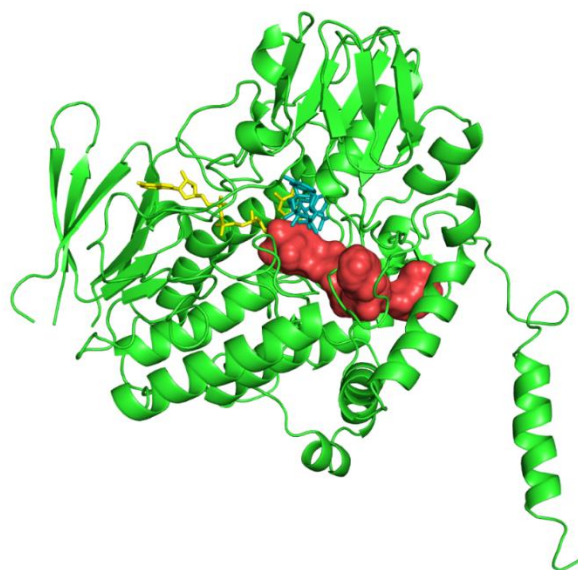




IUSS

Scuola Universitaria Superiore Pavia



Scuola Universitaria Superiore IUSS Pavia

**The Structural and Functional Characterization of Mammalian
Flavin-containing Monooxygenases using Ancestral Sequence
Reconstruction**

A Thesis Submitted in Partial Fulfilment of the Requirements

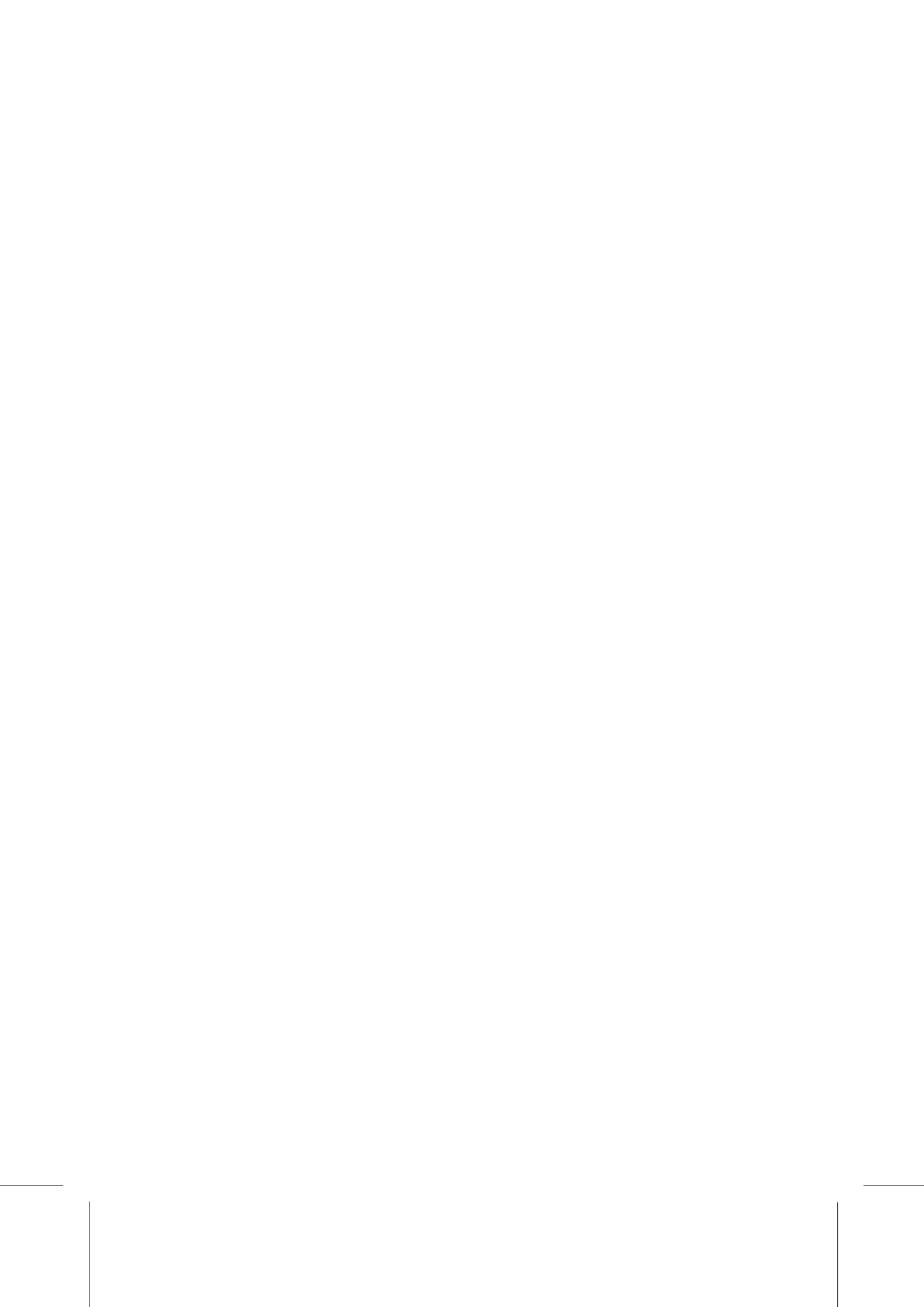
for the Degree of Doctor of Philosophy in

BIOMOLECULAR SCIENCE AND BIOTECHNOLOGY

by

Callum Robert Nicoll

November, 2020





IUSS

Scuola Universitaria Superiore Pavia

Scuola Universitaria Superiore IUSS Pavia

**The Structural and Functional Characterization of Mammalian
Flavin-containing Monooxygenases using Ancestral Sequence
Reconstruction**

A Thesis Submitted in Partial Fulfilment of the Requirements

for the Degree of Doctor of Philosophy in

BIOMOLECULAR SCIENCE AND BIOTECHNOLOGY

by

Callum Robert Nicoll

Supervisor: Prof. Andrea Mattevi, Dep. of Biology and Biotechnology "L. Spallanzani", University
of Pavia

Mentor: Dr. Filippo Fiorentini, Dep. of Biology and Biotechnology "L. Spallanzani", University of
Pavia;

November 2019



ABSTRACT

Flavin-containing monooxygenases (FMOs) are ubiquitous and prevalent in all domains of life. They are essential class B flavin-dependent monooxygenases involved in the metabolism of a swath of xenobiotics. Over the last 352 million years, there has been an explosion in the number of FMOs present in tetrapods – mammals, aves, reptiles and amphibians. Humans possess five different paralogs, FMOs 1-5, and have been demonstrated to oxidize a range of different compounds including drugs and pharmaceuticals. Typically, they oxidize soft nucleophilic centres, including sulfur and nitrogen atoms, through oxygen insertion. This increases the hydrophilicity of substrate and promotes their eventual excretion from the host. Human FMO5, however, was recently demonstrated to perform an alternative mode of action – Baeyer-Villiger monooxygenation – breaking a C-C bond and inserting an oxygen atom adjacent to a carbonyl functional group. Despite this, their low stabilities and poor expression levels have impeded structural characterization and thwarted our understanding behind their broad substrate profiles and mixed enzymatic activity.

Herein this body of work, Ancestral Sequence Reconstruction (ASR) was employed to improve our understanding behind the structural and biochemical properties of the FMO family. Employing ASR, we sought out to expand our evolutionary understanding of the FMOs by delineating the emergence of the paralogs throughout the jawed vertebrates (gnathostomata) and characterizing their various substrate profiles. Additionally, with the technique being prone to generate thermostable systems for a range of proteins, we exploited the technique with the possibility of generating thermostable FMOs that may possess a greater tendency to crystallize.

Four FMO ancestral sequences were inferred. Each represented the mammalian ancestor of FMOs 1, 2, 3 and 6 collectively (with FMO6 being a pseudogene in humans,) and 5, denoted as AncFMO1, AncFMO2, AncFMO3-6 and AncFMO5, respectively. They were successfully cloned and expressed in *E. coli* and purified displaying enzymatic activity. Remarkably, all the AncFMOs were successfully crystallized and conveyed resolutions from 2.7 – 3.2 Å, respectively. These AncFMOs, thus, represent the first human-like structures described hitherto. Each enzyme displayed a highly conserved paired-Rossmann fold motif accommodating FAD and NADP⁺ molecules in their crystal structures. Furthermore, they collectively portrayed a large C-terminal helix that anchors onto the membrane and a large unprecedented 80-residue length insertion that facilitates membrane binding and

orchestrates substrate transit. Fascinatingly, each FMO orchestrated unique tunnels for substrate passage. We speculate that their differing tunnel-residue linings consequently discriminates the various substrate profiles exhibited by each FMO. Furthermore, the crystal structures highlighted key residues within the active site that likely stabilize various flavin intermediates. Curiously, a key switch between a glutamate residue for a histidine was observed for AncFMO5. We postulate that this switch likely represents a pivotal transformation in activity that favours Bayer-Villiger monooxygenation.

This thesis documents the power behind ASR for protein crystallization and verifies the strength behind the approach for examining enzymatic activities. By structurally characterizing four FMOs, we can rationalise what features are essential for catalysis, membrane adherence and substrate uptake within the entire family. Furthermore, the heightened stabilities generated for each FMO corroborates the susceptibility ASR exhibits towards increasing protein thermostability. Henceforth, we implore the use of this methodology as an incentive for other recalcitrant and non-crystallizing protein families.

ACKNOWLEDGEMENTS

During this PhD project I have had the pleasure of meeting and working alongside many inspiring and supporting colleagues. The OXYtrain network is built up of 12 projects dispersed across the European continent with additional global collaborations. As I began to establish and build a foundation for my career, this network has allowed me to make many professional connections, build my confidence and appreciate the asset of partnerships. During this time, I was able to visit many countries within Europe that consequently allowed me to experience many different cultures and create unforgettable memories with newfound friends. A special shout out goes to my hilarious OXYtrain colleague, mentor and friend, Gautier Bailleul who has had to put up with me since day one. Throughout this project, Gautier was extremely available whenever I had questions and a fantastic host when I visited his laboratory as a visiting student. He has taught me many skills regarding gene synthesis, cloning and expression techniques and his efforts and dedication towards his work are admirable and inspiring. I am confident that we will maintain fruitful relationships, professionally and personally, in the future.

Coming from a purely Chemistry based background, my PhD project that focussed principally on structural biology and biochemistry initially seemed daunting. Fortunately, I had the pleasure of being supervised and mentored by, in particular, two dedicated researchers, Filippo Fiorentini and Maria Laura Mascotti, who both show unbridled enthusiasm towards their work that I would like to thank. Filippo spent a considerable amount of time during the early stages of my research project teaching me various skills necessary to tackle membrane proteins and protein crystallography. He is a patient and methodical researcher who I am everlastingly grateful to considering his teachings and counsel spurred on the enthusiasm I now hold towards my work. Laura similarly has been a great mentor. I had the pleasure of visiting Laura for three weeks in a small, but vibrant city in Argentina, San Luis. There I was able to learn the intricacies behind the methodology that fuelled the project. Laura is very talented in her field of research and takes great interest in all aspects of the project. She is always available for questions regarding the project we share but equally also for new projects and ideas that are passed her way. I am very grateful to her for correcting this thesis and more so, her zeal towards using ancestral sequence reconstruction for unravelling staggering biochemical questions has inspired me to pursue a career in this field.

The Biocry laboratory has been a great host over this PhD project and I am fortunate enough to spend another year there. Past and current lab members have been more than welcoming and have made my stay in the lab a treat. Elisa and Chiara have shown me that if you really put your mind to it, entire protein purifications can be performed in a matter of hours and should be considered as a sport. Biagina took great pride in her work and was always willing to help me with new instruments and was always so nice to speak to. Marta Ceccon is a resilient and inspirational researcher with a profound determination to overcome problems. She always managed to put a smile on everyone's faces and despite her absence she is still an integral member of our lab community.

In the laboratory I have the pleasure to be working with many kind (, crazy) and supporting colleagues. I started my PhD alongside Luca who has been a good friend to me during the project. He is very precise in his thought process and was always available regarding advice with protein crystallography and data analysis. I wish him all the best in his next career step and look forward to seeing how Swiss culture can mould a *Napolitano*. Laura Rotilio may have the most contagious laugh I have ever had the pleasure of hearing. She is always so cheerful and even when research gets tough, she always finds a way to make us laugh. Sara is unbelievably kind, finds time for everyone and is more than happy to listen to our concerns and problems. As such, she is a fountain of knowledge and we can always count on her for advice (and gossip). I will always cherish our enjoyable coffee breaks and will try to integrate them into my future career. Joana is also very kind and is a top hugger. I will never forget that when meeting her for the first time, she insisted on giving me a hug and welcoming me into the group. She is very easy to talk to and is always willing to help regardless of the situation. She is very motivated and takes her work very seriously and even when the workload seems immense, pulls through. Laura Alvigini is a relatively recent addition to the laboratory, and I have had great pleasure working more closely with her on a new project. Her dedication to her work is formidable and she doesn't take no for an answer. Her work ethic is meticulous, and she will stop at nothing until she reaches a conclusion. If she wasn't a PhD student, I would not be surprised if she was an acting general for the *Comando interforze per le operazioni delle forze speciali*. Friso is also relatively new to the laboratory and despite COVID-19 disrupting our time together, he is always available. It should be said that he is very skilled in his field of research and was always willing to spend time with us in the laboratory (or over email) teaching us the tricks behind molecular cloning. Finally, Stefano has been the joker of the laboratory and always manages to make us smile. He is always available and helped me many times with protein crystallization experiments.

A big shout out goes to all the masters and bachelor students who I have had a pleasure of working with. Their ambition for research and dedication to learning was and continues to inspire me as an aspiring academic. I should mention that they all helped me understand the ins and outs of a thriving Italian culture. In particular, I would like to thank Dana for her help during my PhD as a master's student. She was the first student I ever had and was a pleasure to teach. Having her as student helped me develop teaching skills that will inevitably be useful in the future. I wish her all the best for her future endeavours and look forward to seeing what she is capable of.

Of course, I must thank my supervisor, Andrea Mattevi, who has taught me a great deal during my time here. He has watched over me closely during my work, helped me stay on track, whilst at the same time giving me a lot of freedom to tackle the problems as I saw fit. He is mindbogglingly passionate about science and is always ready with new elevator-pitch ideas to keep the momentum going. His tenacity and curiosities towards research and biochemical phenomena are inspiring and I hope that I can continue to carry his working and research ideologies forward in my career. Even being abroad, my family have also been extremely supporting during my career. They never once held me back or made me re-question my career and, in that sense, have been totally selfless. I truly can't thank them enough. Finally, I would like to thank my dear, Marta. Throughout this entire project she has been a rock and has helped me excel as researcher, but more importantly, giving me the confidence as an individual to aim big in all aspects of life. She has showed me the true essence of Italian culture and transformed my life in Italy completely. These words merely resemble a small piece of what she means to me as I cannot thank her enough. I cannot wait to see what the future holds for us.





Table of Contents

ABSTRACT	v
ACKNOWLEDGEMENTS	vii
LIST OF FIGURES	xv
LIST OF TABLES	xxv
LIST OF SCHEMES	xxvii
LIST OF SYMBOLS	xxix
1. INTRODUCTION.....	1
1.1 FLAVIN ENZYMOLOGY	1
1.2 HARNESSING MOLECULAR OXYGEN FOR ENZYMATIC ACTIVITY	6
1.3 FLAVIN-DEPENDENT MONOOXYGENASES.....	7
1.3.1 Overall Catalysis and known flavin intermediates	8
1.3.2 Classes of the Flavin-dependent monooxygenases.....	10
1.4. CLASS B FLAVIN-DEPENDENT MONOOXYGENASES	16
1.4.1 Shared structural and enzymatic features	16
1.4.2 Baeyer-Villiger Monooxygenases – Structure and Function.....	17
1.4.3 Nitrogen-hydroxylating monooxygenases – Structure and function.....	21
1.4.4 YUCCAs – Function.....	23
1.5 FLAVIN-CONTAINING MONOOXYGENASES – FMOS.....	24
1.5.1 Overview – mode of action and substrate preference	24
1.5.2 Prevalence and Biodiversity – Bacteria and fungi.....	25
1.5.3 Human FMOS: localization, distribution, polymorphisms and paralogs.....	28

1.5.4 Endogenous roles governed by human FMOs.....	30
1.5.5 Kinetic characterization of pig FMO1.....	33
1.5.6 Substrate metabolism and conversion.....	35
1.6 ANCESTRAL SEQUENCE RECONSTRUCTION.....	39
1.6.1 Overview.....	39
1.6.2. Case Study.....	40
1.6.3. The development of ASR.....	42
1.6.4. Performing ASR.....	44
2. MATERIALS AND METHODS.....	45
2.1 CLONING AND EXPRESSION OF THE ANCFMOS.....	45
2.2 SITE-DIRECTED MUTAGENESIS OF MUTANTS ANCFMO2-E281H, ANCFMO3-6-E281H, ANCFMO5-H282E AND ANCFMO2-E281A MUTANTS.....	46
2.3 CELL DISRUPTION, EXTRACTION, AND PURIFICATION OF ANCFMOS.....	46
2.4 WESTERN BLOT FOR ANCFMO4 EXPRESSION.....	47
2.5 DETERGENT SCREENING FOR ANCFMO EXTRACTION.....	48
2.6 PREPARATION OF HUMAN FMO3 AND HUMAN FMO5.....	48
2.7 KINETIC CHARACTERIZATION.....	48
2.7.1 Steady state kinetics of AncFMOs 2, 3-6 and 5.....	48
2.7.2 Steady state kinetics of AncFMO2-E281A.....	49
2.8 MELTING TEMPERATURE ASSAYS.....	49
2.8.1 ThermoFAD.....	49
2.8.2 TychoTMNT.6.....	49
2.9 CRYSTALLIZATION AND STRUCTURAL DETERMINATION.....	50
2.9.1 AncFMOs 2, 3-6 and 5.....	50
2.9.2 AncFMO1.....	51
2.10 ANCESTRAL SEQUENCE RECONSTRUCTION.....	51
3. RESULTS.....	53
3.1. ANCESTRAL FLAVIN-CONTAINING MONOOXYGENASE CLONING AND TRANSFORMATION	53
3.2 ANCESTRAL FLAVIN-CONTAINING MONOOXYGENASE EXPRESSION, PURIFICATION & CRYSTALLIZATION.....	53
3.2.1 Initial Purification – Determining solubility of Ancestral FMOs.....	54
3.2.2 His-trap Nickel 1 Purification.....	55

3.2.3 Detergent Screening for efficient AncFMO extraction	56
3.2.4 6xHis-SUMO Tag Cleavage and reverse Ni purification	57
3.2.5 Size Exclusion Chromatography for AncFMO2 and AncFMO3-6	60
3.2.6. Purification of AncFMO5	62
3.2.7 Purification of AncFMO1	67
3.2.8 AncFMO4 expression evaluation	74
3.2.9 Crystallization screenings, optimizations and structural elucidation	75
3.3 ANCESTRAL-SEQUENCE RECONSTRUCTION UNVEILS THE STRUCTURAL BASIS OF FUNCTION IN MAMMALIAN FMOs	83
3.4 ENZYMATIC ANALYSIS OF ANCFMO2-E281A	115
3.4.1 Thermostability Assessment	115
3.4.2 Steady-state Kinetics	116
3.5 ANCESTRAL RECONSTRUCTION OF MAMMALIAN FMO1: UNIQUE STRUCTURAL FEATURES EXPLAIN ITS CATALYTIC PROPERTIES	117
3.6 PHYLOGENETIC ANALYSIS OF THE FMO PROTEIN FAMILY IN JAWED VERTEBRATES	185
3.6.1 Phylogenetic Analysis of Flavin-containing monooxygenases	185
3.6.2 Evolution of the Flavin-containing monooxygenases	187
3.6.3 Ancestral Sequence Reconstruction	189
3.6.4 The role of FMO4	193
4. DISCUSSION	199
5. CONCLUDING REMARKS	205
REFERENCES	207
APPENDIX A. PHYLOGENETIC INFERENCE OF THE FLAVIN-CONTAINING MONOOXYGENASES IN JAWED-VERTEBRATES	223



LIST OF FIGURES

Figure 1.1.1: Isoalloxazine-derived cofactors and their various quinone states. Riboflavin represents the essential building block that is transformed into the organic cofactors FMN and FAD. The key aromatic isoalloxazine ring that executes catalysis is shown in yellow on the riboflavin molecule.

Figure 1.1.2: The various quinone states occupied by Flavins: A: Numbering of the isoalloxazine ring. The lower case a notation refers to the carbon atoms fused between each ring. B: various quinone states. The three representations shown on the left represent the protonated states, whilst those on the right represent their deprotonated forms. Each row from top to bottom portray the quinone (oxidized), semiquinone (one-electron reduced) and hydroquinone (two-electron reduced) states, respectively. Vertical chemical transformations describe proton coupled electron transfers (PCET), whilst horizontal steps highlight protonation conversions.

Figure 1.1.3: The UV/Vis spectra for the Flavin redox states: A. The UV/Vis spectrum illustrates the absorption bands, with their respective molar absorption coefficients, for the oxidized FAD (FADox), two-electron reduced FAD (FADred), the neutral one-electron reduced semiquinone (SQneutral) and the anionic one-electron reduced semiquinone (SQanionic). Spectra for each state is shown with an arrow. Whilst the FADred state is colourless, the FADox, SQneutral and SQanionic are yellow, blue and red in colour, respectively. These distinct absorption spectra allow the quinone states to be easily tracked through catalysis. B. The Molecular orbital representation of the π HOMO and π LUMO states of FADox that give rise to the principal electronic transitions shown in the spectra. The UV/Vis spectra have been adapted from Schwinn et al [Schwinn, 2020].

Figure 1.2.1: Molecular Orbital Diagram of molecular oxygen and its activation by reduced Flavin. A. The molecular orbital diagram is shown of molecular oxygen. The atomic orbitals of the single oxygen atoms, 2s and 2p (with the 1s orbital omitted for simplicity), are shown either side of the molecular orbitals. Each molecular orbital is derived from a combination of atomic orbitals, indicated with dashed lines. The arrows represent the spin states of each electron that occupies the orbital. Molecular orbitals are labelled in the following manner: firstly, with a numerical coefficient to represent the quantity of a molecular orbital, with 1 being the first, 2 the second etc. For example, $3\sigma_g^*$ represents the third σ_g molecular orbital in the molecule. The second term describes the type of molecular orbital, i.e. π or σ , with the term in subscript, g or u, representing the symmetry describing the orbital. The asterisk shown in superscript indicates antibonding orbitals. B. The reaction

of reduced Flavin with dioxygen in terms of spin states. Initially, the reduced flavin occupies a singlet state (paired electrons), whilst dioxygen describes a triplet state (two unpaired electrons with identical spin states) as shown in A. In step 1, one electron is transferred from the reduced flavin to dioxygen with a weak redox potential for the generation of superoxide (-330 mV). The caged intermediates can now spin flip their electrons, step 2, to populate a new covalent bond and generate a covalent intermediate as shown in step 3.

Figure 1.3.1.1: Catalytic cycle for Flavin-based substrate monooxygenation. A reducing agent is first utilized to activate the isoalloxazine ring, FAD_{red}, step I. Step II corresponds to the generation of an oxygen-based FAD intermediate, FAD_{int1}. Step III follows and involves the addition of a substrate which is oxidized with the FAD_{int1} being converted to FAD_{int2}. The final step, IV, includes the release of the newly oxidized product from the enzyme complex and the breakdown of FAD_{int2} to reproduce the oxidized FAD, FAD_{ox}. It should be noted that for simplicity, substrate is shown to enter at step III, albeit the substrate may enter in steps II or III. Importantly however, Flavin oxidation by molecular oxygen takes place first.

Figure 1.3.1.2: Flavin-based catalytic intermediates. The number of different intermediates observed during oxidative catalysis are documented and listed regarding their tendencies to perform either nucleophilic-, radical- and electrophilic-based reactions.

Figure 1.3.2.1: Reaction pathways carried out by the different flavin-dependent monooxygenases. Each reaction pathway is shown in a different colour. Pathways shown in red and yellow illustrate the one-component monooxygenase systems whereby all reaction steps are performed by a single enzyme. These include Classes A, B and G. The green pathway corresponds to the two-component monooxygenases whereby the reduced flavin species, FAD_{red}, is generated in a separate enzyme and depicted in dashed lines. FAD_{red} is then transferred into the monooxygenase for catalysis. These include classes C, D, E and H. Pathways in blue and violet represent internal monooxygenases type I and type II, respectively. In these scenarios, the substrate generates FAD_{red} and is then further modified. This includes classes G and H. Type I internal monooxygenases use the generated hydrogen peroxide to modify the initially product further outside the enzyme, depicted as a dashed line. Type II internal monooxygenases are generated as N5-oxide bearing FAD intermediates first and then react with the substrate to produce an intermediate product, Pro1, that then reduces the FAD_{ox} to FAD_{red}, releasing the final product, Pro2. Molecular oxygen regenerates the FADN5-O- to restart the catalysis. Note that in some instances the FAD is representing FMN (Table 1.3.1).

Figure 1.4.2.1: Crystal structure of a cyclohexanone monooxygenase. The protein has been split into three sections that represent FAD, NADPH and substrate binding domains that are coloured in yellow, cornflower blue and orange, respectively. The black ring highlights the entry point for substrates.

Molecules of FAD and NADP⁺ are shown in grey and dark pink, respectively. PDB ID: 4RG3 [Yachnin, 2014].....19

Figure 1.4.2.2: Active site of the CHMO from *Rhodococcus sp. HI-31*. Residues are labelled and coloured with respect to the domain they are found in. FAD, NADPH and substrate binding domains are coloured in yellow, cornflower blue and orange, respectively. The black ring highlights the entry point for substrates. Molecules of FAD and NADP⁺ are shown in grey and dark pink, respectively. PDB ID: 4RG3 [Yachnin, 2014].

Figure 1.4.3.1: Three-dimensional structure of an NMO and its active site. Left, overall three-dimensional structure of the NMO, SidA from *Aspergillus fumigatus* (PDB: 4B63) [Franceschini, 2012]. Right, active site of SidA complexed with the substrate, ornithine, shown in green. Active site residues and FAD are shown in white and dark grey, respectively.

Figure 1.4.3.2: Protein surface of SidA from *Aspergillus fumigatus*. Surface of SidA is depicted using a hydrophobicity scale whereby dark orange is very hydrophobic and dark blue is very hydrophilic. Ornithine is shown in white, PDB: 4B63 [Franceschini, 2012].

Figure 1.5.2.1: Structural features of a bacterial FMO. Left, overall structural topology of a bacterial FMO from *Methylophaga sp. strain SK1*. Middle, active site of the FMO. Residues are shown in white and labelled accordingly. Right, entrance site for substrates of the FMO. FAD and NADP⁺ molecules are shown in yellow and cornflower blue, respectively. Tertiary structure is shown in dark green and the surface of protein is depicted as charge, with blue, white and red representing positive, neutral and negative charges, respectively. PDB ID: 2vq7 [Alfieri, 2008].....26

Figure 1.5.2.2: Structural features of a yeast FMO. Left, overall structural topology of a yeast FMO from *S. pombe*. Right, active site of the yeast FMO in the presence of methimazole (Met). Residues are shown in white and labelled accordingly. PDB ID: 1vqw [Bonanno, 2017].....27

Figure 1.5.6.1: FMO1 substrates un-oxidized by FMO2. Left, Imipramine. Right, Chlorpromazine. Dashed circles highlight the site of oxidation.....37

Figure 1.6.1: Overview of the Ancestral Sequence Reconstruction process. The individual steps are shown with available programmes and servers described. The maroon boxes portray the sequence collection and multiple sequence alignment process required to generate the phylogeny. The turquoise boxes describe the phylogenetic tree curation necessary to illustrate the evolution of the protein of interest. The light green boxes describe how the ancient sequence are constructed.....40

Figure 1.6.2.1: Case Study - Ancestral Sequence Reconstruction for Class B monooxygenases. A phylogenetic tree of a protein family that display BVMO and FMO members are shown. Enzymes that display BVMO and FMO activities are shown in orange and blue squares, respectively.

Enzymes exhibiting mixed functionality are shown with a chequered blue and orange box. Ancestral intermediates for class B monooxygenases are described as AncBMOs. To determine the origin and development of the BVMO and FMO activities, AncBMO1 and AncBMO2 would be extracted for catalysis.

Figure 3.2.1.1: SDS-PAGE evaluating the solubility of AncFMO3-6. Both membrane fractions (solubilized in Triton-X 100) and aqueous fractions were run on a gel, represented as MEM and SN, respectively. Both lanes conveyed a 73 kDa band that may represent the enzyme of interest. Values 75 and 50 depict Molecular weights shown in the REF lane (reference markers) in kDa.

Figure 3.2.2.1: Spectral features and purity after the first step of affinity chromatography purification for AncFMO2 and AncFMO3-6. Left and right panels describe the UV/Vis spectra and SDS-PAGEs for AncFMO2 and AncFMO3-6, respectively. Each spectrum illustrated distinct FAD features with absorption maxima at 449 and 444 nm for AncFMO2 and AncFMO3-6, respectively. Gels portray the washing steps performed during the purification and the purity obtained. REF refers to the standard molecular weight markers with 75 and 50 being described in kDa. FT is the Flow through. Values 0, 5, 30 and E refer to imidazole washes of 0, 5, 30 and 300 mM, respectively.

Figure 3.2.3.1: SDS-PAGE analysing the extent of protein extraction exhibited by a series of different detergents for AncFMO3-6. Octyl glucoside (OG); dodecyl beta-maltoside (DDM); Triton-X-100 reduced (TRX); Lauryl dimethylamine N-oxide (LDAO); Lauryl Maltose Neopentyl Glycol (LMNG); Sodium dodecyl sulfate (SDS); control membranes without addition of detergent (CTRL); reference markers (REF). Molecular weight markers indicated by values 75 and 50 expressed in kDa.....59

Figure 3.2.4.1: Ni-based affinity chromatography step for the purification of SUMO-His-tag cleaved AncFMO2 and AncFMO3-6. Chromatograms A and B show the elution profiles for AncFMO2 and AncFMO3-6, respectively, with peaks I and II depicted. Peaks I were concentrated down and displayed classical Flavoprotein UV/Vis spectra for both AncFMO2 and AncFMO3-6 as portrayed in panels C and D, respectively. SDS-PAGEs of both peaks I and II from each chromatogram, A and B, respectively, document their high purities. Elution Chromatograms show the wavelengths 280, 370 and 450 nm depicted in blue, red and pink lines, respectively. Pressure, conductivity and percentage of elution buffer (300 mM imidazole) used are shown in colours dark green, brown and light green, respectively. Values 75 and 50 represent standard molecular weights in kDa.....62

Figure 3.2.5.1: Gel filtration purifications for AncFMO2 and AncFMO3-6. Top panels show elution chromatograms for AncFMO3-6 (top) and AncFMO2 (bottom). Each chromatogram has their corresponding purities illustrated on an SDS-PAGE. Molecular weights are described in kDa with respect to the reference lane. Final UV/Vis spectra for both AncFMO2 and AncFMO3-6 after concentrating are shown on the left and right, respectively. Traces

coloured blue, red and pink in the chromatogram describe absorptions at wavelengths of 280, 370 and 450 nm, respectively. Traces in dark green and brown represent the pressure (Mpa) and conductivity (mS/cm) during the purification, respectively.

Figure 3.2.6.1: SDS-PAGE of AncFMO5 Ni1 purification. Cell free extract (supernatant before membrane solubilization) (CFE); concentrated CFE (CFEc); Flow through (FT); 0 mM Imidazole (0); 5 mM Imidazole (5); 30 mM Imidazole (30); 300 mM Imidazole (E).

Figure 3.2.6.2: UV/Vis spectrum and SDS-PAGE after Ni1 purification of AncFMO5 expressed at 17 C. Left and right panels describe the UV/Vis spectra and SDS-PAGEs for AncFMO5, respectively during the first affinity-based chromatography purification step. SDS-PAGE portrays the washing steps performed during the purification and the purity obtained. REF refers to the standard molecular weight markers with 75 and 50 being described in kDa. FT is the Flow through. Values 0, 5, 30 and E refer to imidazole washes of 0, 5, 30 and 300 mM, respectively.

Figure 3.2.6.3: Second Ni affinity-based chromatography for the His-SUMO-tag cleaved AncFMO5. The chromatogram describing the elution profile is shown above with blue, red and pink traces representing the wavelengths at 280, 370 and 450 nm, respectively. The trace in light green portrays the stepwise gradient of buffer B (300 mM imidazole) used as washing steps during purification. The bottom panels represent the final UV/Vis spectrum of the concentrated peak I and the SDS-PAGE of both I and II peaks, shown on the left and right, respectively. REF represents the molecular weight markers with values 75 and 50 indicating molecular weight bands in kDa.66

Figure 3.2.6.4: Gel filtration of AncFMO5. The chromatogram describing the elution profile is shown above with blue, red and pink traces representing the wavelengths at 280, 370 and 450 nm, respectively. The bottom panels represent the final UV/Vis spectrum of the concentrated peak I and its respective SDS-PAGE, shown on the left and right, respectively. REF represents the molecular weight markers with values 75 and 50 indicating molecular weight bands in kDa.

Figure 3.2.7.1: Initial affinity-based chromatographic step of AncFMO1 using a Ni resin. The UV/Vis spectrum of the concentrated eluent, E and the imidazole washes conducted are displayed in an SDS-PAGE. Lanes FT, A, 30, E and REF refer to flow-through, buffer A, 30 mM imidazole, elution buffer (300 mM imidazole) and reference markers, respectively. Values 75 and 50 show the molecular weight bands given in kDa.....69

Figure 3.2.7.2: AncFMO1 Ni2 purification after 6xHis-SUMO cleavage. The chromatogram describing the elution profile is shown above with blue, red and pink traces representing the wavelengths at 280, 370 and 450 nm, respectively. The green trace highlights the concentration of buffer B (possessing 300 mM imidazole) used in a stepwise fashion. The bottom panels represent the final UV/Vis spectrum of the concentrated peak I and

its respective SDS-PAGE, shown on the left and right, respectively. REF represents the molecular weight markers with values 75 and 50 indicating molecular weight bands in kDa.....70

Figure 3.2.7.3: Size exclusion chromatography purification step of AncFMO1 solubilized in DDM (0.03% w/v). The dashed line highlights the elution volume of the protein, 12.6 mL. Traces shown in blue, red and pink depict wavelengths of 280, 370 and 450 nm, respectively.

Figure 3.2.7.4: 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, octyl glucoside (OG). CTRL 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.

Error! Reference source not found.....74

Figure 3.2.7.6: AncFMO1 gel filtration after protein expression at 17 C. The top panel shows the chromatogram representing the elution profile. The blue, red and pink traces portray the absorptions at wavelengths 280, 370 and 450 nm, respectively. The bottom panels represent the final UV/Vis spectrum of the concentrated peaks A and B with their respective SDS-PAGE, shown on the left and right, respectively. REF represents the molecular weight markers with values 75 and 50 indicating molecular weight bands in kDa.

Figure 3.2.8.1: AncFMO4 expression and solubilization screening. Left, SDS-PAGE corresponding to the various detergents used to solubilize AncFMO4 from the membranes, including the aqueous supernatant. Right, 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). RE refers to molecular weight markers with weights 75 and 50 indicated in kDa. For clarity, certain lanes are indicated with black lines.....77

Figure 3.2.9.1.1: Crystals of AncFMO3-6. Crystals formed after 4 days at 20 °C in 0.1 M HEPES.Na pH 7.5; 5% (v/v) Isopropanol; 10% (w/v) PEG 4000 conditions.....78

Figure 3.2.9.1.2: Crystals of AncFMO3-6. Crystals formed after 2 days at 20 °C in 0.1 M HEPES.Na pH 7.5; 5% (v/v) Isopropanol; 10% (w/v) PEG 4000 conditions in the presence of 20% OG.

Figure 3.2.9.1.3: Crystals of AncFMO2. Crystals formed after 2 days at 20 °C in 0.1 M Tris pH 8.5; 0.01 M NiCl₂; 10% (w/v) PEG 2000 MME conditions.....80

- Figure 3.2.9.1.4: Detergents used for crystallization. Three detergents that were crucial for crystallization growth are shown. CYMAL-6, octyl-glucoside (OG) and dodecyl-beta-D-maltoside.....81
- Figure 3.2.9.2.1: Crystals of AncFMO3-6, solubilized in CYMAL-6, formed after 2 days at 20 °C in 0.1 M Sodium Acetate pH 4.6 and 10% (w/v) PEG 4000 conditions.
- Figure 3.2.9.2.2: Crystals of AncFMO5 solubilized in DDM. Crystals formed after 1 day at 20 °C in 0.1 M HEPES pH 6.8-7.0 and 10% (w/v) PEG 4000 conditions.
- Figure 3.2.9.2.3: Crystals of AncFMO1. Crystals formed after 2 days at 20 °C in 0.1 M HEPES pH 7.5; 8% (w/v) PEG 4000.....84
- Figure 3.3.1: Crystal structures of each Ancestral FMO. The crystallographic dimers of AncFMO2, AncFMO3-6 and AncFMO5 are shown in gree, purple and orange, respectively. Cofactors FAD and NADP+ are shown in tallow and cornflower blue, respectively. Organic molecules are shown in green including CYMAL-6, DDM and HEPES.....87
- Figure 3.3.2: Distribution of charge around the surface of AncFMO2, with red, white and blue representing negative, neutral and positive, respectively. On rotation about 90°, the large parallel hydrophobic strips across the bottom of the dimer are visible, lined by a ring of positively charged residues indicated by black dashed boxes.
- Figure 3.4.2.2: Michaelis-Menten plot for the NADPH oxidase activity of mutant AncFMO2 E281A. Readings were plotted as triplicates. Vmax and KM were $0.076 \pm 0.003 \text{ s}^{-1}$ and $12.3 \pm 2.3 \mu\text{M}$, respectively. Protein concentration was $1 \mu\text{M}$119
- Figure 3.5.1: Principal structural features of AncFMO1. Top panel describes the crystallographic dimer of AncFMO1 resolved at 3 Å resolution. The bottom left panel depicts the active site cavity of AncFMO1 with a solvent accessible path portrayed with glycerol molecules shown in green. The bottom right panel shows a superposition of AncFMO1 (red) with AncFMO3-6 (purple) with changes in structural motifs emphasised. Organic molecules representing DDM or glycerol are shown in green, respectively. Molecules of FAD and NADP+ are shown in yellow and cornflower blue, respectively.
- Figure 3.6.1.1: Schematic representation of the evolution of the Flavin-containing monooxygenases in vertebrates. Clades representing the FMO paralogs 5, 4, 3/6, 2 and 1 are shown in orange, dark green, purple, light green and red triangles, respectively. Ancient FMOs extracted for characterization and their respective origins are described in the tree. These Ancient FMOs are denoted as ElderFMOs - EldFMOs. EldFMOs extracted include: EldFMO1-5, EldFMO1-4, EldFMO1-3/6, EldFMO5 and EldFMOF. Unclassified FMO sequences are claded as triangles and labelled according to their distribution and origin with X or Y. For example, FMO clades that diverge after the duplication event described by node 541, are denoted as X, whilst those before are labelled as Y. The preceding term highlights their

location within the species tree. For example, Reptilia X are FMOs found in reptilia and diverged after the EldFMO1-5-duplication event. Clades of Amphibia Y, Amphibia X and Reptilia X are shown as yellow, light blue and brown triangles, respectively. The root of the tree is shown as a dark blue triangle that represents fish FMOs represented with a fish silhouette.....188

Figure 3.6.1.2: Evolution of the BVMO and FMO activities in jawed vertebrates. In this simplified tree, the evolution of the FMOs are shown with respect to the EldFMOs selected for biochemical characterization. The orange and blue colours describe BVMO and FMO activities respectively. Branches and clades coloured in blue and orange represent unknown activity. FMO5 is shown in orange as human FMO5 was shown to display BVMO activity. FMOs 1, 2 and 3/6 were shown to perform conventional FMO oxidation and are coloured in blue. FMO4 has not been characterized to date and remains multicoloured.

Figure 3.6.2.1: Distribution of the FMOs throughout the classes of tetrapoda. The species tree for the tetrapoda super class are shown with the emerging classes shown and labelled. On the right, the number of FMOs possessed in each class are listed. FMOs XA, XR and Y are described as FMOX amphibia, FMOX reptilia and FMOY shown only in amphibia (see Figure 3.6.1.1).....192

Figure 3.6.3.1: Structure-based sequence alignment of all AncFMOs and EldFMOs. The alignment was generated by ESript 3.0 using AncFMO3-6, PDB ID: 6SE3. Residues highlighted in red represent a site with high similarity among the submitted sequences. Residues boxed in blue outlines imply highly conserved regions. Residues shown in white with red background are completely conserved. The secondary structure of the AncFMO3-6 sequence is conveyed above the alignment. The black arrow points at site 281.

Figure 3.6.3.2: Key residues among the ancestral FMOs. Top and bottom panels depict key residues that mutate during gene duplications, illustrated by AncFMO2 and AncFMO5, respectively, posing as models. Residues that change from EldFMO1-5 through to EldFMO1-3/6 are labelled with sidechains shown, except for P273/4 that changes in AncFMO4 (see later). Residues shown in white and orange represent AncFMO2 and AncFMO5, respectively. Residues making up loop 3 are shown in ribbon form to emphasise their differing conformations with key side chains displayed. Residues situated above the isoalloxazine ring are shown as atoms. FAD and NADP+ molecules are shown in yellow and cornflower blue colours, respectively.

Figure 3.6.4.1: Structure-based sequence alignment of all mammalian ancestral FMO paralogs and their respective human forms. The alignment was generated by ESript 3.0 using the PDB ID: 6SE3. Residues highlighted in red represent a site with high similarity among the submitted sequences. Residues boxed in blue outlines imply highly conserved regions. Residues shown in white with red background are completely conserved. The

numbering above shows the site number based on the PDB reference structure, with its secondary structures conveyed above.

Figure 3.6.4.2: Sequence alignment of all the mammalian ancestral FMO paralogs including the sauria ancestral FMO4 focussing on loop 3. Alignment was generated using ClustalW with the increasing conservation portrayed with single dots, double dots and asterisks, respectively. The dashed red box highlights key residues leading up towards the active site from sites 273-283. The numbers on the right describe the final site for each sequence in this region. AncFMO4 mammalian and saurian ancestors are listed as AncFMO4m and AncFMO4s, respectively.



LIST OF TABLES

Table 1.3.2.1: Classification of the Flavin-dependent monooxygenases.....	12
Table 3.4.1.1: Melting temperatures of AncFMO2 wildtype and mutant in the presence and absence of oxidized cofactor, NADP ⁺ . Concentrations were 4 and 200 μ M for protein and NADP ⁺ , respectively.....	117
Table 3.4.2.1: Steady-state kinetics and binding constants of NADPH oxidase activities for various FMO constructs. AncFMO2, Human FMO3 and Human FMO5 data was extrapolated from section 3.3.....	118
Table 3.6.1.1: Elder FMO sequences to be reconstructed based on the expanded phylogeny	



LIST OF SCHEMES

Scheme 1.1.1: Overall reaction scheme for oxygen-utilizing flavoproteins. Each reaction can be split into halves, reductive (shown in red) and oxidative (shown in, blue and green) boxes respectively. The reductive-half reaction activates the isoalloxazine ring for catalysis and then follows two possible oxidative-half reactions – ping-pong or sequential. The difference between the two mechanisms is simply at what reaction step does the product, P, produced from the oxidation of the substrate (which in-turn reduces the FAD), S, exits the catalytic site. If it leaves immediately after flavin reduction the system displays ping-pong activity. Alternatively, if P leaves during the final step, the mode of catalysis is sequential. In this pathway the presence of P during oxidation of the reduced Flavin system is key for catalysis and the reactions proceeds through a ternary complex consisting of the electron acceptor (O₂), product and reduced Flavin.

Scheme 1.4.1.1: Mode of catalysis for Class B Flavin-dependent monooxygenases – NMOs, FMOs, BVMOS and YUCCAs. For simplicity, one substrate is shown for each enzyme. Ornithine, trimethylamine, cyclohexanone and indole-3-pyruvate are shown for NMOs, FMOs, BVMOs and YUCCAs, respectively..... 17



LIST OF SYMBOLS

k_{cat}	= Turnover number
K_d	= Dissociation constant
K_M	= Michaelis constant
k_{off}	= Rate off
k_{on}	= Rate on
T_m	= Melting temperature
ε	= Molar absorption coefficient



1.INTRODUCTION

This section is designed to introduce the reader to the field of flavin enzymology with an increasing specificity towards human Flavin-containing monooxygenases. The section starts by introducing flavoproteins in general, followed by the integration of molecular oxygen into catalysis. An overview of flavin-dependent monooxygenases will then be discussed to document the breadth of monooxygenases present in literature to date and their differing classifications. The reader is then directed to the Class B Flavin-dependent monooxygenases to portray key elements of catalysis and structural topology that then leads to human Flavin-containing monooxygenases that represent the centre piece of the thesis.

1.1 FLAVIN ENZYMOLOGY

Since the revolutionary discovery of the Old Yellow enzyme in 1932 [Wellner, 1967] by Warburg and Christian, our understanding and portrayal of flavoproteins has continuously surmounted our expectations. To our knowledge, approximately 1% of all known enzymes are indeed, flavoproteins [Piano, 2017], implying that their biochemistry and roles in biological processes are integral to life. These proteins have been demonstrated to fuel multiple processes including, but not limited to, energy production, light emission, DNA repair, apoptosis, detoxification, oxygenation and many more [Joosten, 2007]. All these enzymes share a common feature; they utilize an isoalloxazine ring derived from the prosthetic groups Flavin adenine dinucleotide (FAD) and Flavin mononucleotide (FMN) that are built from the precursor, riboflavin (Figure 1.1.1) [Romero, 2018]. Initially, the idea was conceived that proteins containing these prosthetic groups would conduct identical mechanisms and biochemistries and were expected to instigate very similar reactions [Wellner, 1967]. However, over the last century, it has come to fruition that a myriad of different modes of action are conducted.

The Flavin cofactor is an organic molecule that can occupy three different redox states: the oxidized flavin, Flox (also denoted as the quinone), the one-electron reduced state, Flsq (semiquinone), and finally the two-electron reduced state, Flred (hydroquinone) [Romero, 2018] (Figure 1.1.2). These different states fine tune the redox potential of the cofactor and whilst most flavoproteins non-covalently bind the prosthetic group, some utilise histidine, serine and tyrosine residues to establish covalent linkages that further manipulate the redox potential [Joosten, 2007; Caldinelli, 2008]. The oxidized flavin species possesses the famous yellow colour which led to the first flavoprotein being described as the 'Old Yellow enzyme'. This feature can be better described by its UV/Vis absorption spectrum and each

quinone-like flavin species has its own unique spectrum (Figure 1.1.3). In FAD_{ox} , the highest occupied molecular orbital (HOMO) exhibits a strong bonding interaction over C4a, C10a and N10 atoms, whereas the lowest unoccupied molecular orbital (LUMO) introduces antibonding interactions between the N5 and C4a, and N10 with C10a atoms, respectively [Schwinn, 2020]. It is this electronic transition between the HOMO and LUMO orbitals that gives rise to the absorbance measured at 450 nm. Furthermore, as electrons populate the LUMO, the π bonds between N5 and C4a, and N1 and C10a atoms break in the FAD_{red} state. The band at 370 nm is derived from multiple electronic transitions [Schwinn, 2020].

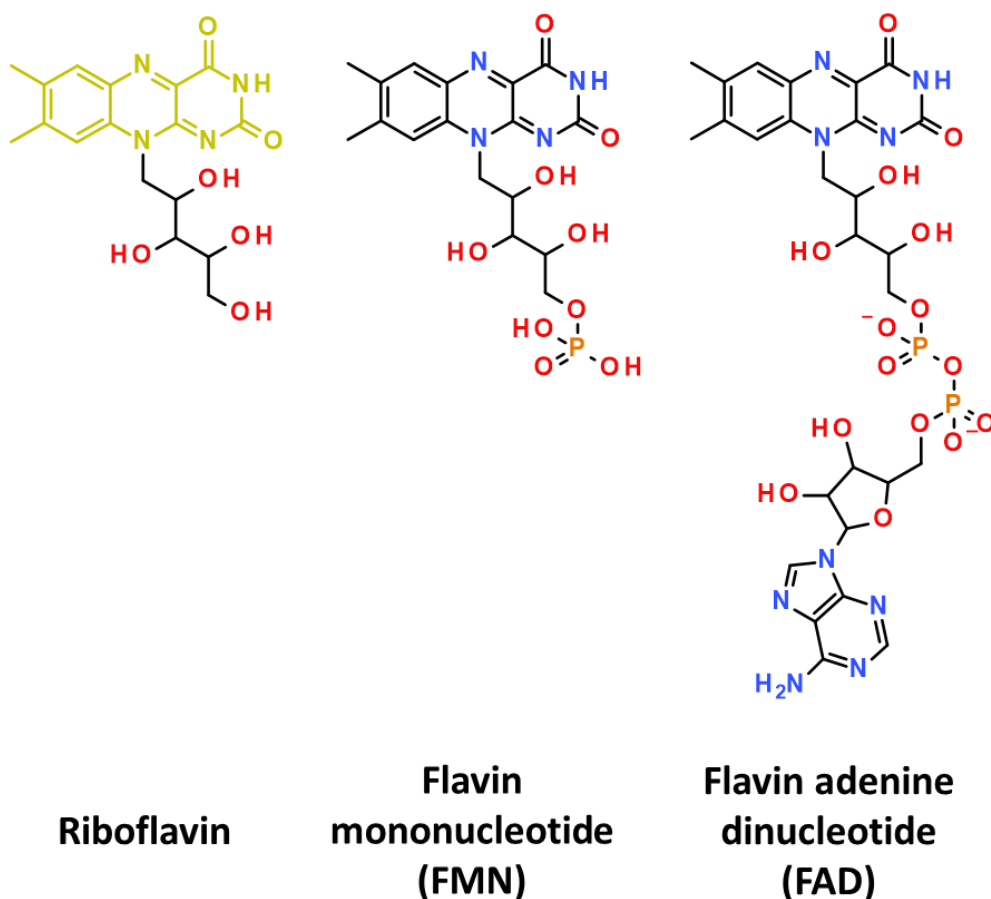


Figure 1.1.1: Isoalloxazine-derived cofactors and their various quinone states. Riboflavin represents the essential building block that is transformed into the organic cofactors FMN and FAD. The key aromatic isoalloxazine ring that executes catalysis is shown in yellow on the riboflavin molecule.

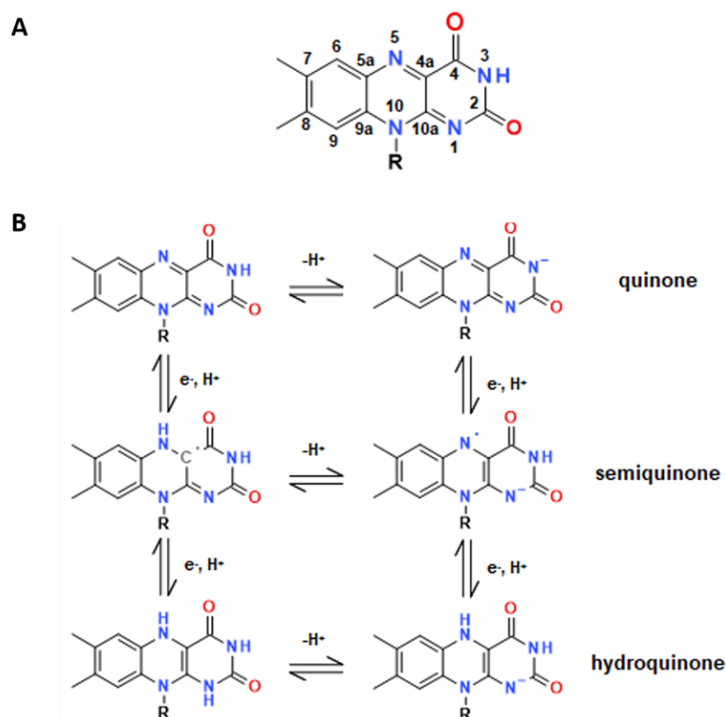
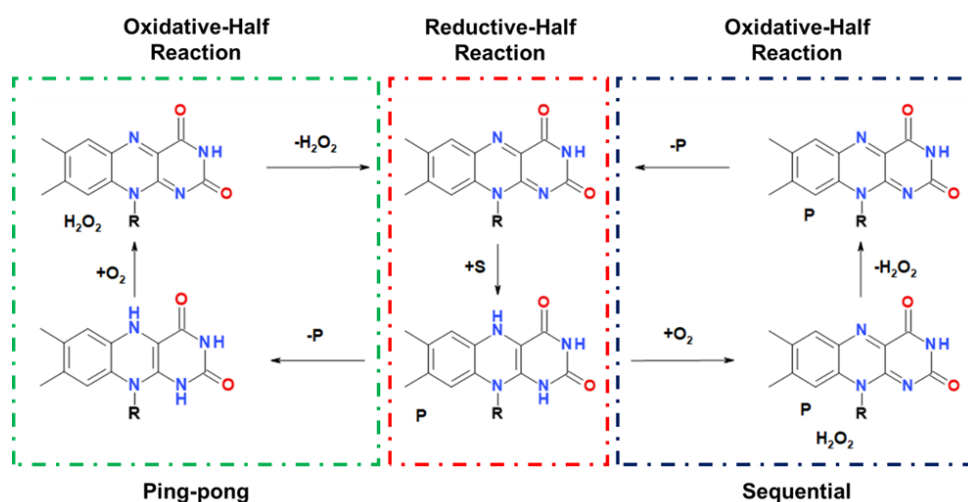


Figure 1.1.2: The various quinone states occupied by Flavins: **A**: Numbering of the isoalloxazine ring. The lower case a notation refers to the carbon atoms fused between each ring. **B**: various quinone states. The three representations shown on the left represent the protonated states, whilst those on the right represent their deprotonated forms. Each row from top to bottom portray the quinone (oxidized), semiquinone (one-electron reduced) and hydroquinone (two-electron reduced) states, respectively. Vertical chemical transformations describe proton coupled electron transfers (PCET), whilst horizontal steps highlight protonation conversions.

Single-electron reduction results in a one electron population of the LUMO orbital, converting it to a singly occupied molecular orbital (SOMO). The orbital possesses a lower energy, resulting in a red-shift of the 450 nm band observed in both semiquinone states. These new SOMO orbitals undergo further excitation to a higher antibonding π -orbital described at 370 nm [Schwinn, 2020]. The addition of a proton to SQanionic state further stabilises the SOMO orbital, resulting in a broad red-shifted band from 500 nm to 600 nm. The fully reduced FAD system, however, fully populates the once π LUMO orbital (making it now the π HOMO) removing all possible electronic transitions between the molecular orbitals shown that promote colouration of the isoalloxazine ring. The UV/Vis spectrum and molecular orbitals are derived from the thorough work carried out by Henriques et al., and Kar et al., respectively [Henriques, 2010; Kar, 2019].

Scheme 1.1.1: Overall reaction scheme for oxygen-utilizing flavoproteins. Each reaction can be split into halves, reductive (shown in red) and oxidative (shown in, blue and green) boxes respectively. The reductive-half reaction activates the isoalloxazine ring for catalysis and then follows two possible oxidative-half reactions – ping-pong or sequential. The difference between the two mechanisms is simply at what reaction step does the product, P, produced from the oxidation of the substrate (which in turn reduces the FAD), S, exits the catalytic site. If it leaves immediately after flavin reduction the system displays ping-pong activity. Alternatively, if P leaves during the final step, the mode of catalysis is sequential. In this pathway the presence of P during oxidation of the reduced Flavin system is key for catalysis and the reactions proceeds through a ternary complex consisting of the electron acceptor (O_2), product and reduced Flavin.



or not at all, with dioxygen, and use other electron acceptors such as quinones to relieve the reduced flavin state [Romero, 2018; Frébortová, 2004; Wang, 1991]. The latter groups all employ molecular oxygen as an electron acceptor: the flavoprotein oxidases use molecular oxygen, for the most part, to produce hydrogen peroxide (H_2O_2) as a by-product, whilst the monooxygenases activate the oxidizing agent in a range of covalent-based intermediates in order to insert an oxygen atom into a substrate (discussed below in section 1.3 Flavin-dependent monooxygenases) [Romero, 2018; Gadda, 2012].

Reducing the isoalloxazine ring often takes place through direct electron transfer, or hydride transfer, with Nitric Oxide Synthases being a beautiful example exhibiting both characteristics in concert [Stuehr, 2009]. Moreover, whilst a large proportion of flavoproteins exchange electrons with molecular oxygen, the catalytic scope is greatly enlarged by cooperating with other cofactors such as nicotinamides, Iron-Sulfur clusters and heme in order to execute catalysis [Piano, 2017; Peters, 2016]. In doing so, these

enzymes can transfer reducing equivalents across protein scaffolds and direct catalysis at different sites [Magnani, 2017]. Therefore, it can be surmised that two catalytic roles have been envisioned by flavin systems over evolution: one scenario describes the isoalloxazine-containing prosthetic group as an electron shuttle, transmitting electrons across individual proteins, or multiple domains, enabling chemical transformations. The second situation revolves around the FAD or FMN cofactor acting as the key benefactor, at the heart of the active site, that chemically converts the substrate into product.

1.2 HARNESSING MOLECULAR OXYGEN FOR ENZYMATIC ACTIVITY

Today, there is approximately 21% of O₂ in our atmosphere. Over millennia, nature has succeeded in utilizing this abundant resource for metabolic processes. In fact, approximately 300 million years ago (Mya) the predicted abundance of O₂ in the atmosphere sharply peaked to 33% [Kumar, 2017]. During this period, possessing enzymes that can fuel oxygen-based catalysis may have been a competitive advantage for the host. Within this sub-section, I describe how molecular oxygen is harnessed for enzymatic activity by flavin-dependent enzymes.

O₂ has a triplet ground state derived from two unpaired electrons occupying equivalent spin states in two separate antibonding orbitals (Figure 1.2.1, A) [Gadda, 2012]. In order to maintain the total spin state of the reacting system, the reduced isoalloxazine ring first donates an electron to O₂ to generate superoxide and semiquinone isoalloxazine ring caged intermediates (Figure 1.2.1, B) [Romero, 2018]. This step creates a doublet spin state superoxide molecule, with one unpaired electron that is paired with the single electron occupied isoalloxazine ring semiquinone intermediate. The singly paired electron can now undergo spin flipping as the overall energy of the intermediate no longer changes. As such, the newly spin flipped electron can now receive an additional electron from the semiquinone isoalloxazine ring intermediate in two modes: through outer-sphere electron transfer supplemented with protons to generate hydrogen peroxide as observed for the flavin-dependent oxidases [Romero, 20018; Gadda, 2012]. Alternatively, via the formation of a covalent bond to create a transient C4a(hydro)peroxy flavin adduct (Figure 1.2.1, B).

When comparing the reduction potentials of O₂/H₂O₂ and FAD_{ox}/FAD_{red} couples, respectively, the process described above is thermodynamically favourable: the former couple possesses a redox potential of +281 mV, whilst the latter is -207 mV in aqueous solution (pH 7.0) and ranges from -400 - +150 mV in flavoproteins [Mayhew, 1999; Massey, 1995; Heuts, 2009]. The range of reduction potentials observed for flavoproteins greatly illustrates the importance the residue environments and active sites play during catalysis. Further scrutiny however on the individual redox processes for O₂/O₂⁻ and FAD_{sq}/FAD_{red}, -330 mV and between -400 - -50 mV for flavoproteins, respectively, implies that for many flavin-containing systems the first step of the reaction (Figure 1.2.1,

mentioned, the flavin-dependent monooxygenases, unlike flavin-dependent oxidases, utilize molecular oxygen to generate a potent intermediate for catalysis. The key intermediate that will be discussed predominantly in this thesis is the well renowned C4a-(hydro)peroxy flavin intermediate. The number of oxygen-based flavin intermediates detected, however, is continuously expanding and makes the chemistry behind this prosthetic group increasingly fascinating [Leys, 2020]. Within this section, I will document the various flavin-dependent monooxygenases currently and describe their classifications.

1.3.1 Overall Catalysis and known flavin intermediates

The flavin-dependent monooxygenases mode of catalysis can be summarised in four steps (Scheme 1.1.1 and Figure 1.3.1.1). The isoalloxazine ring must first be reduced by a two-electron reducing agent to generate the pivotal hydroquinone intermediate that can react with molecular oxygen [Romero, 2018]. Depending on the protonation state, the reduced flavin is typically either singly charged (protonated once) or neutral (protonated twice). This reduction step can be performed by different reducing agents, such as NADPH, NADH or internally with substrate, and may take place inside or outside the monooxygenase peptide scaffold (see section 1.3.2 and Figure 1.3.2.1) [Huijbers, 2014; van Berkel, 2006]. This step is referred to as the reductive-half reaction (Scheme 1.1.1) [Palfey, 2010].

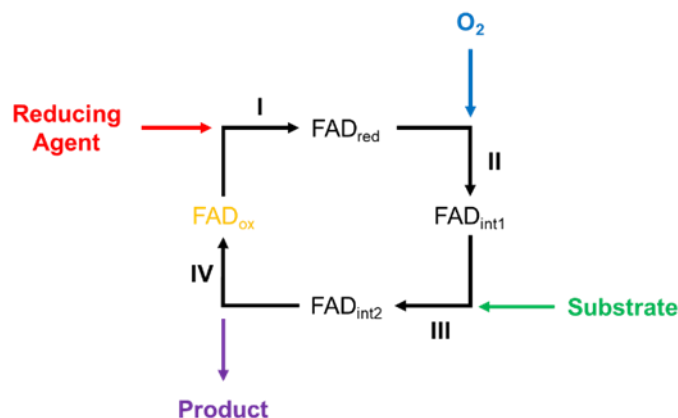


Figure 1.3.1.1: Catalytic cycle for Flavin-based substrate monooxygenation. A reducing agent is first utilized to activate the isoalloxazine ring, FAD_{red}, step I. Step II corresponds to the generation of an oxygen-based FAD intermediate, FAD_{int1}. Step III follows and involves the addition of a substrate which is oxidized with the FAD_{int1} being converted to FAD_{int2}. The final step, IV, includes the release of the newly oxidized product from the enzyme complex and the breakdown of FAD_{int2} to reproduce the oxidized FAD, FAD_{ox}. It should be noted that for simplicity, substrate is shown to enter at step III, albeit the substrate may enter in steps II or III. Importantly however, Flavin oxidation by molecular oxygen takes place first.

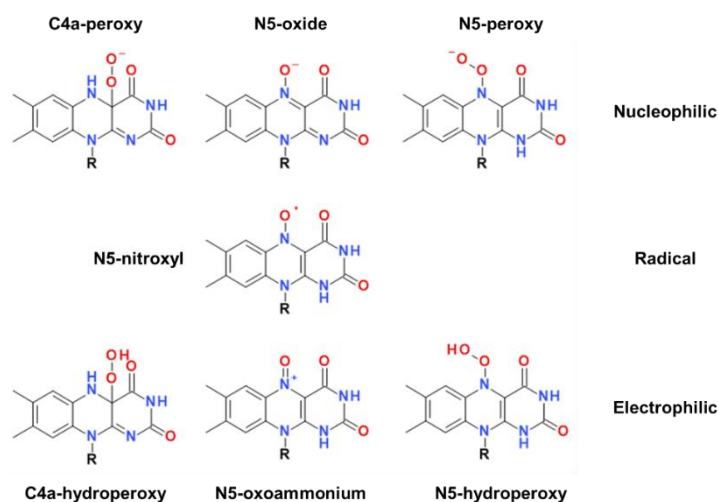


Figure 1.3.1.2: Flavin-based catalytic intermediates. The number of different intermediates observed during oxidative catalysis are documented and listed regarding their tendencies to perform either nucleophilic-, radical- and electrophilic-based reactions.

Once reduced, the cofactor's new reduction potential and electron-rich isoalloxazine ring can reduce molecular oxygen to form a range of intermediates which can then oxidise the substrate of interest (Figure 1.3.1.2) or uncouple to form hydrogen peroxide (Figure 1.3.1.1) [Mattevi, 2006]. The generated flavin intermediate will then harness one of the two oxygen atoms for substrate oxidation, whilst the second oxygen atom is then reduced to water. This is referred to as the oxidative-half reaction and can undergo either 'ping-pong' or 'sequential' mechanisms (Scheme 1.1.1) [Romero, 2018]. In the absence of substrate, the monooxygenase will act as an oxidase, performing two proton-coupled electron transfer, reducing molecular oxygen to hydrogen peroxide (Scheme 1.1.1) [Mattevi, 2006].

Electron directionality for flavin-dependent monooxygenases can be defined as binomial. Electrons can either be transported towards the activated flavin intermediates - electrophilic attack - or alternatively, directed onto an oncoming substrate - nucleophilic [Mattevi, 2006]. Importantly, these pathways are governed by the residue environment shrouding the isoalloxazine ring. For example, the C4a-(hydro)peroxy flavin intermediate (Figure 1.3.1.2) represents one of these states that has been shown to pursue both electron transfer modes [Romero, 2018]. Typically, when the state is protonated, the hydroperoxyl moiety becomes susceptible to electrophilic attack. This typically involves electron-rich centres like nitrogen and sulfur atoms that can afford to donate lone pairs of electrons. On the other hand, when deprotonated, the electron-rich and negatively charged peroxy group engages with electrophilic centres, such as carbonyls [Huijbers, 2014; van Berkel, 2006].

The protonation state is therefore pivotal to the biochemistry ensued by the enzyme. Additionally, with this subtle difference allowing entirely different reactions to transpire illustrates and substantiates the versatility of the isoalloxazine ring.

For many years, the oxygen-derived flavin intermediates were perceived to be built solely at the C4a position of the isoalloxazine ring. However, in the last decade, several systems have been shown to utilise the N5 position for catalysis. Firstly, Teufel et al documented in 2013, that an enzyme, EncM, was involved in the synthesis of the antibiotic, enterocin in *Streptomyces maritimus*, produced an N5-oxide flavin-intermediate [Teufel, 2013]. More specifically, it is utilized this intermediate to perform the dual oxidation of a large polyketide chain that subsequently triggers several chemical transformations, including Favorskii-type rearrangements. Recently, Matthews et al, [Matthews, 2020] illustrated an additional mode of catalysis governed by monooxygenases. The authors portray the formation of an N5-(hydro)peroxy intermediate that harnesses the N5 position over the conventional C4a. Similarly, however, the N5-(hydro)peroxy intermediate inserts an oxygen atom into an organic substrate, but in this scenario, the oxygen atom substitutes a leaving group instead of direct addition.

As illustrated in Figure 1.3.1.2, there are multiple intermediates used during catalysis [Romero, 2018; Leys, 2020]. Each brings different benefits for the reaction in question. For one, the intermediates are usually either electrophilic or nucleophilic with the only exception being the N5-nitroso radical which is advantageous for the substrate in question. The choice of intermediate utilised greatly depends on the reaction and the dimensions of the active site, and the positioning of the substrate and the relevant coenzymes. Regarding the C4a-(hydro)peroxy intermediates, the C4a-O bond is projected outwards from the *si*-face of the flavin ring, suggesting that both substrates and molecular oxygen bind and approach perpendicularly to the face of the ring. The recently documented N5-(hydro)peroxy intermediate moreover was illustrated to interact first with molecular oxygen on the *re*-face, before then positioning the N5-O bond on *si*-face to interact with the oncoming substrate [Matthews, 2020].

1.3.2 Classes of the Flavin-dependent monooxygenases

The flavin-dependent monooxygenases make up a significant proportion of flavoproteins and can pursue a whole range of different chemical transformations. These reactions include, but are not limited to: sulfoxidations, epoxidations, aromatic hydroxylations and oxidative decarboxylations [Huijbers, 2014; van Berkel, 2006; Mascotti, 2016]. Consequently, these enzymatically driven conversions have been exploited for a multitude of biological processes including steroid synthesis, xenobiotic degradation and natural product catabolism, to name a few [van Berkel, 2006]. Considering the range of transformations encompassing this set of flavoproteins, the enzymes have been subjected

and partitioned into a series of classes based on: the type of cofactor (FAD vs FMN), reducing agent (coenzyme), structural topology and the need for protein partners for activation and catalysis [Mascotti, 2016]. So far, eight classes, A-H, of flavin-dependent monooxygenases have been documented (Table 1.3.2.1) [Huijbers, 2014; Mascotti, 2016]. These classes can be divided into groups, 1-4, based on their fold: for example, the largest group, 1, are built with an overall Rossmann-fold topology. However, despite groups 2 and 4 possessing enzymes that display TIM-barrel folds, their overall domain organisations are quite distinct and are grouped separately [Mascotti, 2016]. Intriguingly, the table conveys that multiple enzymes from different classes perform similar conversions, thereby nicely illustrating a degree of convergent evolution among these systems.

The remainder of this section will be referring to Figure 1.3.2.1 and Table 1.3.1. Group A monooxygenases have a Rossmann-fold and are classified as group 1 monooxygenases [Mascotti, 2016]. They are one-component systems and follow catalysis according to the red trace shown in Figure 1.3.2.1. These enzymes require the substrate and reducing agent to initiate catalysis [Huijbers, 2014; Palfey, 2010]. Once the FAD ring has been reduced, the coenzyme exits the protein scaffold and the C4a-hydroperoxy intermediate is generated. This mode of catalysis resembles a ping-pong mechanism (Scheme 1.1.1). The most well-known and studied class A flavin-dependent monooxygenase is the para-Hydroxybenzoate hydroxylase (PHBH, EC 1.14.13.2) that inserts a hydroxyl group ortho to the para-hydroxy group.

Table 1.3.2.1: Classification of the Flavin-dependent monooxygenases

Group	Class	Enzymes	Cofactor	Reducing Agent	Organization	Protein Fold	E.C. and CATH Domain	Reactions
1	A	Aromatic hydroxylases	FAD	NAD(P)H	One-component	Rossmann fold (GR-2)	1.14.13, 3.50.50.60	Hydroxylation, Sulfoxidation
	B	Baeyer-Villiger monooxygenases (BVMOs), Flavin-containing monooxygenases (FMOs), N-hydroxylating monooxygenases (NMOs), YUCCAs	FAD	NADPH	One-component	Rossmann fold (FMO)	1.14.13, 3.50.50.60	Baeyer-Villiger oxidation, Heteroatom oxygenation, N-Hydroxylation, Oxidative decarboxylation
	G	Amino acid decarboxylases	FAD	Substrate	One-component	Rossmann fold (MAO)	1.13.12, 3.50.50.60	Oxidative decarboxylation
	E	Epoxidases	FAD	FADH ₂	Two-component	Rossmann fold (GR-2)	1.14.14, 3.50.50.60	Epoxidation
	F	Amino acid halogenases	FAD	FADH ₂	Two-component	Rossmann fold (GR-2)	1.14.14, 3.50.50.60	Halogenation
2	C	Luciferases, BVMOs type II, Hydroxylases	FMN	FMNH ₂	Two-component	TIM-barrel (luciferase)	1.14.13, 3.20.20.30	Light emission, Baeyer-Villiger oxidation, Epoxidation, Desulfurization, Sulfoxidation, Hydroxylation
3	D	Hydroxylases, Epoxidases	FAD/FMN	FADH ₂ /FMNH ₂	Two-component	Acyl-CoA dehydrogenase	1.14.14, 1.10.540.10 ^x ; 2.40.110.10 ^y	Hydroxylation, N-Hydroxylation
4	H	Decarboxylases, Denitrases	FMN	Substrate	Two-component	TIM-barrel (glycolate oxidase)	1.13.12, 3.20.20.70	Oxidative decarboxylation, Oxidative denitration

x: CATH domain of the group 3 hydroxylases; y: CATH domain of the group 3 epoxidases

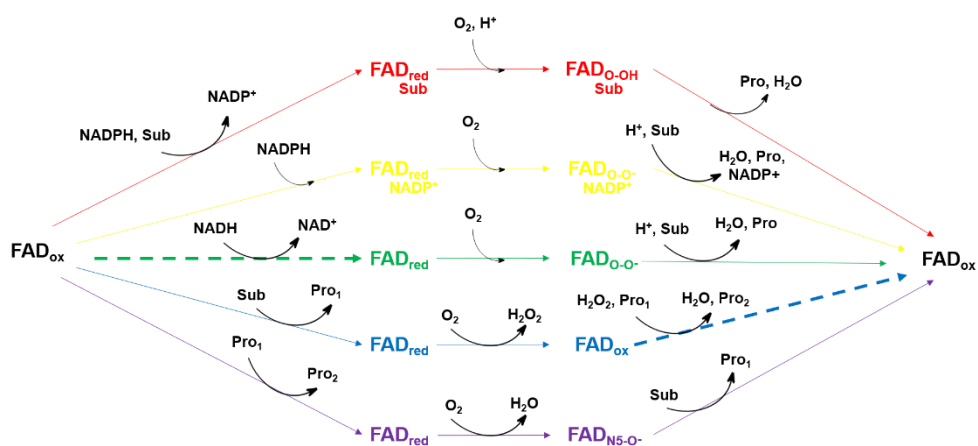


Figure 1.3.2.1: Reaction pathways carried out by the different flavin-dependent monooxygenases. Each reaction pathway is shown in a different colour. Pathways shown in red and yellow illustrate the one-component monooxygenase systems whereby all reaction steps are performed by a single enzyme. These include Classes A, B and G. The green pathway corresponds to the two-component monooxygenases whereby the reduced flavin species, FAD_{red} , is generated in a separate enzyme and depicted in dashed lines. FAD_{red} is then transferred into the monooxygenase for catalysis. These include classes C, D, E and H. Pathways in blue and violet represent internal monooxygenases type I and type II, respectively. In these scenarios, the substrate generates FAD_{red} and is then further modified. This includes classes G and H. Type I internal monooxygenases use the generated hydrogen peroxide to modify the initially product further outside the enzyme, depicted as a dashed line. Type II internal monooxygenases are generated as N5-oxide bearing FAD intermediates first and then react with the substrate to produce an intermediate product, Pro_1 , that then reduces the FAD_{ox} to FAD_{red} , releasing the final product, Pro_2 . Molecular oxygen regenerates the FAD_{N5-O-} to restart the catalysis. Note that in some instances the FAD is representing FMN (Table 1.3.1).

Alternatively, group B monooxygenases (see section 1.4) perform catalysis according to the yellow trace in Figure 1.3.2.1 [Huijbers, 2014]. Here, the enzyme can be activated in the absence of substrate and then releases the oxidized coenzyme in the final catalytic steps. Importantly, in contrast to class A monooxygenases, these systems require the presence of the coenzyme for substrate turnover and can accommodate both FAD and NADPH during catalysis [van Berkel, 2006]. Therefore, their mode of function depicts that of a sequential mechanism (Scheme 1.1.1). They are assigned under the group 1 classification of flavin-dependent monooxygenases owing to their Rossmann-fold and are one-component monooxygenases. These systems are discussed in more detail in the next section.

Class C monooxygenases are particularly unique because they emit light during catalysis [Huijbers, 2014; Tinikul, 2013]. Bacterial luciferases represent the prototype for this class

and are involved in oxidizing long-chain aldehydes into carboxylates and carboxylic acids. Additionally, unlike the other two classes mentioned above, this class requires an additional component in order to generate the reduced flavin – a flavin reductase [Mascotti, 2016; Tinikul, 2013]. Thus, these systems can be depicted by the green trace in Figure 1.3.2.1. Furthermore, they possess a unique TIM-barrel fold (CATH domain 3.20.20.30) and are separately classified as group 2 flavin-dependent monooxygenases.

Group D monooxygenases are also two-component systems and perform catalysis according to the green tree in Figure 1.3.2.1. They possess a unique Acyl-CoA dehydrogenase fold with various systems showing alternative domain organizations (Table 1.3.1). As such, they are classified independently as group 3 flavin-dependent monooxygenase [Huijbers, 2014; Mascotti, 2016]. The canonical Class D flavoprotein is the 4-Hydroxyphenylacetate 3-hydroxylase (HPAH EC 1.14.14.3), that similarly to Class A perform aromatic hydroxylations, but also N-hydroxylations alike the Class B monooxygenases [Huijbers, 2014; Sucharitakul, 2006; Chakraborty, 2010].

Group E monooxygenases are commonly referred to as epoxidases as they have only been conveyed to exhibit this one reaction type [Huijbers, 2014]. They fall into the group 1 class of flavin-dependent monooxygenases owing to their Rossmann-fold structural topology. However, unlike groups A and B, they require an additional flavin reductase component, making them two-component monooxygenases and follow the green trace in Figure 1.3.2.1. The most extensively characterized Class E system is the Styrene monooxygenase (SMO) (E.C 1.14.14 and CATH 3.50.50.60) which is split into the reductase and monooxygenase components, StyB and StyA, respectively [Huijbers, 2014; Kantz, 2005; Morrison, 2013].

Class F flavoprotein monooxygenases are classified as group 1 flavin-dependent monooxygenases due to their Rossmann-fold and are two-component monooxygenases that require an additional reductase for activity. They are commonly referred to as Amino acid halogenases [Huijbers, 2014; Mascotti, 2016; Yeh, 2007; Dong, 2005]. These enzymes play a pivotal role in synthesising antibiotics and various other natural organic products that bear chlorine or bromine atoms [Huijbers, 2014]. The classical class F monooxygenase is the Tryptophan 7-halogenase (EC 1.14.14.7) which adds a chlorine atom onto tryptophan at position 7 through electrophilic aromatic substitution [Huijbers, 2014; Mascotti, 2016; Yeh, 2007]. These systems also follow catalysis according to the green trace shown in Figure 1.3.2.1.

Classes G and H represent the final groups of flavin-dependent monooxygenases and are both internal monooxygenases. In other words, they use their substrate to reduce the FAD ring (Figure 1.3.2.1). Despite the disparity between this mode of action and the other monooxygenase discussed herein, Class G monooxygenases possess a Rossmann-fold and are classified within group 1 monooxygenases [Huijbers, 2014; Mascotti, 2016]. These

proteins oxygenate amino acids via decarboxylation reactions with the prototypical enzyme for this class being the Tryptophan-2-monooxygenase (EC 1.13.12, CATH 3.50.50.60). Class G FMOs intriguingly perform catalysis as a mixture of type I internal monooxygenases (blue trace, Figure 1.3.2.1) and one-component monooxygenases (red and yellow traces, Figure 1.3.2.1) [Romero, 2018]. After using the substrate as a reducing agent, it generates a C4a-(hydro)peroxy intermediate to promote decarboxylation [Huijbers, 2014].

Alternatively, group H has a unique TIM-barrel fold and are assigned as group 4 monooxygenases. This class of systems are more commonly referred to as oxidative decarboxylases and oxidative denitrases [Huijbers, 2014]. Their mode of action, moreover, is still under scrutiny [Smitherman, 2013; Mijatovic, 2008]. Gadda and co-workers demonstrated that the reductive half-reaction resulted in the formation of a reduced semiquinone flavin intermediate and a one-electron oxidised substrate [Gadda, 2010]. Molecular oxygen is then reduced by either: the substrate, forming a peroxy-based intermediate; or the anionic semiquinone, forming a superoxide intermediate and the oxidized flavin [Smitherman, 2013; Mijatovic, 2008]. However, in contrast to class G monooxygenases, these systems display mixed modes of actions with type I internal monooxygenases and two-component systems (green trace, Figure 1.3.2.1). Nitrogenate monooxygenase (NMO, EC 1.13.12) represents a well-known member of this class catalyses the conversion of the metabolic poison, propionate 3-nitronate, into 3-oxopropionate and nitrite.

The final violet trace, shown in Figure 1.3.2.1 corresponds to the newly found internal monooxygenases, EncM [Teufel, 2013]. The initial generation of the N5-oxide for catalysis is still under investigation, but as it requires the substrate alone for FAD reduction classifies it under the internal monooxygenase nomenclature. Furthermore, its unique catalytic mode assigns this system as a type II internal monooxygenase. EncM shows an exclusive structural topology that closely resembles vanillyl-alcohol oxidases which are more commonly associated to flavoprotein dehydrogenases and oxidases [Romero, 2018; Huijbers, 2014; Teufel, 2013; Mascotti, 2016]. Based on its structural fold, EncM may represent a new group of monooxygenases. Moreover, the recently discovered N5-(hydro)peroxy harnessing systems discovered by Matthews et al follow the traces in red and yellow shown in Figure 1.3.2.1 [Matthews, 2020]. They are all one component systems that require reducing agents for catalysis and use very similar oxidative intermediates for catalysis [Matthews, 2020]. All in all, despite decades of extensive biochemical and structural analysis on this family of flavin-dependent systems, new reactions continue to emerge and epitomize the versatility of this cofactor.

1.4. CLASS B FLAVIN-DEPENDENT MONOOXYGENASES

In the previous sub-section, we divulged into the range of activities governed by the flavin-dependent monooxygenase family. Herein this section, we will expand and explore more closely the Class B flavin-dependent monooxygenases that include the Baeyer-Villiger- and Nitrogen-hydroxylating monooxygenases, and YUCCAs. As this thesis is predominantly based on Flavin-containing monooxygenases (FMOs), these systems will be addressed in the next section (see section 1.5).

1.4.1 Shared structural and enzymatic features

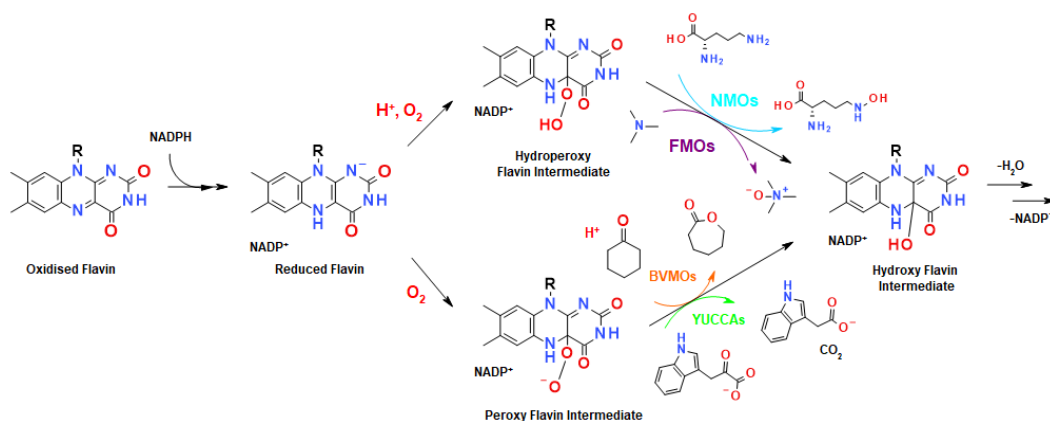
The key structural feature for this class that distinguishes these one-component systems from its counterparts is its paired Rossmann-fold. These enzymes have evolved to incorporate two adenine nucleotide binding domains that collectively accommodate both flavin adenine dinucleotide, FAD, and reduced nicotinamide adenine dinucleotide phosphate, NADPH [Mascotti, 2016]. Additionally, each of these domains have well conserved sequence motifs: the FAD and NADPH binding domains both possess the motif, (GXGXXG), whilst FMOs also consist of an additional sequence, (FXGXXXHXXXY/F), which represents the linker region connecting the two domains together [Thodberg, 2020; Fraaije, 2002].

Whilst FAD resembles the key cofactor, the coenzyme NADPH plays a crucial role in activating the FAD for oxidative catalysis. As previously mentioned, the FAD must be first reduced by NADPH to generate a molecule with the reductive potential to oxidize a molecule of dioxygen. As the reduced coenzyme docks into the active site, a charge-transfer complex rapidly forms between NADPH and oxidized FAD followed by the *proR*-hydride of NADPH being donated to the N5 position of the FAD creating a reduced FAD-oxidized NADP⁺ system [Palfey, 2010].

For many flavin monooxygenases, once the coenzyme's reducing power has been transferred, the molecule dissociates from the enzyme. However, for this class the molecule has been implicated to exhibit 'moonlighting roles', stabilizing key catalytic intermediates through its ribose-based 2'-OH moiety [Alfieri, 2008]. This idea rationalises why the coenzyme would remain bound throughout the catalytic cycle. Unlike class A monooxygenases that require the presence of substrate for catalysis, [Palfey, 2010] Class B monooxygenases do not require substrate for NADPH oxidation. On the one hand, this feature could be disadvantageous for the host as NADPH is metabolized without being effectively utilized to turnover a substrate. Furthermore, the resultant uncoupling reaction produces hydrogen peroxide that could be damaging. Alternatively, the enzyme harnesses this feature in order to always be primed for catalysis. The coenzyme then remains bound to stabilize the flavin intermediates and limit the rate of their decay and hydrogen peroxide production. There has been speculation however that certain class B systems, such as the

FMOs, could function as NADPH oxidases as well, producing reactive oxygen species for purposes such as cell signalling [Fiorentini, 2016]. The overall mode of catalysis for these systems is shown below (Scheme 1.4.1.1).

Scheme 1.4.1.1: Mode of catalysis for Class B Flavin-dependent monooxygenases – NMOs, FMOs, BVMOs and YUCCAs. For simplicity, one substrate is shown for each enzyme. Ornithine, trimethylamine, cyclohexanone and indole-3-pyruvate are shown for NMOs, FMOs, BVMOs and YUCCAs, respectively.



1.4.2 Baeyer-Villiger Monooxygenases – Structure and Function

Baeyer-Villiger oxidation consists of the insertion of an oxygen atom adjacent to a carbonyl functional group [Huijbers, 2014; van Berkel, 2006; Torres, 2010; Tolmie, 2019]. Therefore, carboxylic acids and esters can be generated from aldehydes and ketones. Since its discovery in 1899 by Adolf Baeyer and Victor Villiger, the reaction has been extensively applied in the chemical industry for organic synthesis [van Berkel, 2006; Torres, 2010]. However, a significant bottleneck for organic chemists is the lack of selectivity observed during synthesis. Intriguingly, in 1976, Donoghue and co-workers, [Donoghue, 1976] isolated and characterized the first protein shown to perform Baeyer-Villiger monooxygenation from *Acinetobacter sp. NCIB 9871*. This prototype was named, Cyclohexanone monooxygenase (CHMO) due to its role in cyclohexanone decomposition in bacteria. Moreover, CHMO was demonstrated to turnover several different substrates, indicating that it possesses a broad substrate scope, and that the reactions consisted of high chemo-, regio-, and enantio-selectivities [Torres, 2010; Tolmie, 2019]. The improvement and sophistication of gene mining and gene synthesizing techniques has resulted in the discovery and characterization of many more BVMOs including Phenylacetone monooxygenase (PAMO), steroid

monooxygenase (STMO) and ethionamide monooxygenase, to name a few [Malito, 2004; Martinoli, 2013; Franceschini, 2012; Fraaije, 2004].

Despite a lot of focus on these systems has been regarding their biocatalytic application, understanding their roles within hosts has also been of growing interest [Torres, 2010]. To date, the myriad of complex natural products produced by these systems have been documented to exhibit anti-bacterial, anti-cancer and anti-proliferative activity, making them of great interest in clinical applications [Tolmie, 2019]. Additionally, they have been shown to produce microbial toxins as well. Whilst all these functions are related to the secondary metabolism, certain BVMOs have been demonstrated to metabolize alternative carbon sources, suggesting that they have key primary metabolic roles too [Tolmie, 2019].

Until recently, BVMOs were thought to only be found in bacteria and fungi, implying that during evolution, metazoan and plants lost this enzyme. However, in 2013, Bergantino and co-workers [Beneventi, 2013] discovered a BVMO in moss (*Physcomitrella patens*), suggesting that perhaps BVMOs were lost during the development of the Metazoa. Whilst the specific role of this BVMO was not confirmed, it was postulated that it could be implemented in specialized metabolism or modifications of photosynthetic pigments [Thodberg, 2020; Beneventi, 2013]. As more sequencing information becomes available for plants, perhaps the essential roles governed by these systems will be elucidated. Moreover, Mascotti et al in 2015 carried out an extensive phylogenetic analysis of the BVMO family [Mascotti, 2015]. Their work illustrates that there are many BVMOs located within metazoans and as such, paves the way for new research into their functionality. It should be noted that recent work by Fiorentini et al illustrated that FMO5 is able to partake in Baeyer-Villiger monooxygenation [Fiorentini, 2016; Fiorentini, 2017]. Thereby demonstrating that metazoan species have accessed Baeyer-Villiger monooxygenation through convergent evolution.

With CHMO being the first extensively characterized BVMO, in terms of substrate turnover and portfolio, we will use this enzyme model to depict the key structural features that establish activity for BVMOs [Yachnin, 2014]. The main components of all Class B monooxygenases however, can essentially be split into three domains: FAD and NADPH binding domains, followed by a domain that helps build the pathway and tunnel architecture for substrate transit which we will refer to as, the substrate binding domain (Figure 1.4.2.1). Each of these three domains fixate on the active site where all domains meet. The coenzyme NADPH sits at the base of the NADPH binding domain, whereas the FAD cofactor is slightly more integrated into the protein scaffold. This finding is justified considering that NADPH and/or NADP⁺, must be able to diffuse in and out of the structure in order to restart catalysis, whilst the FAD does not require any additional proteins for activation and can remain bound in the enzyme. The orientation of CHMO as depicted in Figure 1.4.2.1 highlights the trajectory the substrates would voyage towards the

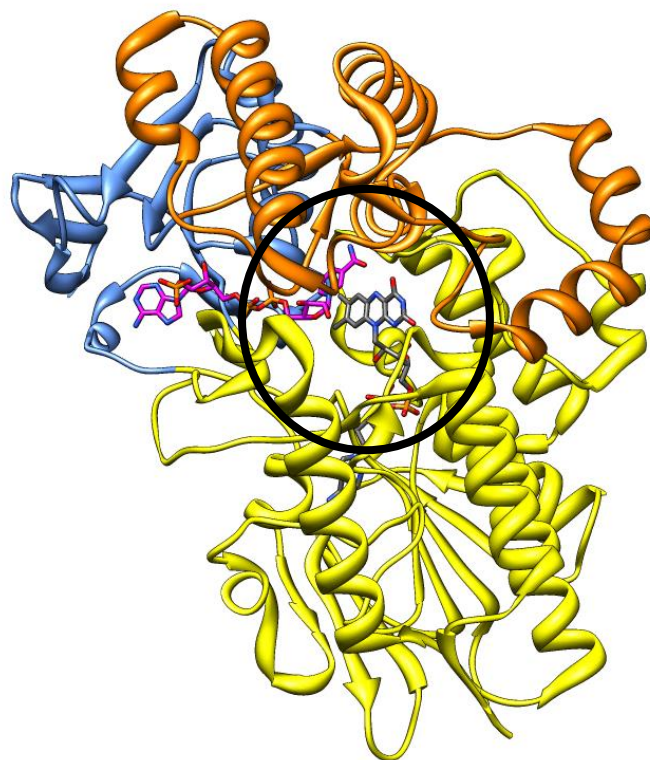


Figure 1.4.2.1: Crystal structure of a cyclohexanone monooxygenase. The protein has been split into three sections that represent FAD, NADPH and substrate binding domains that are coloured in yellow, cornflower blue and orange, respectively. The black ring highlights the entry point for substrates. Molecules of FAD and NADP⁺ are shown in grey and dark pink, respectively. PDB ID: 4RG3 [Yachnin, 2014].

active site (black ring) [Yachnin, 2014]. In this scenario molecules pass through a straight hydrophobic tunnel towards the active site cavity at an angle that is almost planar to the face of the isoalloxazine ring [Romero, 2016]. Pioneering work by Fürst et al., investigated the impact of mutating residues along the substrate binding tunnel into alanines, in order to assess the selectivity of CHMO and to determine which residues were critical for catalysis [Fürst, 2018]. To their surprise, mutating these residues only resulted in a decreased activity as opposed to a switch in substrate selectivity. This finding suggested that the enzyme does not have a selective substrate binding mechanism and instead, the enzymes outer shell is constructed in order to stabilize highly reactive intermediates.

The active site of CHMO is very similar to other BVMOs and consists of key elements for catalysis. Importantly, all three domains mentioned above contribute residues for the active

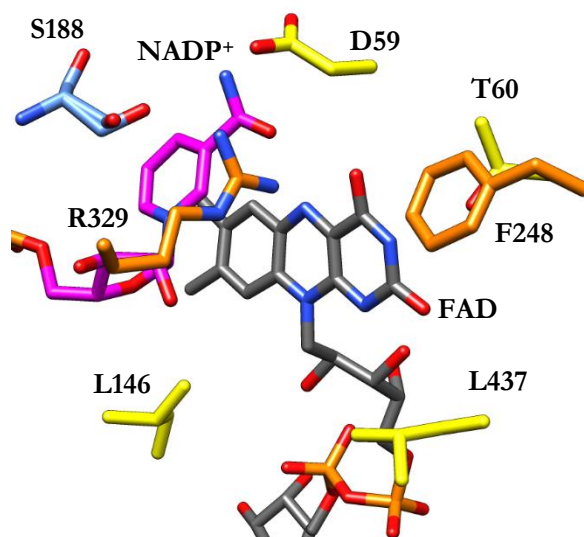


Figure 1.4.2.2: Active site of the CHMO from *Rhodococcus sp. HI-31*. Residues are labelled and coloured with respect to the domain they are found in. FAD, NADPH and substrate binding domains are coloured in yellow, cornflower blue and orange, respectively. The black ring highlights the entry point for substrates. Molecules of FAD and NADP⁺ are shown in grey and dark pink, respectively. PDB ID: 4RG3 [Yachnin, 2014].

site (Figure 1.4.2.2). At the base of the active site are two leucines (146 and 437) that create a hydrophobic pocket for the substrate to bind. In addition, a Phe248 represents a hydrophobic steric blockade that forms part of the tunnel that directs substrates. Asp59 and Arg329 however fulfil purposes attributed to the enzymatic intermediates during catalysis. Asp59 has been demonstrated to be essential for stabilizing the C4a-peroxy intermediate that performs nucleophilic attack on the electrophilic carbonyl centre of the substrate, most likely through hydrogen bond donation. Arg329 however, is postulated to form key intermolecular interactions with the negatively charged, covalent tetrahedral Criegee intermediate formed after nucleophilic addition. Arg329 is observed conveying multiple hydrogen bonds with D59 and the amide moiety of the coenzyme. Romero and co-workers, [Romero, 2016] imply that this residue will occupy a new conformational state in the presence of the reduced, and thus negatively charged, flavin species, placing its positively charged guanidinium group close to isoalloxazine ring. Once the flavin is oxygenated it would form a new conformation that opens up space for the substrate and peroxy moiety of the intermediate [Yachnin, 2014; Romero, 2016; Fürst, 2018].

1.4.3 Nitrogen-hydroxylating monooxygenases – Structure and function

Nitrogen-hydroxylating monooxygenases, or NMOs, are well known for their involvement in siderophore production in bacteria and fungi [Mügge, 2020]. More specifically, they oxidize soft nucleophilic Nitrogen atoms, usually primary amino groups, by directly inserting hydroxyl groups to create hydroxyamines [Huijbers, 2014; van Berkel, 2006]. These can then be further modified and fused together to form hydroxamate-based siderophores that can chelate iron centres. Such compounds include lysine, ornithine, arginine and aliphatic diamines including 1,3-diaminopropane and cadaverine [Mügge, 2020].

Whilst NMOs have been generally associated to metabolize and oxidize small organic molecules, there are some catalysts that can convert much more complex compounds. In 2017, Zhang and Liu were able to demonstrate that an NMO found in *Streptomyces* was able to oxidize xiamycin A and was involved in indolosesquiterpene biosynthesis [Zhang, 2017]. Furthermore, many NMOs contribute towards the biosynthesis of single bonded N-N bonds and also double bonded N-N bonds. For example, the NMO, KtzI, is involved in piperazic acid biosynthesis, a structural component found in molecules from antibiotics to immunosuppressants [Mügge, 2020; Neumann, 2012; Du, 2017]. Finally, these systems have also been shown to be involved in the production of nitrous acid and nitro functional groups. After the initial N-hydroxylation step is completed, the newly formed amino-hydroxy group can act as a substrate again, further oxidizing the nitrogen centre and creating a nitro group. In the presence of a lyase, this new group can be cleaved off the molecule's backbone and released in its protonated form as nitrous acid. [Mügge, 2020; Huang, 2016; Waldman, 2015]

As previously observed with CHMO, NMOs consist of essentially the same three domains: NADPH and FAD binding domains, and then a substrate binding domain. In Figure 1.4.3.1, the protein is orientated with the view corresponding to the route of entry pursued by the coenzyme. Differently from BVMOs that tend to have buried active sites and a long breaching tunnel, the more soluble nature of the NMO's substrates results in a small pocket forming next to the active site (Figure 1.4.3.2) [Francheschini, 2012]. Here with ornithine as the substrate, the amino terminus can be seen orientated towards the FAD molecule, whilst the base is anchored in to place and exposed to the solvent. Importantly however,

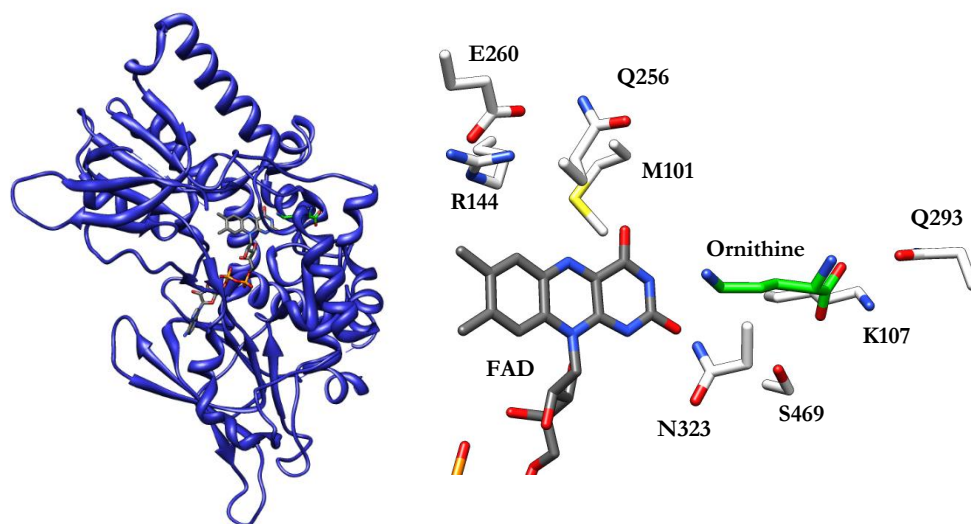


Figure 1.4.3.1: Three-dimensional structure of an NMO and its active site. Left, overall three-dimensional structure of the NMO, SidA from *Aspergillus fumigatus* (PDB: 4B63) [Franceschini, 2012]. Right, active site of SidA complexed with the substrate, ornithine, shown in green. Active site residues and FAD are shown in white and dark grey, respectively.

the FAD molecule is still well protected from the exterior environment, therefore shielding and preserving the reactive flavin intermediates formed during catalysis.

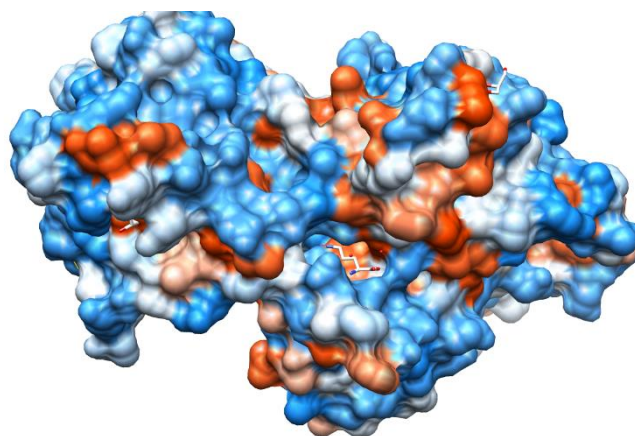


Figure 1.4.3.2: Protein surface of SidA from *Aspergillus fumigatus*. Surface of SidA is depicted using a hydrophobicity scale whereby dark orange is very hydrophobic and dark blue is very hydrophilic. Ornithine is shown in white, PDB: 4B63 [Franceschini, 2012].

During catalysis the C4a-hydroperoxy flavin intermediate is used instead of the negatively charged peroxy. Interestingly, inspection of the coenzyme through solvent kinetic isotope effects, density functional theory and structural analysis deemed that the 2'-OH group of the NADPH ribose would protonate the C4a-peroxy intermediate in order to generate the protonated equivalent utilized for oxygenation [Mügge, 2020; Robinson, 2013]. Moreover, the need for the substrate to produce the oxygenated flavin intermediate during the first step of the oxidative half reaction conveys mixed results. For PvdA, the presence of substrate is required for the generation of the C4a-hydroperoxy intermediate, whilst for SidA, the intermediate will form regardless of substrate [Mügge, 2020; Frederick, 2011; Meneely, 2009]. This suggests that subtle differences in the active site fine tune the protonation event of the peroxy flavin intermediate.

As shown in Figure 1.4.3.1, there are a series of residues that work in concert to orientate and position the substrate in the correct place for catalysis. Whilst K107, Q293 and S469 fixate the backbone of the ornithine molecule (amino and carboxy functional groups), N323 helps to position the amino terminus towards the C4a site. Residues E260 and R144 form a salt bridge above the isoalloxazine ring and are in proximity to the nicotinamide head group of the coenzyme [Franceschini, 2012]. It could be speculated that they have a role in positioning the coenzyme for flavin reduction, however, Q256 will most likely act as a 'back door' residue, sandwiching the face of the nicotinamide ring between itself and the FAD. Interestingly, Olucha and co-workers state that the nicotinamide ring changes its conformation after fulfilling its reductive role, thereby presenting space for the stabilization and generation of the C4a-hydroperoxy flavin-intermediate [Mügge, 2020, Olucha, 2011a; Olucha, 2011b]. Q102 resides above the C4a position and is speculated to be critical for stabilizing the C4a-hydroperoxy intermediate.

1.4.4 YUCCAs – Function

YUCCAs represent the most recent addition to the Class B monooxygenases and are prevalent in plants. BVMOs, NMOs and FMOs are all mainly devoted towards secondary metabolism and are involved in natural product synthesis, toxin production and xenobiotic degradation, to name a few [Huijbers, 2014; van Berkel, 2006; Torres, 2010; Tolmie, 2019; Fiorentini, 2016; Mügge, 2020]. On the contrary, YUCCAs are key enzymes attributed to the production of the growth hormone auxin, regulating embryogenesis as well as the development of several organs, all of which are essential components for plant survival [Thodberg, 2020; Cheng, 2006; Cheng, 2007]. Interestingly, alternative protein roles were demonstrated for these systems that are distinct from auxin biosynthesis. YUC6 from *Arabidopsis* conveyed novel NADPH-dependent thiol-reductase activity, enhanced peroxidase activity and a heightened scavenging ability for removing reactive oxygen species [Cha, 2015].

In addition to their novelty from the other Class B systems, YUCCAs perform decarboxylation reactions. Extensive biochemical characterization of YUC6 highlighted that the enzyme uses FAD, NADPH and in-line with BVMOs, a C4a-peroxy flavin intermediate for catalysis [Thodberg, 2020]. Unlike the other class B monooxygenases that portray intermediates that have half-lives of roughly 30 mins, this system conveyed a mere 20 seconds [Thodberg, 2020; Mayfield, 2010; Chocklett, 2010]. Thus, we can speculate that there could be considerable hydrogen peroxide production present in the cell if the FAD molecule is prematurely reduced. On the contrary, considering the importance of the precursor molecule, Indole-3-pyruvate, for auxin production, it may well be in plentiful excess within the cell. During catalysis, the generated flavin C4a-peroxy intermediate is predicted to perform a nucleophilic attack at the carbonyl adjacent to the terminal carboxyl group, forming a highly reactive dually negatively charged tetrahedral Criegee intermediate [Thodberg, 2010; Dai, 2013]. This intermediate would then collapse with the entropically favourable release of carbon dioxide as electrons are delocalized onto the newly formed C-O bond and the proximal oxygen atom of the peroxy moiety. Lack of current structural information impedes the residual interrogation of the active site that would rationalise the formation of such a charged intermediate. Furthermore, the absence of structural data hinders our understanding of the breadth of substrates that these systems could capitulate. So far, phenyl pyruvate and tryptamine are additional substrates found to be catalysed by YUCCAs [Thodberg, 2010; Dai, 2013; Tivendale, 2014].

1.5 FLAVIN-CONTAINING MONOOXYGENASES – FMOS

1.5.1 Overview – mode of action and substrate preference

Flavin-containing monooxygenases (FMOs) are ubiquitous and thrive in all domains of life [Krueger, 2005; Sehmeyer, 2010; Hao, 2009]. This fact alone emphasizes the significance of these systems for the survival of organisms. Extensive research over the last century has brought about a remarkable uproar in our understanding of flavin containing systems: In 1932 the first FAD containing enzyme, termed the ‘Old Yellow enzyme’ was discovered and then 50 years later, Daniel M. Ziegler and co-workers ushered in the key distinction and resultant classification that distinguished NADPH-cytochrome c oxidases from mixed amine oxidases [Masters, 1971; Ziegler, 1978; Sofer, 1978]. The latter of the two would then become denoted as flavin-containing monooxygenases and pioneering kinetic and substrate analysis by Ziegler, Poulsen, Beaty and Ballou all played integral roles, laying down the foundation as to how these enzymes function in the late 20th century [Beaty, 1981a; Beaty, 1981b; Ziegler, 1990; Poulsen, 1979].

The original name for these class B monooxygenases, mixed amine oxidases, is owed to their broad substrate profiles [Jakoby, 1990; Ziegler, 2002]. To this extent, work was carried out to document the range of substrates that could be accommodated by these proteins. The discovery of several mammalian FMO paralogs substantially increased the breadth of

their collective substrate scope and their individual specificity for certain compounds drew great attention (see sub-section 1.5.6) [Hao, 2009; Ziegler, 1990; Jakoby, 1990; Cashman, 2006]. More specifically, they oxidize compounds that possess soft nucleophilic centres, such as nitrogen and sulfur atoms, that are found in a vast quantity of biological molecules [Ziegler, 1990].

One can imagine that employing these systems as ‘industrial cleaning operatives’ to remove and filter out unwanted compounds is an advantageous feat. With constant biological warfare taking place among a plethora of organisms, the exposure to toxic compounds is a considerable threat and commensurately, all domains of life have recruited this workforce for defence. Throughout this section, we will divulge into the key traits and features that define this subclass of systems, and more so, highlight the principal questions that were posed for these proteins.

1.5.2 Prevalence and Biodiversity – Bacteria and fungi

Before focussing on the mammalian FMOs that represent the principal systems analysed within this thesis, we will first discuss the roles and features of phylogenetically divergent systems from fungi and bacteria. Bacterial FMOs were recently discovered with the first being found from *Methylophaga* in 2003 [Choi, 2003]. Despite their infancy, this enzyme later became the first depicted crystal structure of an FMO, a milestone for FMO enzymology [Alfieri, 2008]. Alongside the human FMOs this system also portrayed broad substrate profiles with activity towards molecules both bearing nitrogen and sulfur nucleophilic centres. This finding implicated that bacterial FMOs were also employed as xenobiotic degraders.

The structure of this bacterial FMO (Figure 1.5.2.1) became the first potential model that could be exploited to rationalise the structural-biochemical features of human FMOs. However, with a sequence identity of only 30%, many predictions and comparisons were met with caution. Despite this, many residues provided key insights into the functionality behind mammalian FMOs. One of the key findings was the localization of Asn78 at the core of the active site. This residue is extensively conserved among all FMOs and commensurately, it was conveyed to be paramount for C4a-hydroperoxide stability [Alfieri, 2008; Cashman, 2006]. Particularly, mutating this residue in human FMO3 results in the disease, trimethylaminuria, whereby patients are unable to metabolize trimethylamine, a product from choline degradation originating in fish [Krueger, 2005; Cashman, 2006; Phillips, 2019; Zhang, 2003]. The accumulation of trimethylamine results in a pungent

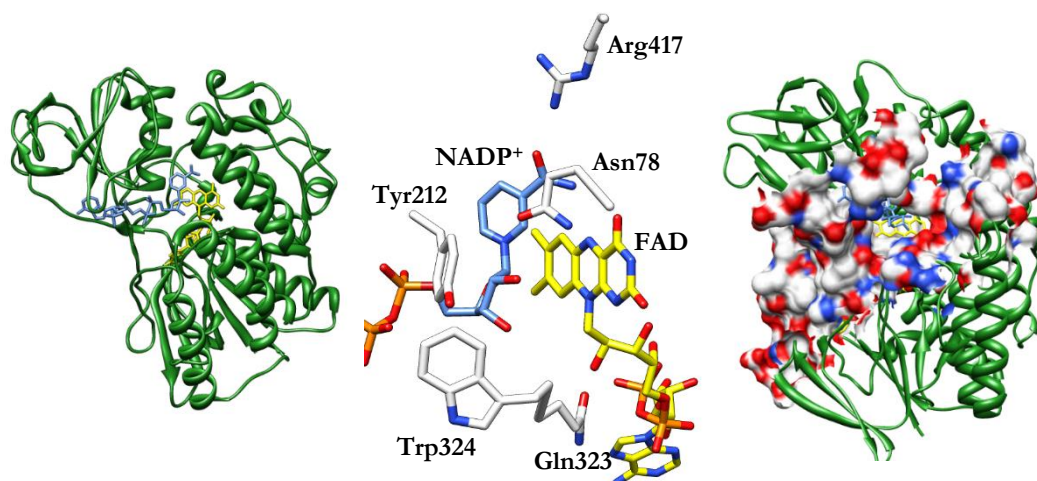


Figure 1.5.2.1: Structural features of a bacterial FMO. Left, overall structural topology of a bacterial FMO from *Methylophaga sp. strain SK1*. Middle, active site of the FMO. Residues are shown in white and labelled accordingly. Right, entrance site for substrates of the FMO. FAD and NADP⁺ molecules are shown in yellow and cornflower blue, respectively. Tertiary structure is shown in dark green and the surface of protein is depicted as charge, with blue, white and red representing positive, neutral and negative charges, respectively. PDB ID: 2vq7 [Alfieri, 2008].

odour that when excreted out in sweat creates a rotten fish-like smell. Therefore, the elucidation of these structural features validated the occurrence of the disease after mutating this key residue.

Similarly, the aspartate residue that occupies the same position in CHMO is postulated to exhibit a hydrogen bond with the C4a-peroxy intermediate for stability [Yachnin, 2014; Romero, 2016]. The overhanging Arg417 dangles down towards the active site and forms key hydrogen bond interactions with the nicotinamide moiety of the coenzyme [Alfieri, 2008]. The residue likely plays a role in anchoring and positioning the molecule. The tyrosine moreover has been denoted as a ‘backdoor’ residue, sandwiching the coenzyme into place against the isoalloxazine ring [Alfieri, 2008]. The hydroxyl group of Tyr212 likely donates/accepts hydrogen bonds with oncoming substrates and may assist in binding within the cavity.

A noticeable feature regarding the bacterial FMO is the active sites accessibility. As shown in Figure 1.5.2.1 (right), the core of the enzyme can be clearly seen from outside the protein scaffold. Curiously, the enzyme adopts a wide entrance site that narrows down at the interface between the three domains, somewhat like a funnel [Alfieri, 2008]. Attractively,

the enzyme has manufactured a broad substrate passageway that can then permit the transit of many xenobiotics that vary in size. This is a common feature observed for many xenobiotic degrading enzymes such as P450s that are ostensibly built to have a large substrate capacity [Naumann, 2002; Petriello, 2017].

Yeast FMOs also share similar characteristics (Figure 1.5.2.2) [Bonanno, 2017]. Additionally, they consist of an asparagine residue sitting above the isoalloxazine ring in the active site that provides stability for the C4a-hydroperoxy intermediate. Furthermore, the active site cavity is similar sized to the bacterial FMO and with an entrance point consisting of parallel dimensions for substrate transit. In 1999, Suh and co-workers showed that a yeast FMO was surprisingly unable to metabolize Nitrogen-based soft nucleophilic centres [Suh, 1999]. Instead, it could only perform sulfur oxidations. This difference represents a key finding regarding the function of the yeast FMO [Suh, 1999].

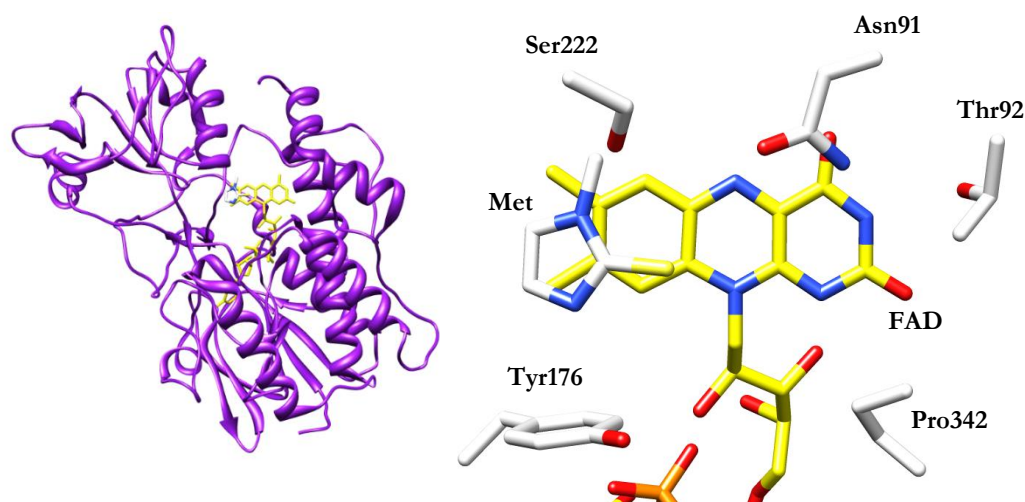


Figure 1.5.2.2: Structural features of a yeast FMO. Left, overall structural topology of a yeast FMO from *S. pombe*. Right, active site of the yeast FMO in the presence of methimazole (Met). Residues are shown in white and labelled accordingly. PDB ID: 1vqw [Bonanno, 2017].

Additionally, Suh and co-workers illustrated that yeast FMOs play important roles in balancing the overall redox potential of endoplasmic reticulum [Suh, 1999]. They showed that yeast FMOs had a significant impact on the amount of glutathione disulphide formed from oxidized glutathione. The yeast FMO ensures that the oxidizing environment of the endoplasmic reticulum remains high to promote the formation of disulphide bonds that are imperative for certain proteins. Therefore, it can be speculated that yeast FMO has an

integral role that directly impacts the function of many other proteins. Furthermore, yeast cells were shown to increase the expression of this gene when the cells were under reductive stress [Suh, 1999]. It should be noted that the enzyme is also able to oxidize other sulfur containing compounds including cysteamine and cysteine, but the purpose of this has not yet been resolved.

1.5.3 Human FMOs: localization, distribution, polymorphisms and paralogs

N- and S- enzymatic oxidations have been well documented and was initially thought to be principally carried out by cytochrome P450s. Research led by Ziegler and Poulsen resulted in the characterization of a hepatic microsomal dimethylaniline N-oxidizing system that was then later classified as a FMO [Masters, 1971; Ziegler, 1978; Sofer, 1978]. The specific and disparate substrate profile between the FMOs and cytochrome P450s spurred on a great deal of interest regarding the FMOs enzymology.

Much of the pioneering work related to gene localization and paralog classification was carried out by the researchers Cashman, Shephard, Phillips and Zhang. Collectively, they documented that there are five human FMO paralogs located on chromosome 1, with the FMO1–FMO4 genes clustering over 220 kilobases (kb) whilst FMO5 is found on a separate chromosome region [Cashman, 2006; Cashman, 1995; Ziegler, 1972; Cashman, 2000]. These five genes represent the functional FMOs that portray high mRNA levels, implying that they are transcribed and likely expressed. Intriguingly however, there is an additional FMO denoted as FMO6 that is a pseudo gene in humans and located in the same region as the FMO1-FMO4 cluster [Cashman, 2006; Hines, 2002]. Whilst not being expressed in humans, in mice, FMO6 is well expressed, insinuating that it may play a significant and specific role in mice or mammals in general? [Cashman, 2006]. Additionally, Shephard and Phillips discovered five more pseudo genes (FMO7-11) that were located approximately 4 Mb telomeric from the original cluster [Hernandez, 2014].

The abundance of these FMOs corroborates their xenobiotic degrading role, maximising the potential substrate scope that this family can tackle. Alternatively, despite having large substrate profiles, each FMO may in fact have been selectively enhanced to detoxify a particular toxin. For example, FMO3 that is well expressed in the liver, is exclusively involved in oxidizing trimethylamine and when inactive results in the disease, trimethylaminuria [Krueger, 2005; Cashman, 2006; Phillips, 2019; Zhang, 2003].

To support the xenobiotic detoxifying notion, the FMOs are illustrated to be expressed throughout the body in many different tissues [Cashman, 2006; Koukouritaki, 2002; Janmohamed, 2004]. Collectively, they act as ‘security’ for the host and metabolize foreign compounds that may reach different regions. FMO1 is expressed predominantly in the kidney and is the most prevalent isoform in the organ. Curiously, when comparing foetal

organs with their adult equivalents, FMO1 is the only isoform observed to show decreased expression levels in adult organs [Cashman, 2006; Janmohamed, 2004]. For example, the expression level of FMO1 in the foetal brain and liver, decrease approximately 20 and 10-fold, respectively, when compared to the adult brain and liver. This finding conveys an important developmental role for this FMO that perhaps is built to metabolize a common xenobiotic found during growth.

FMO2 is found principally in the lung. This paralog is the most expressed member of the family in the human body, with mRNA levels recorded to be more than five times higher than FMO3 or FMO5 in the liver [Cashman, 2006; Cashman, 1995; Cashman, 2000]. Furthermore, it is well expressed in the kidney (with mRNA levels 1.5-fold lower than FMO1) and is the most expressed FMO in the brain. Despite this, most humans possess a truncated and inactive form of FMO2 [Cashman, 2006; Dolphin, 1998; Whetstine, 2000]. Most Caucasian and Asian populations are homozygous for a Q472X mutation that results in an 80-residue truncation that nominally does not bind FAD [Cashman, 2006; Dolphin, 1998; Whetstine, 2000]. However, there are some populations that do express the full lengthened equivalent that has been shown in monkeys and rabbits to metabolize a range of sulfur and nitrogen containing compounds, including nicotine [Krueger, 2005; Cashman, 2006; Williams, 1990]. Certain groups have shown that FMO2 is able to produce toxic sulfenic acids from thiourea compounds [Krueger, 2005; Henderson, 2004]. This may have had negative repercussions for humans, and it is speculated that the inactive isoform may have been naturally selected.

FMO3 is ostensibly the most studied human FMO due to its overexpression in the liver and its direct relation to trimethylaminuria. FMO3 is the third highest expressed FMO in the body and the second highest expressed isoform in the liver [Krueger, 2005; Cashman, 2006]. The enzyme is well expressed in the lung and with most humans not expressing the full length FMO2, makes it (alongside FMO5) one of the most expressed paralog in the lung.

FMO4 is particularly interesting for several reasons. Firstly, whilst all the other FMOs are approximately 530 residues in length, this isoform possesses a 20-residue extension at the C-terminus that has been suggested to be essential for membrane insertion [Krueger, 2005; Cashman, 2006]. Whether this insertion provides any key advantages or features regarding substrate extraction or portfolio is yet to be clarified. The reason for this so far is underlined by the lack of experimental data and unsuccessful expression; despite the relative quantities of detected mRNA for FMO4, no expression has been detected thus far. There is a chance that the insertion results in poor membrane adherence or the enzyme requires protein partners for correct folding. Nevertheless, mRNA levels have been observed at relatively

high quantities in many different tissues including the liver (3rd highest), kidney (3rd highest), lung (4th highest) and small intestine (4th highest) [Cashman, 2006].

FMO5 for many years remained enigmatic in the FMO family due to its unusual substrate scope. This paralog does not metabolize typical nitrogen and sulfur containing compounds like the other FMOs [Krueger, 2005]. Moreover, pioneering work carried out by Fiorentini and co-workers were able to illustrate that this FMO possessed novel BVMO activity and could in fact metabolise a plethora of different aldehydes, carboxylic acids and ketones [Fiorentini, 2016; Fiorentini, 2017]. To date, this isoform is the only known human enzyme able to perform Baeyer-Villiger monooxygenation. Its unique activity substantiates the abundance and ubiquity of this paralog throughout the body. FMO5 is the most expressed paralog in the liver and the small intestine, moreover, with FMO2 being inactive in many populations, this isoform would also be the most expressed member in the brain and the lung [Cashman, 2006]. Importantly, all of the FMOs seem to be well or slightly expressed in each of the detoxification ‘centres’ of the body. Albeit low for some paralogs, the presence of each maximizes the ability to remove and excrete a multitude of compounds that the organ may be exposed to.

There are many single nucleotide polymorphisms (SNPs) observed for each FMO [Krueger, 2005; Cashman, 2006; Cashman, 2003]. Certain mutations have been found to have a large impact on activity. For example, some mutations have been observed for FMO3 that are related to trimethylaminuria [Krueger, 2005; Cashman, 2006; Cashman, 1997; Hernandez, 2003]. However, many other mutations observed did not seem to have a great impact on overall activity, despite showing some slightly different conversion rates. Similarly, this finding was also observed for FMO1 whereby H97Q and I303T variants conveyed increased para-tolylsulfide S-oxygenation and slightly less S-oxygenation of methimazole [Cashman, 2006]. To reiterate however, catalysis was still maintained. FMO2, FMO4 and FMO5 have not been documented to exhibit many SNPs. With FMO2 being inactive in most populations along with unknown activity for FMO4, it is difficult to determine the impact of particular SNPs. No SNPs detected for FMO5 have been demonstrated to govern catalysis [Scott, 2017].

1.5.4 Endogenous roles governed by human FMOs

The extensive substrate profiles exhibited by FMOs (see section 1.5.6) has for many years supported the idea that these systems were exclusively participating in detoxification. Nonetheless, the identification of endogenous compounds as substrates for these systems has instigated the postulation that these systems may carry out additional functions. The limited roles of human FMO2 and FMO4 for the majority of the population has parted attention away from these paralogs. As such, FMO1, FMO3 and FMO5 has received a lot

of scrutiny and their potential endogenous roles will be discussed in turn (see section 1.5.6. for substrate conversion) [Phillips, 2019].

1.5.4.1 Human Flavin-containing monooxygenase 1

Principally, through knock out lines in mice, FMO1 has been attributed to two key processes: taurine production, via the oxidation of hypotaurine, and acting as a novel energy regulator [Phillips, 2019; Veeravalli, 2020; Veeravalli, 2014]. Taurine is one of the most abundant aminosulfonic acids in mammals. It can either be accumulated through the diet, or biologically synthesized from cysteamic acid or hypotaurine [Phillips, 2019]. Knockout lines consisting of mice without FMO1, FMO2 or FMO4, collectively, exhibited urine samples with high levels of hypotaurine. Resultant in vitro studies validated that FMO1 was able to metabolize hypotaurine through S-oxidation to produce taurine, albeit with high K_M values [Phillips, 2019,128]. Taurine has been described as an organic osmolyte and plays a crucial role in stabilizing and maintaining cell volumes. Furthermore, it is involved in bile salt production, modulates intracellular calcium concentrations and in certain tissues it is speculated to display cytoprotective and developmental roles [Phillips, 2019; Lombardini, 1983; Huxtable, 1992]. Whilst the roles of taurine and its precursor are still under investigation, delineating its role will aid in distinguishing the additional features governed by FMO1.

Mice knockout-lines described above were also used to convey the energy balancing function. These mice that lacked FMOs 1, 2 and 4 exhibited significantly less quantities of storage fat in white adipose tissues and consequently leaner phenotypes [Veeravalli, 2014]. This phenomenon has been associated to high rates of fatty acid beta-oxidation and the resultant depletion of triglyceride stores due to high whole-body energy expenditure. This effect has been deemed futile and a wasteful consumption of energy, as the products generated from broken down triglycerides, glycerol and non-esterified fatty acids, are then simply re-esterified back into their original triglyceride states. The authors insinuate that due to the lack of activity regarding FMO4 and the inactive truncated version of FMO2 being prevalent in most humans, this fundamental regulative mode for energy expenditure is likely conducted in a similar manner for humans and thereby governed by FMO1 [Veeravalli, 2014].

1.5.4.2 Human Flavin-containing monooxygenase 3

This enzyme gained considerable attention after being discovered to induce the unpleasant rotten-fish odour, found in patients displaying trimethylaminuria, when mutated [Krueger, 2005; Cashman, 2006; Phillips, 2019; Zhang, 2003]. The mutated isoform is unable to turnover trimethylamine that is derived from dietary components such as choline and carnitine that are principally found in fish. The lack of N-oxygenation results in the compound being excreted out through the skin as sweat, or in the breath [Ayesh, 1993].

The pungent smell is distressing for social interactions and can result in patients developing several psychological conditions. Nevertheless, research has shown that closely monitoring the diet and limiting the intake of particular food groups has greatly reduced the odour.

Fascinatingly, the product, trimethylamine N-oxide (TMAO), that forms after trimethylamine oxidation is currently under scrutiny regarding its role in metabolic processes and diseases. High plasma concentrations of TMAO has been correlated to higher risks of developing cardiovascular diseases including atherosclerosis [Veeravalli, 2018]. Albeit, these results are based on mice being fed with diets consisting of untypically high quantities of trimethylamine precursors. As such, the role of TMAO in health is still under dispute and some groups have implored that the compound actually provides beneficiary health effects [Phillips, 2019; Ufnal, 2015; Ussher, 2013].

1.5.4.3 Human Flavin-containing monooxygenase 5

Remarkably, mice that have the FMO5 gene knocked out, display similar phenotypes to mice deprived of FMO1 as mentioned above [Phillips, 2019]. These mice also display an increased energy expenditure and a lean phenotype corresponding to their depleted triglyceride stores [Phillips, 2019; Gonzalez, 2015]. Furthermore, in these mice, cholesterol levels remain similar to 10-week-old mice, despite eating more than wild type mice that display increasing plasma cholesterol concentrations. Similarly, to FMO1 knock-out mice the rate of fatty acid beta oxidation is increased in white adipose tissue. Contrastingly, the rate of fatty acid beta-oxidation in skeletal muscle is lower in FMO5 knock-out mice than the wild types, suggesting that these mice use an increased amount of carbohydrate as food source [Phillips, 2019; Gonzalez, 2015].

Mice lacking FMO5 also show reduction in the production of several proteins that are involved in carbohydrate metabolism and lipid or cholesterol biosynthesis [Phillips, 2019]. The former is impacted by the downregulation of three enzymes: aldolase B, ketohexokinase (both involved in glucose and fructose metabolism,) and glycerol 3-phosphate dehydrogenase-1 that produces NAD⁺ (that is imperative for glycolysis) and glycerol 3-phosphate which can be utilized for triglyceride production [Phillips, 2019]. Regarding lipid or cholesterol biosynthesis, expression of two enzymes is greatly impacted: cytosolic malic enzyme 1 and β -hydroxy- β -methylglutaryl-CoA synthase 1. The former performs oxidative decarboxylation on its substrate, malate, producing pyruvate. During catalysis however, the by-product NADPH can be used as a coenzyme for alternative steps for cholesterol biosynthesis. The latter enzyme is directly involved in the first step of isoprenoid biosynthesis which, when downregulated, greatly impedes cholesterol biosynthesis [Phillips, 2019].

Finally, FMO5 has been associated to an increased insulin sensitivity and glucose tolerance [Scott, 2017]. Wild type mice exhibited enlarged concentrations of both glucose and insulin with age. Mice that had FMO5 knocked out maintained these respective concentrations as if they were young mice, insinuating that FMO5 may have a key role in metabolic ageing. Furthermore, whilst wild type mice showed significant weight gain and decreased insulin sensitivity when fed a high fat diet, mutant mice lacking FMO5 did not display these characteristics [Scott, 2017]. These mutant mice also have lower amounts of tumour necrosis factor alpha and complement component 3 in epididymal white adipose tissue. Both these elements have major negative impacts on insulin sensitivity and thus provide rationale behind the mutant mice's high insulin selectivity [Scott, 2017].

1.5.5 Kinetic characterization of pig FMO1.

In this section, we will discuss the extensive kinetic analysis carried out by Beaty and Ballou that essentially mapped out the enzymatic activity of the Class B FMOs in the early 1980s [Beaty, 1981a; Beaty, 1981b]. The research was carried out on microsomal FMO1 isolated from hog liver. Due to this research still underpinning the key steps of catalysis after many decades, these two papers will be discussed in detail. Using stopped flow spectroscopy, authors characterized the intermediates formed during catalysis, the spectral features they possessed, the importance of substrates and coenzymes for oxidation and the stability of these FAD intermediate states.

1.5.5.1 *Reductive-half reaction*

In this sub-section, we will analyse the first half of the reaction whereby a reducing equivalent, reacts with the oxidized FAD for activation [Beaty, 1981a]. Firstly, they demonstrate that the reductive half reaction is biphasic and can take place with either NADPH or NADH as coenzyme. Importantly however, in line with all Class B monooxygenases, the dissociation constant for NADPH is considerably lower (approximately 8 μM) than compared to NADH (167 μM), suggesting that this coenzyme is the preferred substrate. Moreover, the maximum velocity is essentially equal for both. For both coenzymes used, they observe clear primary deuterium isotope effects during the transfer of the pro-R hydride from the nicotinamide ring. They were unable to visualize a spectral charge-transfer complexed intermediate between the oxidized FAD and the NAD(P)H, implying that rapid reduction takes place first, corroborated by the large primary deuterium isotope effects. Furthermore, the generation of such a charge-transfer complex would be expected if the rate of reduction changed greatly at different NADPH concentrations, which it did not. Additionally, the rates of each phase were directly proportional to NADPH concentration. These comments collectively suggested that a rapidly formed Michaelis-Menten-complex must come prior to the reduction step.

Once the FAD had been reduced, they observed small spectral changes occurring at longer wavelengths (500-600 nm) corresponding to a charge-transfer complex between the reduced FAD and the oxidized NADPH, NADP⁺. Throughout the catalytic cycle, they were unable to document the production of semiquinone intermediates when using the conventional coenzymes, NADPH and NADH. However, the use of one electron reducing agents such as dithionite, did produce blue semiquinone FAD intermediates. Thus, the catalytic cycle appears to utilize a two-electron based reducing system through direct addition in the form of a hydride.

Importantly, it was illustrated that the concentration or the presence of substrate had no impact on the rate of FAD reduction. In addition, FAD reduction was not rate-limiting. These findings emphasise the “cocked-gun” notation used by Henry Kamin of Duke University whereby the enzyme is prepared for action even in the absence of substrate [Krueger, 2005]. Once the substrate has been able to voyage through the protein scaffold, it will then be immediately oxidized and released into the cytosol. Finally, they illustrated that the concentration of substrate was directly proportional to the rate of reoxidation.

1.5.5.2 Oxidative Half-reaction

This half of catalysis focusses on the oxygen-based flavin intermediates, the impact of substrate and oxidized coenzyme on flavin-based intermediates and their resulting decay to regenerate oxidized FAD [Beaty, 1981b]. An important question that was addressed, concerned the order of binding between substrate and molecular oxygen. They documented that at different substrate concentrations, there was no observed change in the rate of C4a-hydroperoxy intermediate formation and implied that molecular oxygen bound first. This consequently brought attention regarding the role of NADP⁺ during catalysis and the stage at which it dissociates. Intriguingly, the rate of enzyme reoxidation was unaffected by substrate concentration in the absence of the oxidized coenzyme, NADP⁺. However, considering their previous work portrayed that substrate concentration directly impacted the reoxidation rate of NADPH reduced FMO1, it is likely that NADP⁺ remains bound throughout. Furthermore, in the absence of NADP⁺, the rate of reoxidation was 25-fold higher than the rate of turnover in the presence of substrate, implying that without NADP⁺ there is considerable uncoupling and that it is required for substrate oxidation. The importance of NADP⁺ is supported by the result that in its absence, the C4a-hydroperoxy intermediate is not detected, suggesting that its decay is faster than its formation. This implicates the coenzyme’s role in intermediate stabilization. Interestingly, the substrate binding capacity had no impact on the turnover velocity.

Beaty and Ballou were able to isolate the C4a-hydroperoxy intermediate. After reducing the FMO in an anaerobic environment, mixing the solution with an oxygenated buffer containing NADP⁺ resulted in the generation of a stable C4a-hydroperoxy intermediate.

The intermediate displayed a spectrum commensurate with p-hydroxybenzoate hydroxylase and luciferases with a canonical blue shift in the shoulder at 370 nm to 360 nm. To their surprise, the intermediate remained stable for approximately 30 minutes, before slowly decaying. Nevertheless, partial reoxidation of the FAD system spanned over approximately two hours which epitomized the stability of the intermediate. On the addition of methimazole the cofactor was completely oxidized in approximately four minutes. Preceding the C4a-hydroperoxy intermediate formation lies the ternary complex between reduced FAD, NADP⁺ and molecular oxygen, illustrated by saturation behaviour.

To date, the identification of the rate-limiting step is still a challenging feat. The authors suggest that as the overall rate of catalysis is very similar among a broad range of substrates the rate-determining step is not substrate dependent and corresponds to the decay of a flavin-based intermediate. Therefore, they speculate that the decay of the pseudo base, flavin-hydroxy intermediate, and the release of water to regenerate the oxidized flavin species is the rate-determining step. However, whilst many recent articles today still state that the release of NADP⁺ from the binding domain could be rate-limiting, Beaty and Ballou imply that this step cannot be rate-determining. They infer that at millimolar concentrations, NADP⁺ has completely bound to the enzyme within the deadtime of the spectrophotometer instrument, approximately 3 ms, with no further change in the UV/Vis spectrum. Taking into account that the K_d is 194 μM and the k_{on} is roughly $106 \text{ M}^{-1} \text{ s}^{-1}$, the k_{off} is at a minimum, 200 s^{-1} , which is not rate-limiting.

1.5.6 Substrate metabolism and conversion

Daniel Ziegler was one of the leading pioneers in the field of flavin enzymology and made the following statement [Ziegler, 1990]:

“Although the vast majority of these compounds are structural analogues bearing the same functional group, there is little question that this enzyme has the broadest substrate specificity of any single monooxygenase described to date.”

Herein this section, we will divulge into the various compounds that are metabolized by the FMO family and portray the extensive substrate profiles employed by these paralogs. With the FMOs being particularly unfriendly *in vitro*, the initial research focussed on substrate conversion to differentiate and distinguish various FMOs from each other. Work by Fiorentini et al [Fiorentini, 2016; Fiorentini, 2017], was able to use substrate profiling and screening to confirm that FMO5 possessed unique enzymatic activity and performed Baeyer-Villiger monooxygenase activity. For more information regarding the individual compounds documented to be FMO substrates, the reader is directed to the extensive review written by Krueger and Williams [Krueger, 2005] and the article disseminated on human FMO5 [Fiorentini, 2016].

The majority of the work was carried out on pig liver FMO1 and rabbit pulmonary FMO2 for substrate profiling [Krueger, 2005; Ziegler, 1990; Jakoby, 1990; Cashman, 1995]. Collectively, the research validated that soft nucleophilic centres, such as Sulfur and Nitrogen atoms, were the key elements of the substrate that were transformed and oxidized during catalysis [Ziegler, 1990]. Extensive profiling elucidated essential properties that the substrate of interest must possess to be a potential candidate. Firstly, it became clear that substrates possessing a single positive charge, such as tertiary amines, were excellent substrates [Ziegler, 1990]. On the contrary, substrates possessing negative charges were typically poor substrates except for certain acidic substrates such as lipoic acid. Additionally, zwitterions such as amino acids are also considered poor substrates, however, methionine did show oxygenation but with a K_M value that was too high for physiological relevance [Ziegler, 1990; Phillips, 2019]. Intriguingly, despite amine centres being considered optimal functional groups for oxygenation, substrates possessing multiple amino groups were not transformed by FMOs.

Pig liver FMO1 was demonstrated to be well relaxed regarding substrate size. This enzyme could metabolize small substrates such as thiourea and trimethylamine, to large bulky substrates including trifluoperazine and sulindac sulfide [Krueger, 2005; Ziegler, 1990; Williams, 1984; Tynes, 1985; Nagata, 1990]. Rabbit lung FMO2 however exhibited stark differences regarding its substrate profile and typically could not cope with such large substrates [Williams, 1984; Nagata, 1990]. For example, chlorpromazine and imipramine that are good substrates for pig liver FMO1 (Figure 1.5.6.1), were not metabolized by rabbit lung FMO2. Both these compounds contain a bulky head group attached to a thin long aliphatic saturated tail with a terminal amino group. Williams and co-workers, [Williams, 1984] carried out detailed substrate analysis with similar substrates (10-(N,N-dimethylaminoalkyl)-2-(trifluoromethyl) phenothiazines) by varying the length of this aliphatic tail from 2 to 7 carbons. These experiments were conducted to validate whether the width of the substrate tunnel for FMO2 was the bottleneck for substrate transformation. They rationalised that the analogues possessing five-carbons or more were actively metabolized by FMO2, whilst shorter chain-based substrates were not [Krueger, 2005; Williams, 1984]. All of these substrates were metabolized by FMO1. This result led to Williams and authors to hypothesize that the substrate tunnel or vestibule for FMO2 was considerably deeper and thinner than FMO1, meaning that these substrates needed to be thin and small for conversion to be apparent, or otherwise bulky but with a long extended tail [Williams, 1984].

Interpreting the reasons behind why certain substrates are better accepted for other FMOs is still difficult to conclude. For example, even though Williams et al, provided evidence for the narrow substrate tunnel for FMO2, [Williams, 1984] it is unclear as to why primary aliphatic amines are considerably better substrates for FMO2 when compared to FMO1. The latter paralog seems to be better suited for more substituted amines, suggesting that the electrophilic capacity of the C4a-hydroperoxy intermediate in the active of FMO2 is higher than that of FMO1. Hence, it is likely that subtle differences in the active site result in differing selectivity [Krueger, 2005]. There is also noticeably higher selectivity for trimethylamine in FMO3 than the other FMOs suggesting slight variations among the family. Unfortunately, due to extremely low expression levels of FMO4, still no evidence has been extrapolated regarding the substrate profiles and activities of FMO4. Overall, the substrate profiles documented thus far portray FMO1 to be the most promiscuous enzyme of the ‘canonical’ FMOs that pursue N- and S-based oxidations, with FMO2 being the most restrictive and FMO3 being somewhat intermediate [Krueger, 2005; Ziegler, 1990].

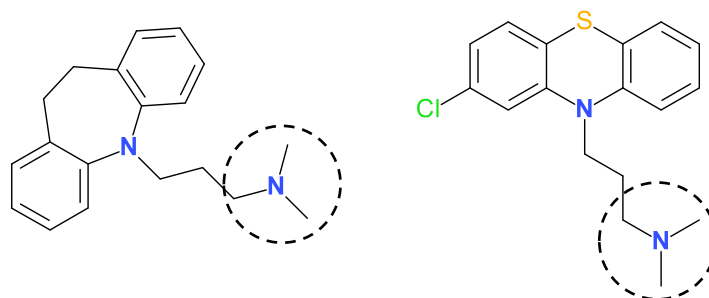


Figure 1.5.6.1: FMO1 substrates un-oxidized by FMO2. Left, Imipramine. Right, Chlorpromazine. Dashed circles highlight the site of oxidation.

For many years, the activity of FMO5 was not well understood and its reputation became very similar to FMO4 such that it did not seem to play a fundamental role within humans [Krueger, 2005; Rettie, 1994]. FMO5 gained attention however due to its high expression levels in both foetal and adult liver, lung and small intestine. Only certain substrates were suggested to be transformed by this FMO including methimazole, meaning that the enzyme seemed to portray slight and extremely limited FMO activity [Krueger, 2005; Overby, 1995]. Fiorentini et al., were able to successfully express human FMO5 in *E. coli*, solubilize it in various detergents and diagnose its governed mode of action [Fiorentini, 2016; Fiorentini, 2017]. After screening multiple classical FMO substrates including trimethylamine and thioanisole they demonstrated that conventional Nitrogen-containing compounds were not substrates, whilst highly soft nucleophilic centres, such as sulfur atoms, could be oxidized [Fiorentini, 2016]. This finding was in line with previous reports of human FMO5 [Leoni, 2008; Ohmi, 2003]. This initial finding spurred on the idea that

FMO5 is more similar to yeast FMOs that are also unable to metabolise N-containing soft nucleophiles [Bonanno, 2007]. On the addition of typical Baeyer-Villiger substrates, FMO5 was demonstrated to oxidize a range of compounds. Both cyclic and aliphatic ketones were excellent substrates and formed esters and lactones, respectively. Furthermore, aldehydes were also good substrates suggesting that essentially all ketone and aldehyde compounds, that fulfil steric and charge parameters for substrate entry, are potential substrates [Fiorentini, 2017]. Importantly, authors could illustrate the importance that this family member plays in drug metabolism [Fiorentini, 2017]. Nabumetone, a non-steroidal anti-inflammatory and pentoxifylline, currently used as a hemorheological agent, are two well used clinically applied pharmaceuticals. Each of these compounds were metabolized to a corresponding ester by FMO5 and delineates a principal mode for their detoxification within the liver [Fiorentini, 2017].

1.5.6.1 Endogenous substrates

A mounting interest in the FMO substrate selectivity revolves around potential endogenous compounds. This includes soft nucleophilic centres found on amino acids such as methionine and cysteine. A recent review by Phillips and Shephard discusses the endogenous substrates metabolized by FMOs and the importance of these transformations for metabolism [Phillips, 2019]. In this section we will consider the roles FMOs play towards the metabolism of tyramine, phenethylamine, cysteamine, methionine and lipoic acid.

While the tyramine and phenethylamine are principally oxidized by monoamine oxidases (MAOs), FMO3 has been demonstrated to oxidize both in the liver [Lin, 1997a; Lin, 1997b]. These substrates can bind the trace amine associated receptor 1, mimicking other sympathomimetic molecules including dopamine and epinephrine [khan, 2016; Broadley, 2010]. Binding this receptor activates the signalling pathway through kinases A and C that results in phosphorylation of the dopamine transporter. This consequently promotes the uptake of monoamine neurotransmitters by presynaptic neurons. Tyramine has also been shown to act upon the peripheral nervous system, inducing vasoconstriction and raising blood pressure [Broadley, 2010]. Importantly however, the oxidized and transformed substrates are poorer substrates for the receptor and therefore, FMO3 may represent a crucial detoxification route and can alter neuronal signalling.

Cysteamine is produced through the degradation cycle of Coenzyme A and represents a key precursor in the biosynthesis of the neurotransmitter, hypotaurine [Phillips, 2019; Dominy, 2007]. Cysteamine has been illustrated to be pivotal for the conversion of cystine into cysteine-based compounds that unlike cystine, cannot crystallize at high concentrations and cause cellular damage [Phillips, 2019; Gahl, 2002]. However, cysteamine at high concentrations in the cell is extremely toxic to cells and both FMO1 and FMO2

have been shown to oxidize this compound at the sulfur position [Krueger, 2005]. Perhaps, these FMOs are able to fine tune the concentration of cysteamine in the cell. They ensure that the concentration is not too high in order for the substrate to be detrimental towards cell prosperity, whilst still promoting its positive role in cystine degradation.

Methionine is an essential amino acid that is not produced by animals. It is utilized as a building block for many essential physiological compounds including carnitine, taurine and several phospholipids [Phillips, 2019]. Additionally, it is used for the synthesis of S-adenosyl methionine that acts as a key methylating agent for a range of enzymatic processes [Ripp, 1999]. Despite human FMO3 being able to metabolize this compound, the K_M is extremely high, approximately 4 mM which suggests that the binding is too low for physiological relevance [Mato, 1997]. The role of such a turnover is still under investigation.

Lipoic acid is an essential molecule utilized in many different metabolic processes and can be synthesised in animals [Phillips, 2019]. It is used as a prosthetic group for several systems including 2-oxoacid dehydrogenases that are involved in the citric acid cycle [Estrada, 1996]. Additionally, it is crucial for amino acid metabolism that include other dehydrogenases. It also plays a key role in producing 5,10-methylene tetrahydrofolate through a glycine cleavage system that is integral for nucleic acid synthesis [Phillips, 2019]. The compound has been attributed to other key processes including acting as an activator for the insulin signalling cascade [Diesel, 2007]. Human FMO2 has been observed to metabolize both the lipoic acid and lipoamide with a K_M of approximately 72 μ M which is physiologically relevant [Krueger, 2005]. However, the products of the reaction have not been characterized and it is unclear whether FMOs are able to metabolize other methyl sulfide derivatives of lipoic acid.

1.6 ANCESTRAL SEQUENCE RECONSTRUCTION

1.6.1 Overview

Ancestral Sequence Reconstruction (ASR) represents one of the key historical biochemistry techniques used for dissecting protein evolution [Hochberg, 2017; Hadzipasic, 2020; Risso, 2018; Eick, 2017; Starr, 2017; Starr, 2016]. Essentially, the approach allows researchers to delineate and distinguish the features behind the protein scaffold that determine and dictate enzymatic activity [Merkel, 2016a, Merkel, 2016b]. Typically, protein sequences derived from a given family are collected, curated and using different phylogenetic inference approaches, the evolutionary history is inferred. An evolutionary tree maps out the entire history of a given protein family within a time period. When following the trajectories of enzyme activity of substrate preference, this snapshot usually illustrates the point in time whereby the two different functionalities emerged. Using ASR, users are able to infer the most likely sequences that correspond to the ancestral states selected. The sequences are then synthesized, expressed as fully functioning proteins which are then biochemically

inspected. Depending on the results, the user can then perform various mutagenesis experiments to elucidate the roles of certain residues for a given chemical activity.

Recently, multiple research groups have started to appreciate the significant melting temperature enhancement observed for ancestrally reconstructed systems [Risso, 2018; Wheeler, 2016]. Accordingly, many biotechnological applications have been envisaged regarding the use of ancestral enzymes. Herein this thesis, we will explore the potential ASR plays on structural elucidation. For now, this section aims to dissect and convey the fundamental steps involved in ASR and describe some theory behind the analysis (Figure 1.6.1).

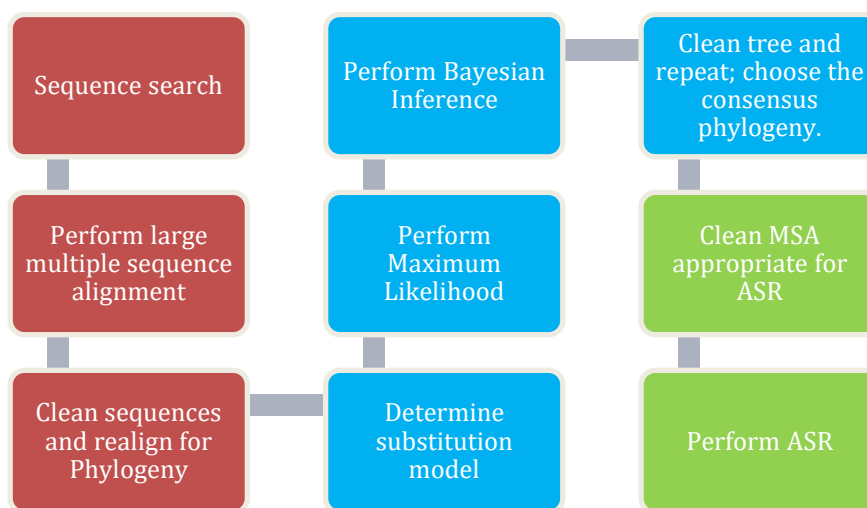


Figure 1.6.1: Overview of the Ancestral Sequence Reconstruction process. The individual steps are shown with available programmes and servers described. The maroon boxes portray the sequence collection and multiple sequence alignment process required to generate the phylogeny. The turquoise boxes describe the phylogenetic tree curation necessary to illustrate the evolution of the protein of interest. The light green boxes describe how the ancient sequence are constructed.

1.6.2. Case Study

To assist the reader, and provide a brief overview on the process, a theoretical example is shown (Figure 1.6.2.1). In this scenario a protein family, that represent class B flavin-dependent monooxygenases, has three different paralogs resulting from two gene duplications with either FMO or BVMO activity. In this study, the aim is to determine

what residues give rise to the differing activities and the origin of the BVMO/FMO functionality. One could simply perform site-directed mutagenesis by comparing each active site in turn, for each paralog, and exchanging their respective active sites to see what core carries out a certain role. This approach is referred to as ‘horizontal analysis’ and involves switching residues across horizontal space whereby the proteins in question are of the same age in sequence time (in this case, extant) [Hochberg, 2017]. For example, residues in the active site of the BVMO clade, would be exchanged with those in the FMO clades (Figure 1.6.2.1). ASR, moreover, utilizes a vertical approach, assessing changes that occur throughout various ancestors and their respective successors. These are represented by class B monooxygenase ancestors, AncBMO1 and AncBMO2 (Figure 1.6.2.1). Using the horizontal approach has been successful in many situations but the vertical approach can be very powerful by not only indicating one or two key residues, but also highlighting peripheral residues that may encourage a certain activity (see section 3.6) [Hochberg, 2017]. Furthermore, this technique considers potential epistasis that could underpin certain activities and phenotypes. For simplicity, we can define epistasis as the promotion or restriction of a given enzymatic activity based on the presence of certain residues within the peptide chain [Starr, 2016].

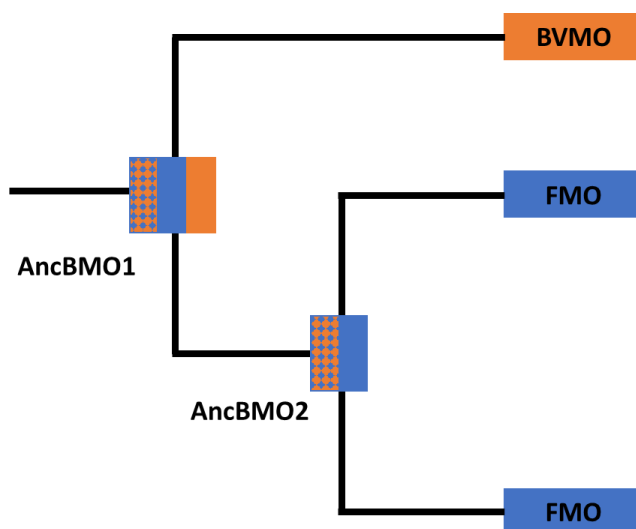


Figure 1.6.2.1: Case Study - Ancestral Sequence Reconstruction for Class B monooxygenases. A phylogenetic tree of a protein family that display BVMO and FMO members are shown. Enzymes that display BVMO and FMO activities are shown in orange and blue squares, respectively. Enzymes exhibiting mixed functionality are shown with a chequered blue and orange box. Ancestral intermediates for class B monooxygenases are described as AncBMOs. To determine the origin and development of the BVMO and FMO activities, AncBMO1 and AncBMO2 would be extracted for catalysis.

In our scenario, AncBMO1 and AncBMO2 would be extracted for biochemical characterization. Each substrate profile would be determined by using FMO and BVMO substrates. According to the phylogenetic tree shown in Figure 1.6.2.1, AncBMO1 could display three different activities – FMO, BVMO or mixed. If AncBMO1 displayed BVMO activity, then the emergence of FMO activity would have taken place during the development of AncBMO2 (Figure 1.6.2.1). Importantly, we cannot rule out that AncBMO2 may still display some BVMO activity that was then lost during the duplication. Hence, AncBMO2 is depicted with two differing potential catalytic modes. Alternatively, if AncBMO1 was a true FMO, then the changes along the branch giving rise to the canonical BVMOs would be inspected. If AncBMO1 exhibited mixed functionality, then all the branches would need clarification and more ancient ancestors would need to be constructed to reveal the origin.

The newly resolved phenotypes can then be inspected more closely to evaluate what changes induced new phenotypes. Substitutions along each branch of interest are then evaluated and synthesized using site-directed mutagenesis. After multiple experiments, the critical residues that confer new activities are revealed. With these traits elucidated, researchers can delineate with the upmost specificity what residues are required for particular activities and the results can be translated to other similar systems.

1.6.3. The development of ASR

The idea behind ASR was initially conceived by Pauling and Zuckerkandl in 1963 [Pauling, 1963]. They postulated that it could be possible to infer the sequences of ancestral proteins, synthesize them and determine the physio-chemical properties of these ancient systems. This idea spurred on the development of evolutionary analysis techniques that could effectively map out the evolution of a family of proteins based on a given set of sequences. The improvement and advancement in genome sequencing have enabled researchers to build thorough and detailed phylogenetic analyses of large divergent protein families [Hadzipasic, 2020; Voordeckers, 2012; Weber, 2016]. An important remark regarding the approach is that its aim is to reliably determine the overall phenotype for a particular ancestor and not its precise sequence [Hochberg, 2017].

Many key features of a protein family make them worthy candidates for ASR: the availability of sequences in a range of phyla, multiple paralogs emerging over time in different phyla or families that display several substrate portfolios, rates of catalysis or enzymatic activities [Voordeckers, 2012]. With these parameters in mind, one can create a range of ancestral states and probe their differing activities. Certain examples of protein families used in ASR include alpha glucosidases, dehydrogenases, steroid receptors and Ser-Thr kinases [Hadzipasic, 2012; Voordeckers, 2012; Weber, 2016]. As an example, Verstappen et al., [Voordeckers, 2012] were able to reconstruct a range of different alpha

glucosidases and delineate the key residue switches that promoted turnover of isomalate sugars over malate-based sugars. These subtle differences have significant effects and by using ASR, researchers can discriminate between the differing activities and rationalize alternative substrate conversions observed within a protein family.

A frequent feature observed for these enzymes is a heightened thermal stability [Risso, 2018; Merkl, 2016b; Risso, 2013]. Many ancestral protein reconstructions have been demonstrated to be significantly more thermostable than their extant counterpart. Reasons behind this trait are currently being explored but also exploited for biocatalytic means. Considering the extreme environments that were previously prevalent on a much younger earth, one could imagine that the stabilized protein produced from the ASR approach is simply an accurate representation of the primordial state [Risso, 2015; Ashenberg, 2013]. Whilst this may be true for some studies, there have been several examples whereby the reconstructed ancestors do not originate from such climates. In a recent review by Risso et al, [Risso, 2018] they postulate that overtime, as temperatures cooled, proteins accepted several destabilizing mutations as they did not greatly impact activity or substrate acceptance. Moreover, these enzymes still have the propensity to reconstitute their original thermostable state if the climate were to unexpectedly change due to natural causes. The authors imply that the stability brought about by this process is derived from this primordial state coming back into focus during reconstruction. Furthermore, they imply that the increased stability is significantly higher than predictions involving computational statistical estimates and therefore cannot be considered as an artifact derived from sequence bias [Risso, 2018; Williams, 2006].

There has also been great interest regarding the promiscuity of ancestral proteins [Risso, 2018; Copley, 2015; Sheehan, 2001]. From a biocatalytic point of view, being able to find thermostable enzymes that can then be fine-tuned to perform a given catalytic function can be very valuable. A well renowned hypothesis is that ancestral enzymes needed to be promiscuous so that their limited genome was able to perform multiple chemical processes to fuel metabolism [Hochberg, 2017; Risso, 2018; Siddiq, 2016]. Despite this key trait, little known information has been captured documenting these properties. Many studies have been directed towards highlighting the activity of the enzyme with respect to known substrates for their respective extant equivalent. However, the few studies that have been undertaken suggest that there are mixed results regarding the promiscuity-specificity trade-off. Whilst some studies document the promiscuity to specificity switch, others highlight switches between two specific substrates, switching between intermediates and then also the contrary – losing specificity with sequence time [Hochberg, 2017; Risso, 2018; Siddiq, 2016]. Nevertheless, extensive work carried out by Thornton and co-workers illustrated that most enzyme families evolved from a multi-functional ancestor, whilst de novo activities are far less common [Siddiq, 2016]. They argue that evolving new specificity from

a multi-functional ancestor is more probable as the moonlighting function has a higher chance of survival as it is protected by the main function of the gene.

1.6.4. Performing ASR

Once the evolutionary history of a protein family is known, we can turn to the reconstruction. ASR has to determine the probability for a given state (residue) at each site in a sequence based on an extensive multiple sequence alignment (MSA). The probability of a given residue is determined through maximum likelihood (ML) under a specific substitution model [Yang, 1995]. The posterior probability of each site for a particular residue can be described as follows (Equation 1.1):

$$L_n(\mathbf{x} | \text{data, tree, model}) = P_n(\text{data} | \text{tree, model, } \mathbf{x}) \quad (1.1)$$

L_n describes the likelihood of finding amino acid \mathbf{x} at each given site by the resultant data (MSA), given tree and substitution model. This is in turn equal to the probability of observing the MSA (data) based on the tree, model and the amino acid at a given site [Yang, 1995]. The residue that gives the highest probability at each site is described as the maximum a posteriori (MAP) state and will be found in the ML sequence for a given node. As such, each node will consist of an ancestral sequence made up by all the MAP states. The user may then extract their ancestral sequences from the output file generated by ASR.

Importantly, in order to assess the robustness of the reconstruction, it is advisable to consider also the ‘alternative ancestor’. This consists of a sequence made up of the second most likely residue at each state that was ambiguously reconstructed. It should be noted that, the ancestral sequence represents the most likely ancestral state and may not represent the exact sequence. The objective of ASR is to reconstruct the phenotype of the ancestor in question, such as substrate profile, meaning that the exact sequence is not fundamental [Merkel, 2016a; Risso, 2013]. It is this property that underpins the reason behind the importance of the alternative ancestor because if it illustrates characteristics commensurate with the ML ancestor, it confirms the robustness of the reconstruction. At this stage, ASR has been completed and the research can now turn to experimental characterization. Desirable ancestral sequences can be ordered and cloned into plasmids and vectors of choice, expressed, purified and characterized [Merkel, 2016a; Risso, 2013].

2.MATERIALS AND METHODS

Herein this section lists all the methods and materials that I performed and used during this thesis. Additional materials and methods, including the initial construction of the FMO tree and stopped-flow kinetics that I was not a part of are described in the articles/manuscripts included in sections 3.3 and 3.5.

2.1 CLONING AND EXPRESSION OF THE ANCFMOS.

Complementary DNAs were ordered from Genescript containing BsaI sites at both the 5' and 3' ends of the insert. The insert contained overhangs TGGT and CAAG at the 5' and 3' ends, respectively, to then be inserted into common pBAD-NK destination vectors with the following modifications: three BsaI sites were eliminated and two were introduced to facilitate the cloning that incorporated SUMO and 6×His-tag regions to the amino-terminus. Inserts were fused into the destination vectors through Golden Gate cloning. The sample was prepared with the following: 75 ng of Golden Gate entry vector (a molar ratio of 2:1 between insert and vector), BsaI-HF (15 U), 30 WU T4 DNA ligase (15 U), T4 DNA ligase buffer (1×) and nuclease-free water added to a final volume of 20 µl. During the cloning procedure, a negative control was prepared with the fragments and inserts omitted. The Golden Gate assembly was conducted in the following manner, where maximum efficiency was desired: a cycle of 37 °C for 5 min followed by 16 °C for 10 min was repeated 30 times; followed by 55 °C for 10 min, 65 °C for a further 20 min; finishing with 8 °C for 20 min. Once cloned, the plasmids were transformed by heat shock into *E. coli*. BL21 cells (25 s, 42 °C). Cells from the resulting colonies were pre-inoculated into 100 ml of LB broth containing 100 µg ml⁻¹ of ampicillin and grown overnight at 37 °C. The cultures were then transferred to 1 l Terrific Broth cultures (15 ml) and grown at 24 °C, 180 r.p.m. for 5–6 h until the optical density (OD) reached 0.3 – 0.5. 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 (AncFMOs 2 and 3-6) or 17 °C (AncFMOs 5 and 1), 180 r.p.m. for an additional 24 h. Cells were then harvested by centrifugation (5,000g, 15 min, 10 °C), flash frozen in liquid nitrogen and stored at –80 °C.

2.2 SITE-DIRECTED MUTAGENESIS OF MUTANTS ANCFMO2-E281H, ANCFMO3-6-E281H, ANCFMO5-H282E AND ANCFMO2-E281A MUTANTS.

AncFMO2 and AncFMO3-6 E281H constructs: PCR-reaction mixture was prepared with 10 μ M primer, forward and reverse, 100 ng of template DNA, 1.6 % DMSO, 0.8 mM MgCl₂ and 1 \times Pfu Ultra II Hotstart Master Mix (Agilent). The Quickchange PCR cycle was performed using the following method: first a 5-min incubation at 95 °C, then cycles (95 °C for 5 min, 60 °C for 30 s, 72 °C for 6 min) were repeated 25 times; followed by 72 °C for 10 min and finishing with 8 °C on hold. The PCR mixture was digested with DpnI overnight and transformed into *E. coli*. AncFMO5-H282E construct: PCR-reaction mixture was prepared with 10 μ M primer, forward and reverse, 100 ng of template DNA, 3.0 % DMSO and 1 \times 2xPhusion Masters mix (Thermo Scientific): first a 30 second incubation at 98 °C, then cycles (98 °C for 8 seconds, 64 °C for 30 s, 72 °C for 90 seconds) were repeated 35 times; followed by 72 °C for 10 min and finishing with 8 °C on hold. The PCR mixture was digested with DpnI overnight and transformed into *E. coli*. AncFMO2-E281A construct: PCR-reaction mixture was prepared with 10 μ M primer, forward and reverse, 100 ng of template DNA, 3.0 % DMSO and 1 \times 2xPhusion Masters mix (Thermo Scientific): first a 30 second incubation at 98 °C, then cycles (98 °C for 8 seconds, 64 °C for 20 s, 72 °C for 90 seconds) were repeated 30 times; followed by 72 °C for 10 min and finishing with 8 °C on hold. The PCR mixture was digested with DpnI overnight and transformed into *E. coli*.

2.3 CELL DISRUPTION, EXTRACTION, AND PURIFICATION OF ANCFMOS.

All procedures were carried out in ice or at 4 °C. Cells (approximately 20 g) were resuspended (1:5) in buffer A (250 mM NaCl, 50 mM KH₂PO₄, pH 7.8) and included additional protease inhibitors: phenylmethylsulfonyl fluoride (1 mM), leupeptin (10 μ M), Pepstatin (10 μ M), and DNase I (5 μ g g⁻¹ of cell paste). The solution was stirred and incubated at 4 °C for 45 min before cell lysis was conducted using sonication or a high-pressure homogenizer. Sonication was conducted using the following conditions: 50 ml solution, 5 s on, 5 s off, with a total sonication time of 20 min using a microtip (70% amplitude). Cells were passed through a high-pressure homogenizer twice. Lysed cells were then spun down (1,200g, 12 min, 4 °C) to remove the cell debris. The resultant supernatant was then centrifuged further (56,000g, 1 h and 40 min, 4 °C) to collect the membrane pellet, which was then re-homogenized in buffer A (15 ml) and centrifuged again (56,000g, 1 h and 20 min, 4 °C) to further purify the insoluble material. The resulting pellet was re-homogenized in buffer A (7 ml) and diluted to a final concentration of 13 mg ml⁻¹ (assayed using Biuret reagent). Triton X-100 Reduced (TRX-100-R) (Sigma-Aldrich) was then added to the solution (0.5% (v/v) final concentration) and mixed overnight at 4 °C. The detergent-solubilized fraction containing the AncFMOs was then abstracted by collecting the supernatant after centrifugation (56,000g, 1 h and 20 min, 4 °C). The supernatant was then transferred to a pre-equilibrated (with buffer A and 0.05% (v/v) TRX-100-R) gravity

column containing Ni-resin (GE Healthcare). The supernatant was washed with buffer A, containing 0.05% TRX-100-R, and then with increasing concentrations of buffer B (50 mM KH_2PO_4 , 500 mM NaCl, 300 mM imidazole, pH 7.8), also containing 0.05% (v/v) TRX-100-R, in step-by-step fashion: 5 mM imidazole wash, 30 mM imidazole wash and finally a 300 mM imidazole wash, where the protein then eluted. The buffers were then exchanged using a centrifugal filter unit (50-kDa cut-off) and multiple washes with buffer A with 0.05% (v/v) TRX-100-R. This step was important for removing high concentrations of imidazole employed during the elution. The protein sample was then concentrated down to a final volume between 500 and 1,000 μl . The sample was then mixed with a 6 \times His-tagged SUMO protease (1.2 mg ml^{-1}) to a volume ratio of 10:1 and incubated overnight at 4 °C. The sample was then loaded onto an Äkta purification system (GE Healthcare) endowed with a multiwavelength detector (set at 280, 370 and 450 nm) and then onto a Ni-affinity His-trap column (GE Healthcare). The column was pre-equilibrated with buffer A containing 0.05% (v/v) TRX-100-R, as stated before, with the proteins eluting in the presence of 6 mM imidazole, derived from buffer B (2%) containing 0.05% (v/v) TRX-100-R. The SUMO-His-tag cleaved protein was then concentrated, and buffer exchanged using a concentrating centrifugal filter unit (50-kDa cut-off) to a final volume between 250 and 500 μl . The sample was incubated for 1 h with 100 μM FAD at 4 °C and then loaded onto a gel filtration column (Superdex 200 10/300, GE Healthcare) pre-equilibrated with a storage buffer (50 mM Tris-HCl, pH 8.5, at 4 °C, 10 mM NaCl) and a detergent of choice to obtain a higher degree of purity (obtained from Anatrace). Typically, dodecyl- β -D-maltoside was used (0.03% (w/v), analytical grade), but other detergents were used for crystallization screenings at 3 \times their respective critical micelle concentration (CMC). The protein eluted with a very high purity and homogeneity (evaluated by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and the shape of the peak in the chromatogram, respectively) with an elution volume of 10.5–11 ml. The sample was concentrated to 100 μl using a centrifugal filter unit (50-kDa cut-off) with a final concentration ranging from 5 to 30 mg ml^{-1} .

2.4 WESTERN BLOT FOR ANCFMO4 EXPRESSION.

The following detergents were screened for protein extraction: 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). Resuspended membranes, generated by the typical cell lysis procedure described above, were incubated and mixed separately with different detergents (final concentration of 1%) at 4 °C overnight. The solution was then spun down at 100 000g, at 4 °C for 1 hour to pellet the insoluble components. The supernatants were then collected and run on an SDS-PAGE Gel. The gel was then transblotted onto a membrane and washed with a milky solution comprising milk powder (2.5% (w/v)), TWEEN-20 (0.05% (v/v)) and TBS (final volume 50 ml) for

one hour. The solution was then washed with the same solution including an anti-6xHis-tag antibody attached to a horseradish peroxidase (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 then washed with a peroxidase solution and Enhanced chemiluminescence (ECL) substrate (BIO-RAD) to initiate fluorescence.

2.5 DETERGENT SCREENING FOR ANCFMO EXTRACTION.

Evaluating which detergents extracted the protein from the membrane fraction is paramount for purification. The following detergents were screened: Triton-X-100 reduced (TRX); dodecyl maltoside (DDM); Octyl glucoside (OG); Lauryl dimethylamine N-oxide (LDAO); Lauryl Maltose Neopentyl Glycol (LMNG). A frozen fraction of membrane pellet was thawed (540 μ L, 75 mg/mL) and incubated with a detergent (60 μ L, 10% (w/v)/(v/v)) at 4 °C for one hour. The solution was then spun down at 16 200g, 4 °C for 1 hour to pellet the insoluble components. The supernatant was then collected and run on an SDS-PAGE Gel. The samples that showed (relatively) large bands at approximately 74 kDa, illustrated efficient protein extraction.

2.6 PREPARATION OF HUMAN FMO3 AND HUMAN FMO5.

Full-length cDNA encoding for *Homo sapiens* FMO3 (UniProt P31513) and FMO5 (Genbank Z47553) were cloned into a modified pET-SUMO vector (Invitrogen) to allow insertion of a cleavable N-terminal 8 \times His-SUMO tag. Expression, cell disruption, extraction and purification were performed according to the methods previously described for human FMO5 [Fiorentini, 2016].

2.7 KINETIC CHARACTERIZATION

2.7.1 Steady state kinetics of AncFMOs 2, 3-6 and 5.

Steady-state kinetics assays were performed in technical duplicates on a Jasco V-660 spectrophotometer. Enzyme activity of the ancestral proteins was measured by monitoring NADPH consumption (absorbance at 340 nm, $\epsilon_{340} = 6.22 \text{ mM}^{-1} \text{ cm}^{-1}$ 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. For the determination of the K_M of the substrates, 100 μ M, 100 μ M and 50 μ M NADPH were used for AncFMO2 (0.1 μ M), AncFMO3-6 (0.1 μ M) and AncFMO5 (2.0 μ M), respectively. For the determination of the K_M for NADPH, 1 mM trimethylamine was used as the oxygen-accepting substrate for AncFMO2 and AncFMO3-6, and 30 μ M of heptan-2-one was used for AncFMO5. The spectrophotometer was set at 37 °C and the NADPH and substrate mix were also incubated at 37 °C for 5 min before starting the reaction by adding the enzyme. The pH and temperature conditions were set based on literature studies for a fair comparison with

previously reported properties of mammalian FMOs. NADPH uncoupling rates were determined in the absence of substrates.

2.7.2 Steady state kinetics of AncFMO2-E281A.

Steady-state kinetics assays were performed on a Varian spectrophotometer (Cary 100 Bio) equipped with a thermostatic cell compartment. 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. For the determination of the K_M for NADPH, no oxygenating substrate was added. Concentration of protein was 1 μM and assayed using FAD absorbance at 450 (absorbance at 450 nm, $\epsilon_{450} = 12.0 \text{ mM}^{-1} \text{ cm}^{-1}$ for FAD). The spectrophotometer was set at 37 °C. The pH and temperature conditions were set based on literature studies for a fair comparison with previously reported properties of mammalian FMOs.

2.8 MELTING TEMPERATURE ASSAYS

2.8.1 ThermoFAD

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 $485 \pm 30 \text{ nm}$ excitation and $625 \pm 30 \text{ nm}$ emission for 5 s) [Forneris, 2009]. Concentrations were determined using a molar extinction coefficient of $12 \text{ mM}^{-1} \text{ cm}^{-1}$ for the FAD band at 442 nm. Experiments were performed in triplicate using human FMO5 and the AncFMOs, in the presence or absence of NADP⁺. Each sample contained the protein of interest (4 μM), with or without NADP⁺ (200 μM), made to a final volume of 20 μl using the storage buffer and incubated in ice for 1 h. The melting temperature, T_m , for human FMO3 (0.05% (v/v) TRX-100-R) was determined in technical duplicates and with a final protein concentration of 5 μM in buffer (100 mM), with varying pH values (pH 6–6.5 MES, pH 7–8 HEPES, pH 8–9 Bicine, pH 9.5 CHES), KCl concentrations (0–500 mM in HEPES pH 8) and NADP⁺ concentrations (5–500 μM in HEPES pH 8, 10 mM KCl, 0.05% (v/v) TRX-100-R), in an attempt to generate optimal storage buffer conditions.

2.8.2 TychoTMNT.6

The 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 μM NADP⁺, respectively. Concentrations of AncFMO1 were determined using $\epsilon_{\text{FAD}} = 12.0 \text{ mM}^{-1} \text{ cm}^{-1}$ at 442 nm. Experiments were performed in triplicate, with each sample containing AncFMO1 (1.0 mg ml⁻¹, determined using the calculated molecular weight of AncFMO1, 61 kDa), with or without NADP⁺ (200 μM), made to a final volume of 10 μl 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 a control experiment was performed using AncFMO2, which corroborated the previously observed T_m values (data not shown).

2.9 CRYSTALLIZATION AND STRUCTURAL DETERMINATION

2.9.1 AncFMOs 2, 3-6 and 5

Each AncFMO crystallized in a range of conditions with multiple detergents. Typically, PEG 4000 was optimal for crystallization. The highest diffracting crystallization conditions for each AncFMO are described below. AncFMO2 (with and without NADP⁺): 12–15 mg ml⁻¹ of AncFMO2 (in storage buffer and CYMAL-6 (0.09% (w/v))) was incubated with crystallization conditions of HEPES buffer (0.1 M, pH 7.5) and PEG 4000 (10%) at 20 °C with a ratio of 1:1 in a sitting drop. Sitting drop was 2 µl after mixing and the reservoir was 1 ml. Prior to crystallization, NADP⁺ (1 mM final) was incubated with 12–15 mg ml⁻¹ of AncFMO2 for 1 h at 4 °C. After 2 days, large yellow crystals formed. AncFMO3-6: 12–15 mg ml⁻¹ of AncFMO3-6 (in storage buffer and CYMAL-6 (0.09% (w/v))) was incubated with crystallization conditions of sodium acetate buffer (0.1 M, pH 5.5) and PEG 4000 (7.5%) at 20 °C with a ratio of 1:1 in a sitting drop. Sitting drop was 2 µl after mixing and the reservoir was 1 ml. Prior to crystallization, NADP⁺ (1 mM final) was incubated with 12–15 mg ml⁻¹ of AncFMO3-6 for 1 h at 4 °C. After 1 d, large yellow crystals formed. AncFMO5: 12 mg ml⁻¹ of AncFMO5 (in storage buffer and dodecyl-β-d-maltoside (0.03% (w/v))) was incubated with crystallization conditions of HEPES buffer (0.1 M, pH 6.9) and PEG 4000 (9%) at 20 °C with a ratio of 1:1 in a sitting drop. The sitting drop was 2 µl after mixing and the reservoir was 1 ml. Prior to crystallization, NADP⁺ (1 mM final) was incubated with 12–15 mg ml⁻¹ of AncFMO5 for 1 h at 4 °C. After 1 d, large yellow hexagon-shaped crystals formed. During crystal fishing, a cryo-protectant was prepared containing modified crystallization conditions with 20% glycerol and PEG 4000 (15%). Data were collected at the European Synchrotron Radiation Facility (Grenoble, France) and the Swiss Light Source (Villigen, Switzerland) and processed with the XDS and CCP4 packages. Aimless was used to merge the observations into average densities. STARANISO was additionally used for AncFMO2, which suffered greatly from anisotropy [Vonnrhein, 2018]. The phase problem was solved by molecular replacement using a recently solved insect FMO (PDB 5NMW) as a search model, and then AncFMO2 for the subsequent AncFMOs, using Phaser and Molrep. The phases were greatly improved by density averaging with DM. Model building and refinement were then conducted using COOT, Buccaneer and Refmac (see section 3.3). The outliers in the Ramachandran plots were 4.4%, 6.6%, 2.3% and 0.4% of the residues for AncFMO2, AncFMO2 bound to NADP⁺, AncFMO3 and AncFMO5, respectively. Figures were then generated using UCSF Chimera, PyMOL (DeLano Scientific; www.pymol.org) and CCP4mg.

2.9.2 AncFMO1

The crystallization condition that resulted in AncFMO1 crystals which displayed the highest diffraction is described below. Prior to crystallization, NADP⁺ (1 mM final) was incubated with AncFMO1 (12.5 mg ml⁻¹, in storage buffer conditions containing 0.03% (w/v) DDM) for 1 h at 4 °C. 1 µl protein-containing solution was then mixed with 1 µL 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 and CCP4 packages. Aimless was then used to merge the observations into average densities for two datasets to maximize the total number of observed reflections. STARANISO was used as, similarly to AncFMO2, these crystals suffered severely from anisotropy. 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. The phases were greatly improved by density averaging with DM. Model building and refinement were then conducted using COOT, Buccaneer and Refmac5. The residue outliers in the Ramachandran plots were 5.9% for AncFMO1. Figures were generated using UCSF Chimera, PyMOL (DeLano Scientific; www.pymol.org) and CCP4mg.

2.10 ANCESTRAL SEQUENCE RECONSTRUCTION

The initial FMO phylogenies that gave rise to the mammalian ancestral FMOs (AncFMOs) that are described in sections 3.3, 3.4 and 3.5 are described in the materials and methods section of 3.3. These trees were built by Dr. Maria Laura Mascotti. During the tree construction, emphasis was focused on the mammalian ancestral states leading to the synthesis of five AncFMOs: AncFMO1, AncFMO2, AncFMO3-6 (that represents the mammalian ancestor proceeding the gene duplication event between FMO3 and FMO6), AncFMO4 and AncFMO5. In this section, the work carried out during my secondment in San Luis, Argentina is described.

To build a thorough phylogeny that depicts the evolution of FMOs in jawed vertebrates, the tree used to generate the AncFMOs was selected for improvement. Human FMO sequences were used in homology searches: FMO1 (NP_002012), FMO2 (Q99518.5) FMO3 (NP_001002294.1) FMO4 (XP_005245102) and FMO5 (NP_001452.2). To ensure that a broad and representative phylogeny, sequences were collected in order to maximize sequence coverage among the tetrapod classes. When available, an FMO paralog was collected for each organism listed per class. The total number of sequences collected was

536 and they were aligned using MAFFT v7 resulting in a MSA of 537 sites. Amino acid substitution model and best-fit model parameters were obtained by the Akaike information criterion in ProtTest v3.4 [Darriba, 2011]. A substitution matrix using the JTT model was selected with a gamma value of 1.144. Phylogenies were inferred through Maximum Likelihood approaches including PhyML and RAxML and Bayesian inference by using MrBayes [Guindon, 2010; Stamatakis, 2014; Ronquist, 2012]. After analyzing each phylogeny in turn, the highest supported tree was carried forward. The phylogeny used by RAxML v0.6.0 (500 bootstraps) was submitted to the BOOSTER to transform the bootstrap values into transfer bootstrap expectation (TBE) values [Lemoine, 2018]. Ancestral sequence reconstruction was then carried out using PAMLX v.4.9. using the Maximum likelihood approach [Yang, 2007]. An empirical substitution matrix was employed alongside equilibrium amino acid frequencies (model = 3). During the reconstruction, four gamma rate categories were included and a JTT substitution matrix. Posterior probabilities were extracted for nodes, 541, 542, 604, 895 and 1007 for final reconstruction: Node 541 represents the first FMO ancestor to appear in tetrapods and is denoted as EldFMO1-5. Nodes 542 and 604 depict the ancestors of FMOs 1-4 and FMOs1-3/6, respectively, and are called EldFMO1-4 and EldFMO1-3/6, respectively. These sequences represent the ancestral intermediates giving rise to the FMO mode of action. Node 895 describes the first FMO5 sequence, denoted as EldFMO5. Node 1007 represents the first fish FMO, EldFMOF. Ambiguity was assigned to states that displayed posterior probabilities greater than or equal to 0.2. Constructs 541, 542, 604, 895 and 1007 showed 30, 46, 29, 19 and 59 ambiguously reconstructed sites, respectively.

3.RESULTS

3.1. ANCESTRAL FLAVIN-CONTAINING MONOOXYGENASE CLONING AND TRANSFORMATION

Each AncFMO was successfully cloned into modified pBAD-NK destination vectors possessing both N-terminal 6xHis and SUMO tags. The insertion of these two tags at the N-terminus provided multiple advantages. With FMOs being membrane bound, it is important to distinguish which regions, based on the sequence, are likely to embed within the membrane. All the FMOs display a large range of hydrophobic residues at the C-terminus, suggesting that they possess large transmembrane helices, similar to Monoamine oxidases [Binda, 2002]. With this in mind, we speculated that adding these tags on the N-terminus would most likely preserve their membrane-binding property, thus stabilizing the peptide scaffold and native conformation. Alternatively, the first residues of the sequence are easily compared to other Class-B monooxygenases, displaying the formation of a key Beta sheet involved in FAD binding. The His tag is a well-known affinity-based chromatographic feature that is relatively short (2 kDa) with high affinity for Ni-resin based columns. Furthermore, the SUMO tag (11 kDa) has been shown to improve protein solubility and may stabilize the system during expression and during purification. Importantly, the SUMO tag can be cleaved by SUMO protease, which in this construct means that both tags are removed.

After performing the Golden Gate cloning method (see materials and methods), the new construct was transformed in bacterial strains such as DH5alpha, TOP10 and BL21 (DE3) to determine whether they possessed antibiotic selectivity. Each AncFMO was successfully transformed in several *E. coli* strains and colonies selected for sequencing correctly displayed the correct amino acid chain. The constructs were then used for expression and purification studies.

3.2 ANCESTRAL FLAVIN-CONTAINING MONOOXYGENASE EXPRESSION, PURIFICATION & CRYSTALLIZATION

AncFMOs 1, 2, 3-6 and 5 could all be expressed in *E. coli* strains, whilst AncFMO4 did not display any expression signal. Both AncFMO2 and AncFMO3-6 were the first constructs to be explored and paved the way for AncFMO1 and AncFMO5 preparations (see later). In line with conditions employed for a bacterial FMO previously characterized, we

expressed both constructs (AncFMO2 and AncFMO3-6) at 24 °C using arabinose (0.02% (v/v) final) as inducing agent with cell culture densities portraying OD₆₀₀ values of approximately 0.3. Typically, the expression would produce approximately 10 g/L after 24 hours. Exchanging the flask however for baffled flasks increased the yield to approximately 20 g/L.

3.2.1 Initial Purification – Determining solubility of Ancestral FMOs

Considering the protein was likely membrane bound, cell lysis was performed using a pressurized homogenizer, Emulsiflex, that breaks the cells using pressure rather than vibrations generated via sonication. It is suggested that the latter technique is too aggressive, breaking not only cell walls but also membranes that may consequently damage the FMOs [Hjelmeland, 1990]. Henceforth, throughout the purifications, Emulsiflex based cell lysis was conducted (see material and methods). Lysed cells were then spun down to concentrate the native membranes for solubilization. It was important to determine whether the protein resided in the aqueous layer, or in the membrane fraction. The supernatant representing the soluble, aqueous fraction was submitted to SDS-PAGE Gel electrophoresis alongside a fraction of the membrane pellet which had been re-homogenised and incubated with SDS (1%) (Figure 3.2.1.1). The gel indicated that there were two bands present at approximately 73 kDa, potentially corresponding to the SUMO-His tagged AncFMO3-6, in both the soluble and insoluble fractions, respectively.

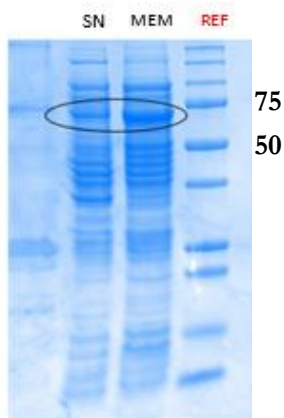


Figure 3.2.1.1: SDS-PAGE evaluating the solubility of AncFMO3-6. Both membrane fractions (solubilized in Triton-X 100) and aqueous fractions were run on a gel, represented as MEM and SN, respectively. Both lanes conveyed a 73 kDa band that may represent the enzyme of interest. Values 75 and 50 depict Molecular weights shown in the REF lane (reference markers) in kDa.

3.2.2 His-trap Nickel 1 Purification

Both fractions were purified using a Ni-resin based gravity column. With the soluble fraction, after several increasing imidazole wash steps, no yellow coloured protein after concentrating (data not shown) was observed. This suggested that the protein may have been poorly folded as it did not retain FAD. Moreover, the purification of the insoluble fraction in the presence of the soft detergent, reduced Triton-X 100 (TRX-100-R,) resulted in the accumulation of a large quantity of a yellow coloured solution that eluted with a high concentration of imidazole (300 mM). This finding suggested the presence of a Flavoprotein with a high affinity for a Ni-resin. Furthermore, running an SDS-PAGE of this sample corroborated the presence of an overly expressed protein with our desired weight (Figure 3.2.2.1). The high affinity of the protein for the column, the extent of expression followed by the yellow colour, implicates that the band corresponds to the desired AncFMO. Thus, our results show that the protein is indeed membrane-bound. Importantly, this result illustrates that the ancestor (AncFMO3-6) continues to display the membrane-binding features exhibited by that of the current mammalian FMOs. With regards to AncFMO2, the protein was not subjected to this preliminary assay and was assumed to be present in the membrane fraction; this was validated after the first Ni purification step (Figure 3.2.2.1). Overall, the protein is well expressed with clear FAD

spectral features visible. The yields after the first purification step for AncFMO3-6 and AncFMO2 were similar, approximating between 3.0 and 5.0 mg.

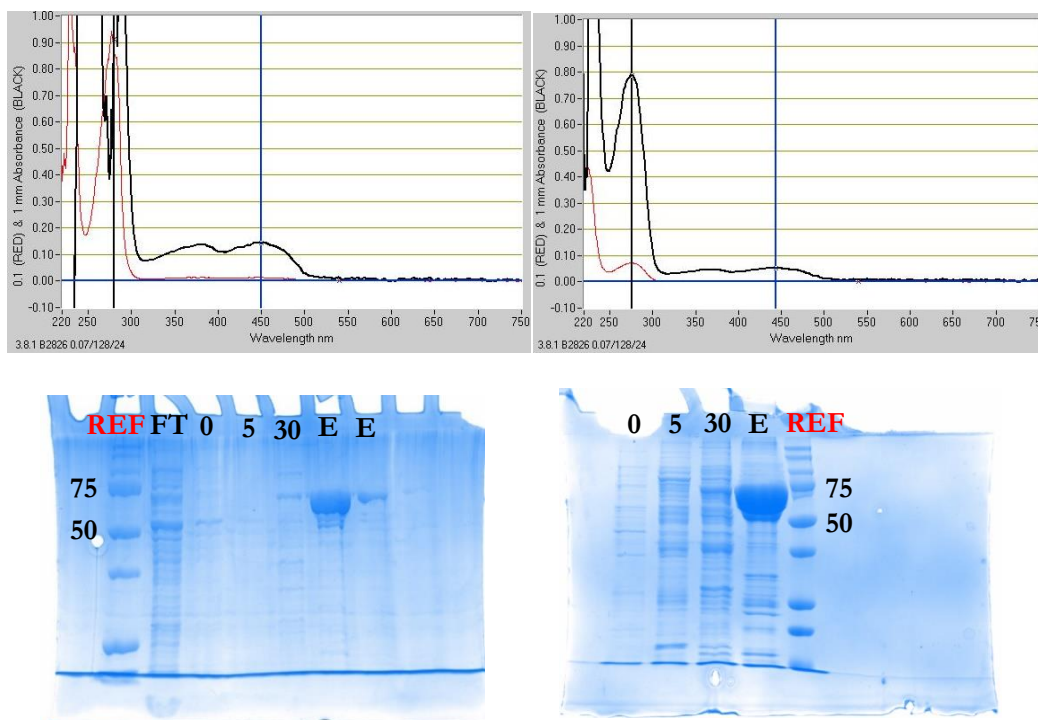


Figure 3.2.2.1: Spectral features and purity after the first step of affinity chromatography purification for AncFMO2 and AncFMO3-6. Left and right panels describe the UV/Vis spectra and SDS-PAGEs for AncFMO2 and AncFMO3-6, respectively. Each spectrum illustrated distinct FAD features with absorption maxima at 449 and 444 nm for AncFMO2 and AncFMO3-6, respectively. Gels portray the washing steps performed during the purification and the purity obtained. REF refers to the standard molecular weight markers with 75 and 50 being described in kDa. FT is the Flow through. Values 0, 5, 30 and E refer to imidazole washes of 0, 5, 30 and 300 mM, respectively.

3.2.3 Detergent Screening for efficient AncFMO extraction

After confirming that the proteins were membrane-bound, we analysed what detergents were able to sufficiently extract them from the membrane pellet. The following detergents were trialled for AncFMO3-6: TRX-100-R, dodecyl maltoside (DDM), Octyl glucoside (OG), Lauryl dimethylamine N-oxide (LDAO), Lauryl Maltose Neopentyl Glycol (LMNG) (Figure 3.2.3.1). All detergents were able to extract the protein and illustrated the range of different detergent-protein complexes that could form. This information may become useful when screening for protein crystallization conditions. With previous results in our laboratory showing good extraction of human FMO5 with TRX-100-R, we decided to use

this detergent for protein solubilization. Furthermore, its oxidized equivalent (TRX-100) (that absorbs at 280 nm) can also be used for extraction where needed.

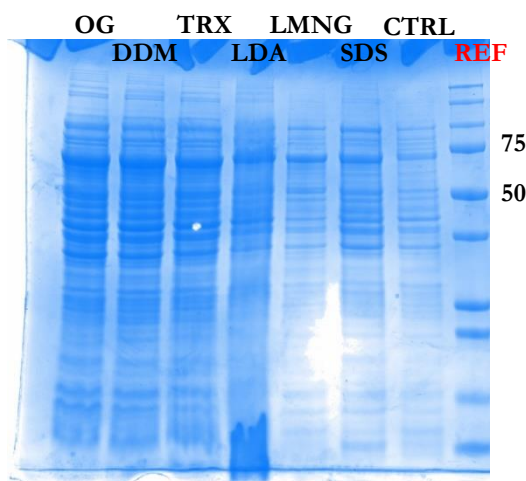


Figure 3.2.3.1: SDS-PAGE analysing the extent of protein extraction exhibited by a series of different detergents for AncFMO3-6. Octyl glucoside (OG); dodecyl beta-maltoside (DDM); Triton-X-100 reduced (TRX); Lauryl dimethylamine N-oxide (LDAO); Lauryl Maltose Neopentyl Glycol (LMNG); Sodium dodecyl sulfate (SDS); control membranes without addition of detergent (CTRL); reference markers (REF). Molecular weight markers indicated by values 75 and 50 expressed in kDa.

3.2.4 6xHis-SUMO Tag Cleavage and reverse Ni purification

Once the protein had been extracted from the membranes using TRX-100-R and purified using a Ni resin, the protein was incubated with SUMO protease overnight to remove the 6xHis-SUMO-tag. After incubation, the solution was transferred to an Akta purification system and run on a 5 mL His-trap, Ni-resin based column. Purifications were performed in step-by-step imidazole washes at 0-, 5-, 30- and 300-mM concentrations in a manner consistent with the initial purification step. Both AncFMO2 and AncFMO3-6 did not elute in the Flow through and required imidazole for elution. This suggests that the enzyme has some native affinity towards the resin. The resulting chromatograms are shown below (Figure 3.2.4.1) for both AncFMO3-6 and AncFMO2. During the purification procedure, a small amount of imidazole (5 mM) was needed to elute the proteins and produced broad peaks. Moreover, after increasing imidazole concentrations to 30 mM, additional AncFMO eluted. For AncFMO3-6 a large peak would elute homogeneously whilst this feature was not seen with AncFMO2 that produced a very small peak (Figure 3.2.4.1). Importantly, the SDS-PAGE confirmed that His-SUMO-tag cleavage was performed on both systems, and

that the first fraction eluted with very high purity (Figure 3.2.4.1). Regarding the extensive purity observed for this fraction, this solution was collected and taken forward for homogeneity assessment using Size exclusion chromatography and eventual crystallization screenings; the second peak (II) is concentrated down and stored at -80 °C for eventual kinetic assays. The yields after this step decreased slightly from the previous purification due to the protein eluting in separate peaks and slight precipitation overnight, but the purity improved significantly. The ratio between 280 and 450 decreased from approximately 15:1 after Ni1, to roughly 12/11:1 for Ni2. A ratio of approximately 10:1 or less is considered to be a very pure flavoprotein sample.

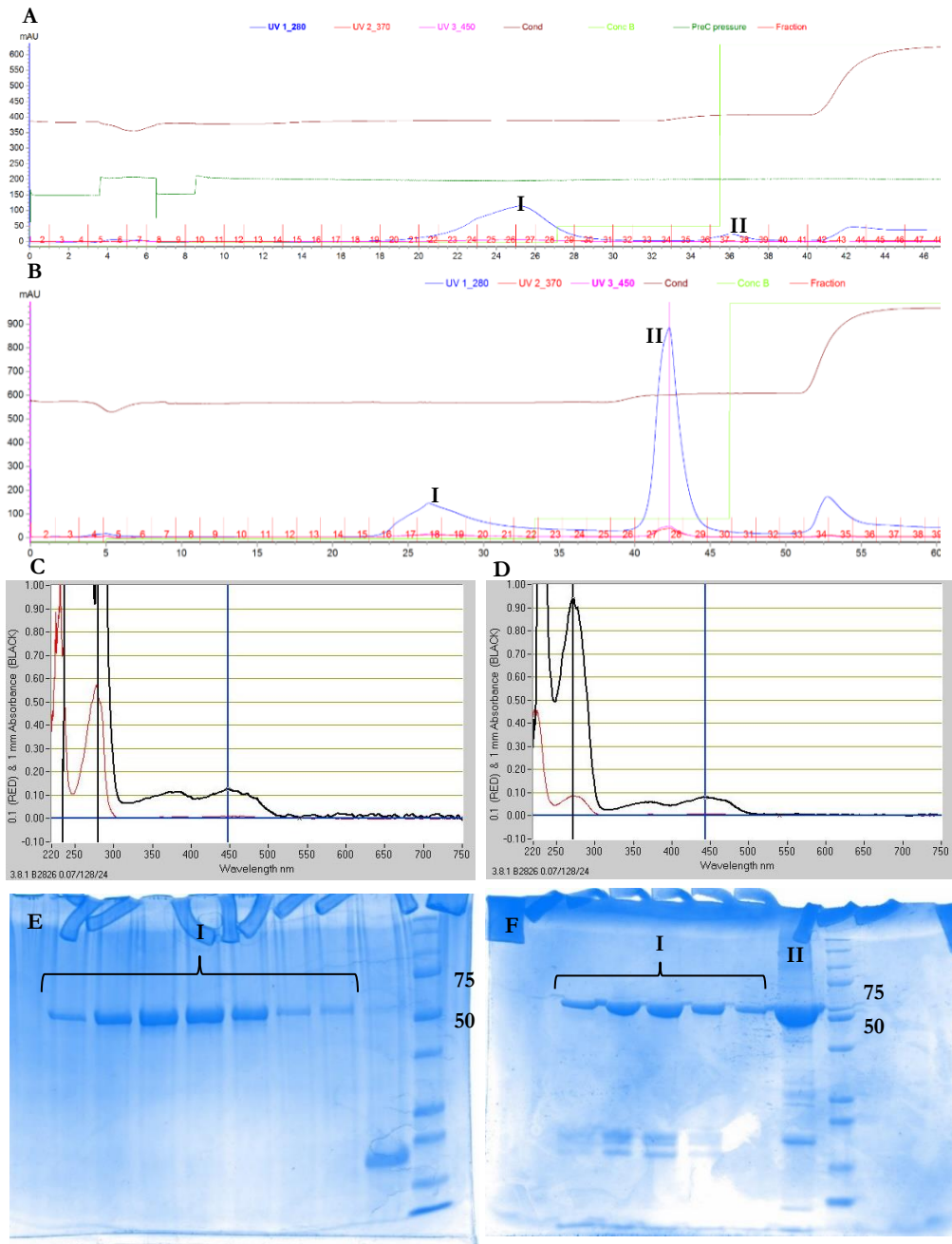


Figure 3.2.4.1: Ni-based affinity chromatography step for the purification of SUMO-His-tag cleaved AncFMO2 and AncFMO3-6. Chromatograms A and B show the elution profiles for AncFMO2 and AncFMO3-6, respectively, with peaks I and II depicted. Peaks I were concentrated down and displayed classical Flavoprotein UV/Vis spectra for both AncFMO2 and AncFMO3-6 as portrayed in panels C and D, respectively. SDS-PAGES of both peaks I and II from each chromatogram, A and B, respectively, document their high purities. Elution Chromatograms show the wavelengths 280, 370 and 450 nm depicted in blue, red and pink lines, respectively. Pressure, conductivity and percentage of elution buffer (300 mM imidazole) used are shown in colours dark green, brown and light green, respectively. Values 75 and 50 represent standard molecular weights in kDa.

3.2.5 Size Exclusion Chromatography for AncFMO2 and AncFMO3-6

Once the His-SUMO tag cleaved protein was collected and concentrated, the protein was incubated with FAD (ca. 300 μ M) for around 40 minutes to ensure that any lost FAD during purification was restored. The solution was then passed through a Superdex 200 10/300 GL column to evaluate the homogeneity. At this stage the detergents and buffers were exchanged, from TRX-100-R to several other detergents including, DDM and CYMAL-6, and from a high salt concentration to a low salt concentration. The change of detergent is executed to ensure that the protein is solubilized in a highly pure and homogenous system to promote crystallization. The low salt concentration is based on the speculation that high salt conditions interfere with the ionic and polar non-covalent interactions established between the protein and detergent micelle. The peak elutes at a volume of 10.2 mL with high homogeneity for both AncFMO3-6 and AncFMO2, respectively with high purities (Figure 3.2.5.1). The elution volume was unable to accurately predict the oligomerization state of the protein as the detergent micelle influences the overall size and shape of the complex that is able to pass through the resin mesh. Nevertheless, it reflects results observed with human FMO5 thereby implying that the Ancestral constructs represent good models of the FMOs [Fiorentini, 2016]. After purification, the sample was concentrated down to 12 mg/mL for crystallization screenings (measured using the FAD absorption at 445 nm). Typically, after the final purification step the yield ranged from 2 – 3 mg, with ratios between the 280 and 450 varying between 8 and 10, further highlighting the purity and uptake of the FAD.

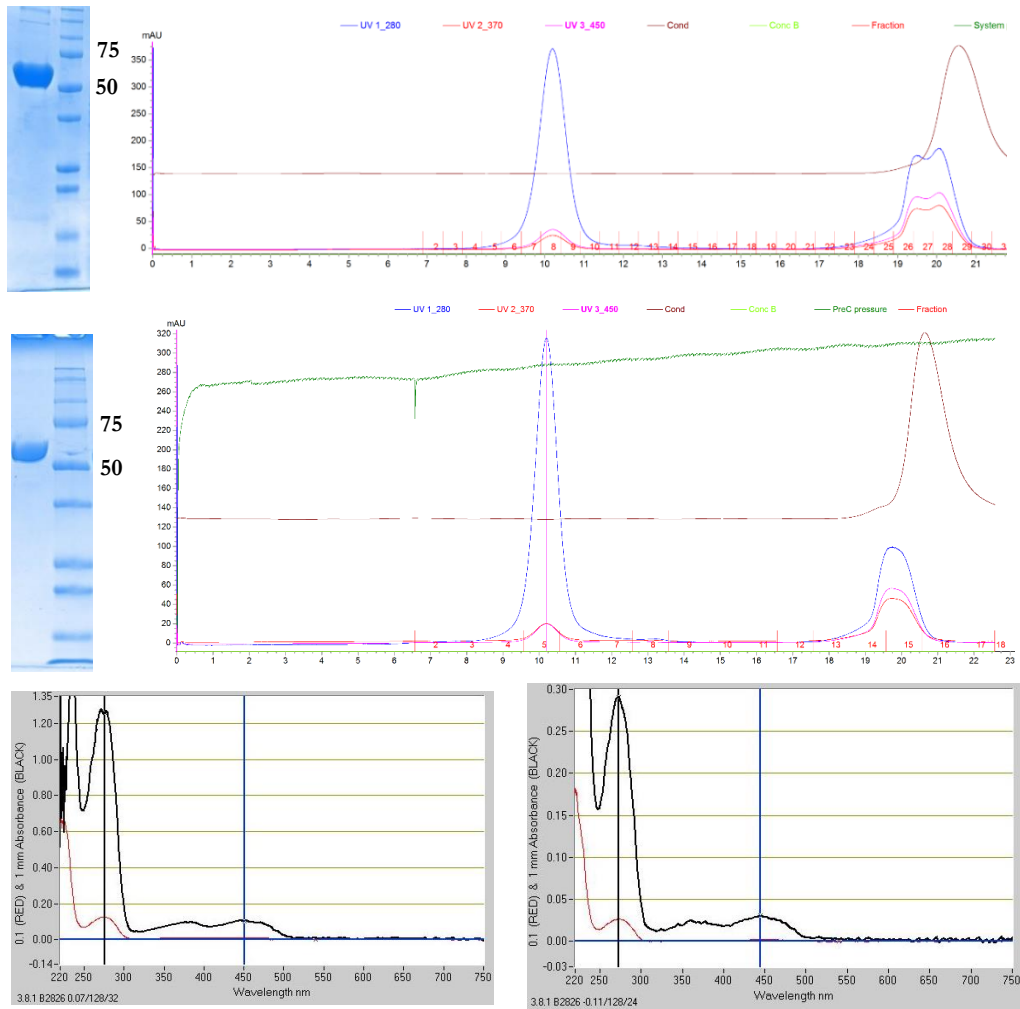


Figure 3.2.5.1: Gel filtration purifications for AncFMO2 and AncFMO3-6. Top panels show elution chromatograms for AncFMO3-6 (top) and AncFMO2 (bottom). Each chromatogram has their corresponding purities illustrated on an SDS-PAGE. Molecular weights are described in kDa with respect to the reference lane. Final UV/Vis spectra for both AncFMO2 and AncFMO3-6 after concentrating are shown on the left and right, respectively. Traces coloured blue, red and pink in the chromatogram describe absorptions at wavelengths of 280, 370 and 450 nm, respectively. Traces in dark green and brown represent the pressure (Mpa) and conductivity (mS/cm) during the purification, respectively.

3.2.6. Purification of AncFMO5

With the purification protocols of AncFMO2 and AncFMO3-6 being well established, the method was applied to AncFMO5. Surprisingly, the yield of protein was very low and concentrated eluents from the initial purification step failed to give rise to an FAD spectrum or distinctly overexpressed molecular weight band in an SDS-PAGE (Figure 3.2.6.1). With such low quantities being obtained, with no substantial enhancement when using alternative buffers, the expression conditions were reconsidered.

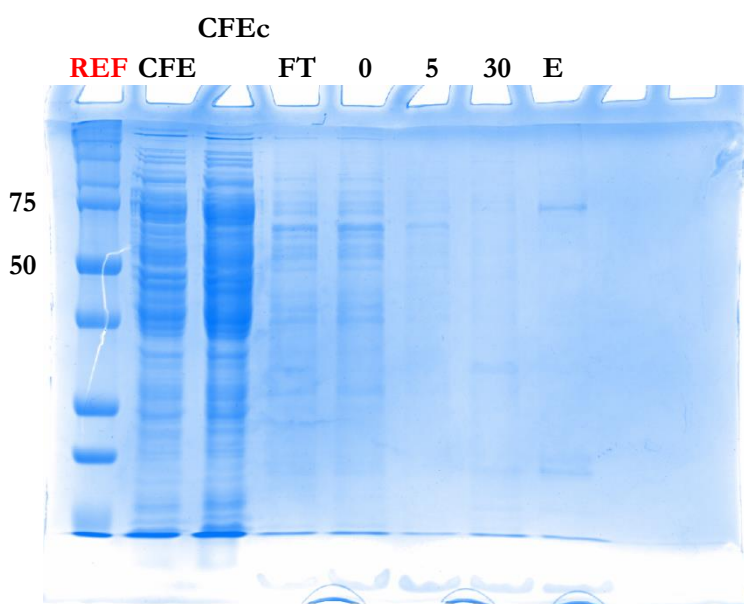


Figure 3.2.6.1: SDS-PAGE of AncFMO5 Ni1 purification. Cell free extract (supernatant before membrane solubilization) (CFE); concentrated CFE (CFEc); Flow through (FT); 0 mM Imidazole (0); 5 mM Imidazole (5); 30 mM Imidazole (30); 300 mM Imidazole (E).

AncFMOs 2 and 3-6 were previously expressed at 24 °C in baffled flasks to ensure high oxygen uptake and improve yields. However, we hypothesized that the protein was more unstable than its ancestral paralogs. Therefore, the enzyme was expressed at 17 °C, to promote protein folding. Remarkably, this change alone greatly improved the yield and after a Ni purification step, the canonical flavin spectrum was detectable (Figure 3.2.6.2). As with all AncFMO purifications, there was always a heme-containing contaminant that at lower yields was very clear (data not shown). For AncFMO5, as shown in Figure 3.2.6.2, the FAD UV/Vis is tainted by the appearance of a Heme absorption at 414 nm. Based on

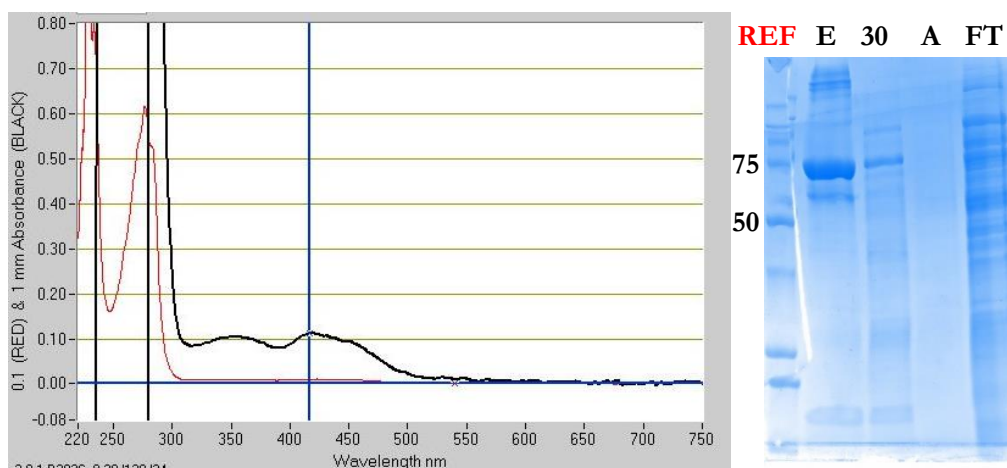


Figure 3.2.6.2: UV/Vis spectrum and SDS-PAGE after Ni1 purification of AncFMO5 expressed at 17 C. Left and right panels describe the UV/Vis spectra and SDS-PAGEs for AncFMO5, respectively during the first affinity-based chromatography purification step. SDS-PAGE portrays the washing steps performed during the purification and the purity obtained. REF refers to the standard molecular weight markers with 75 and 50 being described in kDa. FT is the Flow through. Values 0, 5, 30 and E refer to imidazole washes of 0, 5, 30 and 300 mM, respectively.

this initial result, further purifications were required to reveal the complete FAD spectrum and to accurately deduce the yield.

The His-SUMO tag was cleaved using SUMO protease and the protein was submitted for a second Ni-based affinity chromatography purification. The chromatogram showed some slight differences between the purifications previously observed for the other AncFMOs (Figure 3.2.6.3). Usually, the AncFMOs require some additional imidazole in order to elute the protein from the column. However, for this construct, we observed that the protein elutes out in the flow through with very high purity (Figure 3.2.6.3, peak I). Albeit, significant quantities of the protein eluted out in the 30 mM imidazole wash (Figure 3.2.6.3, peak II). Tempting as it seemed to combine all the protein and fractions to maximize yield (with no cost of adding impurities,) the fractions were kept separate. We speculated that the protein may occupy two different oligomerizations states that consequently dictate the ability to adhere to the resin. Therefore, protein eluting out in peak II were kept for kinetic assays, whilst peak I was subjected to size exclusion chromatography. Overall, the yield of AncFMO5 ranged between 1 and 2 mg after this step, significantly less than AncFMOs 2 and 3-6.

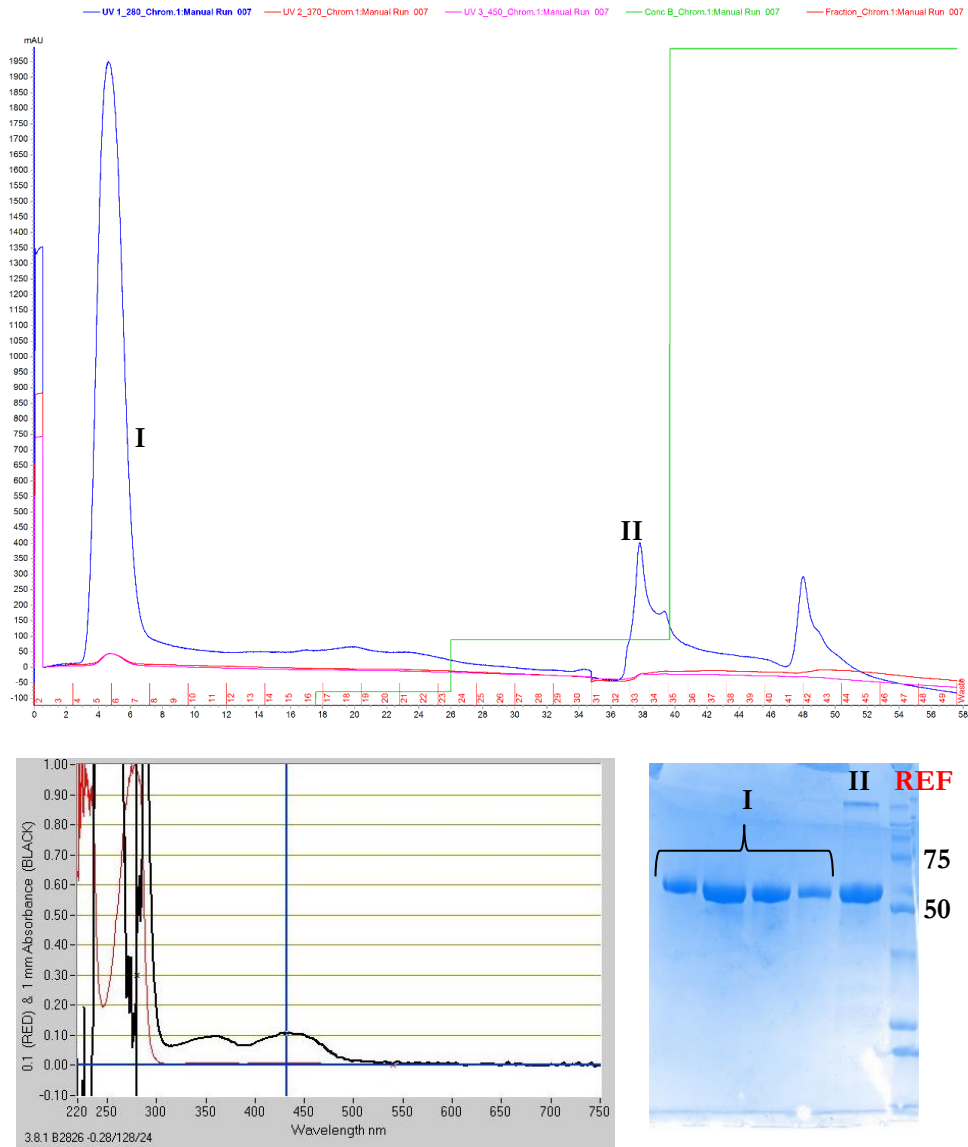


Figure 3.2.6.3: Second Ni affinity-based chromatography for the His-SUMO-tag cleaved AncFMO5. The chromatogram describing the elution profile is shown above with blue, red and pink traces representing the wavelengths at 280, 370 and 450 nm, respectively. The trace in light green portrays the stepwise gradient of buffer B (300 mM imidazole) used as washing steps during purification. The bottom panels represent the final UV/Vis spectrum of the concentrated peak I and the SDS-PAGE of both I and II peaks, shown on the left and right, respectively. REF represents the molecular weight markers with values 75 and 50 indicating molecular weight bands in kDa.

AncFMO5 was subject to a gel filtration purification step to ensure exchange between the

detergents, Triton-X 100 and DDM, but also buffer exchange. The gel filtration for AncFMO5 was very similar to the AncFMOs 2 and 3-6 with some minor discrepancies. Firstly, the elution volume for the protein was shifted slightly from 10.2 mL to 11 mL, respectively, implying that the shape and size of the micelle-protein complex may be slightly smaller. Nevertheless, AncFMO5 elutes with high homogeneity as a monodisperse peak (Figure 3.2.6.4).

Secondly, after the main peak had eluted, a large intense peak with an elution volume of around 13.7 mL appeared. Running these fractions on an SDS-PAGE did not confer any molecular weights (data not shown) and the lack of absorption at wavelengths 370 and 450 nm implied that there was no FAD present. The high purity observed after the second Ni purification suggests that the band cannot be derived from an impurity. However, the original detergent used during this preparation, Triton-X 100, absorbs intensely at 280 nm. We speculate that in the previous preparations, the use of the reduced equivalent of Triton-X 100 (Triton-X 100 R) meant that this peak was not observed during gel filtration as it does not absorb at 280 nm. Thus, we propose that Triton-X 100 (non-reduced) produces the band observed at 13.7 mL. This finding serendipitously displays the detergent exchange during the purification.

After concentrating the protein, a distinct FAD spectrum was observed with yields ranging from 0.5 to 1.5 mg (Figure 3.2.6.4). During the concentration procedure of the system, noticeable precipitation was observed. Despite being able to observe a high 280:450 nm ratio, 9:1, the protein portrayed noticeable instability in these storage conditions, precipitating after 12 hours in ice. This result indicated that AncFMO5 was far less stable to the procured storage buffer conditions with respect to its paralog partners. We speculate that the change in detergent and/or the dramatic decrease in salt concentration (250 mM to 10 mM) may have encouraged denaturation.

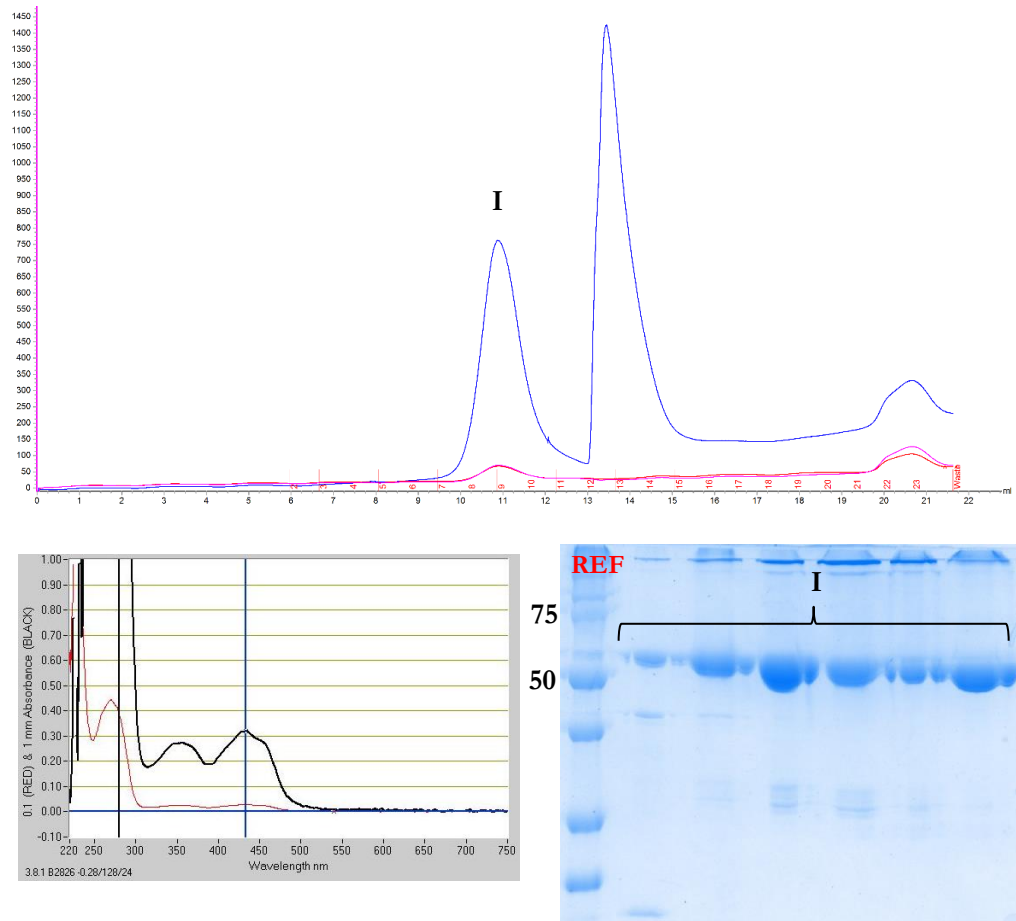


Figure 3.2.6.4: Gel filtration of AncFMO5. The chromatogram describing the elution profile is shown above with blue, red and pink traces representing the wavelengths at 280, 370 and 450 nm, respectively. The bottom panels represent the final UV/Vis spectrum of the concentrated peak I and its respective SDS-PAGE, shown on the left and right, respectively. REF represents the molecular weight markers with values 75 and 50 indicating molecular weight bands in kDa.

3.2.7 Purification of AncFMO1

Initial purifications were conducted with the enzyme expressed at 24 °C, similarly to the AncFMOs 2 and 3-6. Unlike AncFMO5 that was significantly influenced by this higher expression temperature, AncFMO1 was visible after the initial purification step with a distinct flavin spectrum (Figure 3.2.7.1). Furthermore, the yield of protein was considerably higher ranging from 5 to 7 mg after the first Ni column. Once concentrated and incubated with SUMO protease, the protein was run on a His-trap Ni-resin. Consistently with AncFMOs 2 and 3-6, the protein eluted out with the addition of 5-6 mM imidazole and at 30 mM imidazole. Similarly, to AncFMO2, the majority of the protein eluted in the in the first peak, with very little in the second (Figure 3.2.7.2).

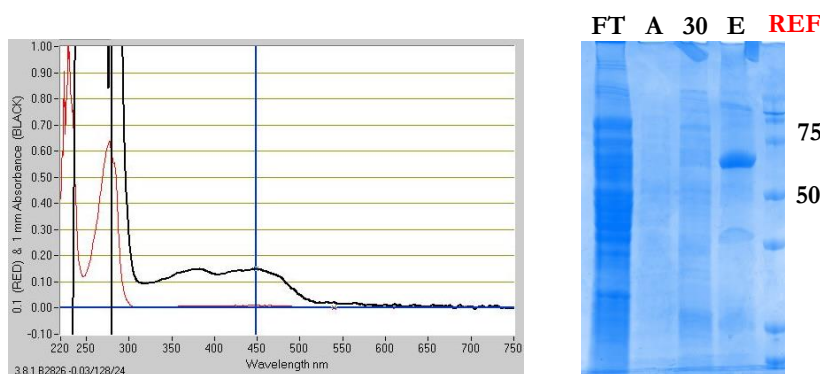


Figure 3.2.7.1: Initial affinity-based chromatographic step of AncFMO1 using a Ni resin. The UV/Vis spectrum of the concentrated eluent, E and the imidazole washes conducted are displayed in an SDS-PAGE. Lanes FT, A, 30, E and REF refer to flow-through, buffer A, 30 mM imidazole, elution buffer (300 mM imidazole) and reference markers, respectively. Values 75 and 50 show the molecular weight bands given in kDa.

The well expressed ancestral FMO however produced issues during the final purification step. During gel filtration, the exchange of detergent from Triton-X 100 to DDM did not take place (indicated by the high 280:450 ratio), and the resulting elution volume was higher, ca. 12.8 mL (Figure 3.2.7.3). Despite high purity (not shown) the protein did not crystallize in conditions similar to the other AncFMOs (see section 3.2). Considering the different elution volume after gel filtration, it was likely that the system was different. We speculated that exchanging the detergents was no longer viable during gel filtration and that the entire

protein purification should be conducted in one detergent. Furthermore, we hypothesized

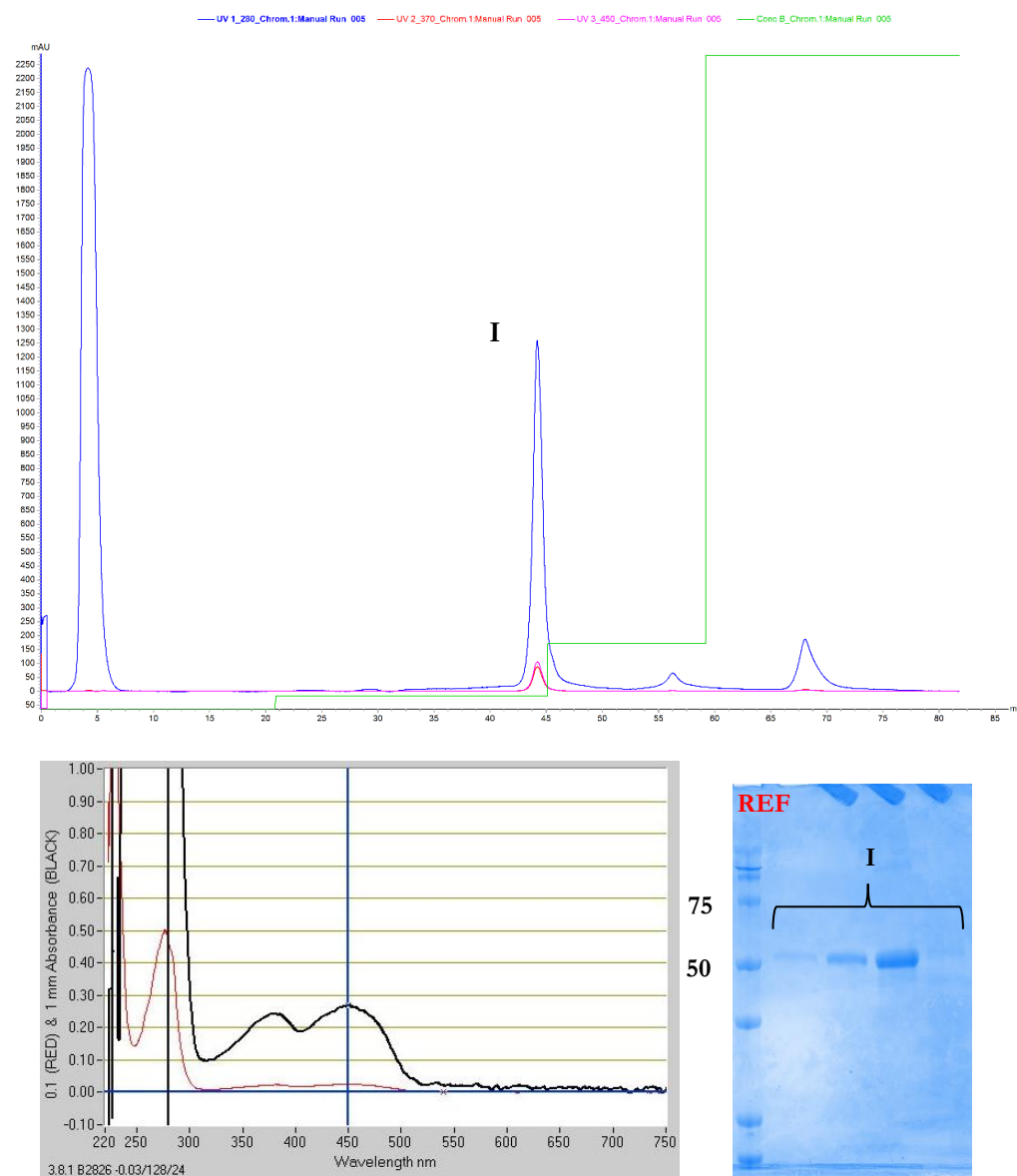


Figure 3.2.7.2: AncFMO1 Ni² purification after 6xHis-SUMO cleavage. The chromatogram describing the elution profile is shown above with blue, red and pink traces representing the wavelengths at 280, 370 and 450 nm, respectively. The green trace highlights the concentration of buffer B (possessing 300 mM imidazole) used in a stepwise fashion. The bottom panels represent the final UV/Vis spectrum of the concentrated peak I and its respective SDS-PAGE, shown on the left and right, respectively. REF represents the molecular weight markers with values 75 and 50 indicating molecular weight bands in kDa.

that perhaps the protein was not soluble in DDM and required alternative detergents

AncFMO1 native membranes were collected and separated into a series of different detergent-containing solutions and mixed overnight for efficient solubilization. The detergents inspected were: octyl-glucoside (OG) which has a significantly shorter aliphatic tail compared to DDM (8 versus 12 methylene groups, respectively), CYMAL-6, DDM, to see if it could extract the protein from within the membranes, Triton-X 100 as a control, Sodium dodecyl sulfate (SDS) which acts as a harsh detergent that will extract almost everything from within the membranes, and finally, a control being without detergent

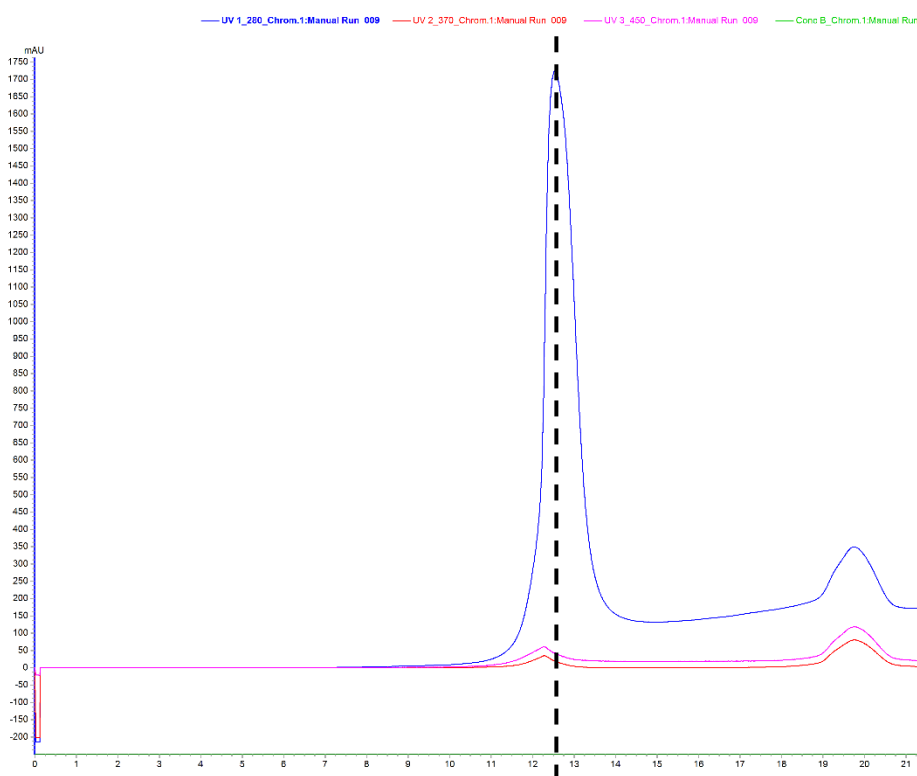


Figure 3.2.7.3: Size exclusion chromatography purification step of AncFMO1 solubilized in DDM (0.03% w/v). The dashed line highlights the elution volume of the protein, 12.6 mL. Traces shown in blue, red and pink depict wavelengths of 280, 370 and 450 nm, respectively.

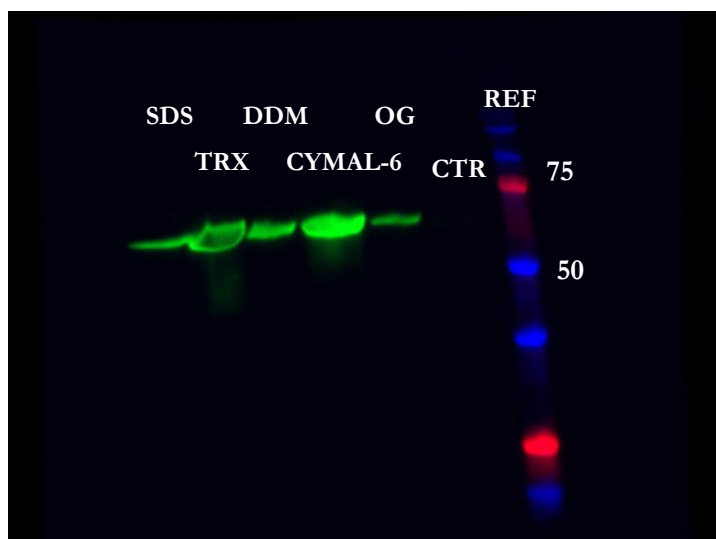


Figure 3.2.7.4: Western-blot detergent screening for AncFMO1 extraction. Native membranes collected during purification were resuspended and incubated overnight with various detergents including sodium dodecyl sulfate (SDS), Triton-X 100 (TRX), dodecyl beta-maltoside (DDM), CYMAL-6, octyl glucoside (OG). CTRL 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.

addition. All detergents were able to extract AncFMO1 from the membrane and was observable through western-blotting utilizing a 6xHis-tag antibody (Figure 3.2.7.4). The western-blot confirmed that the shorter length detergent, OG, is not as efficient at protein extraction. Despite western-blotting being qualitative, Triton-X 100 and CYMAL-6 are the best for extraction. With FMO5 successfully crystallizing in DDM (see section 3.2), we pursued the entire purification first with DDM. Similar yields were obtained (not shown) but the fundamental bottleneck we previously observed persisted. The gel filtration chromatogram mirrored the before seen results obtained with Triton-X 100 as the protein solubilizer (Figure 3.2.7.5). This suggested that the problem may lie entirely with DDM as the detergent of choice. Therefore, the purification was performed solely with Triton-X 100 reduced to see whether the same elution profile was observed. Indeed, the same results were obtained suggesting the protein itself presents these issues and not the detergent. After performing melting temperature assays (see section 3.5) on purified AncFMO1, it was noticeable that the protein possessed a lower melting temperature than AncFMO2 and AncFMO3-6, whilst similar to AncFMO5. This urged us to attempt to express the protein at a lower temperature, to validate whether poor protein folding is the defining property that inhibits detergent exchange.

The Structural and Functional Characterization of mammalian Flavin-containing Monooxygenase
using Ancestral Sequence Reconstruction

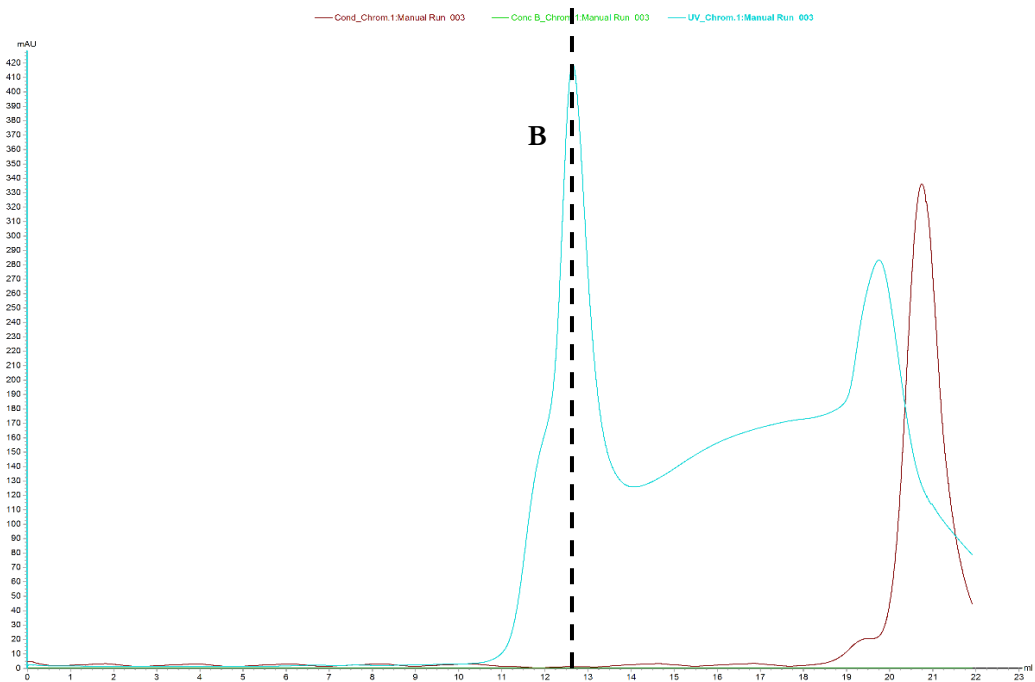
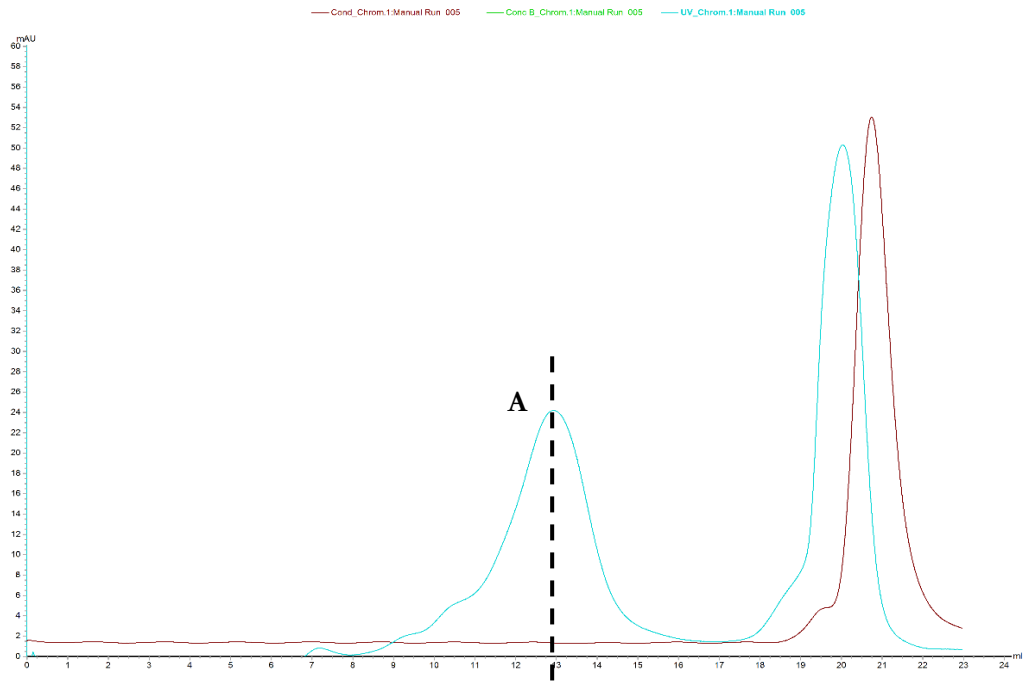


Figure 3.2.7.5: Gel filtration chromatograms of AncFMO1 grown at 24 °C. Upper and lower chromatograms describe the elution profiles of AncFMO1 in the presence of DDM (0.03% w/v) and Triton-X 100 R (0.72 mM), respectively. Peaks A and B have elution volumes of 12.9 and 12.6 mL, respectively. Traces shown in light blue and brown represent the absorption at 280 nm and the conductivity, respectively.

AncFMO1 was expressed at 17 °C and displayed protein yields comparable to higher expression temperatures (data not shown). The enzyme was solubilized with Triton-X 100, purified as previously described and then loaded onto a gel filtration column pre-equilibrated with DDM and the storage buffer. To our surprise, the protein eluted as two oligomers with peaks overlapping during elution (Figure 3.2.7.6). The larger peak elutes out with an elution volume slightly lower than observed previously, ca. 12 mL (compared to 12.8 mL). Intriguingly, the smaller peak that shoulders the former, elutes at ca. 10.8 mL which is very similar to the previous elution volumes portrayed by the other AncFMOs (10.2 and 11 mL). Therefore, these new expression conditions promote the formation of the ‘canonical’ AncFMO-detergent complex. Regarding crystallization, it was important to extract and pool the protein that possessed the correct oligomeric state for crystallization. However, the closely eluting species could not easily be separated based on the fractionation. Nevertheless, we postulated that if the systems obeyed a dynamic equilibrium and were therefore interchangeable between each other, it could be possible that as one crystallizes out of solution (oligomer A), the alternative species (oligomer B) would be pulled into the alternative oligomer (A) and crystallize further (Equation 3.1).



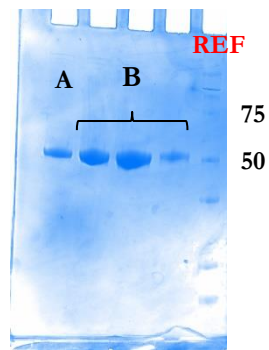
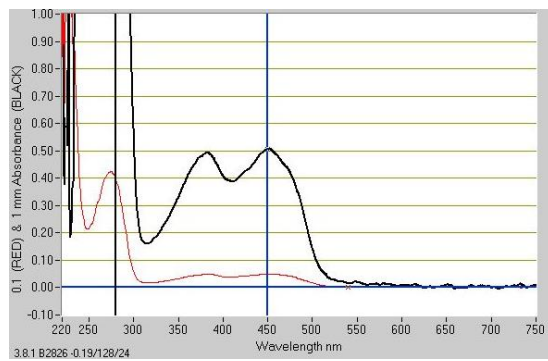
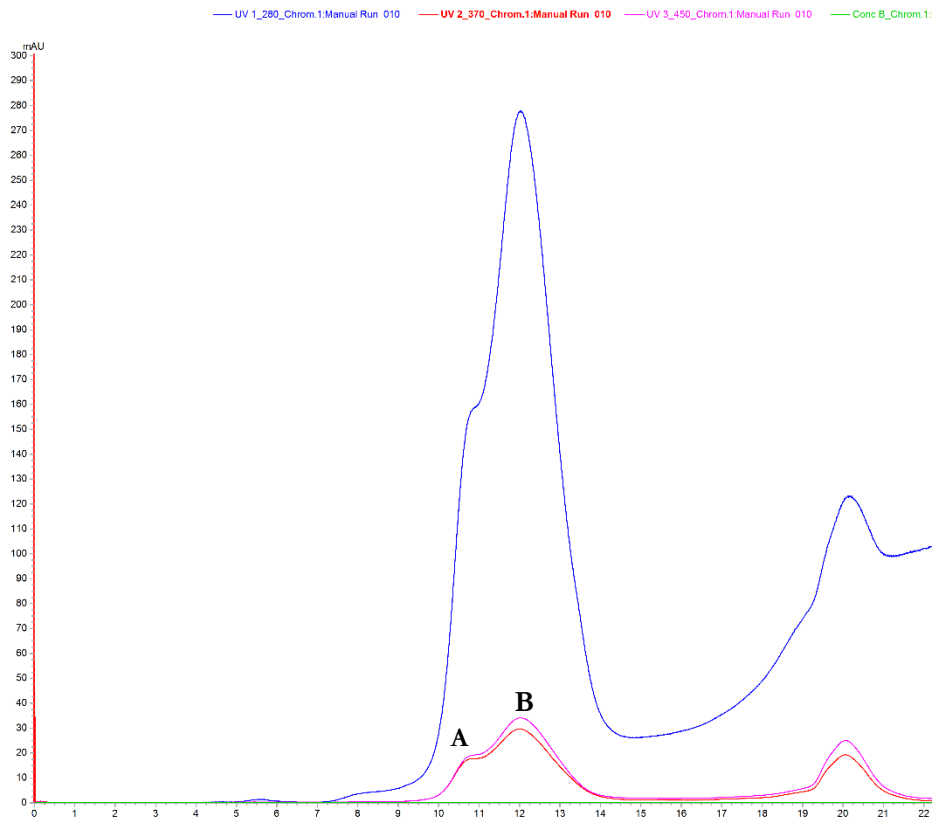


Figure 3.2.7.6: AncFMO1 gel filtration after protein expression at 17 °C. The top panel shows the chromatogram representing the elution profile. The blue, red and pink traces portray the absorptions at wavelengths 280, 370 and 450 nm, respectively. The bottom panels represent the final UV/Vis spectrum of the concentrated peaks A and B with their respective SDS-PAGE, shown on the left and right, respectively. REF represents the molecular weight markers with values 75 and 50 indicating molecular weight bands in kDa.

3.2.8 AncFMO4 expression evaluation

This system was arguably the most exciting out of all the ancestors considering its lack of enzymatic characterization and poor expression levels. Collaborators at the University of Groningen cloned the construct into identical vectors used previously and were able to successfully transform various bacterial strains. However, expressing the enzyme at 24 °C in baffled flasks resulted in a lack of expression and no protein was obtained during the purification. With this in mind, we decided to express the protein at 17 °C, similarly to AncFMO1 and AncFMO5, to see if protein expression was encouraged. Once the cells were collected and lysed using a pressure homogenizer, the lysates were spun down. The resulting supernatant was collected and stored, whilst the membrane fraction was resuspended and incubated with various detergents to see whether protein extraction required different detergents than previously employed. These detergents included: TRX-100, OG, DDM, FOS-Choline 8, glyco-diosgenin (GDN), Lauryldimethylamine oxide (LDAO), dimethyldecylphosphine oxide (DMDPPO) and SDS. After solubilization overnight, all solutions were spun down and ran on an SDS-PAGE to see if there was any sign of overexpressed protein at the expected molecular weight (Figure 3.2.8.1). Including the supernatant that consists of aqueous components, there was no sign of the system. These fractions were then exposed to 6xHis-tag antibodies and a western blot was performed. With this technique being very sensitive, if the protein was expressed, it should be detected. However, the blot could not detect any protein (Figure 3.2.8.1). The failure to observe any protein even at 17 °C, in addition to the lack of expression observed throughout the literature implies that ASR may not be enough to induce expression.

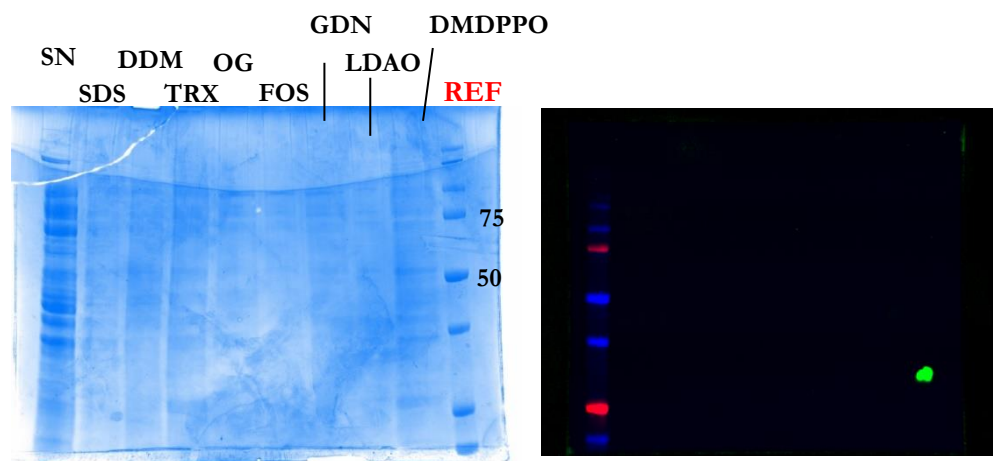


Figure 3.2.8.1: AncFMO4 expression and solubilization screening. Left, SDS-PAGE corresponding to the various detergents used to solubilize AncFMO4 from the membranes, including the aqueous supernatant. Right, Western blot of the SDS-PAGE, with the image flipped with the lanes in reverse. Detergents used; 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). RE refers to molecular weight markers with weights 75 and 50 indicated in kDa. For clarity, certain lanes are indicated with black lines.

3.2.9 Crystallization screenings, optimizations and structural elucidation

Herein this section, we will discuss the various crystallization conditions and optimizations explored and conducted on the AncFMOs. The results consist mostly on the pioneering crystallization screens performed on the first purified ancestral constructs AncFMO2 and AncFMO3-6 that paved the way for both AncFMO5 and AncFMO1. Due to their well conserved structural topology, the resultant crystallization conditions that led to the well diffracting crystals consisted of similar conditions with minor differences among each paralog. Subtle discrepancies between pH and the detergent used were all pivotal to the final optimized condition. Moreover, certain crystallization conditions resulted in highly anisotropic crystals that required anisotropic correction methods.

3.2.9.1 Initial hits, Detergent Screenings and AncFMO2 diffraction optimization

(a) AncFMO3-6

AncFMO3-6 was first exposed to a Classic Lite screening. This set of conditions contains various precipitants ranging from PEGs to AmSO₄, but at a relatively low concentration (0.2 – 1.0 M) compared to alternative screenings. After four days at 20 °C, slightly yellow crystals formed in the following conditions: 0.1 M HEPES.Na pH 7.5; 5% (v/v) Isopropanol; 10% (w/v) PEG 4000 (Figure 3.2.9.1.1). The crystals however raised concerns because they looked stacked, with multiple thin plates orientated upon one another, which can create issues during data collection and resolving phases. Removing isopropanol from the crystallization condition resulted in identical crystal formation, implying it was unnecessary for crystal nucleation.



Figure 3.2.9.1.1: Crystals of AncFMO3-6. Crystals formed after 4 days at 20 °C in 0.1 M HEPES.Na pH 7.5; 5% (v/v) Isopropanol; 10% (w/v) PEG 4000 conditions.

Other crystallization conditions that produced small yellow crystals include: KH₂PO₄ (0.05 M) and PEG 8000 (10% w/v), with 13.57 mg/ml of protein and NADP⁺ (1 mM) at 20 °C; MES (pH 6.5, 0.1 M) and PEG 20,000 (6% w/v), in the presence of NADP⁺ (1 mM) at 20 °C with 13.57 mg/ml protein (not shown). None of these crystallization conditions were reproducible.

The crystals shown in Figure 3.2.9.1.1 that formed in the condition: 0.1 M HEPES.Na pH 7.5; 5% (v/v) Isopropanol; 10% (w/v) PEG 4000, were protein-based crystals and diffracted to resolutions reaching 3.4 Å. These crystals therefore represented the first mammalian FMO protein crystals observed to date. Each crystal varied regarding their diffraction resolution. Multiple crystals had to be screened in order to obtain several well diffracting datasets. The low sequence identity observed for this construct (28% to a

bacterial FMO, PDB:2vq7 [Alfieri, 2008]) and relatively low-resolution halted phase determination using molecular replacement techniques.

Current screening techniques were being applied to AncFMO3-6 solubilized in DDM (0.03% (w/v)). In an attempt to improve the crystal packing and density, a shorter detergent with a similar polar head group was selected, octyl-glucoside (OG). The hypothesis was based on the idea that the shorter hydrophobic tail would shrink the overall micelle size and promote protein-protein contacts. In similar conditions to those that diffracted to 3.4 Å (Figure 3.2.9.1.1), this protein-detergent complex could also crystallize (Figure 3.2.9.1.2). They grew quickly over a two-day period and formed small yellow crystals. However, despite their crystalline shape and yellow colour, protein diffraction only reached 5 Å after screening 60 crystals.

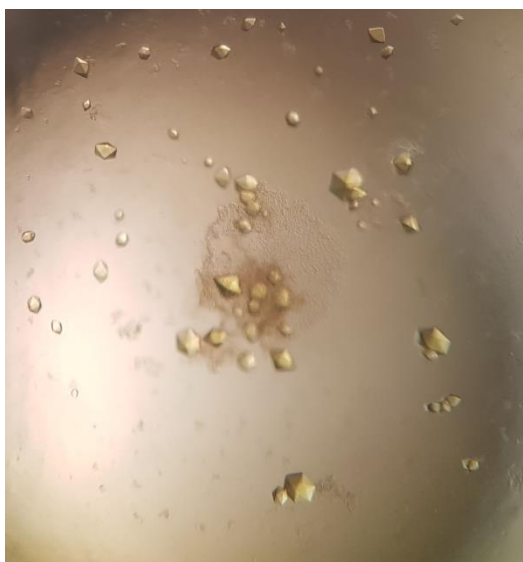


Figure 3.2.9.1.2: Crystals of AncFMO3-6. Crystals formed after 2 days at 20 °C in 0.1 M HEPES.Na pH 7.5; 5% (v/v) Isopropanol; 10% (w/v) PEG 4000 conditions in the presence of 20% OG.

(b) AncFMO2

With AncFMO3-6 producing several crystals in multiple conditions in the Classic LITE screening, the same method was applied to AncFMO2. Intriguingly, the crystallization condition consisting of 0.1 M HEPES.Na pH 7.5; 5% (v/v) Isopropanol; 10% (w/v) PEG 4000 also resulted in yellow crystals, identical to that of AncFMO3-6 but with a higher resolution. The second condition that produced small, thick yellow crystals was (Figure 3.2.9.1.3): NiCl₂ (0.01 M), Tris-HCl (pH 8.5, 0.1 M), PEG 2000 MME (10% w/v). All these conditions mentioned did not contain any NADP⁺ but diffracted to 3.0 Å resolution. The

latter condition could not be reproduced and even at resolutions of 3.0 Å, molecular replacement was not sufficient to resolve the phase problem.



Figure 3.2.9.1.3: Crystals of AncFMO2. Crystals formed after 2 days at 20 °C in 0.1 M Tris pH 8.5; 0.01 M NiCl₂; 10% (w/v) PEG 2000 MME conditions.

With a crystallization condition giving rise to reproducible crystals for both AncFMOs 2 and 3-6, we decided to conduct a detergent crystallization screening to see whether a particular detergent could alter the crystal packing and promote the formation of better diffracting crystals. Essentially, a reproducible crystallization condition is chosen, and an excess of detergent is added, differing in each drop. The excess of the new detergent will promote the formation of a new protein-detergent micellular complex that may produce different crystals. The crystallization screening highlighted a range of different crystallization conditions that could be used for crystal growth. Detergents that promoted crystallization include: Dodecyltrimethyl ammonium chloride, Cetyltrimethylammonium bromide (CTAB), ANAPOE-C12E9, ANAPOE-20, Polyethylene glycol 400 dodecyl ether, CYMAL-6, n-Undecyl-beta-delta maltoside, N,N-Dimethyldodecylamine N-oxide (DDAO). All crystals were fished and sent to synchrotrons for diffraction quality assessment. The only crystals that diffracted to high resolutions included ANAPOE-20, CYMAL-6, n-Undecyl-beta-delta maltoside and DDAO, however, they still could only diffract to approximately 3.0 Å.

The detergent screening method, whilst highlighting some alternative detergents, could not produce diffracting crystals with a high enough resolution. We speculated that perhaps the mixture of detergent in the crystallization condition might perturb the crystal packing and limit the resolution potential. With this in mind, we sought out to solubilize the protein in CYMAL-6 after gel filtration. CYMAL-6 possesses an identical polar head group to both DDM and OG, with an aliphatic hydrocarbon chain length (9 carbons long) intermediate to both DDM (12 carbons) and OG (8 carbons) (Figure 3.2.9.1.4). AncFMO2 crystallized

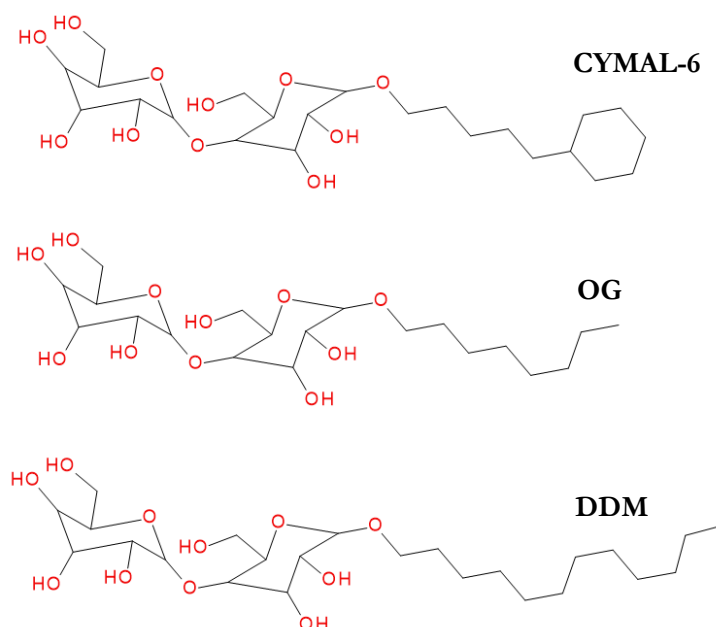


Figure 3.2.9.1.4: Detergents used for crystallization. Three detergents that were crucial for crystallization growth are shown. CYMAL-6, octyl-glucoside (OG) and dodecyl-beta-D-maltoside.

well and reproducibly in this new detergent and after multiple crystal screenings, we obtained a dataset of approximately 2.7 Å. Molecular replacement however failed to solve the phase problem even with this 0.3 Å improvement in resolution. Serendipitously however, a recently deposited PDB structure of an insect FMO, that possessed a higher sequence identity of roughly 32% (PDB:5nmw), became available for molecular replacement assessment. To aid the programme, we performed a technique named, chainsaw [Stein, 2008], which trimmed the reference model down to the key sequence-structure features that are predicted to highly resemble the protein in question (i.e. the FAD and NADP(H) binding domains). Collectively, the new model and chainsaw could resolve the phase problem, and we could finally see the electron density for our protein. This feat represented a large milestone for the structural biology underpinning this xenobiotic degrading family.

The newly collected datasets, moreover, consisted of a high degree of anisotropy. They contained regions of electron density that were poorly mapped and deducing the tertiary structure became arduous and initially we attempted to resolve the model by hand. In order to improve the electron density maps, we took advantage of an anisotropy-correction

server, STARANISO [Vonrhein, C, 2018]. In brief, STARANISO redefines the cut-off sphere or shell used to extract information from the diffraction pattern and reflection images. Usually, the statistical significance of the intensities observed is determined by the mean signal to noise ratio, $I/\sigma(I)$. If the intensities are isotropic in reciprocal space, applying spherical cut-offs ensures that all useful information is collected. However, if the intensities are in-fact anisotropic, applying spherical cut-offs greatly decreases the amount of useful information collected and subsequently increases the amount of noise extracted. With this in mind, STARANISO applies an anisotropic cut-off based on a locally averaged $I/\sigma(I)$. Thus, in all directions in reciprocal space, a cut-off is defined to maximise the extraction of relevant data. Applying this correction to the merged intensities of the unique reflections drastically improved the electron density maps and structure building programs, such as Buccaneer, were successful. Henceforth, AncFMO2 was the first mammalian FMO to be structurally characterized (see section 3.3).

3.2.9.2 Crystal generation of AncFMO3-6, AncFMO5 and AncFMO1

The success of AncFMO2 with CYMAL-6 urged us to investigate the potential of using this detergent for AncFMO3-6. We observed however, the same crystal formation previously seen with DDM under the same conditions with the crystals poorly diffracting. To optimize the conditions, we alternated precipitant and protein concentrations, temperature and pH. Whilst the former parameters played no role in optimization, the latter resulted in a change in crystal formation. Performing a pH range from 4.4 to 9.0 with a range of buffers, Sodium Acetate at pH 4.6 produced thicker crystals, with intensified yellow colour (Figure 3.2.9.2.1).



Figure 3.2.9.2.1: Crystals of AncFMO3-6, solubilized in CYMAL-6, formed after 2 days at 20 °C in 0.1 M Sodium Acetate pH 4.6 and 10% (w/v) PEG 4000 conditions.

These crystals were sent to the synchrotron and diffracted with a resolution of approximately 2.7 Å without any high degree of anisotropy. With the new structure of AncFMO2 already resolved, molecular replacement could sufficiently resolve the phase problem. Similarly, to the other AncFMO2 crystals, each of these crystals varied significantly regarding their diffraction potential. All the crystals diffracted to a minimum of 5 Å, but only 10% of those screened could hit resolutions between 2.7 and 3.0 Å. Adding substrates and cofactors such as NADP⁺ did not perturb crystal growth significantly, but only the latter was observed in the final electron density.

These results emphasized the similarity among the two AncFMOs and beckoned similar crystallization conditions should be trialled on the remaining paralogs. AncFMO5 proved difficult to purify at first (see section 3.1) but after decreasing the temperature during expression, high yields could be obtained for crystallization trials. Curiously, after gel filtration, concentrating the protein in order to reach similar protein concentrations used for crystallization, resulted in a lot of precipitation. The protein solution became sticky and highly viscous and after 24 hours in ice, high degrees of white precipitate were observed. Due to this concerning property, the protein was purified in the presence of DDM instead of CYMAL-6 in order to verify whether this observation was detergent dependent.

After purifying AncFMO5 in the presence of DDM, the same precipitation observation was documented. This feature was not conveyed after the affinity-based purifications, implying that the storage buffer used for crystallizations may be detrimental towards stability. The decrease in NaCl concentration (250 mM to 10 mM) and the increase in pH (7.8 to 8.5) are likely fundamental to protein stability. Considering the success with AncFMO2 and AncFMO3-6 under these conditions, pH and PEG 4000 screenings were conducted immediately after the gel filtration purification step, to see whether the protein could crystallize. Remarkably, AncFMO5 crystallized in conditions commensurate to those of AncFMO2. Nevertheless, the crystal shape was different to those previously seen, with cuboidal morphology (Figure 3.2.9.2.2). This difference although did not alter the reproducibility of diffraction resolutions, reaching 2.8 Å after multiple crystal screenings. Crystal growth was highly reproducible with fresh protein and crystallized in the presence of NADP⁺.

AncFMO1 crystallization presented an alternative challenge regarding purification. As AncFMO5 crystallized well in DDM, we decided to use the same detergent for crystallization trials for AncFMO1. Considering that they are expressed at low

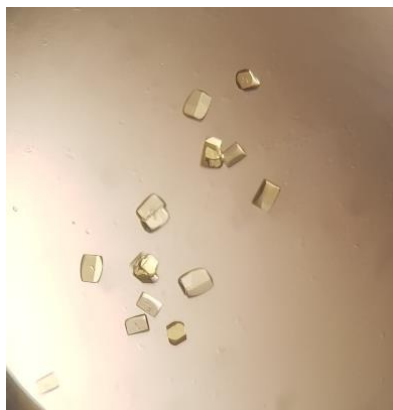


Figure 3.2.9.2.2: Crystals of AncFMO5 solubilized in DDM. Crystals formed after 1 day at 20 °C in 0.1 M HEPES pH 6.8-7.0 and 10% (w/v) PEG 4000 conditions.

temperatures, albeit for different reasons during the purification, its likely they possess similar thermostabilities and may crystallize similarly with the same detergent. The purification of AncFMO1 results in a shouldered peak that eluted out during gel filtration with two different protein-detergent complexed states (see previous section). Assuming the protein is in a dynamic equilibrium between these two states, the entire peak was pooled, including the shoulder, for crystallization trials. Remarkably, AncFMO1 managed to crystallize in conditions similar to AncFMO2 and AncFMO5 but grew uniquely shaped crystals (Figure 3.2.9.2.3).

These crystals displayed high anisotropy, similarly to AncFMO2 and structural elucidation was greatly enhanced by using Staraniso (see section 3.4). These crystals exhibited low diffraction reproducibility, meaning that multiple crystals were shot in order to obtain a reasonable dataset. Furthermore, these crystals would only diffract to a maximum



Figure 3.2.9.2.3: Crystals of AncFMO1. Crystals formed after 2 days at 20 °C in 0.1 M HEPES pH 7.5; 8% (w/v) PEG 4000.

resolution of approximately 3 Å. The low resolution and highly anisotropic diffraction patterns resulted in this paralog being the weakest crystallographic model of the four in total. Nevertheless, the crystals were reproducible and grew in the presence of the key oxidized cofactor, NADP⁺.

3.3 ANCESTRAL-SEQUENCE RECONSTRUCTION UNVEILS THE STRUCTURAL BASIS OF FUNCTION IN MAMMALIAN FMOs

In this section, the recently published work entitled Ancestral-sequence reconstruction unveils the structural basis of function in mammalian FMOs will be discussed with my contributions detailed.

Species survival and flourishing are dictated by many elements. Whilst scavenging capabilities and resourcefulness represent two key assets, being able to thrive is heavily governed by their means to disarm and eliminate potential threats. One such threat are xenobiotics, or foreign material, that organisms are exposed to greatly on a daily basis. If not treated, many molecules can exhibit toxic traits with some even causing death.

Flavin-containing monooxygenases (FMOs) represent one of the protein families that have evolved to tackle such dangers. They are involved in phase I of xenobiotic metabolism and oxidise a myriad of compounds in order to promote their eventual excretion. The importance of these systems is underlined by their prevalence within all domains of life. Additionally, humans have been shown to possess five FMOs with many more present as pseudo genes. Documenting their substrate profiles, tissue distribution, modes of action and catalytic rates have been a major undertaking in the field of flavin enzymology over the last few decades [Krueger, 2005; Jakoby, 1990]. The role of FMOs in the pharmaceutical industry as drug metabolizing agents has drawn considerable attention.

The structural identities of the human FMOs are bereft in the literature. Great advancements in our understanding of the structural features of class B flavin-dependent monooxygenases were made through bacterial and insect structures, but chordate FMOs were absent. Whilst these structures illuminated the key paired-Rossmann-fold that orchestrates NADP(H) and FAD cofactor binding, they could not portray the membrane-anchoring or substrate tunnelled features associated to these FMOs. Multiple attempts to solubilize and purify human FMOs have all but failed with no structural information provided.

Recently, there has been a surge in publications regarding the thermostability enhancement of ancestrally reconstructed proteins generated by ancestral sequence reconstruction (ASR). With thermostability being considered to play an integral role in protein crystallization, we wanted to see whether reconstructing the ancestral equivalents of the human FMOs could

be structurally resolved. However, the basis of this technique is to delineate the evolution of the protein family and allow us to synthesise a range of ancestral intermediates that may display either BVMO or FMO activities. In the long run, if successful, the technique could then be harnessed to generate a range of states that could underline the evolution of FMO or BVMO activities.

Herein this section, I present the first study that depicts the structural features of three ancestral mammalian FMOs that rationalise the modes of action observed for these systems. More specifically, the work that I carried out portrayed the structural features that permit membrane-binding, the active sites that illuminate rationale behind the BVMO and FMO functionalities and the substrate passages that siphon both aqueous and insoluble substrates.

Three mammalian FMO ancestors denoted as AncFMO2, AncFMO3-6 and AncFMO5, that respectively represent the predecessors of the human FMO paralogs, were successfully expressed and purified in *E. coli*. Each FMO displayed activity commensurate with the human FMOs including substrate profiles, catalytic rates and ostensibly more important, the C4a-(hydro)peroxy intermediate was detected. This finding principally verified the reliability behind the model construction and warranted structural investigation. Using ThermoFAD as melting temperature assay, all the AncFMOs conveyed heightened stabilities ranging from 4 to 11 °C more than the maximum stability documented for a human FMO (human FMO5 49 °C), to even higher 21 °C (70 °C) in the presence of the oxidized cofactor, NADP⁺. Thus, we were able to substantiate the application of ASR for thermostable protein development.

Strikingly, each AncFMO was successfully crystallized and their structures elucidated (Figure 3.3.1). Each FMO crystallized as a dimer, with large C-terminal transmembrane helices packing against one another within the unit cell. These large helices depicted the structural components of the peptide scaffold that theoretically drill and anchor within the expected phospholipid bilayer. Appropriately, a ring of positively charged residues resides on the same face of the transmembrane helices and portray the interface between the protein and negatively charged phosphates capping the bilayer (Figure 3.3.2). A stretch of hydrophobic residues was also shown to reside within the membrane that corroborates the insolubility characteristic observed for C-terminally truncated FMOs.

Mammalian FMOs possess an insertion of approximately 80 residues which through sequence analysis alone, could not determine the role of this motif. The crystal structure revealed that this stretch of residues in fact constituted a major role in many elements towards protein functionality. The insertion orchestrates a kinked alpha helix that creates openings for substrate transit and grapples onto the membrane. In conjunction with residues located prior to the C-terminal helix, the insertion part of a large hydrophobic

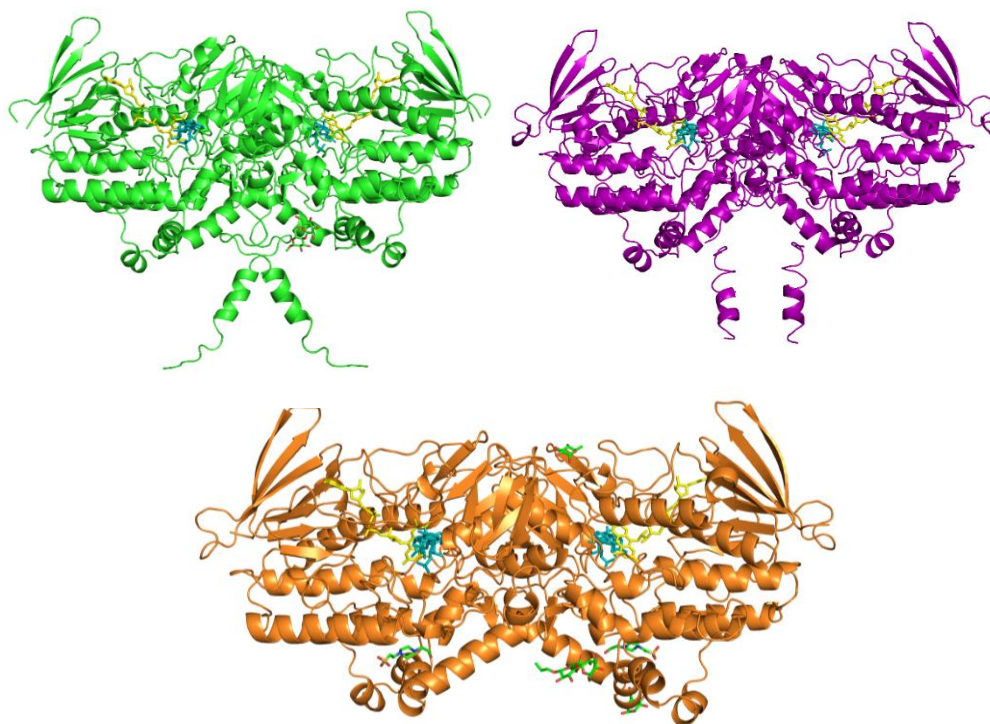


Figure 3.3.1: Crystal structures of each Ancestral FMO. The crystallographic dimers of AncFMO2, AncFMO3-6 and AncFMO5 are shown in green, purple and orange, respectively. Cofactors FAD and NADP⁺ are shown in yellow and cornflower blue, respectively. Organic molecules are shown in green including CYMAL-6, DDM and HEPES.

ridge that promotes monotopic membrane association. Additionally, the crystal structure portrays a highly flexible loop that donates key residues to the active site cavity and shields the vestibule from the solvent. This loop represents the final stretch of residues derived from the insertion. The crystal structures illuminate two distinct residues, a glutamate and a histidine, derived from this insertion that may rationalise the differing enzymatic activities observed between FMO1-3 and FMO5, respectively. Preliminary mutagenesis experiments revealed that switching between these two residues alone could not reverse the activities and additional residues may be critical for catalysis.

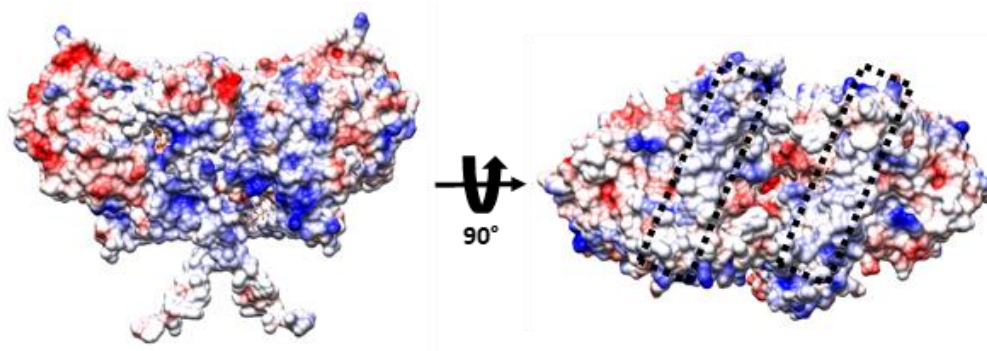


Figure 3.3.2: Distribution of charge around the surface of AncFMO2, with red, white and blue representing negative, neutral and positive, respectively. On rotation about 90° , the large parallel hydrophobic strips across the bottom of the dimer are visible, lined by a ring of positively charged residues indicated by black dashed boxes.

Each AncFMO crystal structure exposed a common substrate tunnel that courses towards the membrane. Curiously, each AncFMO curated both conserved and unique tunnels that partition separately as they reach the membrane. Two entry points lie above and below the kinked alpha helix, derived from the 80-residue insertion, that are membrane accessible. Whilst some structures only showed one of these two access points, we speculate that in solution, residues acting as blockades would be more flexible and susceptible to substrates. Despite this, both tunnels were clearly visible for AncFMO5. AncFMO2 and AncFMO3-6 however presented two exclusive passages that passed towards the solvent. As such, it is likely that these systems have the means to extract additional substrates from the solvent or alternatively, utilise these tunnels as product expelling tracts, releasing products into the cytosol.

In summary, the structural elucidation of three mammalian FMOs has rationalised the key features involved in cofactor binding and active site construction for this enzyme family and class. These constructions illuminate how the protein family utilizes almost identical active sites that possess broad substrate acceptances in order to efficiently degrade copious xenobiotic compounds. Nevertheless, the evolution of inimitable tunnels throughout the family enables the enzymes to collectively turnover extensive substrate portfolios, making them formidable xenobiotic degrading systems. Finally, these findings highlight the power behind ASR for protein crystallization and we implore its use as a structural elucidating incentive.

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

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

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

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

Humans possess five FMO isoforms that are differentially expressed in many tissues such as the kidney, lung, and liver^{2,10,11,14}. The human FMO genes are found on chromosome 1, with FMO1–FMO4 clustering over 220 kilobases (kb), and FMO5 found on a separate chromosome region^{15,16}. The human FMO family contains six non-expressed pseudogenes, which are also located on chromosome 1 (ref. 16). FMOs are involved in phase I of xenobiotic detoxification^{2,3}. They oxidize an array of compounds bearing soft nucleophilic centers, such as nitrogen and sulfur atoms^{17–19}, making them clinically important in regard to drug metabolism^{3,6,12,15,17,20,21}. The most extensively characterized FMO is human FMO3, renowned for its production of trimethylamine *N*-oxide^{22–26}. FMO3 deactivation upon mutation induces trimethylaminuria ('fish-odor syndrome'), whereby the body has an unpleasant smell due to the accumulation of trimethylamine^{27–30}. Whilst FMO4 has not been extensively characterized, FMO1 and FMO3 were shown to have broad substrate ranges, metabolizing substrates as diverse as itopride (acetylcholine esterase inhibitor) and tamoxifen (anti-breast-cancer drug)^{31–34}. FMO2 features a rather broad substrate profile, acting on

pesticides such as naphthylthiourea¹⁹, although its role in human metabolism remains partly unknown because FMO2 is not active in the majority of humans due to a mutation that introduces a premature stop codon^{35,36}. FMO5 is distinct from the other FMOs because it is able to perform Baeyer–Villiger oxidations (Fig. 1)³⁷, metabolizing ketone-containing drugs such as pentoxifylline (a muscle-pain killer)³⁷. Recent literature documents that FMOs are associated to diseases such as atherosclerosis and diabetes^{33,36}, promote longevity³⁸ and regulate cholesterol and glucose levels^{36,39–41}. Despite their discovery over 30 years ago, the determinants underlying the existence of five isoforms remain unexplored, and even more strikingly, no mammalian FMO has been structurally elucidated. This gap in our knowledge on these key enzymes of human drug metabolism likely reflects their distinctive feature: unlike bacterial, fungal, and insect FMOs that are soluble, mammalian FMOs are insoluble and reside in the membranes of the endoplasmic reticulum⁴².

To gain insight into the historical events leading to the paralogues' divergence in mammals, we generated three ancestral FMOs (the last common ancestors of extant mammalian FMO2s, FMO3s–FMO6s and FMO5s; herein referred to as AncFMOs) using ancestral-sequence reconstruction^{43,44}. These enzymes were successfully expressed in *Escherichia coli* and purified as holoenzymes (containing FAD) and active enzymes. Despite countless failed crystallization attempts of human FMO3 and human FMO5, we were able to crystallize and structurally elucidate each AncFMO. In this article, we describe the unprecedented membrane-binding features associated with the mammalian FMO, and we illustrate that substrate specificity is controlled by tunnel design rather than catalytic-site architecture. Furthermore, we demonstrate that the biochemistry of FMOs has been strictly conserved and that ancestral-sequence reconstruction is a powerful tool that can facilitate crystallization.

Results

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

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

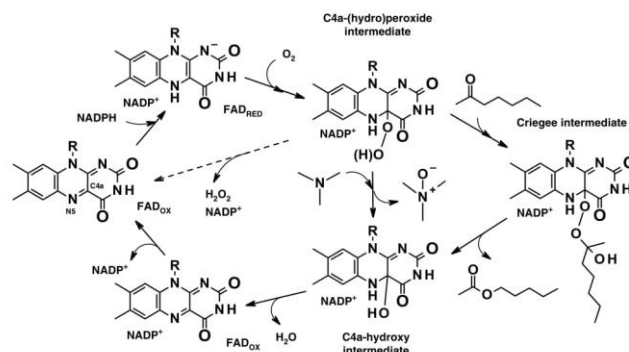


Fig. 1 | The catalytic mechanism of FMOs. Oxidized FAD (FAD_{ox}) is reduced (FAD_{red}) by NADPH. FAD_{red} is consequently oxidized by a molecule of dioxygen to generate the C4a-(hydro)peroxide intermediate. The typical mode of action of FMOs, with the distal oxygen atom from the intermediate being inserted onto a soft nucleophile through nucleophilic addition, is shown with reference to trimethylamine. The Baeyer-Villiger monooxygenation activity conducted by human FMO5 is shown on the right with reference to heptan-2-one. The dotted arrow indicates the uncoupling reaction whereby the C4a-(hydro)peroxide intermediate decays with the release of NADP^+ and hydrogen peroxide. R represents the ribityl-adenosine-diphosphate.

constructed by including experimentally characterized enzymes from Bacteria and Eukarya, plus sequences found by extensive sequence homology searching and hidden Markov model (HMM) profiling (Supplementary Fig. 1). Our work confirmed the findings of the previous studies by Hernandez et al.¹⁵ and Hao et al.¹⁴: (1) jawed vertebrate FMOs are monophyletic and are derived from a single common ancestor (Fig. 2a and Supplementary Fig. 2); (2) several duplication events occurred in the terrestrial vertebrates; (3) the ancestor of mammals already encoded the five FMO paralogues resulting from four major gene-duplication events (Fig. 2a and Supplementary Fig. 2); and (4) a sixth mammalian paralogue (FMO6) resulted from a late gene-duplication event. FMO6 has been described as a pseudogene in humans¹⁷, but it might be functional in mice¹⁸ and its nature is unknown in other mammals.

By performing ancestral-sequence reconstruction, we obtained the protein sequences of AncFMOs from mammals with high posterior probabilities (ranging from 0.98 to 0.99) (Fig. 2b and Supplementary Fig. 3). In the phylogeny, we observed that FMO5 diverged earlier than the other FMOs, in agreement with previous reports¹⁴, and was followed by FMO2, FMO1, FMO4 and the FMO3-6 paralogue couple. This topology suggests that the gene-duplication events took place simultaneously, rendering no clear paralogue couples as has been previously proposed (Fig. 2a and Supplementary Fig. 2)^{14,15}. Among the whole clade of present-day FMOs, 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 Grantham¹⁹). In the case of AncFMO3-6, the ancestor underwent an early duplication event, originating the FMO3 and FMO6 paralogues in mammals. As a general trend, 70% of sites are conserved between the preduplication ancestors and modern FMO3 and FMO6. Along each branch to the human FMO3 or human FMO6 sequences, 94–98 substitutions occurred, with 28–30 of them being conservative. The lower degree of conservation is not surprising considering the duplication scenario. Finally, FMO5 is the most enigmatic of all extant FMOs owing to its Baeyer–Villiger oxidation activity¹⁷. AncFMO5 shows 44 changes along the branch to human FMO5, with 19 conservative substitutions. In light of this historical

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

Catalytic rates of AncFMOs resemble extant mammalian FMOs. Critically for our project, the hitherto-generated AncFMO sequences proved to encode stable proteins that can be effectively produced and purified as recombinant, FAD-loaded and catalytically competent enzymes in *E. coli*. Thus, the first relevant result was that a convenient bacterial expression system for the study and biocatalytic exploitation of close homologues to human FMOs was established (see Methods). We next verified whether these enzymes retained enzymatic activities by performing steady-state kinetics experiments using a NADPH-depletion spectrophotometric assay. The NADPH oxidase activity was initially tested (NADPH consumption in the absence of an organic substrate; $\text{NADPH}_{\text{uncoupling}}$ in Table 1). This was followed by measurements of the kinetics of the reaction in the presence of known oxygen-accepting substrates of FMO2 and FMO3 (methimazole, trimethylamine and thioanisole), and FMO5 (heptan-2-one). The results were reassuring in that AncFMO2, AncFMO3-6 and AncFMO5 proved to be enzymatically active, with kinetic parameters very similar with those reported for their extant human-derived enzymes^{21,23,34,37,47–58}. The k_{cat} , $K_{\text{M,NADPH}}$ and uncoupling values ranged between 0.03–0.32 s^{-1} , 3.5–7.8 μM and 0.016–0.03 s^{-1} , respectively (Table 1). It was especially noticeable that the AncFMOs displayed a high affinity towards the coenzyme NADPH and a substantially higher NADPH consumption rate when a suitable substrate was present. This result is in full agreement with the canonical catalytic mechanism observed for FMOs and sequence-related flavoprotein monooxygenases (Fig. 1). These features were further demonstrated by stopped-flow kinetic studies. NADPH-reduced AncFMO2 and AncFMO3-6 were found to react rapidly with oxygen to form a stable and detectable C4a-(hydro)peroxyflavin intermediate with its well-defined spectroscopic properties (Fig. 3). Based on the steady-state kinetics data, AncFMO5 is assumed to behave similarly. Collectively, these experiments convincingly demonstrated that our AncFMO2, AncFMO3-6 and AncFMO5 enzymes are enzymatically competent and exhibit the typical catalytic features of class B flavoprotein monooxygenases.

ARTICLES NATURE STRUCTURAL & MOLECULAR BIOLOGY

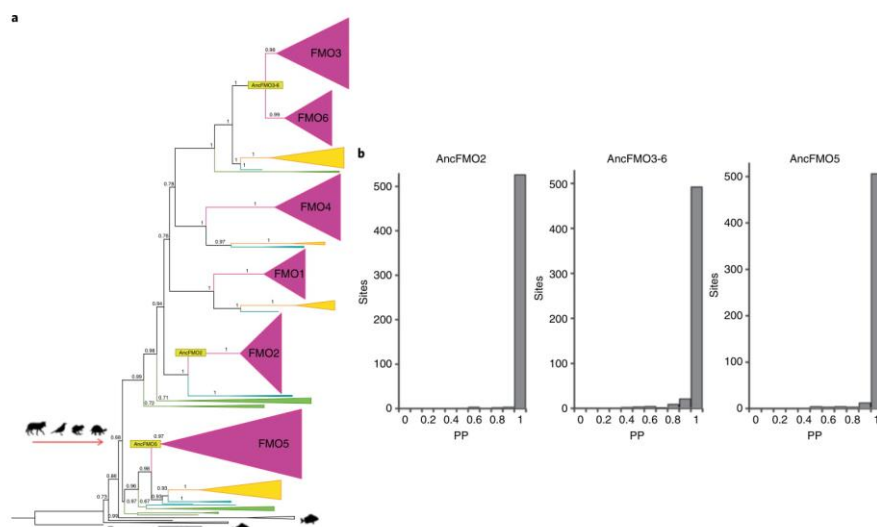


Fig. 2 | Ancestral-sequence reconstruction of FMOs. **a**, Condensed maximum-likelihood phylogeny of FMOs from jawed vertebrates. Clades are colored according to tetrapod classes: Mammalia (magenta), Aves (light orange), Amphibia (green) and Testudines (teal). Clades on the base are from other non-terrestrial gnathostomes (black). Rooting was performed according to the species tree. Above the branches, the transfer bootstrap expectation values are shown. The emergence of terrestrial vertebrates (tetrapods, 352 million years ago)⁷² is marked with an arrow and cartoons on the left. The three ancestral nodes that were experimentally characterized are labeled with yellow squares. The fully annotated phylogeny is presented in Supplementary Fig. 2. **b**, Statistical confidence of ancestral amino acid states. The highest posterior probability (PP) for each of the inferred ancestral states (sites) in AncFMOs is shown. Average accuracy for AncFMO2 was 0.994, for AncFMO3-6 was 0.982 and for AncFMO5 was 0.987. Accession codes to sequences can be found in the Source Data available with the paper online.

AncFMOs are dimers with extensive membrane-binding features. To investigate the role of the AncFMOs in detail, the crystal structures of each AncFMO in the presence of NADP⁺ were determined (Fig. 4). AncFMO2 was also crystallized in the absence of NADP⁺ but, akin to class B flavin-dependent monooxygenases, no major conformational changes were observed between the apo- or holoenzyme crystal structures (Extended Data Fig. 1). The structures were solved at 2.7-, 3.0-, 2.8- and 2.7-Å resolution for AncFMO2 (without NADP⁺), AncFMO2, AncFMO3-6 and AncFMO5 (all including NADP⁺), respectively (Table 2 and Extended Data Fig. 2). For the purpose of the structural analysis, it must be highlighted that the AncFMOs display high sequence identities to their extant human FMO counterparts: 92%, 83% and 92% for AncFMO2, AncFMO3-6 and AncFMO5 respectively, making them excellent structural models of human FMOs (Supplementary Fig. 4).

Our crystal structures depict the AncFMOs as dimers: they possess an extensive monomer–monomer interface over approximately 2,000 Å² (calculated by the PISA server)³¹. Furthermore, their well-conserved FAD and NADP(H) dinucleotide-binding domains are accompanied by two large transmembrane helices (one from each monomer) that project outwards, approximately parallel to the twofold axis (Fig. 4a–d). Pairwise structural superpositions of AncFMO2, AncFMO3-6 and AncFMO5 show that their ordered ~480 Cα atoms overlap with root-mean-square deviations of less than 1 Å. This result reveals a high degree of structural

similarity among the FMO structures. We additionally notice that the dimerization interface of the AncFMOs is different from the dimer interfaces exhibited by soluble FMOs (for example, FMO from *Roseovarius rubinhibens*, PDB entry 5IPY; FMO from *Methylophaga aminisulfivorans*, PDB entry 2VQ7).

The mammalian FMOs were predicted to contain a highly hydrophobic carboxy-terminal transmembrane helix (residues 510–532 in AncFMO3-6; Supplementary Fig. 4). The crystal structures of AncFMO2 and AncFMO3-6 perfectly confirmed this prediction, as both enzymes possess C-terminal transmembrane helices that span 30 Å in length and are decorated with many hydrophobic residues (Fig. 4a,b,d). Of notice, these α-helical scaffolds represent the key protein–protein interactions established within the crystal packing (Extended Data Fig. 3). High disorder rendered the C-terminal residues untraceable in the crystal structure of AncFMO5. The transmembrane helices of AncFMO2 and AncFMO3-6 root themselves deep within the phospholipid bilayer through a bitopic membrane binding mode, whereby the final C-terminal residues exit the other side of the membrane. These two helices anchor the protein firmly into the membrane. Thus, Fig. 4 depicts each enzyme as if it were sitting on the membrane.

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

Table 1 | Steady-state kinetics

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

^aRates were determined by following NADPH consumption (absorbance decrease at 340 nm). The increase ('bursts') in NADPH consumption rates upon addition of the substrates demonstrate that the AncFMOs are highly coupled and effectively oxygenate their substrates. ^bHeptan-2-one is a typical ketone substrate for the Baeyer-Villiger oxidation catalyzed by FMO5 (Fig. 1). ^cThe 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. ^dThe data for human FMO3 are shown in Extended Data Fig. 5. The data for human FMO5 are taken from Fiorentini et al.¹⁷

the charge distribution on the protein surface was scrutinized. Intriguingly on the underside of the dimer, two large hydrophobic strips, about 30 Å in length, extend across the enzyme surface (Fig. 4d). These strips are lined by an extensive ring of positively charged residues. Collectively, these features equip the enzyme for binding to the membrane surface. The array of hydrophobic residues penetrate, 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, respectively, that were used for protein solubilization and crystallisation (Fig. 4a,c,d). These molecules delineate the membrane-enzyme interface and further validate that the hydrophobic strips monotonically embed within the membrane. These findings rationalize the extensive membrane-binding nature employed for this class of enzymes and corroborate that the truncation of the C-terminal helix alone is not sufficient to facilitate protein solubilization⁵². FMOs employ both bitopic and monotopic membrane-binding features to grapple onto the membrane effectively and abstract lipophilic substrates from within the membrane.

An 80-residue insertion promotes membrane association. To comprehend the unique and distinct structural features associated with mammalian FMOs, we compared them with structurally characterized soluble FMOs. Consistent with class B flavin-dependent monooxygenases, the AncFMOs have two well-conserved dinucleotide-binding domains for cofactors FAD (residues 2–154 and 331–442) and NADP(H) (residues 155–213 and 296–330) respectively, known as the paired Rossmann fold (Extended Data Fig. 4a)^{12,54}. Superposition of a bacterial FMO (PDB ID: 2VQ7, 29% identity)

from *Methylophaga* sp. strain SK1 (ref. 12), shows a root-mean-square deviation of 1.1 Å over 205 C α atom pairs, verifying a strict evolutionary conservation of the dinucleotide-binding domains. However, close inspection of the structures reveals very substantial differences. In soluble FMOs, the FAD cofactor is exposed to the solvent and readily accessible by substrates. By contrast, an 80-residue insertion (214–295 in AncFMO3-6; Extended Data Fig. 4b,c) shields the AncFMOs' active site from the cytosol and creates closed substrate-binding cavities. This insertion is comprised of a subdomain orchestrated by three small α -helical units that form a ridge-like, triangular fold. Additionally, this subdomain forms the first half of the hydrophobic strip mentioned above. Despite the FAD and the catalytic center being buried by the insertion, this subdomain provides a series of tunnels that branch out from the membrane towards the active site (see 'AncFMOs show conserved membrane-accessible substrate tunnels'). This finding implies that substrates navigate through tunnels manufactured by the insertion to access the closed catalytic cavity.

AncFMOs possess a buried active site. With the AncFMOs' active sites no longer being open clefts like their soluble homologues, we scrutinized each closely to determine the functions of each residue and whether the mode of NADP(H) binding is akin to class B FMOs. Notably, most residues in the active and NADPH-binding sites are conserved with near-identical conformations (Fig. 5a–c). Thr62, Ser62 and Thr63 for AncFMO2, AncFMO3-6 and AncFMO5, respectively, are within hydrogen-bonding distance to the N3 atom of the isoalloxazine ring and orientate the FAD towards the substrate pocket. Additionally, Asn61, Asn61 and Asn62 are observed in all active sites, and an Asn residue at this position is strictly conserved among human FMOs (Supplementary Fig. 4). This residue situates close to the C4a of the isoalloxazine ring (4.6 Å) and is likely fundamental for the stabilization of the flavin-(hydro)peroxide intermediate (Fig. 1). Consistently, mutating this residue in human FMO3 causes trimethylaminuria, further verifying its integral role within the active site^{12,28}.

The binding mode of NADP⁺ observed in the crystal structures is iconic to class B FMOs (Fig. 5d)⁴. The overhanging Arg223, Arg223 and Arg224 are within hydrogen-bonding distance of the carbonyl derived from the carbamide of NADP⁺. Additionally, the amino group of the same carbamide forms a hydrogen bond with the N5 atom of the isoalloxazine ring. Moreover, the nicotinamide ring is sterically held in place by a well-conserved Asn194, Asn194 and Asn195, which acts like a backdoor for the cofactor (Fig. 5a–c). This feature is not uncommon and is portrayed in some soluble FMOs by a protruding tyrosine^{12,55,56}. The hydroxyl groups of the ribose form part of an intricate hydrogen-bonding network. The 2'-OH group is within hydrogen-bonding distance of the backdoor residue Asn194, Asn194 and Asn195 (3.0 Å) in AncFMO3-6 and AncFMO5, and Glu281 (2.9 Å) in AncFMO2 and AncFMO3-6. Additionally, the conserved Gln373 among the AncFMOs is within hydrogen-bonding distance of the 3'-OH group. Collectively, these hydrogen bonds and the steric interactions orientate the nicotinamide and the ribose in a manner characteristic to this class of enzymes and reiterate a significant role of NADP⁺ in catalysis, most likely in C4a-(hydro)peroxyflavin formation and substrate oxygenation (Fig. 1)^{12,55,57,58}.

A Glu-to-His mutation may promote Baeyer-Villiger oxidation. Considering that AncFMO5 is structurally very similar to AncFMO2 and AncFMO3-6, but at the same time is functionally divergent, we sought to clarify which features gave rise to its Baeyer-Villiger oxidation activity. Inspecting the active site alone, the differing mode of action is likely derived from a Glu-to-His substitution. In AncFMO2 and AncFMO3-6, Glu 281, derived from the above-described mammalian FMO-specific 80-residue insertion, points towards the flavin ring. With positively charged substrates

ARTICLES NATURE STRUCTURAL & MOLECULAR BIOLOGY

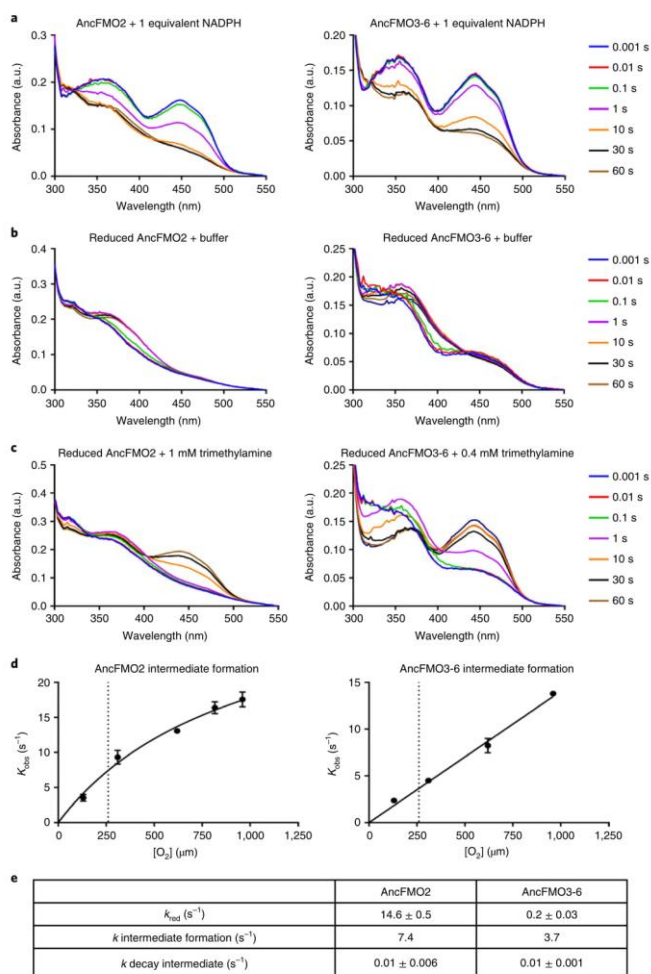


Fig. 3 | Stopped-flow kinetics studies on AncFMO2 and AncFMO3-6. **a**, Enzyme reduction upon the anaerobic addition of NADPH. **b**, Mixing reduced enzyme with dioxygen (0.13 mM) reveals the appearance of a peak at 360 nm, which is characteristic for a C4a-(hydro)peroxyflavin intermediate (Fig. 1). **c**, Mixing reduced enzyme with dioxygen (0.13 mM) and trimethylamine (1 mM or 0.4 mM for AncFMO2 and 3-6, respectively) reveals again a rapid formation of the C4a-(hydro)peroxyflavin intermediate, which subsequently decays to form the reoxidized flavin species. **d**, Dependence of the rate of C4a-(hydro)peroxyflavin formation ($A_{360\text{nm}}$) 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 FMO1². **e**, Rates of reduction, C4a-(hydro)peroxyflavin formation and C4a-(hydro)peroxyflavin decay in the absence of substrate (0.26 mM dioxygen; dotted lines in **d**). Measurements were performed in technical duplicates. a.u., arbitrary units.

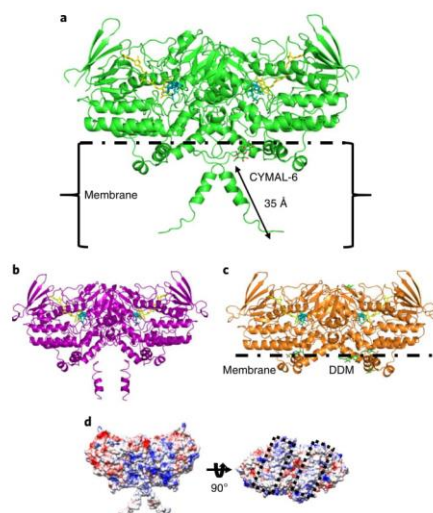


Fig. 4 | Crystal structures of the AncFMOs. **a–c.** Crystallographic dimers of AncFMO2, AncFMO3-6 and AncFMO5 are shown in lime green (**a**), dark magenta (**b**) and orange (**c**). FAD and NADP⁺ are shown in yellow and cornflower blue, respectively. The orientations of the AncFMOs are identical, depicting their structures as if they were sitting on the phospholipid bilayer. In **a**, the lengths of the transmembrane helices are portrayed at 35 Å, with the membrane cross-section indicated with brackets. In **c**, the membrane-protein interface is indicated by a horizontal dashed line, mapped with respect to the polar head group of the dodecyl- β -D-maltoside detergent molecule. Additionally, a molecule of HEPES buffer is observed entering the enzyme at the membrane-protein interface. **d.** Distribution of charge around the surface of AncFMO2, with red, white and blue representing negative, neutral and positive, respectively. On rotation about 90°, the large parallel hydrophobic strips across the bottom of the dimer are visible, lined by a ring of positively charged residues indicated by black dashed boxes.

being preferred by FMOs¹. Glu 281 is probably deprotonated and negatively charged within the cavity. In AncFMO5 however, this residue is substituted for His282, which optimally positions the N₂-H of its imidazole ring towards the substrate pocket (Fig. 5c) and likely serves as a hydrogen-bond donor. This function is commensurate with Baeyer–Villiger monooxygenases, whereby hydrogen-bond-donating residues (that is, Arg; Fig. 5d) are prevalent in the vicinity of the FAD ring to activate the carbonyl functional group of the substrate for electrophilic attack by the flavin-peroxy intermediate and stabilize the Criegee intermediate formed during Baeyer–Villiger oxidation catalysis (Fig. 1)^{39,40}. These observations rationalize the functional convergence observed among the FMO5 clade and Baeyer–Villiger monooxygenases. To probe the importance of His282 in AncFMO5, the H282E mutant was prepared and analyzed. This revealed that the AncFMO5-H282E mutant fully lost its activity. Analogously, the AncFMO2-E281H and AncFMO3-6-E281H mutant enzymes were prepared, which were found to retain FMO activity toward thioanisole, yet they were not able to perform

Baeyer–Villiger oxidations (Supplementary Table 1). This could be due to the fact that the fine structural and geometric features for formation and stabilization of the Criegee intermediate (Fig. 1) need further mutations, for example in the second shell of active-site residues.

AncFMOs show conserved membrane-accessible substrate tunnels. As the mammalian FMOs are notorious for their broad substrate profiles, we conducted extensive research to elaborate how the substrates navigate through the enzyme using the HOLLOW server²¹. The conserved tunnel is roughly perpendicular to the face of the isoalloxazine ring and extends outwards (approximately 16 Å) towards the membrane, before deviating in multiple directions (Fig. 6). In all three structures, the inner segment of the tunnel features a conserved leucine that acts as a gate keeper (Leu375 in AncFMO3-6; Fig. 6b, lower panel): in an upward position, it creates a closed cavity at the active site (AncFMO3-6), and in the downward position, it opens the tunnel to the protein–membrane interface (AncFMO2 and AncFMO5). This leucine is also conserved in human FMO1, 2, 3 and 5, implying an integral role in gating the inner ‘catalytic’ part of the tunnel and affording a solvent-protected environment for catalysis (Supplementary Fig. 4).

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

AncFMOs are reliable and thermostable models for human FMOs. Allegedly, highly thermostable enzymes are highly prone to crystallization⁴¹. Considering that all three AncFMOs crystallized, melting temperature (T_m) assays were conducted using the ThermoFAD technique⁴² to investigate the thermal stability of AncFMOs compared with that of human FMOs. Remarkably, our AncFMOs in storage buffer conditions (see 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 °C and +11 °C, respectively (Supplementary Table 2 and Extended Data Fig. 5)³⁷. Generally, the differences between the AncFMOs and their respective human equivalents are found dispersed across the protein (Extended Data Fig. 6). These patterns of highly distributed and

ARTICLES NATURE STRUCTURAL & MOLECULAR BIOLOGY

Table 2 | Data collection and refinement statistics

	AncFMO2 without NADP ⁺ (PDB 6SEM)		AncFMO2 with NADP ⁺ (PDB 6SFO)		AncFMO3-6 (PDB 6SE3)	AncFMO5 (PDB 6SEK)
	Aimless	Staraniso	Aimless	Staraniso	Aimless	Aimless
Data collection^a						
Space group	C ₂		C ₂		P3 ₂ 2 ₁	P2 ₂ 2 ₁
Cell dimensions						
<i>a</i> , <i>b</i> , <i>c</i> (Å)	152.96, 147.78, 144.93		153.75, 148.68, 139.12		156.09, 156.09, 370.61	96.78, 100.07, 143.15
α , β , γ (°)	90, 96.91, 90		90, 97.08, 90		90, 120, 120	90, 90, 90
Resolution (Å) ^b	48.39–2.75 (2.94–2.75)		138.06–3.01 (3.34–3.01)		49.30–2.80 (2.85–2.80)	49.89–2.70 (2.82–2.70)
<i>R</i> _{merge}	0.08 (3.85)		0.054 (0.85)		0.17 (2.37)	0.23 (2.88)
<i>I</i> / σ (<i>I</i>)	8.8 (0.3)		13.7 (1.3)		3.3 (0.4)	5.3 (1.5)
CC _{1/2}	0.99 (0.19)		0.99 (0.47)		0.987 (0.36)	0.986 (0.63)
Completeness (%) ^b	99.1 (98.0)		58.3 (16.2)		99.4 (99.9)	52.1 (13.9)
Ellipsoidal (%) ^b	–		93.6 (76.9)		–	92.4 (64.1)
Redundancy ^b	3.0 (2.9)		3.0 (2.7)		3.4 (3.5)	3.4 (3.3)
Refinement						
Resolution (Å)	2.75		3.01		2.80	2.70
No. reflections	48,490		32,052		129,389	38,832
<i>R</i> _{work} / <i>R</i> _{free}	0.21/0.256		0.211/0.285		0.217/0.266	0.222/0.266
No. of atoms ^c						
Protein	16639		17072		25188	8020
FAD	212		212		318	106
NADP ⁺	–		192		288	96
Other ligands	–		24		12	87
Water	51		41		357	27
B-factors						
Protein	86.5		99.5		73.6	90.6
FAD	78.1		98.2		56.8	98.1
NADP ⁺	–		85.0		67.6	75.1
Other ligands	–		102.3		57.0	126.6
Water	54.0		48.9		56.5	62.4
R.m.s. deviations						
Bond lengths (Å)	0.008		0.008		0.010	0.007
Bond angles (°)	1.311		1.310		1.530	1.553

^aEach dataset was measured from one crystal. ^bValues in parentheses are for highest-resolution shell. ^cThe asymmetric unit contained 4, 4, 6 and 2 protein molecules for AncFMO2, AncFMO2 bound to NADP⁺, AncFMO3-6 and AncFMO5, respectively.

non-systematic amino acid replacements between ancestral and extant enzymes validate the notion that AncFMOs are very reliable models for the human FMOs. Noticeably, at the periphery of the active site, the small Ala232 in human FMO5 is mutated to a bulky Phe232 in AncFMO5 (Fig. 6d, lower panel). This substitution may well allow larger substrates in human FMO5. Intriguingly, Fiorentini et al.¹⁹ documented that NADP⁺ has no effect on the *T_m* of human FMO5, a result also observed for human FMO3 (Extended Data Fig. 5). However, the melting temperatures of all three AncFMOs increased in the presence of NADP⁺ by +17 °C, +7 °C and +4 °C for AncFMO2, AncFMO3-6 and AncFMO5, respectively (Supplementary Table 2). With AncFMOs exhibiting a low degree of uncoupling, it corroborates that the tight binding of NADP⁺ is necessary for highly coupled reactions (Table 1 and Fig. 2).

Discussion

Our work supports the notion that the number of FMOs in vertebrates substantially increased by successive gene-duplication events,

leading to the multiple paralogues observed in mammals today⁴¹. Tetrapods encode for four (amphibians, testudines and birds) or six (mammals) different FMOs, suggesting defined roles for each of these variants. Analyzing the different paralogues, we observed that FMO3 and FMO6 followed a common evolutionary path preceded by the diversification of FMO4, FMO1 and FMO2. FMO5 originated from the earliest gene-duplication event and, intriguingly, is encoded by all the aforementioned terrestrial vertebrate classes. With AncFMOs exhibiting similar substrate profiles and catalytic rates as their FMO successors, we propose that this class of enzymes has an evolutionarily conserved mode of action. Moreover, two new features are derived from ancestral-sequence reconstruction: (1) increased *T_m* and (2) the stabilizing effect induced by NADP⁺ (see Supplementary Table 2). With the mutations scattered across the protein, it is unlikely that individual mutations stabilize the enzyme greatly. Their summation however, enhances stability tremendously. Whether this higher thermal stability of the AncFMOs has a biological meaning remains unclear⁴⁵.

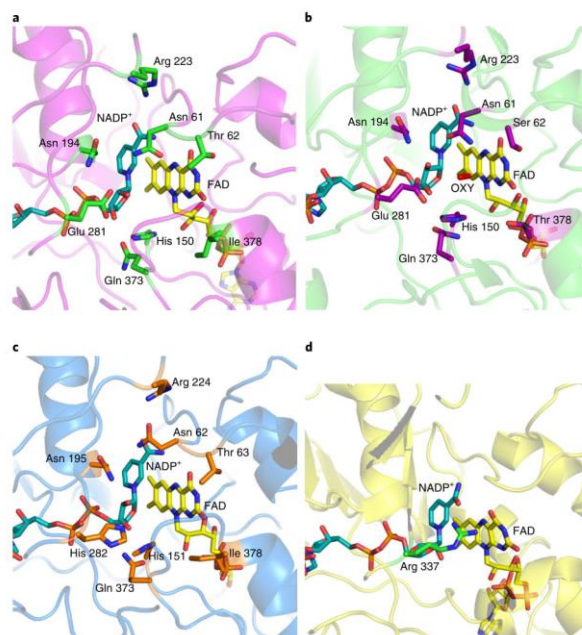


Fig. 5 | Active sites of the AncFMOs. a–c. The active sites for AncFMO2, AncFMO3-6 and AncFMO5 are shown in **a**, **b** and **c**, respectively. All three bear a high degree of similarity, with most amino acids being strictly conserved and displaying identical conformations. The differing residues are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

Our research has resulted in the unveiling of the first structures until now of mammalian FMOs. Together, they demonstrate the extensive membrane-binding features employed by this enzyme class. The literature had always speculated that the C-terminus was involved in membrane association^{37,52,53,66}, but the roles of the large insertions present in human FMOs were ostensibly more enigmatic. The dimerization observed in the crystal structure is not uncommon to membrane proteins and is now attributed to mammalian FMOs⁶⁷. Specifically, the oligomerization state aids membrane insertion, as the protein occupies a larger membrane-surface area⁶⁷. The inserted residues together form a large monotopic binding feature, which constitutively holds the enzyme in the membrane, ensuring constant uptake and release of substrates and products from and to the membrane. These molecules are then siphoned through the enzyme via a series of tunnels implemented by this subdomain. These routes also open to the cytosolic side of the enzyme structures. Presumably, all FMOs are thereby capable of accepting and expelling soluble compounds from and into the cytosolic solvent as well as lipophilic compounds from and into the membrane bilayer.

With the AncFMOs all accommodating very similar active sites, substrate profiles are likely differentiated by the tunnels penetrating the scaffold. FMO2 is generally known to be slightly more restrictive in terms of substrate size, mostly metabolizing molecules

possessing amino groups attached to large aliphatic tails^{2,68}. FMO3 and FMO1 are understood to occupy a breadth of substrate sizes^{2,18,69}. The tunnels hereby depicted do not allow us to confidently rationalize these phenomena specifically. For example, the high activity of FMO3 towards trimethylamine likely arises from the combination of subtle factors including: the distribution of charged residues, the partition of hydrophobic versus hydrophilic residues at the entrance and inside the FMO3 tunnel and the flexibility of the residues that gate tunnel access and substrate diffusion. Nevertheless, the overall architecture of the catalytic site and of the access tunnels fully explains the broad substrate scopes of mammalian FMOs. Above all, the catalytic site promotes the flavin-mediated activation of oxygen through the formation of the flavin-(hydro)peroxide intermediate, as observed in soluble FMOs as well as in Baeyer–Villiger monooxygenases^{70,71}. After flavin reduction, the nicotinamide-ribose moiety of NADP⁺ relocates to give access to the oxygen-reacting C4a atom of the flavin and thereby promote formation of the flavin-(hydro)peroxide that awaits a substrate to be monooxygenated (Fig. 1). Along these lines, no elements for the specific recognition of the substrate can be identified. The tunnel and the inner catalytic cavity of FMOs are rather designed to allow a ‘controlled’ access to the flavin-(hydro)peroxide without any strict or rigorous binding selectivity. It can easily be envisioned that the tunnels can adapt themselves depending on the

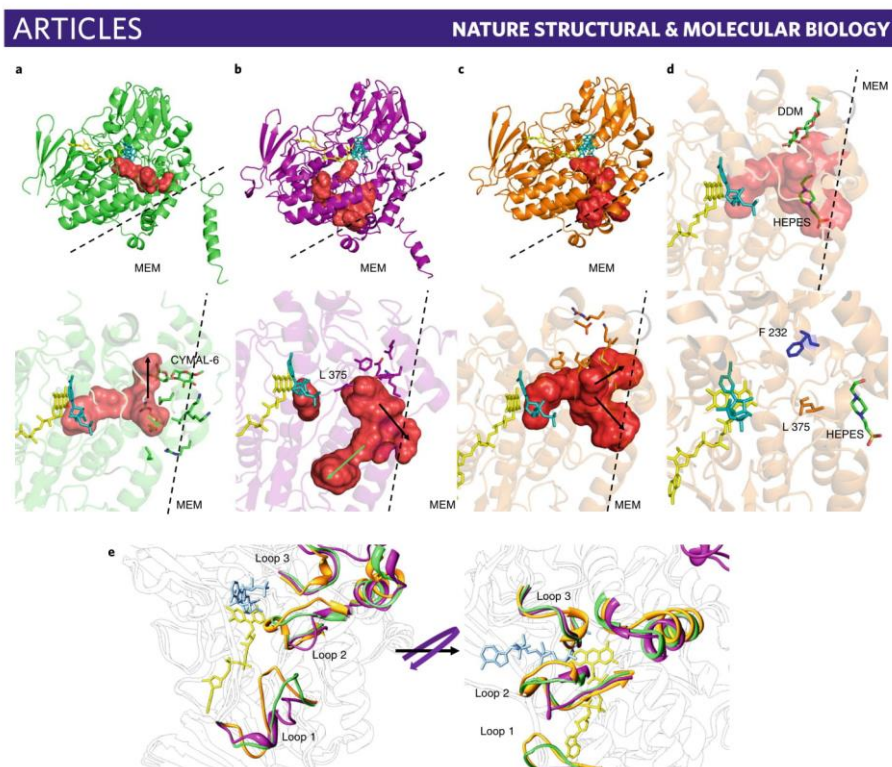


Fig. 6 | Substrate tunnels and structural differences in the AncFMOs. **a–c.** Top, the tunnels of AncFMO2 (**a**), AncFMO3-6 (**b**) and AncFMO5 (**c**) are shown, with the protein-membrane interface labeled as MEM. Bottom, The directions of the tunnels for AncFMO2 (**a**), AncFMO3-6 (**b**) and AncFMO5 (**c**) are shown, with their directions towards the membrane or the cytosol depicted by black or green arrows, respectively. The residues that block tunnel routes based on their conformations are shown. AncFMO2 and AncFMO3-6 contain two tunnel exits: one leading towards the membrane (black arrow) and one to the aqueous environment (green arrow). AncFMO5 contains two tunnels which both lead to membrane (black arrows). **d.** Top, a molecule of dodecyl-β-D-maltoside (DDM) found above the α-helical triad to emphasize the protein-membrane interface in AncFMO5 is shown. Additionally, a molecule of HEPES is present in the tunnel passing below the helix, demonstrating a substrate accessible pathway. Bottom, the Phe232 in AncFMO5 is shown with respect to gatekeeper Leu375, implying its vicinity to the FAD and how the change to alanine in human FMO5 is predicted to open the cavity. **e.** The conformational differences observed among the AncFMOs with AncFMO2, AncFMO3-6 and AncFMO5 depicted in lime green, dark magenta and orange respectively. Loop 1 contains residues 337–352 for AncFMO2 and AncFMO3-6, with 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.

bulkiness of the ligands. The gating elements (for example Leu375) may well seal the active-site cavity when small substrates are bound (Fig. 6). Likewise, the same elements could enable the binding of bulky molecules whose non-reactive groups extend along the tunnel. Thus, FMOs exhibit typical features of enzymes that broadly function in xenobiotic detoxification. Their preference for nitrogen- and sulfur-containing substrates primarily reflects the pronounced reactivity of the flavin-(hydro)peroxide towards these soft nucleophiles.

In conclusion, we have unveiled the first mammalian FMO structures through the approach of ancestral-sequence reconstruction. Additionally, our work adds to our understanding of the evolutionary history leading to the expansion of FMOs in terrestrial

vertebrates. The elucidation of three ancient FMOs has allowed us to map the differences between FMOs, providing excellent templates for structure-based drug design. Furthermore, the thermostable but functionally and structurally conserved proteins delivered by this method should be seriously and routinely considered as a tool for protein crystallization.

Online content

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

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

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

References

- Jakoby, W. B. & Ziegler, D. M. The enzymes of detoxication. *J. Biol. Chem.* **265**, 20715–20718 (1990).
- 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).
- Cruciani, G. et al. Flavin monooxygenase metabolism: why medicinal chemists should matter. *J. Med. Chem.* **57**, 6183–6196 (2014).
- 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).
- Ziegler, D. M. Flavin-containing monooxygenases: enzymes adapted for multisubstrate specificity. *Trends Pharmacol. Sci.* **11**, 321–324 (1990).
- Cashman, J. R. Some distinctions between flavin-containing and cytochrome P450 monooxygenases. *Biochem. Biophys. Res. Commun.* **338**, 599–604 (2005).
- Mascotti, M. L., Lapadula, W. J. & Ayub, M. J. The origin and evolution of Baeyer–Villiger Monooxygenases (BVMOs): an ancestral family of flavin monooxygenases. *PLoS One* **10**, e0132689 (2015).
- Ziegler, D. M. & 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).
- 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).
- Cashman, J. R. & Zhang, J. Human flavin-containing monooxygenases. *Annu. Rev. Pharmacol. Toxicol.* **46**, 65–100 (2006).
- 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).
- Aliferi, A., Malito, E., Orru, R., Fraaije, M. W. & Mattevi, A. Revealing the moonlighting role of NADP in the structure of a flavin-containing monooxygenase. *Proc. Natl Acad. Sci. USA* **105**, 6572–6577 (2008).
- Zhang, J. Quantitative analysis of FMO gene mRNA levels in human tissues. *Drug Metab. Dispos.* **34**, 19–26 (2005).
- 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).
- 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).
- 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).
- Fiorentini, F. et al. Baeyer–Villiger monooxygenase FMO5 as entry point in drug metabolism. *ACS Chem. Biol.* **12**, 2379–2387 (2017).
- Poulsen, L. L. & Ziegler, D. M. Multisubstrate flavin-containing monooxygenases: applications of mechanism to specificity. *Chem. Biol. Interact.* **96**, 57–73 (1995).
- 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).
- Cashman, J. R. Role of flavin-containing monooxygenase in drug development. *Expert Opin. Drug Metab. Toxicol.* **4**, 1507–1521 (2008).
- 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).
- Li, C. Y. et al. Structural mechanism for bacterial oxidation of oceanic trimethylamine into trimethylamine N-oxide. *Mol. Microbiol.* **103**, 992–1003 (2017).
- Miao, J. et al. Flavin-containing monooxygenase 3 as a potential player in diabetes-associated atherosclerosis. *Nat. Commun.* **6**, 6498 (2015).
- Krueger, S. K. et al. Genetic polymorphisms of flavin-containing monooxygenase (FMO). *Drug Metab. Rev.* **34**, 523–532 (2002).
- 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).
- 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).
- Phillips, I. R. & Shephard, E. A. Flavin-containing monooxygenases: mutations, disease and drug response. *Trends Pharmacol. Sci.* **29**, 294–301 (2008).
- Hernandez, D. et al. Trimethylaminuria and a human FMO3 mutation database. *Hum. Mutat.* **22**, 209–213 (2003).
- Dolphin, C. T., Janmohamed, A., Smith, R. L., Shephard, E. A. & Phillips, I. R. Missense mutation in flavin-containing monooxygenase 3 gene, FMO3, underlies fish-odour syndrome. *Nat. Genet.* **17**, 491–494 (1997).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- Fiorentini, F. et al. Biocatalytic characterization of human FMO5: unearthing Baeyer–Villiger reactions in humans. *ACS Chem. Biol.* **11**, 1039–1048 (2016).
- Leiser, S. F. et al. Cell nonautonomous activation of flavin-containing monooxygenase promotes longevity and health span. *Science* **350**, 1375–1378 (2015).
- Warrier, M. et al. The TMAO-generating enzyme flavin monooxygenase 3 is a central regulator of cholesterol balance. *Cell Rep.* **10**, 326–338 (2015).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- Scott, F. et al. Identification of a difference formula to help explain protein evolution. *Science* **185**, 862–864 (1974).
- Krueger, S. K. et al. Characterization of sulfoxidation and structural implications of human flavin-containing monooxygenase isoform 2 (FMO2.1) variants N195L and N413K. *Drug Metab. Dispos.* **37**, 1785–1791 (2009).
- 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).
- 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).
- 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).
- Krissinel, E. Stock-based detection of protein oligomeric states in jSPISA. *Nucleic Acids Res.* **43**, W314–W319 (2015).
- Geier, M. et al. Human FMO2-based microbial whole-cell catalysts for drug metabolite synthesis. *Microb. Cell Fact.* **14**, 1–10 (2015).

ARTICLES

NATURE STRUCTURAL & MOLECULAR BIOLOGY

53. Korsmeyer, K. K. et al. N-glycosylation of pig flavin-containing monooxygenase form 1: determination of the site of protein modification by mass spectrometry. *Chem. Res. Toxicol.* **11**, 1145–1153 (1998).
54. Wierenga, R. K., De Maeyer, M. C. H. & Hoi, W. G. J. Interaction of pyrophosphate moieties with α -helices in dinucleotide binding proteins. *Biochemistry* **24**, 1346–1357 (1985).
55. Orru, R., Torres Pazmiño, D. E., Fraaije, M. W. & Mattevi, A. Joint functions of protein residues and NADPH in oxygen activation by flavin-containing monooxygenase. *J. Biol. Chem.* **285**, 35021–35028 (2010).
56. Lončar, N. et al. Characterization of a thermostable flavin-containing monooxygenase from *Nitricola lacsaponensis* (NiFMO). *Appl. Microbiol. Biotechnol.* **103**, 1755–1764 (2019).
57. Beaty, N. B. & Ballou, D. P. The reductive half-reaction of liver microsomal FAD-containing monooxygenase. *J. Biol. Chem.* **256**, 4611–4618 (1981).
58. Fürst, M. J., Fiorentini, F. & Fraaije, M. W. Beyond active site residues: overall structural dynamics control catalysis in flavin-containing and heme-containing monooxygenases. *Curr. Opin. Struct. Biol.* **59**, 29–37 (2019).
59. Romero, E., Castellanos, J. R. G., Mattevi, A. & Fraaije, M. W. Characterization and crystal structure of a robust cyclohexanone monooxygenase. *Angew. Chem. Int. Ed. Engl.* **55**, 15852–15855 (2016).
60. Torres Pazmiño, D. E., Baas, B. J., Janssen, D. B. & Fraaije, M. W. Kinetic mechanism of phenylacetone monooxygenase from *Thermobifida fusca*. *Biochemistry* **47**, 4082–4093 (2008).
61. Criegee, R. Die umlagerung der dekalin-peroxydesters als folge von kationischem sauerstoff. *Justus Liebigs Ann. Chem.* **560**, 127–135 (1948).
62. Ho, B. K. & Gruswitz, F. HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* **8**, 1–6 (2008).
63. Gumulya, Y. et al. Engineering highly functional thermostable proteins using ancestral sequence reconstruction. *Nat. Catal.* **1**, 878–888 (2018).
64. Forneris, F., Orru, R., Bonivento, D., Chiarelli, L. R. & Mattevi, A. ThermoFAD, a ThermoFluor[®]-adapted flavin ad hoc detection system for protein folding and ligand binding. *FEBS J.* **276**, 2833–2840 (2009).
65. Wheeler, L. C., Lim, S. A., Marqusee, S. & Harms, M. J. The thermostability and specificity of ancient proteins. *Curr. Opin. Struct. Biol.* **38**, 37–43 (2016).
66. Lawton, M. P. & Philpot, R. M. Functional characterization of flavin-containing monooxygenase 1B1 expressed in *Saccharomyces cerevisiae* and *Escherichia coli* and analysis of proposed FAD- and membrane-binding domains. *J. Biol. Chem.* **268**, 5728–5734 (1993).
67. Allen, K. N., Entova, S., Ray, L. C. & Imperiali, B. Monotopic membrane proteins join the fold. *Trends Biochem. Sci.* **44**, 7–20 (2019).
68. Nagata, T., Williams, D. E. & Ziegler, D. M. Substrate specificities of rabbit lung and porcine liver flavin-containing monooxygenases: differences due to substrate size. *Chem. Res. Toxicol.* **3**, 372–376 (1990).
69. Kim, Y. M. & Ziegler, D. M. Size limits of thiocarbamides accepted as substrates by human flavin-containing monooxygenase 1. *Drug Metab. Dispos.* **28**, 1003–1006 (2000).
70. Beaty, N. B. & Ballou, D. P. Transient kinetic study of liver. *J. Biol. Chem.* **255**, 3817–3819 (1980).
71. Beaty, N. B. & Ballou, D. P. The oxidative half-reaction of liver microsomal FAD-containing monooxygenase. *J. Biol. Chem.* **256**, 4619–4625 (1981).
72. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019

Methods

Phylogenetic inference and ancestral-sequence reconstruction. To obtain a robust and representative phylogeny of FMOs, sequences from Bacteria and Eukarya were collected by homology searches using BLAST and HMM profiling. We collected 310 sequences, which were aligned with MAFFT v7³¹. Best-fit model parameters were obtained by the Akaike information criterion in ProtTest v3.4. Phylogenies were inferred by employing the maximum-likelihood method in PhyML v3.0 or RAxML v0.6.0 with 500–1,000 bootstraps and transfer bootstrap expectation (TBE) subsequently³². As FMOs are not monophyletic, derived clades Baeyer–Villiger monooxygenases (BVMOs) and *N*-hydroxylating monooxygenases (NMOs) were included in the phylogeny³³. Later, the gnathostomata FMOs phylogeny was constructed for ancestral-sequence reconstruction. To do this, a dataset of 361 sequences was collected, including a cephalochordate sequence, to root the tree according to species tree (Supplementary Fig. 2)³⁴.

Ancestral-sequence reconstruction was performed using the maximum-likelihood inference method in PAMLX v4.9 (ref. 35³⁵). Sequences were analyzed using an empirical amino acid substitution model (model = 3), four gamma categories and LG substitution matrix. The posterior probability distribution of ancestral states at each site was analyzed at nodes AncFMO2, AncFMO3-6 and AncFMO5. Sites were considered ambiguously reconstructed if alternative states displayed posterior probability > 0.2 (ref. 36). Alternative sites were found to be 2 for AncFMO2, 15 for AncFMO3-6 and 11 for AncFMO5 (Supplementary Table 3). These are mostly conservative amino acid substitutions, and after mapping in the crystal structures, it was evident that they all lay in the periphery of the protein, not affecting the catalytic core.

Cloning and expression of the AncFMOs. Complementary DNAs were ordered from Genescript containing BsaI sites at both the 5' and 3' ends of the insert. The insert contained overhangs TGGT and CAAG at the 5' and 3' ends, respectively, to then be inserted into common pBAD-NK destination vectors with the following modifications: three BsaI sites were eliminated and two were introduced to facilitate the cloning that incorporated SUMO and 6xHis-tag regions to the amino-terminus. Inserts were fused into the destination vectors through Golden Gate cloning. The sample was prepared with the following: 75 ng of Golden Gate entry vector (a molar ratio of 2:1 between insert and vector), BsaI-HF (15 U), 30 WU T4 DNA ligase (15 U), T4 DNA ligase buffer (1x) and nuclease-free water added to a final volume of 20 µl. During the cloning procedure, a negative control was prepared with the fragments and inserts omitted. The Golden Gate assembly was conducted in the following manner, where maximum efficiency was desired: a cycle of 37°C for 5 min followed by 16°C for 10 min was repeated 30 times; followed by 55°C for 10 min, 65°C for a further 20 min; finishing with 8°C for 20 min. Once cloned, the plasmids were transformed by heat shock into *E. coli* BL21 cells (25 s, 42°C). Cells from the resulting colonies were pre-inoculated into 100 ml of LB broth containing 100 µg ml⁻¹ of ampicillin and grown overnight at 37°C. The cultures were then transferred to 11 Terrific Broth cultures (15 ml) and grown at 24°C, 180 r.p.m. for 5–6 h until the optical density (OD) reached 0.3. The cultures were then induced with a sterilized arabinose solution (20% w/v), final concentration of 0.02% (w/v) and incubated at 24°C, 180 r.p.m. for an additional 24 h. Cells were then harvested by centrifugation (5,000g, 15 min, 10°C), flash frozen in liquid nitrogen and stored at -80°C. For site-directed mutagenesis, a PCR-reaction mixture was prepared with 10 µM primer, forward and reverse, 100 ng of template DNA, 1.6 mM DMSO, 0.8 mM MgCl₂ and 1 × Pfu Ultra II Hotstart Master Mix (Agilent). The Quickchange PCR cycle was performed using the following method: first a 5-min incubation at 95°C, then cycles (95°C for 5 min, 60°C for 30 s, 72°C for 6 min) were repeated 25 times; followed by 72°C for 10 min and finishing with 8°C on hold. The PCR mixture was digested with DpnI overnight and transformed into *E. coli*.

Cell disruption, extraction, and purification of AncFMOs. All procedures were carried out in ice or at 4°C. Cells (approximately 20 g) were resuspended (1.5 l) in buffer A (250 mM NaCl, 50 mM KH₂PO₄, pH 7.8) and included additional protease inhibitors: phenylmethylsulfonyl fluoride (1 mM), leupeptin (10 µM), pepstatin (10 µM), and DNase I (5 µg g⁻¹ cell paste). The solution was stirred and incubated at 4°C for 45 min before cell lysis was conducted using sonication or a high-pressure homogenizer. Sonication was conducted using the following conditions: 50 ml solution, 5 s on, 5 s off, with a total sonication time of 20 min using a microtip (70% amplitude). Cells were passed through a high-pressure homogenizer twice. Lysed cells were then spun down (1,200g, 12 min, 4°C) to remove the cell debris. The resultant supernatant was then centrifuged further (56,000g, 1 h and 40 min, 4°C) to collect the membrane pellet, which was then re-homogenized in buffer A (15 ml) and centrifuged again (56,000g, 1 h and 20 min, 4°C) to further purify the insoluble material. The resulting pellet was re-homogenized in buffer A (7 ml) and diluted to a final concentration of 13 mg ml⁻¹ (assayed using Biuret reagent). Triton X-100 Reduced (TRX-100-R) (Sigma-Aldrich) was then added to the solution (0.5% (v/v) final concentration) and mixed overnight at 4°C. The detergent-solubilized fraction containing the AncFMOs was then abstracted by collecting the supernatant after centrifugation (56,000g, 1 h and 20 min, 4°C). The supernatant was then transferred to a pre-equilibrated (with buffer A and 0.05% (v/v) TRX-100-R) gravity column containing Ni-resin (GE Healthcare). The supernatant was washed with buffer A, containing 0.05% TRX-100-R, and then with increasing concentrations of buffer

B (50 mM KH₂PO₄, 500 mM NaCl, 300 mM imidazole, pH 7.8), also containing 0.05% (v/v) TRX-100-R, in step-by-step fashion: 5 mM imidazole wash, 30 mM imidazole wash and finally a 300 mM imidazole wash, where the protein then eluted. The buffers were then exchanged using a centrifugal filter unit (50-kDa cut-off) and multiple washes with buffer A with 0.05% (v/v) TRX-100-R. This step was important for removing high concentrations of imidazole employed during the elution. The protein sample was then concentrated down to a final volume between 500 and 1,000 µl. The sample was then mixed with a 6xHis-tagged SUMO protease (1.2 mg ml⁻¹) to a volume ratio of 10:1 and incubated overnight at 4°C. The sample was then loaded onto an Äkta purification system (GE Healthcare) endowed with a multiwavelength detector (set at 280, 370 and 450 nm) and then onto a Ni-affinity His-trap column (GE Healthcare). The column was pre-equilibrated with buffer A containing 0.05% (v/v) TRX-100-R, as stated before, with the proteins eluting in the presence of 6 mM imidazole, derived from buffer B (2%) containing 0.05% (v/v) TRX-100-R. The SUMO-His-tag cleaved protein was then concentrated and buffer exchanged using a concentrating centrifugal filter unit (50-kDa cut-off) to a final volume between 250 and 500 µl. The sample was incubated for 1 h with 100 µM FAD at 4°C and then loaded onto a gel filtration column (Superdex 200 10/300, GE Healthcare) pre-equilibrated with a storage buffer (50 mM Tris-HCl, pH 8.5, at 4°C, 10 mM NaCl) and a detergent of choice to obtain a higher degree of purity (obtained from Anatrace). Typically, dodecyl-β-D-maltoside was used (0.03% (w/v), analytical grade), but other detergents were used for crystallization screenings at 3× their respective critical micelle concentration (CMC). The protein eluted with a very high purity and homogeneity (evaluated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and the shape of the peak in the chromatogram, respectively) with an elution volume of 10.5–11 ml. The sample was concentrated to 100 µl using a centrifugal filter unit (50-kDa cut-off) with a final concentration ranging from 5 to 30 mg ml⁻¹.

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

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

Kinetic assays of human FMO3. Kinetic assays were performed in technical duplicates at 25°C on a Varian spectrophotometer (Cary 100 Bio) equipped with a thermostatic cell compartment. The apparent K_M for NADPH was measured by varying NADPH concentrations from 20 to 400 µM in aerated reaction mixtures (150 µl) containing 2.5 µM human FMO3, 50 mM HEPES pH 7.5, 10 mM KCl, 5% (v/v) glycerol and 0.05% (v/v) TRX-100-R. Enzyme activity was measured by monitoring NADPH consumption (absorbance at 340 nm, $\epsilon_{340} = 6.22 \text{ mM}^{-1} \text{ cm}^{-1}$ 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 µl of two solutions in single mixing mode. All solutions were prepared in 50 mM potassium phosphate, 10 mM NaCl and 0.05% TRX-100-R, pH 7.5 at 25°C. For every reaction, a concentration of 10–15 µM enzyme was used, and measurements were done in technical duplicates. When needed, the solutions were supplemented with 5.0 mM glucose. Enzyme and substrate solutions were made anaerobic by flushing solutions for 10 min with nitrogen, followed by adding 0.3 µM glucose oxidase (*Aspergillus niger*, type VII, Sigma-Aldrich) to consume the leftover oxygen. The monitoring of the reductive half reaction was done by mixing the anaerobic enzyme solution with anaerobic buffer containing increasing concentrations of NADPH (Oriental Yeast Co. LTD). The rate of flavin reduction was determined by measuring the decrease of absorbance at 447 nm or 442 nm for AncFMO2 and AncFMO3-6, respectively. In order to reduce the flavin cofactor in the FMOs for the oxidative half reaction, NADPH was added to an anaerobic solution containing an equivalent amount of AncFMO. The resulting solution was incubated on ice until the bleaching of the FAD was complete, indicating complete reduction to FADH₂. The anaerobically reduced FMOs were mixed with air-saturated buffer first and then with air-saturated

ARTICLES

NATURE STRUCTURAL & MOLECULAR BIOLOGY

buffer containing 1.0 mM or 0.4 mM trimethylamine, respectively. This allowed us to follow the spectral changes during the oxidative half-reactions. The C4a-(hydro)peroxyflavin intermediate formation and decay were specifically monitored using the PMT module at 360 nm. For determining the rates of reoxidation, the reduced enzymes were mixed with buffers containing different concentrations of dioxygen. The final concentrations of dioxygen (0.13, 0.31, 0.61, 0.96 mM after mixing) were achieved by mixing the anaerobic enzyme solution with (1) air-saturated buffer; (2) equal volumes of 100% argon buffer and 100% O₂ buffer; (3) 100% O₂ buffer; (4) 100% O₂ buffer on ice. All solutions were bubbled for 10 min at room temperature, except the last one which was done on ice. In the case of AncFMO2, in order to confirm its saturating behavior, an additional measurement was performed at 0.816 mM of O₂ by mixing 100% O₂ buffer on ice with 100% argon buffer.

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

ThermoFAD Assays⁴¹. A Bio-Rad MiniOpticon Real-Time PCR System was employed to perform ThermoFAD screenings (temperature gradient 25–70 °C, fluorescence detection every 0.5 °C at 485 ± 30 nm excitation and 625 ± 30 nm emission for 5 s). Concentrations were determined using a molar extinction coefficient of 12 mM⁻¹ cm⁻¹ for the FAD band at 442 nm. Experiments were performed in triplicate using human FMO5 and the AncFMOs, in the presence or absence of NADP⁺. Each sample contained the protein of interest (4 μM), with or without NADP⁺ (200 μM), made to a final volume of 20 μl using the storage buffer and incubated in ice for 1 h. T_{m} for human FMO3 (0.05% (v/v) TRX-100-R) was determined in technical duplicates and with a final protein concentration of 5 μM in buffer (100 mM), with varying pH values (pH 6–6.5 MES, pH 7–8 HEPES, pH 8–9 Bicine, pH 9.5 CHES), KCl concentrations (0–500 mM in HEPES pH 8) and NADP⁺ concentrations (5–500 μM in HEPES pH 8, 10 mM KCl 0.05% (v/v) TRX-100-R), in an attempt to generate optimal storage buffer conditions.

Conversions. Conversions were performed using AncFMO2 and AncFMO3-6 and their respective E281H mutants (Supplementary Table 1). Reaction mixtures (1.0 ml) contained 5.0 mM substrate (c. 1% ethanol), 0.1 mM NADPH, 2.0 μM enzyme, 5.0 μM phosphite dehydrogenase, 20 mM phosphite, 50 mM KP_i pH 7.5, 250 mM NaCl and 0.05% (v/v) TRX-100-R. The mixtures were incubated for 18 h at 30 °C and subsequently extracted with 1.0 ml ethyl acetate. The organic phase was passed through anhydrous sulfate magnesium to remove residual water. Analysis was carried out using a GCMS-QP2010 Ultra (Shimadzu) equipped with a HP-1 column, using electron ionization MS detection.

Crystallization and structural determination of the AncFMOs. Each AncFMO crystallized in a range of conditions with multiple detergents. Typically, PEG 4000 was optimal for crystallization. The highest diffracting crystallization conditions for each AncFMO are described below. AncFMO2 (with and without NADP⁺): 12–15 mg ml⁻¹ of AncFMO2 (in storage buffer and CYMAL-6 (0.09% (w/v))) was incubated with crystallization conditions of HEPES buffer (0.1 M, pH 7.5) and PEG 4000 (10%) at 20 °C with a ratio of 1:1 in a sitting drop. Sitting drop was 2 μl after mixing and the reservoir was 1 ml. Prior to crystallization, NADP⁺ (1 mM final) was incubated with 12–15 mg ml⁻¹ of AncFMO3-6 for 1 h at 4 °C. After 1 d, large yellow crystals formed. AncFMO3-6: 12–15 mg ml⁻¹ of AncFMO3-6 (in storage buffer and CYMAL-6 (0.09% (w/v))) was incubated with crystallization conditions of sodium acetate buffer (0.1 M, pH 5.5) and PEG 4000 (7.5%) at 20 °C with a ratio of 1:1 in a sitting drop. Sitting drop was 2 μl after mixing and the reservoir was 1 ml. Prior to crystallization, NADP⁺ (1 mM final) was incubated with 12–15 mg ml⁻¹ of AncFMO5: 12 mg ml⁻¹ of AncFMO5 (in storage buffer and dodecyl-β-D-maltoside (0.03% (w/v))) was incubated with crystallization conditions of HEPES buffer (0.1 M, pH 6.9) and PEG 4000 (9%) at 20 °C with a ratio of 1:1 in a sitting drop. The sitting drop was 2 μl after mixing and the reservoir was 1 ml. Prior to crystallization, NADP⁺ (1 mM final) was incubated with 12–15 mg ml⁻¹ of AncFMOs for 1 h at 4 °C. After 1 d, large yellow hexagon-shaped crystals formed. During crystal fishing, a cryo-protectant was prepared containing modified crystallization conditions with 20% glycerol and PEG 4000 (15%).

Data were collected at the European Synchrotron Radiation Facility (Grenoble, France) and the Swiss Light Source (Villigen, Switzerland) and processed with the XDS⁴² and CCP4 packages⁴³. Aimless was used to merge the observations into average densities (Table 2). STARANISO was additionally used for AncFMO2, which suffered greatly from anisotropy⁴⁴. The phase problem was solved by molecular replacement using a recently solved insect FMO (PDB 5NMW)⁴⁵ as a search model, and then AncFMO2 for the subsequent AncFMOs, using Phaser and Molrep⁴⁶. The phases were greatly improved by density averaging with DM^{3.0b}. Model building and refinement were then conducted using COOT⁴⁷, Buccaneer⁴⁸ and Refmac5 (ref. 49) (Table 2). The outliers in the Ramachandran plots were 4.4%, 6.6%, 2.3% and 0.4% of the residues for AncFMO2, AncFMO2 bound to NADP⁺, AncFMO3 and AncFMO5, respectively. Figures were then generated using UCSF Chimera⁴⁸, PyMOL (DeLano Scientific; www.pymol.org) and CCP4mg⁴⁹.

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

Data availability

Coordinates and structure factors have been deposited with the Protein Data Bank with accession codes 6SEM (AncFMO2), 6SF0 (AncFMO2 in complex with NADP⁺), 6SE3 (AncFMO3-6), 6SEK (AncFMO5). Source data for Figs. 2 and 3, and Table 1, are available with the paper online.

References

- Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1166 (2017).
- Lemoine, F. et al. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* **556**, 452–456 (2018).
- Hug, L. A. et al. A new view of the tree of life. *Nat. Microbiol.* **1**, 1–6 (2016).
- Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
- 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).
- Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
- Project, C. C. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50**, 760–763 (1994).
- 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).
- 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).
- McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- 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).
- 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).
- 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).
- Petersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Acknowledgements

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

Author contributions

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

Competing interests

The authors declare no competing interests.

Additional information

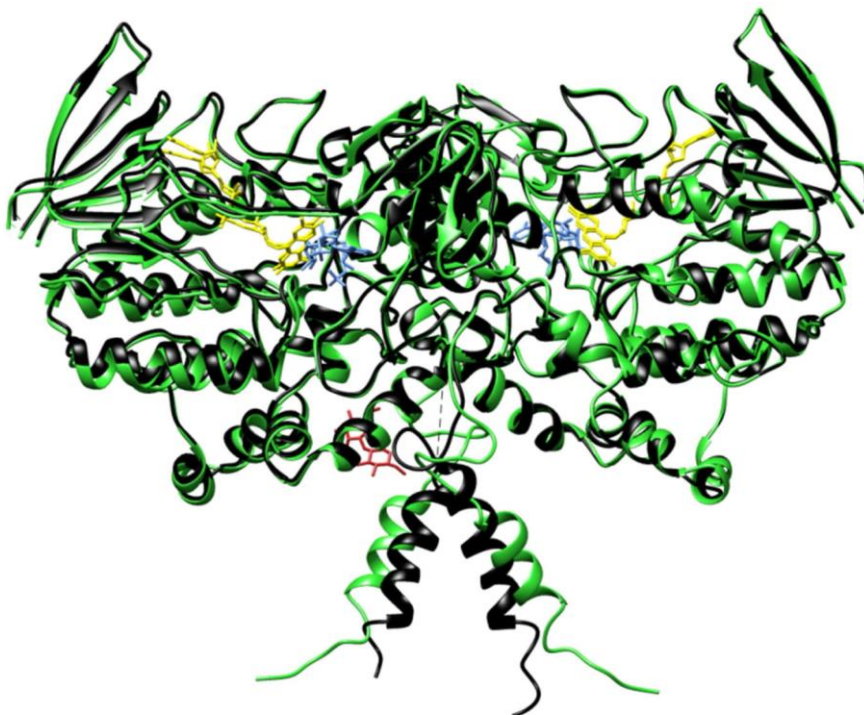
Extended data is available for this paper at <https://doi.org/10.1038/s41594-019-0347-2>.

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

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

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

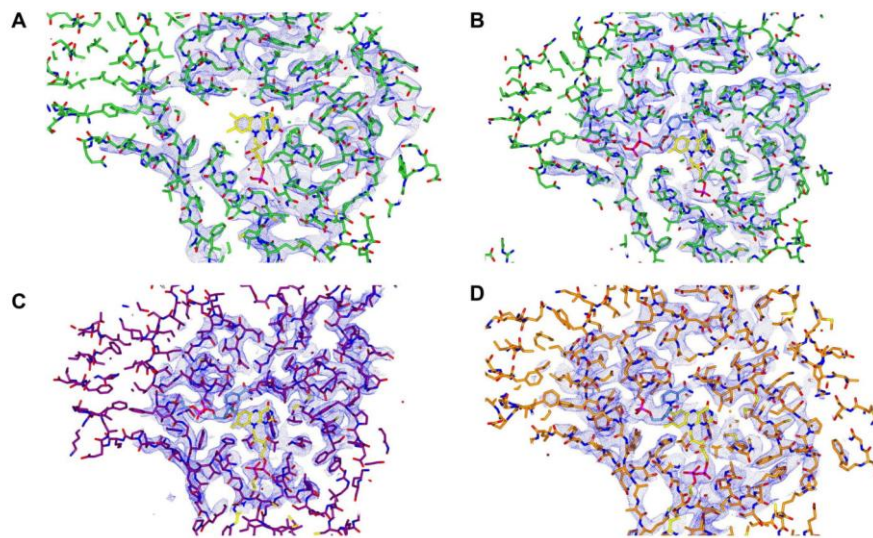
Reprints and permissions information is available at www.nature.com/reprints.



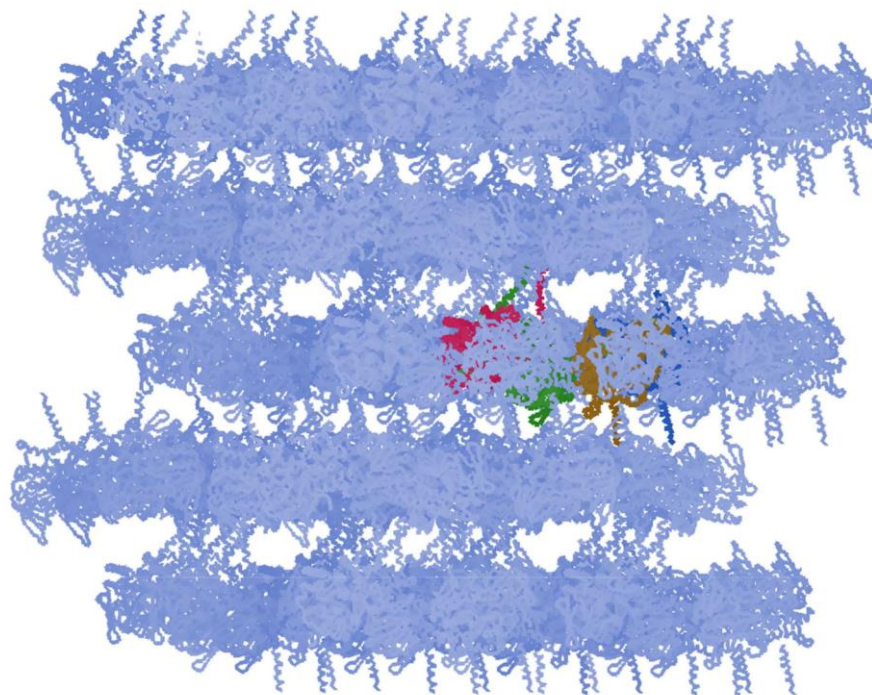
Extended Data Fig. 1 | Crystal Structure of AncFMO2 in the absence of NADP⁺ (green) superposed to the structure bound to NADP⁺ (dark green). AncFMO2 crystallizes in an identical manner with or without NADP⁺. The root-mean-square deviation between the native AncFMO2 and its NADP⁺ complex is 0.23 Å over 530 Cα atom pairs. The orientation of the dimer depicts the structure sitting on top of the phospholipid bilayer as shown in the other structures (see Fig. 4).

ARTICLES

NATURE STRUCTURAL & MOLECULAR BIOLOGY

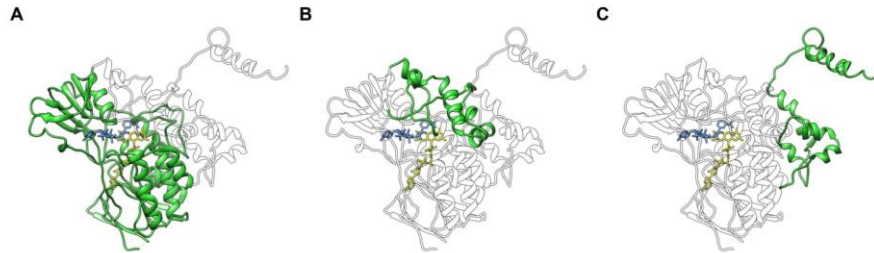


Extended Data Fig. 2 | Electron density maps of the AncFMOs. Structure determination was greatly facilitated by density averaging because the asymmetric unit of the crystals of AncFMO2, AncFMO3-6 and AncFMO5 contain four, six and two protein molecules, respectively. The depicted 2Fo-Fc maps were calculated by averaging the electron density maps obtained after molecular replacement (shown in blue). **a.** AncFMO2 shown in lime green, without NADP⁺. **b.** AncFMO2 shown in lime green, with NADP⁺. **c.** AncFMO3-6 shown in dark magenta, with NADP⁺. **d.** AncFMO5 shown in dark orange, with NADP⁺. Cofactors FAD and NADP⁺ are shown in yellow and cornflower blue respectively. The contour level is 1.4 σ .

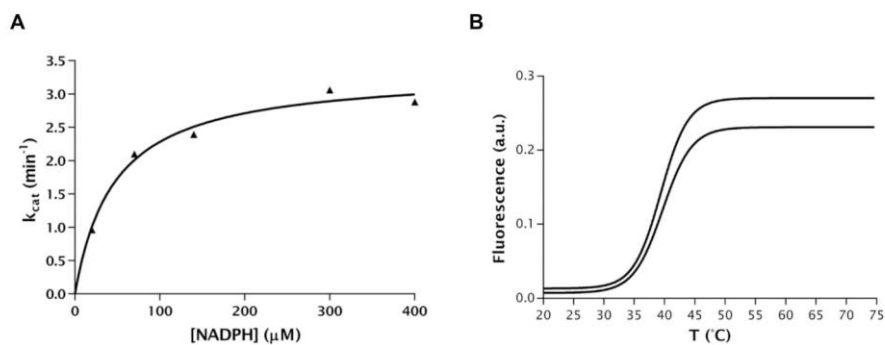


Extended Data Fig. 3 | The crystal packing of AncFMO2 forms multiple planes of soluble dimer-dimer interactions that extend across the lattice. The asymmetric unit is depicted by the four differently colored monomer units of dark yellow, dark red, dark green and dark blue. In between each plane, we see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

ARTICLES NATURE STRUCTURAL & MOLECULAR BIOLOGY



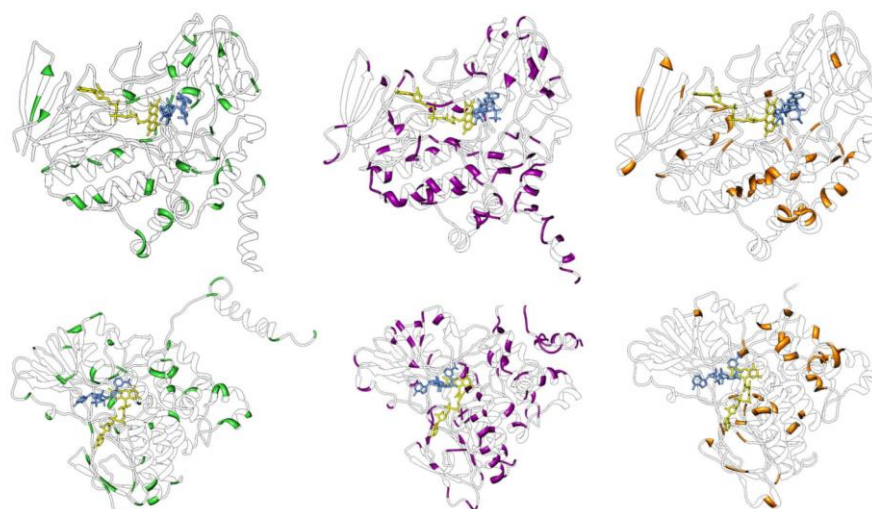
Extended Data Fig. 4 | Topological features of the mammalian FMOs. a. Highly conserved NADP(H) and FAD dinucleotide-binding domains that are observed in all FMOs. **b.** The characteristic 80-residue insertion (residues 214–295 in AncFMO3-6) that covers the FAD and binds to the membrane monotonically through an α -helical triad. **c.** The additional C-terminal (residues 443–528) that orchestrates both monotopic and bitopic membrane binding features through an α -helical triad and a C-terminal helix respectively.



Extended Data Fig. 5 | NADPH oxidase activity and melting temperature of human FMO3. **a**, NADPH consumption was not altered by the presence of substrate and so a Michaelis-Menten curve was plotted at differing NADPH concentrations. The K_M and k_{cat} for NADPH were determined at $46 \pm 9 \mu\text{M}$ and $0.06 \pm 0.16 \text{ s}^{-1}$, respectively. **b**, Extensive buffer screenings for human FMO3 resulted in a maximum melting temperature of 44.5°C (with and without $200 \mu\text{M}$ NADP $^+$ in the upper and lower curves, respectively) in buffer conditions of 100 mM HEPES pH 7.5, 10 mM KCl and 0.05% (v/v) TRX-100-R. All measurements were performed in technical duplicates.

ARTICLES

NATURE STRUCTURAL & MOLECULAR BIOLOGY



Extended Data Fig. 6 | Differing residues between AncFMOs and human FMOs. The upper and lower panels display the structures in two orientations. The changes exhibited by AncFMO2, AncFMO3-6 and AncFMO5 compared to human FMOs are shown in lime green, dark magenta and orange, respectively. A close systematic analysis of the changes does not reveal any clear pattern of amino acid substitutions. Most are localized in the membrane-binding regions, implying that the enzyme can undergo multiple mutations in these parts of the protein as long as the hydrophobic nature of the side chains is conserved. This finding is further corroborated by the sequence alignment of the human FMOs and the AncFMOs sequences, with the sequences at the subdomains and C-termini varying substantially (Supplementary Fig. 4). The mutations, however, are bereft in the FAD and NADP(H) binding domains, describing well-conserved sequence motifs among FMOs. Furthermore, the residues inside the enzyme and more importantly, lining the tunnels, are also well conserved. Only one overwhelming change in the core of the enzyme is observed in AncFMO5, as shown in Fig. 6d (lower panel).

In the format provided by the authors and unedited.

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

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

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

Supplementary Information

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

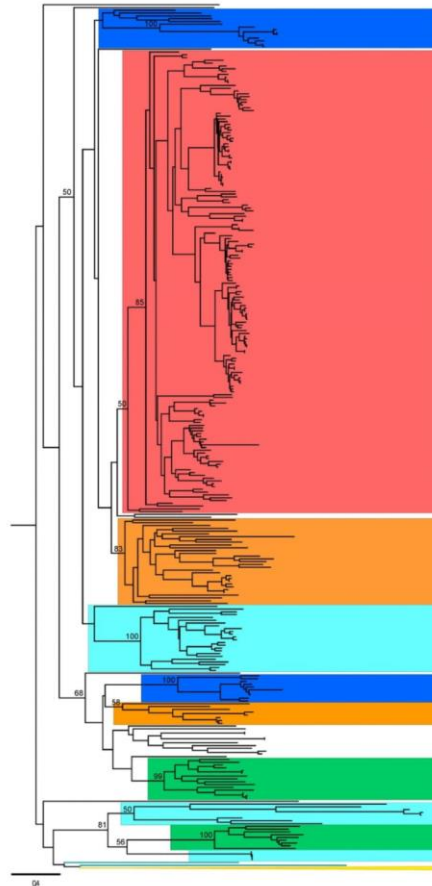
Callum R. Nicoll¹, Gautier Bailleul², Filippo Fiorentini¹,
María Laura Mascotti^{3,*}, Marco W. Fraaije^{2,*} and Andrea Mattevi^{1,*}

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

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

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

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

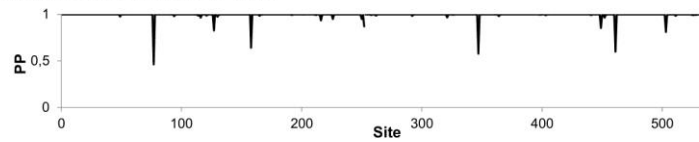


Supplementary Figure 1. Full phylogeny of FMOs. Tree was constructed in PhyML v3.0, 500 bootstraps were run, and best-fit model parameters were obtained with ProtTest v3.4. The employed multiple sequence alignment was trimmed in single sequence extensions and contained 341 taxa and 501 sites. Tree topology 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. 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.

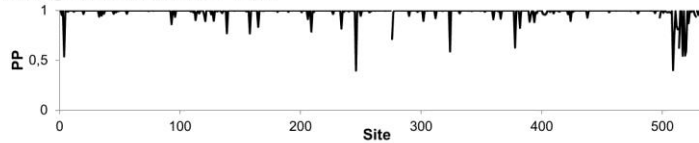

```
>AncFMO2
MAKKVAVIGAGVSLISLKCCVDEGLEPTCFERTEDIGGLWRFKENVEDGRASIYQSVITNTSKEMSCFSDPFMPPEHFPN
FLHNSKLLLEYFRIFAKKFDLLKYIQFQTTVLSVKKCPDFSSSGQWEIVTESNGKEQSAVFDAMVCSGHHILPHIPLQSF
PGIERFKGQYFHSRQYKHPEGFEGKRI LVI GIGNSASDIAVELSKKAAQVFI STRHGSWVMSRISDDGYPDMVVFHTRFS
SMLRNVLPRTVVKKWMMEQMNRWFNHNENYGLEPQNKYLMKEPVLNDDLPSRLLYGAIKVKSRVKELTETSATIFEDGTVEE
DIDVIVFATGYTFSPFLEDLSLVKVENNMVSLYKYMFPHPLEKPTLACIGLIQPLGSI FPTVELQARWVTRVFKGLCTLP
SESTMADI IKRNEKRIDLFGESQSQILQTNYIDYLDLALALEIGAKPDLLSLLKDPKIAMKLYFGPCNSYQYRLVGPQG
WEGARNAIFTQQRILKPLKTRALKASSNFPVSVLLKILGLLAVVVAFFQQLWF
```

Average of reconstruction= 0.994



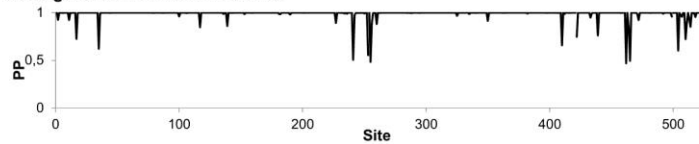
```
>AncFMO3-6
MGKKVAIIGAGVSLASIRSCLEEGLEPTCFERSDDIGGLWKFSDHAEGRASIYQSVFTNSSKEMMCFDPDFYPDDFFN
FMHNSKLQEYITAFAKEKNLLKYIQFKTLVSSVNRKPDFSVTQWDVTEKDGKESAVFDAMVICSGHHVYVNLPKESF
PGLKHFKGKCFHSRDYKEPGIFKGRVLLVIGLNSGCDIATELSHTAEKVI ISSRSGSWMSRVWDDGYPDMMLFITRFE
TFLKNSLPTAISDWYMKQMNARFKHENYGLMPLNGTLRKEPVFNDELPARILCGTVSIKPNVKEFTETSATIFEDGTVEE
AIDCVIFATGYGYAYPFLDDSI IKSRNNEVTLFKGIFPPLLEKPTLAVIGLVQSLGATIPTDQLARWAAKVFANSCTLP
TTNEMDDDI DEKMGKLLKWFQSQTLQTDYITMYDELGSFI GAKPNIPWLFITDPLALEVEFFGPCSPYQFRLMGPQKWD
GARNALLTQWDRITLKPTRTRAVGEAKRPSLFYNLLKILLFPVLLAVLLAFY
```

Average of reconstruction= 0.982

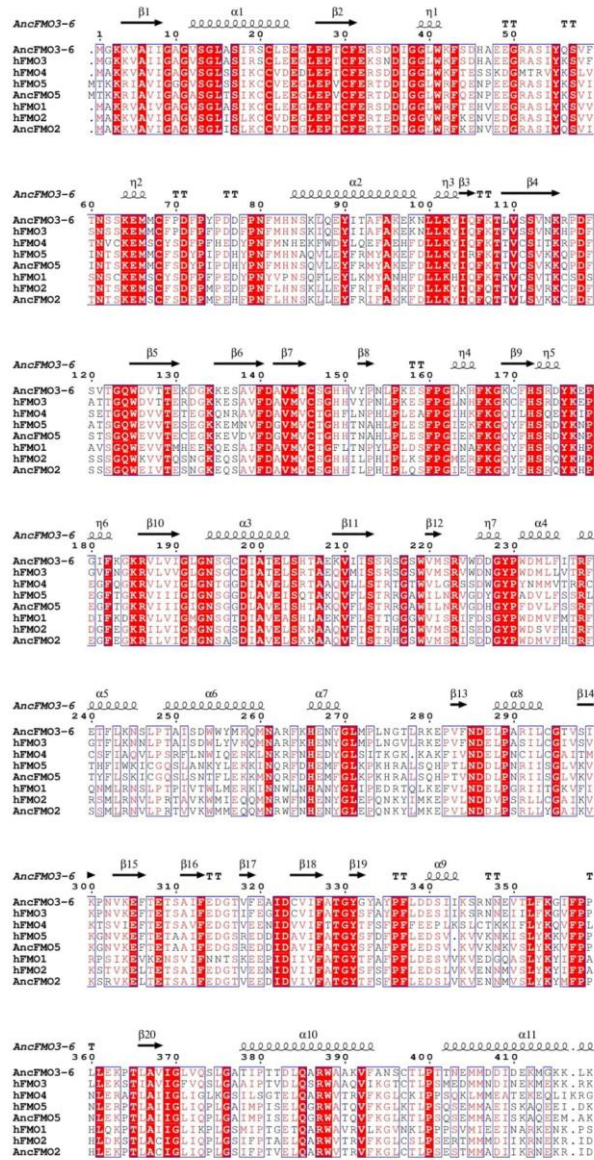


```
>AncFMO5
MTKKRIAVIGAGASGLTSIKCCLEEGLEPVCVFERTDDIGGLWRFQENPEEGRASIYKSVIINTSKEMMCFSDYPIPDHYP
NFMHNSQVLEYFRMYAKEFDLLKYIQFKTTVCSVKKQPDFSTSGQWEVVTECEGKKEVDVFDGVMVCTGHHTNAHLPLES
FPGIEKFKGQYFHSRDYKKNPEGFTGKRVIIIGIGNSGGDLAVEISHTAKQVFLSTRRGAWILNRVGDHGYPPDVLFSRF
TYFLSKICGQSLNFTLEKKNQRFDEHMFGLKPKHRALSQHPTVNDLPNRI ISGLVKVGNVKEFTETAIFEDGSRE
DDIDAVIFATGYSFAPFLEDVSVKVNKNSLYKVFPPNLEKPTLAIIGLIQPLGAIMPISELQGRWATQVFKGLKTLPL
SQSEMMAEISKAQEEMAKRYVDSQRHTIQGDYIDTMEEIDLVGVRPNLLSLAFTDPKLALKLFPGPCTPVQYRLQGPQK
WDGARKTILTTEDRIRKPLMTRVIEKSNSTMTMGRFMLAVVFFAIMAYF
```

Average of reconstruction= 0.9873



Supplementary Figure 3. The amino acid sequence of the AncFMOs. 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.



Supplementary Table 3. Ambiguously reconstructed sites in AncFMOs

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

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

Supplementary Table 3. Ambiguously reconstructed sites in AncFMOs

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

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

3.4 ENZYMATIC ANALYSIS OF ANCFMO2-E281A

Herein this section of the results, the kinetics and thermostability characteristics of a mutant AncFMO2 E281A are explored. With the glutamate residue at E281 being implicated as a potential driving force that encourages canonical FMO-like activity, we decided to remove this residue to see how it could influence activity. The mutant was designed using site-directed mutagenesis and we could successfully transform *E. coli* strains with the new construct that was correctly sequenced.

3.4.1 Thermostability Assessment

There were essentially no differences in expression levels and yields obtained with the wildtype and the mutant. Consistently, the protein sample after the final gel filtration purification step was very pure. The thermostability of the purified sample was then assessed using the ThermoFAD technique also used for the other AncFMO constructs (see section 3.3). The sample was incubated with (200 μ M) and without NADP⁺ to assess whether the stabilizing effect of the cofactor was commensurate with the wildtype. In the absence of the oxidized cofactor, the stability of the enzyme did not change. We speculate that the flexible loop does not form key non-covalent interactions that stabilize the protein scaffold. Interestingly, in the presence of NADP⁺ the mutant still exhibits a significant increase in stability, similar to the wildtype, however the extent is less (Table 3.4.1.1). The increase in stability for the wildtype was 17 °C, whereas the mutant was only 15 °C. This finding implies that glutamate 281 may form a key hydrogen bonding interaction that stabilizes the protein. In the crystal structure of AncFMO2, the distance between the 3'-OH group found on the ribose ring of the nicotinamide cofactor to E281 is 3 Å. Thus, the glutamate residue could form a hydrogen bond with the hydroxyl group that stabilizes the complex.

Table 3.4.1.1: Melting temperatures of AncFMO2 wildtype and mutant in the presence and absence of oxidized cofactor, NADP⁺. Concentrations were 4 and 200 μ M for protein and NADP⁺, respectively.

AncFMO2 construct	T _m without NADP ⁺ [°C]	T _m with NADP ⁺ [°C]
WT	53	70
E281A	53	68

We decided to resolve the crystal structure of the mutant to verify whether the loop adopted a different conformation. We were able to successfully crystallize the protein and obtained a dataset with a resolution of 3.2 Å, similar to the wildtype in the presence of

NADP⁺ (see section 3.3). The electron density revealed that the peptide backbone of the mutant did not exhibit an alternative conformation compared to the wildtype. Furthermore, residue A281, lied in the same position as E281 in the wildtype. This result implies that the conformation of the loop is not influenced by this residue and corroborates the presence of a hydrogen bonding interaction between E281 and the 3'-OH.

3.4.2 Steady-state Kinetics

The activity of the mutant was inspected to determine the impact of the glutamate for catalysis. The enzyme was first incubated with NADPH alone to evaluate the rate of uncoupling. We plotted a Michaelis-Menten curve for the NADPH oxidase activity and observed that the rate of turnover was approximately four-fold higher than for the wildtype and similar to human FMO3 (Table 3.4.2.1). With many articles documenting the role of NADP⁺ towards stabilizing the C4a-hydroperoxy intermediate, the loss of a hydrogen bond may promote the release of the cofactor, destabilize the intermediate, resulting in a higher rate of uncoupling. In contrast however, the K_M measured for NADPH was unexpectedly lower than the wildtype.

Table 3.4.2.1: Steady-state kinetics and binding constants of NADPH oxidase activities for various FMO constructs. AncFMO2, Human FMO3 and Human FMO5 data was extrapolated from section 3.3

FMO construct	k_{cat} [s ⁻¹]	K_M [μM]
AncFMO2 WT (uncoupled)	0.02 ± 0.001	20.0 ± 5.4
AncFMO2 WT (coupled)	0.32 ± 0.05	7.8 ± 1.4
AncFMO2_E281A	0.076 ± 0.003	12.3 ± 2.3
Human FMO3	0.06 ± 0.16	46 ± 9
Human FMO5	0.197 ± 0.009	59 ± 8

Measuring the rate of NADPH depletion in the presence of substrate did not alter the rate. With AncFMO5 carrying out BVMO activity with a histidine residue at position 282 (analogous to site 281 in AncFMO2), we tried trimethylamine as an FMO substrate, heptan-2-one as a BVMO substrate and thioanisole which is oxidized in all FMO paralogs. Interestingly, no burst in NADPH oxidation was observed, making it difficult to determine whether substrate oxidation occurred.

Based on bacterial and insect FMO structures, the loop enclosing the active site of the mammalian FMOs is entirely specific to vertebrate species (see section 3.3). Therefore, we

postulated that these residues are not integral to intermediate generation but may perturb the mechanism employed - BVMO vs FMO oxidation. Thus, substrate oxidation is likely occurring, nevertheless, the loss of the glutamate destabilizes the intermediate, promotes uncoupling and reduces the rate of substrate turnover. It should be noted that E281 can still donate or accept a hydrogen bond using the second carbonyl oxygen found on the sidechain (with the other hydrogen-bonding to NADP⁺). Thus, potentially playing a role in substrate binding which could influence catalytic rate.

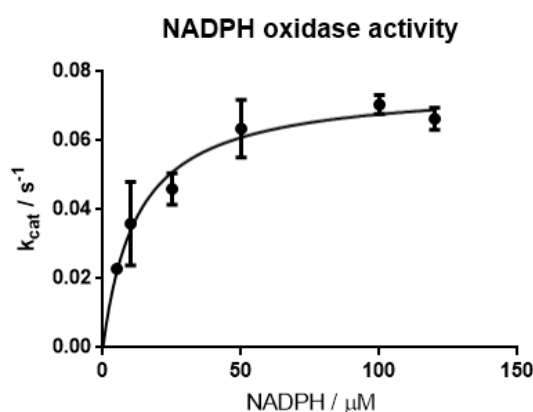


Figure 3.4.2.2: Michaelis-Menten plot for the NADPH oxidase activity of mutant AncFMO2 E281A. Readings were plotted as triplicates. V_{max} and K_{M} were $0.076 \pm 0.003 \text{ s}^{-1}$ and $12.3 \pm 2.3 \mu\text{M}$, respectively. Protein concentration was $1 \mu\text{M}$.

3.5 ANCESTRAL RECONSTRUCTION OF MAMMALIAN FMO1: UNIQUE STRUCTURAL FEATURES EXPLAIN ITS CATALYTIC PROPERTIES

In this section, the recently submitted work entitled Ancestral reconstruction of mammalian FMO1: unique structural features explain its catalytic properties is discussed with my contributions highlighted.

The recent success regarding the structural elucidation of three mammalian FMOs urged us to continue with the remaining FMO paralogs. FMO1 is predominantly expressed in the kidney and constitutes one of the main players in the FMO family regarding drug metabolism. With no literature thus far reporting successful expression of FMO4 and many populations not possessing the full-length active form of FMO2, FMO1 represents one of three key FMOs in secondary metabolism. Furthermore, the paralog has been illustrated to metabolize a vast array of drug compounds and the pig orthologue was greatly exploited in order to map out the enzyme's mode of action and kinetic profiles.

During this research, I was able to successfully purify and crystallize the protein in a manner consistent with the other AncFMOs. The crystal structure resembles a classical mammalian FMO that was previously characterized by this method including the paired Rossmann-fold and membrane binding-domains (Figure 3.5.1). Furthermore, the protein conveyed a membrane-accessible substrate tunnel, confirming its tendency to metabolize hydrophobic compounds.

Intriguingly however, AncFMO1 displayed some structurally contrasting features with respect to the other members of the family. Firstly, multiple changes at the base of the active site abolishes a key hydrogen bond that originally enclosed the core from the solvent, depicted by two glycerol molecules breaching the core (Figure 3.5.1).

Secondly, a large conformational change is observed whereby a once-alpha helix secondary structure unravels into a large arched loop (Figure 3.5.1). This new conformation blocks one of the membrane accessible substrate entry points that sits underneath the kinked alpha helix that situates within the membrane. This alteration alongside the porous active site collectively tears open the protein scaffold and the active site to the aqueous environment, thereby describing new substrate entry points.

Overall, this article provides a thorough biochemical characterization of the mammalian FMO1 paralog and reiterates the propensity of ASR to generate crystallizable proteins. Moreover, we could not express the mammalian ancestral FMO4. This suggests that there is an intrinsic factor within the sequence that dictates and limits protein expression. Nevertheless, the addition of AncFMO1 to the class B flavin-dependent monooxygenases provides further information regarding enzymatic activities and substrate uptake. Finally, our findings illustrate and convey how this family have evolved copious substrate pathways in order to maximise their substrate capacities.

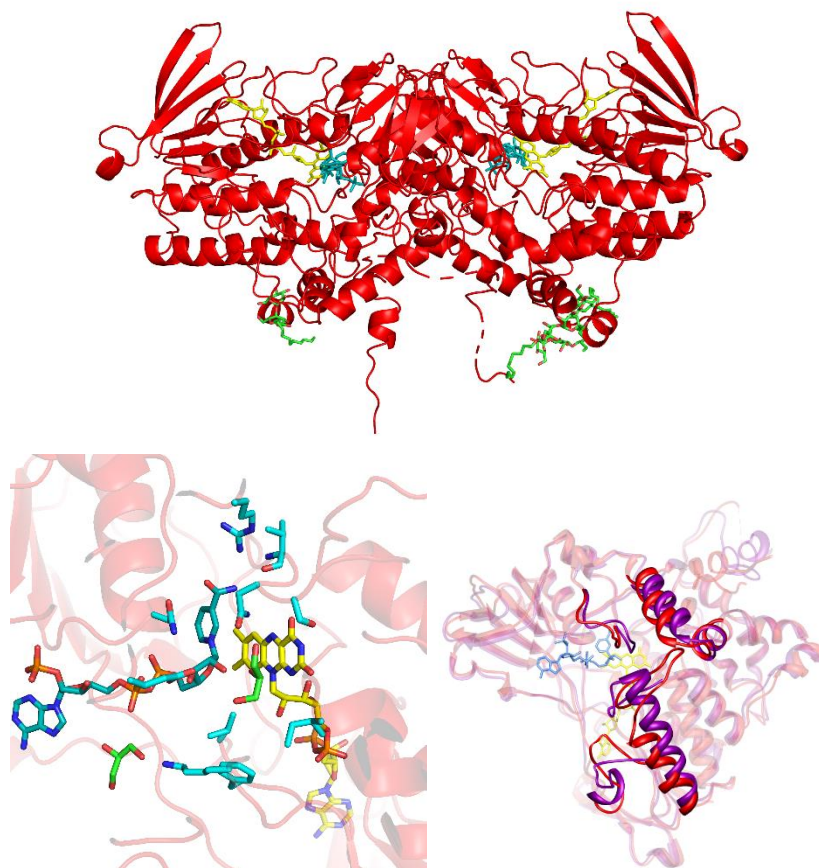


Figure 3.5.1: Principal structural features of AncFMO1. Top panel describes the crystallographic dimer of AncFMO1 resolved at 3 Å resolution. The bottom left panel depicts the active site cavity of AncFMO1 with a solvent accessible path portrayed with glycerol molecules shown in green. The bottom right panel shows a superposition of AncFMO1 (red) with AncFMO3-6 (purple) with changes in structural motifs emphasised. Organic molecules representing DDM or glycerol are shown in green, respectively. Molecules of FAD and NADP⁺ are shown in yellow and cornflower blue, respectively.

Ancestral reconstruction of mammalian FMO1: unique structural features explain its catalytic properties

Gautier Bailleul^a, Callum R. Nicoll^b, María Laura Mascotti^{a,c}, Andrea Mattevi^b, Marco W. Fraaije^{a,*}

^a Molecular Enzymology group, University of Groningen, Nijenborgh 4, 9747AG, Groningen, The Netherlands

^b Department of Biology and Biotechnology “Lazzaro Spallanzani”, University of Pavia, Pavia, Italy

^c IMIBIO-SL CONICET, Facultad de Química Bioquímica y Farmacia, Universidad Nacional de San Luis, San Luis, Argentina

* corresponding authors: A. Mattevi, M.W. Fraaije
E-mail : andrea.mattevi@unipv.it, m.w.fraaije@rug.nl

G. Bailleul: ORCID: 0000-0001-5083-2727
C. R. Nicoll: ORCID: 0000-0002-2122-9387
M.L. Mascotti: ORCID: 0000-0002-5817-4631
A. Mattevi: ORCID: 0000-0002-9523-7128
M.W. Fraaije: ORCID: 0000-0001-6346-5014

Running title: AncFMO1 structural and catalytic features

Keywords: Flavin-containing monooxygenase (FMO), flavin adenine dinucleotide (FAD), ancestral sequence reconstruction (ASR), NAD(P)H, enzyme kinetics, stopped-flow, crystal structure

Abstract

As a means to deal with numerous and potentially deleterious xenobiotics, mammals rely on flavin-containing monooxygenases (FMOs) as oxidative systems to detoxify these compounds. However, despite their high pharmacological relevance, our knowledge regarding these membrane-bound enzymes is greatly impeded by the lack of structural information. Using ancestral sequence reconstruction, 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, could be successfully expressed as functional enzyme. It displayed typical FMO activity by oxygenating benzydamine, tamoxifen and thioanisole, drug-related compounds known to be also accepted by human FMO1. 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 portrayed some unique features, including a significantly porous and exposed active site and NADPH occupying a new conformation with the 2'-phosphate being pushed inside the NADP⁺ 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 (1). 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 recombinants and their challenging isolation. It is nonetheless well established that hFMOs play a crucial role in xenobiotic metabolism(2, 3).

There are 5 FMO paralogs in the human genome (FMO1-5), all clustered on chromosome 1(4, 5). A sixth gene, called FMO6, is also present but is not translated and thus referred to as a pseudogene(6). 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(7, 8). 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(9). 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(10).

FMOs present a tightly bound flavin adenine dinucleotide (FAD) as prosthetic group(11, 12). 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(13). The biochemical properties of mammalian FMOs were first described using enzymes from pig liver microsomes by Ziegler & Poulson in 1970(14). In two extensive studies, Beatty and Ballou unraveled the catalytic mechanism involving the formation of a C(4 α)-hydroperoxyflavin intermediate that performs the oxygenation(15, 16). Formation of this reactive enzyme intermediate is achieved by a stepwise process in which first the FAD cofactor is reduced by a hydride transfer from the nicotinamide adenine dinucleotide phosphate (NADPH) (Figure Aa). With the oxidized nicotinamide still bound, the reduced flavin is able to react

with molecular oxygen resulting in a relatively stable C(4 α)-hydroperoxyflavin intermediate (Figure Ab). In the next step, when substrate approaches the flavin, the peroxy moiety transfers an oxygen atom to a soft-nucleophilic substrate (Figure Ad) while the other oxygen atom is released as water (Figure Ae). The last step is the release of NADP⁺ (Figure Af). 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 Ac).

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(17). 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(18, 19). 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(20, 21). In our previous study, using ASR we managed to express, purify and crystallize the mammalian ancestors of FMO2, FMO3 and FMO5(22). 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(15, 16). Critically, it has been established that hFMO1 displays a broad range of substrates, accepting drugs like benzydamine, ethionamide, tamoxifen or voriconazole(2). 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(23) (Figure 1, 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) when compared with 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). 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 enzyme possesses a characteristic C-terminal extension after the transmembrane domain which may prevent its proper expression in agreement with previous literature (24). 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. 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 exhibits a relatively low melting temperature of 47 °C that increased to 52 °C in the presence of 200 μ M NADP⁺ (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(25), exhibiting typical Michaelis-Menten kinetic behavior (Table 1, Figure B). A k_{cat} of 0.16 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})(26). AncFMO1 was also found to accept the prototypical FMO substrate, thioanisole (0.26 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 is about 5-fold lower than the K_M for NADH ($13 \mu\text{M}$ vs. $73 \mu\text{M}$) (Table 1). In the absence of any organic substrate, AncFMO1 displayed the same uncoupling (*i.e.* the oxidation of NAD(P)H leading to the release of hydrogen peroxide) rate of 0.07 s^{-1} , regardless of the cofactor. However, this time the K_M for NADH was lower than that for NADPH (Table 1). It should be noted that also hFMO1 was shown to be highly uncoupled, producing H_2O_2 in higher amounts compared to hFMO2 and hFMO3(27).

Having established that AncFMO1 is enzymatically active, we sought to further characterize the enzymatic conversions and the resulting products (Table S2). Benzydamine is converted to didesmethyl-benzydamine by CYPs whereas hFMO1 transforms it into the corresponding N-oxide(25, 28). Using HPLC analysis, we found that also AncFMO1 converts benzydamine into its N-oxide. To probe a rather bulky and known FMO substrate, tamoxifen (an anti-estrogen drug(29)) was tested as substrate. Tamoxifen was converted into the N-oxide derivative, as previously reported for hFMO1(30). Thioanisole was converted into the corresponding sulfoxide with high enantioselectivity, producing mainly the (*R*)-enantiomer. 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(31, 32). 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*(33). 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 left incubated for 72 hours at room temperature with indole and a NADPH regeneration system, which indeed resulted in formation of indigo (data not shown). 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 spectrophotometrically 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 2a). Fitting the data resulted in a relatively fast reduction rate constant with $k_{\text{red}} = 15.7 \text{ s}^{-1}$ and a dissociation constant of $99 \mu\text{M}$. 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 $953 \mu\text{M}$, which is almost 10 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(13).

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 results in the formation of a C(4 α)-hydroperoxyflavin intermediate with absorbance features around 360-380 nm(34). The intermediate subsequently decays 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 were observed and we were unable to distinguish intermediate formation from the full reoxidation of the flavin (Figure 2b). This shows that there was only a minor build-up of the C(4 α)-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 μ M dioxygen was around 2 s⁻¹ (Figure 2c) while the steady state kinetics of uncoupling suggested a slower reaction with a rate only reaching 0.07 s⁻¹ (Table 1).

Overall three-dimensional structure of FMO1

We were able to successfully crystallize AncFMO1 and to solve its structure at 3.0 Å resolution (Figure 3a, Table 2, Figure S3). 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 observed for the dimers documented by the other AncFMO crystal structures(22)(Figure 3a). Consistently, the structure of AncFMO1 possesses key structural domains and folds canonical to the mammalian FMOs (Figure 3a). 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 α -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 S4)(22). The oxidized coenzyme, NADP⁺, 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(22). 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 over 449 C α atoms, corroborating a high protein scaffold similarity. The crystal structure revealed that the ambiguous sites were all peripheral from the active site and not expected to impact enzymatic activity. This is supported by the enzyme assays performed on AltAncFMO1.

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 3b, 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 S5), 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, it 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 3b, 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⁺. Together, these two hydrogen bonds formed a key steric blockade that maintained the closed cavity feature that was a characteristic of mammalian FMOs(22). 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⁺, thereby further enlarging the active site (Figure 3b). 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(35). 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 4a). 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⁺ 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⁺ molecule. Additionally, FMO2 and FMO5 possess additional residues that can also engage in intermolecular interactions(22). 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, it establishes hydrogen bonds with the sidechains of Arg280 and Thr214, and the peptide of Met192 (Figure 4a). 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.

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(36), and unlike FMO2, is not limited or restricted to substrate size or length(37). 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 images of proteins(38). AncFMO1 exhibits a narrow tunnel that passes through the α -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 4b, c). Moreover, the structural topology choreographed by residues 416-425 provides additional space for soluble substrates (Figure 4b). This motif (residues 416-425) positions itself directly underneath the kinked α -helical triad that siphons substrates from within the membrane. 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 S6). As a result, this new fold creates a steric blockade underneath the α -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. 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(26, 35, 39). 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 C(4 α)-hydroperoxy flavin intermediate and a theoretical substrate. The distal hydroperoxy oxygen atom would engage with the oncoming soft nucleophile's sp³-hybridised orbital, here represented as the central hydroxyl group of the glycerol molecule (Figure 3b, c). Additionally, the 2'-OH group of the nicotinamide ribose of NADP⁺, that was speculated(40, 41) 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 C(4 α)-hydroperoxy flavin stabilization, thereby suffering from relatively high uncoupling.

Structural inspection of the NADP⁺ 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⁺. 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(42), on the evolutionary adaptation of isocitrate dehydrogenases for NAD⁺ over NADP⁺, 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(43). 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 (22). New homology searches were performed using human FMO1 (NP_002012) and FMO4 (XP_005245102) sequences as queries to improve the

representation of those clades. Dataset included 37 FMO1-like and 49 FMO4-like sequences from mammals. A multiple sequence alignment was built in MAFFT v7 and 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 β cells strain and DNA ligase were from New England Biolabs. NADPH and NADP⁺ 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×), and nuclease-free water added to a final volume of 20 μ 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 β CaCl₂ competent cells. 5.0 μ l of plasmid DNA was added to 100 μ l CaCl₂ 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 μ l LB-SOC(44) was added to allow the cells to recover at 37°C for 1 h. The resuspended cells pellet was then plated on LB-agar(44) containing 100 μ g.ml⁻¹ 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 μ g.ml⁻¹) was grown overnight at 37°C and used to inoculate 2 liter baffled flasks containing 400 ml of Terrific-Broth medium(44), supplemented with 50 mg.l⁻¹ 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 OD₆₀₀ 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 protein 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 β -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 work²². Purification of AncFMO1 was performed using dodecyl- β -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 of AncFMO4 expression

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): SDS (sodium dodecyl sulfate), DDM (dodecyl-beta-maltoside), TRX (Triton X100), OG (octyl

glucoside), FOS (FOS-Choline 8), GDN (glycodyosgenin), LDAO (lauryldimethylamine oxide) and DMDPPO (dimethyldecyl phosphine oxide). Detergent solutions were pelleted to remove the insoluble components by spinning down the samples at 100,000g for 30 mins at 4 °C. 20 ul from each sample (including the supernatant representing the aqueous fractions) were then submitted to sodium dodecyl sulfate–polyacrylamide gel electrophoresis, SDS-PAGE analysis. The resulting gel was then transferred to a Mini Format, 0.2 µm 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 then 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 µM NADP⁺, respectively. Concentrations of AncFMO1 were determined using $\epsilon_{FAD} = 12.0 \text{ mM}^{-1} \cdot \text{cm}^{-1}$ at 442 nm. Experiments were performed in triplicate, with each sample containing AncFMO1 (1.0 mg ml⁻¹, determined using the calculated molecular weight of AncFMO1, 61 kDa), with or without NADP⁺ (200 µM), made to a final volume of 10 µl 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(22) 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 duplicates. 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 µM substrate and 1.0 µM 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_{cat} values were calculated by fitting the data with the Michaelis-Menten equation using GraphPad 6.07 (La Jolla, CA, USA).

Conversions

Substrate conversions were done at pH 7.5, using 5.0 mM substrate (1% MeOH), 0.10 mM NADPH, 2.0 µM enzyme, 5.0 µM 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 µm). 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 μ 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 μ l of the sample into 400 μ l acetonitrile, vortexing it for 20 s and centrifuging. Analysis was performed using reverse phase HPLC. Samples were injected with a volume of 10 μ l onto a JASCO HPLC system, equipped with a Grace Alltima HP C18 column (5 μ m, 4.6 \times 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⁻¹. 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 μ m, 4.6 \times 250 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⁻¹. 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(22). Experiments were run at 25 °C and the k_{red} and K_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 μ 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 allow 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⁺. Dithionite was titrated until the loss of the yellow color of the oxidized FAD was complete, indicating complete reduction to FADH₂. Mixing of reduced AncFMO1 with dioxygen-containing buffers (130 μ M dioxygen, final) was done both with and without 100 μ M NADP⁺.

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⁻¹ cm⁻¹. Prior to crystallization, NADP⁺ (1.0 mM final) was incubated with AncFMO1 (12.5 mg ml⁻¹, in storage buffer conditions containing 0.03% (w/v) DDM) for 1 h at 4 °C. 1 μ l protein-containing solution was then mixed with 1 μ 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(45) and CCP4 packages(46). Aimless was then used to merge the observations into average densities for two datasets(46) to maximize the total number of

observed reflections, (Table S3). STARANISO was used as, similarly to AncFMO2, these crystals suffered severely from anisotropy(47, 48). 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(22, 49). The phases were greatly improved by density averaging with DM(46, 49). Model building and refinement were then conducted using COOT(50), Buccaneer(46) and Refmac5(51). The residue outliers in the Ramachandran plots were 5.9% for AncFMO1. Figures were generated using UCSF Chimera(52), PyMOL (DeLano Scientific; www.pymol.org) and CCP4mg(46). 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 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 and ANPCyT (Argentina) PICT 2016-2839 to M.L.M. M.L.M. is a member of the Research Career of CONICET, Argentina.

Conflict of interest

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

References

1. Krueger, S. K., and Williams, D. E. (2005) Mammalian flavin-containing monooxygenases: Structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacol. Ther.* **106**, 357–387
2. Phillips, I. R., and Shephard, E. A. (2017) Drug metabolism by flavin-containing monooxygenases of human and mouse. *Expert Opin. Drug Metab. Toxicol.* **13**, 167–181
3. Cruciani, G., Valeri, A., Goracci, L., Pellegrino, R. M., Buonerba, F., and Baroni, M. (2014) Flavin monooxygenase metabolism: Why medicinal chemists should matter. *J. Med. Chem.* **57**, 6183–6196
4. Shephard, E. A., Dolphin, C. T., Fox, M. F., Povey, S., Smith, R., and Phillips, I. R. (1993) Localization of genes encoding three distinct flavin-containing monooxygenases to human chromosome 1q. *Genomics.* **16**, 85–89
5. McCombie, R. R., Dolphin, C. T., Povey, S., Phillips, I. R., and Shephard, E. A. (1996) Localization of human flavin-containing monooxygenase genes FMO2 and FMO5 to chromosome 1q. *Genomics.* **34**, 426–429
6. Hines, R. N., Hopp, K. A., Franco, J., Saeian, K., and Begun, F. P. (2002) Alternative processing of the human FMO6 gene renders transcripts incapable of encoding a functional flavin-containing monooxygenase. *Mol. Pharmacol.* **62**, 320–325
7. Dolphin, C. T., Cullingford, T. E., Shephard, E. A., Smith, R. L., and Phillips, I. R. (1996) 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
8. Zhang, J., and Cashman, J. R. (2006) Quantitative analysis of FMO gene mRNA levels in human tissues. *Drug Metab. Dispos.* **34**, 19–26
9. Basarab, T., Ashton, G. H. S., Menagé, H. D. P., and McGrath, J. A. (1999) Sequence variations in the flavin-containing mono-oxygenase 3 gene (FMO3) in fish odour syndrome. *Br. J. Dermatol.* **140**, 164–167

10. Phillips, I. R., and Shephard, E. A. (2020) Flavin-containing monooxygenase 3 (FMO3): genetic variants and their consequences for drug metabolism and disease. *Xenobiotica*. **50**, 19–33
11. Cashman, J. R., and Zhang, J. (2006) Human flavin-containing monooxygenases. *Annu. Rev. Pharmacol. Toxicol.* **46**, 65–100
12. Ziegler, D. M. (2002) An overview of the mechanism, substrate specificities, and structure of FMOs. *Drug Metab. Rev.* **34**, 503–511
13. van Berkel, W. J. H., Kamerbeek, N. M., and Fraaije, M. W. (2006) Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *J. Biotechnol.* **124**, 670–689
14. Ziegler, D. M., and Poulsen, L. L. (1978) Hepatic microsomal mixed-function amine oxidase. *Methods Enzymol.* **52**, 142–151
15. Beaty, N. B., and Ballou, D. P. (1981) The reductive half-reaction of liver microsomal FAD-containing monooxygenase. *J. Biol. Chem.* **256**, 4611–4618
16. Beaty, N. B., and Ballou, D. P. (1981) The oxidative half-reaction of liver microsomal FAD-containing monooxygenase. *J. Biol. Chem.* **256**, 4619–4625
17. Siddiq, M. A., Hochberg, G. K., and Thornton, J. W. (2017) Evolution of protein specificity: insights from ancestral protein reconstruction. *Curr. Opin. Struct. Biol.* **47**, 113–122
18. Hochberg, G. K. A., and Thornton, J. W. (2017) Reconstructing Ancient Proteins to Understand the Causes of Structure and Function. *Annu. Rev. Biophys.* **46**, 247–269
19. Merkl, R., and Sterner, R. (2016) Ancestral protein reconstruction: Techniques and applications. *Biol. Chem.* **397**, 1–21
20. Voordeckers, K., Brown, C. A., Vanneste, K., van der Zande, E., Voet, A., Maere, S., and Verstrepen, K. J. (2012) Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. *PLoS Biol.* 10.1371/journal.pbio.1001446
21. Perez-Jimenez, R., Inglés-Prieto, A., Zhao, Z. M., Sanchez-Romero, I., Alegre-Cebollada, J., Kosuri, P., Garcia-Manyes, S., Kappock, T. J., Tanokura, M., Holmgren, A., Sanchez-Ruiz, J. M., Gaucher, E. A., and Fernandez, J. M. (2011) Single-molecule

paleoenzymology probes the chemistry of resurrected enzymes. *Nat. Struct. Mol. Biol.* **18**, 592–596

22. Nicoll, C. R., Bailleul, G., Fiorentini, F., Mascotti, M. L., Fraaije, M. W., and Mattevi, A. (2020) Ancestral-sequence reconstruction unveils the structural basis of function in mammalian FMOs. *Nat. Struct. Mol. Biol.* **27**, 14–24

23. Kumar, S., Stecher, G., Suleski, M., and Hedges, S. B. (2017) TimeTree: A resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819

24. Itagaki, K., Carver, G. T., and Philpot, R. M. (1996) Expression and characterization of a modified flavin-containing monooxygenase 4 from humans. *J. Biol. Chem.* **271**, 20102–20107

25. Hu, S. X. (2018) Hepatic Flavin-containing Monooxygenase and Aldehyde Oxidase Activities in Male Domestic Pigs at Different Ages. *Drug Metab. Lett.* **12**, 125–131

26. Uno, Y., Shimizu, M., and Yamazaki, H. (2013) Molecular and functional characterization of flavin-containing monooxygenases in cynomolgus macaque. *Biochem. Pharmacol.* **85**, 1837–1847

27. Siddens, L. K., Krueger, S. K., Henderson, M. C., and Williams, D. E. (2014) Mammalian flavin-containing monooxygenase (FMO) as a source of hydrogen peroxide. *Biochem. Pharmacol.* **89**, 141–147

28. Kawaji, A., Ohara, K., and Takabatake, E. (1993) An assay of flavin-containing monooxygenase activity with benzydamine N-oxidation. *Anal. Biochem.* **214**, 409–412

29. Fisher, B., Costantino, J. P., Wickerham, D. L., Cecchini, R. S., Cronin, W. M., Robidoux, A., Bevers, T. B., Kavanah, M. T., Atkins, J. N., Margolese, R. G., Runowicz, C. D., James, J. M., Ford, L. G., and Wolmark, N. (2005) 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

30. Chitra, M., Ernest, H., and David, K. (1993) Metabolism of the cancer agent tamoxifen : Flavin-Containing Monooxygenase-Mediated N-Oxidation. *Drug Metab. Dispos.* **21**, 657-661

31. Fiorentini, F., Romero, E., Fraaije, M. W., Faber, K., Hall, M., and Mattevi, A. (2017) Baeyer-Villiger Monooxygenase FMO5 as entry point in drug metabolism. *ACS Chem. Biol.* **12**, 2379–2387
32. Fiorentini, F., Geier, M., Binda, C., Winkler, M., Faber, K., Hall, M., and Mattevi, A. (2016) Biocatalytic characterization of human FMO5: Eneathing Baeyer-Villiger reactions in humans. *ACS Chem. Biol.* **11**, 1039–1048
33. Fabara, A. N., and Fraaije, M. W. (2020) An overview of microbial indigo-forming enzymes. *Appl. Microbiol. Biotechnol.* **104**, 925–933
34. Romero, E., Robinson, R., and Sobrado, P. (2012) Monitoring the reductive and oxidative half-reactions of a flavin-dependent monooxygenase using stopped-flow spectrophotometry. *J. Vis. Exp.* 10.3791/3803
35. Veeravalli, S., Phillips, I. R., Freire, R. T., Varshavi, D., Everett, J. R., and Shephard, E. A. (2020) Flavin-containing monooxygenase 1 catalyzes the production of taurine from hypotaurine. *Drug Metab. Dispos.* **48**, 378–385
36. Sofer, S. S., and Ziegler, M. (1978) Microsomal mixed-function amine oxidase. Oxidation products of piperazine-substituted phenothiazine drugs. *Drug Metab. Dispos.* **7**, 404–410.
37. Nagata, T., Williams, D. E., and Ziegler, D. M. (1990) Substrate specificities of rabbit lung and porcine liver flavin-containing monooxygenases: Differences due to substrate size. *Chem. Res. Toxicol.* **3**, 372–376
38. Ho, B. K., and Gruswitz, F. (2008) HOLLOW: Generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* **8**, 1–6
39. Parte, P., and Kupfer, D. (2005) 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
40. Alfieri, A., Malito, E., Orru, R., Fraaije, M. W., and Mattevi, A. (2008) Revealing the moonlighting role of NADP in the structure of a flavin-containing monooxygenase. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 6572–6577

41. Robinson, R., Badieyan, S., and Sobrado, P. (2013) C4a-hydroperoxyflavin formation in N-hydroxylating flavin monooxygenases is mediated by the 2'-OH of the nicotinamide ribose of NADP+. *Biochemistry*. **52**, 9089–9091
42. Dean, A. M., and Golding, G. B. (1997) Protein engineering reveals ancient adaptive replacements in isocitrate dehydrogenase. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 3104–3109
43. Expression Atlas query ensg00000010932 FMO1, last accession 2020-09-20. *EMBL-EBI*
44. Sambrook, J., Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*, third ed., CSHL press, New York
45. Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800
46. Project, C. C. (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50**, 760–763
47. Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214
48. Vonrhein, C., Tickle, I. J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., and Bricogne, G. (2018) 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
49. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674
50. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2256–2268
51. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53**, 240–255

52. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612

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- β -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_m Melting temperature

Figures

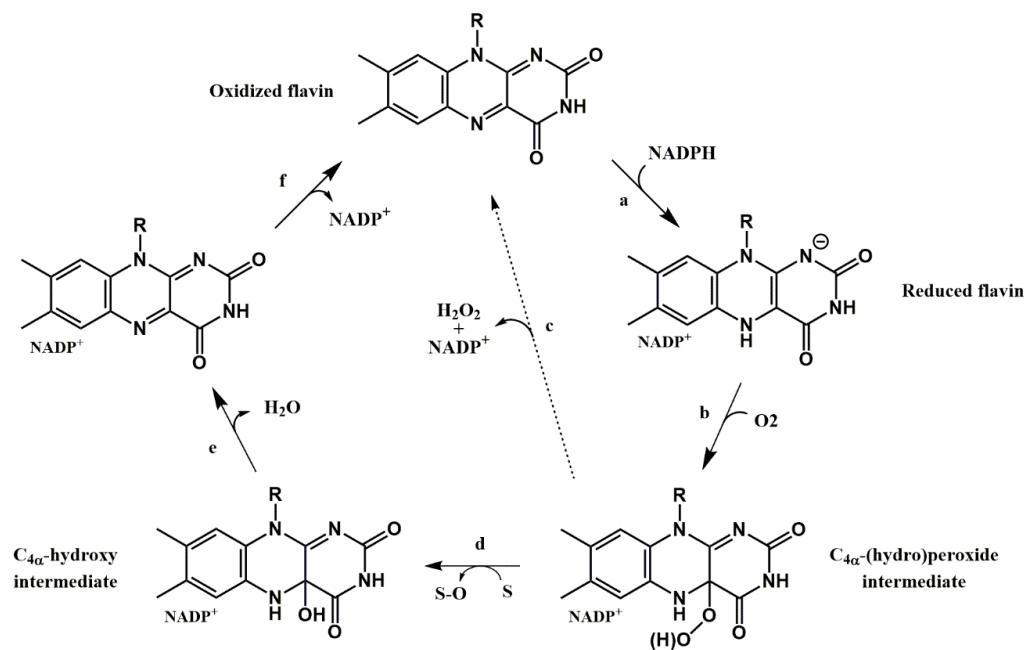


Figure A. Catalytic cycle of flavin-containing monooxygenases. a, binding of NADPH and subsequent reduction of the flavin. b, Reaction with molecular oxygen and C(4 α)-hydroperoxyflavin intermediate formation. c, Uncoupling: release of hydrogen peroxide and NADP⁺. d, Oxygen atom transfer to a substrate.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

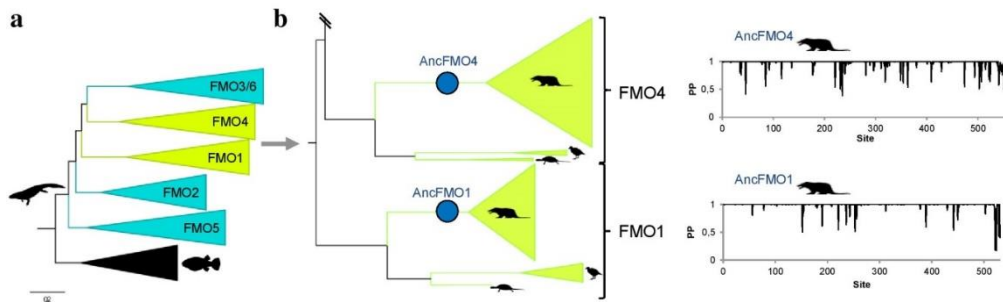


Figure 1. Ancestral sequence reconstruction of mammalian FMOs. a, Compressed tree of FMOs from jawed vertebrates. The tree depicts the explosion of FMOs in tetrapods. The five FMO clades are shown and those colored in green have been analyzed in this work. b, Close up of the FMO4 and FMO1 clades. Taxonomic distribution is depicted with silhouettes as follows: ancestral tetrapod (silhouette), bony fishes (silhouette), mammals (silhouette), aves (silhouette) and testudines (silhouette). Reconstructed mammalian ancestors are shown with blue circles. On the right the corresponding graphs of the posterior probability distribution per site are shown.

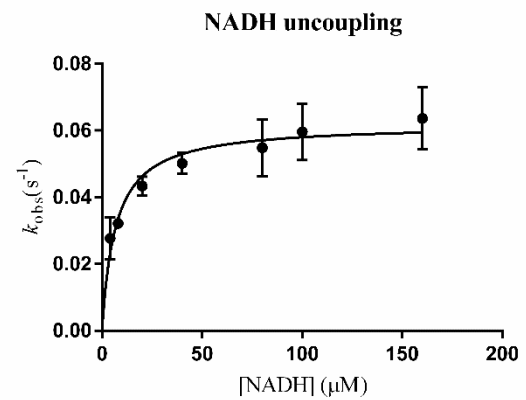
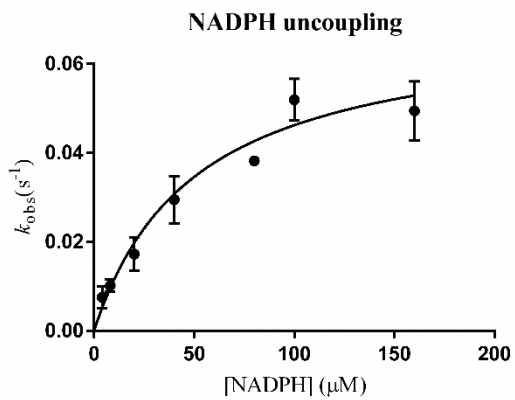
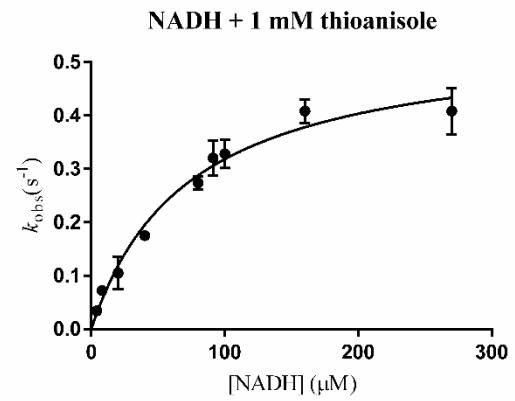
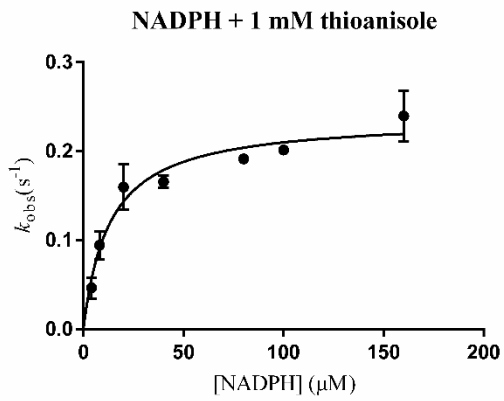
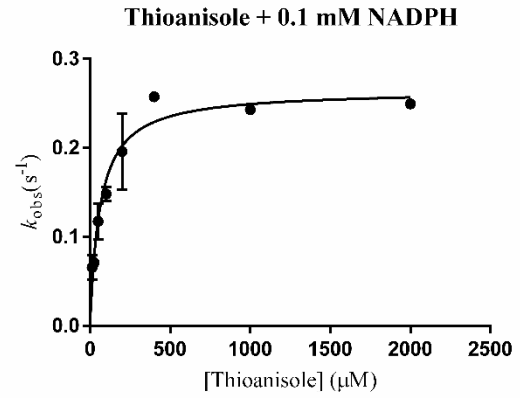
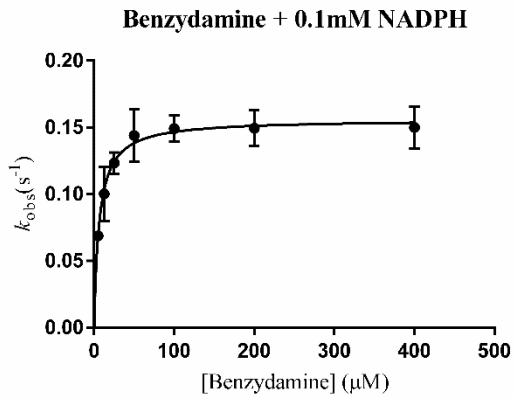


Figure B. Steady state kinetic plots for AncFMO1. Reactions were followed by measuring the NAD(P)H consumption at 340 nm and rates calculated using NAD(P)H extinction coefficient $6.22 \text{ mM}^{-1}\cdot\text{cm}^{-1}$. The plots were obtained by fitting the measured NAD(P)H depletion rates to GraphPad.

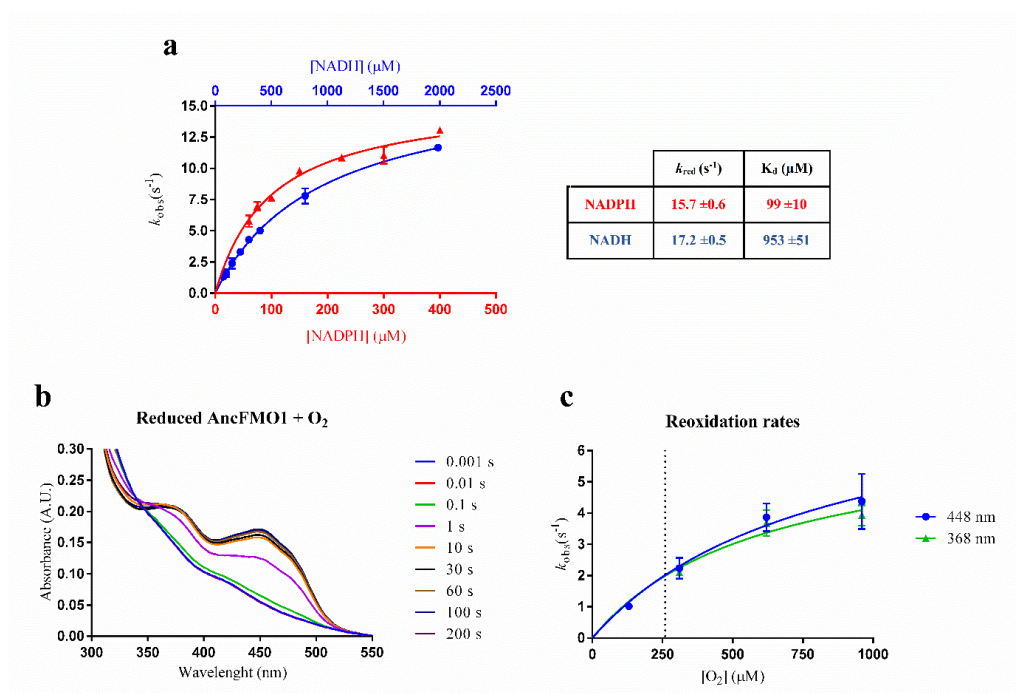


Figure 2. AncFMO1 stopped-flow kinetics. a, Reductive half reaction rates measured in anaerobic conditions with increasing concentrations of NADH (blue points) or NADPH (red triangles). The binding constants K_d and k_{red} values were calculated by fitting the reductive half-reaction to the Michaelis-Menten equation. 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). The dotted lines correspond to the atmospheric concentration of dioxygen ($260 \mu\text{M}$), present during steady state kinetics.

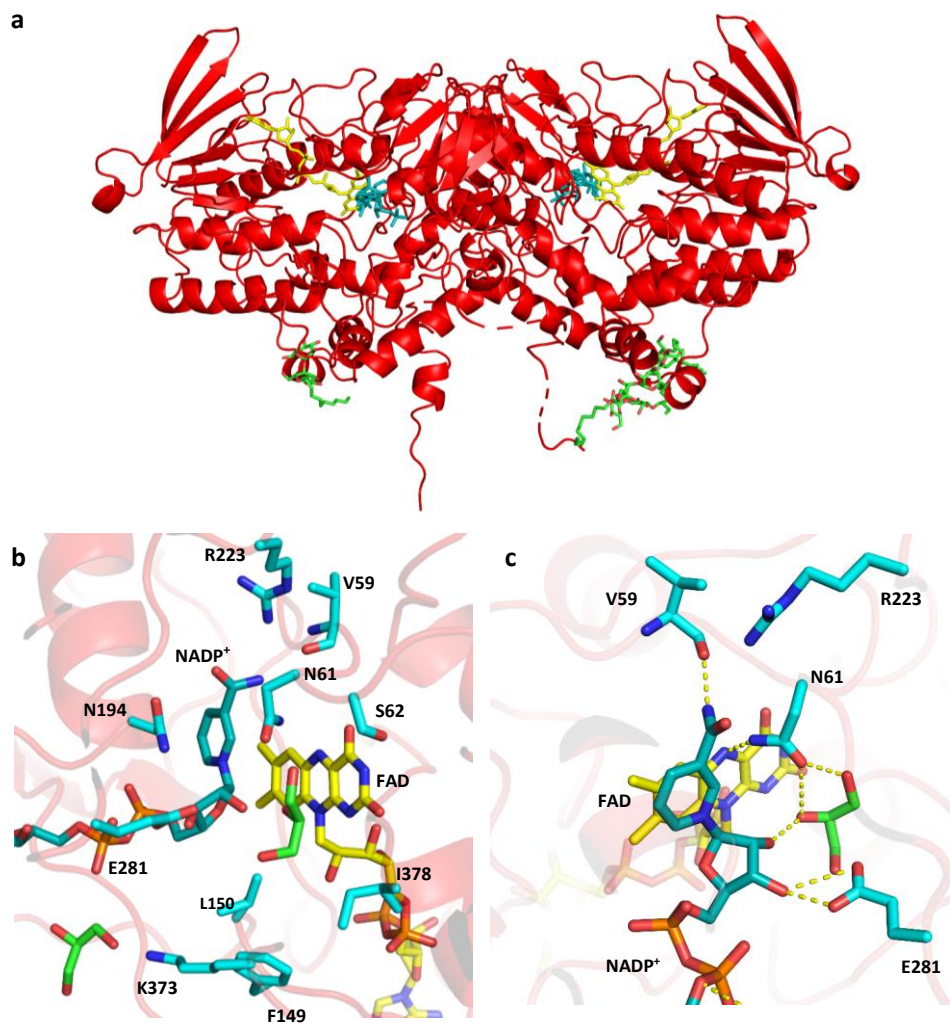


Figure 3. Crystal structure of AncFMO1 and its active site. a, Dimeric AncFMO1 with its partially mapped C-terminal helices pointing downwards towards the membrane. b, The active site of AncFMO1 in the presence of oxidized coenzyme, NADP⁺ and two glycerol molecules (green) is depicted, with the key residues labelled. c, Extensive hydrogen bond interactions between E281, N61, V59, FAD, NADP⁺ and a glycerol molecule are shown to illustrate key intermolecular interactions and potential substrate binding modes, represented by dashed yellow lines. FAD, NADP⁺ and DDM molecules are colored in yellow, cornflower blue and white respectively.

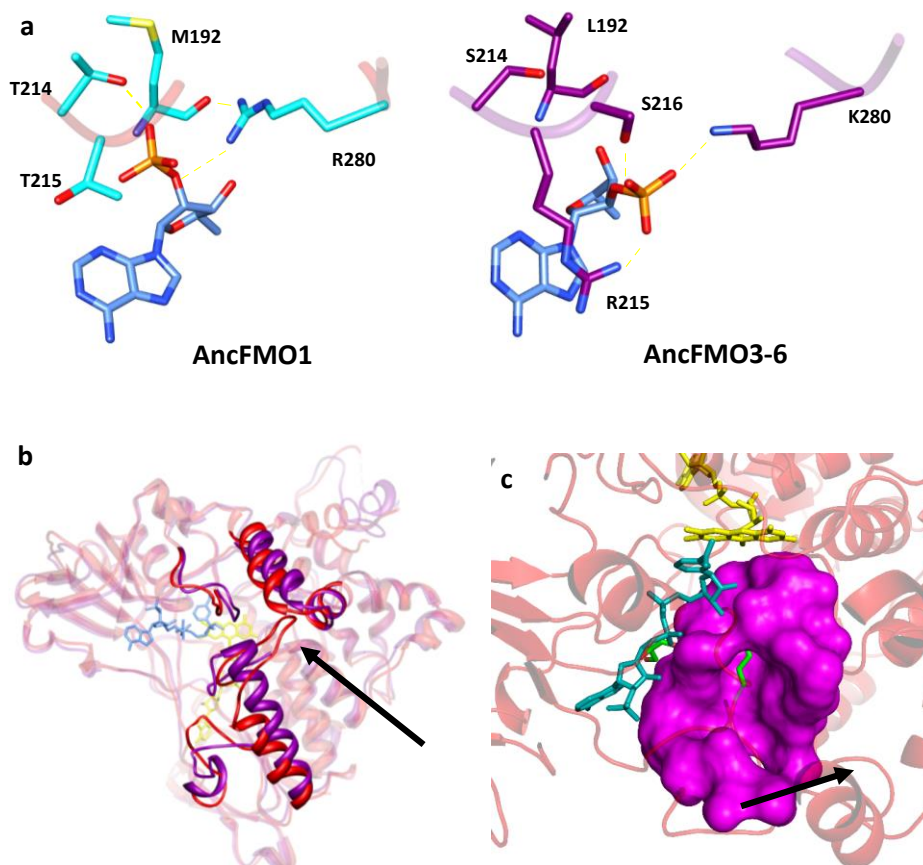


Figure 4. Unique structural features of AncFMO1. a, The differing 2'-phosphate binding site of NADP⁺ for AncFMO1 (left, side chains in cyan) compared to AncFMO3-6 (PDB:6SE3) (right, side chains in dark purple) is shown with key hydrogen bonding interactions shown as yellow dashed lines. NADP⁺ is shown in cornflower blue. b, The conformation adopted by residues 416-425 that reaches out towards the α -helical triad in a large arched conformation is indicated by the black arrow. Super positioning 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⁺ and glycerol molecules are shown in yellow, cornflower blue and green, respectively.

Tables

substrate	fixed substrate	k_{cat} (s ⁻¹)	K_M (μ M)	k_{cat}/K_M (s ⁻¹ /M)
^a benzylamine	0.1 mM NADPH	0.16 ±0.01	6.4 ±1.2	25000
^a thioanisole	0.1 mM NADPH	0.26 ±0.01	62 ±10	4200
^b NADPH	1.0 mM thioanisole	0.24 ±0.01	13 ±2.6	18500
^b NADH	1.0 mM thioanisole	0.55 ±0.04	73 ±12	7500
^c NADPH	-	0.07 ±0.01	51 ±15	1400
^c NADH	-	0.07 ±0.01	6.8 ±1.6	10300

Table 1. Steady state kinetic parameters for AncFMO1. ^aReactions were followed by measuring the NADPH consumption at 340 nm and rates calculated using NADPH extinction coefficient 6.22 mM⁻¹.cm⁻¹. ^bNAD(P)H steady-state kinetics in presence of 1.0 mM thioanisole, rates followed at 340 nm. ^cAncFMO1 uncoupling reaction with NAD(P)H, in absence of substrate.

AncFMO1 (PDB: 7AL4)		
Number of crystals	2	
	Aimless	Staraniso
Data collection ^a		
Resolution Range (Å) ^b	49.100 - 3.000 (3.041 - 3.000)	
Space Group	P 2 ₁	
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 _{merge}	0.245 (7.993)	-
CC _{1/2}	0.994 (0.028)	-
Refinement		
R-work (%)	-	22.8
R-free (%)	-	26.4
Number of non-hydrogen atoms ^c	-	17079
RMS (bonds) Å	-	0.004

**The Structural and Functional Characterization of mammalian Flavin-containing Monooxygenase
using Ancestral Sequence Reconstruction** **149**

RMS (angles) °	-	1.359
Ramachandran favoured (%)	-	82.6
Ramachandran allowed (%)	-	11.2
Ramachandran outliers (%)	-	6.2
Average B-factor	-	82

^aThe dataset was measured after merging reflection files from two crystals. ^bValues in parentheses are for highest-resolution shell. ^cThe asymmetric unit contained four protein molecules 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 Staraniso⁴⁷. The refined data are derived from the anisotropic corrected data. Values in parentheses are for the highest-resolution shell. The asymmetric unit contained 4 protein molecules for AncFMO1.

Supplementary Information

Ancestral reconstruction of mammalian FMO1: unique structural features explain its catalytic properties

Gautier Bailleul^a, Callum R. Nicoll^b, María Laura Mascotti^{a,c}, Andrea Mattevi^b, Marco W. Fraaije^{a,*}

^a Molecular Enzymology group, University of Groningen, Nijenborgh 4, 9747AG, Groningen, The Netherlands

^b Department of Biology and Biotechnology “Lazzaro Spallanzani”, University of Pavia, Pavia, Italy

^c IMIBIO-SL CONICET, Facultad de Química Bioquímica y Farmacia, Universidad Nacional de San Luis, San Luis, Argentina

Figure S1. Jawed vertebrates phylogeny of FMOs. 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/>.

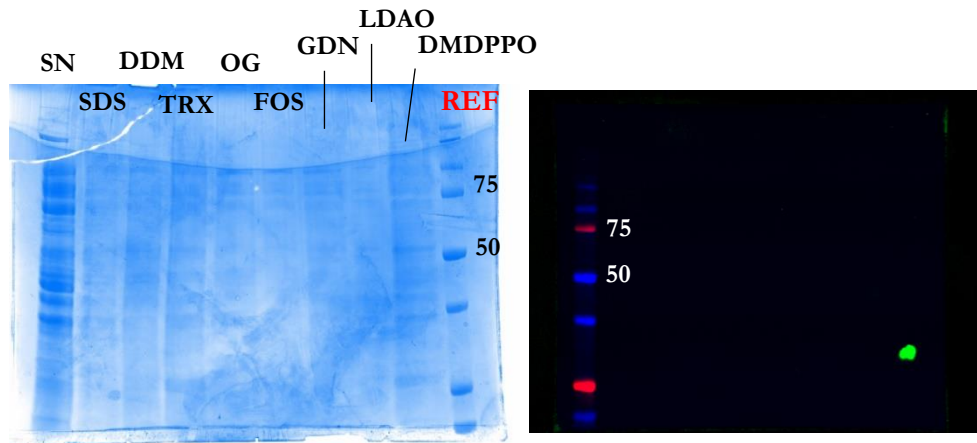


Figure S2: AncFMO4 expression and solubilization screening. Left, SDS-PAGE corresponding to the various detergents used to solubilize AncFMO4 from the membranes, including the aqueous supernatant. Right, Western-blot of the SDS-PAGE, with the image flipped with the lanes in reverse. Detergents used; SN (supernatant), SDS (sodium dodecyl sulfate), DDM (dodecyl-beta-maltoside), TRX (Triton-X 100), OG (octyl glucoside), FOS (FOS-Choline 8), GDN (glycodiosgenin), LDAO (lauryldimethylamine oxide) and DMDPPO (dimethyldecyl phosphine oxide). REF refers to molecular weight markers with weights 75 and 50 indicated in kDa. For clarity, certain lanes are indicated with black lines.

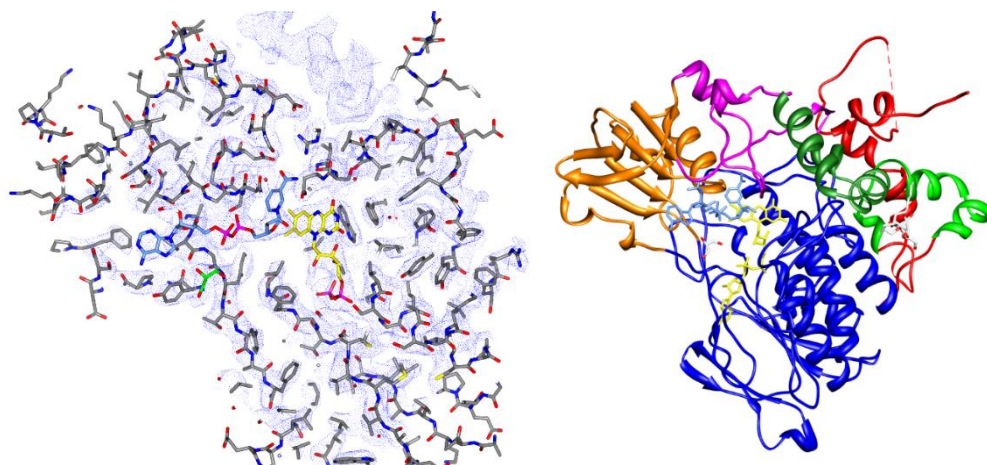


Figure S3. Crystal structure of AncFMO1. **Left**, Electron density maps of the AncFMOs. The depicted 2Fo-Fc maps (shown in blue) were calculated by averaging the electron density maps obtained after molecular replacement, and anisotropic correction using Staraniso. The crystal structure of AncFMO3-6 (PDB: 6SE3) was used as model for molecular replacement after the scaled reflections had been corrected for anisotropy using Staraniso. FAD and NADP⁺ are shown in yellow and cornflower blue respectively. The contour level is 1.4 σ . 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⁺ are colored in yellow and cornflower blue, respectively.

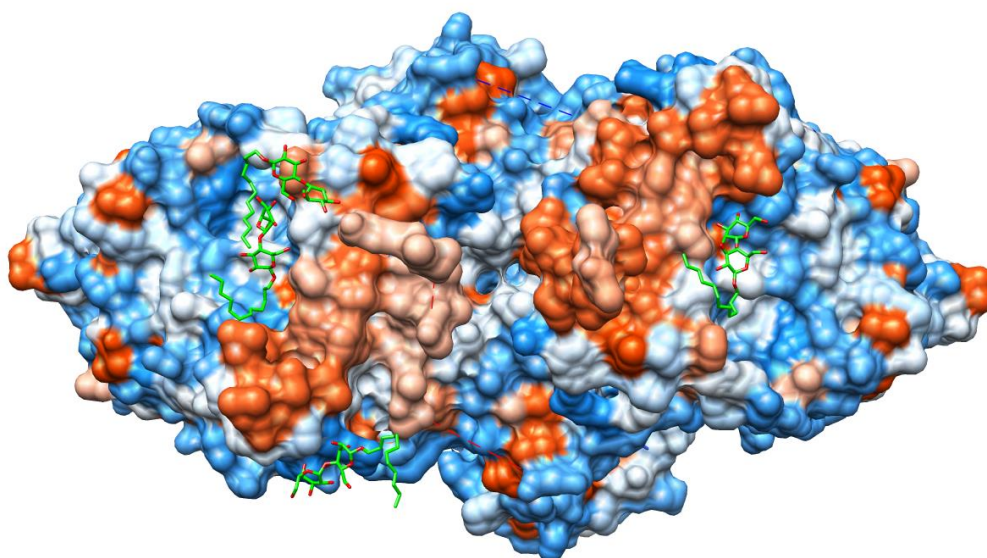


Figure S4. Charged surface distribution for the crystallographic dimer of AncFMO1. Distribution of charge around the surface of AncFMO1, with orange, white and blue representing hydrophobic, neutral and hydrophilic residues, respectively. Darker and lighter colors portray the extent of the characteristic, 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.

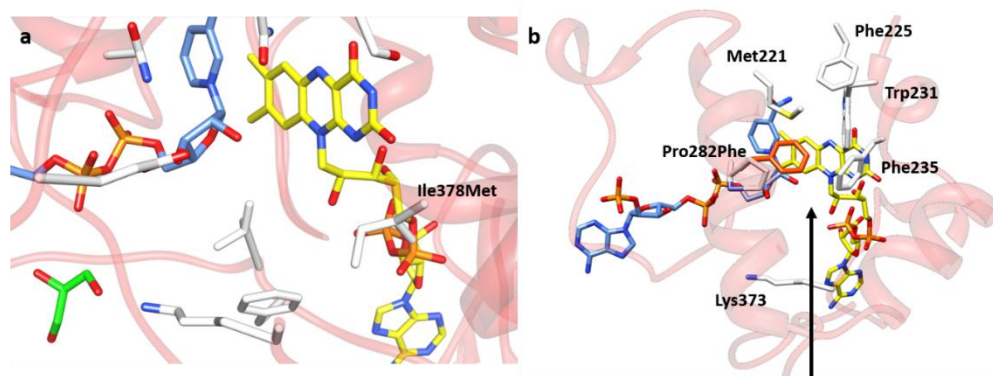


Figure S5. Key residue changes between AncFMO1 and hFMO1. **a**, The active site of AncFMO1 is shown with emphasis directed towards Ile378. This site is a methionine residue in hFMO1. **b**, The trajectory of the hydrophobic substrates with respect to the α -helical triad is shown by a black arrow. FAD, NADP⁺, Glycerol, AncFMO1 residues and theoretical hFMO1 residues are shown in yellow, cornflower blue, green, white and orange, respectively.

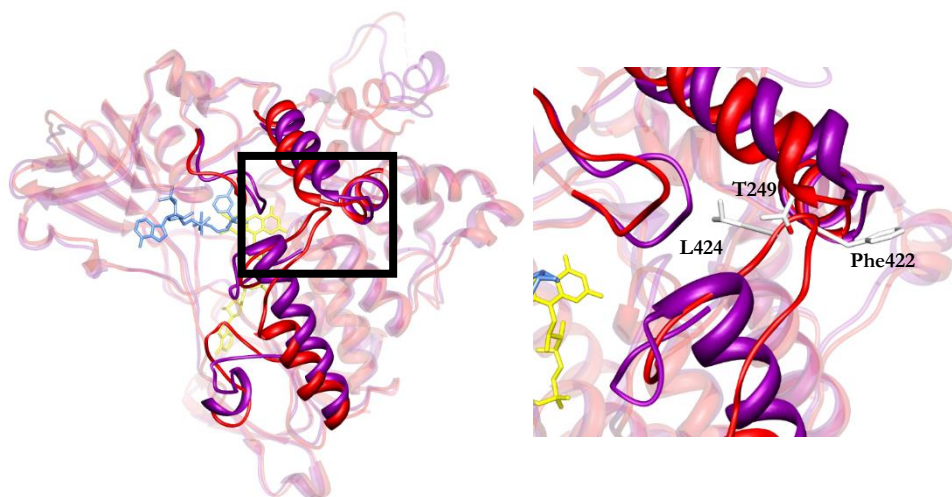


Figure S6. Overarching loop. The AncFMO1 monomer (red) is superimposed with the monomer of AncFMO3-6 (dark magenta). The unique loop feature exhibited by AncFMO1 is shown in the expanded black boxes. FAD, NADP⁺ and residue sidechains are shown in yellow, cornflower blue and white, respectively.

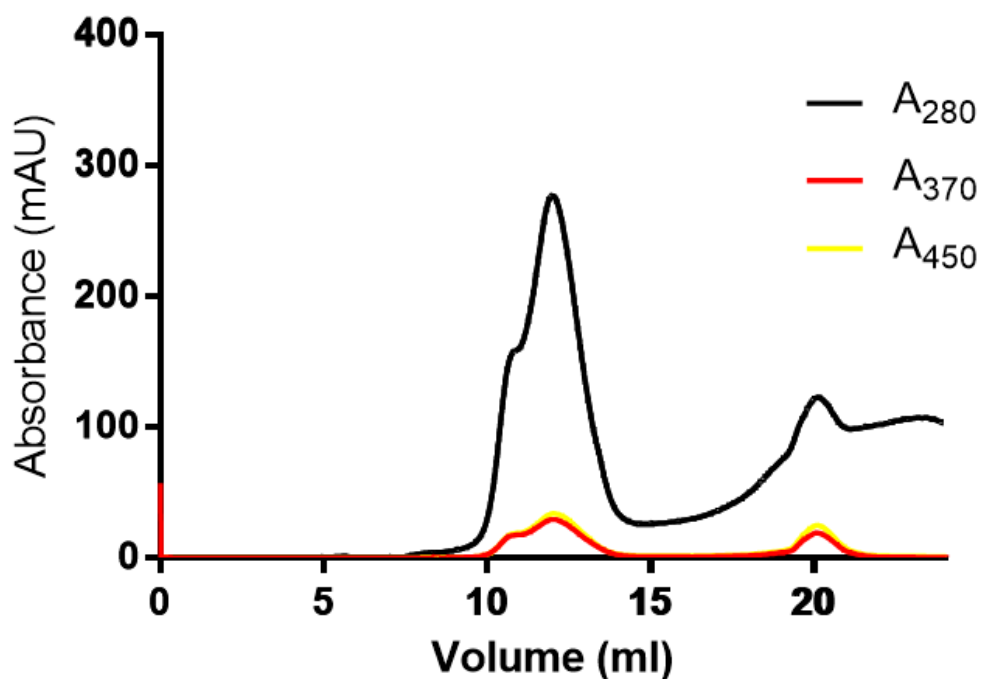


Figure S7. Gel filtration chromatogram of AncFMO1. Previously characterized AncFMOs eluted in a bell-shaped manner between 10.5 and 11 mL(2), as a monodisperse peak. AncFMO1 exhibits two elution volumes at 10.8 and 12 mL represented by peaks A and B respectively. These two peaks overlap with, A, acting 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 blue, pink and red represent wavelengths 280, 450 and 370 nm, respectively, measured in absorbance, mAU. Band C corresponds to the excess of FAD added to the protein solution before size exclusion chromatography.

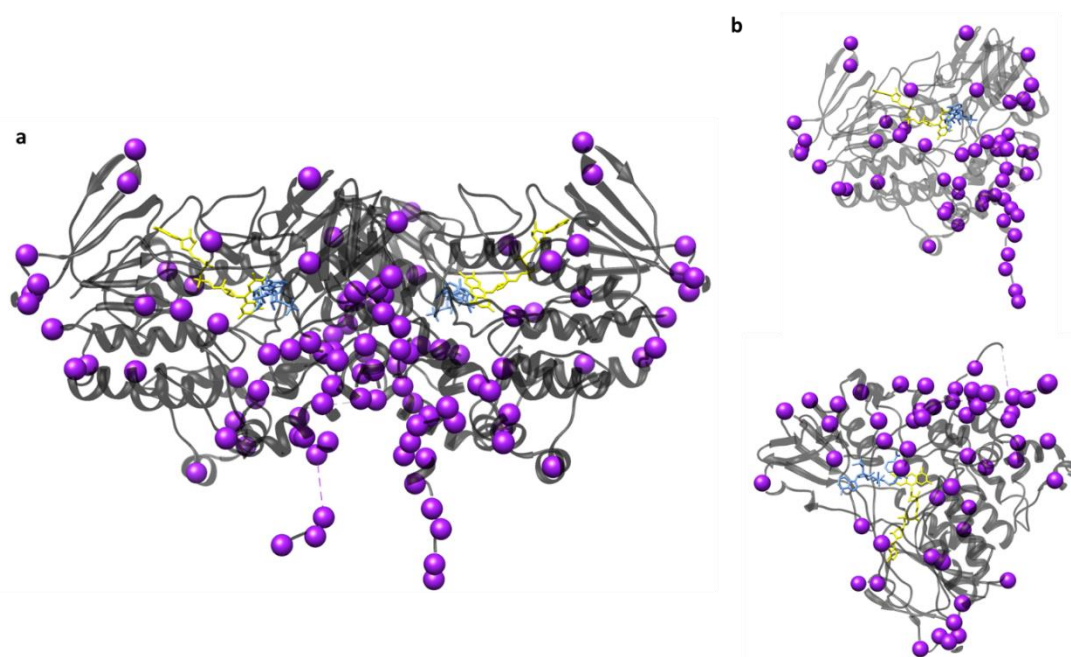


Figure S8. Differing residues between AncFMO1 and hFMO1. **a**, The differences between AncFMO1 and hFMO1 with respect to the crystallographic dimer are shown with purple spheres. AncFMO1 and hFMO1 share a percentage sequence identity and percentage sequence similarity of 89.9% (54 residues) and 94.2% (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.

AncFMO1

>AncFMO1m

MAKRVAIVGAGVSGLASIKCCLEEGLEPTCFERSDDLGLWRFTEHVEEGRASLYKSVVNSNC
 KEMSCYSDFPFPEYDYPNYVPNSQFLEYLKMAYANRFNLLKHIQFKTKVCSVTKCPDFTVTGQWE
 VVTQHEGKQESAI F DAVMVCTGFLTDPYLP L D S F P G I N T F K G Q Y F H S R Q Y K H P D I F K D K R V L V
 VGMGNSGTDIAVEASHLAKKVFLSTTGGAWVMSRVFDSGYPWDMVFTTRFQNMRLRNSLPTPIV
 TWLMARKMNSWFNHANYGLVPEDRTQLREPVLNDELPGCIITGKVLIKPSIKEVKENSVIFNN
 TPKEEPIDIIVFATGYTFAFPFLDESUVKVENGOASLYKYIFPAHLPKPTLAVIGLIKPLGSI
 IPTGETQARWAVRVLKGINKLPPQSVMIIEEVNARKENKPSGFGLCYCKALQSDYITYIDELLT
 YINAKPNLLSMLLTDPRLALTIFFGPCTPYQFRLTGPGKWE GARNAILTQWDRTFKVTKTRIV
 QESPSPFASLLKLLSLPVLLLALLLMC

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

>AncFMO4m

MAKRVAVIGAGVSGLSSIKCLDEDLEPTCFERSDDFGGLWKFTSSKDG MTRVYRSLVTNVC
KEMSCYSDFPFQEDYPNFMNHAKFWDYLREFAEHFDLLKYIQFKTTVCSVTKCPDFSETGQWD
VVTETEGKQDRAVFDAMVCTGHFLNPHLPLESFPGIHKFKGQILHSQEYRTPEAFQGKRVLV
IGLGNTGGDIAVELSR TAAQVFLSTRGTWVLSRSSDGGYPFNMMTTRCHNFIAQVLPSCFL
NWIQERQMKNRFNHENYGLSITKGKKKKAIVNDELPTC ILCGTVTVKTSVKEFTETS AVFEDG
TVEANIDVVI FT TGYTFSFPFLEEPLKSLCTKKIFLYKRVFPNLEKATLAI IGLISLTGSIL
AGTELQARWATR VFKGLCKIPPSQKLMAEVTKKEQLIKRGVIKDT SQDKLDYISYMDLAACI
GAKPNIPLLF LKDPRLAWEVFFGPCTPYQYRLVGP GKWDGARNAILTQWDR TLKPLKTRIVAD
SSKPASMSHYLKVWGAPVLLASLL LICKSSLFLKLV RDKLQDRISPYLISLWRGS

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

Data S1. AncFMO sequences and lists of ambiguously reconstructed sites.

Accession code	Organism	Taxonomic rank	FMO clade
L5LMT2	<i>Myotis davidii</i>	Mammalia	FMO3/6
W5PRR0	<i>Ovis aries</i>		
Q8HYJ9	<i>Bos taurus</i>		
L818K7	<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>		
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 sabaeus</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 sabaesus</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>Dasybus novemcinctus</i>		
XP_006887525	<i>Elephantulus edwardii</i>		
XP_004706886	<i>Echinops telfairi</i>		

The Structural and Functional Characterization of mammalian Flavin-containing Monooxygenase
using Ancestral Sequence Reconstruction

167

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>	Aves	
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>	Amphibia	
F7CV72	<i>Xenopus tropicalis</i>		
XP_005245102	<i>Homo sapiens</i>	Mammalia	FMO4
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 sabaesus</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>		
F6ZTZ1	<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>		

The Structural and Functional Characterization of mammalian Flavin-containing Monooxygenase
using Ancestral Sequence Reconstruction

171

Q8K4B7	<i>Rattus norvegicus</i>		
A0A1U8BUR7	<i>Mesocricetus auratus</i>		
L9KV99	<i>Tupaia chinensis</i>		
A0A1S2ZVIO	<i>Erinaceus europaeus</i>		
A0A1V4JY45	<i>Patagioneas fasciatus</i>		
A0A093FGE7	<i>Gavia stellata</i>	Aves	
U3I700	<i>Anas platyrhynchos</i>		
M7BBI8	<i>Chelonia mydas</i>	Testudines	
K7FTQ9	<i>Pelodiscus sinensis</i>		
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>	Mammalia	FMO1
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 sabaesus</i>		
A0A2K6S2A8	<i>Saimiri boliviensis</i>		
A0A2K5DNS9	<i>Aotus nancymaae</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>		
L9KVX0	<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>	Aves	
A0A091LMM3	<i>Chlamydotis macqueenii</i>		
A0A0A0B033	<i>Charadrius vociferus</i>		
A0A091JJZ4	<i>Egretta garzetta</i>		
A0A091KNM4	<i>Colius striatus</i>		
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>	Mammalia	FMO2
G1RYT1	<i>Nomascus leucogenys</i>		
Q28505	<i>Macaca mulatta</i>		
A0A023JBW5	<i>Macaca fascicularis</i>		
F7FJA6	<i>Callithrix jacchus</i>		

XP_007987720	<i>Chlorocebus sabaesus</i>		
XP_011826204	<i>Mandrillus leucophaeus</i>		
XP_017722374	<i>Rhinopithecus bieti</i>		
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>		

The Structural and Functional Characterization of mammalian Flavin-containing Monooxygenase
using Ancestral Sequence Reconstruction

177

Q5BKK2	<i>Xenopus tropicalis</i>		
F6TLP5	<i>Xenopus tropicalis</i>		
F6TKY1	<i>Xenopus tropicalis</i>		
NP_001075714	<i>Oryctolagus cuniculus</i>	Mammalia	FMO5
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>		

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>		

The Structural and Functional Characterization of mammalian Flavin-containing Monooxygenase
using Ancestral Sequence Reconstruction

179

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>		
A0A1S3A3I9	<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>	Aves	
A0A091EER2	<i>Corvus brachyrhynchos</i>		
H1A3R9	<i>Taeniopygia guttata</i>		
U3JME7	<i>Ficedula albicollis</i>		
A0A091UWC8	<i>Nipponia nippon</i>		
A0A087QQJ0	<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>	Amphibia	
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>		

H3A7E3	<i>Latimeria chalumnae</i>	Sarcopterygii/ Coelacanthimorpha	
V9KV64	<i>Callorhinchus milii</i>	Chondrichthyes	
V9KI45	<i>Callorhinchus milii</i>		
C3ZT14	<i>Branchiostoma floridae</i>	Cephalochordata	

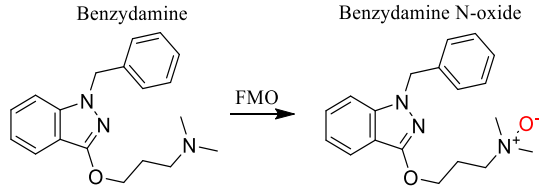
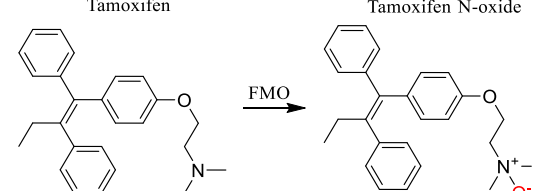
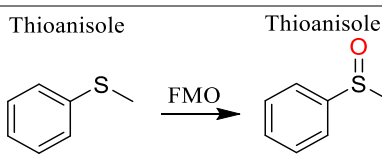
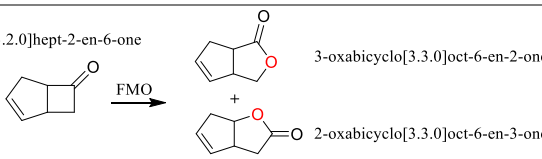
Table S1. Jawed vertebrates FMOs dataset. The protein accession codes (Genbank or Uniprot) are given for each of the sequences collected with the corresponding species names. The taxonomic rank is given as the class or order (for terrestrial tetrapods) and superclass/class or phylum (for the root sequences). 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.

The dataset was constructed according as follows: (i) previously experimentally characterized sequences of the different FMO paralogs (*e.g. Homo sapiens* FMOs) were employed as queries in BLASTp searches in non-redundant protein sequences (nr) or 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 (1). Specifically the 26 orders of mammals were carefully scrutinized. (ii) Partial sequences or poor quality ones were manually excluded. (iii) All collected sequences were gathered and analyzed in a MSA, 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 to build a representative and non-redundant dataset (*i.e.* a robust dataset) including 365 sequences.

protein	melting temperature (T_m) / °C	
	no NADP ⁺	200 μ M NADP ⁺
AncFMO1	47	52
AncFMO2(2)	53	70
AncFMO3-6(2)	60	66.5
AncFMO5(2)	55	59
hFMO3(2)	44.5	44.5
hFMO5(2)	49	48.5

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 Tycho™MNT.6 system for AncFMO1 (see materials and methods). The hFMO3 melting temperature was obtained using Triton X-100 reduced.

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 available.

type of oxidation	Reaction	conversion ^a	ee (%) ^b
N-oxidation	<p>Benzydamine</p>  <p>Benzydamine N-oxide</p>	>99%	Not applicable.
	<p>Tamoxifen</p>  <p>Tamoxifen N-oxide</p>	>99%	Not applicable.
S-oxidation	<p>Thioanisole</p>  <p>Thioanisole sulfoxide</p>	>99%	96 (R)
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%	Not analyzed.

1. S. Kumar, G. Stecher, M. Suleski, and S.B. Hedges. (2017). [TimeTree: a resource for timelines, timetrees, and divergence times](#). *Molecular Biology and Evolution* **34**, 1812-1819, DOI: 10.1093/molbev/msx116.
2. Nicoll, C. R., Bailleul, G., Fiorentini, F., Mascotti, M. L., Fraaije, M. W., and Mattevi, A. (2020) Ancestral-sequence reconstruction unveils the structural basis of function in mammalian FMOs. *Nat. Struct. Mol. Biol.* **27**, 14–24.

3.6 PHYLOGENETIC ANALYSIS OF THE FMO PROTEIN FAMILY IN JAWED VERTEBRATES

The differing enzymatic activities between the FMO5, that acts as a non-conventional BVMO, and FMOs 1-3 that have been characterized to perform standard FMO activities, raises an intriguing question – what functional feature arose first? The initial analysis performed in order to reconstruct the mammalian ancestors conveys the early divergence of FMO5 from the other paralogs. Paralogs 1-6 (excluding 5) however, are all derived from a common ancestor and are more similar. Considering that FMO5 emerged first suggests that the ancestor of all FMOs could in fact be more BVMO-like and that the FMO activity emerged later via subsequent gene duplication events. Alternatively, the ancestor could possess mixed functionality, being able to convert both forms of substrate, with optimizations and specializations taking place through successive gene duplications. Herein this section, I discuss the curation of the phylogenetic tree, the distribution of FMOs throughout the species tree and the ASR process used to generate key ancestral sequence intermediates.

3.6.1 Phylogenetic Analysis of Flavin-containing monooxygenases

All the AncFMO sequences were generated using the phylogeny described in section 3.3 of the Results section. These constructs represented the mammalian ancestral FMOs that existed approximately 177 million years ago (Mya) [Kumar, 2017]. Importantly for the reconstructions, the number of mammalian sequences collected in order to determine the mammalian ancestor was substantial and reliable for reconstruction. Moreover, to generate ancestral sequences that represent the first emerging ancestors of the FMOs collectively, additional sequence information was collected. These include species derived from aves, reptiles, testudines (including turtles and crocodiles) and amphibia in order to improve the sequence representation of these phyla (see materials and methods). This step is crucial because a detailed and broad dataset of sequences will ensure that the ancestral sequences reconstructed and reliably represent the ancestors being depicted. The final obtained tree displayed a topology in agreement with the previously obtained by ML (Appendix A). For simplicity, the phylogenetic tree was condensed down into the schematic shown below (Figure 3.6.1.1).

The nodes labelled according to Figure 3.6.1.1. correspond to the ancestral states that were reconstructed and extracted for biochemical characterization. These ancestral states are described as Elder FMOs (EldFMOs). The EldFMO corresponding to the ancestor of FMOs 1, 2, 3, 4, 5 and 6, is denoted as EldFMO1-5. This ancestor corresponds to the first FMO to emerge in tetrapods. However, it should be noted that the Amphibia Y FMO could in fact be derived from the first gene duplication that took place (Figure 3.6.1.1). However, with this FMO currently being uncharacterized, we focussed our investigation towards the classified FMOs. The posterior probability for the reconstruction of EldFMO1-5 was high, 0.96, and represents a reliable sequence. EldFMO1-5 duplicated and produced the ancestor of FMOs 1-4, here denoted as EldFMO1-4, and eventually the

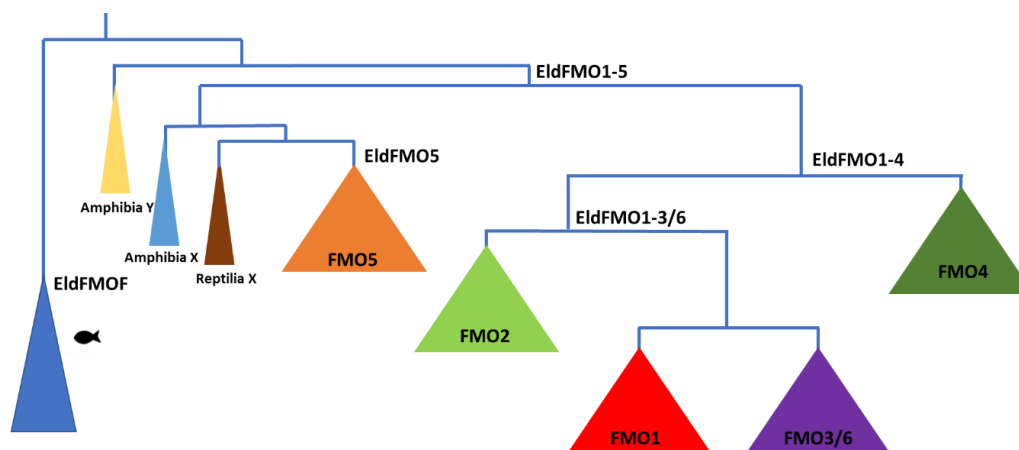


Figure 3.6.1.1: Schematic representation of the evolution of the Flavin-containing monooxygenases in vertebrates. Clades representing the FMO paralogs 5, 4, 3/6, 2 and 1 are shown in orange, dark green, purple, light green and red triangles, respectively. Ancient FMOs extracted for characterization and their respective origins are described in the tree. These Ancient FMOs are denoted as ElderFMOs - EldFMOs. EldFMOs extracted include: EldFMO1-5, EldFMO1-4, EldFMO1-3/6, EldFMO5 and EldFMOF. Unclassified FMO sequences are claded as triangles and labelled according to their distribution and origin with X or Y. For example, FMO clades that diverge after the duplication event described by node 541, are denoted as X, whilst those before are labelled as Y. The preceding term highlights their location within the species tree. For example, Reptilia X are FMOs found in reptilia and diverged after the EldFMO1-5-duplication event. Clades of Amphibia Y, Amphibia X and Reptilia X are shown as yellow, light blue and brown triangles, respectively. The root of the tree is shown as a dark blue triangle that represents fish FMOs represented with a fish silhouette.

ancestor of all FMO5s, EldFMO5. EldFMO1-4 underwent another gene duplication that generated the ancestor of all FMO4s and the ancestor of FMOs 1-3/6, herein called EldFMO1-3/6. EldFMOs 1-4, 1-3/6 and 5 conveyed high posterior probabilities of 0.94, 0.95 and 0.97, respectively. Fish only possess one FMO and its ancestor, EldFMOF (posterior probability of 0.91), emerged before all other FMOs and depending on the results obtained from the other ancestral states, the sequence may be useful for validating the emergence of the enzymatic activities.

These ancestral states were selected because they provide two clear pathways to evaluate the emergence of BVMO and FMO activities (Figure 3.6.1.2). EldFMO1-5 was first selected in order to determine whether the first FMO in tetrapods displayed BVMO or FMO activity. With this clarified, the ancestral intermediates EldFMO1-4 and EldFMO5 would then be biochemically characterized to determine which branch generated the switch in activity (Figure 3.6.1.2). If each EldFMO displayed similar activity, EldFMO1-3/6 would then be evaluated. Considering that FMOs 1, 2 and 3-6 all show typical FMO activity,

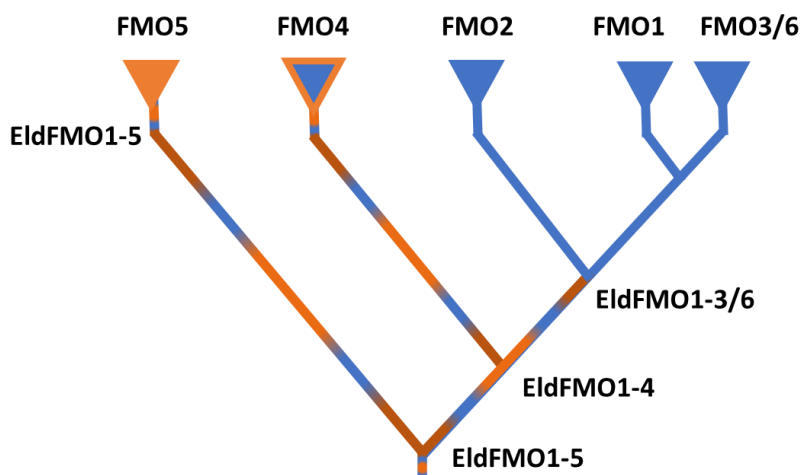


Figure 3.6.1.2: Evolution of the BVMO and FMO activities in jawed vertebrates. In this simplified tree, the evolution of the FMOs are shown with respect to the EldFMOs selected for biochemical characterization. The orange and blue colours describe BVMO and FMO activities respectively. Branches and clades coloured in blue and orange represent unknown activity. FMO5 is shown in orange as human FMO5 was shown to display BVMO activity. FMOs 1, 2 and 3/6 were shown to perform conventional FMO oxidation and are coloured in blue. FMO4 has not been characterized to date and remains multicoloured.

EldFMO1-3/6 is highly likely to be a canonical FMO. If all EldFMOs display FMO activity, the BVMO activity displayed in AncFMO5 would have emerged within the emergence of the mammalian FMO5. The changes taking place between AncFMO5 and EldFMO5 would then be evaluated to determine the origin of the BVMO functionality. The ancestral sequences of the EldFMOs are shown below in Table 3.6.1.1.

3.6.2 Evolution of the Flavin-containing monooxygenases

The final phylogeny was scrutinized to evaluate whether the addition of more sequences influenced the final tree and the gene duplication events. In the previous tree, FMO5 emerged first, followed by FMO2, FMO1, FMO4 and finally FMO3 and FMO6 (which duplicated only from mammalian FMO3). The new phylogeny however has rearranged the gene duplication events to different paralogs (Figure 3.6.1.1). In the new scenario, FMO4 was the second FMO to emerge, followed by FMO2, FMO1 and then FMO3/6. Intriguingly, the tree portrays that FMOs 1-3/6 are more similar to one another with FMO4 emerging earlier. Current information regarding the enzymatic activity of FMO4 is bereft in literature due to lack of known expression conditions. Therefore, it is difficult to predict whether the enzymatic activity of the ancestor of FMO 1-3/6 is the same as the ancestor of FMO4.

Table 3.6.1.1: Elder FMO sequences to be reconstructed based on the expanded phylogeny

Ancestral Node	Sequence
EldFMO1-5	MAKRVAVIGAGASGLTAIKCCLEDEGLEPTCFERSDDIGGLWRFKENPEDGRASIYKSVII NTSKEMMCYSDFPI PDDFPNYMHNSKIMEYFRMYAKHFDDLKYIRFKTTVCSVKKRPDFS TTGQWDVVTE TDGKQESAI FDAVLVCTGHHTDPHLPLDSFPGIEKFKGQYFHSRDYKNPE EFQGKRVLVIGIGNSGGDIAVELSRTAKQVFLSTRRGSWILNVRVSDNGYPLDMVHSTRFK NLLKHILPSSLVNWMAEKKMNQRFNHNENYGLKQHRIFSQHPMVNDLDPNRIISGTVLVK PNVKEFTETS AIFEDGTVEENIDVVI FATGYSFSFPFLEESVLKVQNNKVS LYKYVFPFH LEKPTLAVIGLIQPLGAIMPISELQARWATR VFKGLNKLPSANTMMADI AKKKEEMEKRY VTSQRHTIQVDY IEYMDELASLIGVKPNLLSFLTDPKLALEVFFGCPCTPYQYRLTGPGK WDGARKAILTQWDRI IKPTKTRVLQDNSSSSSMPLLLKILGLALLSALYYFL
EldFMO1-4	MAKRVAVIGAGVSGLTSIKCCLEDEGLEPTCFERSDDIGGLWRF TENPEDGRASIYKSVIT NTSKEMMCYSDFPFPEDFPNYMHNSKFLEYLRMYAKHFDDLKYIQFKTTVCSVKKRPDFS TTGQWDVVTE TDGKQESAI FDAVMVCTGHHTDPHLPLDSFPGIEKFKGQYFHSREYKNPE EFQGKRVLVIGIGNSGGDIAVELSRTAKQVFLSTRRGSWVLSRVSDNGYPLDMVFTTRFQ NLLRNILPSSLVNWMT EKKMNQWFNHNENYGLVQNRILSQHPVNDLDP SRIICGTVVVK PNVKEFTETS AIFEDGTVEENIDVVI FATGYSFSFPFLEESVIKVENNKVS LYKYVFPFH LEKPTLAVIGLIQPLGAIMPTAELQARWATR VFKGLCKLPSANTMMADI AKKKEKMIKRF VTSQRNTIQTDY IEYMDELASF IGVKPNILSFLTDPKLALEVFFGCPCTPYQYRLTGPGK WDGARNAILTQWDRILKPTKTRVLQDDSSSSNSMPFLLKILGLALLSALYYFL
EldFMO1-3/6	MAKRVAVIGAGVSGLTSIKCCLEDEGLEPTCFERSDDIGGLWRF TENVEDGRASIYKSVIT NTSKEMMCYSDFPFPEDFPNYMHNSKFLEYLRMYAKHFDDLKYIQFKTTVCSVKKCPDFS TTGQWDVVTE TDGKQESAI FDAVMVCTGHHTDPYLPPLDSFPGIEKFKGQYFHSREYKNPE GFQGKRVLVIGMNGSGADI AVELSHTAKQVFLSTRRGSWVMSRVSDNGYPWDMVFTTRFQ NLLRNVLPSLVNWMTEKKMNQWFNHNENYGLVQNRITLMKEPVFNDDLDP SRIICGTVVVK PNVKEFTETS AIFEDGTVEENIDVVI FATGYTFSFPFLDESVIKVENNKVS LYKYVFPFH LEKPTLAVIGLIQPLGAIMPTAELQARWATR VFKGLCKLPSANTMMEDI AKKKEKKIKWF GTSQSNTLQTDY IEYMDELASF IGAKPNILSFLTDPKLALEVFFGCPCTPYQYRLTGPGK WDGARNAILTQWDRILKPTKTRVLQDNSSSSNSVPFLLKILGLALLSALYYFL
EldFMO5	MAKRVAVIGAGSSGLTAIKCCLEDEGLEPTCFERSDDIGGLWRFKENPEDGRASIYKSVII NTSKEMMCFSDFPI PDDFPNYMHNSKIMEYFRMYAKHFDDLKYIRFKTTVCSVKKRPDFS TTGQWDVVTE TDGKQESAI FDAVLVCTGHHTNPHLPLDSFPGIEKFKGQYFHSRDYKNPQ EFQGKRVI VIGIGNSGGDL AVELSHTAKQVFLSTRRGAWILNVRVSDNGYPLDVVLSTRFK NLLKQILP TSMVNRWAEKKMNARFNHNENYGLKQHRILSQHPVNDLDPNRIISGKVLVK PNVKEFTETA AIFEDGTVEENIDVVI FATGYSFSFPFLEESVLKVQNNKVS LYKFVFPFH LEKPTLAI IGLIQPLGAIMPISELQARWATR VFKGLNKLPSANDMMADI AKKKEEMEKRY VTSQRHTIQVDY IEYMDELASLIGVKPNLLSFLTDPKLAWEVFFGCPCTPYQYRLTGPGK WDGARKAILTQRDRI IKPTKTRVLQDNNTSSSMPLLLKILGLALLSALYYFL
EldFMOF	MAKRVAVIGAGASGLTCIKCCLEDEGLEPVC FERSDDIGGLWRFKENPEDGRASIYHSVII NTSKEMMCYSDFPI PDDFPNYMHNSKIMEYFRMYAQHFDDLRYIRFQT TVRSVKKRPDFS RSGQWDVVTESSDGQQESAI FDAVMVCTGHHTH PHLPLQDFPGIETFKGQYFHSRDYKSP EEFQGKRVVVIGIGNSGGDIAVELSRVAEQVFLSTRRGSWV LNVRVSDNGFPLDMIYSTRF TNLLKKILPFSFLNWMGEKKLNKRFNHKLYGLQPKHRIFSQHPMVNDLDPNRIISGTVQV KPNVKEFRGSSVVFEDGTVEDNIDVVVFATGYTFSFPFLDSSVLAVSDNKVS LYKYVFPFH ALERPTLAVIGLIQPLGAIMPISELQARWATR VFKGLNKLPSAKTMLKDI EKKKEAMAKR YVTSQRHTIQVDY IPYMDELASQIGVRPNLLWLF TDPRLALQVLF GCPCTPYQYRLCGPG QWAGARKAILTQWDRVAKPLKTRVLQAVENPRSSSMPLLLKVSAAALLSALYHRAFLPL FLPNPPVLLARIVAFLPLL

This expanded tree also highlights some additional uncharacterized FMOs found in Amphibia and Reptilia (Figure 3.6.1.1). Three new clades of FMOs are observed: Amphibia and Reptilia both contain a group of FMOs that cluster towards FMO5, labelled as X, that are derived from EldFMO1-5. Additionally, amphibia possess another group of FMOs that emerge before EldFMO1-5, labelled as Y. All three groups of these unclassified FMOs are well represented within their classes and show high support values for their divergence: The emergences of Amphibia Y, Amphibia X and Reptilia X have TBEs of 1, 0.95 and 0.97, respectively. These FMOs likely represent new uncharacterized FMOs with the tetrapod superclass. Based on their phylogeny and similarity to FMO5, they could represent putative BVMO systems.

Inspecting the presence of the FMOs throughout the different classes within the superclass tetrapod reveals some key insights regarding their distribution. Fish only possess one FMO, meaning that at some point during evolution, as organisms took to the land, an explosion of gene duplications took place that resultantly produced the five (six) classified FMO paralogs observed today. Knowing that FMOs 1, 2 and 3/6, are derived from FMO4, using the most parsimonious method (i.e. the smallest of changes/events), we can predict that the first tetrapod species possessed all 5 FMOs. Intriguingly, not all FMO paralogs were kept in each class, implying significance for certain FMOs (Figure 3.6.2.1). Amphibia do not possess either FMO1 or FMO4. Thus, one could speculate that these paralogs are not crucial for survival. The low expression levels observed for FMO4 in humans substantiates this observation. Based on the gene duplication events illustrated, we can infer that amphibia lost FMO4 and FMO1 during evolution. All five FMO paralogs on the other hand are kept throughout the other classes, except for Aves that lose FMO2. However, many humans do not possess an active FMO2 and therefore lose out on the paralogs detoxifying capacity.

The lack of enzymatic activity present in literature for human FMO4 has been suggested to be a consequence of the C-terminal extension that reduces its expression. Itagaki et al., showed that removing this extension resulted in expression of the truncated construct with some activity detected for thioanisole conversion [Itagaki, 1994]. As such, this finding corroborates the impact the C-terminus plays towards successful expression. The lack of substrate profiling of this truncated equivalent, moreover, means that classifying the enzyme under BVMO-like or FMO-like activity is difficult. Interestingly, this C-terminal extension is solely mammal-based and is absent in all other classes. Based on this observation, we speculate that FMO4 is likely well expressed in the saurian class and could partake in an integral role for xenobiotic degradation.

3.6.3 Ancestral Sequence Reconstruction

If one assumes that the activity of FMO4 is exemplary to that of the FMOs 1, 2 and 3, it could be postulated that the emergence of BVMO or FMO activity took place at the branches connecting the ancestors EldFMO1-5 and EldFMO5, or EldFMO1-5 and

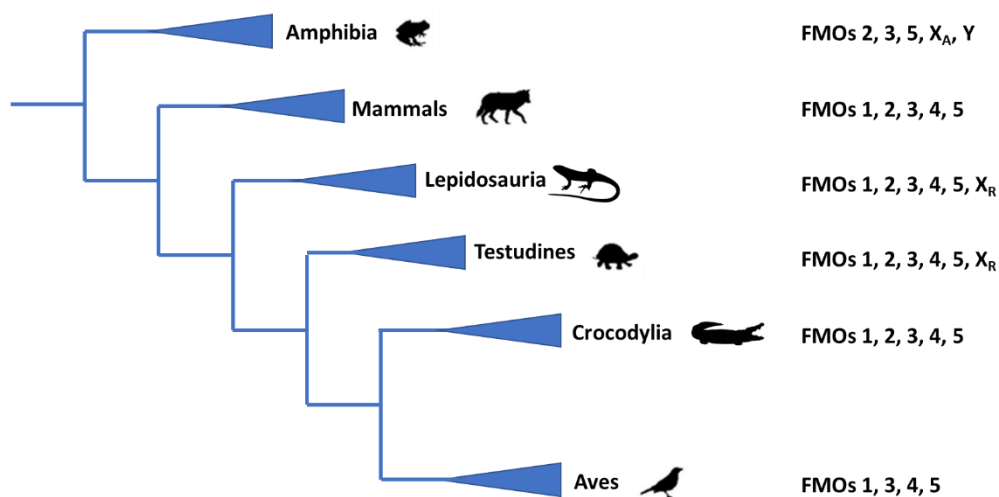


Figure 3.6.2.1: Distribution of the FMOs throughout the classes of tetrapoda. The species tree for the tetrapoda super class are shown with the emerging classes shown and labelled. On the right, the number of FMOs possessed in each class are listed. FMOs X_A, X_R and Y are described as FMOX amphibia, FMOX reptilia and FMOY shown only in amphibia (see Figure 3.6.1.1).

EldFMO1-4, respectively (Figure 3.6.1.2). Analysing the changes that took place between these nodes would therefore be crucial to determining the switch in activity. However, if the ancestor of FMO4 possesses BVMO activity, then the switch could in fact take place between EldFMO1-4 and EldFMO1-3/6 and so these changes need to be considered (Figure 3.6.1.2). As such, all ancestors at nodes EldFMO1-5, EldFMO5, EldFMO1-4 and EldFMO1-3/6 were collected for biochemical characterization.

Each ancestral sequence was extracted and cleaned so that the final sequences represented not only the most likely sequence but also the most parsimonious length. For example, regarding the C-terminal extension of the ancestor of all FMO4s, it could contain the extension. However, considering that all other FMOs don't possess this extension, the probability of the ancestor possessing it is lower. Therefore, EldFMO1-4 was reconstructed without the extension. We scrutinized the key active site residues E281 and H282 for AncFMOs 1, 2 and 3-6, and 5, respectively, to determine the activity expected for each ancestral state. We performed a structure-based sequence alignment employing the structure of AncFMO3-6 as a reference to infer which residues were retained within the active site (Figure 3.6.3.1). Interestingly, the alignment indicated that the glutamate residue at site 281 was only conserved for the ancestor EldFMO1-3/6, whilst ancestors EldFMO1-5, EldFMO1-4, EldFMO5 and EldFMOF all consist of the histidine residues (Figure 3.6.3.1). Based on this evidence, it would suggest that the FMO activity emerged during the

gene duplication event that generated FMO4 and the ancestor of FMOs 1-3/6 (Figure 3.6.1.2).

Analysing the active site alone, we mapped the changes occurring for different ancestral states (Figure 3.6.3.2). Firstly, inspecting the changes observed between EldFMO1-5 and EldFMO1-4, the only active site residues were: Ile60 to Thr60 and Met283 to Val283. Regarding the former change, the side chain does not lie in the active site cavity directly because it sits behind the isoalloxazine ring. Intriguingly however, the Isoleucine is observed in AncFMO5, EldFMOF, EldFMO1-5 and EldFMO5. Threonine on the other hand is observed in AncFMOs 1, 2, 3-6 and ancestors EldFMO1-4 and EldFMO1-3/6. With the former three possessing the threonine and the latter predicted to be an FMO, this switch insinuates that it may play a role in activity or encouraging a switch in mechanistic action.

Regarding the second change at site 283, there are three different residues observed throughout the FMO sequences: methionine, valine and threonine (Figures 3.6.3.1 and 3.6.3.2). Methionine is observed in the oldest FMOs - EldFMOF and EldFMO1-5; valine is observed in AncFMOs 2, 3-6, and EldFMO1-4 and EldFMO1-3/6; threonine is observed in AncFMO5 and node EldFMO5. With threonine being observed only in the BVMO-like systems (assuming EldFMO5 is a BVMO), it may suggest that this residue is key for the specific activity.

Alternatively, the valine residue at node EldFMO1-3/6 has a high probability, whilst at node EldFMO1-4, the probability weakens and has a similar probability for a leucine at this site (0.46 vs 0.41, respectively). This trend could suggest that through evolution the necessity for a valine strengthened at this position. For example, if EldFMO1-3/6 was a true FMO, whilst EldFMO1-4 displayed a mixture of BVMO and FMO activity, the



Figure 3.6.3.1: Structure-based sequence alignment of all AncFMOs and EldFMOs. The alignment was generated by ESript 3.0 using AncFMO3-6, PDB ID: 6SE3. Residues highlighted in red represent a site with high similarity among the submitted sequences. Residues boxed in blue outlines imply highly conserved regions. Residues shown in white with red background are completely conserved. The secondary structure of the AncFMO3-6 sequence is conveyed above the alignment. The black arrow points at site 281.

introduction of the valine in EldFMO1-4 could have promoted FMO activity which was later highly retained after subsequent gene duplications. Remarkably, similar to residue 60,

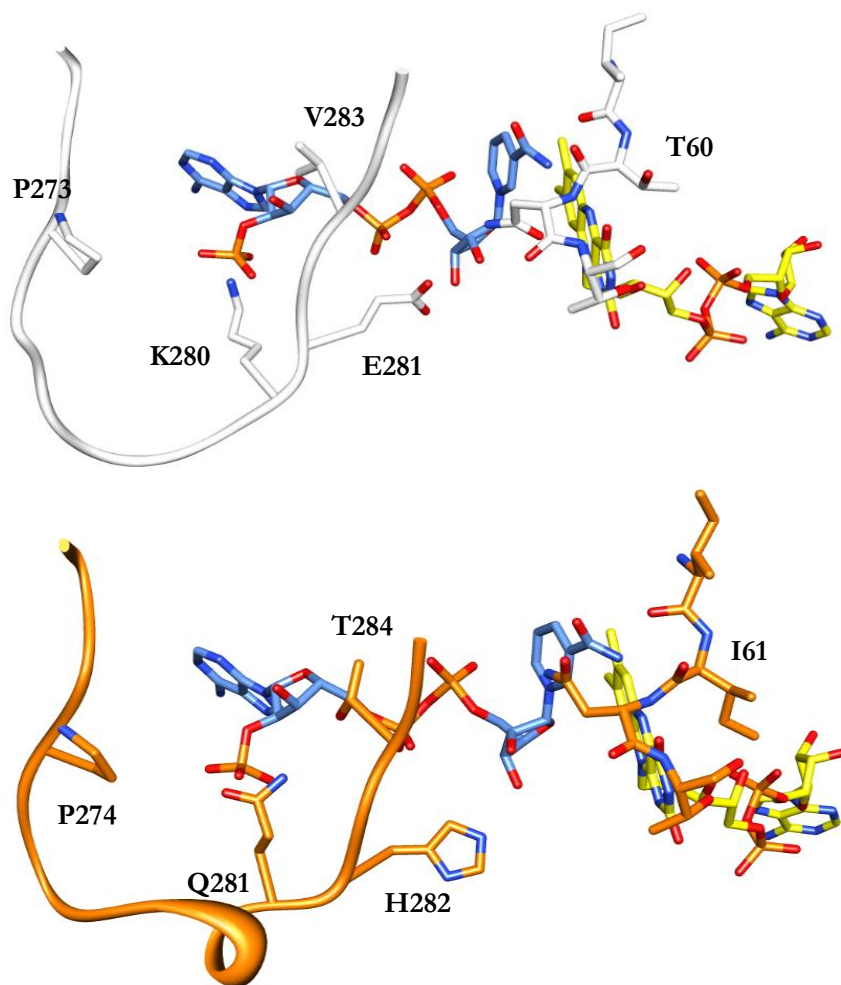


Figure 3.6.3.2: Key residues among the ancestral FMOs. Top and bottom panels depict key residues that mutate during gene duplications, illustrated by AncFMO2 and AncFMO5, respectively, posing as models. Residues that change from EldFMO1-5 through to EldFMO1-3/6 are labelled with sidechains shown, except for P273/4 that changes in AncFMO4 (see later). Residues shown in white and orange represent AncFMO2 and AncFMO5, respectively. Residues making up loop 3 are shown in ribbon form to emphasise their differing confirmations with key side chains displayed. Residues situated above the isoalloxazine ring are shown as atoms. FAD and NADP⁺ molecules are shown in yellow and cornflower blue colours, respectively.

this sidechain points outwards towards the solvent and does not seem to have a direct role within the active site cavity. It should be noted that these changes may be involved in epistasis, directly influencing the capacity for new activities to appear.

Continuing to assess the evolution, the changes within the active site were documented between EldFMO1-4 and EldFMO1-3/6. Two changes within the vicinity of the active site are positions 280 and 281 that change from a glutamine and a histidine to a lysine and a glutamate, respectively (Figures 3.6.3.1 and 3.6.3.2). Through hydrogen bonds, the glutamine kinks loop 3 (see section 3.3) that lines the active site into a small alpha helix that blocks a substrate entry path observed for AncFMO2. The lysine that substitutes this residue instead points outwards towards the NADP⁺ cofactor and untangles the loop, opening it up to substrate passage. The change in chemical property, more specifically charge, between lysine and glutamine may perturb the overall chemical environment within the active site. The lysine residue may preferentially favour more polar substrates, such as positively charged amines, that are typically metabolized by FMOs.

The change at position 281 from histidine to glutamate is a significant change situated in the active site core. The residue is observed to be within the vicinity of a theoretical C4a-hydroperoxy intermediate and within hydrogen-bonding distance to the ribose ring of the co-factor molecule (see section 3.4). Between them, the change in pK_as regarding the site of ionization, varies by approximately two units, with glutamic acid and histidine bearing pK_as of approximately 4 and 6, respectively [Pahari, 2019]. This difference may significantly alter the enthalpic contribution towards the generation of a protonated versus deprotonated C4a-based flavin intermediate, which resultantly dictates whether classic FMO N-oxidation or Baeyer-Villiger oxidation occurs. However, these pK_as are only estimations and further pH analysis would need to be pursued.

3.6.4 The role of FMO4

FMO4 is observed in all classes of the tetrapods except amphibia and has been under scrutiny for many years regarding its function in the body. Many attempts to successfully express and purify the enzyme have been thwarted and the underlying reason is speculated to be the result of the C-terminal extension observed in mammals. However, the active site residues present both in human FMO4 and the predicted mammalian ancestral FMO4 highlights some stark differences between these constructs and the other FMO paralogs (Figures 3.6.4.1 and 3.6.4.2).

Based on the structural-based sequence alignment (Figure 3.6.4.1), multiple lysines are seen in loop 3 preceding site 281 (275-281 contains five lysines). The abundance of this residue, the one-residue truncation (site 272 for AncFMO3-6) and the absence of a well conserved proline residue at site 273 (Figures 3.6.3.2 and 3.6.4.1) raises questions regarding the loop's conformation and the construction of the active site. The alignment interestingly proposes the presence of a lysine residue at site 281, instead of the well observed glutamate or

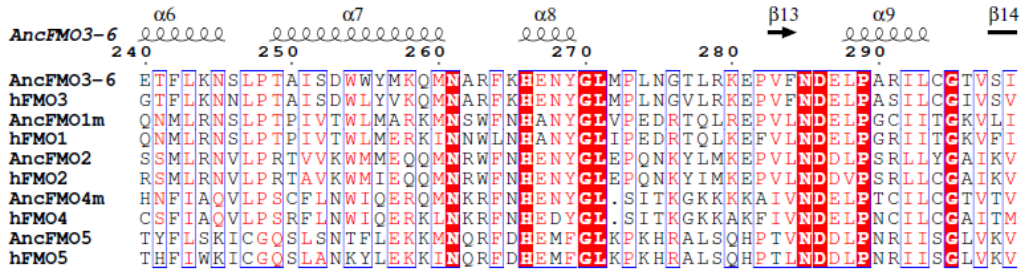


Figure 3.6.4.1: Structure-based sequence alignment of all mammalian ancestral FMO paralogs and their respective human forms. The alignment was generated by ESript 3.0 using the PDB ID: 6SE3. Residues highlighted in red represent a site with high similarity among the submitted sequences. Residues boxed in blue outlines imply highly conserved regions. Residues shown in white with red background are completely conserved. The numbering above shows the site number based on the PDB reference structure, with its secondary structures conveyed above.

histidine. This prediction suggests that the enzyme would likely act as a BVMO with a potentially positively charged hydrogen bond donor being able to stabilize the negatively charged Crigee intermediate, similarly to CHMO [Yachnin, 2014].

As mentioned above, a closer inspection on the FMO4 sequences reveals that the C-terminal extension is only observed for mammals. With this in mind, we examined the active site sequences of the ancestor of FMO4 that directly diverged from the mammalian ancestor, AncFMO4s, to determine whether this ancestor portrayed the same residue features (Figure 3.6.4.2). Curiously, the well conserved proline at site 273 was re-introduced and the abundance of lysine residues was lost. Whilst there is still a one residue truncation, sequence alignments expect a histidine residue would prevail at site 281. As such, one could propose that FMO4 would perform Baeyer-Villiger oxidations for these organisms. It should be noted however that a potential restrictive epistatic Isoleucine residue at site 60,

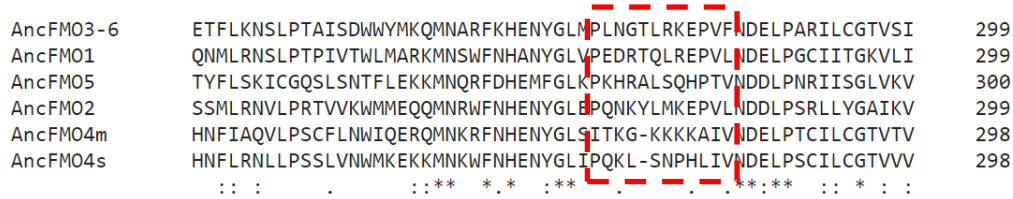


Figure 3.6.4.2: Sequence alignment of all the mammalian ancestral FMO paralogs including the sauria ancestral FMO4 focussing on loop 3. Alignment was generated using ClustalW with the increasing conservation portrayed with single dots, double dots and asterisks, respectively. The dashed red box highlights key residues leading up towards the active site from sites 273-283. The numbers on the right describe the final site for each sequence in this region. AncFMO4 mammalian and saurian ancestors are listed as AncFMO4m and AncFMO4s, respectively.

is absent in the ancestor of FMO4s and substituted with the well observed Threonine residue among the FMOs



4.DISCUSSION

Class B flavin-dependent monooxygenases have been identified as integral oxidizing agents for a range of biochemical transformations. Each protein bears a conserved paired-Rossmann-fold that enables the capture of both critical cofactors for catalysis and are thus, self-reliant. By stabilizing a C4a-(hydro)peroxy flavin intermediate, these systems have built a basis for catalysis [Joosten, 2007; Leys, 2020; Huijbers, 2014; van Berkel, 2006]. NMOs and YUCCAs have fine-tuned their substrate processing malware to engage with few substrates, whereas FMOs and BVMOs alternatively have built extensive portfolios [Huijbers, 2014; Fürst, 2018; Ziegler, 1990].

Pioneering research conducted in the late 20th century revealed key enzymatic activity executed by the mammalian membrane binding FMOs [Beaty, 1981a; Beaty, 1981b]. It became apparent that these systems could produce and maintain the C4a-hydroperoxy flavin intermediate and oxidize a range of compounds which later showed to be similar to their soluble bacterial homologues. Nonetheless, structural features rationalising their modes of action were greatly restricted to soluble FMOs crystal structures.

Harnessing ancestral sequence reconstruction has allowed our studies to reveal key insights into the evolution of FMOs. Moreover, thorough sequence analysis has provided insight towards the diverging enzymatic activities observed throughout the family. The evolutionary analysis envisages that approximately 350 Mya, the first tetrapod species (origin of mammals, amphibians and aves) roamed on land with five different FMOs. As these organisms emerged, they were likely exposed to an onslaught of xenobiotics and new compounds. The high quantities of gaseous oxygen in the atmosphere at this time (between 26 and 33%, compared to 21% now) [Kumar, 2017], would have encouraged oxidative-based catalysis. Therefore, organisms that possessed copious FMOs, consequently contained large catalytic potentials and could thrive in this competitive climate.

FMOs 3 and 5 are well preserved throughout all classes of tetrapoda. FMO5 represents the eldest FMO of the family and is the only member capable of performing BVMO oxidations. This feature makes it an exclusive and desirable enzyme. Alternatively, while FMO3 is likely conserved due to its large breadth in substrate profile, it may be involved with crucial metabolic processes [Krueger, 2005; Phillips, 2019]. FMO4 is the eldest and first diverging member to appear among FMOs 1-4 and is present in all classes except amphibia. Our sequence analysis implies that mammalian FMO4 is substantially different

from sauria (aves, reptiles and testudines) FMO4. This speculation is attributed to the mammalian specific C-terminal extension, an accumulation of lysine residues shrouding the active site and a highly conserved proline at site 273. These unique features could rationalise why limited expression so far has been detected for human FMO4 and correspondingly may not be integral for xenobiotic metabolism in mammals.

Following from FMO4, a gene duplication resulted in the development of FMO2 that is present in all classes except aves. This FMO was the first to display the key glutamate residue at position 281 and, with the absence of mechanistic evidence for FMO4, could represent the first canonical FMO. Although shown to turnover many compounds in other species, this enzyme is not active in the majority of humans due to a prematurely inserted stop codon. Finally, FMOs 1 and 3 embody the youngest of the FMO family with FMO6 emerging solely within mammals from a further gene duplication [Hines, 2002; Hernandez, 2004]. Over the last 350 M years or so, many classes have lost paralogs. Likely, certain FMOs may no longer play critical roles within certain environments.

Herein this thesis, using Ancestral Sequence Reconstruction as a foothold, we have been able to successfully characterize and elucidate the structures of four mammalian FMOs. Firstly, in line with the conventional class B nomenclature, each system contained a paired Rossmann-fold that could successfully accommodate both cofactors, FAD and NADP⁺, within the crystal structures. Focussing more closely with respect to class B FMOs, from all domains of life, each FMO displayed a paramount asparagine residue, Asn61 (AncFMO2 site numbering), hanging over the C4a centre of the isoalloxazine ring. Loss of this residue has been shown to be damaging for substrate conversion and induce trimethylaminuria when absent in FMO3. Clearly, the residue is a vital component for function. Intriguingly however, other Class B monooxygenases have been shown to employ alternative residues at this position such as aspartate residues (BVMOs) and glutamines (NMOs). Remarkably, each system has utilized a given residue to stabilize a particular intermediate for a specific chemical environment.

Mammalian FMOs have employed large C-terminal transmembrane helices to latch onto the membrane. These components fortify membrane binding and keep the scaffold tightly bound against the phospholipid bilayer. Additionally, mammalian FMOs have incorporated supplementary hydrophobic regions to enforce membrane complexation. Ridges derived from two alpha helical clusters across the protein surface stick to the phospholipid bilayer through monotopic association [Allen, 2019]. This interface is derived from the tetrapod-specific 80-residue insertion. By donating one of these alpha helical clusters, it stretches out the hydrophobic lining of the ridge across the protein surface, that in-turn, encourages membrane association and provides solvent free passages for substrates.

In conjunction with the hydrophobic ridge that presents pathways for oncoming substrates, the insertion encloses the active site cavity. The 80-residue insertion feeds a large loop that lines the active site core. Curiously, a similar loop is also observed for BVMOs. Here, the

protein presents the pivotal arginine residue believed to stabilize the negatively charged Crigee intermediate and protects it from the solvent. Thus, the loop not only shields the intermediate from the solvent, it stabilizes its formation. Furthermore, we can speculate that BVMOs use this stretch of residues as a handle for promoting BVMO activity.

Our investigation reveals that these systems harbour a high degree of redundancy throughout the family. Each member has integrated a set of highly conserved residues including Arg223, Asn194 and Asn61 (using AncFMO2 site labelling). Whilst the former two residues are implied to facilitate NADP⁺ binding, Asn61 has been indicated to be paramount for stabilizing the C4a-hydroperoxy intermediate. Differing residues among the family do not appear to distort the shape of the active site and suggests that the FMOs have not evolved its active site to be particularly well fit for a set of substrates [Krueger, 2005; Ziegler, 1990; Jakoby, 1990]. Despite this, AncFMO1 displays an active site that infers an element of optimization. The torn active site cavity and the new peptide conformation collectively complement one another and expand the entrance point for soluble compounds. It has built a cavity that can easily accommodate solvent-based molecules at the expense of uncoupling which is approximately two-fold greater for AncFMO1 (see section 3.5).

The tunnels integrated by each FMO provides new specificities towards different substrates. This builds on the notion first remarked by Henry Kamin of Duke University, regarding the cocked-gun analogy that suggests that once the substrate reaches the active site, it is immediately oxidized. This idea is supported by the redundancy observed within the active site among the AncFMOs. In other words, it is the tracts themselves and their respective member-specific tunnel-residue linings that permit and restrict substrate profiles, not the active site architecture. The extraordinary substrate extraction capacity orchestrated by this enzyme family, coupled with their extensive distribution throughout the body manifests them as formidable xenobiotic detoxifying agents.



5.CONCLUDING REMARKS

Over the last several decades, exhaustive research describing the oxidative prowess of Flavin-containing monooxygenases has been undertaken. This resulted in a swath of compounds being classified as substrates for this family. However, discriminating what features specified and governed their relative substrate profiles was greatly hindered by sequence analysis alone. Stimulating work carried out by Fiorentini et al [Fiorentini, 2016], brought further intrigue to the FMO family as they demonstrated the Baeyer-Villiger activity orchestrated by FMO5. The breadth of substrate scope coupled with the installation of an alternative mode of action coined these systems as promiscuous and enticing xenobiotic detoxifying agents.

Our findings highlight unprecedented structural topologies for membrane binding class B flavin-dependent monooxygenases. Primarily, these crystal structures allowed us to map out the components that give rise to alternative enzymatic activity and encourage substrate diversity. They are comprised of a highly conserved enzymatic core that can accommodate and catalyse innumerable molecules. Furthermore, our work documents a slight but significant switch in an active site residue that may well be the culprit that promotes an alternative functional mechanism. While the dimensions of the core remain relatively fixed with time, we also portray a well conserved tunnel building block observed among all the AncFMOs. Remarkably, each FMO then expands the tunnel's reach by splitting off in different directions. In doing so, the enzymes are able to maximise their substrate uptake.

In summary, this research has demonstrated the significant power behind Ancestral Sequence Reconstruction. Encompassing all the FMOs in the workflow built a strong foundation in which we could compare and dissect each in turn and build a blueprint regarding the enzymes' functionality. It has enabled us to delineate key changes throughout the family that provide rationale behind their enzymatic activities. Henceforth, we can start to evaluate the function behind certain residues and unravel the evolutionary emergence of the FMO activity in tetrapods. Ostensibly more important, the approach shows promise for protein crystallization. Unlike the human FMOs, the ancestral FMOs were well overexpressed in bacterial strains, displayed 'textbook' purifications and crystallized in multiple detergent-containing conditions. Applying this approach to other recalcitrant and non-crystallizing protein families may provide large benefit for a range of systems.

REFERENCES

- Alfieri, A., Malito, E., Orru, R., Fraaije, M. W. & Mattevi, A. [2008] "Revealing the moonlighting role of NADP in the structure of a flavin-containing monooxygenase," *Proc. Natl. Acad. Sci.*, Vol. 105, pp. 6572–6577.
- Allen, K. N., Entova, S., Ray, L. C. & Imperiali, B. [2019] "Monotopic Membrane Proteins Join the Fold," *Trends Biochem. Sci.*, Vol. 44, pp. 7–20.
- Ashenberg, O., Gong, L. I. & Bloom, J. D. [2013] "Mutational effects on stability are largely conserved during protein evolution," *Proc. Natl. Acad. Sci.*, Vol. 110, pp. 21071–21076.
- Ayesh, R., Mitchell, S. C., Zhang, A. & Smith, R. L. [1993] "The fish odour syndrome: biochemical, familial, and clinical aspects," *BMJ*, Vol. 307, pp. 655–657.
- Beaty, N. B. & Ballou, D. P. [1981a] "The reductive half-reaction of liver microsomal FAD-containing monooxygenase," *J. Biol. Chem.*, Vol. 256, pp. 4611–4618.
- Beaty, N. B. & Ballou, D. P. [1981b] "The Oxidative Half-reaction of Liver Microsomal FAD-containing Monooxygenase," *J. Biol. Chem.*, Vol. 256, pp. 4619–4625.
- Beneventi, E., Niero, M., Motterle, R., Fraaije, M. & Bergantino, E. [2013] "Discovery of Baeyer-Villiger monooxygenases from photosynthetic eukaryotes," *J. Mol. Catal. B Enzym.*, Vol. 98, pp. 145–154.
- Binda, C., Newton-Vinson, P., Hubálek, F., Edmondson, D. E. & Mattevi, A. [2002] "Structure of human monoamine oxidase B, a drug target for the treatment of neurological disorders," *Nat. Struct. Biol.*, Vol. 9, pp. 22–26.
- Bonanno, J. B. *et al.* [2007] Mechanism of action of a flavin-containing monooxygenase. *Proc. Natl. Acad. Sci. U. S. A.*, Vol. 104, pp. 9832–9837.
- Broadley, K. J. [2010] "The vascular effects of trace amines and amphetamines," *Pharmacol. Ther.*, Vol. 125, pp. 363–375.

- Caldinelli, L. *et al.* [2008] "Relevance of the flavin binding to the stability and folding of engineered cholesterol oxidase containing noncovalently bound FAD," *Protein Sci.*, Vol. 17, pp. 409–419.
- Cashman, J. R. & Zhang, J. [2006] "Human Flavin-Containing Monooxygenases," *Annu. Rev. Pharmacol. Toxicol.*, Vol. 46, pp. 65–100.
- Cashman, J. R. [1995] "Structural and Catalytic Properties of the Mammalian Flavin-Containing Monooxygenase," *Chem. Res. Toxicol.*, Vol. 8, pp. 165–181.
- Cashman, J. R. [2000] "Human flavin-containing monooxygenase: substrate specificity and role in drug metabolism," *Curr. Drug Metab.*, Vol. 1, pp. 181–191.
- Cashman, J. R. [2003] "The role of flavin-containing monooxygenases in drug metabolism and development," *Curr. Opin. Drug Discov. Devel.*, Vol. 6, pp. 486–493.
- Cashman, J. R. *et al.* [1997] "Human Flavin-Containing Monooxygenase Form 3: cDNA Expression of the Enzymes Containing Amino Acid Substitutions Observed in Individuals with Trimethylaminuria," *Chem. Res. Toxicol.*, Vol. 10, pp. 837–841.
- Cha, J.-Y. *et al.* [2015] "A novel thiol-reductase activity of Arabidopsis YUC6 confers drought tolerance independently of auxin biosynthesis," *Nat. Commun.* Vol. 6, pp. 8041.
- Chakraborty, S., Ortiz-Maldonado, M., Entsch B. & Ballou D. P. [2010] "Studies on the Mechanism of p-Hydroxyphenylacetate 3-Hydroxylase from *Pseudomonas aeruginosa* – a System Composed of a Small Flavin Reductase and a Large Flavin-Dependent Oxygenase," *Biochemistry*, Vol. 49, pp. 372–385.
- Cheng, Y., Dai, X. & Zhao, Y. [2006] "Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in Arabidopsis," *Genes Dev.*, Vol. 20, 1790–1799.
- Cheng, Y., Dai, X. & Zhao, Y. [2007] "Auxin synthesized by the YUCCA flavin monooxygenases is essential for embryogenesis and leaf formation in Arabidopsis," *Plant Cell*, Vol. 19, pp. 2430–2439.
- Chocklett, S. W. & Sobrado, P. [2010] "Aspergillus fumigatus SidA is a highly specific ornithine hydroxylase with bound flavin cofactor," *Biochemistry*, Vol. 49, pp. 6777–6783.
- Choi, H. S. *et al.* [2003] "A novel flavin-containing monooxygenase from *Methylophaga* sp. strain SK1 and its indigo synthesis in *Escherichia coli*," *Biochem. Biophys. Res. Commun.*, Vol. 306, pp. 930–936.

- Copley S. D. [2015] “An Evolutionary Biochemist’s Perspective on Promiscuity,” *Trends Biochem. Sci.*, Vol. 40, pp. 72–78.
- Dai, X. *et al.* [2013] “The biochemical mechanism of auxin biosynthesis by an arabidopsis YUCCA flavin-containing monooxygenase,” *J. Biol. Chem.*, Vol. 288, pp. 1448–1457.
- Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. [2011] “ProfTest 3: fast selection of best-fit models of protein evolution,” *Bioinformatics*, Vol. 27, pp. 1164–1165.
- Diesel, B. *et al.* [2007] “ α -Lipoic Acid as a Directly Binding Activator of the Insulin Receptor: Protection from Hepatocyte Apoptosis,” *Biochemistry*, Vol. 46, pp. 2146–2155.
- DM., Z. [2002] “An overview of the mechanism, substrate specificities, and structure of FMOs,” *Drug Metab. Rev.*, Vol. 34, pp. 503–511.
- Dolphin, C. T. *et al.* [1998] “The flavin-containing monooxygenase 2 gene (FMO2) of humans, but not of other primates, encodes a truncated, nonfunctional protein,” *J. Biol. Chem.*, Vol. 273, pp. 30599–30607.
- Dominy, J. E. *et al.* [2007] “Discovery and characterization of a second mammalian thiol dioxygenase, cysteamine dioxygenase,” *J. Biol. Chem.*, Vol. 282, pp. 25189–25198.
- Dong, C. *et al.* [2005] “Structural biology: Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination,” *Science*, Vol. 309, pp. 2216–2219.
- Donoghue, N. A., Norris, D. B. & Trudgill, P. W. [1976] “The Purification and Properties of Cyclohexanone Oxygenase from *Nocardia globerulea* CL1 and *Acinetobacter* NCIB 9871,” *Eur. J. Biochem.*, Vol. 63, pp. 175–192.
- Du, Y.-L., He, H.-Y., Higgins, M. A. & Ryan, K. S. [2017] “A heme-dependent enzyme forms the nitrogen–nitrogen bond in piperazate,” *Nat. Chem. Biol.*, Vol. 13, pp. 836–838.
- Eick, G. N., Bridgham, J. T., Anderson, D. P., Harms, M. J. & Thornton, J. W. [2017] “Robustness of Reconstructed Ancestral Protein Functions to Statistical Uncertainty,” *Mol. Biol. Evol.*, Vol. 34, pp. 247–261.
- Estrada, D. E. *et al.* [1996] “Stimulation of Glucose Uptake by the Natural Coenzyme α -Lipoic Acid/Thioctic Acid: Participation of Elements of the Insulin Signaling Pathway,” *Diabetes*, Vol. 45, 1798–1804.

- Fiorentini, F. *et al.* [2016] “Biocatalytic Characterization of Human FMO5: Unearthing Baeyer-Villiger Reactions in Humans,” *ACS Chem. Biol.*, Vol. 11, pp. 1039–1048.
- Fiorentini, F. *et al.* [2017] “Baeyer-Villiger Monooxygenase FMO5 as Entry Point in Drug Metabolism,” *ACS Chem. Biol.*, Vol. 12, pp. 2379–2387.
- Fornieris, F., Orru, R., Bonivento, D., Chiarelli, L. R. & Mattevi, A. [2009] “ThermoFAD, a ThermoFluor®-adapted flavin ad hoc detection system for protein folding and ligand binding,” *FEBS J.*, Vol. 276, pp. 2833–2840.
- Fraaije, M. W., Kamerbeek, N. M., Heidekamp, A. J., Fortin, R. & Janssen, D. B. [2004] “The Prodrug Activator EtaA from *Mycobacterium tuberculosis* Is a Baeyer-Villiger Monooxygenase,” *J. Biol. Chem.*, Vol. 279, pp. 3354–3360.
- Fraaije, M. W., Kamerbeek, N. M., van Berkel, W. J. H. & Janssen, D. B. [2002] “Identification of a Baeyer-Villiger monooxygenase sequence motif,” *FEBS Lett.*, Vol. 518, pp. 43–47.
- Franceschini, S. *et al.* [2012] “Structural insight into the mechanism of oxygen activation and substrate selectivity of flavin-dependent N-hydroxylating monooxygenases,” *Biochemistry*, Vol. 51, pp. 7043–7045.
- Frébortová, J. *et al.* [2004] “Catalytic reaction of cytokinin dehydrogenase: preference for quinones as electron acceptors,” *Biochem. J.*, Vol. 380, pp. 121–130.
- Frederick, R. E., Mayfield, J. A. & DuBois, J. L. [2011] “Regulated O₂ Activation in Flavin-Dependent Monooxygenases,” *J. Am. Chem. Soc.*, Vol. 133, pp. 12338–12341.
- Fürst, M. J. L. J., Romero, E., Gómez Castellanos, J. R., Fraaije, M. W. & Mattevi, A. [2018] “Side-Chain Pruning Has Limited Impact on Substrate Preference in a Promiscuous Enzyme,” *ACS Catal.*, Vol. 8, pp. 11648–11656.
- Gadda, G. & Francis, K. [2010] “Nitronate monooxygenase, a model for anionic flavin semiquinone intermediates in oxidative catalysis,” *Arch. Biochem. Biophys.*, Vol. 493, pp. 53–61.
- Gadda, G. [2012] “Oxygen Activation in Flavoprotein Oxidases: The Importance of Being Positive,” *Biochemistry*, Vol. 51, pp. 2662–2669.
- Gahl, W. A., Thoene, J. G. & Schneider, J. A. [2002] “Cystinosis,” *N. Engl. J. Med.*, Vol. 347, pp. 111–121.

- Gonzalez Malagon, S. G. *et al.* [2015] “The phenotype of a knockout mouse identifies flavin-containing monooxygenase 5 (FMO5) as a regulator of metabolic ageing,” *Biochem. Pharmacol.*, Vol. 96, pp. 267–277.
- Guindon, S. *et al.* [2010] “New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0,” *Syst. Biol.*, Vol. 59, pp. 307–321.
- Hadzipasic, A. *et al.* [2020] “Ancient origins of allosteric activation in a Ser-Thr kinase,” *Science*, Vol. 367, pp. 912–917.
- Hao, D. C., Chen, S. L., Mu, J. & Xiao, P. G. [2009] “Molecular phylogeny, long-term evolution, and functional divergence of flavin-containing monooxygenases,” *Genetica*, Vol. 137, pp. 173–187.
- Henderson, M. C., Krueger, S. K., Stevens, J. F. & Williams, D. E. [2004] “Human flavin-containing monooxygenase form 2 S-oxygenation: sulfenic acid formation from thioureas and oxidation of glutathione,” *Chem. Res. Toxicol.*, Vol. 17, pp. 633–640.
- Hernandez, D. *et al.* [2003] “Trimethylaminuria and a human FMO3 mutation database,” *Hum. Mutat.*, Vol. 22, pp. 209–213.
- Hernandez, D., Janmohamed, A., Chandan, P., Phillips, I. R. & Shephard, E. A. [2004] “Organization and evolution of the flavin-containing monooxygenase genes of human and mouse,” *Pharmacogenetics*, Vol. 14, pp. 117–130.
- Heuts, D. P. H. M., Scrutton, N. S., McIntire, W. S. & Fraaije, M. W. [2009] “What’s in a covalent bond?: On the role and formation of covalently bound flavin cofactors,” *FEBS J.*, Vol. 276, pp. 3405–3427.
- Hines, R. N. [2002] “Alternative Processing of the Human FMO6 Gene Renders Transcripts Incapable of Encoding a Functional Flavin-Containing Monooxygenase,” *Mol. Pharmacol.*, Vol. 62, pp. 320–325.
- Hjelmeland, L. M. [1990] “Solubilization of native membrane proteins,” *Methods Enzymol.*, Vol. 182, pp. 253–264.
- Hochberg, G. K. A. & Thornton, J. W. [2017] “Reconstructing Ancient Proteins to Understand the Causes of Structure and Function,” *Annu. Rev. Biophys.*, Vol. 46, pp. 247–269.

- Huang, Z., Wang, K.-K. A. & van der Donk, W. A. [2016] "New insights into the biosynthesis of fosfazinomycin," *Chem. Sci.*, Vol. 7, pp. 5219–5223.
- Huijbers, M. M. E., Montersino, S., Westphal, A. H., Tischler, D. & Van Berkel, W. J. H. [2014] "Flavin dependent monooxygenases," *Arch. Biochem. Biophys.*, Vol. 544, pp. 2–17.
- Huxtable, R. J. [1992] "Physiological actions of taurine," *Physiol. Rev.*, Vol. 72, pp. 101–163.
- Itagaki, K., Carver, G. T. & Philpot, R. M. [1996] "Expression and characterization of a modified flavin-containing monooxygenase 4 from humans," *J. Biol. Chem.*, Vol. 271, pp. 20102–20107.
- J. Henriques, B., K. Olsen, R., Bross, P. & M. Gomes, C. [2010] "Emerging Roles for Riboflavin in Functional Rescue of Mitochondrial & Oxidation Flavoenzymes," *Curr. Med. Chem.*, Vol. 17, pp. 3842–3854.
- Jakoby, W. B. & Ziegler, D. M. [1990] "The enzymes of detoxication," *J. Biol. Chem.*, Vol. 265, pp. 20715–20718.
- Janmohamed, A., Hernandez, D., Phillips, I. R. & Shephard, E. A. [2004] "Cell-, tissue-, sex- and developmental stage-specific expression of mouse flavin-containing monooxygenases (Fmos)," *Biochem. Pharmacol.*, Vol. 68, pp. 73–83.
- Joosten, V. & van Berkel, W. J. [2007] "Flavoenzymes," *Curr. Opin. Chem. Biol.*, Vol. 11, pp. 195–202.
- Kantz, A., Chin, F., Nallamothu, N., Nguyen, T. & Gassner, G. T. [2005] "Mechanism of flavin transfer and oxygen activation by the two-component flavoenzyme styrene monooxygenase," *Arch. Biochem. Biophys.*, Vol. 442, pp. 102–116.
- Kar, R. K., Borin, V. A., Ding, Y., Matysik, J. & Schapiro, I. [2019] "Spectroscopic Properties of Lumiflavin: A Quantum Chemical Study," *Photochem. Photobiol.*, Vol. 95, pp. 662–674.
- khan, M. zahid & nawaz, W. [2016] "The emerging roles of human trace amines and human trace amine-associated receptors (hTAARs) in central nervous system," *Biomed. Pharmacother.*, Vol. 83, pp. 439–449.
- Koukouritaki, S. B., Simpson, P., Yeung, C. K., Rettie, A. E. & Hines, R. N. [2002] "Human Hepatic Flavin-Containing Monooxygenases 1 (FMO1) and 3 (FMO3) Developmental Expression," *Pediatr. Res.*, Vol. 51, pp. 236–243.

- Krueger, S. K. & Williams, D. E. [2005] “Mammalian flavin-containing monooxygenases: Structure/function, genetic polymorphisms and role in drug metabolism,” *Pharmacol. Ther.*, Vol. 106, pp. 357–387.
- Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. [2017] “TimeTree: A Resource for Timelines, Timetrees, and Divergence Times,” *Mol. Biol. Evol.*, Vol. 34, pp. 1812–1819.
- Lemoine, F. *et al.* [2018] “Renewing Felsenstein’s phylogenetic bootstrap in the era of big data,” *Nature*, Vol. 556, pp. 452–456.
- Leoni, C., Buratti, F. M. & Testai, E. [2008] “The participation of human hepatic P450 isoforms, flavin-containing monooxygenases and aldehyde oxidase in the biotransformation of the insecticide fenthion,” *Toxicol. Appl. Pharmacol.*, Vol. 233, pp. 343–352.
- Leys, D. & Scrutton, N. S. [2020] “Flavin doesn’t put all oxygens in one basket,” *Nat. Chem. Biol.*, Vol. 16, pp. 485–486.
- Lin, J. & Cashman, J. R. [1997] “Detoxication of tyramine by the flavin-containing monooxygenase: stereoselective formation of the trans oxime,” *Chem. Res. Toxicol.*, Vol. 10, pp. 842–852.
- Lin, J. & Cashman, J. R. [1997] “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.*, Vol. 282, pp. 1269–1279.
- Lombardini, J. B. [1983] “Effects of ATP and taurine on calcium uptake by membrane preparations of the rat retina,” *J. Neurochem.*, Vol. 40, pp. 402–406.
- Magnani, F. *et al.* [2017] “Crystal structures and atomic model of NADPH oxidase,” *Proc. Natl. Acad. Sci. U. S. A.*, Vol. 114, pp. 6764–6769.
- Malito, E., Alfieri, A., Fraaije, M. W. & Mattevi, A. [2004] “Crystal structure of a Baeyer – Villiger monooxygenase,” *Proc. Natl. Acad. Sci. U. S. A.*, Vol. 101, pp. 13157–13162.
- Martinoli, C. *et al.* [2013] “Beyond the protein matrix: Probing cofactor variants in a Baeyer-Villiger oxygenation reaction,” *ACS Catal.*, Vol. 3, pp. 3058–3062.
- Mascotti, M. L., Juri Ayub, M., Furnham, N., Thornton, J. M. & Laskowski, R. A. [2016] “Chopping and Changing: the Evolution of the Flavin-dependent Monooxygenases,” *J. Mol. Biol.*, Vol. 428, pp. 3131–3146.

- Mascotti, M. L., Lapadula, W. & Ayub, M. “The origin and evolution of Baeyer - Villiger Monooxygenases (BVMOs): An ancestral family of flavin monooxygenases,” *PLoS ONE*, Vol. 10, pp. 1-16.
- Massey, V. [1995] “Introduction: Flavoprotein structure and mechanism,” *FASEB J.*, Vol. 9, pp. 473–475.
- Masters, B. S. S. & Ziegler, D. M. [1971] “The Distinct Nature and Function of NADPH-Cytochrome c Reductase and the NADPH-Dependent Mixed-Function Amine Oxidase of Porcine Liver Microsomes,” *Arch. Biochem.*, Vol. 145, pp. 358-364.
- Mato, J., Alvarez, L., Ortiz, P. & Pajares, M. A. [1997] “S-adenosylmethionine synthesis: Molecular mechanisms and clinical implications,” *Pharmacol. Ther.*, Vol. 73, pp. 265–280.
- Mattevi, A. [2006] “To be or not to be an oxidase: challenging the oxygen reactivity of flavoenzymes,” *Trends Biochem. Sci.*, Vol. 31, pp. 276–283.
- Matthews, A. *et al.* [2020] “Aminoperoxide adducts expand the catalytic repertoire of flavin monooxygenases,” *Nat. Chem. Biol.*, Vol. 16, pp. 556–563.
- Mayfield, J. A. *et al.* [2010] “Comprehensive spectroscopic, steady state, and transient kinetic studies of a representative siderophore-associated flavin monooxygenase,” *J. Biol. Chem.*, Vol. 285, pp. 30375—30388.
- Mayhew, S. G. [1999] “The effects of pH and semiquinone formation on the oxidation-reduction potentials of flavin mononucleotide. A reappraisal,” *Eur. J. Biochem.*, Vol. 265, pp. 698–702.
- Meneely, K. M., Barr, E. W., Bollinger, J. M. & Lamb, A. L. [2009] “Kinetic Mechanism of Ornithine Hydroxylase (PvdA) from *Pseudomonas aeruginosa*: Substrate Triggering of O₂ Addition but Not Flavin Reduction,” *Biochemistry*, Vol. 48, pp. 4371–4376.
- Merkl, R. & Sterner, R. [2016] “Ancestral protein reconstruction: Techniques and applications,” *Biol. Chem.*, Vol. 397, pp. 1–21.
- Merkl, R. & Sterner, R. [2016] “Reconstruction of ancestral enzymes,” *Perspect. Sci.*, Vol. 9, pp. 17–23.
- Mijatovic, S. & Gadda, G. [2008] “Oxidation of alkyl nitronates catalyzed by 2-nitropropane dioxygenase from *Hansenula mrakii*,” *Arch. Biochem. Biophys.*, Vol. 473, pp. 61—68.

- Morrison, E., Kantz, A., Gassner, G. T. & Sazinsky, M. H. [2013] "Structure and mechanism of styrene monooxygenase reductase: New insight into the FAD-transfer reaction," *Biochemistry*, Vol. 52, pp. 6063–6075.
- Mügge, C. *et al.* [2020] "Flavin-dependent N-hydroxylating enzymes: distribution and application," *Appl. Microbiol. Biotechnol.*, Vol. 104, pp. 6481–6499.
- Nagata, T., Williams, D. E. & Ziegler, D. M. [1990] "Substrate Specificities of Rabbit Lung and Porcine Liver Flavin-Containing Monooxygenases: Differences due to Substrate Size," *Chem. Res. Toxicol.*, Vol. 3, pp. 372–376.
- Naumann, C., Hartmann, T. & Ober, D. [2002] "Evolutionary recruitment of a flavin-dependent monooxygenase for the detoxification of host plant-acquired pyrrolizidine alkaloids in the alkaloid-defended arctiid moth *Tyria jacobaeae*," *Proc. Natl. Acad. Sci. U. S. A.*, Vol. 99, pp. 6085–6090.
- Neumann, C. S. *et al.* [2012] "Biosynthesis of piperazic acid via N5-hydroxy-ornithine in *Kutzneria* spp. 744," *ChemBiochem*, Vol. 13, pp. 972–976.
- Ohmi, N. *et al.* [2003] "S-oxidation of S-methyl-esonarimod by flavin-containing monooxygenases in human liver microsomes," *Xenobiotica*, Vol. 33, pp. 1221–1231.
- Olucha, J. & Lamb, A. L. [2011a] "Mechanistic and structural studies of the N-hydroxylating flavoprotein monooxygenases," *Bioorg. Chem.*, Vol. 39, pp. 171–177.
- Olucha, J., Meneely, K. M., Chilton, A. S. & Lamb, A. L. [2011b] "Two structures of an N-hydroxylating flavoprotein monooxygenase: Ornithine hydroxylase from *Pseudomonas aeruginosa*," *J. Biol. Chem.*, Vol. 286, pp. 31789–31798.
- Overby, L. H. *et al.* [1995] "Characterization of flavin-containing monooxygenase 5 (FMO5) cloned from human and guinea pig: evidence that the unique catalytic properties of FMO5 are not confined to the rabbit ortholog," *Arch. Biochem. Biophys.*, Vol. 317, pp. 275–284.
- Pahari, S., Sun, L. & Alexov, E. [2019] "PKAD: a database of experimentally measured pKa values of ionizable groups in proteins," *Database*, Vol. 2019.
- Palfey, B. A. & McDonald, C. A. [2010] "Control of catalysis in flavin-dependent monooxygenases," *Arch. Biochem. Biophys.*, Vol. 493, pp. 26–36.

- Pauling, L. & Zuckerkandl E. [1963] "Chemical Paleogenetics Molecular "Restoration Studies" of Extinct Forms of Life," *Acta. Chem. Scand.*, Vol. 17, pp. 9-16.
- Peters, J. W., Miller, A. F., Jones, A. K., King, P. W. & Adams, M. W. W. [2016] "Electron bifurcation," *Curr. Opin. Chem. Biol.*, Vol. 31, pp. 146–152.
- Petriello, M. C., Hoffman, J. B., Morris, A. J. & Hennig, B. [2017] "Emerging roles of xenobiotic detoxification enzymes in metabolic diseases," *Rev. Environ. Health*, Vol. 32, pp. 105–110.
- Phillips, I. R. & Shephard, E. A. [2019] "Endogenous roles of mammalian flavin-containing monooxygenases," *Catalysts*, Vol. 9, pp. 1001.
- Piano, V., Palfey, B. A. & Mattevi, A. [2017] "Flavins as Covalent Catalysts: New Mechanisms Emerge," *Trends Biochem. Sci.*, Vol. 42, pp. 457–469.
- Poulsen, L. & Ziegler, D. M. [1979] "The Liver Microsomal FAD-containing Monooxygenase," *J. Biol. Chem.*, Vol. 254, pp. 6449–6455.
- Rettie, A. E., Lawton, M. P., Sadeque, A. J., Meier, G. P. & Philpot, R. M. [1994] "Prochiral sulfoxidation as a probe for multiple forms of the microsomal flavin-containing monooxygenase: studies with rabbit FMO1, FMO2, FMO3, and FMO5 expressed in *Escherichia coli*," *Arch. Biochem. Biophys.*, Vol. 311, pp. 369–377.
- Ripp, S. L., Itagaki, K., Philpot, R. M. & Elfarra, A. A. [1999] "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.*, Vol. 367, pp. 322–332.
- Risso, V. A. *et al.* [2015] "Mutational studies on resurrected ancestral proteins reveal conservation of site-specific amino acid preferences throughout evolutionary history," *Mol. Biol. Evol.*, Vol. 32, pp. 440–455.
- Risso, V. A., Gavira, J. A., Mejia-Carmona, D. F., Gaucher, E. A. & Sanchez-Ruiz, J. M. [2013] "Hyperstability and substrate promiscuity in laboratory resurrections of precambrian β -lactamases," *J. Am. Chem. Soc.*, Vol. 135, pp. 2899–2902.
- Risso, V. A., Sanchez-Ruiz, J. M. & Ozkan, S. B. [2018] "Biotechnological and protein-engineering implications of ancestral protein resurrection," *Curr. Opin. Struct. Biol.*, Vol. 51, pp. 106–115.

- Robinson, R., Badiyan, S. & Sobrado, P. [2013] "C4a-hydroperoxyflavin formation in N - hydroxylating flavin monooxygenases is mediated by the 2'-OH of the nicotinamide ribose of NADP+," *Biochemistry*, Vol. 52, pp. 9089–9091.
- Romero, E., Castellanos, J. R. G., Mattevi, A. & Fraaije, M. W. [2016] "Characterization and Crystal Structure of a Robust Cyclohexanone Monooxygenase," *Angew. Chemie - Int. Ed.*, Vol. 55, pp. 15852–15855.
- Romero, E., Gómez Castellanos, J. R., Gadda, G., Fraaije, M. W. & Mattevi, A. [2018] "Same Substrate, Many Reactions: Oxygen Activation in Flavoenzymes," *Chem. Rev.*, Vol. 118, pp. 1742–1769.
- Ronquist, F. *et al.* [2012] "MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space," *Syst. Biol.*, Vol. 61, pp. 539–542.
- Schwinn, K., Ferré, N. & Huix-Rotllant, M. [2020] "UV-visible absorption spectrum of FAD and its reduced forms embedded in a cryptochrome protein," *Phys. Chem. Chem. Phys.*, Vol. 22, pp. 12447–12455.
- Scott, F. *et al.* [2017] "Identification of flavin-containing monooxygenase 5 (FMO5) as a regulator of glucose homeostasis and a potential sensor of gut bacteria," *Drug Metab. Dispos.*, Vol. 45, pp. 982–989.
- Sehlmeyer, S. *et al.* [2010] "Flavin-dependent monooxygenases as a detoxification mechanism in insects: New insights from the arctiids (Lepidoptera)," *PLoS One*, Vol. 5, pp. e10435.
- Sheehan, D., Meade, G., Foley, V. M. & Dowd, C. A. [2001] "Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily," *Biochem. J.*, Vol. 360, pp. 1–16.
- Siddiq, M. A., Hochberg, G. K. & Thornton, J. W. [2017] "Evolution of protein specificity: insights from ancestral protein reconstruction," *Curr. Opin. Struct. Biol.*, Vol. 47, pp. 113–122.
- Smitherman, C. & Gadda, G. [2013] "Evidence for a Transient Peroxynitro Acid in the Reaction Catalyzed by Nitronate Monooxygenase with Propionate 3-Nitronate," *Biochemistry*, Vol. 52, pp. 2694–2704.
- Sofer, S. S. & Ziegler, M. [1978] "Microsomal mixed-function amine oxidase. Oxidation products of piperazine-substituted phenothiazine drugs," *Drug Metab. Dispos.*, Vol. 6, pp. 232–239.

- Stamatakis, A. [2014] “RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies,” *Bioinformatics*, Vol. 30, pp. 1312–1313.
- Starr, T. N. & Thornton, J. W. [2016] “Epistasis in protein evolution,” *Protein Sci.*, Vol. 25, pp. 1204–1218.
- Starr, T. N., Picton, L. K. & Thornton, J. W. [2017] “Alternative evolutionary histories in the sequence space of an ancient protein,” *Nature*, Vol. 549, pp. 409–413.
- Stein, N. [2008] “CHAINS AW: a program for mutating pdb files used as templates in molecular replacement,” *J. Appl. Crystallogr.*, Vol. 41, pp. 641–643.
- Stuehr, D. J., Tejero, J. & Haque, M. M. [2009] “Structural and mechanistic aspects of flavoproteins: Electron transfer through the nitric oxide synthase flavoprotein domain,” *FEBS J.*, Vol. 276, pp. 3959–3974.
- Sucharitakul, J., Chaiyen, P., Entsch, B. & Ballou, D. P. [2006] “Kinetic mechanisms of the oxygenase from a two-component enzyme, p-hydroxyphenylacetate 3-hydroxylase from *Acinetobacter baumannii*,” *J. Biol. Chem.*, Vol. 281, pp. 17044–17053.
- Suh, J. K., Poulsen, L. L., Ziegler, D. M. & Robertus, J. D. [1999] “Yeast flavin-containing monooxygenase generates oxidizing equivalents that control protein folding in the endoplasmic reticulum,” *Proc. Natl. Acad. Sci. U. S. A.*, Vol. 96, pp. 2687–2691.
- Teufel, R. *et al.* [2013] “Flavin-mediated dual oxidation controls an enzymatic Favorskii-type rearrangement,” *Nature*, Vol. 503, pp. 552–556.
- Thodberg, S. & Neilson, E. H. J. [2020] “The “Green” FMOs: Diversity, Functionality and Application of Plant Flavoproteins,” *Catalysts*, Vol. 10, pp. 329.
- Tinikul, R. *et al.* [2013] “The Transfer of Reduced Flavin Mononucleotide from LuxG Oxidoreductase to Luciferase Occurs via Free Diffusion,” *Biochemistry*, Vol. 52, pp. 6834–6843.
- Tivendale, N. D., Ross, J. J. & Cohen, J. D. [2014] “The shifting paradigms of auxin biosynthesis,” *Trends Plant Sci.*, Vol. 19, pp. 44–51.
- Tolmie, C., Smit, M. S. & Opperman, D. J. [2019] “Native roles of Baeyer-Villiger monooxygenases in the microbial metabolism of natural compounds,” *Nat. Prod. Rep.*, Vol. 36, pp. 326–353.
- Torres Pazmiño, D. E., Dudek, H. M. & Fraaije, M. W. [2010] “Baeyer-Villiger monooxygenases: recent advances and future challenges,” *Curr. Opin. Chem. Biol.*, Vol. 14, pp. 138–144.

- Tynes, R. E. & Hodgson, E. [1985] “Catalytic activity and substrate specificity of the flavin-containing monooxygenase in microsomal systems: characterization of the hepatic, pulmonary and renal enzymes of the mouse, rabbit, and rat,” *Arch. Biochem. Biophys.*, Vol. 240, pp. 77–93.
- Ufnal, M., Zadlo, A. & Ostaszewski, R. [2015] “TMAO: A small molecule of great expectations,” *Nutrition*, Vol. 31, pp. 1317–1323.
- Ussher, J. R., Lopaschuk, G. D. & Arduini, A. [2013] “Gut microbiota metabolism of l-carnitine and cardiovascular risk,” *Atherosclerosis*, Vol. 231, pp. 456–461.
- van Berkel, W. J. H., Kamerbeek, N. M. & Fraaije, M. W. [2006] “Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts,” *J. Biotechnol.*, Vol. 124, pp. 670–689.
- Veeravalli, S. *et al.* [2014] “The phenotype of a flavin-containing monooxygenase knockout mouse implicates the drug-metabolizing enzyme FMO1 as a novel regulator of energy balance,” *Biochem. Pharmacol.*, Vol. 90, pp. 88–95.
- Veeravalli, S. *et al.* [2018] “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.*, Vol. 46, pp. 20–25.
- Veeravalli, S. *et al.* [2020] “Flavin-containing monooxygenase 1 catalyzes the production of taurine from hypotaurine,” *Drug Metab. Dispos.*, Vol. 48, pp. 378–385.
- Vonrhein, C. *et al.* [2018] “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.*, Vol. 74, pp. a360–a360.
- Voordeckers, K. *et al.* [2012] “Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication,” *PLoS Biol.*, Vol. 10, pp. e1001446.
- Waldman, A. J., Pechersky, Y., Wang, P., Wang, J. X. & Balskus, E. P. [2015] “The Cremeomycin Biosynthetic Gene Cluster Encodes a Pathway for Diazo Formation,” *Chembiochem*, Vol. 16, pp. 2172–2175.
- Wang, R. & Thorpe, C. [1991] “Reactivity of Medium-Chain Acyl-CoA Dehydrogenase toward Molecular Oxygen,” *Biochemistry*, Vol. 30, pp. 7895–7901.

- Weber, R. E. *et al.* [2016] “Predictable convergence in hemoglobin function has unpredictable molecular underpinnings,” *Science*, Vol. 354, pp. 336–339.
- Wellner, D. [1967] “Flavoproteins,” *Annu. Rev. Biochem.*, Vol. 36, pp. 669–690.
- Wheeler, L. C., Lim, S. A., Marqusee, S. & Harms, M. J. [2016] “The thermostability and specificity of ancient proteins,” *Curr. Opin. Struct. Biol.*, Vol. 38, pp. 37–43.
- Whetstone, J. R. *et al.* [2000] “Ethnic differences in human flavin-containing monooxygenase 2 (FMO2) polymorphisms: Detection of expressed protein in African-Americans,” *Toxicol. Appl. Pharmacol.*, Vol. 168, pp. 216–224.
- Williams, D. E., Shigenaga, M. K. & Castagnoli, N. J. [1990] “The role of cytochromes P-450 and flavin-containing monooxygenase in the metabolism of (S)-nicotine by rabbit lung,” *Drug Metab. Dispos.*, Vol. 18, pp. 418–428.
- Williams, D. E., Ziegler, D. M., Nordin, D. J., Hale, S. E. & Masters, B. S. [1984] “Rabbit lung flavin-containing monooxygenase is immunochemically and catalytically distinct from the liver enzyme,” *Biochem. Biophys. Res. Commun.*, Vol. 125, pp. 116–122.
- Williams, P. D., Pollock, D. D., Blackburne, B. P. & Goldstein, R. A. [2006] “Assessing the accuracy of ancestral protein reconstruction methods,” *PLoS Comput. Biol.*, Vol. 2, pp. 0598–0605.
- Yachnin, B. J. *et al.* [2014] “Lactone-Bound Structures of Cyclohexanone Monooxygenase Provide Insight into the Stereochemistry of Catalysis,” *ACS Chem. Biol.*, Vol. 9, pp. 2843–2851.
- Yang, Z. [2007] “PAML 4: Phylogenetic Analysis by Maximum Likelihood,” *Mol. Biol. Evol.*, Vol. 24, pp. 1586–1591.
- Yang, Z., Kumar, S. & Nei, M. [1995] “A New Method of Inference of Ancestral Nucleotide and Amino Acid Sequences,” *Genetics*, Vol. 141, pp. 1641–1650.
- Yeh, E., Blasiak, L. C., Koglin, A., Drennan, C. L. & Walsh, C. T. [2007] “Chlorination by a Long-Lived Intermediate in the Mechanism of Flavin-Dependent Halogenases,” *Biochemistry*, Vol. 46, pp. 1284–1292.
- Zhang, J., Tran, Q., Lattard, V. & Cashman, J. R. [2003] “Deleterious mutations in the flavin-containing monooxygenase 3 (FMO3) gene causing trimethylaminuria,” *Pharmacogenetics*, Vol. 13, pp. 495–500.

Zhang, Q. *et al.* [2017] "Characterization of the flavoenzyme XiaK as an N-hydroxylase and implications in indolosesquiterpene diversification," *Chem. Sci.*, Vol. 8, pp. 5067–5077.

Ziegler, D. M. & Mitchell, C. H. [1972] "Microsomal oxidase. IV. Properties of a mixed-function amine oxidase isolated from pig liver microsomes," *Arch. Biochem. Biophys.*, Vol. 150, pp. 116–125.

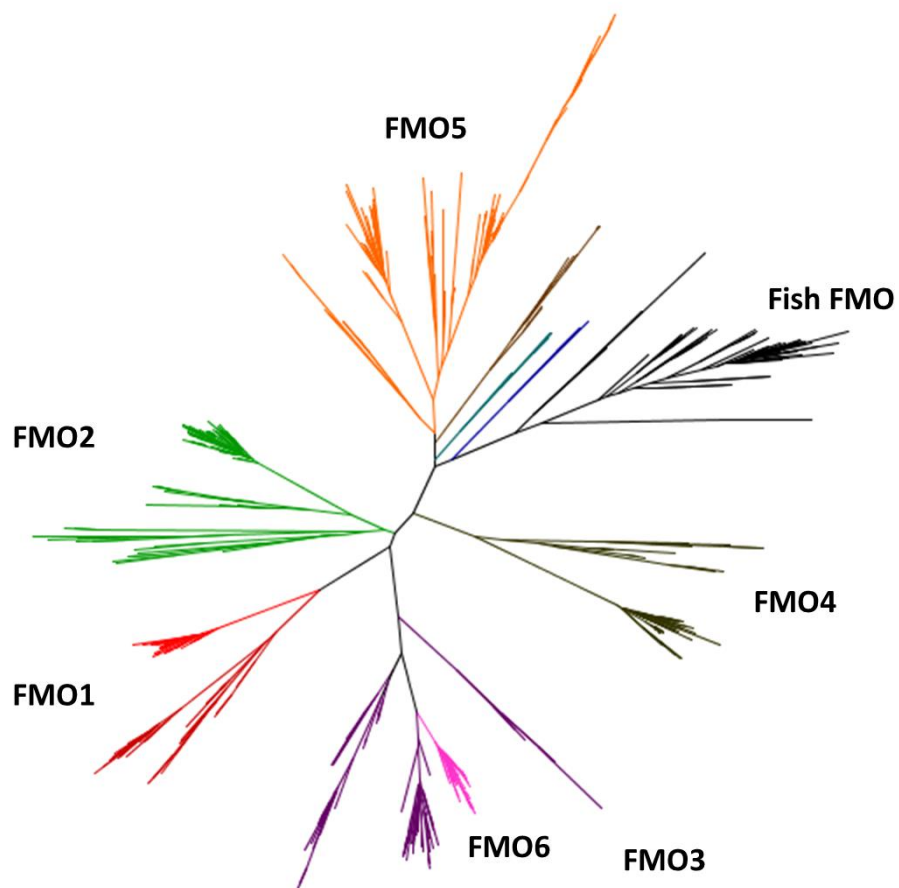
Ziegler, D. M. & Poulsen, L. L. [1978] "Hepatic microsomal mixed-function amine oxidase," *Methods Enzymol.*, Vol. 52, pp. 142–151.

Ziegler, D. M. [1990] "Flavin-containing monooxygenases: enzymes adapted for multisubstrate specificity," *Trends Pharmacol. Sci.*, Vol. 11, pp. 321–324.



APPENDIX A. PHYLOGENETIC INFERENCE OF THE FLAVIN-CONTAINING MONOOXYGENASES IN JAWED-VERTEBRATES

Phylogenetic tree – Radial Tree layout. FMO1, 2, 3, 4, 5 and 6 are coloured in red, light green, purple, dark green, orange and pink, respectively. The black branches represent the Fish FMO sequences that act as the root. Traces shown in dark blue, turquoise and brown represent the Amphibia FMOY, Amphibia FMOX and Reptilia FMOX, respectively.



Phylogenetic tree – Polar Tree layout. In this depiction the separate classes within the FMOs are visible. FMO1, 2, 3, 4, 5 and 6 are coloured in red, light green, purple, dark green, orange and pink, respectively. The black branches represent the Fish FMO sequences that act as the root. Traces shown in dark blue, turquoise and brown represent the Amphibia FMOY, Amphibia FMOX and Reptilia FMOX, respectively.

