MODELING AND MOLECULAR DYNAMICS SIMULATIONS ON THE IN SITU MURINE CYTOCHROME P450 4F SYSTEM

_____

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 Science

_____

by

Jerome W. Butler

August, 2020

MODELING AND MOLECULAR DYNAMICS SIMULATIONS ON THE IN SITU MURINE CYTOCHROME P450 4F SYSTEM

by

Jerome W. Butler

_____

APPROVED:

Donovan Haines, PhD
Thesis Director

Darren Williams, PhD
Committee Member

Christopher Zall, PhD
Committee Member

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

DEDICATION

This thesis is dedicated to all of those who have shaped me into the person I have become. It is dedicated to all the lessons given and received in the lazy afternoons spent in the usual spot for my fellow chemistry graduate students.

It is dedicated to all the mornings of training spent advancing my journey into chemistry, as well as martial arts and to all the evenings spent sharing my enjoyment of the biochemical cosmos with less-than-impressed undergraduate students.

It is dedicated to all the dissolved problems that melted away when entering the office of Dr. Donovan Haines and to the relief I experienced on the phone when talking with my family, located across the country.

It is dedicated to all the encouragement, words of wisdom, and advice given to me from my family, friends, and research idols. To the written words of the past generations of scientists, researchers, and thinkers, as their shoulders are those of which I surely stand upon.

Finally, I would like to distinguish one last time before presenting the work in this thesis, I dedicate my effort to my Mother and Father

Everything I do or ever will do, will be done for all of humankind.

ABSTRACT

Butler, Jerome W., *Modeling and Molecular Dynamics Simulations on the in situ Murine Cytochrome P450 4F System* . Master of Science (Chemistry), August, 2020, Sam Houston State University, Huntsville, Texas.

Cytochrome P450s are major participants in the maintenance and well-being of cellular function and have important roles in the health and disease of living creatures. The ω-hydroxylation, catalyzed by CYP4 family members, has been observed to be an important metabolic pathway for the homeostasis of mammalian cells as it regulates inflammatory processes with the eicosanoid cascade of metabolites of the ω-6 polyunsaturated fatty acid, arachidonic acid. Many human CYP4F and murine Cyp4f subfamily members have recently gained interest for their usage as potential cancer biomarkers as the expression of these proteins are modified in tumor cells. 20-HETE, the ω-hydroxylated product of arachidonic acid, has gained attention for being the chief metabolic product of interest in vascular function, tumor progression and propagation. Whether or not individual Cyp4f isoforms are responsible for the production of this metabolite is of great interest to medicine as such insight could provide researchers with new avenues of study in the fight against cancer. One particular Cyp4f isozyme, Cyp4f13, has received relatively little study until only very recently and is the focus of the work presented in this thesis, as it has not fully had its role in eicosanoid metabolism understood. Using a combination of computational chemistry approaches, this study focuses on exploring the murine cytochrome P450 4f13 system and its active site using all-atomistic Molecular Dynamics Simulation of a homology model. With the embedded protein solvated and *in situ* environment replicated, the resting state of the substrate-free Cyp4f13 system was generated. Solvation of the active site was performed to explore the

inner active cavity of the P450 system, with subsequent molecular docking and mutation of active site residues performed in order to gain insight into the interactions present in the protein-substrate complex. Protonation state changes were observed to have significant effects on both protein structure and arachidonate binding through electrostatic interactions. Leu137, Arg237, and Gly327 were modified and displayed drastic effects on predicted regiospecificity on the P450 substrate. With the insights obtained, we hope to further the understanding of murine Cyp4f13-catalyzed ω-hydroxylation of arachidonic acid.

KEY WORDS: Cytochrome P450; CYP4F; Fatty acids; Eicosanoids; AA, Arachidonic Acid; 20-HETE, 20-Hydroxyeicosatetraenoic Acid; Omega-hydroxylases; Inflammation; Cancer; MD, Molecular dynamics; Homology model

ACKNOWLEDGEMENTS

There are far too many people to thank. As I can only fill up so much of this thesis with direct mentions of acknowledgement, those of whom who are not mentioned will be thanked here.

I would like to start off by thanking my parents, Rocio and Jerome Butler both for their contributions in my growth as a young child scientist, even if I was originally set on being a paleontologist. My mother's nursing study material made great feed for my budding curiosity. My father's wrinkled and creased US ARMY-branded periodic table of elements did wonders for my development as well. I thank them both for the LEGO building sets, as they made for an excellent foundation in my intrigue into the building blocks of the universe, atoms. Those LEGO sets were considerably one of their more expensive contributions.

I would like to acknowledge my brother, Joshua, and my sister, Mekayla, for entertaining my imagination as well as being genuine friends. I would also like to acknowledge my girlfriend, Peyton Brent, who has shown me that there is much more to life than proteins, even if proteins are life's favorite way of expression.

I would like to acknowledge my first chemistry teacher in high school, James Sammons, for showing me the chemical consequences of thermite on asphalt. I would also like to acknowledge my English teacher in high school, Beth Anthony, for showing me the spiritual consequences of a good book and storytelling. I would additionally like to acknowledge my martial arts instructor, Jeremiah Talley, for showing me the physical consequences of a good rear naked choke hold.

I would like to acknowledge the chemistry graduate class of 2020 here at Sam Houston State University, as well as the entire Department of Chemistry. Larger institutions could have spelled an early death for my academic career, as their environment is nothing to compare to the close-knit family of our university. I have a high bar to hold my future post-graduate universities to, and I thank every one of the faculty, staff, and custodial members for solidifying my desire to work in such an environment professionally. I would also like to thank all my friends that I have made in my journey across the United States of America, as the proud military brat that I am.

Lastly, I would like to thank Dr. Donovan Haines specifically. Like only the best mentors can, he has provided me with a peace of being; allowing me just enough time to run far away with my ideas, thoughts, worries, and fears before arriving at a moment of clarity, returning with a lesson in hand. I have entered his office door with such a wide net of problems, issues, doubts, and qualms over the years. In every one of those situations, he taught me how to breathe, take a step back, to look at the problem from every view imaginable, find what was able to be changed and then suggest I do so. His incredible expertise and care for his department and his fellow faculty members will surely mark him as among the best professors of chemistry and chairs at Sam Houston State University.

Most importantly, he's the guy who introduced me to the wicked cool biological masterpieces called cytochrome P450s.

PREFACE

As I write this, I am settled in my parked Ford Fusion, Aphelion, aptly named after the science-fictional star cruiser. I sit beside the terabytes of my research data, stored cozily inside of a small hard drive. My faithful gaming laptop, 8 years old and still running like an champ, packed away in its bag.

It has been an absolutely crazy past two years and I feel justified in saying so. As of today, several months have passed since the announcement of a global threat in the form of the COVID-19 pandemic, caused by the SARS-CoV-2 coronavirus in December of 2019. In a time like this, the words of many great past thinkers come to mind. In 1995, Carl Sagan warned against the celebration of ignorance that he had started to observe in America. In an era of misinformation and growing obliviousness, the study of biological systems, as with COVID-19, performed by researchers such as myself and my peers grows in its importance. Not the work we might be performing currently, but our contributions to the world as scientists in the form of new knowledge and understanding to the field of chemistry and biology. Pseudoscience and ignorance cannot have a place in the general public, as it makes for a poor environment for the future scientists of the world. The next guard of researchers, thinkers, and teachers depend on us to do right in these times of uncertainty and terror.

I have grown so much as both a researcher, and as a person in these years of my time at Sam Houston. This thesis has caused me to have a lot of inner reflection as it has certainly run me ragged in the wake of my defense. It seems hard to imagine that only a few weeks ago I stared at a skeletal outline of the document, horrified and heavily discouraged. How to begin? What had I done in the past two years?

Fast forward to this week, where I have logged almost 60 whole hours dedicated to nothing but writing, discussing, and evaluating my results. Seeing it all cobbled together, yet slowly taking shape, I was driven to tearful laughter at one point.

The work presented in this thesis has my own hard-keyed, hunch-backed, sworn and cursed upon, sweat and tears and I couldn't be happier in being able to share it with you. The simulations and calculations I ran were nasty feats; some lasted for weeks and mind you, if I had done something wrong in the creation process, I wouldn't be able to find out until the computational desktop at the Chemistry Department crashed out. Do you have any idea how horrifying it is to crash something worth more than your entire car? It is not pleasant to say the very least.

To further my anxiety, a lot of my work was done through brute-force, trial and error on my part. If something broke in the simulation, it was because of me. I had to figure out what the problem was and chalk up ideas on how to fix it, as resources available online were extremely limited. It is not very useful to just google search, "How to fix W problem, with X protein, in Y force field, using Z program?"

Not to mention the fact that the computer I was using had an entirely different Operating System, Linux, that I had *never* used before in my life. As a result, the hard drive devoted to my studies contains 15 folders: 8 of which were successful workspaces, and you can guess what the rest were.

I must say, however, that it was fun. It was exciting to learn the method as it was everything I had hoped for in learning about computational chemistry from my undergraduate Physical Chemistry course. The research process taught me a lot about myself, certainly that I work better under the pressure, with nothing to compare than that

of graduate school. This past week has been a sort-of ‘writer’s high’ for me as I fervently took to the task to empty my mind, heart, and soul onto the thesis document.

This experience cannot be replaced; it has set me on a path that I am stoked to be a part of. The challenging world of academia, while foreboding in nature, is where I continue to set my sights. With the research I have performed and present in this thesis, it has opened so many doors that seemed like unimaginable only so long ago. Having been accepted into the Ph.D. program for Medicinal Chemistry at the University of Washington, I have had an opportunity to lend my skills in molecular docking and enzymology as I was able to work in collaboration with UW in the testing of drug candidates to aide in the fight against the current COVID-19 outbreak.

I am excited to see what the future has in store for me, so without further ado, here is the story of my two-year long adventure with my painstakingly crafted virtual cytochrome P450, Cyp4f13.

Hobey-ho, let’s go.

In an empty parking lot on a damp Sunday evening

May 17, 2020

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

CHAPTER I

Introduction

Cytochrome P450

General Background and Nomenclature

The cytochrome P450 (CYP, P450) family of monooxygenases are the enzymatic powerhouses accountable for a large percentage of molecular biotransformation in the human body.[1–3] These enzyme systems serve as excellent biological model proteins to study and are responsible for a wide variety of physiological roles in living species; from drug metabolism, to hormone synthesis and cellular signaling.[4] Modern biochemistry, biophysics and molecular pharmacology have extensively studied these systems for their attractive potentials as therapeutic targets due to their dominance in human drug biotransformation.[5]

Eukaryotic P450s exist mostly as membrane-bound heme-thiolate monooxygenases that increase the rate of oxidation of a variety of molecular targets such as non-polar compounds, bioactive or inactive pharmaceuticals, and other noxious xenobiotics.[6] They primarily behave as monooxygenase enzymes and, through the activation of molecular oxygen, are tasked with modification of molecular functional groups in attempt to aid a biological organism in the clearance of potentially toxic compounds.[7] The P450 enzyme system is usually comprised of several components: the heme-thiolate catalytic domain, a hydrophobic N-terminal anchor, and an electron source in the form of flavin-bound reductase domains with their own membrane anchor with which these proteins are found affixed to a phospholipid bilayer.[8]

The name ‘Cytochrome P450’ first appeared in 1962 as the biomolecule was first characterized as being a novel “microsomal carbon-monoxide binding pigment”.[9] This was before any physiological functions were known for the protein system and had derived from the characteristic peak shift in optical absorption when carbon monoxide was present from around 420 nm to 450 nm as its name describes.[10] As research expanded surrounding the hemeprotein’s function as an oxidase, a large amount of literature has been devoted to this superfamily. The genes coding for P450 enzymes span all branches of life; their expression can vary between tissue types, developmental stages, and even between individuals within a species in the form of polymorphisms.[11,12]

The nomenclature of cytochrome P450 enzymes was established in order to keep categorical indices of different, yet similar, relatives and clans of the superfamily.[13] The root symbol CYP is what is used to designate the hemeprotein as being a cytochrome P450 enzyme. The case of the designation is related to the species being referred to. The case is completely uppercase for all species (i.e. human CYP3A4) except for murine and fruit fly systems where the designation is in sentence case (i.e. mouse Cyp4f13). This is an artifact of historical significance and has no other implications.[14]

Cytochromes are organized by several characters following the root symbol CYP. The first is a number denoting the family of the individual P450 protein. The second is a letter that denotes the subfamily which might give insight into the function or substrate specificity of individual isoforms. The last portion is another numeric value given to individuals within a subfamily. The designation of CYP3A4 refers to the human cytochrome P450 belonging to the 3rd family, ‘A’ subfamily, 4th individual P450 discovered within that subfamily. The naming convention is useful in comparisons

between different isoforms within and between species in the use of understanding substrate specificity, function and physiological role for each P450 enzyme.

Mechanism of the P450 Catalytic Heme Domain

The catalytic domain of cytochrome P450s houses the heme-thiolate prosthetic group and is covalently bound by a conserved cysteinyl thiolate ligand as opposed to the imidazolyl nitrogen as found in hemoglobin and other cytochromes.[15]

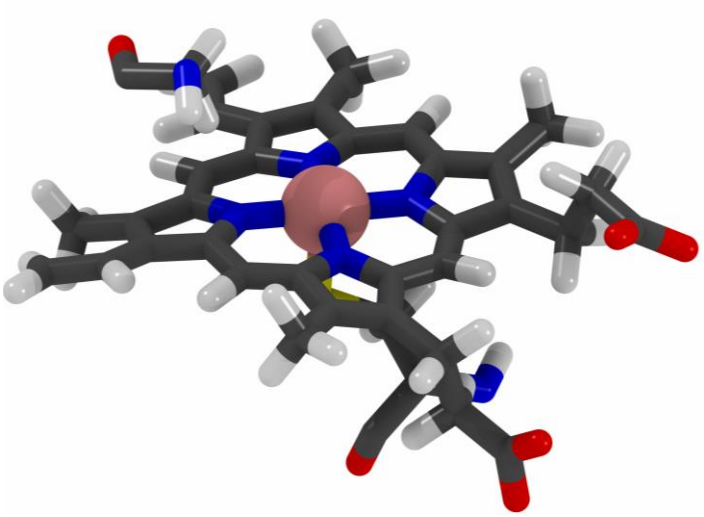

Figure 1. Licorice representation of the heme-thiolate cofactor in cytochrome P450 enzymes. Slightly obscured is the thiolate cysteine residue that serves as the distal axial ligand to the heme-iron center.

This different electronic environment, in collaboration with the conjugated pi network of the surrounding porphyrin ring system, gives the heme-iron an interesting variety of redox chemistry it can perform. Neighboring amino acid residues and backbone elements serve to stabilize protein structure and dictate substrate specificities between P450 isoforms.[15] The only conserved residue between all cytochrome P450 proteins is the heme-thiolate cysteine while common motifs of amino acid residues are conserved within familial clans.[16,17] After extensive research on the active site of more

highly studied P450s, the catalytic site has been observed to have a large amount of substrate plasticity, meaning it is to adapt to bind different molecular structures when certain conditions are met. Individual residues play important roles and vary widely between individual isozymes, even within the same subfamily.[1,11] It has become clear through site-directed mutagenesis and experimentation that the residue makeup of the active site in these systems is key in substrate binding, catalytic function, and successful enzymatic turnover.[18,19]

In the P450 reaction mechanism, the activation of oxygen is required in order to insert a single oxygen atom into a C-H bond, leading to the designation of monooxygenase. The source of the oxygen is atmospheric dioxygen, and results in the reduction of the remaining oxygen atom into water.[20] In the case of microsomal P450s, the monooxygenase reaction catalyzed by the heme domain of cytochrome P450 systems requires an auxiliary electron source usually in the form of membrane-bound flavin-containing reductase units. These cytochrome P450 reductase (CPR) proteins have two separate flavin domains, one containing flavin adenine dinucleotide (FAD), and another containing flavin mononucleotide (FMN) cofactors. The reductase takes electrons from the cytosol in the form of the phosphorylated reducing agent, nicotinamide adenine dinucleotide phosphate (NADPH).

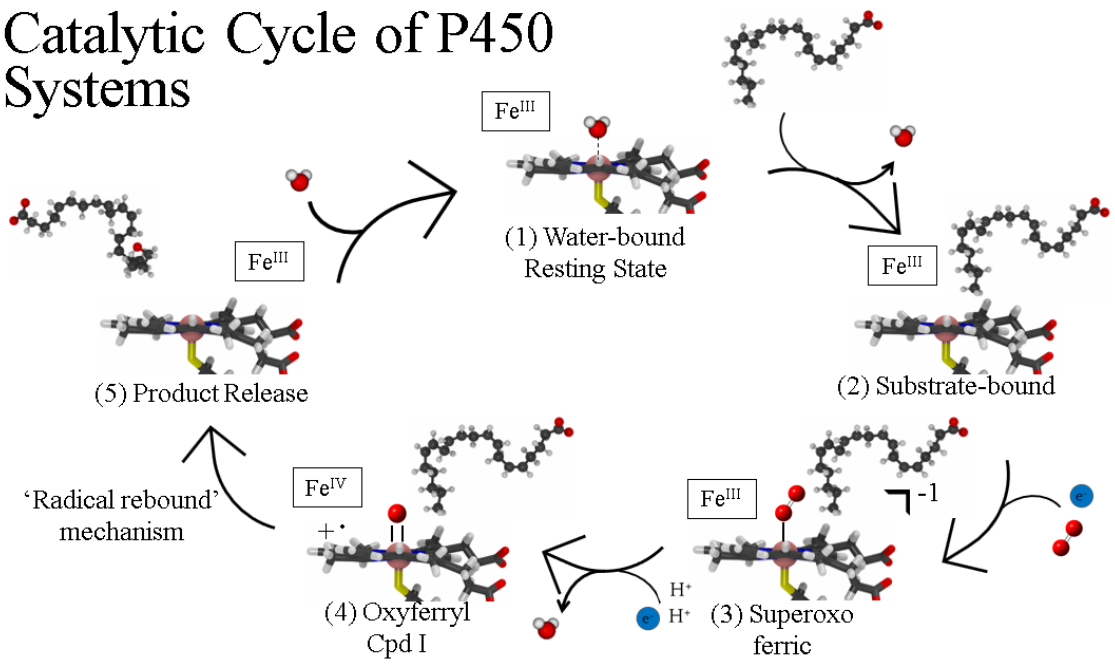

Figure 2. Schematic of the P450 catalytic cycle. The hydroxylation of an aliphatic C-H bond is depicted.

The catalytic cycle of cytochrome P450 monooxygenase system begins with the catalytic heme domain in a resting state on the surface of the endoplasmic reticulum membrane of a liver cell. In this +3 oxidized resting state, as seen in the top of Figure 2, the heme-thiolate iron is coordinated by an axial water molecule trans to the proximal cysteinato ligand. This coordination site is occupied by different water molecules that access and egress from the solvated protein using pathways throughout the folded tertiary structure. These 'channels' are used by small molecules, substrates included, to access the catalytic site of the protein and are unique between isoforms.[21]

Depending on the nature of the substrate and the individual P450, substrate binding is the next step and requires the navigation and positioning of a substrate into the active site via hydrophobic forces, de-solvation effects, and favorable electrostatic interactions.[22] This binding to substrate shifts the absorption peak of the heme-thiolate

metal center by changing the geometrical arrangement of the iron electronic environment. This slight change in configuration has a large effect on the system as a whole, as the hydrogen bonding network of the neighboring helices are disrupted and changed as well. The tugging on the P450 I-helix changes the folded structure and consequently tightens the protein's hold onto the substrate.[22]

The displacement of water from the iron center changes the relative energies of the metal ion orbitals.[19,23,24] The loss of the water from the heme iron results in a five coordinate species that is subsequently easier to reduce as its redox potential shifts from the change in spin state. The removal of a pi basic ligand like water from the low spin (S = 1/2) octahedral resting state of the heme iron results in the lessening of the destabilization of the metal d orbitals which results in the reconfiguration of the electrons of the iron into a high spin state (S = 5/2).

The high spin iron metal center with the substrate nearby, as shown in Figure 2, can now be acted upon by the flavin-bound reductase domain through an electron transfer that reduces the iron(III) to iron(II). What follows is the binding of dioxygen to the open coordination site of the iron center, immediately resulting in yet another change in the electronics of the iron metal center. The resultant 'oxyferrous' complex, more appropriately referred to as an ferric-superoxo species, contains an iron(III) center with a superoxide (O_2^-) ligand. The second electron from the reductase domain is transferred to the iron complex which reduces the superoxo species into a peroxo complex.

The aforementioned hydrogen-bonding network acts as a proton relay that facilitates the next step of the mechanism: the protonation of the peroxide ligand. This protonation results in a hydroperoxyl-ligated iron species that contains a lengthened

oxygen-oxygen bond. The species is protonated again from water molecules in the active pocket and the resulting oxonium intermediate breaks down into the final reduction product, water, leaving behind the ultimate oxidant species: the oxyferryl complex.

Such a complex has garnered a lot of discourse and discussion as formal counting procedures place the iron center in an oxidation state of +5 [25]; however, with such a highly oxidized metal center in the center of delocalized electronic pi system of the porphyrin ligand, this seems unlikely. The intricacies present in the P450 catalytic cycle have been put through thorough investigation using computational methods by Shaik et al.[26] The general consensus among P450 researchers is the occurrence of a ligand-metal transfer of an electron, resulting in an iron(IV) metal center with a radical cation delocalized into the porphyrin system. This was found to be reinforced by theoretical and computational chemistry methods on the system using quantum mechanics approaches.[26]

This oxyferryl complex is extremely electrophilic: in a sense, it serves as a molecular blowtorch ready to oxidize or ‘burn’ nearby substrates or even active site residues in its immediate vicinity. The electrophilic oxo ligand abstracts the nearby hydrogen atom from a substrate (or side chain) leaving behind a radical on the alkyl substrate. In span of femtoseconds, alkyl groups can either rearrange or have the hydroxyl group reattach to the carbon radical, resulting in a hydroxylated product in a process referred to as ‘radical rebound’. The polar product serves as a weakly bound ligand to the iron metal center and is swiftly replaced by a water molecule restoring the enzyme back to its resting state, ready to start the cycle anew.

At multiple steps along this catalytic cycle reside chances for side reactions to occur; potentially harmful reactive oxygen species (ROS) can dissociate from the iron

center, to be lost to the cytosol and wreak havoc on organic structures.[27] Some suggest that the evolutionary development of the superoxide dismutase enzyme family originated in part from the production of reactive oxygen species from cytochromes.

This catalytic cycle generalization cannot sufficiently detail the intricacies that the P450 structure plays during the mechanism. Several highly conserved amino acid residues dictate substrate binding, control selectivity, act as salt bridges, and aide in the activation of dioxygen. These sequence motifs, or residue patterns, are conserved between the clans of the P450 superfamily.[16] Some important conserved sequence motifs between P450s have been discussed and mutations in these regions lead to pronouncedly interesting effects. Some effects observed include enhanced activity, altered substrate specificity, increased protein stability, increased generation of reactive oxygen species, and even novel activity.[19]

Structure of Cytochrome P450s

It is often stated that half of the structure of P450s is conserved between all known crystal structures, however it should be noted that the non-conserved regions show just as much significance.[16,17] The variance allotted by individual mutations have led to the coining of the term “plastic” when describing these regions.[28] This plasticity permits the wide substrate specificity and enhanced environmental tolerance for some isoforms observed across all branches of life.

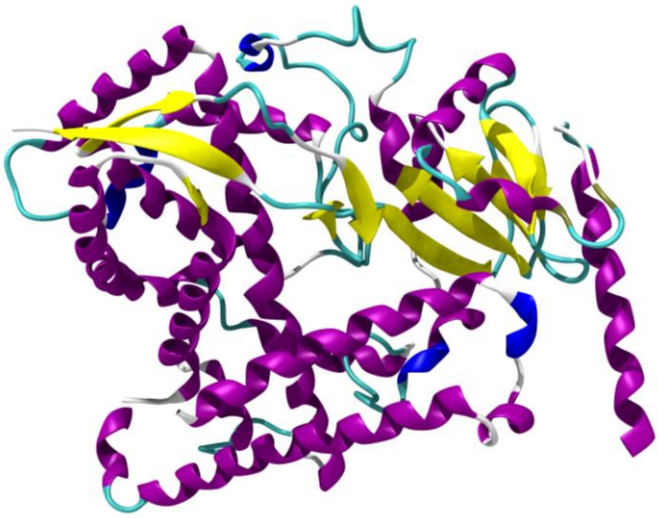

Figure 3. Secondary structure of rabbit CYP4B1 represented by ribbon. Color used to distinguish the individual secondary structural elements: yellow for B-sheets, purple for alpha helices, blue for 3-10 helices.

The structure of eukaryotic P450s consists of a globular heme domain with a transmembrane anchor in the form of an N-terminal alpha helix. The globular heme domain, an example of which shown in Figure 3, is built from several alpha helices and several beta sheets in parallel, antiparallel, or even mixed forms. The transmembrane anchor is comprised of nonpolar residues that allow it to integrate into lipid bilayers. The heme prosthetic group is sandwiched in the hydrophobic core of the protein in a pocket between two alpha helices, with solvated access channels allowing for passage to or from the active site. The type B heme cofactor is not solely an innocent bystander; the propionate "substituents" or groups act as gates for water influx into the active pocket.[19]

One extensively researched system in the study of cytochrome P450 structure is $P450_{BM\text{-}3}$ (CYP102A1) of the bacterium *Bacillus megaterium*.[29] Unlike eukaryotic P450s, the bacterial cytochrome is a self-sufficient protein, capable of efficient electron transfer and high catalytic turnover. This is owed to the backbone linkage of the two flavin-bound

reductase and its heme domain.[30–32] It serves as an excellent representative of the class of enzymes due to its unique fused architecture.[33–35] This allows the system to be entirely soluble and thus not localized to a membrane bilayer. Another unique feature of this bacterial enzyme system is the dimerization that enhances the choreographed catalytic sequence. $P450_{BM-3}$ will combine with another of its kind, tightly placing the FMN binding domain between their partner's heme and FAD domains.[36] This greatly enhances the coupling of the electron transfer between the enzymes' subunits thus leading to a higher amount of successful turnover.

This soluble and self-sufficient nature has led to this system being easier to express as recombinant protein in other bacterial models such as *Escherichia coli* than its eukaryotic counterparts.[23,30,37] This makes it a more attractive alternative than the other monooxygenases of the superfamily. Such ease of use and high catalytic performance has led to study and usage of $P450_{BM-3}$ as its application in protein engineering has garnered a large amount attention from biochemical, biotechnical, medicinal, and even material chemistry fields.[19,38,39] Furthermore, directed evolution on this protein was used as the grounds for pioneering the use of engineered enzymes for catalysis leading to the development of the work that earned the Nobel prize in 2018 in the field of chemistry.[40] It is regarded as one of the most studied enzyme systems and serves as a model system in the understanding and elucidation of structural and function of P450 enzymes.

The structure of $P450_{BM-3}$ has garnered a large amount of attention and is commonly studied through a variety of spectroscopic and mutagenic methods.[19] Several amino acid residues have been singled out for their strategic purpose and function in successful monooxygenation of the P450's substrate.[41–43] Not every side chain residue

can have its role reduced to being that of purely structural; many active site residues have been targeted for their involvement in substrate binding, selectivity and catalytic activity. Within the active pocket of $P450_{BM-3}$, the alcohol-acid pairing between Glu267 and Thr268 behave as organizers for a proton transfer to the activated oxygen of the catalytic cycle.[19] Another noteworthy residue is that of Phe-87, which has been studied extensively due to its impact on substrate oxidation by mutagenesis.[19] Modification of this residue has been observed to shift substrate specificity, making it an important mutation for study in its functional role on modulating P450 activity and catalytic efficiency.

Mutations in $P450_{BM-3}$ on the Phe87 and Ala82 positions have led to their identification as key residues in binding and regioselectivity in fatty acid substrate binding. It is thought that they achieve this by moderating access to the heme center and altering and diversifying the substrate selectivity however the mechanism is still not well understood. Most of the information gained about the influence of these mutations are acquired through X-ray crystallographic methods. These methods depend on considerations of the experimental method, such as solvent tolerance of the protein and other experimental parameters. A theoretical approach might be needed to further explain and analyze experimental observations and the questions raised by crystallographic structures. Molecular Dynamics (MD) simulations serve as an excellent prospective method as they have shown to be powerful approaches when combined with experimental data in investigation of $P450_{BM-3}$ for their function and properties.[44–46]

Localization and Lipid Membrane Topology of P450 Systems

In mammalian cells, cytochrome P450s are located in either the inner membrane of mitochondria or found affixed to the endoplasmic reticulum surface. The globular heme domain protrudes into the cytosol while its N-terminal helix and a portion of its surface remain buried in the membrane of the organelle.

Biomembranes contain, separate, and form barriers between cells and their inner components. The major composition of biological membranes found throughout human tissue cells are the glycerophospholipids. These lipids contain two hydrophobic acyl fatty acid chains, attached to a phosphorylated glycerol hydrophilic "head" group. Due to their amphiphilic nature, containing both polar and nonpolar groups, these compounds readily arrange and form bilayers which give them important roles in biology. A prime example of these compounds found within the endoplasmic reticulum are the di-oleoyl phosphatidylcholine (DOPC) lipids, as it contains a large composition of them (65%) in addition to a variety of other phospholipids, both charged and neutral.[47,48]

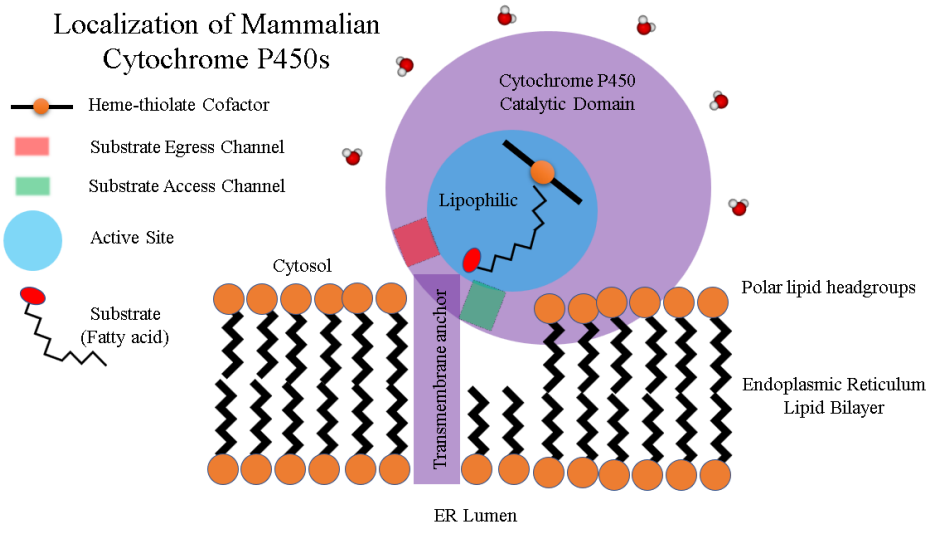

Figure 4. Cartoon depiction of the localization of mammalian cytochrome P450 system. Only the P450 catalytic heme domain and membrane anchor is shown for brevity.

The hydrophobic transmembrane (TM) anchor of eukaryotic P450s is the first aspect to consider when discussing membrane localization, however, it is not the only one. N-terminal TM-truncated cytochrome P450 enzymes have been found associated to membranes, binding even in the absence of the transmembrane helix that normally spans the bilayer.[49–53] This phenomenon was studied by a variety of methods and it is understood that a portion of the protein surface contains several nonpolar loops or secondary structures that assist in the protein in adhering to the lipid bilayer. This results in the slight descent of the heme domain beneath the membrane surface, potentially aiding in substrate access and even optimizing contact with its likewise membrane-bound reductase domains.[54–56]

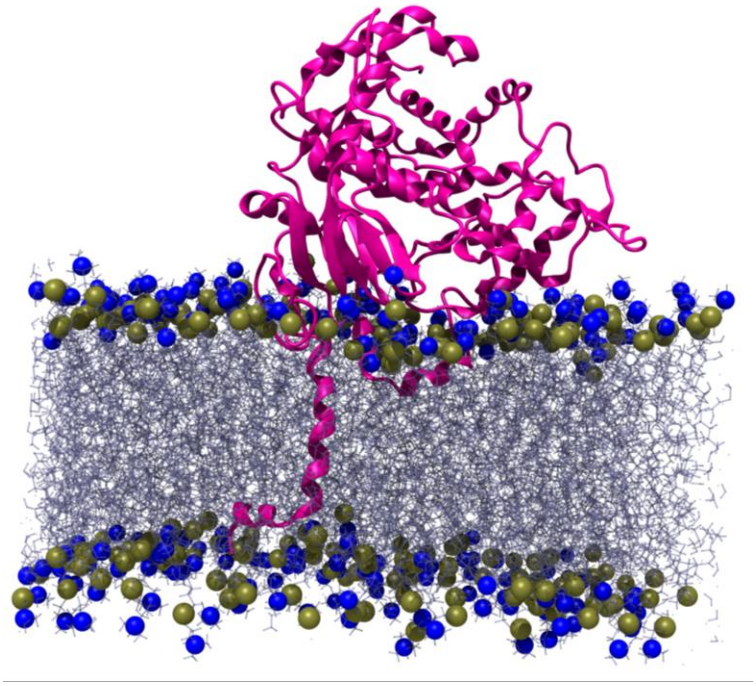

Figure 5. Example of a membrane-localized P450, human CYP3A4 embedded into a phospholipid bilayer. Pink ribbon is used to represent the protein structure. Blue spheres represent the ammonium of the choline groups, and brown for the phosphate groups. Structural coordinates were generated in a the study conducted on CYP3A4 and lipid charge by Otyepka et al.[47]

The membrane bilayer is not innocent in enzymatic catalysis as almost every aspect, from its composure to its charge, shows impact in some way or another on protein function.[57] The composition of the membrane around the P450 system is of interest to researchers as it has been observed to influence catalytic activity for individual isoforms. Different compositions have been observed to have measurable effect on P450 localization and catalysis, such as the net charge and the amount of cholesterol in the membrane.[48]

Lipid composition and membrane orientation of the P450s have been studied using a variety of experimental and theoretical methods.[47,48] Both have concluded that the lipid membrane is not a purely passive environment for the P450 enzyme. Membrane characteristics affect the orientation, localization, ligand binding and even catalytic activity of the cytochrome P450 heme domain.[48] It is even understood that certain membrane interactions facilitate better or worse electron transfer between the catalytic domain and their redox partners due to electrostatic interactions. The order or disorder of a membrane has a standing impact on cytochrome behavior as certain isoform preferences dictate better catalytic efficiency.

While the protein-protein interaction between the cytochrome P450 catalytic domain and the reductase domains are important, they are beyond the scope of this thesis. For further reading, these protein-protein and protein-membrane interactions are discussed at great length by many of the top researchers in the P450 symposium report in 2016.[57] Understanding the interactions of mammalian cytochrome P450s and their membrane environment would provide valuable information in the study of these membrane-bound proteins and their lipid substrates.

Physiological Roles Performed by P450 Superfamily

Chemical Toxicology and Drug Metabolism

The study of chemical toxicology can best be understood as the biochemical basis of survival that stems from an organism's need of nutrients from its natural, and commonly dangerous, environment. The oxidative respiration that cells undergo in order to provide essential nutrients and necessary energetic factors, such as adenosine triphosphate (ATP), is the biochemical drive that steers every branch of life that has spawned after the oxygenation of Earth's atmosphere. This respiration is commonly achieved through the oxidation of carbon-based compounds, carbohydrate in nature or otherwise. Lower life form organisms, such as plants, make for excellent sources for these compounds as they lack the ability to maneuver or find safety from natural predators. Complex organisms can be observed seeking lesser or weakened forms of life in order to satisfy their own nutritional needs.[7]

In order to circumvent this unavoidable power play in an ecological system, certain lifeforms developed chemical means of defense, countermeasures, in the form of toxins. These compounds, commonly of low molecular weight, produce a negative, altering, or potentially lethal effect on a predatory target. This led to the evolutionary drive to develop biochemical methods of detoxification and elimination of these compounds for an organism. Such an interesting back-and-forth in nature has garnered what might be one of the most studied corners of biochemistry: molecular biotransformation and drug metabolism.

In this field, categorization serves as an excellent resource for scientists to begin to understand why, and the also interesting question when, a species engineers a method to deal with environmental stressors in the form of chemical xenobiotics.[2,58] The general flow of xenobiotic drug metabolism consists of absorption, distribution, metabolism, and finally excretion of a small molecule. The field of pharmacokinetics and pharmacology refer to this series as the acronym, "ADME" and serves as the framework for almost any foreign compound, natural or otherwise, introduced to a species' homeostatic system.[58,59]

The enzyme factors responsible for these steps in the metabolism series have been separated into two phases: Phase I enzymes are generally oxidizing enzymes utilizing atmospheric dioxygen and consist of a final reduction product being water and an oxidized substrate, and Phase II enzymes are responsible for preparing of a polar metabolite formed from the first phase for elimination.[7]

The Phase I enzymes include Flavin-containing Monooxygenases (FMOs), cytochrome P450s within mitochondria or endoplasmic reticulum, and other enzymes such as epoxide hydrolases (EHs). Each of these serves a purpose as either transforming a functional group or activating generally unreactive groups into more reactive ones. Examples of these sorts will be discussed throughout this thesis as the substrates of cytochrome P450s contain electronically stable C-H sigma and C-C pi bonds that are transformed into hydroxyl or epoxide groups.[59] The major goal of these enzymes is to begin to make nonpolar toxins more water soluble by these oxidative means, to aid the second phase of metabolizing enzymes.

The Phase II enzymes are commonly distinguished as being 'transferases', those that transfer additional chemical groups such as acetyl, sulfate, and glutathionyl groups to

the sites generated by the Phase I enzymes.[7] These bulky, more water-soluble groups almost always inactivate the harmful effect employed by a xenobiotic toxin due to interruptions in the intermolecular interactions that govern biomolecular recognition. They also prepare the compound for conjugation by an organism and assist in the ensuing elimination.

Cytochrome P450s make for excellent models to study in their roles in drug metabolism due to their relatively elegant system that is not dependent on too many external factors. However, their study comes with a cost of complexity; due to their numbers and prevalence across species, tissue, sex, life stage, etc., their individual activities and functions are easily lost to obscurity. Furthermore, many of these enzymes are considered promiscuous in nature, contorting their active site volumes in order to metabolize large collection of structurally distinct substrates.[1,2,58] This attractive feature serves as a great tool for nature in order to cleverly aid in the biochemical warfare in ecological realms, however it serves additionally as a nightmare for a researcher curious about individual roles and function of these P450 systems.

A recent avenue of research in the field of molecular pharmacology has been in the development of personalized pharmacotherapy based on genetic expression of these cytochrome P450s in individuals. Within a population, a large variety of polymorphisms (varied forms of the gene with slightly different sequences) can exist. The stage of life of an individual is another factor as well. The different expression levels of CYPs leads to widely varied effects from prescribed medication.[11,60] The existence of ‘orphan’ cytochrome P450s with catalytic functions currently unknown or not fully elucidated could prove useful to further these prospective targets in the fight against disease.[3]

Hormone Biosynthesis and Signal Clearance

Hormones, such as steroids, and the large number of derived or related compounds play an indispensable role in cellular and biological function. While at first glance, they appear to play benign roles as purely non-participatory components of membrane bilayers, serving only physical roles as modulatory factors for membrane fluidity.[7,57] However, one look at how much energy nature invests into the creation of compounds like cholesterol, as is observed in human cells, might imply a deeper connected role in cellular homeostasis.

Steroids and other biologically-important sterols are characterized by their polycyclic structure with additional modifications and accessories, such as aliphatic hydrocarbon groups, distinguishing them from each other in name and in function.[7] To be considered a sterol, the nonpolar lipids contain a hydroxyl group either on the root ring structure or on aliphatic attachments.

These compounds make for exceptional human P450 substrates, favorably binding the nonpolar substrates with their relatively hydrophobic active sites.[61–63] The interactions between these compounds and P450 subfamily members has been explored and new information is still being gathered on the roles these molecules play on the physiology of living beings.

Several cytochrome P450 family isoforms functionalize the lipophilic compounds in membrane environments, making them more water soluble and possibly influencing the lipids' localization in a membrane by adding polar functional groups to them.[64] These interactions might further modulate cytochrome P450 activity and chemistry as they are

observed to have an effect on the structure of membrane-attached drug metabolizing P450s.[65]

Human Cytochrome P450s and their Roles in Disease

Proteins are the molecular basis for life and failures in their activity lead to the dysfunctions that we observe as disease in biological organisms.[1] It can be through the seemingly insignificant action of a small molecule interacting with its physiological target that results in a positive effect or pathology for an organism. Furthermore, disturbing intermolecular interactions between protein-protein complexes or other biomolecules can have disastrous or altering effects. The role of proteins in maintaining cellular function cannot be understated: from regulation of growth and differentiation to homeostasis and other biological processes, it remains clear that the key to improve understanding and treatment of disease lies with the study of proteins.

When diving into the literature surrounding human health and disease, one cannot go far without coming across a drug-drug interaction, or a compound activated or inactivated through catalysis by a cytochrome P450 enzyme. The role of cytochrome P450 enzymes in disease are studied and understood in general, with several families namely CYP1, CYP2, CYP3, and CYP4 enzymes showing more prevalence in study.[5,60,66] However, there are smaller clades and groups of P450s that have only causal or implied interactions with disease through either variability in gene product or mutation.

Some of these mutations that appear benign or unimposing for life function, however, are clearly present and impactful in other situations. Some examples exhibit

vastly impaired growth factors such as seen in XY chromosomal sex reversal and hypercalcemia[67], while others impact specific regions of the body like the heart tissue and coronary artery disease. More impactful examples are Bietti crystalline corneoretinal dystrophy, and lamellar ichthyosis type 3, as the process of normal human development is detrimentally affected or even prevented, as in the case of lethality in embryo.[68–70]

Such a wide variance in pathology indicates that the roles these enzymes play in maintaining and moderating signal-molecules and other endogenous or exogenous compounds is essential for a long, healthy life for an organism. These enzymes also perhaps play another more intangible role in the lifespan of organisms by the production of reactive oxidant species (ROS) as mentioned before.[27]

These factors make up the framework for why the cytochrome P450 enzyme system is so rigorously studied in its roles in drug metabolism and disease as they are very well-defined therapeutic targets for drug design. The major conflicting factor is the sheer number of individual proteins within each tissue. Individual studies on P450 isoforms provide a wealth of information for researchers, however, the insight gained often appears incompatible with the promiscuity (some of these enzymes will oxidize an incredible number of different substrates) and active site flexibility that these proteins display.

Eicosanoid Metabolism

This thesis will focus on the metabolism of eicosanoids, a class of endogenous compounds which are used as signaling molecules in several tissue types belonging to mammalian organisms. Eicosanoids are essential biomolecules important for maintaining

homeostasis in biological systems.[71] Their metabolism involves an elaborate network of metabolic pathways from multiple enzymatic systems such as cyclooxygenase (COX), lipoxygenase (LOX), and finally the cytochrome P450 monooxygenase pathway (CYP) which will be explored in detail in this work.[5,72,73] Eicosanoids are lipid-based compounds derived from dietary sources and endogenous phospholipids. Most eicosanoids originate from arachidonic acid (Figure 6), also known as (5Z,8Z,11Z,14Z)-5,8,11,14-eicosatetraenoic acid, a polyunsaturated ω-6 fatty acid with four cis double bonds located at the 5, 8, 11, 14 positions.[74]

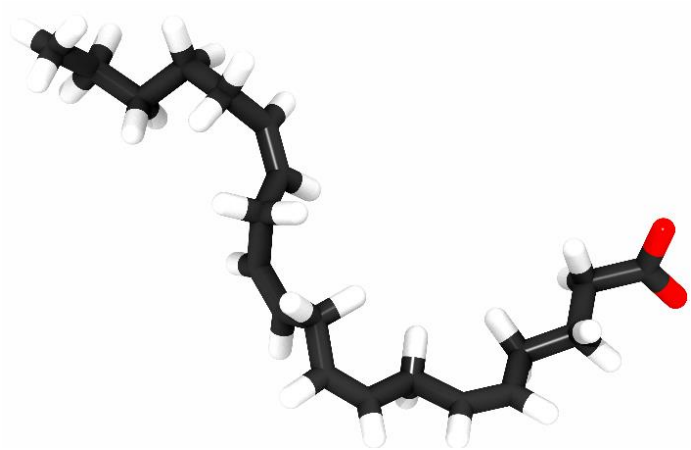

Figure 6. Licorice representation of arachidonic acid. Shown is the deprotonated form, arachidonate, as is present in physiological conditions. Of note, degrees of unsaturation are not shown in this graphical representation.

The usage of phospholipid-liberated arachidonic acid is commonly observed in intracellular signaling, its purpose is generally reserved as a secondary messenger molecule for inflammatory processes and cellular function[71,74–77]. The production of these physiologically important derivatives is referred to as the "arachidonic acid cascade" owing to the parent molecule of the metabolites.[71,78]

Eicosanoid metabolism is extensively studied for its role in human disease and illness with the more prevalent examples being their roles in human diabetes,

hypertension, renal disease, and cancer progression. Different enzymatic pathways result in a large variety of metabolites formed by oxidation reactions performed on arachidonic acid. Cyclooxygenase (COX) and lipoxygenase (LOX) pathways generally result in the formation of prostaglandins, thromboxanes, leukotrienes, and internal hydroxylation products.[72,79–81] The CYP pathways result in terminal or near terminal hydroxylation and epoxidation products.[82–84] These products have been observed to have a high amount of importance in the maintenance of organ function and their relative ratios are mediated by specific cytochrome P450 subfamily members. The CYP 4A and 4F subfamilies are responsible for most of these oxidations and only very recently have been explored as potential therapeutic targets for disease.[85–87]

Cytochrome P450 4 Clan and their Roles in Human Health

The Cytochrome P450 4 (CYP4) Family

The CYP4 family of cytochrome P450s have a preference for oxidation of terminal carbon atoms on saturated and unsaturated fatty acids of varying lengths.[88–91] This catalytic process is referred to as ω-hydroxylation, as the C-H bond on a primary carbon atom furthest from the carboxylate is transformed and functionalized into a terminal hydroxyl group.

This specificity is in direct contrast to what is commonly observed in other P450 enzymes and is especially the case for most drug metabolizing P450 systems that regularly hydroxylate internal or even more exotic C-H bonds such as those found on aromatic compounds.[92] The C-H bonding orbital of a primary carbon compared to that of

a secondary or tertiary species is lower in energy, making it relatively unreactive.[93,94] In a typical fatty acid, there are approximately ten times as many of the more reactive secondary C-H bonds in the fatty acid substrate, so specificity for oxidizing the ω-carbon is remarkable. This is generally explained by the relative stability of the resulting secondary radical being formed in the radical rebound mechanism of the P450 being more than that of a primary radical. With the ω-regiospecificity of CYP4 enzymes commonly facilitating these kinds of reactions, it makes it clear that the enzyme's active site plays a crucial role in overcoming these energetic hurdles.

More importantly to the physiology of organisms, the site of oxidation plays a critical role in a real-world sense. The CYP4 family members act as the chief enzymes in ω-hydroxylation of long chain fatty acids and their derivatives such as the eicosanoids[74,95,96], as well as a number of catabolic roles with essential vitamins[97–99] and several xenobiotic biotransformations.[100–102] These compounds can have vastly contrasting roles in the body. With one oxidation performed, a CYP4 enzyme can transform an endogenous polyunsaturated fatty acid into either a potent vasoconstrictor or a vasodilatory signal, producing major changes in blood pressure and blood flow.[103–105]

Structurally distinct from the high-plasticity drug-metabolizing P450s, the CYP4 family of enzymes constrain access to their catalytic sites through residue interactions that result in a narrow channel that allows for only the terminal carbon atom of an aliphatic substrate to reach the heme center.[106,107] Their binding of long chain and polar headgroup containing compounds makes them remarkably different from most P450 enzymes which have hydrophobic, small molecular weight target substrates.

Another method that these CYP4 enzymes employ to achieve this preference for ω-hydroxylation is through modification of their heme group. The CYP4 family of monooxygenases covalently bind their heme cofactors through an additional ester linkage through a heme methyl group and a conserved carboxylic acid sidechain positioned within the individual CYP4 isoform. This feature is an important aspect of this family for the work in this study and will be discussed in greater detail in the next section.

The differences in expression between individuals is an important aspect of their roles in health and disease. Within humans, there is a total of 12 genes coding for 13 CYP4 enzymes.

Table 1

Subfamily members of the human CYP4 clan, their genes, and the major tissue distribution

CYP4 Gene	CYP4 Enzyme	Tissue Distribution
CYP4A11	CYP4A11	Liver, kidney
CYP4A22	CYP4A22	Liver, small amounts in other tissues
CYP4B1	CYP4B1	Liver, respiratory system, urinary tract
CYP4F2	CYP4F2	Liver, kidney
CYP4F3	CYP4F3A	Leukocytes, bone marrow
	CYP4F3B	Liver, kidney, gastrointestinal tract
CYP4F8	CYP4F8	Epithelium throughout the body
CYP4F11	CYP4F11	Liver, kidney, heart, muscle
CYP4F12	CYP4F12	Small intestine, liver, colon, heart
CYP4F22	CYP4F22	Esophagus, skin, small amounts in other tissues
CYP4X1	CYP4X1	Brain, heart, kidney, skin, small amounts in other tissues

(continued)

CYP4 Gene	CYP4 Enzyme	Tissue Distribution
CYP4V2	CYP4V2	Throughout the body, among various tissues
CYP4Z1	CYP4Z1	Mammary glands

Note. Tissue distribution information is from the protein and gene knowledgebase, UniProt. (https://www.uniprot.org/)

Of these, only a fraction shows fatty acid ω-hydroxylation activities and are studied as such. The more commonly studied human isoforms are CYP4A11, CYP4F2, and both splicing forms of CYP4F3 (4F3A and 4F3B). While CYP4A11, and CYP4F2 are understood to be involved in renal and heart function[108], CYP4F3A isoforms are expressed in neutrophils and are directly crucial in the way chemical signals are moderated by the inflammatory system.[109]

CYP4F8 is expressed outside of liver tissue and has not shown any activity in ω-hydroxylation of fatty acids, in contrast to CYP4V2 which did show some activity.[70,110] CYP4F11 and 4F12 performed more traditional P450 roles as xenobiotic Phase I enzymes, however 4F11 showed preference for ω-hydroxylation of long chain fatty acids and vitamin K.[97,111] CYP4F22, infamous for its role in the development of lamellar ichthyosis type 3, is a specialized CYP4 with a preference of very long chain fatty acids (VLCFA) and it is thought of to be important in the development and formation of skin-layering bilayers.[69]

Covalent Heme Linkages in the CYP4 Family

The chemistry of the heme-thiolate prosthetic group characterized by the covalent attachment through a cysteinyl residue of a cytochrome P450 has been well studied through a variety of methods and is well understood as the most universal feature of P450

cytochromes. Regardless of how the active site environment of individual isoforms may change, they all must still have the signature cysteine residue covalently attached to the heme cofactor to be consider a part of the cytochrome P450 superfamily.

This principle, seemingly solidified through years of rigorous study in P450 systems, is not without exception. In the study of several CYP4 enzymes, an analysis through the digestion of protein resulted in the discovery of hydroxylated heme cofactors.[112,113] Initially thought to have occurred through a side reaction or oxidation after the digestion process, researchers discovered through experimental methods that the heme cofactor was not only covalently held in place by its conserved cysteine, but also through an additional covalent linkage.

The theorized formation of this linkage is through an autocatalytic reaction of the P450s oxyferryl catalytic center. The ultimate oxidant oxyferryl species seemingly self-catalyzes the hydroxylation of one of the methyl substituents on the iron protoporphyrin IX ring. This results in either 5- or 8- hydroxylation, as its site placement is reliant on the porphyrin ring orientation in the catalytic pocket.[114] Interestingly enough, through a conserved glutamic acid residue on the I-helix in the active site pocket, an esterification takes place that results in the formation of a covalent bond between the hydroxymethyl group of the porphyrin ring substituent and the carboxylate of the glutamate residue. This distinctly changes the chemical environment of the active catalytic site and changes the protein's binding activity and specificity.

This modification, first observed in experiment, was confirmed through quantitative means (mass spectrometry) to occur naturally *in vivo* and not to be taken as an artifact of method or expression through non-native host systems such as humanized

rat CYP4s.[112] The covalent binding of the heme cofactor has been confirmed for a number of CYP4 enzymes, however some appear to have a normal heme prosthetic group.[112,113,115] The presence of a glutamate residue at the correct position in the amino acid sequence almost always results in the formation of the unusual heme linkage.

Through the addition of their energetic cofactors (NADPH) and their redox partners (CPR), CYP4 enzymes with the appropriate glutamate were observed to form covalent bonds to their heme. This indicates that the auto-catalytic process occurs through a fairly typical mechanism of P450 oxidation, in that it requires reducing equivalents (and, of course, molecular oxygen) to occur.[113] This process was first observed on incubation of recombinant protein under ideal conditions for some rat CYP4 isoforms.

Mutations in the active site of CYP4 P450s resulted in the elucidation of interesting attributes to the covalent heme attachment. When the conserved CYP4 glutamic acid residue is mutated into aspartic acid, which is a similar amino acid with a shorter chain length, the yield of covalent heme linkage was reduced in CYP4 enzymes yet was still present. Interestingly, experiments showed that within the CYP4 proteins that did not covalently bind their heme, a hydroxylated heme cofactor was still isolated, indicating that the auto-catalysis continued to occur, however the esterification was hindered by the reduced reach of the aspartate residue side chain. Isotopic labelling and chromatographic methods indicated that this was due to the trapping of the electrophilic intermediate by solvent water in the active site, preventing the carboxyl group from reacting.[114–116]

The position of the glutamate is occupied commonly by glycine in the CYP4 family and studied for its impact on secondary structure and the catalytic heme environment. When site-directed mutagenesis is performed to change the I-helix residue into an acidic side chain, covalent heme binding is observed by CYP4 members that do not characteristically contain them.[117]

This covalent modification might be a protective feature of the P450 enzyme. The presence of the ester carbonyl group might act to modify the chemical environment of the heme, preventing additional auto-catalytic side reactions. Reactive oxidant species are commonly generated by these systems, and additional oxidation of the organic moiety of the heme would result in drastically altered catalysis or complete cessation of catalytic function. This sort of fortification can be founded on the protection of N-arylated heme products as seen in reactions between drug-metabolizing cytochrome P450s and reactive arylhydrazines.[118–120]

An alternative, and perhaps more physiologically important, theory on why the covalent linkages evolved is the regiochemical argument. The CYP4 family, which are predominantly fatty acid oxidases as discussed before, have high levels of regioselectivity and when modified to disrupt their heme linkage, no longer show the same ratios of product distribution and have lower regioselectivity.[113] Another pertinent example of this was observed in studies on the rabbit CYP4B1 enzyme, which is the structure in which the models developed in this study will be based upon.[116] When its covalent linkage was disrupted with the glycine mutation in place of the glutamate, the normal ω-hydroxylation products shifted to ω-1 and ω-2 hydroxylated metabolites.

The explanation that heme covalent binding evolved to promote ω-hydroxylation has yet to be fully validated as it would be difficult to prove through experiment that the presence of the heme modification tunes the reactivity of the oxyferryl species to be more favorable for the attack of the primary terminal carbon as opposed to secondary C-H bonds in fatty acid catalysis. While its direct impact on the electronics of the heme iron metal center are not fully elucidated, the steric interactions imposed by the linkage are more recognized. The presence of the linkage restricts access to the oxidant heme species in a way that facilitates ω-hydroxylation, or at the very least, limits access of any other atoms other than the terminal methyl group of a substrate. This suggests a prerequisite condition of a covalent heme linkage and a rigid active site volume for CYP4 enzymes to have specificity for ω-hydroxylation, however this cannot be solely due to the heme modification. Experiments on covalently-bound CYP4s with similar active site construction resulted in a dominating amount of ω-hydroxylated product in the distribution of hydroxylated products of lauric acid, a 12-carbon chain saturated fatty acid.[121]

Further research, like what is presented in this thesis, is important to fully elucidate the impacts of the covalent linkage and their relevance in determining physiological roles of individual CYP4 isoforms in a species.

CYP4 Enzymes and Inflammation

In human physiology, the body walks a fine line of balance for the inflammatory cascade. The eicosanoid metabolites of arachidonic acid can be grouped into several categories, either resulting in proinflammatory or anti-inflammatory signals. These

compounds lead to an inflammation amplification event or to the ensuing relaxation for an organism. Prostaglandins (PGs), leukotrienes (LTs), and 20-hydroxyeicosatetraenoic acid (20-HETE) are the metabolites responsible for promoting inflammation in the body, while the deactivated 20-hydroxyleukotriene-B4 (LTB4), lipoxins (LXs), and epoxyeicosatrienoic acids (EETs) are observed in having anti-inflammatory effects.[79,122–124]

20-HETE is an ω-oxidized product of interest for studying the biological crosstalk of eicosanoid metabolism and inflammation. The terminally hydroxylated product of arachidonic acid by metabolism through P450 activity, it acts as a powerful vasoconstrictor signal molecule and is understood in having a role in renal vascular regulation and the inhibition of ion transport leading to organ failure for an organism.[103,125–128]

In human liver and kidney cells, CYP4F2 and 4F3 are the major producer of this metabolite which indicate an inherent role in inflammatory exacerbation by these enzymes. However, more is to be studied and understood as they also deactivate proinflammatory sub-terminal hydroxylated eicosanoids through ω-hydroxylation as well.[79,82,129] The formation of epoxyeicosatrienoic acids is mediated predominantly by the CYP2C and CYP2J subfamilies, however these are also well-known substrates for metabolism and breakdown by CYP4 enzymes.[75] ω-Hydroxylation to their respective hydroxyepoxyeicosatrienoic acids (HEETs) are catalyzed by the CYP4F2, and 4F3 isoforms.[130] This dual role of mediating vasodilation and constriction has led to a large amount of study in the CYP4F subfamily as the ratio of eicosanoid metabolites (20-HETEs:EETs) can be used as gauges of inflammation in a clinic.[108]

CYP4 Enzymes and Cancer

Pharmacological interest in the CYP4-derived eicosanoid metabolites has increased over the past few years as the demand for new screening methods has grown. The need for early cancer detection methods has led scientists to identify both the CYP4 ω-hydroxylase enzymes and their primary metabolite, 20-HETE, as potent cancer biomarkers.[83,87,131,132]

The expression of P450s is heavily modified in cancer cells which has led to significant study. Due to the increase in inflammatory moderators by tumorous tissue cells, most drug metabolizing P450s are downregulated, showing drastically decreased expression and transcription.[125,126] However, in stark contrast, several CYP4 enzymes are unaffected or even experience enhanced expression and gene upregulation.[87] This has been theorized to be due to a not-yet understood relationship between 20-HETE, angiogenesis and subsequent cancer cell proliferation.

CYP4F members have shown a large amount of promise as biomarkers for cancer progression. In several human cancer cell tissue types, CYP4F2 expression upregulation was observed at the level of mature messenger RNA, which is the form ready to be translated by a ribosome into protein. Actual translation of the mRNA varies widely among cell types and individuals, thus mRNA could not be used to definitively indicate enhanced expression.[83] This upregulation was also observed for other CYP4 members such as CYP4A11 and CYP4F3, further demonstrating the ability of CYP4 ω-hydroxylase enzymes as biomarkers for tumor growth.[127] The rat model also provided for

this conclusion with the expression of CYP4F1 measured and notably increased in rat liver cells with aflatoxin B_1-induced tumors.[128]

20-HETE and its production through the CYP4 family has been heavily linked to cancer cell metastasis through neovascularization, the formation of new blood vessels. This process gives precious resources and nutrients from the host's systematic supply directly to new tumor cells further worsening the prognosis and status of the cancer; this has an additional effect of allowing the transport of mobilized tumor cells to different localized regions of the body.[133] This spread, referred to as metastasis in medicine, almost always results in a higher rate of mortality for the host and makes the treatment of a condition extremely difficult.[134]

While a direct protein target receptor for 20-HETE has not been elucidated, 20-HETE has been shown to interact with growth factors and promotes the proliferation of cells, making it more difficult for a cell to undergo apoptosis and controlled destruction.[135] The production of arachidonic acid-based eicosanoid metabolites by CYP4 enzymes is a currently developing story and using both experimental and theoretical methods, medicinal chemists and biochemists alike are working to map out these interactions as they relate to cancer prognosis and treatment.

Functions of Murine Cytochrome P450 4f Subfamily Members

Murine models make up a large percentage of modern-day drug design, and medicinal chemistry. Through the use of recombinant human P450s in murine models, and 'humanized' mouse methods that involve transplantation of (P450-containing) human liver cells into murine systems, medical research seeks to expand the advance of

medicine through research on these hybridized systems.[136] Furthermore, mouse models are studied as they can still be used to gain insight into metabolic function and physiological roles of related human P450s. The study of the evolutionary relationship of CYP4F subfamily members and their roles in health and disease in both humans and mice serves in the development of medicine.

In the study of proteomes, it is commonplace to compare the collection of proteins that are or can be expressed by individual cell types, tissues, or organisms. Genes coding for protein that originate from a common ancestor are said to be homologous, while genes that share origin but are from different species are said to be orthologous. With humans having 6 individual 4F isoforms, and mice having 9, locus comparisons have been made however the pairings are difficult to make reliably, as shown in Table 2.[137]

Figure 7 shows an evolutionary tree of the Cyp4f isoforms, and as to be expected, only a small amount of evolutionary relationship is suggested between some human CYP4F isoforms and murine P450s. One distinction is found between human and mice as only a single orthologous relationship is observed between the genes of human *CYP4F22* and murine *Cyp4f39*. In contrast, murine cytochrome P450 *4f13-4f18*, *4f37*, and *4f40* genes all have no clear orthologous pair with human P450s. This number of individual genes not having a clearly defined analogous partner between species indicate the probable existence of a single ancestral P450 gene that resulted in independent evolution into multiple genes across the mammalian CYP4F genes.

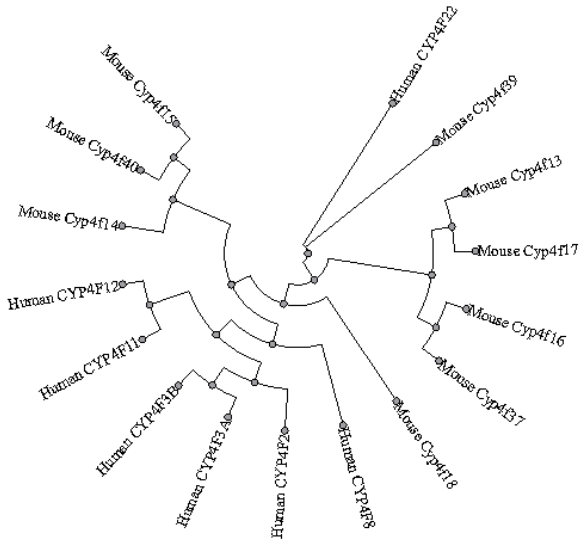

Figure 7. Circular phylogenetic tree of the genes coding for the human and murine cytochrome P450 4F subfamily. The PHYLIP tree building method was used to create the phylogenetic tree for several of the CYP4F subfamily members and resulted in a diagram that indicates evolutionary relationship between isoforms.[138]

Table 2

Pairings of individual human CYP4F and murine Cyp4f genes, excluding pseudogenes or detritus genes

Human Gene	Mouse Gene	Orthologous Pair
CYP4F2	*Cyp4f13*	None known
CYP4F3	*Cyp4f14*	None known
CYP4F8	*Cyp4f15*	None known
CYP4F11	*Cyp4f16*	None known
CYP4F12	*Cyp4f17*	None known
CYP4F22	*Cyp4f18*	None known
	Cyp4f37	None known
	Cyp4f39	*Human CYP4F22*
	Cyp4f40	None known

Note. The CYP4F cluster and gene listing is from the work by Nelson et al. 2003.[137]

Tissue-specific, gender-specific, and species-specific expression of murine Cyp4f subfamily members vary widely. In the tissue distribution of Cyp4f13, the highest amount of expression was found in the liver of both sexes of mice, with relatively high amounts also observed in the ovaries, uterus, kidneys, and intestines. In contrast, Cyp4f14 was detected in moderate amounts only in the small intestine but was 50% higher in females than male mice.[139]

Much is still unknown about the role of individual murine Cyp4f isoforms. Among the members, only the catalytic function of Cyp4f14 and 4f18 is known, both catalyzing the hydroxylation of leukotriene B4.[122,140,141] The catalytic functions of the other mouse Cyp4f isoforms are currently unknown; however it is theorized that there may be unique roles for some individual isoforms in the formation of 20-HETE based on structural patterns and level of expression.

The murine Cyp4f family members have been studied for their physiological substrate targets, as well as their regulators of expression.[142] Some degree of substrate commonality has been found for the 4f subfamily members with individual isoforms showing high amounts of substrate specificity for several classes of compounds, possibly serving as targets for drug development and medicinal research. The activity of peroxisome proliferators like the fibrate family of compounds have been observed in regulating the CYP4F enzymes in an isoform specific manner.[143] Some data has been collected on the catalysis of oxidation of long chain polyunsaturated fatty acids and eicosanoid metabolites like leukotriene B4 by murine Cyp4fs, allowing researchers to study the inactivation pathways involved in inflammation by these CYP4F enzymes.[129]

One notable exception is the murine Cyp4f13 which until very recently had no known substrates, inducers nor repressors. In 2016, the murine Cyp4f13 had been observed in influencing the production of oxidized metabolite formation of specialized lipids in the brain indicating some role in neurological and metabolic disorders. Overexpression or complete knockout of murine P450 4f13 isozymes exhibited drastic effects to the formation of downstream metabolites.[144] An established P450 CYP4F inhibitor, HET0016, has been used to prevent the formation of ω-hydroxylated products such as the 20-HETE formed by human CYP4F isoforms and appears to show activity as an inhibitor on murine Cyp4f13.[145]

Scope of this Thesis

Cytochrome P450-dependent metabolism plays a key role in human health and disease. Examining even one subfamily of related P450s results in a web of interconnected metabolic pathways that cells employ in order to maintain homeostasis. As a major Phase I enzyme responsible for the numerous biotransformations in the body, P450s exist as targets for study for many fields of chemistry as their interactions with novel therapeutic agents can lead to progress in drug design and discovery.

In contrast to earlier studies, which have been focused on the more physiologically prevalent isoforms that are responsible for drug metabolism, the work in this study has been focused on more elusive P450 members that still show potent roles in health and disease. This aspect is what inspired the work presented in this thesis, as the physiological roles played by CYP4F enzymes are not yet fully understood. The primary enzyme investigated in this research is the murine cytochrome P450, Cyp4f13.

Expression and purification of these delicate enzyme complexes is very difficult as the membrane-localized proteins become dysfunctional in the formation of microsomal bodies. Little progress has been made in common-practice bacterial models as they commonly fail to produce significant amounts of protein, or any at all. Insect cell-based protein expression systems have been employed in studying these protein systems to some success, however, still have trouble producing quantitative amounts.[146]

With a better understanding needed; theoretical chemistry methods and simulations are another means of attempting to answer these questions. Theoretical methods allow chemists to probe and observe the biomolecular cosmos on a level not available to experimentalists. Modern computational methods have shown proficiency in investigating some of the more subtle aspects of the P450 isoforms. Both computational methods and direct experimental methods are needed in order to piece together the puzzle posed by nature. One aspect of these P450 systems of particular interest is their ability to shift role and seamlessly moderate the production or degradation of crucial cellular signaling molecules. Experimentally, through active site mutagenesis and other means of probing, the substrate flexibility of these enzymes has been observed but not fully explained by routine instrumental methods of analysis.

Therefore, it is of great interest for chemists to exploit the powerful field of theoretical chemistry in order to elucidate the important interactions of active site residues, or substrate that facilitate such a phenomenon. In this thesis, I plan to discuss my observations of the interactions between arachidonic acid and the murine cytochrome P450 Cyp4f13 system embedded into a lipid membrane bilayer. Through the construction of a structural ensemble resembling that of its native environment, this study will provide

insight into the structural forces at play in enzymatic production of ω-hydroxylated oxidized metabolites.

Because the metabolites of these enzymes play a crucial role in disease and cancer, one of the main goals of this work is to provide understanding in the CYP4F-catalyzed fate of the physiologically important eicosanoid, arachidonic acid. In addition, we have also carried out additional simulations that act as informed probes of the active site by examining the role of residues in positions found to be important in substrate interactions on more well-studied cytochromes such as $P450_{BM-3}$.

One of the primary motivations of this work is the fact that the expression of murine Cyp4fs is altered in several forms of cancer manifestation and due to the common usage of mouse models in medicinal research, a more clear understanding of murine biochemistry is needed. While there has been some success in the study of recombinant enzymes and their metabolite profiles from subcloning, there has been relatively little computational study on active site characterization on individual murine Cyp4f members. No crystal structures of any CYP4F subfamily members have been isolated and the finicky nature of these membrane-embedded proteins make it very difficult to express them in useful amounts without modifications that can alter substrate specificity and other properties.

Specific amino acid residues will be chosen and mutated within the Cyp4f13 active site, selected based on their predicted effect of the mutated amino acid-residues on the catalytic function and binding specificity of the enzyme. To further characterize the murine Cyp4f13 active site, the role of covalent heme linkage will also be studied.

Through these simulations and resulting insights obtained, we hope to further the understanding of murine Cyp4f13-catalyzed ω-hydroxylation of arachidonic acid.

CHAPTER II

Experimental Method and Theory

This chapter presents the theoretical methods and foundations used in this work. First, an introduction to physical theories and the different computational approaches, such as quantum chemical methods (QM), Density Functional Theory (DFT), with Molecular Mechanics (MM) and statistical mechanics. Next, Molecular Dynamic (MD) Simulations will be introduced as well as a description of potential energy forms (force fields), parameterization, thermodynamic ensembles, and a summary of the process behind running molecular dynamic simulations. Finally, a comprehensive description of the homology modeling process used to develop the murine cytochrome P450 4f13 structural ensemble will be provided.

Physical Theory and Computation

Physical theoretical methods can be resolved to a combination of quantum physics and statistical thermodynamics. The first component describes the interactions between atoms and subatomic particles, whereas the second explains the thermodynamic behaviors exhibited by large systems. Computational chemists utilize these theoretical methods in order to arrive at quantum mechanical descriptions of the electronic states for a given molecular system. Currently there are three major approaches, and all are based on approximations as electronic repulsion makes any system far too complicated for exact solutions.

Quantum Mechanical Methods (QM)

Quantum mechanics refer to the behavior of matter being described by discrete physical phenomena such as the properties exhibited by waves or particles like energy and momentum. The mathematical formalisms for quantum mechanics were developed at the beginning of the 20th century and resulted in functions that could explain for almost all the shortcomings of classical physics. The cornerstone of quantum chemistry lies with the famed Schrödinger's equation and is the basis for most of theoretical chemistry; all ground-state quantum chemistry is founded on solving the time-independent Schrödinger's equation for an atomic system, given by the following equation.[147]

$$\hat{H}\psi = E\psi \qquad (1)$$

Where $\hat{H}$ is the Hamiltonian operator,

E is the energy eigenvalue,

ψ is the electronic wavefunction

This mathematical function can only be exactly solved for single electron systems and would result in a scalar multiple of the input with a clear significance: the energy levels of the electronic system that the operation was applied to. This equation would pave the way for the advances in quantum chemistry to come. Given a system with a single particle as an input, distinct energy eigenstates, which can be considered as 'atomic orbitals', could be resolved. However, one of the major consideration was that the ease of the integration was heavily dependent on the size of the system. While Schrödinger's equation results in the exact solution for single-electron atomic systems,

such as in the case of a hydrogen atom or a helium cation, the equation becomes unwieldy for multi-electron systems and to a greater extent for molecules. For every other instance, assumptions and numerical approximations are necessary. The general time-independent Schrödinger's equation can be separated into kinetic and potential terms, as seen in the following equation for spherically symmetric systems as seen in the 1s orbital of a hydrogen atom.

$$\frac{-\hbar^2}{2m}\nabla^2\psi + V(r,\theta,\phi)\psi(r,\theta,\phi) = E\psi(r,\theta,\phi) \qquad (2)$$

Where r, θ, ϕ are the spherical coordinates of an electron in a system,

$\frac{-\hbar^2}{2m}\nabla^2\psi$ is the Laplacian of the kinetic energy term,

$V(r,\theta,\phi)\psi(r,\theta,\phi)$ is the potential energy term

These terms can be further broken down to subsequent terms that describe each of the interactions between subatomic particles and lead to one of the key principles for quantum mechanics: the Born-Oppenheimer approximation.[148] When solving for electronic wavefunctions, it is generally understood to consider nuclei as static particles with external potentials. Protons are about 1,800 times more massive than electrons and their kinetic energy terms are unnecessary to numerically describe which reduces the amount of integration. This approximation also serves useful in fully separating the kinetic and potential energies of systems and their subatomic elements. Eq. 3 illustrates the three remaining components considered in quantum chemistry.

$$\hat{H}(r,R) = \hat{T}_{nuc}(R) + \frac{e^2}{4\pi\epsilon_0}\frac{Z_A Z_B}{R} + \hat{H}_{elec}(r,R) \quad (3)$$

Where r and R are electron and nuclear positions,

$\hat{T}_{nuc}(R)$ refers to the kinetic energy of individual nuclei,

$\frac{e^2}{4\pi\epsilon_0}\frac{Z_A Z_B}{R}$ is the coulombic repulsion of the two nuclei,

$\hat{H}_{elec}(r,R)$ is the summation of the kinetic, repulsion, and mutual attraction to nuclei from each electron in a system.

The numerical approach is the use of approximate wavefunction forms to describe molecular orbital theory. Numerical approximations use ordinary differential equations that can be considered solutions to the complex partial differential equations of quantum theory. This works precisely and accurately only for the hydrogen atom and works reasonably so for very small molecular systems but is impractical and computationally expensive for most larger systems. This approach uses approximate wavefunction forms of electrons for interacting systems, which represents one of the drawbacks with this level of theory. For example, Slater determinants are expressions used to describe many-particle systems and act as educated guesses for the Schrödinger's equation but lack an electron repulsion term.[149] The Hartree-Fock method[150] and later combinations of other theoretical methods would tackle some of these problems but their usages on large biomolecular systems are currently held back by the computing power of modern-day technology.

Density Functional Theory (DFT)

The quantum mechanical approach employed by most computational methods is based on the use of the electron density, as opposed to the electronic wavefunction, as the independent variable for a function. This ‘functional’, defined as a function of a function, is the basis for Density Functional Theory (DFT).[151] This form of theory is widely used throughout computational chemistry and has led to a vast amount of research and understanding in the last century. Density Functional Theory uses the ground-state electron density for a system as all the electronic information of the ground-state wavefunction.

While being a powerful method of studying the electrochemical, chemical bonding and reactivity in small molecules, it remains unfeasible to use for the dynamics of proteins as the time scale necessary for physiological relevance is in the nanosecond ranges. This limitation relegates DFT to being used to study specific biochemical ensembles, such as catalytic sites of metalloproteins, and can be used as a powerful probe for the mechanistic aspects behind enzymatic reactions. When used with experimentally-derived data, this method is commonly used to generate parameters for less rigorous forms of simulation as will be discussed in the next section.

Molecular Mechanics (MM)

The final approach is the Molecular Mechanics (MM) method which bridges the divide between massive biological systems and computational chemistry.[152] Molecular mechanics approximates the energy of a system using functions with empirically obtained parameters referred to as force fields. In quantum mechanics, a wavefunction

describes the kinetic and potential energy of a molecular system, while in MM, a set of potential functions describe the energy of a molecular system with empirical information and classic phenomena, such as a harmonic oscillator for bond vibrations.

Computer simulation of massive biomolecular environments, such as seen in bacterial representations, are the next logical leap in progression for the field of computational chemistry.[153] There are several challenges for the field, namely the cost of calculation. As the amount of physical detail increases in a simulation, so too does the computational power necessary to perform such calculations. More detailed theories can describe complex phenomena and offer higher accuracy. Molecular mechanics is less detailed which allows for simulations of larger systems and for longer timescales; as a result, MM-based force fields are the method of choice in the simulation of biomolecules. Computational chemistry packages such as Gaussian[154] and GAMESS[155] are commonly used in order to perform quantum mechanics, DFT, and molecular mechanics calculations.

Statistical Mechanics

Statistical mechanics, while normally used to study macroscopic physical phenomena, is commonly used to explain thermodynamic behavior in complex microscopic systems such as the states of biomolecules.[156] The foundation of statistical mechanics is based on the existence of a state of a system, biochemical or otherwise, and the integration of motion that acts upon the state, progressing it in time. Examples of these motions are those found in classical mechanics such as Newton’s laws of motion, or quantum mechanics through use of the Schrödinger equation.

Calculation of the probabilities of these states is commonly performed in a variety of methods in the form of statistical ensembles, which are collections of all virtual states a system can possibly exist in. As microscopic variables differ between individual states, these aptly named microstates are collected and the density of states is used to draw conclusions about the system. The average of the states over time, circumstances, and mechanistic differences result in vastly different outcomes from seemingly similar initial conditions. This probabilistic method results in an infinite space of outcomes from which sampling is performed in order to approximate solutions.

Computer simulations of molecular dynamics rely on the averages of all possible states of a molecular system with a specified internal energy.[157] These so-called statistical ensembles, also known as micro-canonical ensembles, also consist of a fixed composition (set number of particles) as the commonly allotted degrees of freedom are either in pressure or volume.

Many methods of integrating randomness or stochastic behavior in these systems are performed through additional modifications to a system. While this is actively changing or possibly destroying information about a system with its chaotic or pseudorandom influences, in practice it provides researchers a way to get a more realistic sampling of ensembles. These practices are necessary to reduce abnormal chemical phenomena to subtle correlations. Commonly used methods of this in theoretical chemistry are the reinitialization of atomic trajectories and velocities, and the cycling of thermal excitation with pseudo-randomly generated seeds.

Molecular Dynamic Simulations

QM approaches are used in certain fields, photochemistry for instance, as the time steps covered span femtoseconds and can be performed on several atoms at a time. Such a small timespan allows for the resolution of minute changes in the atomic scale such as bond stretching and vibrational modes. DFT could be used on fast chemical reactions as it has a practical timescale of picoseconds and can reliably approximate the behavior of many atoms. Molecular mechanics are used on protein dynamics and drug binding as it has the capacity to process microseconds to nanoseconds and is able to handle thousands of atoms without being too computationally heavy.

Molecular Dynamics (MD) combines QM and MM with statistical mechanics and acts as the leading approach of computational methods for biological systems.[158,159] MD simulations both utilize potential functions and numerically integrate Newtonian motion to large biomolecular systems that can be used to gain information about biochemical events through thermodynamic ensembles. These potential functions, referred to as force fields, can vary widely between molecular mechanics simulations, such as that of all-atom approaches used in this work, to that of coarse-grained systems which are used for even larger molecular complexes. Molecular Dynamic Simulations have been commonly used to evaluate the substrate binding interactions and specificities in P450s.[160–163]

Several MD software packages are available and are designed specifically to handle biomolecular systems, such as AMBER[164] and CHARMM[165]. These two originated as merely force fields for the study of biomolecules, eventually leading to the development of entire simulation software packages. Later software implementations sought to capitalize in the advances made in computing by the strides taken in the late

20th century. Some simulation software packages such as LAMMPS[166] were created to investigate macroscopic material modeling and have shown exceptional use in chemical engineering and material science. Others, like GROMACS[167] and NAMD[168], were designed to utilize preexisting force fields as functional forms of potential energy to study large biomolecular systems, and now are a commonplace method in the simulation of biochemical phenomena.

Functional Forms of Potential Energy

Force fields are the empirically determined representation to the potential energy function of a molecular system. Many force fields have been developed in the field of computational chemistry and many have been tailored to meet the needs of the biomolecular systems. All interatomic interactions are approximations based and are obtained through experimental data, thus the term empirical. However, these methods prove to be too reductive in nature and resolve chemical reactivity to simplistic mechanical movement, leading to failures to fully replicate electrostatic potential charge distributions and the polarization of atomistic environments. Furthermore, protein structure predicted or created by these force field methods are heavily reliant on rigorous validation and refinement. A common phrase uttered throughout the field of computer science is the famed, ‘Garbage in, garbage out’. In essence, if initial molecular structure based on threading algorithms is not refined through human or machine-based learning, the simulations and the subsequent results can be taken to be unusable. Despite these deficiencies, the fields of computational chemistry and structural biology have seen a large amount of use and work being produced through these methods as the force field

methods have done much to develop understanding of molecular interactions leading to advances in biomaterial science, drug design and medicine.

Several force fields are designed with different considerations in mind and are used for different purposes. Classical, polarizable, real-time reactive, and more reductive coarse-grained functional forms of potential energy are all examples of developed methods in molecular design software. GROMOS, AMBER, and CHARMM exist as some of the forerunners of molecular dynamics of macromolecules; developed primarily for energy minimization and dynamic simulations for peptides, small molecules and macromolecules.

The development of force fields in the field of biomolecular modeling has grown extensively over the years. One of the most commonly used in the study of protein structure is the previously mentioned Chemistry at HARvard Macromolecular Mechanics (CHARMM) force field.[169]

The following equations (4-6) for the potential energy function is what is used in the CHARMM22 force field, which was released in 1991.[170]

$$V_{total} = V_{bonded} + V_{nonbonded} \quad (4)$$

$$V_{bonded} = V_{bonds} + V_{angles} + V_{Urey-Bradley} + V_{impropers} + V_{dihedrals} \quad (5)$$

$$V_{nonbonded} = V_{vdW} + V_{Coulombic} \quad (6)$$

In what is commonly referred to as the functional form, the potential energy function for a system can be defined as the summation of the bonding and non-bonding interactions between the constituent atoms. For the bonding interactions, different pair wise atom-atom interactions can be described as functions of displacement of atomic

distances, angle and energy in form of force constants. The non-bonding interactions are contained within electrostatic potentials and van der Waals interactions.

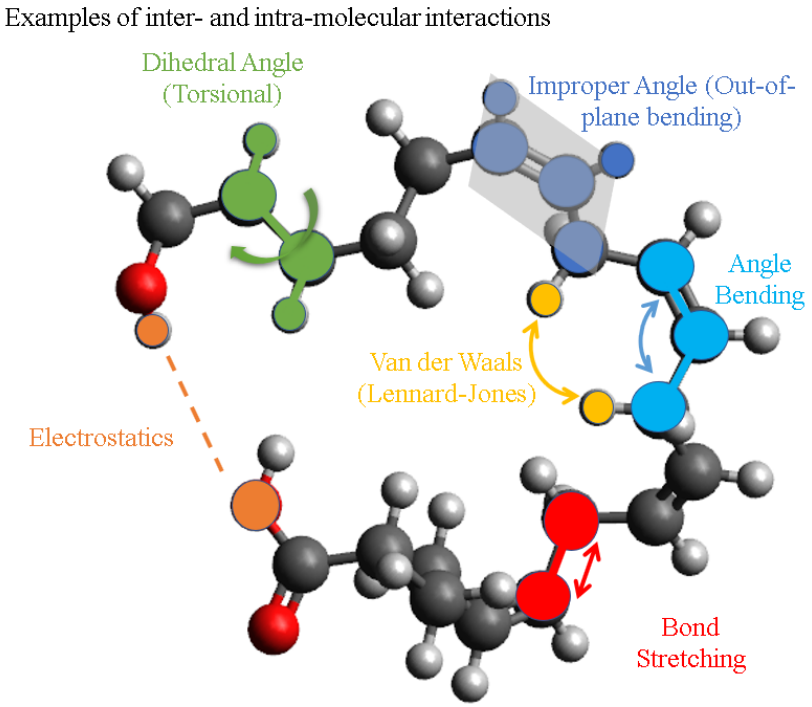

Figure 8. Schematic depiction of the interactions that the CHARMM force field utilizes. 20-Hydroxyeicosatetraenoic acid is the molecule used in the representative scheme.

These functions are designed to reproduce molecular geometry and properties using empirical methods and are used to describe the time evolution of the different interactions between atoms in a system, as shown in Figure 8. Constants allowing calculation of the energies of these interactions are referred to as parameters and are obtained from experiment and re-verified through successive optimization cycles.

Bonding Interactions

The bonding interactions encompass all the covalent bonding in a molecular system, taking into consideration certain attributes such as hybridization through use of

simple harmonic terms that describe stretching and bending modes. Torsional strain introduced by rotation about single bonds can be described as well as deviations in the planarity of groups of atoms.

Bond Stretching

$$V_{bonds} = \sum_{bonds} k_b(b - b_0)^2 \quad (7)$$

Eq.7 refers to the oscillations of the covalent interaction between two atoms in a molecular system. These terms are assumed to be approximately harmonic and derived from Hooke's law which linearly relates the force required to extend or compress a stiff spring by a distance to a constant factor determined by the character of the spring. In the situation of that between atoms, the magnitude of displacement away from equilibrium bond distance is an energetic penalty. This is analogous to that observed in other elastic bodies in response to applied forces. However, while this a reliable first-order approximation, it has limitations and does not exactly model the Morse potential of a diatomic molecule. The values for the force constants and equilibrium bond lengths are obtained from experiment, such as crystallographic diffraction data, spectroscopic data, or even theoretically from quantum mechanics calculations.

Angle Bending

$$V_{angles} = \sum_{angles} k_\theta(\theta - \theta_0)^2 \quad (8)$$

Eq. 8 represents angle bending in a similar approach; equilibrium bond angles between three-body systems with defined bending force constants result in energetic penalties when a system is acted upon by external forces.

$$V_{Urey-Bradley} = \sum_{Urey-Bradley} k_u(u - u_0)^2 \quad (9)$$

Eq. 9 shows another related parameter, the Urey-Bradley component. It is used for 1,3 non-bonded interactions for atoms separated by four bonds like as seen in allylic strained compounds. This component is generally underused in the study of biomolecule simulation as it is not easily analogized and is poorly transferable. However, this term is largely inconsequential for overall simulation sampling as its impact is more modest than incorrect charges or poorly optimized dihedral angles.

$$V_{impropers} = \sum_{impropers} k_\omega(\omega - \omega_0)^2 \quad (10)$$

Eq.10 represents the improper angle bending, produced by atoms bending out-of-plane formed by groups of atoms. It is modelled by this function in a similar fashion to the other representations of optimal bond angle displacement.

Torsional Rotation

$$V_{dihedrals} = \sum_{dihedrals} k_\phi[1 + \cos(n\phi - \delta)] \quad (11)$$

Eq. 11 represents torsional rotation about a dihedral bond, which are common features in peptide structure, are important stereochemical considerations when modeling protein structure. In the refinement process of homology models, it is commonplace to create Ramachandran plots that plot all the dihedral torsional angles of the constituent amino acid residues within a protein. Failure to establish optimized and permitted dihedral angles leads to incorrect protein structure prediction and is a major consideration for scientists as they hope to obtain insight into the structure of proteins.

Non-bonding Interactions

Van der Waals Interactions

$$V_{vdW} = \sum_{12-6\ Lennard-Jones} \epsilon \left[\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{6} \right] \quad (12)$$

Van der Waals interactions are represented as a complex 12-6 Lennard-Jones potential term. This term controls the Lennard-Jones well-depth to conform to the Morse potential and also contains two inner components; a steep repulsive term, and smoother attractive term, representing the London dispersion forces.

Electrostatic Interactions

$$V_{Couloumbic} = \frac{q_i q_j}{\epsilon r_{ij}} \quad (13)$$

Electrostatic interactions are modelled with a single Coulombic potential term. Attraction or repulsion is decided by the signs of the individual atoms interacting, with distance between them influencing the magnitude of the effect. The dielectric constant acts to simulate the effective nuclear charge and the shielding effect that atomic nuclei exhibit.

Polarization is simulated by NAMD through using a Drude model oscillator to perform its calculations.[171] Model oscillators are used to simulate the effects of electronic polarizability in the context of a classical molecular mechanics force field. They are inspired by the Drude model of mobile electrons and are used in the computational study of proteins, nucleic acids, and other biomolecules. The Drude model of electrical conduction treats electrons like pinballs that ping from nuclei in a crystal lattice. This

results in nonbonded interactions being the most-time consuming part of energy calculation.

Integration of Newtonian Motion

The integration of motion is the key step in simulating the progression of time for molecular systems. The methods, while abundant in number, generally resort to the calculation of Newton's second law of motion for each individual particle of a simulation.

$$\vec{F} = m\vec{a} \tag{14}$$

The classical equations of motion used are deterministic by nature and allow a system to describe the positions and momenta of every one of its particles at any given point in time. Numerical integration is founded in the basis that given the initial position and velocity, a particle's instantaneous position and velocity can be found as a function of time.

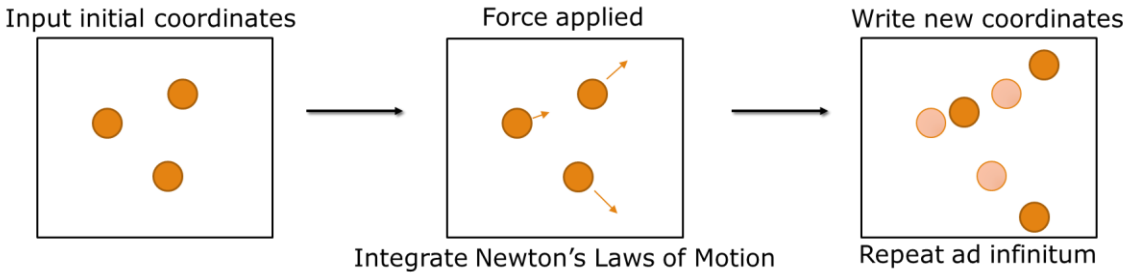

Figure 9. Cartoon depiction of the integrator process. A symplectic integrator is being demonstrated, with each repeating cycle of integration requiring the set of previous position and velocity for each particle.

For the simulation of biomolecules, these integrations would benefit from high order approximation, with trajectory correction methods and adaptive time steps so that non-important interactions would be given larger timesteps while more biochemically

important ones would receive smaller ones. These developing methods would allow for faster calculation, and more reliable results with more efficient computer resource use yet however, are currently unfeasible for the massive systems commonly simulated by MD. Biomolecular simulations have large numbers of degrees of freedom, are computational expensive, do not benefit from exact (fully minimized) solutions, and are usually focused on long-term averages of states changing chaotically from small perturbations as opposed to a final set of coordinates.

Integrators used in MD simulations benefit from some constraints, such as the conservation of both energy and volume within a space. This allows the preservation of physical properties and allows systems to be time-reversible. This is useful in calculation as the Newtonian equations are reversible and this property allows an integrator to eliminate the need for adaptive time stepping.

The symplectic integrator used in MD simulation, as is demonstrated in Figure 9, is a pattern of position updating as a function of previous position and momentum. The integrator is given position from atomic nuclear coordinates and a momentum is assigned in order to establish a conserved total energy for the system. The integrator then determines the new atomic position with the momentum of the previous coordinate used and updates the position. This process is performed theoretically ad infinitum, but in practice is dependent on the desired time frame of a simulation, and memory storage. The positions and charges are used to calculate potentials for each set of positions and velocities. These trajectories are stored as output and can be combined with visualization software in order to analyze and interpret. One major consideration is the chaotic nature

of these simulations. Small differences in the initial conditions quickly lead to very different trajectories.

Velocity Verlet is an integrator method that is based on the simultaneous calculation of both position and velocity at the same value of the time variable.[172] This provides for a more accurate representation than older methods but is still too unreliable and inaccurate. Additionally, there is the Beeman predictor-corrector method which can be used on systems of larger number of particles.[173] It is similar to the Verlet method, but requires the tracking of position, velocity, acceleration and previous acceleration vectors per particle which can get very expensive for a computation.

Some parts of the force field are expensive to calculate such as bond and angle vibrations within multibody systems, the likes of which are found in catalytic active sites of proteins. Others are relatively simpler yet slowly changing, such as long-distance electrostatic interactions. Some methods of efficiency have been implemented and are used commonly in NAMD simulations such as the SHAKE/Rattle algorithm.[174]

The Rattle algorithm employed by NAMD is based on the Velocity Verlet algorithm that calculates and predicts the next set of position and velocities of a particle from a present point, without requiring information from previous time steps. This reduces the amount of memory needed to make these large calculations and satisfies the constraints of the system while maintaining adequate precision. Requiring nothing more than an input of velocity and atomic position means that velocity rescaling, and system modifications to simulate constant temperature and pressure are possible.

Rigid bonds and fixed atoms also cut down computational costs, as fixing the movement of the fastest but non-essential bonds (non-polar bonds with hydrogen atoms) or entire portions of a system greatly reduces the calculation load.

Integrators such like the ones mentioned are algorithms that accelerate the atoms in the direction of the force designated by the force field. More sophisticated algorithms include higher order terms for better accuracy. There is one major limiting factor to these calculations, and it is due to their dependence on time or more specifically time steps. Time-steps are limited by fast degrees of freedom like bond vibrations, which can occur in femtoseconds. The use of small time-steps leads to large amounts of computational power necessary to simulate and increase the sheer amount of data recorded.

Periodic Boundary Conditions

Several techniques have been developed to close the gaps between simulations and the conditions encountered in real experiments. An important example of these improvements are periodic boundary conditions. Periodic boundary conditions are a method that increases the effective size of a simulation to theoretically infinite span. The defined arrangement of atoms in a simulation can be treated as an image and repeated as an endless array of the same image. Each of these periodic units are referred to as ‘cells’, an example of which is displayed in Figure 10.[175]

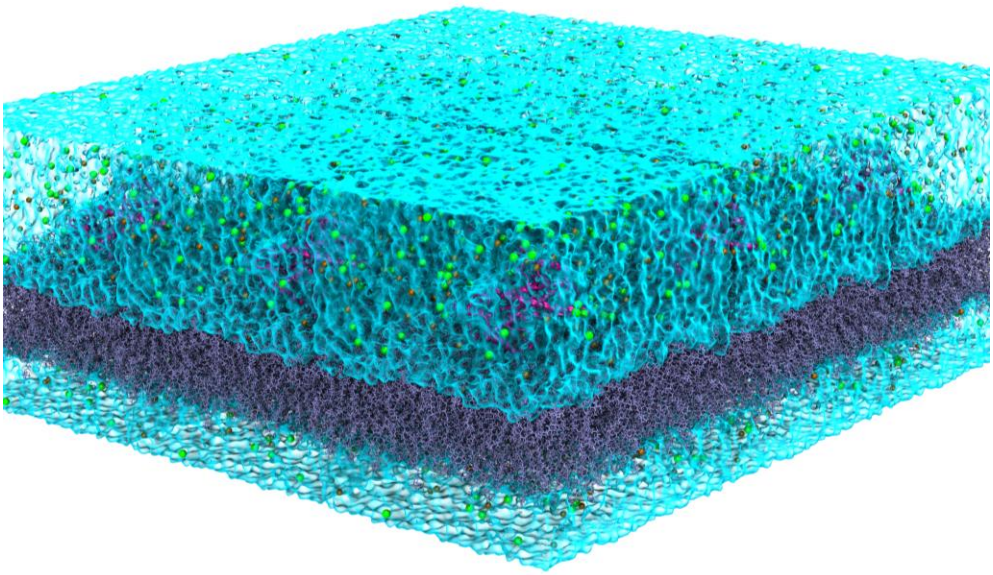

Figure 10. Periodic images of a P450 structural ensemble consisting of protein, solvent, ions, and lipid bilayer. Model used was the substrate-free Cyp4f13 homology model, whose construction is described in this work.

This allows the simulation to produce results that are more relevant to real biochemical behavior, as they are able to be observed in thermodynamic quantities. This conservation of processing power and particles also acts as a way to improve computational efficiency. Any particle that leaves the defined box of an image is simulated as entering from the opposite side of the image. This allows proteins to experience force similar to an actual bulk lipid bilayer packed with protein.

Proper construction of a structural ensemble should be large enough to avoid close contact between periodic images to reduce the interactions between proteins and their neighboring images. This is done to avoid 'finite size' effects by these images. Padding in the form of solvent is often employed to prevent such interactions, however must be competently constructed as long-range electrostatic interactions play a crucial role in many biomolecular processes such as entry/exit channels, and active sites.[176]

NAMD employs methods of calculating electrostatic interactions across images through the use of additional solvers: the Multilevel Summation Method (MSM)[177] and the Particle Mesh Ewald (PME) method.[178] Multilevel Summation Method can tackle periodic or non-periodic boundaries and is often preferred for systems with non- or semi-periodic boundaries. Particle Mesh Ewald method uses the periodic boundary conditions and behaves well for constant pressure simulations. It consists of the summation of long-range interactions between particles, first used in studying electrostatics in ionic crystals. It consists of a short-range potential component and a long-range summation of Fourier transformations of charge density, dependent on the spacing between the particles. This serves as an excellent model for forces obeying inverse-square laws such as gravity and electrostatics. It requires periodicity as it converges the long-range interaction over all space.[179]

Thermostats and Theoretical Pistons

Thermostat and barostat algorithms ensure that the temperature and pressure of the system fluctuate around a target value and allow for the MD simulation to sample different thermodynamic ensembles. Since kinetic energy and temperature are directly related, control of these physical properties is very important for a simulation.

The Berendsen thermostat method uses velocity scaling but does not produce canonical results; it works by increasing the velocities of slow-moving particles and vice versa. This leads to a too narrow distribution of velocities and is not representative of reality.[180,181]

The Andersen thermostat method uses velocity randomization by randomly resetting the momenta of particles.[182] This achieves a canonical ensemble but contains a major drawback: any kinetic properties and calculations are affected by this methodology. This means that it is ineffective for biological systems which are governed by kinetics more so than thermodynamics.

NAMD employs the use of the Langevin dynamics also known as stochastic dynamics, in the form of a theoretical piston.[183,184] It is very similar to Newton's equations, containing a couple additional components. The components of the Langevin piston maintain and control kinetic energy, therefore controlling the temperature and pressure.

Langevin dynamics involve a friction damping term and with a random force that acts on the particle to simulate solvent interactions. The amount of damping is used to maintain temperature and changed by the discretion of the experiment. If there is too high of a damping constant used in a calculation, it may slow the system's dynamics significantly. This is extremely effective at modeling a system wherein a protein is being buffeted by solvent molecules in its environment.

Thermodynamic Ensembles

Sampling from thermodynamic ensembles is where the statistical mechanics come into play. A statistical mechanical ensemble allows for simulations to exchange energy with an external environment. An ensemble represents all the microstates that are accessible to the simulation and provides the probability for a system to be in any given microstate. A microstate is a specific configuration of a thermodynamic system that the

system can occupy with a certain probability in the course of its thermal fluctuations. The canonical ensemble represents only the possible states of a system in a thermal equilibrium at a fixed temperature.

An ideal MD simulation conserves the total energy and entropy, and samples the microcanonical ensemble (NVE) where the number of particles, the volume, and the energy of the system is held constant. Used commonly for annealing simulations, this ensemble sees less use in simulation due to the fact that it far less computationally intensive to keep the temperature of a system held constant as seen in canonical, constant-volume ensembles (NVT). Another commonly used ensemble is the Gibbs ensemble or isobaric-isothermic ensemble (NPT) which requires the use of theoretical pistons for varying volume. Conserving both pressure and energy is difficult and computationally heavy, even more so if a system is allowed to interchange particles with the surroundings.

Nanoscale Molecular Dynamics (NAMD) Simulations

Nanoscale Molecular Dynamics (NAMD), the package used in this work, is commonly used in protein folding simulations and *ab initio* calculations in which only initial physical constants and constraints are given in order to simulate the dynamics of biomolecular systems such as protein-protein or protein-substrate interactions.[168,185] NAMD was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign. NAMD has the capacity to read input from a large collection of force fields available and comes with its own visualization software in the form of the

Visual Molecular Dynamics programming suite (VMD). This visualization software is used to generate graphical representations, prepare molecular structure for simulation, and to analyze with a variety of plugins and tools for the examination of protein structure.

In order to perform Molecular Dynamic simulations, a few forms of input are required by the NAMD software. These inputs come in the form of several file types: atomic nuclear coordinates files, structure files, electronic topology files, and parameter files.

The Protein Data Bank file format (PDB) is the source of atomic coordinates and does not require bonding information or secondary structure for proteins. They can be obtained from the shared public database, the Protein Databank (www.rscb.org) available worldwide. This service collects and stores a large amount of protein structural data procured from a variety of methods such as X-ray diffraction, NMR, as well as the recently expanding vista of Cryo-electron microscopy methods.

Structures obtained from the PDB resource come with drawbacks. Most structures originate from crystallographic data and due to the size of hydrogen atoms and the wavelengths generally used in X-ray crystallography falling between 1-2 angstroms, hydrogen atoms are not able to be resolved in molecular structure. The lack of core electrons for hydrogen atoms also means that no X-rays can be diffracted by sub-valence electrons.[186]

The topology and parameter files are dependent on the force field being used and are available for CHARMM for any class of molecule from the University of Maryland School of Pharmacy. A compendium of biomolecular components and their necessary force field files are available from the MacKerell lab homepage

(http://mackerell.umaryland.edu/charmm_ff.shtml). The topology files define how and which atoms are connected to one another. Equally as important, partial charges are defined in this file type. Parameter files contain the information on the numerical constants that NAMD needs in order to generate forces and energies. In molecular dynamics, bonds are treated with stiff degrees of freedom; the energy obtained from these calculations are accurate only under the assumption that the bond lengths are near equilibrium length. Every single type of bond will have a description with the topology file and every single bonding interaction will be described by the parameter file. With the combination of the two, a completed image of a protein can be created with all the structural information in tow.

The last designation needed to run simulations is the Protein Structure File format (PSF). It is a constructed file type that is generated from provided topological and atomic positional information. Built-in plugins in the NAMD software exist, like Automatic PSF Builder, and generate a structure file based solely on a raw PDB file with initial guesses as the states of each of the residues within the structure.

This can lead to problems for researchers like incorrect protonation state assignment on residues or improper chemical bonding. It is the responsibility of the researcher to investigate the reliability and accuracy of the guess generated by these methods.

Molecular dynamics simulations are carried out in four major steps that exist as their own individual simulation runs: minimization, heating, equilibration, and production. Each require initial atomic coordinates, protein structure, electronic topology,

and parameters to be input by a researcher and must be constant through each step of the procedure.

The minimization step is to relieve excess strain on a system, resulting in a minimum in potential as bond distances and angles achieve equilibrium value. The timesteps needed for this step is dependent on the system but generally within a few picoseconds of model time. The minimization process used by NAMD is performed using a conjugated gradient method combined with a line search algorithm that finds successive search directs along the potential energy gradient that result in a total minimization of a system with the limitation of not being allowed to move along same directions per each time step. The method then converges these steps in order to arrive at a (theoretically) global minima of the system.

The heating step is performed to raise the total temperature of a system to the designated temperature of the experiment. Using the Langevin piston as a control for temperature, and allowing pressure to vary, a system can reassign values of velocity to achieve a desired temperature. These velocity rescaling methods reinitialize the velocities and trajectories of individual atoms, with the resulting change in thermal energy calculated for the system at each time step until a set temperature is reached.[187] This is sometimes performed in annealing loops that cycle above and then back to a target temperature in attempt to cross any unpredicted local minima and to decrease the number of defects in the model. The actual heating process is performed in hundreds of time steps; however, the cycle loop generally takes the total time for this step into the several picosecond model-time range.

The equilibration step is performed in a constant pressure NPT thermodynamic ensemble that is done to equilibrate the system before properties or states can be measured. The output of these steps is useful for making decisions on the production run but are otherwise are not suitable for drawing conclusion upon. These simulations usually last for several nanoseconds and can be shorter or longer, depending on qualitative properties of the system, such as membrane bilayer association or active site solvation.

Once the model has been minimized, heated, and equilibrated, production runs are carried out to convey information about the system being studied. These simulations are done in the canonical ensemble (NVT) and can typically range from nanoseconds to even fractions of microseconds given the computational power available.

Methodological challenges define the movement towards refinement and improvement in Molecular Dynamics. The challenges in MD are simulation timescale and accuracy. Statistical mechanics efficiently samples the ‘correct’ thermodynamic ensemble. Algorithms and advances in computer science make the simulations run faster by designing faster algorithms and taking better advantage of current hardware, like the 10x increase in speed brought in recent years for some packages by that advent of graphical processing unit (GPU) based computing. Force fields are achieving higher accuracy without unduly increasing the complexity of a calculation aided by the further development of better parameterization methods. Data analysis research leads to being able to process larger volumes of data and help researchers in drawing scientific conclusions from their findings.

Applications of these methods are of great interest to the modern-day scientist. With these simulations of biological systems, there is a variety of pharmacological,

medical, and toxicological uses. The binding and dynamics of possible substrates to and within their potential protein target might provide the foundation for research in both experiment and theory. With simulations calculations requiring a ligand repeatedly visiting a binding pocket can be made, something that was not feasible with purely QM methods. Lots of sampling is required but is a lower throughput than ‘docking’ studies and incorporates many more physical effects.

Homology Modeling and Molecular Docking of Biomolecules

As was previously discussed, homology models are generally constructed through the use of crystallographic data. However, with the difficulties that arise from generating a crystalline protein sample, the number of structures is limited. There are thousands of different cytochrome P450 structures available, originating from a variety of species and with several substrates and modifications. However, there are many P450 isoforms with no structural data.[188]

This vacancy leads to the necessary task of homology modeling. Homology modeling refers to the sequence of procedures biochemists take to get fairly accurate guesses at the structural makeup of unsolved proteins using related species known as target or template structures.[189]

This method is generally resolved to a seven-step process:

1. Template identification and initial sequence alignment
2. Alignment optimization
3. Backbone modeling
4. Loops or deletion modeling

5. Sidechain modeling
6. Model optimization
7. Model validation

These steps can be performed by protein structure scientists in separate steps with bioinformatics tools available such as sequence alignment algorithms like those used in the Unipro UGENE[190] software or all together through structure prediction and threading services like the I-TASSER[191–193] server and SWISS-MODEL[194] modeling server.

Sequence Alignment and Template Selection

For researchers to identify regions and patterns in proteins or genetic information that might be functional, structural, or imply evolutionary relation, the unique sequence of biomolecules is commonly arranged and used to compare. This arrangement is typically performed in order to align residues or nucleotide base pairs between sequences so that identical or similarities can be more readily observed.

The amount, or degree, of similarity is a common metric for studying bioinformatics as it can be taken to be a rough measure of how much a sequence has been maintained by natural selection or by lineage. This level of sequence similarity, usually reported as a percentage, gives biomolecular scientists an ability to quickly observe patterns of repeated residue motifs and chart an evolutionary history for an enzyme's protein sequence. These generational sequences are powerful tools in the study of proteomics as the groups of highly conserved regions can indicate structural or functional importance. Absence or mutation is also important for understanding biochemical

properties for a protein, as modifications that are retained across species show potential for research or investigation.

Aligning sequences by hand was important at one time but has generally fallen to computer programming and data science, however, can still be done on smaller or highly similar sequences. Alignment can be done locally or globally referring to the selection of the alignment; local alignment consists of smaller stretches of sequence, while global alignment considers an alignment of every item in a sequence. Other alignment methods of protein sequenced include those that are based on 3D structure and 'best-fit' superpositions of protein backbones, such as seen in the Structural Alignment of Many Proteins (STAMP) package.

Iterative algorithms are used to speed up the alignment process, rapidly comparing pair-wise residues of a protein sequence to another of interest. Similarities of chemical environment such as charge, polarity, and acidity are used as metrics to analyze the individual differences between protein sequences. Motifs are commonly distinguished by these methods and can indicate secondary structure, binding domains, to analogous matches across species. They are used by the field of bioinformatics to quickly compare possibly related protein sequences, or even to predict possible relation or function.

Alignment methods in modern day bioinformatics are heavily reliant on sequence and structural alignment software. Many are built on hybrid fundamentals that integrate local and global searches, with multiple pairwise sequencing such as BLAST[195], MUSCLE[196], or CLUSTALW.[197] Some integrate computational phylogenetics software such as PHYLIP to use comparative methods of cross-analysis to study the convergent or divergent evolution of proteins and genes.[138]

Analysis and assessment of these methods are still reliant on researcher intervention as they can be prone to error. Credibility and significance are common metrics used to evaluate alignment and are represented in data science as scoring functions. Scoring functions, in general, assign penalties for sequence elements that are either misaligned, have a changed chemical environment, such as a polar residue in place of a hydrophobic one, or if a gap is necessary to achieve a higher overall rating for the alignment.

Template selection for homology modeling requires two major considerations: a high rate of similarity between the target protein sequence and a template sequence, and the need of an experimentally obtained structure (usually X-ray crystallography or NMR) for said template. Alignment and subsequent selection of a template sequence can be performed with ease by a protein structure scientist; however, it is limited by the amount of crystallographic or structural data present in online protein databases. Homology models themselves are poor templates as the propagation of error theoretically increases by a large amount but can still provide rough estimates for protein structure.

Structural Threading and Refinement

Generally, this method is performed through the use of online services provided by dedicated web-servers or institutions such as the I-TASSER (Iterative Threading ASSEmbly Refinement) server created by the Zhang lab of the University of Michigan.[191–193,198,199]

Many of these servers go through an iterative algorithm-based model prediction of 3D protein structure. The target protein sequence to undergo the homology modeling

process is given as input to the service and solutions usually takes several days or weeks, dependent on the length or complexity of a sequence and the number of job requests in line.

The protein structure prediction protocol begins with an initial backbone threading approach. The method used by the I-TASSER server is that of a locally installed meta-threading approach (LOMETS) where the protein sequence of interest is compared to similar structures within all available PDB libraries using a variety of threading methods.[200,201] Once several target protein structures have been identified, the amino acid sequence of the homology target can be threaded along the top models and result in a large number of templates generated. These go on to be processed by restraints such as secondary structural confinements based on residue torsional angle strain before the next step can proceed. Clusters of conformationally satisfied templates are collected and put through another restraint process, in the case of I-TASSER defined by energetics.

This method of alignment must be done carefully as such work is prone to error or inconsistencies in the prediction of structure. Typically, programs are used that optimize hydrogen bonding networks to have the lowest energetic state, such as the REconstruct atomic MOdel (REMO) package used in I-TASSER, to verify and refine initial structures.[202] Clusters of these now-refined structures are now ready to be further refined through cycles of energetic restraints and other refinement simulations.

The resulting cluster of predicted optimized structures are ready to be compared again to known crystal structure using a pairwise structural alignment algorithm, such as the TM-align algorithm used by the I-TASSER service.[203] This results in a root-mean-squared-deviations in atomic coordinates to be calculated for each model created. A set of

final models are then sent to the user along with additional data about their predicted function, enzyme classification (EC), and even binding pocket or site information.

A scoring element is commonplace for this method as well, with some examples such as I-TASSER modeling service's 'C-score 'and 'Z-score'. C-score is a confidence value used to estimate the quality of the prediction models generated by the I-TASSER platform. It is calculated by the significance of the threading template used and the convergence of parameters of the structure assembly simulations. The Z-score functions as a more threading program specific value; this scoring unit is based on the difference between the raw and average scores of each threading technique in the units of standard deviation. Now that the backbone of the homology model is in place on what is to be assumed to be a reliable 3D protein structure, the next problem to overcome is that of structure refinement and patching.

A common refinement technique is through measuring the displacement of atomic positions from a point of reference, either from crystallographic data, or from template structures. The root-mean-squared-deviation of atomic coordinates (RMSD) is a metric commonly used to evaluate a threading technique, as large displacements indicate a poor result of the homology modeling procedure.

$$\mathrm{RMSD}(\mathrm{r},\mathrm{s}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\|r_i - s_i\|^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}((r_{ix}-s_{ix})^2 + \left(r_{iy}-s_{iy}\right)^2 + (r_{iz}-s_{iz})^2)} \quad (15)$$

Equation 15 is a common representation of the RMSD used in structural biology, where '*r*' and '*s*' are the positions of atomic positions for the backbone heavy atoms of a

template and of a target structure. RMSD is expressed in units of angstroms for displacement and has an associated error in the same units of measure. Issues observed here generally originate from artifacts in the protein crystallography method, such as truncation or poor resolution.

The final template provided by some service platforms, such as I-TASSER, are based on crystallographic information which leaves many templates without hydrogen atoms due to their poor resolution in data collection. This may seem trivial for situations where the hydrogens can be placed reasonably so such as amide backbone hydrogen atoms or aliphatic residue sidechains. However, this becomes more problematic for residue sidechains with multiple protonation states, such as L-histidine with its three possible forms, (protonated on either side, or both).

Another refinement technique that is used to solve this problem are predictive services, used to determine the protonation states of ionizable groups in a protein structure. PROPKA is an example of a web service that utilizes the 3D electrostatic environment of the protein structure to make its predictions.[204]

Post-translational modifications and unique covalent linkages such as phosphorylation, protein glycation, or metal binding must also be affixed in this step of the homology modeling process. This is done through a patching process in which specific chemical bonds are designated in protein structural files, through use of explicit coding or through integrated plugins on molecular dynamics software packages. One last consideration for these modifications is that their exotic nature usually leads to the occurrence of improper or missing force field parameters.

The I-TASSER server provides functional prediction services in addition to some ligand binding predictions based on a method of comparing the threaded models with establish protein structures in order to draw conclusions on possible function, ligand binding sites, and classification. One of the protein characterization services used by I-TASSER and through its own web-server is the Computed Atlas of Surface Topography of proteins (CASTp) which goes through a similar process of comparing against curated protein databases.[205,206]

Parameterization using Quantum Mechanics and Molecular Mechanics

As mentioned previously, each term of the potential energy function used in the CHARMM force field describes a value designated to belong between atoms and the constants for each interaction are dependent on which atoms are interacting. Each interaction between specific atom types must have an empirically determined force constant and an equilibrium distance or angle. These values are referred to as parameter sets and the process to create them is called “parameterization”.

Force field parameterization is necessary to complete molecular models and their simplified description of reality. Most models have incomplete physics; many models place fixed point charges on atoms with no ability to demonstrate electronic polarization. Classical mechanics are currently unable to consider isotopic effects as is observed in experiment. Fixed bond topology prevents the fundamental chemistry of molecular systems from occurring. Bonds are rigid; the breaking and forming of bonds are disallowed. Despite these drawbacks, much can be recovered through the use of parameters. Modification of partial charges can recover some semblance of polarization

effects, additionally by tuning van der Waals parameters some systems show improvement in predicted density or other physical phenomena. Some force fields have even been observed to exceed the accuracy of quantum methods: the use of QM methods will result in thermodynamic phase behavior of water that is not seen in experiment, whereas the correct behavior is observed using multifaceted MD calculations.[207]

As an example displayed in Figure 11, the bond angle found on the center sp^3 carbon atom of propane would need a much different potential energy function to represent it than that of a ring-strained cyclopropane carbon.

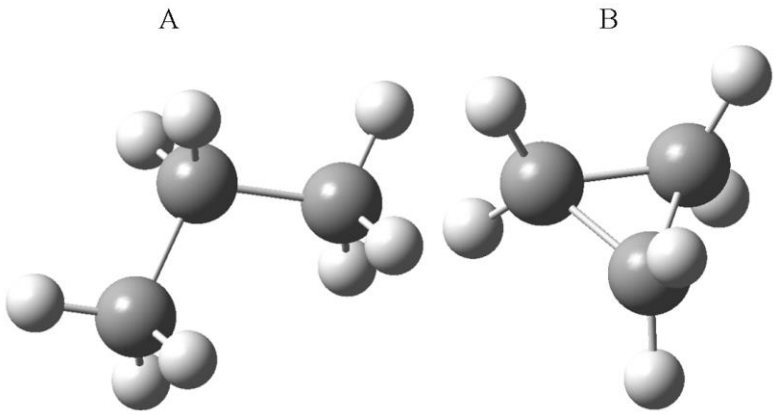

Figure 11. Ball-and-stick representations of propane (A) and cyclopropane (B). The angle strain imposed by the C-C-C bonding within cyclopropane molecule is apparent when compared to the linear propane molecule.

The parameterization process is reliant on high levels of model chemistry and incorporates almost all forms of computational chemistry in order to generate appropriate results. Quantum theory is used for the electronic aspects of bonding arrangements such as partial charge and bonding optimization. Density Functional Theory is used as an initial guess on geometry optimization with novel parameters, as it is less computationally expensive and will be further refined with more robust levels of theory. Multilayered calculations utilizing both quantum mechanical and molecular mechanical

methods are used to quickly converge on equilibrium bond distances and angles. These QM/MM output values can be added into simulations as additional force field parameter files.

Within CHARMM 36 format, there is a collection of a variety of chemical bonding interactions parameterized for almost every amino acid side chain imaginable. There is only need for parameterization of novel protein modifications, such as the covalent heme linkage of the CYP4 family of enzymes. This parameterization is commonly performed through computational chemical program suites like Gaussian and molecular dynamics software packages such as the Force Field Toolkit plugin of VMD and can be performed using the modular CHARMM general force field (CGenFF).[154,208,209] The CGenFF is a force field consisting of a general variety of generic atom types and their non-bonded interaction parameters for simulation of new biomolecular systems.[209]

The parameterization process can be broken down into 5 overarching steps:

1. Identification of Missing Parameters
2. Initial Geometry Optimization by DFT
3. Partial Charge Optimization through QM Water Interaction
4. Bond and Angle Optimizations through QM/MM
5. Dihedral Optimization by Torsional Scanning with QM/MM

After this process is performed, the results are compared to analogous compounds to ensure significance and relevance. Mistakes in this portion of the MD simulations will

not appear until after calculations are completed, making it a very important step in the modeling of biochemical ensembles.

Automatic PSF Builder, Solvate and Autoionize Plugins

Standard examples of peripheral plugins used in the simulation of biomolecules are those of Automatic PSF builder, Solvate and the Autoionize plugins contained in the VMD program suite.

Automatic PSF builder takes the atom types and charges in the form of the topology file and the atomic coordinates of a biomolecule that is segmented into individual portions (protein, co-enzymes, lipid bilayers, etc.) and generates a structural file that encompasses all of the individual bonding interactions in a program-friendly format. Written in a format more friendly for computing, this filetype serves as a part of the key files necessary to perform molecular dynamics simulations in addition to CHARMM force field parameters that dictate atom-atom interactions.

Proteins do not behave well in *in-vacuo* environments; therefore it is necessary for computational simulations to have solvent present. Through either implicit or explicit representation of solvents, proteins require the polar and collisional influences facilitated by solvent molecules. In the case of almost all biological realms found on planet Earth, water is the major solvent.

Several theoretical models for water have been designed with many considerations in mind. Some have been created with multiple representations of polarity through the addition of dipole functions. Others have dummy atoms to represent centers of mass for physical chemical purposes, and even rigid or flexible bonds to allow for

vibrational modes to be simulated. In the CHARMM force field, the TIP3P water model is used as an explicit solvent and can be added to protein systems in a myriad of ways and volumes. Spheres and cuboids of water solvent can be modelled and generated in the concentration designated by a user. The Solvate plugin calls for user input in the placement and geometrical arrangement of the added water, generating segmented atomic coordinate files for use in simulation.

Proteins are inherently susceptible and sensitive to the ionic strength of surrounding polar solvents. In order to buffer the electrostatic interactions of the structural ensemble, ions are added to completely neutralize the system. Additional ions are added to replicate the cytosolic concentration of ions in biological systems.

This is also is a requirement for the way that certain MD simulations are performed. The summation of electrostatic interactions, as calculated by the particle-mesh Ewald (PME) summation, requires the system to have a charge of zero. The Autoionize plugin queries from the user what ions are to be used, allowing for full customization of both cations and anions, as well as the preferred final concentration of salt in the final solution produced.

Molecular Docking through Automated Docking Tools

The process of molecular docking is imperative for understanding intermolecular forces, such as noncovalent interactions like electrostatic or steric, and is commonly used to study proteins and their ligands for use in rational drug design and discovery. Molecular docking methods utilize binding free energy to evaluate the stability of protein-substrate complexes, as well as identify conformations that might be of

importance to a researcher. They are performed using automated docking software or program suites such as AutoDockTools.[210,211]

Much like MD simulations, molecular docking uses its own set of parameters to describe the rigid protein structure, the flexible ligand to be docked, and the protein-ligand complex. The parameters typically consist of binding free energy and is determined through the summation of the net interactions within both the individual molecules as well as their interactions with one another. This includes bond distance, angle and torsional free energy while also taking into account total internal energy of the system before and after binding. The more biologically relevant interactions are calculated, with the highest amount of resources being used to approximate hydrogen bonding, electrostatic, dispersion and desolvation energies.

The general docking process consists of utilizing either a crystallographic protein structure or a homology model to define a protein as a receptor macromolecule. Next, a ligand must be chosen to dock to the receptor and can be collected from libraries and repositories online in a multitude of file formats.

To perform the binding of the substrate, the creation of a ‘grid’ of the target macromolecule is required to be mapped in the next step. This calculation results in a topological mapping of the receptor and is used to dock the target ligand, as the input only contain partial charges and atomic coordinates. Instead, it uses grid maps that describe desolvation and electrostatic interactions. These interactions are calculated using a probe atom of each specific atom type of the ligand, to record the energies and produce a corresponding grid map.

Search algorithms are used to find conformations of a ligand docked to the grid maps generated. Search algorithms seek to optimize the binding interactions between the ligand and grid maps, exploring different conformations of the ligand to result in better solutions. It searches the designated grid selection for the best conformations of the ligand and categorizes the results based on binding energy, produced from nonbonded interactions and desolvation.

Some examples such as genetic algorithms are iterative methods, based on 'mutation' and 'recombination' of previous solutions. Another example is that of the Monte Carlo simulated annealing, which is based on probabilistic optimization with neighboring conformations evaluated as possible solutions.[212] Hybrid search algorithms exist that combine searching techniques such as with Lamarckian Genetic Algorithm (LGA) which is a hybrid of a genetic algorithm and a Monte Carlo method.[213]

The two general approaches to molecular docking are shape complementary methods and simulation approaches. Shape complementary methods are quick and involve rapid scanning of possible orientations of a ligand within a predefined boundary assigned to a target molecular receptor. Simulation approaches are more intensive and involve placement of the ligand a distance away from the targeted binding site and performing calculated movement through time as the ligand is allowed to bind to the target receptor. Both methods have provided researchers with invaluable information on the usages of small ligand molecules in the study of *in silico* drug discovery.

Flexibility of both the receptor's binding site and the ligand are challenges present in the molecular docking process. The convention used in many methods is to designate a rigid receptor and allow for the small molecule ligand to be flexible around designated

bonds. This raises an issue in the study of biomolecules as the concept of the induced fit model of enzymes displays subtle changes in protein structure to bind a substrate.

More often than not, automated docking methods result in non-physiologically important results such as the binding of a ligand to the non-active surface of a protein receptor. The prevalence of these results is based heavily on the grid map work performed on a receptor, as larger grids result in more impractical binding conformations of a ligand.

Assessment of molecular docking results, like in homology modeling and molecular simulation, is required in order to determine the confidence in its predictive ability. One challenge that hinders automated docking methods is the lack of a standardized scoring function. What is commonly used is binding free energy, however this metric varies widely in value between methods and techniques. Benchmarking with confirmed binding modes produced by X-ray crystallography is a common method of evaluation of docking programs.

AutoDockTools is a program suite of automated molecular docking tools.[210] It is commonly used across the fields of study and has shown application in chemical mechanism studies, protein-protein interactions, structure-based drug design, and even organic synthetic chemistry. AutoDockTools is an example of shape complementary methods and encompasses several docking programs.

AutoDock Vina is a quicker method of docking for most proteins, as the automated program creates predefined grip maps for the user, limiting the amount of research influence and customization possible.[214] The major drawback is that due to these predefined maps, only the canonical 20 amino acids can be used in the receptor, allowing

no room for post-translational modifications, metal ions, or adducts to be represented. This makes it largely unusable for phosphorylated, glycated or metal-binding proteins.

AutoDock is the more modular method and has less limitations placed on it, allowing a user to use a gridding software, AutoGrid, to create custom grid maps for any macromolecule imaginable. This gives the researcher the ability to model binding interactions with metal cations present, or covalently bound modifications through explicit descriptions of the receptor.

These methods seek to further the understanding in specific interactions between small molecule ligands their biomolecular targets. Their roles in the activation or in-activation might lead to computational insight gained in the prospective pharmacological importance of small compounds. Molecular Docking is used to screen large databases to rapidly detect potential drug interactions *in silico* with confidence. It is also can be used to predict and improve how future modifications to a drug might impact and improve its physiological activity; thus, leading to the creation of more potent and selective analogs.

This method was used in the pandemic of the novel coronavirus disease 2019 (COVID-19) as the interaction between the small molecule remdesivir was flagged as an early potential therapeutic as it acted as a nucleotide analog to disrupt the interactions between the SARS-CoV-2 viral RNA-dependent RNA polymerase and their target viral nuclear information.[215]

CHAPTER III

Molecular Dynamics Simulations of Substrate-free Cyp4f13 in DOPC bilayer

Introduction

This chapter presents the process, results, and findings in the creation and simulations of the murine substrate-free Cyp4f13 structural ensemble. The resting state of the murine P450 catalytic cycle needed to be modelled and put through simulation to obtain a structure ready for the molecular docking procedure in the next set of experiments. This work was necessary as the molecular docking of the substrate of interest required use of a sufficiently equilibrated model of the protein's catalytic site, able to portray physiological significance. Another focus of this work was to gain qualitative understanding on the dynamics that the substrate-free P450 exhibits while being embedded in its native environment of a DOPC lipid bilayer. The intention is to develop insights that will allow for a better understanding of the dynamic binding interactions between the murine P450 Cyp4f catalytic domain and its fatty acid substrate.

Computational Method and Details

Template Selection with use of UGENE Sequence Alignment Toolkit

The free bioinformatics software UGENE by Unipro was used to perform multiple sequence alignment, analysis, and subsequent template selection for use in homology modeling. Sequence similarity consisted of Hirschberg (KAlign) alignment, a pairwise method that bases its sequence similarity off of a string-matching algorithm.[216]

The alignment of P450 CYP4F subfamily members was performed in order to elucidate common structural motifs amongst the isoforms as well as across species. Further alignment of murine P450 Cyp4f13 was performed on a set of more studied P450s within the PDB database in order to identify possibly target templates for homology modeling. The specific P450 isoforms used in the alignment will be presented in the results.

The alignment was performed with the MUltiple Sequence Comparison by Log-Expectation (MUSCLE) method of multiple sequence alignment. A phylogenetic tree for the CYP4F subfamily was built using the PHYLogeny Inference Package (PHYLIP) neighbor-joining tree building method. Consensus and phylogenetic analysis were performed using the UGENE software interface.

Initial Alignment and Backbone Threading using I-TASSER

In the creation of the cytochrome P450 structural ensemble, several components need to be constructed and prepared for molecular dynamic simulation. The first and foremost moiety of this ensemble is the protein structure of the homology model of murine cytochrome P450 4f13.

The murine cytochrome P450 4f13 model was assembled from structural analogs in the PDB library using the I-TASSER service for protein structure prediction. The best structural analogs were evaluated for their relevance and similarity then were used to thread the Cyp4f13 sequence to, resulting in an initial structure. After refinement and the usage of sequence similarities to predict function, the best model is determined by the criteria of RMSD of atomic positions and transmembrane anchor position.

Using the FASTA format of the Cyp4f13 protein sequence, a request was sent to the I-TASSER server in the early months of this research. After a brief waiting period, several homology models were produced by the service. The models were individually assessed using a variety of criteria, one of which was the state of the transmembrane anchor. The position of the transmembrane anchor being parallel with bilayer leaflets would result in a shorter equilibration period for subsequent molecular dynamics simulations, so the model with the most appropriately positioned anchor was selected for the homology model. This model would be used as the P450 catalytic domain segment (CYP) in all the ensemble construction and simulations.

Homology Model Refinement

The homology model produced by the I-TASSER process would have predicted coordinates for the position and protonation of the residues based on crystallographic data which do not contain information on the nuclei of hydrogen atoms. This would lead to results that might not be representative of the protonation states present in the biological environment. The protonation states of the homology model would need to be refined and were externally predicted by the PROPKA 2.0 web-server (http://www.propka.org). This holoenzyme model would then be ready for covalent attachment with its heme cofactor through patching. The Computed Atlas of Surface Topography of proteins (CASTp 3.0) web-server was used in addition to visual inspection methods to assign residue make-up of channels within the Cyp4f13 homology model active site. Several sites were identified, including the heme cofactor binding site.

The prosthetic heme cofactor must be additionally supplied in the form of atomic coordinates, force field parameters, and electronic topology in order to prepare the protein for simulation. As was previously discussed, the covalent attachment of the heme moiety and the acidic residue in the I-helix in other CYP4 enzymes is absent in native murine Cyp4f13 and was not necessary for this series of simulations. The parameter files and topology necessary for the heme-thiolate patch were derived and developed from previous work adapted from the AMBER force field and determined by quantum chemical calculations on simplified heme models.[217]

To achieve an optimal fitting of the heme moiety, the X-ray crystal data available for the heme domain of $P450_{BM-3}$ was used with STAMP structural alignment in order to collect atomic coordinates. The STAMP structural alignment, although outputting sequence alignment information, only uses 3D structure in its alignment and is an appropriate method to align the Cyp4f13 and $P450_{BM-3}$ heme domains. Once positioned, the atomic positions of the heme cofactor were copied from the $P450_{BM-3}$ coordinate file and placed into the Cyp4f13 active site with no steric conflicts. Ligand docking methods were available, however spatial alignment gave no steric conflicts and served as an acceptable alternative.

Embedding of Homology Model into Lipid Bilayer

There are several methods such as the VMD plugin Membrane Builder that can be used to create membrane lipid layers and bilayers. After successful creation, membrane-anchoring alpha helices can have their atomic coordinates placed into the crafted membrane and conflicting lipid molecules can be removed through command line in the

Tk console in the VMD program. Brute-force methods like these are necessary in the absence of crystallographic data of these biological ensembles.

Fortunately, through a personal communication, a cytochrome P450 embedded into a lipid bilayer was obtained by Professor Michael Otyepka from Palacky University Olomouc of the Czech Republic. The structure contained the cytochrome P450 3A4 catalytic heme domain integrated with a lipid bilayer composed purely of di-oleoyl phosphatidylcholine (DOPC).

Molecular Dynamic Simulations of Substrate-free Enzyme System

Each step of the molecular dynamics was performed in separate simulations, run in bulk through basic shell commands used in the Linux operating system. The parameters and topology used were the CHARMM36 protein, lipid and solvent files obtained through the MacKerell lab homepage of the University of Maryland School of Pharmacy. Modification of the topological file for hemeproteins was performed in order to simulate the heme-thiolate with an iron(III) oxidation state.

After removal of the overlapping or colliding nuclear coordinates and proper segmentation of the lipid bilayer, the solvent plugin for VMD was applied. This action was performed to the designated to-be simulated region of space that contained the protein and membrane bilayer. It extended past the bilayer for 10 Å and above the protein catalytic unit by 25 angstroms in the Z axis while confined to the dimensions of the lipid bilayer in both the X and Y directions. This would allow for the proper application of periodic boundary conditions in order to simulate a more real impression of a P450 embedded into the endoplasmic reticulum, which is crowded with neighboring proteins.

The periodic boundary conditions applied to a 130 by 130 by 150 angstrom defined space size. These dimensions were determined after the solvation step through a script that uses the volume of the solvent box in order to find cell basis vectors and a point of origin.

A minimization process of 20 picoseconds was used on the system. Afterwards, a heating step of 30,000 timesteps was used to climb to 310 K using the velocity rescaling temperature coupling scheme for 12 ns before locking in temperature for a total of 60 ns of simulation time. This was performed in an NVE thermodynamic ensemble, controlling the volume with the use of the Langevin dynamics used by NAMD. Next, the equilibration step of 30 ns was performed using an NPT isothermal-isobaric ensemble and was immediately followed an unrestrained 300 ns NVT canonical ensemble production simulation.

An integrator timestep of 2 was used (2 fs per timestep), and rigid hydrogen bonds were applied to the entire system, including solvent TIP3P water molecules. This was necessary as rigid bonds are required by the integrator when using larger timesteps. The cutoff for non-bonded interactions was 12 angstroms for all simulations. The Particle Mesh Ewald method was used in order to retain full-system periodic electrostatics and recover long-range interactions between the periodic images. The grid spacing for the PME summation was set to 1 Å as it is the default for this method. The theoretical Langevin piston was set to have a pressure of 1 atm (1.01325 bar) for the equilibration run with a default dampening coefficient set to the same temperature of 310 K. The trajectory of every particle was saved every 2 picoseconds and all velocities for the simulation were generated with pseudo-random seeds.

Results and Discussion

Template Selection of Cyp4f13 Homology Model

The sequence alignment of the I-helix of cytochrome P450 4F subfamily members was performed to verify which of the CYP4 proteins contained a covalently bound heme. This portion of the I-helix, as seen in Figure 12, corresponds to the location of a conserved glutamic acid residue in CYP4 enzymes, the residue that forms the ester linkage, indicated by the abbreviation of E in the alignment.

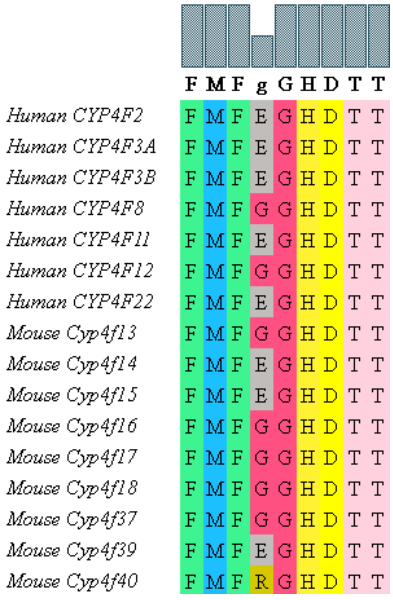

Figure 12. Sequence alignment of the I-helical region between cytochrome P450 4F enzymes. Enzymes from human and murine origins are included, along with a simple consensus of the sequence above the alignment. The case of the consensus indicates the level of conservation for a residue.

Important to note is that several mouse isoforms have a glycine residue, indicated as a G, with the notable example being murine Cyp4f13. This is important as CYP4

enzymes are generally considered ω-hydroxylases in function, and its catalytic ability to do so is dependent on the presence or lack thereof the crosslinked heme.

Drug metabolizing and bacterial P450s are the majority of P450 structures available within the Protein Databank due to their use in drug discovery and biomolecular sciences. A portion of the alignment of these structures with murine Cyp4f13 is shown in Figure 13. While a large variety were selected, some individual isoforms, such as rabbit CYP4B1, were selected due to the property of maintaining a high amount of structural similarity (>40%) owing to belonging to the same family.

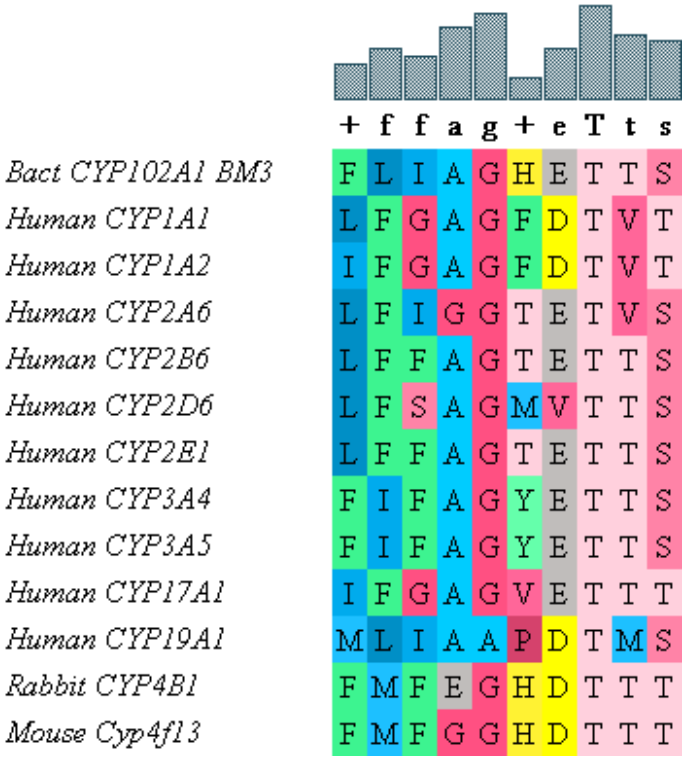

Figure 13. Sequence alignment of murine Cyp4f13 and several P450s with crystal structures available. The "+" consensus indicates that there are multiple residues with high levels of frequency in the alignment.

Table 3

The Similarity Percentages of P450s with crystallographic data and murine Cyp4f13

P450 enzyme	Sequence Similarity Percentage
Bacterial CYP102A1$_{BM-3}$	20%
Human CYP1A1	19%
Human CYP1A2	18%
Human CYP2A6	16%
Human CYP2B6	17%
Human CYP2D6	18%
Human CYP2E1	18%
Human CYP3A4	22%
Human CYP3A5	22%
Rabbit CYP4B1	42%
Human CYP17A1	16%
Human CYP19A1	17%

Note. The entire sequence of P450$_{BM-3}$ was used in the MUSCLE sequence alignment. Sequence Similarity Percentage was calculated using the KAlign sequence matching algorithm.

Table 3 displays a surprising amount of similarity for the bacterial CYP102A1 (P450$_{BM-3}$) as the fused protein contains a much larger sequence with its redox active domains a part of its residue chain. The heme domain of P450$_{BM-3}$ shows a relatively high amount of sequence similarity (>20%) between itself and the murine P450 4f13 isoform, though this is reasonable as P450$_{BM-3}$ has been shown to be an active participant of fatty acid hydroxylation similar to the CYP4 family. This observation is important as the soluble bacterial protein structure has more in common with the murine cytochrome P450 4f13 than many of the membrane-localized eukaryotic P450s, making it an important metric in order to investigate individual residue roles in catalysis as mutant BM-3s have been extensively studied.

Threading of the Murine Cyp4f13 Protein Sequence

The top three identified analogs were the mammalian rabbit CYP4B1[218], human microsomal CYP3A4[219], and the heme domain of bacterial cytochrome $P450_{BM-3}$[21].These were used to generate several possible models with varying levels of viability. Their scoring, according to the TM-align algorithm for sequence independent structure comparisons where 1 is a perfect fit, were 0.898, 0.791, and 0.790 respectively.

The model with both the highest amount of statistical probability and optimal transmembrane anchor position was chosen; this model (Model1) was used for the molecular dynamic simulation runs in the experiment.

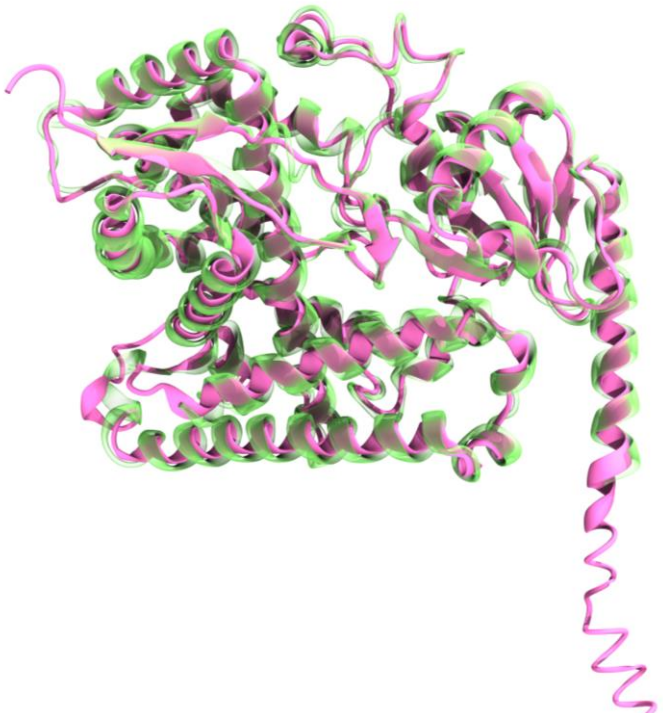

Figure 14. Structural overlay of the template P450, CYP4B1 and Model1 threaded by the I-TASSER service. The template structure and homology model are represented in green and pink ribbon, respectively. The RMSD difference between the N-terminal transmembrane anchor positioning between the two is relatively small with optimal positioning.

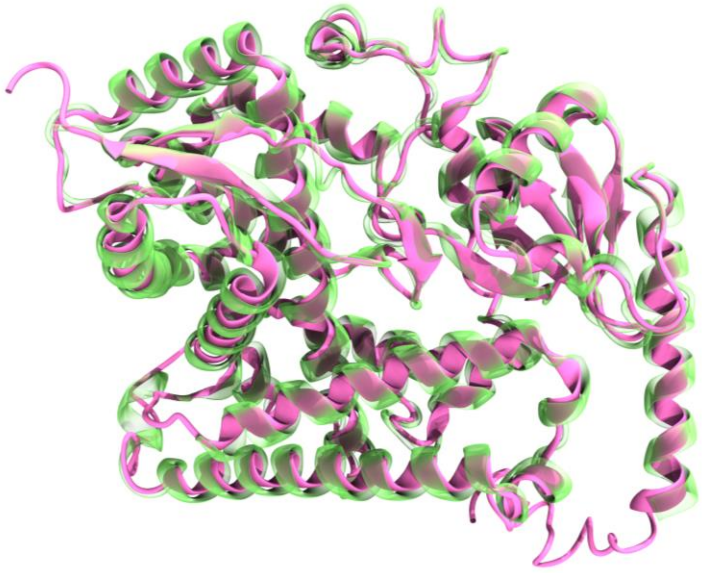

Figure 15. Structural overlay of the template P450, CYP4B1 and Model2 threaded by the I-TASSER service. The template structure and homology model are represented in green and pink ribbon, respectively. A tucked-in N-terminal transmembrane anchor is observed, indicating it as a poor choice for the homology model.

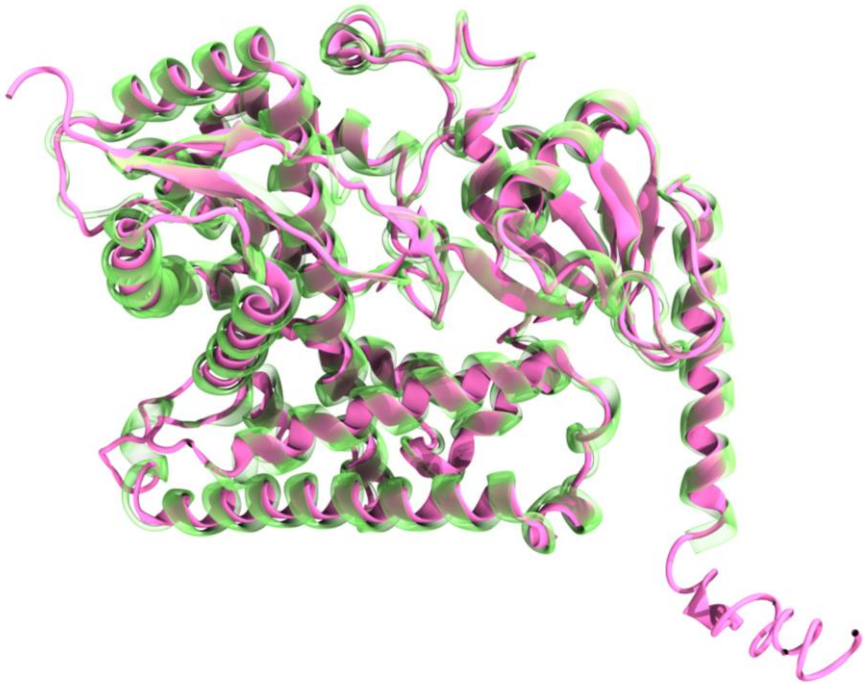

Figure 16. Structural overlay of the template P450, CYP4B1 and Model3 threaded by the I-TASSER service. Poor fitting is displayed along some of the secondary structure with an outwardly extended N-terminal transmembrane anchor.

The C-score for Model1, shown in Figure 14, is 0.05 with a range of -5 to 2 and the best threading method indicated by Z-scoring through LOMETS was the MUSTER method.[220] The RMSD of Model1 superimposed to the 3D structure of the rabbit CYP4B1 is 0.43 Å; this indicates a good fit of the backbone to the template and ready for heme-thiolate patching and bilayer anchoring.

Several ligand binding locations were identified for the model, predicted by the TM-Align algorithm of the I-TASSER service. Notable examples were that of a heme binding domain, to little surprise, and that of a few small molecule binding predictions.

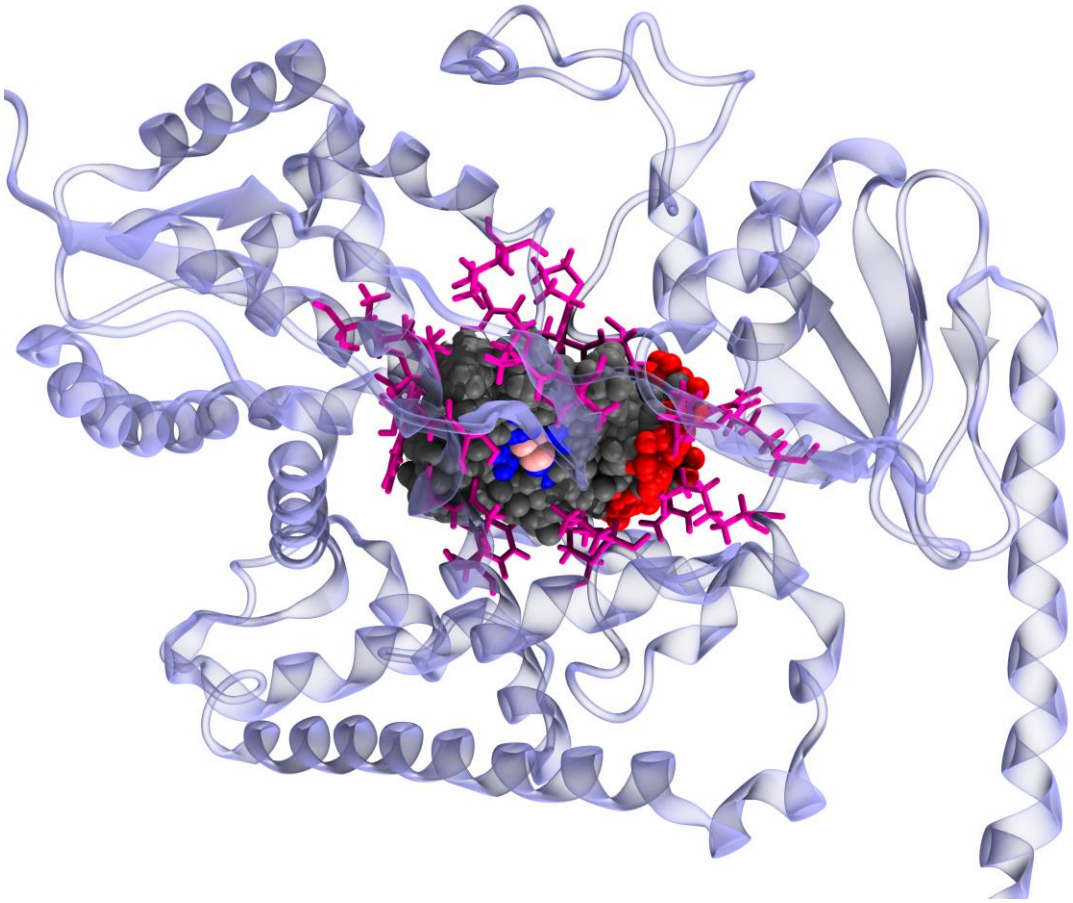

Figure 17. Cluster of the predicted heme binding sites on Model1 I-TASSER generated Cyp4f13 homology model.

While providing as an excellent template for homology modeling, the crystal structure used in the threading process was based on the rabbit cytochrome P450 4B1, which contains the covalent heme linkage indicated as being important in substrate binding. While serving as the best choice, the presence of that linkage should be considered as it might impact the threading results produced.

Initial template selection was performed before a 2018 study resulted in marginally better crystallographic data of a mutant CYP4B1 enzyme without the covalent heme linkage, which could have provided as a better initial structure for the native Cyp4f13 isoform.[221] Later Cyp4f13 homology models produced in this study would utilize this newer model as a template and the differences observed appear to not be significant or are resolved by the simulation process.

Model Refinement: Protonation States and Heme-thiolate Prosthetic Group

As determined by the PROPKA service, several histidine residues were predicted to have differing protonation states (HSD, HSE, or HSP) than the default (HIS) residues that the homology model process generated. A listing of these predicted protonation state changes are in Table 4.

Table 4

Listing of Histidine Residues within Cyp4f13 and their Protonation State Changes

Residue Index in P450 Cyp4f13	Protonation State
20	HSD
63	HSE
79	HSP

(continued)

Residue Index in P450 Cyp4f13	Protonation State
82	HSP
85	HSD
88	HSE
103	HSE
146	HSE
147	HSD
156	HSE
175	HSD
194	HSE
236	HSE
264	HSE
308	HSD
329	HSD
346	HSD
391	HSE
427	HSD
428	HSD

Note. HSD refers to the protonation of the delta nitrogen, HSE refers to the protonation of the epsilon nitrogen, HSP refers to the bi-protonation of a histidine residue.

With the homology model of the holoenzyme produced, the heme cofactor would need atomic coordinates in appropriate positions within the heme-binding catalytic domain of the P450. The conserved heme-thiolate bond found in all P450s would have crystallographic data within wild-type protein structures and its spatial coordinates should vary amongst them due to different catalytic states being captured by the crystallization process. These variations, while important, are not too large in magnitude and can be resolved with subsequent MD simulation in order to refine their positions. The heme-domain of cytochrome $P450_{BM-3}$ was used with STAMP structural alignment capability

of the Multi-seq plugin that is commonly used to compare protein structure in the VMD program suite.

In the covalent attachment of the characteristic heme-thiolate prosthetic group, ‘patching’ was performed to create an appropriate protein structure file. Patching is a process built into model generation where residues can be fused or otherwise altered. The patching was done through three distinct steps: the removal and reassignment of charge of the conserved cysteine residue (Cys467) in the murine Cyp4f13 enzyme, the reassignment of charges for the entire heme-thiolate cofactor, and then the addition of a new Fe-S covalent bond formation between the biomolecule and the cofactor with a final charge reassignment. This was done using a script in the Tk console of the VMD program suite and the successful heme-thiolate cofactor is shown in Figure 18.

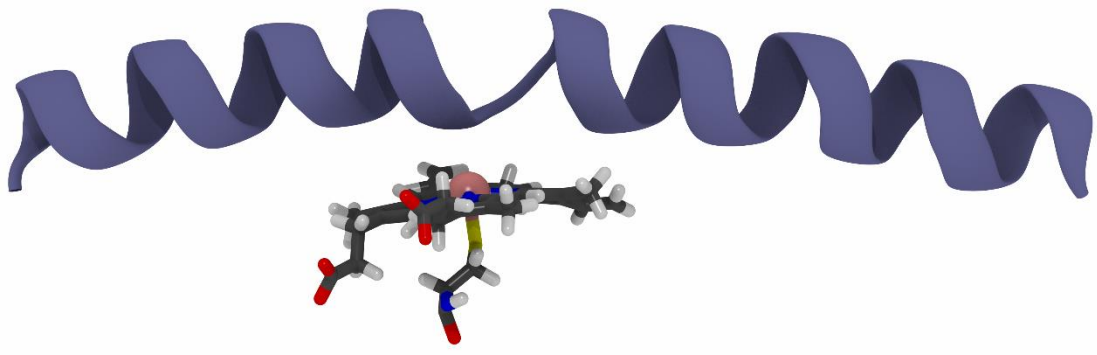

Figure 18. Structural view of the murine Cyp4f13 apoprotein with the heme-thiolate prosthetic group represented in a licorice cartoon to show distinction. The bond between the heme-iron and the Cys467 is observed, indicating proper patching of the cofactor.

Positioning of Transmembrane Anchor into Membrane Bilayer

Using STAMP structural alignment, Model1 was aligned to the position of the CYP3A4 embedded in the DOPC membrane that was received from P450 researchers, pictured in Figure 19. While this contribution was significant in helping the progression of the ensemble creation process, it came with a few difficulties. The membrane bilayer

had truncation at the ends of the lipid region and contained missing segments and indexes. These would need to be removed or modified as their presence would result in a failure in the creation of the protein structure file. This problem was overcome by using atomselect tcl commands and removal of offending lipid units.

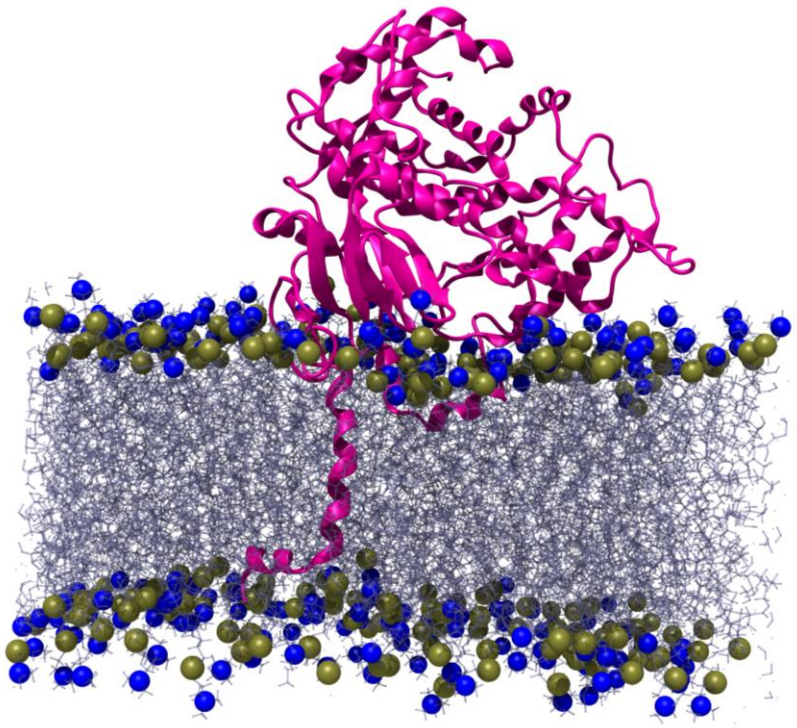

Figure 19. Structural view of the CYP3A4 catalytic heme domain model embedded in DOPC bilayer. The coordinates of the CYP3A4 enzyme and bilayer were received through personal communication with the authors from the 2018 paper on interactions between CYP3A4 and the lipid bilayer.[47] Protein structure shown in pink ribbon, and lipid in modified ice blue licorice models. Phosphate and ammonium groups are represented by yellow and blue van der Waals sphere respectively.

Another problem was that the large size of the lipid bilayer coordinate file did not allow for the re-indexing by Automatic PSF Builder to perform correctly and resulted in a fatal error in the construction process. To overcome this obstacle, a script was run that segmented the file into chunks of 8-10 DOPC lipids and then afterwards output their

coordinates through use of the atomselect and writepdb functions in the Tk console of VMD.

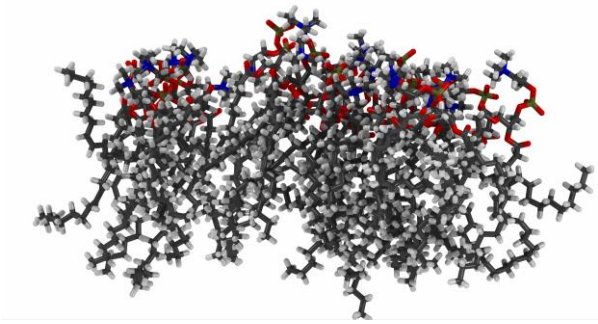

Figure 20. Structural view of a single DOPC lipid segment (PC10) containing 10 di-oleoyl phosphatidylcholine lipids.

Each of these segments, an example of which demonstrated in Figure 20, were localized to the working directory of the constructed model and would be used by Automatic PSF Builder to generate the membrane embedded murine P450 model.

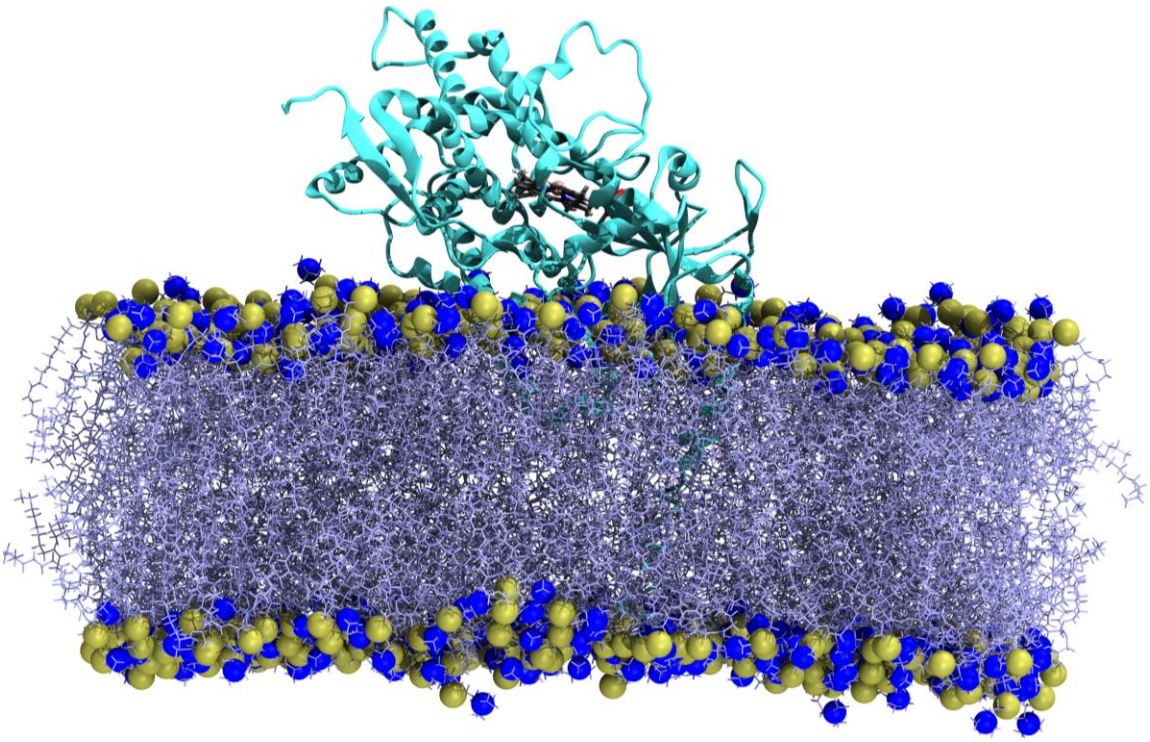

Figure 21. Structural view of the membrane embedded murine Cyp4f13 homology model. vdW radii were added to distinguish phosphates (yellow) and ammonium (blue) groups.

The next step was to remove lipids that overlapped with the embedded protein, occupying the same space. The transmembrane anchor of the murine P450 packed into the same region as the human CYP3A4 did, and thus resulted in a low amount of the membrane lipids needing to be removed. Removal of entire lipid molecules was performed, as opposed to individual atoms, as the later would have resulted in errors in the indexing of the files and missing parameters.

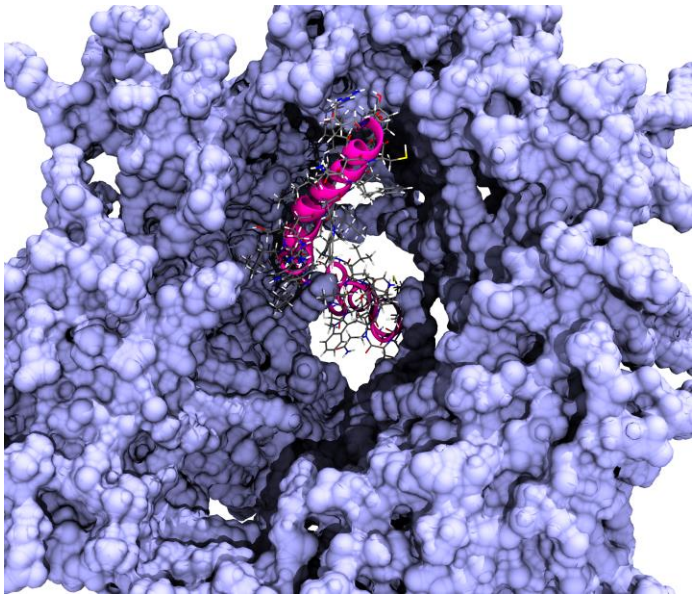

Figure 22. View of the excision region surrounding the truncated transmembrane anchor of the murine Cyp4f13 homology model before solvation and ionization procedures were applied. The protein ribbon is colored in magenta, with DOPC lipids displayed as quicksurf models in an ice blue coloring and represent the calculated electronic isosurface.

As shown in Figure 22, large portion was carved and removed from the atomic coordinates. This cylindric region, shown in Figure 22 and measuring roughly 10 angstroms in diameter, would need to be taken into account as the solvation protocol would result in solvent being placed in the absence of the removed DOPC lipids.

Solvation and Ionization of Cyp4f13 Embedded in DOPC Membrane

The solvent plugin for VMD was applied and successfully added 27 additional segments of water molecules, each containing several molecules of TIP3P water models. The pre-equilibration structure file is shown in Figure 23. The solvent box was set slightly larger than the dimensions of the lipid bilayer, resulting in extraneous water being added to nonpolar regions, which would need to be allowed to vacate during the equilibration run.

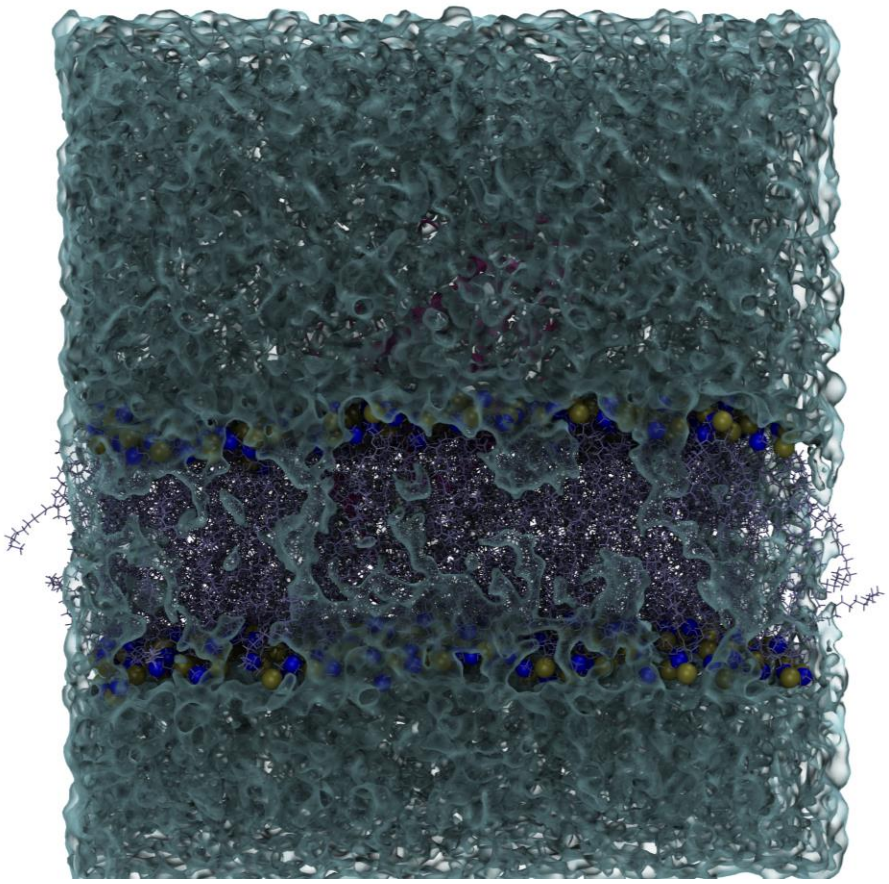

Figure 23. Side view of Cyp4f13 homology model after TIP3P solvent box was added. A surface was added with light blue coloring to all water molecules with transparency. Van der Waals radii were added to phosphate and ammonium groups of the DOPC lipids. Licorice representations were used on the di-oleoyl acyl groups.

The autoionize plugin of the VMD software was used to add counter ions to the system. The plugin was set to add potassium and chloride ions to balance charges to a total of 0 net charge for the system and to result in a concentration of 100 mM to approximate physiological conditions. This resulted in a non-integer total net charge but was approximately 0 with 108 K^+ and 108 Cl^- ions being added, pictured in Figure 24.

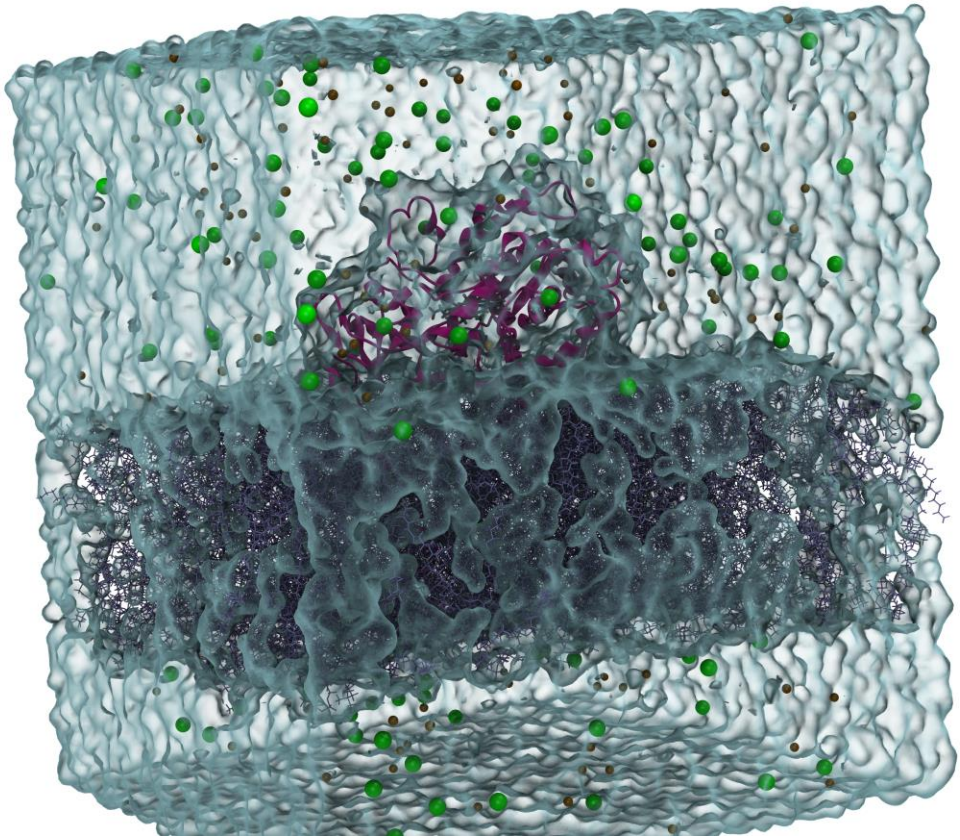

Figure 24. The locations of counter ions in the Cyp4f13 system with van der Waals radii representations for individual potassium and chloride ions.

Molecular Dynamics Simulations on Substrate-free Cyp4f13 Structural Ensemble

To gain computational insight into ω-hydroxylation of arachidonic acid by murine cytochrome P450 4f13, a representative snapshot with the ligand-free Cyp4f13 needed to

be created from the homology model structure. The final coordinates of the production simulation on the murine Cyp4f13 homology model were to be used as this snapshot, provided that adequate equilibration and active site solvation was achieved. The entire simulation procedure consisted of the structural ensemble, which contained explicit solvent, membrane bilayer, P450, and ions being subjected to an initial minimization, heating, equilibration and long timeframe production simulation.

The minimization simulation was observed to be a success as relative total energy plateaued to convergence within several hundred 2 fs timesteps of the calculation. The entire process lasts for 20 picoseconds of simulation time, with the first 100 femtoseconds shown in Figure 25.

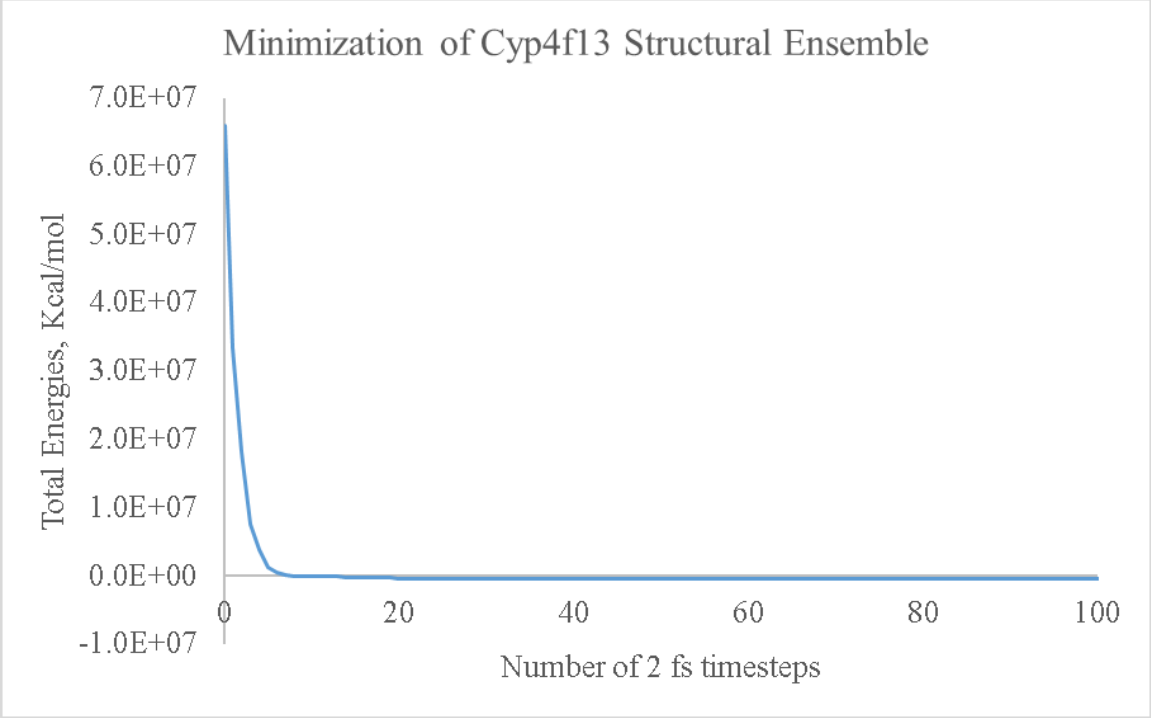

Figure 25. NAMD Plot of the calculated total energy for the system during the minimization stage. Total energy consists of each of the bonded and non-bonded energetic components summated for their contributions.

To enhance the mimicry of the natural physiological environment of the murine protein, heating of the system was performed to ensure that the appropriate biological temperature (310 K) was reached. Average temperature varied but remained consistent as the system reached body temperature, as is observed in Figure 26.

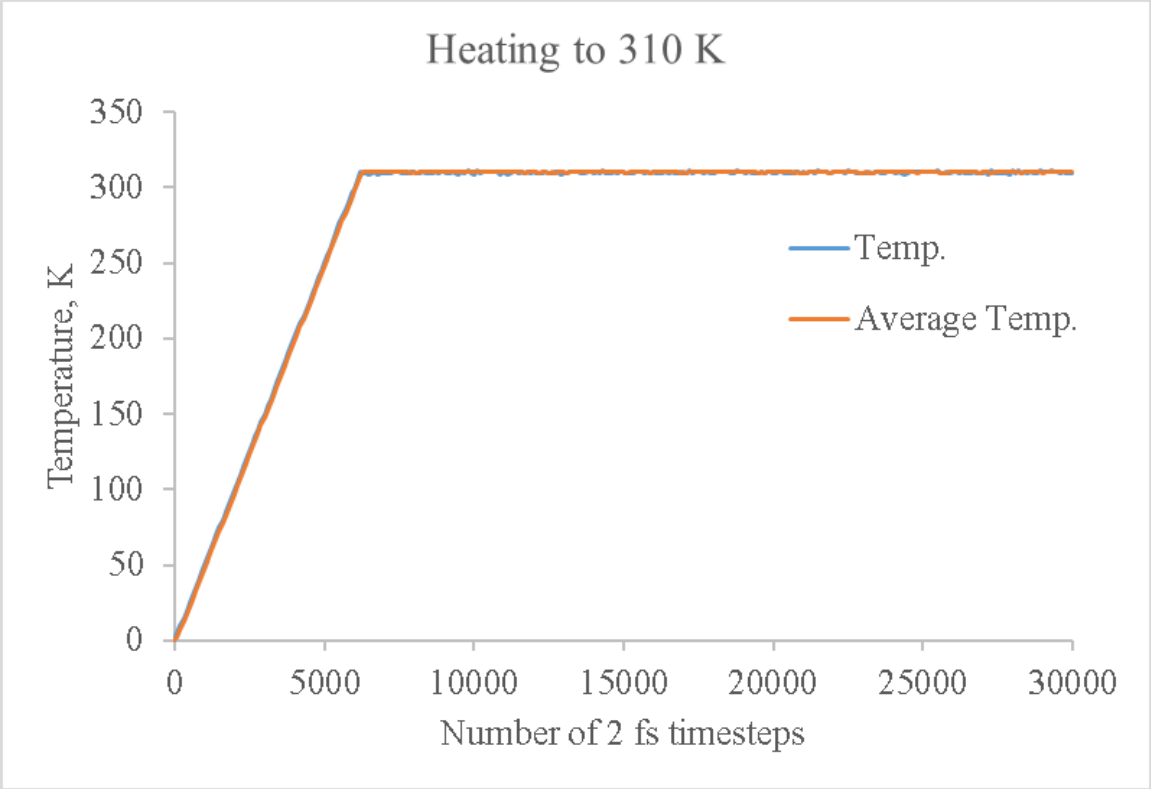

Figure 26. NAMD Plot of the theoretical temperature of the system during the heating stage. This simulation was performed in an NVE thermodynamic ensemble with volume as the constant variable. A constant rise to 310 K is observed with reasonable fluctuations after the initial climbing procedure.

To achieve best results in the molecular docking portion of the experiment, a well-equilibrated model should be used as a best first approximation to use as input for the docking procedure. Another consideration is the presence of a vacuum generated by the membrane positioning from the ensemble creation process. This sort of event is not commonly something that occurs in a cellular membrane as it would be expose an

organelle to a change in pressure or environment and needed to be worked out in the NPT equilibration step of the simulation.

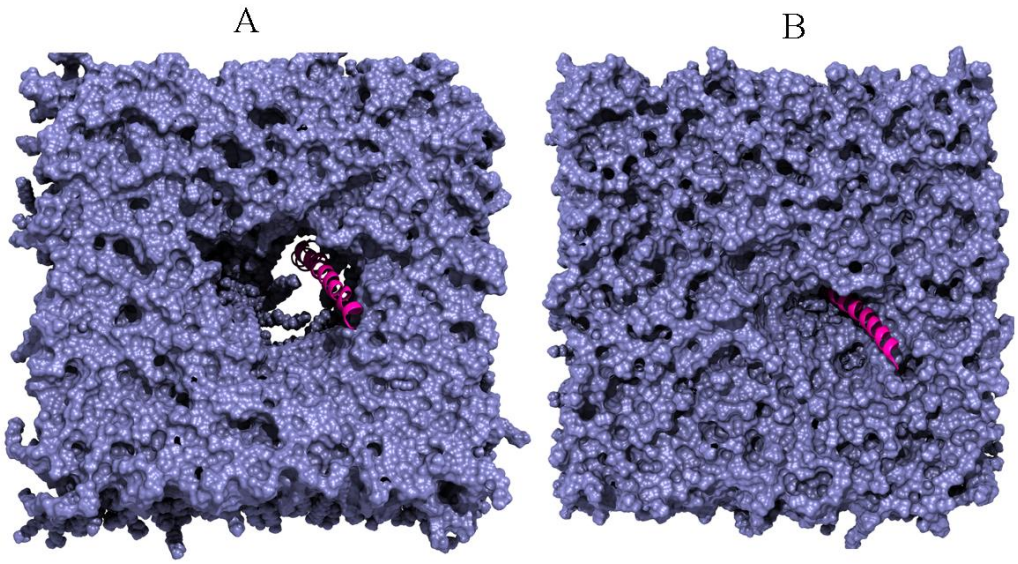

Figure 27. Top view of the membrane bilayer before (A) and following (B) the NPT equilibration simulation run on the Cyp4f13 homology model. The excision vacuum produced by removal of the conflicting DOPC lipids has closed around the protein structure, represented by the truncated Cyp4f13 transmembrane anchor in a ribbon representation.

The equilibration shown in Figure 27, appeared successful as the lipid membrane exuded the water molecules from around the transmembrane anchor of the Cyp4f13 homology model. This meant that the system now adhered to proper requirements necessary for the periodic boundary conditions.

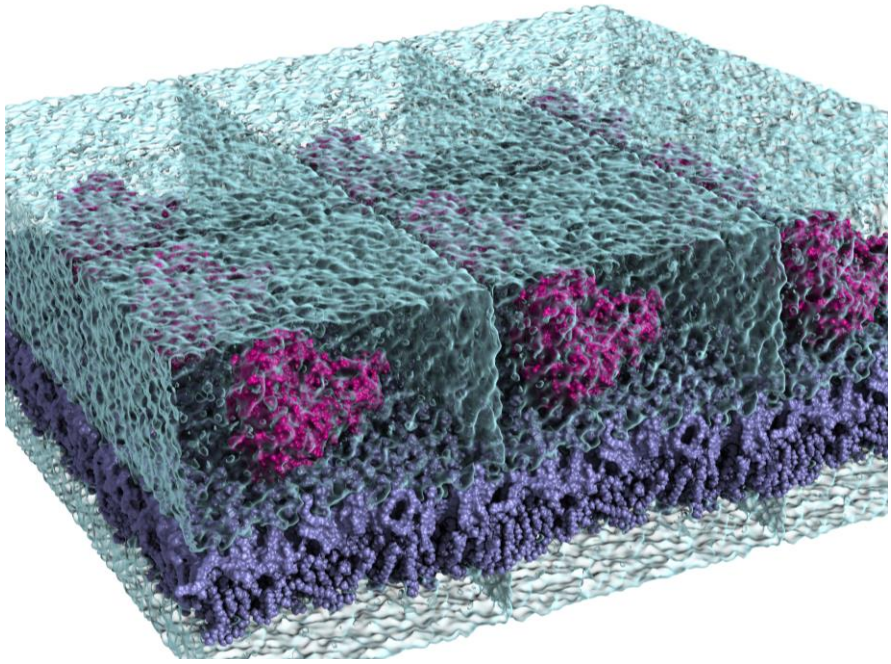

Figure 28. View of the murine Cyp4f13 structural ensemble images after equilibration simulation. Note that water has completely vacated the nonpolar region, as periodic boundary conditions have led to the complete formation of a lipid bilayer

After some optimization of cell boundaries, periodic boundary conditions were successfully achieved and no breaches in the membrane layer were observed after equilibration run. Important to note that water has completely vacated the nonpolar region, as is observed in Figure 28, as periodic boundary conditions have led to the complete formation of a lipid bilayer.

In a previous run, some periodic cell boundary distances resulted in the formation of lipid nanodiscs and are available as supplementary material. While amusing, these P450 nanodisc ensembles do not make for a good representation of their physiological environment in the endoplasmic reticulum. An error made in creating the periodic cell boundaries appeared to be the causing factor, as a solvent box much larger than the membrane patch resulted in hydrophobic forces to dominate the interactions. The interactions of the water with the polar headgroups of the lipids caused the charged

phosphate and ammonium groups to face the bulk water while pressuring the nonpolar oleoyl substituents towards the center, resulting in the formation of a nanodisc-like lipid capsule. Reassuringly, these geometries are what one would expect in the case of phospholipid confined to a small space without enough lipid to form a full lipid bilayer, providing some validation of the lipid dynamics in the system. A script was run to generate more refined descriptions for cell origin and periodic boundary dimensions based on the solvent water box and not the lipid bilayer and these optimized boundaries were used for all subsequent simulations.

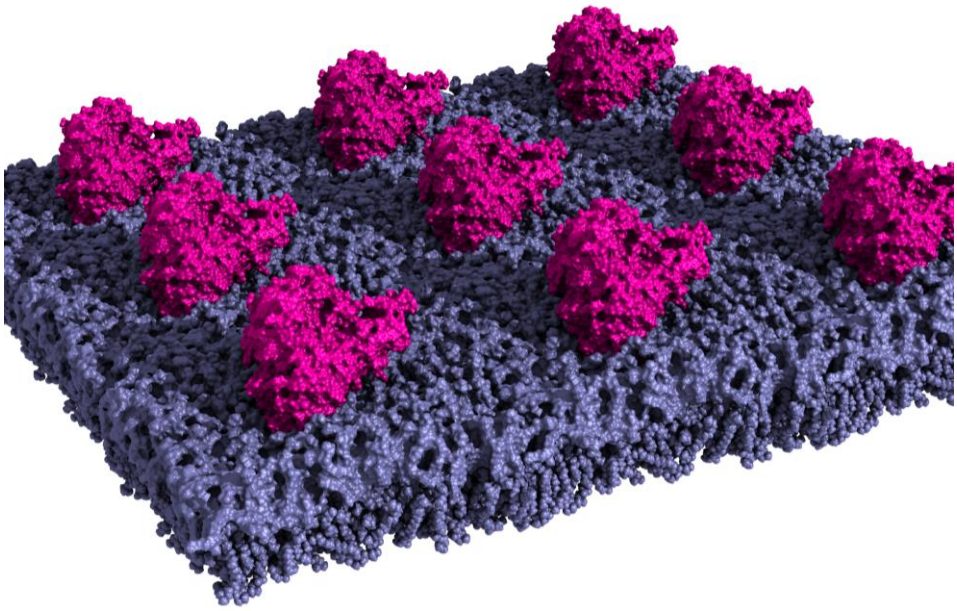

Figure 29. Structural view of the properly produced periodic images of the murine cytochrome P450 4f13 heme domain embedded into the DOPC bilayer. Periodic images in the Z dimension, additional repeating units in the X and Y dimensions, counter ions, and water not shown. The periodic cell repeats infinitely in the X, Y, and Z dimensions in the full calculation model.

Arrival at the Resting State of the P450 Catalytic Cycle

The next goal was to use the production run of the simulation to solvate the active site of the Cyp4f13 homology model, to produce the resting state of the P450 catalytic

cycle. In this state, a solvent water molecule has been found to occupy the open coordination site of the heme-thiolate metal center.[21] What would also need to be performed is solvation of the protein's natural internal access and egress channels. This was be achieved by allowing the simulated solvent water molecules to map and cave the protein's native channels. To visually inspect this, the trajectories of every TIP3P water molecule in a region around the heme cofactor segment was represented as van der Waal spheres and tracked over the 300 ns timeframe of the production simulation, as seen in Figure 30.

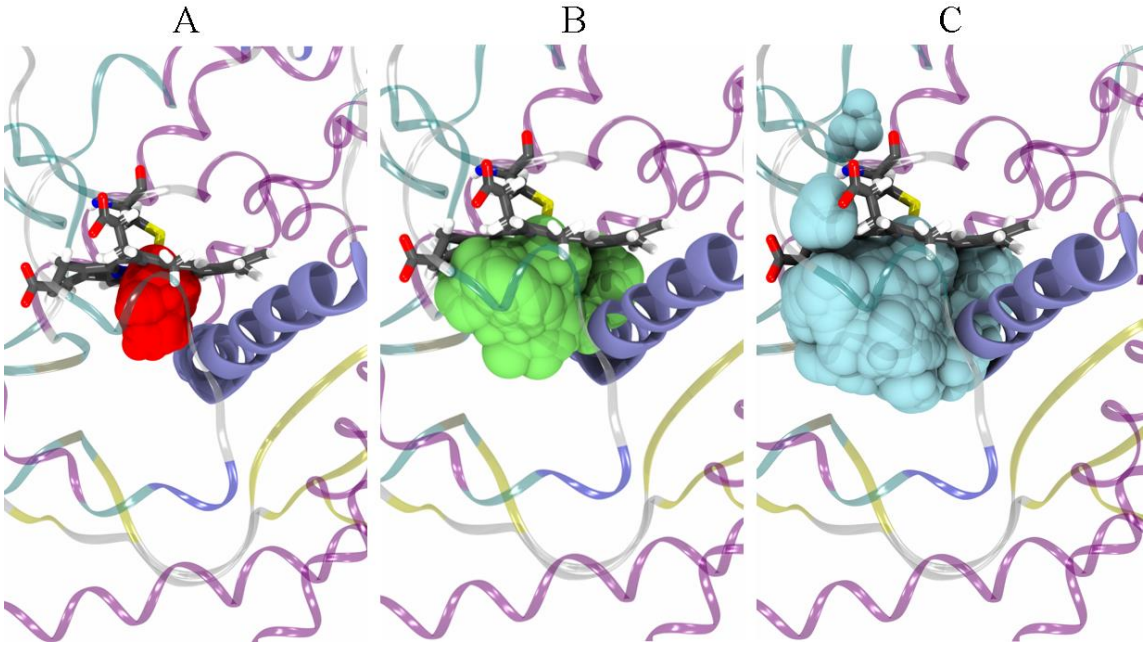

Figure 30. Pocket produced by simulation of water molecules within the Cyp4f13 active site of the homology model. The trajectories of all water molecules within a distance of 3, 5, and 7 angstroms from the heme cofactor were mapped in strides of 50 frame intervals for the entire 300-ns simulation and are indicated in red (A), green (B) and blue (C) respectively.

Solvent accessibility is a commonly used metric to study the characteristics of protein active sites. With many P450s having integral function dependent on the vicinity of nearby water molecules next to the substrate, proper channel evaluation would need to

be assessed. Visually inspecting water trajectory is a common method used in theory; however, prediction services are also used by researchers in addition to expedite the process.

Several channels of the Cyp4f13 homology model were found utilizing both methods of visual inspection and through software by inputting the final coordinates of the Cyp4f13 homology model into the CASTp 3.0 web-service. Figure 31 shows the amino acid residues lining the assigned access channel (A) and the solvent egress channel (B).

Table 5

Channel assignment and residue makeup of the Cyp4f13 homology model made by visual inspection

Channel Assignment and Opening	Residue Name and Index
Access Channel (Into lipid bilayer)	Trp37, Asn44, Pro55, Trp59, Phe60, Trp61, Gly62, Leu64, Leu66, Met67, Lys68, Ile77, Leu80, Ser90, Trp91, Val92, Gly93, Pro94, Tyr96, Pro97, Ile98, Arg100, Lys120, Glu121, Thr123, Leu124, Tyr125, Phe127, Leu128, Leu137, Val232, Val233, Arg235, Arg237, Gln238, Pro239, Tyr242, Phe326, Gly327, Thr331, Val394, Leu396, Ile397, Ser398, Arg399, Cys400, Gly416, Asn417, Ile418, Val420, Pro501, Glu502, Leu503, and Ile504.
Solvent Egress Channel (Out to cytosol)	Leu128, Trp131, Leu132, Leu136, Leu137, Leu151, Phe155, Leu160, Tyr163, Val164, Phe167, Met191, Phe192, Ile195, Ser196, Met198, Thr199, Leu200, Leu203, Ile207, Phe208, Tyr221, Ile222, Ile225, Leu226, Ser229, Val263, Thr267, Ile271, Phe297, Ile298, Leu301, Ile317, Arg318, Glu320, Ala321, Asp322, Thr323, Phe324, Met325, Phe326, Gly328, Asp330, Thr332, Thr333, Ser334, Leu336, Ser337, Ile468, Phe472, Ala473, Ser475, Glu476, Ile477, Val479, Ala480, Arg506, Ala507, Gly510, and Leu511.

Note. Residue identification was performed by using solvent accessibility of residues within the active cavity and network of accessible channels out of the protein structure

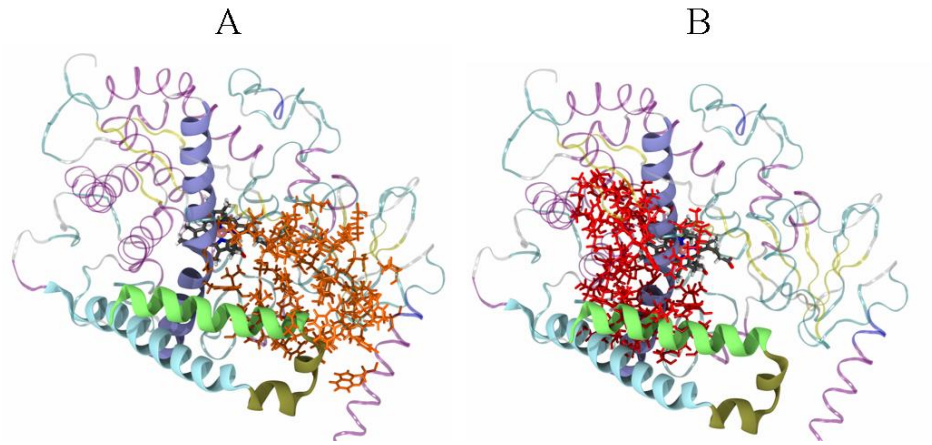

Figure 31. Distribution of amino acid residues selected by visual inspection as channel residues. Residues for the substrate access and solvent channels are placed along the ribbon protein structure and are represented as orange and red licorice models respectively.

Several residues were indicated by the CASTp service as predicted to have an important role in substrate binding, however further study would be done to investigate their potentials, as the docking with substrate would be more indicative than the solvated substrate-free active site.

Table 6

Predicted residue makeup of the Cyp4f13 active pocket by the CASTp 3.0 analyst service

Binding Site	Residue Name and Index
Predicted Active Site Cavity	Trp37, Asn44, Phe60, Trp61, Gly62, Leu64, Leu66, Met67, Lys68, Phe76, Ile77, Leu80, Ser90, Trp91, Val92, Gly93, Pro94, Tyr96, Pro97, Ile98, Arg100, Val102, Lys120, Glu121, Thr123, Leu124, Tyr125, Phe127, Leu128, Trp131, Leu132, Leu137, Phe192, Ser196, Thr199, Leu200, Leu203, Ile 222, Ile225, Leu226, Leu228, Ser229, Ser230, Val232, Val233, Lys234, Arg235, Arg237, Gln238, Pro239, Tyr242, Val263, Thr267, Asp322, Met325, Phe326, Gly327, Gly328, Asp330, Thr331, Thr333, Val394, Leu395, Leu396, Ile397, Gly416, Asn417, Ile418, Val420, Ser422, Lys500, Pro501, Glu502, Leu503, Ile504, and Arg506.

Note. The CASTp service predicted a large number of mouth and pockets within the Cyp4f13 structure, only the highest volume pocket was chosen to be assessed.

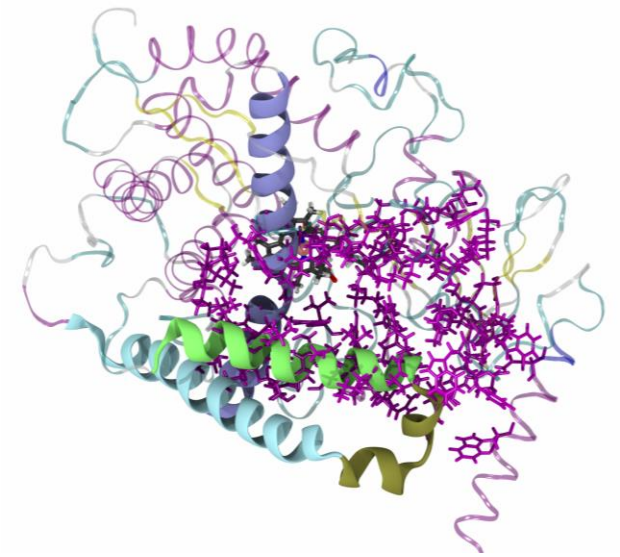

Figure 32. Visual representation of the Cyp4f13 homology model active site pocket through the CASTp 3.0 analyst service interface. A pocket volume of about 2000 cubic angstroms is predicted by the analyst service. Residues predicted to be involved in formation of protein mouth or pocket structure are represented in purple licorice models.

The last goal of this section of the experiment was to arrive at a model of the first step of the P450 catalytic cycle which is generally considered to include the ligation of a solvent water molecule to the open coordination site of the iron(III) species of the heme-thiolate.

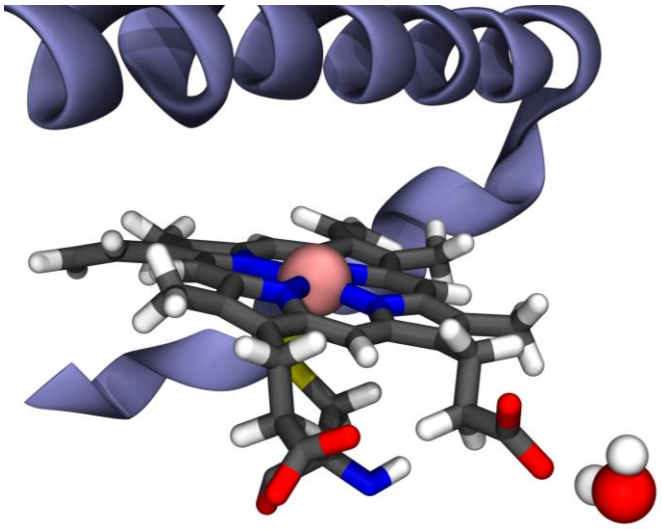

Figure 33. Close-up view of the Cyp4f13 active site before the equilibration procedure. Licorice and van der Waals representations were used for the heme, heme-iron and water

molecules within 4 angstroms from the cofactor. A vacant coordination site on the heme cofactor is observed.

As seen in Figure 33, before NPT simulation was performed the catalytic site of the heme cofactor displayed an empty coordination site at the axial position of its metal coordination sphere. Coordination of water molecule was observed quickly through the simulation process, resulting in the ligand association seen in Figure 34.

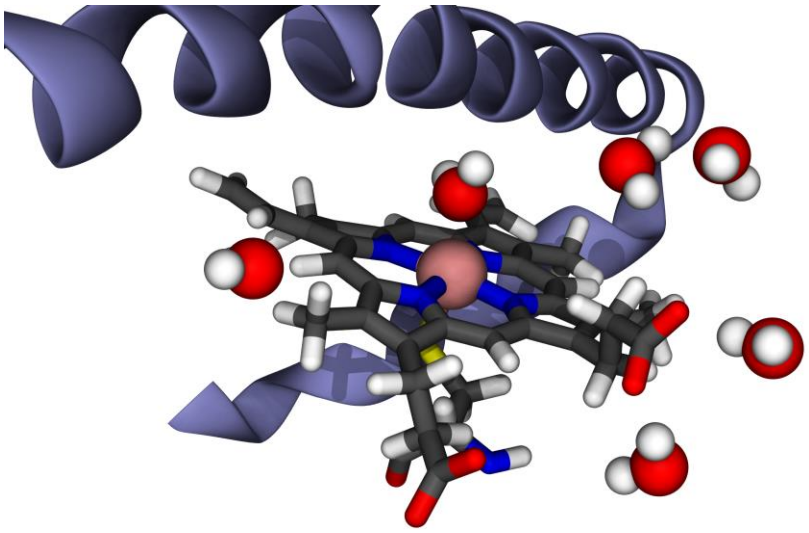

Figure 34. Same close-up view after the equilibration procedure on the Cyp4f13 homology model. The vacant site is now occupied by a ligand water molecule after the equilibration MD simulation.

It is important to note that MD simulations, such as the ones performed in this study, do not take phenomena such as orbital interactions into consideration. The force field can replicate the formation of coordinate bonds through Lewis acid and base interactions based on only coulombic electrostatic interactions. The orientation of the TIP3P water molecule, while reassuring for the relevance of this model, produced by the simulation is due to nonbonded interactions between the iron(III) and the oxygen of the water, with no information available about orbital interactions between the two.

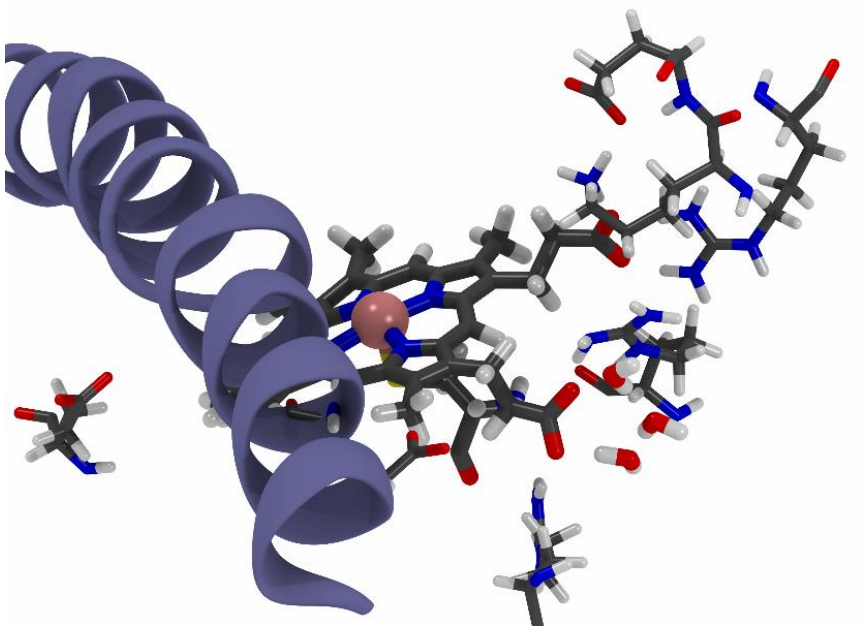

Figure 35. Structural representations of the heme cofactor and the interactions of its propionate groups with active site residues. Licorice representations for the nearby charged protein residue sidechains are colored by element.

The interactions between the propionate groups of the heme-thiolate prosthetic group are known to be very important in the formation of channels for water to enter and access the catalytic site. These aqueducts are known to shift and change formation due to the arrival and binding of substrate and will be important to monitor as the simulations of the docked arachidonic acid are performed. Salt bridging interactions between the heme propionates, Arg399, Arg465, Lys120, and several water molecules are observed in the substrate-free state.

With the first stage of the catalytic cycle modeled and channels and tunnels of the protein structure mapped and investigated, the substrate arachidonic acid could then be docked to the simulated murine Cyp4f13 homology model.

Summary

In this chapter the constructed substrate-free murine cytochrome P450 4f13 structural ensemble is discussed and studied in two components: the globular catalytic domain embedded into the pure-phospholipid bilayer composed of DOPC, and the active site.

The simulations of the substrate-free cytochrome P450 4f13 system were performed to prepare the structure for molecular docking, and as a refinement technique for subsequent simulations. The substrate-free enzyme also acts as a control in order to compare the successive simulation results with to study the binding interactions, and potential tertiary structure effects of the substrate on its biomolecular target. Simulations of the binding interactions of the P450 system and its substrate can be produced and studied with model systems. Both experimental and computational studies of model cytochrome P450 systems have elucidated the nature of interactions that occur in real systems. The study of substrate interactions with P450s with Molecular Dynamics simulations has been the subject of much discussion in the literature.

Arguments are made against using these methods to model protein systems that include concerns about the difficult to model nature of the complex native physiological environment. The work in this thesis hopes to address some of these considerations through the construction of this model and its validation by comparison to results predicted by chemical theory or by experimental findings available in the literature. The resultant findings for this new model have been discussed in this chapter, indicating success for the simulation as the water-coordinated, heme catalytic resting state was found to be prevalent within the allotted simulation and equilibration time. The protein

structure remained intact, did not destabilize, and the protein's embedding in the membrane had the expected orientation and general geometry. However, the conclusions made on the production simulation require consideration of the nuances in using this computational method. The simulations are only a thermodynamic sampling of the possible solvated states of the Cyp4f13 active site. Longer timeframe simulations, in the microsecond range or longer, would be necessary to reduce the impact of the potential differences in protein structure and conformation of sampled states.

CHAPTER IV

Molecular Dynamics Simulations of Substrate-bound Cyp4f13 in DOPC bilayer

Introduction

This chapter presents the process, results, and findings in the simulation of the substrate-bound murine Cyp4f13 structural ensemble. This work was necessary as the probing of the active site relies on a deep level of understanding of the molecular interactions behind substrate binding on an atomistic scale. The substrate-free state of the murine P450 catalytic cycle was to be modelled appropriately and docked with several conformations of the substrate of interest, arachidonic acid. Additionally, the oxidized eicosanoid metabolite, 20-hydroxyeicosatetraenoic acid (20-HETE), was used to investigate predicted product interactions with both active site residues and solvent.

With substrates docked, several hundred nanoseconds of simulations were performed on two physiologically relevant conformations of the arachidonate substrate, with one performed on the oxidized product. The focus of this work was to investigate residue interactions of importance within the active pocket, in addition to exploring the structural characteristics of the murine Cyp4f enzyme. The intention is to develop insights that will allow for well-informed decisions to be made on mutations that might alter catalysis and binding activity. This is done to further the understanding of the binding interactions between the murine Cyp4f13 catalytic domain, its fatty acid substrate, and its products.

Computational Method and Details

Substrate Docking of the Cyp4f13 Homology Model

The representative structural ensemble from the first production run was selected for analysis and subsequent molecular docking. The final coordinates of the 300-ns production simulation on the substrate-free P450 system (filename: Cyp4f13-memb-solv-ion) were stored in order to be used as the rigid macromolecule to be used in the docking procedure employed by the AutoDockTools (ADT) software.

The target ligand for molecular docking would be that of the polyunsaturated fatty acid substrate of interest, arachidonic acid. The molecule was constructed in Gaussian 09w using the GaussView 05 graphical user interface. Initial geometry optimizations were carried out at the Density Function Theory (DFT) level, using the B3LYP functional with the 6-31G(d) basis set. [222]

Preparation of the rigid Cyp4f13 receptor for docking consisted of several steps to generate and calculate the grid maps for the protein-ligand interactions. Foremost, the water, lipid bilayer segments, and counter ions were removed to ensure proper ligand binding interaction energy calculation. Second, all nonpolar hydrogen atoms were merged or removed, and polar hydrogen atoms were added or remain unchanged for every residue in the structure, as the failure to do this results in a mismatch of atom types and will result in a crash. Next, charges are calculated and added through the Gasteiger method which is based on electronegativity equilibration rather than distinct template values used by other methods. These charges and coordinates are saved into a new filetype (PDBQT) which is the input AutoDock uses for the molecular docking process.

Finally, a similar formatting method is applied to the target ligand, with the addition of rotatable bond designations on the ligand to be docked. For arachidonic acid, 14 rotatable bonds were found. A root is identified by AutoDockTools as the 'central' atom that the rest of the molecule is built from. This has more of an application on larger or rigid-body aromatic ligands but was still necessary to be performed in order to generate an acceptable PDBQT file.

With the rigid receptor PDBQT file of the Cyp4f13 catalytic domain selected, the location and extent of the search space could be designated. A 3000 $Å^3$ region was determined based on the CASTp predicted active site volume. The atom types of the ligand were identified, and the grid map creation process was performed by the AutoGrid component of ADT. The resultant set of grid maps for the rigid receptor was obtained and saved as a Grid Parameter File (GPF) for use in the searching step. The search algorithm used was the Lamarckian Genetic Algorithm (LGA) to generate the top 10 conformations with a population size of 150 solutions. Between each generation of conformations evaluated, only one conformation was set to automatically survive as the best individual.

Several conformations were predicted for the binding of arachidonate to the cytochrome P450 holoenzyme coordinates. These conformations were organized in order of their calculated binding energies by the ADT software. Of the 10 conformations produced by the searching algorithm, several conformations were rejected due to illogical conformations of the arachidonate substrate for oxidation by the heme cofactor. The top two conformations of physiological relevance were selected, and their coordinates were

used for MD simulations on the Cyp4f13-AA complex (referred to as D1 and D2-Cyp4f13-AA).

Product Docking to Cyp4f13 Homology Model

The oxidized product, 20-hydroxyeicosatetraenoate, was also targeted for docking and subsequent simulation. The final coordinates of best conformation of arachidonic acid were used as input into Gaussian 09w program suite and hand-modified to add a hydroxyl group to the terminal carbon position. As with the substrate, initial geometry optimizations were carried out at the DFT level, using the B3LYP functional with the 6-31G(d) basis set.

These set of coordinates were once again sent through a docking process using the ADT software package and through the same specifications as the arachidonate ligand. Of the conformations predicted, only the top individual chosen to be relevant for study was selected. The criterion used was that the hydroxylated product would be required to have similar carboxylate interactions with the same residues present in the binding of arachidonate, as this would allow for comparison between the docked homology models. This resulted in a single reasonable docking conformation of the 20-HETE ligand to be chosen for simulation (D3-Cyp4f13-20-HETE).

Molecular Dynamic Simulations of Substrate-bound Enzyme System

After the initial refinement simulations were performed, the 100-ns MD simulations were conducted for the first docked Cyp4f13-AA complex. All refinement and longer timeframe production simulations was performed using the same

specifications as the substrate-free simulations. A total of 200-ns of production simulation time was carried out for the D1-Cyp4f13 model. Three consecutive 100-ns simulations were performed on both the D2-Cyp4f13-AA and D3-Cyp4f13-20-HETE models. The simulation on the last two substrate-bound Cyp4f13 complexes were performed for identical lengths of time to make comparison easier. The reasoning, results and findings will be discussed in the appropriate sections.

The CHARMM36 lipid/detergent topology and parameter file contained the arachidonate compound (ARAN), which was fortunate, as this meant that it was not necessary to create topology or parameters for this residue from scratch. The simulation of the D3-Cyp4f13-20-HETE differed slightly and required more work to set up. The difference was the usage of additionally modified lipid parameter files as the hydroxylated product, 20-HETE, contained a unique topology that the CHARMM36 package did not contain. The creation of a HETE residue in the parameter file was added through direct addition by chemical analogy. Ethanol was the analogous compound used in the instance of a hydroxylated primary carbon belonging to an aliphatic chain, as is present in 20-HETE. The modified parameter file was only used for the simulation of the Cyp4f13-20-HETE complex.

Results and Discussion

Molecular Docking of Target Substrates using AutoDockTools

After the successful creation of the substrate-free resting state for the murine Cyp4f13 homology model, the coordinates were used as the rigid molecular target for

which to utilize molecular docking procedures upon. This was done to find relevant conformations of arachidonate within the active site in the interest of investigating ω-hydroxylation catalysis by Cyp4f13.

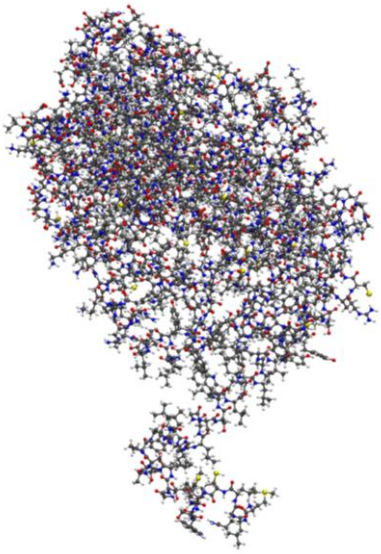

Figure 36. Ball-and-stick model view of the final coordinates of the substrate-free Cyp4f13 homology model used in the docking before charges were added and rigging was performed. Lipid bilayer, ions, solvent water, and heme cofactor were removed in this process.

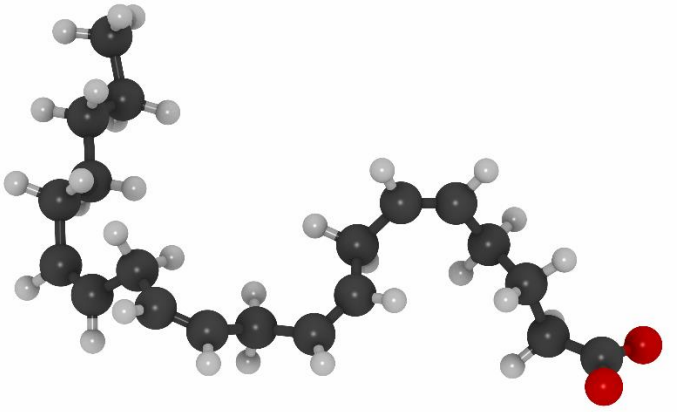

Figure 37. A ball-and-stick representation of the manually constructed arachidonate ligand. GaussView 5.0 was used to create the nuclear coordinates along with Gaussian 09w for initial geometry optimizations.

Shown in Figure 37, the geometry optimized structure of arachidonate resulted in a non-hairpin conformation with appropriate number of explicit hydrogens by the computational method used to generate it. 14 torsional degrees of freedom were detected, Gasteiger charges were added, and the C10 carbon was selected as the root by AutoDockTools.

The ten results of the docking procedure ranged in values of -3.7 to -2.0 kcal per mole in binding free energy, however many of them were considered not physiologically important. Some carboxylate head groups of the arachidonic acid coordinated to the heme center. This was an interesting observation, however, did not provide useful for investigating the potential role of Cyp4f13 as an ω-hydroxylase of arachidonic acid.

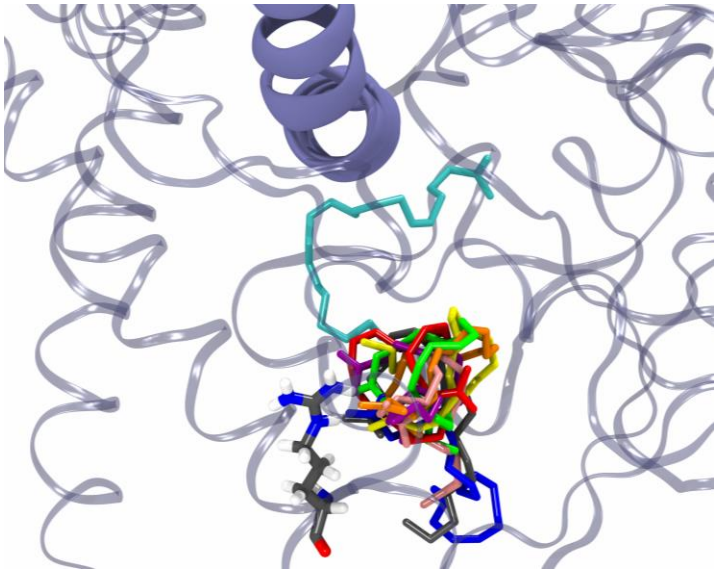

Figure 38. Clustering of the rejected conformational results from the ADT docking procedure. VMD was used to visualize and each licorice representation is presented in different color codes.

The two best conformations were chosen based on their orientation and interactions within the Cyp4f13 homology model active pocket. The best conformation,

shown in Figure 39, shows carboxylate binding interactions with only Arg237 and a near proximity of the terminal carbon unit of the substrate to the heme-thiolate cofactor.

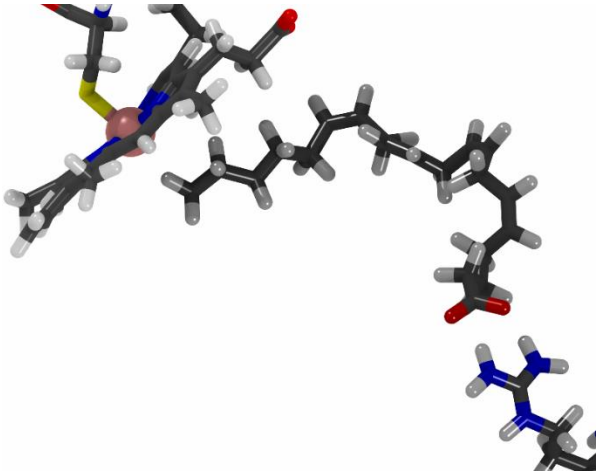

Figure 39. Structural view of the best conformation (D1-Cyp4f13-AA). Interactions between arachidonate carboxylate headgroup and Arg237 are shown in licorice graphical representations.

The second-best conformation resulted in an interesting interaction between the ε-proton on the His236 residue and the carboxylate headgroup of the arachidonic acid. This residue was modified by the PROPKA protonation prediction service, and its true protonation state remains unknown; not able to be resolved by even crystallographic means. What was decided was that in order to investigate whether the protonation state of the His-236 residue was important in the binding of arachidonate, the histidine was changed to be protonated on both sides of the imidazole sidechain (residue type HSE to HSP), resulting in a net positive charge for the residue. This protonation, picture in Figure 40, was performed to investigate the substrate-histidine 236 interactions present in the second-best conformational model that were absent from the first conformation.

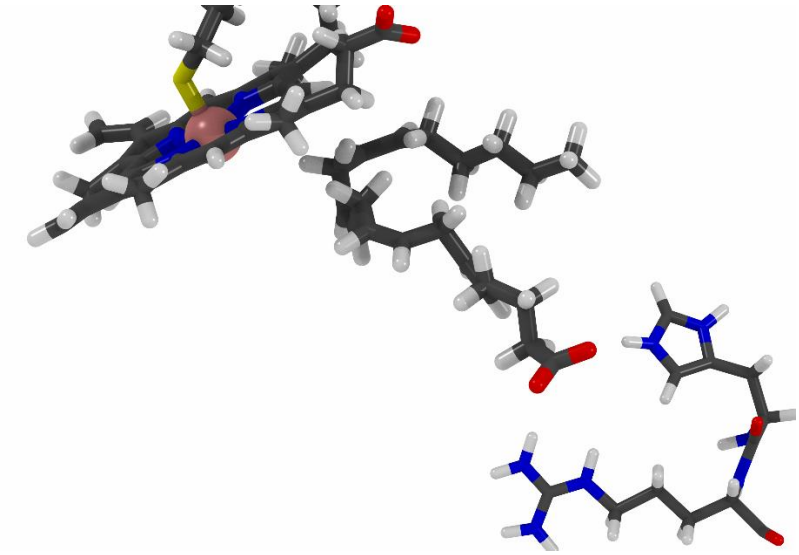

Figure 40. Structural view of the second-best conformation (D2-Cyp4f13-AA) with the modified histidine 236 (HSP) residue. Interactions between arachidonate carboxylate headgroup with Arg-237 and the protonated His-236 are represented with licorice models.

To clarify, this protonation was modified after the molecular docking procedure and was not present in the rigid protein file used to dock the arachidonate ligand. This was done with the intention that after simulations were performed, convergence of binding interactions would be able to discern whether this modification to protonation state had an impact on carboxylate binding.

The docking of 20-HETE to the Cyp4f13 homology model resulted in several conformations, like what was observed with arachidonate. The best conformation was selected by the metric of having the greatest binding free energy and having similar binding mode to arachidonic acid, pictured in Figure 41.

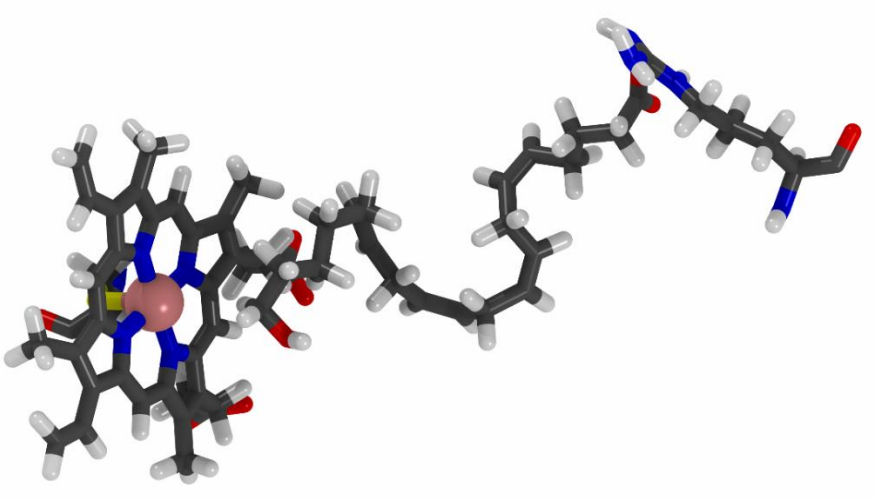

Figure 41. Structural view of the top resultant conformation of docked 20-hydroxyeicosatetraenoate (20-HETE) to the Cyp4f13 homology model. The 20-HETE hydroxyl group is pointed away from the heme in the initial docking structure as the heme was not present during the docking process.

MD Simulations on D1-Cyp4f13-AA complex

In order to interpret the frames of the dynamic simulations of the D1-Cyp4f13-AA complex, an analogous criterion was used as described in the paper by Li et al. on a similar system (CYP4F2) in 2018.[223]

All 100,000 frames of the simulation were collected and analyzed for their significance in the indication of ω-hydroxylation of arachidonic acid by the murine Cyp4f13 homology model. The atomic distances and angles between: the heme iron, a theoretical oxo ligand as is present in the ultimate oxidant state of the P450 catalytic cycle, and a hydrogen atom belonging to the terminal carbon unit of the arachidonate substrate were used as the metrics for the criteria. The P450 oxyferryl species has a bond length of about 1.64 angstroms, slightly shorter than other ferryl species due to its electronic environment.[223] With the furthest theoretical distance of oxidation by the oxidant P450 species being less than 2.8 angstroms, the obtuse triangle formed by the

three atoms (Fe-O-H) should have a maximum distance of 4-4.3 angstroms between the heme iron and hydrogen atom of the ω-carbon, with a smaller distance improving the chances of ω-hydroxylation. This obtuse triangle is demonstrated in Figure 42.

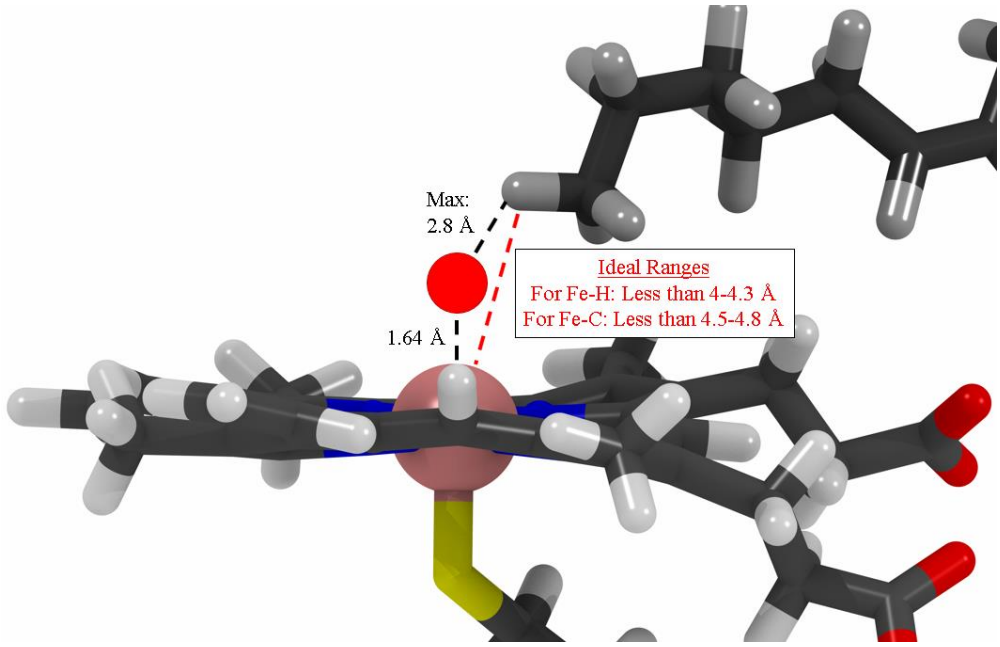

Figure 42. A schematic illustration of the bonding and angles assessed in the criterion for ω-hydroxylation of arachidonic acid by Cyp4f13.

The trapezoidal shape formed by the iron, oxo-ligand, and C-H bond would have a theoretical limit to distance for its larger base of about 4.5-4.8 angstroms apart. The tight (4.5 Å) and loose (4.8 Å) criteria would be used to analyze the large number of simulation frames to assess the likelihood of each site containing appropriate distance and orientation for oxidation. This criterion was tightened based on the number of frames that succeeded to show acceptable distance of the heme iron to each of the oxidation sites belonging to the polyunsaturated fatty acid substrate.

To also investigate the occurrence of sub-terminal or possibly olefin oxidation of the fatty acid substrate, distances between individual carbon units were collected as well. Some indication observed in the results implies the possible formation of an epoxide

metabolite, in addition to several sub-terminal hydroxylation products, as seen in Figure 43 and Figure 44.

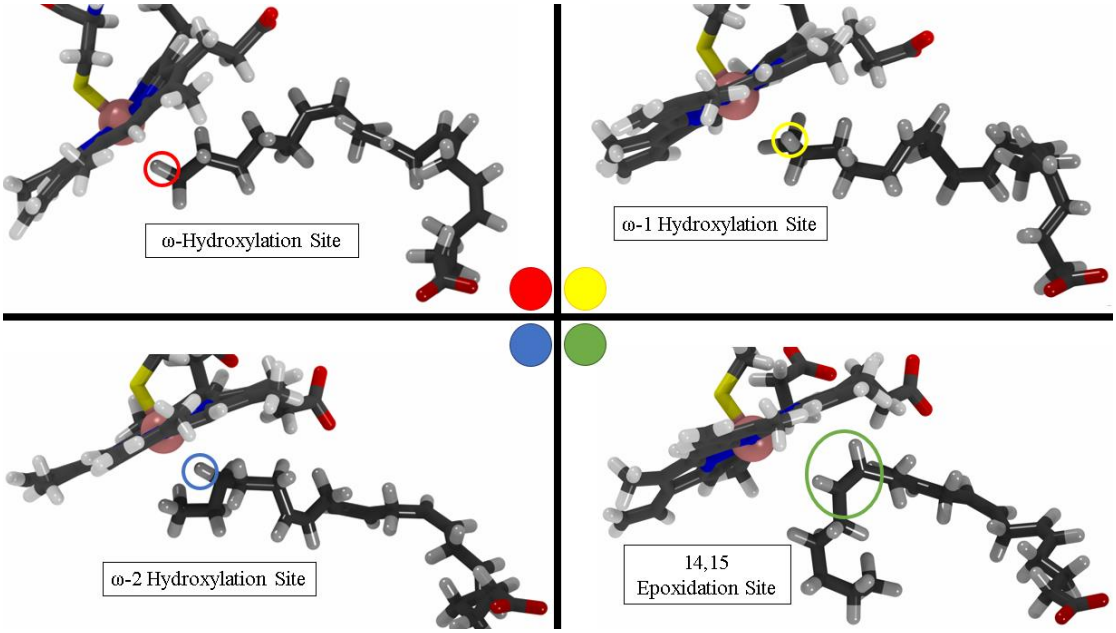

Figure 43. Several orientations of arachidonic acid within the active site of the Cyp4f13 homology model. Hydroxylation products and epoxide products are labeled and colored for distinction.

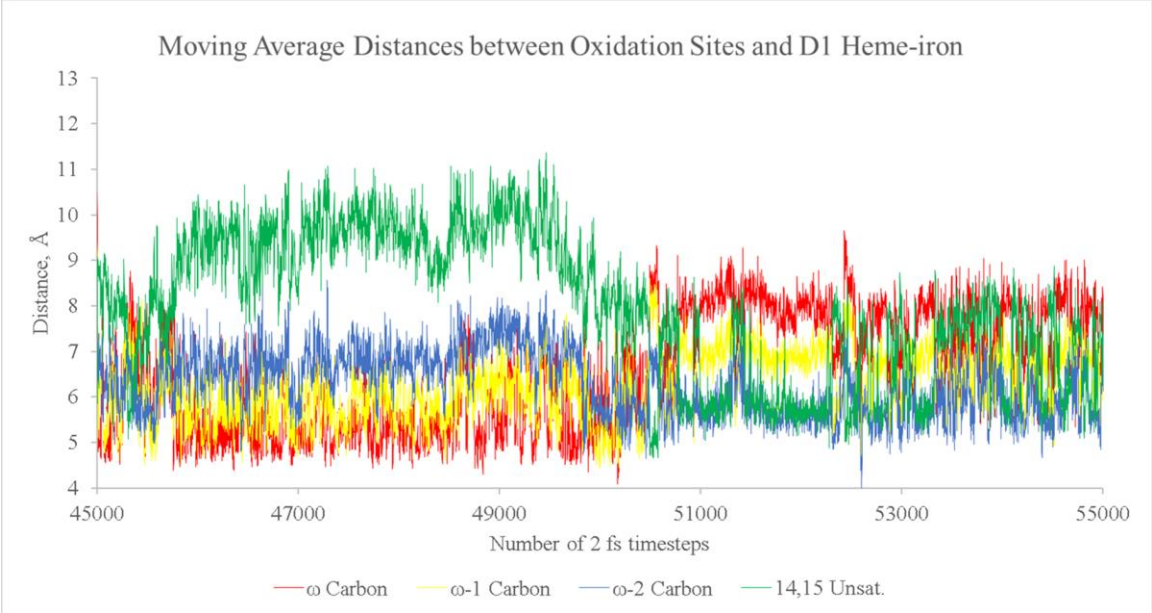

Figure 44. Moving average distances between the Cyp4f13 heme-iron and different oxidation sites for a portion of the D1 production simulation. Trajectory frames are recorded every 1000 timesteps with the total timeframe of 200-ns in length

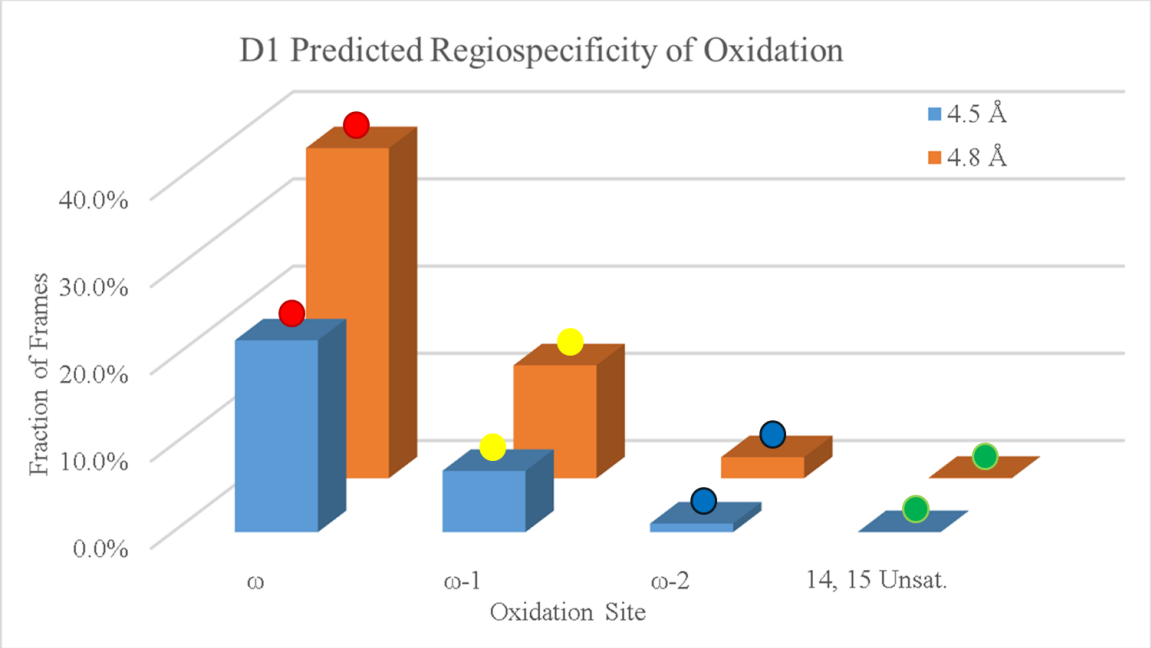

Figure 45. Predicted regiospecificity plot based on the frames that fit the criteria in both a tight (>4.5 A) and a loose definition (>4.8 A). The fraction of fit frames, their oxidation product formed, and the criterion used are shown. Colored dots are used for reference to Figure 43.

The result of the criteria selection, shown in Figure 45, indicate that the ω-position is the most favorable oxidation site exposed to the heme-iron. Most of the frames that satisfied the distance criteria belonged to the terminal carbon site, with substantially lower amounts of successful frames from the subterminal sites (ω-1, ω-2) and 14,15 epoxide formation site. However, some chemical and thermodynamic aspects are not well represented by these results. While the steric accessibility remains to be the factor assessed in this study, the energetics are important to discuss.

Sub-terminal hydroxylation reactions on the aliphatic fatty acid tail would be more thermodynamically favored based on being secondary carbon units. The secondary C-H bond would have a lower relative bond strength than the primary C-H bonds in the

terminal unit, leading to a lower activation energy needed to abstract the hydrogen atom in the radical rebound mechanism of the P450-catalyzed oxidation.

The P450-catalyzed olefin epoxidation is understood to occur readily for several P450 isoform substrates, as it undergoes a similar yet distinct process than the radical rebound process necessary for hydroxylation. The process generally relies on the same concept of attack by the electrophilic oxyferryl or hydroperoxo-iron species belonging to a P450. This reaction would not need to abide by the same distance restrictions, as the pi-orbitals that would be involved in the reaction reach beyond the atomic nuclei by a considerable extent. The orbital interactions necessary for the reorganization would happen at a quicker rate than the hydrogen atom abstraction used in hydroxylation.

The vinylic hydrogen positions on the substrate were not considered as targets for hydroxylation, as they have a much higher bond strength that that of aliphatic carbon units, and thusly, are outcompeted by epoxide production formation.

MD Simulations on D2-Cyp4f13-AA complex

The same criteria used in the previous section was used to gain insight in potential site of oxidation for the arachidonic acid with the noteworthy differences between docking 1 and docking 2 of arachidonate conformations was the protonation state of the binding residues interacting with the carboxylate head group of the substrate. Histidine 236 of the murine Cyp4f13 homology model was protonated through a by-hand modification, with Automatic PSF builder used to generate atomic coordinates for the proton. Charges were reassigned in this step and resulted in a positively charged,

protonated histidine that is observed in forming interactions with the arachidonate in D2-Cyp4f13-AA while not being observed in D1.

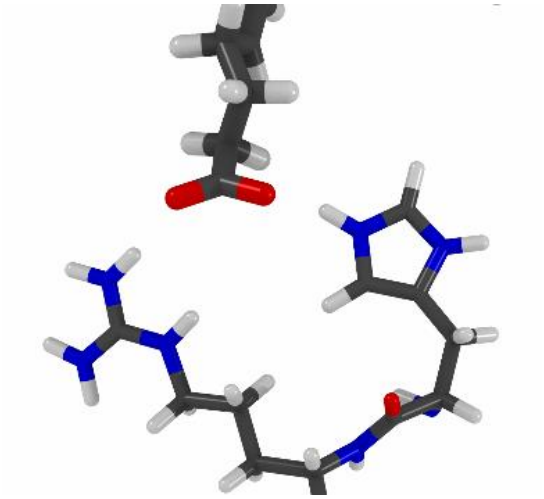

Figure 46. Close-up view of the arachidonate residue interacting with the protonated His-236 (HSP) of D2-Cyp4f13-AA model. The residue type of target histidine was changed from HSE to HSP in the structural files.

The desired effect of the additional simulation series on the different conformational state of the substrate-bound Cyp4f13 homology model was to have the interactions in the binding site converge to a single set of interactions as opposed to the several orientations predicted. The successive MD simulations gave a very promising result as the interactions between the carboxylate of the arachidonate appeared to converge on hydrogen bonding with the Arg237 and Tyr96 residues, as was observed in the D1-Cyp4f13-AA runs. The addition of the protonated histidine residue did, however, appear to alter some of the binding activity for the arachidonate substrate, as the different binding site interactions introduced an observable effect on the active site.

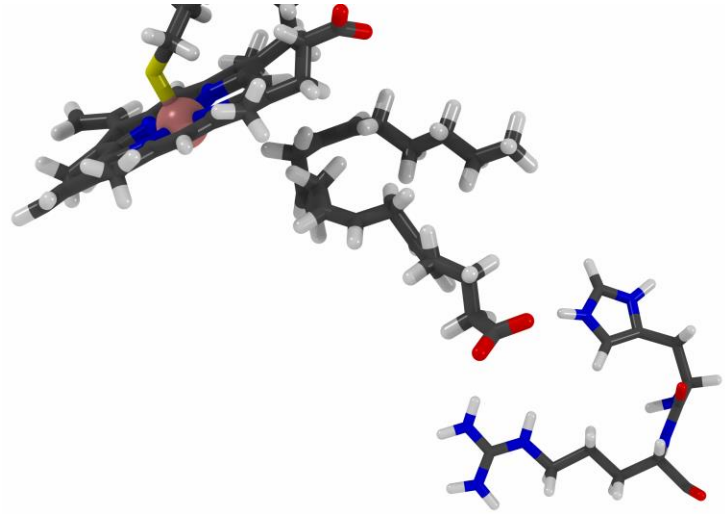

Figure 47. Different angle of substrate binding interactions of D2-Cyp4f13-AA homology model partway through the production simulation. The complete displacement of the aliphatic tail region of the arachidonate, indicating the formation of mid-chain oxidation products.

The simulation of D2-Cyp4f13-AA indicated a complete change in substrate conformation within the catalytic site. As seen in Figure 47, the aliphatic tail of the polyunsaturated substrate repositioned within the active cavity, exposing its internal unsaturation sites to the heme-iron, as opposed to terminal or sub-terminal sites as seen in the previous simulation. This unexpected conformational change resulted in different oxidation sites being presented to the heme-iron, possibly indicating the conversion of substrate to internally oxidized product.

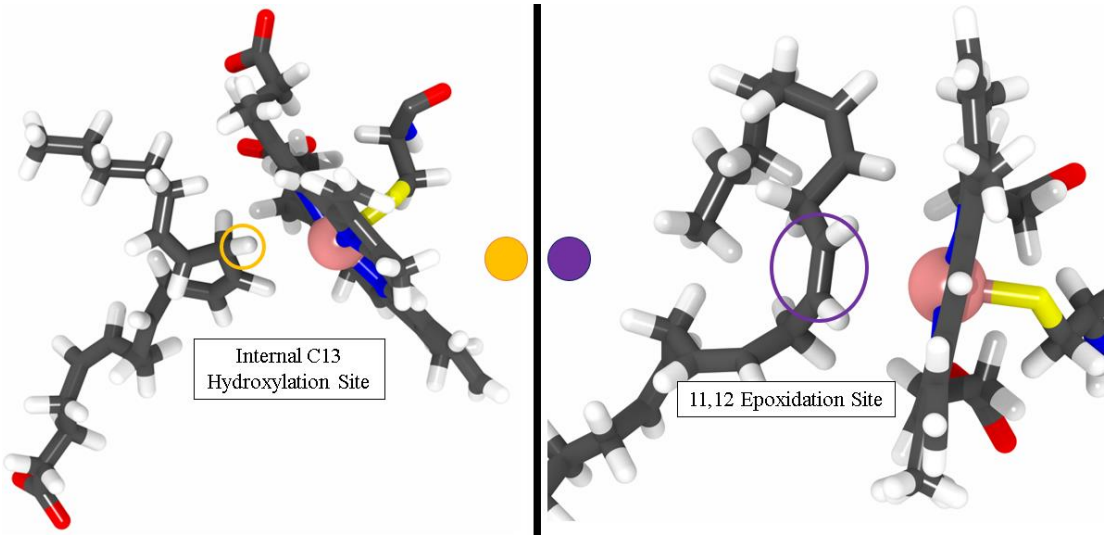

Figure 48. Two additional oxidation sites along the arachidonic acid substrate in the simulation of the D2-Cyp4f13-AA system. The bisallylic carbon (C13) and the penultimate unsaturation (11, 12) site are labeled for their product formation.

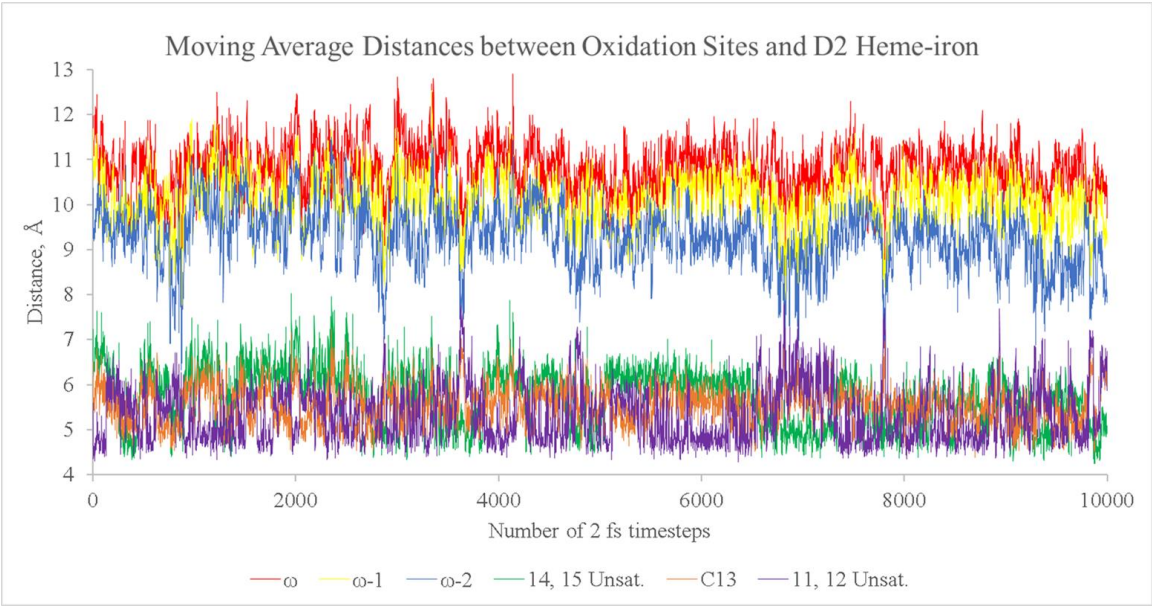

Figure 49. Line plot of the distances of the simulation trials of D2-Cyp4f13-AA heme iron and different carbon units. Each set of 100-ns trajectory frames are stacked and labeled with their implied metabolite formation.

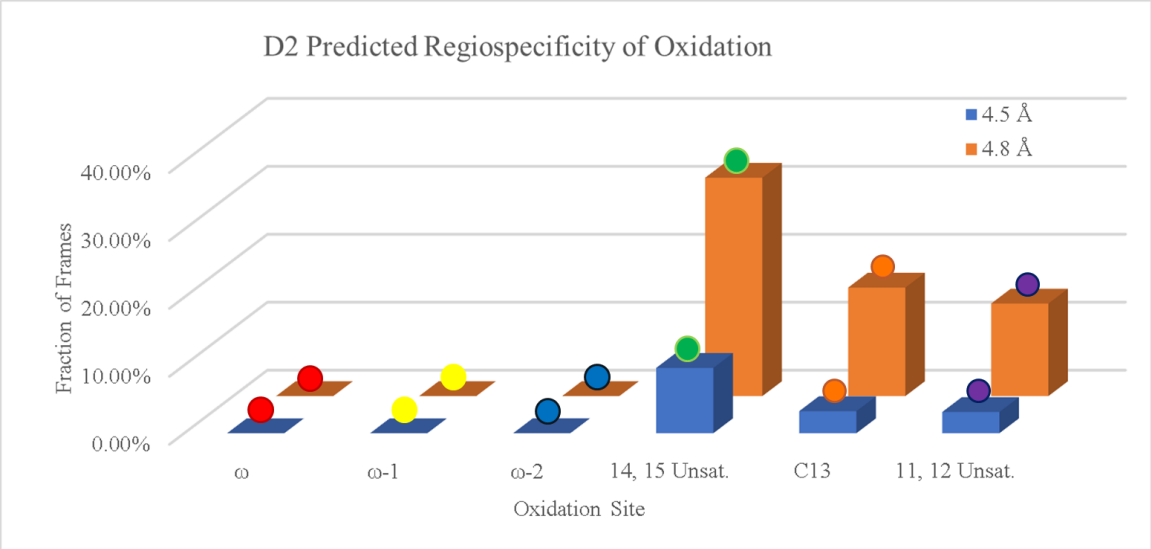

Figure 50. Predicted regiospecificity plot of the simulation frames that met criteria for the D2-Cyp4f13-AA system. A collection of all previously mentioned possible oxidation sites are shown and colored accordingly.

The high frequency of internal epoxy- or hydroxylation sites that met the same set of distance criterion as before indicating that there were several orientations of the substrate present and within appropriate distance for the respective oxidation reaction to occur. The increased frequency of frames for different oxidation sites that fulfilled the criteria suggests a fair amount of flexibility in the active site of the murine Cyp4f13 system, as the open active site allows for a variety of conformational states for its substrate.

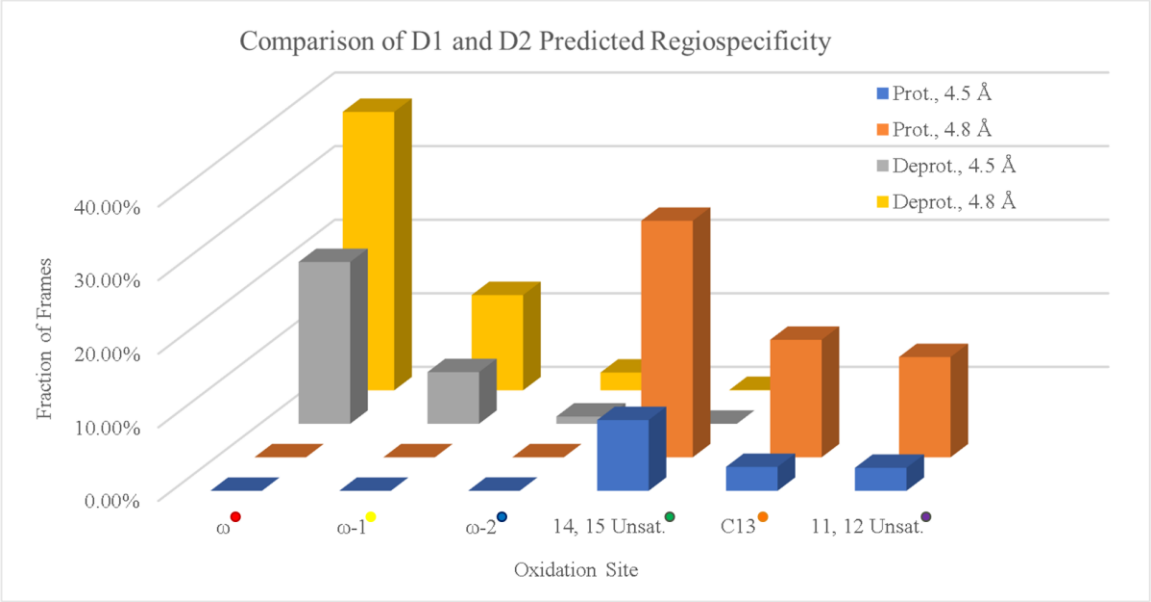

Figure 51. Comparison of the predicted regiospecificity of the protein-substrate simulations. The neutral His236 Cyp4f13 is represented by D1-Cyp4f13-AA, while the protonated His236 Cyp4f13 is represented by D2-Cyp4f13-AA.

When comparing all the relative frequencies of the oxidation sites across the two simulation sets, as portrayed in Figure 51, it becomes apparent that the ω-hydroxylated site is not site-specific for the Cyp4f13 system. The curling of the arachidonate tail presented different sites for possible oxidation to the heme center. What is to be understood is that the metabolite profile reported for Cyp4f enzymes is highly selective but not specific. Meaning that it is mostly favored for the ω-hydroxylation but not entirely dominated by, as these simulations are in agreement with this fact.[224]

One interesting observation was that of the frequency of the appropriate distance to carbon 13 of the arachidonate chain to the heme-iron. The position of this carbon, between two isolated unsaturation points within the polyunsaturated compound, exists as a special target for hydrogen abstraction, P450-catalyzed or not. After hydrogen abstraction, this position results in the formation of a 1,5-pentadienyl radical species that

can rearrange to several isomers that can readily react as sites of hydroxylation. The relative stability of this radical indicates that the formation of the species would make it very thermodynamically favorable for the enzyme system, as the hydrogen abstraction and radical rebound process is a notable feature of the P450 catalytic cycle.

Role of Binding Site Residues and their Protonation States

Observed in the first 100-ns trial for the D2-Cyp4f13-AA model, solvent water molecules maneuvered their way into the active-site pocket, coordinating to the heme center, displacing the fatty acid substrate enough to curl the tail towards a hydrophobic fold in the interior of the Cyp4f13 catalytic cavity.

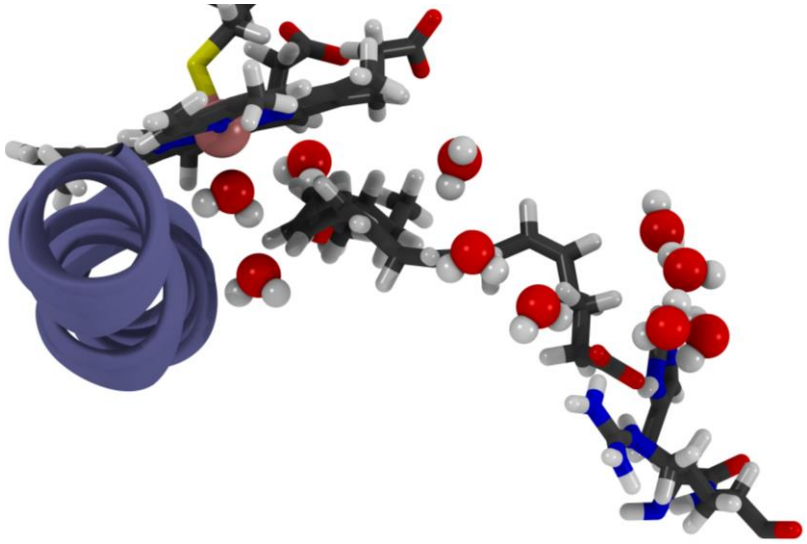

Figure 52. Structural view of the substrate-bound Cyp4f13 homology model with water taking the place on the empty coordination site for the heme-thiolate. Increased solvation around His236 is observed in the late stages of the simulation.

One theory as to why solvent water interactions, viewed in Figure 52, overtook the hydrophobic interactions presented by the catalytic site was the introduction of the charged histidine residue within the binding site of the Cyp4f13 model. Several factors could originate from this inclusion and result in a different amount of solvation being

observed. With the positive charge producing a more polar environment, solvent water interactions could have entered the binding site, pictured in Figure 53, which would normally have had minor amounts of positively charged residues, possibly undermining the binding interactions between the protein and substrate.

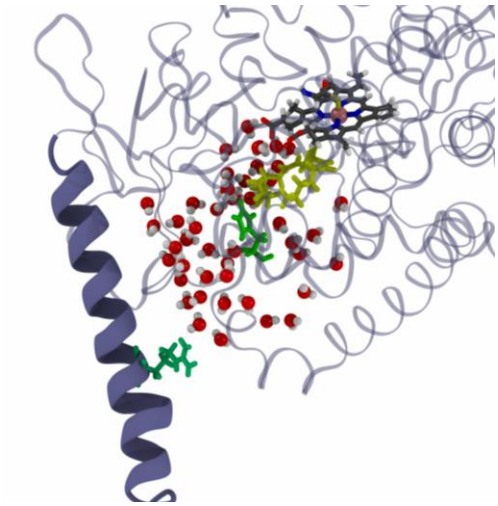

Figure 53. Bolus of water surrounding the transmembrane hinge near the protonated His236 of the D2-Cyp4f13 homology model. Arg-33 and the protonated His-236 are shown in green licorice representations, with yellow used on the arachidonic acid.

However, this theory appears to have flawed logic as through investigating the same location in the previous simulation, D1-Cyp4f13-AA, showed a large amount of water activity in that very same vicinity. The similar amounts of protein solvation would not have produced the differing effect on the solvation of the substrate observed in the simulation.

Another theory is that of unanticipated tertiary structural impacts. The presence of a positive residue in a normally neutral position might have a more impactful influence on the secondary structure of the neighboring protein composition than what was originally intended. The positive Arg33 of the transmembrane anchor sits in the middle of the N-terminal alpha helix and the globular heme domain. It appears to interact

repulsively with the protonated histidine, causing a disturbance in protein conformation around the hinge of the globular domain. The formation of a gap in the protein structure could explain for the increased solvation of the carboxylate binding site, leading to further disturbances for binding interactions between the protein and its substrate.

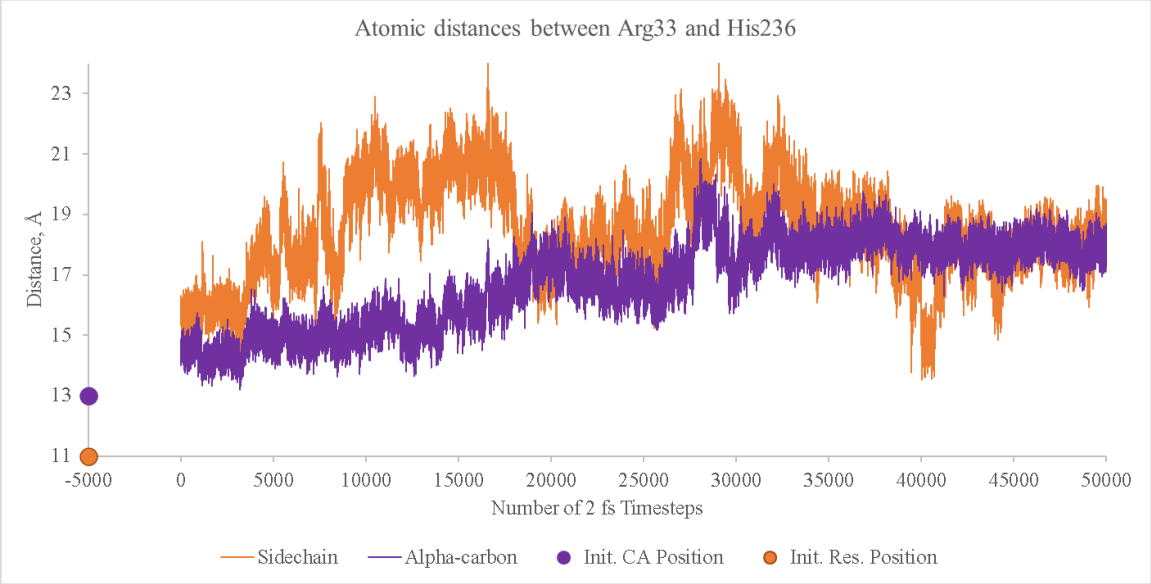

Figure 54. Distance plot of the guanidinium group of Arg33 and the protonated His236 over the 300-ns simulation timeframe. Shown are the distances between the α carbon and the sidechain of the residues in angstroms over simulation timesteps, with pre-minimization, heating, and equilibration positions indicated by markers.

This interesting artifact led to the possibility of a more pertinent role for the His236. This varying gap between the membrane anchor and His236 is right where the fatty acid substrate is expected to enter the substrate access channel to approach the heme iron. With the only actual difference between D1 and D2 homology models being the protonation state of His236, more pertinent were the differences observed in structural conformation for the two P450 complexes. A large amount of structural alteration was observed due to residue interactions between the protonated His236 and nearby charged residues.

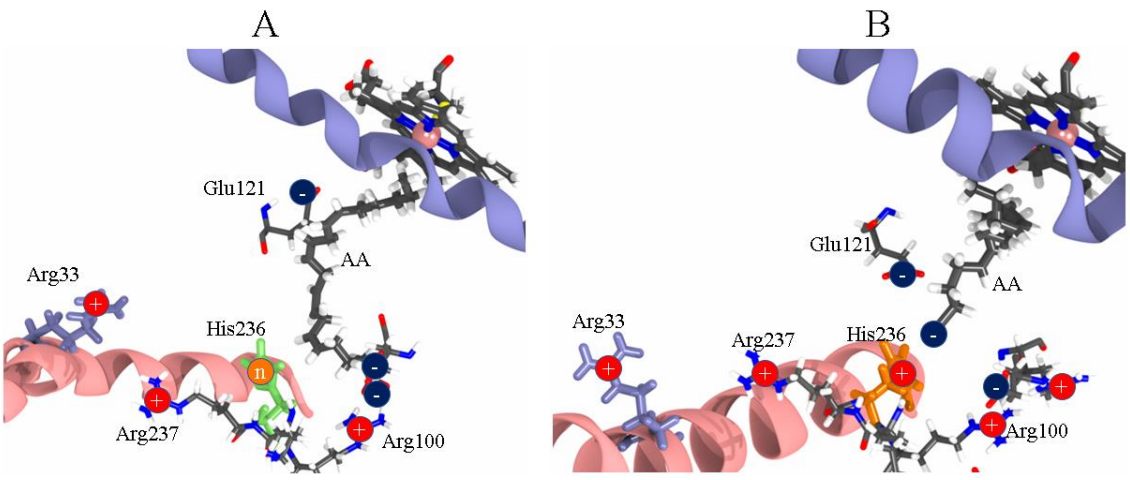

Figure 55. Comparison of frames from the production simulation of D1-Cyp4f13 (A) and D2-Cyp4f13 (B). The transmembrane alpha helix and I-helix are represented as pink and ice blue cartoons respectively. The neutral histidine (HSE) of the unmodified Cyp4f13 is represented in lime and the protonated (HSP) histidine is colored in orange. All charged residues within 8 angstroms of His236 in both simulations are shown in licorice representations.

As seen in Figure 55a, the Cyp4f13 with the neutral histidine remains largely inactive with the carboxylate binding of the fatty acid substrate. It remains turned towards solvent, occasionally rotating to influence the binding through purely steric interactions. The Arg33 of the transmembrane anchor appears to point towards the inner folds of the protein structure, interacting with residues and solvent in the D1-Cyp4f13 complex. For the D2 complex, some differences are immediately observed when comparing the positions of the same residue side chains. Protonated His236 now is shown interacting with the carboxylate headgroup of the substrate, which is to be expected by electrostatic considerations. The acidic residue nearby (Glu121) has changed position, possibly from the newfound presence of a positively charged amino acid in the vicinity. Arg33 has turned towards the lipid bilayer and could possibly have changed its orientation due to repulsive electrostatic interactions with the positively charged His236 residue as well.

What can be gained from these observations is the possible physiological significance they might imply. It is possible that the Cyp4f13 enzyme is utilizing pH-mediated gating activity of the protein to aid or modify substrate binding. Under more acidic conditions, the protein may have had to evolve to allow or disallow substrate entering the active cavity through a gating mechanism. Such a finding might also indicate the presence of possible interaction changes with the lipid bilayer under different pH conditions as well. The charge difference near the lipid-submerged F-G loop of the protein might interact differently to changes in lipid composition. The di-oleoyl phosphatidylcholine bilayer used in this simulation is a net neutral lipid; perhaps with a different bilayer composition, a net-negative lipid such as phosphatidylglycerol, a different orientation for the membrane-embedded P450 would be observed.

MD Simulations on D3-Cyp4f13-20-HETE complex

Like with the D2-Cyp4f13-AA model, three successive 100-ns MD simulation runs were performed on the docked oxidized metabolite bound to the Cyp4f13 homology model (D3-Cyp4f13). The model indicated the formation of hydrogen bonding interactions between the carboxylate of the 20-HETE compound and the very same Arg100 and Arg237 residues as the other docking models displayed, pictured in Figure 56.

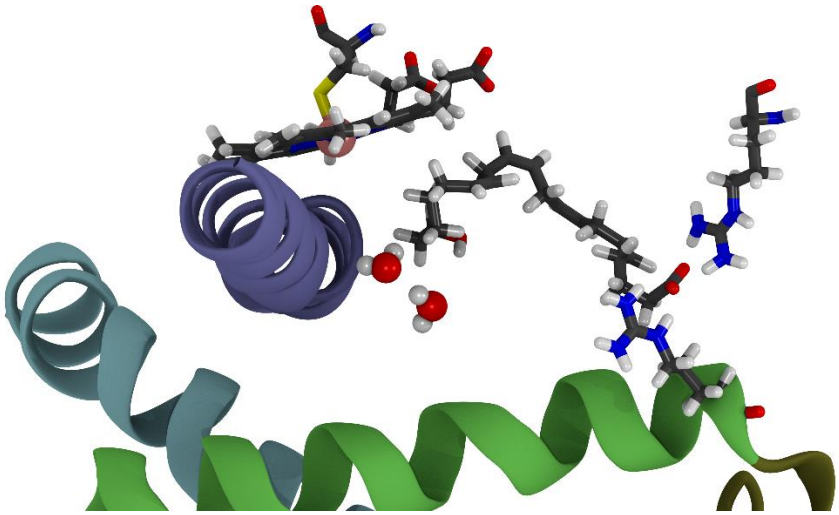

Figure 56. Structural view of the D3-Cyp4f13-20-HETE model active site before the production simulation. Licorice representations for Arg100, Arg237, the heme cofactor, and the 20-HETE substrate.

Since the ω-hydroxylated product was added for dynamic simulations, some of the expectations made about the simulations were that solvent water molecules would enter the active pocket, as they had done so in the previous substrate-bound simulations. The favorable water interactions with the hydroxyl group attached to the polyunsaturated fatty acid would result in a disruption and solvation of hydrophobic residues residing in the active pocket, such as Leu137. These disturbances would compound over the theoretical time steps and could possibly result in the movement of the oxidized product in a manner that would be indicative of eventual egression or exiting from the catalytic site. However, the computable timeframe necessary for this to be observed is expected to span fractions of milliseconds.

What could be done to facilitate substrate egression would be through the use of Steered Molecular Dynamics (SMD) simulations. SMD or force probe simulations consist of calculations where outside forces or velocities are applied to molecules in a studied system in order to manipulate and observe changes in protein structure and

dynamics. By tugging on the substrate along a desirable vector, SMD could be used to simulate the mechanical egression event.

Instead of measuring the relative distances between the heme-iron of the cofactor and carbon units, the distance between the hydroxyl group oxygen atom and the heme-iron were used as a metric in order to observe product displacement. However, once again the simulations provided for another interesting result with its own fascinating share of considerations.

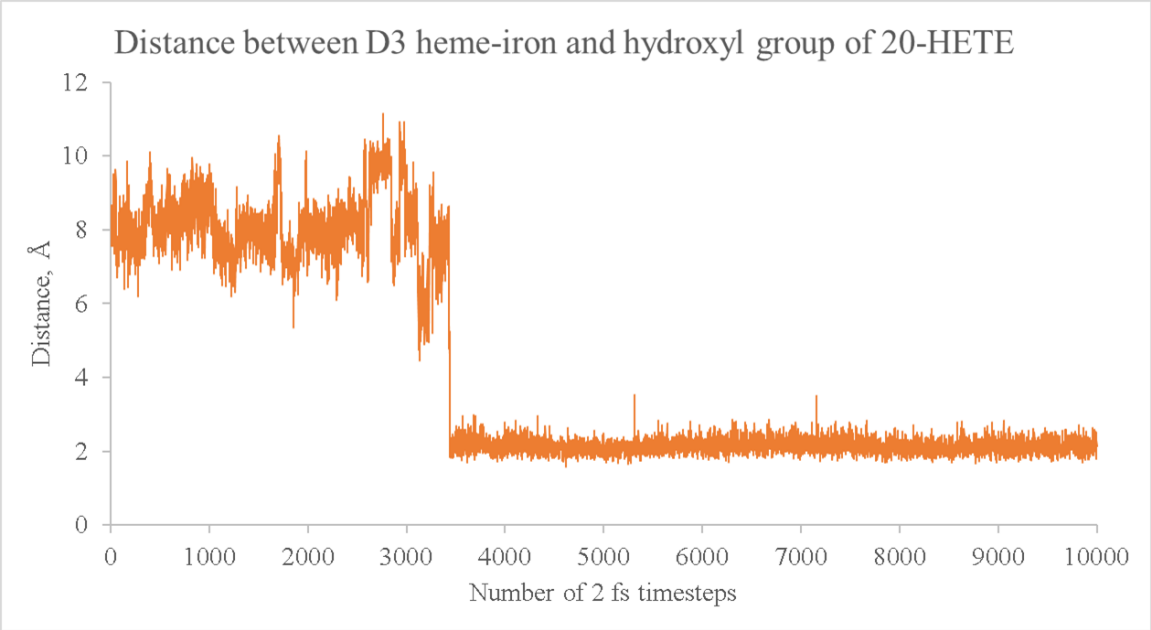

Figure 57. Line plot of the distances of the first 100-ns of the simulation of D3-Cyp4f13-20-HETE heme iron and the hydroxyl oxygen of the 20-HETE substrate. A ligand association event is observed at around one-third of the production simulation timeframe.

Figure 57 shows a distinct drop in distance observed one-third of the way of the first 100-ns simulation run between the primary alcohol of the ω-hydroxylated product and the iron center. The open coordination site of the heme-iron underwent ligand association with the hydroxyl group of the 20-HETE metabolite. In an effort to increase the sample size with additional simulation time, the model was set to run for two more

additional 100-ns simulation with the intention to allow water to integrate more within the active pocket, possibly still achieving some form of displacement for the oxidized product. The ligand coordination between the heme-iron and the hydroxyl group was maintained throughout the remainder of the simulation. The ligand association of the 20-hydroxyeicosatetraenoate will be thoroughly explored and discussed in the next section.

Due to the polar hydroxyl group disrupting the hydrophobic forces that are usually necessary to facilitate P450-mediated metabolism of polyunsaturated fatty acids, it seems unlikely that the Cyp4f13 enzyme is responsible for the binding of these oxidized metabolites. Furthermore, the oxidized metabolites would need to dissociate from the protein in order to regenerate the resting state of the heme cofactor, as is observed in successful enzymatic turnover.

However, these interesting observations from simulation should be explored experimentally, as many P450s catalyze the formation of dicarboxylic acid metabolites of fatty acids by sequential oxidation of fatty acids.[225] An occurrence early in the simulation presented the substrate in an orientation that might result in the formation of dihydroxyl, hydroperoxyl, and even dicarboxylic acid metabolites. These further oxidized derivatives of arachidonic acid are amongst the eicosanoid metabolites of physiological and pharmacological interest; however, these results are weakly indicated by the Cyp4f13 homology model simulation runs.

Coordination of Hydroxyl Group to Heme-thiolate Metal Center

The ligand association of a hydroxyl group over a solvent water molecule is an interesting observation. The factors that influence the association and dissociation of

ligands are complex and owe to the study of coordination chemistry, as everything from orbital overlap, neighboring ligand influence, steric limitations, and bonding angle can alter and drive the preference of a metal center for whether or not a ligand can coordinate. In the situation of the substrate-bound murine Cyp4f13 homology model, the only possibilities for coordination to the metal center would be that of a solvent water molecule and the alcohol group on its hydroxylated substrate, as seen in Figure 58.

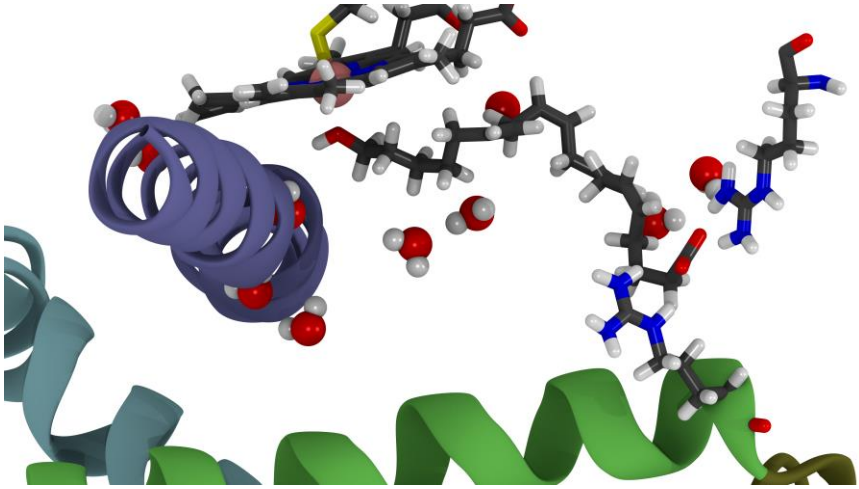

Figure 58. Structural view of the D3-Cyp4f13-20-HETE model active site displayed after 300-ns of simulation. The coordination of the hydroxyl group of the 20-HETE substrate is preserved up to the end of the simulation.

The first coordination chemistry standpoint to consider is the orbital overlap aspect. The difference in electronic environment between the oxygen of an aliphatic hydroxyl group and that of a water molecule is relatively small, yet present. Both exhibit appropriate hybridization to bear four electron pairs, either in the form of chemical bonds or through lone pairs. In the theory of orbital hybridization, this requires a sp^3 configuration of the oxygen atom's atomic orbitals, resulting in the canonical Lewis dot structure representation. However, this representation does a poor job in describing the chemical environment of the two possible ligands. Molecular orbital theory is commonly

used to explain coordination chemistry phenomena, often employed as Ligand Field Theory, as it can describe orbital arrangement, symmetry, and other characteristics in the coordination of ligands to metal centers.

A quick look into the frontier molecular orbitals involved in metal dative bonding for a water molecule and a primary alcohol is provided in Figure 59 and would result in very similar expectations.

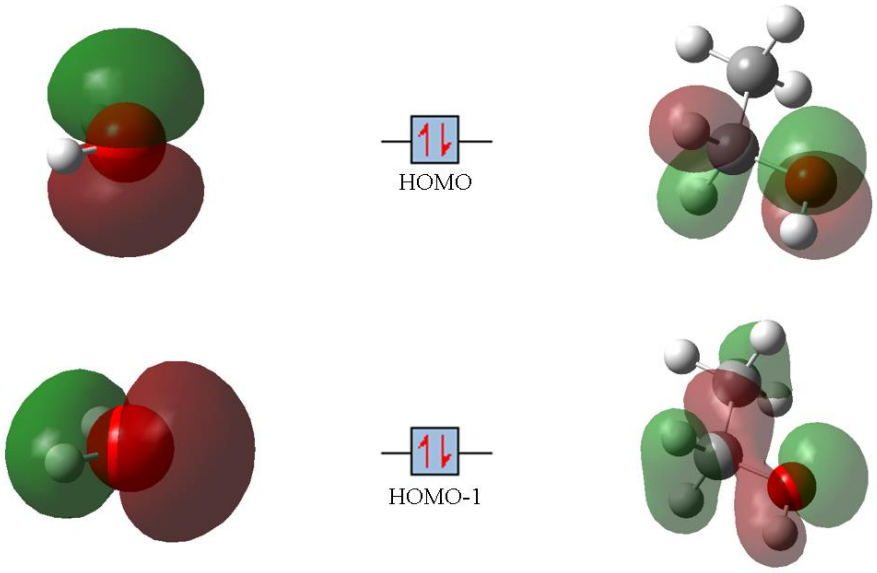

Figure 59. Simplified frontier molecular orbitals for water and ethanol as an example of a primary alcohol. Shown are the Highest Occupied Molecular Orbitals (HOMO and HOMO-1).

The coordination occurs through primarily a weak sigma donation of electron density from the A_1-symmetric molecular orbital on water that is below its highest occupied molecular orbital (HOMO), the B_2-symmetric non-bonding orbital, originating from a p orbital of the same symmetry on oxygen. This p orbital is of the right symmetry to donate to the metal center and its strength is based on a number of factors such as orbital overlap and electronegativity of the donor atom. The oxygen atom contains conflicting attributes in this regard as it has 2p valence orbitals that exhibit great overlap

with the 3d valence orbitals found on first row transition metals, like the heme-iron in the case of the Cyp4f13 cofactor. However, oxygen is also one of the most electronegative atoms with small amount of nuclear shielding due to its low number of inner electron shells in comparison to its group members. This results in a high amount of effective nuclear charge, resulting in an oxygen atom's tenacity to oxidize other atoms it comes into contact with.

These factors result in the donation ability of oxygen to a metal center to be relatively weak, however, these interactions are still present and must be accounted for. The symmetrical aspects lend to a weak pi donation, raising the appropriately orientated pi orbitals of the metal center resulting in a smaller splitting between its metal d orbitals in an octahedral configuration.

The coordination for a primary alcohol, such as ethanol, occurs in a similar fashion. An orbital of primarily oxygen character, below the HOMO, is used primarily to sigma donate to the Lewis acidic metal center, with its remaining lone pair able to weakly pi donate to the pi symmetry orbitals on the metal center. However, the interactions between pi donors and metal centers are generally weaker interaction than the dative sigma donating ability of uncharged, monodentate, oxygen donor ligands.

The next considerations should be made for the electronic environment of the oxygen in between the two species. In comparison to the hydrogen atoms in water, primary alcohols have a single organic substituent. The presence of a carbon atom, with a higher electronegativity (2.55 compared to 2.2 for hydrogen in the Pauling scale), means that there is a more covalent interaction in C-O than that of the polar bond between H-O. This has an impact on the oxygen's tendency to be a donor, attenuating its ability to act as

a ligand for a metal center. This difference, however, is very minute for the case of monodentate primary alcohols, as the organic moiety only changes the electronic environment by a small degree. This can be observed by comparing acidity dissociation constants (pK_a) of the two species (15.74 for water and 16.0 for ethanol). This measure indicates relatively little change between the Lewis basicity or donation capabilities of the oxygen donor atom as both a sigma and pi donor.

Considering from a viewpoint of an octahedral ligand field, this ultimately results in both water and a primary alcohol acting as moderately weak ligands, able to be replaced by stronger or more abundant ligands such as additional water molecules.

With the orbital comparisons made, what must be understood and reiterated is the system being analyzed, an MD simulation result. MD simulations do not contain any information on the orbital interactions of a molecular system, other than that of purely bonded or non-bonded physical characteristics. Orbital overlap and ligand field theory cannot be used to explain the results from an MD simulation. Electrostatics and van der Waals interactions, taken from force field parameters, are the factors that are modelled within a biological MD simulation and should be the considered method of analysis used in order to investigate theoretical phenomena.

The next aspect to consider is the possibility of an error in the parameterization of the 20-HETE ligand. As was previously stated, the modified parameter was hard-keyed into the CHARMM 36 lipid file and was made through analogy using ethanol, found in the CHARMM format. There could have been a mistake made in the partial charge assignment to the hydroxyl group of the oxidized substrate. However this does not appear to be the case, as the assigned partial charges for both oxygen types (OHL in 20-HETE

and OT for TIP3P water) indicate a lower partial charge assigned to the oxygen (-0.65 in 20-HETE and -0.834 in TIP3P water). This was derived from the CHARMM topology values from ethanol and placed a lower amount of charge on the oxygen atom.

With the possibility of a mis-parameterization ruled unlikely, the last aspect to consider is the sheer probabilistic character of the interaction. Water, having a relatively miniscule molecular size and weight, is in bulk amount in any biological ensemble. Water and small molecule substrates having a similar affinity to form electrostatic interactions with an open metal center site would easily result in the much higher concentration solvent water molecule occupying sites they compete for. Water should be found coordinated in a far more simulation states than the few predicted with a single organic alcohol hydroxyl group.

What is not accounted for in this interpretation, however, is the protein architecture of the active cavity. This aspect is what appears to be the key as to why the hydroxylated tail of the 20-HETE was found to coordinate and persisted through the simulation. An effect of the 20-HETE binding is related to the ‘chelate effect’ observed in metal coordination spheres, where through the coordination of a single ligand at multiple sites results in a more energetically stable environment. This is performed through both a kinetic and thermodynamic manner as the formation of such a complex is both entropically favored and driven by the effective concentration of a ligand. This same effect is observed, however, executed in reverse in the simulation. The binding of 20-HETE at multiple locations throughout the protein, the arginine residues of the binding site and heme iron center, resulted in a protein-substrate complex portraying very favorable energetics.

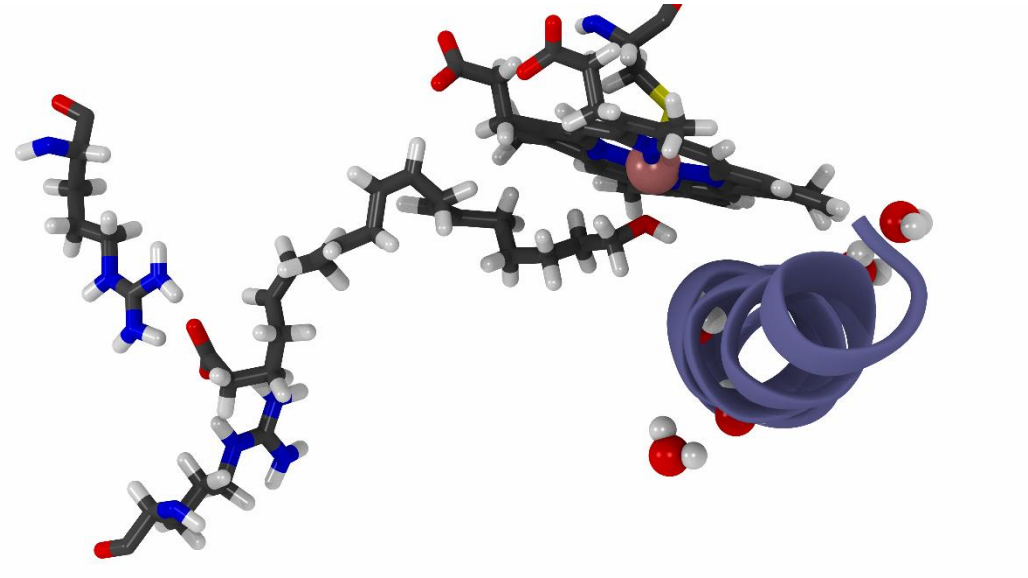

Figure 60. Another look at the structural view of the last frames of the simulation of D3-Cyp4f13-20-HETE. Water molecule coordinates are drawn within a radius of 5 angstroms around the I-helix. Several points of interaction are observed for the 20-HETE substrate and the protein active site, indicating the formation of a 'chelate' to the protein structure.

Due to the desolvation of the catalytic site, in combination with appropriate substrate binding interactions, there appears to be no access to the heme-iron open coordination site from solvent water molecules. The hydrophobic interactions acted upon by the active site nonpolar residues have succeeded in shielding the bound substrate from the TIP3P water molecules. Aqueducts are still present throughout the protein, as seen in Figure 60, however, direct access to the coordination site is restricted in the substrate bound state of fatty acid oxidizing P450.

This resulted in a sufficiently locked coordination of the hydroxylated fatty acid, as the active site structure made it near impossible to dissociate and be replaced by water, at least in the simulation time allotted. However, it is understood that the spontaneous egression of an oxidized fatty acid metabolite is an event on a timescale of several

hundred nanoseconds, which implies that the simulation time presented in this work was not adequate enough to allow this to occur.[226]

Active Site Residues of Interest for the Cyp4f13 Homology Model

Investigations of the active site also concluded with examination of some of the amino acid residues predicted for having a role in substrate binding in the substrate-free simulations. The substrate-bound murine Cyp4f13 active site contained many important residue side chains, several of which will be used as probes for role in the next portion of the experiment. The three residues of significance are the Leu137, Arg237, and Gly327 residues.

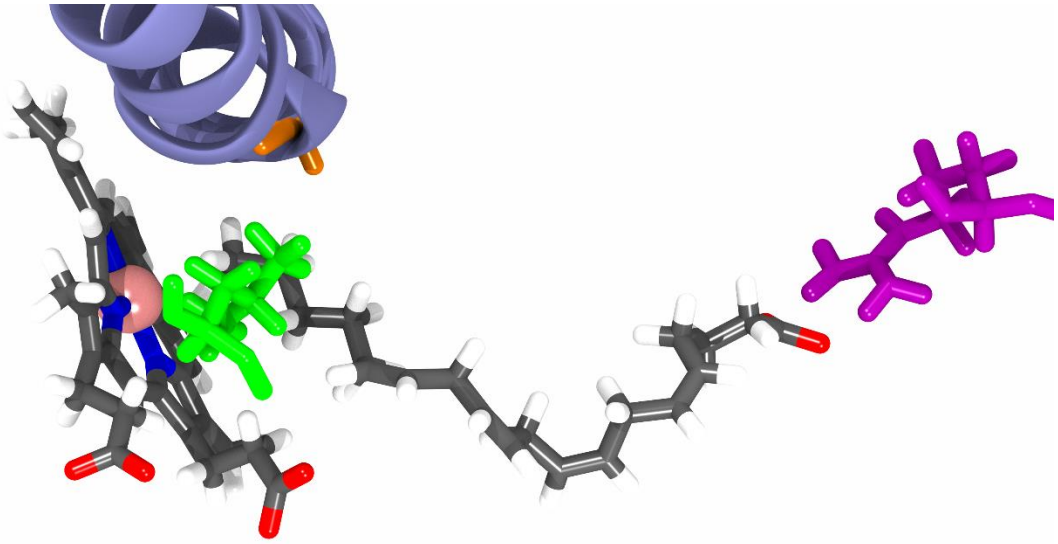

Figure 61. Licorice representation of selected amino acid residues within the Cyp4f13 homology model. Gly327 is represented in orange coloring, Arg237 is in a purple, and Leu137 in green.

Glycine 327, as seen in Figure 61, performs no observable role in the simulations other than acting as a flexible point for the α helical I-helix secondary structure of the P450. While this residue is seemingly unimportant for catalysis at a first glance, it is in

the same position as the conserved glutamate residue found in CYP4 enzymes with covalent heme linkage through an ester bond. This would be the first residue of interest to mutate and modify to observe the effects of heme linkage on the murine Cyp4f13 system. Mutation into glutamate, as opposed to the shorter-chain aspartate, in combination with covalent linkage with the heme, could result in a significantly modified steric environment for the Cyp4f13 heme and could shift the preference towards ω-hydroxylation, as is observed in other CYP4 enzymes.

Arginine 237, shown coordinating to arachidonate in Figure 61, was directly involved in carboxylate binding in all docked substrate models, always showing a role in the interactions within the substrate binding site and possibly with the retrieval of membrane-localized fatty acids. Mutation of this long-chain polar residue to similar length, non-polar amino acids is common throughout the study of P450 mutants. As seen in the work by Richards et al., arginine-leucine mutations are commonly used, as many $P450_{BM-3}$ mutants employ the modification in order to change substrate specificity.[227]

A large number of positive residues sit at the polar headgroup level of the membrane around the submerged loops of Cyp4f13. Some quantum chemical methods have suggested that the guanidium group belonging to arginine residues has superior carboxylate binding ability over other positively charged side chains, such as lysine or histidine.[228] One possible mechanism for substrate entering the active pocket is through a ladder-ratcheting technique as the carboxylate head group of the fatty acid substrate climbs its way from surface to interior arginine sidechains present along the P450s structure. More information on the uptake of substrate could be gained through mutation of the carboxylate-binding arginine to disrupt the hydrogen bonding interactions that hold

the fatty acid substrate in place. This could prove interesting in the investigation of substrate binding specificity and regiospecificity of arachidonic acid oxidation.

Leucine 137, green in the Cyp4f13 homology model pictured in Figure 61, acts as a non-polar residue confining the active pocket volume into the shape necessary to expose the aliphatic tail of a substrate to the oxidant oxyferryl species. Also observed in the work by Richards et al., leucine is commonly mutated to more polar side chains, such as asparagine or glutamine in order to change the polarity of the chemical environment of the active site.[227] Differing residue chain length could possibly result in different interactions with the substrate or impact the network of hydrogen bonds orchestrated by active site residues and solvent water molecules.

Some other residues targeted for their potential roles in binding were the phenylalanine 326 positioned in manner similar to that of Phe87 of cytochrome $P450_{BM-3}$, isoleucine 504 that appeared to assist in the positioning of substrate through steric interactions, tyrosine 330 which is the conserved tyrosine residue that is important in the catalytic cycle as a member of the proton network, and leucine 396 which appears to have a role in water channel dynamics, acting as a gate.

MD simulations have been used to aid in the development of $P450_{BM-3}$ mutants.[46] Hydrophobic active site residues were observed in interacting with substrate allowing researchers to characterize their functional and structural relationships within the active cavity of $P450_{BM-3}$. Using mutant libraries of these substrate interacting residue mutations led to the rational design of P450s with shifted or improved catalytic efficiency and regio- and stereoselectivity.[19,42,43,45] The previous work inspired the method of selection of residues for the mutation of the Cyp4f13 homology model in this study.

Summary

In this chapter the constructed substrate-free murine cytochrome P450 4f13 structural ensemble was used to dock two metabolites of ω-hydroxylation of arachidonic acid: arachidonate and 20-hydroxyeicosatetraenoate.

Two conformations of arachidonate were simulated with different protein protonation states and used to probe the substrate carboxylate binding interaction of the Cyp4f13 system, with the second one containing an extra positively charged His residue in attempts to influence the negatively charged polar headgroup of the fatty acid substrate. Both sets of simulations converged on a comparable binding mode for the carboxylate headgroup, showing favorable interactions with Arg237 and Tyr96.

The protonation of a residue within the binding site led to interesting observations as to the potential evolutionary role of certain specific residue locations and states for the Cyp4f13 protein. When protonated, a histidine residue was observed in influencing significant displacement of structural elements within the murine P450. More specifically, that of the transmembrane anchor in relation to the heme domain, possibly indicating an evolutionary pH-gated mechanism for changing the binding interactions with its carboxylate substrates. It was theorized that this change in structural conformation resulted in a larger opening for solvent water to enter the active cavity, changing the interactions between the enzyme and its substrate, however, was found to be present in the D1 homology model. Several positively charged residues were observed to have changed interactions from this protonation state change and their roles in substrate binding should be investigated by future studies.

Simulations of the polar hydroxylated metabolite of arachidonic acid resulted in the hydroxyl group on the terminal carbon unit of the polyunsaturated fatty acid chain coordinating to the metal center and this interaction persevered throughout the lengthened simulation time. Coordination chemistry and thermodynamic arguments were used to explain these results on the basis that the desolvation of the Cyp4f13 homology model active site resulted in a high amount of binding specificity for the 20-HETE substrate. This was enough to restrict access to the heme-iron center from the more favorable solvent water ligand, with electrostatic charge parameters exceeding that of the hydroxyl group.

Simulations of the binding interactions of the P450 system and its substrates were produced and studied with model systems. Relative distances between the carbon units of the fatty acid substrate indicate possible sites of oxidation, and potentially metabolite profiles for the native protein. However, findings such as those produced by this work would need experimental evidence in order to corroborate these conclusions. What can be gained is insight into important residue interactions between the substrate and its physiological enzymatic target. The residues that were identified were Gly327, Arg237, and Leu137.

The resultant findings have been discussed in this chapter, indicating success in proper substrate docking as both simulations of the Cyp4f13-arachidonate complex converged on a single binding mode for the substrate. The coordination of the hydroxylated product, while interesting, also acted as another indication of proper methodology, as the coordination to the heme catalytic resting state was produced. What would be expected though, through longer simulation timeframe, would be the egress of

the polar substrate back out into the system, or at least to a channel of the protein through the solvation of the active pocket. Evidence for preference for an ω-hydroxylated product was unable to be obtained from this simulation, as the increase solvation did not ever allow the metal center to undergo ligand dissociation with water. The second state of the P450 catalytic cycle portrays the displacement of water upon the binding of substrate. The complete catalytic cycle is an oxygen-dependent process where dioxygen must bind to the reduced iron(II) species in order to produce the oxyferryl state capable of oxidizing a C-H bond in the P450 catalytic cycle. This simulation uses the iron(III) species having its electronic character only described by a modified parameter and topological files in the CHARMM force field format. This would be able to recover some of the electrostatic forces, however, can do nothing to resolve the lack of dioxygen simulated in the system, oxidation state changes, nor its changes of spin in the catalytic cycle. While the mechanism behind the formation of the oxidized product was not the focus of this study, it made the data obtained more difficult to draw conclusions from in deciding the role Cyp4f13 plays in ω-hydroxylation of its substrates.

As was mentioned in the previous chapter, the simulations are first and foremost, a thermodynamic sampling of the possible states of the substrate-or product-bound Cyp4f13 protein. Longer timeframe simulations, such as that of the microseconds range, would still be necessary to explore differences in sampled states, such as those proposed for the oxidized metabolite. The simulations of the substrate-and product-bound cytochrome P450 4f13 system were performed to gain computational insight into the active site and to prepare for the structural probing of the system through mutation.

CHAPTER V

Molecular Dynamics Simulations of Mutated Cyp4f13

Introduction

This chapter presents the process, results, and findings in the simulation of the mutated substrate-bound murine Cyp4f13 structural ensemble. This work is necessary to improve the understanding of ω-hydroxylation catalyzed by the murine Cyp4f13 system which might possibly provide insight into the physiological role of the murine Cyp4f enzyme where no clear analogous pairing has been made to a human CYP4F.

The substrate-bound state of the murine P450 catalytic cycle docked with the converged conformation of the substrate of interest (D1-Cyp4f13-AA), was used as the basis to generate mutants that were determined to be likely to be relevant to binding specificity in the previous chapter. Furthermore, the addition of the covalent heme linkage through an esterification by a hydroxymethyl group on the heme and neighboring glutamate residue was also generated and patched into the ensemble. As noted earlier in this thesis, this is known to have significant impact on regiospecificity of oxidation in related P450 CYP4 family members. Parameterization was performed by analogy through (1H-pyrrol-3-yl) methyl acetate. Two independent mutations were also generated on active site residues deemed important in substrate binding of arachidonic acid, as determined in the previous chapter. Drastic changes to chemical environment through mutation was carried out to investigate interactions within the active pocket for both the substrate and the solvated protein. Hundreds of nanoseconds of simulation were performed on all three mutants of arachidonate-bound murine Cyp4f13. The focus of this

work was to investigate the impact of specific site mutations on key residues responsible for different aspects of substrate binding.

Also noted are the impact of structural differences within the active pocket as electronic conclusions should not be made based on molecular mechanistic behavior. The intention is to develop insights that will allow for understanding to be made on isoform-specific characteristics of the murine Cyp4f13 active pocket. This is done to further the understanding of the physiological role and substrate specificities of the murine Cyp4f13 system. With the information gained in this chapter, future studies can be designed and performed to elucidate their metabolic roles, and consequently, as possible therapeutic targets for medicine.

Computational Method and Details

Mutagenesis through VMD Mutator Plugin

Mutagenesis was performed through the VMD Mutator Plugin which requires the atomic coordinates of the protein, and its protein structure file. Its requirement of a protein structure file means that the data from a previously constructed homology model must be used and cannot be performed on purely crystallographic data.

At this point, a newer rabbit CYP4B1 crystal structure that looked like a more optimal homology model template became available.[221] A request was sent to the I-TASSER server and the best resulting homology model produced was used in the R237L and L137Q mutations of the Cyp4f13 system. This new structure was of a non-linked heme cofactor mutant of the rabbit CYP4B1, making it an improved template for native

Cyp4f13. This, however, was not necessary for the modeling of the G327E mutant of Cyp4f13 as the covalent heme linkage was the ultimate goal of that portion of the experiment. Thusly, the original native Cyp4f13 homology model "D1-Cyp4f13-AA" was used for the covalently linked G327E mutant, while the newer model "MX-Cyp4f13-AA" was used for the R237L and L137Q Cyp4f13 mutants.

Parameterization of Covalent Heme Linkage

With the covalent heme linkage not having a set of topology or parameter values through analogy or otherwise, it would need to undergo the full parameterization process.

The software used to parameterize the novel linkage in this study is the Force Field Toolkit Plugin (ffTK) of the Visual Molecular Dynamics (VMD) program package in addition to the Gaussian 09w computational chemistry program suite. This toolkit utilizes Nanoscale Molecular Dynamics (NAMD) calculations in addition to QM calculations performed by the Gaussian 09w program along with a graphical interface to generate CHARMM parameter constants and topological information for theoretically any novel residue in the CHARMM format using the CGenFF force field.

The first step in parameterization of a novel residue is the identification of missing parameters. This is done by chemical analogy and is an important step as the parameters obtained are heavily dependent on how similar an analogous compound is to the target novel residue. The analogous compound used was the (1H-pyrrol-3-yl) methyl acetate species, pictured in Figure 62. The species contains the aromatic environment of the heme porphyrin moiety in addition to the ethanoate ester formed by the proximal glutamate residue *in situ*.

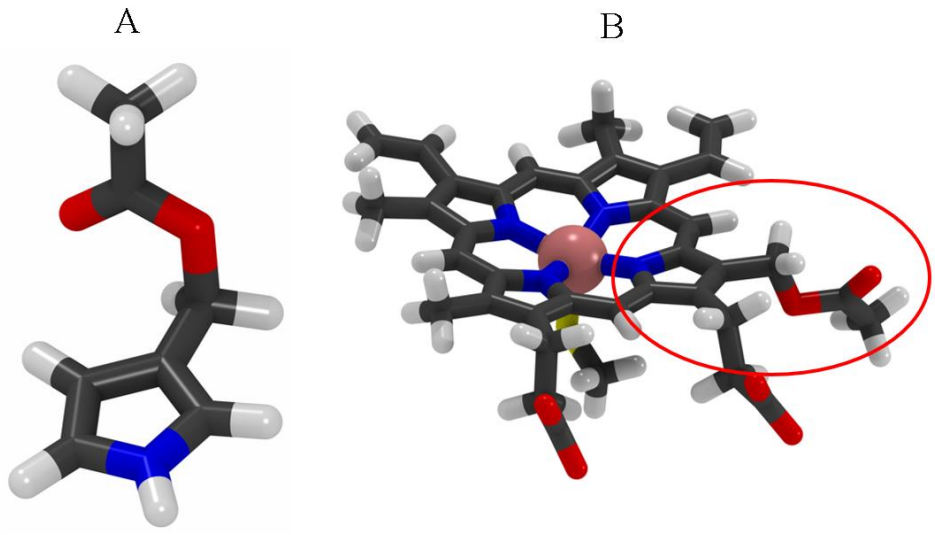

Figure 62. Side-by-side comparison of the novel residue (A) and the covalent heme linkage (B). Both models are represented as licorice models with the analogous chemical components encircled on the heme cofactor

The compound was constructed using the GaussView 5.0 interface, its coordinates exported in a PDB format to be used, analyzed and processed by the ffTK program. For each unique bonding between atom types, a blank template was created for the final parameter set.

The CGenFF force field topology and parameter files were used to define van der Waals (LJ) constant values for the atom types discovered in the target residue. Each atom type discovered by the ffTK was manually updated with the corresponding value found in the CGenFF topology file was set from reference. After all the possible values to be parameterized were assigned, the parameterization process was commenced using a variety of QM methods.

The initial geometry optimization was performed with DFT on the novel residue to arrive at an energetically-minimized, equilibrated state. The level of theory used is reliant on the system being measured and is dependent on the level of detail needed by

the researcher. Higher levels of theory are generally avoided for initial optimization, as molecular mechanistic (MM or MD) approaches result consistently with appropriate results with the except of sophisticated transition metal complexes.

The model chemistry used to create the initial geometry optimization was the density functional theory with the B3LYP functional. This functional is commonly used in heme-protein systems and provides for a fairly accurate representation of the organic moieties of metallo-protein complexes. The Pople 6-31G double zeta split-valence basis set was used on the novel residue in combination with a diffuse function on heavy atoms, and polarization functions for both heavy atoms and hydrogen. This model chemistry, 6-31G+(2d, p), was chosen to be a moderate level of rigor to achieve a good starting point for the equilibrium geometry for the residue and would be refined using the quadratically convergent self-consistent field (SCF) method to assist in convergence. DFT was chosen due to its ability to converge more readily than *ab initio* Hartree-Fock calculations.

The optimization of partial charges was produced through QM-derived interactions with theoretical water models. After assigned hydrogen-bonding donors and acceptor atoms on the novel residue, ffTK assigned partial charges to heteroatoms based on their optimized interactions with TIP3P water molecules. Single point energy calculations were used to optimize the interactions between the polar charge points on the rigid water model. The resulting output from the calculation was used to deduct what partial charges to assign individual atom types to collect topological information for the novel residue.

The fully optimized bond distances and bond angles were determined through combined QM and MM methods. In the search to find the minimized bond strain on the

to-be parameterized system, perturbations are performed on the conformer geometries that resulted in a potential energy surface (PES). In order to avoid local minima or saddle points in the surface, a Hessian (or force) matrix is used to distinguish between stationary points on the PES and was performed at a higher model of chemistry.

The model chemistry used was the Møller–Plesset model (MP2), a second order perturbation theory. This *ab initio* method includes a corrective term for electron correlation effects that is neglected in Hartree-Fock calculations. The 'tight' self-consistent field option was used in the Gaussian job so to narrow the convergence criteria, allowing for less room for trivial or unwanted solutions to the Hessian calculation.

After the Hessian calculation was performed, the QM portion of the bond distance and angle analysis was handed off to molecular mechanics. The ffTK plugin uses several quick iterations of MD simulations through the NAMD software, with the values obtained from the Hessian QM calculation as a starting point for its parameters. After numerous femtosecond simulations are completed, the ffTK plugin saves the resulting equilibrium bond distances and angles as another output file to update the in-progress parameter file.

The final step of the parameterization was the dihedral and improper angle optimization through torsional scanning. This portion of the method consisted of another combination of QM and MM calculations. The ffTK plugin generated two Gaussian job input files for every torsional angle between heavy atoms. A 90-degree scan was taken for the rotation in the positive (clockwise) and negative (counterclockwise) directions. The plugin then used molecular mechanistic simulations through NAMD once more to

find a local minimum in the potential energy surface created by the QM portion. The values that ffTK designated were used as a local minimum that was set as the initial point for additional MD calculations. This process is referred to as ‘refitting’ and this iterative method was performed until the energy presented by ffTK increased from the last refinement run, indicating the arrival at a global minimum. Finally, with the updated final set of parameters optimized, the resulting values for the (1H-pyrrol-3-yl) methyl acetate residue could be assessed by comparison to related systems and then added to the CHARMM parameter files for MD simulation.

Molecular Dynamic Simulations of Mutant Cyp4f13

A different set of specifications was used in the refinement simulations and longer timeframe production simulations than the previous simulations. In order to simulate a more unrestrained environment to explore interaction changes of the individual mutations, a timestep of 1 was used in the integrator, meaning that atomic velocities were calculated at twice the rate than previous simulations. Rigid hydrogen bonds were only applied to solvent water molecules, allowing for dynamic C-H, N-H, and O-H bond motion of the protein structure. This would result in a larger calculation wall time but was done in attempt to achieve more reliable results.

Table 7

Table of mutant homology models for the Cyp4f13 structural ensemble

Model Name	Residue Mutation	Covalent-Bound Heme
M1Cyp4f13	G327E	Yes
M2Cyp4f13	R237L	No
M3Cyp4f13	L137Q	No

Note. The D1-Cyp4f13 parameter files were used for M2 and M3 homology models with M1 receiving the additional parameterized values to be used in the simulation.

A total of 100-ns production simulation time was allotted for each of the mutant Cyp4f13 homology models. The model used to generate the G327E Cyp4f13 mutant was the D1-Cyp4f13-AA structure, based off the relevant rabbit CYP4B1 template with a covalent heme linkage. A new homology model was obtained using a newer mutant rabbit CYP4B1 crystal structure with its covalent heme linkage removed. This was used as the basis for the R237L and L137Q models, as they did not contain covalently bound heme cofactors. The reasoning, results and findings will be discussed in the appropriate sections.

Results and Discussion

Parameterization of (1H-pyrrol-3-yl) Methyl Acetate

The results of the parameterization of the (1H-pyrrol-3-yl) methyl acetate residue, acting as an analogous representative for the covalent heme linkage, was successful in producing values for all the identified missing parameters. Whether or not the values are accurate, or at least representative, was examined by comparing the QM-derived bonding

geometries and partial charge distribution to similarly related compounds and through literature review.

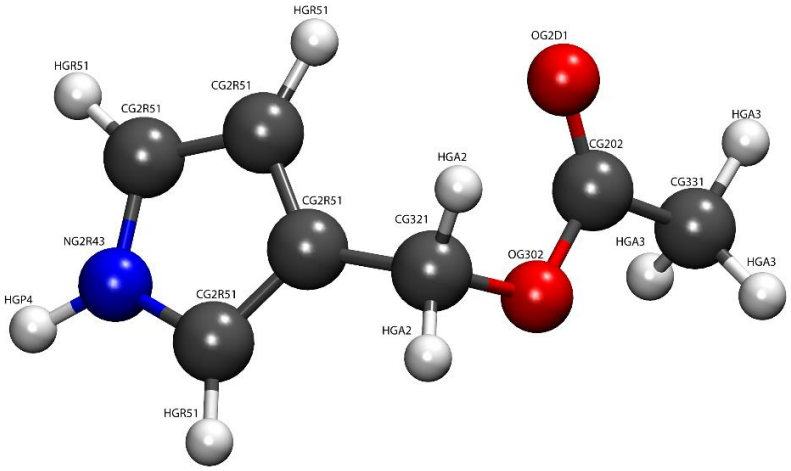

Figure 63. Atom designations for the (1H-pyrrol-3-yl) methyl acetate residue in the CGenFF format. A ball-and-stick representation is used for the analogous chemical component of the heme covalent linkage.

Using the related amino acid residues, which are contained in the CHARMM 36 protein parameter files, results obtained from the parameterization of the porphyrin moiety were within reason.

Table 8

Comparison of parameterization results for aromatic C-C bonds and similar chemical bonding environments

Residue	Aromatic C-C Bond Force Constant (kcal/mol/Å^2)	Aromatic C-C Bond Distance (Å)
Novel Residue	365	1.39
Phenylalanine	440	1.34
Tryptophan	350	1.43

Note. Phenylalanine and tryptophan parameter values were obtained from the CHARMM36 force field

The aromatic carbon-carbon bond on the porphyrin portion of the covalent heme linkage residue results are present in Table 8 and show an equilibrium bond distance and force constant between that of the CHARMM parameters for benzene of a phenylalanine structure and that found in the aromatic indole group in tryptophan. The differences could be explained to be due to the different bond angle strain found in the pyrrole ring analogy of the porphyrin ring compared to the planar 6-membered benzene, and of the fused indole ring of tryptophan.

Table 9

Comparison of parameterization results for the C-N-C bond angle and similar chemical environments

Residue	C-N-C Angle Force Constant (kcal/mol/rad^2)	C-N-C Bond Angle (Degrees)
Novel Residue	180	112
Histidine	145	108
Tryptophan	85	112

Note. Histidine and tryptophan parameter values were obtained from the CHARMM36 force field

In comparing the C-N-C bond angle for the novel residue in Table 9, similar species were chosen from CHARMM protein parameters. The nitrogen-containing aromatic amino acids tryptophan and histidine show very similar equilibrium bond angles but have varying force constants associated with them. The novel residue shows the highest force constant, which could be rationalized by the difference to the larger indole ring of the tryptophan and the two-nitrogen containing imidazole found in histidine. This would imply a larger energetic component necessary to bend the porphyrin C-N-C bond

as opposed to that found of the tryptophan residue. This might be an artifact of the usage of low-frequency vibrational modes from experiment to create more accurate physical parameters in the CHARMM36 force field.

The mutation and glutamate residue linkage appeared to have been properly parameterized. As an example, the equilibrated angle of the ester C-O-C bond between the glutamate and hydroxymethyl group on the heme fell within reason in comparing against comparable ester bonds in the glyceride esters in the CHARMM lipid parameters (109.6 degrees in the parameterized residue vs. 110 in acyl-acetate ester and 109 in methyl acetate).

Calculated by their interactions as either hydrogen-bonding acceptors or donors, the atoms within the residue displayed very reasonable results for partial charges. The carbonyl and ester oxygen atoms received a partial charge of -0.625, and -0.503 respectively. This appeared within reason when compared to the -0.63 and -0.49 to the ester found within the di-oleoyl phosphatidylcholine lipid.

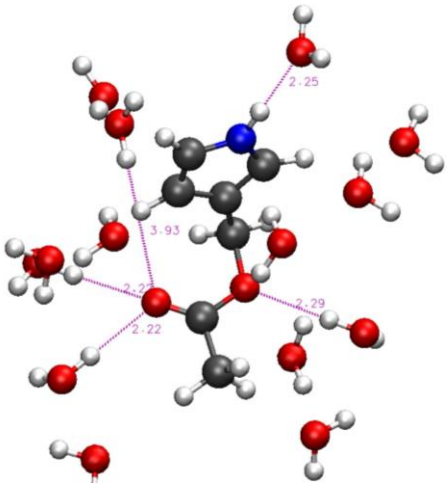

Figure 64. Ball-and-stick representation of the TIP3P water-based interactions with the novel residue to develop the partial charges for the force field topology files. Hydrogen-bond distances between the covalent ester linkage are indicated with red lines and labels.

Comparing to literature gave satisfactory results; when comparing to a parallel study on the homology modeling and MD simulation of a CYP4F system, Li et al. reported their parameter values in their supplementary information.[223] Their parameter values, while in different units and format than used in CHARMM, match up fairly well with the results in partial charge and bond geometries produced through ffTK. Their team employed used of multilayered QM/MM calculation through the Our own N-layered Integrated molecular Orbital and molecular Mechanics (ONIOM) method. This was originally going to be within the scope of this study, however calculations of the truncated models to be used in the QM/MM calculation never converged and a different methodology was considered due to time considerations.

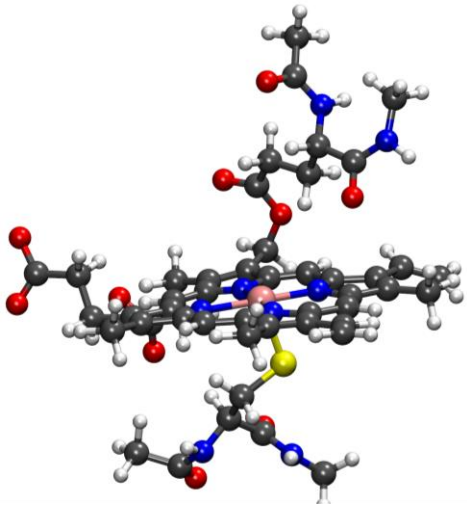

Figure 65. Proposed truncated model of the heme-thiolate covalent linkage. The model contains the usual heme porphyrin IX substituent groups, in addition to both the heme-thiolate cysteinyl residue and the esterified glutamate residue, capped with an N-terminal acetyl group and a C-terminal methylamine group.

Figure 65 depicts the more appropriate chemical analogy of the covalent heme linkage than the novel residue, however it became too unwieldy to parameterize and much of the chemical bonding considerations of the ester moiety could be recovered with the parameterization process. With the parameterization of the (1H-pyrrol-3-yl) methyl

acetate residue completed, the modifications were added to the CHARMM topology and parameter files along with three additional patches, found in the appendix.

The first patch (HEMED) facilitates the removal of one of the hydrogens belonging to the 5- methyl substituent of the porphyrin IX cofactor. It also serves to reassign partial charges to that of a methoxy from its methylene origin. With physiological glutamate expected in the protein, there is no proton to remove for the glutamic acid residue. Instead, the second patch (GUU) does a similar procedure in charge reassignment, changing the carboxylate partial charges to that of an ester. It also reassigns the atom types of the two carboxylate oxygens to one as a carbonyl oxygen of an ester, and the other as the bridging ester oxygen. The last patch (GUUH) directly bonds the now ester oxygen of the glutamate residue straight to the carbon of the methylene group of the heme. With this, preparation for the MD simulations of the covalent linked G327E mutant Cyp4f13 homology model could commence.

MD Simulation of Gly327Glu Cyp4f13 Mutant with Covalent-bound Heme

A preparatory model for the mutant was created with the removal of the lipid bilayer, arachidonate substrate, solvent, and ions of the D1-Cyp4f13-AA model. Using the Mutator Plugin of the VMD software, glycine 327 was selected as target for mutation. The nuclear coordinates of the new glutamate residue atoms were guessed by Automatic PSF builder and assessed by visual inspection. A script was used to load up individual sections of the structural ensemble where the patches were incorporated after patching of the heme-thiolate cofactor and applied to the resultant glutamate 327 of the mutant Cyp4f13 protein.

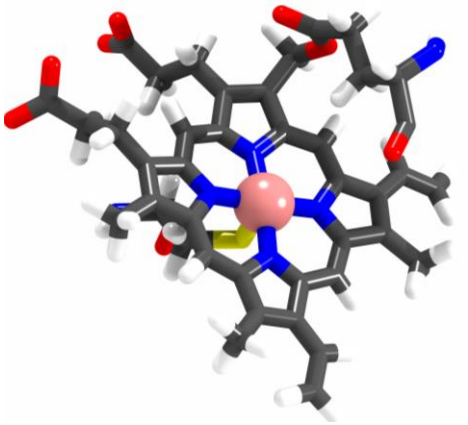

Figure 66. Structural view of the final covalently attached heme moiety of the minimized G327E Cyp4f13 mutant homology model. Licorice representations of the involved amino acid residues are added for emphasis.

An additional script, similar to the previous preparation scripts, was used to create the same DOPC bilayer structural ensemble with the docked arachidonic acid substrate in-tow. This final model was simulated for an NVT production run of 100-ns, using a smaller integrator timestep and vibrational limitations eased for protein hydrogen bonds.

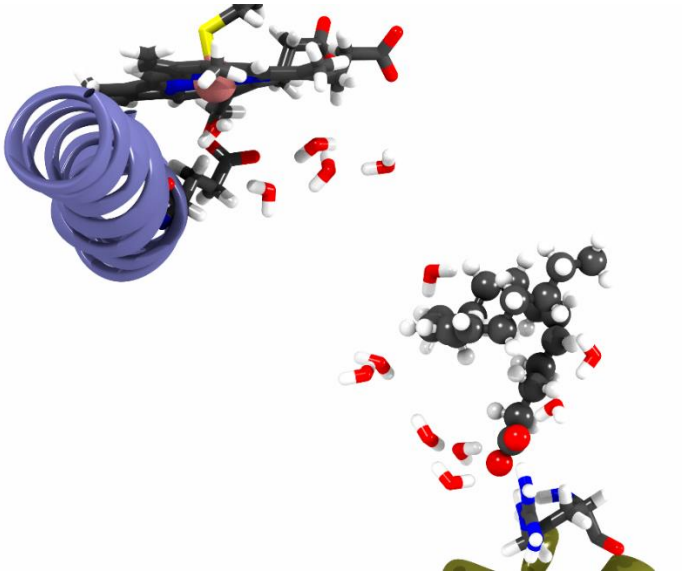

Figure 67. Structural view of the G327E Cyp4f13 mutant after 100-ns production simulation. Water is represented as licorice models and is displayed for those within 3 angstroms of the heme iron or the arachidonate substrate.

The results were less than ideal for the desired effect to be studied in this research. The substrate was almost completely displaced from the active pocket by solvent TIP3P water molecules, depicted in Figure 67, as their hydrogen bonding network sufficiently disturbed the non-polar forces from binding tightly to the arachidonic acid. The fatty acid substrate was located too far from the heme iron to be reasonably expected to be oxidized.

Between these results and the past runs, several differences in the simulation process were considered and could explain the events observed. The first major different was the addition of the glutamate ester linkage of the covalent-bound heme cofactor. With the addition of a residue sidechain with sufficiently higher hydrogen bonding capability (glycine vs. glutamine), the carbonyl oxygen atom might be extending out into the active site enough to have caused sufficient disruption to the non-polar interactions necessary for binding specificity. Water found to be hydrogen bond donating to the carbonyl group would provide evidence for this hypothesis. However, what might appear to be the trivial solution does not appear to be the case as no observable increase in hydrogen bond formation with solvent water molecules observed in its structure compared to the substrate-free state.

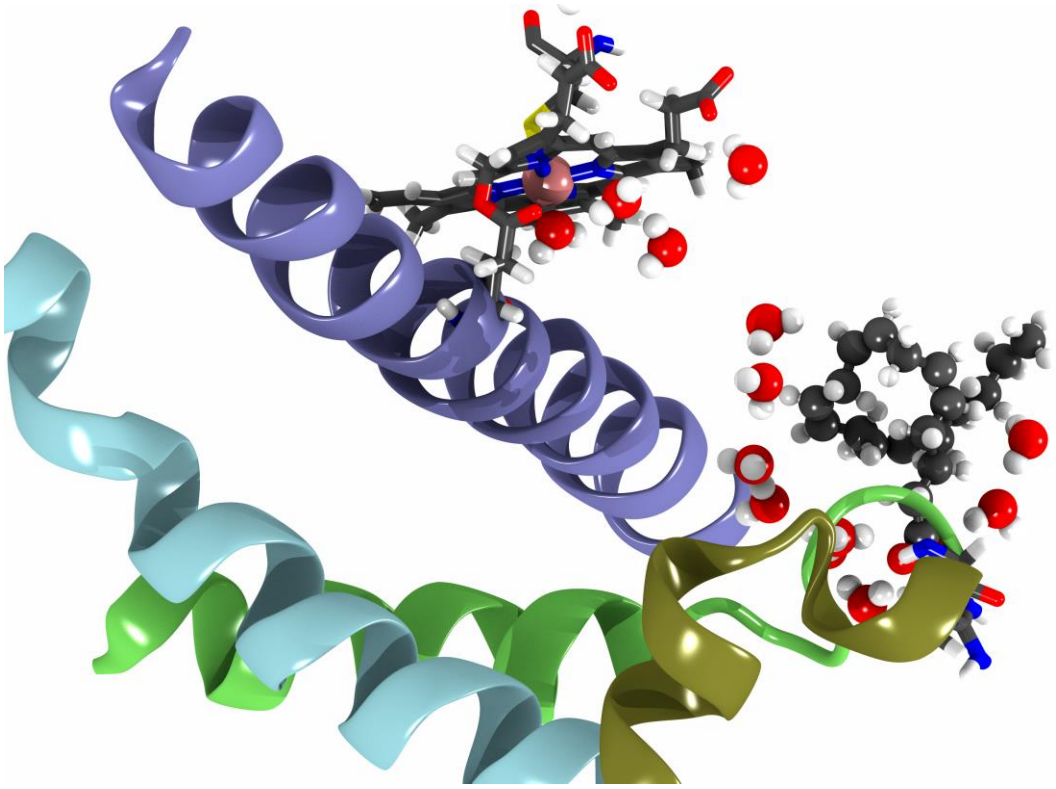

Figure 68. Different structural view of a frame from the production simulation of the G327E Cyp4f13 mutant. Ball-and-stick representations for the all water molecules within a 3-angstrom distance from the iron and substrate are present.

The next difference to be discussed is that of the change in simulation specifications. With the usage of smaller timesteps and removal of the hydrogen bond vibration constraint for the protein structure, it could be possible that the degrees of freedom previously held back water from entering the active pocket due to rigid amino acid hydrogen bonds. An additional simulation was run at the newer specifications on the older D1-Cyp4f13-AA model to investigate the differences, however, resulted in a very similar simulation of proper arachidonic acid binding, with no advanced displacement or egress of the substrate. Such a phenomenon was also not observed in the R237L nor the L137Q mutant simulations, ruling out this possibility.

The final hypothesis presented in this thesis is that the polar carbonyl group influenced the active site's hydrogen bonding network, and perhaps impacted the secondary structure in a subtle way as to change the interactions necessary for arachidonic acid binding. Further to mention is that of the result of thermodynamic sampling, as this could possibly be a result of 'bad' initial trajectories. More simulations with different initial conditions would need to be performed in order to eliminate this possibility.

MD Simulation of Arg237Leu Cyp4f13 Mutant

Arginine 237 was selected for mutation using the Mutator plugin of the VMD software, as the principal residue involved in carboxylate binding of fatty acid substrates for the Cyp4f13 system.

The R237L Cyp4f13 mutant was created using the same preparatory script as the previous simulations, however, they did not use the modified parameter files present in the G237E Cyp4f13 model. With the glutamate mutation not necessary and covalent heme linkage not the focus of this portion of the study, the original parameters used in Cyp4f13 simulation were used instead. As was previously noted, this model was constructed with the more recent rabbit CYP4B1 crystal structure with a non-covalently bound heme. The impact of this mutant P450 and its crystallization is not fully known and would provide for some interesting thought as its modification might have unpredicted effects on the crystallographic data obtained from it.

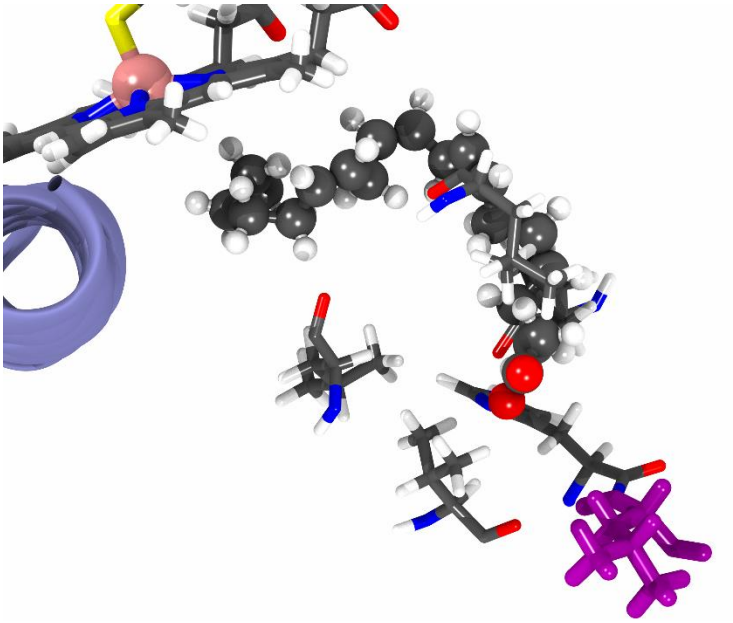

Figure 69. Structural view of the pre-minimized M2Cyp4f13 homology model structure. Purple licorice representation of the mutant Leu237 residue was added for emphasis. Residue sidechains within 4 angstroms of the carboxylate headgroup of arachidonate are represented as well.

Shown in purple in Figure 69, the R237L mutation of the long, positively charged arginine residue to the shorter, non-polar, branched leucine was the first used on the newer Cyp4f13 homology model. The expected effect would be that of a drastic change to chemical environment for the binding of arachidonic acid for the Cyp4f13 enzyme. After 100-ns of simulation time was allotted, the results indicated an interesting finding.

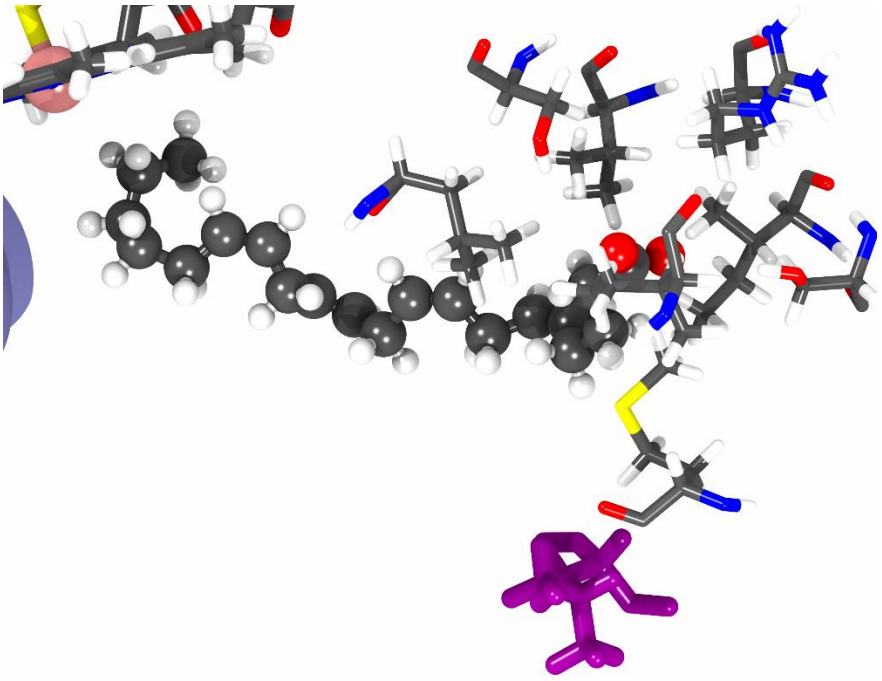

Figure 70. Structural view of R237L Cyp4f13 homology model bound to arachidonic acid after 60-fs of minimization. Representations were added for the residues near the carboxylate of the substrate in the same manner as the previous figure. An arginine residue (Arg100) in observed in substrate binding, as the substrate has migrated away from the mutated leucine residue.

A change in binding modes is observed as the carboxylate headgroup migrated several Å away to another arginine residue, Arg100. Depicted in Figure 70, Arginine 100 appears to be the new principal residue responsible for the binding interactions with the carboxylate of the substrate. This is not entirely surprising as the simulation on the 20-HETE metabolite within the D3-Cyp4f13 homology model in the previous experiment was also found to form hydrogen-bonding interactions with both Arg100 and Arg237.

This mutation displayed additional effects that are of interest in the scope of this research. Better orientation of the polyunsaturated fatty acid resulted in the favorable positioning to facilitate the ω-hydroxylation of its substrate. The distance between the terminal carbon and the heme-iron indicate an appropriate distance to have an oxidation event take place, as the oxygen of the oxyferryl species would protrude out from the

plane of the heme and is sometimes observed to be within close proximity of the primary carbon of the substrate. However, the space between the heme-iron and the ω-hydroxylation site is impeded by several water molecules.

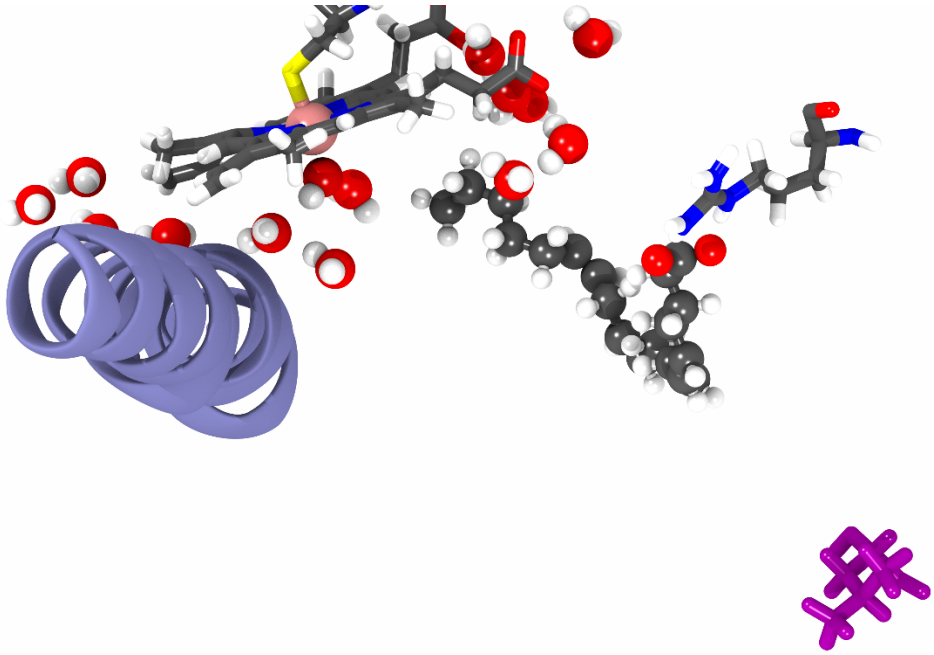

Figure 71. Structural view of the solvent environment within the R237L Cyp4f13 homology model active site near the end of the production simulation. Substrate and water molecules within a distance of 5 angstroms of the heme cofactor are represented in van der Waals models.

Figure 71 shows the final frames of the simulation which reveal that several TIP3P water molecules are observed in close proximity to the catalytic site; with one directly coordinating to the heme-iron, and others observed in a cluster forming hydrogen-bonding interactions with the protein structure and heme propionate groups. The majority of the active pocket of cytochrome P450s are lined with non-polar residues like leucine, isoleucine and non-polar aromatics to interact favorably with their hydrophobic substrates. These interactions, however, appear disturbed by hydrophilic forces on the binding of substrate and are observed in the simulation.

This is not a grim omen for the simulation, as the P450 catalytic cycle *in vivo* requires water to be accessible to the active site to have a protonation event to occur. This happens through water entering via a gated aqueduct near one of the propionate groups on the heme.[22] The gated mechanism occurs through a concerted hydrogen bonding interaction between the backbone of the protein and the salt bridge formed by the heme propionate with the arginine of the conserved P450 heme-binding motif. The water channel and proton relay formation of P450s is well studied with computational methods as seen in the work by Dubey and Shaik.[22]

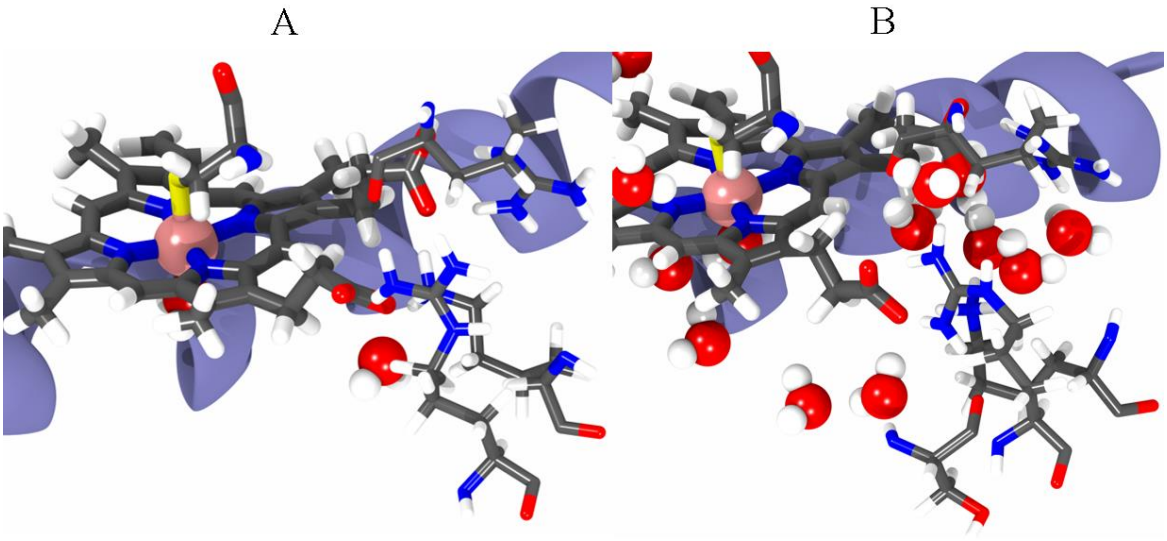

Figure 72. Side-by-side comparison of the R237L Cyp4f13 homology model heme propionate binding site. The frames shown are from the beginning (A) and end (B) of the production simulation of 100-ns. Water is drawn in van der Waals models within 4 angstroms of the heme cofactor in both frames.

The position of the salt bridge between the heme propionates and positive residues, as seen in Figure 72, within the heme binding site still indicate substantial interactions with the positive residues of the mutant Cyp4f13 model. This makes it difficult to conclude that the proton relay formation is what is being portrayed by the

simulation, as this interaction is dissolved by the tugging of the arginine residue by a backbone amide hydrogen.

However, what is not being observed in this simulation is the departing of water molecules from the coordination site due to the binding of a substrate. As reported by Dubey and coworkers, a simulation time of 350 ns was used to simulate the $P450_{BM-3}$ system and its interactions with *N*-palmitoylglycine. This might mean that the results presented in this thesis are too early of a sampling in the simulation of these interactions, indicating the need for additional simulation time to be performed on the system in order to observe such an effect.

The distinct channel relocation observed in P450 systems when binding to substrate appears to be still underway in the simulation of the mutant Cyp4f13 model, and conclusions are hard to make on the impact of the R237L mutation on substrate specificity. What can be said, however is the definitive rearrangement of hydrogen-bonding interactions for the binding of the carboxylate head group of arachidonic acid to the protein structure. The clear path forward for this research that could provide insight into the observations of these trajectories would be running much longer calculations. Channel relocations and proton relay formation could be observed in microsecond timescales, as opposed to the nanoseconds used in these simulations. With the necessary resources, long timeframe simulations with unrestricted TIP3P water models could result in a desolvation of the active pocket and formation of appropriate aqueducts around the heme group, indicating complete substrate binding.

One additional consideration that is implied from the results of this mutation is the uptake of the substrate into the active pocket. Lined with many positive residues, the

Cyp4f13 access channel to the membrane bilayer serves as the mode of entry for its potential eicosanoid substrates. Whether or not Cyp4f13 catalyzes oxidations on free fatty acids or acyl arachidonate esters from phospholipids has not been elucidated as of yet. However, how its substrates are taken up by these membrane-bound proteins is an interesting question. With the removal of the arginine at the membrane surface, the simulations might imply better substrate binding; however, they might indicate a worsening of substrate (kinetic) uptake.

It is proposed that the way hydrophobic substrates are taken up by microsomal P450s is through becoming dissolved in the membrane bilayer. A nonpolar region of the P450 heme domain sits beneath the plane formed by the polar head groups of the phospholipids in the membrane and positive residues line the surface of the protein adjacent to the hinge of its transmembrane anchor.[229]

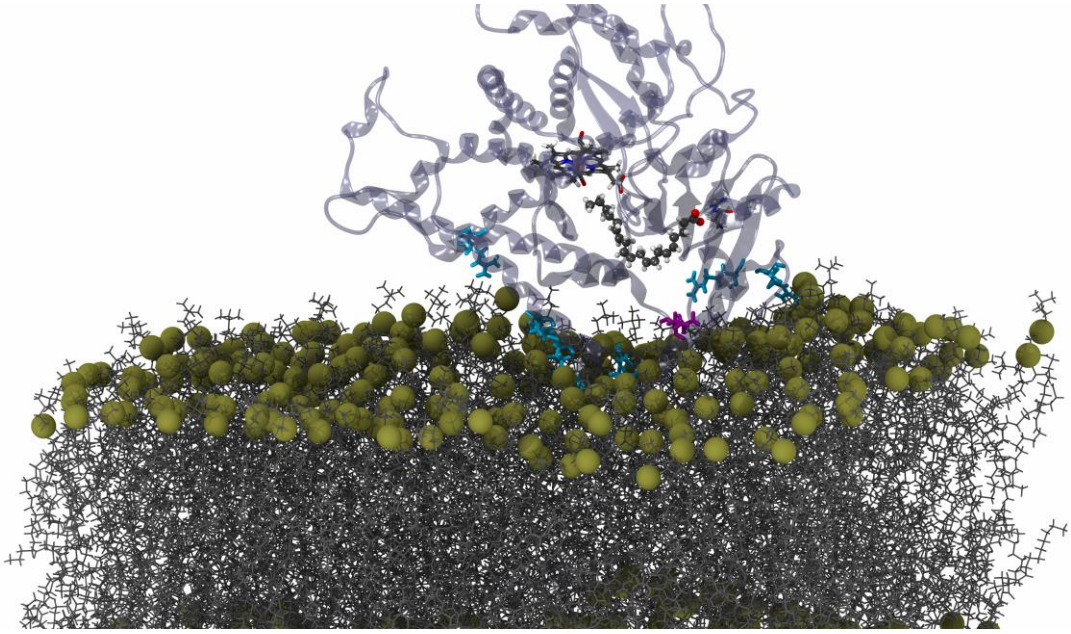

Figure 73. Zoomed out view of the mutant membrane-embedded Cyp4f13 homology model and nearby DOPC lipids. The R237L mutant in represented in purple and blue licorice was used to distinguish the arginine residues near the DOPC lipids.

The F-G loop region in the murine Cyp4f13 is enriched in arginine sidechains, such as observed in Figure 73, which serve as binding coordinators for substrates. This is performed through electrostatic attractions to dissolved carboxylate headgroups of free fatty acids, or possibly negatively charged phosphate groups of the surrounding phospholipids. It is currently unknown whether cytochrome P450s can oxidize esterified long chain (>20 carbon units) fatty acids still attached to their diacylglyceride phospholipid carriers. The angle in which the heme-domain of cytochrome P450s is studied and the work performed by Otyepka et al. has been referred to in prior sections. With the mutation event changing the charged environment of the P450s F-G loop, significant changes in membrane-orientation might be observed over simulation time.

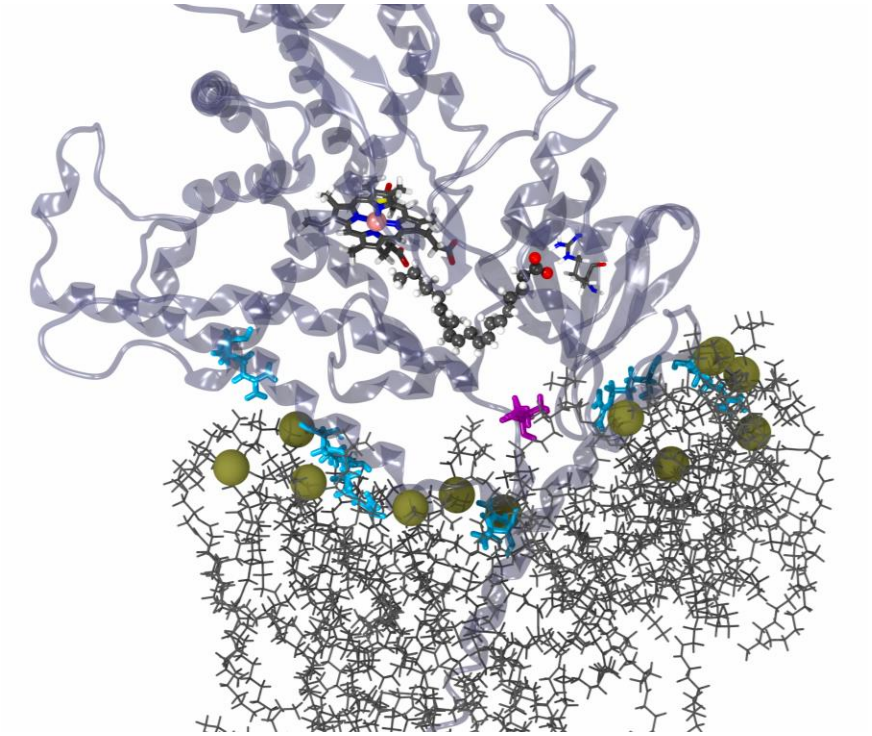

Figure 74. Close-up view of the membrane-embedded portion of the mutant Cyp4f13 system. Phosphate groups are indicated by yellow van der Waals sphere. The R237L mutant is represented in purple and the arginine residues in proximity to lipid bilayer are in blue.

The postulated model for fatty acid binding could require the presence of that mutated arginine 237 residue, as the molecular docking and subsequent simulation produced enough evidence to suggest the residue as being important in arachidonic acid binding to the P450 system. While the binding observed in the mutant, showing favorable interactions with the carboxylate of the substrate and arginine 100, is preserved, the uptake of the fatty acid by the protein might have been negatively affected. One possible experiment that could be used to answer the question posed by this mutation would be to simulate the coordinates of the arachidonic acid substrate a small distance away from the transmembrane anchor hinge and access channel of the protein, and allow sufficient time to observe uptake into the nonpolar active pocket of the mutant Cyp4f13 model. This mutation and its effects on the membrane-orientation, protein-membrane interaction, and substrate binding could provide useful in the study of these mammalian cytochromes.

MD Simulation of Leu137Gln Cyp4f13 Mutant

As one of the non-polar residues responsible for hydrophobic interactions with the cytochrome P450 4f13 substrate, arachidonic acid, leucine 137 was chosen to be mutated in order to observe its effects on binding interactions.

The L137Q Cyp4f13 homology model was created using the same preparatory script as the previous set of simulations and like with the R237L mutant, did not use the modified parameter files present in the G327E Cyp4f13 model. The original CHARMM parameter files were used instead. As before, this model was constructed with the more recent rabbit CYP4B1 crystal structure with a non-covalently bound heme.

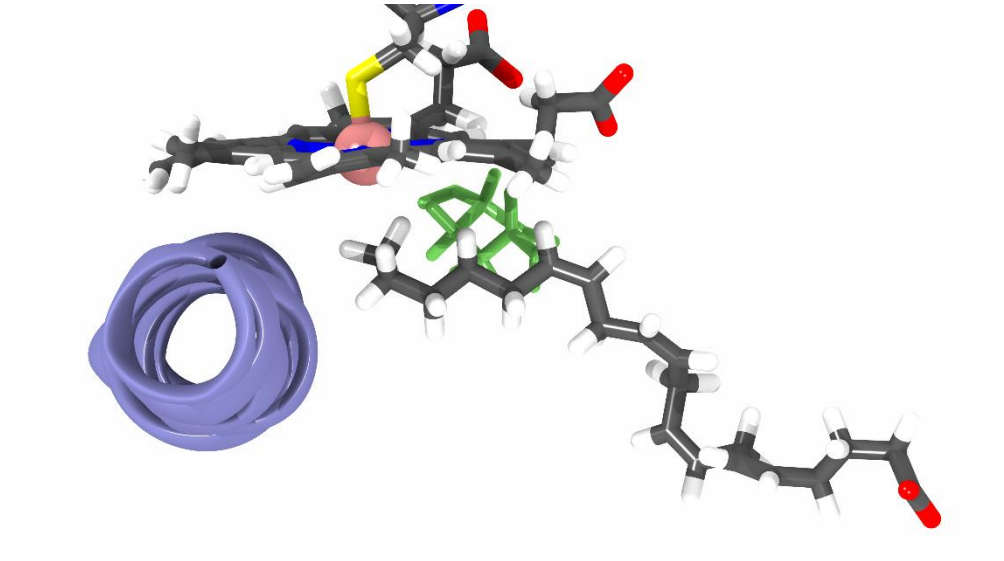

Figure 75. Structural view of the minimized L137Q Cyp4f13 homology model structure. Ball-and-stick representations of the changed residue were added for emphasis.

The expected effect of this L137Q mutation, depicted in green in Figure 75, of the nonpolar aliphatic leucine residue to the polar glutamine is that it should produce a considerably different chemical environment for the binding of arachidonic acid for the Cyp4f13 enzyme. With the amide group of the glutamine sidechain providing for a large increase in hydrogen-bonding capability for the active site residue, a perturbation of the substrate interactions by solvent water molecules is predicted, as there would be increase in polar interactions present in the predominantly nonpolar catalytic site. Since the previous two simulations on mutant Cyp4f13 homology models were observed in having pronounced solvation of the active cavity, as seen in Figure 68 and Figure 71, a more exaggerated event was predicted to occur with this L137Q mutation.

The 100 ns NVT simulation gave contrary results, as the modification did not appear to increase the solvation of the active site any more so than the previous mutations, and furthermore resulted in a relative increase in desolvation of the protein's active site around the hydrophobic substrate. The simulation displayed less water activity

within the cavity, shown in Figure 76, which goes against initial predictions. The glutamine mutation did however find interaction with water molecules and this persisted throughout the simulation.

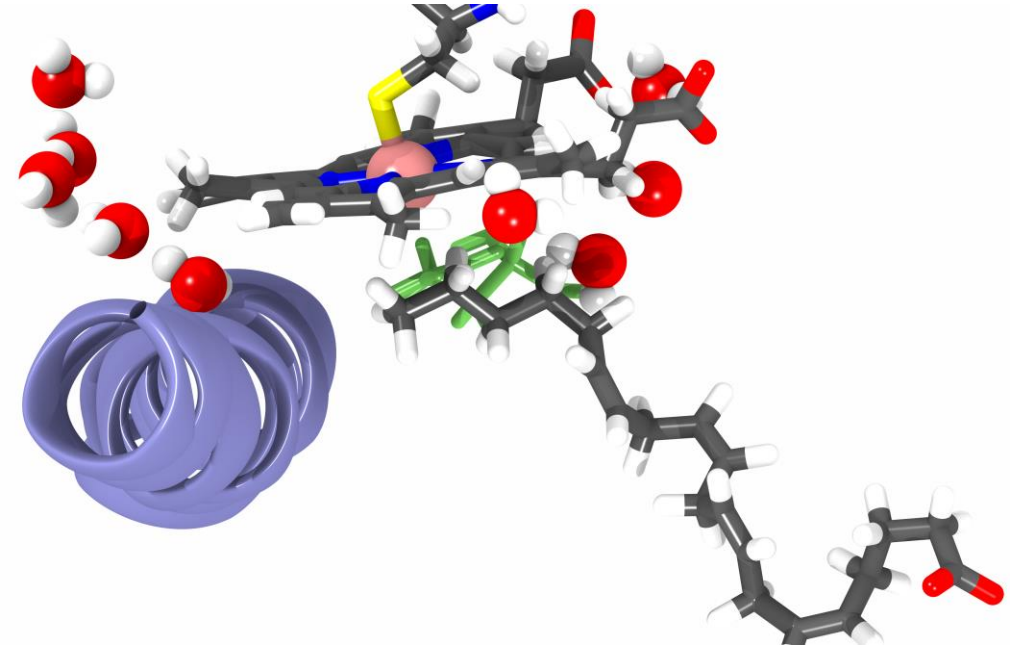

*Figure 76.*Water within active site for the L137Q Cyp4f13 homology model after production simulation. Licorice representations for the heme cofactor, arachidonate substrate, mutant glutamine and vdW water molecules within 4 angstroms of the heme are present.

Water was found coordinating to the amide carbonyl of the mutant L137Q residue, interacting with the backbone of the protein as well. Shown using large trajectory stride steps along the production simulation in Figure 77, the polyunsaturated fatty acid substrate remained confined from solvent interaction for the vast majority of the simulation frames.

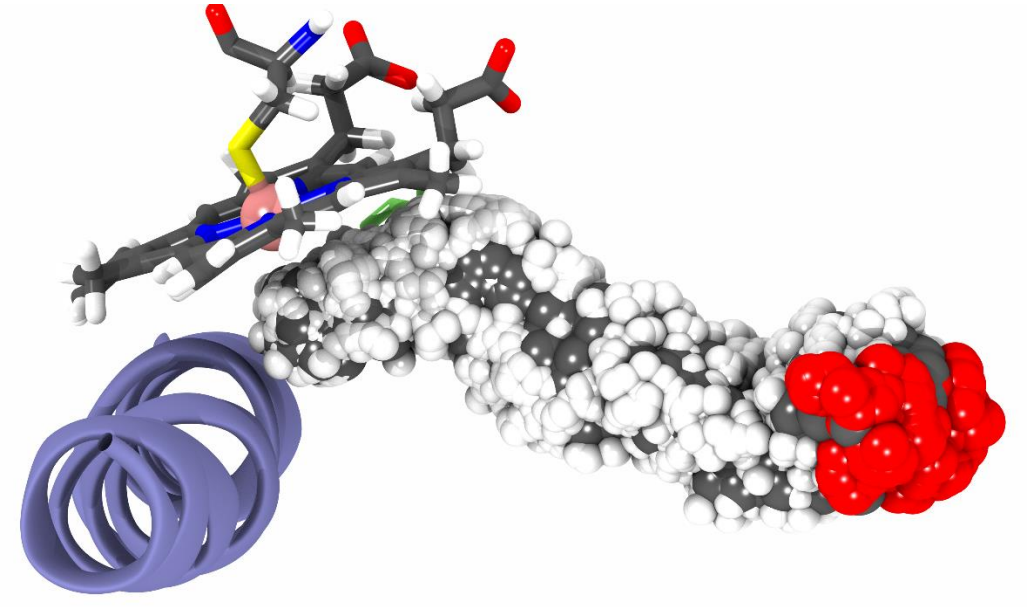

Figure 77. Clustering of arachidonic acid binding conformations for the L137Q Cyp4f13 homology model during production simulation. The trajectories are drawn in strides of 100 for the 10,000 simulation frames. Ball-and-stick representations were added for the substrate and the mutated glutamine residue, in addition to licorice representations for the amino acids involved in substrate binding.

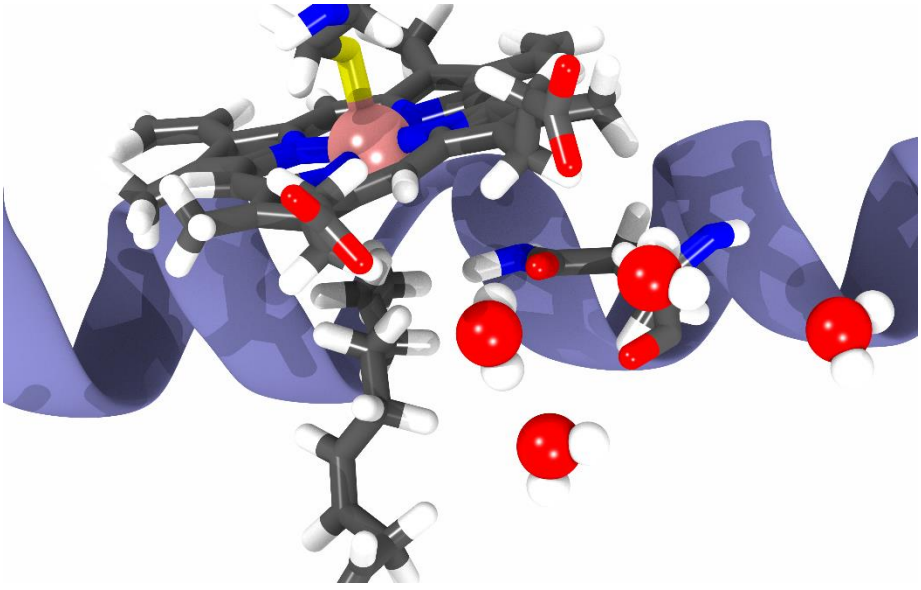

Figure 78. Structural view of the L137Q Cyp4f13 homology model active site. The mutated glutamine residue is shown with a licorice representation along with nearby vdW water molecules (within 4 Å) showcasing several hydrogen-bonding interactions.

Only a relatively small number of water molecules entered the catalytic site, as observed in Figure 78, appearing to interact with the glutamine or heme propionate groups, however, the majority of the water interactions were observed around the carboxylate binding domain of the protein. The 100 ns of simulation indicated a significant amount of desolvation for the substrate. The spatial region occupied by the mutation appears to have an influence on the access to the heme catalytic site, as only the substrate and a few errant solvent water molecules show ability to enter the vicinity of where the oxidant species would be located.

Metabolite Product Insights from Simulation Studies

Managing and analyzing the large amount of data in these trajectory files comes with its own set of challenges as it comes time to present them in a comprehensible manner. Analyzing trajectories by standard parameters like RMSD from backbone nuclear positions are not very useful in determining product formation or much behind the biomolecular interactions in a system. Some structural insight can be gained by analyzing displacements of the protein backbone, however, are noisy and prone to error propagation.

Analyzing the results of each individual mutation and their impacts on the metabolite production implied by the simulation would be performed by a similar method of analysis as the previous chapter. Using bond distances and the angle formed by a theoretical oxygen atom attached to the heme-iron as is observed in the ultimate oxidant state of the P450 catalytic cycle and aliphatic carbon-hydrogen bonds on the substrate, correlations can be made on the different mutations. These correlations will come with

their own sets of considerations in the sampling indicated by the 100 ns MD simulations of the mutant Cyp4f13 structural ensembles.

Unfortunately, in the case of G327E mutant, not much is to be said about predicted metabolite formation. Within the timeframe of the production simulation, water molecules completely invaded the active cavity and the arachidonic acid was beginning to fully egress from the site. This made it difficult to gain predicted oxidation potential as most of the trajectory frames failed to meet neither the tight nor loose established distance criteria. This was disappointing, however in the scope of this thesis, these result appear to agree with the null hypothesis that the presence of a covalently bound heme cofactor through the glutamate residue was the not the primary factor in ω-hydroxylation of arachidonic acid by murine Cyp4f isoforms. There would be need to be a deeper explanation for the evolutionary advantages to the lacking of the covalent heme linkage for the Cyp4f13 protein, and should be studied further.

However, what can be said is that through introducing the hydrogen bond accepting residue into the native Cyp4f13 active site, a significant amount of solvation facilitated the clearance of the fatty acid substrate from the catalytic site. This disturbance in the protein cavity architecture might be a factor in the mutation event that resulted in glycine being the residue selected for in the evolution of the murine Cyp4f13 system. This also implies that there might be additional underlying factors that influence the site of oxidation of the arachidonic acid by Cyp4f isoforms.

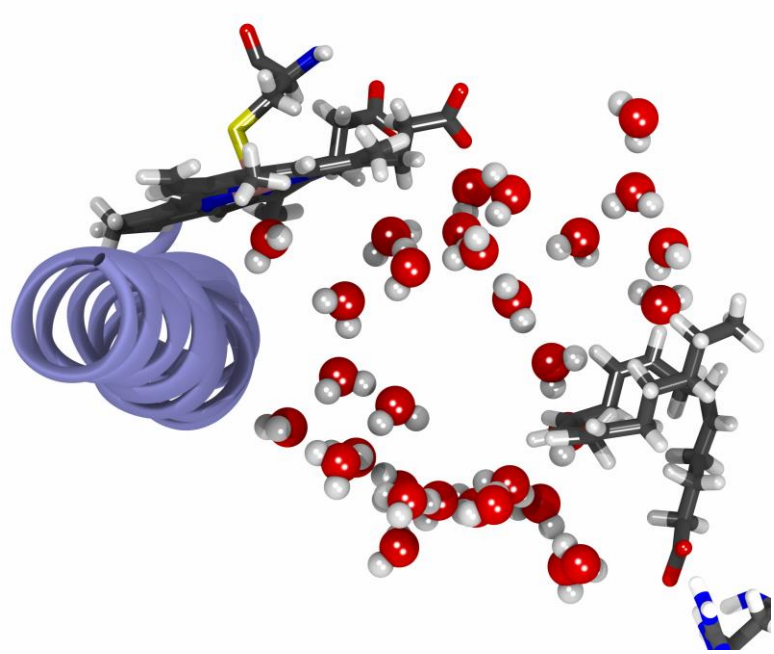

Figure 79. Structural view of the bolus of TIP3P water molecules filling the active cavity of the mutant G327E Cyp4f13 homology model. Water molecules within a region of 8 angstroms from the core of the protein active site are represented by vdW spheres.

While disappointing, it is worth mentioning that this result indicates the need of more simulation time, perhaps with different initial velocities through a different pseudo-randomly generated seed for the NAMD integrator. A more favorable steric environment was envisioned by this mutation, as it should have resulted in a more potent orientation for substrate to facilitate ω-hydroxylation by the mutant Cyp4f13 enzyme.

For the R237L mutant, the terminal carbon unit was indeed the closest to the heme-iron than any other site on the fatty acid chain, however, the presence of several water molecules coordinating to the active site prevented the substrate from progressing deeper into the cavity. However, this is observed in other P450 systems, as the substrate binding tightly leads to a series of structural changes that result in the displacement of water, not full desolvation of the catalytic site.

Another consequence from this mutation was now that the fatty acid carboxylate headgroup had migrated over to arginine 100, it had a different conformation within the active cavity, and due to the steric influence of leucine 395 and leucine 396, the polyunsaturated fatty acid wrapped around the nonpolar sidechains, limiting its ability to reach the catalytic site of the heme-iron.

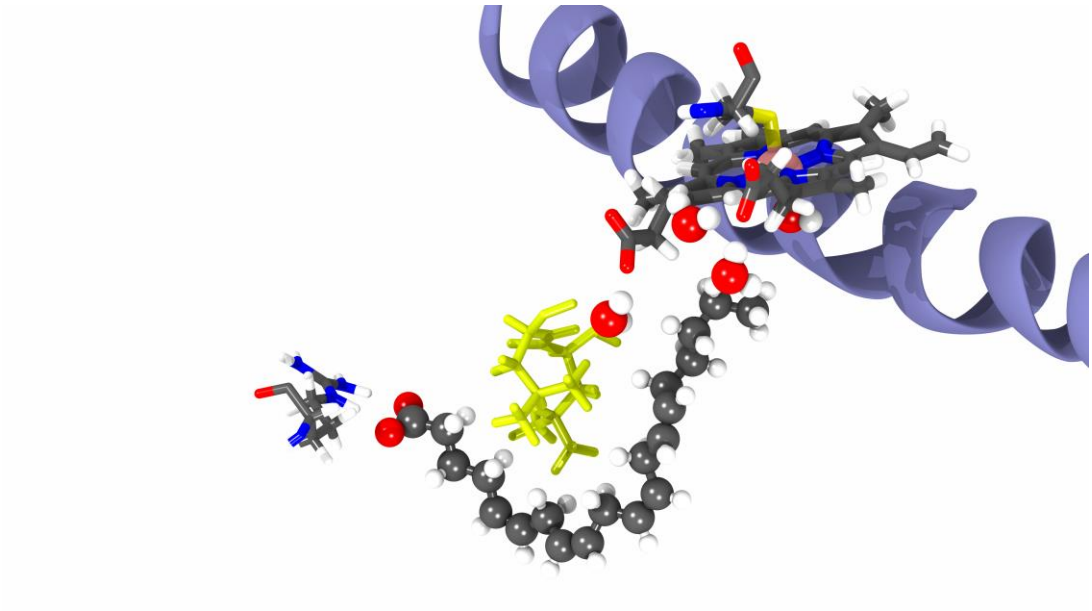

Figure 80. Structural view of the R237L mutant Cyp4f13 catalytic site with leucine representations emphasize their impact on arachidonic acid binding. Leucine 395 and 396 are shown in yellow licorice, in addition to the Arg100 residue binding the carboxylate of the arachidonate substrate.

While these interactions are normally present in the native protein, they become intrusive in this mutant as the carboxylate head group must reach around them to interact with the arginine 100 within the interior of the protein, as seen in Figure 80. This limits the amount of exposure for the terminal carbon unit has to the oxidant state of the P450 and could impact the product formation in a detrimental way as well. While binding may be enhanced in an overall sense, it is possible that the rate of successful product conversion and egression would be negatively affected by this mutation.

For the L137Q mutant, the simulation provided for interesting results in terms of proper substrate orientation for ω-hydroxylation of the arachidonic acid. The simulation frames were aligned, pictured in Figure 81, and showed a relative distance between the heme-iron and the hydrogen atoms of the terminal carbon fluctuating around the optimal distance (>2.8 angstroms) to where the electrophilic oxyferryl species would have its oxygen located. Predicted regiospecificity is shown in Figure 82 and indicates enhanced ω, ω-1, and ω-2 hydroxylation, with negligible frames of epoxide formation present. With both the terminal and sub-terminal site so readily available, it seems likely that the thermodynamic product of secondary C-H bond oxidation would be preferred by the system.

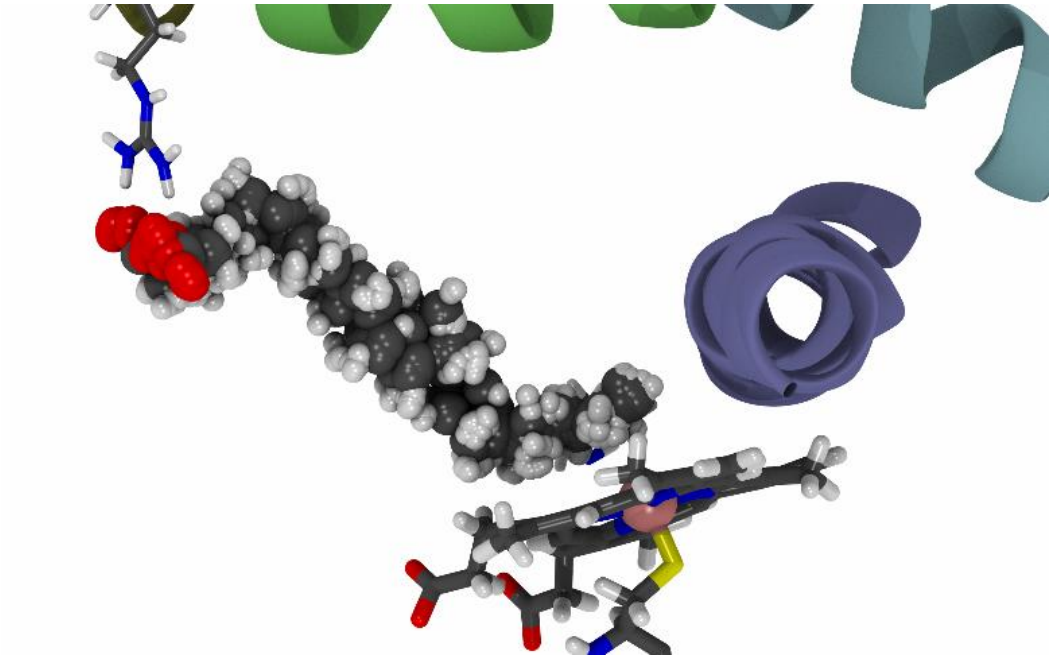

Figure 81. Overlay cluster of arachidonic acid binding conformations for the L137Q Cyp4f13 homology model during the final 100-ns of NVT simulation. The frames are drawn in strides of 100 for the 5000 simulation frames. Ball-and-stick representations were added for the substrate and the mutated glutamine residue, in addition to licorice representations for the amino acids involved in substrate binding.

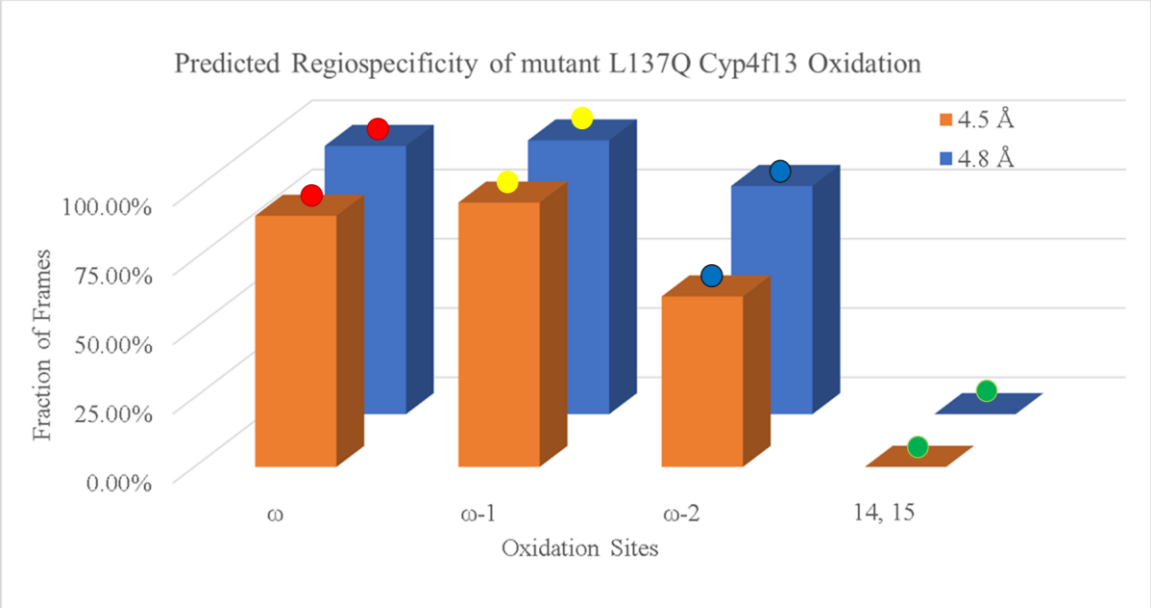

Figure 82. Predicted regiospecificity for the L137Q mutant Cyp4f13 homology model. The substrate consistently remained exposed to the heme in the same vicinity throughout the entire simulation, as evidenced by the relatively dominant amounts of terminal or subterminal oxidation.

While initially theorized to produce the most amount of perturbations to orientations in substrate binding to catalyze ω-hydroxylation, this mutant proved to be the least disruptive. The L137Q mutation introduced a large amount of hydrogen bonding through the amide of its mutated sidechain, however its usage is questionable when comparing the modified residue to its predecessor in the native protein.

Leucine has three heavy atoms (non-hydrogen) along its residue sidechain extending from the alpha carbon of its amino acid backbone. This makes glutamine a dissimilar residue to use as it has four heavy atoms along its sidechain. Asparagine, a related amino acid structure with a three heavy atom sidechain as well, would make for a better choice in mutation. The selection of the glutamine over the more appropriate asparagine would have an impact in volume taken up by the residue by a small but prominent amount. The volume added by the additional methylene side chain would have

a different steric environment in the protein's native cavity. Water channels could be disturbed in a way that the asparagine might have impacted less so.

This residue was chosen based off previous studies on mutant P450s. In the paper by Richards et al., the team of researchers used a library of $P450_{BM-3}$ mutants. Included in almost every construct was the L188Q mutation shown to have a significant effect on the ability of P450 to bind its non-polar drug substrate noscapine. This might have also been an artifact of the generation of cloned mutants, as mutation was made easier through a L-to-Q codon change. This appears to be the case as the codon switch between the two amino acids is one base pair (CUA to CAA). If the researchers wanted to change the mutation to the more appropriate-length asparagine, they would need to perform at least 2-3 base pairs in order to mutate the sidechain from leucine to asparagine.

As computational mutation does not share this experimental hardship, potential future work should consist of mutation of the leucine to asparagine and perform a similar simulation to observe its possible effects on the structural formation of the Cyp4f13-arachidonic acid complex.

Summary

In this work, we demonstrate that the residues within the active site of the Cyp4f13 homology model have important interactions that control substrate binding and metabolic product formation. The homology model simulations revealed that Cyp4f13 interacts with the arachidonate substrate though several side chains, including the Arg237 and Leu137 residues. Additionally, the lack of a covalently bound heme through an ester bond formed by a proximal glutamate residue and a hydroxylated heme methyl

substituent at the Gly327 position, has a role in the site for oxidation of the P450s fatty acid substrate. Importantly, these interactions were preserved in the previous set of MD simulations when different conformations and substrates were introduced.

In these series of experiments, the effects of mutating the putative interacting residues Gly327, Arg237, and Leu137 were studied. These mutations significantly altered the interactions of the active cavity whether through introducing hydrophilic interactions in the form of solvent water presence, or through enhancing hydrophobic forces, leading to better desolvation of the active site. To theoretically corroborate these findings, longer time frame simulation would be necessary to gain a large amount of sample size in order to draw more reliable conclusions. Notably, the G327E mutation with the parameterized novel heme linkage had an interesting effect, as it was theorized that its steric influence would be the primary mode of action of altering substrate binding and oxidation regiochemistry of the arachidonic acid substrate. The G327E constructs had a substantial increase of polarity of the active site as solvent TIP3P water models filled the cavity volume in the 100 ns time frame of the constant volume MD simulation. The R237L construct resulted in a change of the binding mode of the fatty acid substrate, as the carboxylate head group of the arachidonate migrated to Arg100, preserving throughout the remainder of the simulation. The terminal carbon remained the nearest site for oxidation, however mostly remained out of theoretical reach for the P450 radical rebound catalytic mechanism to possibly occur (>4.8 angstroms). Lastly, the L137Q construct resulted in an enhanced desolvation of the active pocket, indicating a higher amount of binding of the substrate, as solvent water molecules interacted only with the additional hydrogen bond acceptors and donors of the amide moiety from Gln137. While all mutant

Cyp4f13 constructs demonstrated sufficient substrate binding of arachidonic acid, we observed significant differences in their predicted formation of product. Wild-type interactions were heavily altered by the introduction of the additional covalent heme linkage and the three site mutations: G327E, R237L, and L137Q. Thus, we theoretically demonstrate that the mutation of several important residues of the native Cyp4f13 influences the binding mode, binding specificity, active cavity solvation, and alters water channel formation. Additional work should incorporate longer time frame simulation, different initial thermodynamic conditions, in addition to further mutation of key residues within both the binding and catalytic sites of the murine Cyp4f13 enzyme.

CHAPTER VI

Conclusions and Future Work

In this thesis, all-atom molecular dynamics (MD) simulations have been used to investigate the binding of arachidonic acid to a constructed P450 structural ensemble for the murine Cyp4f13 enzyme. We employed use of homology modeling, parameterization, residue mutation and MD simulations to investigate the active site and different conformational states of prospective metabolites for the murine Cyp4f13 system. The homology model for Cyp4f13 was produced using I-TASSER service and was subjected to consequent MD refinement techniques to generate a model ready for molecular docking. Molecular docking was employed to determine conformations of the target substrate, arachidonic acid, for the Cyp4f13 homology model. A pair of conformations were selected and used to investigate the binding mode of the polyunsaturated fatty acid within the enzyme's active site. Additionally, the ω-hydroxylated metabolite of clinical significance, 20-HETE, was also modelled and put through simulation in order to observe its possible evacuation from the active site. Through individual mutation, the active site was probed for its role in substrate binding and oxidative regiochemistry with three specific mutations: G327E, R237L, and L137Q. For the G327E mutant, a covalent heme linkage, observed to have an important role in ω-hydroxylation of substrates, was parameterized in the CHARMM force field format, and patched into the system in order to observe its effects on binding and site of oxidation.

Template selection was made through sequence alignment and similarity comparison. Refinement consisted of method validation through scoring of threading methods, RMSD analysis of the protein backbone, protonation state prediction, prosthetic

group patching, energy minimization, annealing, and thorough equilibration to ensure proper embedding into simulated DOPC lipid membrane bilayer. Solvation and counter ions were added to the system in order to produce an electrostatic environment of the virtual protein in its chemical setting. Molecular Dynamics Simulations were calculated by the Nanoscale Molecular Dynamics (NAMD) software. Molecular docking was performed using the AutoDockTools automated docking tool package. Quantum chemical calculations, QM or otherwise, were performed using the Gaussian 09w program suite. Visualization was performed using Visual Molecular Dynamics (VMD) program package and the GaussView 5.0 interface. Trajectories were analyzed using the RMSD Trajectory tool and NAMDPlot plugins of the VMD package. Mutation, structural alignment, and residue selection was performed using the Mutator, Multiseq and Sequence Viewer addons for VMD.

Homology modeling of the Cyp4f13 system presented in this work appeared to be successful, as the initial resting state of microsomal P450s was produced in the substrate-free model and persisted throughout its constant volume production simulation. Proper docking of the arachidonic acid resulted in a small collection of conformations of the arachidonate ligand to the Cyp4f13 model, with only relevant states chosen to have significance to the ω-hydroxylated product formation of the eicosanoid substrate. Computational chemistry software was used to modify the arachidonic acid to produce the necessary coordinates for the simulation of the 20-HETE metabolite as well. Simulation of these docking conformations resulted in the identification of several amino acid residues that were observed to have important interactions with the eicosanoid substrate with the binding and catalytic sites of the Cyp4f13 homology model. Mutation

was introduced to the Cyp4f13 system as a method of probing the active site and resulted in different binding activity for the protein-substrate complex. The modifications to the polarity of the active pocket resulted in altered solvation of the inner cavity of the Cyp4f13 active site and perturbation of the native protein's tertiary structure. The interactions observed with the active site and solvent water molecules between the native protein and the mutants were remarkably different, as the hydrophobic interactions with the protein and substrate were disturbed by a pronounced presence of water, or lack thereof in the case of the L137Q mutant.

The simulation timescales and sampling required to obtain physiologically relevant binding activity for the Cyp4f13 homology models and mutant variants bound to the eicosanoid substrate serve as the limiting factor in the accuracy of these results. Conclusions were drawn from thermodynamic sampling of the dynamics presented by the wild-type and mutant Cyp4f13 structural ensembles. Relative distances, non-bonding interactions, and energy considerations were used to assess and characterize the Cyp4f13 active site. Simulations such as those presented in this thesis are always difficult to precisely reproduce due to the pseudo-random nature of initial velocity integration. Caution should be exercised with extracting properties and characteristics of molecular systems from trajectory calculations such as these. An original goal of this study was to achieve a method to predict the metabolite profiles based on the simulation frames, however, this became unrealistic based on the issue of computational time and of reproducibility of these calculations. It is hoped, however, that the simulations and outcomes presented in this thesis will provide the foundations for future work that could

be performed that would provide better insight into the dynamics of both murine and human CYP4F subfamily of P450 monooxygenases.

There is much to be done in terms of future work. The most pressing consideration that deserves study would be the simulation of the ultimate oxidant, Compound I, of the P450 catalytic cycle. The proper parameterization would make it an interesting model for use in investigating the potential fates of oxidized eicosanoid metabolites by the CYP4F enzymes. The ferric iron(III) species used in this study acts as a great model of the resting state, however, its electronic environment greatly changes with the reduction of the metal center by its redox partners, and subsequent spin-state change orchestrated by substrate binding. This is especially the case after dioxygen binding and successive protonation results in the extremely electrophilic oxyferryl species.

The introduction of the cytochrome P450 reductase FAD and FMN domains would also lead to an interesting avenue of research. The addition of the two structures and their interactions with the heme domain should result in a change of conformational state for the P450, as is observed to be the case in several microsomal P450s. This might alter substrate binding with its protein-protein interaction that is not observed in the solitary globular heme domain of the P450.

A known CYP4F inhibitor, HET0016, has been observed to have an observable effect on murine Cyp4f13, making it one of the only known compounds confirmed to inhibit the enzyme.[145] Modeling and docking of the inhibitor into a homology model of the murine Cyp4f could prove useful in drug design, as these proteins are theorized to play a role in inflammation, cancer, and could potentially be used as a biomarker

Understanding the interactions of small molecules and the active sites of their physiological targets is imperative in the engineering and design of novel therapeutics.[87]

As was previously mentioned, further work on mutant Cyp4f13 models would also be interesting to perform, as the work presented in this thesis indicated a drastic change in substrate binding activity. Possible mutation of leucine 137 to the more appropriately length asparagine residue could produce a less dramatic alteration to the active site while maintaining the same change in polarity that the L137Q mutation introduced.

More simulation time might be necessary to investigate the direct impact in the addition of the heme covalent linkage of the G327E mutant. Longer time frame simulations, such as in the microsecond range, could result in a desolvation of the active site and might indicate the possible formation of an ω-hydroxylated product of the eicosanoid substrate by the Cyp4f13 isoform.

The individual roles of murine Cyp4f proteins are currently still being investigated, as their roles in inflammation and cancer are increasingly suggested by mounting evidence in experimental methods.[90] What remains to be seen are the computational insights that could be gained using all-atom dynamic simulation to understand what clinical results are found using subcloning and recombinant methods. It is hoped that such developments may well lead to the more common use of these computational methods, such as MD, to aid in the interpretation of metabolomic data by researchers and as such, provide for a better understanding of the biochemical cosmos for the fields of biochemistry, and medicinal chemistry.

REFERENCES

(1) Guengerich, F. P. Cytochromes P450, Drugs, and Diseases. *Mol. Interv.* **2003**, *3* (4), 194–204.

(2) Guengerich, F. P. Cytochrome P450s and Other Enzymes in Drug Metabolism and Toxicity. *AAPS J.* **2006**, *8* (1), E101–E111.

(3) Guengerich, F. P.; Wu, Z. L.; Bartleson, C. J. Function of Human Cytochrome P450s: Characterization of the Orphans. *Biochem. Biophys. Res. Commun.* **2005**, *338* (1), 465–469.

(4) Manikandan, P.; Nagini, S. Cytochrome P450 Structure, Function and Clinical Significance: A Review. *Curr. Drug Targets* **2017**, *19* (1), 38–54.

(5) Nebert, D. W.; Russell, D. W. Clinical Importance of the Cytochromes P450. *Lancet* **2002**, *360* (9340), 1155–1162.

(6) Lewis, D. F. *Cytochromes P450. Structure, Function, and Mechanism*; Taylor & Francis, 1996; Vol. 40.

(7) Stryer, L.; Berg, J. M.; Tymoczko, J. L. *Biochemistry*, Ninth ed.; W. H. Freeman, 2019.

(8) Ortiz de Montellano, P. R. *Cytochrome P450: Structure, Mechanism, and Biochemistry, Fourth Edition*; 2015.

(9) Omura, T.; Sato, R. A New Cytochrome in Liver Microsomes. *J. Biol. Chem.* **1962**, *237* (4), 1375–1376.

(10) Estabrook, R. W. A Passion for P450s (Remembrances of the Early History of Research on Cytochrome P450). *Drug Metab. Dispos.* **2003**, *31* (12), 1461–1473.

(11) Wang, J.-F.; Zhang, C.-C.; Chou, K.-C.; Wei, D.-Q. Structure of Cytochrome

P450s and Personalized Drug. *Curr. Med. Chem.* **2008**, *16* (2), 232–244.

(12) Nelson, D. R. Comparison of P450s from Human and Fugu: 420 Million Years of Vertebrate P450 Evolution. *Arch. Biochem. Biophys.* **2003**, *409* (1), 18–24.

(13) Nelson, D. R. Cytochrome P450 Nomenclature, 2004. In *Methods in Molecular Biology (Clifton, N.J.)*; Humana Press: New Jersey, 2006; Vol. 320, pp 1–10.

(14) Nelson, D. R.; Koymans, L.; Kamataki, T.; Stegeman, J. J.; Feyereisen, R.; Waxman, D. J.; Waterman, M. R.; Gotoh, O.; Coon, M. J.; Estabrook, R. W.; Gunsalus, I. C.; Nebert, D. W. P450 Superfamily: Update on New Sequences, Gene Mapping, Accession Numbers and Nomenclature. *Pharmacogenetics* **1996**, *6* (1), 1–42.

(15) Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I. Structure and Chemistry of Cytochrome P450. *Chem. Rev.* **2005**, *105* (6), 2253–2277.

(16) Mestres, J. Structure Conservation in Cytochromes P450. *Proteins Struct. Funct. Genet.* **2005**, *58* (3), 596–609.

(17) Peterson, J. A.; Graham, S. E. A Close Family Resemblance: The Importance of Structure in Understanding Cytochromes P450. *Structure* **1998**, *6* (9), 1079–1085.

(18) Domanski, T.; Halpert, J. Analysis of Mammalian Cytochrome P450 Structure and Function by Site-Directed Mutagenesis. *Curr. Drug Metab.* **2005**, *2* (2), 117–137.

(19) Whitehouse, C. J. C.; Bell, S. G.; Wong, L. L. P450 BM3 (CYP102A1): Connecting the Dots. *Chem. Soc. Rev.* **2012**, *41* (3), 1218–1260.

(20) Guengerich, F. P. Mechanisms of Cytochrome P450-Catalyzed Oxidations. *ACS Catal.* **2018**, *8* (12), 10964–10976.

(21) 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.

(22) 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.

(23) Noble, M. A.; Miles, C. S.; Chapman, S. K.; Lysek, D. A.; Mackay, A. C.; Reid, G. A.; Hanzlik, R. P.; Munro, A. W. Roles of Key Active-Site Residues in Flavocytochrome P450 BM3. *Biochem. J.* **1999**, *339* (2), 371–379.

(24) Munro, A. W.; Leys, D. G.; McLean, K. J.; Marshall, K. R.; Ost, T. W. B.; Daff, S.; Miles, C. S.; Chapman, S. K.; Lysek, D. A.; Moser, C. C.; Page, C. C.; Dutton, P. L. P450 BM3: The Very Model of a Modern Flavocytochrome. *Trends Biochem. Sci.* **2002**, *27* (5), 250–257.

(25) Ogliaro, F.; Harris, N.; Cohen, S.; Filatov, M.; De Visser, S. P.; Shaik, S. A Model "Rebound" Mechanism of Hydroxylation by Cytochrome P450: Stepwise and Effectively Concerted Pathways, and Their Reactivity Patterns. *J. Am. Chem. Soc.* **2000**, *122* (37), 8977–8989.

(26) Fishelovitch, D.; Hazan, C.; Hirao, H.; Wolfson, H. J.; Nussinov, R.; Shaik, S. QM/MM Study of the Active Species of the Human Cytochrome P450 3A4, and the Influence Thereof of the Multiple Substrate Binding. *J. Phys. Chem. B* **2007**, *111* (49), 13822–13832.

(27) Hrycay, E. G.; Bandiera, S. M. Involvement of Cytochrome P450 in Reactive Oxygen Species Formation and Cancer. In *Advances in Pharmacology*; 2015; Vol. 74, pp 35–84.

(28) Pochapsky, T. C.; Kazanis, S.; Dang, M. Conformational Plasticity and

Structure/Function Relationships in Cytochromes P450. *Antioxidants Redox Signal.* **2010**, *13* (8), 1273–1296.

(29) Narhi, L. O.; Fulco, A. J. Characterization of a Catalytically Self-Sufficient 119,000-Dalton Cytochrome P-450 Monooxygenase Induced by Barbiturates in Bacillus Megaterium. *J. Biol. Chem.* **1986**, *261* (16), 7160–7169.

(30) Warman, A. J.; Roitel, O.; Neeli, R.; Girvan, H. M.; Seward, H. E.; Murray, S. A.; McLean, K. J.; Joyce, M. G.; Toogood, H.; Holt, R. A.; Leys, D.; Scrutton, N. S.; Munro, A. W. Flavocytochrome P450 BM3: An Update on Structure and Mechanism of a Biotechnologically Important Enzyme. *Biochem. Soc. Trans.* **2005**, *33* (4), 747–753.

(31) Haines, D. C.; Chen, B.; Tomchick, D. R.; Bondlela, M.; Hegde, A.; Machius, M.; Peterson, J. A. Crystal Structure of Inhibitor-Bound P450BM-3 Reveals Open Conformation of Substrate Access Channel. *Biochemistry* **2008**, *47* (12), 3662–3670.

(32) Mouri, T.; Shimizu, T.; Kamiya, N.; Goto, M.; Ichinose, H. Design of a Cytochrome P450BM3 Reaction System Linked by Two-Step Cofactor Regeneration Catalyzed by a Soluble Transhydrogenase and Glycerol Dehydrogenase. *Biotechnol. Prog.* **2009**, *25* (5), 1372–1378.

(33) Wong, L. L. Cytochrome P450 Monooxygenases. *Curr. Opin. Chem. Biol.* **1998**, *2* (2), 263–268.

(34) Guengerich, F. P. Common and Uncommon Cytochrome P450 Reactions Related to Metabolism and Chemical Toxicity. *Chem. Res. Toxicol.* **2001**, *14* (6), 611–650.

(35) Chefson, A.; Auclair, K. Progress towards the Easier Use of P450 Enzymes. *Mol.*

Biosyst. **2006**, *2* (10), 462–469.

(36) Sevrioukova, I. F.; Li, H.; Zhang, H.; Peterson, J. A.; Poulos, T. L. Structure of a Cytochrome P450-Redox Partner Electron-Transfer Complex. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (5), 1863–1868.

(37) Daff, S. N.; Chapman, S. K.; Turner, K. L.; Holt, R. A.; Govindaraj, S.; Poulos, T. L.; Munro, A. W. Redox Control of the Catalytic Cycle of Flavocytochrome P-450 BM3. *Biochemistry* **1997**, *36* (45), 13816–13823.

(38) McLean, K. J.; Girvan, H. M.; Munro, A. W. Cytochrome P450/Redox Partner Fusion Enzymes: Biotechnological and Toxicological Prospects. *Expert Opin. Drug Metab. Toxicol.* **2007**, *3* (6), 847–863.

(39) Di Nardo, G.; Gilardi, G. Optimization of the Bacterial Cytochrome P450 BM3 System for the Production of Human Drug Metabolites. *Int. J. Mol. Sci.* **2012**, *13* (12), 15901–15924.

(40) Arnold, F. H. Directed Evolution: Bringing New Chemistry to Life. *Angew. Chemie - Int. Ed.* **2018**, *57* (16), 4143–4148.

(41) Seifert, A.; Pleiss, J. Identification of Selectivity Determinants in CYP Monooxygenases by Modelling and Systematic Analysis of Sequence and Structure. *Curr. Drug Metab.* **2012**, *13* (2), 197–202.

(42) Seifert, A.; Pleiss, J. Identification of Selectivity-Determining Residues in Cytochrome P450 Monooxygenases: A Systematic Analysis of the Substrate Recognition Site 5. *Proteins Struct. Funct. Bioinforma.* **2009**, *74* (4), 1028–1035.

(43) Weber, E.; Seifert, A.; Antonovici, M.; Geinitz, C.; Pleiss, J.; Urlacher, V. B. Screening of a Minimal Enriched P450 BM3 Mutant Library for Hydroxylation of

Cyclic and Acyclic Alkanes. *Chem. Commun.* **2011**, *47* (3), 944–946.

(44) Lonsdale, R.; Rouse, S. L.; Sansom, M. S. P.; Mulholland, A. J. A Multiscale Approach to Modelling Drug Metabolism by Membrane-Bound Cytochrome P450 Enzymes. *PLoS Comput. Biol.* **2014**, *10* (7).

(45) Seifert, A.; Tatzel, S.; Schmid, R. B.; Pleiss, J. Multiple Molecular Dynamics Simulations of Human P450 Monooxygenase CYP2C9: The Molecular Basis of Substrate Binding and Regioselectivity toward Warfarin. *Proteins Struct. Funct. Genet.* **2006**, *64* (1), 147–155.

(46) Pleiss, J. Systematic Analysis of Large Enzyme Families: Identification of Specificity- and Selectivity-Determining Hotspots. *ChemCatChem* **2014**, *6* (4), 944–950.

(47) Navrátilová, V.; Paloncýová, M.; Berka, K.; Otyepka, M. Effect of Lipid Charge on Membrane Immersion of Cytochrome P450 3A4. *J. Phys. Chem. B* **2016**, *120* (43), 11205–11213.

(48) Šrejber, M.; Navrátilová, V.; Paloncýová, M.; Bazgier, V.; Berka, K.; Anzenbacher, P.; Otyepka, M. Membrane-Attached Mammalian Cytochromes P450: An Overview of the Membrane's Effects on Structure, Drug Binding, and Interactions with Redox Partners. *J. Inorg. Biochem.* **2018**, *183* (December 2017), 117–136.

(49) Larson, J. R.; Coon, M. J.; Porter, T. D. Alcohol-Inducible Cytochrome P-450IIE1 Lacking the Hydrophobic NH2-Terminal Segment Retains Catalytic Activity and Is Membrane-Bound When Expressed in *Escherichia coli*. *J. Biol. Chem.* **1991**, *266* (12), 7321–7324.

(50) Cullin, C. Two Distinct Sequences Control the Targeting and Anchoring of the Mouse P450 1A1 into the Yeast Endoplasmic Reticulum Membrane. *Biochem. Biophys. Res. Commun.* **1992**, *184* (3), 1490–1495.

(51) Pernecky, S. J.; Larson, J. R.; Philpot, R. M.; Coon, M. J. Expression of Truncated Forms of Liver Microsomal P450 Cytochromes 2B4 and 2E1 in *Escherichia coli*: Influence of NH2-Terminal Region on Localization in Cytosol and Membranes. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (7), 2651–2655.

(52) Gillam, E. M. J.; Baba, T.; Kim, B. R.; Ohmori, S.; Guengerich, F. P. Expression of Modified Human Cytochrome P450 3A4 in *Escherichia coli* and Purification and Reconstitution of the Enzyme. *Arch. Biochem. Biophys.* **1993**, *305* (1), 123–131.

(53) Sagara, Y.; Barnes, H. J.; Waterman, M. R. Expression in *Escherichia coli* of Functional Cytochrome P450c17 Lacking Its Hydrophobic Amino-Terminal Signal Anchor. *Arch. Biochem. Biophys.* **1993**, *304* (1), 272–278.

(54) Shank-Retzlaff, M. L.; Raner, G. M.; Coon, M. J.; Sligar, S. G. Membrane Topology of Cytochrome P450 2B4 in Langmuir-Blodgett Monolayers. *Arch. Biochem. Biophys.* **1998**, *359* (1), 82–88.

(55) Kawato, S.; Gut, J.; Cherry, R. J.; Winterhalter, K. H.; Richter, C. Rotation of Cytochrome P-450. *J. Biol. Chem.* **1982**, *257* (12), 7023–7029.

(56) Etter, H. U.; Richter, C.; Ohta, Y.; Winterhalter, K. H.; Sasabe, H.; Kawato, S. Rotation and Interaction with Epoxide Hydrase of Cytochrome P-450 in Proteoliposomes. *J. Biol. Chem.* **1991**, *266* (28), 18600–18605.

(57) Scott, E. E.; Wolf, C. R.; Otyepka, M.; Humphreys, S. C.; Reed, J. R.; Henderson,

C. J.; Mclaughlin, L. A.; Paloncýová, M.; Navrátilová, V.; Berka, K.; Anzenbacher, P.; Dahal, U. P.; Barnaba, C.; Brozik, J. A.; Jones, J. P.; Estrada, D. F.; Laurence, J. S.; Park, J. W.; Backes, W. L. The Role of Protein-Protein and Protein-Membrane Interactions on P450 Function. *Drug Metab. Dispos.* **2016**, *44* (4), 576–590.

(58) Guengerich, F. P. Cytochrome P450 and Chemical Toxicology. *Chem. Res. Toxicol.* **2008**, *21* (1), 70–83.

(59) Meunier, B.; de Visser, S. P.; Shaik, S. Mechanism of Oxidation Reactions Catalyzed by Cytochrome P450 Enzymes. *Chem. Rev.* **2004**, *104* (9), 3947–3980.

(60) Nebert, D. W.; Wikvall, K.; Miller, W. L. Human Cytochromes P450 in Health and Disease. *Philos. Trans. R. Soc. B Biol. Sci.* **2013**, *368* (1612).

(61) Häggström, M.; Richfield, D. Diagram of the Pathways of Human Steroidogenesis. *WikiJournal Med.* **2014**, *1* (1).

(62) Auchus, R. J.; Miller, W. L. P450 Enzymes in Steroid Processing. In *Cytochrome P450: Structure, Mechanism, and Biochemistry, Fourth Edition*; Springer International Publishing: Cham, 2015; pp 851–879.

(63) Zhang, Y. Y.; Yang, L. Interactions between Human Cytochrome P450 Enzymes and Steroids: Physiological and Pharmacological Implications. *Expert Opin. Drug Metab. Toxicol.* **2009**, *5* (6), 621–629.

(64) Peter Guengerich, F.; Wilkey, C. J.; Glass, S. M.; Reddish, M. J. Conformational Selection Dominates Binding of Steroids to Human Cytochrome P450 17A1. *J. Biol. Chem.* **2019**, *294* (26), 10028–10041.

(65) Navrátilová, V.; Paloncýová, M.; Kajšová, M.; Berka, K.; Otyepka, M. Effect of

Cholesterol on the Structure of Membrane-Attached Cytochrome P450 3A4. *J. Chem. Inf. Model.* **2015**, *55* (3), 628–635.

(66) Edson, K.; Rettie, A. CYP4 Enzymes As Potential Drug Targets: Focus on Enzyme Multiplicity, Inducers and Inhibitors, and Therapeutic Modulation of 20-Hydroxyeicosatetraenoic Acid (20-HETE) Synthase and Fatty Acid ω-Hydroxylase Activities. *Curr. Top. Med. Chem.* **2013**, *13* (12), 1429–1440.

(67) Cappellani, D.; Brancatella, A.; Kaufmann, M.; Minucci, A.; Vignali, E.; Canale, D.; De Paolis, E.; Capoluongo, E.; Cetani, F.; Jones, G.; Marcocci, C. Hereditary Hypercalcemia Caused by a Homozygous Pathogenic Variant in the CYP24A1 Gene: A Case Report and Review of the Literature. *Case Rep. Endocrinol.* **2019**, *2019*, 1–7.

(68) Nakano, M.; Kelly, E. J.; Wiek, C.; Hanenberg, H.; Rettie, A. E. CYP4V2 in Bietti's Crystalline Dystrophy: Ocular Localization, Metabolism of ω-3-Polyunsaturated Fatty Acids, and Functional Deficit of the p.H331p Variant. *Mol. Pharmacol.* **2012**, *82* (4), 679–686.

(69) 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.

(70) Kelly, E. J.; Nakano, M.; Rohatgi, P.; Yarov-Yarovoy, V.; Rettie, A. E. Finding Homes for Orphan Cytochrome P450s: CYP4V2 and CYP4F22 in Disease States. *Mol. Interv.* **2011**, *11* (2), 124–132.

(71) Hanna, V. S.; Hafez, E. A. A. Synopsis of Arachidonic Acid Metabolism: A Review. *J. Adv. Res.* **2018**, *11*, 23–32.

(72) Luo, P.; Wang, M. H. Eicosanoids, β-Cell Function, and Diabetes. *Prostaglandins Other Lipid Mediat.* **2011**, *95* (1–4), 1–10.

(73) Dennis, E. A.; Norris, P. C. Eicosanoid Storm in Infection and Inflammation. *Nat. Rev. Immunol.* **2015**, *15* (8), 511–523.

(74) Powell, P. K.; Wolf, I.; Jin, R.; Lasker, J. M. Metabolism of Arachidonic Acid to 20-Hydroxy-5,8,11,14-Eicosatetraenoic Acid by P450 Enzymes in Human Liver: Involvement of CYP4F2 and CYP4A11. *J. Pharmacol. Exp. Ther.* **1998**, *285* (3), 1327–1336.

(75) Spector, A. A. Arachidonic Acid Cytochrome P450 Epoxygenase Pathway. *J. Lipid Res.* **2009**, *50* (SUPPL.), S52-6.

(76) Zeldin, D. C. Epoxygenase Pathways of Arachidonic Acid Metabolism. *J. Biol. Chem.* **2001**, *276* (39), 36059–36062.

(77) Laniado-Schwartzman, M.; Abraham, N. G. The Renal Cytochrome P-450 Arachidonic Acid System. *Pediatr. Nephrol.* **1992**, *6* (5), 490–498.

(78) Steuck, M.; Hellhake, S.; Schebb, N. H. Food Polyphenol Apigenin Inhibits the Cytochrome P450 Monoxygenase Branch of the Arachidonic Acid Cascade. *J. Agric. Food Chem.* **2016**, *64* (47), 8973–8976.

(79) Kikuta, Y.; Kusunose, E.; Kusunose, M. Prostaglandin and Leukotriene ω-Hydroxylases. *Prostaglandins Other Lipid Mediat.* **2002**, *68–69*, 345–362.

(80) Quiroga, I.; Scior, T. Structure-Function Analysis of the Cytochromes P450, Responsible for Phenprocoumon Metabolism. *J. Mex. Chem. Soc.* **2017**, *61* (4),

349–360.

(81) Corcos, L.; Le Lucas, D.; Le Jossic-Corcos, C.; Dréano, Y.; Simon, B.; Plée-Gautier, E.; Amet, Y.; Salaüna, J. P. Human Cytochrome P450 4F3: Structure, Functions, and Prospects. *Drug Metabol. Drug Interact.* **2012**, *27* (2), 63–71.

(82) Fan, F.; Muroya, Y.; Roman, R. J. Cytochrome P450 Eicosanoids in Hypertension and Renal Disease. *Curr. Opin. Nephrol. Hypertens.* **2015**, *24* (1), 37–46.

(83) Alexanian, A.; Miller, B.; Roman, R. J.; Sorokin, A. 20-HETE-Producing Enzymes Are up-Regulated in Human Cancers. *Cancer Genom. Proteom.* **2012**, *9* (4), 163–169.

(84) Pearson, T.; Warren, A. Y.; Barrett, D. A.; Khan, R. N. Detection of EETs and HETE-Generating Cytochrome P-450 Enzymes and the Effects of Their Metabolites on Myometrial and Vascular Function. *Am. J. Physiol. - Endocrinol. Metab.* **2009**, *297* (3).

(85) Imig, J. D.; Simpkins, A. N.; Renic, M.; Harder, D. R. Cytochrome P450 Eicosanoids and Cerebral Vascular Function. *Expert Rev. Mol. Med.* **2011**, *13* (1), e7.

(86) 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.

(87) Alexanian, A.; Sorokin, A. Targeting 20-HETE Producing Enzymes in Cancer - Rationale, Pharmacology, and Clinical Potential. *Onco. Targets. Ther.* **2013**, *6*, 243–255.

(88) Fer, M.; Corcos, L.; Dréano, Y.; Plée-Gautier, E.; Salaün, J. P.; Berthou, F.; Amet,

Y. Cytochromes P450 from Family 4 Are the Main Omega Hydroxylating Enzymes in Humans: CYP4F3B Is the Prominent Player in PUFA Metabolism. *J. Lipid Res.* **2008**, *49* (11), 2379–2389.

(89) Imaoka, S. Biological Functions of Cytochrome P450s in the CYP4 Family. *Drug Metab. Pharmacokinet.* **1999**, *14* (2), 139–147.

(90) 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.

(91) Hardwick, J. P. Cytochrome P450 Omega Hydroxylase (CYP4) Function in Fatty Acid Metabolism and Metabolic Diseases. *Biochem. Pharmacol.* **2008**, *75* (12), 2263–2275.

(92) Adas, F.; Salaün, J. P.; Berthou, F.; Picart, D.; Simon, B.; Amet, Y. Requirement for ω and (ω-1)-Hydroxylations of Fatty Acids by Human Cytochromes P450 2E1 and 4A11. *J. Lipid Res.* **1999**, *40* (11), 1990–1997.

(93) De Visser, S. P.; Kumar, D.; Cohen, S.; Shacham, R.; Shaik, S. A Predictive Pattern of Computed Barriers for C-H Hydroxylation by Compound I of Cytochrome P450. *J. Am. Chem. Soc.* **2004**, *126* (27), 8362–8363.

(94) Fisher, M. B.; Zheng, Y. M.; Rettie, A. E. Positional Specificity of Rabbit CYP4B1 for ω-Hydroxylation of Short-Medium Chain Fatty Acids and Hydrocarbons. *Biochem. Biophys. Res. Commun.* **1998**, *248* (2), 352–355.

(95) Kroetz, D. L.; Xu, F. Regulation and Inhibition of Arachidonic Acid ω-Hydroxylases and 20-HETE Formation. *Annu. Rev. Pharmacol. Toxicol.* **2005**, *45* (1), 413–438.

(96) Lasker, J. M.; Chen, W. B.; Wolf, I.; Bloswick, B. P.; Wilson, P. D.; Powell, P. K.

Formation of 20-Hydroxyeicosatetraenoic Acid, a Vasoactive and Natriuretic Eicosanoid, in Human Kidney. Role of CYP4F2 and CYP4A11. *J. Biol. Chem.* **2000**, *275* (6), 4118–4126.

(97) Edson, K. Z.; Prasad, B.; Unadkat, J. D.; Suhara, Y.; Okano, T.; Peter Guengerich, F.; Rettie, A. E. Cytochrome P450-Dependent Catabolism of Vitamin K: ω-Hydroxylation Catalyzed by Human CYP4F2 and CYP4F11. *Biochemistry* **2013**, *52* (46), 8276–8285.

(98) McDonald, M. G.; Rieder, M. J.; Nakano, M.; Hsia, C. K.; Rettie, A. E. CYP4F2 Is a Vitamin K1 Oxidase: An Explanation for Altered Warfarin Dose in Carriers of the V433M Variant. *Mol. Pharmacol.* **2009**, *75* (6), 1337–1346.

(99) Parker, R. S.; Sontag, T. J.; Swanson, J. E.; McCormick, C. C. Discovery, Characterization, and Significance of the Cytochrome P450 ω-Hydroxylase Pathway of Vitamin E Catabolism. *Ann. N. Y. Acad. Sci.* **2004**, *1031* (1), 13–21.

(100) Hashizume, T.; Imaoka, S.; Mise, M.; Terauchi, Y.; Fujii, T.; Miyazaki, H.; Kamataki, T.; Funae, Y. Involvement of CYP2J2 and CYP4F12 in the Metabolism of Ebastine in Human Intestinal Microsomes. *J. Pharmacol. Exp. Ther.* **2002**, *300* (1), 298–304.

(101) Kovarik, J. M.; Dole, K.; Riviere, G. J.; Pommier, F.; Maton, S.; Jin, Y.; Lasseter, K. C.; Schmouder, R. L. Ketoconazole Increases Fingolimod Blood Levels in a Drug Interaction via CYP4F2 Inhibition. *J. Clin. Pharmacol.* **2009**, *49* (2), 212–218.

(102) Wang, M. Z.; Saulter, J. Y.; Usuki, E.; Cheung, Y. L.; Hall, M.; Bridges, A. S.; Loewen, G.; Parkinson, O. T.; Stephens, C. E.; Allen, J. L.; Zeldin, D. C.; Boykin,

D. W.; Tidwell, R. R.; Parkinson, A.; Paine, M. F.; Hall, J. E. CYP4F Enzymes Are the Major Enzymes in Human Liver Microsomes That Catalyze the O-Demethylation of the Antiparasitic Prodrug DB289 [2,5-Bis(4-Amidinophenyl)Furan-Bis-O-Methylamidoxime]. *Drug Metab. Dispos.* **2006**, *34* (12), 1985–1994.

(103) Imig, J. D.; Falck, J. R.; Inscho, E. W. Contribution of Cytochrome P450 Epoxygenase and Hydroxylase Pathways to Afferent Arteriolar Autoregulatory Responsiveness. *Br. J. Pharmacol.* **1999**, *127* (6), 1399–1405.

(104) Gainer, J. V.; Bellamine, A.; Dawson, E. P.; Womble, K. E.; Grant, S. W.; Wang, Y.; Cupples, L. A.; Guo, C. Y.; Demissie, S.; O'Donnell, C. J.; Brown, N. J.; Waterman, M. R.; Capdevila, J. H. Functional Variant of CYP4A11 20-Hydroxyeicosatetraenoic Acid Synthase Is Associated with Essential Hypertension. *Circulation* **2005**, *111* (1), 63–69.

(105) Escalante, B.; Falck, J. R.; Yadagiri, P.; Sun, L.; Laniado-Schwartzman, M. 19(S)-Hydroxyeicosatetraenoic Acid Is a Potent Stimulator of Renal Na+-K+-ATPase. *Biochem. Biophys. Res. Commun.* **1988**, *152* (3), 1269–1274.

(106) He, X.; Cryle, M. J.; De Voss, J. J.; Ortiz De Montellano, P. R. Calibration of the Channel That Determines the ω-Hydroxylation Regiospecificity of Cytochrome P4504A1: Catalytic Oxidation of 12-Halododecanoic Acids. *J. Biol. Chem.* **2005**, *280* (24), 22697–22705.

(107) Lewis, D. F. V.; Lake, B. G. Molecular Modelling of CYP4A Subfamily Members Based on Sequence Homology with CYP102. *Xenobiotica* **1999**, *29* (8), 763–781.

(108) Theken, K. N.; Deng, Y.; Alison Kannon, M.; Miller, T. M.; Poloyac, S. M.; Lee,

C. R. Activation of the Acute Inflammatory Response Alters Cytochrome P450 Expression and Eicosanoid Metabolism. *Drug Metab. Dispos.* **2011**, *39* (1), 22–29.

(109) Hsu, M. H.; Savas, Ü.; Griffin, K. J.; Johnson, E. F. Human Cytochrome P450 Family 4 Enzymes: Function, Genetic Variation and Regulation. *Drug Metab. Rev.* **2007**, *39* (2–3), 515–538.

(110) Nakano, M.; Kelly, E. J.; Rettie, A. E. Expression and Characterization of CYP4V2 as a Fatty Acid ω-Hydroxylase. *Drug Metab. Dispos.* **2009**, *37* (11), 2119–2122.

(111) Dhar, M.; Sepkovic, D. W.; Hirani, V.; Magnusson, R. P.; Lasker, J. M. Omega Oxidation of 3-Hydroxy Fatty Acids by the Human CYP4F Gene Subfamily Enzyme CYP4F11. *J. Lipid Res.* **2008**, *49* (3), 612–624.

(112) Hoch, U.; Ortiz De Montellano, P. R. Covalently Linked Heme in Cytochrome P4504A Fatty Acid Hydroxylases. *J. Biol. Chem.* **2001**, *276* (14), 11339–11346.

(113) LeBrun, L. A.; Xu, F.; Kroetz, D. L.; Ortiz de Montellano, P. R. Covalent Attachment of the Heme Prosthetic Group in the CYP4F Cytochrome P450 Family. *Biochemistry* **2002**, *41* (18), 5931–5937.

(114) Baer, B. R.; Schuman, J. T.; Campbell, A. P.; Cheesman, M. J.; Nakano, M.; Moguilevsky, N.; Kunze, K. L.; Rettie, A. E. Sites of Covalent Attachment of CYP4 Enzymes to Heme: Evidence for Microheterogeneity of P450 Heme Orientation. *Biochemistry* **2005**, *44* (42), 13914–13920.

(115) Henne, K. R.; Kunze, K. L.; Zheng, Y. M.; Christmas, P.; Soberman, R. J.; Rettie, A. E. Covalent Linkage of Prosthetic Heme to CYP4 Family P450 Enzymes. *Biochemistry* **2001**, *40* (43), 12925–12931.

(116) Zheng, Y. M.; Baer, B. R.; Kneller, M. B.; Henne, K. R.; Kunze, K. L.; Rettie, A. E. Covalent Heme Binding to CYP4B1 via Glu310 and a Carbocation Porphyrin Intermediate. *Biochemistry* **2003**, *42* (15), 4601–4606.

(117) Limburg, J.; LeBrun, L. A.; Ortiz De Montellano, P. R. The P450cam G248E Mutant Covalently Binds Its Prosthetic Heme Group. *Biochemistry* **2005**, *44* (10), 4091–4099.

(118) Ortiz de Montellano, P. R.; Kunze, K. L. Cytochrome P-450 Inactivation: Structure of the Prosthetic Heme Adduct with Propyne. *Biochemistry* **1981**, *20* (25), 7266–7271.

(119) Dierks, E. A.; Davis, S. C.; Ortiz De Montellano, P. R. Glu-320 and Asp-323 Are Determinants of the CYP4A1 Hydroxylation Regiospecificity and Resistance to Inactivation by 1-Aminobenzotriazole. *Biochemistry* **1998**, *37* (7), 1839–1847.

(120) Henne, K. R.; Fisher, M. B.; Iyer, K. R.; Lang, D. H.; Trager, W. F.; Rettie, A. E. Active Site Characteristics of CYP4B1 Probed with Aromatic Ligands. *Biochemistry* **2001**, *40* (29), 8597–8605.

(121) Kim, D.; Cha, G. S.; Nagy, L. D.; Yun, C. H.; Guengerich, F. P. Kinetic Analysis of Lauric Acid Hydroxylation by Human Cytochrome P450 4A11. *Biochemistry* **2014**, *53* (39), 6161–6172.

(122) Kikuta, Y.; Kusunose, E.; Kusunose, M. Characterization of Human Liver Leukotriene B4 ω-Hydroxylase P450 (CYP4F2). *J. Biochem.* **2000**, *127* (6), 1047–1052.

(123) Kikuta, Y.; Kusunose, E.; Endo, K.; Yamamoto, S.; Sogawa, K.; Fujii-Kuriyama, Y.; Kusunose, M. A Novel Form of Cytochrome P-450 Family 4 in Human

Polymorphonuclear Leukocytes. CDNA Cloning and Expression of Leukotriene B4 ω-Hydroxylase. *J. Biol. Chem.* **1993**, *268* (13), 9376–9380.

(124) Khanapure, S.; Garvey, D.; Janero, D.; Gordon Letts, L. Eicosanoids in Inflammation: Biosynthesis, Pharmacology, and Therapeutic Frontiers. *Curr. Top. Med. Chem.* **2007**, *7* (3), 311–340.

(125) Harvey, R. D.; Morgan, E. T. Cancer, Inflammation, and Therapy: Effects on Cytochrome P450-Mediated Drug Metabolism and Implications for Novel Immunotherapeutic Agents. *Clin. Pharmacol. Ther.* **2014**, *96* (4), 449–457.

(126) Rivory, L. P.; Slaviero, K. A.; Clarke, S. J. Hepatic Cytochrome P450 3A Drug Metabolism Is Reduced in Cancer Patients Who Have an Acute-Phase Response. *Br. J. Cancer* **2002**, *87* (3), 277–280.

(127) Gandhi, A. V.; Saxena, S.; Relles, D.; Sarosiek, K.; Kang, C. Y.; Chipitsyna, G.; Sendecki, J. A.; Yeo, C. J.; Arafat, H. A. Differential Expression of Cytochrome P450 Omega-Hydroxylase Isoforms and Their Association with Clinicopathological Features in Pancreatic Ductal Adenocarcinoma. *Ann. Surg. Oncol.* **2013**, *20* (3 SUPPL.), 636–643.

(128) Chen, L.; Hardwick, J. P. Identification of a New P450 Subfamily, CYP4F1, Expressed in Rat Hepatic Tumors. *Arch. Biochem. Biophys.* **1993**, *300* (1), 18–23.

(129) Kalsotra, A.; Strobel, H. W. Cytochrome P450 4F Subfamily: At the Crossroads of Eicosanoid and Drug Metabolism. *Pharmacol. Ther.* **2006**, *112* (3), 589–611.

(130) Le Quéré, V.; Plée-Gautier, E.; Potin, P.; Madec, S.; Salaün, J. P. Human CYP4F3s Are the Main Catalysts in the Oxidation of Fatty Acid Epoxides. *J. Lipid Res.* **2004**, *45* (8), 1446–1458.

(131) Panigrahy, D.; Kaipainen, A.; Greene, E. R.; Huang, S. Cytochrome P450-Derived Eicosanoids: The Neglected Pathway in Cancer. *Cancer Metastasis Rev.* **2010**, *29* (4), 723–735.

(132) Schuck, R. N.; Zha, W.; Edin, M. L.; Gruzdev, A.; Vendrov, K. C.; Miller, T. M.; Xu, Z.; Lih, F. B.; DeGraff, L. M.; Tomer, K. B.; Jones, H. M.; Makowski, L.; Huang, L.; Poloyac, S. M.; Zeldin, D. C.; Lee, C. R. The Cytochrome P450 Epoxygenase Pathway Regulates the Hepatic Inflammatory Response in Fatty Liver Disease. *PLoS One* **2014**, *9* (10), e110162.

(133) Sherwood, L. M.; Parris, E. E.; Folkman, J. Tumor Angiogenesis: Therapeutic Implications. *N. Engl. J. Med.* **1971**, *285* (21), 1182–1186.

(134) Weigelt, B.; Peterse, J. L.; Van't Veer, L. J. Breast Cancer Metastasis: Markers and Models. *Nat. Rev. Cancer* **2005**, *5* (8), 591–602.

(135) Roman, R. J. P-450 Metabolites of Arachidonic Acid in the Control of Cardiovascular Function. *Physiol. Rev.* **2002**, *82* (1), 131–185.

(136) Tateno, C.; Yoshizane, Y.; Saito, N.; Kataoka, M.; Utoh, R.; Yamasaki, C.; Tachibana, A.; Soeno, Y.; Asahina, K.; Hino, H.; Asahara, T.; Yokoi, T.; Furukawa, T.; Yoshizato, K. Near Completely Humanized Liver in Mice Shows Human-Type Metabolic Responses to Drugs. *Am. J. Pathol.* **2004**, *165* (3), 901–912.

(137) Nelson, D. R.; Zeldin, D. C.; Hoffman, S. M. G.; Maltais, L. J.; Wain, H. M.; Nebert, D. W. Comparison of Cytochrome P450 (CYP) Genes from the Mouse and Human Genomes, Including Nomenclature Recommendations for Genes, Pseudogenes and Alternative-Splice Variants. *Pharmacogenetics* **2004**, *14* (1), 1–

18.

(138) Hancock, J. M.; Zvelebil, M. J.; Cummings, M. P. PHYLIP (PHYLogeny Inference Package). In *Dictionary of Bioinformatics and Computational Biology*; 2004.

(139) Renaud, H. J.; Cui, J. Y.; Khan, M.; Klaassen, C. D. Tissue Distribution and Gender-Divergent Expression of 78 Cytochrome P450 MRNAs in Mice. *Toxicol. Sci.* **2011**, *124* (2), 261–277.

(140) Christmas, P.; Jones, J. P.; Patten, C. J.; Rock, D. A.; Zheng, Y.; Cheng, S. M.; Weber, B. M.; Carlesso, N.; Scadden, D. T.; Rettie, A. E.; Soberman, R. J. Alternative Splicing Determines the Function of CYP4F3 by Switching Substrate Specificity. *J. Biol. Chem.* **2001**, *276* (41), 38166–38172.

(141) Christmas, P.; Tolentino, K.; Primo, V.; Berry, K. Z.; Murphy, R. C.; Chen, M.; Lee, D. M.; Soberman, R. J. Cytochrome P-450 4F18 Is the Leukotriene B4 ω-1/ω-2 Hydroxylase in Mouse Polymorphonuclear Leukocytes: Identification as the Functional Orthologue of Human Polymorphonuclear Leukocyte CYP4F3A in the down-Regulation of Responses to LTB4. *J. Biol. Chem.* **2006**, *281* (11), 7189–7196.

(142) Cui, X.; Kawashima, H.; Barclay, T. B.; Peters, J. M.; Gonzalez, F. J.; Morgan, E. T.; Strobel, H. W. Molecular Cloning and Regulation of Expression of Two Novel Mouse CYP4F Genes: Expression in Peroxisome Proliferator-Activated Receptor Alpha-Deficient Mice upon Lipopolysaccharide and Clofibrate Challenges. *J. Pharmacol. Exp. Ther.* **2001**, *296* (2), 542–550.

(143) Hardwick, J. P.; Osei-Hyiaman, D.; Wiland, H.; Abdelmegeed, M. A.; Song, B. J.

PPAR/RXR Regulation of Fatty Acid Metabolism and Fatty Acid ω-Hydroxylase (CYP4) Isozymes: Implications for Prevention of Lipotoxicity in Fatty Liver Disease. *PPAR Res.* **2009**, *2009*, 952734.

(144) Alecu, I.; Othman, A.; Penno, A.; Saied, E. M.; Arenz, C.; Von Eckardstein, A.; Hornemann, A. T. Cytotoxic 1-Deoxysphingolipids Are Metabolized by a Cytochrome P450-Dependent Pathway. *J. Lipid Res.* **2017**, *58* (1), 60–71.

(145) Miyata, N.; Taniguchi, K.; Seki, T.; Ishimoto, T.; Sato-Watanabe, M.; Yasuda, Y.; Doi, M.; Kametani, S.; Tomishima, Y.; Ueki, T.; Sato, M.; Kameo, K. HET0016, a Potent and Selective Inhibitor of 20-HETE Synthesizing Enzyme. *Br. J. Pharmacol.* **2001**, *133* (3), 325–329.

(146) Lee, C. A.; Kadwell, S. H.; Kost, T. A.; Serabjit-Singh, C. J. CYP3A4 Expressed by Insect Cells Infected with a Recombinant Baculovirus Containing Both CYP3A4 and Human NADPH-Cytochrome P450 Reductase Is Catalytically Similar to Human Liver Microsomal CYP3A4. *Arch. Biochem. Biophys.* **1995**, *319* (1), 157–167.

(147) Griffiths, D. J.; Schroeter, D. F. *Introduction to Quantum Mechanics*; 2018.

(148) Born, M.; Oppenheimer, R. Zur Quantentheorie Der Molekeln. *Ann. Phys.* **1927**, *389* (20), 457–484.

(149) Slater, J. C. The Theory of Complex Spectra. *Phys. Rev.* **1929**, *34* (10), 1293–1322.

(150) Froese Fischer, C. General Hartree-Fock Program. *Comput. Phys. Commun.* **1987**, *43* (3), 355–365.

(151) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct

Asymptotic Behavior. *Phys. Rev. A* **1988**, *38* (6), 3098–3100.

(152) Hubbard, R. E. Molecular Graphics and Modeling: Tools of the Trade. In *Guidebook on Molecular Modeling in Drug Design*; Elsevier, 1996; pp 19–54.

(153) Karr, J. R.; Takahashi, K.; Funahashi, A. The Principles of Whole-Cell Modeling. *Curr. Opin. Microbiol.* **2015**, *27*, 18–24.

(154) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian˜09 Revision B.01. 2016.

(155) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993**, *14* (11), 1347–1363.

(156) Zwanzig, R. W. Statistical Mechanics. *J. Am. Chem. Soc.* **1962**, *84* (18), 3602–3603.

(157) Wereszczynski, J.; McCammon, J. A. Statistical Mechanics and Molecular Dynamics in Evaluating Thermodynamic Properties of Biomolecular Recognition. *Q. Rev. Biophys.* **2012**, *45* (1), 1–25.

(158) Alder, B. J.; Wainwright, T. E. Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys.* **1959**, *31* (2), 459–466.

(159) Koehl, P.; Levitt, M. A Brighter Future for Protein Structure Prediction. In *Nature Structural Biology*; 1999; Vol. 6, pp 108–111.

(160) Moors, S. L. C.; Vos, A. M.; Cummings, M. D.; Van Vlijmen, H.; Ceulemans, A. Structure-Based Site of Metabolism Prediction for Cytochrome P450 2D6. *J. Med. Chem.* **2011**, *54* (17), 6098–6105.

(161) Kingsley, L. J.; Wilson, G. L.; Essex, M. E.; Lill, M. A. Combining Structure- and Ligand-Based Approaches to Improve Site of Metabolism Prediction in CYP2C9 Substrates. *Pharm. Res.* **2015**, *32* (3), 986–1001.

(162) Bonomo, S.; Jørgensen, F. S.; Olsen, L. Dissecting the Cytochrome P450 1A2- and 3A4-Mediated Metabolism of Aflatoxin B1 in Ligand and Protein Contributions. *Chem. - A Eur. J.* **2017**, *23* (12), 2884–2893.

(163) Li, J.; Cai, J.; Su, H.; Du, H.; Zhang, J.; Ding, S.; Liu, G.; Tang, Y.; Li, W. Effects of Protein Flexibility and Active Site Water Molecules on the Prediction of Sites of Metabolism for Cytochrome P450 2C19 Substrates. *Mol. Biosyst.* **2016**, *12* (3), 868–878.

(164) D.A. Case; K. Belfon; I.Y. Ben-Shalom; S.R. Brozell; D.S. Cerutti; T.E.

Cheatham; III; V.W.D. Cruzeiro; T.A. Darden; R.E. Duke; G. Giambasu; M.K. Gilson; H. Gohlke; A.W. Goetz; R. Harris; S. Izadi; S.A. Izmailov; K. Kasavajhala; A. Kovalenko; R. Krasny; T. Kurtzman; T.S. Lee; S. LeGrand; P. Li; C. Lin; J. Liu; T. Luchko; R. Luo; V. Man; K.M. Merz; Y. Miao; O. Mikhailovskii; G. Monard; H. Nguyen; A. Onufriev; F.Pan; S. Pantano; R. Qi; D.R. Roe; A. Roitberg; C. Sagui; S. Schott-Verdugo; J. Shen; C.L. Simmerling; N.R.Skrynnikov; J. Smith; J. Swails; R.C. Walker; J. Wang; L. Wilson; R.M. Wolf; X. Wu; Y. Xiong; Y. Xue; D.M. York; P.A. Kollman. AMBER 2020.

(165) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30* (10), 1545–1614.

(166) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117* (1), 1–19.

(167) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26* (16), 1701–1718.

(168) Nelson, M. T.; Humphrey, W.; Gursoy, A.; Dalke, A.; Kale, L. V.; Skeel, R. D.; Schulten, K. NAMD: A Parallel, Object-Oriented Molecular Dynamics Program. *Int. J. High Perform. Comput. Appl.* **1996**, *10* (4), 251–268.

(169) Huang, J.; Mackerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34* (25), 2135–2145.

(170) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586–3616.

(171) Harder, E.; Anisimov, V. M.; Vorobyov, I. V.; Lopes, P. E. M.; Noskov, S. Y.; MacKerell, A. D.; Roux, B. Atomic Level Anisotropy in the Electrostatic Modeling of Lone Pairs for a Polarizable Force Field Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2006**, *2* (6), 1587–1597.

(172) Martys, N. S.; Mountain, R. D. Velocity Verlet Algorithm for Dissipative-Particle-Dynamics-Based Models of Suspensions. *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.* **1999**, *59* (3), 3733–3736.

(173) Beeman, D. Some Multistep Methods for Use in Molecular Dynamics Calculations. *J. Comput. Phys.* **1976**, *20* (2), 130–139.

(174) Andersen, H. C. Rattle: A "Velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations. *J. Comput. Phys.* **1983**, *52* (1), 24–34.

(175) Valverde, J. R. Molecular Modelling: Principles and Applications. *Brief. Bioinform.* **2001**, *2* (2), 199–200.

(176) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; 1996.

(177) Hardy, D. J.; Wu, Z.; Phillips, J. C.; Stone, J. E.; Skeel, R. D.; Schulten, K. Multilevel Summation Method for Electrostatic Force Evaluation. *J. Chem. Theory Comput.* **2015**, *11* (2), 766–779.

(178) Di Pierro, M.; Elber, R.; Leimkuhler, B. A Stochastic Algorithm for the Isobaric-Isothermal Ensemble with Ewald Summations for All Long Range Forces. *J. Chem. Theory Comput.* **2015**, *11* (12), 5624–5637.

(179) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593.

(180) van Gunsteren, W. F.; Berendsen, H. J. C. Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. *Angew. Chem. Int. Ed. Engl.* **1990**, *29* (9), 992–1023.

(181) Okonechnikov, K.; Golosova, O.; Fursov, M.; UGENE team. Unipro UGENE: A Unified Bioinformatics Toolkit. *Bioinformatics* **2012**, *28* (8), 1166–1167.

(182) Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *J. Chem. Phys.* **1980**, *72* (4), 2384–2393.

(183) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103* (11), 4613–4621.

(184) Pastor, R. W.; Brooks, B. R.; Szabo, A. An Analysis of the Accuracy of Langevin and Molecular Dynamics Algorithms. *Mol. Phys.* **1988**, *65* (6), 1409–1419.

(185) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.;

Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781–1802.

(186) Branden, C.; Tooze, J. *Introduction to Protein Structure*; Garland Publishing Company: New York, 1991.

(187) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101.

(188) Werck-Reichhart, D.; Feyereisen, R. Cytochromes P450: A Success Story. *Genome Biol.* **2000**, *1* (6), reviews3003.1-9.

(189) Krieger, E.; Nabuurs, S. B.; Vriend, G. Homology Modeling. In *Structural Bioinformatics*; 2005; pp 509–523.

(190) Okonechnikov, K.; Golosova, O.; Fursov, M.; Varlamov, A.; Vaskin, Y.; Efremov, I.; German Grehov, O. G.; Kandrov, D.; Rasputin, K.; Syabro, M.; Tleukenov, T. Unipro UGENE: A Unified Bioinformatics Toolkit. *Bioinformatics* **2012**, *28* (8), 1166–1167.

(191) Zhang, Y. I-TASSER: Fully Automated Protein Structure Prediction in CASP8. *Proteins Struct. Funct. Bioinforma.* **2009**, *77* (SUPPL. 9), 100–113.

(192) Yang, J.; Zhang, Y. I-TASSER Server: New Development for Protein Structure and Function Predictions. *Nucleic Acids Res.* **2015**, *43* (W1), W174–W181.

(193) Roy, A.; Yang, J.; Zhang, Y. COFACTOR: An Accurate Comparative Algorithm for Structure-Based Protein Function Annotation. *Nucleic Acids Res.* **2012**, *40* (W1), 471–477.

(194) Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C. SWISS-MODEL: An Automated Protein Homology-Modeling Server. *Nucleic Acids Res.* **2003**, *31* (13), 3381–

3385.

(195) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410.

(196) Edgar, R. C. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res.* **2004**, *32* (5), 1792–1797.

(197) Thompson, J. D.; Gibson, T. J.; Higgins, D. G. Multiple Sequence Alignment Using ClustalW and ClustalX. *Curr. Protoc. Bioinforma.* **2003**, *00* (1), 2.3.1-2.3.22.

(198) Zhang, Y. I-TASSER Server for Protein 3D Structure Prediction. *BMC Bioinformatics* **2008**, *9*, 40.

(199) Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A Unified Platform for Automated Protein Structure and Function Prediction. *Nat. Protoc.* **2010**, *5* (4), 725–738.

(200) Wu, S.; Zhang, Y. LOMETS: A Local Meta-Threading-Server for Protein Structure Prediction. *Nucleic Acids Res.* **2007**, *35* (10), 3375–3382.

(201) Zheng, W.; Zhang, C.; Wuyun, Q.; Pearce, R.; Li, Y.; Zhang, Y. LOMETS2: Improved Meta-Threading Server for Fold-Recognition and Structure-Based Function Annotation for Distant-Homology Proteins. *Nucleic Acids Res.* **2019**, *47* (W1), W429–W436.

(202) Li, Y.; Zhang, Y. REMO: A New Protocol to Refine Full Atomic Protein Models from C-Alpha Traces by Optimizing Hydrogen-Bonding Networks. *Proteins Struct. Funct. Bioinforma.* **2009**, *76* (3), 665–676.

(203) Zhang, Y.; Skolnick, J. TM-Align: A Protein Structure Alignment Algorithm Based on the TM-Score. *Nucleic Acids Res.* **2005**, *33* (7), 2302–2309.

(204) Bas, D. C.; Rogers, D. M.; Jensen, J. H. Very Fast Prediction and Rationalization of pK_a Values for Protein-Ligand Complexes. *Proteins Struct. Funct. Genet.* **2008**, *73* (3), 765–783.

(205) Binkowski, T. A.; Naghibzadeh, S.; Liang, J. CASTp: Computed Atlas of Surface Topography of Proteins. *Nucleic Acids Res.* **2003**, *31* (13), 3352–3355.

(206) Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J. CASTp 3.0: Computed Atlas of Surface Topography of Proteins. *Nucleic Acids Res.* **2018**, *46* (W1), W363–W367.

(207) Laury, M. L.; Wang, L. P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W. Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *J. Phys. Chem. B* **2015**, *119* (29), 9423–9437.

(208) Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. Rapid Parameterization of Small Molecules Using the Force Field Toolkit. *J. Comput. Chem.* **2013**, *34* (32), 2757–2770.

(209) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31* (4), 671–690.

(210) Morris, G. M.; Ruth, H.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. Software News and Updates AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785–2791.

(211) Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated Docking of Flexible

Ligands: Applications of AutoDock. *J. Mol. Recognit.* **1996**, *9* (1), 1–5.

(212) Pincus, M. Letter to the Editor—A Monte Carlo Method for the Approximate Solution of Certain Types of Constrained Optimization Problems. *Oper. Res.* **1970**, *18* (6), 1225–1228.

(213) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19* (14), 1639–1662.

(214) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461.

(215) Elfiky, A. A. Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA Dependent RNA Polymerase (RdRp): A Molecular Docking Study. *Life Sci.* **2020**, *253*, 117592.

(216) Lassmann, T.; Sonnhammer, E. L. L. Kalign - An Accurate and Fast Multiple Sequence Alignment Algorithm. *BMC Bioinformatics* **2005**, *6*, 1–9.

(217) 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.

(218) 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.

(219) Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. The Structure of Human Microsomal Cytochrome P450 3A4 Determined by X-ray Crystallography to 2.05-Å Resolution. *J. Biol. Chem.* **2004**, *279* (37), 38091–38094.

(220) Wu, S.; Zhang, Y. MUSTER: Improving Protein Sequence Profile-Profile Alignments by Using Multiple Sources of Structure Information. *Proteins Struct. Funct. Genet.* **2008**, *72* (2), 547–556.

(221) Jennings, G. K.; Hsu, M. H.; Shock, L. S.; Johnson, E. F.; Hackett, J. C. Noncovalent Interactions Dominate Dynamic Heme Distortion in Cytochrome P450 4B1. *J. Biol. Chem.* **2018**, *293* (29), 11433–11446.

(222) Dennington, R.; Keith, T. A.; Millam, J. M. GaussView Version 5. 2019.

(223) Li, J.; Zhang, H.; Liu, G.; Tang, Y.; Tu, Y.; Li, W. Computational Insight into Vitamin K1 ω-Hydroxylation by Cytochrome P450 4F2. *Front. Pharmacol.* **2018**, *9* (SEP), 1–10.

(224) El-Sherbeni, A. A.; El-Kadi, A. O. S. Repurposing Resveratrol and Fluconazole to Modulate Human Cytochrome P450-Mediated Arachidonic Acid Metabolism. *Mol. Pharm.* **2016**, *13* (4), 1278–1288.

(225) Eschenfeldt, W. H.; Zhang, Y.; Samaha, H.; Stols, L.; Eirich, L. D.; Wilson, C. R.; Donnelly, M. I. Transformation of Fatty Acids Catalyzed by Cytochrome P450 Monooxygenase Enzymes of *Candida tropicalis*. *Appl. Environ. Microbiol.* **2003**, *69* (10), 5992–5999.

(226) Feher, V. A.; Baldwin, E. P.; Dahlquist, F. W. Access of Ligands to Cavities within the Core of a Protein Is Rapid. *Nat. Struct. Biol.* **1996**, *3* (6), 516–521.

(227) Richards, L.; Lutz, A.; Chalmers, D. K.; Jarrold, A.; Bowser, T.; Stevens, G. W.; Gras, S. L. Production of Metabolites of the Anti-Cancer Drug Noscapine Using a P450BM3 Mutant Library. *Biotechnol. Reports* **2019**, *24*, e00372.

(228) Melo, A.; Ramos, M. J.; Floriano, W. B.; Gomes, J. A. N. F.; Leao, J. F. R.; Magalhaes, A. L.; Maigret, B.; Nascimento, M. C.; Reuter, N. Theoretical Study of Arginine-Carboxylate Interactions. *J. Mol. Struct. THEOCHEM* **1999**, *463* (1–2), 81–90.

(229) Hasemann, C. A.; Kurumbail, R. G.; Boddupalli, S. S.; Peterson, J. A.; Deisenhofer, J. Structure and Function of Cytochromes P450:A Comparative Analysis of Three Crystal Structures. *Structure* **1995**, *3* (1), 41–62.

VITA

Jerome W. Butler III

EDUCATION

Master of Science Student in Chemistry at Sam Houston State University, August 2020 – present. Thesis title: "Modeling and Molecular Dynamics simulations on the *in situ* Murine Cytochrome P450 4f System"

Bachelor of Science (May 2018) in Chemistry, Sam Houston State University, Huntsville, Texas.

Associate of Science (May 2015), Lonestar College System, Houston, Texas.

ACADEMIC EMPLOYMENT

Graduate Teaching Assistant, Department of Chemistry, Sam Houston State University, September 2018 – present. Responsibilities include: leading and directing the laboratory portion of many chemistry courses, proofreading and review of writing-enhanced formal laboratory reports submitted to the course, guiding and informing potentially future scientists and researchers of safety precautions and common laboratory procedures. Laboratory sections covered: Biochemistry I (CHEM 3438), Physical Chemistry I (CHEM 4448), General Chemistry II (CHEM 1412), Intro Organic and Biochemistry (CHEM 1407), Organic Chemistry I Laboratory (CHEM 2123).

Graduate Research Assistant to Dr. Donovan Haines, Department of Chemistry, Sam Houston State University, September 2018 – present. Research activities include: the study of cytochrome P450 enzymes and their functions and roles in human health and disease using computational chemistry methods. Experienced in homology modeling, molecular docking, molecular dynamics simulation and other computational approaches.

PRESENTATIONS AT PROFESSIONAL MEETINGS

Butler, J. W. and D.C. Haines. *Molecular modeling and molecular dynamics simulations on murine Cyp4f13: Insight into a homolog of human ω-hydroxylase CYP4F3, a potentiator of tumor cell metastasis.* ACS Southwest & Rocky Mountain Regional Meeting (ACS SWRM), El Paso, Texas, November 14, 2019.

Butler, J. W. and D.C. Haines. *Investigation into Murine ω-Hydroxylase Cyp4f13, a Tumor Cell Potentiator, using Homology Modeling and Molecular Dynamics*

Simulations. Texas Academy of Science 2020 Annual Meeting (TAS 2020) Stephen F. Austin State University, Nacogdoches, Texas, February 28, 2019.

ACADEMIC AWARDS

Raven Scholar, Department of Chemistry, Sam Houston State University 2018-2019

COSET Graduate Achievement Scholarship, Sam Houston State University 2018, 2019

Robert A. Welch Fellowship, Sam Houston State University Summer 2018, 2019