STRUCTURE-FUNCTION AND EXPRESSION STUDIES OF MOUSE CYP4F14 AND CYP4F39 AND ANALYSIS OF ODOR LEAKAGE FROM HUMAN REMAINS CONTAINMENT UNITS FOR USE IN SPACE

_____

A Thesis

Presented to

The Faculty of the Department of Chemistry

Sam Houston State University

_____

In Partial Fulfillment

of the Requirements for the Degree of

Master of Chemistry

_____

by

Madushika Madri Jayakody

August, 2020

STRUCTURE-FUNCTION AND EXPRESSION STUDIES OF MOUSE CYP4F14 AND CYP4F39 AND ANALYSIS OF ODOR LEAKAGE FROM HUMAN REMAINS CONTAINMENT UNITS FOR USE IN SPACE

by

Madushika Madri Jayakody

_____

APPROVED:

Donovan C Haines, PhD
Thesis Director

David E Thompson, PhD
Committee Member

Richard E Norman, PhD
Committee Member

John B Pascarella, PhD
Dean, College of Science and Engineering Technology

DEDICATION

Dedicated to my parents, Kanthi Kaushalya Perera and Cicil Ananda Parakrama Jayakody, who brought me to this world, believe in me always and give me strength.

Dr. Donovan Haines, my dear mentor, supervisor, none of this would have been possible without you. Your dedication towards science and work motivates me every day to be a better scientist and a person. I hope to be like you one day!

My dear husband Yushadha J. Bogoda, the support and encouragement you gave me in the past two years is unmatchable.

ABSTRACT

Student, Jayakody M. M., Structure-function and expression studies of mouse Cyp4f14 and Cyp4f39 and analysis of odor leakage from human remains containment units for use in space. Master of Science (Chemistry), August 2020, Sam Houston State University, Huntsville, Texas.

The availability of both mouse and human genome sequences along with the fact that mouse is a good model organism to study mammals, especially humans, has prompted comparisons between the two species at multiple levels. Cytochrome P450 4F is a relatively new CYP subfamily that has gained a great deal of attention due to its involvement in inflammation and lung cancer. Mouse Cyp4f14 and Cyp4f39 show significant sequence similarity to human CYP4F2, CYP4F3 and CYP4F22 and are involved in limiting neuroinflammation and autosomal recessive congenital ichthyosis, respectively.

Nevertheless, there are not many studies done on the structure and substrate binding by these enzymes. The first part of this thesis focuses on the mouse cytochrome P450 4F subfamily enzymes Cyp4f14 and Cyp4f39, their structure, substrate binding and expression by using a combination of experimental and computational chemistry approaches. Here we report initial expression studies of Cyp4f14 that were carried out by subcloning mouse *Cyp4f14* (from Origene #MR208397) into pEx-N-His-GST vector and optimizing expression in *E. coli* as a function of OD_{600} at different isopropyl β-D-1-thiogalactopyranoside (IPTG) levels. Based on *p*-nitrocarboxylic acid (*p*NCA) oxidation assays and spectroscopy carried out with crude enzyme, induction at an OD_{600} of 0.6 with 0.5-0.8 mM of IPTG seems to be optimal for expression in the presence of cytochrome P450 reductase. Analysis of sequence alignments and preliminary work with homology models yielded insights into how mouse Cyp4f14 interacts with its substrate. Molecular

docking studies revealed the possible binding sites and modes for substrates such as leukotriene B_4, 8-*p*NCA, 11-*p*NCA and arachidonic acid. It is hoped that these results will provide insights into enzyme-substrate interactions and form a basis for future research. This study also reports a detailed structural analysis of mouse Cyp4f39 which is the ortholog of human CYP4F22. Our goal is to provide a strong basis to help future studies of these important enzymes.

The second part of this thesis involves testing human remains containment units (HRCUs), or 'body bags'. Volatile organic compounds emitted by decomposing cadavers were analyzed by polydimethylsiloxane/divinylbenzene solid phase microextraction (PDMS/DVB SPME) fibers and Gas Chromatography/Mass Spectrometry (GC/MS). These HRCUs will be used in the International Space Station (ISS) and potentially in colonies on the moon or Mars to store cadavers in the situation of a death (including return of the cadaver to Earth). The ability of bags containing non-refrigerated cadavers to contain volatiles was tested over time, with particular focus on volatiles causing human odor sensation. Different types of bags were found to have very different breakthrough times for odorous volatiles.

KEY WORDS: Cytochrome P450; CYP4F; BL21; Neuroinflammation; Arachidonic Acid; Omega-hydroxylases; Molecular docking; SPME.

ACKNOWLEDGEMENTS

Dr. Donovan Haines, brilliant scientist, engaging teacher, caring mentor and generous human being, I am indebted to you. Thank you for all your encouragement, time and inspiration. As a teacher, you have given me more than anyone else. I am grateful to have had you as my research supervisor.

I am thankful to Dr. David E. Thompson for whom I was a Teacher's Assistant. Teaching Analytical chemistry lab was a wonderful experience that gave me a taste of something that I would like to do more of in the future. I am also thankful to Dr. Dustin E. Gross, Dr. Christopher Zall, who provided the initial guidance that enabled me, coming from a Molecular Biology background to get back into Chemistry. I am thankful to Dr. Richard Norman, Dr. Darren Williams and Dr. Ilona Petrikovics who also furthered my chemistry education. My gratitude goes to my colleague Jerome Butler, whose intense dedication to studies and incredible capacity for learning inspired me and consistently kept me on my toes. Sam Houston State University was an excellent institution to complete this research in. I am thankful to Dr. Rukma Basnayake and Dr. Ananda Manage for giving me your support and looking into my well-being while in Huntsville. My special gratitude goes to Dr. Joan Bytheway from the Southeast Texas Applied Forensic Science at SHSU, Dr. Rachell Smith from the Texas Research Institute for Environmental Studies at SHSU and the NASA team who gave me a once in a lifetime opportunity.

I would like to thank the people of Walker county (Huntsville, Tx) who made me feel safe and welcome in their community where I completed most of my studies. I also thank the great people of Texas and the great people of the United States for creating the

opportunity for someone from a developing country to contribute to scientific research and learn with your finest academics. There is no better example of how you spread the light of progress to the world, share in the bounty that you have been blessed with and bring together the world's different people in friendship.

PREFACE

This thesis involves two projects, as during preparation of the prospectus an opportunity arose to carry out a collaboration with NASA researchers and the Southeast Texas Forensic Science Facility in addition to the previously planned project. As the two projects are quite different in nature, they are presented here as separate chapters.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

CHAPTER I

Expression, structural analysis and substrate interactions of mouse Cyp4f14 and Cyp4f39

Introduction

Cytochrome P450 4F3 (*CYP4F3*) and cytochrome P450 4F22 (*CYP4F22*) have been identified to be associated with neuroinflammation, lung cancers and Autosomal Recessive Congenital Ichthyosis (ARCI). These enzymes belong to a CYP gene family that comprises several members.[1–3] These enzymes participate in the metabolism of various endobiotics and xenobiotics. The enzymes are involved in Phase I drug metabolism. They are also likely to metabolize hormones known as hydroxyeicosatetraenoic acids (HETEs) and 5-oxo-epoxyeicosatetraenoic acids (5-oxo-ETEs) by adding a hydroxyl group to each hormone's C-20 terminal carbon.[4–6]

During brain inflammation, inflammatory mediators such as chemokine cytokines are released by microglia. Microglia produce eicosanoids such as the immunomodulatory hormone leukotriene B_4 (LTB4) and prostaglandins which have important roles in inflammation of the Central Nervous System (CNS) and pathogenesis of neurodegenerative diseases like Alzheimer's, Parkinson's and cancer.[7] Studies have shown that the CYP4F family enzymes Cyp4f13, Cyp4f14, Cyp4f15, Cyp4f18 are capable of hydroxylating LTB4 and limiting neuroinflammation.[1] Cyp4f14 is capable of controlling neuroinflammation by hydroxylating LTB4 into 20-hydroxy LTB4. This process is important as LTB4 is capable of entering the neurons and triggering inflammation leading to chronic inflammation. By inactivating LTB4, the reaction cascade that leads to neuroinflammation can be controlled (Figure 1).

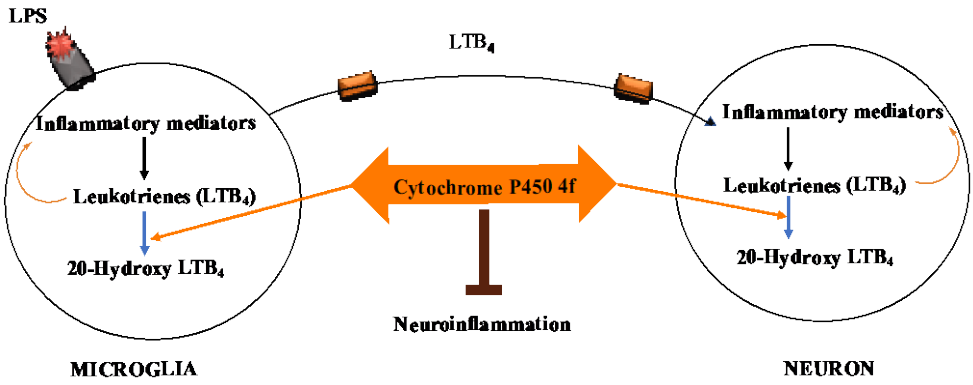


Figure 1. Role of Leukotriene B_4 in Neuroinflammation. Ligands such as lipopolysaccharides (LPS) bind to pattern recognizing receptors on microglia which leads to the activation of an inflammatory cascade, releasing LTB4. LTB4 enters neurons amplifying the response. Brain cytochrome CYP4f family enzymes limit continuous neuroinflammation by converting LTB4 in to 20-hydroxy LTB4.

Because of the similarity between CYP genes in humans and mice, many biomedical studies are carried out in mouse models. Having 82% sequence identity to human genes *CYP4F2* and *CYP4F3*, which are responsible for the production of LTB4 hydroxylases in humans, mouse gene *Cyp4f14* is an ideal candidate to study.

Human CYP4F2 and CYP4F3 and mouse Cyp4f14 are known to hydroxylate LTB4. Rat CYP4F1, CYP4F4, CTP4F5, and CYP4F6 and Mouse Cyp4f18 are also known to hydroxylate LTB4.[8] Our gene of interest, Cyp4f14, is known to hydroxylate lipoxins and hydroxyeicosatetraeonic acids (Figure 2).[8]

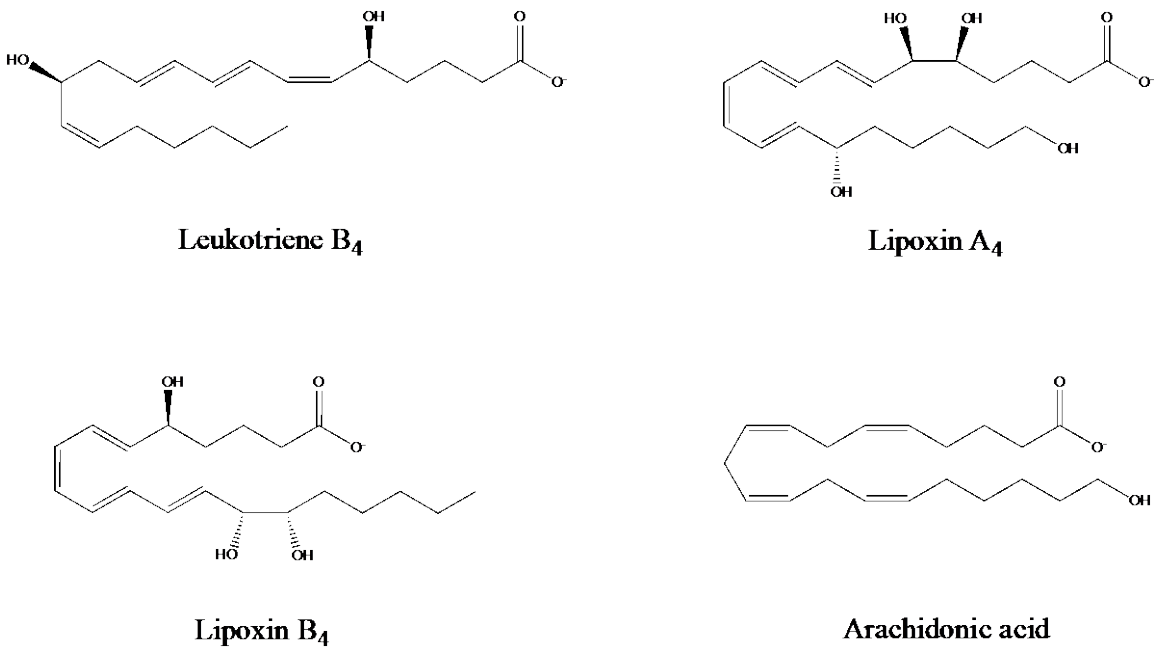


Figure 2. Chemical Structures of Eicosanoids Metabolized by Cyp4f14.[8]

A related P450 enzyme of interest, Cyp4f39, is the human ortholog of CYP4F22. This gene encodes an epidermal ω-hydroxylase which is important for the synthesis of epidermal ω-hydroxyceramides. Epidermal ω-hydroxyceramides are essential for skin-barrier function and are involved in the pathway of acylceramide production.[3] Reduced enzymatic activity of CYP4F22 caused by mutations is correlated to autosomal recessive congenital ichthyosis (ARCI).

ARCI is a rare group of keratinization disorders that affects around 1 in 100,000 people, characterized by abnormal scaling of the skin.[5] Patients with ARCI are born in a tight translucent sheath of skin called a collodion membrane and their skin becomes scaly in appearance as they mature.[9] Affected individuals often suffer from heat loss and

excessive loss of fluid (dehydration). They also develop breathing problems and are prone to infections due to broken and blistered skin.[10]

The cause of the disease involves mutations in genes required for formation of the skin permeability barrier. These mutations however, vary by the disease subtype.[11] In the structure of the skin, multilayer lipid membranes are crossinked with protein bricks.[12] Mutations in the proteins catalyzing formation of the crosslinks, called transglutaminases, affect the structural integrity of the skin.[13] This is the cause for a majority of cases. Later, it was found that several mutations in the gene for CYP4F22 lead to lamellar ichthyosis.

Long chain O-acylceramides are required for skin permeability barrier formation(*Figure 3*).[12] CYPF22 ω -hydroxylates long chain ceramide (~30 carbon long) containing acyl groups.[13] During the formation of lamellar bodies, ω -hydroxylated ceramide is conjugated to linoleic acid, which is then incorporated into the skin permeability barrier. Transglutaminase displaces linoleic acid with protein amino acid side chains from keratin to combine the strength of the protein component with the water barrier of the lipid component.

To date, there are limited studies on both Cyp4f14 and Cyp4f39, mouse orthologs of human CYP4F2 and CYP4F22. The structures of these enzymes have not been determined experimentally. Thus, we have attempted to create an expression system for both these enzymes to help facilitate further studies. Both Cyp4f14 and Cyp4f39 show 80% and 87% sequence identity to their human counterparts CYP4F2 and CYP4F22, respectively (APPENDIX A). Thus, these two mouse enzymes are good candidates for preliminary studies of these enzymes.

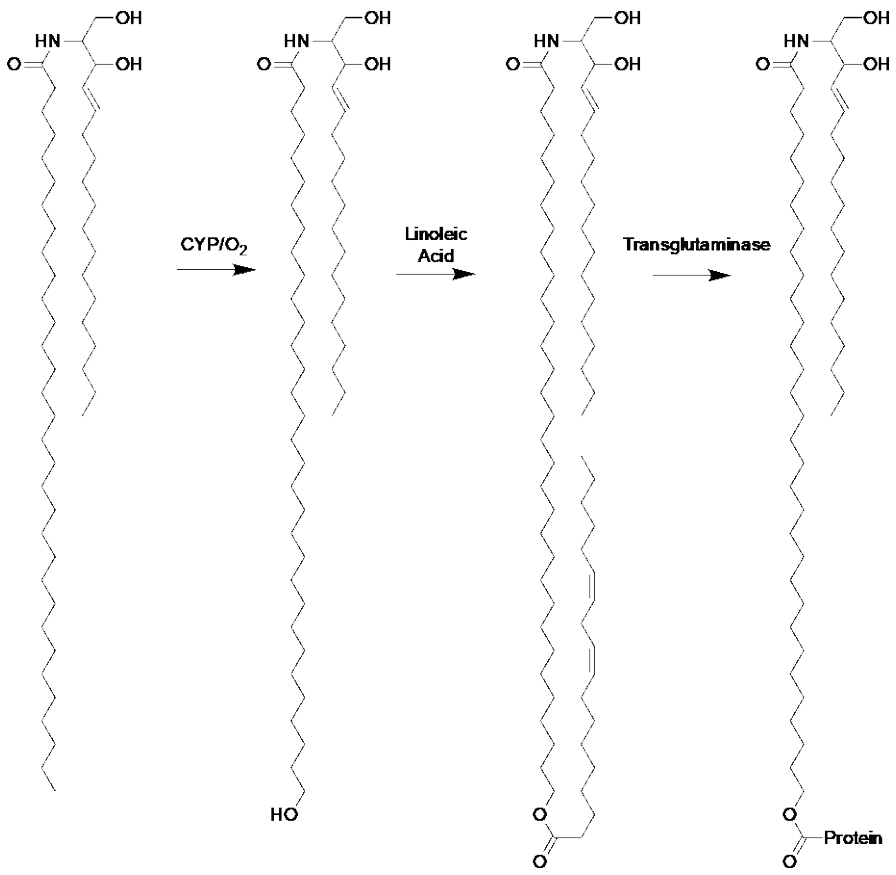


Figure 3. Role of CYP4F22 in Production of Skin Permeability Barrier.

Expression of these genes in *E. coli* can offer a simple and inexpensive way to study these enzymes and their byproducts. The study of byproducts is highly important in drug development and characterization. Drug development via classical chemical synthesis can be expensive and can present significant challenges.[14] It is challenging to express membrane bound enzymes like CYP4Fs, as purification is required. Sometimes membrane preparation is also needed to obtain the active enzyme.[14] Membrane preparation includes the disruption of the cell membrane using mechanical methods and centrifugation. These factors limit the yield of active enzyme in traditional purification methods.

In silico studies can be used to study substrate binding and further enhance the activity of the enzyme. If developed, an *E. coli* expression system could be extremely useful. An expression system can provide a readily available source of the enzyme. The reasons discussed above highlight the importance of studying CYP family enzymes and expressing them in *E. coli*.

Cloning/Plasmid vectors/ Expression vectors

Plasmids are small circular DNA molecules found in bacteria and some other organisms. Plasmids can replicate independently of the host chromosome. In bacteria, plasmids always carry at least one gene, giving useful features such as tolerance of harsh environmental conditions, chemicals, etc. Plasmids are often used as vehicles or vectors to transfer genes from one organism to another.

When a standard vector is used to transform foreign DNA into an *E. coli* cell, it is very unlikely the cell will be able to produce sufficient amounts of recombinant protein. Standard vectors lack small nucleotide sequences that signal the cell to begin transcription and translation of foreign DNA.

Also, genes of higher organisms are surrounded by expression signals that are not found in *E. coli* systems. Thus *E. coli* RNA polymerase cannot attach to a foreign gene (foreign promoter) simply because a bacterial cell does not have the technology to recognize the foreign promoter.

As a solution to this problem, expression vectors have been developed. The expression vectors have the capability to place the foreign DNA under control of *E. coli*

expression signals, allowing the transcription and translation of the gene in the bacteria system.

The anatomy of an expression vector includes three main components: promoter, terminator and ribosome binding site.[15] The promoter is where the RNA polymerase (σ subunit) binds and initiates transcription. An important property that comes from the origin of replication is the copy number. Copy number indicates the number of copies of the plasmid per cell.[16] For example, vectors with a pMB1 origin can produce 15-60 copies per cell and vectors with a pUC origin (mutated version of pMB1) can produce 500-700 copies per cell.[15]

As the name indicates, transcription stops at the transcriptional terminator sequence. The ribosome binding site (RBS) is a short nucleotide sequence recognized in RNA by the ribosome. The RBS is located a few nucleotides upstream of the initiation codon.

Apart from the origin of replication, RBS and terminator, there are promoters, affinity tags, cloning sites and selection markers present in typical expression vectors to enhance the production of protein (*Figure 4).*

Figure 4. Structure of an Expression Vector. The figure depicts the major structural features present in common expression vectors. AT= Affinity Tag, MCS= Multiple Cloning Site and T= Terminator

Subcloning

Gene subcloning requires that the DNA molecule be cut in a very precise and reproducible manner. The foreign DNA is then ligated into the parent strand and finally, the recombinant plasmid inserted into an *E. coli* cell (transformation). To cut the DNA, type II restriction endonucleases are used. All restriction endonucleases are able to recognize DNA sequences at very specific sites and cleave the phosphodiester bonds that connect two adjacent deoxyribonucleotides within a DNA molecule.[17]

In the reaction mechanism, a special histidine residue in the active site of the restriction endonuclease first attacks the phosphorus of the scissile bond giving rise to a covalent intermediate and release of the 3′-DNA (Figure 5). The water present in the reaction medium then attacks the newly formed phosphorylated-histidine intermediate, liberating the 5′-phosphorylated DNA.

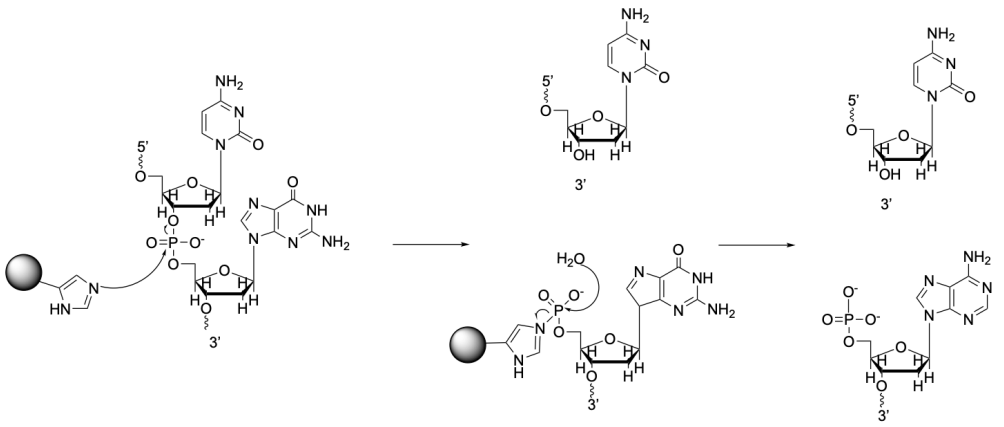


Figure 5. Reaction Mechanism of Restriction Endonuclease Bfil.

The next step is DNA ligation which occurs in three essential nucleotide transfer steps (Figure 6.). In the first step of DNA ligation, Lysine 151 of the DNA ligase attacks the α-phosphorus group of ATP (or NAD). This nucleophilic attack results in release of PPi (or NMN in the case of NAD) and formation of a P-N bond between AMP and the lysine residue of the ligase. In the next step, DNA adenylate is formed when the 5' phosphate oxygen of nicked DNA attacks the phosphorus of the ligase-adenylate intermediate. During this nucleophilic attack, the ligase lysine residue is a leaving group. In the finale step, DNA ligase catalyzes the attack of the 3' OH of nicked DNA on DNA-adenylate which results in joining of the two strands. AMP leaves as a biproduct.[18]

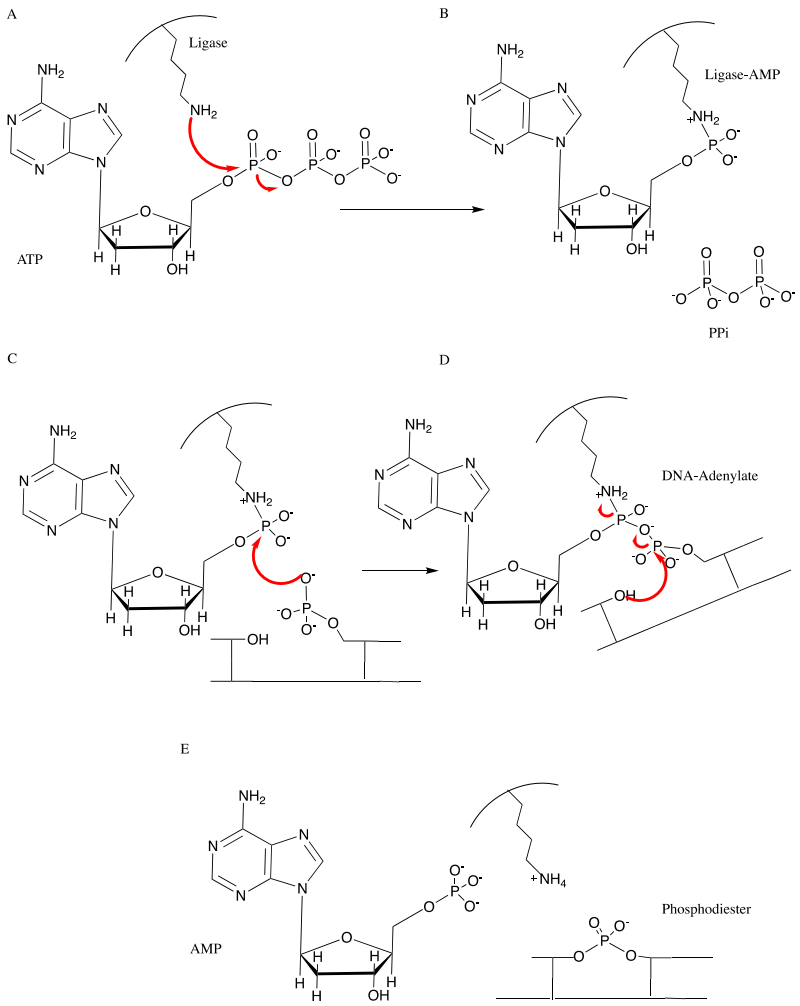


Figure 6. Reaction Mechanism of DNA Ligase.[18]

The final step of the subcloning process is transformation of ligated DNA into *E. coli* cells. Transformation is one of three methods by which bacteria exchange genetic information. Conjugation and transduction are the other two methods. During transformation naked DNA passes through the cell membrane. One of the most widely used transformation methods is the heat shock method. Competent bacterial cells, cells that are able to readily take up DNA, are prepared by treatment with $CaCl_2$ followed by a brief exposure to increased temperature.[19]

BL21(DE3) Cells

BL21(DE3) cells are specialized *E. coli* cells often used for membrane protein overexpression. These strains contain the λDE3 lysogen which carries the gene encoding for T7 RNA polymerase under control of the lacUV5 promoter.[18] This is a strong variant of a lac promoter. T7 RNA polymerase, a high efficiency viral enzyme, exclusively recognizes the lacUV5 promoter and transcribes any gene placed after this promoter. This system produces messenger RNA 8 times faster than traditional *E. coli* expression systems.[19,20] This results in high messenger RNA yield and high subsequent protein production.

UV-visible spectroscopy

UV-Visible spectroscopy is one of the most important techniques for identifying and detecting the presence of intact cytochrome P450. All cytochrome P450s contain a porphyrin IX as the prosthetic group and are categorized as b-type cytochromes.

Many biological systems involve enzymes that have porphyrins located in their active site. These are highly conjugated molecules composed of 4 pyrrole rings and 4 methine bridges. The electronic absorption spectra of the porphyrin consists of a strong Soret band (B-band) at around 400 nm and weaker transitions (2-3) in the 500-600 nm range (Q-band).[21,22] The absorption spectra of porphyrin including Q-band absorption is best described using Gouterman's model (Figure 7).[23] This is a combination of Hückel and Configuration Interaction theory. According to this model, the lowest unoccupied orbital, e_g (LUMO), is described as accidently degenerate. The highest occupied

molecular orbitals a_{1u} (HOMO) and a_{2u} ($HOMO_{-1}$) are described as nearly degenerate.

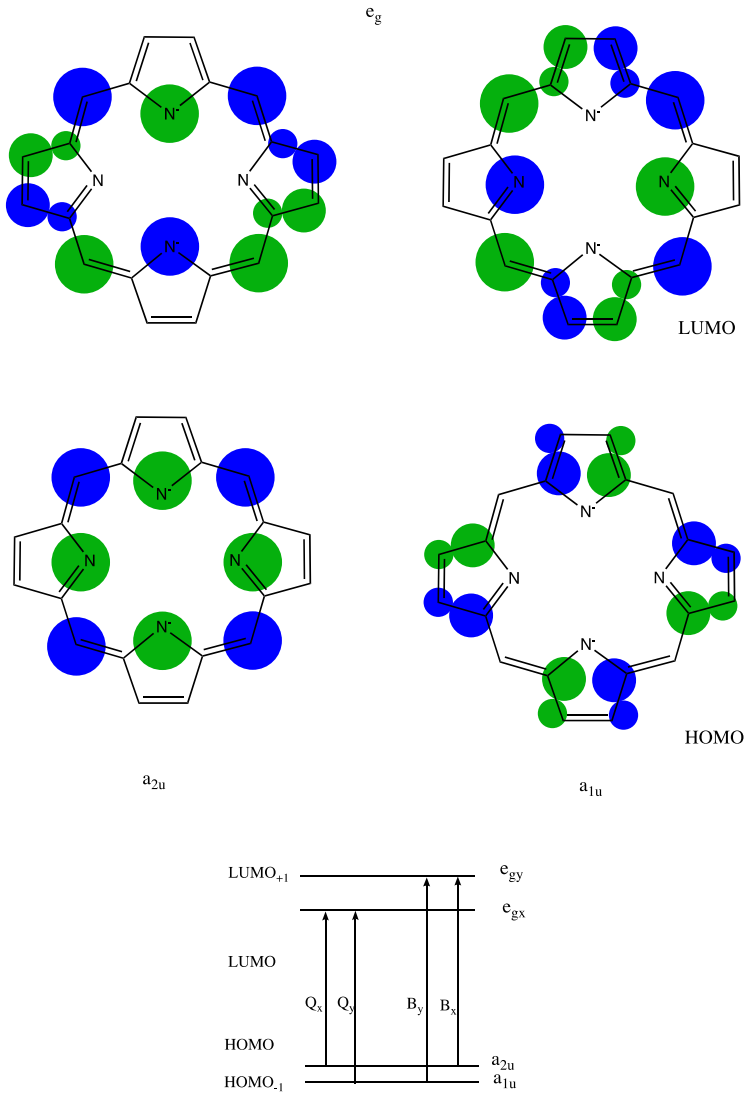


Figure 7. Porphyrins HOMO and LUMO as Described in Gouterman Theory. Four Gouterman orbitals in porphyrins and a simplified representation of electronic transitions from HOMO to LUMO are shown.[21]

Depending on the electronic environment created by the kind and positions of substituents on the porphyrin, the relative intensities of the Q-bands change.

The Soret absorption peak of ligand-free cytochrome P450 is observed at 420 nm.[19,20] However, upon being reduced by an electron donor and binding to carbon monoxide, the absorption peak shifts from 420 nm to 450 nm. Reduction of Fe changes the spin state of the heme center from a high spin to low spin.[24] This change results in a shift in the Soret peak and is characteristic for all P450s with a cysteinyl *trans*-axial ligand.[25]

UV-Vis spectroscopy is capable of detecting changes that occur in the distal substrate binding pocket of cytochrome P450, ligand affinities and mode(s) of ligand binding, distinguishing between different mutant proteins, detecting protein-protein interactions with reductase partners, assessing P450 concentrations using the extinction coefficient for the ferrous CO complex, and diagnosing mechanisms of enzyme inactivation.

In this study, UV-Vis Spectroscopy is used to assess the growth of a bacterial culture by taking optical density (OD) values at 600 nm and to study expression of the recombinant P450.

Optical density at 600 nm

Optical density is a measure of the amount of light scattered by the sample not the absorption.[26] The optical density values for a sample varies from instrument to instrument depending on their optical pathlength and depends on the sizes of the cells in the sample.

Generally, when detecting bacterial growth, OD_{600} (optical density at 600 nm) is preferred as it falls in the visible region and does not cause any harm to the bacteria or its DNA. A wavelength below 400 nm can be used, however, exposure to UV light causes

the bacterial DNA to mutate. Also, most of the bacteriological media are orange in color. Thus, they will absorb a portion of the incident light and interfere with the readings. Most bacteriological media show an absorption peak around 300-400 nm and no absorption at 600 nm, making it an ideal wavelength for optical density measurements.

p-Nitrophenoxycarboxylate assay

A *p*NCA assay is a rapid and continuous spectrophotometric assay system which uses the chromogenic product *p*-nitrophenolate to detect $P450_{BM-3}$ activity in solutions.[27,28] $P450_{BM-3}$, a highly studied bacterial fatty acid hydroxylating enzyme, is found in *Bacillus megaterium*.

In the reaction scheme shown below (Figure 8), in the presence of active P450 along with the P450's reducing partner plus oxygen and NADPH, $P450_{BM-3}$ will catalyze the hydroxylation of a long-chain *p*-nitrophenoxycarboxylate (n= 6-8, 10-12)[27,29], which is a fatty acid analog. The resulting unstable hemiacetal will then dissociate into products, an ω -oxycarboxylate and the chromophore *p*-nitrophenolate.[27]

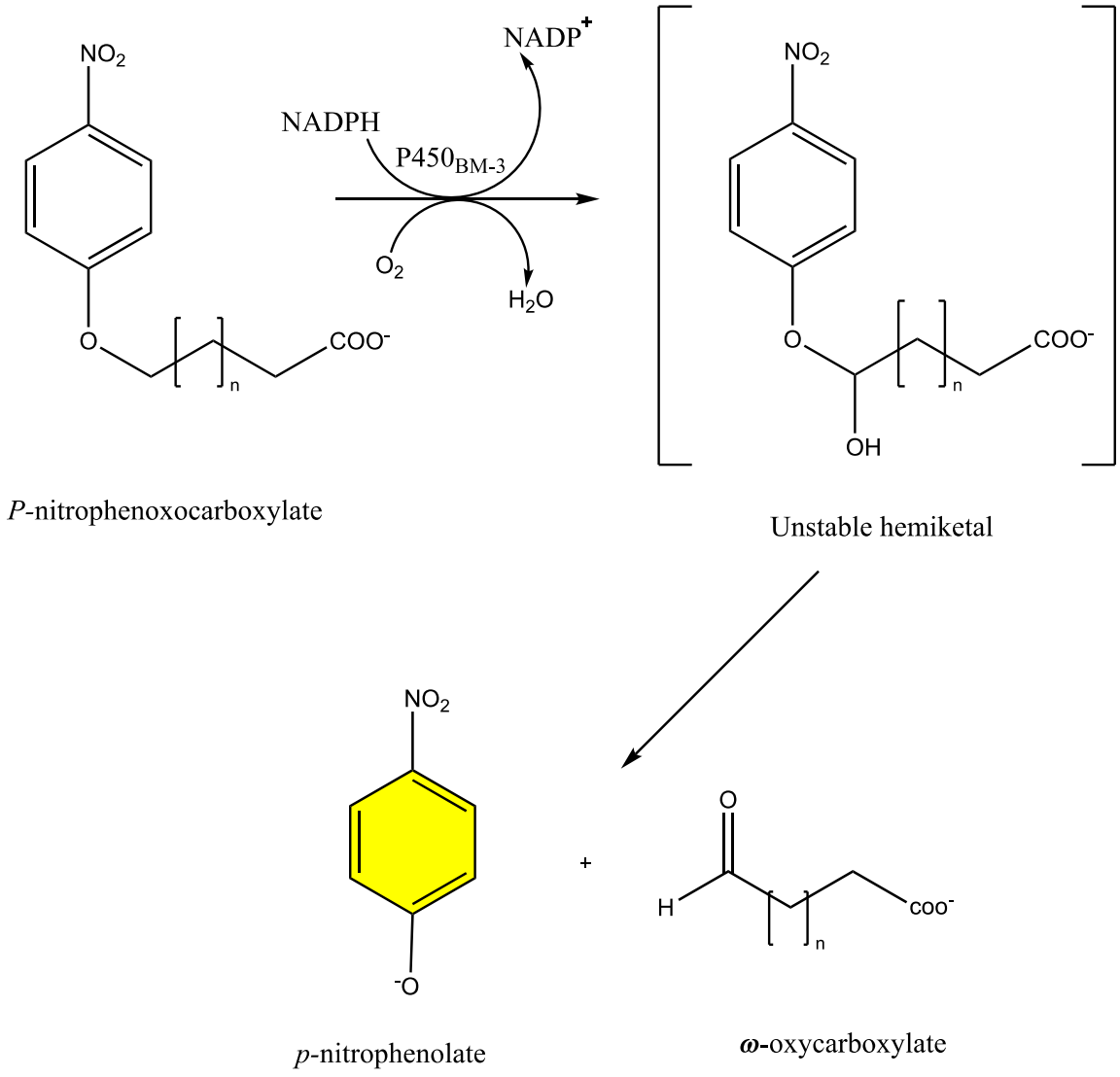


Figure 8. pNCA Oxidation Assay.[27] n=6-12, 15

Although *p*NCA is stable in alkaline solutions, and is more intensely colored in the deprotonated form, $P450_{BM-3}$ is not stable in basic solutions. It is reported that at pH 10 and 30 ºC, $P450_{BM-3}$ loses its activity by 80%.[27]At pH> 9 *p*-nitrophenolate shows maximum absorption due to the high amount of deprotonated form present. The absorption maximum of *p*-nitrophenolate lies at 400 nm. However, because both heme

and NADPH show absorption maxima around 350-450 nm and 260-340 nm,[30] respectively, 410 nm is used for activity measurements at pH 8.1-8.2. The selected pH range retains both enzyme activity and assay sensitivity. $P450_{BM-3}$ is known to hydroxylate *p*NCA chain lengths of 6 to 8 and 10-12.[25]

Catalytic cycle of cytochrome P450

Cytochrome P450 (CYP450) is a super family of Phase I drug metabolizing enzymes that are mainly found in liver endoplasmic reticulum.[31] Some cytochrome P450s are also found in the brain. It's been over 60 years since the structure of P450 was first established by using X-ray crystallography and now there are 29,899 crystal structures in the Protein Databank (www.pdb.org).

The cytochrome P450 oxidation cycle (Figure 9.) initiates when the substrate (R-H) enters the active site and interacts with the Fe^{+3} center. Cytochrome P450 in the resting state (substrate free state) has a water molecule attached to the Fe^{+3} center (I).[29,32]

In the resting state, P450 (I) has a 6-coordinate low-spin ferric atom in the active center, with 4 equatorial N ligands of the porphyrin and an axial water ligand *trans* to the proximal thiolato ligand (Cysteine). Cysteine is a highly conserved amino acid that occupies the proximal *trans* position. Entry of the substrate detaches this water molecule from the heme center, forming high spin Fe^{+3} (II).[29] The entry of the substrate (R-H) causes the axial water molecule to leave. This is likely to happen due to the conformational changes induced in the active site.[32]

Displacement of water increases the ferric spin state from S=1/2 (low spin) to S=5/2 (high spin). The resulting Fe^{+3} complex has a more positive redox potential (II),

which makes complex II more prone to receive electrons from the cytochrome P450 reductase (CPR).[20,29] The reduced Fe^{+2} complex (III) has a high affinity for O_2, and forms ferric superoxide complex (IV). Complex IV is again a good electron acceptor, thus is reduced again by CPR making a ferric peroxy complex (V). In the next step, protonation of complex (V) results in complex (VI), also known as compound 0 (Cpd 0). A second protonation occurs as Cpd 0 is a good Lewis base and is capable of accepting an additional proton. This leads to release of water via heterolytic cleavage of the O-O bond and formation of Compound I (Cpd I, complex VII). Compound I is thought to be the ultimate oxidant of P450s.[33]

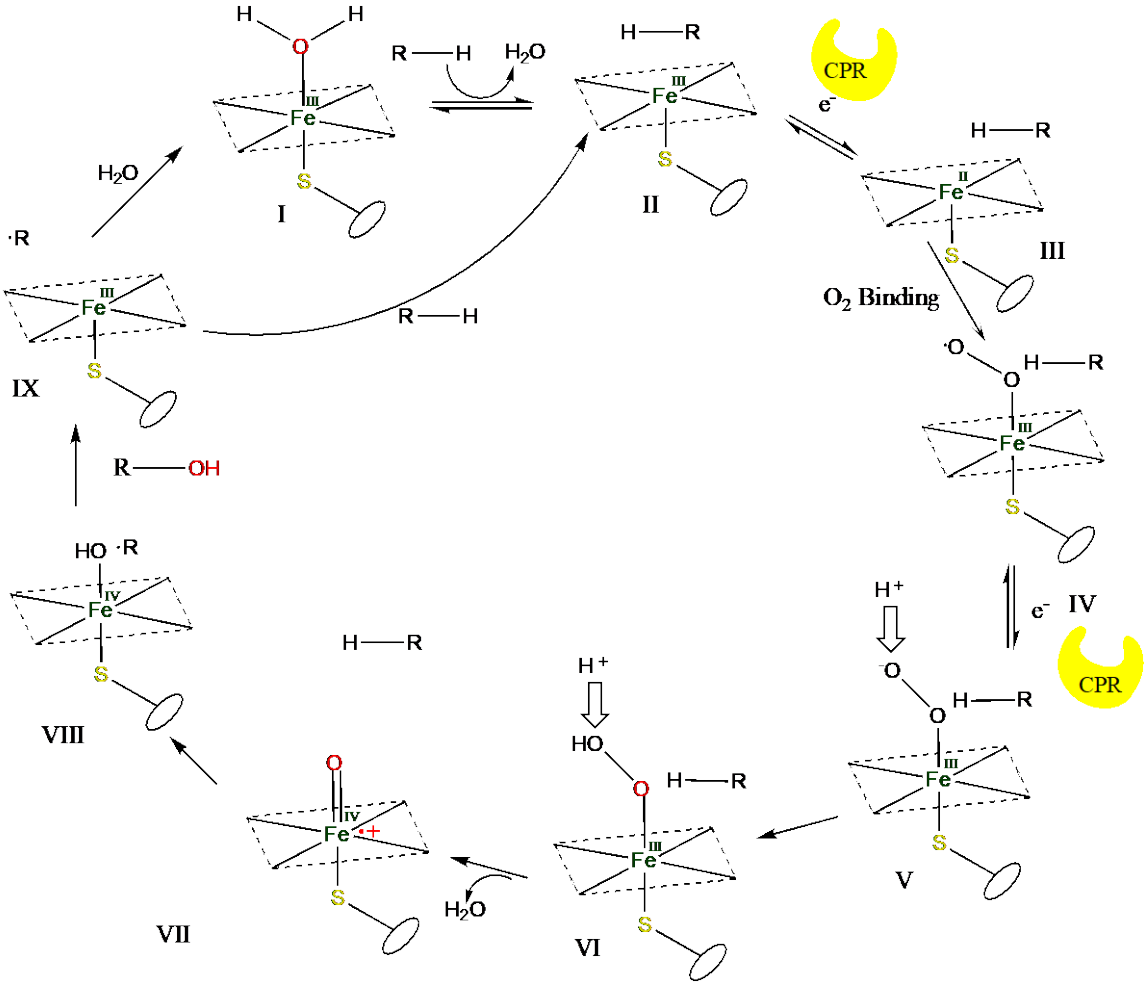


Figure 9. Catalytic cycle of Cytochrome P450.

Compound I (complex VII in the figure) abstracts a hydrogen atom from the substrate omega carbon leaving a carbon radical. Radical rebound generates the ROH as shown in the Figure 9. Hydroxylated substrate then gets replaced by water to start the catalytic cycle all over again.

Active Site of P450$_{BM-3}$

CYP450 has a heme porphyrin that is located in the hydrophobic core of the enzyme. This is where the active site is.[32] Substrates enter the active site via access channels that connect the heme center and the surface.

Below is a depiction of a CYP450 active site (Figure 10.). Two residues at the top left corner, Glu267 and Thr268, are known as the acid-alcohol pair. Thr268 is one of the most highly conserved amino acids among P450 enzymes.[31] It plays a role in proton donation and interacts with the heme-bound peroxide intermediate.

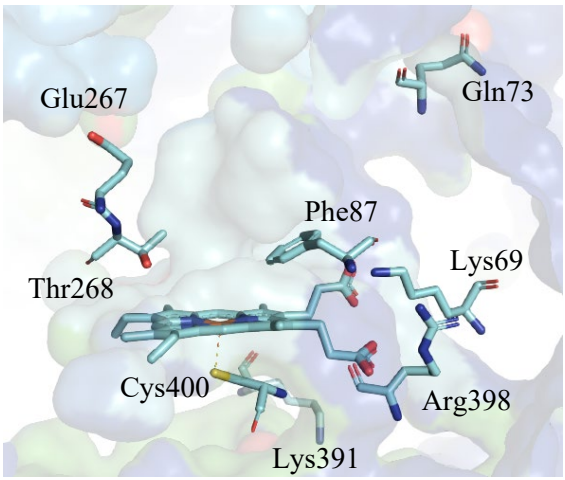


Figure 10. Active Site of P450BM-3. Active site amino acids are represented as stick models and the rest of the protein is seen in the background (surface representation). Carbon atoms are colored in teal. Oxygen, nitrogen and sulfur atoms are colored in red, blue and yellow, respectively. This figure was prepared using PyMOL software.

In the upper right, Gln73 can be seen. It's one of the three residues that binds the carboxylic head of substrate. The other two residues that interact with the substrate's

carboxylic head are Ala74 and Tyr51. The steric hindrance postulated by the Phe87 residue in the active site accounts for the curling of the substrate's fatty acid chain. This exposes the ω-1 and ω-3 positions to the heme center to for oxidation. It is believed that the Arg398 interaction helps gating water into the active site. Near highly conserved Cys400 there are positively charged residues such as Lys391.[32] This is believed to be where the reductase attaches and transfers electrons to the heme center.

Autodock simulations

Autodock is molecular docking software that is widely used in drug design. This software evaluates binding in two steps.[34,35] The ligand and protein are assumed to start in an unbound conformation. The intramolecular energetics are estimated for transition from these unbound states to the conformation of the ligand and protein in the bound state. The second step then evaluates the intermolecular energetics of combining the ligand and protein in their bound conformations. Both of these calculations are done in a user-defined grid surrounding the binding area.[36]

The Autodock forcefield is composed of six pair-wise energy calculations (V) of bound and unbound state protein (P) and ligand (L), and the entropy lost due to binding (DS_{conf}).[34] (Equation 1) (Figure 11.)

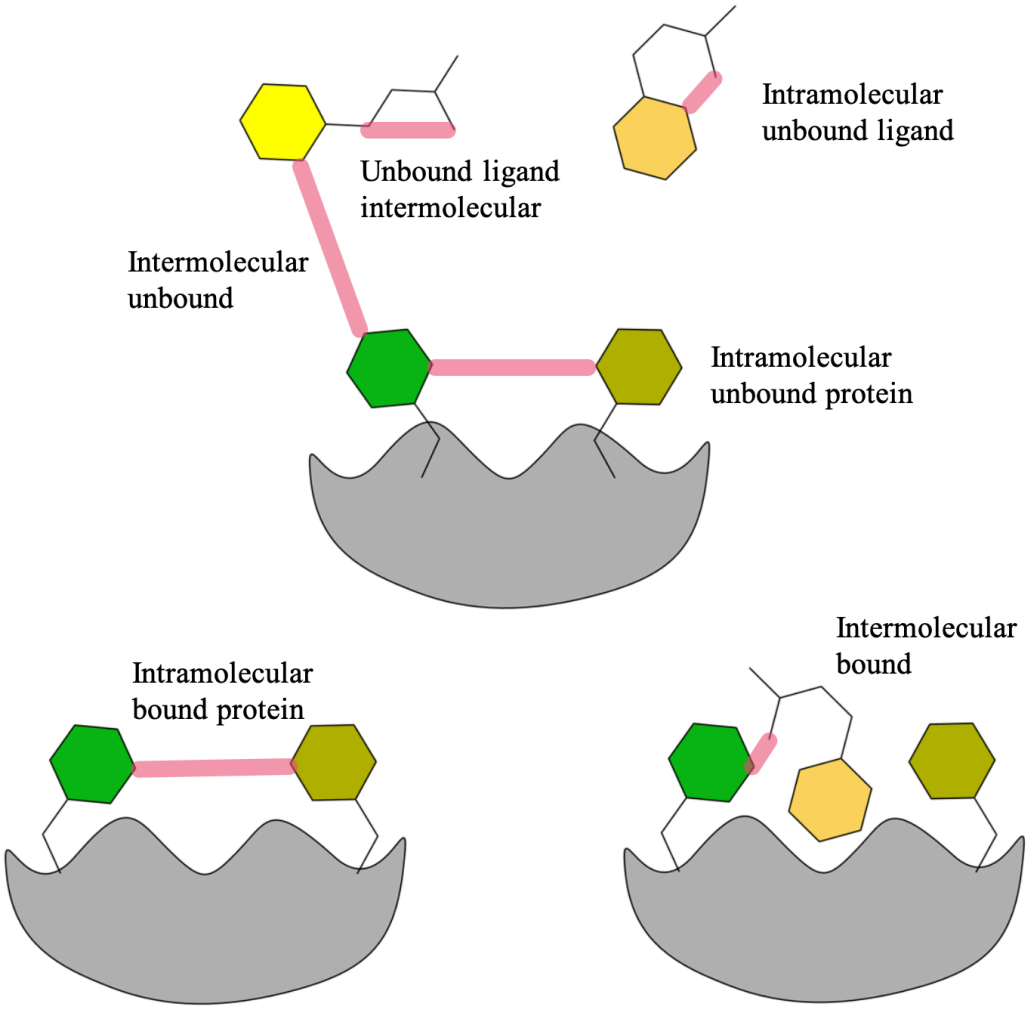


Figure 11. Intermolecular and Intramolecular Interactions Between Protein and Ligand.The grey shapes are a minimalized representation of the macromolecule. Yellow hexagon shapes represent the ligand molecules and green hexagon shapes represents the amino acids in the protein. Pink bars represent intra- and inter-molecular interactions between protein and ligand.

Each pair wise energy scoring function is calculated by taking the summations of all inter molecular potentials of all pairs of ligand atoms, i, and protein atoms, j, as a function of their distances, r (Equation 2).

Equation 1

$$DG = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + \left(V_{bound}^{P-L} - V_{unbound}^{P-L} + DS_{conf}\right)$$

Equation 2

$$V = W_{Vdw} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + W_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) + W_{elec} \sum_{i,j} \left(\frac{q_i q_j}{e(r_{ij}) r_{ij}}\right) + W_{sol} \sum_{i,j} \left(S_i V_j + S_j V_i\right) e^{(-r_{ij}^2/2\sigma^2)}$$

The van der Waals term (W_{Vdw}) refers to the Lennard-Jones 6-12 potential for dispersion/repulsion interactions, where the parameters A and B refer to attraction and repulsion forces, respectively. These parameters are based on the Amber force field. Overall, this term refers to the shape complementarity between the ligand and the receptor.[36]

The second term is the directional hydrogen bond term (W_{hbond}) based on a 10/12 potential. "Here the parameters C, D are assigned to give a maximal well depth of 5 kcal/mol at 1.9 $\AA$ for Hydrogen bonds with oxygen and nitrogen, and a well depth of 1 kcal/mol at 2.5 $\AA$ for hydrogen bonds with sulfur." [34] The function E(t) is the directional angle-based weight from an ideal H-bonding geometry. Possible electrostatic interactions are calculated using a coulombic potential term (W_{elec}) where $q_i q_j$ terms are charges of the interacting atoms, r is the distance between two interacting atoms and e is the charge of an electron. Finally, the desolvation term (W_{sol}) is defined, accounting for both

entropy and enthalpy changes that occur due to the release of water molecules that surround the ligand and fill the protein binding site upon the ligand binding.[36]

As explained in Olson *et al.,* 2012 "the desolvation term is a function of the solvent-accessible surfaces of ligand (S_i) and protein (S_j)[37] and accounts for the implicit bulk waters present in the docking volume (V)". An exponential term with distance-weighting factor $\sigma = 3.5$ Å is also used.

Autodock calculations

Autodock calculations are done in four steps. Step 1 is preparing an Autodock coordination file, also known as an Autodock structure file (PDBQT), using AutodockTools software. PDBQT is an extension of Protein Data Bank (PDB) files. PDBQT includes atomic partial charges and atom types in addition to x,y,z coordinates. PDBQT files also contain information on the torsion degrees of freedom. AutodockTools is a graphical user interface designed to carry out Autodock calculations. It can also be used to create PDBQT files from PDB files.[34] Step 2 is performing Autogrid calculations. In Autogrid, the protein is embedded in a 3-dimensional grid of points at which interaction energies will be calculated (generally as fields). During the docking process, these precalculated grid interactions are referenced to speed up ligand energy evaluation. Thus, the protein itself is discarded. Only the field grids calculated from the protein structure are used further on.

In step 3, Autodock is used to search conformational space for the ligand. By this point, the energy functions have already been calculated using the force field and grid information.[34] The search of conformational space is carried out by several search

methods. Though traditional genetic algorithms and simulated annealing methods are available, the Lamarckian genetic algorithm (LGA) is known to be the most efficient one. During the docking process, Autodock is run several times to give many docking conformations. These docking conformations are then analyzed for binding energy and consistency of results, to identify the minimum energy structures.

The final step of docking calculation is analysis using AutodockTools. AutodockTools provides a number of methods to analyze results including visualizing conformations, interactions between ligand and protein, and visualizing the affinity potential created by Autogrid and clustering tools.

Methods

Materials: Expression vector pEX-N-His-GST and Entry vectors, Cyp4f14 (NM_022434) Mouse Tagged ORF Clone and Cyp4f39 (NM_177307) and Mouse Tagged ORF Clone were purchased from Origene (Rockville, MD, USA). Restriction endonucleases; *AsiSI*, *MluI* and *EcoR1,* Antarctic phosphatase, CutSmart buffer, T4 DNA Ligase, 10x T4 DNA ligase reaction buffer, 6x DNA loading buffer, 100 bp DNA reference ladder, 1 Kb DNA reference ladder, high efficiency competent *E. coli* cells for Transformation Protocol (C2992) and Expression (*E. coli* BL21: C2530H) were purchased from New England BioLabs (Ipswich, MA, USA). Ultrapure water for all the DNA work was taken from a Milli-Q ® Direct 8/16 system from Millipore SAS (Molsheim, France). SOC medium (Super Optimal broth with Catabolite repression) that was used in transformation contains 20 g/L tryptone, 5 g/L yeast extract, 4.8 g/L $MgSO_4$, 3.603 g/L dextrose, 0.5g/L NaCl , and 0.186 g/L KCl. Difco™ Luria Bertani broth (LB) was used to prepare culture media. Isopropyl β-D-1-thiogalactopyranoside (IPTG) and ampicillin sodium salt were purchased from (St. Louis, MO, USA). Lysis Buffer used in expression studies contains 100 mM KPi with 100 mM KCl, pH 8.0. 11-*p*NCA was synthesized by our lab prior to this project. Agarose was purchased from Mallinckrodt Baker Inc. (Phillipsburg, NJ, USA). GelRed Nucleic Acid Gel Stain (10000X) was purchased from Biotium (Hayward, CA, USA). TAE (Tris, Acetic acid, EDTA pH 8.3) buffer was prepared as a 50X stock solution. DNA purification was done by the Promega PureYield™ Plasmid Midiprep DNA purification kit purchased by Promega (Madison, WI, USA). B-PER™ Bacterial Protein Extraction Reagent (Catalog number: 78243) was

purchased from Thermo Scientific™ (Rockford, lL, USA). DNA Sequencing was done using Seqwright (Woodlands, TX, USA).

Equipment: Eppendorf Centrifuge 5810 R from Eppendorf (Hamburg, Germany) was used to carry out centrifugation of cell cultures with volumes higher than 1.5 mL. Centrifugations of volumes < 1.5 mL were carried out in Legend Micro 21 Centrifuge from Thermo Fisher Scientific (Waltham, Massachusetts, USA). Incubation of cell cultures and reagents was carried out with a VWR Shaking Incubator 1575 from VWR™ International (Suwanee, Georgia, USA). DNA concentrations were measured using the Eppendorf Biophotometer. For spectroscopic studies, Jasco V-750 double beam UV-visible spectrophotometer from Jasco Incorporated (Easton, MD, USA) and Varioskan® Flash Spectral Scanning Multimode Reader and SkanIt Software 2.4.3 RE for Varioskan Flash. Volatile organic compounds were analyzed using Agilent Technologies 7890A series gas chromatograph (GC) coupled to the 5975C series mass selective detector (MSD, located at the TRIES Lab, Sam Houston State University, Huntsville, Texas).

Software: For structural analysis and molecular visualization PyMOL and Biovia Discovery Studio Viewer was used. Autodock was used for molecular docking studies. GaussView 09 was used to optimize the ligands. Uniport and NCBI (National Center for Biotechnology Information) were used to retrieve protein and DNA sequence data. MEGA and Snapgene 2.0 programs were used for sequence alignments and visualization. OpenBabel software was used to convert PDBQT files in to PDB files. Python 3.8 was used for GC-MS data analysis and Visual studio and Anaconda were used as Python IDE (Integrated development environment).

Subcloning

For TrueORF Destination vector pEX-N-His-GST (Figure 12) digestion, 2 µL 10x CutSmart restriction buffer, 0.6 µL *AsiSI*, 0.6 µL *MluI*, 14.8 µL nuclease-free water and 2 µL TrueORF Destination vector were mixed in an Eppendorf tube and then incubated at 37 ℃ for 3 hours.

For TrueORF entry vector (Figure 13.) digestion 2 µL 10x restriction buffer, 0.6 µL AsiSI, 0.6 µL MluI 13.8 µL nuclease-free water and 3 µL TrueORF Entry Vector were mixed in an eppendorf tube and incubated at 37 ℃ for 3 hours.

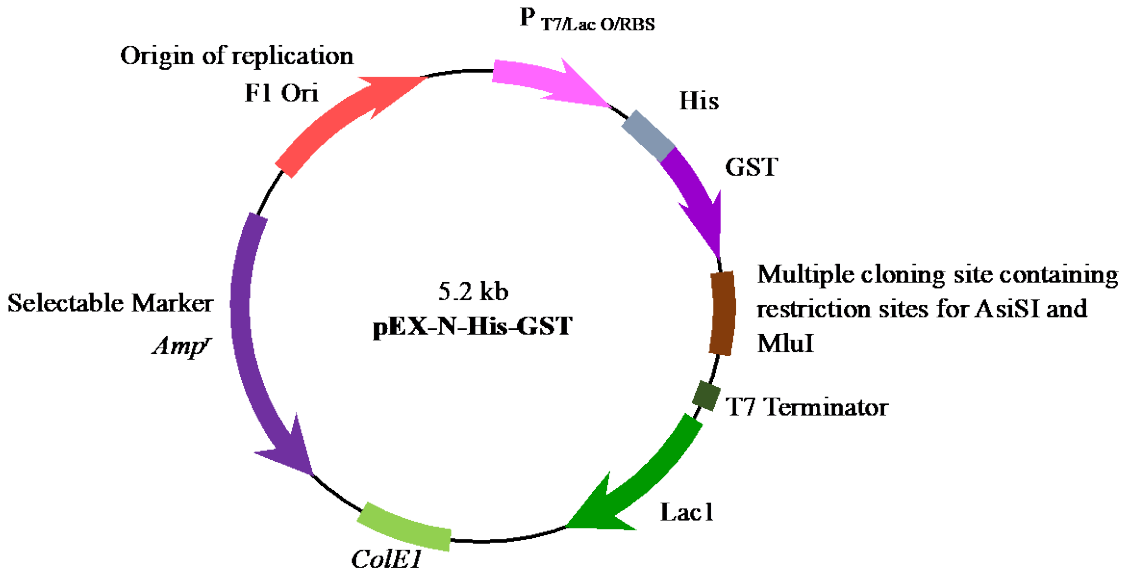


Figure 12. Origene pEX-N-His-GST Expression Vector. [37]

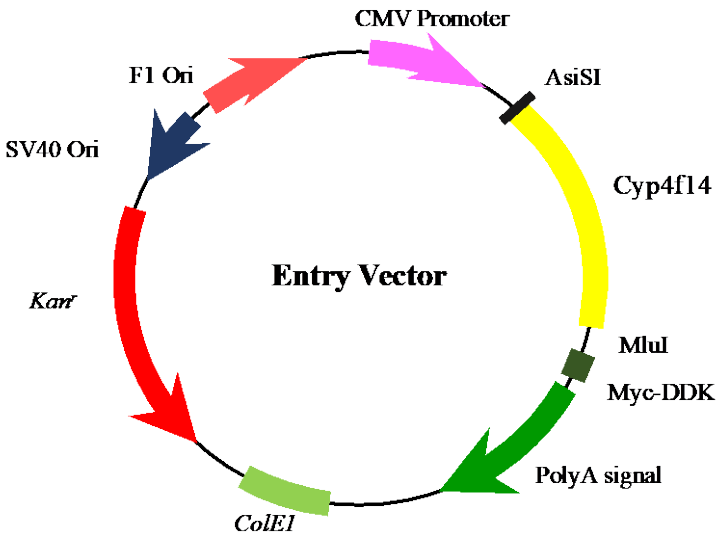


Figure 13. Entry Vector Containing Cyp4f14 (Myc-DDK-tagged) - Mouse cytochrome P450, family 4, subfamily f, polypeptide 14 (Cyp4f14), transcript variant 1.

Antarctic phosphatase

Antarctic phosphatase (1 µL) was added to 20 µL TrueORF destination vector digestion mixture and incubated in 37 ℃ for 30 minutes. This heat labile phosphatase will dephosphorylate the 5′ end of digested destination vector, making it unable to self-ligate. After a 30 minute incubation, products were purified using Illustra GFX PCR DNA and Gel Band Purification Kits.

DNA ligation

In an Eppendorf tube, 1 µL 10 x T4 DNA ligation buffer, 0.5 µL T4 DNA Ligase (400,000 U/mL), 3.5 µL nuclease-free water, 2 µL digested ORF clone, and 3 µL

digested destination vector were mixed. Then the mixture was incubated at room temperature overnight. This procedure was repeated for Cyp4f39.

Transformation into competent E. coli cells

In order to extract DNA for further analysis, pEX-N-His-GST-Cyp4f14 and pEX-N-His-GST-Cyp4f39 recombinant plasmid vectors needed to be transformed into competent *E. coli* cells. For transformation, first a tube of competent *E. coli* cells was thawed on ice for 10 minutes and then 1–5 µL containing (depending on DNA concentration) 0.001–100 ng of plasmid DNA was added to the cell mixture. The tube was carefully flicked 4–5 times to mix cells and DNA. The mixture was placed on ice for 30 minutes. After that, a 45-second-long heat shock[38] was applied by exposing the bacteria sample to 42 °C by placement of the sample in a water bath. This sudden increase in the temperature was expected to cause the bacteria to open pores in the membrane and take up plasmid DNA.

Then the mixture was transferred immediately to ice for 2 minutes. After that, 250 µL of room temperature SOC (Super Optimal broth with Catabolite repression) medium was added into the mixture and incubated in a shaker (250 rpm) incubator at 37 °C for 55 minutes. SOC is a nutrition rich medium that maximizes the transformation efficacy of *E. coli* cells and represses transcription from lac-derived promoters to avoid complications in the event that the expressed protein is toxic to the host cell.

After incubation, 50–100 µL of cell culture was spread onto a selection plate (LB+ ampicillin) and incubated overnight at 37 °C. A single colony was selected and inoculated with 100 mL liquid broth media (LB+ ampicillin) and incubated overnight at

37 °C. After overnight incubation, broth cultures were then transferred into 50 mL falcon tubes and DNA extraction was done using Promega PureYield™ Plasmid Midiprep DNA purification kit according to the manufacturer's manual.

Agarose gel electrophoresis

Agarose gel (1 % (w/v)) was prepared by dissolving 0.50 g of agarose in 1X TAE buffer (50 mL) with microwave heating for 20-30 seconds. The heated mixture was allowed to cool for one minute and then GelRed nucleic acid stain (0.5 µL) was added and mixed well to obtain a homogenous solution. The solution was poured onto the assembled gel tray. Any air bubbles formed were removed using a pipet tip. Then the gel comb was inserted, and the gel was set aside to set. Once the gel solidified, the comb was slowly removed, and 30 mL of TAE buffer was poured into the gel container slightly covering the entire gel and the wells. After that, one by one samples (1 µL GelRed loading dye solution + 4 µL sample) were transferred into wells. The gel was covered with 1X TAE buffer. The gel was electrophoresed at a constant voltage of 80 V for 60 minutes. Finally, the gel was observed under UV transillumination.

Transformation in to BL21 cells

In order to purify Cyp4f14 (Mouse cytochrome P450, family 4, subfamily f, polypeptide 14, transcript variant 1), pEX-N-His-GST-Cyp4f14 recombinant plasmid vector needs to be transformed into competent BL21 *E. coli* cells. For transformation, first, a tube of BL21 competent *E. coli* cells was thawed on ice for 10 minutes and then 1–5 µL containing (depending on DNA concentration) 0.001–100 ng of plasmid DNA

was added to the cell mixture. The tube was carefully flicked 4–5 times to mix cells and DNA. The mixture was placed on ice for 30 minutes. After that, a 45-second-long heat shock was applied by exposing the bacteria sample to 42 °C by placement of the sample in a water bath. This sudden increase in the temperature was expected to cause the bacteria to open pores in the membrane and take up plasmid DNA.

Then the mixture was transferred immediately to ice for 2 minutes. After that 250 µL of room temperature SOC (Super Optimal broth with Catabolite repression) medium was added into the mixture and incubated in a shaker (250 rpm) incubator at 37 °C for 55 minutes. SOC is a nutrition rich medium that maximizes the transformation efficacy of *E. coli* cells.

After incubation, 50–100 µL of cell culture was spread onto a selection plate (LB+ ampicillin) and incubated overnight at 37 °C. A starter culture with lysogeny broth (LB) liquid broth media with 100 µg/mL ampicillin was then inoculated with a single colony and left to culture until slightly cloudy. When the culture started to get cloudy, optical density (OD) measurements at 600 nm were taken using JASCO-750 UV-Vis spectrophotometer.

UV-visible spectroscopy

The quantitative measurement program on the JASCO 750 was used to assess the growth of a bacteria sample by measuring the optical density at 600 nm. Two cuvettes were filled with the solvent (LB broth). It is important to keep the outside of the cuvette clean and free of fingerprints. Also, there should not be any bubbles in the solution. Dirt,

including fingerprints, on the outside walls of cuvette and water bubbles in the solution can interrupt the light beam and affect the reading.

The cuvettes should be aligned properly. Often the cuvette has two sides which are meant for handling (may be grooved) and not meant to shine light through. Once the cuvettes containing the solvents are inside, the sample chamber was closed, and measurement was taken for the blank sample.

Once a measurement was taken, the cuvette in the sample chamber was taken out and refilled with the bacterial cell culture. The cuvette was wiped properly to make sure there were no water bubbles inside or out. Once all the walls of cuvette were clean, it was inserted into the sample compartment and a new measurement was taken.

Measurement of the optical density of the culturing medium continued periodically as the bacteria grew. At the first measurement where the OD_{600} equaled or exceeded 0.5, 0.1 mL of 1 M isopropyl β-D-1-thiogalactopyranoside[39,40] (IPTG) was added to the growing *E. coli* culture to induce Cyp4f14 gene transcription and subsequent protein synthesis. After adding IPTG, the culture was allowed to grow for an additional 48-72 hours in 37 °C while being monitored for a color change. The appearance of red color in the growth medium suggested that the Cyp4f14 gene was successfully being transcribed into the Cyp4f14 RNA and then translated into enzyme. Different IPTG concentrations (0.5 mM, 0.8 mM, 1.0 mM and 1.2 mM) were used in different culture samples for optimization under different OD_{600} values.

Isolation of recombinant protein from the transformation culture.

Frozen pellets of *E. coli* expressing Cyp4f14 were thawed in ice. Then, 4 mL of B-PER™ Bacterial Protein Extraction Reagent[41] per gram of cell pellet and 4 mL of Lysis buffer (100 mM KPi with 100 mM KCl, pH 8.0) were added (50/50 mixture of B-PER and Lysis buffer) and pipetted up and down to get a homogeneous mixture. This mixture was incubated for 10-15 minutes in room temperature. Finally, the lysate was centrifugated at 15,000 g for 5 minutes to separate soluble protein from the insoluble protein.

After an optimum level of growth was achieved for 100 mL cultures, it was scaled up to 250 mL and eventually 500 mL cultures to increase yields. It is comparatively difficult to extract membrane bound protein while retaining their activity or correct folding.

pNCA assay

First, 10 mM of 11-*p*NCA ($C_{16}H_{23}NO_5$) was dissolved in 2.13 mL 50 mM K_2CO_3. To 1 mL of lysate buffer, 20 µL of 10 mM NADPH was added using a micropipette. Then 10 µL of *p*NCA was added using a syringe. Finally, 1 µL of cell lysate was added to the mixture and inverted 3 times to mix. Assay activity was measured using a Varioskan Flash double monochromator microplate-reading UV-Visible spectrophotometer.

Generation of homology models

Cyp4f14 and Cyp4f39 protein structures were modeled using the Iterative Threading Assembly Refinement (I-TASSER)[42] server based on the amino acid sequences obtained from Universal Protein Resource (UniProt). Structural artifacts of the developed model were removed by adding the missing heme residue (APPENDIX B). Coordinates for the crystal structure 1jpz were retrieved from Protein Data Bank (PDB) and the heme was extracted. PDB structure Ijpz was used for comparison and structure determination.

Autodock 4.2

PDB structures of Cyp4f14 and Cyp4f39 were generated using i-TASSER, a server for protein structure and function prediction. Missing heme extracted from PDB structure 1jpz, was incorporated into the i-TASSER generated Cyp4f14 and Cyp4f39 models by PyMOL software.

Semi empirical/ PM6 calculations were performed to optimize *p*NCA and 20-HETE using Gaussian 09 software. The LTB4 geometry for the ligand file was extracted from PDB structures using PyMOL software.

Autodock Tools 1.5.6 software was used to calculate the docking scores. Polar hydrogens were added and Gasteiger charges were incorporated into the macromolecules Cyp4f14 and Cyp4f39. These were then saved as a PDBQT file. Ligands were prepared by adding Gasteiger charges, polar H-bonds, and all the rotatable torsions (Figure 14).

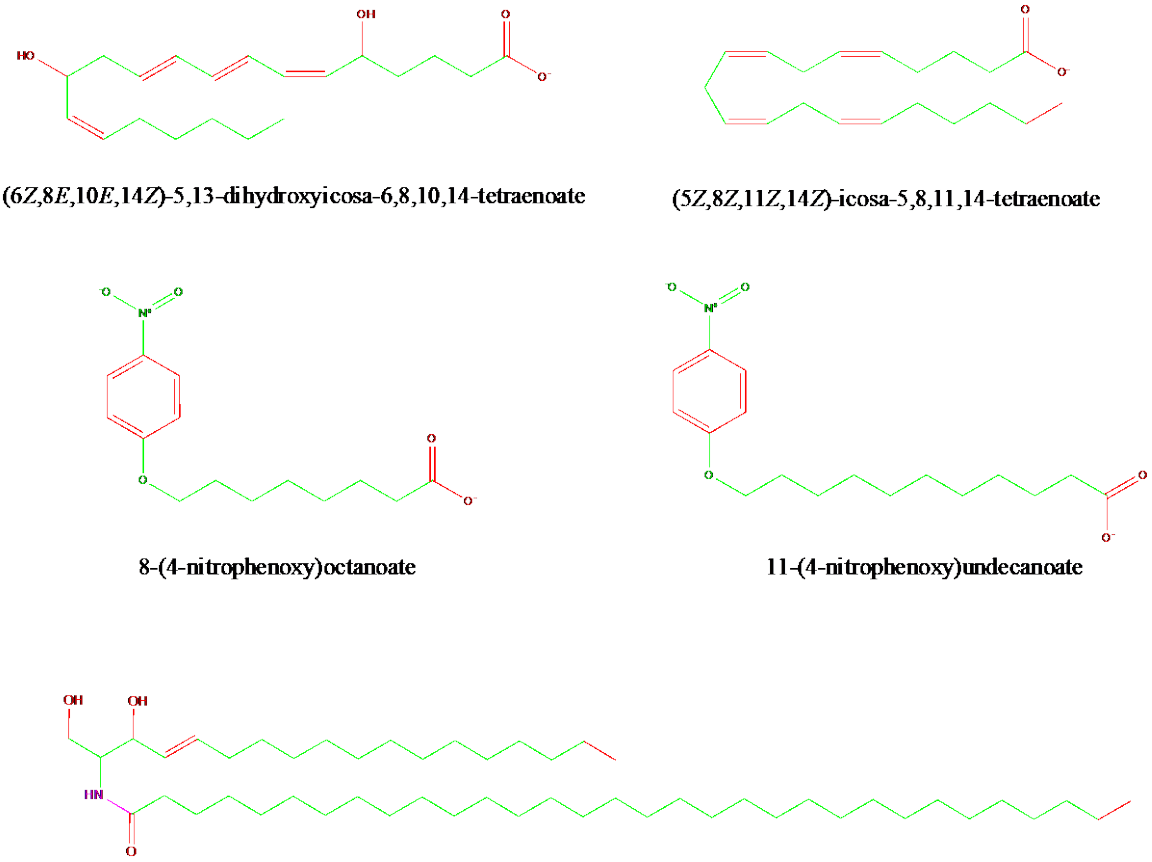


Figure 14. Torsions/ Ligands. Rotatable bonds are shown in green, unrottable bonds are in red and non-rotatable bonds are in magenta. LTB4 has 15/32 rotatable bonds, arachidonic acid has 14/32 rotatable bonds, 8-(4-nitrophenoxy)octanoate (8-*p*NCA) has 12/32 rotatable bonds and 11-(4-nitrophenoxy)octanoate (11-*p*NCA) has 15/32 rotatable bonds and (E)-N-(1,3-dihydroxyoctadec-4-en-2-yl)dotriacontanamide (long chain ceramide) has 47/32 rotatable bonds, and 1 non-rotatable bond.

Next, Autogrid calculations were performed using a user-defined grid box (Figure 15.), which was prepared for each receptor molecule covering the active site amino acids. Grid parameters for Cyp4f14 and Cyp4f39 were selected based on the spread of their

active site. Apart from this, several mutations were made using the PyMOL mutagenesis tool and used in docking.

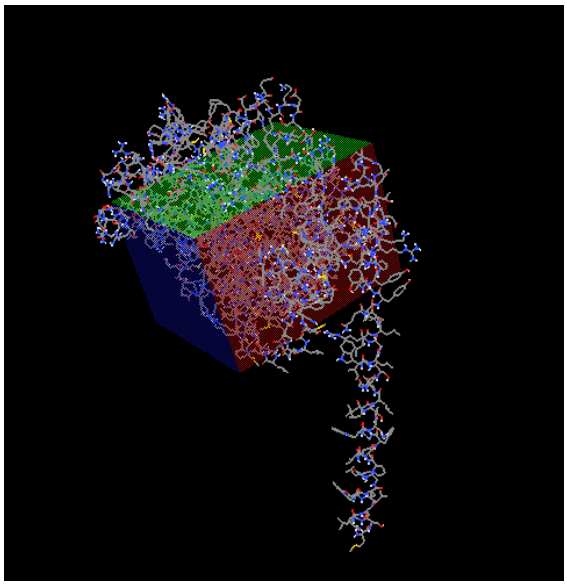


Figure 15. Screen Capture of the Grid Box Created Using Autodock 4.2. The protein in the image is Cyp4f14.

Results

Subcloning

Recombinant *E. coli* cells containing pEX-N-His-GST-Cyp4f14 vector were grown on LB agar plates containing ampicillin (Figure 16). These plates only allow successfully transformed cells to grow. This is because the pEX-N-His-GST vector contains an ampicillin resistance gene, whereas the entry vector does not. As can be seen in Figure 16, white colonies indicate successful transformations. However, this does not necessarily indicate that the vector has incorporated Cyp4f14 correctly.

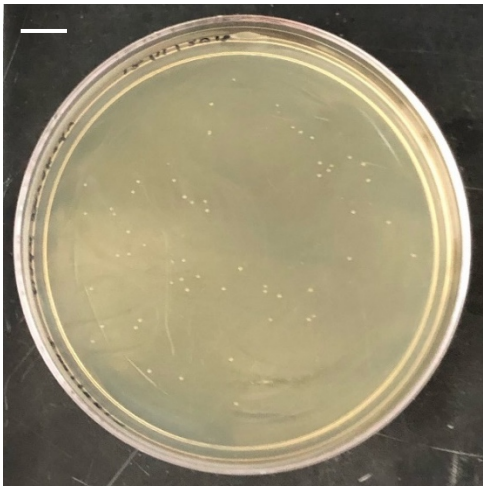


Figure 16. Transformation Result. Image of white, oval shape pEX-N-His-GST-Cyp4f14 recombinant *E. coli* colonies on LB-amp media. Scale bar = 1 cm.

Therefore, to further confirm the results, recombinant *E. coli* cells were grown in ampicillin containing LB broth media and their DNA was purified using the Promega PureYield™ Plasmid Midiprep System. After that, purified recombinant pEX-N-His-GST-Cyp4f14 plasmid and the destination vector were digested with *EcoRI* restriction

enzyme. The predicted sequence of the recombinant plasmid would have only three restriction sites for *EcoRI* (2 in the Cyp4f14 insert and 1 in the expression vector). Thus, we expected to see three bands in the agarose gel electrophoresis including a band around 100 bp. However, the agarose gel electrophoresis showed unusual results by showing four bands (Figure 17). This could be due to some DNA impurities present in the sample. We also saw three bands for *EcoR1* digested destination vector.

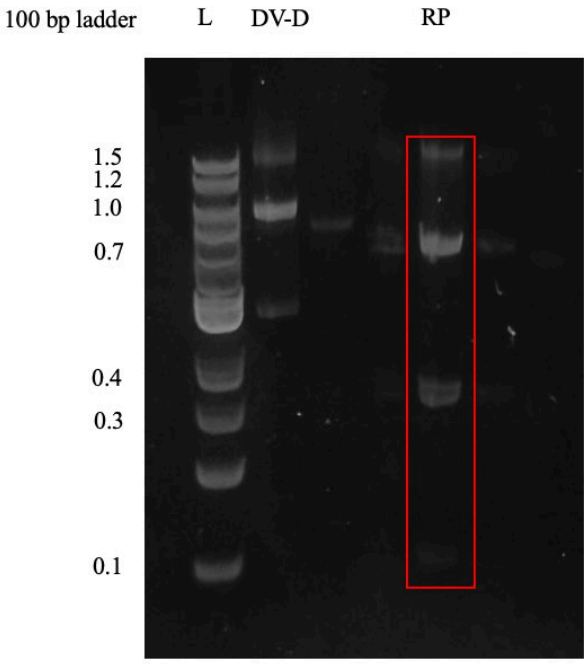


Figure 17. 1% Agarose Gel Verification of Cyp4f14 Recombinant Vector After Digestion with *EcoRI* Restriction Enzyme. Lane one contains a 100 bp DNA ladder, lane two contains destination vector digested by *EcoRI* and lane five contains the digested destination vector RP (Recombinant Plasmid). Lane 3 and 4 do not contain anything related to this experiment. The cleaved Cyp4f14 DNA fragment is seen at 100 bp. DNA was stained using GelRed dye and visualized under 365 nm UV light.

Thus, the recombinant vector was digested with the *AsiSI a*nd *MluI* restriction enzymes. These are the original restriction enzymes used to digest the destination vector and the entry vector. Therefore, the enzymes would remove the insert that was added to the plasmid. This digestion should result in two bands, the destination vector (5.2 kb) and the Cyp4f14 DNA insert (1.5 kb). As expected, we observed two bands, one around 5-6 kb and a band at 1.5 kb (Figure 18.). The identity of the 1.5 kb band as Cyp4f14 was confirmed by commercial DNA sequencing and a BLAST search.

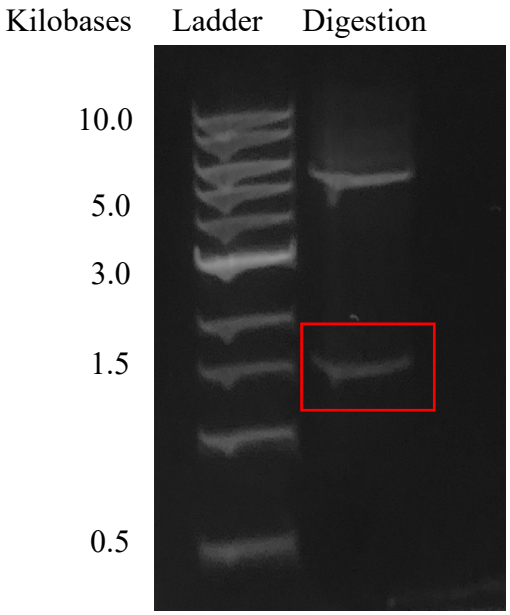


Figure 18. 1% Agarose Gel Verification of Cyp4f14 Recombinant Vector After Digestion with *AsiSI* and *MluI* Restriction Enzymes. Lane 1 contains a 10 kb DNA ladder and lane two contains the digested DNA. The cleaved Cyp4f14 DNA fragment is seen at 1.5 kb. DNA was stained using GelRed dye and visualized under 365 nm UV light.

UV-visible spectroscopy of Cyp4f14

In contrast to prokaryotic CYPs, which are soluble, mammalian CYPs are anchored to the phospholipid bilayer of the endoplasmic reticulum (ER) via an N-terminal hydrophobic transmembrane -helix (A′- helix). Many studies, including the first reported crystal structure of mammalian CYP, have been published with truncated enzyme structures that lack the N-terminal transmembrane helix.[43]

In this study, the untruncated Cyp4f14 was used and initial expression studies with Cyp4f14 lysate in 50/50 B-PER and lysis buffer solution is reported (Figure 19 and Figure 20.). Low expression of Cyp4f14 was observed in both UV-Vis data for substrate free Cyp4f14 lysate and for the *p*NCA assay. Low levels of expression could result from protein misfolding and protein denaturation that may have occurred during purification steps. The B-PER reagent used in the purification process, showed a peak around 250-300 nm. Because we used the cell lysate to measure the spectrum, a lot of interference may have been present. Therefore, the UV-Vis spectra observed was enhanced using polynomial background subtraction applied through the LINEST function on MS Excel.

After removing background interference, the UV Vis spectrum of Cyp4f14 shows evidence of characteristic heme spectra (Figure 20.). The Soret band is visible at 400-420 nm. The Q-bands can be seen at 535 nm and at 560 nm. The broad Soret peak indicates overlapping of many peaks. This is due to the presence of denatured protein in the cell lysate and other forms of heme and heme synthesis and degradation products.

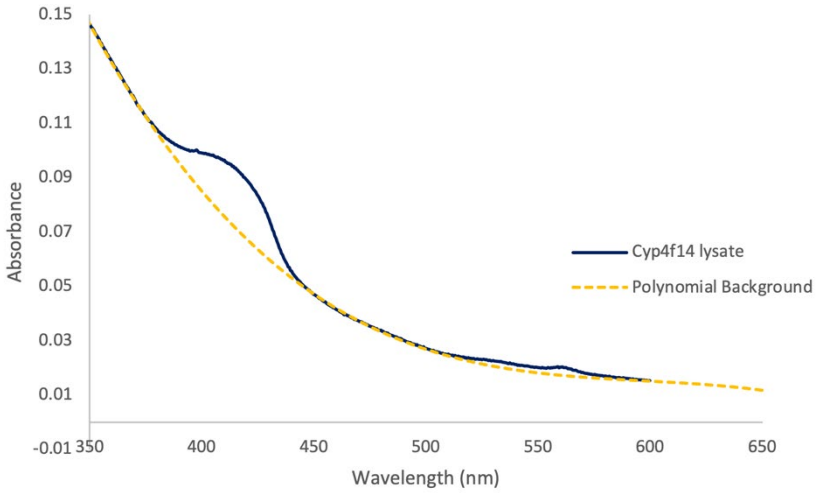


Figure 19. UV-Visible Spectrum of Cyp4f14. Substrate free Cyp4f14 lysate in 50/50 mixture of lysis buffer and B-PER reagent (*blue solid line*) with the Soret maximum at 416-420 nm. Polynomial background generated using MS Excel is shown as a yellow dotted line.

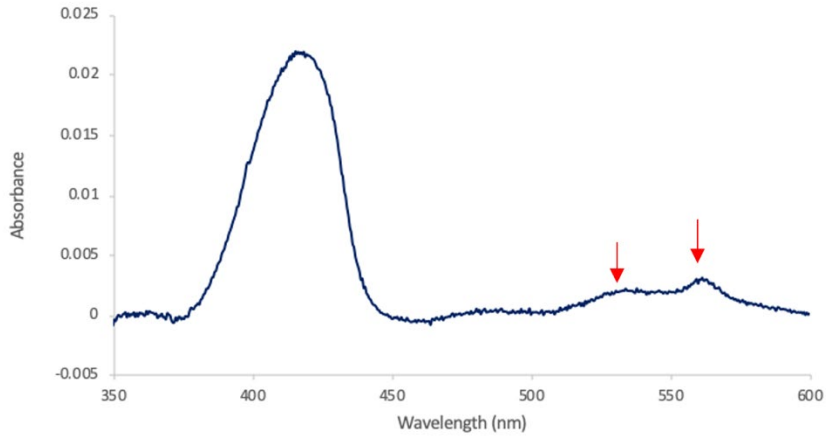


Figure 20. Background-Subtracted UV-Visible Spectrum of Cyp4f14. Substrate free Cyp4f14 lysate in 50/50 mixture of lysis buffer and B-PER reagent after background correction using LINEST function (MS Excel). Red arrows indicate the Q-bands at 535 nm and 560 nm, respectively.

pNCA assay of Cyp4f14

To measure the activity of the enzyme, we used the *p*NCA assay. This assay has been previously used with $P450_{BM-3}$ and $P450_{BM-3}$ mutants (F87A), successfully.[27,28] In the study done by Schwaneberg and team, they reported that the *p*NCA assay works for both purified and lysed cell extract of $P450_{BM-3}$. This assay has major advantages due to the chromogenic nature of the substrate, and we decided to try it with this enzyme since its reported specificity is fairly similar to that reported for $P450_{BM-3}$. Schwaneberg also reported that chain lengths of 8, 10-12 and 15 work best with $P450_{BM-3}$ and its mutants.

As mentioned in the introduction, *p*NCA is stable in alkaline solution and the leaving *p*-nitrophenolate is more intensely colored than the protonated form prevalent in neutral or acidic solutions. However, P450s lose most of their activity in alkaline solutions. Thus, to avoid this, we used a pH of 8.0. This is the pH of the sample in the 50/50 lysis buffer and B-PER containing mixture. Calculations using the Henderson-Hasselbalch equation indicate that at this pH, 88% of all *p*-nitrophenol molecules are chromogenic (exhibit a yellow color) as they are in the deprotonated form. The pK_a of *p*-nitrophenol is 7.1[27]

The absorbance at 410 nm was recorded because this wavelength would not have interference from characteristic heme porphyrin peak and NADPH absorption maxima which arise around 350-450 nm and 260-340 nm respectively. Figure 21 shows the results of 11-*p*NCA assays for IPTG-induced Cyp4f14 lysates.

Generally, the activity we observed for the samples induced at OD_{600}=0.6 are higher than those for samples induced at OD_{600}=1.0. Further, samples induced with 0.5 mM-0.8 mM IPTG seem to have the highest activity. Both 0.5 mM data sets for 0.6 OD

induction show appropriate curving in the graph and high absorbance. This suggest the activity went up and then leveled off. A similar pattern was observed with 1.0 mM IPTG concentration added at OD_{600}=0.6. However, this increase observed around 7-8 minutes was not a smooth increase. Thus, it may not be true activity. The wobbling observed in the 0.8 mM (red line) and 1.0 mM (blue line) could be indicative of air bubbles. Overall, our data indicates the presence of trace amounts of *p*NCA. Although ideally the experiment would be repeated to avoid air bubbles, complications of the COVID-19 epidemic prevented further expression studies.

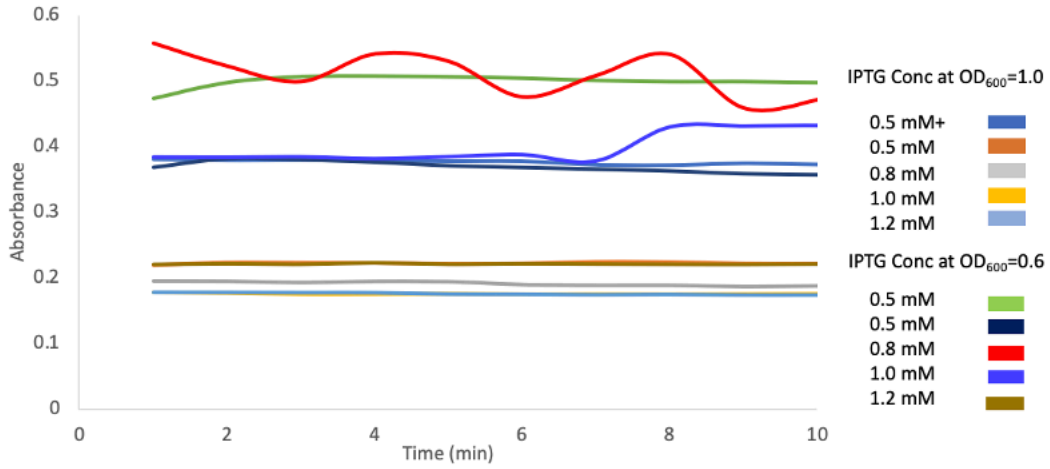


Figure 21. pNCA Oxidation Assay Results. Absorbance vs Time, ω-hydroxylation of 10 mM *p*-nitrophenoxyundecanoic acid by Cyp4f14 cell lysate in 50/50 lysis buffer (pH= 8.0) and B-PER reagent. Four different IPTG concentrations was used to measure the activity of samples that were induced at OD_{600} = 1.0 and OD_{600} =0.6. Absorbance was measured at 410 nm. The first sample (0.5 mM+ IPTG) contains extra IPTG, which was added directly to the microplate. OD_{600}= 0.6 contains 0.5 mM added twice to balance both sides of the microplate.

We also carried out optimization with different IPTG concentrations. Our results show 1.0 mM IPTG to be optimal for bacteria growth. The other IPTG concentrations that were used are 0.5 mM, 0.8 mM, 1.0 mM, 1.2 mM and 2.0 mM. Higher IPTG concentrations (> 2.0 mM) resulted in clear solutions, indicating no bacterial growth. We also noticed that cultures incubated at room temperature tend to grow more slowly and resulted in lower yields than the cultures incubated at 37 °C. This indicates the importance of temperature for growth of the bacteria. The low activity observed using UV-Vis spectra data and the *p*NCA assay, combined with complications of the COVID-19 epidemic that prevented us from continuing studies that required access to the research laboratory, caused us to switch strategies and pursue computational studies of these enzymes.

Sequence alignment CYP4F family

Cytochrome P450s are a diverse family of monooxygenases with high sequence variability. Therefore, P450s are capable of catalyzing oxidation of a huge number of different substrates. CYP4F is a relatively recently identified CYP subfamily.

This study used a number of CYP4 family proteins including six human genes (CYP4F2, 4F3, 4F8, 4F11, 4F12 and 4F22), 4 rat genes (CYP4F1, 4F4, 4F5 and 4F6), and 6 mouse genes (Cyp4f13, 4f14, 4f15, 4f16, 4f18 and 4f39)[2]. A bacterial P450, $P450_{BM-3}$, also oxidizes fatty acids and eicosanoids and is one of the most widely studied P450s historically, so it was included as well. Sequence alignment of these genes with $P450_{BM-3}$'s heme domain sequence was used to identify important motifs in this family.

Human CYP4F2 and CYP4F3, rat CYP4F1, CYP4F4, CYP4F5 and CYP4F6, and mouse Cyp4f14 and Cyp4f18, are all known to metabolize LTB4. Mouse Cyp4f39 and CYP4F22, other P450s of interest that are known or expected to metabolize very long fatty acid chains, were included in this alignment as well.

The alignment shows the motif FXXGXRXCXG is fully conserved among the CYP4F family enzymes we studied.[44] CXG in the 'β−bulge' region, also known as the cysteine pocket, is conserved in all P450s. Interestingly, all the sequences had a CIG. The glycine in the 4th position makes a β hairpin turn with the last (10th position) glycine ligand. The alignment below shows the motif FXXGXRXCXG is fully conserved among the CYP4F family enzymes we studied (*Figure 22*).[44] FSAGPRNCIG is fully conserved among the CYP4F family enzymes we tested except Cyp4f13 (and it is not present in $P450_{BM-3}$).

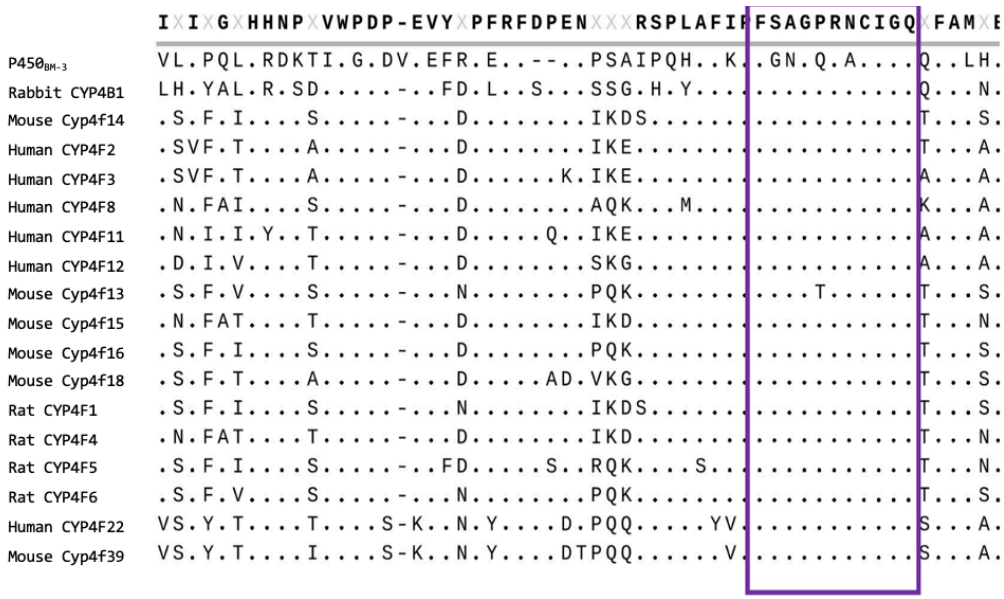


Figure 22. A Portion of the Amino Acid Sequence Alignment Including the FXXGXRXCXG Motif. Aligned amino acids are indicated in dots. Purple box was used to highlight the FXXGXRXCXG motif. Dashes represents gaps, periods indicate that the consensus residue is present at that position.

The other signature motif we explored is the EXXR motif[44] with glutamic acid and arginine conserved. EXXR is found at 321-324 positions in $P450_{BM-3}$ and at 387-390 and 395-398 in Cyp4f14 and Cyp4f39, respectively. EXXR is located in the K-helix and this region is known as meander loop. This positively charged region is proposed to be the site where reductase binds (Figure 23) Apart from $P450_{BM-3}$ and CYP4B1 all tested sequences have ESLR conserved. $P450_{BM-3}$ and CYP4B1 have EALR and ECFR instead of ESLR.

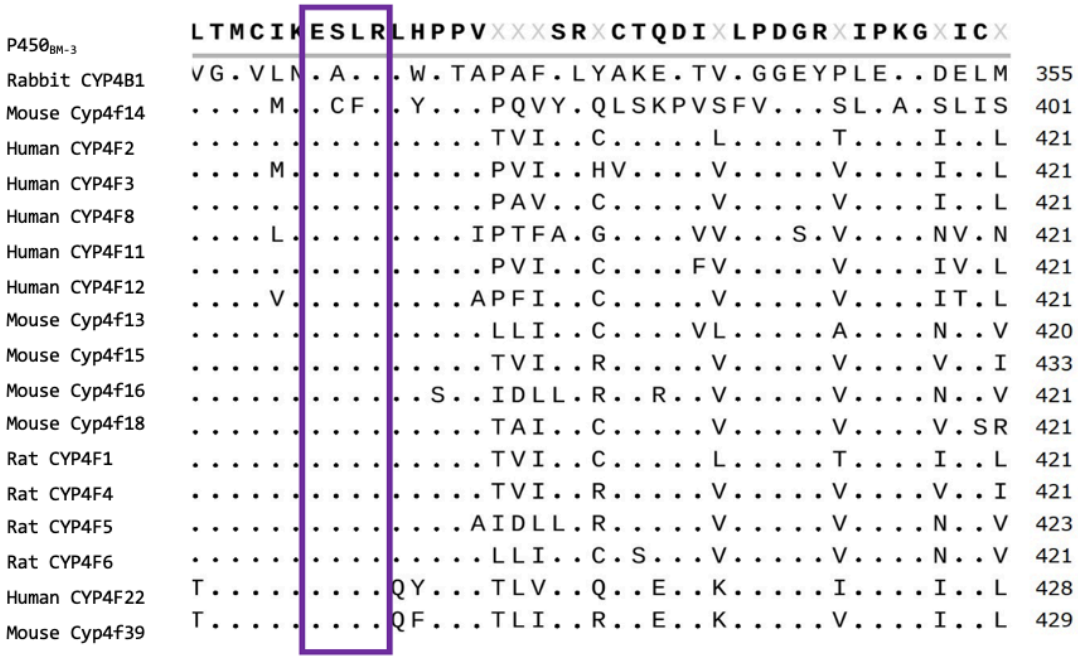


Figure 23. A Portion of the Amino Acid Sequence Alignment Including the EXXR Motif. Aligned amino acids are indicated in dots. The purple box was used to highlight the EXXR motif. Dashes represents gaps, periods indicate that consensus residue is present at that position.

The alignment also shows that residues in I- Helix sequence are almost fully conserved among the CYP4f subfamily. A complete sequence alignment is available in APPENDIX C.

I-TASSER generated model

The I-TASSER (Iterative Threading ASSEmbly Refinement) server predicted 5 models corresponding to the five largest structure clusters for both Cyp4f14 and Cyp4f39. For each target, I-TASSER simulations generate a large ensemble of structural conformations called decoys. Final models are selected using the SPICKER program

where all the decoys get clustered depending on their pairwise structure similarity. The five models that correspond to the five largest clusters are reported. The confidence of each model is measured by the C-score (Confidence score).[42] C-score is a quantitative measure, calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. Typically, C-score varies from -5 to 2. Higher values signify a model with a higher confidence and vice-versa. Based on correlation between C-score and protein length, TM-score and RMSD[45,46] scores are estimated.

TM score is a quantitative assessment metric for measuring protein similarity. TM score varies from 0 to 1, where a perfect match gets a score of 1. RMSD is root mean square deviation of the coordinates. TM score and RMSD are usually used to compare structural similarities between unknown and known protein. However, in this case there are no known structures, so predicting the quality of the model is necessary. Thus, the closest model with highest C-scores was used to predict TM-score and RMSD values. A TM-score higher than 0.5 indicates a model of correct topology and a TM-score less than 0.17 means random similarity.[42] These cutoff values are independent of protein length.

Initially, I-TASSER resulted in 4 models for Cyp4f14 and 3 models for Cyp4f39. For Cyp4f14, model 1 had the highest C-score and TM-score, 0.66 and 0.8± 0.09 . The rest of the Cyp4f14 models had negative C-score values (-0.87, -0.91, -0.38 respectively). For the Cyp4f39, model 1 had a C-score of 0.42, and the TM-score of 0.66±0.13. The rest of the models had negative C-scores, -1.10 and -1.01 respectively.

According to the results I-TASSER provided, CYP4B1 is the template structure that shows the highest structure similarity with both enzymes. CYP4B1 belongs to the

same CYP family, CYP4 as both Cyp4f14 and Cyp4f39. Although this has been in discussion for years, only in 2017 was the first crystal structure of CYP4B1 determined.[47] However, we carried out most of our study aligning our proteins with $P450_{BM-3}$ (1jpzb) (Figure 24). $P450_{BM-3}$ is much more extensively studied than CYP4B1. Serendipitously, $P450_{BM-3}$ (1jpzb) also showed the third highest and second highest TM and RMSD scores for Cyp4f14 and Cyp4f39, respectively. The TM (and RMSD score) reported for Cyp4f14 and Cyp4f39 are, 0.789 (2.72) and 0.782 (2.58), respectively.

Figure 24. Alignment of Models with Cyp4B1. Cyp4f14 model 1(left, C score 0.66, TM score = 0.896) and Cyp4f39 model 1 (right, C score 0.42, TM Score 0.885) with CYP4B1(5t6qA.pdb). CYP4B1 is colored in blue and Cyp4f39 (right) and Cyp4f14 (left) are in red. Alignment was directly imported from I-TASSER results and colored using PyMOL software.

P450s are known to oxidize a wide range of substrates. But interestingly, the structure of the protein is very much conserved among all P450s that have been

discovered and sequenced to date. This remarkable feature enabled us to compare Cyp4f14 and Cyp4f39 with the $P450_{BM-3}$ heme domain (PDB code 1jpz).[31] The table below (Table 1) compares the secondary structure positions of Cyp4f14 and Cyp4f39 determined by alignment with $P450_{BM-3.}$ Alignment was done using PyMOL software.

Since $P450_{BM-3}$ is a bacterial protein, it is cytosolic and not membrane bound. In contrast, all mammalian P450s are membrane bound. Most mammalian P450s are found in liver endoplasmic reticulum (ER), embedded in the phospholipid bilayer by a transmembrane anchor.[48] This transmembrane anchor was identified and named as A'-helix in both Cyp4f14 and Cyp4f39. All the other secondary structure features were identified by alignment with $P450_{BM-3.}$

A summary of the secondary structure elements presents in Cyp4f14 and Cyp4f39 is included in Table 1 with sequence positions. The images below show the I-TASSER generated models of Cyp4f14 (Figure 25) and Cyp4f39 (*Figure 26*) We also identified key amino acids in the active site using these alignments (Figure 27. and Figure 28).

Table 1: Secondary structure positions

	$P450_{BM-3}$[43]	Cyp4f14	Cyp4f39		$P450_{BM-3}$	Cyp4f14	Cyp4f39
α-helix				c	-	-	-
A'	-	2-46	36-57	d	163-167	206-217	215-224
A	25-37	71-81	78-92	f	381-385	442-451	450-457
B	54-62	103-113	110-120	β-sheet			
B'	72-83	123-134	130-141	1-1	38-44	87-93	95-99
C	93-110*	141-154	150-162	1-2	47-53	96-101	105-109
D	115-132	157-185	167-193	1-3	350-355	417-421	424-430
E	142-159	191-208	200-216	1-4	329-336	398-402	406-410
F	172-189	220-239	226-247	1-5	66-70	116-120	-
G	197-226	250-276	259-286	2-1	339-341	406-408	414-416
H	232-239	298-302	306-311	2-2	334-347	409-416	417-421
I	250-282	321-342	322-354	3-1	139-141	-	-
J	283-298	346-362	356-370	3-2	444-451	-	-
J'	304-310	372-377	379-385	3-3	420-424	-	-
K	311-325	379-391	387-398	4-1	432-436	-	-
K'	356-361	424-429	432-437	4-2	438-442	-	-
K"	363-368	-	-	5-1	240-242	-	-
L	402-419	472-488	478-495	5-2	246-249	-	-
3_{10}-helix							
a'	-	1	1-36	β-bulge			
a	16-21	47-73	57-79		393-396	461-466	467-474
b	-	-	-		399-400	467-468	475-476

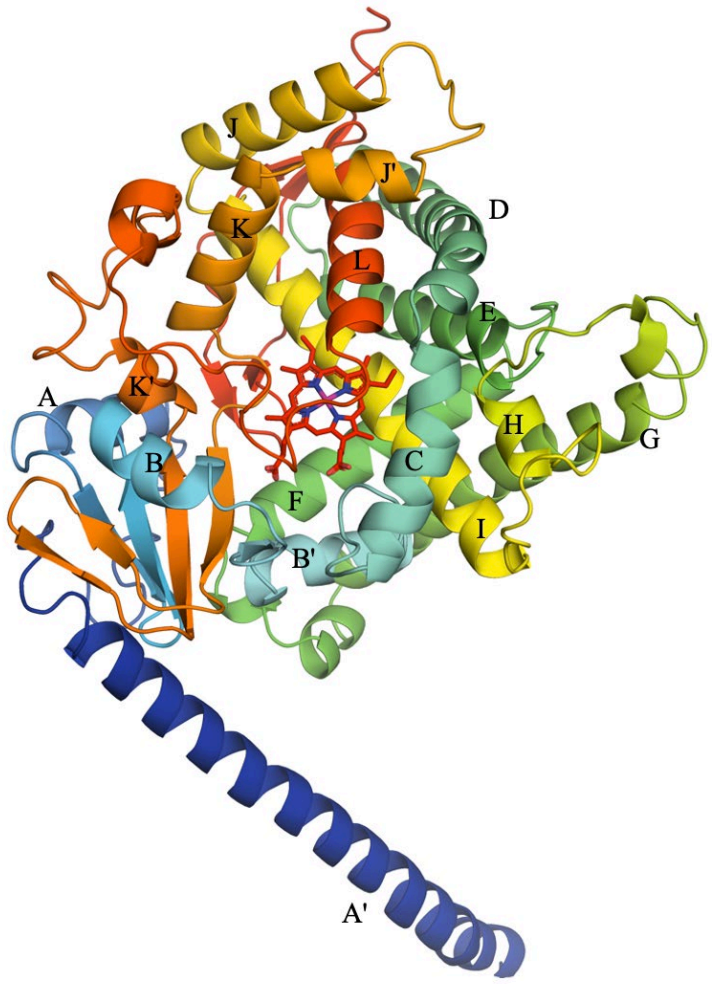


Figure 25. Structure of I-TASSER Generated Model of Mouse Leukotriene B4 ***ω***-Hydroxylase (Cyp4f14) as Viewed from the Distal Face. Secondary structural features are labeled in analogy to structure of P450BM-3 (1jpz.pdb). Secondary structural features are color-coded from N-terminal (blue) to C-terminal(red) using PyMOL software.

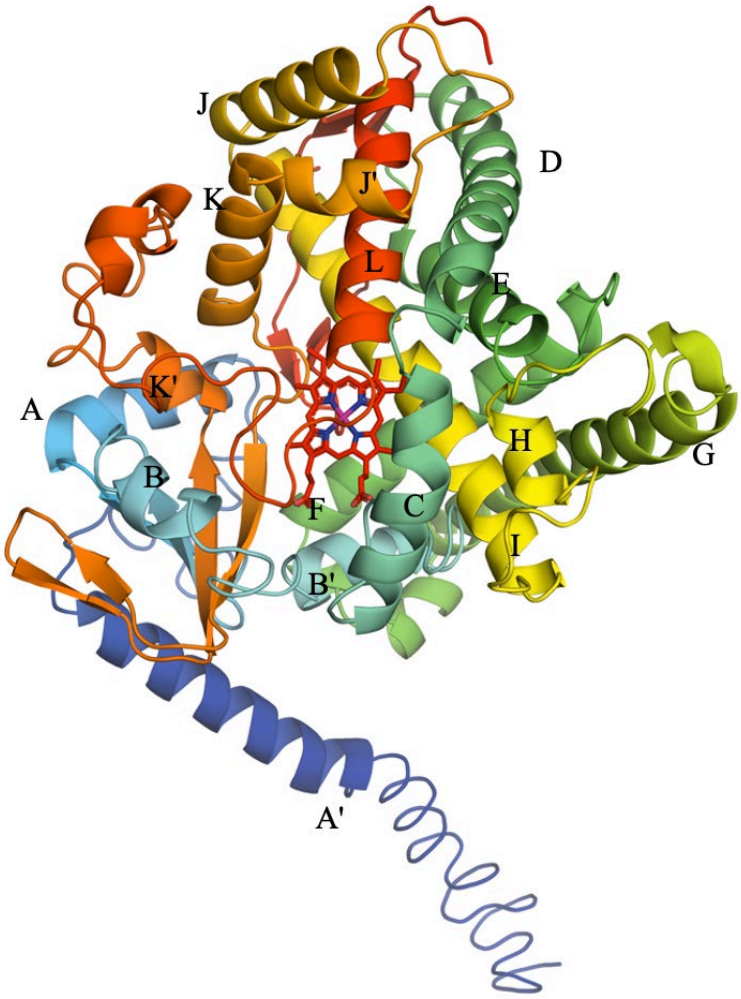


Figure 26. Structure of I-TASSER Generated Model of Mouse Epidermal Ceramide ω-Hydroxylase (Cyp4f39) as Viewed from the Distal Face. Secondary structural features are labeled in analogy to structure of P450BM-3 (1jpz.pdb). Secondary structural features are color-coded from N-terminal (blue) to C-terminal(red) using PyMOL software.

Active site of Cyp4f14

A comparison of the active site residues of Cyp4f14 and $P450_{BM-3}$ (1jpz) heme domains is presented below. These residues were identified using the PyMOL alignment tool. In $P450_{BM-3,}$ amino acids Arg47, Tyr51 and Gln73 are two residues that interact with the substrate's carboxylate acid head. In Cyp4f14, we identified these corresponding residues to be: Phe124, Ile123 and Thr100 for Gln73, Ala74 and Try51, respectively.

In Figure 27, Thr332 takes the top left position of the active site. Thr267 in $P450_{BM-3}$ forms an acid alcohol pair with Glu267. In Cyp4f14, the glutamic acid residue is replaced by an aspartic acid residue. This works as both Glu328 and Asp331 are negatively charged. However, as aspartic acid is one methylene group shorter, it can be speculated that this pair might not be close enough to the heme center to transfer the necessary proton. Both Thr332 and Asp331 are part of the I-helix.

The other important amino acid that seems to be missing in Cyp4f14 is $P450_{BM-3}$'s Lys391. However, in close vicinity there are two positive residues, His428 and Arg390, that might just provide a suitable place for reductase to attach and transfer electrons to the heme center. Finally, the highly conserved cysteine ligand is seen at the 468th position.

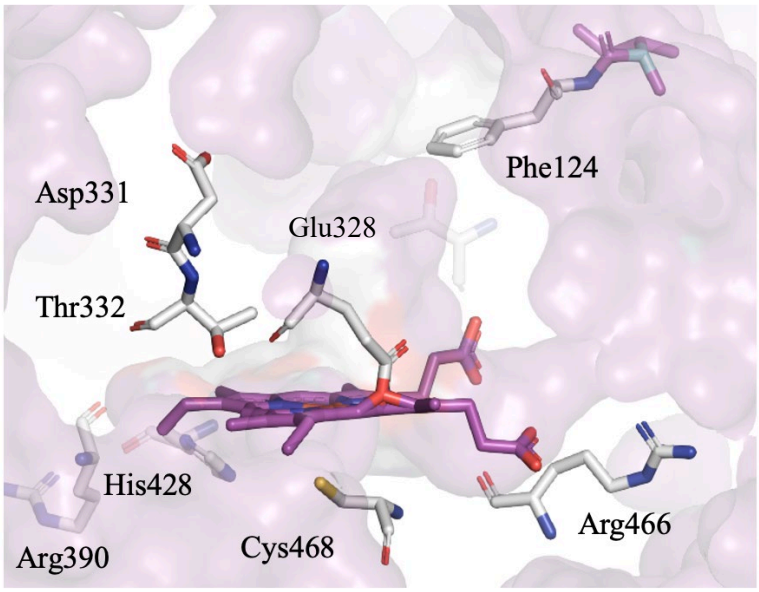


Figure 27. Predicted Active Site of Cyp4f14. Carbon atoms of heme porphyrin are colored in magenta. Carbon atoms of predicated amino acids in the active site are colored in white. Oxygen, nitrogen and sulfur atoms are colored in red, blue and yellow, respectively. The figure was prepared using PyMOL software.

Active site of Cyp4f39

Similar to Cyp4f14, we compared the active site residues of Cyp4f39 and $P450_{BM-3}$ (1jpz) heme domains. The figure below shows the predicted active site of Cyp4f39 (Figure 28). The amino acids identified as lining the active site were Lys128, Phe132, Ser134, Asp339, Thr340, Val467, Arg474 and Cys476. Gln73 of $P450_{BM-3}$ seems to be replaced by a serine residue. Both glutamine and serine are amino acids with polar side chains. However, glutamine is more polar because of the amide group and is also two methylene groups longer.

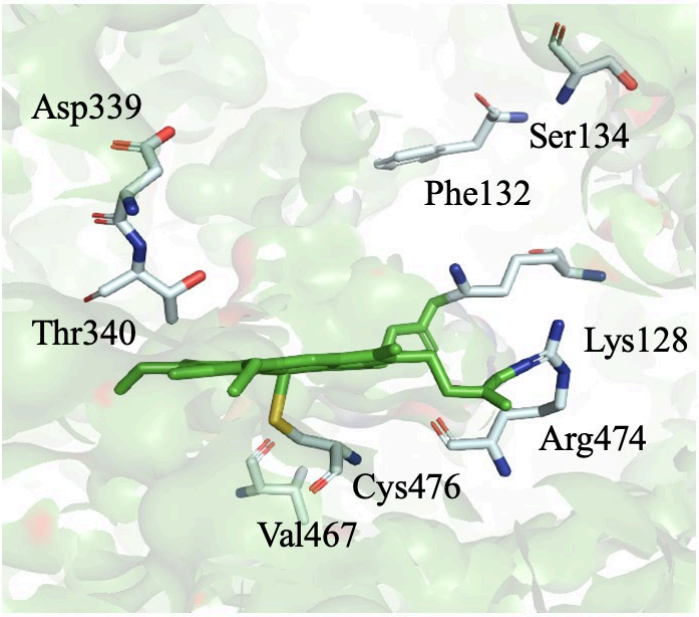


Figure 28. Predicted Active Site of Cyp4f39. Carbon atoms of heme porphyrin are colored in green. Carbon atoms of amino acids predicted to be in the active site are colored in light teal. Oxygen, nitrogen and sulfur atoms are colored in red, blue and yellow, respectively. The figure was prepared using PyMOL software.

Ester linkage between heme and glutamate in CYP4F family enzymes is an experimentally proven phenomenon with important implications for the regiospecificity of oxidation. The heme ion-oxo complex removes an electron from the glutamic acid carboxylate followed by extraction of hydrogen from C-5 position of the heme group. A methylcarbocation formed by intramolecular electron exchange within heme gets quenched by glutamic acid residue (Glu310 in CYP4B1) to form the ester linkage.[49] In the I-TASSER generated model of Cyp4f14, the ester linkage is not clear. However, Glu328 is located extremely close to heme and seems to make a close connection with the adjacent double bond to C-5 methyl group (Figure 29).

The distance between Glu-O and C-5 methyl group of heme was measured for all three species using PyMOL software. These distances appeared to be, 1.5 Å, 2.4 Å and 2.5 Å for CYP4B1, Cyp4f14 and Cyp4f39 respectively. This indicates that they are all close enough to make the Heme-Glu ester linkage.

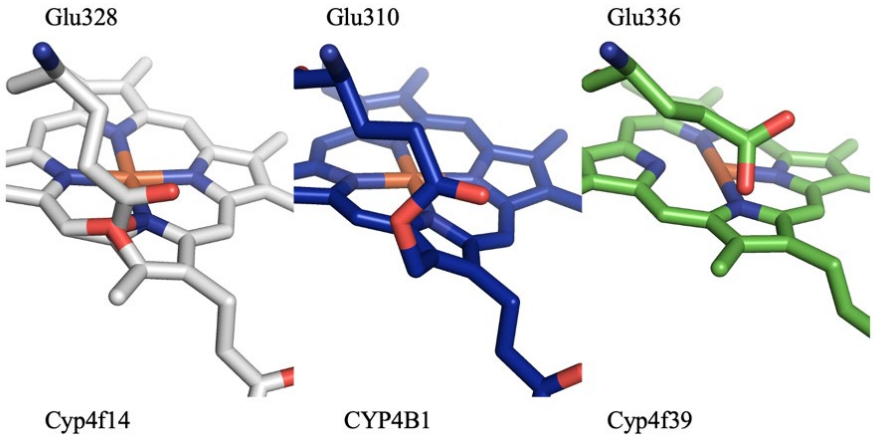


Figure 29. Heme-Glu Linkage of Cyp4f14, CYP4B1, Cyp4f39. Cyp4f14 (left, white) CYP4B1(middle, blue) Cyp4f39 (right, green). These images were generated using PyMOL software.

Molecular docking studies

Molecular docking studies with Cyp4f14 were performed with selected substrates to better understand the active site interactions and the structure of the enzymes. Docking studies predicted several orientations in which the ω−carbon was positioned close to the heme center.

The Autodock-modeled binding of LTB4 to Cyp4f14 reveals important details about the active site and substrate binding. The carboxylate group of LTB4 can be seen hydrogen bonding with Gln237 and Ala233 (Figure 30.). Also, favorable electrostatic interactions between His236 and the substrate could lead to increased binding affinity of the substrate to the active site. Further down from Gln237 and His236, there is Phe124 and Tyr125. Phe124 provides the required steric hindrance to curl the fatty acid chain. Lys120 seems to be making a strong ionic interaction with the heme porphyrin. Asp331-Thr332, present in the I-helix, resemble the Glu267-Thr268 pair in $P450_{BM-3}$.

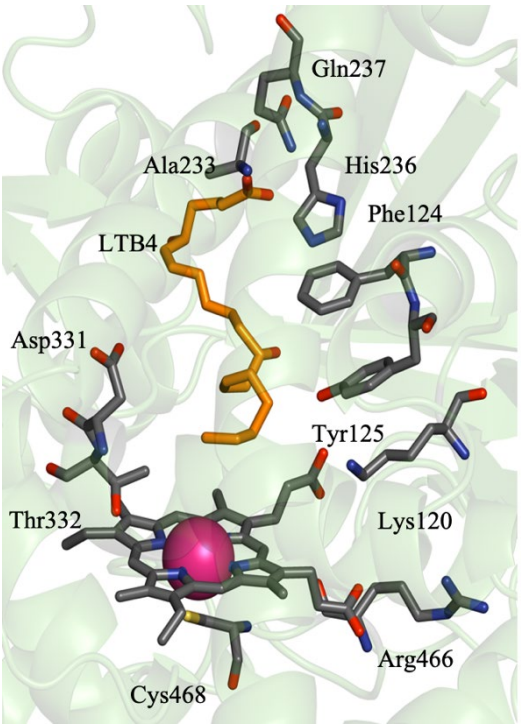


Figure 30. Molecular Docking of Substrate LTB4 to Cyp4f14 Performed via Autodock 4.2. Carbon atoms of interacting amino acids of Cyp4f14 are colored in grey and carbon atoms of the substrate are colored in orange. The green ribbons represent the secondary structure of the Cyp4f14. The figure was prepared using PyMOL software.

The C13-hydroxy group of LTB4 interacts with polar OH- group of Tyr125 in the B`-helix, which is likely significant as it would be expected to help fix the position of this point in the substrate chain and therefore control what carbons can come within reach of the heme iron for oxidation (Figure 31). The carboxylate head of LTB4 also H-bonds with Gln237 and makes electrostatic interactions

with Ala233. These interactions help bridge the F-helix (substrate access channel) and the B´-helix.

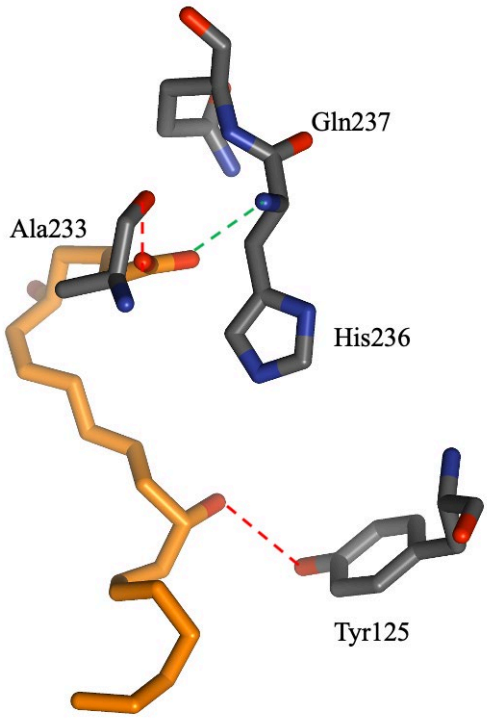


Figure 31. The Hydrogen Bond Network and Electrostatic Interactions in the Carboxylate Binding Pocket with LTB4. Carbon atoms of the amino acids and LTB4 are colored grey and orange respectively. Oxygen is red and nitrogen is blue. The figure was prepared using PyMOL software.

Apart from docking studies, we also performed several mutations on the macromolecule Cyp4f14 to see how they affect substrate binding. We anticipated that removal of bulky amino acids from the substrate access channel would allow the substrate to reach the heme center easily. Most of the mutations we performed are located in the I-helix or C-helix. We hypothesized that the mutations we made would help remove the steric hindrance in the substrate access channel and thereby allow substrate to reach heme center. Although most of these mutations are located outside the active channel, when performing Autodock calculations, we made sure they were located in the gridbox.

The first mutation we discuss here is W144A (Mutation A_1). In this mutation Tryrosine (Tyr144) is replaced by an alanine residue. This is located in the C-helix. To our surprise, there were no binding around the mutated site. However, this mutation reduced the distance between heme and the ω carbon, increasing the chances of ω hydroxylation.

The second mutation we discuss here is W144A/I123Q/T100Y (Mutation A_2) which is an extension of W144A. Here we introduced two new mutations, I123Q and T100Y, on top of W144A. When Cyp4f14 was aligned with $P450_{BM-3}$, Tyr51 and Gln73 of $P450_{BM-3}$ seemed to align with Thr100 and Ile123. In $P450_{BM-3}$ Tyr51 and Gln73 forms a hydrogen bonding network with the substrate carboxylic head. We wanted to see the effect of replacing isoleucine (Ile) with glutamine (Gln) and threonine(Thr) with tyrosin(Tyr). Ile123 is found in B′ helix.

Figure 32 shows a comparison of the binding of LTB4 to wildtype enzyme and mutated enzymes (mutation A and A2, respectively). It is evident that both mutations led

the substrate to approach the heme center in a more favorable conformation that would lead to the hydroxylation of the ω-carbon.

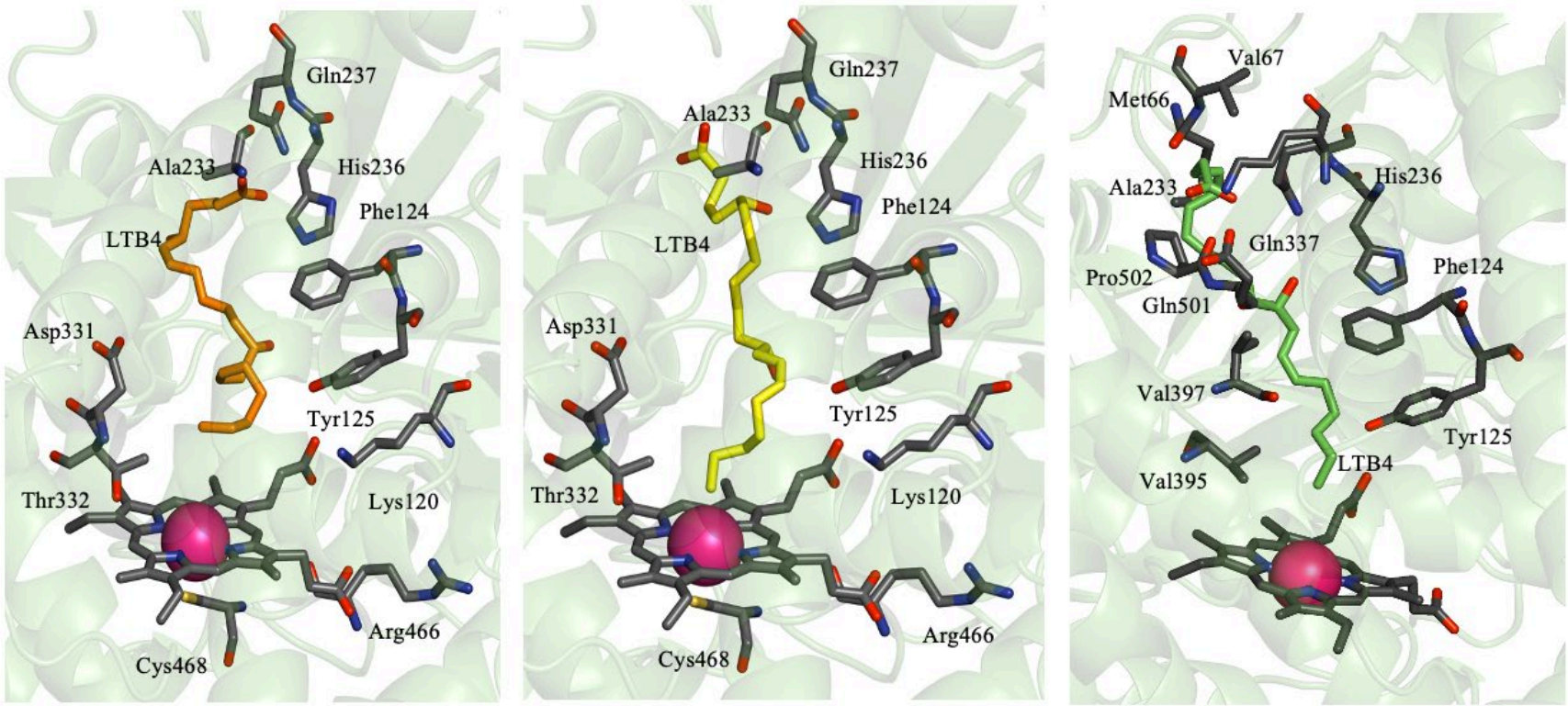


Figure 32. A Comparison of the Substrate Poses from Docking Studies Performed with Cyp4f14 and Leukotriene B4. Docking runs were performed before and after *in silico* mutations. The green background represents the secondary structure of the Cyp4f14. Carbon atoms of interacting amino acids of Cyp4f14 are colored in grey. Carbon atoms of the substrate are colored in yellow (A1) and green (A2). Oxygen is red and nitrogen is blue. The figure was prepared using PyMOL software.

Interactions between LTB4 and macromolecule amino acid residues are also worth discussing. In mutation A_1, strong hydrogen bonding interactions between C5 and C13 hydroxyl groups of LTB4 with Gln237 and Tyr125 are observed, respectively. Gln237 is a part of C-helix and Tyr125 belongs to B` helix. Favorable van der Waals interactions are also made through carboxylate head group of LTB4 with methionine and proline residues.

In mutation A_2 the C5-hydroxyl group of LTB4 is hydrogen bonding with histidine (His236) in the F-helix. (Figure 33). The carboxylate head of LTB4 can be seen making van der Waals interactions with methionine and the backbone carbonyl group of glutamine (Gln237). Proline 502 is seen to be making an unfavorable interaction with the C5-carboxylate group of LTB4.

In order to examine the regiospecificity of hydroxylation by wild type Cyp4f14 and its mutants, the distance from the heme center to the ω, ω_{-1} and ω_{-2} positions of LTB4 was calculated. Cyp4f14 A_1 displayed a dramatically enhanced regiospecificity towards ω-hydroxylation product (Table 2). It also showed enhanced selectivity towards ω-1 and ω-2 hydroxylation products.

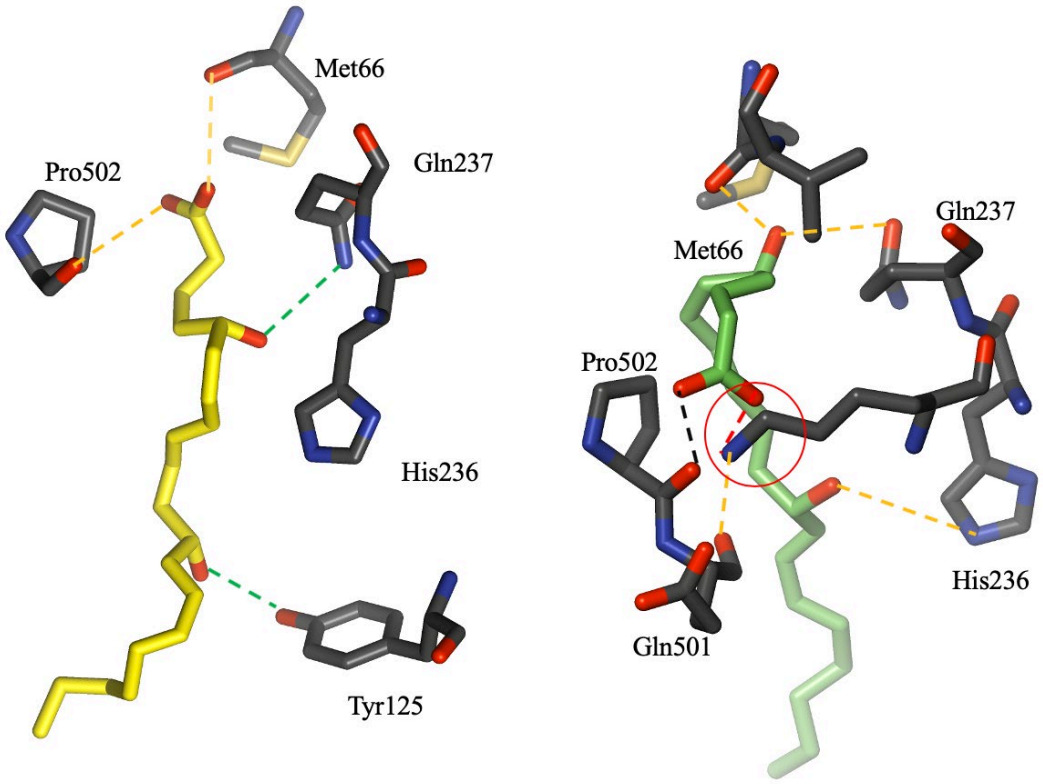


Figure 33. The Hydrogen Bond Network and Electrostatic Interactions in the Carboxylic Binding Pocket with LTB4. Carbon atoms of the amino acids are colored grey. Green dash lines represent hydrogen-bonding interactions. Orange dash lines represent polar interaction. Carbon atoms of the substrate are colored in yellow (A1) and green (A2). Oxygen is red, nitrogen is blue, and sulfur is yellow. The figure was prepared using PyMOL software.

Table 2: Distances in Angstroms from the substrate to heme of Cyp4f14 WT and its mutants

Position	Cyp4f14 WT	Cyp4f14 A_1	Cyp4f14 A_2
ω	5.6	2.7	4.9
ω_{-1}	5.9	4.0	6.2
ω_{-2}	5.8	4.7	7.8

We also used Autodock to simulate the binding of arachidonic acid (Figure 34). Arachidonic acid derivatives are known substrates of CYP4f subfamily enzymes. In arachidonic acid docking, the carboxylate head group is facing the distal end of the heme. In the proximal end ω_3 - ω_5 carbon atoms are directly above the iron center, indicating possible oxidation sites. Phe124, Thr125 and Phe327 are close to the fatty acid tail of arachidonic acid. Their possible role could be to curl the fatty acid chain exposing ω_1-ω_{13} positions to be oxidized. Thr332-Asp331 can be found in the vicinity of the docking site. At the distal end, the carboxylate head is contained in a pocket containing mostly positively charged amino acids.

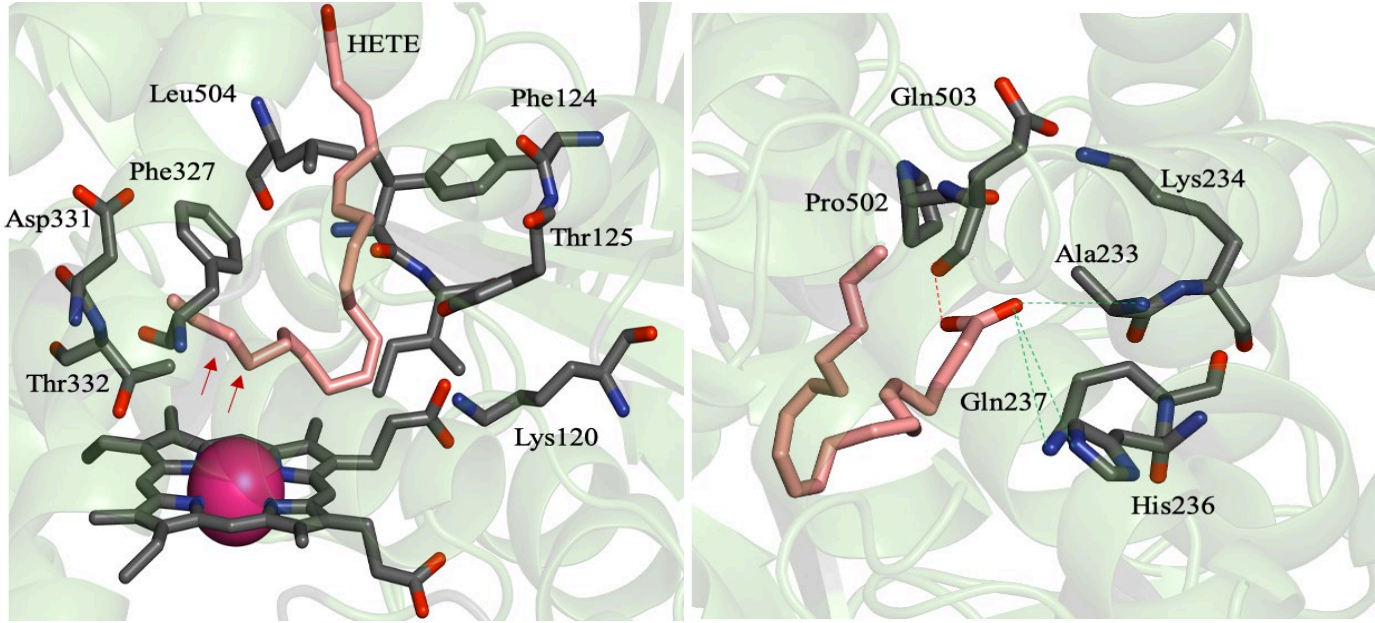


Figure 34. Most Stable Substrate Conformation from Docking Studies Performed with Cyp4f14 and Arachidonic Acid (Substrate). Red arrows are pointing at the expected oxidation positions (Top). Red dotted lines are used to represent the electrostatic interactions between carboxylic head of the substrate and interacting amino acids. Green dotted lines are used to represent H bonding interactions (Bottom) The green background represents the secondary structure of the Cyp4f14. Carbon atoms of interacting amino acids of Cyp4f114 are colored in grey and carbon atoms of the substrate are colored in pink. Oxygen is red and nitrogen is blue. The figure was prepared using PyMOL software.

*p*NCA has been previously used with $P450_{BM-3}$ as an assay system to detect active enzyme in a purified sample and cell lysate.[29] However, to date, it has not been used as a substrate for CYP4F family enzymes. In this study we have used an 11-*p*NCA oxidation assay to check for the presence of active enzyme in the cell lysate. However, our experimental results from the 11-*p*NCA assay show almost no expression. Hence, to follow up, docking studies were performed using 8-*p*NCA and 11-*p*NCA as substrates. *p*NCA is not a natural P450 substrate and a lack of studies on the *p*NCA assay and its ability to work well with the crude enzyme also motivated us to carry out docking studies.

Docking studies performed with 8-*p*NCA are reported below (Figure 35.). As evident in the docking image, the oxygen of the ortho nitro group binds directly with the iron center. In order to get 8-hydroxy-8-4-nitrophenoxyoctanoate, C-8 of the carboxylate chain need to be exposed to heme center. But instead, the heme center interacts with the nitro group (oxygen). As reported by the Schwaneberg team, the $P450_{BM-3}$ wildtype ω-hydroxylated less than 1% of the 8-*p*NCA starting material in their assay.[27] This is parallel with our initial docking studies with 8-*p*NCA. Binding via nitro group (oxygen) will not lead to the formation of desired chromophore, *p*-nitrophenolate. Thus, this can be the reason for poor activity of the *p*NCA assay.

the Schwaneberg team later reported that they were able to increase the catalytical activity of $P450_{BM-3}$ by introducing series of mutations.[28] We also introduced F124A and F327A mutations to see if removing some of the steric bulk in the active site would result in exposing C-8 to the heme center, but they resulted in similar conformations.

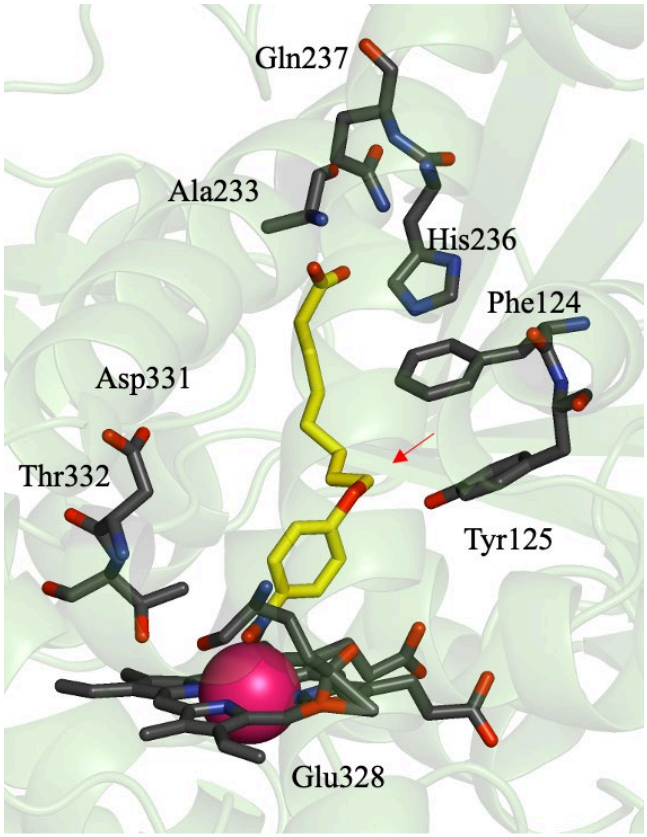


Figure 35. Most Stable Substrate Conformation from Docking Studies Performed with Cyp4f14 and 8- *p*NCA (Substrate). A red arrow points at the C-8 carbon, the anticipated oxidation site for chromophore production. The green background represents the secondary structure of the Cyp4f14. Carbon atoms of interacting amino acids of Cyp4f14 are colored in grey and carbon atoms of the substrate are colored in yellow. Oxygen is red and nitrogen is blue. The figure was prepared using PyMOL software.

$P450_{BM-3}$ wildtype ω-hydroxylated product percentage for 11-*p*NCA is reported to be around 68%.[27] We performed docking studies with 11-*p*NCA. Here we report the closest docking conformation that could possible give rise to *p*NCA chromophore (Figure 36).

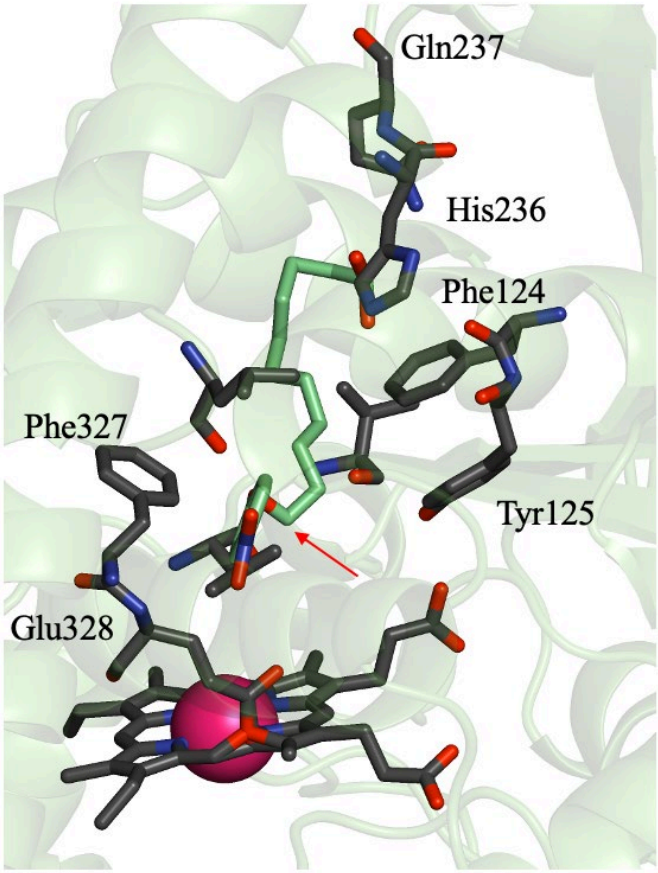


Figure 36. Most Stable Substrate Conformation from Docking Studies Performed with Cyp4f14 and 11- *p*NCA (Substrate). The red arrow is pointing at the expected oxidation site. The green background represents the secondary structure of the Cyp4f14. Carbon atoms of interacting amino acids of Cyp4f114 are colored in grey and carbon atoms of the substrate are colored in green. Oxygen is red and nitrogen is blue. The figure was prepared using PyMOL software.

Long chain ceramides serve as the substrates for Cyp4f39. However, we were not able to perform a successful Autodock run for this molecule even after trying various mutations of the enzyme, different number of torsions (in the ligand) and different sizes

of grid boxes. Normal substrates for P450s are around 12-20 carbon long, but the fatty acid chain of the ceramide substrates are 30-34-carbon-long fatty acids. Binding may require long hydrophobic acyl chain binding to the surface amino acids, which would require reconfigured AutoDock grids. This is something the Haines lab intends to study in the future.

Docking performed with arachidonic acid and Cyp4f39 shows evidence of initial binding to the F-helix and E-helix (Figure 37). The carboxylate of the substrate interacts with Arg516, Lys510, Glu512 and Ile235. Tyr246 seems to hydrogen bond with the carboxylate.

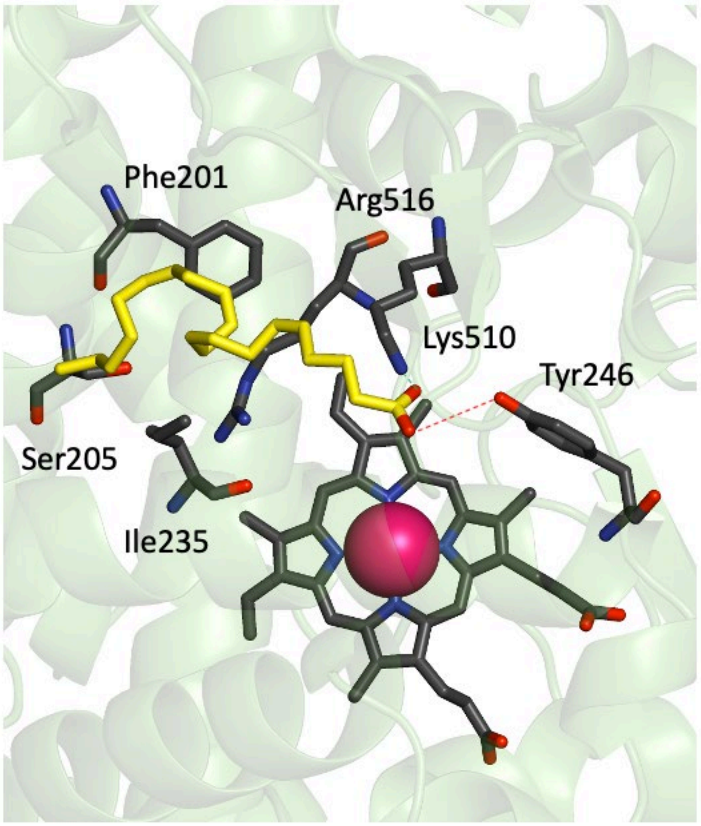


Figure 37. Interactions of Arachidonic Acid with Macromolecule Cyp4f39. The green background represents the secondary structure of the Cyp4f39. The red dotted lines indicates hydrogen bonding between Tyr246 and the substrate carboxylate head. Carbon atoms of interacting amino acids are colored in grey and carbon atoms of the substrate are colored in yellow. Oxygen is red and nitrogen is blue. Figure was prepared using PyMOL software.

We did not compare the binding energies of the substrates for a couple of reasons. Firstly, all of the binding energies from Autodock have a large uncertainty of around 2.5 kcal/mol (for discussion see, for example, https://youtu.be/_Lhjlx3kJxo?t=455).

Secondly, all of these binding calculations consider the macromolecule as a rigid structure. Accordingly, the reported binding energies are unreliable.

Also, the Babel-based method Autodock used to prepare the coordination files does not handle metal charges. Thus, the resulting PDBQT file will always have zero for the charge of iron. This issue was addressed manually by editing the PDBQT file and adding 1.130 as the charge of heme.[50] Since the actual formal charge of +3 may be too strong for the current force field parameterization, 1.130 was selected based on literature.[50]

Conclusion

In conclusion, we studied two mouse CYP4 enzymes, Cyp4f14 and Cyp4f39, in order to to develop an *E. coli* expression system that can produce high-level expression of protein to support further characterization of these important enzymes. We also carried out *in silico* structural analyses in order to have a better understanding of the structure and the active site of these proteins.

Cyp4f14 or leukotriene B4 ω-hydroxylase is known to suppress tumor progression through its metabolism of the pro-inflammatory mediator leukotriene B4 (LTB4), which is also involved in pathogenesis of neurodegenerative diseases such as Alzheimer's disease. Based on spectroscopic studies and the *p*NCA oxidation assay, induction at an OD_{600} of 0.6 with 0.5- 0.8 mM IPTG seems to be optimal for expression in the presence of NADPH.

Structural models and docking studies predicted important amino acids that are important for substrate binding. In Cy4f14, these amino acids were identified as Asp331, Thr332, Phe124, Tyr125, Cys468, Lys120 and Arg466 as these residues align with residues in $P450_{BM-3}$ that are known to have important roles in proton transfer, binding to the heme center, and directing water into the active site.

From the mutations we performed, mutation A_1 (W144A) showed the best results for ω oxidations of LTB4. It is hoped that these modelling will provide a good foundation for future research on this enzyme.

The second enzyme we studied, human Cyp4f39, the human ortholog of CYP4F22. CYP4F22 is an epidermal ω-hydroxylase involved in the synthesis of acylceramide which is an essential lipid for skin barrier formation. Reduced enzymatic

activity of CYP4F22 caused by mutations is correlated to autosomal recessive congenital ichthyosis (ARCI), which is a rare group of keratinization disorders characterized by abnormal scaling of the skin. Our study revealed important structural features of Cyp4f39 including active site amino acids. We were also able to observe the initial binding site of arachidonic acid to the E and F-helix amino acids Phe201, Tyr 246, Ser205 and Ile235.

To improve the expression of Cyp4f14, these studies can be performed in yeast cells or sf9 insect cells. We can also recommend a truncated structure of Cyp4f14 by removing the membrane anchor could be a good way to extend expression studies in *E. coli* based on prior experience.

To improve Cyp4f39, docking studies and identify active site residues that interact with substrate long chain ceramides, a series of molecular mechanics (MM) can be used. If succeeful, more accurate and advanced quantum mechanism and molecular mechanics (QM/MM) can be used to model the enzyme. QM/MM methods can produce more accurate parameters for heme cofactors. Similarly, Cyp4f14 can be studied using MM and QM/MM methods to confirm the results obtained by Autodock in this study.

CHAPTER II

Analysis of odor leakage from Human Remains Containment Units for use in space

Introduction

NASA has been sending humans to space since 1965. So far, thankfully, NASA has not had to deal with the event of a death in space. While this remains a possibility, NASA's focus in the past has been preventing deaths in space rather than providing specialized equipment for managing a corpse. Options for dealing with a corpse in space are limited. Exposing a corpse to the shared atmosphere inside a space station or spacecraft would be predicted to reduce morale and present severe psychological strain, present a biohazard, and reduce air quality. UN guidelines prevent corpses from being released into space.[51] Return of a corpse to earth in a timely manner tends to be costly and resource prohibitive. With the increase in interest of longer-term missions to the Moon and Mars, NASA is revisiting this issue by looking into the viability of body bag technology for use in space stations and future off-earth colonies.

The purpose of this study is to test the ability of bags being tested by NASA (referred to herein as Human Remains Containment Units (HRCU A and B) containing non-refrigerated cadavers to contain volatiles over time. HRCU A is a bag designed to contain remains and hazardous chemical agents that might have contributed to death, including any gases that might accumulate at lower pressures. HRCU B is a bag specialized to contain hazardous biological matter, in this case, human remains, body fluids and microbes or other biological agents of concern that might have contributed to death such as infectious agents. Both bags have carbon-filtered pressure vents to prevent build-up of excess pressure.

Decomposing body

The analysis of volatile organic compounds (VOCs) that are produced from cadavers during decomposition is an important tool for forensic investigations.[52] Its uses include locating human remains and identifying their stage of decomposition. For this project, VOCs are defined as any emissions released from a decomposing body at interior International Space Station temperatures (21 -24°C) that can be detected by gas chromatography coupled with mass spectrometry. Of particular interest are odorants that would be detected by humans.

In the process of decomposition, larger biological macromolecules such as carbohydrates, proteins, and nucleic acids are broken down into their monomers.[53] Internal decomposition accelerates relative to external decomposition by the unchecked internal enzymes of the cadavers, leading to inside-out cell lysis. This process is called autolysis. As a result of this, volatile organic compounds are produced as byproducts. Autolysis also provides ample amounts of nutrients for microbial growth, causing further VOC release.[52]

Solid Phase Micro Extraction (SPME)

SPME is an absorption/desorption technique that allows extractions to be performed in a quick and easy way. The SPME technique does not require noxious organic solvents. A SPME sampling device is portable and reusable.

There are two types of SPME fibers: automated fibers and manual fibers. The fiber core is 1-2 cm long and coated with the stationary phase. The fiber can be either

immersed in a liquid sample or placed above the sample (Headspace SPME) during sample collection.[54,55] The latter approach is more expedient as it protects the fiber from damage caused by external samples which allow the fiber to be used in more than 100 extractions.

A SPME device is composed of a holder and the fiber assembly. The fiber core is contained in a hollow needle where it is attached to a stainless-steel guiding rod. The fiber is contained within a hollowed needle that is used to pierce a septum. The basic principle behind SPME is separation of analytes between the sample matrix and the extraction medium. Commercially available SPME fibers come with different extraction media/fiber coatings depending on the class of compounds to be analyzed. Among these, polydimethylsiloxane (PDMS), polydimethylsiloxane/divinylbenzene (PDMS/DVB) and carboxene/PDMS are widely used in headspace sample analysis. However, the fibers are very fragile and therefore easily breakable. However, stableflex or metal alloy dispersed SPME fibers with fused silica cores make the inner fiber less breakable.

Below is a schematic diagram of a SPME fiber and the GC-MS Setup (Figure 38).

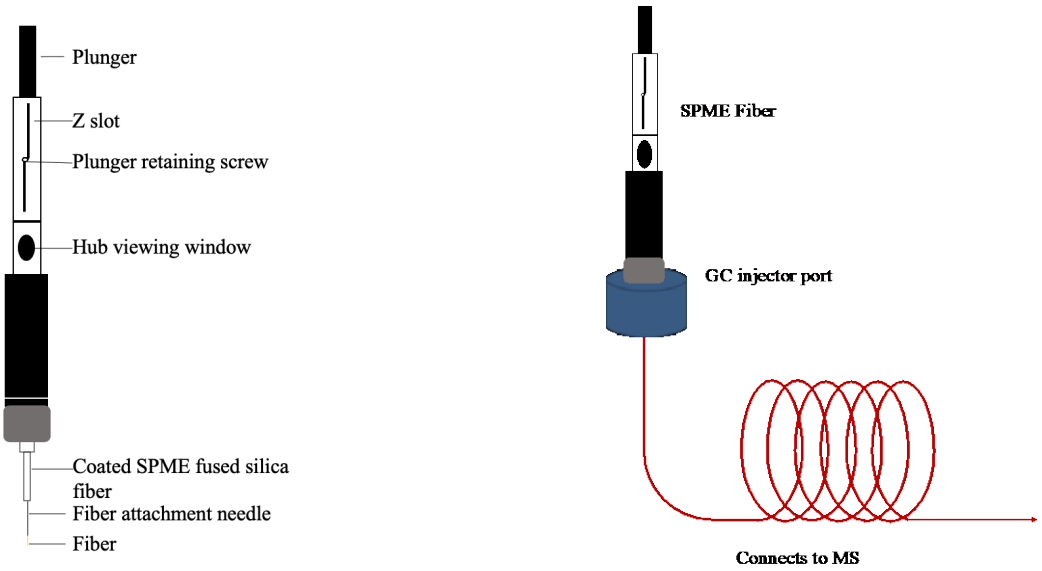


Figure 38. Schematic of SPME Fiber and SPME-GC-MS Setup.

Methods

Human cadaver sampling and analysis

Two cadavers (CADs A and B) were provided by NASA, the Southeast Texas Applied Forensic Science Facility (STAFS) and Sam Houston State University, Huntsville, Texas. Sampling and analyses were part of a broader project involving other simultaneous measurements by other research teams. The studies reported here were a last-minute addition to the broader project. The two cadavers were in the fresh stage of decomposition, were male, and fell in the 95-percentile range for height and weight.

Environmental conditions were maintained to mimic those of the International Space Station. Environmental conditions included a room temperature range of 21 – 24 °C, humidity range of 40 – 50%, and no insect access. The study began September 25, 2019 at 7:30 am, referenced as Day 1. Both cadavers (the broader project included a third cadaver that was not part of our studies and will not be discussed further here) were unclothed and laying supine on gurneys. Each cadaver was placed in an open HRCU and forensically sampled by NASA personnel. The HRCU's were then closed and the cadavers were put in a prep lab at AARC for the duration of the decomposition portion of the study.

The two HRCU bags, bag A and bag B, received cadaver A (CAD A) and cadaver B (CAD B), respectively. CAD A received an abdominal puncture with a tube inserted into the abdomen. This was then secured with medical tape. Both Cadaver A (CAD A) and cadaver B (CAD B) were placed in a supine position until Day 9 of the study. On Day 9, two NASA personnel repositioned CAD B into a loosely flexed position, secured

with ‘spider’ straps. The two HRCUs were not opened until the end of the experiment (Figure 39).

Figure 39. CAD B in the Loosely Flexed Position Secured with ‘Spider’ Straps.

Solid phase microextraction (SPME)

PDMS/DVB SPME fibers (65 µm) were used to sample the volatiles that could potentially leak from the HRCU. These microextraction fibers were placed on CADs A & B at two different positions, one just outside of the zipper closure and the other at the carbon-filtered ventilation port on each bag. Fibers were clamped to a stand and placed near each zipper and port. A control fiber was placed at the opposite end of the room using a similar set up. Fibers were collected every two days during the first phase of the study. In the second phase, fibers were collected once a week.

GC-MS analysis of VOCs

The VOCs were analyzed using an Agilent Technologies 7890A series gas chromatograph (GC) coupled to a 5975C series mass selective detector (MSD), located at the Texas Research Institute for Environmental Studies (TRIES) as Sam Houston State University (Figure 40). Data analyses was done using a combination of ChemStation on the instrument and Python programming language with custom code for detailed analysis (APPENDIX E).

The GC was equipped with an Agilent J&W DB-5ms capillary column with 30 m x 250 µm dimensions and 0.25 µm coating thickness. A SPME injection port liner was used and the GC injector was operated in splitless mode at 250 °C. Hydrogen carrier gas was set to a flow rate of 1.1873 mL/min and the initial temperature for the GC oven was at 30 °C for 2 minute, then ramped to 80 °C at 6 °C per minute, then to 120 °C at 15 °C per minute, and finally to 300 °C at 40 °C per minute and held at 300 °C for 2 minutes. The total time for the run was 19.5 minutes and the MS scan range was 20 to 300 m/z at the rate of 2.94 scans per second.

Figure 40. Full GC-MS Setup (Left) and SPME after Fiber Injection in the Inlet Port (Right).

Results

Two SPME fibers were used at the port and zipper of each cadaver, and data were collected every two days in the first phase of the experiment (Figure 41 and Figure 42). In the second phase, data were collected once a week.

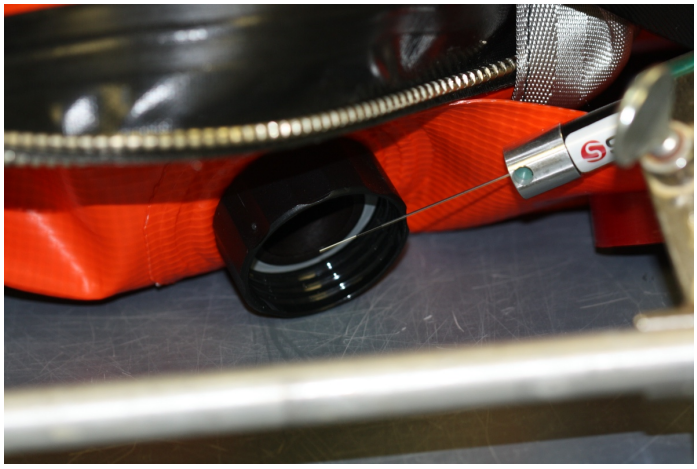


Figure 41. SPME Fiber at Port.

Figure 42. Zipper SPME Fiber Position. Inset shows expanded view of fiber positioning for clarity.

The decomposition process involves the breakdown of a cadaver's macromolecules into monomers and other simple compounds. Previous studies have reported the decompositional release of over 400 volatile or semi-volatile compounds.[56] One group of decomposition processes, proteolysis, results in the breakdown of protein into their amino acid components by bacterial enzymes.[56,57] These amino acids are further broken down into even simpler compounds through deamination, decarboxylation and desulfhydration.[53,56] Foul smelling compounds containing sulfur result from desulfhydration. These include hydrogen sulfide and CH_3SH that are produced in very early stages of decomposition.

Our goal was to test the ability of the NASA provided bags (HRCU A and HRCU B) to contain volatiles from non-refrigerated cadavers over time. Decompositional volatiles causing human odor sensations, such as H_2S and CH_3SH were of particular interest to us. We hypothesized that if either bag failed, we would be able detect changes in the peak patterns observed from by GC-MS analyses. We expected to detect H_2S and CH_3SH along with short chain aldehydes such as nonanal, decanal, etc. These small molecules are commonly detected from decomposition byproducts. Being small and volatile, H_2S and CH_3SH are more easily detected by SPME than larger compounds. Being compounds that form in early stages of decomposition, they[57] would also allow us to identify bag failure early.

Conversely, the challenge of identifying H_2S is that it is a small molecule that comes completely in the void fraction in normal gas chromatography. Interestingly, as we continued with our experiment, the intensity of the void peak started to increase. This indicated that fibers were picking up new volatiles that were not present during early

sampling events. We compared our observations with the NIST (National Institute of Standards and Technology) database and confirmed that what we were observing was consistent with H_2S and CH_3SH (Figure 45).

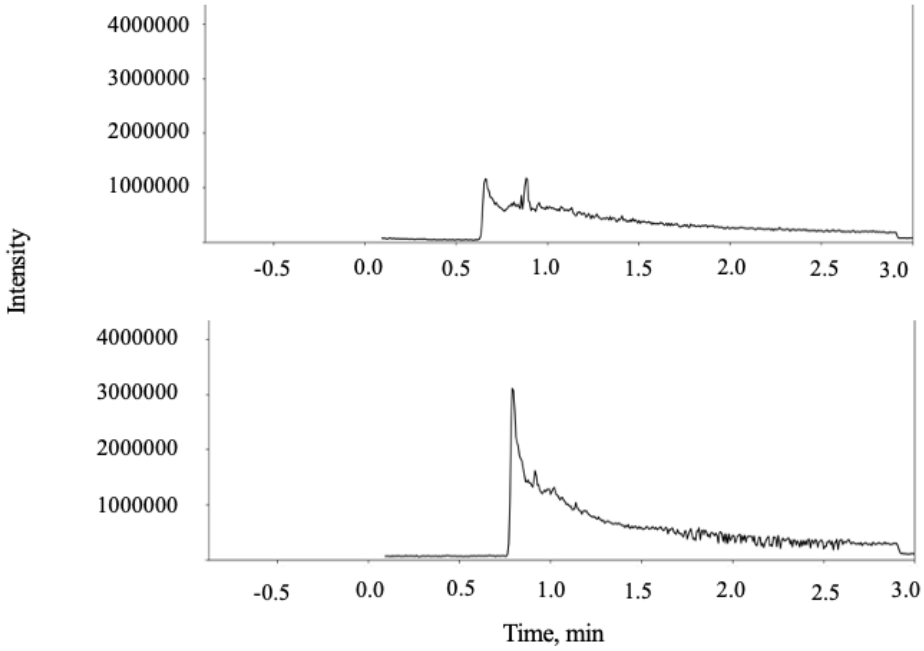


Figure 43. Comparison of Total Ion Current at the Void Peak Over Time vs Intensity of Zipper B samples. Top to bottom correspond to zipper B sample on day 1 and 8, respectively. Y-axis scale is from 0-4000000 with 1000000 data intervals. X-axis scale goes from 0.0 to 3.0 min. Data visualization was done using Python.

Below is a detailed analysis of the void peak (Figure 44). Here we show a detailed comparison of ions of interest for H_2S and CH_3SH that come out in the void region. It is evident in the figure below that around 1 minute (in the void region) major amounts of H_2S and CH_3SH can be seen only in the later samples.

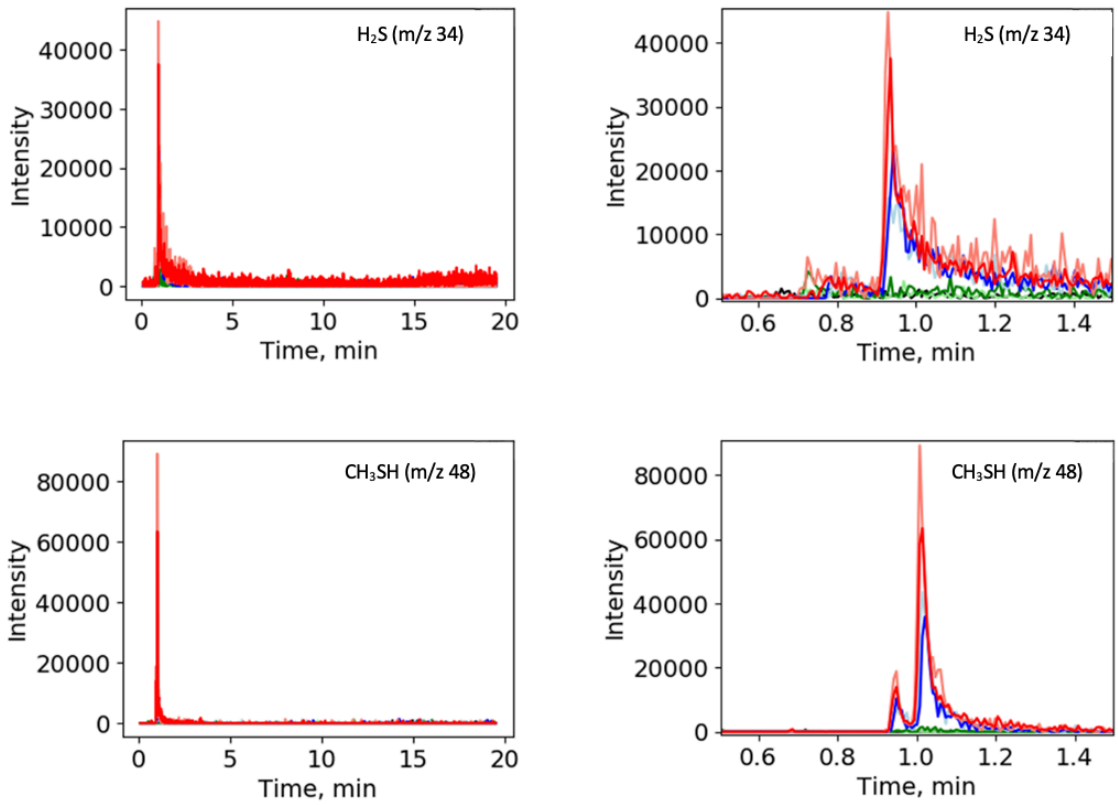


Figure 44. Comparison of Specific Ions 34 and 48 m/z Corresponding to H_2S and CH_3SH, Respectively. Left side shows the extracted ion chromatogram of specific ions 34 and 48 m/z corresponding to H_2S and CH_3SH. The right side is the zoomed in alignment of H_2S (top) and CH_3SH (bottom) around 1 minute. Colors black, pale green, green, pale blue, blue, pale red, red represent data obtained from zipper B sample on day 1, 8, 11, 13, 15, 17, and 19, respectively.

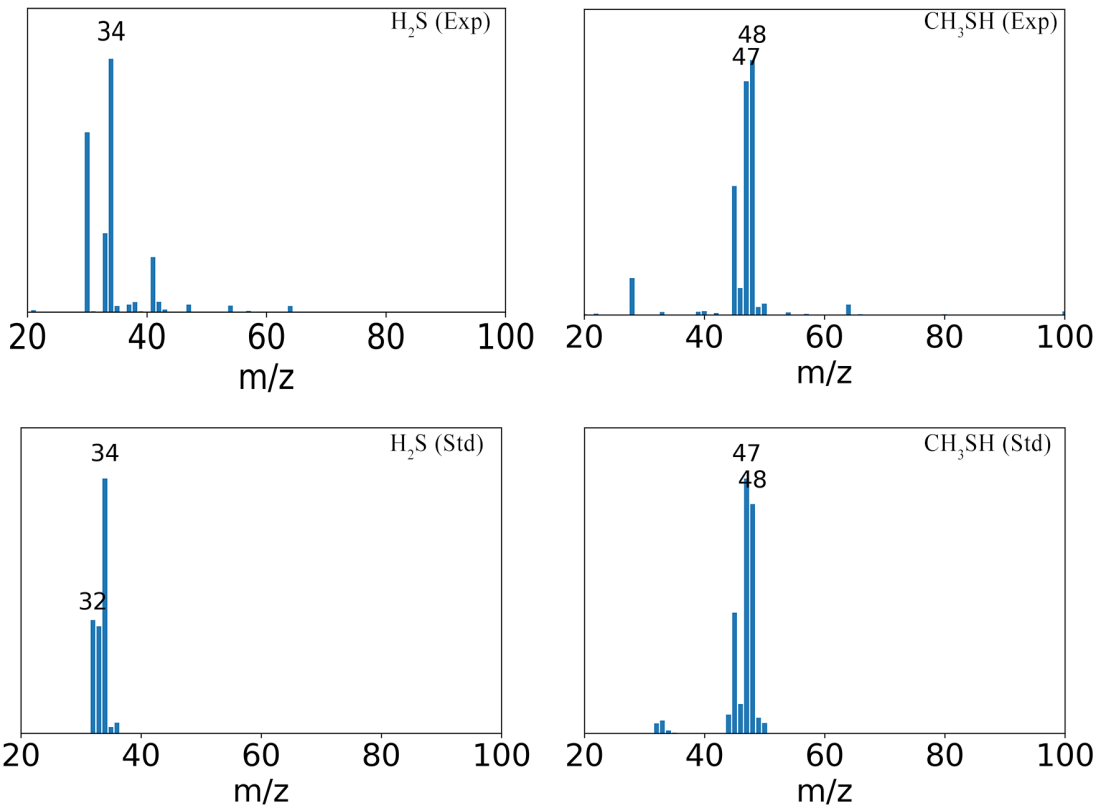


Figure 45. A Comparison of NIST Database Reference Spectra and Observed Spectra of H_2S and CH_3SH. The experimental spectrum of H_2S is top left and the bottom left spectrum right below shows the NIST database reference spectra (bottom). The spectrum at top right represents the observed spectrum of CH_3SH. and the spectrum right below at bottom right shows the NIST database reference spectra. Data visualization was done using Python.

We started detecting both H_2S and CH_3SH from day 8 onwards. Simultaneously, odor was also noticed in the room that increased with time. From day 11 to 23, the CAD B (Port-B and Zipper-B) fiber sensors seemed to detect considerably higher amounts of

H_2S and CH_3SH than the control and CAD A. This indicated that HRCU B was the major source of odor. However, for HRCU A the zipper fibers also seemed to pick up low-level signals of these volatiles. This could be because the zipper A fiber is picking signals from HRCU B, which was in closer proximity to CAD B. As noted previously, odor permeated the room through parts of the experiment. On day 21, the odor in the room was extremely high and it recorded the highest amount of volatiles so far for HRCU B. At this point we (the NASA Team) had no interest in continuing the experiment with HRCU B because of the odor in the room and the increased VOCs detected by SPME-GC-MS both indicated the failure of HRCU B.

Odors started to accumulate around the Zipper A fiber from days 10-12. After day 11, elevated levels of H_2S were detected from the control fiber. This is due to the accumulation of H_2S and CH_3SH in the mostly sealed room (spaces around external doors and windows were sealed with tape). After day 15, we had to discontinue the control fiber; as the Zipper A fiber broke during the course of the experiment and the control was sacrificed to replace the Zipper A fiber. Thus, the discontinuity observed in data of Zipper A fiber on days 13-17 is due to the broken fiber. Data observed for Zipper A on day 15 is unknown and no data observed by the control from ays 15 to 21 due to those issues (and Figure 47).

Similar to HRCU B, on day 53 we stopped the experiment with HRCU A as it started to show a similar spike in VOC detection in the GC-MS analysis.

Below figures show a summary of the total H_2S emission from HRCU A (Figure 46) and HRCU B (Figure 47) and a summary of total CH_3SH emission from HRCU A

and HRCU B. The discontinuity observed in control and Zipper A fibers resulted from broken fibers as described above.

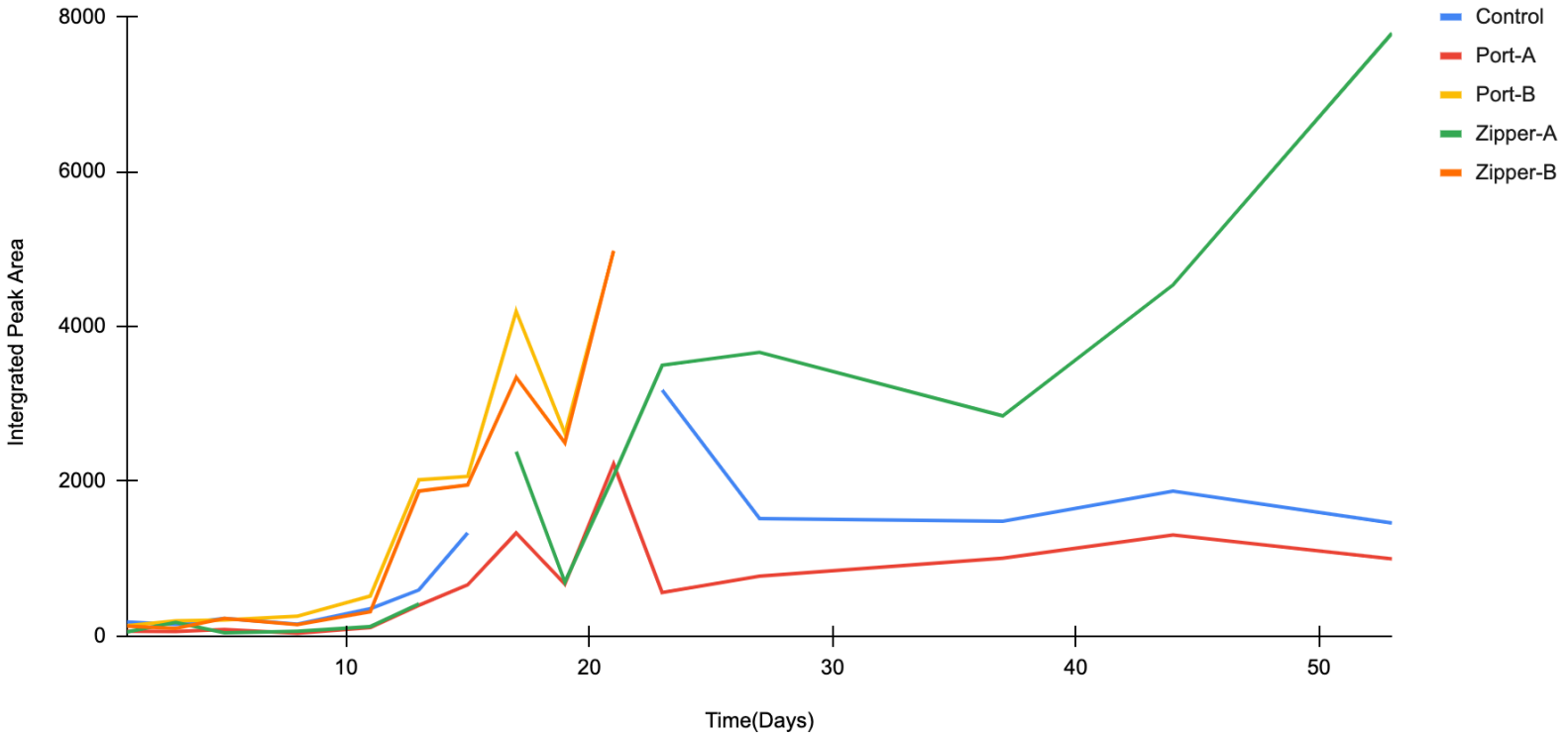


Figure 46. Emission of H_2S Versus Experiment Day. Summary of total H_2S emission from HRCU A, HRCU B and control fiber from day 1-53. Port A and Zipper A fibers are represented in red and green colors respectively. Port B and Zipper B are represented in yellow and orange colors, respectively. Control is blue.

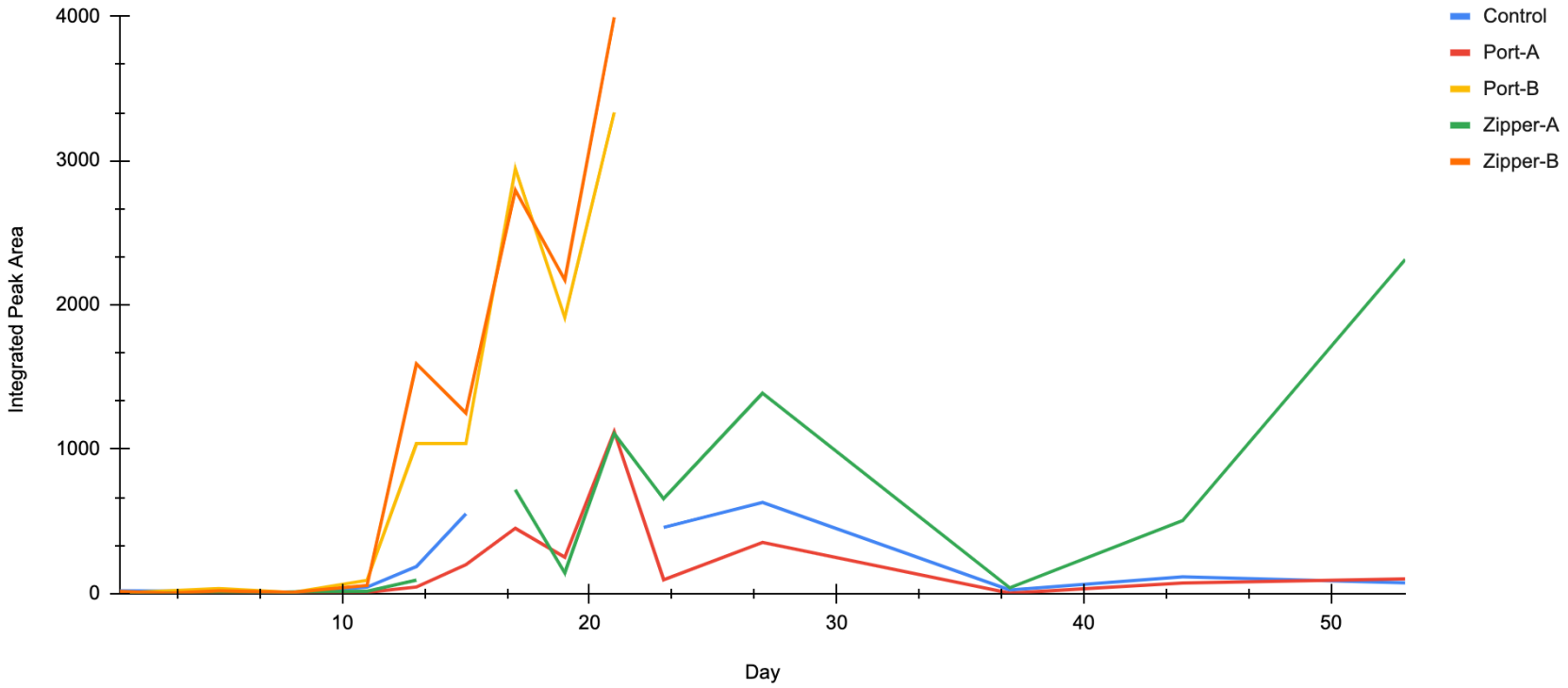


Figure 47. Emission of CH_3SH Versus Experiment Day. Summary of total CH_3SH emission from HRCU A, HRCU B and control fiber from day 1-53. Port A and Zipper A fibers are represented in red and green colors respectively. Port B and Zipper B are represented in yellow and orange colors, respectively. Control is blue.

Apart from H_2S and CH_3SH, we had an interesting discovery on day 11, as the fibers picked up some unusual signals. By comparison with NIST database, we confirmed that these signals were due to N,N-diethyl-meta-toluamide (DEET) (Figure 48, APPENDIX F). DEET is a common insect repellent. The source of DEET is still unclear. However, it was detected about the time that spider straps were brought in by NASA personnel to place CAD B in a seated position. We learned that the spider straps had been used previously in an outdoor camping setting. This may have made them a source of DEET. However, re-examination of the data showed that DEET was present in low quantities even before they were brought in. The source or sources of the DEET are still unknown

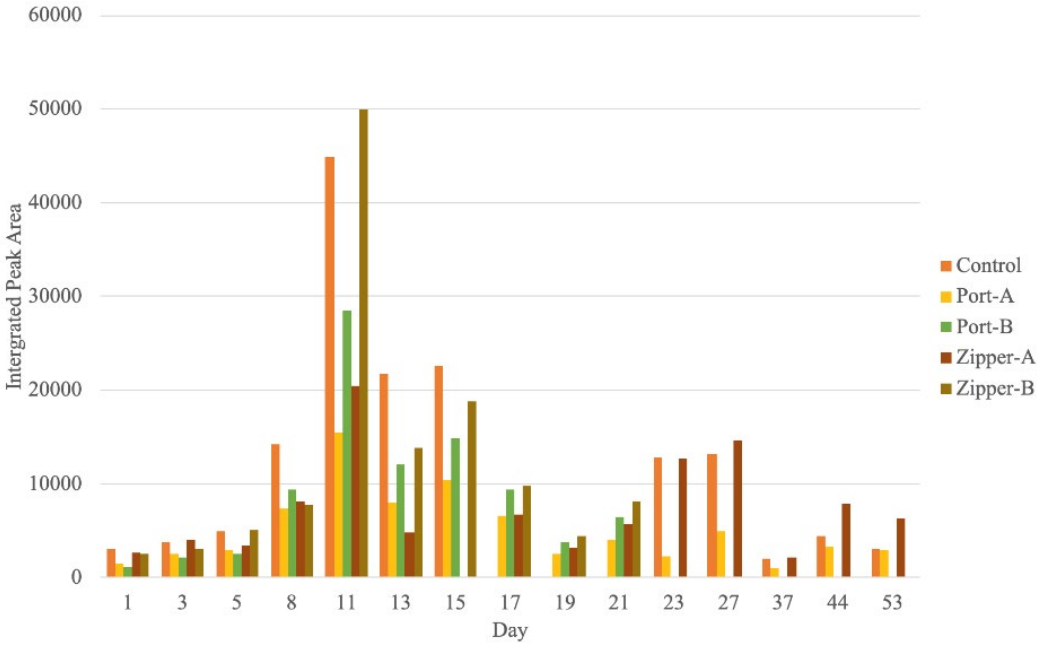


Figure 48. DEET emission over time. Highest DEET was observed on day 11.

Both bags appeared to contain the volatiles emitted by the contained decomposing cadavers until about day 11. After this point HRCU B appeared to emit sulfur-containing

volatiles. HRCU A prevented odor emission significantly longer than HRCU B, appearing to hold major volatiles until around day 44. Although low level volatiles were observed before that, it is believed that these were primarily from carry-over from HRCU B and after its removal those volatiles detected on HRCU A fibers decreased back to baseline. After day 44 a sharp increase was observed for HRCU A as had been seen at days 11-13 for HRCU B. Both HRCU A and HRCU B had activated charcoal filter in the port. However, from the results we obtained. we could clearly see the port filter of HRCU B failed as port B was one of the major sources of VOC emitted.

The inner absorbent fabric in HRCU A has an activated charcoal coating to contain chemical hazards. This was missing in HRCU B. This additional feature would have led to the absorption of considerable amount of a volatiles. However, once saturation was reached, HRCU A started discharging volatiles.

Conclusion

Both HRCU B and HRCU A failed over the course of the experiment. HRCU B's failure was ascertained by the results obtained by GCMS analysis which detected the massive leakage of H_2S and CH_3SH compared to control around days 11-15. As this bag was designed to contain larger biological agents, it is not surprising that it couldn't keep small molecules like H_2S and CH_3SH from escaping.

HRCU A showed promising results until day 37. This bag has an activated charcoal layer, which is expected to have absorbed significant amounts of VOCs.

REFERENCES

(1) Johnson, A. L.; Edson, K. Z.; Totah, R. A.; Rettie, A. E. Cytochrome P450 ω-Hydroxylases in Inflammation and Cancer. *Adv. Pharmacol*. **2015**, 74, 223–262. DOI:/10.1016/bs.apha.2015.05.002

(2) Sehgal, N.; Agarwal, V.; Valli, R. K.; Joshi, S. D.; Antonovic, L.; Strobel, H. W.; Ravindranath, V. Cytochrome P4504f, a Potential Therapeutic Target Limiting Neuroinflammation. *Biochem. Pharmacol.* **2011**, *82* (1), 53–64. DOI:/10.1016/j.bcp.2011.03.025

(3) Ohno, Y.; Nakamichi, S.; Ohkuni, A.; Kamiyama, N.; Naoe, A.; Tsujimura, H.; Yokose, U.; Sugiura, K.; Ishikawa, J.; Akiyama, M.; Kihara, A. Essential Role of the Cytochrome P450 CYP4F22 in the Production of Acylceramide, the Key Lipid for Skin Permeability Barrier Formation. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (25), 7707–7712. DOI:/10.1073/pnas.1503491112

(4) Wang, C.; Tao, Q.; Wang, X.; Wang, X.; Zhang, X. Impact of High-Fat Diet on Liver Genes Expression Profiles in Mice Model of Nonalcoholic Fatty Liver Disease. *Environ. Toxicol. Pharmacol.* **2016**, *45*, 52–62. DOI:/10.1016/j.etap.2016.05.014

(5) Jarrar, Y. B.; Lee, S. -J. Molecular Functionality of Cytochrome P450 4 (CYP4) Genetic Polymorphisms and Their Clinical Implications. *Int. J. Molec. Sci*. **2019**, 20(17), 4274. DOI:/10.3390/ijms20174274

(6) Kikuta, Y.; Kasyu, H.; Kusunose, E.; Kusunose, M. Expression and Catalytic Activity of Mouse Leukotriene B_4 ω-Hydroxylase, CYP4F14. *Arch. Biochem. Biophys.* **2000**, *383* (2), 225–232. DOI:/10.1006/abbi.2000.2078

(7) Biringer, R. G. The Role of Eicosanoids in Alzheimer's Disease. *Int. J. Environ. Res. Public Health* **2019**, *16* (14), 2560. DOI:/10.3390/ijerph16142560

(8) Kalsotra, A.; Strobel, H. W. Cytochrome P450 4F Subfamily: At the Crossroads of Eicosanoid and Drug Metabolism. *Pharmaco. Therap.* **2006**, 112 (3), 589–611. DOI:/10.1016/j.pharmthera.2006.03.008

(9) Hotz, A.; Bourrat, E.; Küsel, J.; Oji, V.; Alter, S.; Hake, L.; Korbi, M.; Ott, H.; Hausser, I.; Zimmer, A. D.; Fischer, J. Mutation Update for CYP4F22 Variants Associated with Autosomal Recessive Congenital Ichthyosis. *Hum. Mutat.* **2018**, *39* (10), 1305–1313. DOI:/10.1002/humu.23594

(10) Elias, P. M.; Williams, M. L.; Holleran, W. M.; Jiang, Y. J.; Schmuth, M. Pathogenesis of Permeability Barrier Abnormalities in the Ichthyoses: Inherited Disorders of Lipid Metabolism. *J. Lipid Res*. **2008**, 49, 697–714. DOI:/10.1194/jlr.R800002-JLR200

(11) Feingold, K. R. The Role of Epidermal Lipids in Cutaneous Permeability Barrier Homeostasis. *J. Lipid Res*. **2007**, 48, 2531–2546. DOI:/10.1194/jlr.R700013-JLR200

(12) Terrinoni, A.; Serra, V.; Codispoti, A.; Talamonti, E.; Bui, L.; Palombo, R.; Sette, M.; Campione, E.; Didona, B.; Annicchiarico-Petruzzelli, M.; Zambruno, G.; Melino, G.; Candi, E. Novel Transglutaminase 1 Mutations in Patients Affected by Lamellar Ichthyosis. *Cell Death Dis.* **2012**, *3* (10), e416. DOI:/10.1038/cddis.2012.152.

(13) Lefèvre, C.; Bouadjar, B.; Ferrand, V.; Tadini, G.; Mégarbané, A.; Lathrop, M.; Prud'homme, J.-F.; Fischer, J. Mutations in a New Cytochrome P450 Gene in

Lamellar Ichthyosis Type 3. *Hum. Mol. Genet.* **2006**, *15* (5), 767–776. DOI:/10.1093/hmg/ddi491.

(14) Quehl, P.; Hollender, J.; Schüürmann, J.; Brossette, T.; Maas, R.; Jose, J. Co-Expression of Active Human Cytochrome P450 1A2 and Cytochrome P450 Reductase on the Cell Surface of *Escherichia Coli*. *Microb. Cell Fact.* **2016**, 26, 1–15. DOI:/10.1186/s12934-016-0427-5.

(15) Rosano, G. L.; Ceccarelli, E. A. Recombinant Protein Expression in *Escherichia coli:* Advances and Challenges. *Front. Microbiol*. **2014**, 5, 172. DOI:/10.3389/fmicb.2014.00172

(16) Del Solar, G.; Espinosa, M. Plasmid Copy Number Control: An Ever-Growing Story. *Mol. Microbiol*. **2000**, 37(3), 492-500. DOI:/10.1046/j.1365-2958.2000.02005.x

(17) Sasnauskas, G.; Connolly, B. A.; Halford, S. E.; Siksnys, V. Site-Specific DNA Transesterification Catalyzed by a Restriction Enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, 104(7), 2115-2120. DOI:/10.1073/pnas.0608689104

(18) Shuman, S. DNA Ligases: Progress and Prospects. *J. Biol. Chem*. **2009**, 284(26), 17365–17369. DOI:/10.1074/jbc.R900017200

(19) Kleber-Janke, T.; Becker, W. -M. Use of Modified BL21(DE3) *Escherichia coli* Cells for High-Level Expression of Recombinant Peanut Allergens Affected by Poor Codon Usage. *Protein Expr. Purif.* **2000**, *19* (3), 419–424. DOI:/10.1006/prep.2000.1265

(20) Luthra, A.; Denisov, I. G.; Sligar, S. G. Spectroscopic Features of Cytochrome P450 Reaction Intermediates. Arch. Biochem. Biophys. **2011**, 507(1), 26–35.

DOI:/10.1016/j.abb.2010.12.008

(21) Senge, M. O.; Ryan, A. A.; Letchford, K. A.; MacGowan, S. A.; Mielke, T. Chlorophylls, Symmetry, Chirality, and Photosynthesis. *Symmetry.* **2014**, *6* (3), 781–843. DOI:/10.3390/sym6030781

(22) Song, Y.; Schubert, A.; Maret, E.; Burdick, R. K.; Dunietz, B. D.; Geva, E.; Ogilvie, J. P. Vibronic Structure of Photosynthetic Pigments Probed by Polarized Two-Dimensional Electronic Spectroscopy and *Ab Initio* Calculations. *Chem. Sci.* **2019**, *10* (35), 8143–8153. DOI:/10.1039/c9sc02329a

(23) Gouterman, M.; Wagnière, G. H.; Snyder, L. C. Spectra of Porphyrins. Part II. Four Orbital Model. *J. Mol. Spectrosc.* **1963**, *11* (1–6), 108–127. DOI:/10.1016/0022-2852(63)90011-0

(24) Mak, P. J.; Denisov, I. G. Spectroscopic Studies of the Cytochrome P450 Reaction Mechanisms. *Biochim. Biophys. Acta*. *Proteins Proteom.* **2018,** 1866(1), 178–204. DOI:/10.1016/j.bbapap.2017.06.021

(25) Haines, D. C.; Hegde, A.; Chen, B.; Zhao, W.; Bondlela, M.; Humphreys, J. M.; Mullin, D. A.; Tomchick, D. R.; Machius, M.; Peterson, J. A. A Single Active-Site Mutation of P450BM-3 Dramatically Enhances Substrate Binding and Rate of Product Formation. *Biochemistry* **2011**, *50* (39), 8333–8341. DOI:/10.1021/bi201099j.

(26) Koch, A. L.; Turbidity Measurements of Bacterial Cultures in Some Available Commercial Instruments. *Anal. Biochem.* **1970**, 38(1), 252-*259*. DOI:/10.1016/0003-2697(70)90174-0

(27) Schwaneberg, U.; Schmidt-Dannert, C.; Schmitt, J.; Schmid, R. D. A Continuous

Spectrophotometric Assay for P450 BM-3, a Fatty Acid Hydroxylating Enzyme, and Its Mutant F87A. *Anal. Biochem.* **1999**, 269(2), 359-*366.* DOI:/10.1006/abio.1999.4047

(28) Li, Q. -S.; Schwaneberg, U.; Fischer, M.; Schmitt, J.; Rgen Pleiss, J.; Lutz-Wahl, S.; Schmid, R. D. Rational Evolution of a Medium Chain-Specific Cytochrome P-450 BM-3 Variant. *Biochim. Biophys. Acta.* **2001**, 1545(1-2),114-121. DOI:/10.1016/s0167-4838(00)00268-5

(29) Whitehouse, C. J. C.; Bell, S. G.; Wong, L. L. P450BM3 (CYP102A1): Connecting the Dots. *Chem. Soc. Rev.* **2012**, *41* (3), 1218–1260. DOI:/10.1039/c1cs15192d.

(30) Murataliev, M. B.; Feyereisen, R. Functional Interactions in Cytochrome P450BM3. Fatty Acid Substrate Binding Alters Electron-Transfer Properties of the Flavoprotein Domain. *Biochemistry* **1996**, 35(47), 15029–15037. DOI:/10.1021/bi961667u

(31) Haines, D. C.; Tomchick, D. R.; Machius, M.; Peterson, J. A. Pivotal Role of Water in the Mechanism of P450BM-3. *Biochemistry* **2001**, *40* (45), 13456–13465. DOI:/10.1021/bi011197q

(32) Dubey, K. D.; Shaik, S. Cytochrome P450 - The Wonderful Nanomachine Revealed through Dynamic Simulations of the Catalytic Cycle. *Acc. Chem. Res.* **2019**, *52* (2), 389–399. DOI:/10.1021/acs.accounts.8b00467

(33) Isin, E. M.; Guengerich, F. P. Complex Reactions Catalyzed by Cytochrome P450 Enzymes. *Biochim. Biophys. Acta, Gen. Sub*. **2007**, 1770 (3), 314–329. DOI:/10.1016/j.bbagen.2006.07.003

(34) Morris, G. M.; Goodsell, D. S.; Pique, M. E.; Lindy, W.; Huey, R.; Forli, S.; Hart, W. E.; Halliday, S.; Belew, R.; Olson, A. J. *User Guide AutoDock Version 4.2 Updated for Version 4.2.5 Automated Docking of Flexible Ligands to Flexible Receptors*. **2012** http://autodock.scripps.edu/faqs-help/manual/autodock-4-2-user-guide/AutoDock4.2_UserGuide.pdf (Accessed on 05.10.2020)

(35) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem*. **2009**, *30* (16), 2785–2791. DOI:/10.1002/jcc.21256

(36) Forli, S.; Olson, A. J. A Force Field with Discrete Displaceable Waters and Desolvation Entropy for Hydrated Ligand Docking. *J. Med. Chem.* **2012**, *55* (2), 623–638. DOI:/10.1021/jm2005145.

(37) TrueORF TM CDNA Clones and PrecisionShuttle TM Vector System Application Guide. https://www.origene.com/products/vectors/precisionshuttle-system (Accessed on 02.20.2020)

(38) Green, R. M; Sambrook, J. Molecular Cloning : A Laboratory Manual. Cold Spring Harbor Laboratory Press, **2012.** 162-168.

(39) Fernández-Castané, A.; Caminal, G.; López-Santín, J. Direct Measurements of IPTG Enable Analysis of the Induction Behavior of *E. coli* in High Cell Density Cultures. *Microb. Cell Fact* **2012**, 58,1-9. DOI:/10.1186/14752859-11-58

(40) Wagner, S.; Klepsch, M. M.; Schlegel, S.; Appel, A.; Draheim, R.; Tarry, M.; Högbom, M.; van Wijk, K. J.; Slotboom, D. J.; Persson, J. O.; de Gier, J. -W.

Tuning *Escherichia coli* for Membrane Protein Overexpression. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (38), 14371–14376. doi.org/10.1073/pnas.0804090105

(41) Saribas, A. S.; Gruenke, L.; Waskell, L. Overexpression and Purification of the Membrane-Bound Cytochrome P450 2B4. *Protein Expr. Purif.* **2001**, *21* (2), 303–309. DOI:/10.1006/prep.2000.1377

(42) Zhang, Y. I-TASSER Server for Protein 3D Structure Prediction. *BMC Bioinformatics* **2008**, *9*, 1–8. doi:/10.1186/1471-2105-9-40

(43) Syed, K.; Mashele, S. S. Comparative Analysis of P450 Signature Motifs EXXR and CXG in the Large and Diverse Kingdom of Fungi: Identification of Evolutionarily Conserved Amino Acid Patterns Characteristic of P450 Family. *PLoS One* **2014**, *9* (4), e95616. DOI:/10.1371/journal.pone.0095616

(44) Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 2010, 5, 725–738. DOI.org/10.1038/nprot.2010.5

(45) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein Structure and Function Prediction. *Nat. Methods* **2014**, *12* (1), 7–8. DOI:/10.1038/nmeth.3213

(46) Hsu, M. -H.; Baer, B. R.; Rettie, A. E.; Johnson, E. F. The Crystal Structure of Cytochrome P450 4B1 (CYP4B1) Monooxygenase Complexed with Octane Discloses Several Structural Adaptations for ω-Hydroxylation. *J. Biol. Chem.* **2017**, *292* (13), 5610–5621. DOI:/10.1074/jbc.M117.775494

(47) Mustafa, G.; Nandekar, P. P.; Camp, T. J.; Bruce, N. J.; Gregory, M. C.; Sligar, S. G.; Wade, R. C. Influence of Transmembrane Helix Mutations on Cytochrome

P450-Membrane Interactions and Function. *Biophys. J.* **2019**, *116* (3), 419–432. DOI:/10.1016/j.bpj.2018.12.014

(48) Li, H.; Poulos, T. L. The Structure of the Cytochrome P450BM-3 Haem Domain Complexed with the Fatty Acid Substrate, Palmitoleic Acid. *Nat. Struct. Biol.***1997**, 4, 140-146. DOI:/ 10.1038/nsb0297-140

(49) Baer, B. R.; Kunze, K. L.; Rettie, A. E. Mechanism of Formation of the Ester `Linkage between Heme and Glu310 of CYP4B1: ^{18}O Protein Labeling Studies. *Biochemistry* **2007**, *46* (41), 11598–11605. DOI:/10.1021/bi701064

(50) Oda, A.; Yamaotsu, N.; Hirono, S. New AMBER Force Field Parameters of Heme Iron for Cytochrome P450s Determined by Quantum Chemical Calculations of Simplified Models. *J. Comput. Chem.* **2005**, *26* (8), 818–826. DOI:/10.1002/jcc.20221

(51) United Nations Office For Outer Space Affairs; Committee on the Peaceful Uses of Outer Space. Space Debris Mitigation Guidelines of the Committee on the Peaceful Uses of Outer Space. https://www.unoosa.org/oosa/documents-and-resolutions/search.jspx?view=documents&match=ST/SPACE/49 (Accessed on 05.01.2020)

(52) Janaway, R. C.; Percival, S. L.; Wilson, A. S. Decomposition of Human Remains. In *Microbio. Aging: Clinical Manifestations*; Human press, **2009**, 313–334. DOI:/10.1007/978-1-59745-327-1_14

(53) Vass, A.A. (2001) Beyond the Grave—Understanding Human Decomposition. *Microbiol. Today*, 28, 190-192 http://www.academia.dk/BiologiskAntropologi/Tafonomi/PDF/ArpadVass_2001.p

df (Accessed on 03.05.2020)

(54) Vas, G.; Vékey, K. Solid-Phase Microextraction: A Powerful Sample Preparation Tool Prior to Mass Spectrometric Analysis. *J. Mass Spectrom.* **2004**, *39* (3), 233–254. DOI:/10.1002/jms.606.

(55) Oliveira, W. d. S.; Monsalve, J. O.; Nerin, C.; Padula, M.; Godoy, H. T. Characterization of Odorants from Baby Bottles by Headspace Solid Phase Microextraction Coupled to Gas Chromatography-Olfactometry-Mass Spectrometry. *Talanta* **2020**, 207, 120301. DOI:/10.1016/j.talanta.2019.120301.

(56) Vass, A. A.; Smith, R. R.; Thompson, C. V.; Burnett, M. N.; Dulgerian, N.; Eckenrode, B. A. Odor Analysis of Decomposing Buried Human Remains. *J. Forensic Sci.* **2008**, *53* (2), 384–391. DOI:/10.1111/j.1556-4029.2008.00680.x.

(57) Hoffman, E. M.; Curran, A. M.; Dulgerian, N.; Stockham, R. A.; Eckenrode, B. A. Characterization of the Volatile Organic Compounds Present in the Headspace of Decomposing Human Remains. *Forensic Sci. Int.* **2009**, *186* (1–3), 6–13. DOI:10.1016/j.forsciint.2008.12.022.

APPENDIX A

Cytochrome P450, family 4, subfamily F, polypeptide 2 [Homo sapiens]

Sequence ID: AAH67440.1 **Length:** 520 **Number of Matches:** 1

Range 1: 1 to 520 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
879 bits(2270)	0.0	Compositional matrix adjust.	416/520(80%)	467/520(89%)	0/520(0%)

```
Query  1    MSQLSLSWLGLGPEVAFPWKTLLLLGASWILARILIQIYAAYRNYRHLHGFPQPPKRNWL  60
            MSQLSLSWLGLGP  A PW  LLL+GASW+LA +L   YA Y N R L  FPQPP+RNW
Sbjct  1    MSQLSLSWLGLGPVAASPWLLLLLVGASWLLAHVLAWTYAFYDNCRRLRCFPQPPRRNWF  60

Query  61   MGHVGMVTPTEQGLKELTRLVGTYPQGFLMWIGPMVPVITLCHSDIVRSILNASAAVALK  120
             GH GMV PTE+G++ LT+LV TYPQGF +W+GP+ P+++LCH DI+RS++NASAA+A K
Sbjct  61   WGHQGMVNPTEEGMRVLTQLVATYPQGFKVWMGPISPLLSLCHPDIIRSVINASAAIAPK  120

Query  121  DVIFYSILKPWLGDGLLVSAGDKWSRHRRMLTPAFHFNILKPYVKIFNDSTNIMHAKWQR  180
            D  FYS L+PWLGDGLL+SAGDKWSRHRRMLTPAFHFNILKPY+KIFNDS NIMHAKWQ
Sbjct  121  DKFFYSFLEPWLGDGLLLSAGDKWSRHRRMLTPAFHFNILKPYMKIFNDSVNIMHAKWQL  180

Query  181  LISDGSARLDMFEHVSLMTLDSLQKCVFSFDSNCQEKSSEYIAAILELSALVAKRHQQPL  240
            L S+GSA LDMFEH+SLMTLDSLQKCVFSFDS+CQEK SEYIAAILELSALV+KRH + L
Sbjct  181  LASEGSACLDMFEHISLMTLDSLQKCVFSFDSHCQEKPSEYIAAILELSALVSKRHHEIL  240

Query  241  MFMDLLYNLTPDGMRFRKACNVVHEFTDAVIRERHRTLPDQGLDDFLKSKAKSKTLDFID  300
            + +D LY LTPDG RFR+AC +VH+FTDAVI+ER RTLP QG+DDFL++KAKSKTLDFID
Sbjct  241  LHIDFLYYLTPDGQRFRRACRLVHDFTDAVIQERRRTLPSQGVDDFLQAKAKSKTLDFID  300

Query  301  VLLLSKDEDGKELSDEDIRAEADTFMFEGHDTTASGLSWILYNLARHPEYQERCRQEVQE  360
            VLLLSKDEDGK+LSDEDIRAEADTFMFEGHDTTASGLSW+LY+LA+HPEYQERCRQEVQE
Sbjct  301  VLLLSKDEDGKKLSDEDIRAEADTFMFEGHDTTASGLSWVLYHLAKHPEYQERCRQEVQE  360

Query  361  LLRGREPEEIEWDDLAQLPFLTMCIKESLRLHPPVTVISRCCTQDILLPDGRTIPKGIIC  420
            LL+ REP+EIEWDDLA LPFLTMC+KESLRLHPPV VISR  TQDI+LPDGR IPKGIIC
Sbjct  361  LLKDREPKEIEWDDLAHLPFLTMCMKESLRLHPPVPVISRHVTQDIVLPDGRVIPKGIIC  420

Query  421  LISIFGIHHNPSVWPDPEVYDPFRFDPENIKDSSPLAFIPFSAGPRNCIGQTFAMSEMKV  480
            LIS+FG HHNP++WPDPEVYDPFRFDPENIK+ SPLAFIPFSAGPRNCIGQTFAM+EMKV
Sbjct  421  LISVFGTHHNPAMWPDPEVYDPFRFDPENIKERSPLAFIPFSAGPRNCIGQTFAMAEMKV  480

Query  481  ALALTLLRFRLLPDDKEPRRQPELILRAEGGLWLRVEPLS  520
             LALTLLRFR+LPD  EPRR+PEL+LRAEGGLWLRVEPLS
Sbjct  481  VLALTLLRFRVLPDHTEPRRKPELVLRAEGGLWLRVEPLS  520
```

Figure A1: BLAST ClustalW Sequence Alignment of Mouse Cyp4f14 and Human CYP4F2. Alignment shows an 80% sequence identity between Cyp4f39 and CYP4F22.

cytochrome P450 4F22 [Homo sapiens]

Sequence ID: NP_775754.2 **Length:** 531 **Number of Matches:** 1

See 7 more title(s)

Range 1: 1 to 531 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
946 bits(2444)	0.0	Compositional matrix adjust.	461/532(87%)	496/532(93%)	1/532(0%)

```
Query  1    MLPLTDYLLQLLGLEKTAFRVYVVSALLLLVLFFFFRLLVRAFKLFSDFRITCRKLSCFP  60
            MLP+TD LL LLGLEKTAFR+Y VS LLL +LFF FRLL+R  +L   F ITCR+L CFP
Sbjct  1    MLPITDRLLHLLGLEKTAFRIYAVSTLLLFLLFFLFRLLLRFLRLCRSFYITCRRLRCFP  60

Query  61   EPPGRHWLLGHMSMYLPNEKGLQNEKKVLDTMHHIILAWVGPFLPLLVLVHPDYIKPVLG  120
            +PP R+WLLGH+ MYLPNE GLQ+EKKVLD MHH++L W+GP LPLLVLVHPDYIKP+LG
Sbjct  61   QPPRRNWLLGHLGMYLPNEAGLQDEKKVLDNMHHVLLVWMGPVLPLLVLVHPDYIKPLLG  120

Query  121  ASAAIAPKDEFFYSFLKPWLGDGLLISKGNKWSRHRRLLTPAFHFDILKPYMKIFNQCTN  180
            ASAAIAPKD+ FY FLKPWLGDGLL+SKG+KWSRHRRLLTPAFHFDILKPYMKIFNQ  +
Sbjct  121  ASAAIAPKDDLFYGFLKPWLGDGLLLSKGDKWSRHRRLLTPAFHFDILKPYMKIFNQSAD  180

Query  181  IMHAKWRRHLAEGSVTSFDMFEHISLMTLDSLQKCVFSYNSDCQERMSDYISSIIELSAL  240
            IMHAKWR HLAEGS  S DMFEHISLMTLDSLQKCVFSYNS+CQE+MSDYIS+IIELSAL
Sbjct  181  IMHAKWR-HLAEGSAVSLDMFEHISLMTLDSLQKCVFSYNSNCQEKMSDYISAIIELSAL  239

Query  241  VVRRQYRLHHYLDFMYYLTADGRRFRQACDTVHNFTTEVIQERRQALRQQGAEAWLKAKQ  300
             VRRQYRLHHYLDF+YY +ADGRRFRQACD VH+FTTEVIQERR+ALRQQGAEAWLKAKQ
Sbjct  240  SVRRQYRLHHYLDFIYYRSADGRRFRQACDMVHHFTTEVIQERRRALRQQGAEAWLKAKQ  299

Query  301  GKTLDFIDVLLLAKDEEGKELSDEDIRAEADTFMFEGHDTTSSGLSWALFNLAKYPEYQE  360
            GKTLDFIDVLLLA+DE+GKELSDEDIRAEADTFMFEGHDTTSSG+SW LFNLAKYPEYQE
Sbjct  300  GKTLDFIDVLLLARDEDGKELSDEDIRAEADTFMFEGHDTTSSGISWMLFNLAKYPEYQE  359

Query  361  KCREEIQEVMKGRELEELDWDDLTQLPFTTMCIKESLRQFPPVTLISRRCTEDIKLPDGR  420
            KCREEIQEVMKGRELEEL+WDDLTQLPFTTMCIKESLRQ+PPVTL+SR+CTEDIKLPDGR
Sbjct  360  KCREEIQEVMKGRELEELEWDDLTQLPFTTMCIKESLRQYPPVTLVSRQCTEDIKLPDGR  419

Query  421  VIPKGIICLVSIYGTHHNPIVWPDSKVYNPYRFDPDTPQQRSPLAFVPFSAGPRNCIGQS  480
            +IPKGIICLVSIYGTHHNP VWPDSKVYNPYRFDPD PQQRSPLA+VPFSAGPRNCIGQS
Sbjct  420  IIPKGIICLVSIYGTHHNPTVWPDSKVYNPYRFDPDNPQQRSPLAYVPFSAGPRNCIGQS  479

Query  481  FAMAEMRVVVALTLLRFRLSVDRTHKVRRKPELILRTENGLWLNVEPLPSRA  532
            FAMAE+RVVVALTLLRFRLSVDRT KVRRKPELILRTENGLWL VEPLP RA
Sbjct  480  FAMAELRVVVALTLLRFRLSVDRTRKVRRKPELILRTENGLWLKVEPLPPRA  531
```

Figure A2. BLAST ClustalW Sequence Alignment of Mouse Cyp4f39 and Human CYP4F22. Alignment shows an 87% sequence identity between Cyp4f39 and CYP4F22.

APPENDIX B

In order to add the missing heme in the active site, first the coordinates for the crystal structure 1JPZ were retrieved from Protein Data Bank (PDB). Then it was aligned with the I-TASSER generated structures using *align 1jpz, model1* command.

Heme was extracted using the PyMOL "extract object" tool (Figure B1 and Figure B2). Finally, the I-TASSER model enzyme was saved as a new molecule.

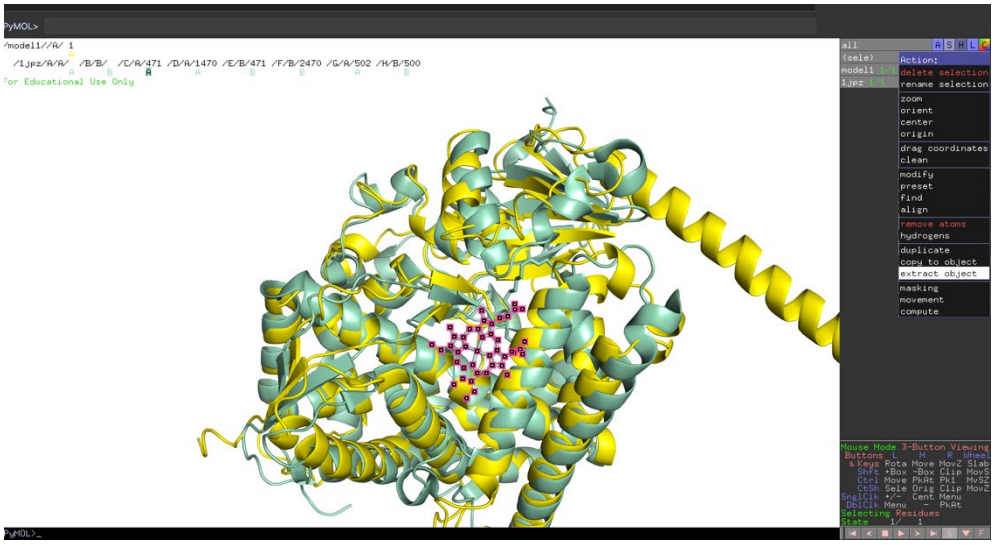


Figure B1. Aligned Cyp4f14 (yellow) and $P450_{BM-3}$ heme domain (teal). Selected residues of heme-porphyrin are highlighted in pink.

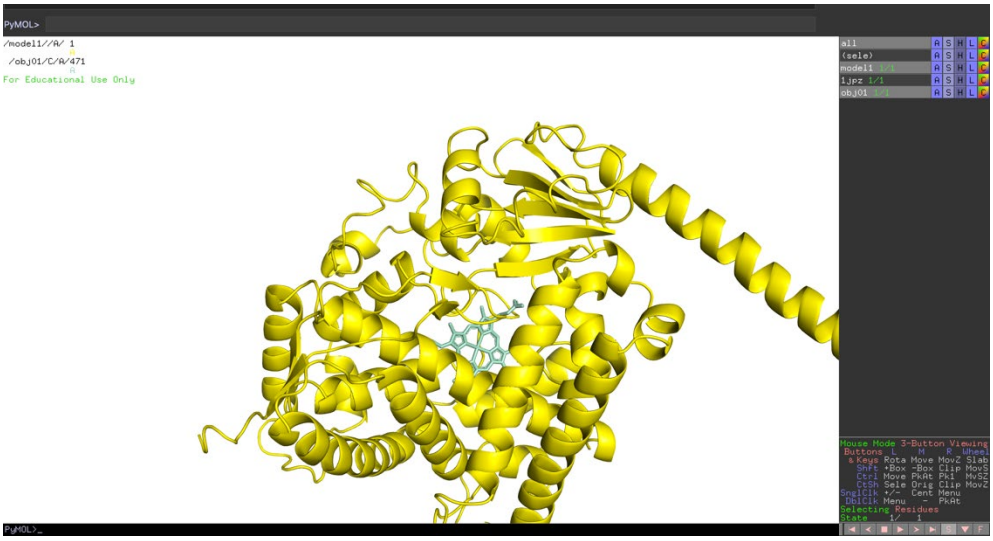
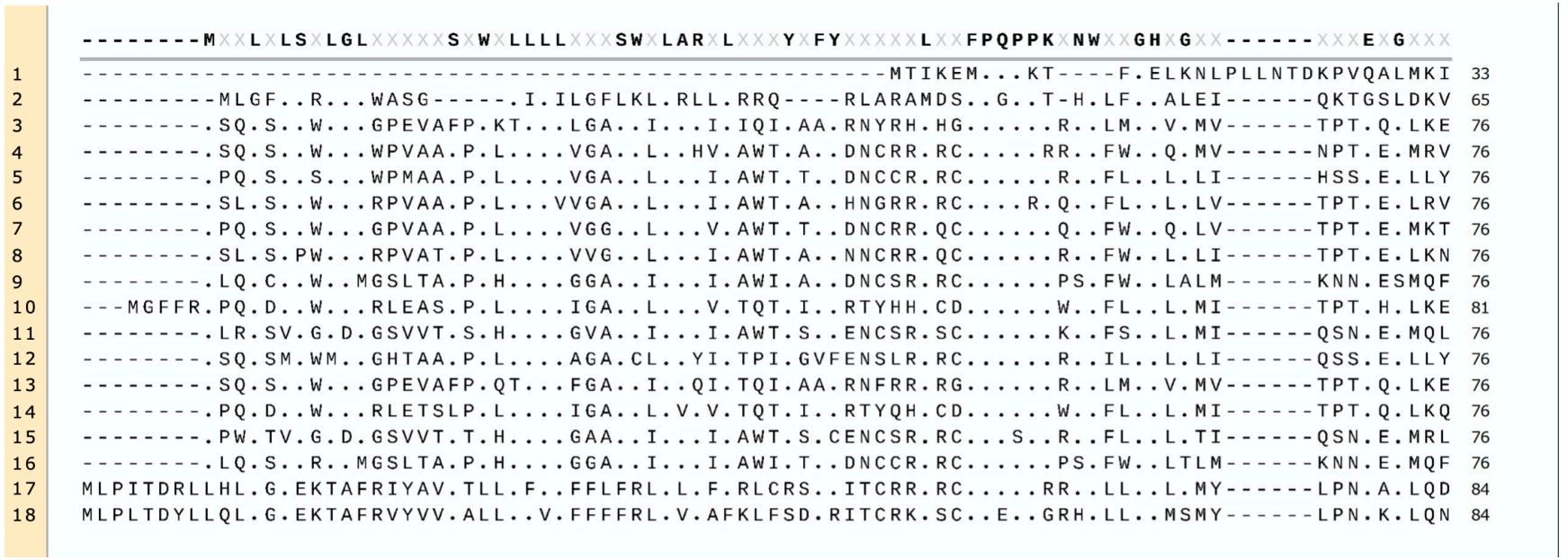


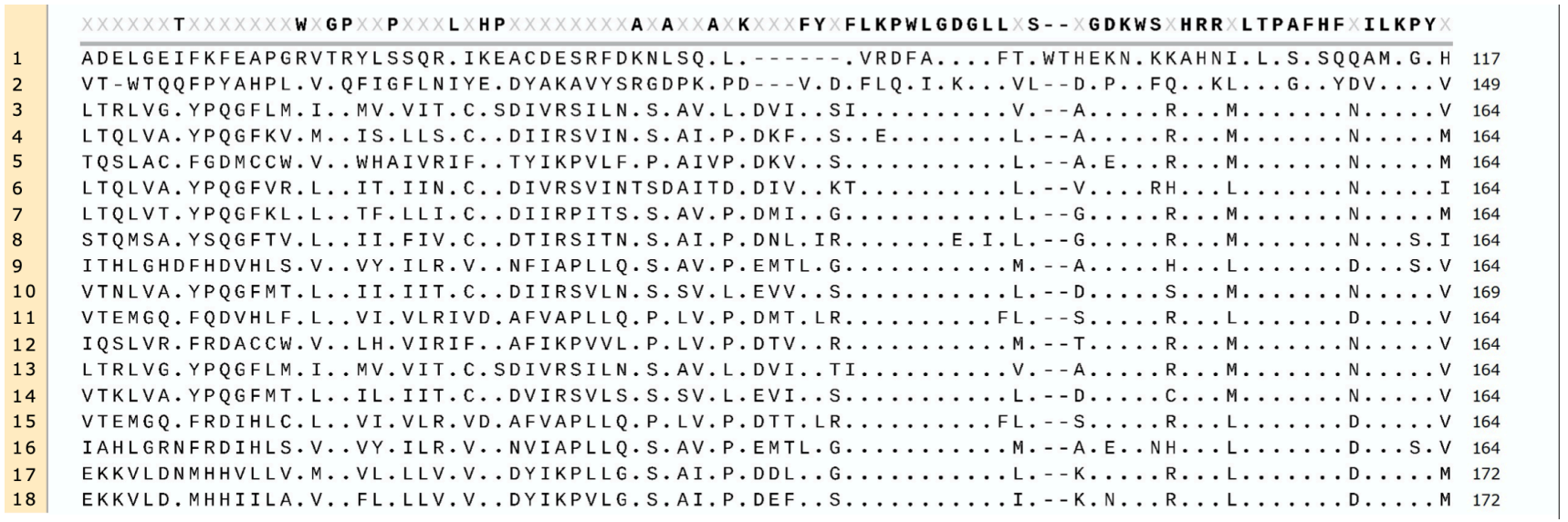
Figure B2. Extracted heme residue in the hydrophobic core of Cyp4f14

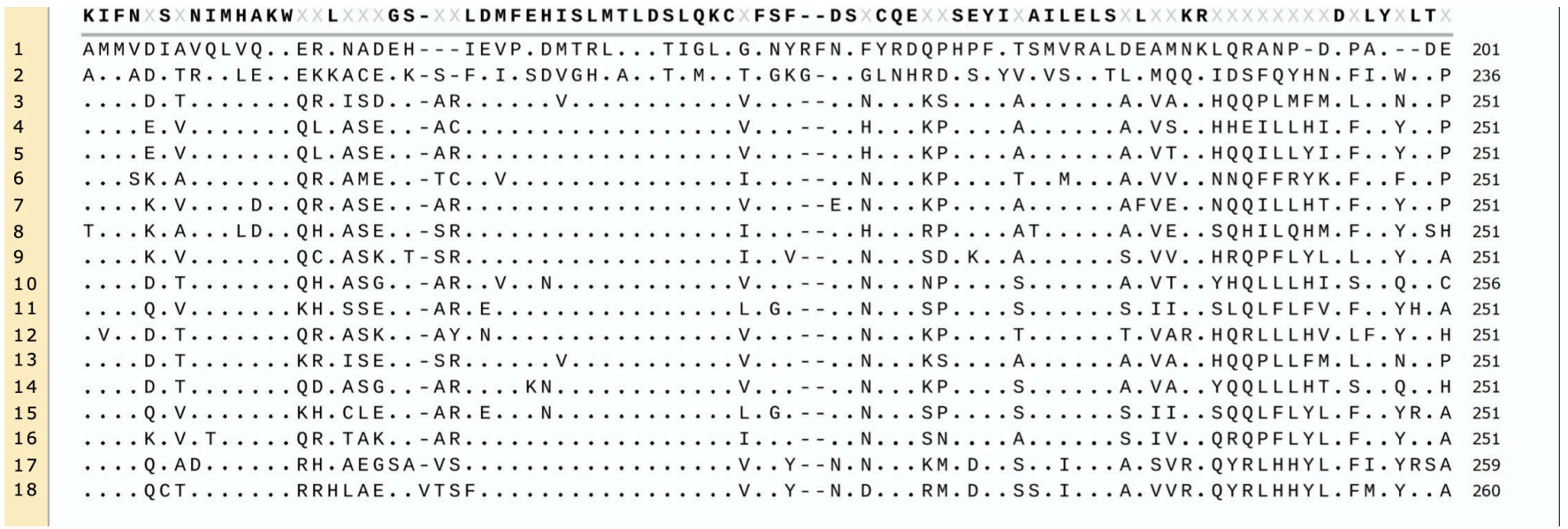
APPENDIX C

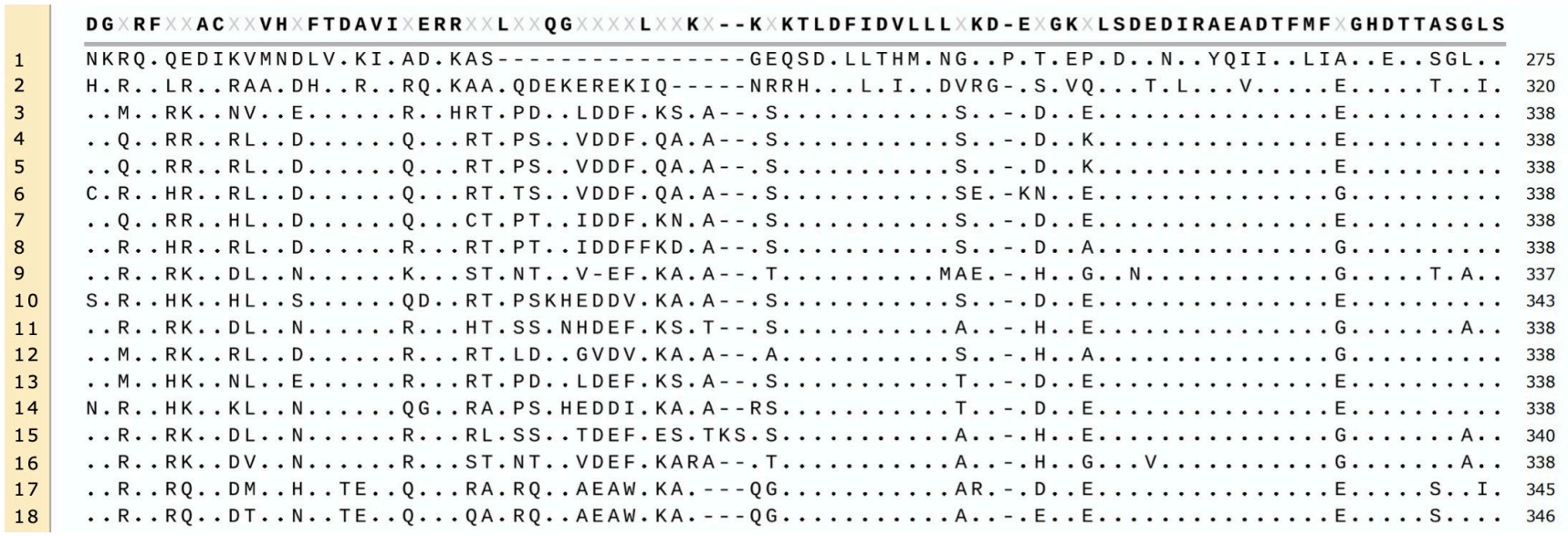
Complete Amino Acid Sequence Alignments of Cytochrome P450 4F family proteins including six human genes (CYP4F2, 4F3, 4F8, 4F11, 4F12 and 4F22), 4 rat genes (CYP4F1, 4F4, 4F5 and 4F6), 6 mouse genes (Cyp4f13, 4f14, 4f15, 4f16, 4f18 and 4f39). This alignment also includes the model enzyme $P450_{BM-3}$ and recently crystalized rabbit CYP4B1 sequence. Alignment was done by the MUSCLE alignment tool. Visualized using Snapgene 2.0.

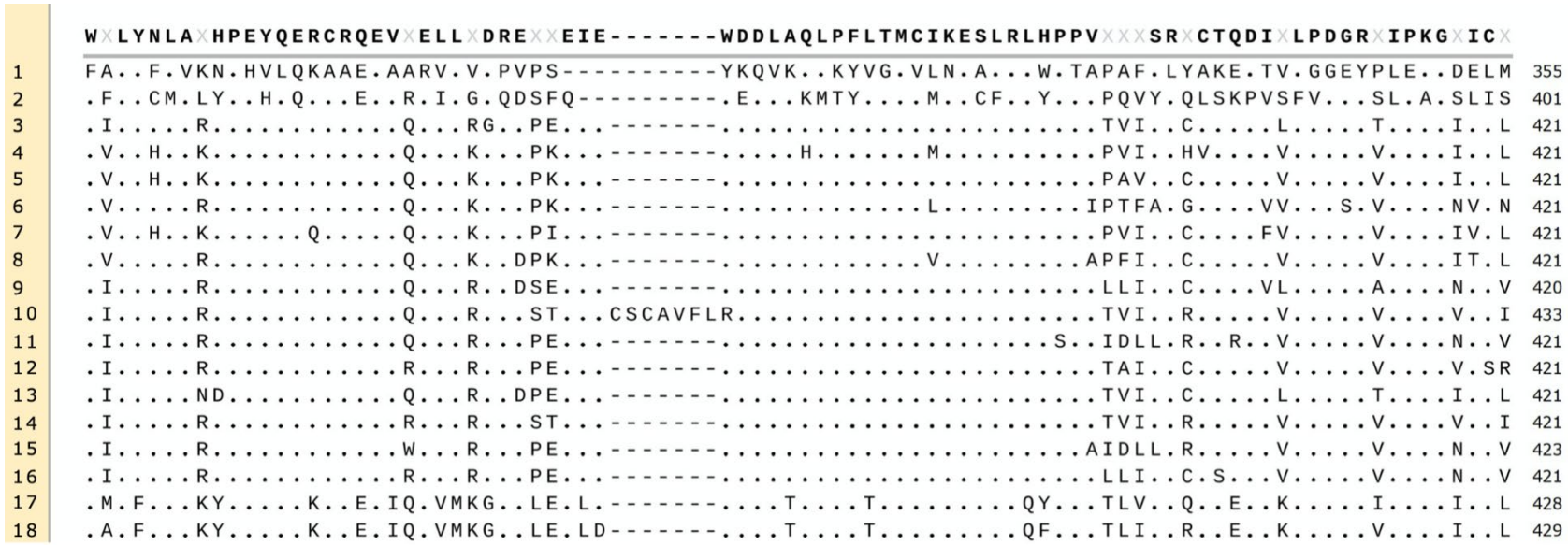
1. $P450_{BM-3}$
2. Rabbit CYP4B1
3. Mouse Cyp4f14
4. Human CYP4F2
5. Human CYP4F3
6. Human CYP4F8
7. Human CYP4F11
8. Human CYP4F12
9. Mouse Cyp4f13
10. Mouse Cyp4f15
11. Mouse Cyp4f16
12. Mouse Cyp4f18
13. Rat CYP4F1
14. Rat CYP4F4
15. Rat CYP4F5
16. Rat CYP4F6
17. Human CYP4F22
18. Mouse Cyp4f39

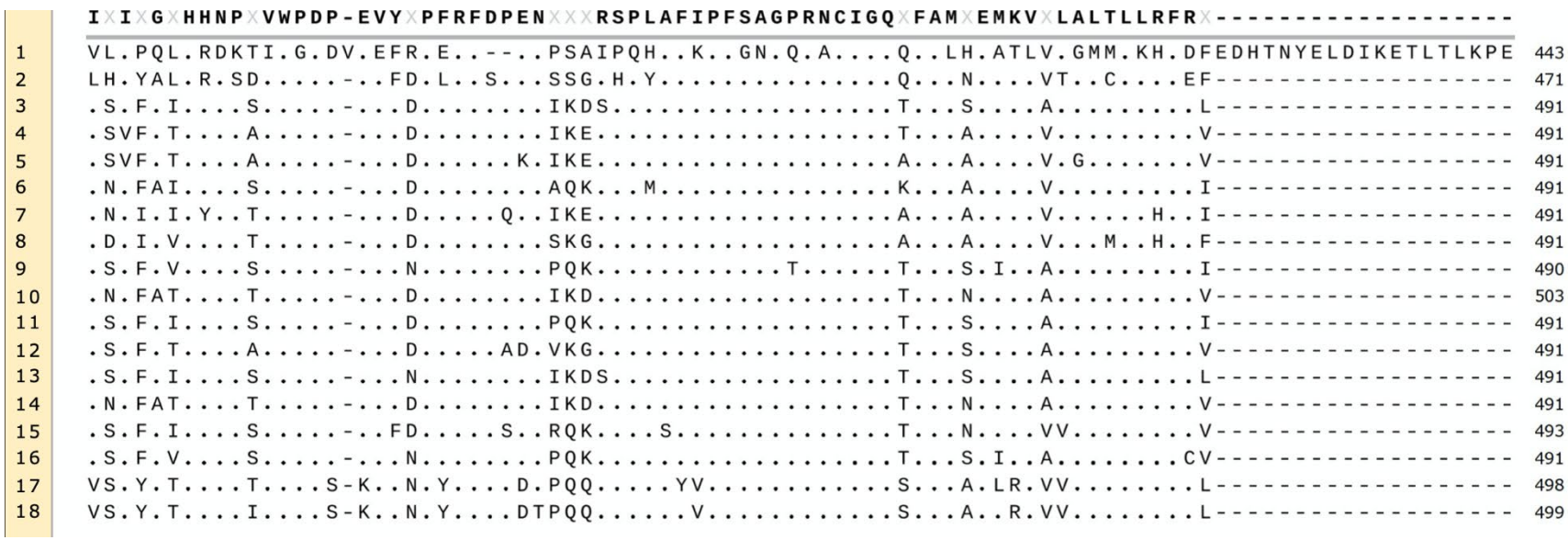

--------MXXLXLSXLGLXXXXXSXWXLLLLXXXSWXLARXLXXXYXFYXXXXXLXXFPQPPKXNWXXGHXGXX------XXXEXGXXX
1 ----------MTIKEM...KT----F.ELKNLPLLNTDKPVQALMKI 33
2 ----------MLGF..R...WASG-----.I.ILGFLKL.RLL.RRQ----RLARAMDS..G..T-H.LF..ALEI------QKTGSLDKV 65
3 ---------.SQ.S..W...GPEVAFP.KT...LGA..I...I.IQI.AA.RNYRH.HG......R..LM..V.MV------TPT.Q.LKE 76
4 ---------.SQ.S..W...WPVAA.P.L....VGA..L..HV.AWT.A..DNCRR.RC.....RR..FW..Q.MV------NPT.E.MRV 76
5 ---------.PQ.S..S...WPMAA.P.L....VGA..L...I.AWT.T..DNCCR.RC......R..FL..L.LI------HSS.E.LLY 76
6 ---------.SL.S..W...RPVAA.P.L...VVGA..L...I.AWT.A..HNGRR.RC....R.Q..FL..L.LV------TPT.E.LRV 76
7 ---------.PQ.S..W...GPVAA.P.L....VGG..L...V.AWT.T..DNCRR.QC......Q..FW..Q.LV------TPT.E.MKT 76
8 ---------.SL.S.PW...RPVAT.P.L....VVG..L...I.AWT.A..NNCRR.QC......R..FW..L.LI------TPT.E.LKN 76
9 ---------.LQ.C..W..MGSLTA.P.H....GGA..I...I.AWI.A..DNCSR.RC......PS.FW..LALM------KNN.ESMQF 76
10 ---MGFFR.PQ.D..W...RLEAS.P.L....IGA..L...V.TQT.I..RTYHH.CD......W..FL..L.MI------TPT.H.LKE 81
11 ---------.LR.SV.G.D.GSVVT.S.H....GVA..I...I.AWT.S..ENCSR.SC......K..FS..L.MI------QSN.E.MQL 76
12 ---------.SQ.SM.WM..GHTAA.P.L....AGA.CL..YI.TPI.GVFENSLR.RC......R..IL..L.LI------QSS.E.LLY 76
13 ---------.SQ.S..W...GPEVAFP.QT...FGA..I..QI.TQI.AA.RNFRR.RG......R..LM..V.MV------TPT.Q.LKE 76
14 ---------.PQ.D..W...RLETSLP.L....IGA..L.V.V.TQT.I..RTYQH.CD......W..FL..L.MI------TPT.Q.LKQ 76
15 ---------.PW.TV.G.D.GSVVT.T.H....GAA..I...I.AWT.S.CENCSR.RC...S..R..FL..L.TI------QSN.E.MRL 76
16 ---------.LQ.S..R..MGSLTA.P.H....GGA..I...I.AWI.T..DNCCR.RC......PS.FW..LTLM------KNN.E.MQF 76
17 MLPITDRLLHL.G.EKTAFRIYAV.TLL.F..FFLFRL.L.F.RLCRS..ITCRR.RC.....RR..LL..L.MY------LPN.A.LQD 84
18 MLPLTDYLLQL.G.EKTAFRVYVV.ALL..V.FFFFRL.V.AFKLFSD.RITCRK.SC..E..GRH.LL..MSMY------LPN.K.LQN 84


XXXXXXTXXXXXXXXWXGPXXPXXXLXHPXXXXXXXXXAXAXXAXKXXXFYXFLKPWLGDGLLXS--XGDKWSXHRRXLTPAFHFXILKPYX
1 ADELGEIFKFEAPGRVTRYLSSQR.IKEACDESRFDKNLSQ.L.-------.VRDFA....FT.WTHEKN.KKAHNI.L.S.SQQAM.G.H 117
2 VT-WTQQFPYAHPL.V.QFIGFLNIYE.DYAKAVYSRGDPK.PD---V.D.FLQ.I.K...VL--D.P..FQ..KL...G..YDV....V 149
3 LTRLVG.YPQGFLM.I..MV.VIT.C.SDIVRSILN.S.AV.L.DVI..SI.........V.--A.....R...M.......N.....V 164
4 LTQLVA.YPQGFKV.M..IS.LLS.C..DIIRSVIN.S.AI.P.DKF..S..E........L.--A.....R...M.......N.....M 164
5 TQSLAC.FGDMCCW.V..WHAIVRIF..TYIKPVLF.P.AIVP.DKV..S.........L.--A.E...R...M.......N.....M 164
6 LTQLVA.YPQGFVR.L..IT.IIN.C..DIVRSVINTSDAITD.DIV..KT.........L.--V....RH...L.......N.....I 164
7 LTQLVT.YPQGFKL.L..TF.LLI.C..DIIRPITS.S.AV.P.DMI..G.........L.--G.....R...M.......N.....M 164
8 STQMSA.YSQGFTV.L..II.FIV.C..DTIRSITN.S.AI.P.DNL.IR.......E.I.L.--G.....R...M.......N...S.I 164
9 ITHLGHDFHDVHLS.V..VY.ILR.V..NFIAPLLQ.S.AV.P.EMTL.G.........M.--A.....H...L.......D...S.V 164
10 VTNLVA.YPQGFMT.L..II.IIT.C..DIIRSVLN.S.SV.L.EVV..S.........L.--D.....S...M.......N.....V 169
11 VTEMGQ.FQDVHLF.L..VI.VLRIVD.AFVAPLLQ.P.LV.P.DMT.LR.........FL.--S.....R...L.......D.....V 164
12 IQSLVR.FRDACCW.V..LH.VIRIF..AFIKPVVL.P.LV.P.DTV..R.........M.--T.....R...M.......N.....V 164
13 LTRLVG.YPQGFLM.I..MV.VIT.C.SDIVRSILN.S.AV.L.DVI..TI.........V.--A.....R...M.......N.....V 164
14 VTKLVA.YPQGFMT.L..IL.IIT.C..DVIRSVLS.S.SV.L.EVI..S.........L.--D.....C...M.......N.....V 164
15 VTEMGQ.FRDIHLC.L..VI.VLR.VD.AFVAPLLQ.P.LV.P.DTT.LR.........FL.--S.....R...L.......D.....V 164
16 IAHLGRNFRDIHLS.V..VY.ILR.V..NVIAPLLQ.S.AV.P.EMTL.G.........M.--A.E..NH...L.......D...S.V 164
17 EKKVLDNMHHVLLV.M..VL.LLV.V..DYIKPLLG.S.AI.P.DDL..G.........L.--K.....R...L.......D.....M 172
18 EKKVLD.MHHIILA.V..FL.LLV.V..DYIKPVLG.S.AI.P.DEF..S.........I.--K.N...R...L.......D.....M 172


KIFN S NIMHAKW L GS- LDMFEHISLMTLDSLQKC FSF--DS CQE SEYI AILELS L KR D LY LT
1 AMMVDIAVQLVQ..ER.NADEH---IEVP.DMTRL...TIGL.G.NYRFN.FYRDQPHPF.TSMVRALDEAMNKLQRANP-D.PA.--DE 201
2 A..AD.TR..LE..EKKACE.K-S-F.I.SDVGH.A..T.M..T.GKG-..GLNHRD.S.YV.VS..TL.MQQ.IDSFQYHN.FI.W..P 236
3 ....D.T.......QR.ISD..-AR......V..........V.....-..N...KS....A......A.VA..HQQPLMFM.L..N..P 251
4 ....E.V.......QL.ASE..-AC..................V.....-..H...KP....A......A.VS..HHEILLHI.F..Y..P 251
5 ....E.V.......QL.ASE..-AR..................V.....-..H...KP....A......A.VT..HQQILLYI.F..Y..P 251
6 ...SK.A.......QR.AME..-TC..V...............I.....-..N...KP....T..M...A.VV..NNQFFRYK.F..F..P 251
7 ....K.V....D..QR.ASE..-AR..................V.....E.N...KP....A......AFVE..NQQILLHT.F..Y..P 251
8 T...K.A...LD..QH.ASE..-SR..................I.....-..H...RP....AT.....A.VE..SQHILQHM.F..Y.SH 251
9 ....K.V.......QC.ASK.T-SR..................I..V--..N...SD.K..A......S.VV..HRQPFLYL.L..Y..A 251
10 ....D.T.......QH.ASG..-AR..V..N............V.....-..N...NP....S......A.VT..YHQLLLHI.S..Q..C 256
11 ....Q.V.......KH.SSE..-AR.E................L.G.---..N...SP....S......S.II..SLQLFLFV.F..YH.A 251
12 .V..D.T.......QR.ASK..-AY.N................V.....-..N...KP....T......T.VAR.HQRLLLHV.LF.Y..H 251
13 ....D.T.......KR.ISE..-SR......V...........V.....-..N...KS....A......A.VA..HQQPLLFM.L..N..P 251
14 ....D.T.......QD.ASG..-AR....KN............V.....-..N...KP....S......A.VA..YQQLLLHT.S..Q..H 251
15 ....Q.V.......KH.CLE..-AR.E...N............L.G.---..N...SP....S......S.II..SQQLFLYL.F..YR.A 251
16 ....K.V.T.....QR.TAK..-AR..................I.....-..N...SN....A......S.IV..QRQPFLYL.F..Y..A 251
17 ....Q.AD......RH.AEGSA-VS..................V..Y--N.N...KM.D..S..I...A.SVR.QYRLHHYL.FI.YRSA 259
18 ....QCT.......RRHLAE..VTSF..................V..Y--N.D...RM.D..SS.I...A.VVR.QYRLHHYL.FM.Y..A 260


DGXRFXXACXXVHXFTDAVIXERRXXLXXQGXXXXLXXKX--KXKTLDFIDVLLLXKD-EXGKXLSDEDIRAEADTFMFXGHDTTASGLS
1 NKRQ.QEDIKVMNDLV.KI.AD.KAS----------GEQSD.LLTHM.NG..P.T.EP.D..N..YQII..LIA..E..SGL.. 275
2 H.R..LR..RAA.DH..R..RQ.KAA.QDEKEREKIQ-----NRRH...L.I..DVRG-.S.VQ...T.L...V.....E.....T..I. 320
3 ..M..RK..NV..E.....R..HRT.PD..LDDF.KS.A--.S..........S...-.D..E.............E.......... 338
4 ..Q..RR..RL..D.....Q...RT.PS..VDDF.QA.A--.S..........S...-.D..K.............E.......... 338
5 ..Q..RR..RL..D.....Q...RT.PS..VDDF.QA.A--.S..........S...-.D..K.............E.......... 338
6 C.R..HR..RL..D.....Q...RT.TS..VDDF.QA.A--.S..........SE.-KN..E.............G.......... 338
7 ..Q..RR..HL..D.....Q...CT.PT..IDDF.KN.A--.S..........S...-.D..E.............E.......... 338
8 ..R..HR..RL..D.....R...RT.PT..IDDFFKD.A--.S..........S...-.D..A.............G.......... 338
9 ..R..RK..DL..N.....K...ST.NT..V-EF.KA.A--.T.........MAE.-.H..G..N...........G.....T.A.. 337
10 S.R..HK..HL..S.....QD..RT.PSKHEDDV.KA.A--.S..........S...-.D..E.............E.......... 343
11 ..R..RK..DL..N.....R...HT.SS.NHDEF.KS.T--.S..........A...-.H..E.............G.......A.. 338
12 ..M..RK..RL..D.....R...RT.LD..GVDV.KA.A--.A..........S...-.H..A.............G.......... 338
13 ..M..HK..NL..E.....R...RT.PD..LDEF.KS.A--.S..........T...-.D..E.............E.......... 338
14 N.R..HK..KL..N.....QG..RA.PS.HEDDI.KA.A--RS..........T...-.D..E.............E.......... 338
15 ..R..RK..DL..N.....R...RL.SS..TDEF.ES.TKS.S..........A...-.H..E.............G.......A.. 340
16 ..R..RK..DV..N.....R...ST.NT..VDEF.KARA--.T..........A...-.H..G...V..........G.......A.. 338
17 ..R..RQ..DM..H..TE..Q...RA.RQ..AEAW.KA.---QG..........AR.-.D..E.............E.....S..I. 345
18 ..R..RQ..DT..N..TE..Q...QA.RQ..AEAW.KA.---QG..........A...-.E..E.............E.....S.... 346


W X LYNLA X HPEYQERCRQEV X ELL X DRE X X EIE-------WDDLAQLPFLTMCIKESLRLHPPV X X X SR X CTQDI X LPDGR X IPKG X IC X
1 FA..F.VKN.HVLQKAAE.AARV.V.PVPS----------YKQVK..KYVG.VLN.A...W.TAPAF.LYAKE.TV.GGEYPLE..DELM 355
2 .F..CM.LY..H.Q...E..R.I.G.QDSFQ----------.E...KMTY....M..CF..Y...PQVY.QLSKPVSFV...SL.A.SLIS 401
3 .I.....R.....Q...RG..PE...-------.....TVI..C.....L.....T....I..L 421
4 .V..H..K.....Q...K...PK...----------.....H.....M.....PVI..HV....V.....V....I..L 421
5 .V..H..K.....Q...K...PK...-------.....PAV..C.....V.....V....I..L 421
6 .V.....R.....Q...K...PK...--------------.....L.....IPTFA.G....VV...S.V....NV.N 421
7 .V..H..K.....Q.....Q...K...PI...-------.....PVI..C....FV.....V....IV.L 421
8 .V.....R.....Q...K..DPK...-------.....V.....APFI..C.....V.....V....IT.L 421
9 .I.....R.....Q...R..DSE...-------.....LLI..C....VL.....A....N..V 420
10 .I.....R.....Q...R...ST...CSCAVFLR.....TVI..R.....V.....V....V..I 433
11 .I.....R.....Q...R...PE...-------.....S..IDLL.R..R..V.....V....N..V 421
12 .I.....R.....R...R...PE...-------.....TAI..C.....V.....V....V.SR 421
13 .I.....ND.....Q...R..DPE...-------.....TVI..C.....L.....T....I..L 421
14 .I.....R.....R...R...ST...-------.....TVI..R.....V.....V....V..I 421
15 .I.....R.....W...R...PE...-------.....AIDLL.R.....V.....V....N..V 423
16 .I.....R.....R...R...PE...-------.....LLI..C.S...V.....V....N..V 421
17 .M.F...KY.....K..E.IQ.VMKG..LE.L.----------....T....T.....QY...TLV..Q..E..K.....I....I..L 428
18 .A.F...KY.....K..E.IQ.VMKG..LE.LD--------....T....T.....QF...TLI..R..E..K.....V....I..L 429


I X I X G X HHNP X VWPDP-EVY X PFRFDPEN X X X RSPLAFIPFSAGPRNCIGQ X FAM X EMKV X LALTLLRFR X ----------
1 VL.PQL.RDKTI.G.DV.EFR.E..--..PSAIPQH..K..GN.Q.A....Q..LH.ATLV.GMM.KH.DFEDHTNYELDIKETLTLKPE 443
2 LH.YAL.R.SD.....-..FD.L..S...SSG.H.Y.............Q...N....VT..C....EF---------- 471
3 .S.F.I....S.....-...D........IKDS.................T...S....A.........L---------- 491
4 .SVF.T....A.....-...D........IKE..................T...A....V.........V---------- 491
5 .SVF.T....A.....-...D......K.IKE..................A...A....V.G.......V---------- 491
6 .N.FAI....S.....-...D........AQK...M.............K...A....V.........I---------- 491
7 .N.I.I.Y..T.....-...D.....Q..IKE..................A...A....V......H..I---------- 491
8 .D.I.V....T.....-...D........SKG..................A...A....V...M..H..F---------- 491
9 .S.F.V....S.....-...N........PQK.............T.....T...S.I..A.........I---------- 490
10 .N.FAT....T.....-...D........IKD..................T...N....A.........V---------- 503
11 .S.F.I....S.....-...D........PQK..................T...S....A.........I---------- 491
12 .S.F.T....A.....-...D.....AD.VKG..................T...S....A.........V---------- 491
13 .S.F.I....S.....-...N........IKDS.................T...S....A.........L---------- 491
14 .N.FAT....T.....-...D........IKD..................T...N....A.........V---------- 491
15 .S.F.I....S.....-..FD.....S..RQK....S.............T...N....VV........V---------- 493
16 .S.F.V....S.....-...N........PQK..................T...S.I..A........CV---------- 491
17 VS.Y.T....T....S-K..N.Y....D.PQQ.....YV...........S...A.LR.VV........L---------- 498
18 VS.Y.T....I....S-K..N.Y....DTPQQ.....V............S...A..R.VV........L---------- 499

APPENDIX D

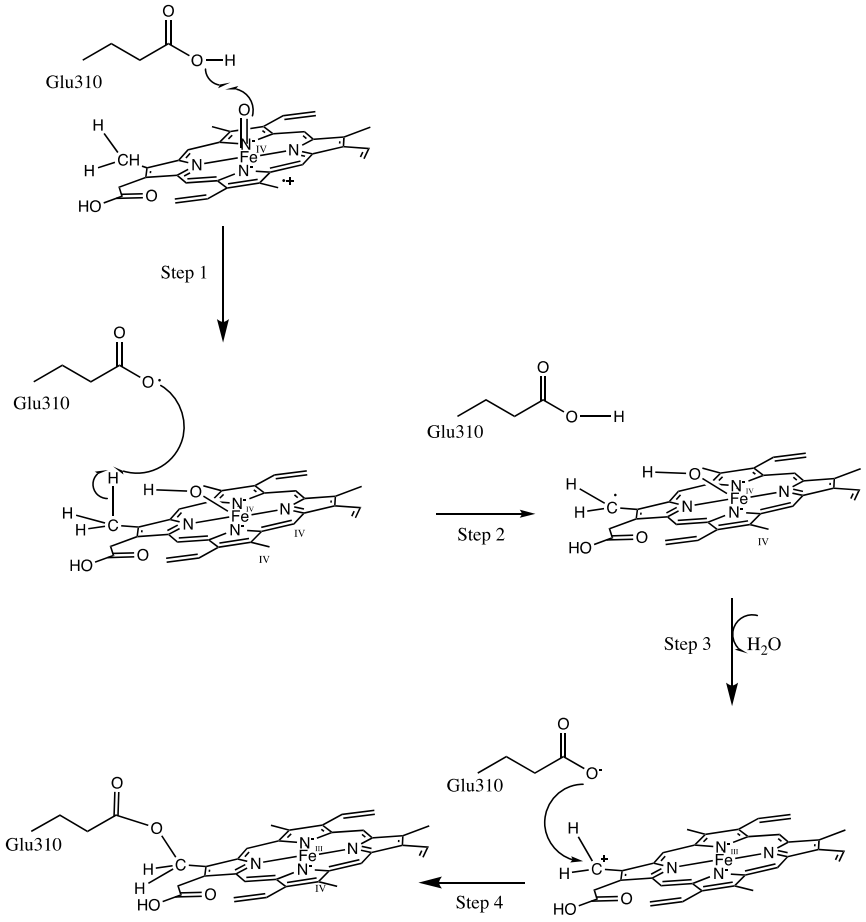


Figure D1. Formation of Covalent Glu310-heme Linkage in CYP4B1. In this mechanism, first a hydrogen was abstracted by highly electrophilic iron-oxo complex leaving a radical on Glu310. The Glu310 radical then abstract a hydrogen from C-5 methyl group and finally water leaves forming the ester linkage.

APPENDIX E

The following script was used in Python to carry out the data analyses. The gcmsanalyze and ms_reference code libraries are available upon request from Dr. Donovan C. Haines, haines@shsu.edu, or from GitHub.

```
import os
import matplotlib.pyplot as plt
import matplotlib.colors as colors
import pandas as pd
path_to_python32 =(r"/Applications/Python
3.8/NASA/PythonNASA")
os.chdir(r"/Applications/Python
3.8/NASA/PythonNASA/gcmsanalyze/Analzye gcms/Analzye gcms")
import gcmsanalyze
import ms_reference
os.chdir(r"/Applications/Python 3.8/NASA/PythonNASA/AMDIS
Analysis")
reference_library =
ms_reference.GCMSReferenceDatabase("nist/")
reference_library.load_retention("retentions.csv")
gcms_col =
gcmsanalyze.GCMSFileCollection(files_to_process="filestopro
cess.csv", verb=150)
gcms_col.attach_reference(reference_library)
colors_temp = colors=gcms_col.files_parameters['color']
colors_temp[56]="black"
colors_temp[59]="lightgreen"
colors_temp[60]="green"
colors_temp[61]="lightblue"
colors_temp[62]="blue"
colors_temp[63]="salmon" #lightred
colors_temp[64]="red"
colors_temp[65]="maroon"

gcms_col.plot_eic_all(spectra=[56,59,60,61,62,63,64],ionlis
t={34}, time1=0.0, time2=2.0, colors=colors_temp)
gcms_col.plot_eic_all(spectra=[56,59,60,61,62,63,64],ionlis
t={48}, time1=0.0, time2=2.0, colors=colors_temp)


#plot refs and mass spectra
import numpy as np
def plot_ms(spectrum=0, time=1.0, mass_low=0,
mass_high=1000, width=0.8, label_top=False,
label_number=10, label_size=6, label_dist=0.03,
```

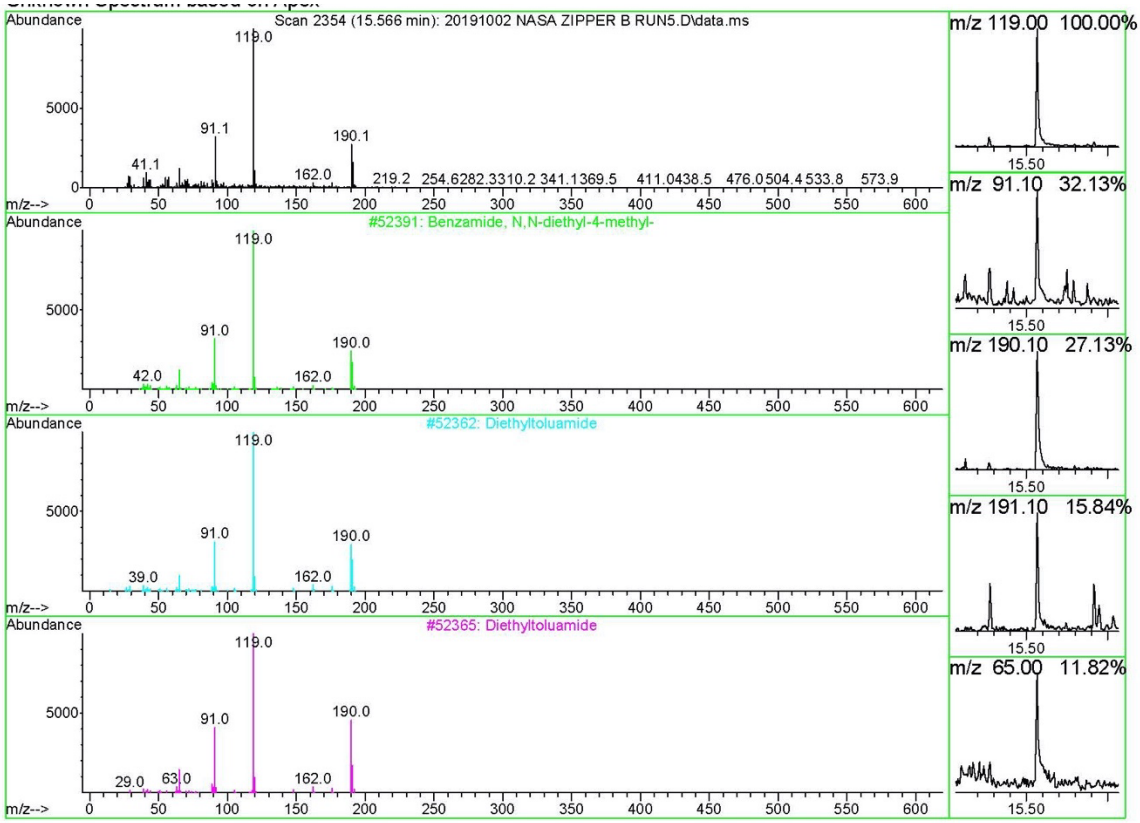
```
label_rot='vertical', subtract_time=-100,
is_reference=False):
     time_index = np.abs(gcms_col.data[spectrum].times-
time).argmin() #find the index of this time entry
     #print(time_index)
     #now get the intensities
     if not is_reference:
         temp_masses= gcms_col.data[spectrum].masses
         temp_int = np.array([ max(0,i) for i in
gcms_col.data[spectrum].intensity[time_index, :] ])
         if subtract_time>=0:
             print("Subtracting time 2 from time 1:")
             print(time)
             print(subtract_time)
             time_index_bg =
np.abs(gcms_col.data[spectrum].times-
subtract_time).argmin() #find the index of this time entry
             temp_int_bg = np.array([ max(0,i) for i in
gcms_col.data[spectrum].intensity[time_index_bg, :] ])
             temp_int=temp_int-temp_int_bg
     if is_reference:
         if not gcms_col.reference_loaded:
             print("Error in plot_ms, reference spectrum
requested but no reference database attached.")
         print("Using reference spectrum for ")

print(gcms_col.reference_database.list_names()['Name'][spec
trum])

temp_masses=gcms_col.reference_database.database[spectrum][
'x']

temp_int=gcms_col.reference_database.database[spectrum]['y'
]
     #print(temp_int)
     #plot as a barplot
     plt.bar(temp_masses, temp_int, width=0.8, )
     plt.xlabel("Intensity")
     plt.ylabel("m/z")
     plt.axis([mass_low, mass_high,0,max(temp_int)*(1.0 +
max(0.05, label_dist*2))])

     if label_top:

points_to_label=gcms_col.identify_labels(temp_masses,
temp_int, max_answers=label_number)
         for label_count in range(0,len(points_to_label)):
```

```
plt.text(temp_masses[points_to_label[label_count]],

temp_int[points_to_label[label_count]]*(1.0+label_dist),

r'{0:.0f}'.format(temp_masses[points_to_label[label_count]]
), rotation=label_rot,
            horizontalalignment='center',
verticalalignment='bottom', fontsize=label_size)

     plt.show()


#plot the MS Spectra
#plot_ms(spectrum=21,time=0.929, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3,
subtract_time=0.914)
#plot_ms(spectrum=46,time=1.01, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3, subtract_time=0.99)
#plot_ms(spectrum=36,time=0.929, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3,
subtract_time=0.914)
#plot_ms(spectrum=47,time=1.01, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3, subtract_time=0.99)
#plot_ms(spectrum=45,time=0.929, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3,
subtract_time=0.914)
#plot_ms(spectrum=48,time=1.01, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3, subtract_time=0.99)
#plot_ms(spectrum=65,time=0.929, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3,
subtract_time=0.914)
#plot_ms(spectrum=49,time=1.01, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3, subtract_time=0.99)
#plot_ms(spectrum=55,time=0.929, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3,
subtract_time=0.914)
```

```
#plot_ms(spectrum=50,time=1.01, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3, subtract_time=0.99)
#plot_ms(spectrum=64,time=0.929, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3,
subtract_time=0.914)
#plot_ms(spectrum=51,time=1.01, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3, subtract_time=0.99)
#plot_ms(spectrum=65,time=0.929, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3,
subtract_time=0.914)
plot_ms(spectrum=33,time=1.01, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3, subtract_time=0.99)
plot_ms(spectrum=34,time=1.01, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3, subtract_time=0.99)
plot_ms(spectrum=35,time=1.01, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3, subtract_time=0.99)
plot_ms(spectrum=36,time=1.01, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3, subtract_time=0.99)
plot_ms(spectrum=62,time=1.01, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3, subtract_time=0.99)
plot_ms(spectrum=63,time=1.01, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3, subtract_time=0.99)
plot_ms(spectrum=64,time=1.01, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3, subtract_time=0.99)
plot_ms(spectrum=65,time=1.01, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3, subtract_time=0.99)
#plot_ms(spectrum=34,time=1.01, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3, subtract_time=0.99)
#plot_ms(spectrum=35,time=1.01, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3, subtract_time=0.99)
#plot_ms(spectrum=36,time=1.01, mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3, subtract_time=0.99)
```

```
#plot the standards
#plot_ms(spectrum=26,mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3, is_reference=True)
     #plot_ms(spectrum=29,mass_low=20,mass_high=100,
label_top=True, label_size=12, label_dist=0.05,
label_rot='horizontal', label_number=3, is_reference=True)
```

APPENDIX F



Data File: C:\MSDCHEM\1\DATA\DCH\NASA\20191002 NASA ZIPPER B RUN5.D
Sample :

Peak Number: 183 at 15.566 min Area: 1103451 Area % 1.98

The 3 best hits from each library.	Ref\#	CAS\#	Qual
C:\Database\HAINES.L No hits were retrieved.			
C:\Database\NIST08.L			
1 Benzamide, N,N-diethyl-4-methyl-	52391	002728-05-4	94
2 Diethyltoluamide	52362	000134-62-3	91
3 Diethyltoluamide	52365	000134-62-3	91

Figure F1. NIST Database Reference Spectra of DEET.

VITA

Madushika Madri Jayakody
Graduate Teaching Assistant
Department of Chemistry
College of Sciences and Engineering Technology
Sam Houston State University

Academic Training

Master of Science, Chemistry, Sam Houston State University
2018- Present

Master of Science, Molecular and Applied Microbiology , University of Peradeniya
2016-2018

Bachelor of Science. Molecular Biology and Biotechnology *(Honors)* University of Peradeniya
2011-2015

Research

Masters Researcher, Sam Houston State University
2018-present

- Analysis of volatile organic compounds using solid phase microextraction. (A project done for NASA)
- Constructed an *E. coli* expression system for mouse Leukotriene B4 ***ω***-Hydroxylase (Cyp4f14)
- Used PyMOL software to study the active site of the mouse Leukotriene B4 ***ω***-Hydroxylase and
  Auto dock software to study its substrate binding
- Analysis of sequence alignments and preliminary work with homology models
- Analyzed Cyp4f14 using JASCO UV-Vis spectrophotometer and VarioskanFlash for its spectrometric properties
- Carried out enzyme kinetics experiments using chromogenic substrate *p*NCA

Techniques and skills: Gas Chromatography-Mass Spectrometer (GC/MS), UV-Vis Spectrophotometer, and VarioskanFlash, Python (Programming Language), Auto dock , Gaussian-G09W, HsPiP and POV-Ray

Research Assistant, University of Peradeniya

2015-2016

- Developed and used a modified CTAB method to remove the nuisance inhibitors in the DNA extracted from papaya seeds.
- Used *psbA-trnH* primer pair for PCR amplification for DNA barcoding

Techniques: PCR, multiplex PCR, DNA purification

Undergraduate Researcher, University of Peradeniya
2013-2015

- Isolated marine bacteria associated with *Ulva Sp* and *Caulerpa Sp.* and checked for their secondary metabolites against common human pathogens for antibacterial activity
- Extracted DNA from positive samples and amplified *16srRNA* gene using PCR technique.
- Sequenced the PCR products using ABI 3500 genetic analyzer

Techniques: PCR, DNA Sequencing

Oral Presentation

- Structural Analysis of Mouse Leukotriene B4 ***ω***-Hydroxylase Cyp4f14, **Madri M Jayakody**Donovan C Haines. Texas Academy of Sciences Annual Meeting, Nacogdoches, TX, Feb 2020.

Poster Presentation

- Expression of Mouse Leukotriene B4 ***ω***-Hydroxylase Cyp4f14 in *E. coli* and Structural Analysis of its Active Site. **Madri M. Jayakody** and Donovan C. Haines. Department of Chemistry, Sam Houston State University. Southwest Regional Meeting of the American Chemical Society, El Paso, TX, November 2019.

- Expression of Cyp4f39 in *E. coli* and Structural analysis of its active site, **Madri M Jayakody**,Donovan C Haines. Texas Academy of Sciences Annual Meeting, Nacogdoches, TX, Feb 2020.

- **Jayakody MM,** Samaraweera P (2015) Screening of Antibacterial Activity of Marine Bacteria Associated with Green Algae *Ulva Sp* And *Caulerpa Sp.* Undergraduate Research Poster Presentation 2015, University of Peradeniya.

Publications

- Dissanayake DRRP, Herath HMPD, Dissanayake MDMIM, Chamikara MDM, **Jayakody MM,** Amarasekara SSC, Kularathna KWTR, Karannagoda NNH, Ishan M, Sooriyathirana SDSS (2015) The length polymorphism of the locus *psbA-trnH* is idyllic to detect adulterations of black pepper with papaya seeds and chili. *Sabaragamuwa Agri Journal, Sabaragamuwa University of Sri Lanka*.

Professional Experience

- **Graduate Teaching Assistant** (Sep 1, 2018- Present), Department of Chemistry, College of Science
  & Engineering Technology, Sam Houston State University, Huntsville, TX. USA.

  Courses: Instrumental Analytical Chemistry, Quantitative Chemical Analysis, General Chemistry 2, Inorganic and Environmental Chemistry (Tutoring), Introductory Organic and Biochemistry (Lab and Tutoring)

  Techniques: UV-Vis Spectrophotometer, Hitachi F-4500 Fluorescence Spectrophotometer, Varian 220FS AAS, Inductively Coupled Plasma-Atomic Emission Spectrometry, Gas Chromatography FID, Gas Chromatography MS (GC/MS)

- **Graduate Teaching and Research Assistant** (Jan 19, 2015- Nov 7, 2015), Department of Molecular Biology and Biotechnology, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka.

  Courses: Recombinant DNA Technology, Biochemistry Laboratory II,

  Techniques: primers design, digestion, ligation, cloning. Western blot, Southern blot, Dot blot, ELISA,enzymatic assays, Cloning: transformation, RNA extraction, reverse transcription, PCR, mutiplex PCR, RT-PCR, DNA purification (CsCl gradient, mini- and maxi-preps), DNA gel analysis, Quantification and extraction, Analytical methods: Chromatography (thin layer, ion exchange), titrations, spectrophotometry

- **Research Assistant** (2016 – 2017), Department of Anatomy, Faculty of Medicine, University of Sri Jayawardhanapura, Colombo, Sri Lanka.

  Tasks: Writing for grants, supporting administrative work. Chemical/consumable/equipment purchasing.