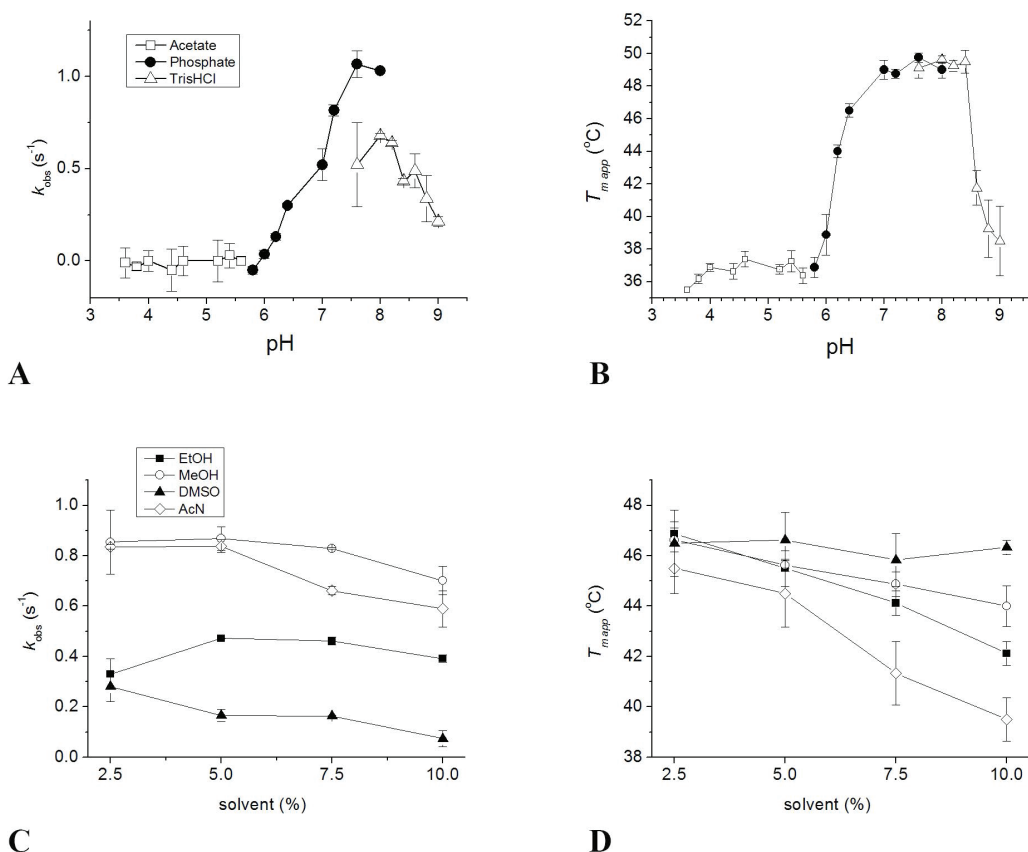
The substrates display rather similar  $k_{\text{cat}}$  values ( $0.09 - 2.01 \text{ s}^{-1}$ ) while the  $K_M$  values vary somewhat more ( $45.6 - 21,500 \text{ }\mu\text{M}$ ). Trimethylamine and methimazole are the best substrates when considering the kinetic parameters. Yet, also more bulky compounds, such as indole, 6-bromoindole, and indoline are accepted by the enzyme. In particular, indigo conversion could be easily observed by the appearance of a blue color indicating the formation of indigo blue. All above mentioned reactions were measured using  $100 \text{ }\mu\text{M}$  NADPH. NADH was not accepted by the enzyme as coenzyme. Using  $3.0 \text{ mM}$  trimethylamine as saturating substrate, the  $K_M$  for NADPH was determined:  $8.2 \text{ }\mu\text{M}$ .

Table 2. Steady-state kinetics ofPTDH-NiFMO

		$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$K_{\text{M}}$ ( $\mu\text{M}$ )	$k_{\text{cat}}/K_{\text{M}}$ ( $\text{s}^{-1} \text{M}^{-1}$ )
Trimethylamine		$2.01 \pm 0.08$	$45.6 \pm 9.6$	44,000
Methimazole		$1.10 \pm 0.08$	$77 \pm 10$	14,000
Pyrrole		$0.69 \pm 0.03$	$146 \pm 17$	4,800
Indoline		$0.70 \pm 0.03$	$98 \pm 14$	7,100
Indole		$0.11 \pm 0.01$	$137 \pm 23$	730
6-Bromoindole		$0.09 \pm 0.03$	$640 \pm 340$	140
Tris(hydroxymethyl)aminomethane (TRIS)		$0.18 \pm 0.02$	$21500 \pm 5600$	9

The activity and stability of the enzyme were evaluated at different pH values and co-solvent conditions both by means of checking the NADPH depletion rates in the presence of trimethylamine, and by detecting the apparent melting temperature using the ThermoFAD method<sup>13</sup>. The effects of pH and co-solvents on NiFMO activity and the apparent melting temperature are shown in Figure 2. The optimal pH for activity was found to be slightly basic, between pH 7 and 8. Also the thermostability (highest  $T_m$  of 51 °C) was highest at slightly alkaline conditions. The solvent tolerance of NiFMO proved to be rather good as the enzyme retains activity even in the presence of up to 10% solvent content. Specifically, high conversion rates were observed with methanol and acetonitrile as co-solvents. However, acetonitrile had a detrimental effect on the  $T_m$  of NiFMO, whereas the enzyme showed high thermostability ( $T_m \sim 47$  °C) in the presence of up to 10% (v/v) DMSO.



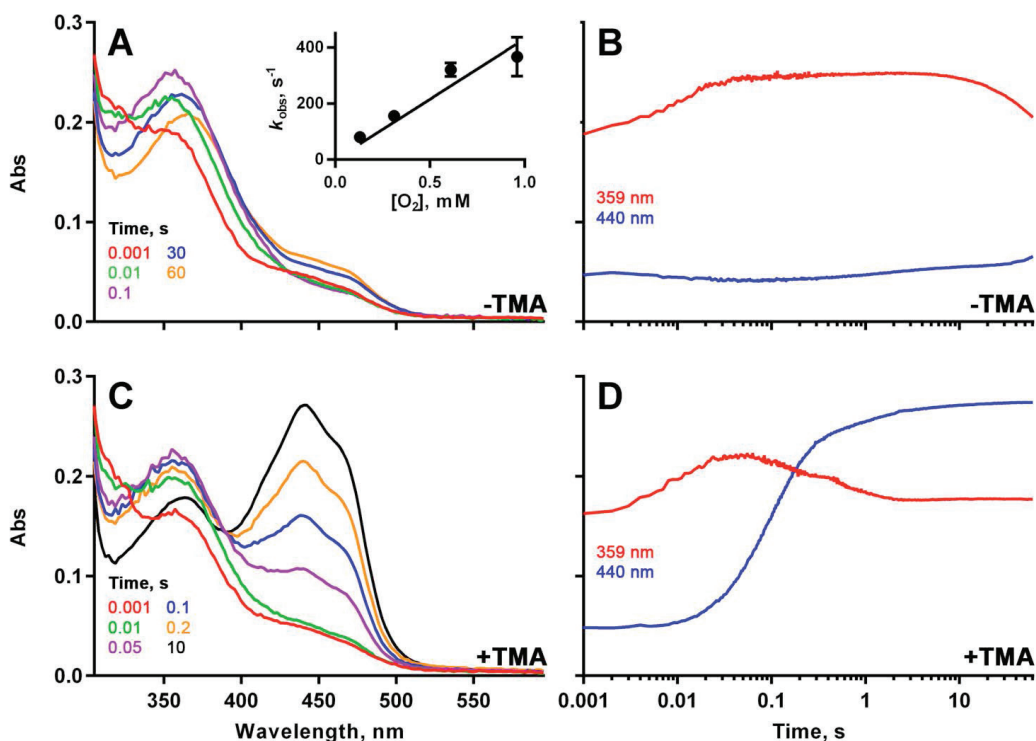


**Figure 2. Stability and co-solvent tolerance of NiFMO.** Effect of the pH (A,B) and co-solvents (C,D) on NiFMO activity on trimethylamine (measured by NADPH depletion) and on the apparent melting temperature ( $T_{m, \text{app}}$ ) of NiFMO (measured by ThermoFAD technique). EtOH is ethanol, MeOH is methanol, and AcN is acetonitrile.

### Rapid kinetics assays reveal formation and stabilization of the C4 $\alpha$ -hydroperoxyflavin intermediate

After identification of a set of compounds oxygenated by NiFMO, we sought to investigate the oxidative half-reaction in the presence and absence of a suitable substrate, trimethylamine, by using the stopped-flow technique. We initially monitored the spectral changes occurring after mixing the fully NADPH-reduced NiFMO with various dioxygen concentrations, at pH 7.5 and 25 °C. In the absence of substrates, the fully reduced NiFMO reacted very rapidly with dioxygen to form a C4 $\alpha$ hydroperoxyflavin intermediate with an absorption maximum of 359 nm (80 s<sup>-1</sup>, at 0.13 mM dioxygen) (Figure 3A, B). The observed rates for these reactions correlated linearly with the dioxygen concentration, giving a second-order rate constant for the reaction with dioxygen of  $0.4 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$

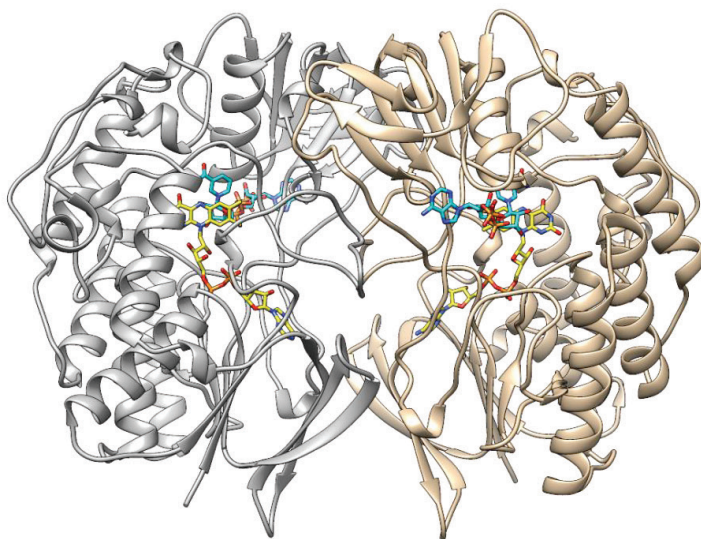
(Figure 3A, inset). The intermediate slowly converted into fully oxidized NiFMO in the absence of trimethylamine ( $0.002 - 0.02 \text{ s}^{-1}$ ). Recovery of only 28% of the absorbance expected for the fully re-oxidized enzyme was observed after 60 s (Figure 3A). This slow recovery of the oxidized enzyme is fully in line with the observed slow uncoupling rate of  $0.015 \text{ s}^{-1}$ . We then tested the effect of trimethylamine on the enzyme re-oxidation in the presence of  $0.13 \text{ mM}$  dioxygen as co-substrate. In the presence of trimethylamine, the rate of formation of the intermediate was fast ( $100 \text{ s}^{-1}$ ). Enzyme re-oxidation was complete within 10 s when trimethylamine is included in the assays (Figure 3C, D). The stopped-flow traces at  $440 \text{ nm}$  for enzyme re-oxidation were best fit to a double exponential function, with an initial fast phase accounting for 87% of the total absorption change ( $4\text{-}9$  and  $6 \text{ s}^{-1}$ ).



**Figure 3.** Kinetics of the oxidative half-reaction of NiFMO. Oxidative half-reaction of NiFMO in the absence (A, B) and the presence (C, D) of trimethylamine. Stopped-flow spectra and traces were recorded after mixing NADPH-reduced enzyme with a solution containing  $0.13 \text{ mM}$  dioxygen. To obtain the plot showed in the inset,  $k_{\text{obs}}$  values at various dioxygen concentrations were calculated from exponential fits of the traces at  $359 \text{ nm}$ .

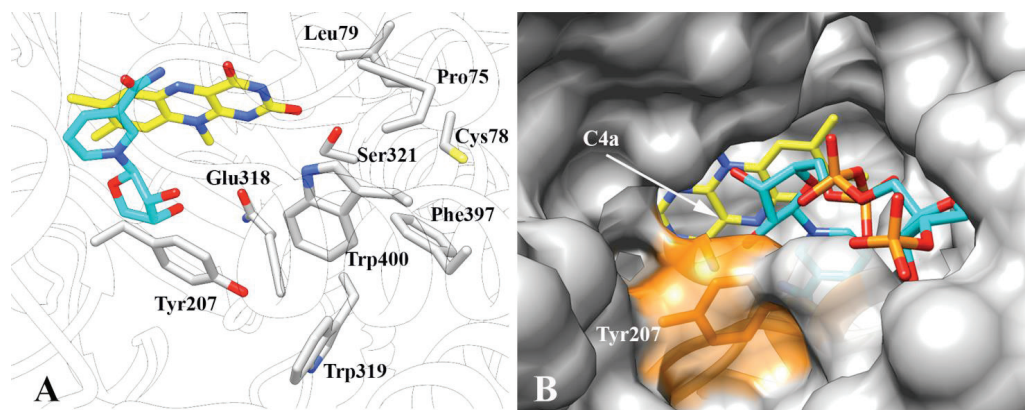
### Crystal structure of NiFMO

To correlate the biochemical profile of this novel bacterial FMO with the structural features, the tag-free purified protein was used for crystallization trials. Crystals were obtained only following incubation of the protein with 1.0 mM NADP<sup>+</sup>. The NiFMO structure was solved at 1.8 Å resolution by molecular replacement, resulting in an electron density of excellent quality (Figure 4 and Table 1). In terms of overall structure, NiFMO is very similar to FMO from *Methylophaga* sp. strain SK1<sup>9</sup>, featuring two distinct domains for the binding of the nucleotide cofactors, FAD and NADP<sup>+</sup>, connected through a double linker.



*Figure 4. Ribbon diagram of the NiFMO crystallographic dimer. Monomers are in gray and light brown. FAD and NADP<sup>+</sup> are depicted with yellow and cyan carbons, respectively. Oxygen atoms are in red, nitrogen atoms in blue and phosphorous atoms in orange.*

The sequence identity between the two proteins is 74% with a root-mean-square deviation of 0.4 Å between the Ca atoms of the two proteins. The identity and the position of the catalytic residues shaping the active site are fully conserved, in agreement with the very similar substrate profiles of these bacterial FMOs. The oxygenation of indole and indole derivatives – a prerogative of bacterial FMO variants – was shown to correlate with the presence of a highly conserved tyrosine (Tyr207 in NiFMO) which positions as a front-door leading suitable compounds to the oxygenating FAD intermediate. Its substitution to asparagine in mammalian FMOs results in a drastically decreased oxygenation of indigoid compounds<sup>22</sup>. The Tyr side chain may favor binding and correct positioning of the indigo-like substrates through favorable aromatic stacking interactions with the indigo aromatic ring (Figure 5).



**Figure 5.** The active site of NiFMO. The left panel shows the active site side chains (carbon atoms in gray) together with NADP<sup>+</sup> (carbon atoms in cyan) and FAD (carbon atoms in yellow). The right panel shows surface representation of NiFMO with a view of the catalytic site from the outside. NiFMO displays an open conformation which has the potential to expose the oxygenating flavin-C4a-OO(H) (the “flavin-(hydro)peroxide”) intermediate to any suitable substrate gaining access to the active site. Tyr207 – highly conserved among bacterial FMOs characterized so far and substituted with an Asn in mammalian FMOs – was hinted to correlate with the capability to accept indole as substrate. FAD is shown as yellow sticks and NADP<sup>+</sup> as blue sticks. The surface corresponding to Tyr207 is highlighted in orange. The oxygen reactive C4a-atom of FAD is labeled.

## Discussion

The bacterial FMOs characterized so far displayed substrate profiles very similar to the ones reported for human FMOs, which - with the exception of FMO5 - are known to catalyze the oxygenation of different compounds bearing soft-nucleophilic heteroatoms, mainly N- and S- atoms<sup>2,23</sup>. The solubility and generally higher stability of the bacterial FMOs has fostered the utilization of the bacterial homologs for screening and/or synthesis of human metabolites and drug molecules<sup>6</sup>. Along this line and considering the potential value of FMO-based processes for the production of indigoid pigments<sup>8</sup>, the compounds for substrate scope exploration were selected based on the following considerations: trimethylamine is a common endogenous substrate for human FMOs<sup>2</sup>; pyrrole is a probe to test NiFMO activity on heterocyclic organic compounds; methimazole is of pharmacological relevance; indoline, indole, and 6-bromoindole are of significance for the bio-based synthesis of indigoid pigments. All these compounds proved to be substrates of NiFMO, with trimethylamine exhibiting the highest catalytic efficiency (Table 2). Unexpectedly, enzymatic activity was also observed on tris(hydroxymethyl) aminomethane, initially used as buffer for the enzymatic assays. This shows that organic buffer components should be used with care when working with enzymes that display

a relaxed substrate specificity. Collectively, these data confirm that NiFMO accepts a broad range of N-containing molecules, featuring a substrate profile similar to those of mammalian FMOs 1-4.

The stopped-flow experiments showed that NiFMO has the typical mechanistic features of the flavin-containing monooxygenases and Baeyer-Villiger monooxygenases: efficient reduction by NADPH and fast reaction of the reduced flavin with dioxygen to form a stable C4 $\alpha$ -hydroperoxyflavin intermediate. This species efficiently reacts with a substrate to form the oxygenated product<sup>24</sup>. In the absence of a suitable substrate, the flavin-peroxide slowly decays to form hydrogen peroxide, resulting in a so-called uncoupling activity. The uncoupling for NiFMO is rather slow (0.015 s<sup>-1</sup>) preventing the nonproductive usage of NADPH.

As observed for *Methylophaga* FMO, the three-dimensional structure of NiFMO shows a ready-to-react conformation which can expose the reactive FAD-OOH to any suitable nucleophile gaining access to the active site (Figure 5)<sup>25</sup>. This reveals how FMOs may exist in the cells in an active form (“cocked gun”) waiting for a suitable substrate with which to react. Substrate selection/accessibility into the active site seems to be mostly guided by the rigid conformation of the crevice leading to it.

From a structure/function point of view, a most relevant observation is that NiFMO features a  $T_m$  value (51 °C), which is 8 °C higher than the  $T_m$  of *Methylophaga* FMO ( $T_m$  43.3 °C)<sup>13</sup>. We have comparatively analyzed the crystal structures of NiFMO and *Methylophaga* FMO searching for hints about factors that can enhance thermal stability. The dimeric arrangement of the two proteins is identical in terms of both extension of the dimer interface (about 10% of the monomer accessible surface buried upon dimer formation) and number of intermolecular hydrogen bonds (22 and 24 in NiFMO and *Methylophaga* FMO, respectively). A more significant difference becomes instead apparent from the inventory of the intra-subunit salt-bridges which are significantly more numerous in NiFMO (82) than in *Methylophaga* FMO (72). The abundant electrostatic interactions might contribute to thermal stability as previously observed in another flavoprotein monooxygenase<sup>26,27</sup>.

In conclusion, here we report the discovery and characterization of the first thermostable FMO. In view of the ability of this bacterial FMO to oxygenate indigoid compounds as well as a wide variety of substrates converted by typical human FMOs, we believe that NiFMO may take part in bio-based processes targeting both the production of dyes and the preparation of drug metabolites.

## Materials and methods

### *Chemicals and reagents*

Bovine liver catalase, indigo, indole and indole derivatives were purchased from Sigma-Aldrich. All other chemicals were analytical grade and obtained from Sigma-Aldrich or Merck. NADPH and NADP<sup>+</sup> were purchased from Oriental Yeast Co. LTD.

### **Strains, plasmids and growth conditions**

A codon-optimized gene (GenBank accession number MK061530) encoding NiFMO (GenBank: WP\_036542404.1) for *E. coli* expression was synthesized by GenScript. Using restriction cloning, the gene was cloned into a pBAD vector resulting in two constructs for expressing NiFMO with an N-terminal fusion partner: (a) NiFMO fused to His6-SUMO (SUMO = small ubiquitin-like modifier), and (b) NiFMO fused to His-tagged phosphite dehydrogenase (His6-PTDH). Primers and plasmid maps are available upon request. Plasmid sequences were verified by sequencing (Eurofins Genomics). Recombinant His6-SUMO-NiFMO and His6-PTDH-NiFMO were overexpressed and purified according to the following procedure: *E. coli* NEB10 $\beta$  cells were freshly transformed with the corresponding plasmid and an overnight culture was diluted 100-fold into 400 mL Terrific Broth medium with 50  $\mu\text{g}/\text{mL}$  ampicillin (TB<sub>amp</sub>) in 2 L baffled flasks. Cells were induced at OD<sub>600</sub> = 1.5 with 0.02% w/v L-arabinose (final concentration) and incubated at 25 °C for 36 h (135 rpm). Cells were harvested at 4 °C and centrifuged at 6000 rpm using JA10.500 rotor for 20 min in the Beckman-Coulter centrifuge. They were then washed and resuspended in 50 mM potassium phosphate (KPi) buffer pH 7.5 containing 0.25 M NaCl and 0.1 mM phenylmethylsulfonyl fluoride. Resuspended cells were disrupted by sonication and centrifuged at 4 °C 16000 rpm using JA17 rotor for 45 min. The cell-free extract was applied on a 5 mL HiTrap Ni-Sepharose HP column pre-equilibrated in 50 mM KPi buffer pH 7.5 with 0.25 M NaCl. Stepwise elution was used to wash non-specifically bound proteins and elution of NiFMO was achieved with 50 mM KPi buffer pH 8.0, 0.25 M NaCl, and 500 mM imidazole. After SDS-PAGE analysis, pure fractions were pooled, buffer-exchanged to 50 mM KPi pH 7.5, and flash frozen in liquid nitrogen. Samples were stored at -20 °C until further use. Concentration of purified NiFMO fusions was determined by using an extinction coefficient of 15.0 mM<sup>-1</sup> cm<sup>-1</sup> at 440 nm. Extinction coefficient was determined as described previously<sup>12</sup>.

### **Thermal stability assays**

The ThermoFAD method<sup>13</sup> was used to determine the apparent melting temperature of His6-PTDH-NiFMO in different pH conditions or in the presence of additives. By means of an RT-PCR machine (CFX96-Touch, Bio-Rad), the fluorescence of the FAD cofactor was monitored using a 450–490 excitation filter and a 515–530 nm emission filter, typically used for SYBR Green based RT-PCR. The temperature was increased with 0.5 °C per step, starting at 25 °C and ending at 90 °C, using a holding time of 10 s at each step. The maximum of the first derivative of the observed flavin fluorescence was taken as the apparent melting temperature. 20  $\mu\text{L}$  of the enzyme solution containing 10  $\mu\text{M}$  His<sub>6</sub>-PTDH-NiFMO and the additive (solvent or buffer of appropriate pH) was put into a 96 wells PCR plate and covered with a transparent adhesive film.

### ***Steady-state kinetic analysis***

Kinetic assays were performed in duplicates for 120 s using a Jasco V-660 spectrophotometer and H1MD microplate reader (BioTek). Different concentrations of substrates (0.05 - 2.50 mM) were dissolved in 50 mM KPi pH 7.5. Stock solutions of indole derivatives (indoline, pyrrole, and 6-Br-indole) were prepared in methanol. The final concentration of methanol in the test reaction was kept below 1% (v/v). His<sub>6</sub>-PTDH-NiFMO (final concentration 0.5  $\mu$ M) was added to the reaction mixture and the reaction was started by the addition of 0.10 mM NADPH (final concentration). NADPH consumption was followed at 340 nm.

### ***pH optimum and activity in presence of co-solvents***

Determination of the pH optimum of His<sub>6</sub>-PTDH-NiFMO activity was performed by using the above-described procedures with a series of 50 mM buffers (sodium-acetate pH 4.0-5.6, KPi pH 5.8-8.0, tris(hydroxymethyl)aminomethane-HCl pH 7.5-9.0). Final concentrations in the reaction mixtures were 1.0 mM trimethylamine, 1.0  $\mu$ M His<sub>6</sub>-PTDH-NiFMO, and 0.10 mM NADPH.

### ***Rapid kinetics analysis***

The stopped-flow experiments were carried out using NiFMO fused to His<sub>6</sub>-SUMO. The reaction of NiFMO with dioxygen was studied using the single-mixing mode of a SX20 stopped-flow spectrophotometer equipped with a photodiode array detector (Applied Photophysics, Surrey, UK). All solutions were prepared in 50 mM KPi pH 7.5. Reactions were run in duplicate by mixing equal volumes of two solutions, at 25 °C. The stopped-flow instrument was made anaerobic by flushing the flow-circuit with a dioxygen-scrubbing solution containing 5.0 mM glucose and 0.3  $\mu$ M glucose oxidase (*Aspergillus niger*, type VII, Sigma-Aldrich). To prepare fully reduced NiFMO, titrations of oxidized NiFMO with NADPH were carried out in a vial under anaerobic conditions. After adding a small aliquot (<10  $\mu$ L) of NADPH (>40 mM) to the enzyme (15-30  $\mu$ M), the mixture was incubated at room temperature until no further enzyme reduction was observed. The titration was then continued until NiFMO was completely reduced as evidenced by the bleaching of the flavin. The resulting reduced enzyme was loaded into the stopped-flow instrument and mixed with buffer containing various dioxygen concentrations to monitor the spectral changes taking place in the stopped-flow cell. To achieve the desired final dioxygen concentration (0.13, 0.31, 0.61, and 0.96 mM), the anaerobic enzyme solution was mixed with: i) air-saturated buffer; ii) solution containing equal amounts of buffer bubbled with 100% nitrogen and buffer bubbled with 100% dioxygen; iii) buffer bubbled with 100% dioxygen; and iv) buffer bubbled with 100% dioxygen in ice. The bubbling time for all solutions was 10 min. The lowest dioxygen concentration was assayed in the presence and the absence of trimethylamine (150  $\mu$ M final concentration). The stopped-flow traces at 359 and 440 nm were fit to exponential

functions to determine the observed rates ( $k_{\text{obs}}$ ). The second-order rate constant for the reaction of the reduced flavin with dioxygen was calculated from the slope of the linear plot of  $k_{\text{obs}}$  versus dioxygen concentration. All data were analyzed using the software Pro-Data (Applied Photophysics, Surrey, UK) or GraphPad Prism 6.05 (La Jolla, CA, USA).

### ***Crystallization of NiFMO***

After His<sub>6</sub>-SUMO tag cleavage using SUMO protease and elution of the unbound native NiFMO from a second His-trap column, the protein was further purified and the oligomeric state was analyzed by size-exclusion chromatography (Superdex 200 10/300, GE Healthcare). The eluted peak was concentrated up to 12 mg/mL and used for crystallization trials (final buffer composition: 25 mM tris(hydroxymethyl)aminomethane-HCl pH 8.0, 200 mM NaCl). NADP<sup>+</sup> (1.0 mM final concentration) was added to the protein solution immediately before crystallization experiments. The crystallization screening was performed by means of an Oryx 8 crystallization robot (Douglas Instruments). NiFMO crystallized in 0.20 M Mg-formate and 20% (w/v) PEG3350 at 4 °C in microbatch configuration.

### ***X-ray methods***

The crystal structure of NiFMO was solved using standard methods. Diffraction data measured at 100 K on the beam line PX-III of the Swiss Light Source (Villigen, CH). Data were processed with XDS and programs of the CCP4 suite<sup>14,15</sup>. The structure was solved by molecular replacement using the PDB entry 2VQ7<sup>9</sup> as search model and the program Phaser<sup>16</sup>. The structure was refined at 1.8 Å resolution using Refmac5 and Coot<sup>17,18</sup>. Crystallographic statistics are listed in Table 1. Model analysis was done with Coot<sup>18</sup>, PISA<sup>19</sup>, and ESBRI<sup>20</sup>.



**Table 1. Crystallographic table for the structure of the wild-type NiFMO in complex with NADP<sup>+</sup>[a]**

[a] Statistics for the highest-resolution shell are shown in parentheses.

[b] A cut-off criterion for resolution limits was applied on the basis of the mean intensity correlation coefficient of half-subsets of each dataset.

PDB code	6HNS
Resolution range	50 – 1.84
Space group	P2 <sub>1</sub> 22 <sub>1</sub>
Unit cell (Å)	59.06, 125.61, 144.72
Total reflections	413067 (19305)
Unique reflections	93414 (4431)
Multiplicity	4.4 (4.4)
Completeness (%)	99.2 (97.1)
Mean I/sigma (I)	14.7 (1.6)
R-merge (%)	0.076 (0.849)
CC1/2 <sup>[b]</sup>	0.99 (0.58)
R-work (%)	0.191
R-free (%)	0.222
Number of non-hydrogen atoms	
Protein	7227
Ligands (FAD/NADP <sup>+</sup> /Mg <sup>++</sup> )	204
Waters	465
RMS (bonds) (Å)	0.011
RMS (angles) (°)	1.61
Ramachandran favored (%)	96
Ramachandran allowed (%)	4
Ramachandran outliers (%)	0
Average B-factor (Å <sup>2</sup> )	26.0

## **Author Contributions**

Nikola Lončar and Filippo Fiorentini contributed equally to this work. All listed authors performed experiments and/or analyzed data. FF and SS crystallized the enzyme, collected the corresponding datasets at ESRF and SLS facilities, performed structural analysis and elucidated the structure. NL cloned the NiFMO genes into the vectors. GB carried out kinetic analysis and validated the substrate profiles. ER conducted rapid kinetics. NL, FF, GB, SS and ER prepared the figures. NL and FF wrote the manuscript, AM, MWF edited it. All authors provided critical feedback and helped shape the research, analysis and manuscript.

## **Funding and acknowledgements**

Nikola Lončar was supported by the NWO-LIFT project Indigreen. Gautier Bailleul received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant Oxytrain, agreement No 722390. The financial support of the Fondazione Cariplo (grant 2015-0406) is kindly acknowledged. We thank Callum R. Nicoll for his help in the project.

## **Compliance with ethical standards**

This article does not contain any studies with human participants or animals performed by any of the authors.

## **Declaration of interest**

The authors declare that they have no conflicts of interest with the contents of this article.

## References

- 1 Cashman, J. R. & Zhang, J. Human flavin-containing monooxygenases. *Annu. Rev. Pharmacol. Toxicol.* 46, 65–100 (2006).
- 2 Krueger, S. K. & Williams, D. E. Mammalian flavin-containing monooxygenases: Structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacol. Ther.* 106, 357–387 (2005).
- 3 Phillips, I. R. & Shephard, E. A. Drug metabolism by flavin-containing monooxygenases of human and mouse. *Expert Opin. Drug Metab. Toxicol.* 13, 167–181 (2017).
- 4 Choi, H. S. et al. A novel flavin-containing monooxygenase from *Methylophaga* sp. strain SK1 and its indigo synthesis in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 306, 930–936 (2003).
- 5 Han, G. H. et al. Bio-indigo production in two different fermentation systems using recombinant *Escherichia coli* cells harboring a flavin-containing monooxygenase gene (*fmo*). *Process Biochem.* 46, 788–791 (2011).
- 6 Gul, T., Krzek, M., Permentier, H. P., Fraaije, M. W. & Bischoff, R. Microbial flavoprotein monooxygenases as mimics of mammalian flavin-containing monooxygenases for the enantioselective preparation of drug metabolites. *Drug Metab. Dispos.* 44, 1270–1276 (2016).
- 7 Ameria, S. P. L. et al. Characterization of a flavin-containing monooxygenase from *Corynebacterium glutamicum* and its application to production of indigo and indirubin. *Biotechnol. Lett.* 37, 1637–1644 (2015).
- 8 Hsu, T. M. et al. Employing a biochemical protecting group for a sustainable indigo dyeing strategy. *Nat. Chem. Biol.* 14, 256–261 (2018).
- 9 Alfieri, A., Malito, E., Orru, R., Fraaije, M. W. & Mattevi, A. Revealing the moonlighting role of NADP in the structure of a flavin-containing monooxygenase. *Proc. Natl. Acad. Sci. U. S. A.* 105, 6572–6577 (2008).
- 10 Orru, R., Torres Pazmin, D. E., Fraaije, M. W. & Mattevi, A. Joint functions of protein residues and NADP(H) in oxygen activation by flavin-containing monooxygenase. *J. Biol. Chem.* 285, 35021–35028 (2010).
- 11 Dimitriu, P. A. et al. *Nitrincola laxisaponensis* gen. nov., sp. nov., a novel alkaliphilic bacterium isolated from an alkaline, saline lake. *Int. J. Syst. Evol. Microbiol.* 55, 2273–2278 (2005).
- 12 Aliverti A, Curti B, V. M. Identifying and quantitating FAD and FMN in simple and in iron-sulfur-containing flavoproteins. *Methods Mol. Biol.* 9–23 (1999) doi:10.1385/1-59259-266-X:9.
- 13 Forneris, F., Orru, R., Bonivento, D., Chiarelli, L. R. & Mattevi, A. ThermofAD, a Thermofluor®-adapted flavin ad hoc detection system for protein folding and ligand binding. *FEBS J.* 276, 2833–2840 (2009).
- 14 Kabsch, W. XDS. *Acta Crystallogr Sect D Biol Crystallogr* 66, (2010).
- 15 Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67, 235–242 (2011).
- 16 McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674 (2007).
- 17 Dodson, E. J., Murshudov, G. N. & Vagin, A. A. Description of program using maximum likelihood residual for macromolecular refinement, illustrated by several examples. *Acta Crystallogr. Sect. A Found. Crystallogr.* 52, C85–C85 (1996).
- 18 Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 60, 2126–2132 (2004).

- 19 Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* 372, 774–797 (2007).
- 20 Costantini, S., Colonna, G. & Facchiano, A. M. ESBRI: A web server for evaluating salt bridges in proteins. *Bioinformatics* 3, 137–138 (2008).
- 21 van Berkel, W. J. H., Kamerbeek, N. M. & Fraaije, M. W. Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *J. Biotechnol.* 124, 670–689 (2006).
- 22 Cho, H. J. et al. Structural and functional analysis of bacterial flavin-containing monooxygenase reveals its ping-pong-type reaction mechanism. *J. Struct. Biol.* 175, 39–48 (2011).
- 23 Fiorentini, F. et al. Biocatalytic characterization of human FMO5: Eneathing Baeyer-Villiger reactions in humans. *ACS Chem. Biol.* 11, 1039–1048 (2016).
- 24 Romero, E., Gómez Castellanos, J. R., Gadda, G., Fraaije, M. W. & Mattevi, A. Same Substrate, Many Reactions: Oxygen Activation in Flavoenzymes. *Chem. Rev.* 118, 1742–1769 (2018).
- 25 Beaty, N. B. & Ballou, D. P. The oxidative half-reaction of liver microsomal FAD-containing monooxygenase\*. *J. Biol. Chem.* 256, 4619–4625 (1981).
- 26 Romero, E., Castellanos, J. R. G., Mattevi, A. & Fraaije, M. W. Characterization and Crystal Structure of a Robust Cyclohexanone Monooxygenase. *Angew. Chemie* 128, 16084–16087 (2016).
- 27 Vogt, G., Woell, S. & Argos, P. Protein thermal stability, hydrogen bonds, and ion pairs. *J. Mol. Biol.* 269, 631–643 (1997).

## Supplementary information

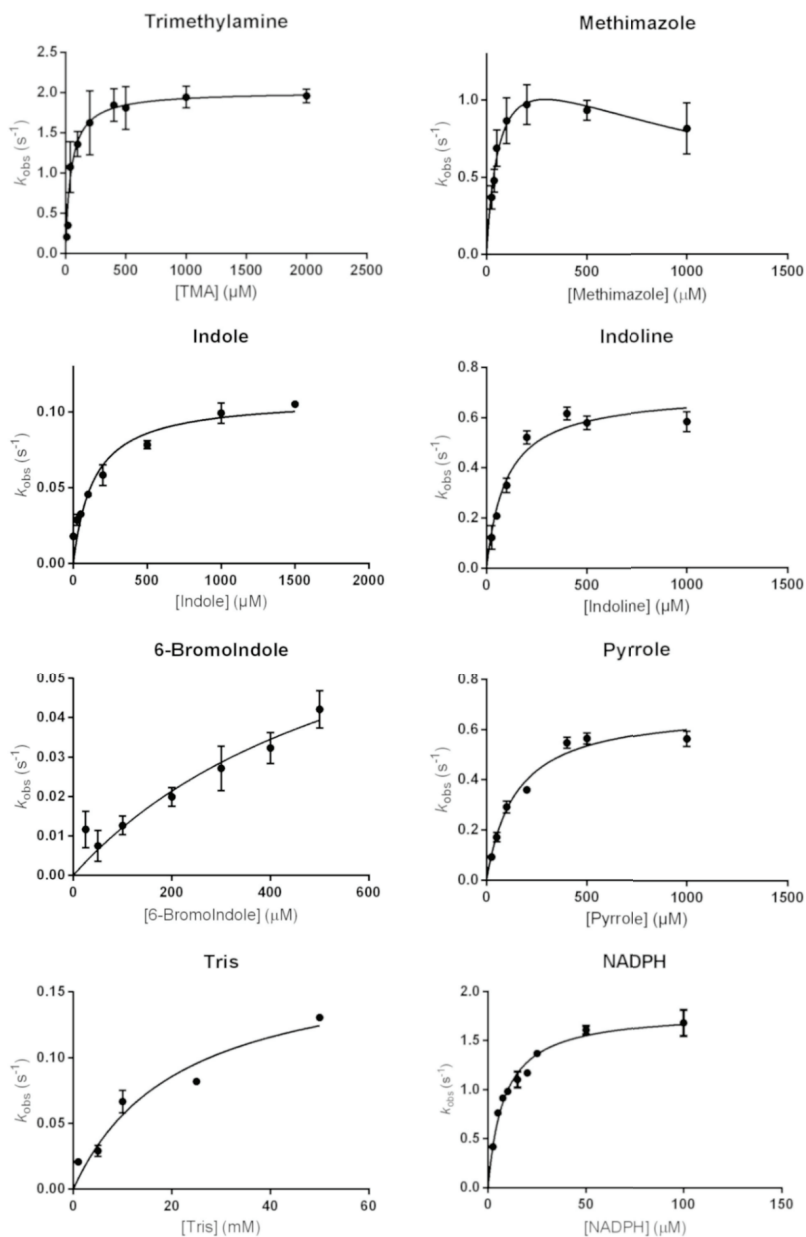
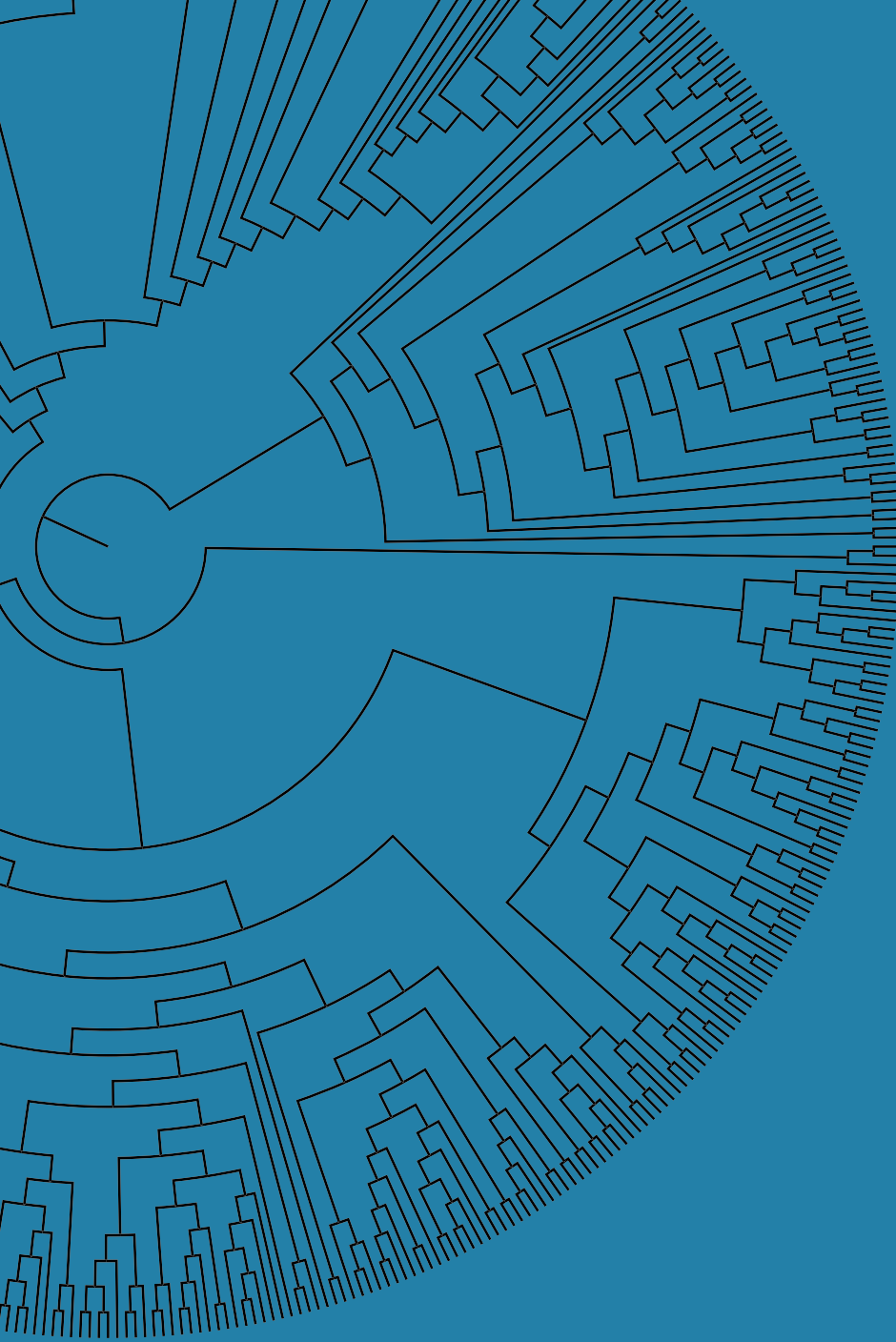


Figure S1. Enzymatic activity of NiFMO on selected substrates. The plots show the  $k_{obs}$  values as a function of substrate concentrations. The activity on each substrate shows Michaelis-Menten behavior. The corresponding apparent steady-state kinetic parameters for the amine substrates are shown in Table 2.





## CHAPTER III

# **Ancestral sequence reconstruction unveils the structural basis of catalysis and membrane binding in mammalian flavin-containing monooxygenases**

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

This chapter is based on a published article: *Nat. Struct. Mol. Biol.* 27, 14–24 (2020).



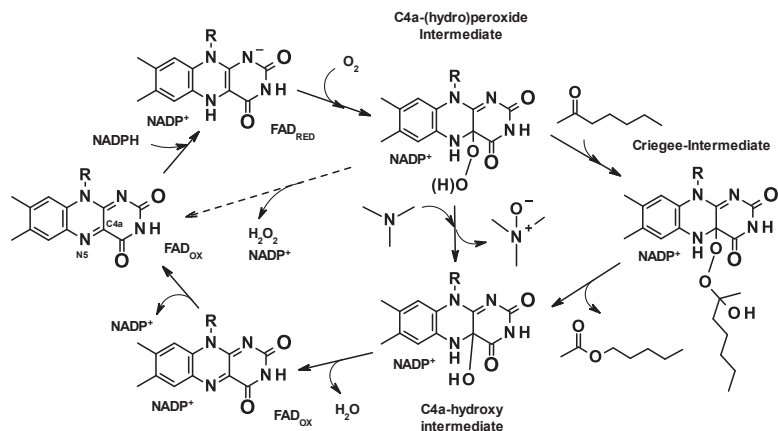
**Abstract**

Flavin-containing monooxygenases (FMOs) are ubiquitous in all domains of life and metabolize a myriad of xenobiotics including toxins, pesticides and drugs. However, despite their pharmacological significance, structural information remains bereft. To further our understanding behind their biochemistry and diversity, we scrutinized three ancient mammalian FMOs: AncFMO<sup>2</sup>, AncFMO<sup>3-6</sup> and AncFMO5, using Ancestral Sequence Reconstruction, kinetic and crystallographic techniques. Remarkably, all AncFMOs could be crystallized, and were structurally resolved between 2.7 and 3.2 Å. These crystal structures depict the unprecedented topology of mammalian FMOs. Each employs extensive membrane-binding features and intricate substrate-profiling tunnel networks through a conspicuous membrane-adhering insertion. Furthermore, a glutamate–histidine switch is speculated to induce the distinctive Baeyer-Villiger oxidation activity of FMO5. The AncFMOs exhibited catalysis akin to human FMOs and, with sequence identities between 82 and 92%, represent excellent models. Our study demonstrates the power of ancestral sequence reconstruction as a strategy for the crystallization of proteins.

## Introduction

Xenobiotic metabolism is an ancient and imperative process pursued by all organisms. With evolution resulting in the production of, and thus exposure to, a vast number of noxious and toxic natural products<sup>1</sup>, organisms have employed a multitude of intricate detoxification systems, to tackle the sheer quantity of diverse chemicals<sup>1-6</sup>. Flavin-containing monooxygenases (FMOs; EC 1.14.13.8) represent one of these detoxifying protein families and are prevalent in all domains of life<sup>4,7</sup>. FMOs are members of the Class B flavin-dependent monooxygenases and utilize the cofactors FAD and NADP(H), and dioxygen for activity<sup>8-10</sup>. Typically, FMOs pursue catalysis as illustrated in Scheme 1, whereby a soft nucleophile (in this work demonstrated with trimethylamine) receives the distal oxygen atom from the C4 $\alpha$ -(hydro)peroxyflavin intermediate<sup>11,12</sup>. The more water-soluble hydroxylated product is then released by the enzyme to be excreted from the host.

Humans possess five FMO isoforms that are differentially expressed in many different tissues such as the kidney, lung, and liver<sup>2,10,13,14</sup>. The human *FMO* genes are found on chromosome 1, with *FMO1-4* clustering over 220 kb, and *FMO5* found on a separate chromosome region<sup>15,16</sup>. The human FMO family contains six non-expressed pseudogenes which are also located on chromosome 1.16 FMOs are involved in phase I of xenobiotic detoxification<sup>2,3</sup>. They oxidize an array of compounds bearing soft nucleophilic centers such as nitrogen and sulfur atoms,<sup>2,17-19</sup> making them clinically important regarding drug metabolism<sup>3,6,12,15,17,20,21</sup>. The most extensively characterized FMO is human FMO3, renowned for its production of trimethylamine N-oxide.<sup>22-26</sup> FMO3 deactivation upon mutation induces trimethylaminuria (“fish odor syndrome”), whereby the body has an unpleasant smell due to the accumulation of trimethylamine<sup>27-30</sup>. Whilst FMO4 has not been extensively characterized, FMO1 and FMO3 were shown to have broad substrate ranges, metabolizing substrates as diverse as itopride (acetylcholine esterase inhibitor), and tamoxifen (anti breast-cancer drug)<sup>2,31-34</sup>. Also FMO2 features a rather broad substrate profile, acting on pesticides such as naphthylthiourea<sup>19</sup>, although its role in human metabolism remains partly unknown because FMO2 is not expressed in the majority of humans due to a mutation<sup>35,36</sup>. FMO5 is distinct from the other FMOs because it is able to perform Baeyer-Villiger oxidations (Scheme 1),<sup>37</sup> metabolizing ketone-containing drugs such as pentoxifylline (a muscle-pain killer)<sup>17</sup>. Recent literature documents that FMOs are associated to diseases such as atherosclerosis and diabetes,<sup>23,26</sup> promote longevity<sup>38</sup>, and regulate cholesterol and glucose levels<sup>26,39-41</sup>. Despite their discovery over 30 years ago, the determinants underlying the existence of five isoforms remain unexplored and, even more strikingly, no mammalian FMO has been structurally elucidated. This gap in our knowledge on these key enzymes of human drug metabolism likely reflects their distinctive feature: unlike bacterial, fungal, and insect FMOs that are soluble, mammalian FMOs are insoluble and reside in the membranes of the endoplasmic reticulum<sup>2,5</sup>.



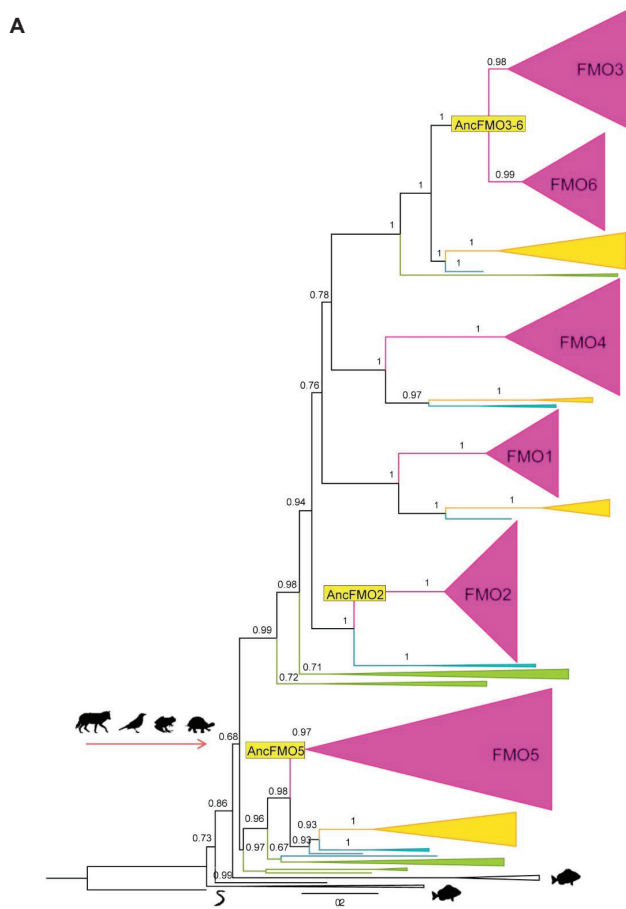
**Scheme 1.** The catalytic mechanism of FMOs. FAD<sub>ox</sub> is reduced by NADPH. FAD<sub>red</sub> is consequently oxidized by a molecule of dioxygen to generate the C4 $\alpha$ -(hydro)peroxide intermediate. The typical mode of action of FMOs with the distal oxygen atom from the intermediate being inserted onto a soft nucleophile through nucleophilic addition is shown with reference to trimethylamine. The Baeyer-Villiger monooxygenation activity conducted by human FMO5 is shown on the right with reference to heptan-2-one. The dotted arrow indicates the uncoupling reaction whereby the C4 $\alpha$ -(hydro)peroxide intermediate decays with the release of NADP<sup>+</sup> and hydrogen peroxide.

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

## Results

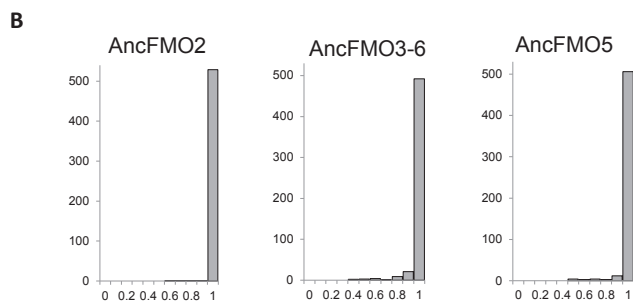
### Ancestral sequence reconstruction of mammalian flavin-containing monooxygenases

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



**Figure 1. Ancestral Sequence Reconstruction of FMOs.**

*A: Condensed Maximum Likelihood phylogeny of FMOs from jawed vertebrates. Clades are colored according to tetrapod classes: mammalia (magenta), aves (light orange), amphibia (green) and testudines (teal). Clades on the base are from other non-terrestrial gnathostomes (black). Rooting was performed according to the species tree. Above the branches transfer bootstrap expectation values are shown. The emergence of terrestrial vertebrates (tetrapods, 352 mya)<sup>46</sup> is marked with an arrow and cartoons on the left. The three ancestral nodes that were experimentally characterized are labeled with yellow squares. Fully annotated phylogeny is presented in Supplementary Figure 2.*



*B: Statistical confidence of ancestral amino acid states. The highest posterior probability (PP) for each of the inferred ancestral states (sites) in AncFMOs is shown. Average accuracy for AncFMO2= 0.994, AncFMO3-6: 0.982 and AncFMO5: 0.987.*

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

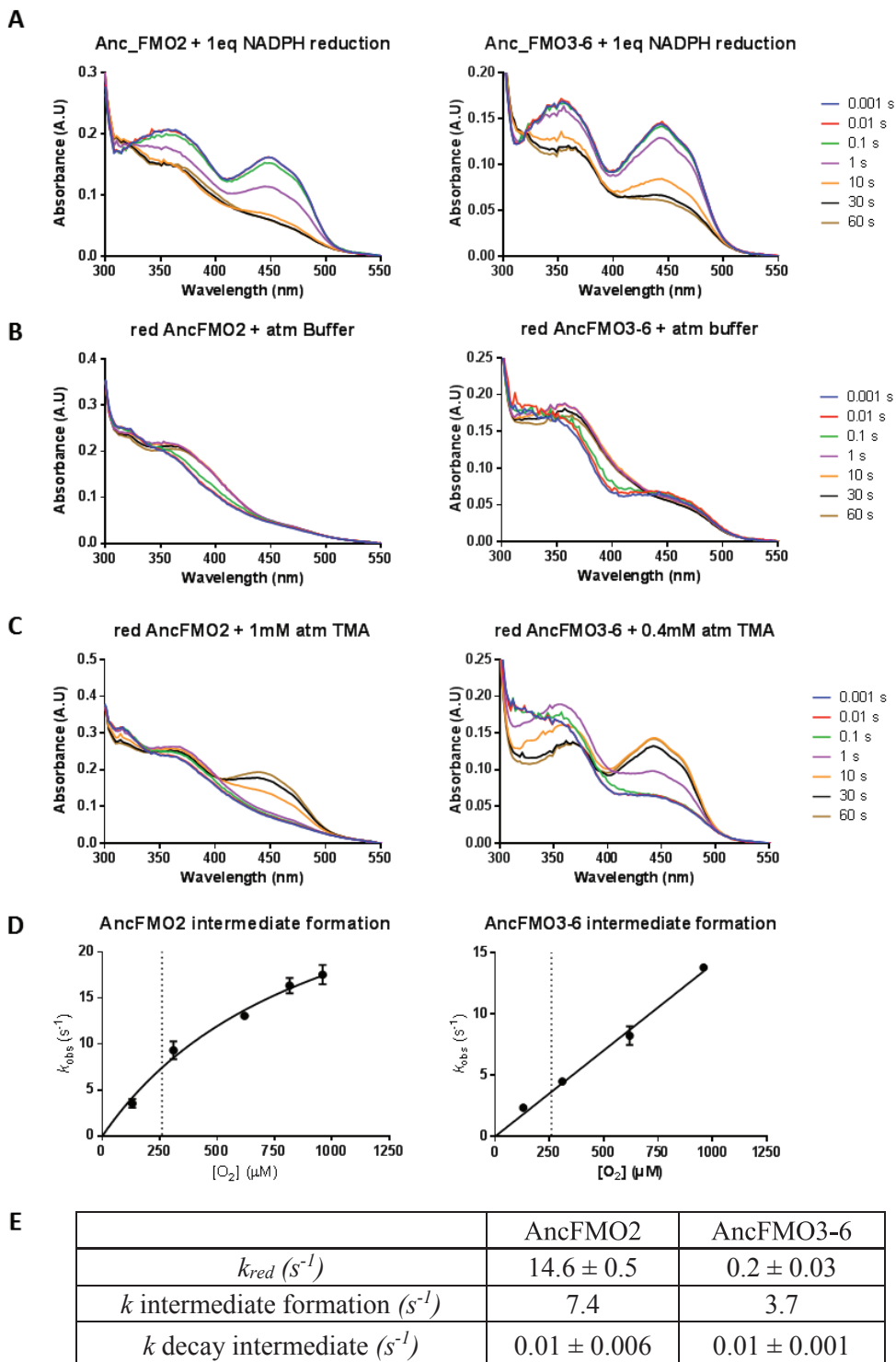
### *AncFMOs portray catalytic rates similar to those of extant mammalian FMOs*

Critically for our project, the hitherto generated AncFMOs sequences proved to encode stable proteins that can be effectively produced and purified as recombinant, FAD-loaded, and catalytically competent enzymes in *E. coli*. Thus, the first relevant result was that a convenient bacterial expression system for the study and biocatalytic exploitation of close homologs to human FMOs was established (see materials and methods). We next verified whether these enzymes retained enzymatic activities by performing steady-state kinetics experiments using a NADPH-depletion spectrophotometric assay. The

NADPH oxidase activity was initially tested (NADPH consumption in absence of an organic substrate;  $\text{NADPH}_{\text{uncoupling}}$  in Table 1). This was followed by the measurements of the reaction kinetics in the reaction in the presence of known oxygen-accepting substrates of FMO2 and FMO3 (methimazole, trimethylamine and thioanisole), and FMO5 (heptan-2-one). The results were reassuring in that AncFMO2, AncFMO3-6, and AncFMO5 proved to be enzymatically active with kinetic parameters very similar with those reported for their extant human-derived enzymes<sup>2,17,34,37,48-51</sup>. The  $k_{\text{cat}}$ ,  $K_{\text{M}}^{\text{NADPH}}$  and uncoupling values ranged between 0.03-0.32  $\text{s}^{-1}$ , 3.5-7.8  $\mu\text{M}$ , and 0.016-0.03  $\text{s}^{-1}$ , respectively (Table 1). It was especially noticeable that the AncFMOs displayed a high affinity towards the coenzyme NADPH and a significantly higher NADPH consumption rate when a suitable substrate was present. This result is in full agreement with the canonical catalytic mechanism observed for FMOs and sequence related flavoprotein monooxygenases (Scheme 1). These features were further demonstrated by stopped-flow kinetic studies. NADPH-reduced AncFMO2 and AncFMO3-6 were found to react rapidly with oxygen to form a stable and detectable C4 $\alpha$ -(hydro)peroxyflavin intermediate with its well-defined spectroscopic properties (Figure 2). Based on the steady-state kinetics data, AncFMO5 is assumed to behave similarly. Collectively, these experiments convincingly demonstrated that our AncFMO2/3-6/5 enzymes are enzymatically competent and exhibit the typical catalytic features of Class B flavoprotein monooxygenases.

**Figure 2. Stopped-flow kinetics studies on AncFMO2 and AncFMO3-6 (opposite page).**

**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 C4 $\alpha$ -(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 C4 $\alpha$ -(hydro)peroxyflavin intermediate which subsequently decays to form the reoxidized flavin species. **D:** Dependence of the rate of C4 $\alpha$ -(hydro)peroxyflavin formation ( $A_{360\text{nm}}$ ) on varying oxygen concentrations. The dotted lines correspond to the atmospheric concentration of dioxygen (0.26 mM). For AncFMO2, the observed saturation behavior suggests a binding event taking place before dioxygen reacts with the reduced flavin. Interestingly, such a saturation behavior was also reported for pig liver FMO1<sup>52</sup>. **E:** Rates of reduction, C4 $\alpha$ -(hydro)peroxyflavin formation, and C4 $\alpha$ -(hydro)peroxyflavin decay in the absence of substrate (0.26 mM dioxygen; dotted line on panels D).





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

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

<sup>b</sup> Heptan-2-one is a typical ketone substrate for the Baeyer-Villiger oxidation catalyzed by FMO5 (Scheme 1).

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

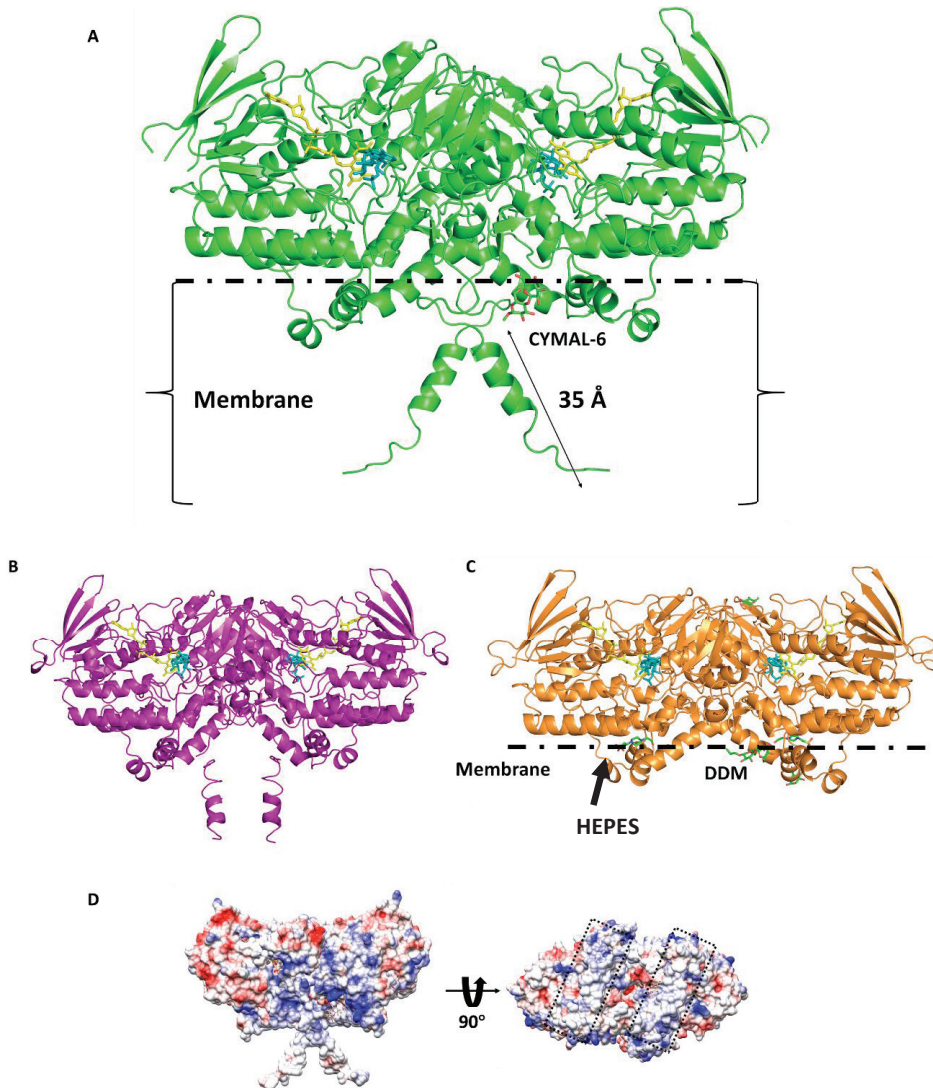
	Substrate <sup>a</sup>	$k_{cat}$ (s <sup>-1</sup> )	$K_M$ ( $\mu$ M)
<b>Ancient FMOs</b>			
<b>AncFMO2</b>	Methimazole	0.19 ± 0.01	106 ± 22
	Thioanisole	0.3 ± 0.02	6.9 ± 1.6
	Trimethylamine	0.16 ± 0.008	445 ± 74
	NADPH	0.32 ± 0.05	7.8 ± 1.4
	NADPH <sub>uncoupling</sub>	0.02 ± 0.001	20 ± 5.4
<b>AncFMO3-6</b>	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 <sub>uncoupling</sub>	0.022 ± 0.002	16 ± 5.4
<b>AncFMO5</b>	Heptan-2-one <sup>b</sup>	0.07 ± 0.003	6.36 ± 1.2
	NADPH	0.06 ± 0.001	6.48 ± 0.38
	NADPH <sub>uncoupling</sub>	0.03 ± 0.001	2.1 ± 0.5
<b>Extant FMOs</b>			
<b>human FMO3</b>	NADPH <sup>c</sup>	0.06 ± 0.16	46 ± 9
<b>human FMO5</b>	NADPH <sup>c</sup>	0.197 ± 0.009	59 ± 8

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

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

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

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



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

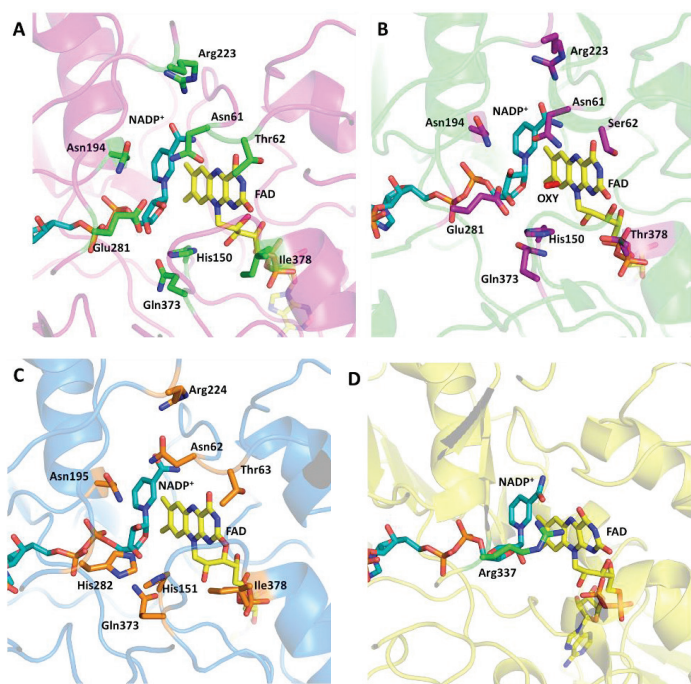
It was reported that truncation of the C-terminal helices was insufficient for protein solubilization<sup>54,55</sup>. This indicated that the enzyme possessed additional membrane-binding features. To understand which elements promote membrane association, the charge distribution on the protein surface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

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

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

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

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



**Figure 4. Active sites of the AncFMOs.** The active sites for AncFMO2, AncFMO3-6 and AncFMO5 are shown in A, B and C, respectively. All three bear a high degree of similarity with most amino acids being strictly conserved and displaying identical conformations. The differing residues are as follows: Thr62/Ser62/Thr63; E281/E281/H282; Ile378/Thr378/Ile378 for AncFMO2/AncFMO3-6/AncFMO5 respectively. AncFMO3-6 also contains a tentatively assigned molecule of dioxygen (OXY). For the sake of comparison, panel D shows the binding of NADP<sup>+</sup> to the active site of phenylacetone monooxygenase, a prototypical class B monooxygenase (PDB code 2ylr). Arg337 is a conserved residue that is essential for the Baeyer-Villiger activity of this and similar enzymes.

The binding mode of NADP<sup>+</sup> observed in the crystal structures is iconic to Class B flavin-dependent monooxygenases (Figure 4D).<sup>4</sup> The overhanging Arg223/223/224 is within hydrogen bonding distance of the carbonyl derived from the carbamide of NADP<sup>+</sup>. Additionally, the amino group of the same carbamide forms a hydrogen bond with the N5 atom of the isoalloxazine ring. More so, the nicotinamide ring is sterically held in place by a well conserved Asn194/194/195 which acts like a back door for the cofactor (Figure 4A-C). This feature is not uncommon and portrayed in some soluble FMOs by a protruding tyrosine.<sup>12,57,58</sup> The hydroxyl groups of the ribose forms part of an intricate hydrogen bonding network. The 2'-OH group is within hydrogen bonding distance of the back-door residue Asn194/194/195 (3.0 Å) in AncFMO3-6 and AncFMO5, and Glu281 (2.9 Å) in AncFMO2 and AncFMO3-6. Additionally, the conserved Gln373 among the AncFMOs is within hydrogen-bonding distance of the 3'-OH group. Collectively, these hydrogen bonds and the steric interactions orientate the nicotinamide and the ribose in a manner characteristic to this class of enzymes and reiterate a significant role of NADP<sup>+</sup> in catalysis, most likely in C4 $\alpha$ -(hydro)peroxyflavin formation/stabilization and substrate oxygenation (Scheme 1)<sup>12,52,57,59</sup>.

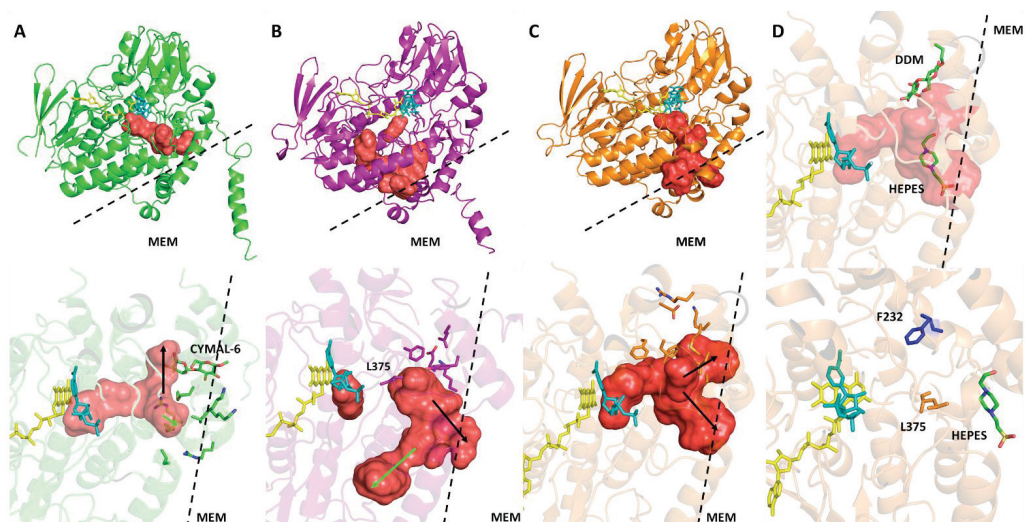
### *A Glu-His mutation in the mammalian specific insertion may promote Baeyer-Villiger oxidation in FMO5*

With AncFMO5 being structurally very similar to the AncFMO2 and AncFMO3-6, but at the same time functionally divergent, we sought out to clarify what features gave rise to its Baeyer Villiger oxidation activity. Inspecting the active site alone, the differing mode of action is likely derived from a Glu-to-His substitution. In AncFMO2 and AncFMO3-6, Glu281, derived from the above-described mammalian FMO-specific 80-residue insertion, points towards the flavin ring. With positively charged substrates being preferred by FMOs<sup>2</sup>, Glu281 is probably deprotonated and negatively charged within the cavity. In AncFMO5 however, this residue is substituted for His282 which optimally positions the N $\epsilon$ -H of its imidazole ring towards the substrate pocket (Figure 4C) and likely serves as a hydrogen bond donor. This function is commensurate with Baeyer-Villiger monooxygenases, whereby hydrogen bond-donating residues (*i.e.* Arg; Figure 4D) are prevalent in the vicinity of the FAD ring to activate the carbonyl functional group of the substrate for electrophilic attack by the flavin-peroxy intermediate and stabilize the Criegee intermediate formed during Baeyer-Villiger oxidation catalysis (Scheme 1)<sup>60-62</sup>. These observations rationalize the functional convergence observed among the FMO5 clade and Baeyer-Villiger monooxygenases. To probe the importance of His282 in AncFMO5, the H282E mutant was prepared and analyzed. This revealed that the H282E AncFMO5 mutant fully lost its activity. Analogously, the E281H AncFMO2 and E281H AncFMO3-6 mutant enzymes were prepared which were found to retain FMO activity toward thioanisole. Yet, they were not able to perform Baeyer-Villiger oxidations (Supplementary Table 2). This could be due to the fact that the fine structural and geometric features for formation and stabilization of the Criegee intermediate (Scheme 1) needs further mutations, e.g. in the second-shell of active site residues.

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

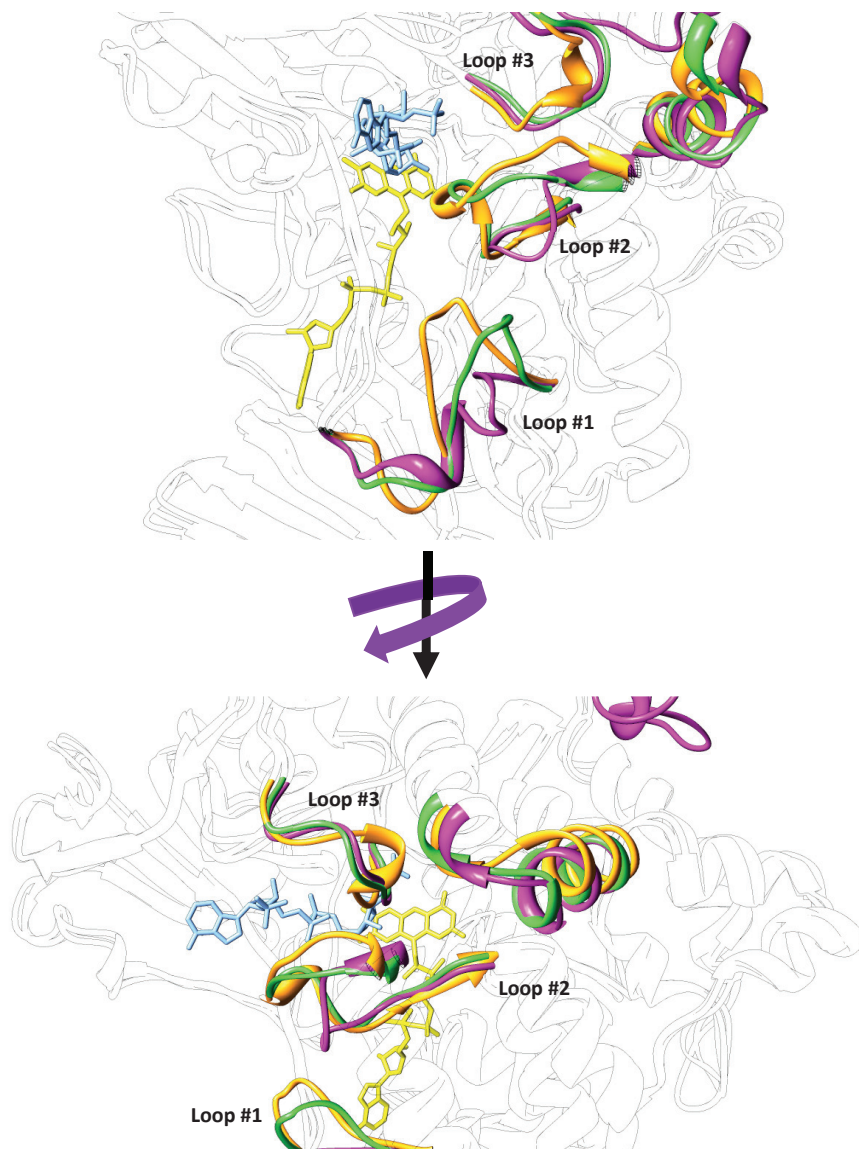
As the mammalian FMOs are notorious for their broad substrate profiles, we conducted extensive research to elaborate how the substrates navigate through the enzyme using the *HOLLOW* server<sup>63</sup>. The conserved tunnel is roughly perpendicular to the face of the isoalloxazine ring and extends outwards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

The substrates/products penetrate/exit the tunnels through the subdomain found in the 80-residue insertion. Here, the paths are heavily dictated by the conformations of the residues in and around the subdomain (Figure 5A-D). Specifically, a few noticeable changes were observed (Figure 6). The largest conformational difference is seen at residues 337-352 and 338-352 for AncFMO2/3-6 and AncFMO5 respectively (herein referred to as loop 1). In AncFMO2 and AncFMO5, loop 1 forms a large arched fold that sits underneath the NADP(H) binding pocket. In AncFMO3-6, loop 1 instead forms a tightly coiled  $\alpha$ -helix creating an open cavity below the NADP<sup>+</sup> binding pocket. This new cavity leads to the cytosolic tunnel observed in AncFMO3-6 (Figure 5B). The second difference observed comprises residues 419-431 for AncFMO2 and AncFMO5 and residues 419-429 for AncFMO3-6 (loop 2) in the neighborhood of the tunnel entrances. The final differences detected concern residues 273-282 of AncFMO2 and AncFMO3-6, and 274-283 of AncFMO5 (loop 3). In AncFMO5, loop 3 features a  $\alpha$ -helical turn that blocks the cytosolic tunnel observed in AncFMO2 and AncFMO3-6. Moreover, AncFMO5 possesses a shorter  $\alpha$ -helix in the subdomain which widens the cavity entrance site. These features have critical implications for the mechanisms of substrate binding and selectivity in FMOs. On the one hand, these structural variations on surface elements at the tunnel entrances are likely to govern the similar but not identical substrate acceptance of the FMOs. On the other hand, despite these differences, all three AncFMO structures show that the tunnels can be accessed by both hydrophilic substrates that predictably diffuse from the cytosol, and by hydrophobic substrates that likely diffuse from the membrane. Likewise, hydrophilic and hydrophobic products can diffuse from the active site to the cytosol and to the membrane, respectively.



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





*Figure 6. Structural differences among the AncFMOs. Upper and lower panels describe the conformational differences observed among the AncFMOs with AncFMO2, AncFMO3-6 and AncFMO5 depicted in lime green, dark magenta and orange respectively. Loop #1 contains residues 337-352 for AncFMO2 and AncFMO3-6, with 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.*

AncFMOs are thermostable enzymes that are stabilized significantly in the presence of NADP<sup>+</sup> and are reliable models for human FMOs.

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

## Discussion

Our work supports the notion that the number of FMOs in vertebrates significantly increased by successive gene duplication events, leading to the multiple paralogs observed in mammals today<sup>44</sup>. Tetrapods encode for four (amphibians, testudines and birds) or six (mammals) different FMOs, suggesting defined roles for each of these variants. Analyzing the different paralogs, we observed that FMO3 and FMO6 followed a common evolutionary path preceded by the diversification of FMO4, FMO1 and FMO2. FMO5 originated from the earliest gene duplication event and, intriguingly, is encoded by all the aforementioned terrestrial vertebrates' classes. With AncFMOs exhibiting similar substrate profiles and catalytic rates as their FMO successors, we propose that this class of enzymes has an evolutionary conserved mode of action. Moreover, two new features are derived from ancestral sequence reconstruction: (i) increased melting temperatures and (ii) the stabilizing effect induced by NADP<sup>+</sup> (Supplementary Table 3). With the mutations scattered across the protein, it is unlikely that individual mutations stabilize the enzyme greatly. Their summation however, enhances stability tremendously. Whether this higher thermal stability of the AncFMOs has a biological meaning remains unclear<sup>66</sup>.

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

With the AncFMOs all accommodating very similar active sites, substrate profiles are likely differentiated by the tunnels penetrating the scaffold. FMO2 is generally known to be slightly more restrictive in terms of substrate size, mostly metabolizing molecules possessing amino groups attached to large aliphatic tails<sup>2,69</sup> whilst, FMO3 and FMO1 are understood to be more promiscuous, occupying a breadth of substrate sizes<sup>2,18,70</sup>. The tunnels hereby depicted do not allow us to confidently rationalize these phenomena specifically. For example, the high activity of FMO3 towards trimethylamine likely arises from the combination of subtle factors including the distribution of charged residues, the partition of hydrophobic versus hydrophilic residues at the entrance and inside the FMO3 tunnel, and the flexibility of the residues that gate tunnel access and substrate diffusion. Nevertheless, the overall architecture of the catalytic site and of the access tunnels fully explains the broad substrate scopes of mammalian FMOs. Above all, the catalytic site promotes the flavin-mediated activation of oxygen through the formation of the flavin-(hydro)peroxide intermediate as observed in soluble FMOs as well as in Baeyer-Villiger monooxygenases<sup>71,72</sup>. After flavin reduction, the nicotinamide-ribose moiety of NADP<sup>+</sup> relocates to give access to the oxygen-reacting C4a atom of the flavin and thereby promote formation of the flavin-(hydro)peroxide that awaits a substrate to be monooxygenated (Scheme 1). Along these lines, no elements for the specific recognition of the substrate can be identified. The tunnel and the inner catalytic cavity of FMOs are rather designed to allow a “controlled” access to the flavin-(hydro)peroxide without any strict or rigorous binding selectivity. It can easily be envisioned that the tunnels can adapt themselves depending on the bulkiness of the ligands. The gating elements (e.g. Leu375) may well seal the active-site cavity when small substrates are bound (Figure 5). Likewise, the same elements could enable the binding of bulky molecules whose non-reactive groups extend along the tunnel. Thus, FMOs exhibit typical features of enzymes that broadly function in xenobiotic detoxification. Their preference for nitrogen- and sulfur-containing substrates primarily reflects the pronounced reactivity of the flavin-(hydro)peroxide towards these soft-nucleophiles.

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

## Materials & Methods

### *Phylogenetic Inference and Ancestral Sequence Reconstruction*

To obtain a robust and representative phylogeny of FMOs, sequences from the Bacteria and Eukarya were collected by homology searches using BLAST and HMM profiling. 310 sequences were collected and aligned with MAFFT v7<sup>73</sup>. Best-fit model parameters were obtained by the Akaike information criterion in ProtTest v3.4. Phylogenies were inferred employing the maximum likelihood method in PhyML v3.0 or RAxML v0.6.0 with 500/1000 bootstraps and transfer bootstrap expectation (TBE) subsequently applied.<sup>74</sup> As FMOs are not monophyletic, derived clades BVMOs and NMOs were included in the phylogeny. Later<sup>7,9</sup>, the Gnathostomata FMOs phylogeny was constructed for ancestral sequence reconstruction. To do this, a dataset of 361 sequences was collected including also a cephalochordate sequence to root the tree according to the species tree (Supplementary Data 1)<sup>75</sup>.

Ancestral sequence reconstruction was performed using the maximum likelihood inference method in PAMLX v.4.9<sup>43,76</sup>. Sequences were analyzed using an empirical amino acid substitution model (model = 3), 4 gamma categories and LG substitution matrix. The posterior probability distribution of ancestral states at each site was analyzed at nodes AncFMO2, AncFMO3-6 and AncFMO5. Sites were considered ambiguously reconstructed if alternative states displayed PP > 0.2.<sup>77</sup> Alternative sites were found to be 2 for AncFMO2, 15 for AncFMO3-6 and 11 for AncFMO5 (Supplementary Table 4). These are mostly conservative amino acid substitutions, and after mapping in the crystal structures, it was evident that they all lay in the periphery of the protein, not affecting the catalytic core.

### *Cloning and Expression of the AncFMOs*

Genes were ordered from Genescript containing Bsal sites at both the 5' and 3' ends of the insert. The insert contained overhangs TGGT and CAAG at the 5' and 3' ends respectively to then be inserted into common pBAD-NK destination vectors with the following modifications: three Bsal sites were eliminated and two were introduced to facilitate the cloning that incorporated SUMO and 8xHis-tag regions to the N-terminus. Inserts were fused into the destination vectors through Golden Gate cloning. The sample was prepared with the following: 75 ng of golden gate entry vector, a molar ratio

of 2:1 between insert and vector, Bsal(-HF) (15 U), 30 WU T4 DNA ligase (15 U), T4 DNA ligase buffer (1x) and Nuclease free water added to a final volume of 20  $\mu$ L. During the cloning procedure, a negative control was prepared with the fragments/inserts omitted. The Golden Gate assembly was conducted in the following manner, where maximum efficiency was desired: A cycle of 37 °C for 5 mins followed by 16 °C for 10 minutes was repeated 30 times; followed by 55 °C for 10 minutes, 65 °C for a further 20 minutes, finishing with 8 °C for 20 minutes. Once cloned, the plasmids were transformed by heat shock into *E. coli*. BL21 cells (25 seconds, 42 °C). Cells from the resulting colonies were pre-inoculated into 100 mL of LB broth containing 100  $\mu$ g/mL of ampicillin and grown overnight at 37 °C. The cultures were then transferred to 1L Terrific Broth cultures (15 mL) and grown at 24 °C, rpm 180 for 5 – 6 hours until the OD reached 0.3. The cultures were then induced with a sterilized arabinose solution (20% w/v), final concentration of 0.02% (w/v) and incubated at 24 °C, 180 rpm for an additional 24 hours. Cells were then harvested by centrifugation (5000g, 15 mins, 10 °C) flash frozen in liquid nitrogen and stored at -80 °C. For site-directed mutagenesis, a PCR-reaction mixture was prepared with 10  $\mu$ M primer forward and reverse, 100 ng of template DNA, 1.6 % DMSO, 0.8 mM MgCl<sub>2</sub> and 1x Pfu Ultra II Hotstart Master Mix (Agilent). The Quickchange PCR cycle was performed using the following method: firstly a 5 minute incubation at 95°C, then cycles (95 °C for 5 minutes - 60 °C for 30 seconds - 72 °C for 6 minutes) were repeated 25 times; followed by 72 °C for 10 minutes and finishing with 8 °C on hold. The PCR mixture was digested with *DpnI* overnight and transformed into *E. coli*.

#### **Cell Disruption, Extraction, and Purification of AncFMOs**

All procedures were carried out in ice or at 4 °C. Cells (ca. 20 g) were resuspended (1:5) in buffer A (250 mM NaCl, 50 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.8) and included additional protease inhibitors: phenylmethylsulfonyl fluoride (1 mM), leupeptin (10  $\mu$ M), pepstatin (10  $\mu$ M) and DNase I (5  $\mu$ g/g of cell paste). The solution was stirred and incubated at 4 °C for 45 minutes before cell lysis was conducted using sonication or a high-pressure homogenizer. Sonication was conducted using the following conditions: 50 mL solution, 5 seconds on, 10 seconds off, 1-minute cycles with a total sonication time of 3 minutes using a microtip (70% amplitude). Cells were passed through a high-pressure homogenizer twice. Lysed cells were then spun down (1200g, 12 mins, 4 °C) to remove the cell debris. The resultant supernatant was then centrifuged further (56,000g, 1 hr. 40 mins, 4 °C) to collect the membrane pellet which was then re-homogenized in buffer A (15 mL) and centrifuged again (56,000g, 1 hr. 20 mins, 4 °C) to further purify the insoluble material. The resulting pellet was re-homogenized in buffer A (7 mL) and diluted to a final concentration of 13 mg/mL (assayed using Biuret reagent). Triton X-100 Reduced (TRX-100-R) (Sigma-Aldrich) was then added to the solution (0.5% (v/v) final concentration) and mixed overnight at 4 °C. The detergent-solubilized fraction containing the AncFMOs was then abstracted by collecting the supernatant after centrifugation (56,000g, 1 hr. 20 mins, 4 °C). The supernatant was then transferred to a pre-equilibrated (with buffer A and 0.05% (v/v) TRX-100-R) gravity column containing a Ni-resin (GE Healthcare).

The supernatant was washed with buffer A, containing 0.05% TRX-100-R, and then with increasing concentrations of buffer B (50 mM KH<sub>2</sub>PO<sub>4</sub>, 500 mM NaCl, 300 mM imidazole, pH 7.8), also containing 0.05% (v/v) TRX-100-R, in step-by-step fashion: 5 mM imidazole wash, 30 mM imidazole wash and finally a 300 mM imidazole wash, where the protein then eluted. The buffers were then exchanged using a centrifugal filter unit (50 kDa cut-off) and multiple washes with buffer A with 0.05% (v/v) TRX-100-R. This step was important to remove high concentrations of imidazole employed during the elution. The protein sample was then concentrated down to a final volume between 500 and 1000  $\mu$ L. The sample was then mixed with a 6xHis-tagged SUMO protease (1.2 mg/mL) to a volume ratio of 10:1 respectively and incubated overnight at 4 °C. The sample was then loaded onto an Akta purification system (GE Healthcare) endowed with a multiwavelength detector (set at 280/370/450 nm) and then onto a nickel-affinity His-trap column (GE Healthcare). The column was pre-equilibrated with buffer A containing 0.05% (v/v) TRX-100-R, as stated before, with the proteins eluting out in the presence of 6 mM imidazole derived from buffer B (2%) containing 0.05% (v/v) TRX-100-R. The SUMO-His-tag cleaved protein was then concentrated, and buffer exchanged using a concentrating centrifugal filter unit (50 kDa cut-off) to a final volume between 250 and 500  $\mu$ L. The sample was incubated for 1 h with 100  $\mu$ M FAD at 4 °C and then loaded onto a gel filtration column (Superdex 200 10/300, GE Healthcare) pre-equilibrated with a storage buffer (50 mM Tris-HCl pH 8.5 at 4 °C, 10 mM NaCl) and a detergent of choice to obtain a higher degree of purity (obtained from anatrace). Typically DDM was used (0.03% (w/v) DDM (analytical grade)), but other detergents were used for crystallization screenings at 3x their respective Critical-Micelle Concentration (CMC). The protein eluted with a very high purity and homogeneity (evaluated by SDS-page electrophoresis and the shape of the peak in the chromatogram respectively) with an elution volume of 10.5 – 11 mL. The sample was concentrated down to 100  $\mu$ L using a centrifugal filter unit (50 kDa cut-off) with a final concentration ranging from 5 to 30 mg/mL.

### ***Preparation of human FMO3 and human FMO5***

Full-length cDNA encoding for Homo sapiens FMO3 (UniProt P31513) and FMO5 (NCBI accession number: Z47553) were cloned into a modified pET-SUMO vector (Invitrogen) to allow insertion of a cleavable N-terminal 8xHis-SUMO tag. Expression, cell disruption, extraction, and purification were performed according to the methods previously described for human FMO5<sup>37</sup>.

### ***Kinetic Assays of the AncFMOs***

Steady-state kinetics assays were performed in duplicates on a Jasco V-660 spectrophotometer. Enzyme activity of the ancestral proteins was measured by monitoring NADPH consumption (absorbance at 340 nm,  $\epsilon_{340} = 6.22 \text{ mM}^{-1} \text{ cm}^{-1}$  for NADPH). The buffer used for kinetic analyses was 50 mM potassium phosphate, 250 mM NaCl, 0.05% TRX-100-R (Sigma-Aldrich), pH 7.5. The reaction volumes were 100  $\mu$ L and contained variable NADPH and substrate (methimazole, thioanisole, trimethylamine,

heptanone) concentrations and 0.01-2.0  $\mu\text{M}$  enzyme. The spectrophotometer was set at 37 °C and the NADPH and substrate mix were also incubated at 37 °C for 5 minutes before starting the reaction by adding enzyme. The pH and temperature conditions were set based on literature studies for a fair comparison with previously reported properties of mammalian FMOs.

### *Kinetic Assays of human FMO3*

Kinetic assays were performed in duplicate at 25 °C on a Varian spectrophotometer (Cary 100 Bio) equipped with a thermostatic cell compartment. The apparent  $K_M$  for NADPH was measured by varying NADPH concentrations from 20 to 400  $\mu\text{M}$  in aerated reaction mixtures (150  $\mu\text{L}$ ) containing 2.5  $\mu\text{M}$  human FMO3, 50 mM HEPES pH 7.5, 10 mM KCl, 5% (v/v) glycerol, 0.05% (v/v) TRX-100-R. Reactions were followed by monitoring the decrease of NADPH concentration (decrease in absorbance) at 340 nm ( $\epsilon_{340} = 6.22 \text{ mM}^{-1} \text{ cm}^{-1}$  for NADPH).

### *Rapid Kinetics Analysis of the AncFMOs*

Stopped-flow experiments were carried out using the SX20 stopped-flow spectrophotometer equipped with either the photodiode array detector or the single channel photomultiplier (PMT) module (Applied Photophysics, Surrey, UK). Results were obtained by mixing 50  $\mu\text{L}$  of two solutions in single mixing mode. All solutions were prepared in 50 mM potassium phosphate, 10 mM NaCl and 0.05% TRX-100-R, pH 7.5 at 25 °C. For every reaction a concentration of 10-15  $\mu\text{M}$  enzyme was used and measurements were done in duplicate. When needed, the solutions were supplemented with 5.0 mM glucose and the enzyme and/or substrate solutions were made anaerobic by flushing solutions for 10 minutes with nitrogen, followed by adding 0.3  $\mu\text{M}$  glucose oxidase (*Aspergillus niger*, type VII, Sigma-Aldrich) to consume the left-over oxygen. The monitoring of the reductive half reaction was done by mixing the anaerobic enzyme solution with anaerobic buffer containing increasing concentrations of NAD(P)H (Oriental Yeast Co. LTD.). The rate of flavin reduction was determined by measuring the decrease of absorbance at 447 nm or 442 nm for AncFMO2 and AncFMO3-6, respectively. In order to reduce the flavin cofactor in the FMOs for the oxidative half reaction, NADPH was added to an anaerobic solution containing an equivalent amount of AncFMO. The resulting solution was incubated on ice until the bleaching of the FAD was complete, indicating complete reduction to  $\text{FADH}^2$ . The anaerobically reduced FMOs were mixed with air-saturated buffer first and then with air-saturated buffer containing 1.0 mM or 0.4 mM of trimethylamine, respectively. This allowed us to follow the spectral changes during the oxidative half-reactions. The C4 $\alpha$ -(hydro)peroxyflavin intermediate formation and decay was specifically monitored using the PMT module at 360 nm. For determining the rates of reoxidation, the reduced enzymes were mixed with buffers containing different concentrations of dioxygen. The final concentrations of dioxygen (0.13, 0.31, 0.61, 0.96 mM after mixing) were achieved by mixing the anaerobic enzyme solution with (i) air-saturated buffer; (ii) mixing equal volumes of 100% argon buffer

and 100% O<sub>2</sub> buffer; (iii) 100% O<sub>2</sub> buffer; (iv) 100% O<sub>2</sub> buffer on ice. All solutions were bubbled for ten minutes at room temperature except the last one which was done on ice. In the case of AncFMO2, in order to confirm its saturating behaviour, an additional measurement was performed at 0.816 mM of O<sub>2</sub> by mixing 100% O<sub>2</sub> buffer on ice with 100% argon buffer.

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

### *ThermoFAD Assays<sup>65</sup>*

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

### *Conversions*

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

### *Crystallization and Structural Determination of the AncFMOs*

Each AncFMO crystallized in a range of conditions with multiple detergents. Typically, PEG 4000 was optimal for crystallization. The highest diffracting crystallization conditions for each AncFMO are described below. AncFMO2 (with and without NADP<sup>+</sup>): 12–15 mg/mL of AncFMO2 (in storage buffer and CYMAL-6 (0.09% (w/v))) was incubated with crystallization conditions of HEPES buffer (0.1 M, pH 7.5) and PEG 4000 (10%) at 20 °C with a ratio of 1:1 on a sitting drop. Sitting drop was 2 µL after mixing and contained a reservoir of 1 mL. Prior to crystallization, NADP<sup>+</sup> (1 mM final) was incubated with 12–



15 mg/mL of AncFMO2 for 1 hour at 4 °C. After two days, large yellow crystals formed. AncFMO3-6: 12-15 mg/mL of AncFMO3-6 (in storage buffer and DDM (0.03% (w/v)) was incubated with crystallization conditions of Sodium Acetate buffer (0.1 M, pH 5.5) and PEG 4000 (7.5%) at 20 °C with a ratio of 1:1 on a sitting drop. Sitting drop was 2 µL after mixing and contained a reservoir of 1 mL. Prior to crystallization, NADP<sup>+</sup> (1 mM final) was incubated with 12-15 mg/mL of AncFMO3-6 for 1 hour at 4 °C. After one day, large yellow crystals formed. AncFMO5: 12 mg/mL of AncFMO5 (in storage buffer and DDM (0.03% (w/v)) was incubated with crystallization conditions of HEPES buffer (0.1 M, pH 6.9) and PEG 4000 (9%) at 20 °C with a ratio of 1:1 on a sitting drop. Sitting drop was 2 µL after mixing and contained a reservoir of 1 mL. Prior to crystallization, NADP<sup>+</sup> (1 mM final) was incubated with 12-15 mg/mL of AncFMO5 for 1 hour at 4 °C. After one day, large yellow hexagon shaped crystals formed. During crystal fishing, a cryo-protectant was prepared containing modified crystallizations conditions with 20% glycerol and PEG 4000 (15%). Data were collected at the European Synchrotron Radiation Facility (Grenoble, France) and the Swiss Light Source, (Villigen, Switzerland) and processed with the XDS<sup>78</sup> and CCP4 packages<sup>79</sup>. Aimless was used to merge the observations into average densities (Supplementary Table 1). STARANISO was additionally used for AncFMO2 which suffered greatly from anisotropy<sup>80,81</sup>. The phase problem was solved by Molecular Replacement using a recently solved insect FMO (PDB:5nmw)<sup>82</sup> as a search model, and then AncFMO2 for the proceeding AncFMOs, using Phaser and Molrep<sup>83</sup>. The phases were greatly improved by density averaging with DM<sup>79,84</sup>. Model Building and Refinement were then conducted using COOT,<sup>85</sup> Buccaneer<sup>79</sup>, and Refmac5<sup>86</sup> respectively

(Supplementary Table 1). Figures were then generated using UCSF Chimera<sup>87</sup>, PyMOL (DeLano Scientific; [www.pymol.org](http://www.pymol.org)) and CCP4mg<sup>79</sup>.

### **Data availability**

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

### **Author Contributions**

All listed authors performed experiments and analyzed data: CRN generated purification protocols, crystallized the AncFMOs, collected the corresponding datasets at ESRF and SLS facilities, performed structural analysis and elucidated the AncFMO structures. GB and CRN performed golden gate cloning to insert the AncFMO genes into their respective vectors, designed by GB. GB carried out mutagenesis and extensive kinetic analysis and validated the substrate profiles using stopped-flow UV/Vis spectroscopy

and GCMS for each AncFMO. MLM conducted thorough evolutionary analyses and performed ancestral sequence reconstruction obtaining AncFMO protein sequences. CRN, GB and MLM prepared the figures. CRN wrote the manuscript and edited by AM, MWF and MLM. All authors provided critical feedback and helped shape the research, analysis and manuscript. AM and MWF conceived the original idea.

### **Funding and acknowledgements**

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

### **Declaration of interest**

The authors declare that they have no conflicts of interest with the contents of this article.

## References

- 1 Jakoby, W. B. & Ziegler, D. M. The enzymes of detoxication. *J. Biol. Chem.* 265, 20715–20718 (1990).
- 2 Krueger, S. K. & Williams, D. E. Mammalian flavin-containing monooxygenases: Structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacol. Ther.* 106, 357–387 (2005).
- 3 Cruciani, G. et al. Flavin monooxygenase metabolism: Why medicinal chemists should matter. *J. Med. Chem.* 57, 6183–6196 (2014).
- 4 Huijbers, M. M. E., Montersino, S., Westphal, A. H., Tischler, D. & Van Berkel, W. J. H. Flavin dependent monooxygenases. *Arch. Biochem. Biophys.* 544, 2–17 (2014).
- 5 Ziegler, D. M. Flavin-containing monooxygenases: enzymes adapted for multisubstrate specificity. *Trends Pharmacol. Sci.* 11, 321–324 (1990).
- 6 Cashman, J. R. Some distinctions between flavin-containing and cytochrome P450 monooxygenases. *Biochem. Biophys. Res. Commun.* 338, 599–604 (2005).
- 7 Mascotti, M. L., Lapadula, W. J. & Ayub, M. J. The origin and evolution of Baeyer - Villiger Monooxygenases (BVMOs): An ancestral family of flavin monooxygenases. *PLoS One* 10, 1–16 (2015).
- 8 D. M. Ziegler & Pettit, F. H. Formation of an intermediate N-oxide in the oxidative demethylation of N,N-dimethylaniline catalyzed by liver microsomes. *Biochem. Biophys. Res. Commun.* 15, 188–193 (1964).
- 9 Mascotti, M. L., Juri Ayub, M., Furnham, N., Thornton, J. M. & Laskowski, R. A. Chopping and Changing: the Evolution of the Flavin-dependent Monooxygenases. *J. Mol. Biol.* 428, 3131–3146 (2016).
- 10 Cashman, J. R. & Zhang, J. Human Flavin-Containing Monooxygenases. *Annu. Rev. Pharmacol. Toxicol.* 46, 65–100 (2006).
- 11 Romero, E., Castellanos, J. R. G., Gadda, G., Fraaije, M. W. & Mattevi, A. The same substrate, many reactions: oxygen activation in flavoenzymes. *Chem. Rev.* 118, 1742–1769 (2017).
- 12 Alfieri, A., Malito, E., Orru, R., Fraaije, M. W. & Mattevi, A. Revealing the moonlighting role of NADP in the structure of a flavin-containing monooxygenase. *Proc. Natl. Acad. Sci.* 105, 6572–6577 (2008).
- 13 Zhang, J. Quantitative analysis of FMO gene mRNA levels in human tissues. *Drug Metab. Dispos.* 34, 19–26 (2005).
- 14 Dolphin, C. T., Cullingford, T. E., Shephard, E. A., Smith, R. L. & Phillips, I. R. Differential developmental and tissue-specific regulation of expression of the genes encoding three members of the flavin-containing monooxygenase family of man, FMO1, FMO3 and FMO4. *Eur. J. Biochem.* 235, 683–689 (1996).
- 15 McCombie, R. R., Dolphin, C. T., Povey, S., Phillips, I. R. & Shephard, E. A. Localization of human flavin-containing monooxygenase genes FMO2 and FMO5 to chromosome 1q. *Genomics* 34, 426–429 (1996).
- 16 Hernandez, D., Janmohamed, A., Chandan, P., Phillips, I. R. & Shephard, E. A. Organization and evolution of the flavin-containing monooxygenase genes of human and mouse. *Pharmacogenetics* 14, 117–130 (2004).
- 17 Fiorentini, F. et al. Baeyer-Villiger Monooxygenase FMO5 as Entry Point in Drug Metabolism. *ACS Chem. Biol.* 12, 2379–2387 (2017).

- 18 Poulsen, L. L. & Ziegler, D. M. Multisubstrate flavin-containing monooxygenases: applications of mechanism to specificity. *Chem. Biol. Interact.* 96, 57–73 (1995).
- 19 Henderson, M. C., Krueger, S. K., Siddens, L. K., Stevens, J. F. & Williams, D. E. S-oxygenation of the thioether organophosphate insecticides phorate and disulfoton by human lung flavin-containing monooxygenase 2. *Biochem. Pharmacol.* 68, 959–967 (2004).
- 20 Cashman, J. R. Role of flavin-containing monooxygenase in drug development. *Expert Opin. Drug Metab. Toxicol.* 4, 1507–1521 (2008).
- 21 Siddens, L. K., Krueger, S. K., Henderson, M. C. & Williams, D. E. Mammalian flavin-containing monooxygenase (FMO) as a source of hydrogen peroxide. *Biochem. Pharmacol.* 89, 141–147 (2014).
- 22 Li, C. Y. et al. Structural mechanism for bacterial oxidation of oceanic trimethylamine into trimethylamine N-oxide. *Mol. Microbiol.* 103, 992–1003 (2017).
- 23 Miao, J. et al. Flavin-containing monooxygenase 3 as a potential player in diabetes-associated atherosclerosis. *Nat. Commun.* 6, 1–10 (2015).
- 24 Krueger, S. K. et al. Genetic polymorphisms of flavin-containing monooxygenase (FMO). 34, 523–532 (2002).
- 25 Dolphin, C. T., Janmohamed, A., Smith, R. L., Shephard, E. A. & Phillips, I. R. Compound heterozygosity for missense mutations in the flavin-containing monooxygenase 3 (FMO3) gene in patients with fish-odour syndrome. *Pharmacogenetics* 10, 799–807 (2000).
- 26 Veeravalli, S. et al. Effect of flavin-containing monooxygenase genotype, mouse strain, and gender on trimethylamine N-oxide production, plasma cholesterol concentration, and an index of atherosclerosis. *Drug Metab. Dispos.* 46, 20–25 (2018).
- 27 Phillips, I. R. & Shephard, E. A. Flavin-containing monooxygenases: mutations, disease and drug response. *Trends Pharmacol. Sci.* 29, 294–301 (2008).
- 28 Hernandez, D. et al. Trimethylaminuria and a human FMO3 mutation database. *Hum. Mutat.* 22, 209–213 (2003).
- 29 Dolphin, C. T., Janmohamed, A., Smith, R. L., Shephard, E. A. & Phillips, I. R. Missense mutation in flavin-containing mono-oxygenase 3 gene, FMO3, underlies fish-odour syndrome. *Nat. Genet.* 17, 491–494 (1997).
- 30 Lang, D. H. et al. Isoform specificity of trimethylamine N-oxygenation by human flavin-containing monooxygenase (FMO) and P450 enzymes Selective catalysis by fmo3. *Biochem. Pharmacol.* 56, 1005–1012 (1998).
- 31 Hodgson, E., Rose, R. L., Cao, Y., Dehal, S. S. & Kupfer, D. Flavin-containing monooxygenase isoform specificity for the N-oxidation of tamoxifen determined by product measurement and NADPH oxidation. *J. Biochem. Mol. Toxicol.* 14, 118–120 (2000).
- 32 Mushiroda, T., Douya, R., Takahara, E. & Nagata, O. The involvement of flavin-containing monooxygenase but not CYP3A4 in metabolism of itopride hydrochloride, a gastroprokinetic agent: Comparison with cisapride and mosapride citrate. *Drug Metab. Dispos.* 28, 1231–1237 (2000).
- 33 Shephard, E. A. & Phillips, I. R. The potential of knockout mouse lines in defining the role of flavin-containing monooxygenases in drug metabolism. *Expert Opin. Drug Metab. Toxicol.* 6, 1083–1094 (2010).
- 34 Krueger, S. K., VanDyke, J. E., Williams, D. E. & Hines, R. N. The role of flavin-containing monooxygenase (FMO) in the metabolism of tamoxifen and other tertiary amines. *Drug Metab. Rev.* 38, 139–147 (2006).

- 35 Veeramah, K. R. et al. The potentially deleterious functional variant flavin-containing monooxygenase 2\*1 is at high frequency throughout sub-Saharan Africa. *Pharmacogenet. Genomics* 18, 877–886 (2008).
- 36 Dolphin, C. T. et al. The flavin-containing monooxygenase 2 gene (FMO2) of humans, but not of other primates, encodes a truncated, nonfunctional protein. *J. Biol. Chem.* 273, 30599–30607 (1998).
- 37 Fiorentini, F. et al. Biocatalytic Characterization of Human FMO5: Unearthing Baeyer-Villiger Reactions in Humans. *ACS Chem. Biol.* 11, 1039–1048 (2016).
- 38 Leiser, S. F. et al. Promotes Longevity and Health Span. *Science*. 350, 1375–1378 (2015).
- 39 Warriar, M. et al. The TMAO-Generating Enzyme Flavin Monooxygenase 3 Is a Central Regulator of Cholesterol Balance. *Cell Rep.* 10, 326–338 (2015).
- 40 Gonzalez Malagon, S. G. et al. The phenotype of a knockout mouse identifies flavin-containing monooxygenase 5 (FMO5) as a regulator of metabolic ageing. *Biochem. Pharmacol.* 96, 267–277 (2015).
- 41 Scott, F. et al. Identification of flavin-containing monooxygenase 5 (FMO5) as a regulator of glucose homeostasis and a potential sensor of gut bacteria. *Drug Metab. Dispos.* 45, 982–989 (2017).
- 42 Risso, V. A., Sanchez-Ruiz, J. M. & Ozkan, S. B. Biotechnological and protein-engineering implications of ancestral protein resurrection. *Curr. Opin. Struct. Biol.* 51, 106–115 (2018).
- 43 Hochberg, G. K. A. & Thornton, J. W. Reconstructing Ancient Proteins to Understand the Causes of Structure and Function. *Annu. Rev. Biophys.* 46, 247–269 (2017).
- 44 Hao, D. C., Chen, S. L., Mu, J. & Xiao, P. G. Molecular phylogeny, long-term evolution, and functional divergence of flavin-containing monooxygenases. *Genetica* 137, 173–187 (2009).
- 45 Hines, R. N. Alternative Processing of the Human FMO6 Gene Renders Transcripts Incapable of Encoding a Functional Flavin-Containing Monooxygenase. *Mol. Pharmacol.* 62, 320–325 (2002).
- 46 Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* 34, 1812–1819 (2017).
- 47 Grantham, R. Amino acid difference formula to help explain protein evolution. *Science*. 185, 862–864 (1974).
- 48 Krueger, S. K. et al. Characterization of sulfoxylation and structural implications of human flavin-containing monooxygenase isoform 2 (FMO2.1) variants S195L and N413K. *Drug Metab. Dispos.* 37, 1785–1791 (2009).
- 49 Overby, L. H., Carver, G. C. & Philpot, R. M. Quantitation and kinetic properties of hepatic microsomal and recombinant flavin-containing monooxygenases 3 and 5 from humans. *Chem. Biol. Interact.* 106, 29–45 (1997).
- 50 Ripp, S. L., Itagaki, K., Philpot, R. M. & Elfarra, A. A. Methionine S-oxidation in human and rabbit liver microsomes: Evidence for a high-affinity methionine S-oxidase activity that is distinct from flavin-containing monooxygenase 3. *Arch. Biochem. Biophys.* 367, 322–332 (1999).
- 51 Lin, J. & Cashman, J. R. N-oxygenation of phenethylamine to the trans-oxime by adult human liver flavin-containing monooxygenase and retroreduction of phenethylamine hydroxylamine by human liver microsomes. *J. Pharmacol. Exp. Ther.* 282, 1269–1279 (1997).
- 52 Beaty, N. B. & Ballou, D. P. The reductive half-reaction of liver microsomal FAD-containing monooxygenase. *J. Biol. Chem.* 256, 4611–4618 (1981).
- 53 Krissinel, E. Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Res.* 43, W314–W319 (2015).

- 54 Geier, M. et al. Human FMO2-based microbial whole-cell catalysts for drug metabolite synthesis. *Microb. Cell Fact.* 14, 1–10 (2015).
- 55 55. Korsmeyer, K. K. et al. N-glycosylation of pig flavin-containing monooxygenase form 1: Determination of the site of protein modification by mass spectrometry. *Chem. Res. Toxicol.* 11, 1145–1153 (1998).
- 56 56. Wierenga, R. K., De Maeyer, M. C. H. & Hoi, W. G. J. Interaction of Pyrophosphate Moieties with  $\alpha$ -Helices in Dinucleotide Binding Proteins. *Biochemistry* 24, 1346–1357 (1985).
- 57 57. Orru, R., Torres Pazmiño, D. E., Fraaije, M. W. & Mattevi, A. Joint functions of protein residues and NADP(H) in oxygen activation by flavin-containing monooxygenase. *J. Biol. Chem.* 285, 35021–35028 (2010).
- 58 Lončar, N. et al. Characterization of a thermostable flavin-containing monooxygenase from *Nitrocola laciasaponensis* (NiFMO). *Appl. Microbiol. Biotechnol.* 103, 1755–1764 (2019).
- 59 Fürst, M. J., Fiorentini, F. & Fraaije, M. W. Beyond active site residues: overall structural dynamics control catalysis in flavin-containing and heme-containing monooxygenases. *Curr. Opin. Struct. Biol.* 59, 29–37 (2019).
- 60 Romero, E., Castellanos, J. R. G., Mattevi, A. & Fraaije, M. W. Characterization and Crystal Structure of a Robust Cyclohexanone Monooxygenase. *Angew. Chemie - Int. Ed.* 55, 15852–15855 (2016).
- 61 Torres Pazmiño, D. E., Baas, B. J., Janssen, D. B. & Fraaije, M. W. Kinetic mechanism of phenylacetone monooxygenase from *Thermobifida fusca*. *Biochemistry* 47, 4082–4093 (2008).
- 62 Criegee, R. Die Umlagerung der Dekalin-peroxydester als Folge von kationischem Sauerstoff. *Justus Liebigs Ann. Chem.* 560, 127–135 (1948).
- 63 Ho, B. K. & Gruswitz, F. HOLLOW: Generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* 8, 1–6 (2008).
- 64 Gumulya, Y. et al. Engineering highly functional thermostable proteins using ancestral sequence reconstruction. *Nat. Catal.* 1, 878–888 (2018).
- 65 Forneris, F., Orru, R., Bonivento, D., Chiarelli, L. R. & Mattevi, A. ThermoFAD, a ThermoFluor®-adapted flavin ad hoc detection system for protein folding and ligand binding. *FEBS J.* 276, 2833–2840 (2009).
- 66 Wheeler, L. C., Lim, S. A., Marqusee, S. & Harms, M. J. The thermostability and specificity of ancient proteins. *Curr. Opin. Struct. Biol.* 38, 37–43 (2016).
- 67 Lawton, M. P. & Philpot, R. M. Functional characterization of flavin-containing monooxygenase 1B1 expressed in *Saccharomyces cerevisiae* and *Escherichia coli* and analysis of proposed FAD- and membrane-binding domains. *J. Biol. Chem.* 268, 5728–5734 (1993).
- 68 Allen, K. N., Entova, S., Ray, L. C. & Imperiali, B. Monotopic Membrane Proteins Join the Fold. *Trends Biochem. Sci.* 44, 7–20 (2019).
- 69 Nagata, T., Williams, D. E. & Ziegler, D. M. Substrate Specificities of Rabbit Lung and Porcine Liver Flavin-Containing Monooxygenases: Differences due to Substrate Size. *Chem. Res. Toxicol.* 3, 372–376 (1990).
- 70 Young Mi Kim & Ziegler, D. M. Size limits of thiocarbamides accepted as substrates by human flavin-containing monooxygenase 1. *Drug Metab. Dispos.* 28, 1003–1006 (2000).
- 71 Beaty, N. B. & Ballou, D. P. Transient Kinetic Study of Liver. *J. Biol. Chem.* 255, 3817–3819 (1980).
- 72 Beaty, N. B. & Ballou, D. P. The Oxidative Half-reaction of Liver Microsomal FAD-containing Monooxygenase. *J. Biol. Chem.* 256, 4619–4625 (1981).
- 73 Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 1–7 (2017).

- 74 Lemoine, F. et al. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* 556, 452–456 (2018).
- 75 Hug, L. A. et al. A new view of the tree of life. *Nat. Microbiol.* 1, 1–6 (2016).
- 76 Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* 24, 1586–1591 (2007).
- 77 Siddiq, M. A., Loehlin, D. W., Montooth, K. L. & Thornton, J. W. Experimental test and refutation of a classic case of molecular adaptation in *Drosophila melanogaster*. *Nat. Ecol. Evol.* 1, 1–6 (2017).
- 78 Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* 26, 795–800 (1993).
- 79 Project, C. C. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 50, 760–763 (1994).
- 80 Tickle, I. J. et al. Staraniso. *Glob. Phasing Ltd* (2018).
- 81 Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* 69, 1204–1214 (2013).
- 82 Kubitzka, C. et al. Crystal structure of pyrrolizidine alkaloid N-oxygenase from the grasshopper *Zonocerus variegatus*. *Acta Crystallogr. Sect. D Struct. Biol.* 74, 422–432 (2018).
- 83 McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674 (2007).
- 84 Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 60, 2256–2268 (2004).
- 85 Jung, W. S., Singh, R. K., Lee, J. K. & Pan, C. H. Crystal Structure and Substrate Specificity of D-Galactose-6-Phosphate Isomerase Complexed with Substrates. *PLoS One* 8, 2–11 (2013).
- 86 Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 53, 240–255 (1997).
- 87 Pettersen, E. F. et al. UCSF Chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–12 (2004).

## Supplementary Information



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



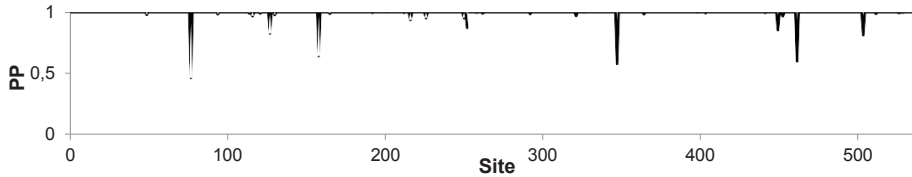


**Supplementary Figure 2. Vertebrates phylogeny of FMOs** (previous page). Tree was constructed in RAxML v0.6.0, 1000 bootstraps were run and best-fit model parameters were obtained with ProtTest v3.4. Bootstrap transfer was applied and TBE values are shown at the nodes. The employed MSA was trimmed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

>AncFM02

MAKKVAVIGAGVSGLSLKKCCVDEGLEPTCFERTEDIGGLWRFKENVEDGRASIQSVITNTSKEMSCFSDPFMPPEHFPN  
FLHNSKLEFYFRIFAKKFDLLKYIQFQTTVLSVKKCPDFSSSQWEIVTESNGKEQSAVFDAMVCSGHILPHIPLQSF  
PGIERFKGQYFHSRQYKHPEGFEGKRLIVIGIGNSASDIIVELSKKAAQVFI STRHGSWVMSRISDDGYPWDMVFHTRFS  
SMLRNVLPRTVVKMMEQMNRWFNHNENYGLEPQNKYLMKEPVLNDDLPSRLLYGAIKVKSRVKELTETSIFEDGTVEE  
DIDVIVFATGYTFSPFFLEDSLKVENNMVSLYKYMFPFHLEKPTLACIGLIQPLGSIFPTVELQARVWTRVFKGLCTLP  
SESTMADI IKRNEKRIDLFGESQSQILQNTNYIDYLBELALEIGAKPDLLSLLKDPKAMKLYFGPCNSYQYRLVGGQ  
WEGARNAIFTQQRILKPLKTRALKASSNFPVSVFLKILGLLAVVVAVFFQLQWF

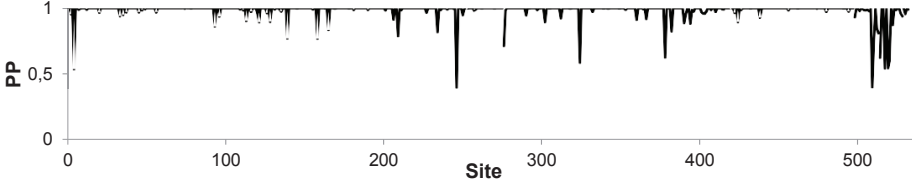
Average of reconstruction= 0.994



>AncFM03-6

MGKKVAIIGAGVSGLASIRSCLEEGLEPTCFERSDDIGGLWKFSDHAEGRASIQSVFTNSSKEMMCFDFPYDDDFPN  
FMHNSKLQYITAFAKEKNLLKYIQFKTLVSSVKNRPFVSTGQWDVTEKDGKESAVFDAMVICSGHHVYPNLPKESF  
PGLKHFKGKCFHSRDYKEPGIFKGRVVLVIGLNSGCDIATELSHTAEKVISSRSGSWVMSRVWDDGYPWDMFLTRFE  
TFLKNSLPATAISDWYMKQMNARFKHENYGLMPLNGTLRKEPVFNDELPARILCGTVSIPKPNVKEFTETSIFEDGTVEE  
AIDCVIFATGYGYAYPFLDDSI IKSRNNEVTLFKGI FPPLEKPTLAVIGLVQSLGATIPPTDLQARWAAKV FANSC TLP  
TTNEMDDI DEKMGGKLLWFGQSQTLDYITYMDELGSFIGAKPNIPWFLTDPQLALEVFFGPCSPYQYRFLMGPQKWD  
GARNAILTQWDRTLKPTRTRAVGEAKRPSLFYNLLKILLFPVLLAVLLAFY

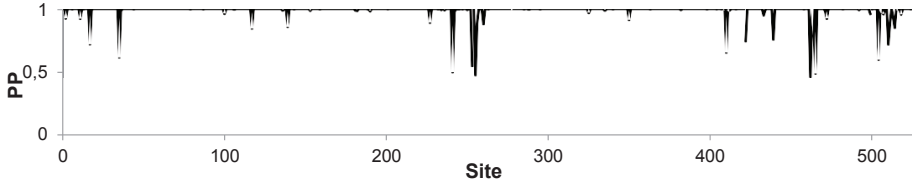
Average of reconstruction= 0.982



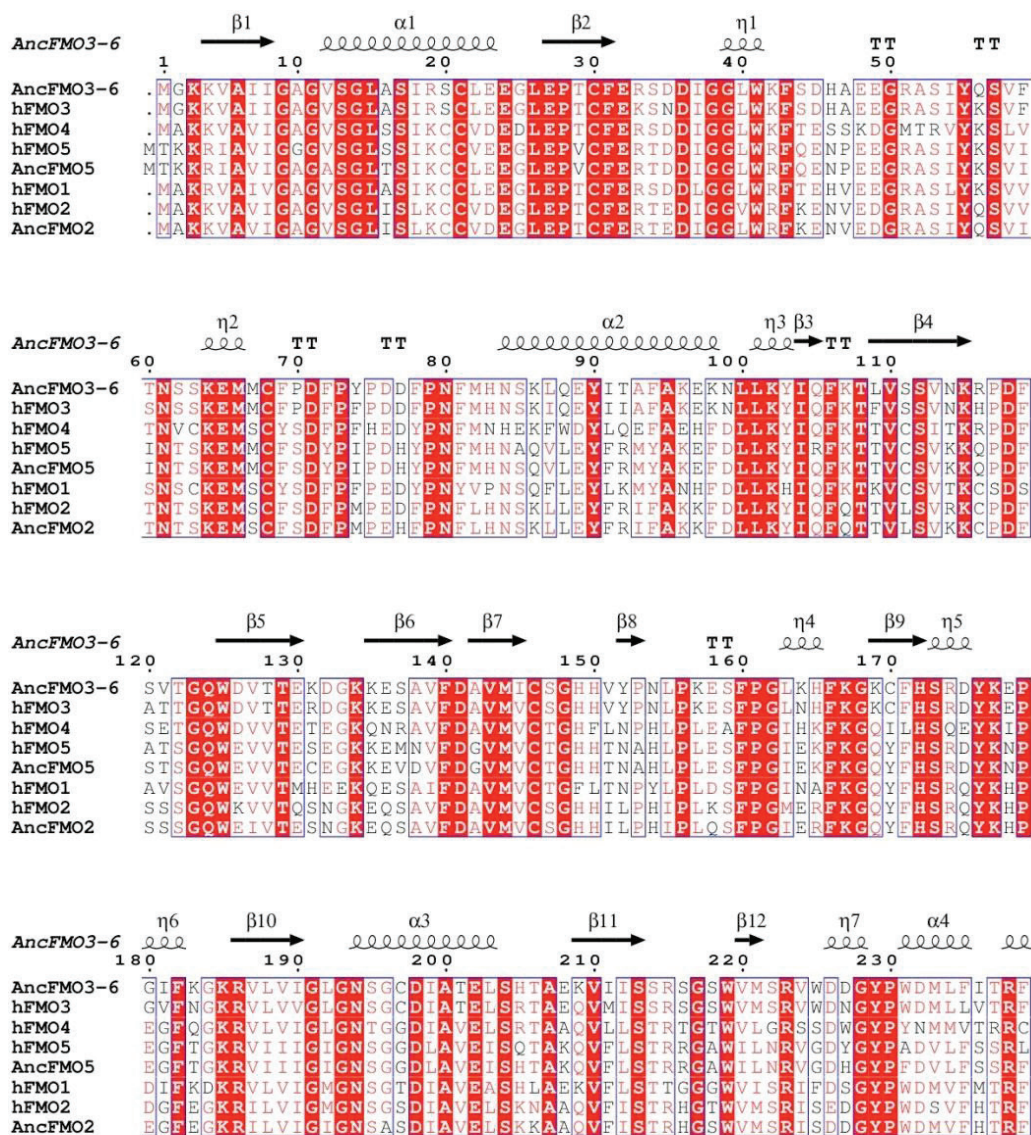
>AncFM05

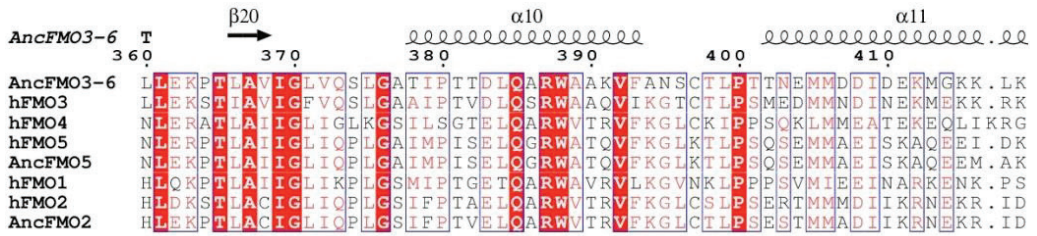
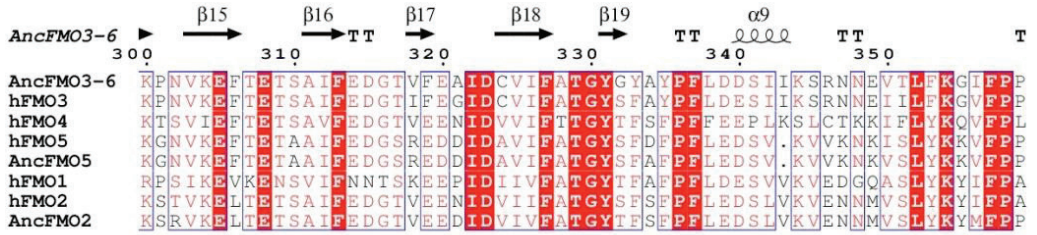
MTKKRIAVIGAGASGLTSIKCCLEEGLEPVCFERTDDIGGLWRVQENPEEGRASIYKSVIINTSKEMMCFSDYPIPDHYP  
NFMHNSQVLEYFRMYAKEFDLLKYIQFKTTVCSVKKQPDFSTSGQWEVVTECEGKKEVDVFDGVMVCTGHHHTNAHLP  
FPGIEKFKGQYFHSRDYKNPEGF TGRV I IIGIGNSGDLAVEISH TAKQVFLSTRRGAWILNVRVDHGYPFDVLFSSRF  
TYFLSKICGQSLNFTLEKKNRQFDHEMFGKPKHRALSQHPTVNDL PNR IISGLVVKGNVKEFTETAAIFEDGSRE  
DDIDAVIFATGYSFAFPFLEDSVKVVKNKVSLYKVFPPNLEKPTLAIIGLIQPLGAIMPISLQGRWATQVFKGLK TLP  
SQSEMMAEISKAQEEMAKRYVDSQRHTIQGDYIDTMEE IADLVGVRPNLLS LAFTDPK LALKLF GFGPCTPVQYRLQGP  
WDGARKTILTTEDRIRKPLMTRVIEKSNMSTMTMGRFMLAVVFFAII MAYF

Average of reconstruction= 0.9873



**Supplementary Figure 3. The amino acid sequence of the AncFMOs (previous page). For the three characterized ancestral proteins, sequences of the ML (maximum likelihood) ancestors are shown. Graphs display the posterior probabilities of each reconstructed site for AncFMO2, AncFMO3-6 and AncFMO5.**



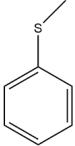
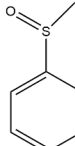
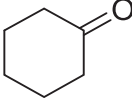
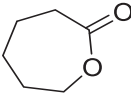


**AncFMO3-6**

AncFMO3-6 .....  
hFMO3 .....  
hFMO4 LKLVRDKLQDRMSPYLVS LWRG  
hFMO5 .....  
AncFMO5 .....  
hFMO1 .....  
hFMO2 WS.....  
AncFMO2 WF.....

**Supplementary Figure 4. Structure-based sequence alignment** (previous pages). The alignment was generated by using AncFMO3-6 pdb structure as input file and inserting the Ancestral FMO and hFMO sequence files. Generated with Esprit 3.0.

**Supplementary Table 1. Conversions by AncFMO2 and AncFMO3-6 variants. Cyclohexanone showed no conversion for all four enzymes (n.d. for not detected).**

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

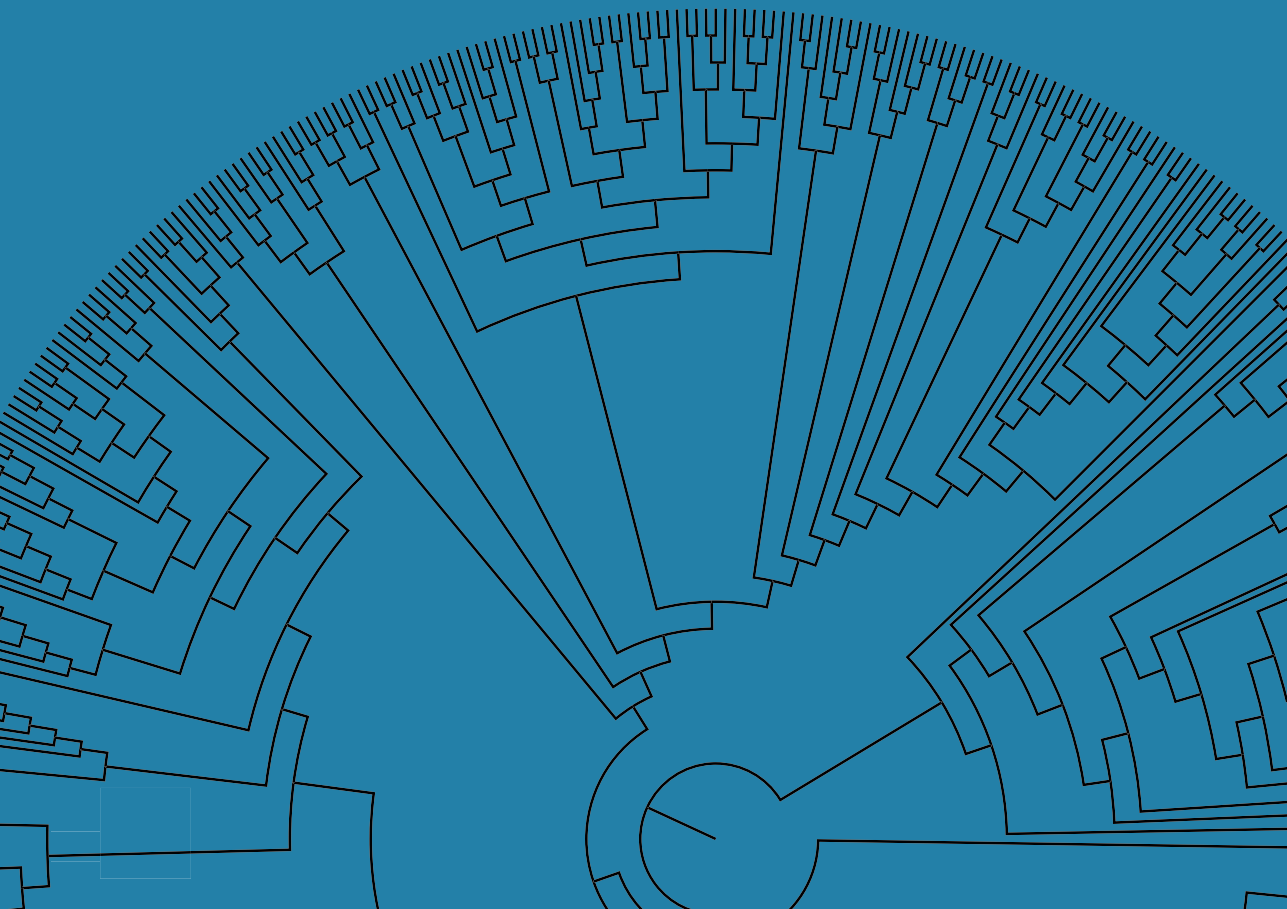
**Supplementary Table 2. Melting Temperatures of each AncFMO. Measured in triplicate by using the ThermoFAD technique (see materials and methods).**

Protein	NADP <sup>+</sup> (μM)	T <sub>m</sub> (°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

**Supplementary Table 3. Ambiguously reconstructed sites in AncFMOs.** Ambiguously reconstructed sites defined by the Alt state highest posterior probability (PP) > 0.2 are presented for each AncFMO.


AncFMO2					
Site	M	State	PP	Alt State	PP
158	L	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





## CHAPTER IV

# Ancestral reconstruction of mammalian FMO1 enables structural determination, revealing unique features that explain its catalytic properties



Gautier Bailleul  
Callum R. Nicoll  
María Laura Mascotti  
Andrea Mattevi  
Marco W. Fraaije.

This chapter is based on a published article: *J. Biol. Chem.* 296, 0–2 (2021).

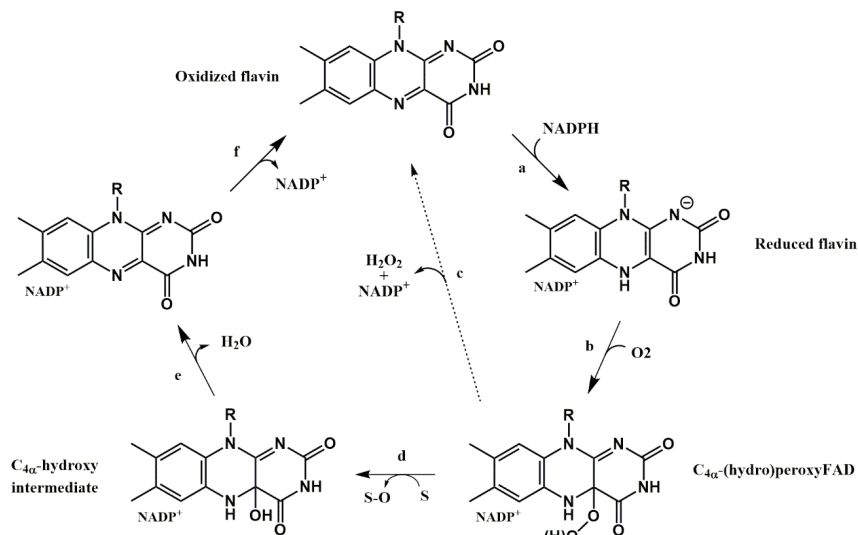
## Abstract

Mammals rely on the oxidative flavin-containing monooxygenases (FMOs) to detoxify numerous and potentially deleterious xenobiotics; this activity extends to many drugs, giving FMOs high pharmacological relevance. However, our knowledge regarding these membrane-bound enzymes has been greatly impeded by the lack of structural information. We anticipated that ancestral sequence reconstruction could help us identify protein sequences that are more amenable to structural analysis. As such, we hereby reconstructed the mammalian ancestral protein sequences of both FMO1 and FMO4, denoted as AncFMO1 and AncFMO4, respectively. AncFMO1, sharing 89.5 % sequence identity with human FMO1, was successfully expressed as functional enzyme. It displayed typical FMO activity as demonstrated by oxygenating benzydamine, tamoxifen and thioanisole, drug-related compounds known to be also accepted by human FMO1, and both NADH and NADPH cofactors could act as electron donors, a feature only described for the FMO1 paralogs. AncFMO1 crystallized as a dimer and was structurally resolved at 3.0 Å resolution. The structure harbors typical FMO aspects with the FAD and NAD(P)H binding domains and a C-terminal transmembrane helix. Intriguingly, AncFMO1 also contains some unique features, including a significantly porous and exposed active site, and NADPH adopting a new conformation with the 2'-phosphate being pushed inside the NADP<sup>+</sup> binding domain instead of being stretched out in the solvent. Overall, the ancestrally reconstructed mammalian AncFMO1 serves as the first structural model to corroborate and rationalize the catalytic properties of FMO1.

## Introduction

For dealing with endogenous and foreign toxic compounds, mammals and other animals developed oxidative systems that clear such potentially harmful elements from cells and tissues<sup>1</sup>. The oxygenation of these molecules allows them to be recognized, degraded, excreted or activated. In humans, this oxidative detoxification system is mostly based on the activity of cytochromes P450 monooxygenases (CYPs) and the flavin-containing monooxygenases (FMOs; EC 1.14.13.8). Significant advances have been made towards the characterization of human FMOs (hFMOs) but these enzymes still remain significantly less studied than CYPs. This may partly be due to the fact that CYPs are more numerous in the human proteome (57 CYPs vs 5 FMOs). Another factor that explains the limited insights into the functioning behind the membrane-bound hFMOs is the difficulty with which they are expressed as recombinant proteins and their challenging isolation. It is nonetheless well established that hFMOs play a crucial role in xenobiotic metabolism<sup>2,3</sup>. There are 5 FMO paralogs in the human genome (FMO1-5), all clustered on chromosome 1<sup>4,5</sup>. A sixth gene, called FMO6, is also present but is not translated and thus referred to as a pseudogene<sup>6</sup>. The expression of hFMOs varies across the developmental stage and tissues. hFMO3 and hFMO5 transcripts are predominant in the liver whereas hFMO2 is mostly found in the lung and hFMO1 in the kidney<sup>7,8</sup>. A good example of the pharmacological relevance of hFMOs is the role of FMO3 in the metabolic disorder called trimethylaminuria, also known as fish odor syndrome<sup>9</sup>. Mutations in the gene encoding hFMO3 result in an inactive form of this FMO, unable to metabolize trimethylamine. The accumulation of this food ingredient in the body results in a strong unpleasant odor which may lead to social isolation and other negative effects<sup>10</sup>.

FMOs present a tightly bound flavin adenine dinucleotide (FAD) as prosthetic group<sup>11,12</sup>. While there are many other flavin-containing monooxygenases that are not sequence-related to FMOs, this abbreviation has persisted in literature to be used to denote this specific class of detoxifying enzymes<sup>13</sup>. The biochemical properties of mammalian FMOs were first described using enzymes from pig liver microsomes by Ziegler & Poulson in 1970<sup>14</sup>. In two extensive studies, Beaty and Ballou unraveled the catalytic mechanism involving the formation of a C4 $\alpha$ -hydroperoxyflavin intermediate that performs the oxygenation<sup>15,16</sup>. Formation of this reactive enzyme intermediate is achieved by a stepwise process in which first the FAD is reduced by a hydride transfer from the nicotinamide adenine dinucleotide phosphate (NADPH) (Figure 1a). With the oxidized nicotinamide still bound, the reduced flavin is able to react with molecular oxygen resulting in a relatively stable C4 $\alpha$ -hydroperoxyflavin intermediate (Figure 1b). In the next step, when substrate approaches the flavin, the peroxy moiety transfers an oxygen atom to a soft-nucleophilic substrate (Figure 1d) while the other oxygen atom is released as water (Figure 1e). The last step is the release of NADP<sup>+</sup> (Figure 1f)<sup>17</sup>. If there is no suitable substrate bound close enough to the flavin cofactor, hydrogen peroxide is formed which is commonly referred to as uncoupling (Figure 1c).



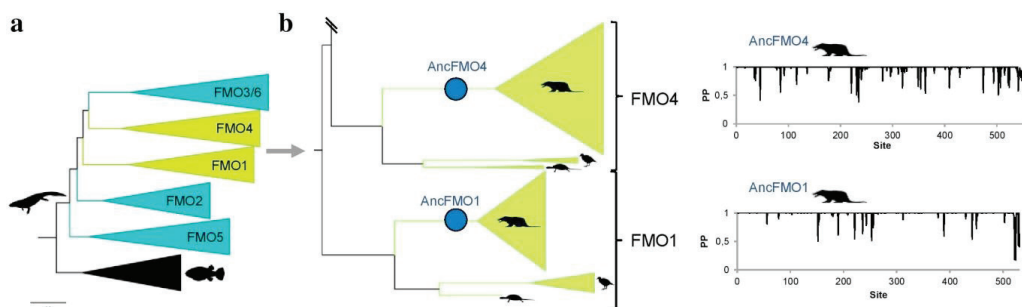
**Figure 1.** Catalytic cycle of flavin-containing monooxygenases. a, binding of NADPH and subsequent reduction of the flavin. b, Reaction with molecular oxygen and C<sub>4</sub>α-hydroperoxyflavin intermediate formation. c, Uncoupling: release of hydrogen peroxide and NADP<sup>+</sup>. d, Oxygen atom transfer to a substrate. e, Release of water. f, Release of NADP<sup>+</sup>.






The burst of genomic information in the last decades plus the ease of gene synthesis has enabled ancestral sequence reconstruction (ASR) studies. ASR consists of identifying the path of changes that occur during the evolution of a specific enzyme family through the inference of the ancestors' sequences<sup>18</sup>. By expressing and isolating ancestrally reconstructed proteins, it is possible to follow their biochemical features or structural modifications over time in comparison to extant enzymes<sup>19,20</sup>. ASR has many applications for enzymologists, from the characterization of ancestral enzymes' promiscuity or stability to the deciphering of evolutionary paths by gene duplications or the refinement of enzyme complexes<sup>21,22</sup>. In our previous study, using ASR we managed to express, purify and crystallize the mammalian ancestors of FMO2, FMO3 and FMO5<sup>23</sup>. This resulted in detailed structural insights of these resurrected FMOs. In this work, we aimed at reconstructing the ancestral mammalian FMO1 and FMO4 proteins. Surprisingly, FMO1 is one of the less studied hFMOs despite the fact that it was used as model FMO in the pioneering studies by Beaty and Ballou<sup>15,16</sup>. Critically, it has been established that hFMO1 displays a broad range of substrates, accepting drugs like benzydamine, ethionamide, tamoxifen or voriconazole<sup>2</sup>. We present for the first time the crystal structure of AncFMO1 together with its catalytic and kinetic features. These data shed light on the different enzymatic behavior of mammalian FMOs and will improve our understanding of their physiological roles.

## Results

### Ancestral sequence reconstruction

Sequence analysis has shown that the FMO paralogs emerged at the time of tetrapod evolution and all five FMO paralogs were encoded in the genome of the mammalian ancestor, 177 mya<sup>24</sup> (Figure 2, Figure S1, Table S1). AncFMO1 was reconstructed with high confidence (overall posterior probability of 0.95). This sequence shows the typical length exhibited by other mammalian FMOs (531 amino acids) and shows 89.8% sequence identity (54 residue changes) to hFMO1. We found that 10 sites were ambiguously reconstructed (Data S1). AncFMO4 was also reconstructed with high posterior probability (0.97) with 24 ambiguously reconstructed sites (Dataset S1). This enzyme exhibits a C-terminal extension of approximately 20 amino acids and its length was 559 amino acids. The sequence identity compared to hFMO4 was 89.8% (57 changes).



**Figure 2. Ancestral sequence reconstruction of mammalian FMOs.** *a*, Compressed tree of FMOs from jawed vertebrates. The tree depicts the explosion of FMOs in tetrapods. The five FMO clades are shown and those colored in green have been analyzed in this work. *b*, Close up of the FMO4 and FMO1 clades. Taxonomic distribution is depicted with silhouettes as follows: ancestral tetrapod (  ), bony fishes (  ), mammals (  ), aves (  ) and testudines (  ). Reconstructed mammalian ancestors are shown with blue circles. On the right the corresponding graphs of the posterior probability distribution per site are shown.

Full length AncFMO1 and AncFMO4 were selected for experimental characterization. Synthetic genes encoding proteins with the second-best state at each of the ambiguously reconstructed sites (alternative amino acids) were also obtained to produce the alternative versions (AltAncFMOs) and assess the robustness of the reconstruction. No expression of AncFMO4 or AltAncFMO4 was observed after growth at 24 °C or 17 °C. Lack of expression was confirmed through Western-blot analysis by utilizing a 6xHis-tag fused peroxidase monoclonal primary antibody. The cell lysate and insoluble membrane pellets incubated with a range of detergents failed to show protein expression (Figure

S2). The characteristic C-terminal extension after the transmembrane domain apparently prevents its proper expression in agreement with previous literature<sup>25</sup>. By contrast, up to 20 mg per liter of culture of AncFMO1 and AltAncFMO1 could be purified as FAD-containing proteins as evidenced by its intense yellow color. Furthermore, through Western-blot analysis, AncFMO1 was demonstrated to be extracted from the membrane fraction using a range of different detergents (Figure S2). The melting temperature of AncFMO1 was evaluated using a TychoTMNT.6 system to assess its stability (see Experimental Materials). The AncFMOs that we studied previously displayed high thermal robustness (melting temperatures of the native enzyme ranged between 53–60 °C). AncFMO1 instead exhibited a relatively low melting temperature of 47 °C that increased to 52 °C in the presence of 200  $\mu$ M NADP<sup>+</sup> (Table S2).

### ***Catalytic properties***

We first verified whether AncFMO1 was functional by testing known substrates of hFMOs. Gratifyingly, we found that AncFMO1 is indeed active on benzydamine, an anti-inflammatory drug 26, exhibiting typical Michaelis-Menten kinetic behavior (Table 1, Figure S3). A  $k_{cat}$  of 0.18  $s^{-1}$  was obtained which is very similar to what was previously determined for FMO1 from the long-tailed macaque (*Cynomolgus macaque*) (0.17  $s^{-1}$ )<sup>27</sup>. AncFMO1 was also found to accept the prototypical FMO substrate, thioanisole (0.19  $s^{-1}$ ). The alternative FMO1 ancestor, AltAncFMO1, displayed a similar activity toward benzydamine and thioanisole with a rate of 0.20  $s^{-1}$  and 0.16  $s^{-1}$  in presence of 0.4 mM substrate, respectively, thus demonstrating the robustness of the reconstruction. The nicotinamide cofactor specificity of AncFMO1 was also explored. The kinetic analysis revealed that both the nicotinamide cofactors, NADH and NADPH, are accepted as electron donors. While similar rates of catalysis were obtained, the  $K_M$  for NADPH was significantly lower than the  $K_M$  for NADH (Table 1). In the absence of any organic substrate, AncFMO1 displayed a similar uncoupling (*i.e.* the oxidation of NAD(P)H leading to the release of hydrogen peroxide) rate regardless of the cofactor, 0.13 and 0.11  $s^{-1}$  for NADPH and NADH, respectively. Once again, the  $K_M$  for NADH was higher than that for NADPH (Table 1). It should be noted that also hFMO1 was shown to be higher amounts compared to hFMO2 and hFMO3<sup>28</sup>.

**Table 1. Steady state kinetic parameters for AncFMO1.** <sup>a</sup>Reactions were followed by measuring the NADPH consumption at 340 nm and rates calculated using NADPH extinction coefficient 6.22 mM<sup>-1</sup>. cm<sup>-1</sup>. <sup>b</sup>NAD(P)H steady-state kinetics in presence of 1.0 mM thioanisole, rates followed at 340 nm. <sup>c</sup>AncFMO1 uncoupling reaction with NAD(P)H, in absence of substrate. The values in brackets refer to standard deviations.

substrate	fixed substrate	$k_{cat}$ (s <sup>-1</sup> )	$K_M$ (μM)	$k_{cat}/K_M$ (s <sup>-1</sup> /M)
<sup>a</sup> benzylamine	0.10 mM NADPH	0.18 (0.004)	< 13	> 14000
<sup>a</sup> thioanisole	0.10 mM NADPH	0.19 (0.007)	< 25	> 7600
<sup>b</sup> NADPH	1.0 mM thioanisole	0.26 (0.009)	< 9	> 29000
<sup>b</sup> NADH	1.0 mM thioanisole	0.53 (0.03)	54 (8)	10000
<sup>c</sup> NADPH	-	0.13 (0.01)	55 (13)	2400
<sup>c</sup> NADH	-	0.11 (0.009)	91 (17)	1200

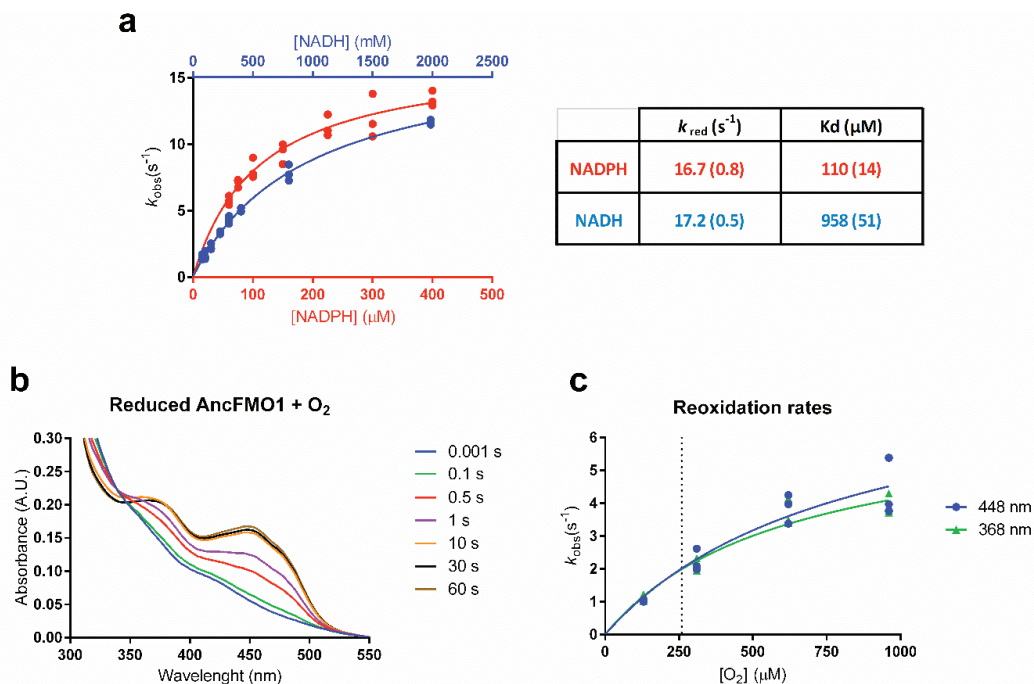
Having established that AncFMO1 is enzymatically active, we sought to further characterize the enzymatic conversions and the resulting products (Table S3). Benzylamine is converted to didesmethyl-benzylamine by CYPs whereas hFMO1 transforms it into the corresponding N-oxide<sup>26,29</sup>. Using HPLC analysis, we found that also AncFMO1 converts benzylamine into its N-oxide. Thio-anisole was converted into the corresponding sulfoxide with high enantioselectivity, producing mainly the (*R*)-enantiomer. To probe the activity of AncFMO1 towards a bulky molecule, tamoxifen (an anti-estrogen drug **30**) was tested as substrate. We found that tamoxifen was converted into the N-oxide derivative, as previously reported for hFMO1<sup>31</sup>. To test whether in addition to N- and S-oxidations AncFMO1 catalyzes also Baeyer-Villiger oxidations, two ketones were assessed as potential substrates; hepta-2-one and bicyclo[3.2.0]hept-2-en-6-one<sup>32,33</sup>. While no conversion of hepta-2-one was observed, bicyclo[3.2.0]hept-2-en-6-one was partly converted, implying some moderate Baeyer-Villiger oxidation activity. Upon harvesting the cells, it became clear that a blue pigment had been formed as evidenced by colored cell pellets upon centrifugation. For microbial FMOs, it has been shown that they form indigo blue when expressed in *E. coli*<sup>34</sup>. Thus, it is most likely that AncFMO1 converts indole into indoxyl which spontaneously dimerizes to form indigo blue. To verify this, the purified enzyme was incubated for 1 hour at 30°C with indole and a NADPH regeneration system, which indeed resulted in formation of indigo (Figure S4). Collectively, these data demonstrated that AncFMO1 featured all the typical properties of FMOs and was able to convert known FMO1 substrates.



### **Rapid kinetics**

Both the reductive and oxidative half reactions were investigated separately using the stopped-flow technique. The first steps of the catalytic cycle of FMOs involve the binding of NADPH after which the FAD is reduced through a direct hydride transfer. The reductive half reaction can be monitored spectro-photometrically by mixing enzyme with NADPH under anoxic conditions since the reduction of the flavin can be easily measured by following the decrease of absorbance at 448 nm. We determined the reduction rates at different NADPH concentrations (Figure 3a). Fitting the data resulted in a relatively fast reduction rate constant with  $k_{\text{red}} = 16.7 \text{ s}^{-1}$  and a dissociation constant of 110  $\mu\text{M}$ . The kinetic and spectral data indicate that flavin reduction takes place in a single irreversible hydride transfer step. We also monitored the reductive half reaction using NADH, resulting in a similar reduction rate constant ( $k_{\text{red}} = 17.2 \text{ s}^{-1}$ ) and a dissociation constant of 958  $\mu\text{M}$ , which is almost 9 times higher compared to NADPH. The presence of 1.0 mM thioanisole did not affect the rate of reduction suggesting that flavin reduction precedes binding of substrate as has been proposed for other class B flavoprotein monooxygenases<sup>13</sup>.

By preparing anaerobically reduced AncFMO1, the oxidative half reaction could be monitored by mixing it with dioxygen-containing buffer. For several FMOs, the reaction of reduced enzyme with dioxygen resulted in the formation of a C4 $\alpha$ -hydroperoxyflavin intermediate with absorbance features around 360-380 nm<sup>35</sup>. The intermediate subsequently decayed at varying rates depending on its stabilization by the enzyme. For AncFMO1, when mixed as reduced enzyme with dioxygen, a rapid increase of absorbance at both 368 nm and 448 nm was observed and we were unable to distinguish intermediate formation from the full reoxidation of the flavin (Figure 3b). This shows that there was only a minor build-up of the C4 $\alpha$ -hydroperoxyflavin intermediate and indicates that AncFMO1 is not good at stabilizing the oxygenating enzyme intermediate. This observation is consistent with the sustained uncoupling featured by AncFMO1 (Table 1). The rate of reoxidation of the reduced AncFMO1 in presence of 260  $\mu\text{M}$  dioxygen was around  $2 \text{ s}^{-1}$  (Figure 3c) while the steady-state kinetics of uncoupling suggested a slower reaction with a rate only reaching  $0.12 \text{ s}^{-1}$  (Table 1). These kinetic data show that AncFMO1 displays a similar kinetic behavior when compared with other FMOs.

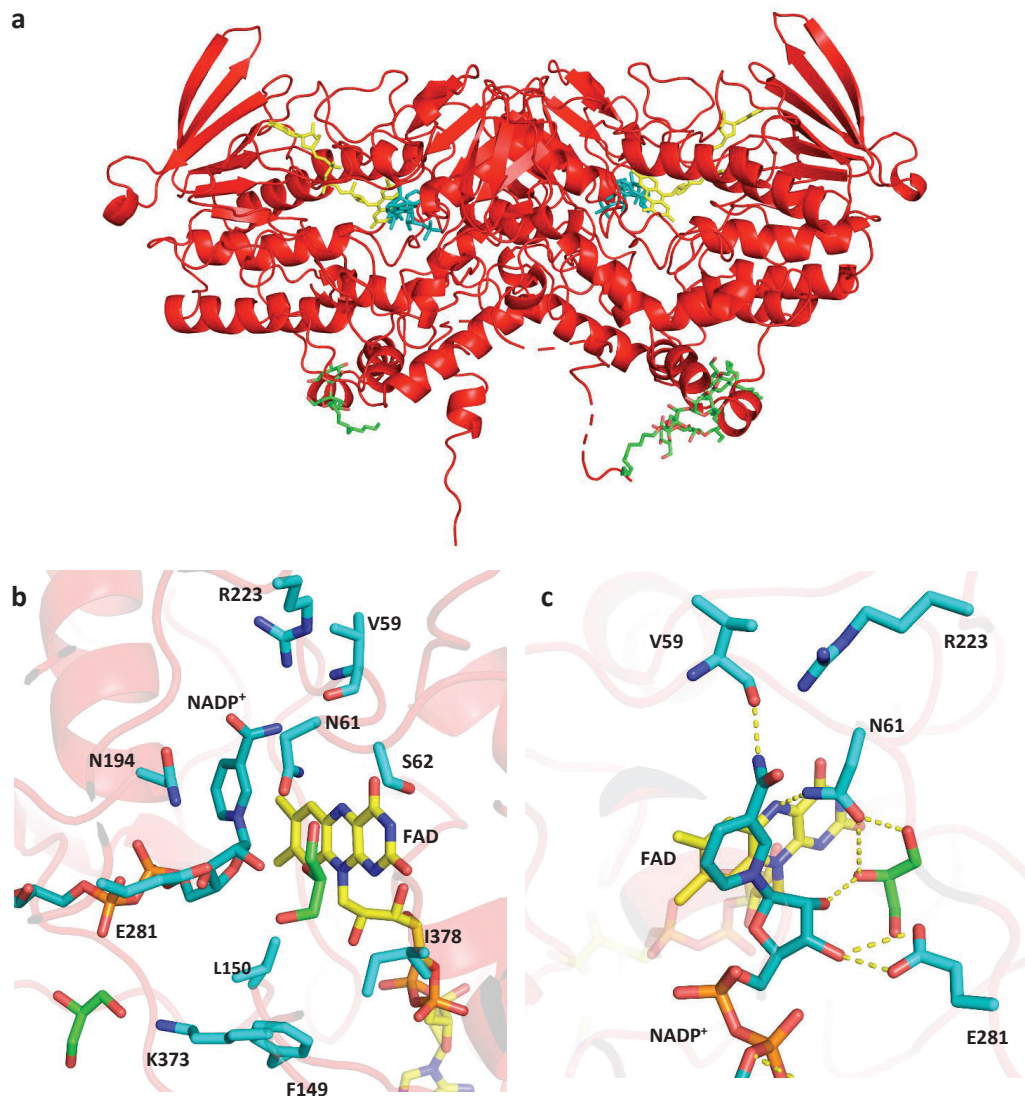


**Figure 3. AncFMO1 stopped-flow kinetics.** *a*, Reductive half reaction rates measured in technical triplicates under anaerobic conditions with increasing concentrations of NADH (blue points) or NADPH (red points). The binding constants  $K_d$  and  $k_{red}$  values were calculated by fitting the reductive half-reaction to the Michaelis-Menten equation with standard deviation reported in brackets. *b*, Reoxidation spectra of the reduced AncFMO1 with oxygenated buffer over time. *c*, Reoxidation rates with increasing concentration of dioxygen, observed at 368 nm (green triangles) and 448 nm (blue points) in technical triplicates. The dotted lines correspond to the atmospheric concentration of dioxygen (260  $\mu M$ ), present during steady state kinetics.

### Overall three-dimensional structure of FMO1

We were able to successfully crystallize AncFMO1 and to solve its structure at 3.0 Å resolution (Figure 4a, Table 2, Figure S5). STARANISO, an anisotropy correction server, was critical to correct for the anisotropy of crystals leading to drastically improved electron density maps. The asymmetric unit consisted of two dimers (Table 2) and the dimerization observed for AncFMO1 is the same as for the dimers documented by the other AncFMO crystal structures<sup>23</sup> (Figure 4a). Consistently, the structure of AncFMO1 possesses key structural domains and folds canonical to the mammalian FMOs (Figure 4a). Alongside the well conserved NAD(P)H and FAD binding domains, known as the paired Rossmann fold, AncFMO1 includes a large 80-residue insertion, two small

$\alpha$ -helical triads that form the hydrophobic ridge that embeds into the phospholipid bilayer, and a large transmembrane C-terminal helix that drills into the membrane. Additionally, in accordance with the previously described AncFMOs, AncFMO1 displayed large hydrophobic strips on the surface of the crystal structure that promote monotopic membrane association (Figure S6) (22). The oxidized coenzyme, NADP<sup>+</sup>, was successfully bound in the crystal structure and two glycerol molecules were found located near/at the active site. Due to weak electron density and a highly disordered peptide-backbone, the C-terminal helices were poorly constructed in the crystal structure. Nevertheless, resolved detergent molecules demarcated the protein-membrane interface that was illustrated by other AncFMOs<sup>23</sup>. In addition, pairwise structural superpositions between AncFMO1 and AncFMO3-6 monomers (sequence identity 59%) portrayed a relative-mean square deviation (RMSD) of 0.8 Å<sup>2</sup> over 449 C $\alpha$  atoms, corroborating a high protein scaffold similarity. The crystal structure revealed that the ambiguous sites of the AltAncFMO1 sequence were all peripheral from the active site and not expected to impact enzymatic activity. This is supported by the enzyme activities observed for AltAncFMO1.



**Figure 4.** Crystal structure of AncFMO1 and its active site. *a*, Dimeric AncFMO1 with its partially mapped C-terminal helices pointing downwards towards the membrane. *b*, The active site of AncFMO1 in the presence of oxidized coenzyme, NADP<sup>+</sup> and two glycerol molecules (green) is depicted, with the key residues labelled. *c*, Extensive hydrogen bond interactions between E281, N61, V59, FAD, NADP<sup>+</sup> and a glycerol molecule are shown to illustrate key intermolecular interactions and potential substrate binding modes, represented by dashed yellow lines. FAD, NADP<sup>+</sup> and DDM molecules are colored in yellow, cornflower blue and white respectively.

**Table 2. Data collection and refinement statistics.** The final dataset was measured by combining two previously merged reflection files, derived from two crystals, which were then corrected for anisotropy using Staraniso<sup>47</sup>. The refined data are derived from the anisotropic corrected data. Values in parentheses are for highest-resolution shell. The asymmetric unit contained 4 protein molecules for AncFMO1.

	AncFMO1 (PDB: 7AL4)			
Number of crystals	2			
	Aimless		Staraniso	
<b>Data collection<sup>a</sup></b>				
Resolution Range (Å) <sup>b</sup>	49.100 - 3.000 (3.041 - 3.000)			
Space Group	P 2 <sub>1</sub>			
Unit Cell (Å), (°)	115.92	92.453	156.69	90 95.122 90
Total Reflections	437615 (22096)		-	
Unique Reflections	66237 (3899)		39983 (567)	
Multiplicity	6.6 (5.0)		-	
Completeness (%)	99.7 (98.8)		60.2 (14.1)	
Completeness elipsoidal (%)	-		99.5 (94.8)	
Mean I/sigma (I)	3.4 (0.3)		5.5 (1.1)	
R <sub>merge</sub>	0.245 (7.993)		-	
CC <sub>1/2</sub>	0.994 (0.028)		-	
<b>Refinement</b>				
R-work (%)	-		22.8	
R-free (%)	-		26.4	
Number of non-hydrogen atoms <sup>c</sup>	-		17079	
RMS (bonds) Å	-		0.004	
RMS (angles) °	-		1.359	
Ramachandran favoured (%)	-		82.6	
Ramachandran allowed (%)	-		11.2	
Ramachandran outliers (%)	-		6.2	
Average B-factor	-		82	

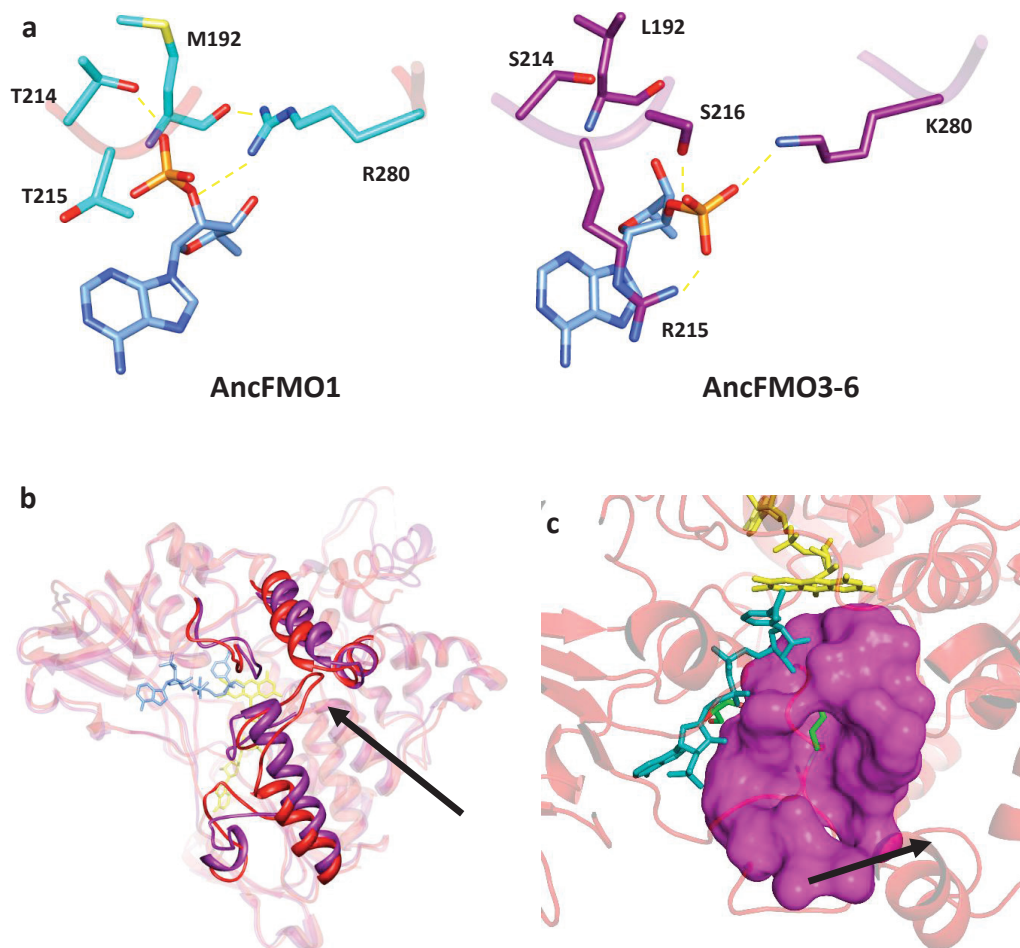
<sup>a</sup>The dataset was measured after merging reflection files from two crystals. <sup>b</sup>Values in parentheses are for highest-resolution shell. <sup>c</sup>The asymmetric unit contained four protein molecules respectively.

### Distinct features of the FMO1 active site

To gain insight into the degree of structural conservation among the mammalian FMOs, we inspected the active site of AncFMO1. Firstly, the underlying architecture that encases the isoalloxazine ring of the FAD and anchors the nicotinamide ring of NAD(P)H, is identical to the AncFMOs (Figure 4b, c). Furthermore, the active site residues are well conserved among the FMOs and occupy similar conformations. Moreover, apart from Ile378 that is a methionine in hFMO1 (Figure S7), all active site residues are conserved between AncFMO1 and hFMO1 making AncFMO1 a reliable model for the human counterpart.

Despite the likeness AncFMO1 shares with the other AncFMOs, the three-dimensional structure reveals some clear differences that render this enzyme somewhat distinct from its paralogs. Residues 149 and 150 which sit close to the ribityl tail of the isoalloxazine ring, are distinct from other FMOs (Figure 4b, c). The previously described AncFMO crystal structures possessed two conserved histidine residues at these positions. His150 was shown to form a hydrogen bond with the highly conserved Gln373, which was observed to be also within hydrogen bond distance to the 3'-OH group of NADP<sup>+</sup>. Together, these two hydrogen bonds formed a key steric blockade that maintained the closed cavity feature that was a characteristic of mammalian FMOs<sup>23</sup>. AncFMO1 possesses a phenylalanine and a leucine at positions 149 and 150 creating a hydrophobic and wider pocket at the base of the active site. Furthermore, with Leu150 not being able to create hydrogen bonds, Lys373 is not tied down and can extend outwards, towards a glycerol molecule situated below the diphosphate moiety of the NADP<sup>+</sup>, thereby further enlarging the active site (Figure 4b). With these residues being conserved between AncFMO1 and hFMO1, it is likely that these features would also be observed for hFMO1, therefore further validating the use of AncFMO1 as a structural model.

Commensurately with AncFMO1, hFMO1 has been previously described to use both NADPH and NADH as reducing coenzymes for oxidative catalysis<sup>36</sup>. With the only difference being that of the 2'-phosphate, the coenzyme-protein interactions in the crystal structure were inspected. Intriguingly, the 2'-phosphate occupies a new conformation unique to AncFMO1 (Figure 5a). Whilst the other AncFMOs crystal structures portray the 2'-phosphate to be extended out into the solvent, in the AncFMO1 crystal structure, the 2'-phosphate is tucked inside the NADP<sup>+</sup> binding domain. Typically, a highly conserved arginine residue (site 215 for AncFMO2 and AncFMO3-6) is observed to be within the vicinity of the 2'-phosphate binding-pocket. A range of intermolecular interactions including hydrogen bonds and ionic interactions are then established between the guanidinium headgroup of the arginine side chain and the 2'-phosphate of the NADP<sup>+</sup> molecule<sup>23</sup>. AncFMO1, however, possesses a threonine at position 215 and a glycine at 216 (His, Ser and Arg for AncFMO2, 3-6 and 5, respectively) which are unable to produce these key non-covalent interactions, resulting in the 2'-phosphate rotating upwards in order to establish contacts. More specifically, the 2'-phosphate establishes hydrogen bonds with the sidechains of Arg280 and Thr214, and the carbonyl peptide of Met192 (Figure 5a). Apart from Arg280 that is a chemically equivalent lysine in hFMO1, these residues are conserved between AncFMO1 and hFMO1, therefore it is likely that the coenzyme would bind in the same conformation. The loss of key residues in the 2'-phosphate binding pocket likely contributes towards the abolishment of the coenzyme selectivity.



**Figure 5. Unique structural features of AncFMO1.** *a*, The differing 2'-phosphate binding site of NADP<sup>+</sup> for AncFMO1 (left, side chains in cyan) compared to AncFMO3-6 (PDB:6SE3) (right, side chains in dark purple) is shown with key hydrogen bonding interactions shown as yellow dashed lines. NADP<sup>+</sup> is shown in cornflower blue. *b*, The conformation adopted by residues 416-425 that reaches out towards the  $\alpha$ -helical triad in a large arched conformation is indicated by the black arrow. Superpositioning AncFMO1 (red) against AncFMO3-6 (dark purple) conveys the new structural topology. *c*, The large active site cavity (approximately 19 Å wide) is depicted in magenta and stretches out towards the solvent and the membrane-protein interface (indicated with a black arrow). The side chain of E281 is shown in green pointing towards the isoalloxazine ring and the active site. FAD, NADP<sup>+</sup> and glycerol molecules are shown in yellow, cornflower blue and green, respectively.

### ***AncFMO1 contains a porous and exposed active site***

FMO1 has been described in literature to be the least selective of the FMO family, metabolizing a range of substrates such as tamoxifen, benzydamine, trifluoperazine and sulindac sulphide<sup>37</sup>, and unlike FMO2, is not limited or restricted to substrate size or length<sup>38</sup>. This finding substantiates that the active site of FMO1 is more susceptible and vulnerable to substrate exposure. To corroborate this speculation we inspected the tunnels and cavities of AncFMO1 using HOLLOW, a program that facilitates the internal and external surface imaging of proteins<sup>39</sup>. AncFMO1 exhibits a narrow tunnel that passes through the  $\alpha$ -helical triad towards the membrane corroborating that this FMO has the capacity to extract substrates from the membrane in a manner consistent with its paralogs. Nonetheless, unlike the other characterized crystal structures that conveyed closed active site cavities, AncFMO1 exhibits a significantly porous active site. First of all, the above-described loss of a key hydrogen bond between residues 150 and 373, results in a *leaky* active site (Figure 5b, c). Moreover, the structural topology choreographed by residues 416-425 provides additional space for soluble substrates (Figure 5b). Two hydrophobic residues, Phe420 and Leu422, protrude upwards either side of the alpha helical triad and are anchored in place through a hydrogen bond interaction between the peptide backbone and the side chain of Thr249 (Figure S8). As a result, this new fold creates a steric blockade underneath the  $\alpha$ -helical triad, greatly restricting the ease of substrate access from the membrane. Despite the motif blocking a hydrophobic access point underneath the alpha helical triad, it could be conceived that the vicinity of this newly formed cavity to the protein-membrane interface (and the flexibility of the loop) may be able to siphon substrates dwelling in the membrane (Figure 4b, c). Moreover, this new conformation also greatly opens up the protein surface to the solvent and its flexibility and proximity to the protein-membrane interface may provide new entry points for substrate transit.

## **Discussion**

The first mammals on Earth were armed with the arsenal of five FMOs to deal with harmful chemicals generated endogenously or from the natural environment. This group of detoxifying enzymes has been maintained across evolution in all species in the order, including humans. In our previous work we characterized in depth the ancestors of human FMO2, FMO3 and FMO5 (Chapter 3). However, FMO1 and FMO4 remained uncharacterized until now. In this study, we performed ancestral sequence reconstruction of FMO1 and FMO4. While we could not express the ancestral form of FMO4, the AncFMO1 was successfully overexpressed and could be purified as stable and functional enzyme. This allowed us to study, for the first time, structural details of this mammalian FMO and correlate the structural features with its catalytic properties. Mammalian FMO1s have been shown to metabolize a wide variety of compounds, including various drug molecules. The conversion and steady-state kinetics conducted with benzydamine, tamoxifen and thioanisole, revealed that AncFMO1 accepts the same substrates as



other mammalian FMOs with a similar rate and affinity<sup>27,36,40</sup>. We therefore confirmed that the reconstructed enzyme is not troublesome to work with and can be used as a proper hFMO1 mimic with 89.5% sequence identity and almost full conservation of active site residues. Overall, the high sequence identity and the similarity in enzymatic catalysis between AncFMO1 and hFMO1, validates the use of AncFMO1 as a trustworthy structural surrogate for enzymatic assays and structure-based drug design. Also, this is the first multi-aspect characterization of a mammalian FMO1.

The active site of AncFMO1 consists of key residues required for intermediate stabilization and coenzyme binding, including Arg223, Asn194 and Asn61. Serendipitously, the presence of a glycerol molecule in the active site highlights key hydrogen bond interactions that could form during catalysis between the generated C4 $\alpha$ -hydroperoxy flavin intermediate and a theoretical substrate. The distal hydroperoxy oxygen atom would engage with the oncoming soft nucleophile's sp<sup>3</sup>-hybridised orbital, here represented as the central hydroxyl group of the glycerol molecule (Figure 4b, c). Additionally, the 2'-OH group of the nicotinamide ribose of NADP<sup>+</sup>, that was speculated<sup>41,42</sup> to stabilize the oxygenating intermediate, establishes a hydrogen bond with the same central hydroxy group of the glycerol. It could be assumed that the 2'-OH group of the nicotinamide ribose provides additional ligand binding roles such as orientating substrates for catalysis and positioning the nucleophilic centre of the substrate for oxidation. The expanded volume of the inner chamber of the active site of FMO1 appears to be relatively inefficient at C4 $\alpha$ -hydroperoxy flavin stabilization, thereby suffering from relatively high uncoupling. Structural inspection of the NADP<sup>+</sup> binding site inferred that AncFMO1 lacks multiple residues that have been illustrated in multiple nucleotide-binding enzymes to form contacts with the 2'-phosphate of NADP<sup>+</sup>. This deprivation promotes a conformational change with the 2'-phosphate relocating inwards towards the NAD(P)H binding domain. Extensive work carried out by Dean and co-workers<sup>43</sup>, on the evolutionary adaptation of isocitrate dehydrogenases for NAD<sup>+</sup> over NADP<sup>+</sup>, demonstrated that removing a similar overhanging arginine residue that hydrogen bonds with the 2'-phosphate contributed towards NADH preference. Remarkably, despite AncFMO1 losing multiple key phosphate binding partners in the NAD(P)H binding domain it is still able to utilize NADPH as a coenzyme. This fact clearly portrays that the 2'-phosphate contacts are not universal. In other words, the structural determinants of the cofactor acceptance are intrinsic to the nature of each enzyme family. During evolution the FMO2, FMO3 and FMO5 paralogs developed, or retained, an exclusive specificity for NADPH. FMO1 might instead have developed a less stringent specificity, or perhaps retained an ancestral dual cofactor usage. This leaves room for speculation with the differential expression of FMO1 at developmental stages in humans<sup>44</sup>. Its ability to use both nicotinamide cofactors renders FMO1 a potent detoxifying enzyme by being able to use both pools of reducing cofactors. Similarly to multiple aqueous xenobiotic degrading enzymes, AncFMO1 displayed a large solvent accessible cavity with multiple entry points. The generation of this chamber is due to two structural features. Firstly, the residue changes at the base of the active site (Leu150, Phe149 and Lys373,) create a hole in the vestibule and a *leaky* cavity.

This is in turn emphasized by a large loop (residues 416-425) that further expands the substrate entry point. Collectively, these adaptations promote the transit of soluble compounds, suggesting FMO1 may have evolved to show a greater selectivity towards aqueous substrates. These elements and the crystal structure provide rationale behind the breadth of substrates documented for this FMO paralog, both hydrophobic and hydrophilic, and accentuate its xenobiotic detoxifying prowess.

Now that the structures of the ancestral forms of mammalian FMO1, FMO2, FMO3 and FMO5 have been elucidated, it becomes clear that these monooxygenases have similar catalytic-site architectures. Only the AncFMO5 structure revealed somewhat different features that may explain that FMO5 is an outlier concerning its catalytic properties, being able to efficiently catalyze Baeyer-Villiger oxidations. The high similarity of the inner part of the active site cavity in AncFMO1, AncFMO2 and AncFMO3 coincides with an overlapping substrate specificity and reactivity (S- and N-oxidations). This seems to result in a set of redundant enzymes. Yet, inspection of their structures suggests that the outer segments of the active-site tunnels differ in their local topology, modulating accessibility and substrate preferences. This strategy has resulted in FMOs that can metabolize different classes of xenobiotics.

## Experimental procedures

### *Ancestral sequence reconstruction*

The FMO phylogeny of jawed vertebrate previously reported was employed as the starting point<sup>23</sup>. The dataset was constructed according as follows: (i) previously experimentally characterized *Homo sapiens* FMO sequences were employed as queries in BLASTp searches in GenBank non-redundant protein sequences (nr) and in Uniprot KB. The searches were restricted by the taxonomy of organisms by classes or orders aiming to mine the whole diversity included in the terrestrial vertebrates (*i.e.* Amphibia, Aves and Mammalia classes and Testudines order) guided by TimeTree knowledge database<sup>24</sup>. Specifically the 26 orders of mammals were carefully scrutinized. Hits were collected when showed > 55% identity (>80% coverage) and E-value= 0.0. (ii) Partial sequences (>510 amino acid length) or poor quality ones were excluded. (iii) All collected sequences were gathered and analyzed in a MSA using MAFFT v7, removing those that were collected more than one time. In the case of amphibian sequences as more than 5 paralogs were detected, the genomic location and context was analyzed to rule out annotation artifacts. Those sequences corresponding to true ORFs were kept in the dataset. This process allowed building a representative and non-redundant dataset. Dataset included 37 FMO1-like and 49 FMO4-like sequences from mammals. The MSA contained 365 sequences, 569 sites. Best-fit model parameters were obtained by the Akaike information criterion in ProtTest v3.4. Phylogeny was inferred by the maximum likelihood method in RAxML v0.6.0 (1000 bootstraps) and subjected to transfer bootstrap expectation (TBE) in BOOSTER. Ancestral sequence reconstruction was performed as marginal reconstruction using the maximum likelihood inference method in PAMLX v.4.9. Sequences were analyzed using an empirical substitution matrix

and empirical equilibrium amino acid frequencies (model = 3), 4 gamma categories and LG substitution matrix. The posterior probability distribution of ancestral states at each site was analyzed at nodes corresponding to mammalian AncFMO1 and AncFMO4. The length of the ancestors was treated by parsimony analyzing the presence/absence of gaps in the targeted nodes on the basis of the length of the derived sequences in each clade. This allowed us to determine that AncFMO4 had a C-terminal extension after the predicted transmembrane domain, which is a unique feature of mammalian FMO4s. Sites were considered ambiguously reconstructed when the alternative states displayed posterior probabilities (PP) > 0.2.

### ***Chemicals***

All chemicals were ordered from Merck, while NEB10 $\beta$  cells strain and DNA ligase were from New England Biolabs. NADPH and NADP<sup>+</sup> were ordered at Oriental Yeast Co.

### ***Cloning, transformation & expression***

Synthetic genes containing BsaI restriction sites at both the 5' and 3' ends were ordered from Integrated DNA Technologies. Genes were cloned following the Golden Gate cloning method. The recipient vector was a pBAD plasmid modified in such a way that the target protein is expressed fused at its N-terminus to a SUMO protein that carries a 6xHis-tag at its N-terminus. The cloning mixture was the following: 55.3 ng and 58.2 ng of AncFMO1 or AncFMO4 inserts, respectively, 75 ng of Golden Gate entry vector (a molar ratio of 2:1 insert:vector), 15 U BsaI-HF, 15 U T4 DNA ligase, T4 DNA ligase buffer (1 $\times$ ), and nuclease-free water added to a final volume of 20  $\mu$ l. A negative control was prepared without any inserts and the number of used cycles aimed for maximum efficiency: the first step with a cycle at 37 °C for 5 min was followed by 16 °C for 10 min, which was repeated 30 times. Then the temperature was set at 55 °C for 10 min and finally at 65 °C for 20 min. The sample was stored at 8 °C until the next day. Once cloned, the pBAD-6xHis-SUMO-AncFMO1/4 plasmids were transformed into NEB10 $\beta$ CaCl<sub>2</sub> competent cells. 5.0  $\mu$ l of plasmid DNA was added to 100  $\mu$ l CaCl<sub>2</sub> competent cells and incubated for 30 min. The cells were then heat shocked at 42°C for 30 s and incubated on ice for 5 min. 250  $\mu$ l LB-SOC45 was added to allow the cells to recover at 37°C for 1 h. The resuspended cells pellet was then plated on LB-agar45 containing 100  $\mu$ g.ml<sup>-1</sup> ampicillin and incubated overnight at 37°C. Plasmids were isolated and verified by sequencing and a 20% glycerol stock was stored at -70°C. A pre-inoculum of 4 ml LB-amp (50  $\mu$ g.ml<sup>-1</sup>) was grown overnight at 37°C and used to inoculate 2 liter baffled flasks containing 400 ml of Terrific-Broth medium45, supplemented with 50 mg.l<sup>-1</sup> ampicillin and incubated at 17 or 24 °C. Expression was induced by adding 0.02% L-arabinose from a sterile 20% stock (w/v) when the OD<sub>600</sub> was between 0.2 and 0.5. Cultures were grown at 24°C with shaking for a total of 30 hours before harvesting. Incubating the cells at 17 °C significantly increased the yield of AncFMO1 and its detergent exchangeability, compared to those grown at 24°C, later found to be crucial for crystallization. Cells were harvested by centrifugation (5,000g, 15 min, 10 °C), flash frozen in liquid nitrogen and stored at -20 °C.

### ***Cell disruption, membrane extraction and purification***

All the following steps were carried out on ice or at 4°C. Cell pellets were resuspended into Buffer A (250 mM NaCl, 50 mM potassium phosphate, pH 7.5) with a 5:1 ratio volume (ml) : mass (g) and supplemented with 0.10 mM phenylmethylsulfonyl fluoride and 1.0 mM  $\beta$ -mercaptoethanol to prevent protein degradation. Cell disruption was done by sonication (70% amplitude, 5 s ON, 5 s OFF, for a total of 20 min) or a high-pressure homogenizer (2 runs). After centrifuging at 18000g for 20 min, the supernatant was removed and the pellet was resuspended into Buffer A2 (250 mM NaCl, 50 mM potassium phosphate, 0.5% Triton X100-reduced, pH 7.5) with the same ratio as before (5:1). The resuspended pellet was mixed overnight at 4°C in order to solubilize the membrane protein and centrifuged at 18000g to collect the supernatant. AncFMO1 and AltAncFMO1 were purified with a metal-ion affinity chromatography that bound the histidine tag attached to the N-terminal part of the fused SUMO protein. The cell free extract was applied to the column and washed with increasing concentrations of imidazole. Buffer B contained (250 mM NaCl, 50 mM potassium phosphate, 300 mM imidazole, 0.5% Triton X100-reduced, pH 7.5). Following the washing steps of 0, 10, and 50 mM imidazole, the protein was finally eluted with 300 mM imidazole. The elution buffer was exchanged with a storage buffer using a desalting column (250 mM NaCl, 50 mM potassium phosphate, 0.05% Triton X100-reduced, pH 7.5).

Purified 6xHis-SUMO tagged enzyme was frozen with liquid nitrogen and kept at -20°C. Enzymatic assays were performed using these aliquots. Crystallization trials required further purification, including 6xHis-SUMO tag cleavage, and was carried out as outlined in our previous work 22. Purification of AncFMO1 was performed using dodecyl- $\beta$ -D-maltoside (DDM) (0.03% w/v analytical grade). Exchanging the detergent from Triton X100 to DDM was only successful for protein that had been produced in cultures grown at 17 °C. Protein produced at 24 °C did not fully exchange Triton X100 and the heterogeneous solution was problematic during crystallization trials.

### ***Western-blot analysis***

Cells were lysed as described above with both the aqueous layer and membrane pellet kept for analysis. The insoluble membrane pellet was resuspended in Buffer A to a final volume of approximately 60 ml. Resuspended membranes were aliquoted in eight 7 ml solutions and incubated individually overnight at 4 °C with the following detergents (1% (v/v) final). AncFMO4: SDS (sodium dodecyl sulfate), DDM (dodecyl-beta-maltoside), TRX (Triton X100), OG (octyl glucoside), FOS (FOS-Choline 8), GDN (glyco-diosgenin), LDAO (lauryldimethylamine oxide) and DMDPPO (dimethyldecyl phosphine oxide). AncFMO1: SDS, TRX, DDM, CYMAL-6 and OG. The control fraction (CTR) represents the membrane fraction without any added detergent. Solutions were pelleted to remove the insoluble components by spinning down the samples at 100,000g for 30 mins at 4 °C. 20  $\mu$ l from each sample (including the aqueous supernatant derived from AncFMO4 expression test representing the aqueous fractions) were then submitted to sodium dodecyl sulfate-polyacrylamide gel electrophoresis, SDS-PAGE analysis. The resulting gels were then transferred to a Mini Format, 0.2  $\mu$ m PVDF, Single application (BIO-RAD) membrane

using a Trans-Blot Turbo Pack and a Trans-Blot Turbo Transfer System (BIO-RAD). The membrane was then washed with a milky solution comprising milk powder (2.5% (w/v) NESTLE), TWEEN-20 (0.05% (v/v)) and a Tris-buffered saline solution (50 mM Tris-Cl, pH 7.5, 150 mM NaCl), final volume 50 ml, for one hour. This step is important for preventing non-specific antibody binding. The solution was then washed with the same solution including an anti-6xHis-tag fused peroxidase monoclonal primary antibody (final volume 15 ml) for one hour. The membrane was then washed with the same buffer excluding the antibody and milk powder to remove any excess milk and antibody unattached. The membrane was finally washed with the Clarity, Western Enhanced chemiluminescence (ECL) substrate (BIO-RAD) to initiate chemiluminescence according to the manufacturer's instructions to visualize the bands.

### *Melting temperature assays*

The melting temperature ( $T_m$ ) of AncFMO1 was assessed and determined using a TychoTMNT.6 system (NanoTemper Technologies GmbH, Munich, Germany) in the absence and presence of 200  $\mu\text{M}$  NADP<sup>+</sup>, respectively. Concentrations of AncFMO1 were determined using  $\epsilon_{\text{FAD}} = 12.0 \text{ mM}^{-1}\cdot\text{cm}^{-1}$  at 442 nm. Experiments were performed in triplicate, with each sample containing AncFMO1 (1.0 mg ml<sup>-1</sup>, determined using the calculated molecular weight of AncFMO1, 61 kDa), with or without NADP<sup>+</sup> (200  $\mu\text{M}$ ), made to a final volume of 10  $\mu\text{l}$  using the storage buffer. To ensure the  $T_m$  of AncFMO1 assessed using the TychoTMNT.6 system was comparable to the ThermoFAD assay performed on the AncFMOs<sup>23</sup> a control experiment was performed using AncFMO2, which corroborated the previously observed  $T_m$  values (data not shown).

### *Enzyme assays*

All reactions components were prepared in the same buffer as the one used for enzyme storage: 250 mM NaCl, 50 mM potassium phosphate, 0.05% Triton X100-reduced, pH 7.5. Steady-state kinetics measurements were done on a Jasco V-660 spectrophotometer in technical triplicates. Unless stated otherwise, AncFMO1 activity was measured by monitoring NADPH consumption at 25°C (at 340 nm,  $\epsilon_{\text{NADPH}} = 6.22 \text{ mM}^{-1}\cdot\text{cm}^{-1}$ ). The reaction mixture comprised 0.10 mM NADPH (or NADH), 2.5 - 1000  $\mu\text{M}$  substrate and 0.01 - 0.1  $\mu\text{M}$  enzyme. Catalytic activity of AltAncFMO1 was measured with the same settings as AncFMO1. The conditions were chosen according to literature in order to have a fair comparison with the extant human FMO. NAD(P)H uncoupling rates were obtained in absence of any substrate.  $K_M$  and  $k_{\text{cat}}$  values were calculated by fitting the data with the Michaelis-Menten equation using GraphPad 6.07 (La Jolla, CA, USA). When it was not possible to measure further, the  $K_M$  was reported as below the lowest substrate concentration for which a rate could be determined.

### Conversions

Substrate conversions were done at pH 7.5, using 5.0 mM substrate (1% MeOH), 0.10 mM NADPH, 2.0  $\mu$ M enzyme, 5.0  $\mu$ M phosphite dehydrogenase and 20 mM sodium phosphite. The last two components were used as a regeneration system for NADPH and the control did not contain any AncFMO1 protein. The final reaction volume was adjusted to 1.0 ml with buffer and put into 4 ml vials before being incubated at 30°C, with shaking, for 18 hours. Conversion of thioanisole, heptan-2-one and bicyclo[3.2.0]hept-2-en-6-one could be analyzed by GC-MS while benzydamine and tamoxifen conversions were monitored by HPLC. To determine the enantioselectivity in the sulfoxidation of thioanisole, chiral HPLC analysis was performed. For GC-MS, compounds were extracted twice by adding one volume of ethyl acetate, vortexing for 20 seconds, centrifuging and eluting the organic phase through anhydrous magnesium sulfate. GC-MS analyses were performed using an HP-1 Agilent column (30 m x 0.25 mm x 0.25  $\mu$ m). For thioanisole, the method was the following: injector and detector temperature at 250 °C, a split ratio of 5.0, and an injection volume of 1  $\mu$ l. The column temperature was held at 50 °C for 4 min, increased by 10 °C/min to 250 °C and held for 5 min. Thioanisole and methyl phenyl sulfoxide had a retention time of 10.05 min and 13.60 min, respectively.

HPLC analyses were performed after diluting 100  $\mu$ l of the sample into 400  $\mu$ l acetonitrile, vortexing it for 20 s and centrifuging. Analysis was performed using reverse phase HPLC. Samples were injected with a volume of 10  $\mu$ l onto a JASCO HPLC system, equipped with a Grace Alltima HP C18 column (5  $\mu$ m, 4.6x250 mm). The solvents used were water with 0.1% v/v formic acid (A) and acetonitrile (B) and the flow rate was 0.8 ml.min<sup>-1</sup>. For benzydamine the method was the following: 45 min on a gradient of 25-95% B, 3 min with 95% B followed by a 3 min decreased gradient of 95-25% B and finally a re-equilibration for 2 min. Benzydamine and benzydamine N-oxide were detected at 308 nm with a retention time of 13.08 min and 13.82 min, respectively. For tamoxifen the method was the following: 30 min on a gradient of 40-95% B, 3 min with 95% B followed by a 5 min decreased gradient of 95-40% B and finally a re-equilibration for 2 min. Tamoxifen and tamoxifen N-oxide were detected at 276 nm with a retention time of 13.06 min and 14.20 min, respectively.

Enantiomeric excess (ee) values were determined by chiral HPLC analysis using a Shimadzu LC-10ADVP HPLC equipped with a Chiralcel OD-H column (5  $\mu$ m, 4.6x250 mm). For the chiral HPLC, the GC-MS sample was evaporated and resuspended into isopropanol. The method was heptane/isopropanol 90:10 with a flow rate of 1 ml.min<sup>-1</sup>. The *R*- and *S*- enantiomers of methyl phenyl sulfoxide were detected at 220 nm with a retention time of 10.34 min and 13.04 min, respectively.

### **Rapid kinetics**

Stopped-flow experiments were carried out on a SX20 stopped-flow spectrophotometer equipped with a single-channel photomultiplier (PMT) or a photodiode array (PDA) detection module (Applied Photophysics, Surrey, UK). Solutions were prepared in 50 mM potassium phosphate, 250 mM NaCl and 0.05% TRX-100-reduced, pH 7.5 and unless mentioned, the experimental design followed the protocol as published previously<sup>23</sup>. Experiments were run at 25 °C, with 5 - 15  $\mu$ M enzyme, in technical triplicates. The  $k_{\text{red}}$  and  $K_{\text{d}}$  for AncFMO1 were determined by mixing anaerobically the enzyme with NAD(P)H and following the absorbance decrease at 448 nm. To probe the effect of substrate on the rate of reduction, the reduction of AncFMO1 by 100  $\mu$ M NAD(P)H was followed both with and without 1.0 mM thioanisole. The oxidative half reaction was monitored using the PDA module which allows collecting absorbance spectra every millisecond. In order to prepare reduced AncFMO1 for the oxidative half reaction, dithionite was added to an anaerobic solution containing AncFMO1 and an equivalent amount of NADP<sup>+</sup>. Dithionite was titrated until the loss of the yellow color of the oxidized FAD was complete, indicating complete reduction to FADH<sub>2</sub>. Mixing of reduced AncFMO1 with dioxygen-containing buffers (130  $\mu$ M dioxygen, final) was done both with and without 100  $\mu$ M NADP<sup>+</sup>.

### **Crystallization and structural determination of AncFMO1**

The crystallization condition that resulted in AncFMO1 crystals which displayed the highest diffraction is described below. Concentrations of protein were measured using the absorbance of the FAD at 448 nm, using an extinction coefficient of 12 mM<sup>-1</sup> cm<sup>-1</sup>. Prior to crystallization, NADP<sup>+</sup> (1.0 mM final) was incubated with AncFMO1 (12.5 mg ml<sup>-1</sup>, in storage buffer conditions containing 0.03% (w/v) DDM) for 1 h at 4 °C. 1  $\mu$ l protein-containing solution was then mixed with 1  $\mu$ l of the crystallization condition comprising 100 mM HEPES (pH 7.5), PEG 4000 (10% v/v) and glycerol (20% v/v) as cryo-protectant, in a sitting drop at 20 °C. The same crystallization solution was used as reservoir solution (1 ml). After 2 days, large yellow crystals formed. Crystals were then fished directly from the drop with no additional cryo-protectants.

Data were collected at the Swiss Light Source (Villigen, Switzerland) and processed with the XDS<sup>46</sup> and CCP4 packages<sup>47</sup>. Aimless was then used to merge the observations into average densities for two datasets<sup>47</sup> to maximize the total number of observed reflections, (Table 2). STARANISO was used as, similarly to AncFMO2, these crystals suffered severely from anisotropy<sup>48,49</sup>. Using STARANISO on the final merged dataset was imperative for elucidating a good electron density map. The phase problem was solved by molecular replacement using AncFMO3-6 (PDB 6SE3) as a search model using Phaser<sup>23,50</sup>. The phases were greatly improved by density averaging with DM<sup>47,50</sup>. Model building and refinement were then conducted using COOT51, Buccaneer<sup>47</sup> and Refmac5<sup>52</sup>. The residue outliers in the Ramachandran plots were 5.9% for AncFMO1. Figures were generated using UCSF Chimera53, PyMOL (DeLano Scientific; www.pymol.org) and CCP4mg<sup>47</sup>. Coordinates of the refined model of AncFMO1 were submitted to the protein data bank with PDB code 7AL4.

## **Data availability**

All data are contained within the manuscript.

## **Authors contribution**

All listed authors performed experiments and/or analyzed data. M.L.M. conducted thorough evolutionary analyses and performed ancestral-sequence reconstruction. G.B. performed Golden Gate cloning to insert the two AncFMOs genes into their respective vectors, carried out extensive kinetic analysis and validated the substrate profiles using spectrophotometer, stopped-flow spectroscopy and GCMS for AncFMO1. C.R.N. crystallized AncFMO1, collected the corresponding datasets at the ESRF and SLS facilities, performed structural analysis and elucidated AncFMO1 structure. G.B., C.R.N. and M.L.M. prepared the figures. G.B., C.R.N., and M.L.M. wrote the manuscript and A.M. and M.W.F. edited it. All authors provided critical feedback and helped shape the research, analysis and manuscript. M.L.M., A.M. and M.W.F. conceived the original idea.

## **Funding and acknowledgements**

The research for this work has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 722390; the Italian Ministry of Education, University and Research (MIUR) under the "Dipartimenti di Eccellenza (2018–2022)" program; The European Union's Horizon 2020 research and innovation programme under grant agreement No 847675 and ANPCyT (Argentina) PICT 2016-2839 to M.L.M. M.L.M. is a member of the Research Career of CONICET, Argentina.

## **Declaration of interest**

The authors declare that they have no conflicts of interest with the contents of this article.



## References

- 1 Krueger, S. K. & Williams, D. E. Mammalian flavin-containing monooxygenases: Structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacol. Ther.* 106, 357–387 (2005).
- 2 Phillips, I. R. & Shephard, E. A. Drug metabolism by flavin-containing monooxygenases of human and mouse. *Expert Opin. Drug Metab. Toxicol.* 13, 167–181 (2017).
- 3 Cruciani, G. et al. Flavin monooxygenase metabolism: Why medicinal chemists should matter. *J. Med. Chem.* 57, 6183–6196 (2014).
- 4 Shephard, E. A. et al. Localization of genes encoding three distinct flavin-containing monooxygenases to human chromosome 1q. *Genomics* 16, 85–89 (1993).
- 5 McCombie, R. R., Dolphin, C. T., Povey, S., Phillips, I. R. & Shephard, E. A. Localization of human flavin-containing monooxygenase genes FMO2 and FMO5 to chromosome 1q. *Genomics* 34, 426–429 (1996).
- 6 Hines, R. N., Hopp, K. A., Franco, J., Saeian, K. & Begun, F. P. Alternative processing of the human FMO6 gene renders transcripts incapable of encoding a functional flavin-containing monooxygenase. *Mol. Pharmacol.* 62, 320–325 (2002).
- 7 Dolphin, C. T., Cullingford, T. E., Shephard, E. A., Smith, R. L. & Phillips, I. R. Differential developmental and tissue-specific regulation of expression of the genes encoding three members of the flavin-containing monooxygenase family of man, FMO1, FMO3 and FMO4. *Eur. J. Biochem.* 235, 683–689 (1996).
- 8 Zhang, J. & Cashman, J. R. Quantitative analysis of FMO gene mRNA levels in human tissues. *Drug Metab. Dispos.* 34, 19–26 (2006).
- 9 Basarab, T., Ashton, G. H. S., Menagé, H. D. P. & McGrath, J. A. Sequence variations in the flavin-containing mono-oxygenase 3 gene (FMO3) in fish odour syndrome. *Br. J. Dermatol.* 140, 164–167 (1999).
- 10 Phillips, I. R. & Shephard, E. A. Flavin-containing monooxygenase 3 (FMO3): genetic variants and their consequences for drug metabolism and disease. *Xenobiotica* 50, 19–33 (2020).
- 11 Cashman, J. R. & Zhang, J. Human flavin-containing monooxygenases. *Annu. Rev. Pharmacol. Toxicol.* 46, 65–100 (2006).
- 12 Ziegler, D. M. An overview of the mechanism, substrate specificities, and structure of FMOs. *Drug Metab. Rev.* 34, 503–511 (2002).
- 13 van Berkel, W. J. H., Kamerbeek, N. M. & Fraaije, M. W. Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *J. Biotechnol.* 124, 670–689 (2006).
- 14 Ziegler, D. M. & Poulsen, L. L. Hepatic microsomal mixed-function amine oxidase. *Methods Enzymol.* 52, 142–151 (1978).
- 15 Beaty, N. B. & Ballou, D. P. The reductive half-reaction of liver microsomal FAD-containing monooxygenase. *J. Biol. Chem.* 256, 4611–4618 (1981).
- 16 Beaty, N. B. & Ballou, D. P. The oxidative half-reaction of liver microsomal FAD-containing monooxygenase\*. *J. Biol. Chem.* 256, 4619–4625 (1981).
- 17 Campbell, A. C., Robinson, R., Mena-aguilar, D., Sobrado, P. & Tanner, J. J. Structural Determinants of Flavin Dynamics in a Class B Monooxygenase. *ACS Biochem.* (2020) doi:10.1021/acs.biochem.0c00783.
- 18 Siddiq, M. A., Hochberg, G. K. & Thornton, J. W. Evolution of protein specificity: insights from ancestral protein reconstruction. *Curr. Opin. Struct. Biol.* 47, 113–122 (2017).

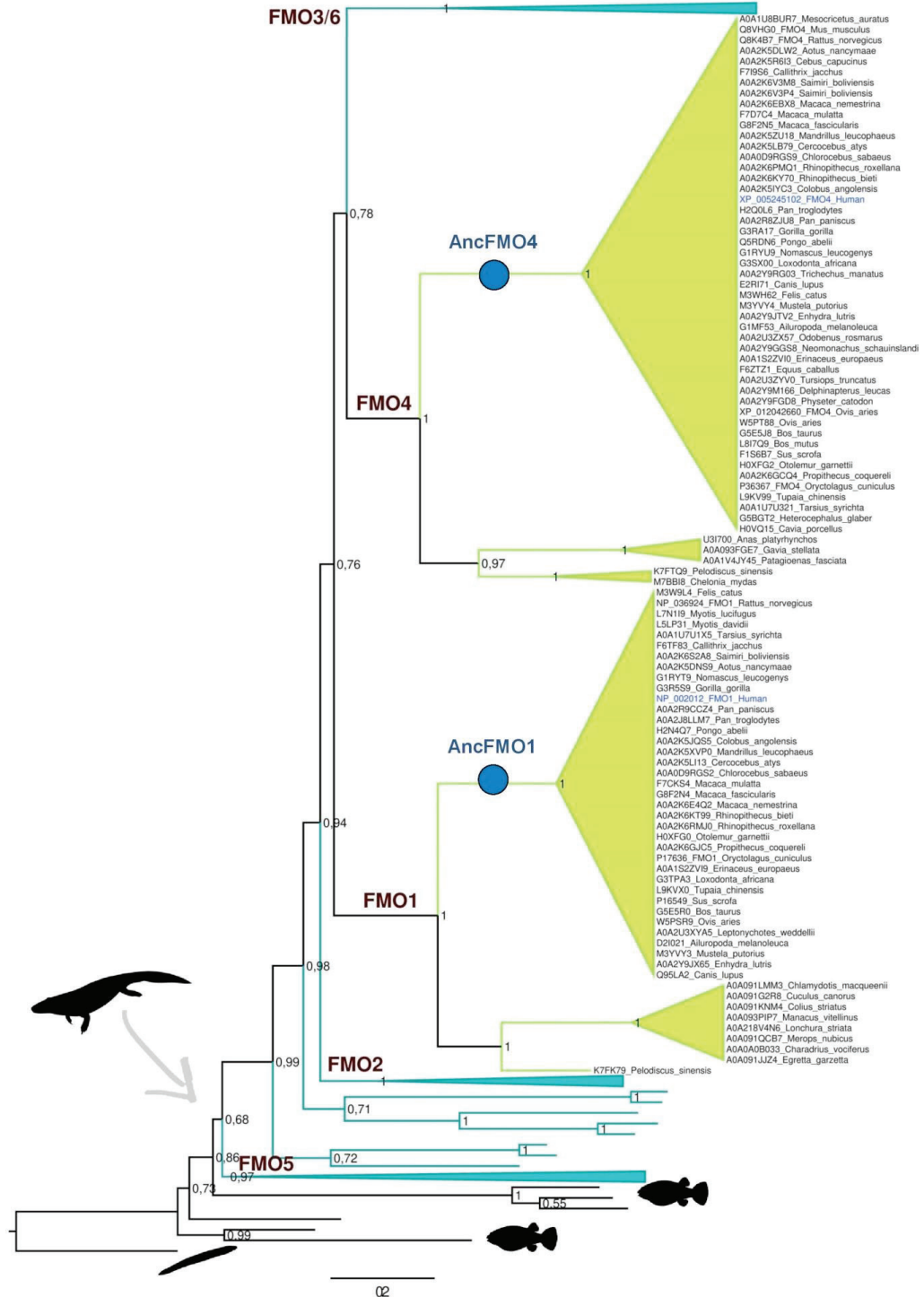
- 19 Hochberg, G. K. A. & Thornton, J. W. Reconstructing Ancient Proteins to Understand the Causes of Structure and Function. *Annu. Rev. Biophys.* 46, 247–269 (2017).
- 20 Merkl, R. & Sterner, R. Ancestral protein reconstruction: Techniques and applications. *Biol. Chem.* 397, 1–21 (2016).
- 21 Voordeckers, K. et al. Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. *PLoS Biol.* 10, (2012).
- 22 Perez-Jimenez, R. et al. Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. *Nat. Struct. Mol. Biol.* 18, 592–596 (2011).
- 23 Nicoll, C. R. et al. Ancestral-sequence reconstruction unveils the structural basis of function in mammalian FMOs. *Nat. Struct. Mol. Biol.* 27, 14–24 (2020).
- 24 Kumar, S., Stecher, G., Suleski, M. & Heddes, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* 34, 1812–1819 (2017).
- 25 Itagaki, K., Carver, G. T. & Philpot, R. M. Expression and characterization of a modified flavin-containing monooxygenase 4 from humans. *J. Biol. Chem.* 271, 20102–20107 (1996).
- 26 Hu, S. X. Hepatic Flavin-containing Monooxygenase and Aldehyde Oxidase Activities in Male Domestic Pigs at Different Ages. *Drug Metab. Lett.* 12, 125–131 (2018).
- 27 Uno, Y., Shimizu, M. & Yamazaki, H. Molecular and functional characterization of flavin-containing monooxygenases in cynomolgus macaque. *Biochem. Pharmacol.* 85, 1837–1847 (2013).
- 28 Siddens, L. K., Krueger, S. K., Henderson, M. C. & Williams, D. E. Mammalian flavin-containing monooxygenase (FMO) as a source of hydrogen peroxide. *Biochem. Pharmacol.* 89, 141–147 (2014).
- 29 Kawaji, A., Ohara, K. & Takabatake, E. An assay of flavin-containing monooxygenase activity with benzydamine N-oxidation. *Anal. Biochem.* 214, 409–412 (1993).
- 30 Fisher, B. et al. Tamoxifen for the prevention of breast cancer: Current status of the National Surgical Adjuvant Breast and Bowel Project P-1 study. *J. Natl. Cancer Inst.* 97, 1652–1662 (2005).
- 31 Chitra, M., Ernest, H. & David, K. Metabolism of the cancer agent tamoxifen : Flavin-Containing Monooxygenase-Mediated N-Oxidation. *Drug Metab. Dispos.* 21, (1993).
- 32 Fiorentini, F. et al. Baeyer-Villiger Monooxygenase FMO5 as entry point in drug metabolism. *ACS Chem. Biol.* 12, 2379–2387 (2017).
- 33 Fiorentini, F. et al. Biocatalytic characterization of human FMO5: Enearthing Baeyer-Villiger reactions in humans. *ACS Chem. Biol.* 11, 1039–1048 (2016).
- 34 Fabara, A. N. & Fraaije, M. W. An overview of microbial indigo-forming enzymes. *Appl. Microbiol. Biotechnol.* 104, 925–933 (2020).
- 35 Romero, E., Robinson, R. & Sobrado, P. Monitoring the reductive and oxidative half-reactions of a flavin-dependent monooxygenase using stopped-flow spectrophotometry. *J. Vis. Exp.* (2012) doi:10.3791/3803.
- 36 Veeravalli, S. et al. Flavin-containing monooxygenase 1 catalyzes the production of taurine from hypotaurine. *Drug Metab. Dispos.* 48, 378–385 (2020).
- 37 Sofer, S. S. & Ziegler, M. Microsomal mixed-function amine oxidase. Oxidation products of piperazine-substituted phenothiazine drugs. *Drug Metab. Dispos.* 232–239 (1978).
- 38 Nagata, T., Williams, D. E. & Ziegler, D. M. Substrate specificities of rabbit lung and porcine liver flavin-containing monooxygenases: Differences due to substrate size. *Chem. Res. Toxicol.* 3, 372–376 (1990).



- 39 Ho, B. K. & Gruswitz, F. HOLLOW: Generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* 8, 1–6 (2008).
- 40 Parte, P. & Kupfer, D. Oxidation of tamoxifen by human flavin-containing monooxygenase (FMO) 1 and FMO3 to tamoxifen-N-oxide and its novel reduction back to tamoxifen by human cytochromes P450 and hemoglobin. *Drug Metab. Dispos.* 33, 1446–1452 (2005).
- 41 Alfieri, A., Malito, E., Orru, R., Fraaije, M. W. & Mattevi, A. Revealing the moonlighting role of NADP in the structure of a flavin-containing monooxygenase. *Proc. Natl. Acad. Sci. U. S. A.* 105, 6572–6577 (2008).
- 42 Robinson, R., Badiyan, S. & Sobrado, P. C4a-hydroperoxyflavin formation in N-hydroxylating flavin monooxygenases is mediated by the 2'-OH of the nicotinamide ribose of NADP<sup>+</sup>. *Biochemistry* 52, 9089–9091 (2013).
- 43 Dean, A. M. & Golding, G. B. Protein engineering reveals ancient adaptive replacements in isocitrate dehydrogenase. *Proc. Natl. Acad. Sci. U. S. A.* 94, 3104–3109 (1997).
- 44 Expression Atlas. query [ensg00000010932 FMO1](https://www.ebi.ac.uk/ena/ExpressionAtlas/ensembl/00000010932/FMO1), last accession 2020-09-20. EMBL-EBI.
- 45 Sambrook, J., Russell, D. W. *Molecular Cloning: A Laboratory Manual*. (CSHL press, 2001).
- 46 Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* 26, 795–800 (1993).
- 47 Project, C. C. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 50, 760–763 (1994).
- 48 Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* 69, 1204–1214 (2013).
- 49 Vonrhein, C. et al. Advances in automated data analysis and processing within autoPROC, combined with improved characterisation, mitigation and visualisation of the anisotropy of diffraction limits using STARANISO. *Acta Crystallogr. Sect. A Found. Adv.* 74, a360–a360 (2018).
- 50 McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674 (2007).
- 51 Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 60, 2256–2268 (2004).
- 52 Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 53, 240–255 (1997).
- 53 Pettersen, E. F. et al. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612 (2004).

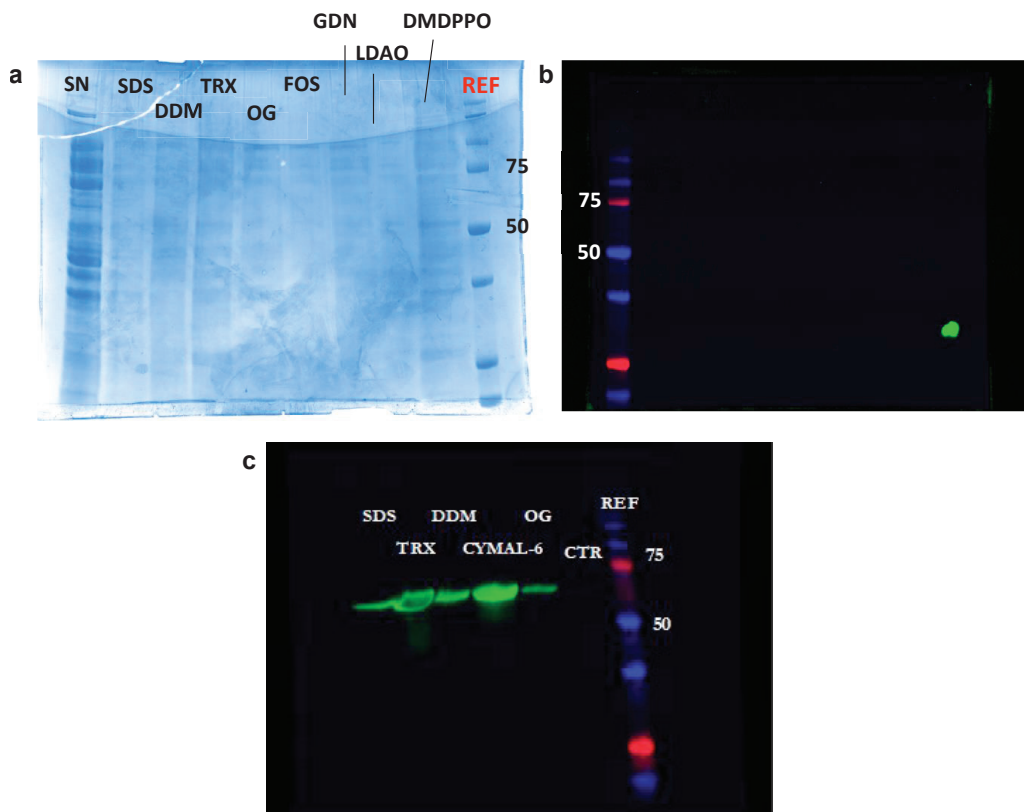
## Abbreviations and nomenclature

AltAncFMO	Alternative versions of the Ancestral flavin-containing monooxygenase
AncFMO	Ancestral flavin-containing monooxygenase
ASR	Ancestral sequence reconstruction
ee	Enantiomeric excess
DDM	Dodecyl- $\beta$ -D-maltoside
FAD	Flavin adenine dinucleotide
FMO	Flavin-containing monooxygenase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hFMO	Human flavin-containing monooxygenase
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
CYP	Cytochrome P450
PEG 4000	Polyethylene glycol 4000
PDA	Photodiode array
PMT	Single-channel photomultiplier
PP	Posterior probability
RMSD	Relative-mean square deviation
TBE	Transfer bootstrap expectation
T <sub>m</sub>	Melting temperature

# Supplementary Information



**Figure S1. Jawed vertebrates phylogeny of FMOs** (previous page). Tree was constructed in RAXML v0.6.0, 1000 bootstraps were run and subjected to TBE (shown at the nodes). The employed MSA was trimmed in single sequence extensions and contained 365 taxa and 569 sites. Clades are collapsed and colored according to the paralog groups: FMO2, FMO3/6 and FMO5 (teal), FMO1 and FMO4 (lime). Uniprot accession codes and species names are provided for all sequences in FMO1 and FMO4 clades. The emergence of terrestrial vertebrates is depicted with an arrow and a silhouette (  ). The tree was rooted according to the species tree in cephalochordates (  ). The tree was prepared in Figtree v1.4.2 and silhouettes obtained from <http://www.phylopic.org/>.



**Figure S2. AncFMO4 expression and solubilization screening of AncFMO1.** a, SDS-PAGE corresponding to the various detergents used to solubilize AncFMO4 from the membranes, including the aqueous supernatant. b, Western-blot of the SDS-PAGE, with the image flipped with the lanes in reverse. Detergents used; SN (supernatant), SDS (sodium dodecyl sulfate), DDM (dodecyl-beta-maltoside), TRX (Triton-X 100), OG (octyl glucoside), FOS (FOS-Choline 8), GDN (glyco-diosgenin), LDAO (lauryldimethylamine oxide) and DMDPPO (dimethyldecyl phosphine oxide). For clarity, certain lanes are indicated with black lines. c, Western-blot detergent screening for AncFMO1 extraction. Native membranes collected during purification were resuspended and incubated overnight with

various detergents including sodium dodecyl sulfate (SDS), Triton-X 100 (TRX), dodecyl beta-maltoside (DDM), CYMAL-6 and octyl glucoside (OG). CTR refers to the control sample without any detergent addition. REF refers to the standard molecular weight markers, with bands of 75 and 50 kDa indicated. 6xHis-tag antibodies were used to generate fluorescence.

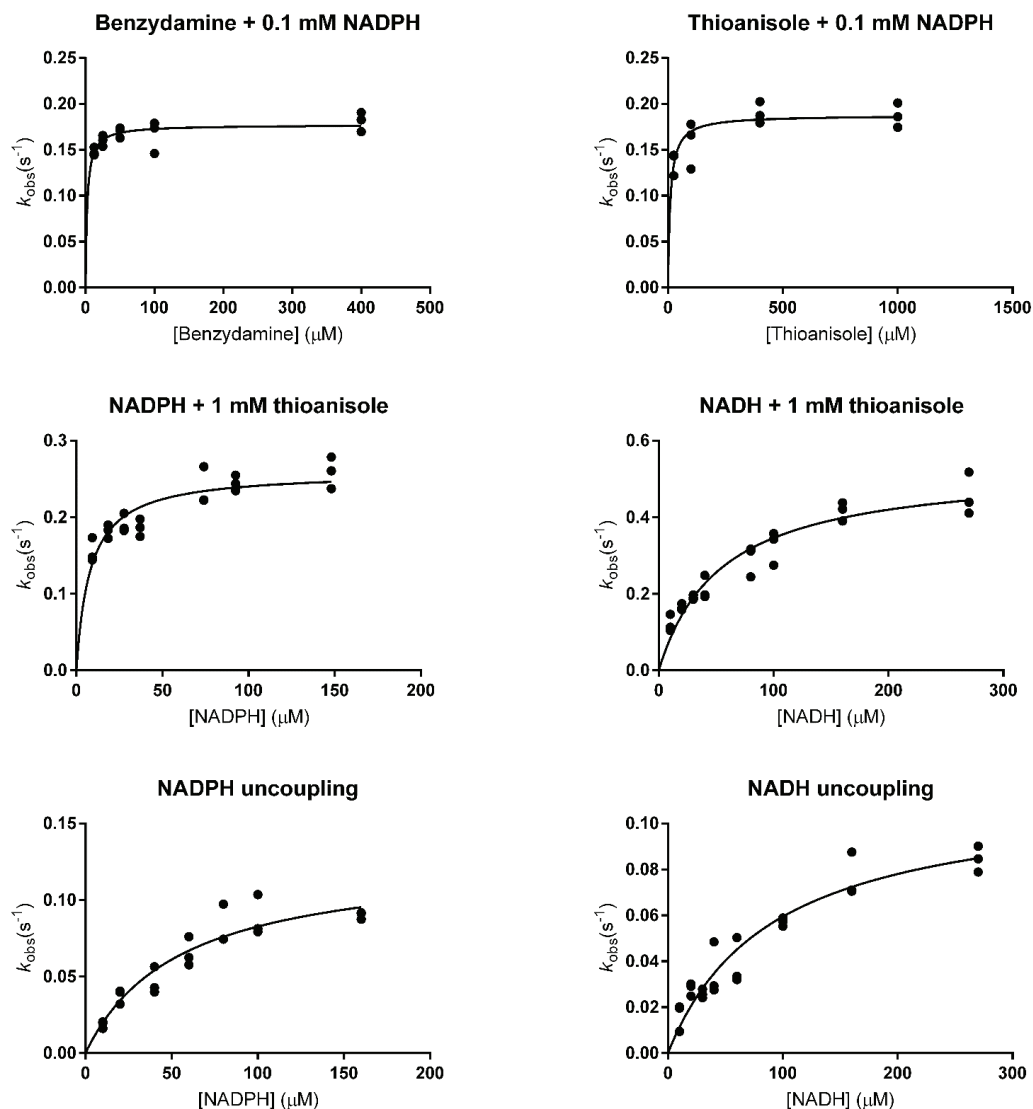
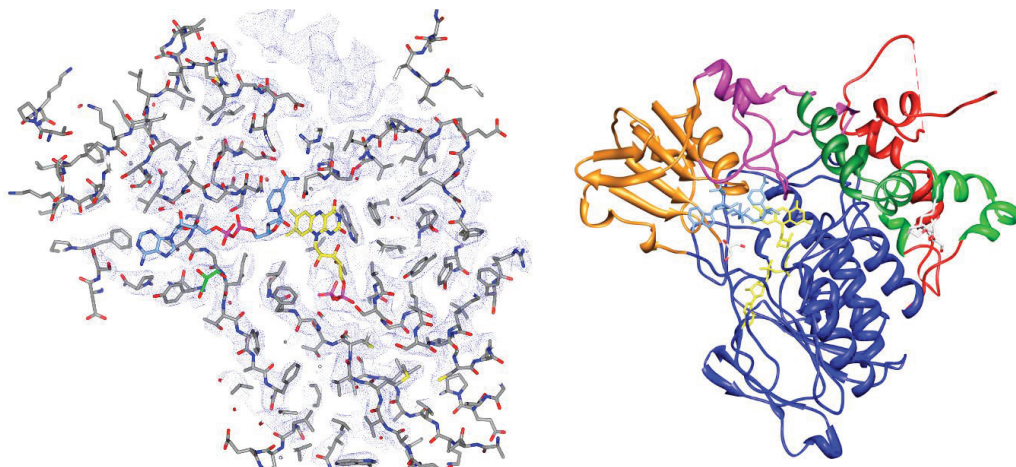


Figure S3. Steady-state kinetic plots for AncFMO1. Reactions were followed by measuring the NAD(P)H consumption at 340 nm and rates calculated using NAD(P)H extinction coefficient  $6.22 \text{ mM}^{-1} \cdot \text{cm}^{-1}$ . The plots were obtained by fitting the measured NAD(P)H depletion rates triplicates to GraphPad. Standard deviations are reported in Table 1.

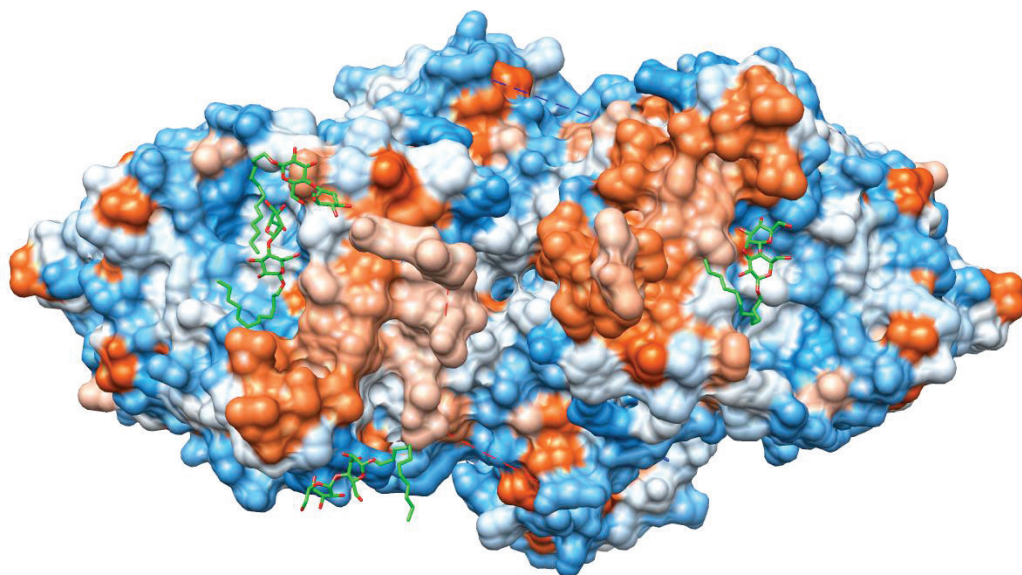


**Figure S4. Indigo formation.** The tube on the left contained 100  $\mu\text{M}$  NADPH, 20 mM phosphite, 5  $\mu\text{M}$  PTDH, 1 mM indole and 2  $\mu\text{M}$  AncFMO1. The control tube on the right only lacks AncFMO1. The picture was taken 1 hour after addition of enzyme, at 30  $^{\circ}\text{C}$ . All components were prepared in buffer (250 mM NaCl, 50 mM potassium phosphate, 0.05% Triton X100-reduced, pH 7.5).

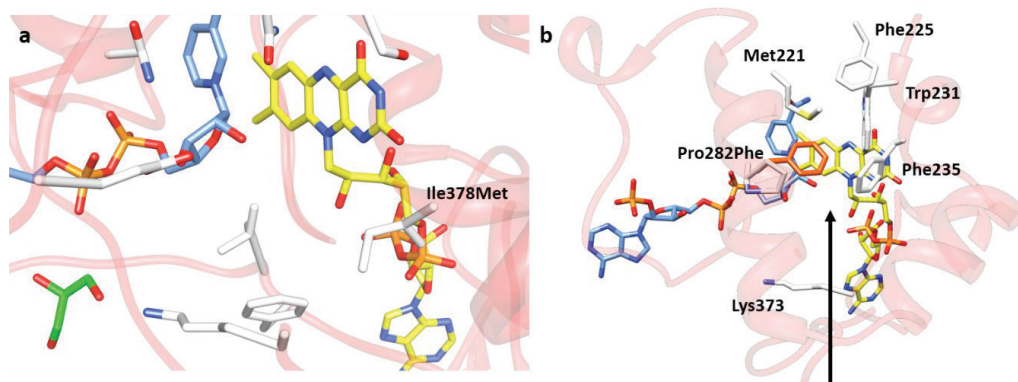




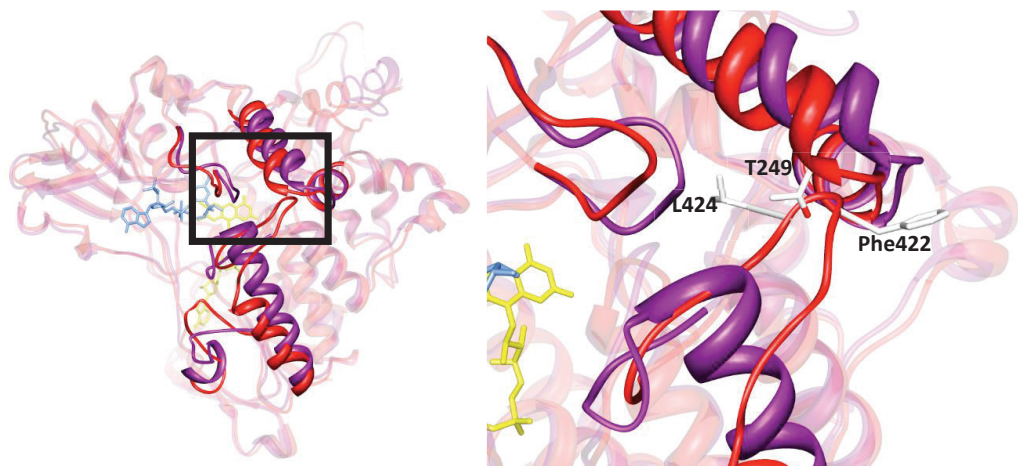
**Figure S5. Crystal structure of AncFMO1.** Left, Electron density maps of the AncFMOs. The depicted 2Fo-Fc maps (shown in blue) were calculated by averaging the electron density maps obtained after molecular replacement, and anisotropic correction using Staraniso. The crystal structure of AncFMO3-6 (PDB: 6SE3) was used as model for molecular replacement after the scaled reflections had been corrected for anisotropy using Staraniso. FAD and NADP<sup>+</sup> are shown in yellow and cornflower blue respectively. The contour level is 1.4  $\sigma$ . The glycerol molecule shown and the amino acids are shown in green and grey, respectively. Right, AncFMO1 depicted as a monomer with its domains and subdomains colored. Blue and orange colors represent the well conserved FAD (residues 2–154 and 331–442) and NAD(P)H (residues 155–213 and 296–330) binding domains, respectively. The subdomains colored in pink and dark green portray the distinct 80-residue insertion (residues 214–295) associated to mammalian FMOs. The dark green subdomain (residues 231–261) represents the hydrophobic substrate access points. Together, the light green (residues 449–473) and dark green subdomains form a large hydrophobic strip that grapples onto the membrane. Glycerol and DDM molecules are shown in white. Finally, the red-colored subdomain (residues 474–532) represents the C-terminus. FAD and NADP<sup>+</sup> are colored in yellow and cornflower blue, respectively.



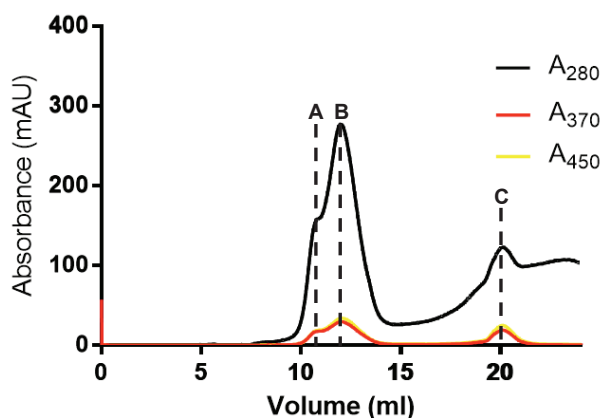
**Figure S6.** Charged surface distribution for the crystallographic dimer of AncFMO1. Distribution of charge around the surface of AncFMO1, with orange, white and blue representing hydrophobic, neutral and hydrophilic residues, respectively. Darker and lighter colors portray the extent of the characteristic, with the former and latter representing higher or lower, respectively. Large parallel hydrophobic strips across the bottom of the dimer are visible (dark orange), lined by a ring of positively polar residues (dark blue). Molecules of detergent, DDM are shown clustering around these hydrophobic strips in green.



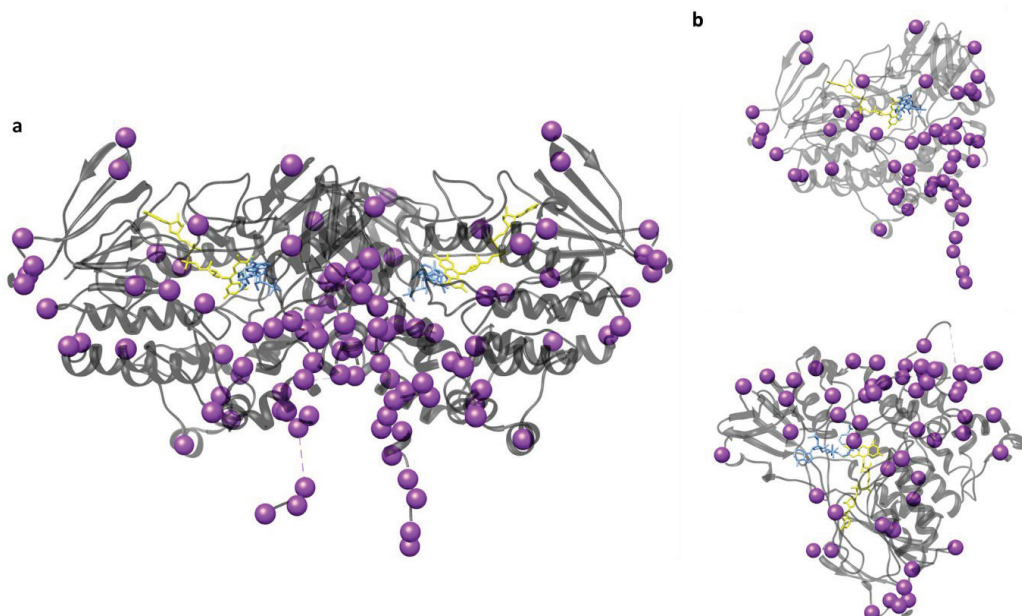
**Figure S7.** Key residue changes between AncFMO1 and hFMO1. *a*, The active site of AncFMO1 is shown with emphasis directed towards Ile378. This site is a methionine residue in hFMO1. *b*, The trajectory of the hydrophobic substrates with respect to the  $\alpha$ -helical triad is shown by a black arrow. FAD, NADP<sup>+</sup>, Glycerol, AncFMO1 residues and theoretical hFMO1 residues are shown in yellow, cornflower blue, green, white and



**Figure S8. Overarching loop.** The AncFMO1 monomer (red) is superimposed with the monomer of AncFMO3-6 (dark magenta). The unique loop feature exhibited by AncFMO1 is shown in the expanded black boxes. FAD, NADP<sup>+</sup> and residue sidechains are shown in yellow, cornflower blue and white, respectively.



**Figure S9. Gel filtration chromatogram of AncFMO1.** Previously characterized AncFMOs eluted in a bell-shaped manner between 10.5 and 11 mL, as a monodisperse peak. AncFMO1 exhibits two elution volumes at 10.8 and 12 mL represented by peaks A and B respectively. These two peaks overlap with, A, acting as a shoulder to the more pronounced peak, B. These two peaks collectively represent two different oligomeric states of AncFMO1. Crystallographic studies were performed with protein derived from both peaks, A and B, concentrated down together with the hypothesis that the oligomeric states A and B (derived from peaks A and B, respectively) were in dynamic equilibrium between each other. With the previously crystallized AncFMOs crystallizing whilst exhibiting the oligomeric state, A, we postulated that the oligomer B would shift to the oligomeric state A as that oligomer precipitates out of solution in crystalline form. Traces in black, red and yellow represent wavelengths 280, 370 and 450 nm, respectively, measured in absorbance, mAU. Band C corresponds to the excess of FAD added to the protein solution before size exclusion chromatography.



*Figure S10. Differing residues between AncFMO1 and hFMO1. a, The differences between AncFMO1 and hFMO1 with respect to the crystallographic dimer are shown with purple spheres. AncFMO1 and hFMO1 share a percentage sequence identity and percentage sequence similarity of 89.9% (54 residues) and 94.2% (29 residues), respectively, demonstrating the extensive likeness between the two enzymes. b, The differences observed with respect to monomer in two different orientations. The FAD and NADP molecules are shown in yellow and cornflower blue, respectively.*

**AncFM01**

&gt;AncFM01m

MAKRVAIVGAGVSGLASIKCCLEEGLEPTCFERSDDLGLWRFTEHVEEGRASLYKSVVSNSCKEMSC  
 YSDFPFPEYDYPNYVPNSQFLEYLKMYPANRFNLLKHIQFKTKVCSVTKCPDFTVTGQWEVVTQHEGKQE  
 SAIFDAVMVCTGFLTDPYLPDLSFPGINTFKGQYFHSRQYKHPDIFKDKRVLVVGMGNSGTDIAVEASH  
 LAKKVFLSTTGGAWVMSRVFDSGYPWDVMTTRFQNLRLNSLPTPIVTWLMARKMNSWFNHANYGL  
 VPEDRTLQREPVLNDELPGCIITGKVLIKPSIKEVKENSIVFNNTPKKEEPIDIIVFATGYTFAPPFLDESUVK  
 VENGQASLYKYIFPAHLPKPTLAVIGLIKPLGSIPTGETQARWAVRVLKGINKLPPQSVMIIEVFNARKEN  
 KPSGFGLCYCKALQSDYITYIDELTYINAKPNLLSMLLTDPRALTIFFGPCTPYQFRLTGPGKWEGAR  
 NAILTQWDRTFKVTKTRIVQESPSPFASLLKLLSLPVLLALLLMC

site	ML-Anc		Alt-Anc	
	State	PP	Alt state	PP
152	D	0.51	N	0.49
190	V	0.62	I	0.38
221	M	0.55	I	0.45
236	T	0.64	M	0.36
243	L	0.78	F	0.21
253	T	0.52	N	0.48
256	M	0.74	L	0.22
389	A	0.59	V	0.41
442	Y	0.54	S	0.41
450	L	0.72	F	0.27

**AncFMO4**

&gt;AncFMO4m

MAKRVAVIGAGVSGLSSIKCCLEDEPLEPTCFERSDDDFGGLWKFTESSKDGMRVYRSLVTNVCKEMSC  
 YSDFPFQEDYPNFMNHAKFWDYLRFAEHFDLLKYIQFKTTVCSVTKCPDFSETGQWDVVTTETEGKQD  
 RAVFDAVMVCTGHFLNPHLPESFPGIHKFKGQILHSQEYRTPEAFQGKRVLVIGLGNTGGDIAVELSRT  
 AAQVFLSTRGTWVLSRSSDGGYPFNMMTTRRCHNFIAQVLPSCFLNWIQERQMNRFNHENYGLSIT  
 KGKKAIVNDELPTCILCGTVTVKTSVKEFTETSAVFEDGTVEANIDVVIFTTGYTFSFPFLEEPLKSLC  
 TKKIFLYKRVPNLEKATLAIIGLISLTGSILAGTELQARWATRVPFKGLCKIPPSQKLMAEVTKKEQLIK  
 RGVIKDTSQDKLDYISYMDELAACIGAKPNIPLFLKDPRLAWEVFFGPCTPYQYRLVGPVKWDGARN  
 AILTQWDRTLKPLKTRIVADSSKPASMSHYLKVWGAPVLLASLLICKSSLFLKLVDRDKLQDRISPYLISL  
 WRGS

site	ML-Anc		Alt-Anc	
	State	PP	State	PP
37	F	0.76	I	0.24
116	C	0.70	R	0.30
177	R	0.76	K	0.24
221	L	0.54	I	0.24
231	F	0.51	W	0.30
235	T	0.38	V	0.36
240	H	0.78	N	0.20
319	A	0.63	E	0.37
349	I	0.66	M	0.32
354	R	0.62	Q	0.29
363	A	0.54	T	0.36
409	T	0.63	V	0.33
474	V	0.54	M	0.40
494	L	0.71	M	0.27
503	A	0.51	P	0.45
507	K	0.64	Q	0.34
522	V	0.68	I	0.32
524	L	0.64	F	0.23
541	R	0.58	Q	0.33
552	L	0.48	R	0.39
553	I	0.49	V	0.42
555	L	0.70	I	0.26
557	R	0.62	Q	0.25
558	G	0.73	N	0.21

Accession code	Organism	Taxonomic rank	FMO clade
L5LMT2	<i>Myotis davidii</i>		
W5PRR0	<i>Ovis aries</i>		
Q8HYJ9	<i>Bos taurus</i>		
L8I8K7	<i>Bos mutus</i>		
A0A140C4U0	<i>Bubalus bubalis</i>		
W5PRR0	<i>Ovis aries</i>		
XP_005690690	<i>Capra hircus</i>		
XP_020742463	<i>Odocoileus virginianus texanus</i>		
XP_007165360	<i>Balaenoptera acutorostrata scammoni</i>		
XP_004269798	<i>Orcinus orca</i>		
XP_019807672	<i>Tursiops truncatus</i>		
XP_022415770	<i>Delphinapterus leucas</i>		
XP_007113108	<i>Physeter catodon</i>		
S9YQ93	<i>Camelus ferus</i>		
XP_010955230	<i>Camelus bactrianus</i>		
XP_006211173	<i>Vicugna pacos</i>		
A0A1R7SPQ2	<i>Sus crofa</i>		
Q95LA1	<i>Canis lupus</i>		
M3YVX5	<i>Mustela putorius</i>		
XP_004408956	<i>Odobenus rosmarus</i>		
XP_021538316	<i>Neomonachus schauinslandi</i>		
XP_008694265	<i>Ursus maritimus</i>		
M3VZK1	<i>Felis catus</i>	Mammalia	<b>FMO3/6</b>
XP_007088793	<i>Panthera tigris</i>		
XP_014917803	<i>Acinonyx jubatus</i>		
XP_012513247	<i>Propithecus coquereli</i>		
XP_012620619	<i>Microcebus murinus</i>		
XP_004688531	<i>Condylura cristata</i>		
XP_006147860	<i>Tupaia chinensis</i>		
A0A1U7TTF6	<i>Tarsius syrichta</i>		
H0XFF2	<i>Otolemur garnettii</i>		
F6QL08	<i>Equus caballus</i>		
XP_014700658	<i>Equus asinus</i>		
F7HH82	<i>Callithrix jacchus</i>		
Q5REM1	<i>Pongo abelii</i>		
Q8SPQ7	<i>Macaca mulatta</i>		
A0A0D9RGQ9	<i>Chlorocebus sabaues</i>		

A0A023JCA1	<i>Macaca fascicularis</i>		
Q53FW5 (FMO3)	<i>Homo sapiens</i>		
A0A0G2JSI0	<i>Rattus norvegicus</i>		
A0A1U7Q764	<i>Mesocricetus auratus</i>		
XP_021054533	<i>Mus pahari</i>		
XP_005364035	<i>Microtus ochrogaster</i>		
XP_015849094	<i>Peromyscus maniculatus bairdii</i>		
XP_008846748	<i>Nannospalax galili</i>		
XP_020025626	<i>Castor canadensis</i>		
XP_006887483	<i>Elephantulus edwardii</i>		
U3KLZ1	<i>Oryctolagus cuniculus</i>		
XP_004853513	<i>Heterocephalus glaber</i>		
XP_004706887	<i>Echinops telfairi</i>		
XP_010835019	<i>Bison bison</i>		
W5PS24	<i>Ovis aries</i>		
XP_005216983	<i>Bos taurus</i>		
XP_004013737	<i>Ovis aries</i>		
XP_005690691	<i>Capra hircus</i>		
XP_005902457	<i>Bos mutus</i>		
XP_006075483	<i>Bubalus bubalis</i>		
I3LIW4	<i>Sus crofa</i>		
XP_006211175	<i>Vicugna pacos</i>		
XP_006173292	<i>Camelus ferus</i>		
J9P0F0	<i>Canis lupus</i>		
M3W9K9	<i>Felis catus</i>		
XP_019280198	<i>Panthera pardus</i>		
M3YVX8	<i>Mustela putorius</i>		
XP_022364137	<i>Enhydra lutris</i>		
XP_019664755	<i>Ailuropoda melanoleuca</i>		
XP_008586856	<i>Galeopterus variegatus</i>		
G7MFB5	<i>Macaca mulatta</i>		
G8F2N3	<i>Macaca fascicularis</i>		
XP_011744241	<i>Macaca nemestrina</i>		
XP_007987722	<i>Chlorocebus sabaues</i>		
XP_017722368	<i>Rhinopithecus bieti</i>		
XP_524962	<i>Pan troglodytes</i>		
O60774 (FMO6)	<i>Homo sapiens</i>		
G1TFY5	<i>Oryctolagus cuniculus</i>		



XP_004454775	<i>Dasypus novemcinctus</i>		
XP_006887525	<i>Elephantulus edwardii</i>		
XP_004706886	<i>Echinops telfairi</i>		
XP_006872703	<i>Chrysochloris asiatica</i>		
M0R553	<i>Rattus norvegicus</i>		
A0A1U7Q721	<i>Mesocricetus auratus</i>		
XP_021054447	<i>Mus pahari</i>		
XP_003505573	<i>Cricetulus griseus</i>		
XP_006974814	<i>Peromyscus maniculatus bairdii</i>		
XP_008846747	<i>Nannospalax galili</i>		
XP_007525282	<i>Erinaceus europaeus</i>		
XP_004688176	<i>Condylura cristata</i>		
I3M344	<i>Ictidomys tridecemlinea</i>		
XP_015347117	<i>Marmota marmota</i>		
XP_006147858	<i>Tupaia chinensis</i>		
XP_008144061	<i>Eptesicus fuscus</i>		
XP_005867703	<i>Myotis brandtii</i>		
G3SQL7	<i>Loxodonta africana</i>		
XP_012786396	<i>Ochotona princeps</i>		
A0A093JJ61	<i>Eurypyga helias</i>		
A0A0A0B1A6	<i>Charadrius vociferus</i>		
A0A087VQN0	<i>Balearica regulorum</i>		
A0A091XN66	<i>Opisthocomus hoazin</i>		
A0A091R3N8	<i>Leptosomus discolor</i>		
A0A091JN32	<i>Egretta garzetta</i>		
A0A091G0C9	<i>Cuculus canorus</i>		
A0A091TAH2	<i>Phaethon lepturus</i>		
A0A093JBD9	<i>Fulmarus glacialis</i>		
A0A091KMX8	<i>Chlamydotis macqueenii</i>		
A0A091MAZ1	<i>Cariama cristata</i>		
A0A093CLM1	<i>Pterocles gutturali</i>		
U3I4R2	<i>Anas platyrhynchos</i>		
Q8QH01	<i>Gallus gallus</i>		
K9UTG7	<i>Coturnix coturnix</i>		
A0A099ZAQ2	<i>Tinamus guttatus</i>		
M7C297	<i>Chelonia mydas</i>	Testudines	
A0A1L8GN41	<i>Xenopus laevis</i>		
F7CV72	<i>Xenopus tropicalis</i>	Amphibia	

XP_005245102	<i>Homo sapiens</i>	Mammalia	<b>FMO4</b>
A0A2R8ZJU8	<i>Pan paniscus</i>		
H2Q0L6	<i>Pan troglodytes</i>		
Q5RDN6	<i>Pongo abelii</i>		
G1RYU9	<i>Nomascus leucogenys</i>		
A0A2K5IYC3	<i>Colobus angolensis palliatus</i>		
A0A2K6PMQ1	<i>Rhinopitecus roxellana</i>		
A0A2K6KY70	<i>Rhinopitecus bieti</i>		
A0A2K5ZU18	<i>Mandrillus leucophaeus</i>		
A0A2K5LB79	<i>Cercocebus atys</i>		
F7D7C4	<i>Macaca mulatta</i>		
G8F2N5	<i>Macaca fascicularis</i>		
A0A2K6EBX8	<i>Macaca nemestrina</i>		
A0A0D9RGS9	<i>Chlorocebus sabaeus</i>		
G3RA17	<i>Gorilla gorilla</i>		
A0A2K6V3P4	<i>Saimiri boliviensis</i>		
A0A2K5DLW2	<i>Aotus nancymaae</i>		
A0A2K5R6I3	<i>Cebus capucinus</i>		
F7I9S6	<i>Callithrix jacchus</i>		
A0A2K6V3M8	<i>Saimiri boliviensis</i>		
XP_012042660	<i>Ovis aries</i>		
L8I7Q9	<i>Bos mutus</i>		
G5E5J8	<i>Bos taurus</i>		
W5PT88	<i>Ovis aries</i>		
A0A2Y9FGD8	<i>Physeter catodon</i>		
A0A2Y9M166	<i>Delphinapterus leucas</i>		
A0A2U3ZYV0	<i>Tursiops truncatus</i>		
F1S6B7	<i>Sus crofa</i>		
G1MF53	<i>Ailuropoda melanoleuca</i>		
M3WH62	<i>Felis catus</i>		
A0A2Y9GGS8	<i>Neomonachus schauinslandi</i>		
A0A2U3ZX57	<i>Odobenus rosmarus</i>		
A0A2Y9JTV	<i>Enhydra lutris</i>		
M3YVY4	<i>Mustela putorius</i>		
E2RI71	<i>Canis lupus</i>		
F6TZ1	<i>Equus caballus</i>		
A0A2K6GCQ4	<i>Propithecus coquereli</i>		

H0XFG2	<i>Otolemur garnettii</i>		
A0A1U7U321	<i>Tarsius syrichta</i>		
P36367	<i>Oryctolagus cuniculus</i>		
A0A2Y9RG03	<i>Trichechus manatus</i>		
G3SX00	<i>Loxodonta africana</i>		
H0VQ15	<i>Cavia porcellus</i>		
G5BGT2	<i>Heterocephalus glaber</i>		
Q8VHG0	<i>Mus musculus</i>		
Q8K4B7	<i>Rattus norvegicus</i>		
A0A1U8BUR7	<i>Mesocricetus auratus</i>		
L9KV99	<i>Tupaia chinensis</i>		
A0A1S2ZVI0	<i>Erinaceus europaeus</i>		
A0A1V4JY45	<i>Patagioneas fasciatus</i>		
A0A093FGE7	<i>Gavia stellata</i>	Aves	
U3I700	<i>Anas platyrhynchos</i>		
M7BB18	<i>Chelonia mydas</i>		
K7FTQ9	<i>Pelodiscus sinensis</i>	Testudines	
NP_036924	<i>Rattus norvegicus</i>		
NP_002012	<i>Homo sapiens</i>		
A0A2R9CCZ4	<i>Pan paniscus</i>		
A0A2J8LLM7	<i>Pan troglodytes</i>		
G1RYT9	<i>Nomascus leucogenys</i>		
H2N4Q7	<i>Pongo abelii</i>		
G3R5S9	<i>Gorilla gorilla</i>		
A0A2K6KT99	<i>Rhinopithecus bieti</i>		
A0A2K6RMJ0	<i>Rhinopithecus roxellana</i>		
A0A2K5JQS5	<i>Colobus angolensis palliatus</i>		
A0A2K6E4Q2	<i>Macaca nemestrina</i>		
G8F2N4	<i>Macaca fascicularis</i>		
F7CKS4	<i>Macaca mulatta</i>		
A0A2K5XVP0	<i>Mandrillus leucophaeus</i>		
A0A2K5LI13	<i>Cercocebus atys</i>		
A0A0D9RGS2	<i>Chlorocebus sabaues</i>		
		Mammalia	<b>FMO1</b>

A0A2K6S2A8	<i>Saimiri boliviensis</i>		
A0A2K5DNS9	<i>Aotus nancymae</i>		
F6TF83	<i>Callithrix jacchus</i>		
A0A2K6GJC5	<i>Propithecus coquereli</i>		
H0XFG0	<i>Otolemur garnettii</i>		
M3YVY3	<i>Mustela putorius</i>		
A0A2Y9JX65	<i>Enhydra lutris</i>		
Q95LA2	<i>Canis lupus</i>		
D2I021	<i>Ailuropoda melanoleuca</i>		
A0A2U3XYA5	<i>Leptonychotes weddellii</i>		
M3W9L4	<i>Felis catus</i>		
P16549	<i>Sus crofa</i>		
G5E5R0	<i>Bos taurus</i>		
W5PSR9	<i>Ovis aries</i>		
L9K VX0	<i>Tupaia chinensis</i>		
A0A1U7U1X5	<i>Tarsius syrichta</i>		
A0A1S2ZVI9	<i>Erinaceus europaeus</i>		
L5LP31	<i>Myotis davidii</i>		
L7N1I9	<i>Myotis lucifugus</i>		
P17636	<i>Oryctolagus cuniculus</i>		
G3TPA3	<i>Loxodonta africana</i>		
A0A218V4N6	<i>Lonchura striata</i>		
A0A091LMM3	<i>Chlamydotis macqueenii</i>		
A0A0A0B033	<i>Charadrius vociferus</i>		
A0A091JJZ4	<i>Egretta garzetta</i>		
A0A091KNM4	<i>Colius striatus</i>	Aves	
A0A091QCB7	<i>Merops nubicus</i>		
A0A093PIP7	<i>Manacus vitellinus</i>		
A0A091G2R8	<i>Cuculus canorus</i>		
K7FK79	<i>Pelodiscus sinensis</i>	Testudines	
Q5REK0	<i>Pongo abelii</i>		
G1RYT1	<i>Nomascus leucogenys</i>		
Q28505	<i>Macaca mulatta</i>		
A0A023JBW5	<i>Macaca fascicularis</i>		
F7FJA6	<i>Callithrix jacchus</i>		
XP_007987720	<i>Chlorocebus sabaeus</i>		
XP_011826204	<i>Mandrillus leucophaeus</i>		

XP_017722374	<i>Rhinopithecus bieti</i>	Mammalia	<b>FMO2</b>
NP_001009008	<i>Pan troglodytes</i>		
NP_001451	<i>Homo sapiens</i>		
XP_004027940	<i>Gorilla gorilla</i>		
XP_010334438	<i>Saimiri boliviensis</i>		
XP_021523737	<i>Aotus nancymaae</i>		
XP_017389638	<i>Cebus capucinus</i>		
M3YVY2	<i>Mustela putorius</i>		
XP_022363881	<i>Enhydra lutris</i>		
G1MF96	<i>Ailuropoda melanoleuca</i>		
XP_008694286	<i>Ursus maritimus</i>		
XP_021538314	<i>Neomonachus schauinslandi</i>		
E2RHC8	<i>Canis lupus</i>		
F6T988	<i>Equus caballus</i>		
XP_014700659	<i>Equus asinus</i>		
XP_014638919	<i>Ceratotherium simum simum</i>		
XP_006872704	<i>Chrysochloris asiatica</i>		
XP_006887484	<i>Elephantulus edwardii</i>		
XP_004454778	<i>Dasypus novemcinctus</i>		
A0A1U7QFV1	<i>Mesocricetus auratus</i>		
XP_003505578	<i>Cricetulus griseus</i>		
XP_006974813	<i>Peromyscus maniculatus bairdii</i>		
G3V6F6	<i>Rattus norvegicus</i>		
Q8K2I3	<i>Mus musculus</i>		
XP_021054156	<i>Mus pahari</i>		
XP_020025620	<i>Castor canadensis</i>		
XP_008846746	<i>Nannospalax galili</i>		
XP_003130152	<i>Sus crofa</i>		
XP_007165362	<i>Balaenoptera acutorostrata scammoni</i>		
XP_010955205	<i>Camelus bactrianus</i>		
W5PS98	<i>Ovis aries</i>		
G5E540	<i>Bos taurus</i>		
XP_005690692	<i>Capra hircus</i>		
XP_010850196	<i>Bison bison</i>		
XP_006075480	<i>Bubalus bubalis</i>		
XP_020759823	<i>Odocoileus virginianus texanus</i>		
G1SND1	<i>Oryctolagus cuniculus</i>		
G3TH74	<i>Loxodonta africana</i>		

XP_012658178	<i>Otolemur garnettii</i>		
XP_019503343	<i>Hipposideros armiger</i>		
S7MVT1	<i>Myotis brandtii</i>		
A0A0P6JH69	<i>Heterocephalus glaber</i>		
A0A091DJC6	<i>Fukomys damarensis</i>		
P36366	<i>Cavia porcellus</i>		
M3W9L1	<i>Felis catus</i>		
XP_007088786	<i>Panthera tigris</i>		
XP_019280219	<i>Panthera pardus</i>		
XP_019601297	<i>Rhinolophus sinicus</i>		
XP_005374888	<i>Chinchilla lanigera</i>		
XP_015347321	<i>Marmota marmota</i>		
L9KUM1	<i>Tupaia chinensis</i>		
K7FEU6	<i>Pelodiscus sinensis</i>	Testudines	
M7BR11	<i>Chelonia mydas</i>		
Q6DF15	<i>Xenopus tropicalis</i>	Amphibia	N/A
Q6PA74	<i>Xenopus laevis</i>		
A0A1L8GN32	<i>Xenopus laevis</i>		
F6Y350	<i>Xenopus tropicalis</i>		
A0A1L8GN11	<i>Xenopus laevis</i>		
Q5BKK2	<i>Xenopus tropicalis</i>		
F6TLP5	<i>Xenopus tropicalis</i>		
F6TKY1	<i>Xenopus tropicalis</i>		
NP_001075714	<i>Oryctolagus cuniculus</i>		
W5QH53	<i>Ovis aries</i>		
A6QLN7	<i>Bos taurus</i>		
NP_001452	<i>Homo sapiens</i>		
H2PZU9	<i>Pan troglodytes</i>		
G3S9Y3	<i>Gorilla gorilla</i>		
G1QH17	<i>Nomascus leucogenys</i>		
A0A2K5NFE4	<i>Cercocebus atys</i>		
A0A096NWS4	<i>Papio anubis</i>		
G7NTQ5	<i>Macaca fascicularis</i>		
A0A1D5Q215	<i>Macaca mulatta</i>		
A0A0D9S0K5	<i>Chlorocebus sabaues</i>		
XP_010382298	<i>Rhinopithecus roxellana</i>		
XP_011806642	<i>Colobus angolensis palliatus</i>		
H2N639	<i>Pongo abelii</i>		
F7ALU6	<i>Callithrix jacchus</i>		

XP_012327405	<i>Aotus nancymaae</i>	Mammalia	<b>FMO5</b>
H0X7S2	<i>Otolemur garnettii</i>		
A0A1U7TPN0	<i>Tarsius syrichta</i>		
E2QUP4	<i>Canis lupus</i>		
XP_023114833	<i>Felis catus</i>		
XP_007092639	<i>Panthera tigris</i>		
XP_014939496	<i>Acinonyx jubatus</i>		
XP_007182187	<i>Balaenoptera acutorostrata scammoni</i>		
XP_004285858	<i>Orcinus orca</i>		
XP_023975388	<i>Physeter catodon</i>		
G1QEL7	<i>Myotis lucifugus</i>		
S7MVU6	<i>Myotis brandtii</i>		
F1SDB7	<i>Sus crofa</i>		
F6PLR9	<i>Equus caballus</i>		
G3T364	<i>Loxodonta africana</i>		
XP_006900388	<i>Elephantulus edwardii</i>		
I3MPC8	<i>Ictidomys tridecemlinea</i>		
P49109	<i>Cavia porcellus</i>		
G5BA06	<i>Heterocephalus glaber</i>		
XP_005413529	<i>Chinchilla lanigera</i>		
XP_004646243	<i>Octodon degus</i>		
A0A091EEE1	<i>Fukomys damarensis</i>		
P97872	<i>Mus musculus</i>		
Q8K4C0	<i>Rattus norvegicus</i>		
A0A1A6H802	<i>Neotoma lepida</i>		
G3HP87	<i>Cricetulus griseus</i>		
S9XQV9	<i>Camelus ferus</i>		
A0A1S3FCC7	<i>Dipodomys ordii</i>		
L5K3R4	<i>Pteropus alecto</i>		
F7FCD2	<i>Monodelphis domestica</i>		
G3WY85	<i>Sarcophilus harris</i>		
A0A1S3A319	<i>Erinaceus europaeus</i>		
XP_004589355	<i>Ochotona princeps</i>		
XP_004651790	<i>Jaculus jaucus</i>		
XP_026633339	<i>Microtus ochrogaster</i>		
E1C7G8	<i>Gallus gallus</i>		
A0A091EER2	<i>Corvus brachyrhynchos</i>		
H1A3R9	<i>Taeniopygia guttata</i>		
U3JME7	<i>Ficedula albicollis</i>		

A0A091UWC8	<i>Nipponia nippon</i>	Aves	
A0A087QJ0	<i>Aptenodytes forster</i>		
A0A091GDX2	<i>Cuculus canorus</i>		
A0A091I0D0	<i>Calypte anna</i>		
G1MZ19	<i>Meleagris gallopavp</i>		
A0A093NXE8	<i>Pygoscelis adeliae</i>		
U3J9N7	<i>Anas platyrhynchos</i>		
M7C2B2	<i>Chelonia mydas</i>	Testudines	
M7BR25	<i>Chelonia</i>		
M7B995	<i>Chelonia</i>		
K7G1T3	<i>Pelodiscus sinensis</i>		
F6T1Q3	<i>Xenopus tropicalis</i>	Amphibia	
Q6PB07	<i>Xenopus laevis</i>		
Q6AX90	<i>Xenopus laevis</i>		
F6T1R3	<i>Xenopus tropicalis</i>		
A0A1L8GN28	<i>Xenopus laevis</i>		
F6S648	<i>Xenopus tropicalis</i>		
A0A1L8GN11	<i>Xenopus laevis</i>		
A0A1L8GN1	<i>Xenopus laevis</i>		
Q4SLV0	<i>Tetraodon nigroviridis</i>	Actinopterygii/ Actinopteri	N/A (root)
A0A0F8AHM2	<i>Larimichthys crocea</i>		
A0A2U9C3W1	<i>Scophthalmus maximus</i>	Sarcopterygii/ Coelacanthimorpha	
H3A7E3	<i>Latimeria chalumnae</i>		
V9KV64	<i>Callorhynchus milii</i>	Chondrichthyes	
V9KI45	<i>Callorhynchus milii</i>		
C3ZT14	<i>Branchiostoma floridae</i>	Cephalochordata	

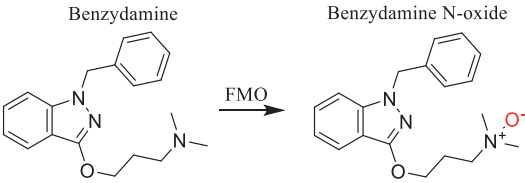
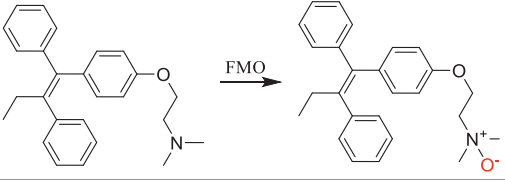
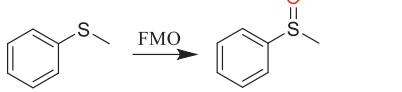
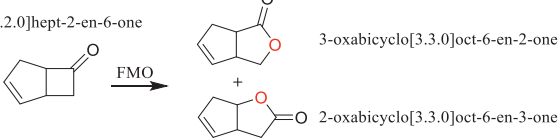
**Table S1. Jawed vertebrates FMOs dataset.** The protein accession codes (Genbank or Uniprot) are given for each of the sequences collected with the corresponding species names. The taxonomic rank is given as the class or order (for terrestrial tetrapods) and superclass/class or phylum (for the root sequences). The FMO paralog clade was assigned for each sequence based on the phylogeny and the sequences from *Homo sapiens* are written in red for reference. N/A: not assigned. The two studied FMO paralog clades FMO1 and FMO4 are highlighted in lime in agreement with the coloring of the phylogeny.



**Table S2. Melting temperatures of the AncFMOs.** Melting temperatures were measured in triplicate by using the ThermoFAD technique for AncFMOs 2, 3-6 and 5, and the TychoTMNT.6 system for AncFMO1 (see materials and methods). ahFMO3 melting temperature in Triton X-100 reduced (personal communication F. Fiorentini).

protein	melting temperature ( $T_m$ ) / °C	
	no NADP <sup>+</sup>	200 $\mu$ M NADP <sup>+</sup>
<b>AncFMO1</b>	47	52
<b>AncFMO2<sup>1</sup></b>	53	70
<b>AncFMO3-6<sup>1</sup></b>	60	66.5
<b>AncFMO5<sup>1</sup></b>	55	59
<b>hFMO3<sup>a</sup></b>	44.5	44.5
<b>hFMO5(1)</b>	49	48.5

**Table S3. Conversions catalyzed by AncFMO1.** Reactions were incubated at 30 °C for 18 hours, at pH 7.5 in the presence of a NADPH recycling system. *a* Conversion of Baeyer-Villiger substrates and thioanisole were measured on GC-MS while N-oxides were identified on HPLC. *b* Enantiomeric excess was measured by HPLC. *n.a.*: not applicable, *n.d.*: not determined.

type of oxidation	Reaction	conversion <sup>a</sup>	<i>ee</i> (%) <sup>b</sup>
N-oxidation	<p>Benzydamine</p>  <p>Benzydamine N-oxide</p>	>99%	n.a.
	<p>Tamoxifen</p>  <p>Tamoxifen N-oxide</p>	>99%	n.a.
S-oxidation	<p>Thioanisole</p>  <p>Thioanisole sulfoxide</p>	>99%	96 ( <i>R</i> )
B-V oxidation	<p>bicyclo[3.2.0]hept-2-en-6-one</p>  <p>3-oxabicyclo[3.3.0]oct-6-en-2-one + 2-oxabicyclo[3.3.0]oct-6-en-3-one</p>	40%	n.d.

## References

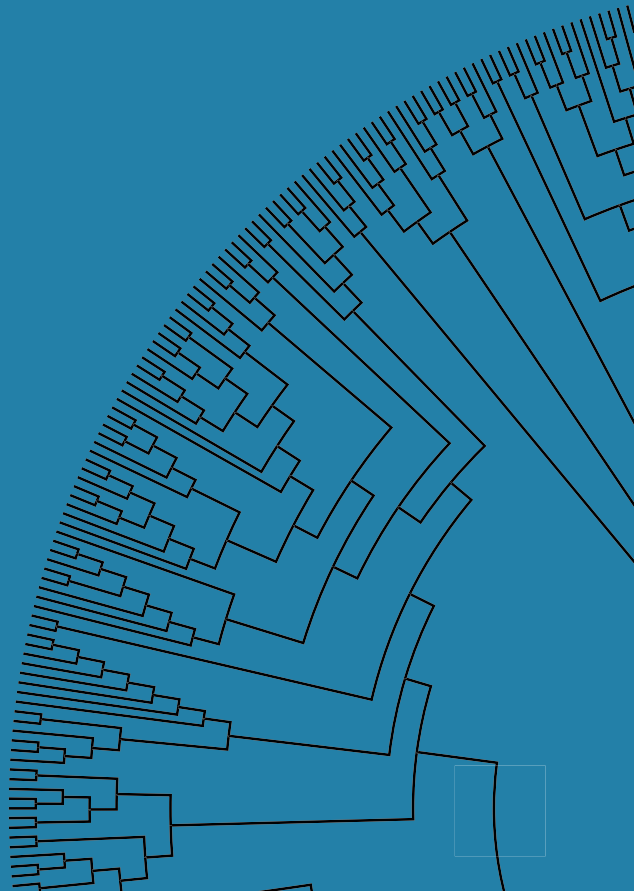
- Nicoll, C. R. et al. Ancestral-sequence reconstruction unveils the structural basis of function in mammalian FMOs. *Nat. Struct. Mol. Biol.* 27, 14–24 (2020).



## CHAPTER V

# Reconstruction and characterization of ancestral tetrapod flavin-containing monooxygenases reveal the evolutionary path to the extant paralogs

Gautier Bailleul  
Callum R. Nicoll  
Andrea Mattevi  
María Laura Mascotti  
Marco W. Fraaije

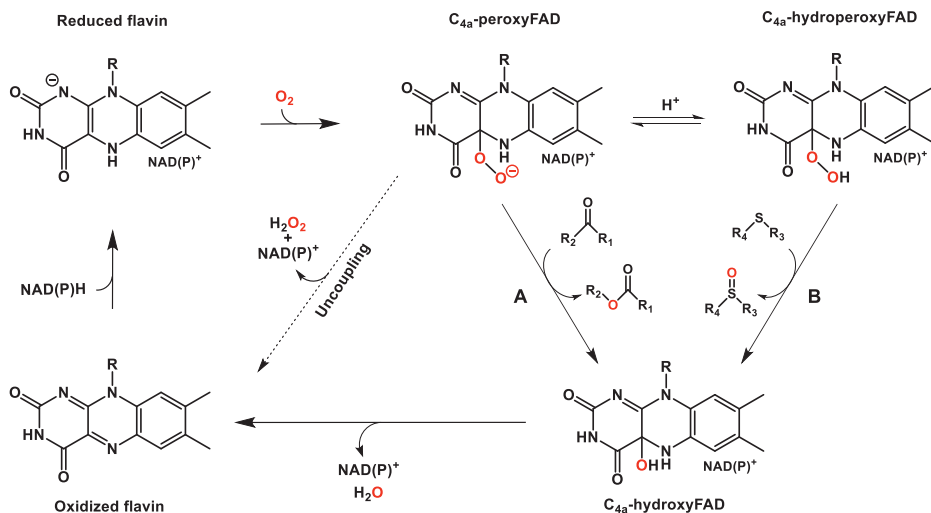


**Abstract**

Mammals typically contain 5 different flavin-containing monooxygenases (FMOs) in their proteome. These FMOs play an essential role in detoxification in mammals and are involved in diseases or drug activity within humans. While the 5 paralogs share a high sequence identity, recent studies have shown that FMO1, FMO2 and FMO3 behave as heteroatom oxidizing enzymes while FMO5 performs Baeyer- Villiger (B-V) oxidations. We have performed an ancestral sequence reconstruction (ASR) of tetrapod FMOs which confirmed a rapid extension of the set of FMOs in tetrapods starting from one parent tetrapod FMO. Several ancestral tetrapod FMO sequences were reconstructed and the respective proteins were expressed and studied concerning their monooxygenase activities. The primordial tetrapod ancestor, tAncFMO1-5, possessed both activities (heteroatom oxidation and B-V oxidation) and the first duplication event resulted in a FMO that lost B-V oxidation activity (tAncFMO1-4) while this atypical oxygenation activity remained in the other FMO descendent, tAncFMO5. Multiple duplication events led to 5 distinct FMO lineages of which 4 display functional overlap (heteroatom oxygenation by FMO1- 4) and a FMO lineage specialized in B-V oxidations (FMO5). A careful analysis of amino acid replacements during evolution and experimentally validating their effects, revealed that a small number of mutations was enough to abolish B-V oxidation activity in the primordial tetrapod FMO, while remaining the heteroatom oxygenation activity. These results shed light on the evolution of human FMOs, revealing the adaptations in time that resulted in different catalytic properties.

## Introduction

Flavoprotein monooxygenases (FPMOs), as their name indicates, possess a flavin cofactor as prosthetic group. By virtue of the oxygen reactivity of the flavin cofactor and using electrons provided by a reduced nicotinamide cofactor, FPMOs incorporate one oxygen atom derived from molecular oxygen into their target substrate<sup>1</sup>. FPMOs are widespread in nature and can be classified into 6 classes, class A-F<sup>2</sup>. Class B FPMOs harbors three subgroups, each sharing a common origin and type of activity: (i) Baeyer-Villiger monooxygenases (BVMOs) that converts ketones into esters or lactones, (ii) N-hydroxylating monooxygenases, and (iii) flavin-containing monooxygenases (FMOs) that perform oxygenations of heteroatoms<sup>2</sup>. The difference in their activities is thought to be due to the protonation state of the C4 $\alpha$ -(hydro)peroxyflavin intermediate that is formed upon reduction of the FAD cofactor and a subsequent reaction with dioxygen. Baeyer-Villiger reactions and epoxidations require this reactive flavin intermediate to be deprotonated for performing the nucleophilic attack of the substrate (Figure 1A). On the other hand, in order to accomplish sulfoxidations or hydroxylations, the protonated C4 $\alpha$ -hydroperoxyflavin is required for an electrophilic attack<sup>3</sup> (Figure 1B).



**Figure 1.** Catalytic cycle of class B flavoprotein monooxygenases. The recurring molecular structure represents the isoalloxazine moiety of the FAD cofactor. A nucleophilic attack (A) occurs during B-V reactions while an electrophilic attack (B) is needed for sulfoxidations and N-oxidations. R corresponds to the ribityl adenosine part of FAD.

While still being understudied compared to their cytochrome P450 monooxygenase relatives<sup>4</sup>, human flavin-containing monooxygenases (hFMOs) are now receiving more attention for their detoxifying role in the human body<sup>5-7</sup>. An increasing amount of studies show that hFMOs are linked to diseases like trimethylaminuria<sup>8,9</sup> and catalyze key steps in drug metabolism<sup>10,11</sup>. Humans possess 5 FMOs paralogs displaying varying tissue-dependent expression patterns: hFMO1 and hFMO2 are mainly present in the kidneys and lungs, respectively, while hFMO3 and hFMO5 are predominant in the liver<sup>12</sup>. Recently, it was proven that, unlike hFMO1, hFMO2 and hFMO3 that perform N- and S-oxidations<sup>13-15</sup>, hFMO5 catalyzes Baeyer-Villiger oxidations<sup>16</sup>. hFMO4 has not yet been characterized, seemingly because of a C-terminus extension that prevents its expression in mammalian cells and in heterologous expression hosts<sup>17</sup>. The biochemical and structural investigation of mammalian FMOs, however, has often been limited due to their challenging isolation and relatively low stability<sup>18-20</sup> likely as a consequence of their membrane-bound nature. As such, the structural features of mammalian FMOs were uncharted. Very recently however, in an attempt to understand the evolution behind mammalian FMOs, ancestrally reconstructed mammalian FMOs were able to provide the first structural models through crystallization<sup>21,22</sup>. The four reconstructed mammalian paralogs behaved as their human extant counterparts and exhibited the same type of activity, *i.e.*: AncFMO5 catalyzing B-V oxidations and AncFMO1, AncFMO2 and AncFMO3 performing N- and S- oxygenations. The structures of all reconstructed enzymes presented many similarities, including a well-conserved dinucleotide-binding domain (Rossmann fold), a buried active site, and several membrane binding features such as the large transmembrane C-terminal helix and an 80-residue insertion that provided tunnel features for substrate transit<sup>22</sup>. However, inspection of the active site revealed a glutamate at position 281 in AncFMO2 (mAncFMO2) and AncFMO3-6 (mAncFMO3-6) which could explain the different functionalities observed when compared with mAncFMO5, that exhibits a histidine in the same structural location and position in sequence. The difference translates in either a deprotonated residue for AncFMO2 and AncFMO3-6 or a positively charged histidine for AncFMO5. The character of this residue may be of key importance to tune the reactivity of the peroxyflavin intermediate. Yet, by substituting this residue in AncFMO2 and AncFMO3-6, we were only partially able to reduce some FMO activity without introducing B-V activity<sup>22</sup>. This suggests that probably multiple residues are responsible for the variation in activities observed within the mammalian FMOs clade.

While our previous ASR studies focused on resurrecting and investigating mammalian FMO ancestors, a more complete evolutionary ASR study of animal FMOs is possible through the availability of a large number of animal genome sequences. Concerning the evolutionary history of metazoan FMOs, the most striking observation was an explosion event whereby one gene from the tetrapod ancestor underwent several duplication events. This resulted in most animals, such as amphibians, birds, reptiles and mammals, to possess up to six different FMOs. In biochemical terms, the main outcome of this “FMO explosion” is a broader phenotypic diversity of more recent and

extant FMOs by functional divergence and specialization, coinciding with redundancy due to overlapping activities. A careful analysis of ancestral FMOs should allow to retrace effects of changes in FMO sequence during evolution. In this work we reconstructed and characterized the ancestral FMOs from tetrapods from before and after the FMO explosion event. By using an integrative approach combining molecular phylogeny, bioinformatics, structural analysis, mutagenesis and biochemical characterization, we determined the properties of the resurrected ancestors and unveiled the sequence determinants for the switch in BVMO (B-V oxidation) – FMO (heteroatom oxidation) activity.

## Results & Discussion

### *Ancestral sequence reconstruction - unveiling the history of tetrapod FMOs*

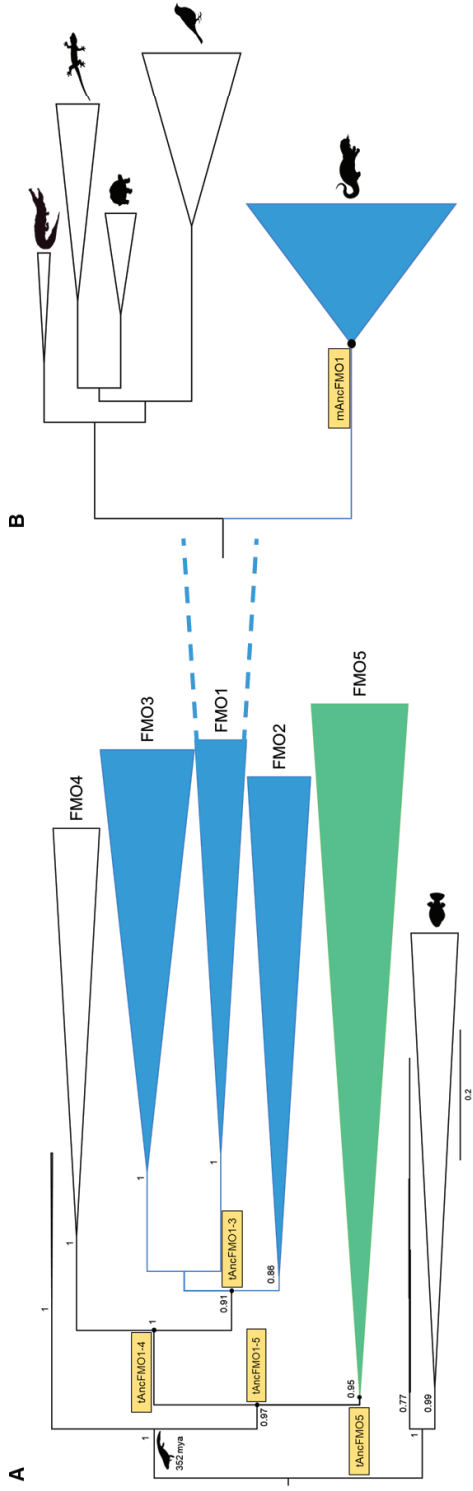
The first step in performing ancestral sequence reconstruction (ASR) is to build a reliable phylogeny. In order to recover the most likely sequences of all tetrapod ancestors, we reinforced the dataset that was previously used for the reconstruction of the mammalian FMOs. Additional sequences were either part of the testudines, crocodylia, aves, or lepidosauria clades and the root was further populated with Actinopterygii (bony fishes) (Figure S1). The new tree confirmed the evolutionary paths taken by the different FMOs<sup>21,22</sup>. Employing the new phylogeny, we recovered 4 ancestral sequences named tAncFMO1-5, tAncFMO5, tAncFMO1-4 and tAncFMO1-3 (“tAnc” stands for tetrapod ancestral) that correspond to the emergence of each of the paralog clades (Figure 2). Noticeably, all the ancestors have the same length of 532 amino acids and were reconstructed with high confidence (overall posterior probabilities, PP > 0.94) (Figure 3). The earliest tetrapod ancestor enzyme, tAncFMO1-5 (PP = 0.95), is the starting point from which all duplication events originate, thus the pre-explosion ancestor. The FMO5 clade is the first to diverge in the family with tAncFMO5 (PP = 0.97) as the corresponding reconstructed ancestor. A second duplication event led to the creation of the FMO4 clade and the most likely ancestors from before and after this split were dubbed tAncFMO1-4 (ancestor of FMOs 1-4) and tAncFMO1-3 (ancestor of FMOs 1-3) with PPs of 0.94 and 0.95, respectively (Figure 2 & 3). The FMO2 clade was the next to emerge, followed by FMO1 and FMO3, with the latter duplicating further in mammals to give rise to FMO6, a pseudogene being present only in mammals<sup>23</sup>.

The first duplication event from tAncFMO1-5 generated the lineages of FMO5 and FMO1-4. tAncFMO5 harbors 28 amino acid substitutions when compared to tAncFMO1-5, while tAncFMO1-4 contains 41 substitutions. The previously reconstructed mammalian FMO5 (mAncFMO5) shares a high identity with tAncFMO5 with 399 identical sites out of 532 (75% identity). Furthermore, the identity between the ancestor of the FMO1 to FMO3 clade (tAncFMO1-3) and its predecessor that contains additionally the FMO4 clade (tAncFMO1-4) was also high (95% identity). Interestingly, unlike the mammalian sequences, the other tetrapod’s FMO4 and hence the tAncFMOs did not contain the classic C-terminal extension, being a unique feature of mammalian FMO4s. The high



conservation between all tetrapod FMOs suggests that fundamental structural features are likely to remain conserved and therefore that the different catalytic traits are due to changes in a specific subset of residues.

Three equally parsimonious hypotheses about the switch between the heteroatom and B-V oxidation activities can be proposed from this phylogeny. First, the tetrapod ancestor was behaving as a BVMO and one of the duplication events introduced the changes that enabled N- and S-oxygenations. Alternatively, the tetrapod ancestor was only able to oxygenate heteroatoms and the ability to perform B-V oxidations only appeared within the emerging FMO5 clade. Third, the tetrapod ancestor was promiscuous, bearing both oxygenation activities, and FMOs specialized over time with duplication events. In order to elucidate the origin of the switch and the residues responsible for it, the four tAncFMOs mentioned above were expressed and functionally characterized.



**Figure 2. Ancestral sequence reconstruction of tetrapod FMOs.** A, Phylogeny of FMOs from jawed vertebrates with the reconstructed ancestors (yellow boxes) at the nodes. The five FMO clades are colored based on the mammalian ancestor activity, in light green for B-V and blue for FMO. The depicted silhouettes represent the ancestral tetrapod ( ) and bony fishes ( ). The support values (TBE) are indicated at the major divergences. B, Close up of the FMO1 clade as a representative of the other 4 paralog clades. Taxonomic distribution is depicted with silhouettes as follows: mammals ( ), aves ( ), testudines ( ), lepidosauria ( ) and crocodylia ( ). The blue color represents the activity of the mammalian ancestor previously reconstructed (mAncFMO1).

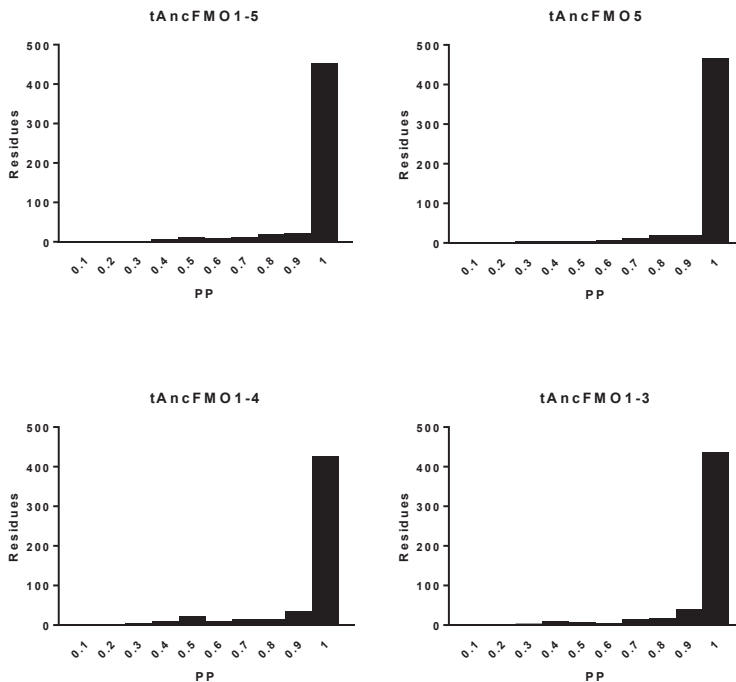


Figure 3. Relative uncertainty in the reconstruction of tAncFMOs. Histograms indicate the repartition of reconstructed residues based on their posterior probabilities (PP).

### The functional portfolio of tAncFMOs

Codon-optimized genes encoding the four ancestral protein sequences were obtained as synthetic DNA fragments and cloned into an expression vector. The FMOs were produced using *E. coli* as expression host. Alike mammalian FMOs, the expressed tetrapod FMOs were localized in the membrane fraction and could be purified with yields ranging between 10 to 20 mg per liter of culture. As expected, the purified enzyme samples were yellow colored, indicative for a bound flavin cofactor. UV-Vis absorbance spectra of all purified enzymes were highly similar to the spectra reported for the ancestral mammalian FMOs<sup>21,22</sup>.

The enzymatic behavior of all reconstructed enzymes was assessed using a specific substrate panel of seven compounds to test the three oxidation activities of interest: N- and S-oxidations and B-V oxidation (Figure S2 & S3). Benzylamine and tamoxifen, anti-inflammatory and antiestrogen drugs, respectively<sup>24,25</sup>, are known to be subject to N-oxidation by FMOs<sup>26,27</sup>. In order to follow the S-oxidation potential, the aromatic thioethers methyl-*p*-tolyl sulfide and the bulky benzyl phenyl sulfide<sup>28,29</sup> were 0.1 0.2 included in the substrate panel. Finally, the three ketones chosen to assess the B-V activity were the aliphatic hepta-2-one, the alicyclic cyclohexanone and the aromatic phenylacetone<sup>16,30,31</sup>.

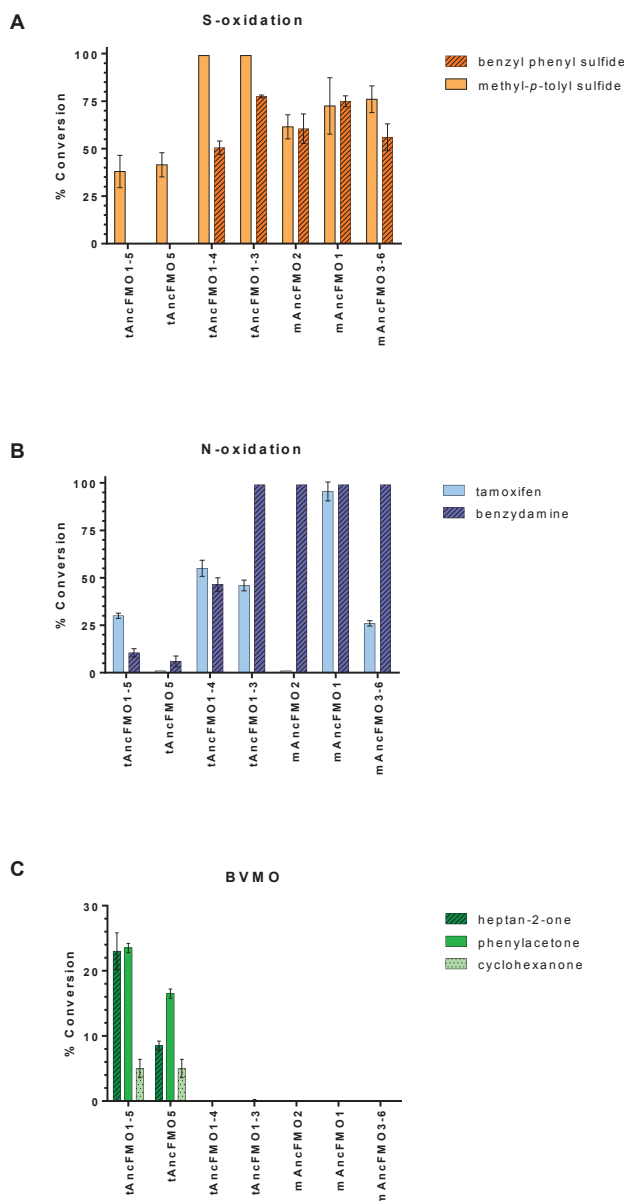
Thioether oxidation conversions (Figure 4A) showed that all tAncFMOs were able to accept methyl-*p*-tolyl sulfide as a substrate but with different efficiencies. While tAncFMO1-5 and tAncFMO5 transformed less than half of the total amount (38% and 42%, respectively), tAncFMO1-4 and tAncFMO1-3 were able to fully convert the aromatic thioether. Benzyl phenyl sulfide did not result in any conversion from either the original tetrapod ancestor, tAncFMO1-5 or the FMO5 ancestor, tAncFMO5. Yet, tAncFMO1-4, tAncFMO1-3 and all tested mammalian ancestors (mAncFMO1, mAncFMO2 and mAncFMO3-6) were found to convert this bulky thioether confirming the trend of heteroatom oxidation specialization during evolution.

Alike the thioethers, the two substrates assessed for N-oxidation were mainly converted by enzymes that emerged from tAncFMO1-4 (Figure 4B). tAncFMO1-5, tAncFMO5 and tAncFMO1-4 converted 10, 6 and 46 % of benzydamine, respectively, whilst tAncFMO1-3 and mammalian ancestral FMO1 to 3 carried out N-oxygenation to completion. Tamoxifen was converted by tAncFMO1-5 and tAncFMO1-4, while the FMO5 tetrapod ancestor, tAncFMO5, was barely able to produce any N-oxide. A clear distinction occurred within mammalian ancestral FMOs as mAncFMO1 was able to fully convert both tamoxifen and benzydamine, as previously demonstrated<sup>21</sup>, while mAncFMO2 did not accept tamoxifen. This difference in the substrate specificity most-likely is a unique feature of FMO2 that emerged after the tAncFMO1-3 duplication event as mAncFMO3-6 and mAncFMO1 are both able to convert tamoxifen.

Intriguingly, all three B-V substrates were only converted by tAncFMO1-5 and tAncFMO5 (Figure 4C). This is in line with the observation that hFMO5 is also able to perform B-V oxidations<sup>32</sup>. None of the other resurrected ancestors, including tAncFMO1-4, were able to perform B-V oxidations.

These results provide the answers to the two main questions of this work: which were the ancestral functions of the tetrapod FMOs and when did the switch between the heteroatom oxygenations and B-V activities arise. The reconstructed tAncFMO1-5 is able to perform both B-V oxidation and heteroatom oxygenation activity. The first duplication event that resulted into tAncFMO5 and tAncFMO1-4 was crucial for the activity switch that we now observe between human and other mammalian FMO5 and their paralogs. Despite the fact that tAncFMO5 is still able to perform B-V reactions, it lost some of its typical FMO-type activity. On the other hand, tAncFMO1-4 fully lost its B-V oxidation ability and instead further improved its S- and N-oxidation capacities. Conversion of tamoxifen and benzydamine increased significantly between tAncFMO1-5 and tAncFMO1-4, respectively. Additionally, unlike tAncFMO1-5, tAncFMO1-4 is able to convert benzyl phenyl sulfide, a typical extant FMO-type activity. The primordial tetrapod B-class FPMO had the ability to catalyze both B-V oxidations and heteroatom oxygenations and this dual-activity was later split between tAncFMO5 and tAncFMO1-4, respectively (Figure 4). This can be described as a case of functional optimization after gene-duplication<sup>33</sup>. Conversions with previously reconstructed mammalian ancestral FMOs support the hypothesis about the specialization of FMOs following the duplication events. By inspecting the substitutions that occurred between tAncFMO1-4 and tAncFMO1-5 it

should be possible to dissect the determinants that enable FMOs to specialize towards distinct oxidation activities, B-V oxidations or heteroatom oxygenations.

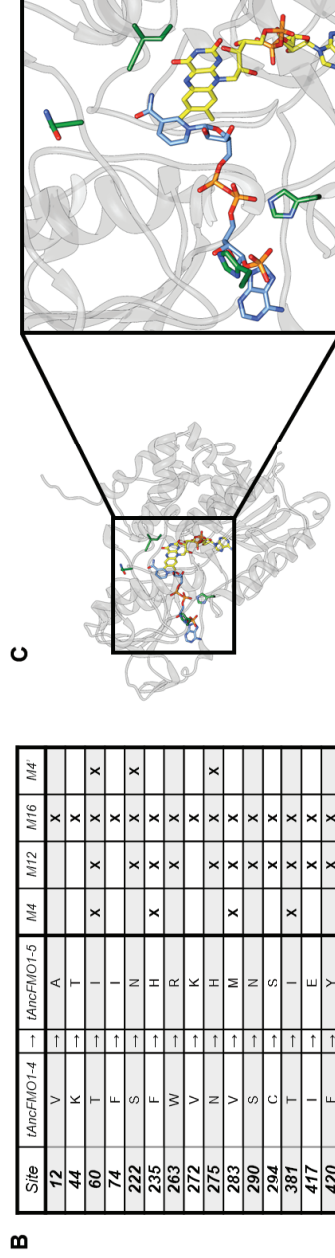


**Figure 4. Conversions by ancestral tetrapod and mammalian FMOs. A, Sulfoxidation of benzyl phenyl sulfide (striped dark orange) and methyl-p-tolyl sulfide (solid light orange). B, N-oxidation of tamoxifen (solid light blue) and benzydamine (striped dark blue). C, Baeyer-Villiger reaction toward heptan-2-one (striped dark green), phenyl acetone (solid green) and cyclohexanone (dotted light green). Error bars correspond to standard deviation.**

***Dissecting the determinants of tAncFMOs function***

With the aim of finding the residues responsible for the functional switch observed among the N- and S-oxidizing FMOs lineage (*i.e.*: the one derived from tAncFMO1-4) and the B-v oxidizing FMO lineage (*i.e.*: the one derived from tAncFMO5) we embarked on a search for amino acid replacements to revert the catalytic profile of tAncFMO1-4 to tAncFMO1-5. This enzyme engineering exercise should unveil the set of residues sufficient for the FMO activity switch. To identify the most relevant residues, mutant enzymes were designed based on (i) the accuracy of reconstruction of each residue, (ii) the degree of conservation within all tetrapod sequences, and (iii) the structural role and importance based on the information from previously solved mAncFMO structures (Figure 5A).

Site	ASR		Degree of conservation			Structural Input
	tAncFMO1-5 MAP	tAncFMO1-4 MAP	Consurf	BVMOs	FMOs	
12	1.00	0.98	4	-	-	In the heart of the FAD binding domain. Follows after GxGxxG motif.
44	1.00	1.00	6	-	-	Surface residue
60	1.00	1.00	8	I	T/S	Sits behind FAD ring in proximity to C=O bond at the 4 position.
74	1.00	0.94	7	I	(variable)	Surface residue with sidechain pointing inwards towards the FAD binding domain.
222	1.00	0.98	7	N	S	Surface residue in the 80-residue insertion. In vicinity of dimerisation interface.
235	0.47	0.97	1	-	-	The histidine sticks inwards towards the main part of tunnel
263	1.00	0.98	5	-	-	Surface residue
272	0.85	0.55	3	-	-	Surface residue of Loop 3
275	1.00	0.72	7	H	(variable)	Surface residue in Loop 3 that covers the FAD face. Histidine shares hydrogen bonds with Gln281 that subsequently coils Loop 3
283	0.80	0.46	7	M/T	V/I	Very close to the active site
290	1.00	0.98	8	N	(variable)	Surface residue in NADP+ binding domain. May be important for dimerization.
394	1.00	1.00	7	S	(variable)	Surface residue that may be useful for dimerisation
381	1.00	0.85	7	I	T/G	Buried residue, not in the tunnel
417	0.72	0.88	2	-	-	Present on large alpha helix preceding loop 2
420	1.00	0.81	7	Y	F/G	Part of Loop 2 that reaches out towards the NADP+ molecule and blocks a tunnel access point in mAncFMO1.
426	1.00	0.77	6	-	-	Part of Loop 2 (see above). Histidine forms hydrogen bonds to Loop 3 that anchors His282 into place



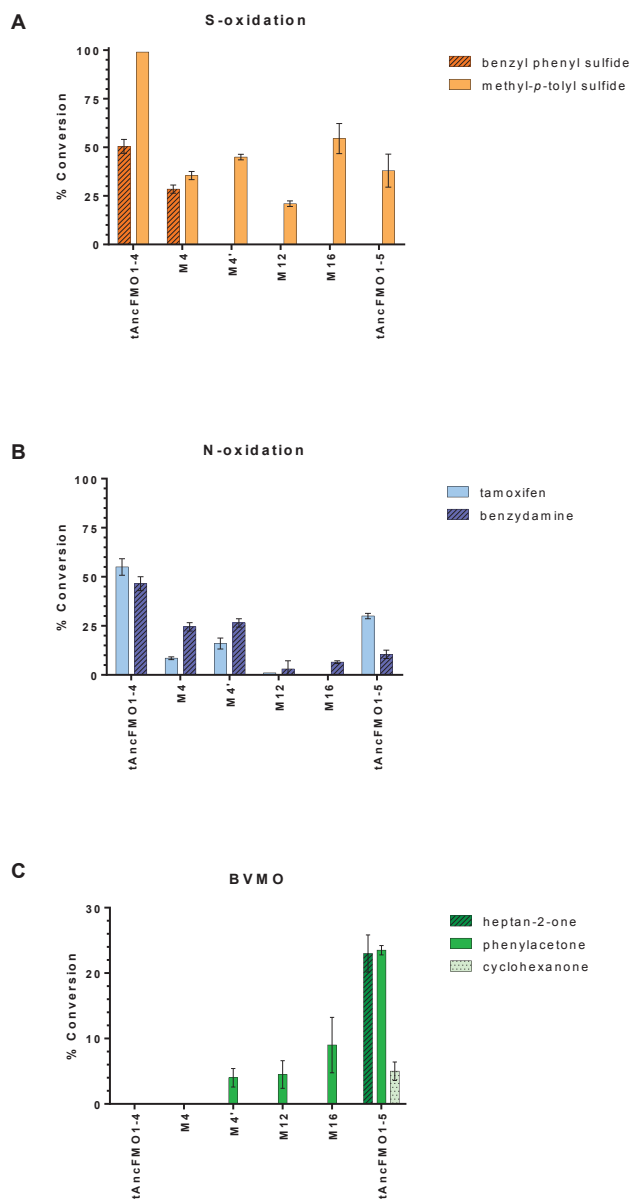
**Figure 5. Mutant design methodology.** A, Analysis of the selected sites for mutations. The ASR column shows the maximum a posteriori state (MAP) of each site and its posterior probabilities of the reconstruction. The degree of conservation was computed by the Consurf server using the multiple sequence alignment of all the gathered tetrapod sequences. The score ranges from 1 to 9 for variable to conserved sites, respectively. When the score is 7 or higher, it indicates that the amino acid in the ancestral enzyme may impact/play a key role in B-V reactions (tAncFMO1-5, mAncFMO5, mAncFMO5) or heteroatoms oxygenation (tAncFMO1-4, tAncFMO1-3, mAncFMO1, 2, 3 & 4). Structural inputs are based on previously resolved crystal structures from mAncFMOs and present relevant information about the structural position of residues and potential impact on catalysis. B, Mutated tAncFMO4 residues in M4, M4', M12 and M16. C, On the left, overall structure of mAncFMO5 with the 4 mutations from M4' highlighted in green and zoomed in the black box. The FAD is represented in yellow and NADP+ in cornflower blue.

This strategy allowed us to propose three mutants using the tAncFMO1-4 sequence as a basis. The mutants were named, M4, M12 and M16, based on the number of substitutions, 4, 12 and 16, respectively, with each subsequent mutant containing additional mutations when compared with the previous mutant (Figure 5B). The three mutants were expressed, purified and assessed with the same substrate panel as used for the previously characterized ancestral FMOs (see above). While M4 was still able to convert benzyl phenyl sulfide, the additional eight mutations in M12 resulted in full loss of activity towards this thioether (Figure 6A). The other tested S-containing substrate, methyl-*p*-tolyl sulfide, was fully converted by tAncFMO1-4 while all mutants were less efficient in oxidizing this thioether (20-60% conversion), similar to tAncFMO1-5 (40% conversion). Conversion of the N-containing test substrates, benzydamine and tamoxifen, by the three mutant FMOs revealed a stepwise decrease in oxygenation efficiency, eventually resulting in complete loss of activity on tamoxifen by M16 (Figure 6B). The conversion level of benzydamine by M16 is similar to the performance of tAncFMO1-5. Yet, tAncFMO1-5 was able to convert up to 28% of tamoxifen, more than any of the mutants, suggesting that one (or several) of the non-selected residues has a positive effect on accepting tamoxifen as substrate.

Next, the capabilities of the designed mutants for performing B-V oxidations were probed. This revealed that cyclohexanone and hepta-2-one, that are substrates for tAncFMO1-5, were not converted by either tAncFMO1-4 or any of the designed mutants (Figure 6C). Nonetheless, low but significant B-V oxidation was observed with phenylacetone: M12 and M16 converted 4 % and 9 % of the substrate, respectively. The fact that the B-V activity was retrieved in M12, albeit partially, indicates that at least some of the residues that trigger B-V oxidation activity are found among the 12 mutations. With four additional mutations, M16, slightly higher phenylacetone conversions were observed (Figure 6C). Based on these results, a second round of mutational analysis was performed based on a structural analysis of the surroundings of the 12 sites of M12 and their potential effect on enzyme activity. Based on a close structural inspection of the residues, we decided to create an alternative 4-fold mutant: M4'. In this variant of tAncFMO1-4, the T60I mutation was kept while the other three mutations were different from M4 (see Figure 5B). Threonine 60 is positioned close to the isoalloxazine ring of the flavin cofactor and its side chain points towards the active site cavity while being in close proximity to the C4-O4 carbonyl bond of the isoalloxazine ring (Figure 4C). Whilst it is not clear as to how the switch would perturb the pKa of the C4 $\alpha$ -peroxy intermediate and thus the protonation state, the isoleucine residue is well conserved among B-V oxidizing FMOs. The threonine is highly conserved among extant and ancestral FMOs 1-4. Therefore, it was decided to keep this residue in M4'. In line with the conservation among B-V oxidizing and heteroatom oxidizing FMOs, serine 222 was selected based on its high conservation in the latter. Additionally, it is present in the 80-residue insertion that form the substrate access tunnels. The other two residues, asparagine 275 and asparagine 426, were selected based on a hydrogen bonding network expected to be introduced when substituting both by a histidine. At both positions, histidine residues are



well conserved in all B-V oxidizing FMOs. Interestingly, the asparagine residues are not completely conserved among heteroatom oxidizing FMOs. M4' was expressed, purified and studied for its catalytic properties. The enzyme presented an N- and S- oxygenation activity profile similar to that of M4 except for the complete disappearance of benzyl phenyl sulfide conversion (Figure 6A & 6B). Interestingly, the four mutations enabled this variant of tAncFMO1-4 to convert phenylacetone into benzyl acetate (Figure 6C). With only four mutations we were able to reintroduce the B-V activity that was still absent in tAncFMO1-4, albeit with a low efficiency (4% conversion). This indicates that one or more residues in the four residues selected for M4' are sufficient to trigger B-V activity.



**Figure 6.** Conversions by *tAncFMO1-4* mutants and *tAncFMO1-5s*. **A**, Sulfoxidation of benzyl phenyl sulfide (striped dark orange) and methyl-*p*-tolyl sulfide (solid light orange). **B**, N-oxidation of tamoxifen (solid light blue) and benzydamine (striped dark blue). **C**, Baeyer-Villiger oxidation of heptan-2-one (striped dark green), phenylacetone (solid green) and cyclohexanone (dotted light green). Error bars correspond to standard deviation.

## Conclusion

A phylogenetic analysis of the FMO protein family has revealed that a rapid expansion of distinct FMO variants occurred during the emergence of terrestrial vertebrates (tetrapods), around 352 million years ago<sup>34</sup>. This “FMO explosion” within the tetrapods is in line with the observation that bony fishes exhibit only one copy of a FMO gene. The emergence of tetrapods coincides with land colonization and the multiple duplication events leading to different FMOs may be due to associated changes in the tetrapod environment<sup>35,36</sup>. A strong selective pressure to deal with different and new metabolites that were not present in the sea might have given an advantage to organisms with various types of FMOs<sup>37</sup>. The higher amount of oxygen present in the air than in water also may have put a selective pressure towards the duplication of oxygenases to deal with xenobiotics and to develop other detoxification processes<sup>38</sup>.

Extant FMOs maintained their catalytic flexibility during evolution and are known to accept a relatively broad range of small or bulky substrates such as trimethylamine, tamoxifen and benzydamine (Figure S3)<sup>11,27,39</sup>.

Characterization of ancestrally reconstructed tetrapod FMOs unveiled how the catalytic behavior evolved over time and is summarized in Figure 7. The tetrapod ancestor possessed one enzyme, tAncFMO1-5, that was able to carry out both electrophilic and nucleophilic attacks, resulting in either S-/N- or B-V oxidation, respectively. This bi-functional ancestor evolved by gene duplication into specialized enzymes that only pursue one of these activities. The members of the first lineage, the tetrapod FMO5s, almost fully lost their ability to catalyze N-oxidations and mainly focus on B-V oxidations. The second lineage originating from tAncFMO1-5 finally resulted in the emergence of the extant FMO1s to FMO4s. The first enzyme emerging from this lineage, tAncFMO1-4, lost its ability to execute oxidation of carbonyl moieties (B-V oxidations) but specialized in heteroatom oxygenations. A second duplication event resulted in the divergence of tAncFMO1-4 into the FMO4 clade and a clade composed of FMO1s, FMO2s and FMO3s. The reconstructed tAncFMO1-3 further increased its monooxygenase activity toward organic sulfides and N-containing molecules without recovering any of the B-V oxidation activity (Figure 7A).

The first tetrapod ancestor, tAncFMO1-5, possessed mixed functionality, acting both as a heteroatom and as a Baeyer-Villiger monooxygenase. During evolution, tAncFMO1-4 evolved by 41 mutations that abolished the B-V activity. Nonetheless, these changes resulted in an over two-fold activity increase in both N- and S- monooxygenation for the tested substrates. This indicates that during evolution from tAncFMO1-5 to tAncFMO1-4, the FMO has evolved such that it became efficient as heteroatom oxidizing monooxygenase while losing BVMO activity. In this work, we also attempted to reverse this event by installing B-V monooxygenation activity using the tAncFMO1-4 backbone. Making 12 changes essentially extinguished both N- and S- heteroatom oxygenations (except for methyl-*p*-tolyl sulfide) highlighting the importance of these residues. At the same time, these 12 residues were also able to introduce B-V oxidation activity. A similar effect could be achieved by changing just four residues. Importantly, however, these

changes were not sufficient to fully restore the activities observed for tAncFMO1-5, implying that the other 37 residues are pivotal for substrate acceptance and turnover. These findings substantiate the notion that epistasis is the underlining culprit behind the functional shift. In order to retrace the determinants that support B-V oxidation catalysis by tAncFMO1-5 and the extant hFMO5, and that are absent in tAncFMO1-4, several mutants of tAncFMO1-4 were studied. It was found that by merely four mutations, introducing the corresponding tAncFMO1-5 residues, some B-V oxidation activity could be restored (Figure 7B). These four residues (Thr60, Ser22, Asn275 and Asn426 in tAncFMO1-4) were chosen based on three criteria. First, the extent of conservation between B-V oxidizing and heteroatom oxygenating FMOs was considered. Thr60 and Ser222 are highly conserved in heteroatom oxygenating FMOs, whilst Ile60 and Asn222 are found in FMOs catalyzing B-V oxidations. Whilst His275 and His426 are conserved in FMOs catalyzing B-V oxidations, the heteroatom oxygenating FMOs display several different residues at these positions. Second, the location of each residue in the FMO structure was considered. The only residue that is close to the active site is Thr60. The side chain of Thr60 is very close to the carbonyl bond (C4-O4) of the FAD. The third criterium concerns the introduction or omission of non-covalent interactions. This was cross-checked using the mammalian AncFMO crystal structures. Both His275 and His426 stood out on account of their cooperating hydrogen bond network. Together, they effectively anchor the highly flexible loop at a focal point represented by the peptide bond of His282 and the sidechain of Gln281. These four residues were enough to induce B-V monooxygenation.



**Figure 7. Summary of evolutionary history and functional divergence of tetrapods FMOs.** *A*, Reconstructed *tAncFMOs* are represented with diamonds at the corresponding nodes. Colors are based on the catalytic behavior of *tAncFMOs* or *mAncFMOs*: red for the oxygenation of heteroatoms and green for the B-V oxidations. *B*, Conversion results of *tAncFMO1-4* mutants. Histograms represent the conversion results with S-oxidations displayed in purple, N-oxidation in pink and B-V reaction in green. The right Y-axis corresponds to the B-V reaction and the left Y-axis to S- and N- conversion percentages.

Uncovering how enzyme properties change during evolution provides insight in how life has adapted during evolution but also can be very informative for guiding enzyme engineering approaches. Enzyme engineering studies often target altering stability or catalytic properties of enzymes<sup>31,40,41</sup>. ASR can be used to track and identify the changes that occurred during the evolution of a specific enzyme family. ASR has been found useful as means to increase substrate promiscuity or thermostability<sup>44,45</sup>. Accordingly, we also noticed that all the ancestral FMOs were relatively thermostable and exhibited melting temperatures ranging from 58 to 65 OC (Table S1). ASR also

allowed us to crystallize ancestral FMOs and elucidate the first (ancestral mammalian) FMO structures<sup>21,22</sup>. Exploring the sequences of human FMO3 (hFMO3) and hFMO5 to determine the key residues for the activity switch would have been nearly impossible as there are 253 differences against 41 between tAncFMO1-5 and tAncFMO1-4. The elucidation of mammalian structures also gave supplementary and crucial inputs that were not available before and that were crucial to identify residues that are involved in dictating whether a FMO acts as heteroatom monooxygenase or BVMO.

To conclude, in this work we tackled the question of how an enzyme function emerges by employing a palaeobiochemistry approach. We found that gene duplication events have allowed two FMO lineages to functionally diverge. This has equipped tetrapods with a set of detoxifying enzymes with some functional overlap (FMOs 1-4) and another FMO type that is effectively a BVMO (FMO5). The specialization in activity was only possible through duplications and subsequent activity specialization. These findings illustrate that abolishing or/and improving a specific activity seems less prodigious than introducing new functions.

## Experimental procedures

### *Ancestral sequence reconstruction*

The FMO phylogeny of tetrapods was constructed employing the previously reported dataset as a starting point<sup>21</sup>. The dataset was reinforced as follows: (i) previously experimentally characterized *Homo sapiens* FMO sequences were employed as queries in BLASTp searches in GenBank non-redundant protein sequences (nr) and in Uniprot KB. The searches were restricted by the taxonomy of organisms by classes or orders aiming to mine the whole diversity included in the terrestrial vertebrates (*i.e.* Amphibia, Aves, Crocodylia, Lepidosauria and Mammalia classes and Testudines order) guided by TimeTree knowledge database<sup>34</sup>. (ii) Partial sequences or poor quality ones were excluded. (iii) All collected sequences were gathered and analyzed in a multiple sequence alignment (MSA) using MAFFT v7, removing those that were collected more than one time. Sequences that corresponded to true ORFs were kept in the dataset.

This process allowed building a representative and non-redundant dataset. Dataset included 35, 272, 77, 55, 31 and 66 FMO-like sequences from amphibians, mammals, birds, reptiles, testudines and bony fishes respectively. The MSA contained 536 sequences and 613 sites. Best-fit model parameters were obtained by the Akaike information criterion in ProtTest v3.4. Phylogeny was inferred by the maximum likelihood method in RAxML v 8.2.10 (1000 bootstraps) and subjected to transfer bootstrap expectation (TBE) in BOOSTER. Ancestral sequence reconstruction was performed as marginal reconstruction using the maximum likelihood inference method in PAMLX v.4.9. Sequences were analyzed using an empirical substitution matrix and empirical equilibrium amino acid frequencies (model = 3), 4 gamma categories and LG substitution matrix. The posterior probability distribution of ancestral states at each

site was analyzed at nodes corresponding to tAncFMO1-5, tAncFMO5, tAncFMO1-4 and tAncFMO1-3. The length of the ancestors was treated by parsimony analyzing the presence/absence of gaps in the targeted nodes on the basis of the length of the derived sequences in each clade. Sites were considered ambiguously reconstructed when the alternative states displayed posterior probabilities (PP) > 0.2.

### ***Bioinformatics analysis***

The design of mutant was based on bioinformatic analysis, including the accuracy of reconstruction and degree of conservation of each residues and structural comparisons with previously solved mAncFMO structures. The multiple sequence alignment of the sequences from our dataset was submitted to the Consurf server (<https://consurf.tau.ac.il/2016/>). The structures used as models were mAncFMO5 (PDB 6SEK) and mAncFMO2 (PDB 6SFO) as ketone (BVMO activity) or heteroatom oxidizing FMOs, respectively.

### ***Chemicals***

*Escherichia coli* NEB10 $\beta$  cells and DNA ligase were from New England Biolabs. NADPH was ordered from Oriental Yeast Co, tamoxifen and benzydamine N-oxides were from TCI chemicals and all other chemicals were from Merck.

### ***Cloning, transformation & expression***

Synthetic genes containing Bsal restriction sites at both the 5' and 3' ends were ordered from Twist Bioscience. Lyophilized genes were residue to a final concentration of 10 ng. $\mu$ l<sup>-1</sup> in sterile 10mM Tris-HCl, pH8. All tAncFMO genes were cloned following the Golden Gate cloning method. The recipient vector was a pBAD plasmid modified in such a way that the target protein is expressed fused at its N-terminus to a SUMO protein that carries a 6xHis-tag at its N-terminus. The cloning mixture was the following: 55.4 ng of tAncFMOs inserts, 75 ng of Golden Gate entry vector (a molar ratio of 2:1 insert:vector), 15 U Bsal-HF V2, 15 U T4 DNA ligase, T4 DNA ligase buffer (1 $\times$ ), and nuclease-free water added to a final volume of 20  $\mu$ l. A negative control was prepared without any insert and the PCR cycles were as it follows: the first step with a cycle at 37 °C for 1 hour was followed by 55 °C for 10 min. Then the temperature was set at 65 °C for 20 min and hold at 8 °C. Once cloned, the pBAD-6xHis-SUMO-tAncFMO plasmids were transformed into CaCl<sub>2</sub>-competent *E. coli* NEB10 $\beta$  cells. 5.0  $\mu$ l of plasmid DNA was added to 100  $\mu$ l CaCl<sub>2</sub> competent cells and incubated for 30 min. The cells were then heat shocked at 42°C for 30 s and incubated on ice for 5 min. 500  $\mu$ l LB-SOC was added to allow the cells to recover at 37°C for 1 h. The resuspended cells pellet was then plated on LB-agar containing 100  $\mu$ g.ml<sup>-1</sup> ampicillin and incubated overnight at 37°C. Plasmids were isolated and verified by sequencing and a 20% glycerol stock was stored at -70°C. Mutants M4, M4', M12 and M16 genes were ordered from IDT and cloned as described above.

### **Expression, cell disruption and purification**

A pre-inoculum of 4 ml LB-amp ( $50 \mu\text{g}\cdot\text{ml}^{-1}$ ) was grown overnight at  $37^\circ\text{C}$  and used to inoculate 2 liter baffled flasks containing 400 ml of Terrific-Broth medium, supplemented with  $50 \text{ mg}\cdot\text{l}^{-1}$  ampicillin and incubated at  $37^\circ\text{C}$ . Expression was induced by adding 0.02% L-arabinose from a sterile 20% stock (w/v) when the OD600 was between 0.2 and 0.5. Cultures were grown at  $24^\circ\text{C}$  with shaking for a total of 30 hours before harvesting. Cells were harvested by centrifugation (2755 g, 25 min). Pellets were resuspended into Buffer A (250 mM NaCl, 50 mM potassium phosphate, pH 7.5) with a 5:1 ratio [volume (ml) : mass (g)] and supplemented with 0.10 mM phenylmethylsulfonyl fluoride and 1.0 mM  $\beta$ -mercaptoethanol. Cell disruption was done by sonication (70% amplitude, 5 s ON, 5 s OFF, for a total of 20 min). After centrifuging at 19500g for 20 min, the supernatant was removed and the pellet was resuspended into Buffer A2 (250 mM NaCl, 50 mM potassium phosphate, 0.5% Triton X100-reduced, pH 7.5) with the same ratio as before (5:1). The resuspended pellet was mixed overnight at  $4^\circ\text{C}$  in order to solubilize the membrane protein and centrifuged at 19500g to collect the supernatant. tAncFMOs were purified with metal-ion affinity chromatography that bound the histidine tag attached to the N-terminal part of the fused SUMO protein. The cell free extract was applied to the column and washed with increasing concentrations of imidazole. Buffer B contained (250 mM NaCl, 50 mM potassium phosphate, 300 mM imidazole, 0.5% Triton X100-reduced, pH 7.5). Following the washing steps with 0, 25, and 50 mM imidazole, the protein was finally eluted with 300 mM imidazole. The elution buffer was exchanged with a storage buffer using a desalting column (250 mM NaCl, 50 mM potassium phosphate, 0.05% Triton X100-reduced, pH 7.5).

Purified 6xHis-SUMO tagged enzyme was frozen with liquid nitrogen and kept at  $-70^\circ\text{C}$ . Experiments were performed using these aliquots. Concentrations of tAncFMOs were determined from frozen samples that were incubated at  $95^\circ\text{C}$ , centrifuged and the supernatant analyzed on Jasco V-660 spectrophotometer. Using  $\epsilon_{\text{FAD}} = 11.3 \text{ mM}^{-1}\cdot\text{cm}^{-1}$  at 450 nm the amount of holoenzyme was quantified.

### **Melting temperature assays**

The melting temperatures ( $T_m$ ) of tAncFMOs were assessed and determined using a CFX96 Real-Time PCR System from Bio-Rad. Each sample contained  $10 \mu\text{M}$  tAncFMO, in presence or absence of  $1 \text{ mM}$   $\text{NAD(P)}^+$ , and topped-up to a final volume of  $10 \mu\text{l}$  with the storage buffer.

### **Conversions**

Substrate conversions were done in duplicates, using 1.0-5.0 mM substrate ( $<1\%$  MeOH),  $0.10 \text{ mM}$  NADPH,  $2.0 \mu\text{M}$  enzyme,  $5.0 \mu\text{M}$  phosphite dehydrogenase and  $20 \text{ mM}$  sodium phosphite. The last two components were used as a regeneration system for NADPH and the control did not contain any enzyme. All compounds were prepared in storage buffer ( $50 \text{ mM}$  KPi,  $250 \text{ mM}$  NaCl,  $0.05 \%$  triton X100-reduced, pH 7.5) and the final reaction volume was adjusted to  $1.0 \text{ ml}$  and put into  $4 \text{ ml}$  vials before being



incubated at 30°C, with shaking, for 16 hours. Conversion of phenyl acetone, heptan-2-one, cyclohexanone, benzyl phenyl sulfide and methyl-*p*-tolyl sulfide could be analyzed by GC-MS while benzydamine and tamoxifen conversions were monitored by HPLC. Tamoxifen, benzydamine and benzyl phenyl sulfide conversions were done using 1.0 mM substrate while the remaining substrates were converted from 5.0 mM.

For GC-MS analysis, compounds were extracted by adding one volume of ethyl acetate with 0.02 % (v/v) mesitylene as external standard. Samples were vortexed for 20 seconds, centrifuged and the organic phase was eluted through anhydrous magnesium sulfate. GC-MS analyses were performed using an HP-1 Agilent column (30 m x 0.25 mm x 0.25 µm). The method was the following: injector and detector temperature at 250 °C, a split ratio of 5.0, and an injection volume of 1 µl. The column temperature was held at 50 °C for 4 min, increased by 10 °C/min to 250 °C and held for 5 min. The retention times of the substrates and products are displayed in the supplementary information (Figure S2) with the corresponding mass spectra. Conversions were calculated based on the substrate depletion and normalized with the external standard.

HPLC analyses were performed after diluting 300 µl of the sample into 1200 µl acetonitrile, vortexing it for 20 s and centrifuging. Analysis was performed using reverse phase HPLC. Samples were injected with a volume of 10 µl onto a JASCO AS2051 Plus HPLC system, equipped with a Grace Alltima HP C18 column (5 µm, 4.6×250 mm). The solvents used were water with 0.1% v/v formic acid (A) and acetonitrile (B) and the flow rate was 0.8 ml.min<sup>-1</sup>. For benzydamine the method corresponded to 8 min on an isocratic flow of 35% B and 65% A. Benzydamine and benzydamine N-oxide were detected at 308 nm with a retention time of 5.3 min and 5.7 min, respectively (Figure S3). For tamoxifen the method was the following: 30 min on a gradient of 40-95% B, 3 min with 95% B followed by a 5 min decreased gradient of 95-40% B and finally a re-equilibration for 2 min. Tamoxifen and tamoxifen N-oxide were detected at 276 nm with a retention time of 10.5 min and 11.7 min, respectively (Figure S3). Both products were confirmed using the corresponding standards and the conversion calculated based on substrate depletion.

### **Data availability**

All data are contained within the manuscript.

### **Author's contribution**

All listed authors performed experiments and/or analyzed data. M.L.M., G.B. and C.R.N. conducted thorough evolutionary analyses and performed ancestral-sequence reconstruction under the direct supervision of M.L.M. G.B. performed Golden Gate cloning to insert tAncFMOs genes into their respective vectors, purified and analyzed the substrate profile of the enzymes. G.B. made the mutants and characterized them as well.

G.B., C.R.N. and M.L.M. prepared the figures. G.B., C.R.N., and M.L.M. wrote the manuscript and A.M. and M.W.F. edited it. All authors provided critical feedback and helped shape the research, analysis and manuscript. M.L.M. conceived the original idea.

### **Funding and additional information**

The research for this work has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 722390; the Italian Ministry of Education, University and Research (MIUR) under the "Dipartimenti di Eccellenza (2018–2022)" program; The European Union's Horizon 2020 research and innovation programme under grant agreement No 847675. The authors thank Kamiel H. Heeres, that helped with the preparation of the first round of mutants, under the direct supervision of G.B.

### **Conflict of interest**

The authors declare that they have no conflicts of interest with the contents of this article.

## References

- 1 Walsh, C. T. & Wenczewicz, T. A. Flavoenzymes: Versatile catalysts in biosynthetic pathways. *Nat. Prod. Rep.* 30, 175–200 (2013).
- 2 van Berkel, W. J. H., Kamerbeek, N. M. & Fraaije, M. W. Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *J. Biotechnol.* 124, 670–689 (2006).
- 3 Walsh, C. T. & Chen, Y. J. Enzymic Baeyer-Villiger Oxidations by Flavin-Dependent Monooxygenases. *Angew. Chemie - Int. Ed.* 333–343 (1988).
- 4 Cashman, J. R. Some distinctions between flavin-containing and cytochrome P450 monooxygenases. *Biochem. Biophys. Res. Commun.* 338, 599–604 (2005).
- 5 Cashman, J. R., Akerman, B. R., Forrest, S. M. & Treacy, E. P. Population-specific polymorphisms of the human FMO3 gene: Significance for detoxication. *Drug Metab. Dispos.* 28, 169–173 (2000).
- 6 Rossner, R., Kaeberlein, M. & Leiser, S. F. Flavin-containing monooxygenases in aging and disease: Emerging roles for ancient enzymes. *J. Biol. Chem.* 292, 11138–11146 (2017).
- 7 Canyelles, M. et al. Trimethylamine N-oxide: A link among diet, gut microbiota, gene regulation of liver and intestine cholesterol homeostasis and HDL function. *Int. J. Mol. Sci.* 19, (2018).
- 8 Phillips, I. R. & Shephard, E. A. Flavin-containing monooxygenases: mutations, disease and drug response. *Trends Pharmacol. Sci.* 29, 294–301 (2008).
- 9 Phillips, I. R. & Shephard, E. A. Flavin-containing monooxygenases: new structures from old proteins. *Nat. Struct. Mol. Biol.* 27, 3–4 (2020).
- 10 Mayatepek, E., Flock, B. & Zschocke, J. Benzydamine metabolism in vivo is impaired in patients with deficiency of flavin-containing monooxygenase 3. *Pharmacogenetics* 14, 775– 777 (2004).
- 11 Parte, P. & Kupfer, D. Oxidation of tamoxifen by human flavin-containing monooxygenase (FMO) 1 and FMO3 to tamoxifen-N-oxide and its novel reduction back to tamoxifen by human cytochromes P450 and hemoglobin. *Drug Metab. Dispos.* 33, 1446–1452 (2005).
- 12 Zhang, J. & Cashman, J. R. Quantitative analysis of FMO gene mRNA levels in human tissues. *Drug Metab. Dispos.* 34, 19–26 (2006).
- 13 Siddens, L. K., Krueger, S. K., Henderson, M. C. & Williams, D. E. Mammalian flavin-containing monooxygenase (FMO) as a source of hydrogen peroxide. *Biochem. Pharmacol.* 89, 141–147 (2014).
- 14 El-Serafi, I. et al. Flavin-containing monooxygenase 3 (FMO3) role in busulphan metabolic pathway. *PLoS One* 12, (2017).
- 15 Veeravalli, S. et al. Flavin-containing monooxygenase 1 catalyzes the production of taurine from hypotaurine. *Drug Metab. Dispos.* 48, 378–385 (2020).
- 16 Fiorentini, F. et al. Biocatalytic characterization of human FMO5: Eneathing Baeyer-Villiger reactions in humans. *ACS Chem. Biol.* 11, 1039–1048 (2016).
- 17 Itagaki, K., Carver, G. T. & Philpot, R. M. Expression and characterization of a modified flavin-containing monooxygenase 4 from humans. *J. Biol. Chem.* 271, 20102–20107 (1996).
- 18 Yanni, S. B. et al. Role of flavin-containing monooxygenase in oxidative metabolism of voriconazole by human liver microsomes. *Drug Metab. Dispos.* 36, 1119–1125 (2008).
- 19 Henderson, M. C., Siddens, L. K., Morré, J. T., Krueger, S. K. & Williams, D. E. Metabolism of the anti-tuberculosis drug ethionamide by mouse and human FMO1, FMO2 and FMO3 and mouse and human lung microsomes. *Toxicol. Appl. Pharmacol.* 233, 420–427 (2008).

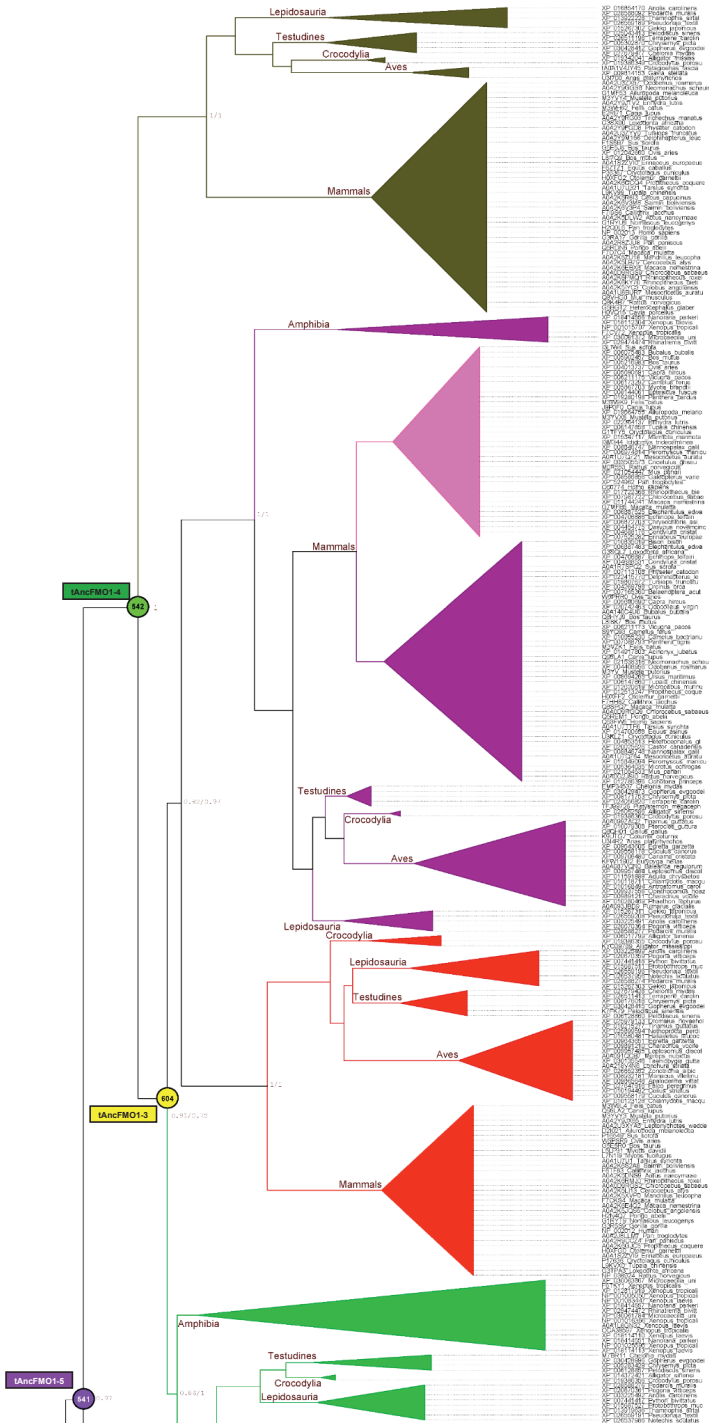
- 20 Indra, R. et al. Identification of human enzymes oxidizing the anti-thyroid-cancer drug vandetanib and explanation of the high efficiency of cytochrome P450 3A4 in its oxidation. *Int. J. Mol. Sci.* 20, (2019).
- 21 Bailleul, G., Nicoll, C. R., Mascotti, M. L., Mattevi, A. & Fraaije, M. W. Ancestral reconstruction of mammalian FMO1 enables structural determination, revealing unique features that explain its catalytic properties. *J. Biol. Chem.* 296, 0–2 (2021).
- 22 Nicoll, C. R. et al. Ancestral-sequence reconstruction unveils the structural basis of function in mammalian FMOs. *Nat. Struct. Mol. Biol.* 27, 14–24 (2020).
- 23 Hines, R. N., Hopp, K. A., Franco, J., Saeian, K. & Begun, F. P. Alternative processing of the human FMO6 gene renders transcripts incapable of encoding a functional flavin-containing monooxygenase. *Mol. Pharmacol.* 62, 320–325 (2002).
- 24 Fisher, B. et al. Tamoxifen for the prevention of breast cancer: Current status of the National Surgical Adjuvant Breast and Bowel Project P-1 study. *J. Natl. Cancer Inst.* 97, 1652–1662 (2005).
- 25 Hu, S. X. Hepatic Flavin-containing Monooxygenase and Aldehyde Oxidase Activities in Male Domestic Pigs at Different Ages. *Drug Metab. Lett.* 12, 125–131 (2018).
- 26 Chitra, M., Ernest, H. & David, K. Metabolism of the cancer agent tamoxifen : Flavin- Containing Monooxygenase-Mediated N-Oxidation. *Drug Metab. Dispos.* 21, (1993).
- 27 Uno, Y., Shimizu, M. & Yamazaki, H. Molecular and functional characterization of flavin-containing monooxygenases in cynomolgus macaque. *Biochem. Pharmacol.* 85, 1837–1847 (2013).
- 28 Jensen, C. N., Ali, S. T., Allen, M. J. & Grogan, G. Mutations of an NAD(P)H-dependent flavoprotein monooxygenase that influence cofactor promiscuity and enantioselectivity. *FEBS Open Bio* 3, 473–478 (2013).
- 29 Gran-Scheuch, A., Trajkovic, M., Parra, L. & Fraaije, M. W. Mining the Genome of *Streptomyces leeuwenhoekii*: Two New Type I Baeyer–Villiger Monooxygenases From Atacama Desert. *Front. Microbiol.* 9, 1–10 (2018).
- 30 Romero, E., Castellanos, J. R. G., Mattevi, A. & Fraaije, M. W. Characterization and Crystal Structure of a Robust Cyclohexanone Monooxygenase. *Angew. Chemie* 128, 16084–16087 (2016).
- 31 Fürst, M. J. L. J., Gran-Scheuch, A., Aalbers, F. S. & Fraaije, M. W. Baeyer–Villiger Monooxygenases: Tunable Oxidative Biocatalysts. *ACS Catal.* 9, 11207–11241 (2019).
- 32 Fiorentini, F. et al. Biocatalytic Characterization of Human FMO5: Unearthing Baeyer–Villiger Reactions in Humans. *ACS Chem. Biol.* 11, 1039–1048 (2016).
- 33 Voordeckers, K. et al. Reconstruction of Ancestral Metabolic Enzymes Reveals Molecular Mechanisms Underlying Evolutionary Innovation through Gene Duplication. *PLoS Biol.* 10, (2012).
- 34 Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* 34, 1812–1819 (2017).
- 35 Clack, J. A. The emergence of early tetrapods. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 232, 167–189 (2006).
- 36 Lee, C. et al. Coelacanth-specific adaptive genes give insights into primitive evolution for water-to-land transition of tetrapods. *Mar. Genomics* 38, 89–95 (2018).
- 37 Ip, Y. K., Chew, S. F. & Randall, D. J. Five tropical air-breathing fishes, six different strategies to defend against ammonia toxicity on land. in *Physiological and Biochemical Zoology* vol. 77 768–782 (The University of Chicago Press, 2004).

- 38 Ching, B. et al. L-gulonolactone oxidase expression and vitamin C synthesis in the brain and kidney of the African lungfish, *Protopterus annectens*. *FASEB J.* 28, 3506–3517 (2014).
- 39 Phillips, I. R. & Shephard, E. A. Flavin-containing monooxygenase 3 (FMO3): genetic variants and their consequences for drug metabolism and disease. *Xenobiotica* 50, 19–33 (2020).
- 40 Ali, M., Ishqi, H. M. & Husain, Q. Enzyme engineering: Reshaping the biocatalytic functions. *Biotechnology and Bioengineering* vol. 117 1877–1894 (2020).
- 41 Chen, K. & Arnold, F. H. Engineering new catalytic activities in enzymes. *Nature Catalysis* vol. 3 203–213 (2020).
- 42 Lončar, N., van Beek, H. L. & Fraaije, M. W. Structure-Based Redesign of a Self-Sufficient Flavin-Containing Monooxygenase towards Indigo Production. *Int. J. Mol. Sci.* 20, 6148 (2019).
- 43 Arabnejad, H. et al. Computational Design of Enantiocomplementary Epoxide Hydrolases for Asymmetric Synthesis of Aliphatic and Aromatic Diols. *ChemBioChem* 1893–1904 (2020) doi:10.1002/cbic.201900726.
- 44 Gaucher, E. A., Govindarajan, S. & Ganesh, O. K. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* 451, 704–707 (2008).
- 45 Risso, V. A., Gavira, J. A., Mejia-Carmona, D. F., Gaucher, E. A. & Sanchez-Ruiz, J. M. Hyperstability and substrate promiscuity in laboratory resurrections of precambrian  $\beta$ -lactamases. *J. Am. Chem. Soc.* 135, 2899–2902 (2013). 167

## Abbreviations and nomenclature

tAncFMO	Tetrapod ancestral flavin-containing monooxygenase
mAncFMO	Mammalian ancestral flavin-containing monooxygenase
ASR	Ancestral sequence reconstruction
FAD	Flavin adenine dinucleotide
FMO	Flavin-containing monooxygenase
hFMO	Human flavin-containing monooxygenase
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
PP	Posterior probability
TBE	Transfer bootstrap expectation
T <sub>m</sub>	Melting temperature

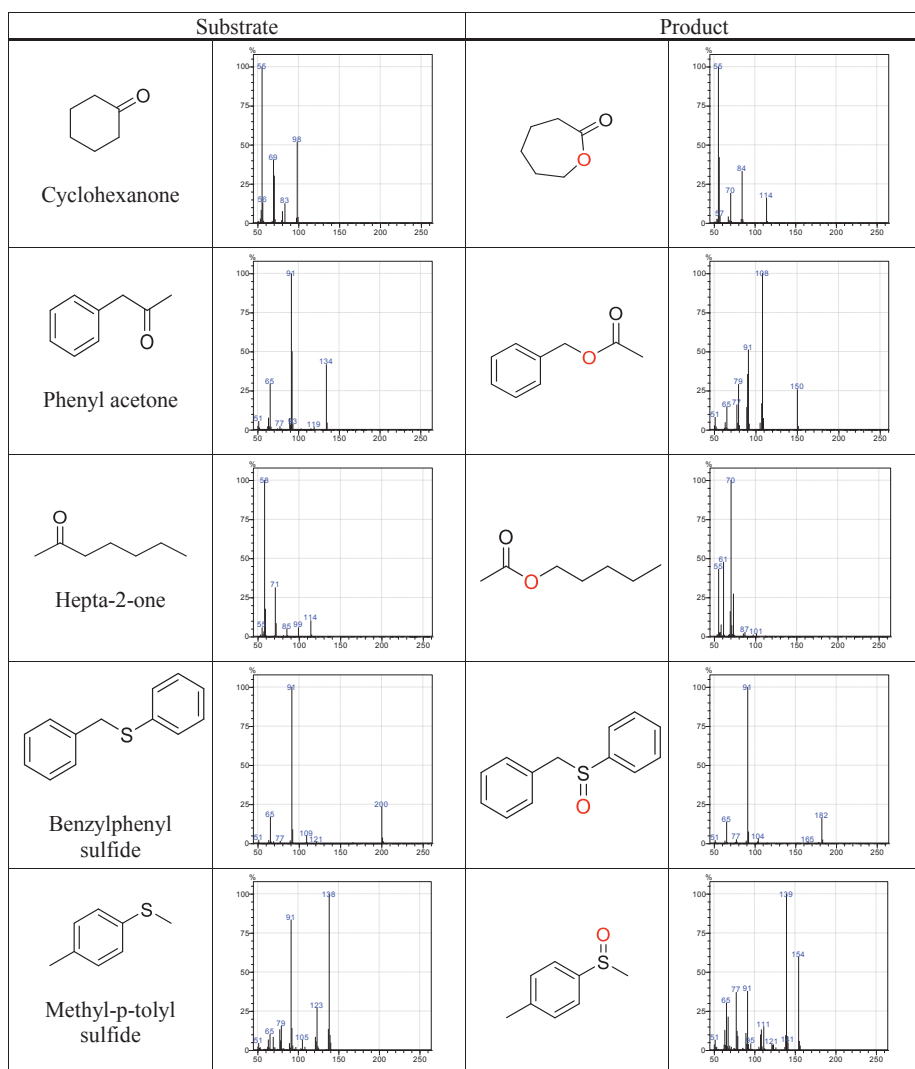
# Supplementary information







**Supplementary Figure 1: Vertebrates phylogeny of FMOs** (previous pages). Tree was constructed in RAxML v8.2.10, 1000 bootstraps were run and best-fit model parameters were obtained with ProtTest v3.4. Bootstrap transfer was applied and both TBE values and PP are shown at the nodes. The employed MSA was trimmed in single sequence extensions and contained 536 taxa and 613 sites. Clades are collapsed and colored according to FMO clades: FMO4 (dark green), FMO3 (purple), FMO6 (pink), FMO1 (red), FMO2 (light green), FMO5 (orange) and amphibian FMOs (both light and dark blue). Actinopterygii, coelacanthimorpha, chondrichthyes (all jawed vertebrates) and the cephalochordate sequence in the root are shown in black. The four ancestral nodes that were experimentally characterized are plotted at the nodes.



**Figure S2.** Structures of substrates and products analyzed by GC-MS, with their respective MS fragmentation patterns.

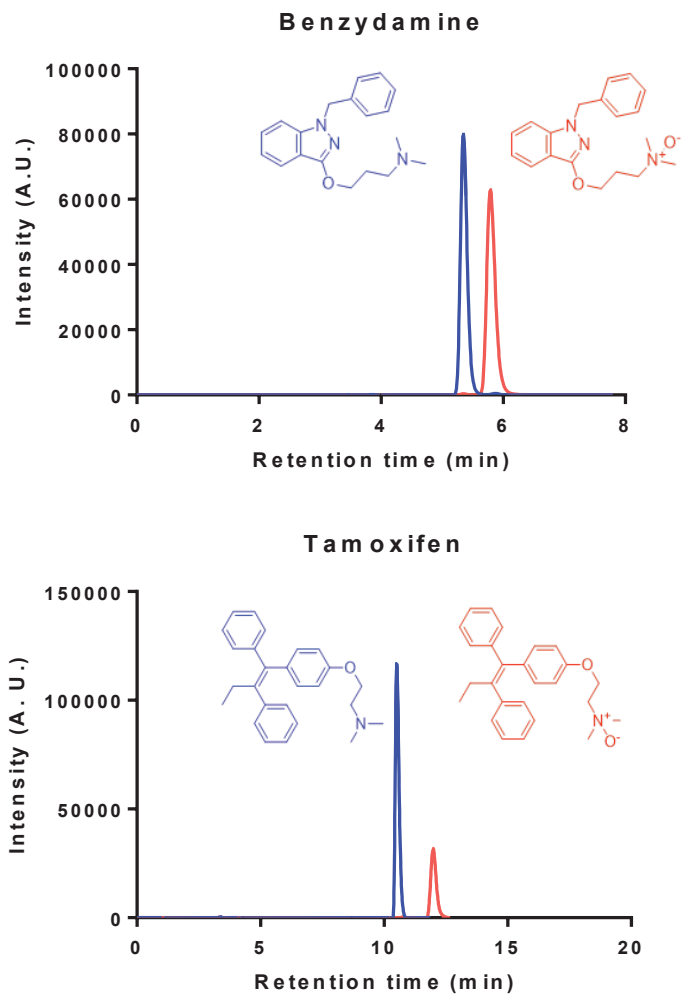
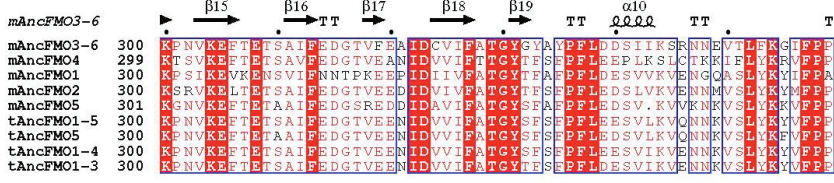
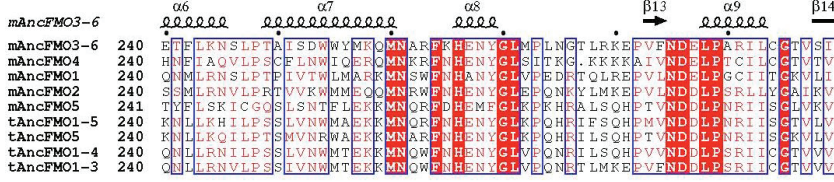
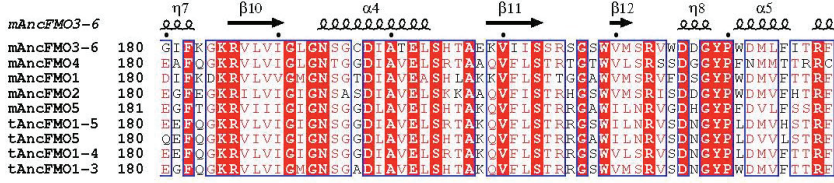
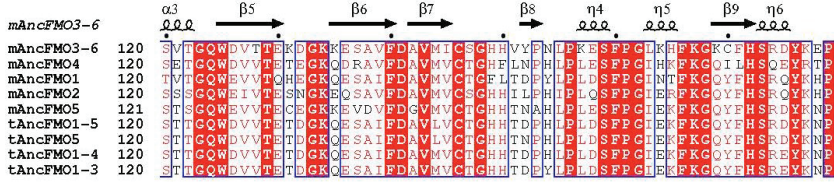
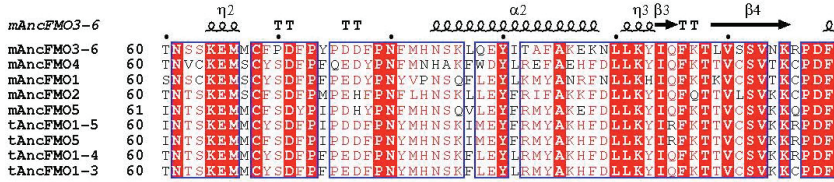
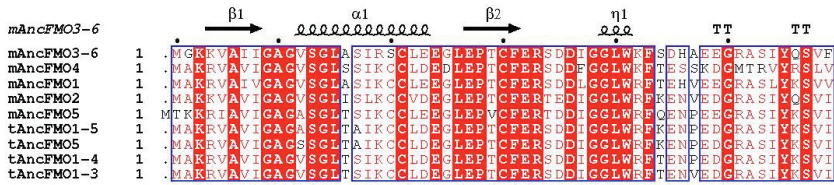


Figure S3. HPLC Chromatograms of benzydamine and tamoxifen biotransformations with the respective substrates (blue) and products (red).



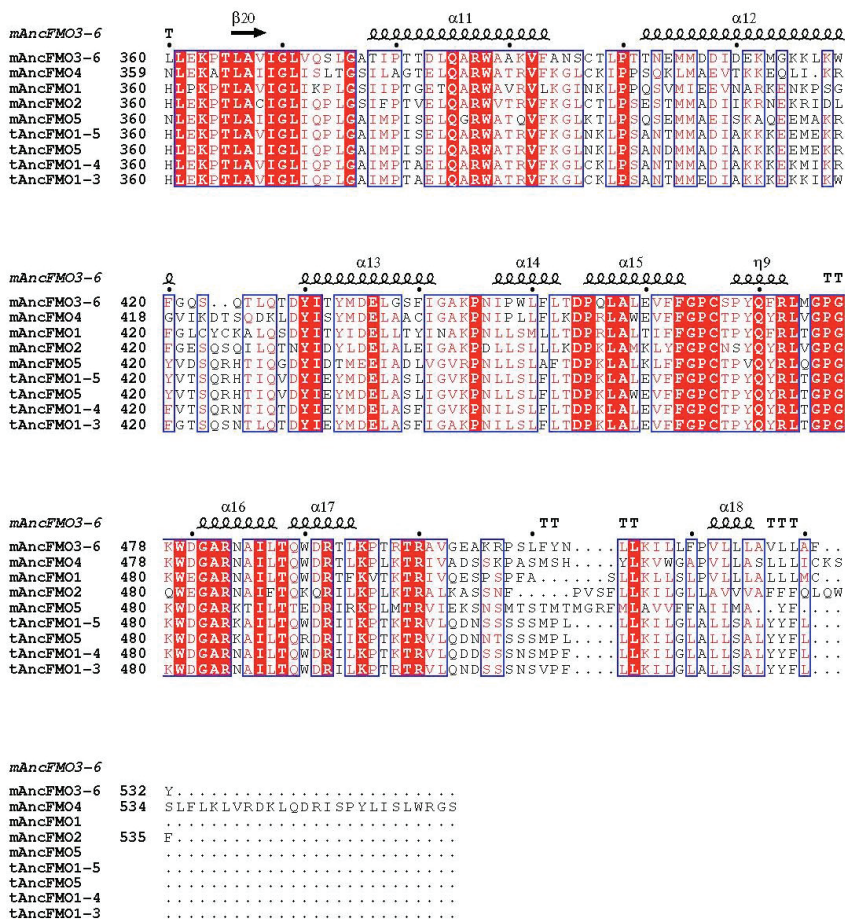


Figure S4. Multiple sequence alignment (MSA) of ancestral FMOs. The MSA was obtained using the ENDscript server1 and the structural features of mAncFMO3-6 are given as reference.

*Table S1. Melting temperatures of the tAncFMOs. Melting temperatures, with and without NAD(P)<sup>+</sup>, were measured in duplicate by using the ThermoFAD technique<sup>2</sup> for all tAncFMOs (see materials and methods).*

Enzyme	Melting temperature (T <sub>m</sub> ) / °C		
	no NAD(P) <sup>+</sup>	100 equivalents NADP <sup>+</sup>	100 equivalents NAD <sup>+</sup>
<b>tAncFMO1-5</b>	60	60	60
<b>tAncFMO5</b>	58	57	58
<b>tAncFMO1-4</b>	60	60	61
<b>tAncFMO1-3</b>	65	65	65

## References

1. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–W324 (2014).
2. Forneris, F., Orru, R., Bonivento, D., Chiarelli, L. R. & Mattevi, A. ThermoFAD, a ThermoFluor®-adapted flavin ad hoc detection system for protein folding and ligand binding. *FEBS J.* 276, 2833–2840 (2009).







CHAPTER VI

**Summary and conclusions**

Gautier Bailleul  
Marco W. Fraaije





## Summary and Conclusions

The biotechnology field grew exponentially during the last decades. This is for a large part due to the fact that it often presents a sustainable and green alternative for both chemical and pharmaceutical industries. Enzymes exhibit high selectivities and catalytic activities and can open new paths towards synthesis of valuable products without the need of harsh chemicals. The malleability of enzymes is a supplementary asset as they can be modified through a plethora of enzyme engineering methods, such as directed evolution, to tune their stability, selectivity, catalytic efficiency or other features. There are seven major classes of enzymes: oxidoreductases, transferases, hydrolases, lyases, isomerases, ligases, and translocases. The oxidoreductases, also referred to as redox enzymes, perform oxidation and reduction reactions and often outcompete chemical approaches. Flavoproteins monooxygenases (FPMOs), introduced in **chapter 1**, are redox enzymes that use a redox-active flavin cofactor (FAD or FMN), responsible of their typical yellow color. FPMO can be classified into six subclasses and the enzymes described in this thesis all belong to the so-called class B FPMOs. Specifically, type I flavin-containing monooxygenases (FMOs) and type I Baeyer-Villiger monooxygenases (BVMOs) have been studied. These enzymes employ a reduced nicotinamide cofactor, nicotinamide adenine dinucleotide (NAD) or nicotinamide adenine dinucleotide phosphate (NADP), as electron donor. Although they are sequence-related, FMOs and BVMOs differ substantially concerning the reactions they catalyze: while FMOs catalyze oxygenations of heteroatoms, BVMO convert ketones via a Baeyer-Villiger oxidation, respectively. Except for an introduction into these FPMOs, **chapter 1** also provides an overview of the industrial prospects, role in metabolism, and pharmaceutical relevance of bacterial, plant and mammalian FMOs.

**Chapter 2** reports on a study of a newly discovered bacterial FMO from *Nitrocola lacisaponensis* (NiFMO), an alkaliphilic extremophile bacterium. The enzyme was identified through a genome mining approach in order to find a more thermostable alternative of the already described bacterial FMO from *Methylophaga aminisulfidivorans* (mFMO)<sup>1,2</sup>. NiFMO was highly expressed (about 100 mg·L<sup>-1</sup>) as a soluble and active FAD-containing enzyme. The 3-dimensional structure of the enzyme was elucidated with a resolution of 1.8 Å. The relatively high thermostability of NiFMO (melting temperature of 51 °C), its high organic solvent tolerance and broad substrate portfolio establish the enzyme as a candidate for applied biocatalysis. These features render NiFMO an interesting candidate for biotechnological applications. A potential application is in the dye industry where it could be used for the production of the dyes indigo and Tyrian purple by oxygenating indole or bromoindole, respectively. Such enzyme-based process would be attractive when compared with current chemical production processes as it would eliminate the need for harsh chemicals and may start from renewable starting material<sup>3,4</sup>. Humans possess five FMOs paralogs, named FMO1, FMO2, FMO3, FMO4 and FMO5, which are involved in the detoxification of numerous deleterious xenobiotics

including toxins and pesticides. The ability of human FMOs (hFMOs) to process (degrade or activate) drugs and their involvement in diseases reveal their high pharmacological significance<sup>5,6</sup>. However, despite their relevance in human health, hFMOs -and any other mammalian FMOs- have been poorly characterized when compared with other human xenobiotic degrading enzymes such as cytochrome P450 monooxygenases. For example, no detailed structural information is available for hFMOs. This is partly explained by the fact that hFMOs and other animal FMOs, all membrane-embedded, have shown to be difficult to purify. **Chapter 3** describes how we used ancestral sequence reconstruction (ASR) to scrutinize three mammalian FMOs ancestors to further our understanding concerning their biochemical properties. The protein sequences of AncFMO2, AncFMO3-6 and AncFMO5 were reconstructed with high posterior probabilities (ranging from 0.98 to 0.99) and the phylogeny of mammalian FMOs was further refined to include the early divergence of FMO5s. The three AncFMOs could be crystallized, and were structurally resolved at 2.7-3.2 Å resolution, revealing their specific membrane-binding features and intricate substrate-profiling tunnel networks. Catalytic studies included fast kinetics using the stopped-flow technique in order to investigate the flavin reduction rate and the formation of a peroxyflavin intermediate formation and its decay during the catalytic cycle. Steady-state kinetics were also conducted by following the NADPH depletion in absence and presence of known mammalian FMO substrates. Remarkably, all AncFMOs exhibited similar catalytic behavior as their extant relatives and can therefore be employed as reliable models for hFMOs with which they share 82-92% of sequence identity.

In **chapter 4**, we supplemented our mammalian ancestral FMO collection by reconstructing the sequences of AncFMO1 and AncFMO4. While the latter possessed a C-terminal extension unique to mammalian FMO4s that prevented its expression, AncFMO1 could be expressed, characterized and crystallized. The crystal structure of AncFMO1 was elucidated at 3.0 Å resolution and, similar to the other AncFMO structures, revealed a structure of a dimeric FAD-containing enzyme. Its structure contains typical FAD and NAD(P)H binding domains and a C-terminal helix for membrane binding. Some unique features were discovered, including a porous and exposed active site, along with NADPH adopting a new conformation that pushes the 2'-phosphate inside the NADP<sup>+</sup> binding domain instead of being stretched out in the solvent. The enzyme displays common FMO activity and oxidizes the drug-related compounds benzydamine, tamoxifen and thioanisole. As previously demonstrated in hFMO1, AncFMO1 is able to accept both NADH and NADPH cofactors as electron donors, a feature only described within the FMO1 clade<sup>7</sup>. Alike to the previously reconstructed mammalian FMOs, AncFMO1 shares 89% sequence identity with hFMO1 and identical catalytic behavior, indicating that the enzyme can be used as a bona fide structural model of FMO1.

**Chapter 3** and **4** demonstrate the power of ancestral sequence reconstruction as a strategy for the crystallization of proteins. Recent research has revealed that within the five mammalian FMOs, FMO5 is an outlier<sup>8</sup>. It was shown that FMO5s behave as

BVMOs, while the other four FMO paralogs only catalyze the oxygenation of heteroatom containing substrates. In **chapter 3**, we speculated that a glutamate-histidine switch is able to induce the Baeyer-Villiger oxidation activity in AncFMO5. The substitution resulted in a total loss of activity in AncFMO5 but did not introduce any Baeyer-Villiger oxidation ability within AncFMO2 and AncFMO3-6. Questions about the catalytic diversity within FMOs that consequently arise are:

- (i) what were the ancestral functions of the original tetrapod FMO?
- (ii) what are the determinants behind the functional switch between Baeyer-Villiger oxidation activities and heteroatom oxidations?

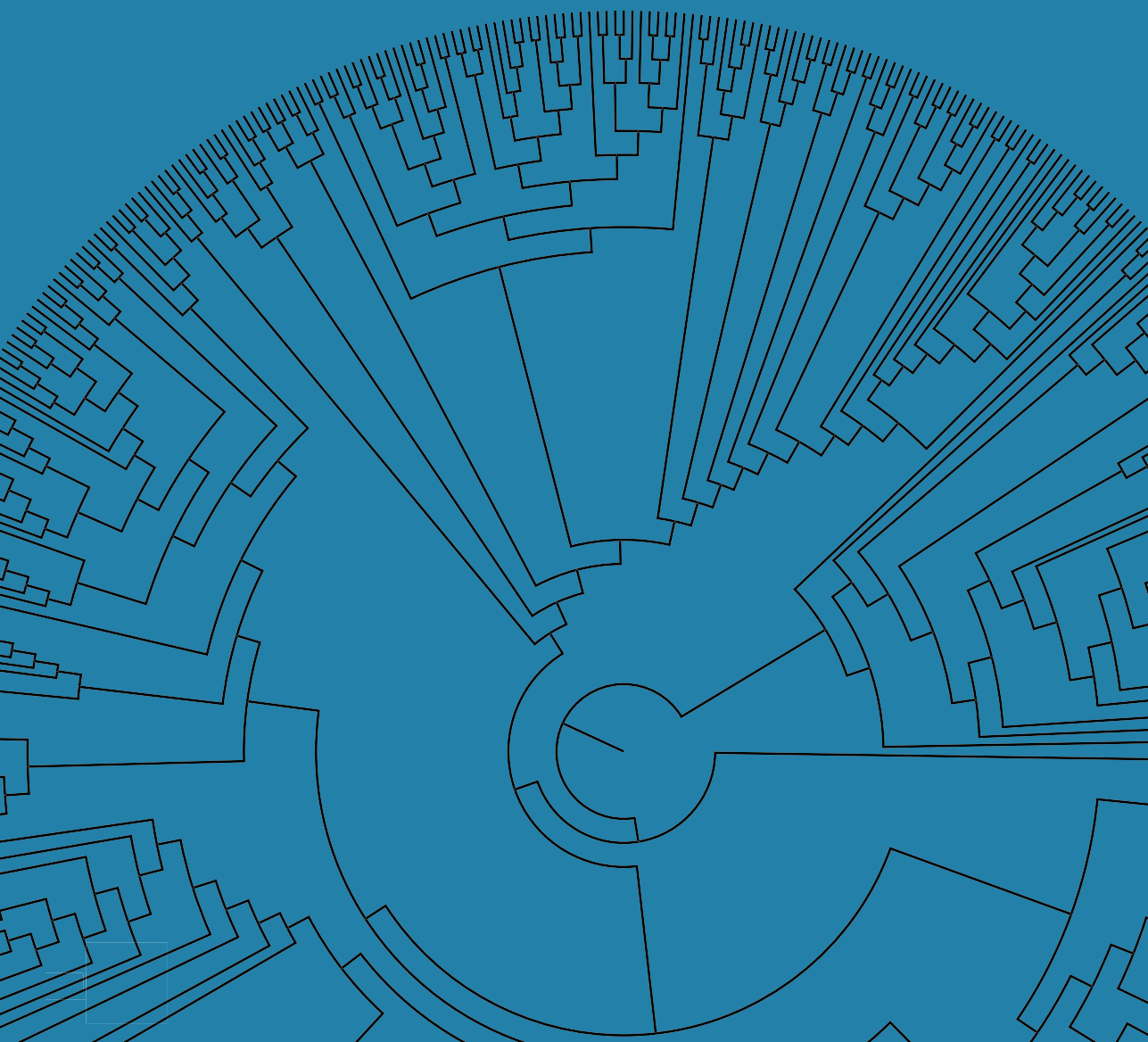
In **Chapter 5**, we reconstructed the tetrapod ancestors of FMOs using ASR. The phylogeny used for the mammalian ancestors was supplemented with additional sequences of Testudines (turtles), Crocodylia, Aves (birds), and Lepidosauria (subclass of reptiles) clades and the root was further populated with Actinopterygii (bony fishes). The generated phylogenetic tree confirms a rapid expansion of FMOs (FMO explosion) that happened within the tetrapod clade. The tetrapod ancestor possessed a single FMO copy that was later followed by multiple duplication events leading to the five different FMO lineages. We recovered four ancestral sequences, named based on the emergence of each of the paralog clade: tAncFMO1-5, tAncFMO5, tAncFMO1-4 and tAncFMO1-3. The biocatalytic profile of tAncFMOs was assessed with a substrate portfolio composed of known BVMO and FMO substrates. The eldest ancestor, tAncFMO1-5, was able to perform both reaction types. This proficiency was lost during the first duplication event, leading to tAncFMO1-4 being only able to catalyze heteroatoms oxygenation. On the other hand, tAncFMO5 kept the ability to oxidize ketones. The residues potentially responsible for the switch in type of oxygenation activity were investigated through a multi-approach method combining the reconstruction score, the conservation degree within tetrapods and the localization of the amino acids in the enzyme structure. The substitution of four residues in tAncFMO1-4 resulted in the recovery of the enzyme to perform Baeyer-Villiger oxidation of phenyl acetone concomitant with an overall loss of activity toward heteroatom-containing compounds. Two residues, Thr60 and Ser222, were highly conserved within heteroatom-oxidizing FMOs and located close to the FAD cofactor, in the active site, or present in the 80-residue insertion that forms the substrate-entry tunnels, respectively. Two histines, His275 and His426, are suggested to be crucial for the oxygenation of ketones via their cooperating hydrogen bond network. By a biochemical reconstruction approach, combining molecular phylogeny, bioinformatics, structural analysis, mutagenesis and biochemical characterization, this chapter unraveled how enzyme functions emerge. Specifically, it sheds light on the evolution of different oxygenation activities within the family of animal FMOs.

The work presented in this thesis deepens our general knowledge about FMOs. First, with the characterization of a newly discovered thermostable bacterial FMO a new biocatalytic tool has been added to the toolbox of available FPMOs (Chapter 2). Then

with the unprecedented elucidation of mammalian FMO structures (Chapter 3 and 4), new insights on the functioning of these detoxifying enzymes has been obtained. The elucidated structures, together with the generated biochemical data, provide pivotal input for future pharmacology studies. Ultimately, by reconstructing a set of carefully selected ancestral FMOs, glimpses of steps in evolution of tetrapod FMOs were obtained for the first time, revealing determinants behind the FMO-BVMO switch (Chapter 5).

## References

- 1 Choi, H. S. et al. A novel flavin-containing monooxygenase from *Methylophaga* sp. strain SK1 and its indigo synthesis in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 306, 930–936 (2003).
- 2 Alfieri, A., Malito, E., Orru, R., Fraaije, M. W. & Mattevi, A. Revealing the moonlighting role of NADP in the structure of a flavin-containing monooxygenase. *Proc. Natl. Acad. Sci. U. S. A.* 105, 6572–6577 (2008).
- 3 Lončar, N., van Beek, H. L. & Fraaije, M. W. Structure-Based Redesign of a Self-Sufficient Flavin-Containing Monooxygenase towards Indigo Production. *Int. J. Mol. Sci.* 20, 6148 (2019).
- 4 Lee, J. et al. Production of Tyrian purple indigoid dye from tryptophan in *Escherichia coli*. *Nat. Chem. Biol.* 17, (2020).
- 5 Phillips, I. R. & Shephard, E. A. Flavin-containing monooxygenases: mutations, disease and drug response. *Trends Pharmacol. Sci.* 29, 294–301 (2008).
- 6 Parte, P. & Kupfer, D. Oxidation of tamoxifen by human flavin-containing monooxygenase (FMO) 1 and FMO3 to tamoxifen-N-oxide and its novel reduction back to tamoxifen by human cytochromes P450 and hemoglobin. *Drug Metab. Dispos.* 33, 1446–1452 (2005).
- 7 Veeravalli, S. et al. Flavin-containing monooxygenase 1 catalyzes the production of taurine from hypotaurine. *Drug Metab. Dispos.* 48, 378–385 (2020).
- 8 Fiorentini, F. et al. Biocatalytic Characterization of Human FMO5: Unearthing Baeyer-Villiger Reactions in Humans. *ACS Chem. Biol.* 11, 1039–1048 (2016).



APPENDICES

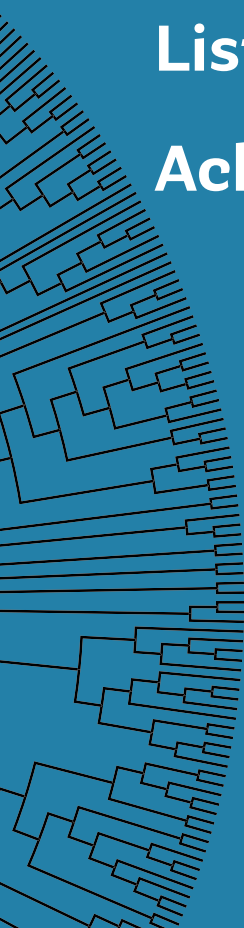
**Nederlandse samenvatting**

**Résumé en Français**

**Curriculum Vitae**

**List of Publications**

**Acknowledgements**







## Nederlandse samenvatting

Vertaald door Jeroen Drenth en Eric de Vries

Gedurende de laatste decennia is het biotechnologische werkveld exponentieel gegroeid. Vooral omdat onder andere het gebruik van enzymen groene en duurzame alternatieven biedt voor de chemische en farmaceutische industrie. Enzymen zijn biologische katalysatoren met hoge activiteit en selectiviteit die nieuwe syntheseroutes kunnen vormen onder milde omstandigheden, waardoor geen schadelijke chemicaliën nodig zijn. De vormbaarheid van enzymen is een bijkomend voordeel voor het gebruik in industriële processen. Hierbij kunnen eigenschappen als selectiviteit, stabiliteit en katalytische efficiëntie getuned worden voor specifieke doeleinden door een veelvoud aan enzymengineeringmethoden, zoals directed evolution. Er zijn verscheidene enzymklassen, namelijk oxidoreductasen, transferasen, hydrolasen, lyasen, isomerasen, ligasen en translocasen. De oxidoreductasen, ook wel redoxenzymen genoemd, katalyseren oxidatie- en reductiereacties, waarbij ze vaak beter werken dan chemische methoden. Flavinebevattende monoöxygenasen (FPMO's), die in **hoofdstuk 1** werden geïntroduceerd, gebruiken de redoxactieve flavine cofactoren FMN en FAD. Deze cofactoren geven de enzymen een karakteristieke gele kleur. De FPMO's kunnen onderverdeeld worden in zes verschillende subklassen. Alle enzymen in dit proefschrift behoren tot de klasse B FPMO's en meer specifiek tot de type I flavine-bevattende monoöxygenasen (FMO's) en type I Baeyer-Villiger monoöxygenasen (BVMO's). Deze enzymen gebruiken de gereduceerde vorm van nicotinamide andenine dinucleotide (NADH) of nicotinamide andenine dinucleotide fosfaat (NADPH) als elektrondonor. Alhoewel FMO's en BVMO's gerelateerde sequenties hebben, verschillen zij substantieel in de reacties die zij katalyseren: FMO's katalyseren de oxygenatie van heteroatomen, terwijl BVMO's juist een Baeyer-Villiger-oxygenatie van ketonen katalyseren. **Hoofdstuk 1** geeft een introductie tot de FPMO's en geeft een overzicht van de (mogelijke) industriële toepassingen, metabole functies en de farmaceutische relevantie van FMO's uit bacteriën, planten en zoogdieren.

**Hoofdstuk 2** beschrijft een nieuwe bacteriële FMO uit *Nitricola lacsaponensis* (NiFMO), een alkalifiele, extremofiele bacterie. Dit enzym werd geïdentificeerd door genoom mining en vormt een thermostabieler alternatief voor de eerder beschreven FMO uit *Methylophaga aminisulfdivorans* (mFMO). NiFMO kon worden geproduceerd in grote hoeveelheden (100 mg L<sup>-1</sup>) als wateroplosbaar en actief FAD-bevattend enzym door overexpressie in *Escherichia coli*. De driedimensionale structuur van dit enzym werd verkregen met een resolutie van 1.8 Å. De relatief hoge thermostabiliteit van NiFMO (smelttemperatuur van 51 °C), de hoge tolerantie voor organische oplosmiddelen en brede substraatportfolio maken dit enzym een goede kandidaat voor toegepaste biokatalyse. Een mogelijke industriële toepassing is de productie van de kleurstoffen indigo en Tyrisch purper door de oxygenatie van, respectievelijk, indol en broomindol. Zo'n enzymatische productiemethode is een lucratief alternatief voor de huidige methoden, doordat hernieuwbare grondstoffen gebruikt kunnen worden en organische oplosmiddelen en andere chemicaliën juist vermeden kunnen worden.

---

Mensen bezitten vijf FMO paralogen, FMO1 tot en met FMO5, welke betrokken zijn bij de afbraak van schadelijke xenobiotica, zoals toxines en pesticiden. Menselijke FMO's (hFMO's) zijn farmacologisch zeer belangrijk, omdat zij zorgen voor de afbraak of activatie van veel medicijnen en soms een actieve rol spelen in bepaalde ziektebeelden. Ondanks hun grote farmacologische en medische relevantie zijn deze hFMO's, evenals FMO's van andere dieren, slecht gekarakteriseerd in vergelijking met andere xenobioticum-afbrekende enzymen, zoals cytochroom P450. Zo zijn er bijvoorbeeld geen gedetailleerde 3D-structuren bekend van deze enzymen. Dit komt deels doordat deze enzymen verankerd zijn aan het celmembraan en moeilijk op te zuiveren zijn. **Hoofdstuk 3** beschrijft het gebruik van ancestrale sequentie reconstructie (ASR) om de biochemische eigenschappen van drie voorouderlijke FMO's uit zoogdieren te bestuderen. De eiwitsequenties van AncFMO2, AncFMO3-6 en AncFMO5 werden gereconstrueerd met hoge A-posteriori waarschijnlijkheden (variërend van 0,98 tot 0,99) en de fylogenie van zoogdierlijke FMO's werd verder verfijnd, waardoor de vroege divergentie van FMO5 zichtbaar werd. De drie AncFMO's konden worden gekristalliseerd en hun structuur werd opgehelderd met een resolutie tussen 32,7 en 3.2 Å. Hierdoor konden hun specifieke membraanbindingseigenschappen en substraat specificerende tunnelnetwerken geïdentificeerd worden. De katalytische eigenschappen van deze voorouderlijke FMO's, zoals flavinereductiesnelheden en de formatie en verval van de peroxyflavine-intermediären, werden bestudeerd door stopped-flow metingen. Steady-state kinetiek werd ook bepaald door het meten van de afname in NADPH in de aanwezigheid of afwezigheid van bekende substraten voor menselijke FMO's. Een opmerkelijke uitkomst van deze studie was de grote overeenkomstigheid tussen het katalytische gedrag van de ancestrale en huidige FMO's. Hierdoor kunnen de voorouders dus dienen als modellen voor studies naar het gedrag en de eigenschappen van de huidige FMO's, waarmee zij 82 – 92% sequentie-identiteit delen.

In **hoofdstuk 4** werd de collectie van voorouderlijke zoogdierlijke enzymen verder uitgebreid met AncFMO1 en AncFMO4. Door een C-terminale extensie, welke uniek is voor mammaliane FMO4, kon AncFMO4 niet tot expressie worden gebracht in *Escherichia coli*. AncFMO1 kon wel tot expressie worden gebracht, waarna deze werd opgezuiverd en gekristalliseerd. De kristalstructuur had een resolutie van 3.0 Å en vormt een FAD-bevattend dimeer, net als de andere AncFMO's. De structuur bevat typische FAD en NAD(P) H bindingsdomeinen en een C-terminale helix voor membraanbinding. Verscheidene unieke eigenschappen werden gekarakteriseerd, waaronder een poreuze en aan het oplosmiddel blootgestelde actief centrum, evenals een nieuwe bindingsconformatie voor NADPH. Het enzym laat de algemene FMO-activiteit zien en oxideert de medicijn gerelateerde stoffen benzydamine, tamoxifen en thioanisole. AncFMO1 accepteert zowel NADPH als NADH als elektrondonor, zoals voorheen beschreven werd voor hFMO1. Deze eigenschap is karakteristiek voor de leden van de FMO1-clade. Net als de andere ancestrale FMO's laat ook AncFMO1 identiek katalytisch gedrag zien met hFMO1, waarmee het 89% van de sequentie-identiteit deelt. Hierdoor is het een bonafide model voor de studie naar de eigenschappen van hFMO1.

**Hoofdstuk 3 en 4** demonstreren de kracht van ancestrale sequentiereconstructie als een strategie voor de kristallisatie van eiwitten. Recent onderzoek liet zien dat FMO5 een uitbijter is binnen de vijf zoogdierlijke FMO's. Onderzoek toonde aan dat FMO5 zich gedraagt als een BVMO, terwijl de andere vier paralogen alleen heteroatomen kunnen oxygeneren. In hoofdstuk 3 speculeerden wij dat een glutamine-histidinesubstitutie Baeyer-Villiger-oxidatie-activiteit zou kunnen induceren. Deze substitutie resulteerde in verlies van alle activiteit bij AncFMO5, maar resulteerde niet in Baeyer-Villiger-activiteit bij AncFMO2 en AncFMO3-6. Hierdoor kwamen de volgende vragen over katalytische diversiteit binnen de FMO's naar voren: (i) Wat is de functie van de originele tetrapod FMO? (ii) Wat zijn de factoren die de switch tussen Baeyer-Villiger-oxidatie en heteroatoomoxygenatie bepalen?

In **hoofdstuk 5** reconstrueerden wij de ancestrale tetrapod FMO's met ASR. De fylogenie die wij eerder gebruikten voor de voorouderlijke enzymen van zoogdieren werd uitgebreid met additionele sequenties van *Testudines* (schildpadden), *Crocodylia* (krokodillen), *Aves* (vogels) en *Lepidosauria* (een subklasse van reptielen) claden en de root werd verder uitgebreid met *Actinopterygii* (beenvissen). De gegenereerde fylogenetische stamboom laat een snelle uitbreiding van FMO's in de tetrapod clade zien (FMO-explosie). De voorouder van de tetrapoden had één enkele kopie van het FMO-gen, welke door verschillende duplicatie-events resulteert in de vijf verschillende FMO afstammingslijnen. Wij reconstrueerden vier ancestrale sequenties, welke genaamd zijn naar de verschijning van de paraloog clade: tAncFMO1-5, tAncFMO5, tAncFMO1-4 en tAncFMO1-3. Het biokatalytische profiel van deze AncFMO's werd getest met een portfolio van bekende BVMO- en FMO-substraten. De oudste voorouder, tAncFMO1-5, kan beide reactietypes katalyseren. Deze vaardigheid is verloren gegaan bij de eerste duplicatie, die leidde tot tAncFMO1-4. Deze FMO kan namelijk alleen heteroatoomoxygenatie katalyseren. Een andere afsplitsing, tAncFMO5, kan echter beide reactietypen katalyseren. De residuen die potentieel verantwoordelijk zouden kunnen zijn voor de switch in oxygenatie-activiteit werden onderzocht door een combinatie van methoden, waarbij gekeken werd naar de reconstructiescore, de mate van behoud van residuen binnen de tetrapoden en de positie van aminozuren in de enzymstructuur. De substitutie van vier aminozuren in tAncFMO1-4 sequentie resulteerde in herstel van Baeyer-Villiger-oxidatie van fenylaceton en tegelijkertijd het verlies van heteroatoomoxygenatie-activiteit. Twee van de vier residuen, namelijk Thr60 en Ser222, zijn sterk geconserveerd binnen de heteroatoom oxiderende FMO's, waarin ze zich dichtbij de FAD cofactor, in het actieve centrum of in de 80-residuen lange insertie voor de substraattunnel bevinden. De twee andere residuen, namelijk His275 en His426, zijn cruciaal voor de oxygenatie van ketonen door hun coöpererende waterstofbrugnetwerk. In **hoofdstuk 5** werd de herkomst van deze katalytische functies herleid door een biochemische reconstructiemethode, die moleculaire fylogenie, bio-informatica, structuuranalyse, mutagenese en biochemische karakterisatie met elkaar combineert. Dit hoofdstuk belicht de evolutie van de verschillende oxygenatie-activiteiten binnen de familie van dierlijke FMO's.

---

Het werk dat beschreven wordt in dit proefschrift vergroot en verdiept onze kennis over FMO's. Allereerst door de karakterisatie van een nieuw ontdekte thermostabiele, bacteriële FMO die kan dienen als biotechnologische tool (hoofdstuk 2). Daarna met de opheldering van FMO-structuren uit zoogdieren, die nieuwe inzichten geven over de functie van deze ontgiftende enzymen (hoofdstuk 3 en 4). De beschreven structurele en biochemische data is van groot belang voor toekomstige farmacologische studies. Ten slotte werd een set van zorgvuldig gekozen ancestrale FMO's gekarakteriseerd, waardoor inzicht werd verkregen in de evolutie van tetrapode FMO's en de determinanten voor de FMO-BVMO-switch (hoofdstuk 5).

## Résumé en français

Traduit par Gautier et corrigé par ses parents

Le domaine de la biotechnologie a connu une croissance exponentielle au cours des dernières décennies. Cela est dû en grande partie au fait qu'elle représente souvent une alternative durable et écologique pour les industries chimiques et pharmaceutiques.

Les enzymes présentent à la fois une spécificité et activité catalytiques élevées et ouvrent de nouvelles voies vers la synthèse de produits à valeur ajoutée. Ces particularités évitent de recourir à des produits chimiques agressifs et nuisibles pour l'environnement. La malléabilité des enzymes est un atout supplémentaire car ces dernières peuvent être modifiées par une multitude de méthodes d'ingénierie moléculaire (évolution dirigée), afin d'ajuster leur sélectivité et d'améliorer leur stabilité ou leur efficacité. Il existe sept grandes classes d'enzymes : oxydoréductases, transférases, hydrolases, lyases, isomères, ligases et translocases. Les oxydoréductases, également appelées enzymes redox, sont capables d'effectuer des réactions d'oxydation et de réduction et surpassent souvent les approches chimiques. Les flavoprotéines mono-oxygénases (FPMOs pour *flavoprotein monooxygenases*), introduites au **chapitre 1**, sont des enzymes redox qui utilisent un cofacteur flavine redox-actif (FAD ou FMN), responsable de leur couleur jaune caractéristique. Les FPMOs sont classées en six sous-classes et les enzymes décrites dans cette thèse appartiennent toutes à la classe B. Les flavin-containing monooxygenases de type I (FMOs) et *Baeyer-Villiger monooxygenases* de type I (BVMOs) y ont été plus précisément étudiées. Ces enzymes emploient un cofacteur réduit de nicotinamide, soit un nicotinamide adénine dinucléotide (NAD) ou un nicotinamide adénine dinucléotide phosphate (NADP), comme donneur d'électrons. Bien qu'elles aient une séquence protéique très proche, les FMOs et les BVMOs diffèrent considérablement en ce qui concerne les réactions qu'elles catalysent: alors que les FMOs catalysent l'oxygénation des hétéroatomes, les BVMOs transforment des cétones via une oxydation dite de Baeyer-Villiger. De manière additionnelle à une introduction des FPMOs, le **chapitre 1** donne également un aperçu des perspectives industrielles, du rôle dans le métabolisme et de la pertinence pharmaceutique des FMOs bactériennes, végétales et mammifères. Le **chapitre 2** rend compte de l'étude d'une FMO bactérienne nouvellement découverte chez *Nitricola lacisaponensis* (NiFMO), une bactérie extrêmophile alcaline. L'enzyme a été identifiée via une approche explorative du génome, afin de trouver une alternative plus thermostable de la FMO bactérienne déjà décrite chez *Methylophaga aminisulfivorans* (mFMO). NiFMO a pu être fortement exprimée (environ 100 mg.l<sup>-1</sup>) en tant qu'enzyme soluble et active contenant de la FAD. La structure à 3 dimensions de l'enzyme a été élucidée avec une résolution de 1,8 Å. La thermo-stabilité relativement élevée de NiFMO (température de fusion de 51 °C), sa tolérance élevée aux solvants organiques et son large portfolio de substrats acceptés établissent l'enzyme comme un candidat viable pour la biocatalyse appliquée. Une application potentielle dans l'industrie des colorants serait possible, car elle peut être utilisée pour la production d'indigo ou de pourpre de

---

Tyr, en oxygénant respectivement l'indole ou le bromoindole. Un tel procédé à base d'enzymes serait attrayant comparé aux procédés de production chimique actuels. En effet il éliminerait le besoin de produits toxiques et nocifs pour l'environnement et pourrait ne dépendre que de matières premières renouvelables.

Les humains possèdent cinq paralogues de FMOs, nommées FMO1, FMO2, FMO3, FMO4 et FMO5, qui sont impliquées dans la désintoxication de nombreux xénobiotiques délétères dans le corps, y compris les toxines et les pesticides. La capacité des FMO humaines (hFMOs) à traiter (dégrader ou activer) les médicaments et leur implication dans certaines maladies révèlent leur grande importance pharmacologique. Cependant, malgré leur pertinence pour la santé humaine, les hFMOs, et tout autre FMO de mammifères, sont mal caractérisées par rapport à d'autres enzymes telles que les monooxygénases du cytochrome P450, capables, elles aussi, de dégrader des xénobiotiques. Par exemple, aucune information structurale détaillée n'est disponible pour les hFMOs. Cela s'explique en partie par le fait que les hFMOs et autres FMOs animales, toutes associées à une membrane, se sont révélés difficiles à purifier. Le **chapitre 3** décrit comment nous avons utilisé la reconstruction de séquences ancestrales (ASR) pour examiner trois ancêtres des FMOs de mammifères afin d'approfondir notre compréhension de leurs propriétés biochimiques. Les séquences protéiques d'AncFMO2, d'AncFMO3-6 et d'AncFMO5 ont été reconstruites avec des probabilités postérieures élevées (allant de 0,98 à 0,99) et la phylogénie des FMO5 de mammifères a été affinée afin d'inclure la divergence précoce des FMO5. Les trois AncFMOs ont pu être cristallisées, et ont été structurellement résolues à 2.7-3.2 Å, révélant leurs caractéristiques spécifiques de liaison membranaire ainsi que des réseaux complexes de tunnels pour le passage des substrats. Les études de cinétiques rapides ont été réalisées en utilisant la technique du *stop-flow*, afin d'étudier le taux de réduction de la flavine ainsi que la formation, puis le déclin de l'état intermédiaire de la peroxyflavine pendant le cycle catalytique. Les cinétiques à l'état d'équilibre ont également été mesurées en suivant la baisse de NADPH en l'absence et présence de substrats connus chez les FMOs des mammifères. De façon remarquable, toutes les AncFMOs ont exhibé un comportement catalytique semblable à leurs parents existants et peuvent donc être employées comme des modèles fiables pour représenter les hFMOs avec lesquelles elles partagent 82-92 % d'identité dans leurs séquences protéiques.

Dans le **chapitre 4**, nous avons complété notre collection de FMOs ancestrales de mammifères en reconstituant les séquences d'AncFMO1 et d'AncFMO4. Malgré le fait que la seconde possède une extension dans la partie C-terminale, unique aux FMO4 des mammifères, qui empêche son expression, AncFMO1 a pu être exprimée, caractérisée et cristallisée. La structure d'AncFMO1 a été élucidée à une résolution de 3,0 Å, révélant une enzyme dimérique associée à une FAD, semblable aux autres AncFMOs. La structure contient des domaines de liaison FAD et NAD(P)H typiques et une hélice en position C-terminale nécessaire à la liaison membranaire. Certaines caractéristiques uniques

ont été découvertes, un site actif poreux et exposé, ainsi qu'une nouvelle conformation du NADPH qui pousse le 2'-phosphate vers l'intérieur du domaine de liaison NADP<sup>+</sup> au lieu d'être étiré dans le solvant. L'enzyme affiche une activité FMO ordinaire et oxyde des composés médicalement utilisés, notamment le benzydamine, tamoxifène et thioanisole. Comme démontré précédemment dans hFMO1, AncFMO1 est capable d'accepter à la fois des cofacteurs NADH et NADPH comme donneurs d'électrons, une caractéristique uniquement décrite dans le clade des FMO1s. À l'identique des FMOs de mammifères précédemment reconstruites, AncFMO1 partage 89 % d'identité de séquence avec hFMO1 et exhibe un comportement catalytique identique, ce qui indique que l'enzyme ancestrale peut être utilisée comme un modèle structural bona fide de FMO1.

Les **chapitres 3 et 4** démontrent le pouvoir de la reconstruction de séquences ancestrales en tant que stratégie pour la cristallisation de protéines. Des récentes recherches ont révélé que parmi les cinq FMO des mammifères, FMO5 est un cas particulier. Il a été démontré que les FMO5s se comportent comme des BVMOs, tandis que les quatre autres paralogues ne font que catalyser l'oxygénation des substrats contenant des hétéroatomes. Dans le **chapitre 3**, nous avons émis l'hypothèse qu'une mutation glutamate-histidine est capable d'induire l'activité d'oxydation de Baeyer-Villiger dans AncFMO5. La substitution a entraîné une perte totale d'activité dans AncFMO5 mais n'a introduit aucune capacité d'oxydation de Baeyer-Villiger dans AncFMO2 et AncFMO3-6. Les questions relatives à la diversité catalytique au sein des FMOs qui se posent en conséquence sont les suivantes:

- (i) Quelles étaient les fonctions ancestrales de la FMO chez le tétrapode original ?
- (ii) Quels sont les déterminants derrière l'altération fonctionnelle entre les activités d'oxydation de Baeyer-Villiger et les oxydations des hétéroatomes?

Dans le **chapitre 5**, nous avons reconstruit les FMO ancestrales des tétrapodes à l'aide de l'ASR. La phylogénie utilisée pour les ancêtres des mammifères a été complétée avec des séquences supplémentaires des clades Testudines (tortues), Crocodylia, Aves (oiseaux) et Lepidosauria (sous-classe des reptiles) et la racine a été peuplée d'Actinopterygii (poissons osseux). L'arbre phylogénique généré confirme une expansion rapide des FMO (explosion) qui s'est produite dans le clade des tétrapodes. L'ancêtre des tétrapodes possédait une seule copie de FMO qui a été suivie plus tard par de multiples événements de duplication, menant aux cinq lignées différentes FMO. Nous avons récupéré quatre séquences ancestrales : tAncFMO1-5, tAncFMO5, tAncFMO1-4 et tAncFMO1-3. Le profil biocatalytique des tAncFMOs a été évalué via un portfolio de substrats composé de substrats connus pour être acceptés par les BVMOs ou FMOs. L'ancêtre aîné, tAncFMO1-5, est capable d'effectuer les deux types de réaction. Cette compétence a été perdue lors du premier événement de duplication, ce qui a conduit tAncFMO1-4 à ne pouvoir catalyser uniquement l'oxygénation des hétéroatomes. D'autre part, tAncFMO5 a gardé la capacité d'oxyder les cétones. Les résidus potentiellement responsables du changement de type d'activité d'oxygénation ont été étudiés par une approche multidisciplinaire, combinant



---

le score de reconstruction, le degré de conservation chez les tétrapodes et la localisation des acides aminés dans la structure enzymatique. La substitution de quatre résidus dans tAncFMO1-4 a eu comme conséquence le retour de l'habilité de l'enzyme d'exécuter l'oxydation de Baeyer-Villiger sur de l'acétone phényle, concomitante avec une perte globale d'activité vers des composés contenant des hétéroatomes.

Deux résidus, Thr60 et Ser222, sont hautement conservés dans les FMOs oxydants des hétéroatomes et sont respectivement situés proches du cofacteur FAD, dans le site actif, ou dans l'insertion de 80 résidus qui forme les tunnels d'entrée des substrats. Deux histines, His275 et His426, sont suggérées comme cruciales pour l'oxygénation des cétones via leur réseau de liaison hydrogène coopérant. Par une approche de reconstruction biochimique, combinant la phylogénie moléculaire, la bioinformatique, l'analyse structurale, la mutagenèse et la caractérisation biochimique, ce chapitre a démêlé la façon dont les fonctions enzymatiques émergent. Plus précisément, il met en lumière l'évolution des différentes activités d'oxygénation au sein de la famille des FMOs animales.

Les travaux présentés dans cette thèse approfondissent nos connaissances générales sur les FMOs. Tout d'abord, avec la caractérisation d'une FMO bactérienne thermostable nouvellement découverte, un nouvel outil biocatalytique a été ajouté à la boîte à outils des FPMOs disponibles (chapitre 2). Puis, avec l'élucidation sans précédent des structures FMO de mammifères (chapitres 3 et 4), de nouvelles connaissances sur le fonctionnement de ces enzymes détoxifiantes ont été obtenues. Les structures élucidées, ainsi que les données biochimiques générées, fournissent un point d'entrée pour de futures études pharmacologiques. In fine, grâce à la reconstruction d'un ensemble de FMOs ancestrales soigneusement sélectionnées, nous avons obtenu pour la première fois un aperçu des étapes de l'évolution des FMOs chez les tétrapodes, révélant ainsi les déterminants de la variation entre FMOs et BVMOs (chapitre 5).

## Curriculum Vitae



Born on the 5<sup>th</sup> May 1993, in Montivilliers (Normandie, France), Gautier started his career in science by obtaining a Bachelor degree in Biology at Université Le Havre in 2014. Afterwards, he did a Master in Biotechnology at University of Lille 1. During these two years, Gautier had the opportunity to travel abroad for international internship in academic labs to gain significant experience. The first internship in 2015 was at the Chalmers University of Technology in Göteborg, Sweden, where he worked for six months on enhancing the hydrolysis yield of thermophilic anaerobic digestion of cattle manure. One year later, new interests in enzyme engineering led him to the Concordia Centre for

Structural and functional genomics (Montréal, Canada) where he worked on developing the chemoenzymatic synthesis of glycosaminoglycan carbohydrates by engineering a glucosidase into a glycosynthase.

It was in 2017 that Gautier was granted the Marie Skłodowska-Curie fellowship (EU funded) as part of the OXYTRAIN Innovative Training Network and started in the lab of Marco Fraaije in Groningen, the Netherlands. From this point, he first focused on engineering flavin-containing monooxygenases (FMOs) to enzymatically produce the indigo dye. The research aim was later deviated to the reconstruction of ancestral FMOs that resulted in the first mammalian structures elucidated to date and a deeper insight of the catalytic behaviors and evolutionary paths of these enzymes.



## List of Publications

Lončar, N., Fiorentini, F., **Bailleul, G.**, Savino, S., Romero, E., Mattevi, A., & Fraaije, M. W.(2019).Characterization of a thermostable flavin-containing monooxygenase from *Nitrocola lacinisaponensis* (NiFMO). *Applied Microbiology and Biotechnology*, 103(4), 1755-1764. <https://doi.org/10.1007/s00253-018-09579-w>

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>

**Bailleul, G.**, Nicoll, C. R., Mascotti, M. L., Mattevi, A., & Fraaije, M. W.(2021).Ancestral reconstruction of mammalian FMO1 enables structural determination, revealing unique features that explain its catalytic properties. *The Journal of Biological Chemistry*, 296. <https://doi.org/10.1074/jbc.RA120.016297>



## Acknowledgments

Dear **Marco**, I guess you are used to reading always the same acknowledgements saying how grateful we are that you gave us this golden opportunity to join your group and here I am reminding you again. Thank you so much for everything, you were the best mentor I could have hoped for and you gave me the perfect amount of managing I needed. Sometimes, well many times actually, I felt completely lost and did not know which way to go, but after only 5-10 minutes discussing with you in your office I knew exactly where I was and where I had to go. You pushed me when it was needed, made me notice when I was stupid, reassured me when I had breakdowns and trusted me enough to not worry about me, especially at the end. One of my favorite moments with you was when you came to rescue me while I was lost in the woods and we managed to find our way back through swamps and brambles (beautiful metaphor for a PhD actually...).

Thank you.

Thank you **Callum**, my PhD mate and professional dabber, I am glad we were put together to work on our old babies. Half of my PhD is based on your MARVELOUS crystals, who would have guessed that a French and a British could work so well together? I am thankful for the laughs about our email and the time you welcomed me to Pavia and taught me about the silly membrane-protein purification. Please don't forget: Pasteur is Asian.

**Laura**, if Callum and I considered our ancestral proteins as our babies, you are for sure their mom. Thank you for the wonderful work you have been doing and your cheerful mind. I am also grateful for the time you spent teaching us about ASR during our time in Argentina, where we met the most welcoming people on Earth! I hope your nice little family will have a great time in Groningen.

Dear **Andrea (Mattevi)**, thank you for your contributions and insights that were a big part of our ancestral working group.

---

Thank you **Dick** for being part of my committee and for your vast knowledge in science (but not only!) that always reflects in relevant questions during work discussions.

When I started, I got lucky and was put next to **Antonija's** office desk, therefore starting one of the best friendships of my life. Thank you for being such a good friend; there are very few people to whom I can talk about so much heavy stuff but you are one of them. You are one of the purest people I know and I am glad I had the chance to meet you. From the bottom of my heart, thank you so much for being you.

**Eric**, the second half of that perfect couple, thank you for the nice times and another big/huge/enormous thanks for proof-reading my thesis ("*He did what??*") that even I didn't want to do. Both of you deserve all the best the world has to offer.

Dear **Misun**, you were part of the people that I got so much attached to. Thank you for letting me put all the music I wanted in the lab and to annoy you while you were working on your bench and ignoring me. I am also glad I was there to make you improve your ass-kicking technique on me. You know, we share quite a unique bond as you are the only person ever of whom I washed the feet. Thanks for the nice *walks* to talk but also to take a break from *work*.

My precious **Yapei**, I am most thankful for the time we had together. I saw you transforming from a shy and lost Chinese student into the lab boss that you are right now and that always made me smile. You know more than you think and you should be proud of yourself. As everyone, I have to thank you for the amazing food you cooked for me but also for always being here for me. You deserve the best and I am looking forward to see you successfully defend your thesis without any fear.

I was one of the younglings that **Elvira** (Romero) spent on quite some time helping, even if she was always so busy and I am more than thankful for that. You are also my stopped-flow mentor which became much more relevant for my thesis than I firstly expected so thank you for this as well.

**Brenda**, my Oxytrain mommy, you took care of the posters I forgot on the train, made sure I did my EU work and even enjoyed the 40€ empty piñata I bought for you. You were there for me during the worst parts of my PhD and I am forever grateful for it.

**Mr Kilvaras/Nikks/Nikolas**, thank you for not letting me be the biggest video game freak in the lab. I also enjoyed every time I was the only one able to understand your jokes and references. Please keep your beard sharp and your crypto wallet full.

My dear **Jeroen**, thank you for writing the Dutch summary, but also the many dances, yodeling and your dirty mind that never failed to make me smile. Even if I constantly had to push back my lab bench invasion, I am glad we were next to each other. I also believe you have one of the butts I slapped the most in my life...

**Estela**, I stayed half of my PhD in Lelystad and therefore did not party in Groningen much. Nevertheless, you were here the few times I was quite “over-enjoying” a party and I would like to thank you for these really fun times even if we may not remember everything properly.

**Simone**, thank you for the time you spent with me while I was just killing hFMO3. I hope to try your semicolon cocktail one day.

**Yiming**, thank you for the laughs about your self-made haircut and for always being smiling.

**Andrea**, my Oxytrain lil’sister. Thank you for the laughs and for giving me the opportunity to speak French, ce qui nous a permis de se plaindre des gens comme **Misun** à côté d’eux en sans qu’ils ne puissent se rendre compte. Merci aussi d’avoir été là quand j’en avait le plus besoin.



---

Thank you **Guang**, the latest newcomer in my office, to whom I believe I was (not yet) successful in transmitting my passion about stopped-flow.

**Cate**, you curly-haired monster, thank you for the time we spent happily insulting each other and for the future career reflections. I also loved learning new Italian slangs every time I scared you in the lab.

Thank you, **Milos**, our lab very demanded chemist. Thank you for teaching me about GC-MS, for the help whenever I had a question regarding my lack of chemistry knowledge and mostly for providing me so many M&Ms.

**Nikola**, our very own lab CEO, thank you for showing me around when I first started and for the talks about basically everything. You are always the person asking people with a true sincerity about how they are doing and I really appreciated that part of you. You and Lur are moving so much that I can't follow properly but I wish you both a really happy life in Groningen, Barcelona or wherever you see the right fit!

**Mohamed** was a nice and smiling guy in the first office, I enjoyed talking with him.

Thank you, **Sandra**, our dear lab secretary, for being so efficient and such a nice person.

I would like to thank all the people that were part of Oxytrain, the PIs and my fellow ESRs: **Federica, Nico, Marie, Gustavo, Monika, Alba, Callum, Eleonora, Matic, Lukas** and **Andrea**. I spent some real quality time with all of you during our many interesting workshops and conferences, with beers and jokes.

I would also like to thank many other people that contributed to a pleasant work environment and nice chats: **Daniël, Hugo, Saif, Christiaan, Hein, Eduardo, Qinlong, Henriette, and Andy.**

I had many issues with **steady-state kinetics** during these 4 years (ask anyone who was there) and that resulted in wonderful back pains. I finally overcame this hate-hate relationship and given how highly involved they were during my PhD, I decided they deserved a spot in the acknowledgement, even though I still cannot stand them.

Merci à aux **Acacias**, un groupe d'amis du collège qui s'est transformé au fur et à mesure en une seconde famille. Je ne vais pas tous vous faire un paragraphe parce que j'ai pas que ça à faire, vous savez déjà à quel point vous êtes importants pour moi, même si parfois vous avez du mal à me supporter (c'est mutuel). Merci donc à vous tous pour juste être vous-mêmes, **Marie, Esteban, Clément, Quentin, Yann, Teddy et Alicia.**

**Bérénice**, si je ne t'avais pas rencontré je n'aurais jamais fait de thèse donc je suppose que tu as plus que mérité ta place dans mes remerciements. Ces dernières années n'ont pas été faciles pour aucun de nous deux mais nous avons enfin réussi, même si ce n'est pas comme on l'avait imaginé, ensemble. De ces années je garde en mémoire les meilleurs moments, le ciel étoilé, la balade dans le parc complètement jetlagués, les cabanes et les très, très nombreux rires ensemble. Merci pour tout.

Je tiens à remercier mon **Papa** et ma **Maman**, qui, à aucun moment n'ont douté de moi et qui m'ont laissé choisir ma propre voie tout en me supportant de loin et en s'assurant que tout se passe pour le mieux. Je sais que ça fait partie de votre "travail" mais ne vous inquiétez pas pour moi, je vais continuer à tout déchirer, quels que soient les obstacles.

**Quentin**, mon Bro, tu ne te rends peut-être pas compte mais tu es un modèle pour moi et j'ai vraiment de la chance de t'avoir pour grand-frère. On partage tellement de choses et d'avis, aussi certains pourraient penser que je te copie avec mon obsession pour le Canada et.. ils ont peut-être raison au final ! Tu as toujours été mon roc et je me sens tellement privilégié par notre relation, du plus profond de mon cœur, merci pour tout.

---

**Angélique**, tu es une des personnes les plus pures que je connaisse, et je suis tellement heureux que tu aies choisi mon frère car cela m'a permis de te rencontrer. Tu donnes tellement de toi-même à tant de gens et étant donné que j'en fais largement parti récemment, je te remercie pour ça.

Vous avez tous les deux été là pour moi, et vous m'avez soutenu de façon inconditionnelle, je vous souhaite à tous les deux tout le bonheur du monde et j'ai hâte de découvrir un nouveau monde en votre compagnie !

I would like to thank my two lovely paranymphs, **Alejandro** and **Elvira** for their awesomeness. I was convinced that you were the best ones I could have chosen as my paranymphs when you picked up the mess I was in in the middle of the night and also later, thank you for that. By the time I am writing this, you don't know this yet but I will ask you to organize most of the defense party because I am too lazy, so thank you for that as well I guess!

**Alito**, my dear twin and secret lover (don't tell Mel), the one I always enjoyed talking and laughing with, thank you for the fun times and quality GIFs/memes. Watching you half-throwing up every time you try speaking French or procrastinating all the time (gladly, we finally got rid of your million bottles btw) always made my day brighter. My only regret is the fact that we only became so close at the second half of my PhD. If only I knew how close we would become and how much time we would spend together... I would have stayed further away from you. **Chupalo**. **Mel**, I am so glad I had the chance to also meet you, good luck to both of my favorite Colinas in Amsterdam, I will miss you!

Elvira, your friendship is sometimes quite exhausting but it is so much worth it. I like seeing you acting as everything is the end of the world without really believing it. I am glad I moved to your office close to the end. Even if my productivity decreased significantly, the *deep* discussions we had were worth it. Going for walks, slapping each other over stupid things we were saying, having dinners and breakdowns together made my problems less heavy and brought me a lot of laughs. I am (not really) sorry I spent so much time making fun of you, but you know I would have not if I didn't like you a lot... As people say *in Italy*, grazie mille.

Finally, I would like to end these acknowledgments with some piece of wisdom that I recently gathered. Many people say that a PhD is more than troublesome (and they are right!), but as is life. During my life, like everyone, I went through really difficult situations and the only way I made it through every time was only thanks to people that surrounded me with love and kindness. These people will of course recognize themselves here and I would never be able to express how much grateful I am to them. I truly believe the best investments one can make are these relationships in which you can give as much as you receive. I was and still am receiving so much from so many people that my goal is to keep giving as much as I can and keep the positivity wheel spinning for others.

Random acts of kindness, whichever their size, matter and make the world a slightly brighter place.

Thank you all.