Selectivity Mechanisms Employed by Flavin-Dependent Monooxygenases

by

Attabey Rodríguez Benítez

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in The University of Michigan
2021

Doctoral Committee:

Assistant Professor Alison R. H. Narayan, Chair
Professor Janet L. Smith Co-Chair
Assistant Professor Jennifer Bridwell-Rabb
Professor Anna K. Mapp
Professor Bruce A. Palfey
Professor Melanie S. Sanford

Attabey Rodríguez-Benítez

attabey@umich.edu

ORCID iD: 0000-0002-5824-2585

To my family and ancestors whose shoulders I stand on, this would not have been

possible without your sacrifice.

ACKNOWLEDGEMENTS

This milestone would not have been possible without the support of my family, friends, mentors over the past few years. First, I would like to thank both my advisors Prof. Janet L. Smith and Prof. Alison R. H. Narayan for their full support and guidance. Both of you provided me with not only great mentorship and guidance but also the space to try new projects and explore non-traditional career paths. I am deeply thankful to Janet and Alison for guiding me on this amazing journey. I also would like to thank the rest of my committee members, Prof. Anna K. Mapp, Prof. Melanie S. Sanford, Prof. Jennifer Bridwell-Rabb, and Prof. Bruce A. Palfey, for their guidance and support throughout my graduate career. I also would like to acknowledge and thank my collaborators at the University of Michigan, Prof. Charles III. Brooks, Dr. Troy Wymore, Sara Tweedy, and Bruce A. Palfey, I appreciate your contributions over the past years.

I also would like to thank the mentors I met during my undergraduate years at the University of Puerto Rico, which lead the way to where I am today. Dr. Alan Rodríguez was my first mentor and the person that showed how beautiful science and chemistry can be and that a career as a scientist was attainable. He introduced me to Prof. John Soderquist who took the chance on the first year me to join his lab. With the guidance of Denisse Alequín Torres, I became a successful synthetic chemist in a short amount of time. I would like to thank Prof. José Rivera and María del Carmen Rivera Sánchez for allowing me to explore the interface between chemistry and biology and guiding me in becoming an independent scientist. Lastly, I would like to thank Prof. Rosa V. Flores and Prof. Ingrid del Carmen Montes González who taught me to never doubt myself and that

I should pursue my curiosity because it might lead to interesting places.

Starting in a brand new lab and my project from the ground and on a topic outside my expertise was challenging, but I am grateful to through this journey with my two-unit of people the Narayan and Smith labs. I consider you all more than my unit of people or friends, but as a family and I am more than honored to have worked with such talented scientists. My journey in graduate school would have been very different without your guidance, mentorship, jokes, and 80s hour during late lab nights. We all have gone through high and lows together and created memories that will last a lifetime and I look forward to seeing you in the future to create some more.

I also would like to thank my friends and partner that made the Michigan tundra like Puerto Rico. You all have been my support network the times I have need it the most. Thank you for being there through the rocky road that graduate school is, for the late nights of Bad Bunny, for sharing all the high and lows, and making Michigan more like home.

Lastly, I would like to acknowledge and thank my family. Words cannot express how thankful I am for all the sacrifices you have made for me to be where I am today. Thank you so much for believing in me and providing your unconditional support throughout all these years. Leaving you all behind has been one of the hardest things I have done in my life. I miss you all with all my heart and I could not have done it without you all.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

FAD flavin adenine dinucleotide

FDMO flavin-dependent monooxygenase

NADH Nicotinamide adenine dinucleotide

NADPH Nicotinamide adenine dinucleotide phosphate

SSN sequence similarity network

VAE variational autoencoder

ee enantiomeric excess

ADP adenosine diphosphate

G6P glucose-6-phosphate

G6PDH glucose-6-phophate dehydrogenase

PHBH *p*-hydroxybenzoate hydroxylase

PHHY phenol 2-monooxygenase

MHBH 3-hydroxybenzoate hydroxylase

SalH Salicylate hydroxylase

MD molecular dynamics

QM/MM quantum mechanical/molecular mechanical

rmsd root-mean-square deviation

IDT Integrated DNA Technologies

TB Terrific Broth

IPTG isopropyl-β-D-1-thiogalactopyranoside

CV column volumes

APS Advanced Photon Source

DMSO Dimethyl sulfoxide

E. coli Escherichia coli

SD Steepest Descent

ABNR Adopted Basis Newton Raphson

GBMV generalized Born molecular volume

fs femtosecond

ns nanoseconds

UPLC-DAD ultra high-performance liquid chromatography coupled to a photodiode array

PDA PhotoDiode Array

er enantiomeric ratio

LB Luria Broth

HRP horseradish peroxidase

EFI Enzyme Function Initiative

ABSTRACT

Nature is an incredible source of inspiration for the discovery and subsequent development of new bioactive compounds. Unfortunately, the synthesis of these molecules is often prohibitively complex, requiring the installation of multiple functional groups with intricate three-dimensional architectures critical to their biological activity. Biocatalytic reactions embody many features of ideal chemical transformations, including the potential for impeccable selectivity, high catalytic efficiency, mild reaction conditions, and the use of environmentally benign reagents. These advantages have created a demand for new biocatalysts that expand the portfolio of complexity-generating reactions available to synthetic chemists. Oxidative dearomatization is a powerful transformation in the synthesis of complex molecules, as it generates a stereocenter and simultaneously producing a compound primed for further reactions. Nature has developed a class of biocatalysts, flavin-dependent monooxygenases, which perform this reaction with an excellent site- and stereoselectivity under mild conditions. Our studies on the TropB-catalyzed hydroxylation of phenolic compounds has defined the substrate scope of these biocatalysts; however, the mechanistic underpinnings were a mystery. Through analysis of class A FAD monooxygenases and biochemical characterization of TropB we determined that the phenolate form of the substrate interacts with Tyr239 and Arg206 to control the site- and stereo-selectivity of the hydroxylation. We then we explore how this control for site- and stereo-selective is translated to a selection of FAD-dependent monooxygenases. Through a sequence-profiling approach we identified the FDMO AfoD with complementary selectivity compared to TropB. We determined by probing similarly positioned residues through muta-

genesis and biochemical characterization that selectivity can be eroded when Tyr118 hydrogen bonding is affected. These findings pave the way in identifying new biocatalysts for reaction development toward natural product synthesis.

CHAPTER I

Introduction

1.1   Biocatalysis

Natural products are an inspiration for the synthesis of therapeutics. The World Health Organization estimates that approximately 80 percent of the emerging world's population relies on traditional medicines and by 2013 one third of the United States Food and Drug Administration (FDA)-approved new molecular entities were based on or derived from natural products. [1, 2] Unfortunately, the synthesis of these molecules is often prohibitively complex, through biocatalysis scientists can synthesize molecules that would not be accessible through small molecule synthesis. [3] Chemists are inspired by the catalytic tools evolved by Nature to develop eco-friendly routes which have enabled chemists to synthesize molecules that would not be possible through traditional chemistry. [4] [5] The bridge between biosynthesis and synthetic chemistry is central in the field of bio-

catalysis where enzymes are leveraged in chemical processes to streamline syntheses and minimize potential hazardous waste. [6] [7] The first challenge is identifying a biocatalyst capable of reacting with a variety of substrates. Currently, there are 218 million annotated protein sequences available in public databases, a number that doubles every 28 months, and just like the deep seafloor, only a minuscule portion of this territory has been explored. [8]

1.2   Flavin Dependent Monooxygenases as Biocatalysts

Each cDNA sequence encodes for a protein with a unique composition and order of amino acids that dictate its fold, and in the case of an enzyme, the reactions it can catalyze. However, predicting a function based on the amino acid sequence is not an easy feat. Typically, the function has been experimentally determined through labor-intensive protein expression and isolation coupled with experimental characterization of enzymes from primary metabolism and natural product biosynthetic pathways. [9, 10]

Enzymes and their function can be distributed into different classes depending on their fold which also dictates the reaction that these enzymes can perform. In Flavin-dependent enzymes, each of the classes have a canonical enzyme used as guide for classification. Well-characterized enzymes have historically served as benchmarks for predicting the function of uncharacterized enzymes. For example, flavin-dependent monooxygenase (FDMO) can mediate various transformations depending on the amino acids in their active site. Class A flavin adenine dinucleotide (FAD)-dependent monooxygenases mediate the hydroxylation of substrates. [11] The fold is characterized by having one nucleotide-binding site occupied by the FAD also reflected in the sequence by the GXGXXG motif. [12,13] Class A FDMOs are defined by being "cautious" enzymes [14], starting catalysis with flavin reduction after substrate binding (Fig. 1.1) and (Fig. 1.2). Accommodating the substrate in the active site before the reaction occurs is critical in mediating the site- and enantioselectivity of the product. This binding will trigger a change in conformation in

the FAD. The FAD is usually located in the active site and this conformation is denotated as the "in" conformation. [15] When the FAD swings by a 30° it gets exposed to solvent and the electron source, the conformation is denoted as "out". [15] After substrate binding, a change in conformation from the "in" to the "out" conformation is triggered. In the "out" conformation the flavin is reduced after accessing the electron source, Nicotinamide adenine dinucleotide phosphate (NADPH), or Nicotinamide adenine dinucleotide (NADH) (Fig. 1.1). [16] Reduced flavin (FADH$_2$) reacts readily with molecular oxygen to form the C4a-hydroperoxyflavin FAD_{HOOH} (**1.4**) intermediate which reacts with the substrate to form the hydroxylated product (Fig. 1.1), or can eliminate a molecule of hydrogen peroxide to form oxidized FAD (Fig. 1.1). The catalytic cycle is then closed by the elimination of water from the C4a-hydroxyflavin to regenerate FAD.
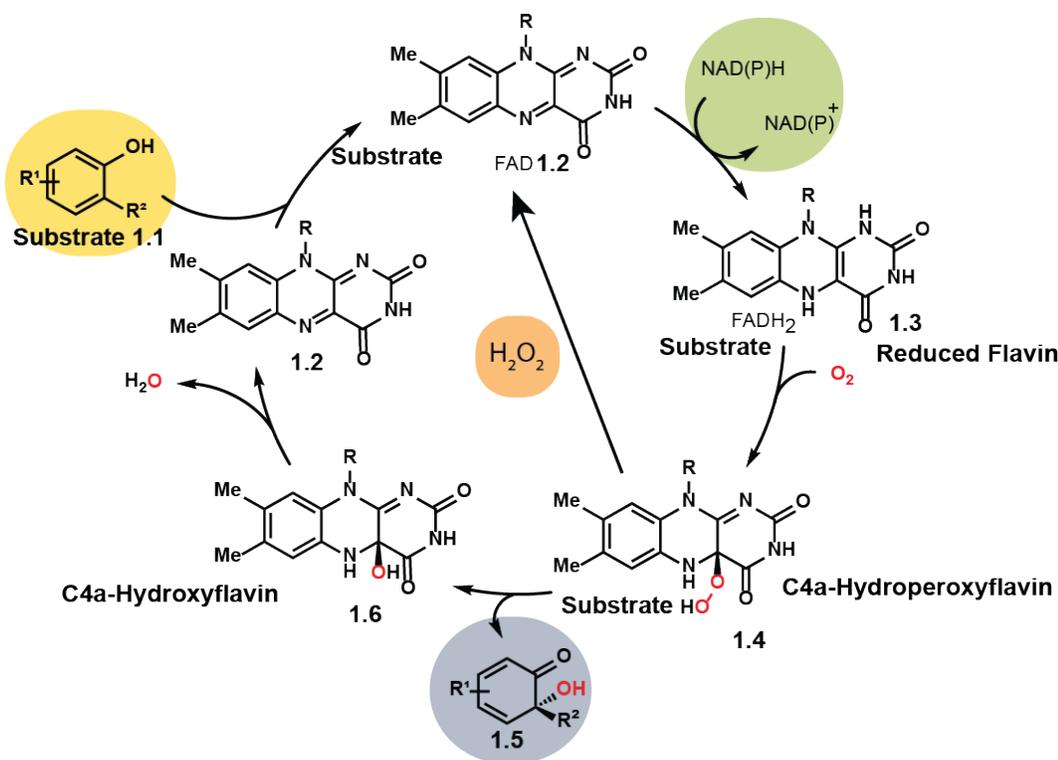
Figure 1.1: Class A Flavin Dependent Monooxygenases Catalytic Cycle.

Class A FDMOs share a structurally similar nucleotide-binding site to the class B FD-MOs; however, they have fundamental structural and mechanistic differences. Class B also

3

known as "bold" enzymes [14] can start catalysis in the absence of the substrate Fig. 1.2. Class B enzymes have two nucleotide-binding sites, for the FAD and the NAD(P)H. The catalytic cycle is started by NAD(P)H binding and subsequent reduction of the FAD. The enzyme-NADP complex quickly reacts with oxygen to form the C4a-hydroperoxyflavin which is remarkably stable (approximately 5 min) compared to the millisecond-scale lifetime found in class A monooxygenases. [14] This stability is due to the NADP-bound in the active site, which is released after the product is formed. After the substrate is bound to the active site, hydroxylation of the substrate is performed through a nucleophilic attack from the FAD_{HOOH} to the substrate.
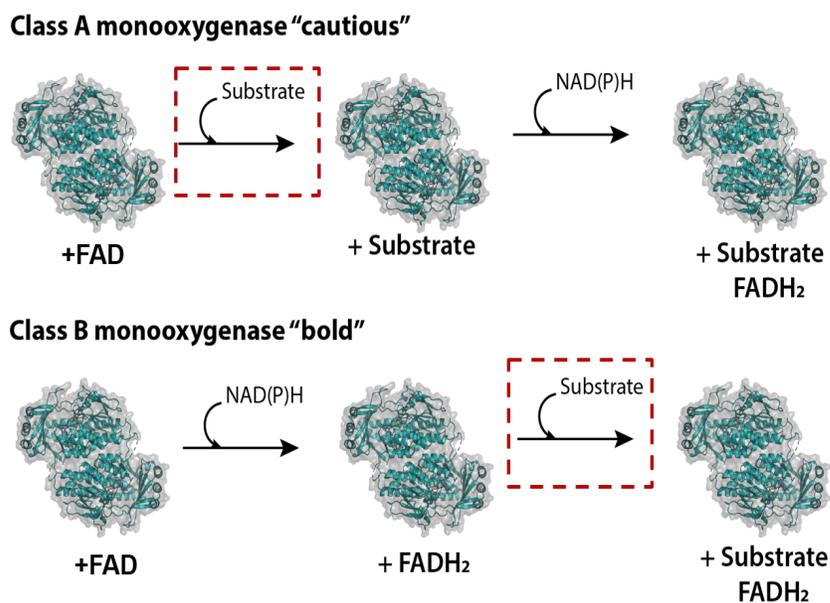
Figure 1.2: Comparison of Class A and B Flavin Dependent Monooxygenases substrate binding. Top panel showcases class A FAD-dependent monooxygenases characterized by the first step, substrate binding. This forms the Enzyme-substrate complex which later is exposed to NAD(P)H to reduce the FAD. Bottom panel showcases Class B FAD-dependent monooxygeases highlighting that substrate binding is preceded by FAD reduction.

1.3 Nature's Tools For Oxidative Dearomatization Of Resorcinol Compounds
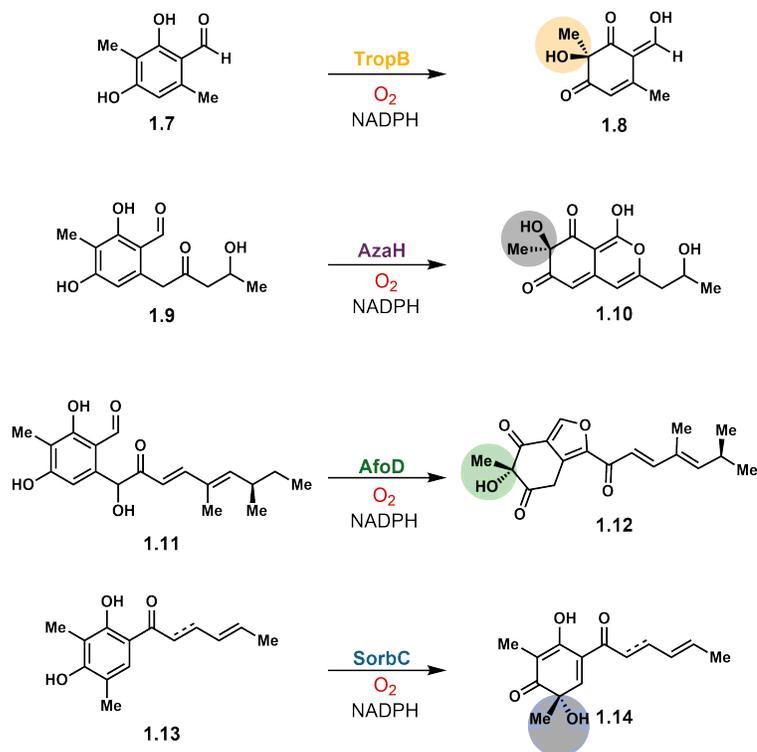
Figure 1.3: Nature's tools for oxidative dearomatization of resorcinol compounds.

The site- and stereoselectivity of class A FMDOs during hydroxylation provides an efficient route to valuable enantio-enriched intermediates for the synthesis of natural products with medicinal value with several significant advantages over conventional synthetic methods. [17, 18] For example, recent work from our group has demonstrated that the flavoproteins TropB [19], AzaH [20], and SorbC [21] achieve the rapid conversion of a number of phenolic substrates to dearomatized products (Fig. 1.3). [22] Although TropB shares a high sequence identity with AzaH and AfoD, 43 percent and 38 percent, respectively, AzaH generates products with the same configuration as TropB, while AfoD installs the oxygen with the opposite stereoselectivity (Fig. 1.3). Additionally, TropB, AzaH, and AfoD have complementary substrate scopes. For example, TropB does not tolerate bulky

substitutions at the C5 or C6 position, except for phenyl at the C6 position, while AzaH accommodates these substrates seamlessly (Fig. 1.3). While AfoD handles similar substrates as AzaH, the opposite stereocenter can be accessed. Lastly, SorbC has evolved orthogonal site- and facial-selectivity, hydroxylating the C5 position to yield the dearomatized product. This orthogonal selectivity and complementary substrate scope provide an excellent opportunity to employ these biocatalysts with a breadth of substrates. In contrast to traditional chemical transformations, in which site-selectivity is dictated by the substrate, these enzymes exhibit catalyst control over site- and stereoselectivity. [23]

1.4   Profiling Function Across Sequence Space

To expand our current library of class A FDMOs we need to predict function and mechanism based on experimental findings on these enzymes coupled with the amino acid sequence fingerprint. Although this approach can lead to accurate function assignments in some cases [24], there other instances in which enzymes possess slightly altered motifs and can be overlooked in such a function assignment. Through our previous work on enzyme characterization, benchmarks for enzyme function prediction can be set to construct a sequence similarity network (SSN). [25] Sequence similarity networks might provide a framework for identifying enzymes that act on specific substrates classes and for surveying regions of sequence space where the substrate preference is unknown.

Recently, there has been an evolution of the tools available for canvassing and identifying the sequence space with untapped synthetic potential. Some of the commonly used tools for sequence profiling and visualization are multiple sequence alignments [26], phylogenetic trees [27], and sequence similarity networks [25](Fig. 1.4). Additionally, other visualization tools are being developed such as the variational autoencoder latent space model. A variational autoencoder (VAE) provides a single value for each encoding dimension in the latent space using probability. This tool coupled with machine learning can organize a set of sequences with certain properties in mind like stability. [28]
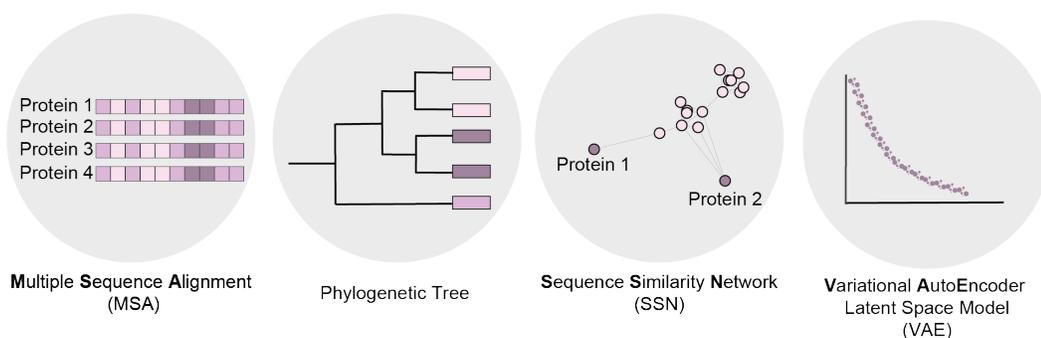
Figure 1.4: Tools for sequence profiling from left to right: multiple sequence alignment, phylogenetic tree, sequence similarity network, and variational autoencoder latent space model.

In a multiple sequence alignment, three or more protein sequences that have some evolutionary connection are aligned (Fig. 1.4). This profiling can be used to identify functional relationships among sequences. This approach highlights conserved residues and motifs that can potentially be used to predict enzyme function or pinpoint residues that might be important for enzyme function. Phylogenetic trees, as previously mentioned, indicate the relationship between sequences across evolution (Fig. 1.4). The branching pattern of these trees reflects how proteins may have evolved from a series of common ancestors. This tool can illuminate which sequences within a family are most related and distinguish close cousins from distant relatives. However, performing family-wide profiling requires an accurate large-scale sequence alignment, which can be challenging. Visualizing sequence relationships for family-wide profiling can be cumbersome with the methods previously outlined. Sequence similarity networks are visual tools that were developed to group protein sequences based on a similarity threshold. [29] Depending on this threshold for similarity, the sequences can be grouped based on their similarity, which can translate to their potential reactivity.

SSNs have proven useful for profiling other classes of FAD-dependent enzymes to identify biocatalysts appropriate for target-oriented synthesis. [24, 30] These examples showcase the synthetic utility of enzymes hiding in plain sight; the sequences are known,

but their reactivity will remain a mystery without dedicated experimental work toward family-wide reactivity profiling. These efforts are guided by tools for visualizing sequence space and have the potential to bring light to the deep seafloor of unexplored enzymes.

1.5   Dissertation Objectives

To develop the FDMOs as synthetic tools, we must define structural features that govern substrate binding, the critical catalytic residues, and how these features translate to other FDMOs. With this information, I will (1) modulate the site- and stereoselectivity of the biocatalyst as needed by altering a relatively small number of amino acids and (2) find and identify other biocatalysts to leverage for the manufacture of natural product building blocks. Such biocatalytic tools will provide an efficient route to valuable enantioenriched intermediates for the synthesis of bioactive molecules. Work toward this goal is outlined in Chapters 2 and 3.

In Chapter 2, I recently demonstrated that a flavin-dependent monooxygenase, TropB, maintains high levels of site- and stereoselectivity across a range of structurally diverse substrates. [22] I outlined the structural basis for substrate binding in TropB, which performs a synthetically challenging asymmetric oxidative dearomatization reaction with exquisite site- and stereoselectivity across a range of phenol substrates, providing a foundation for future protein engineering and reaction development efforts. Our hypothesis for substrate binding is informed by a crystal structure of TropB and molecular dynamics simulations with the corresponding computational TropB model and is supported by experimental data. Our data indicates that the phenolate form of the substrate binds in the active site. Furthermore, the substrate position is controlled through two-point binding of the phenolate oxygen to Arg206 and Tyr239, which are shown to have distinct and essential roles in catalysis. Arg206 is involved in the reduction of the flavin cofactor, suggesting a role in flavin dynamics. Further, QM/MM simulations reveal the interactions that govern the facial selectivity that leads to a highly enantioselective transformation.

Thus, the structural origins of the high levels of site- and stereoselectivity observed in reactions of TropB across a range of substrates are elucidated, providing a foundation for future protein engineering and reaction development efforts. [31]

In Chapter 3, the TropB findings are extrapolated to other class A FAD-dependent monooxygenases. I did a sequence profiling surrounding TropB to predict the potential. From this compendium of enzymes I selected an enzyme with predicted complementary stereoselectivity to that of TropB. A sequence analysis comparison of AfoD and TropB suggested that the residues critical for activity in TropB were in complementary locations in AfoD. I hypothesized that these residues are critical for the complementary stereoselectivity in AfoD. In this chapter, I crystallized AfoD, and through a sequence and structural analysis, then identified and probed the residues that might be critical for stereo- and site-selectivity. Using a sequence similarity network (SSN), several sequences have been selected based on their domain and selectivity. I initially chose seven enzymes to evaluate their reactivity and selectivity. Based on this approach, I identified enzymes with the complementary site- and stereoselectivity. [30] These findings showcase how sequence profiling coupled with experimental results can be leveraged to identify biocatalysts with interesting bioactivity. I predict that this approach can be used to canvas and identify new biocatalysts with a promiscuous substrate scope and complementary site and stereoselectivity.

<div align="center">

CHAPTER II

Structural Basis for Selectivity in Flavin-Dependent Monooxygenase-Catalyzed Oxidative Dearomatization

</div>

Summary

I recently demonstrated that a flavin-dependent monooxygenase, TropB, maintains high levels of site-and stereoselectivity across a range of structurally diverse substrates. In this chapter, I describe the structural basis for substrate binding in TropB, which performs a synthetically challenging asymmetric oxidative dearomatization reaction with exquisite site- and stereoselectivity across a range of resorcinol substrates, providing a foundation for future protein engineering and reaction development efforts. Our hypothesis for substrate binding is informed by a crystal structure of TropB and molecular dynamics sim-

ulations with the corresponding computational TropB model and is supported by experimental data. In contrast to canonical class A FAD-dependent monooxygenases in which substrates bind in a protonated form, the data indicate that the phenolate form of the substrate binds in the active site. Furthermore, the substrate position is controlled through two-point binding of the phenolate oxygen to Arg206 and Tyr239, which are shown to have distinct and essential roles in catalysis. Arg206 is involved in the reduction of the flavin cofactor, suggesting a role in flavin dynamics. Further, QM/MM simulations reveal the interactions that govern the facial selectivity that leads to a highly enantioselective transformation. Thus, the structural origins of the high levels of site- and stereoselectivity observed in reactions of TropB across a range of substrates are elucidated, providing a foundation for future protein engineering and reaction development efforts.

2.1 Introduction

The demand for biocatalytic transformations is growing as chemists recognize the advantages of these processes, including high levels of selectivity, low catalyst loadings, environmentally benign reagents, and excellent sustainability profiles. [7] For transformations that pose a challenge to traditional chemical synthesis, there is a greater interest in developing biocatalytic methods. The oxidative dearomatization of phenols to generate quinol products is a reaction that presents a substantial selectivity challenge using traditional chemical methods see **2.1** to **2.2**, Fig. 2.1. [32] A suite of I(III), I(V), Cu(I), and Pb(IV) reagents have been developed for oxidative dearomatization. [32–36] In these reactions, a phenolic substrate is activated as an electrophile, which is intercepted by an oxygen nucleophile in either an intra- or intermolecular fashion. [37–39] Achieving high levels of selectivity in this reaction is desirable as site-selectivity is often substrate- rather than catalyst-controlled. Further, controlling stereoselectivity under catalytic conditions remains a challenge. [33, 39–43]

For example, Maruoka and co-workers recently reported an Indane-based organoio-
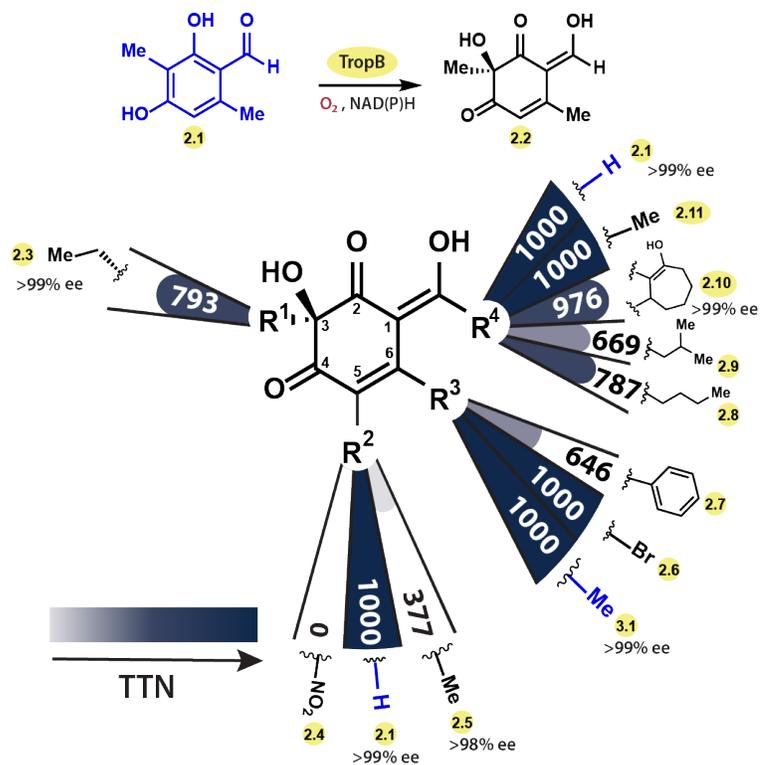
Figure 2.1: Chemistry of flavin-dependent monooxygenase TropB. Oxidative phenol dearomatization mediated by TropB in stipitatic acid biosynthesis (see **2.1** to **2.2**) and substrate scope. Non-native substrates with perturbation of a single substituent compared to the native substrate (**2.1**, blue) were evaluated for reactivity with TropB, measured in total turnover numbers (TTNs); maximum TTN = 1000 in this assay.

dine catalyst that provides p-quinol products in up to 84% enantiomeric excess (ee), the highest stereoselectivity for catalytic oxidative dearomatization reported to date. [43] Anticipating that a biocatalyst could provide exceptional selectivity and a superior sustainability profile, I recently investigated the substrate scope and scalability of a panel of flavin-dependent monooxygenases that mediate oxidative dearomatization reactions in various secondary metabolite pathways. [22] In contrast to traditional chemical transformations in which site-selectivity is governed by the substrate, selectivity in the flavin-dependent monooxygenase reaction is dictated by the orientation of the substrate in the active site relative to the oxidant, C4a-hydroperoxyflavin (**2.12**). [11, 14] I envisioned that a suite of flavin-dependent monooxygenases could provide complementary selectivity and redefine the state-of-the-art method for mediating this transformation. An initial study demonstrated that three biocatalysts, TropB, AzaH, and SorbC, perform well in preparative-scale reactions on a range of substrates while maintaining high levels of site- and stereoselectivity, providing a single product in 94% to >99% ee. [22] For example, TropB oxidizes non-native substrates such as **2.3**-**2.11**, which vary significantly in electronic and steric properties compared to the native TropB substrate **2.1**, yet for each substrate a single product is generated in ≥98% ee Fig. 2.1. [22]

I aim to enhance the synthetic utility of these biocatalysts by expanding the substrate scope, developing a suite of catalysts with complementary selectivities, and accessing non-native mechanistic pathways. However, to accomplish these goals and fully realize the synthetic potential of TropB and related enzymes, a structural and mechanistic understanding of this highly selective biocatalyst is required. I solved a crystal structure of TropB, which indicated that the active site is sufficiently spacious to accommodate a range of substrates but must maintain a conserved binding orientation to achieve high site- and stereoselectivity across an array of substrates. To provide a detailed understanding of the active site architecture, substrate binding pose, and residues required for catalysis, molecular dynamics simulations were carried out with a corresponding computational enzyme
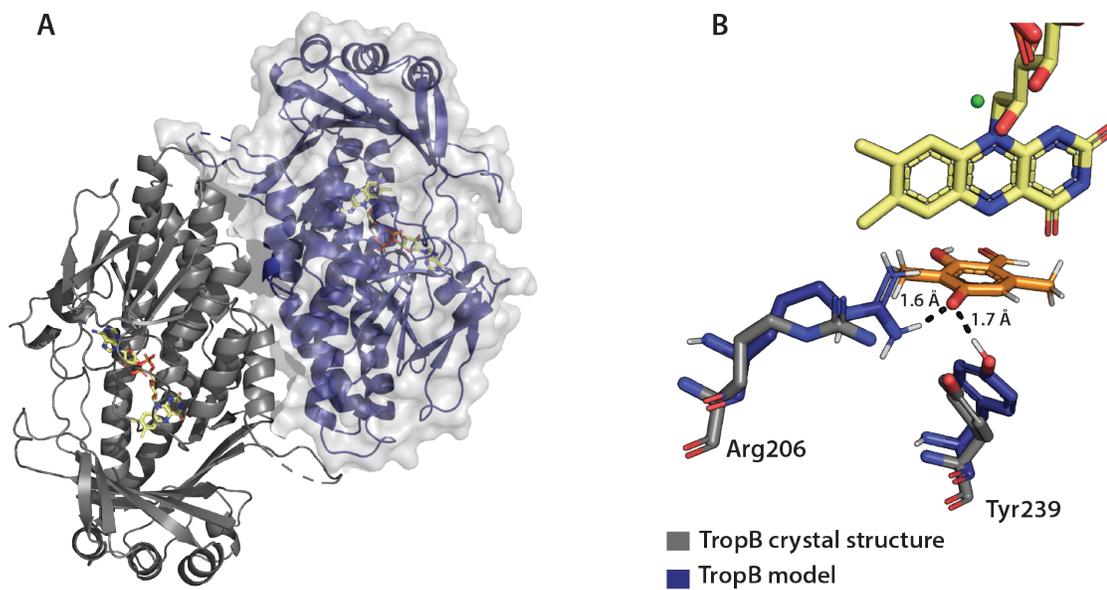
Figure 2.2: A) TropB dimer, with monomer 1 colored in indigo with disordered residues 81-86 shown as a dashed line and monomer 2 shown in gray; (B) TropB crystal structure overlaid with the TropB QM/MM model with substrate **2.1**.

model, and detailed characterization of wild type TropB alongside several variants revealed distinct roles for two amino acids in the active site.

2.2 Structure of TropB

As a first step toward elucidating the structural basis for selectivity, I solved a crystal structure of TropB to a resolution of 1.7 Å with its flavin adenine dinucleotide (FAD) cofactor bound Fig. 2.2 A and Table 2.2. TropB crystallized as a dimer, consistent with a dimeric form observed by size-exclusion chromatography Fig. 2.20. The large dimer interface (1244 Å^2 of buried surface area) includes 14 salt bridges and 6 additional hydrogen bonds. The overall structure comprises two domains: (a) FAD-binding domain 1, a typical class A flavin-dependent monooxygenase with a nucleotide binding site for the adenosine diphosphate (ADP) moiety of the cofactor, and (b) catalytic domain 2 with a binding site for the isoalloxazine ring Fig. 2.24A and Fig. 2.25A [11]. Part of the interdomain loop between α4 (domain 1) and β4 (domain 2) is disordered residues 81-86, Fig. 2.2A. Two long

helices (α11 and α12) also connect domains 1 and 2. A molecule of FAD sits in the narrow interdomain cleft, with a planar isoalloxazine ring that is consistent with the observed yellow color of crystals Fig. 2.24. FAD contacts the protein through a combination of hydrogen bonds and van der Waals interactions. The adenosine and pyrophosphate groups are bound to TropB as in other class A FAD-dependent monooxygenases. [44] The isoalloxazine ring sits in a binding pocket formed by Pro329, His330, and His331 with hydrogen bonds of O4 and N3 to Ala55, N1 to Ala335, and O2 to the Cys336 amide. A chloride ion sits in a tight pocket between the isoalloxazine ring and Pro329, Gly332, and Gly334, as in other FAD class A monooxygenases Fig. 2.24A. [13, 45, 46]

In the structure of TropB crystallized in the absence of substrate, a large cavity adjacent to the isoalloxazine ring is anticipated to be the site of substrate binding. This site is defined by several side chains (Arg206, Leu226, Met228, His235, Ile237, Tyr239, Ala250, Phe252, Pro329, His330, His331, and Tyr397), some of which are conserved in other class A flavin monooxygenases. [45] One molecule of hexanediol, a component of the crystallization solution, was bound in the cavity, based on well-defined electron density Fig. 2.3.
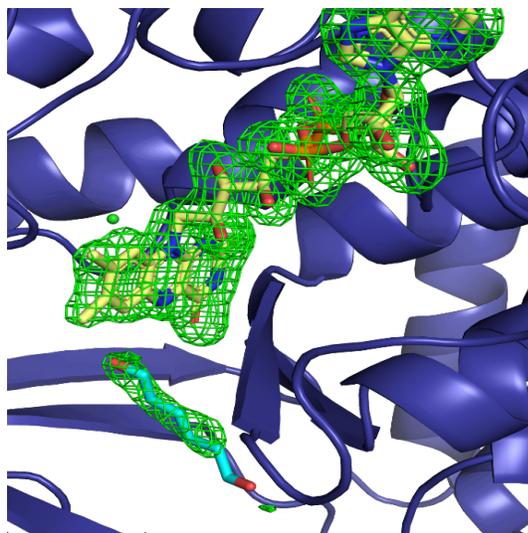
Figure 2.3: Bound FAD with omit electron density (green contours at 3.0 σ)

As crystallization with the flavin cofactor in the oxidized state should permit formation

of a substrate complex that cannot further react in the crystal, significant effort was dedicated to obtaining crystals of a TropB ternary complex with FAD and the native substrate, 3-methylorcinaldehyde (**2.1**). Crystal-soaking and cocrystallization experiments with **2.1** or structural analogues of **2.1** resulted in maps with new electron density in the active site cavity; however, despite extensive experimentation, none of the conditions explored led to well-defined density for a single orientation of the substrate or to the full displacement of hexanediol from the cavity. Although full occupancy of a TropB ternary complex was not captured, crystallization in the presence of **2.1** induced a change in position of the cofactor isoalloxazine ring, consistent with flavin dynamics of class A flavin-dependent monooxygenases Fig. 2.4. [14, 15] Substrate binding in class A flavin-dependent monooxygenases triggers flavin reduction by inducing a "flavin-out" conformation, in which the flavin is accessible to the reduced cofactor, nicotinamide adenine dinucleotide phosphate (NADPH) or nicotinamide adenine dinucleotide (NADH). [15, 47, 48] The TropB crystal structure obtained in the absence of substrate depicts the oxidized cofactor in the "flavin-in" conformation, whereas the structure in the presence of substrate indicated partial occupancy of a flavin-out conformation Fig. 2.4. This change in flavin position upon substrate binding would allow for reduction of FAD to $FADH_2$ by an external reductant (NADH or NADPH), consistent with other class A flavin-dependent monooxygenases. [49, 50]
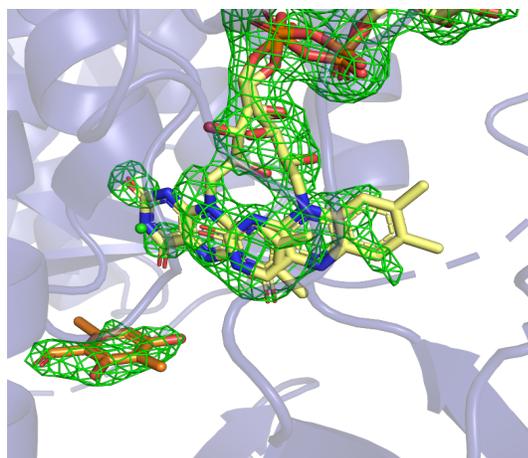
Figure 2.4: Electron density for flavin in the "in" and "out" conformations. As in other class A FAD-dependent monooxygenases, [15, 47, 48] substrate binding triggers flavin reduction by inducing a "flavin-out" conformation in which the flavin is accessible to the reduced nicotinamide adenine dinucleoide phosphate (NADPH). Omit electron density (green contours at 2.5 σ) is shown for FAD and substrate **2.1**.))

2.3   Substrate Binding and Catalysis

Arg206 and Tyr239 are the only amino acids in the TropB active site that possess relevant functional groups and are near the flavin cofactor and partially occupied substrate site. In the absence of clear density to illustrate the contacts of the native substrate (**2.1**), I sought to draw an analogy to substrate complexes of other class A flavin-dependent monooxygenases of known structure. TropB has a low sequence identity to other class A flavin-dependent monooxygenases, except for SalH (31%); [44] nonetheless, structural features involved in either substrate binding or catalysis remain constant across enzymes in this class Figs. 2.14 to 2.15. In *p*-hydroxybenzoate hydroxylase (PHBH) a hydrogen bond network composed of Tyr201, Tyr385, His72, and a conserved water molecule is responsible for substrate binding and further catalysis. [14] This network is absent in TropB; however, Tyr239 and Arg206 are within hydrogen bonding distance of the substrate in the TropB active site. Additionally, the Arg206 guanidinium group is in a position analogous to the Arg220 guanidinium group in PHBH. This is also the case for other class A monooxygenases, phenol 2-monooxygenase (PHHY) (Tyr289, Arg287), [52]
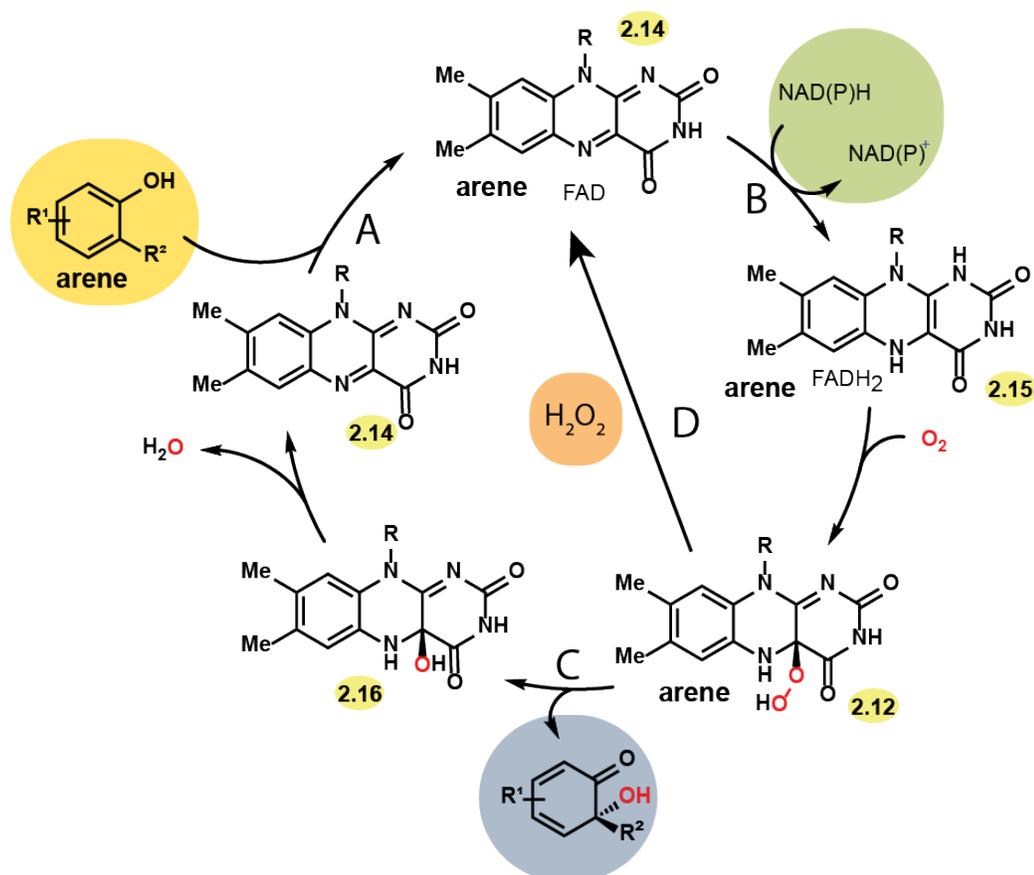
Variant	Product Formation	Specific Activity (umol min⁻¹ mg⁻¹)	Kd (uM)	NADPH Consumed	H₂O₂ formed in the reaction
WT	Yes	0.33	21 ± 2	Yes	0.2 mM
Y239F	No	0	> 300	Yes	1.4 mM
R206Q	No	0	32 ± 4	No	-
R206E	No	0	> 100	No	-

Figure 2.5: Activity of wild type TropB and TropB variants. (A) Substrate **2.1** binding to wild-type TropB; (B) NADPH depletion upon binding of **2.1**; (C) product **2.2** formation; (D) formation of peroxide in shunt pathway. *Reaction conditions: 2.5 mM substrate **2.1**, 2.5 μM TropB variant, 1 mM NADP+, 5 mM glucose-6-phosphate (G6P), 1 U ml⁻¹ glucose-6-phophate dehydrogenase (G6PDH), 50 mM potassium phosphate buffer pH 8.0, 30 °C, 1 h. 0% conversion indicates that no substrate consumption or product formation was detected. **Reaction Conditions: 100 nM TropB variant, 250 μM NADPH, 125 μM 1 monitored at 362 nm over 1000 s. ***Concentration of hydrogen peroxide was measured using an established assay. [51]

3-hydroxybenzoate hydroxylase (MHBH) (Tyr271, Arg262), [53] and Salicylate hydroxylase (SalH) (Arg190). [44] This led us to hypothesize that Tyr239 and Arg206 might be involved in substrate binding and perhaps catalysis.

A set of TropB variants was generated to probe the role of Tyr239 and Arg206 in substrate binding. Arg206 was substituted with either Glu (R206E) or Gln (R206Q) and Tyr239 was replaced with Phe (Y239F) to probe the role of hydrogen bonding in substrate binding 2.2, Fig. 2.18, and Fig. 2.21. TropB R206E, TropB R206Q, and TropB Y239F produced no detectable product by UPLC-PDA analysis under conditions where the wild type TropB led to complete consumption of the substrate with a specific activity of 0.33 μmol $min^{-1}mg^{-1}$ Figs. 2.5 and 2.19 - 2.25, and 2.35. These results indicate that Arg206 and Tyr239 each have a critical role in achieving productive catalysis.

Having established that TropB R206E, TropB R206Q, and TropB Y239F produce no detectable product with the native substrate, 3-methylorcinaldehyde (**2.1**), we sought to identify which step of the catalytic cycle was disrupted. TropB binding of **2.1** was detected by the change in absorbance at 492 nm, revealing a K_d of 21 ± 2 μM for wild-type TropB and 32 ± 4 μM for TropB R206Q Figs. 2.5 and 2.36-2.38. Substituting Arg206 or Tyr239 with an amino acid that cannot act as a hydrogen bond donor led to a significant increase in K_d values, suggesting that hydrogen bond donors at positions 206 and 239 are critical for substrate binding. Next, we investigated whether NADPH can reduce FAD to FADH- in each TropB variant in the presence of substrate by monitoring NADPH absorbance at 362 nm. Complete consumption of NADPH was observed in 800 s with wild type TropB, and partial consumption of NADPH was observed with TropB Y239F over 1000 s Fig. 2.5 and Fig.2.42. Additionally, the reduction rate of TropB Y239F was significantly slower compared to that of the wild-type, but no product formation was detected in Fig. 2.43. However, no depletion of NADPH was detected for TropB R206Q or R206E. Thus, we increased the enzyme concentration by 200-fold, and under forcing conditions we observed partial consumption of NADPH at extended time periods Fig. 2.44. [51] Previously Ballou

and co-workers demonstrated that the corresponding Arg residue (Arg220) in PHBH is involved in the flavin dynamics required for FAD to toggle from the "in" conformation to the "out" conformation for reduction to FADH- by stabilizing the "out" conformation. [47]

To test the ability of each TropB variant to react with molecular oxygen to form the labile C4a- hydroperoxyflavin intermediate **2.12**, an intermediate corroborated by stopped-flow analysis in Fig. 2.47, the hydrogen peroxide concentration was measured in samples containing each TropB variant with 2.5 mM substrate and NADPH recycling system, generating a readout on the level of peroxide formed through uncoupling see **2.12** to **2.14**, Fig. 2.5D. In this assay, relatively low levels of H_2O_2 were measured (0.2 mM) with wild type TropB, whereas, significantly higher concentrations of H_2O_2 were present in reactions with TropB Y239F (1.4 mM), indicating a higher level of uncoupling in the *in vitro* reaction. In reactions with TropB (R206E and R206Q), trace amounts of H_2O_2 were detected (0.03 mM), which did not increase over time, further indicating that these substitutions hindered flavin reduction Fig. 2.46. Characterization of the Arg206 and Tyr239 variants support the role of these two amino acids in substrate binding. These results indicate that Arg206 is also important for FAD reduction to $FADH_2$ and that Tyr239 stabilizes the substrate in a position suitable for rapid reaction with labile C4a-hydroperoxyflavin. Additionally, TropB R206Q and TropB Y239F were crystallized. In the TropB Y239F structure, Phe239 maintains the same position as Tyr239 in the wild-type protein Fig. 2.27, 2.1. In the crystal structure of TropB R206Q, Gln206 is ordered and unavailable to stabilize FAD in the "out" conformation, thus, affecting flavin reduction in the catalytic cycle Fig. 2.26, 2.1).

Next, we considered which additional residues could be critical for a productive reaction. Deprotonation is a common step for other class A flavin-dependent monooxygenases, and it is generally accepted that this class of monooxygenases requires a catalytic base for deprotonation of phenolic substrates in order to obtain a sufficiently nucleophilic reactant. [11, 45, 54] Deprotonation is thought to occur prior to flavin reduction or during substrate hydroxylation. [14] In the case of TropB, His235 has been predicted to act as
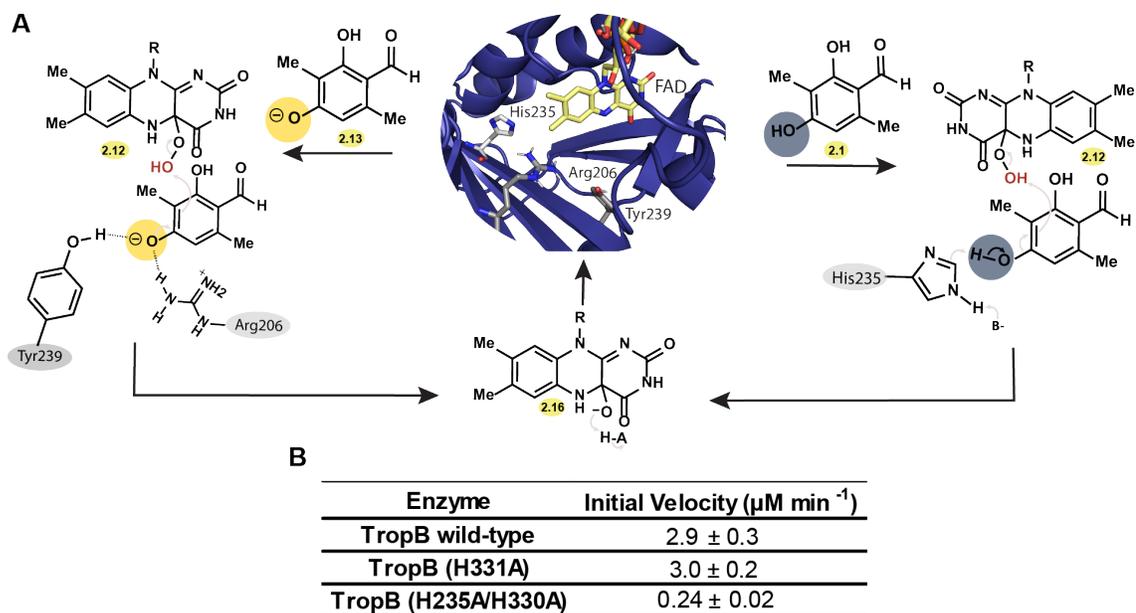
Figure 2.6: Hypotheses for substrate activation and positioning. (A) The mechanism on the right requires deprotonation of the substrate in the active site, which has been proposed to be performed by His235. [55] The mechanism proposed on the left relies on binding of the anionic form of the substrate, including interactions with Arg206 and Tyr239 consistent with our experimental findings. The C4a-hydroperoxoflavin O_{distal} is highlighted in red. (B) Initial velocity rates for TropB wild-type and the histidine variants.

a catalytic base; [55] however, the crystal structure indicates that His235 is too far from the substrate to perform a deprotonation. The apparent lack of a basic residue near the anticipated site of substrate binding led us to hypothesize that the substrate binds as the phenolate, eliminating the requirement a base to initiate catalysis as previously suggested Fig. 2.6 A. [56, 57]

To test this hypothesis, we first investigated the protonation state of the substrate under the reaction conditions, specifically the optimum pH 8.28. The protonation state of the native substrate (**2.1**) was monitored by UV absorbance, as the phenolate absorbs at 336 nm, whereas, the protonated form is characterized by a maximum absorbance at 292 nm in Fig. 2.40 - 2.41. This indicates that, under the reaction conditions at pH 8, the predominant form of the substrate is the anionic form (**2.13**) with an experimentally determined pKa of 7.1. To experimentally test the proposal that H235 is necessary for catalysis Fig.

2.6A, [55] TropB H235A was generated in Fig. 2.19. Substitution of His235 with Ala produced a catalytically active TropB variant with 10% conversion of substrate to product, Figs. 2.7 - 2.8.
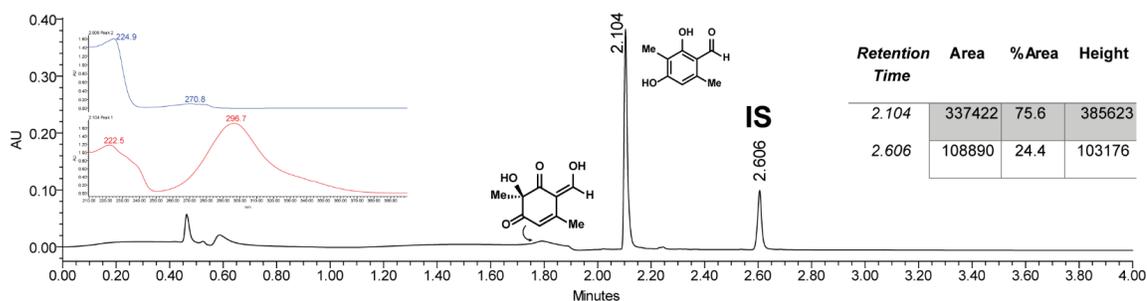
Figure 2.7: Reaction of **2.1** and TropB H235A. PDA traces of enzymatic reaction.
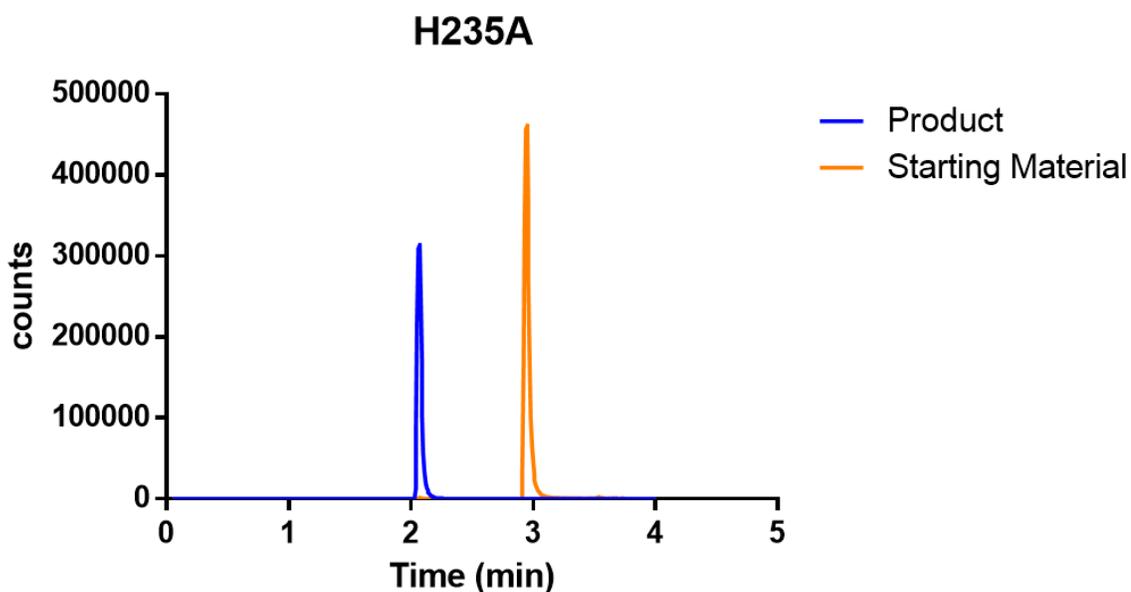
Figure 2.8: Reaction of **2.1** and TropB H235A. LC-MS trace.

To probe the role of neighboring His residues, His330 and His331 were substituted with Ala, affording an active variant capable of 11% and 71% conversion of substrate to dearomatized product, respectively, Figs. 2.9 to 2.12. Since these residues are close together, one could imagine His235 and His330 having a redundant function as the requisite base. Thus, a double substitution was performed to generate H235A/H330A, and the variant remained

active 10% conversion, Figs. 2.13 and 2.14.
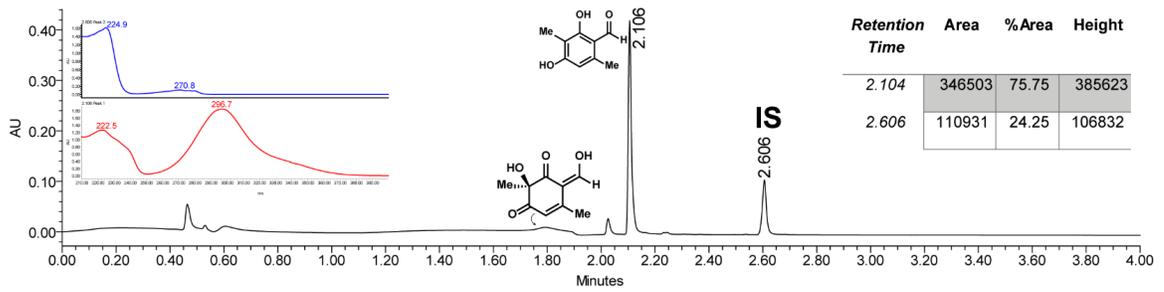
Figure 2.9: Reaction of **2.1** and TropB H330A. PDA traces of enzymatic reaction.
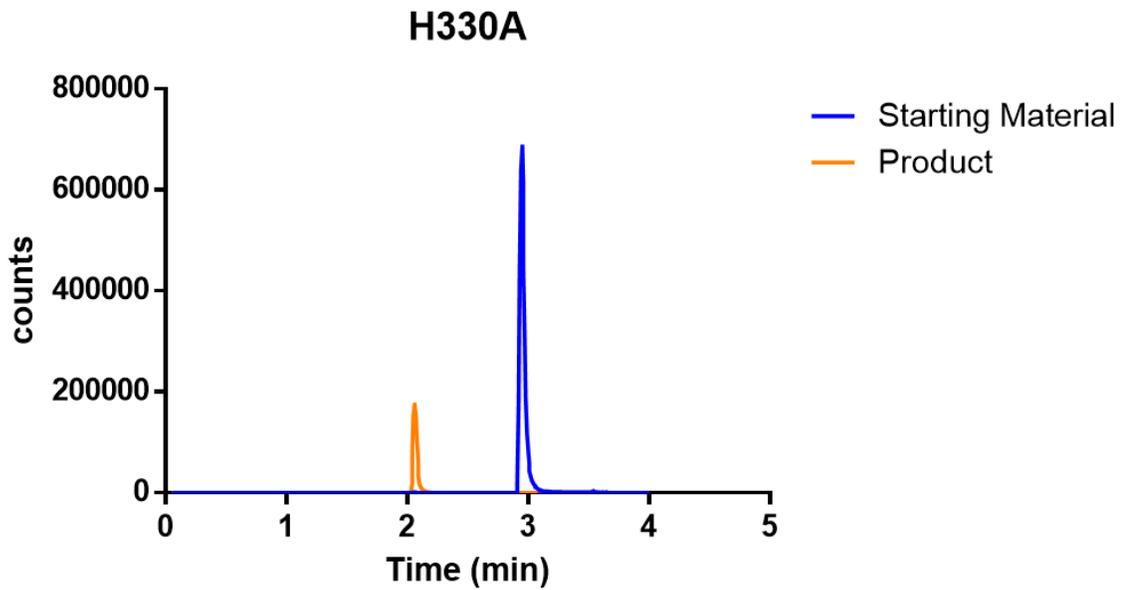
Figure 2.10: Reaction of **2.1** and TropB H330A. LC-MS trace.
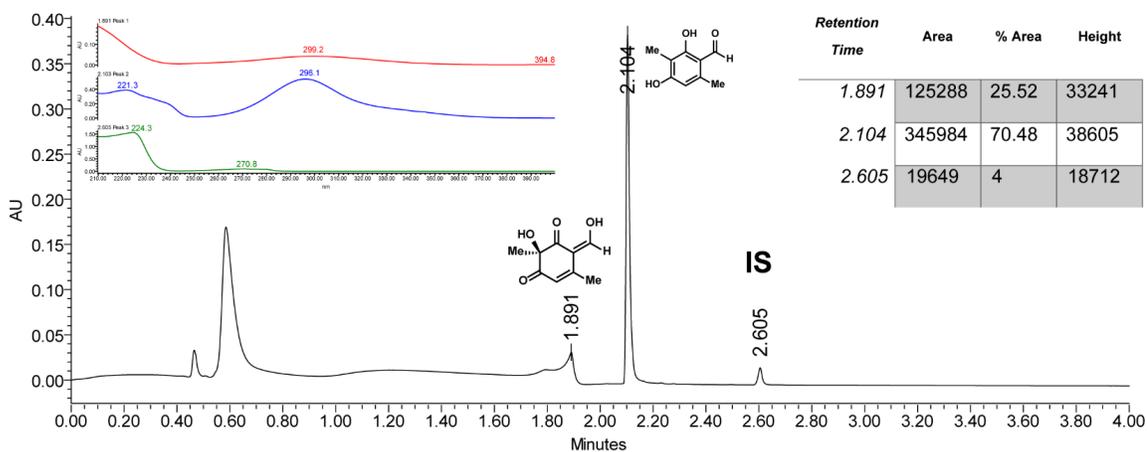
Figure 2.11: Reaction of **2.1** and TropB H331A. PDA traces of enzymatic reaction.
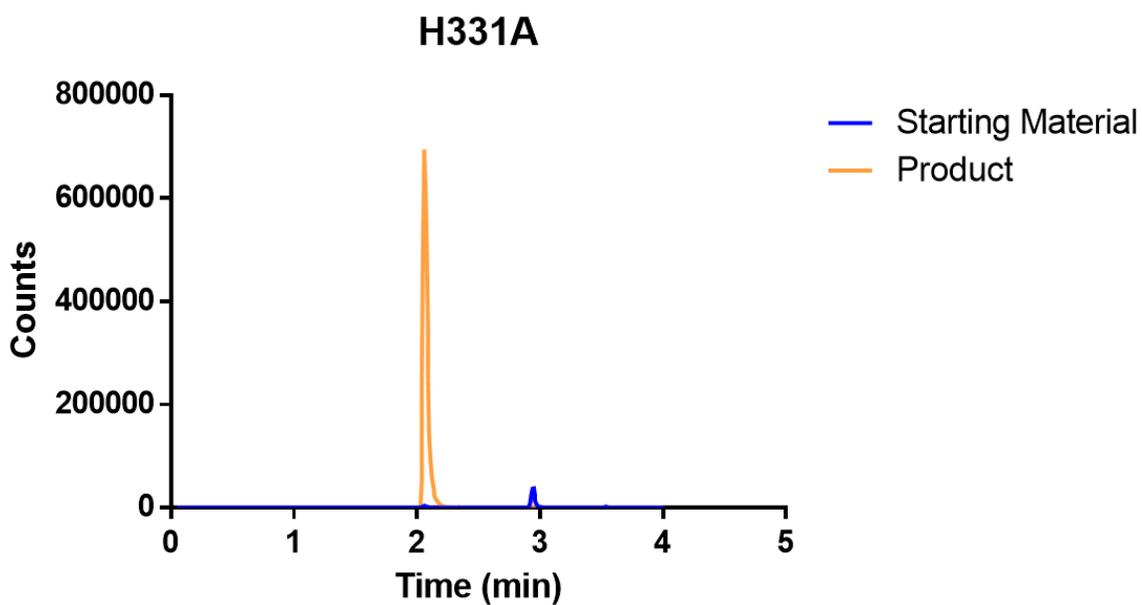
Figure 2.12: Reaction of **2.1** and TropB H331A. LC-MS trace.
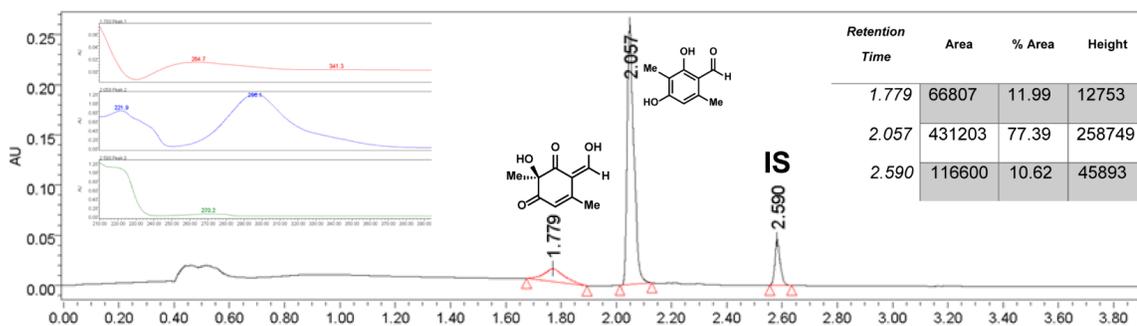
Figure 2.13: Reaction of **2.1** and TropB H235A/H330. PDA traces of enzymatic reaction.
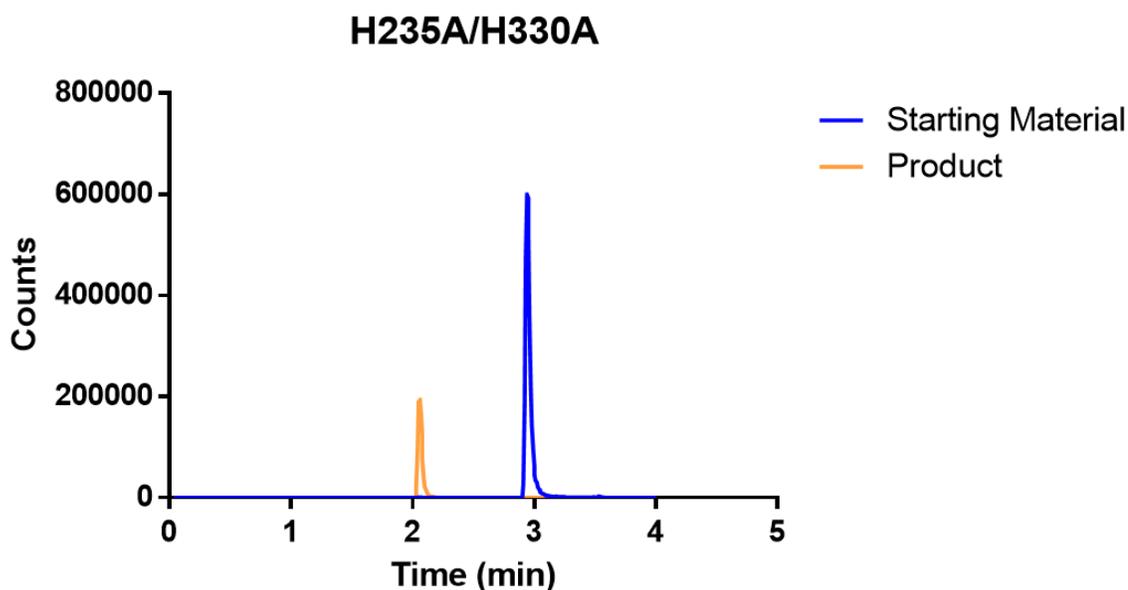
Figure 2.14: Reaction of **2.1** and TropB H235A/H330. LC-MS trace.

These results are consistent with substrate binding in the anionic form, eliminating the need for a basic residue in the active site for substrate deprotonation. Further, the initial rates of reactions with **2.1** and wild-type TropB, TropB H331A, or TropB H235A/H330A follow the reactivity trend with wild type TropB and TropB H331 having initial velocities of 2.9 and 3.0 μM min^{-1}, respectively Fig. 2.6B. The TropB H235A/H330A variant has a decreased initial velocity, 0.24 μM min^{-1}, that corresponds with the lower conversions observed in reactions with this variant Fig. 2.48. These data indicate that the His residues are not required for catalysis but they contribute to an ideal active site environment.

2.4 Computational Approaches To Elucidate the Substrate Binding Pose

To provide additional insight into the orientation of the native substrate **1** in the active site, we turned to computational analysis Experiments and analysis in this section was performed by collaborators Dr. Troy Wymore and Sara Tweedy at Dr. Charles Brooks

III lab at the University of Michigan, Ann Arbor. Several enzyme-substrate models with the flavin cofactor as either FAD (**2.14**) or C4a-hydroperoxyflavin (**2.12**) were evaluated with ligand-protein docking calculations that use the CDocker protocol. [58] The protocol employs (i) flexible active site side chains and a soft-grid representation of the remaining receptor during multiple rounds of simulated annealing followed by (ii) geometry optimizations with the CHARMM molecular mechanical force field and (iii) scoring of the complexes with the force field and an implicit solvent model. In addition, molecular dynamics (MD) simulations were performed using the docked complexes in explicit solvent using classical force fields and hybrid quantum quantum mechanical/molecular mechanical (QM/MM) methods. Flexible docking of **2.1** in the anionic form (see **2.13**) to TropB with the flavin cofactor in either the oxidized or hydroperoxy form led to a favorable, low-energy pose consistent with the known site- and stereoselectivity of the reaction. This docking indicated that both Arg206 and Tyr239 hydrogen bond with the phenolate oxygen of **2.13** Fig. 2.2B and Fig. 2.6A. The pose showed a significant change in the location of the Arg206 side chain from its position in the crystal structure, while the movement of Tyr239 was minimal Fig. 2.2B and Fig. 2.49. The displacement of Arg206 is reasonable given that the enzyme was crystallized with the flavin in the oxidized "in" state and must go through several catalytic steps before reaching the C4a-hydroperoxyflavin state, (**12**) Fig. 2.5. The charge distribution in the molecular mechanics (MM) force field delocalized the negative charge on the substrate in accordance with the gas phase and solvated QM calculations and was not concentrated on the phenolate oxygen. This indicates that the interaction with Arg206 and Tyr239 was not driven by an unphysical electrostatic interaction that could arise if the charges were overly localized by standard charge distribution methods.

To characterize enzyme-substrate contacts, several MD simulations of the identified docking pose were performed. Specifically, models with different tautomeric states of His235, His330, and His331 were constructed since the tautomeric states of His residues

are not directly observed in crystal structures, and these states can be challenging to accurately predict. During each of the 25 ns simulations, the substrate exhibited only minimal fluctuations (0.24 to 0.65 Å root-mean-square deviation (rmsd) from average atomic position), remaining in a bound position (averaging 5.2 to 6.8 Å for the O_{distal}-C3 distance, Figs. 2.53 - 2.55 and 2.3. A reactive configuration was observed in 1.3% of the snapshots as strictly defined by (1) an O_{distal}-C3 distance less than 4 Å and (**2.2**) an O_{distal}-C3-C3$_{methyl}$ angle between 60° and 115° Fig. 2.54. All of these snapshots showed a hydrogen bond between the anionic substrate **2.13** and Arg206, and approximately half of the snapshots also showed a hydrogen bond with Tyr239, supporting the relevance of this interaction. [59, 60] One of the Arg206 terminal amine groups hydrogen bonded to the phenolate oxygen throughout the simulation (average O-H distance 2.0 Å, 88% of frames <2.4 Å). Interestingly, the phenolic proton of Tyr239 formed a tighter hydrogen bond as the simulation progressed. The average distance for this hydrogen bond was 2.7 Å over 24 ns but averaged 2.0 Å in the final 10 ns with 90% of the snapshots showing a distance less than 2.4 Å Fig. 2.57. Thus, MD simulations with a classical force field suggest that both Arg206 and Tyr239 form hydrogen bonding interactions with the substrate. To investigate the interactions between the substrate and the active site environment with a more accurate description, MD simulations were also performed using a QM/MM representation. [61] These simulations were performed by the Dr. Charles Brooks III lab specifically Dr. Troy Wymore and Sara Tweedy. In DFTB3 simulations, water molecules adorn the polar region of the active site, separating His235 from the substrate. Thus, neither a direct nor Grotthaus-like mechanism for substrate activation by His235 would be feasible. The substrate maintained hydrogen bonding interactions with both Tyr239 and Arg206 throughout these simulations.

The combined knowledge from the TropB crystal structure and subsequent MD suggests that the anionic form of the substrate, phenolate **2.13**, is the predominant form under the reaction conditions and that the negative charge is stabilized through interactions

with both Arg206 and Tyr239, thus positioning the substrate for a productive reaction with the C4a-hydroperoxyflavin cofactor **2.12**,Fig. 2.6A) and leading to a highly site- and stereoselective transformation.

2.5   Computational Insights on Facial Selectivity

To identify the factors controlling facial selectivity we turned to computation. Experiments and analysis in this section was performed by collaborators Dr. Troy Wymore and Sara Tweedy at Dr. Charles Brooks III lab at the University of Michigan, Ann Arbor. While QM/MM simulations indicated that hydrogen bonding of Arg206 and Tyr239 with the phenolate form of the substrate is the dominant intermolecular interaction Fig. 2.58, these interactions alone are insufficient to be the basis of for the facial selectivity observed experimentally. Additional interactions between substrate and TropB were observed in MD simulations. Both hydrophobic C3 and C6 methyl groups of substrate **2.1** interacted with hydrophobic residues Fig. 2.58. The C6 methyl group was near Ala55 and the phenyl group of Phe119. In contrast, the C2 hydroxyl group was positioned in a solvated environment, hydrogen bonding exclusively with the aldehyde moiety Fig. 2.61 - 2.61. If the opposite face of the substrate were presented to the C4a-hydroperoxyflavin factor (**2.12**) by rotating along the C1- C4 axis, the C2 hydroxyl group would occupy an unsuitable hydrophobic region and the C6 methyl group would be in a hydrophilic space, placing these two groups in uncomplementary environments Fig. 2.15A and Fig. 2.63. Similarly, the C3 methyl group makes hydro-phobic contacts with Phe252, Ala250, and Ile237. If the substrate rotated about the C1-C4 axis, however, this methyl group would sit near the backbone carbonyl of Gly53 and the hydroperoxy group of the flavin Fig. 2.62. Thus, this region would be an energetically unfavorable position for the hydrophobic C3 methyl group if the substrate presented the opposite face to the cofactor. The bulkier methyl group would also clash with Tyr239, disrupting the hydrogen bond with the substrate. Thus, we conclude that the facial selectivity in the hydroxylation of substrate 1 in TropB arises from
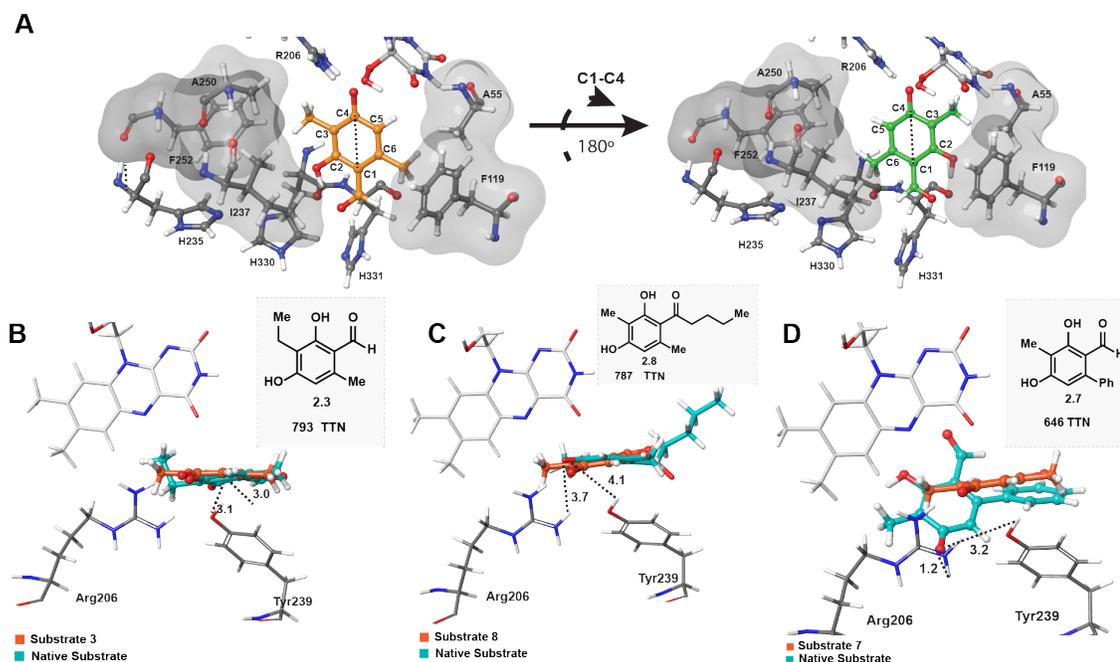
Figure 2.15: QM/MM simulation of the native substrate (**2.1**) in the active site to illustrate facial selectivity. (A) Left: Substrate **2.1** (orange) bound in the active site, showing favorable contacts with Ala250 and Ile237. Right: 180° degree rotation along the C1-C4 axis of **2.1** (green), illustrating unfavorable substrate orientation. (B-D) Substrate docking of **2.3**, **2.8**, and **2.7**, respectively, with the correct face presented to the flavin.

the amphipathic active site environment as illustrated in Fig. 2.15A. Docking of non-native substrates revealed binding poses analogous to native substrate **2.1** Fig. 2.15B and D and Fig. 2.50 - 2.51. For example, substrates **2.3** and **2.8** adopt a similar pose to native substrate **2.1**, occupying the same space in the TropB active site and maintaining hydrogen bond interactions with Arg206 and Tyr239 Fig. 2.15B, C, respectively. Substrates with significantly more sterically demanding substituents such as **2.7** provide docking solutions with the substrate slightly shifted in the active site to accommodate bulky substituents but maintain the relevant hydrogen bonding contacts and facial selectivity as observed for the native substrate Fig. 2.15D. These docking results are consistent with the site- and stereoselectivity observed in reactions with non-native substrates. [7]

2.6   Conclusion

Through structural insights gained from both the crystal structure and molecular dynamics simulations, we have uncovered the origins of the exquisite site- and stereoselectivity exhibited in reactions of TropB with a breadth of substrates. A crystal structure of TropB has enabled computational studies that support binding of the anionic form of the substrate, which is further supported by experimental data. Molecular dynamics simulations and characterization of TropB variants demonstrated that the phenolate substrate has significant interactions with both Arg206 and Tyr239, controlling the pose of the substrate within the active site and the selectivity observed in subsequent reactions. The excellent stereo-selectivity observed in reactions with TropB can be attributed to pronounced facial selectivity in substrate binding. These findings clearly illustrate the structural features that control the site- and stereoselectivity in the complexity-generating reaction mediated by TropB, critical for future protein engineering and reaction development efforts with TropB and related enzymes.

2.7   Experimental

I. Cloning, protein expression and purification

General Considerations

Escherichia coli (E. coli) cloning strains DH5α (Invitrogen) were used for DNA propagation. Phusion HF polymerase was purchased from New England BioLabs. All primers were purchased from Integrated DNA Technologies (IDT) ddH2O was sourced from a MilliQ Biocel water purification unit from Millipore.

Sequence Information

A gene encoding TropB (DAA64700.1) was codon-optimized for overexpression in *E. coli* synthesized by GeneArt (Thermofisher). GeneArt cloned the synthesized sequence

into pET151 vector conferring the T7 expression system, ampicillin resistance, and an N-terminal 6x His tag encoded upstream from the inserted gene. No further modification to the construct was necessary.

Codon-Optimized tropB Sequence

Codon-Optimized TropB Sequence

```
ATGCCTGGTAGCCTGATTGATACCCGTCAGCAGCCGCTGAGCGTTGGTATTGTTGGTGGTGGTATTATTG
GCGTTATTCTGGCAGCAGGTCTGGTTCGTCGTGGTATTGATGTTAAAGTTTTTGAACAGGCACGTGGCTTT
CGTGAAATTGGTGCAGGTATGGCATTTACCGCAAATGCAGTTCGTTGTATGGAAATGCTGGATCCGGCAA
TTGTTTGGGCACTGCGTAGCAGCGGTGCAGTTCCGATTAGCATTGGTGATCATCAGGCCGAAGCACGTGA
TTATCTGCGTTGGGTTGATGGTTATCATGAAAGCAGCAAACGTCTGTATCAGCTGGATGCAGGTATTCGTG
GTTTTGAAGCATGTCGTCGTGATCAGTTTCTGGAAGCACTGGTTAAAGTTCTGCCGGAAGGTATTGTGGAA
TGTCAGAAACGTCTGCAGAAAATCCACGAAAAAAACGAAACCGAAAAAGTGACCCTGGAATTTGCAGATG
GCACCTTTGCACATGTTGATTGTGTTATTGGTGCCGATGGTATTCGTAGCCGTGTTCGTCAGCACCTGTTT
GGTGAAGATAGCCCGTATAGCCATCCGCATTATAGCCATAAATTTGCATTTCGTGGTCTGATCACCATGGA
AAATGCAATTAGCGCACTGGGCGAAGATAAAGCACGTACCCTGAATATGCATGTTGGTCCGAATGCACAT
CTGATTCATTATCCGGTTGCAAATGAAACCATGGTGAATATTGCAGCCTTTGTTAGCGATCCGGAAGAATG
GCCTGATAAACTGAGCCTGGTTGGTCCGGCAACCCGTGAAGAAGCAATGGGTTATTTTGCAAATTGGAAT
CCGGGTCTGCGTGCAGTTCTGGGTTTTATGCCGGAAAATATTGATCGTTGGGCAATGTTCGATACCTATG
ATTATCCGGCACCGTTTTTTAGCCGTGGTAAAATTTGTCTGGTTGGTGATGCAGCACATGCAGCAGTTCCG
CATCATGGTGCCGGTGCATGTATTGGTATTGAAGATGCACTGTGTGCAACCGTTCTGCTGGCAGAAGTTT
TTGTTAGCACCCGTGGCAAAAGCAGCATTGTTCGTAATCGTGCAATTGCCGCAGCATTTGGTAGCTTTAAT
GCAGTGCGTCGTGTTCGTGCACAGTGGTTTGTTGATAGCAGCCGTCGTGTTTGTGATCTGTATCAACAGC
CGGAATGGGCAGATCCGCAGAAACGTATTAAAGCCGAAATTGCTTCGAAGAGATTAAAGATCGCAGCCA
TAAAATCTGGCACTTCGATTATAACTCCATGCTGCAAGAAGCCATCGAAAAATATCGTCATAATATGGGCA
GCTAA
```

Figure 2.16: Codon-Optimized tropB Sequence [31]

TropB Protein Sequence

TropB Protein Sequence

```
MPGSLIDTRQQPLSVGIVGGGIIGVILAAGLVRRGIDVKVFEQARGFREIGAGMAFTANAVRCMEMLDPAIVWAL
RSSGAVPISIGDHQAEARDYLRWVDGYHESSKRLYQLDAGIRGFEACRRDQFLEALVKVLPEGIVECQKRLQKI
HEKNETEKVTLEFADGTFAHVDCVIGADGIRSRVRQHLFGEDSPYSHPHYSHKFAFRGLITMENAISALGEDKA
RTLNMHVGPNAHLIHYPVANETMVNIAAFVSDPEEWPDKLSLVGPATREEAMGYFANWNPGLRAVLGFMPEN
IDRWAMFDTYDYPAPFFSRGKICLVGDAAHAAVPHHGAGACIGIEDALCATVLLAEVFVSTRGKSSIVRNRAIAA
AFGSFNAVRRVRAQWFVDSSRRVCDLYQQPEWADPQKRIKAENCFEEIKDRSHKIWHFDYNSMLQEAIEKYR
HNMGS
```

Figure 2.17: TropB Protein Sequence [31]

Site-directed mutagenesis

TropB(Y239F)	pARB_8	GCACATCTGATTCAT**TTT**CCGGTTGCAAATG
TropB(R206Q)	pARB_25	CATAAATTTGCATTT**CAG**GGTCTGATCACCATGG
TropB(R206E)	pARB_26	CATAAATTTGCATTT**GAG**GGTCTGATCACCATGG
TropB(H235A)	pARB_27	GTTGGTCCGAATGCA**GCG**CTGATTCATTATCCG
TropB(H331A)	pARB_1	GCAGCAGTTCCGCAT**GCT**GGTGCCGGTGCATG
TropB(H330A)	pARB_40	CATGCAGCAGTTCCG**GCG**CATGGTGCCGGTGC

Note: All sequences are 5'-3'. Bold red font indicates respective point mutation performed

Table 2.1: Primer Sequences

TropB(Y239F), TropB(R206Q), TropB(R206E), TropB(R206K), TropB(H235A), TropB(H331A), TropB(H330A), and TropB(H235A/H330A). Substitutions were generated by site-directed mutagenesis on pET151-tropB(WT). 25 μL PCR reaction mixtures contained 5 μL of 5X Phusion HF buffer, 1 ng/μL WT parent plasmid, 0.5 μM of primer, 200 μM dNTPs, 0.5 U μL^{-1} Phusion HF. Amplification was accomplished with the following PCR protocol: 98 °C for 0:30 s, (98 °C 0:10, Tm -5 °C 0:30 s, 72 °C 0:30/kb) for 25 cycles, with a final extension of 72 °C for 10:00 min. This was followed by a 10 μL digestion containing 1 μL of NEB CutSmart buffer, 8 μL of PCR mixture and 20 units of DpnI. The reaction was incubated at 37 °C for 3 h and transformed into chemically competent E. coli DH5α cells.

Protein Expression and Purification

Protein overexpression: TropB WT, Y238F, R206Q, R206E, H331A, H330A, H235A/H330A: Plasmids were transformed into E. coli strain BL21(DE3). 500 mL of Terrific Broth (TB) containing 100 μg mL^{-1} ampicillin was inoculated with 5 mL overnight culture prepared from a single colony in Luria Broth (LB) and 100 μg mL^{-1} ampicillin. Culture was grown at 37 °C and 250 rpm to an OD600 = 0.8. The culture was then cooled to 18 °C for 1 h at 200 rpm and protein expression was induced with 0.1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) and expressed at 20 °C for 18 h at 200 rpm. After overnight expression, cultures were centrifuged at 13,881 x g for 30 min. Cell pellets from overexpression were stored at

-80 °C for long-term storage.

General purification procedure: Cell pellets from overexpression were resuspended in 40 mL of lysis buffer (50 mM Tris:HCl pH 7.8, 300 NaCl, 10 mM imidazole, and 10% (v/v) glycerol) with 0.1 mg mL^{-1} lysozyme, 0.05 mg mL^{-1} DNase, and 0.1 mM flavin adenine dinucleotide (FAD), incubated on a rocker at 4 °C for 45 min, and lysed by sonication. Insoluble material was removed by centrifugation (46,413 x g for 30 min). Supernatant was sterile filtered through a 0.45 μM syringe filter. The clarified lysate was loaded onto a 5 mL HisTrap column (GE Healthcare) on a Healthcare ÄKTA FPLC. The column was washed with 10 column volumes (CV) of lysis buffer, and the protein was eluted with a 8-100% gradient of elution buffer (50 mM Tris:HCl pH 7.8, 300 mM NaCl, 400 mM imidazole, and 10% (v/v) glycerol). Fractions containing protein, which were visibly yellow, were concentrated to 2 mL using a 30 kDa centrifugal cutoff filter at 4,000 x g, 4 °C. Concentrated proteins were further purified by size exclusion chromatography (HiLoad 16/60 Superdex S200, GE Healthcare) in gel filtration buffer (50 mM Tris:HCl pH 7.8, 100 mM NaCl, 10% (v/v) glycerol). Estimated molecular weight including the 6x His tags for TropB and variants, 50.8 kDa, was confirmed by SDS-PAGE analysis Fig. 2.18. TropB eluted as a dimer with an apparent molecular weight of 102 KDa and was concentrated to 20.3 mg mL^{-1} (Pierce 660nm assay, Thermo Scientific). FAD content was measured by the flavin incorporation assay. [62]
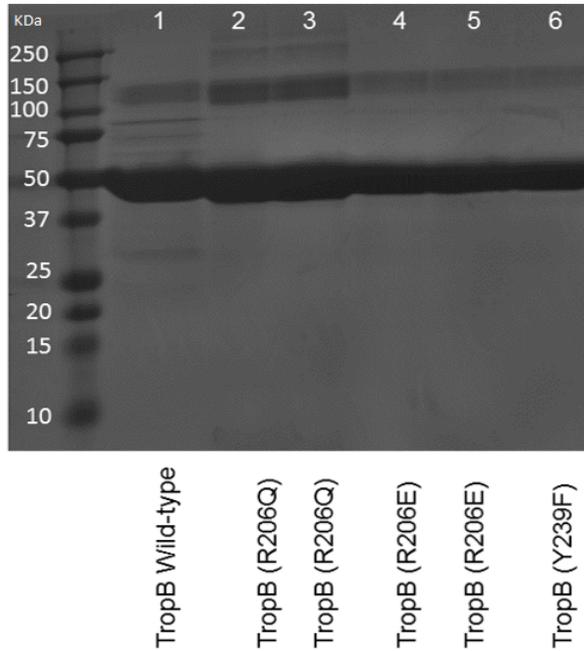
Figure 2.18: Purified TropB, TropB R206Q, TropB 206E, and TropB Y239F, SDS-PAGE gel. The gel was stained with Quick Coomassie stain (Anatrace) and visualized with the Azure Gel Imaging System.
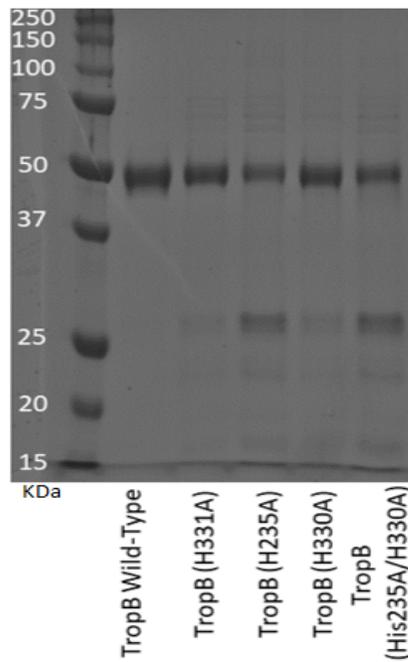
Figure 2.19: Purified TropB Wild-typeTropB, TropB H331A, TropB H235A, and TropB H330A, TropB H330A/H235A SDS-PAGE gel. The gel was stained with Quick Coomassie stain (Anatrace) and visualized with the Azure Gel Imaging System.

TropB H235A: The cell pellet was resuspended in 40 mL lysis buffer (50 mM Tris:HCl pH 7.8, 300 NaCl, 10 mM imidazole, 10% (v/v) glycerol) with 0.1 mg mL^{-1} lysozyme, 0.05 mg mL^{-1} DNase, and 0.1 mM FAD, incubated on a rocker at 4 °C for 45 min, lysed by sonication and cleared by centrifugation (46,413 x g for 30 min). The supernatant was incubated with Ni-NTA on a rocker for 2 h at 4 °C, followed by purification by gravity using a 25-50 mM gradient with increments of 5 mM imidazole. Protein was eluted with 100% elution buffer (50 Mm Tris:HCl pH 7.8, 300 mM NaCl, 400 mM imidazole, 10% (v/v) glycerol). Concentrated protein was desalted over a PD-10 desalting column (GE Healthcare). The protein was concentrated further using a 30 kDa centrifugal concentrator at 4,000 x g, 4 °C. Concentrated protein was divided into 100 μL aliquots, frozen in liquid nitrogen and stored at -80 °C

For crystallization: Protein was dialyzed overnight at 4 °C in 50 mM Tris pH 7.8, 50 mM NaCl. After dialysis, proteins were incubated with 500 μM FAD.

Oligomeric state determination: Separations were conducted as described in respective purification procedures using a Sephacryl S-200 HR gel filtration column. Approximate sizes were determined relative to a calibration curve generated using the Gel Filtration Standards Kit (Sigma). Where the calibration curve is defined by:

$$y = 10^{(-0.7496(x/39.88)+3.068)} \tag{2.1}$$

Where Y volume that is extrapolated to the column calibration curve and X is the peak volume of the desired protein. The apparent molecular weight of TropB was 95 kDa. Based on this analysis, it can be concluded that TropB forms a dimer in solution.
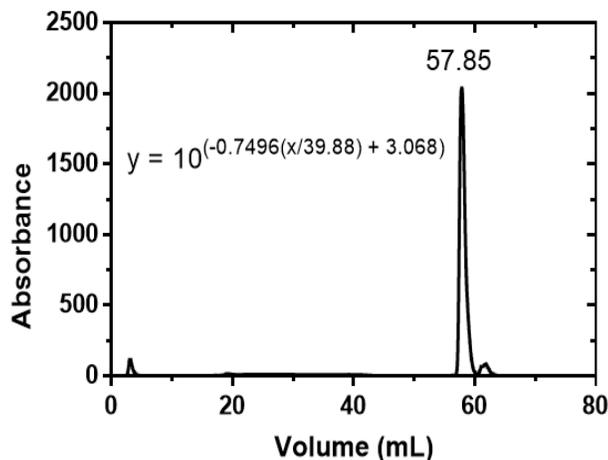
Figure 2.20: Size-exclusion elution profile of TropB

Determination of flavin incorporation: Samples of each protein were diluted to 20 μM using gel filtration buffer (50 mM Tris:HCl pH 7.8, 100 mM NaCl, 10% (v/v) glycerol) for UV-Vis analysis in a disposable poly(methyl 2-methylpropenoate) cuvette. The absorbance spectrum for each protein was taken using a SpectraMax M5 UV-Vis (Molecular Devices). Scans were taken from 300-700 nm in 2 nm increments Fig. 2.21. 40 μL of a fresh 10% SDS solution (w/v) was added to each 1 mL solution, mixed and boiled at 100 °C for 10 min. Absorbance of the denatured enzyme was taken from 300-700 nm in 2 nm increments Fig. 2.21. Absorbance at 450 nm for denatured protein and the extinction coefficient (11,300 $M^{-1}cm^{-1}$) for free FAD were used to determine the concentration of FAD in each protein sample using Beer's law. The typical FAD incorporation for each protein was TropB 80%, TropB Y239F 90%, TropB R206Q 87%, and TropB R2066E 96%. This percentage was multiplied by the value obtained from the Pierce 660 assay to attain the correct protein concentration.
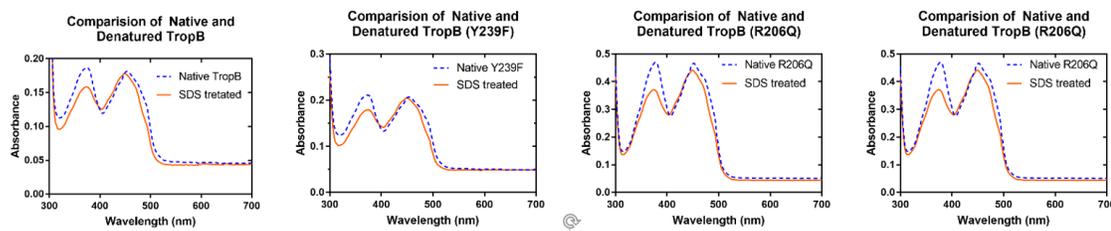
Figure 2.21: FAD incorporation comparison of native and denatured protein.

II. Protein crystallization and structure determination

Crystallization: Diffraction-quality crystals of wild type TropB were grown by sitting-drop vapor diffusion at 20 °C from a solution containing 1 μL protein stock with 0.65% Dimethyl sulfoxide (DMSO) and 2 μL reservoir solution (1.11 M ammonium tartrate dibasic, 6% hexanediol). Crystals of the substrate **2.1** complex were grown under identical conditions with the inclusion of 4 mM substrate **2.1** in the protein stock. TropB Y239F was crystallized as the wild type TropB without DMSO. TropB R0206Q was crystallized also without DMSO in a mixture of 1 μL protein stock and 3 μL reservoir solution. All crystals were flash-cooled in liquid nitrogen without additional cryoprotection.

Structure determination and refinement TropB WT: Diffraction data were recorded at GM/CA beamline 23-ID-B at the Advanced Photon Source (APS), integrated with XDS [63] and scaled in Aimless [64] in the CCP4 suite [64] 2.2. The structure was solved by molecular replacement using Phaser [65] with flavin-dependent salicylate hydroxylase (PDB code 5EVY) [44] as a search model by truncating the non-equivalent side chains to alanine and deleting surface exposed loops. An 87% complete model was obtained using Phenix AutoBuild. [66] Further model building was carried out in Coot. [67] The final model includes amino acids 11-447 in each TropB monomer. Residues 1-10 at the N-terminus and an internal loop (81-86) are disordered and not included in the model 2.2. The dimer model also includes 2 FAD cofactors, 2 hexanediol molecules, 2 chloride ions, and 481 water molecules. Figures were made with Pymol [68], buried surface areas were cal-

culated with the PISA server [69], and the electrostatic surface potential with APBS. [70] The stereochemical quality of the structure was validated with MolProbity. [71] The structures of the substrate **2.1** complex and the R206Q and Y239F variants were solved from the wild type free enzyme structure. Substrate binding was deleterious to crystal quality and limited our ability to interpret details in the active site. Crystals of the R206Q and Y239F variants were significantly smaller than those of wild type TropB, which limited diffraction quality.

Chloride binding: Electron density in a tight pocket between the FAD isoalloxazine ring and Pro329, Gly332 and Gly334 was attributed to the presence of a chloride ion Fig. 2.23. This is the predicted binding site for the molecular oxygen substrate. Several other FAD-dependent monooxygenases can be inhibited by chloride or show competitive inhibition respect to NAD(P)H. [45, 72, 73]

Figure 2.22: TropB sequence alignment with SALH (BAA61829.1), SorbC(XP002567552), and AzaH(G3XMC2.1). TropB amino acids subjected to mutagenesis are starred. Alignment is colored by conservation.
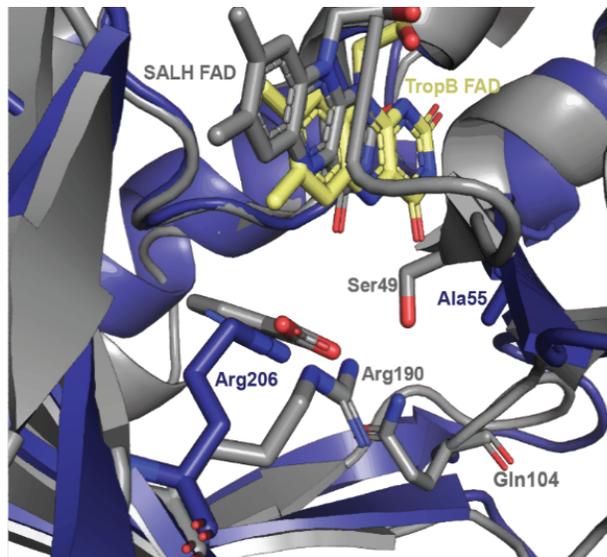
Figure 2.23: Structural alignment of SALH (5EVY) [44] in gray and TropB in deep blue. Highlighting comparison of catalytic residues in SALH compared to TropB.

	TropB WT	TropB + Substrate 1	TropB R206Q	TropB Y239F
Data Collection				
Space group	$C222_1$	$C222_1$	$C222_1$	$C222_1$
Cell dimensions a,b,c (Å)	70.81 184.49 164.13	70.41 183.83 164.23	70.81 184.17 163.52	70.26 184.51 163.26
α,β,γ (°)	90.0 90.0 90.0	90.0 90.0 90.0	90.0 90.0 90.0	90.0 90.0 90.0
X-ray Source	APS 23 ID-B	APS 23 ID-B	APS 23 ID-B	APS 23 ID-B
Wavelength (Å)	1.033	1.033	1.033	1.033
d_{min} (Å)	1.750 (1.813-1.750)[1]	2.245 (2.325-2.245)[1]	2.300 (2.382-2.300)[1]	2.303 (2.386-2.303)[1]
R_{merge}	0.095 (1.952)[1]	0.190 (2.457)[1]	0.192 (1.480)[1]	0.092 (1.092)[1]
R_{merge} inner shell	0.058	0.074	0.058	0.038
Wilson B-factor (Å^2)	34	52	50.7	52.9
Mean I/σ (I)	13.6 (1.1)[1]	9.6 (1.0)[1]	12.49 (1.45)[1]	12.4 (1.44)[1]
Completeness (%)	99.7 (97.7)[1]	99.8 (98.9)[1]	99.7 (98.1)[1]	99.8 (99.1)[1]
Multiplicity	12.8 (12.4)[1]	13.3 (13.6)[1]	12.9 (13.5)[1]	6.7 (6.9)[1]
Total Reflections	1,382,301 (130,270)[1]	680,982 (68,084)[1]	617,329 (62,315)[1]	316,868 (31,739)[1]
$CC_{1/2}$	0.998 (0.602)[1]	0.996 (0.410)[1]	0.998 (0.735)[1]	0.999 (0.793)[1]
CC*	1.000 (0.867)[1]	0.999 (0.763)[1]	1.000 (0.921)[1]	1.000 (0.941)[1]
Refinement				
Data Range (Å)	46.43-1.750	45.96-2.245	46.91-2.300	46.13-2.303
Reflections used in refinement (#)	107,896	47,257	47,766	47,257
R_{work}/R_{free}	0.169/0.195	0.186/0.232	0.190/0.250	0.186/0.233
Non-hydrogen atoms (#)	7,456	7,056	7,082	7,056
protein	6,821	6,830	6,831	6,830
ligand	154	108	114	108
water	481	118	137	118
Amino acid residues	861	862	862	862
Deviation from ideality				
RMSD(bonds) (Å)	0.012	0.011	0.014	0.011
RMSD(angles) (°)	1.12	0.94	1.27	0.95
Average B-factor (Å^2)	44.1	63.2	55.9	63.1
Protein	44.15	63.5	59.3	63.4
ligands	41.8	52.9	46.8	53
solvent	44.9	55.9	49.4	55.9
Ramachandran Plot				
Favored (%)	96.7	96.5	95.4	96.5
Allowed (%)	3.3	3.3	4.6	3.3
Outliers (%)	0	0.2	0	0.2
PDB	**6NES**	**6NET**	**6NEU**	**6NEV**

Table 2.2: Crystallographic Summary, [1]Values in parentheses pertain to the outermost shell data
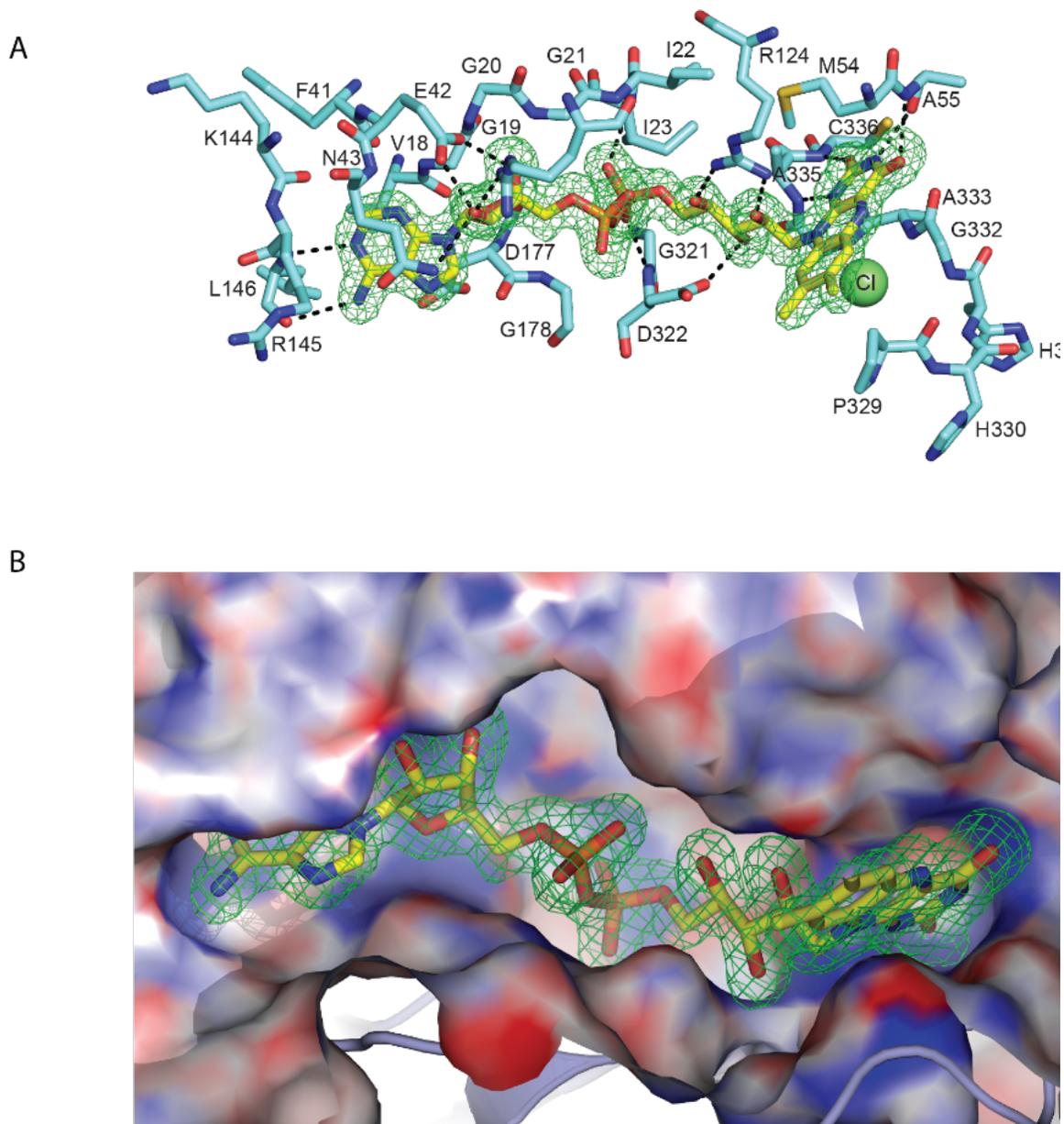
Figure 2.24: FAD binding site. (A) Bound FAD with omit electron density (green contours at 3.0 σ) and hydrogen bonds (black dashed lines) to the protein. (B) TropB flavin binding site colored according to electrostatic potential (red to blue from - kT/e to + kT/e).
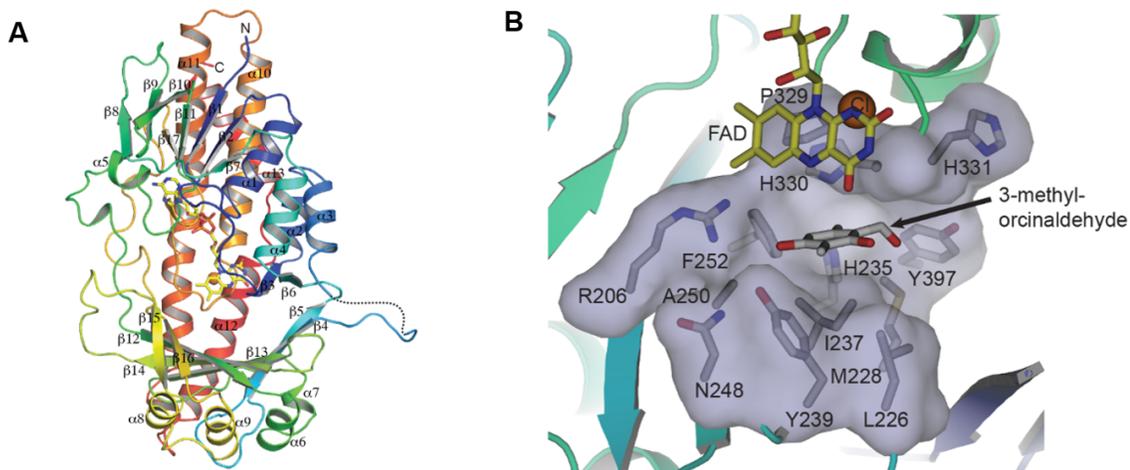
Figure 2.25: FAD binding site. Overall structure of TropB and chloride binding in the active site cavity. (A) Ribbon diagram of the TropB monomer colored as a rainbow from the blue N-terminus to the red C-terminus. (B) TropB active site with substrate **2.1**, its natural substrate, modeled. A chloride ion sits in a tight pocket between the isoalloxazine ring and Pro329, Gly332 and Gly334, as in other FAD class A monooxygenases.
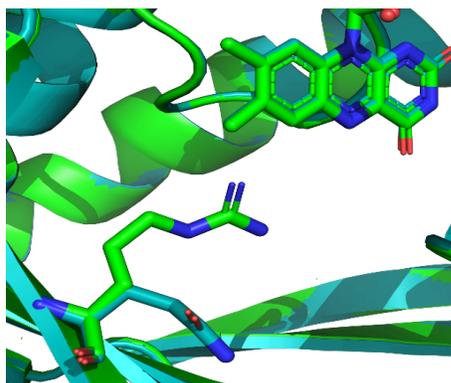
Figure 2.26: Comparison of wild-type TropB (green) and TropB R206Q (cyan). The Arg206 points towards the FAD, while the Gln side chain points towards the backbone and is not able to hydrogen bond with the FAD in the out conformation.
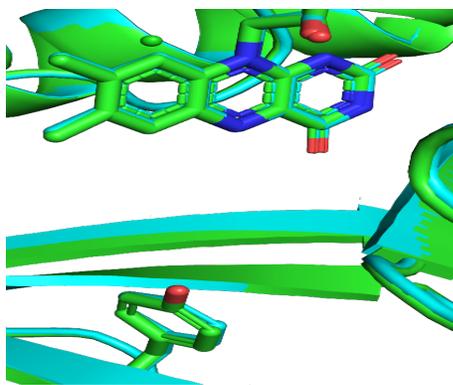
Figure 2.27: Comparison of wild type TropB (green) and TropB Y239F (cyan). The Phe239 points in the same direction as Tyr239 residue.

III. Biocatalytic Reactions

General Considerations: A stock solution of the substrate (50 mM) was prepared by dissolving the substrate in DMSO (analytical grade). Stock solutions of NADP+ (100 mM) and glucose-6-phosphate (G6P, 500 mM) in water were stored at -20 °C. Aliquots of glucose-6-phosphate dehydrogenase (G6P-DH, 100 U mL^{-1}), TropB and variants were stored at -80 °C.

In vitro analytical-scale reactions: Each reaction contained 50 mM of potassium phosphate buffer pH 8.0 (25 μL, 100 mM stock), 2.5 mM substrate (2.5 μL, 50 mM), 1 mM NADP+ (0.5 μL, 100 mM), 1 mM G6P-DH (0.5 μL, 100 u mL^{-1}), 5 mM G6P (5 μL, 5 mM), and Milli-Q water to a final volume of 50 μL. Reactions were carried out at 30 °C for 1 h and quenched with the addition of 70 μL acetonitrile (HPLC grade) containing 2.5 mM pentamethylbenzene as an internal standard. Reactions were cleared by centrifugation at 16,000 x g, 12 min) and the supernatant was analyzed by UPLC-DAD. Consumption of the starting material was quantified by comparison to standard curves of each substrate.

Substrate Calibration Curve:
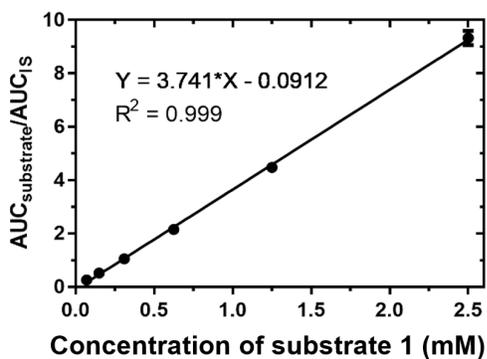
Figure 2.28: Calibration Curve of substrate **2.1**.

UPLC traces of Biocatalytic Reactions

	Retention Time	Area	% Area	Hei
1	1.892	124377	52.10	293
2	2.607	114359	47.90	106
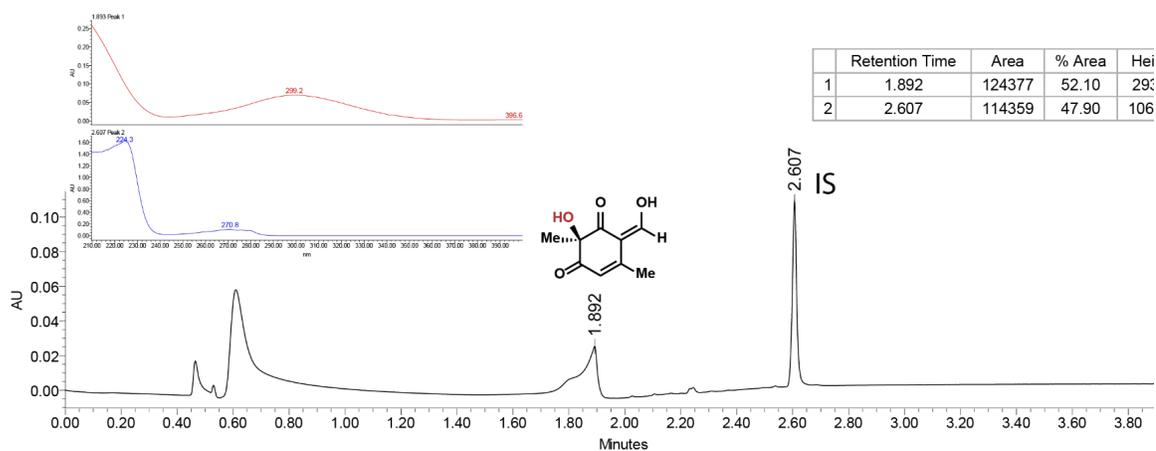
Figure 2.29: Oxidative dearomatization of **2.1** by TropB WT. PDA traces of enzymatic reaction. IS = internal standard.
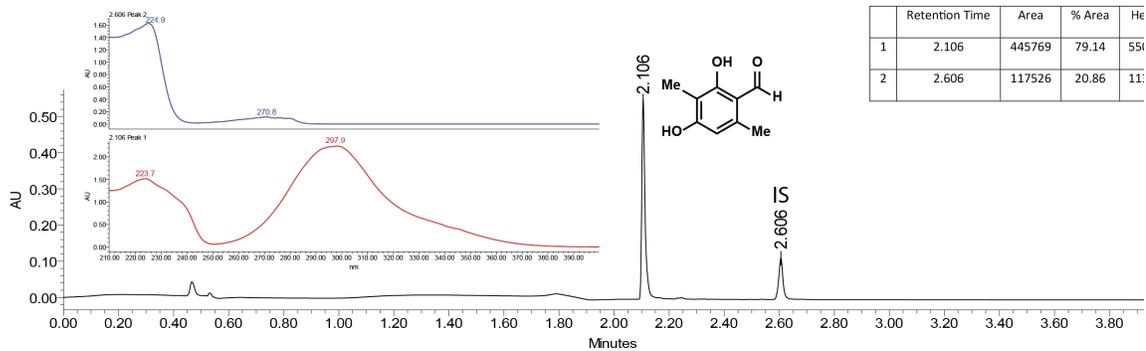
Figure 2.30: No enzyme control of the reaction of **2.1**. PDA traces of enzymatic reaction.

	Retention Time	Area	% Area	He
1	2.106	445769	79.14	550
2	2.606	117526	20.86	11
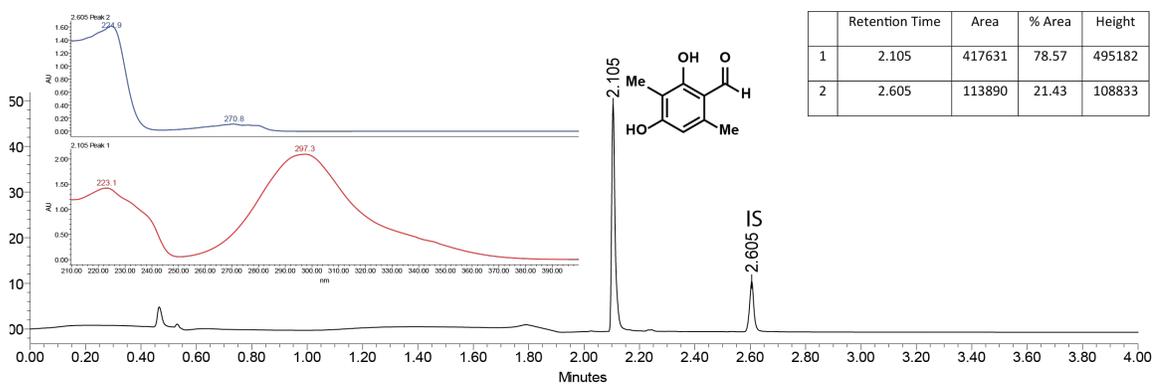
Figure 2.31: Reaction of **2.1** and TropB R206Q. PDA traces of enzymatic reaction.

	Retention Time	Area	% Area	Height
1	2.105	417631	78.57	495182
2	2.605	113890	21.43	108833
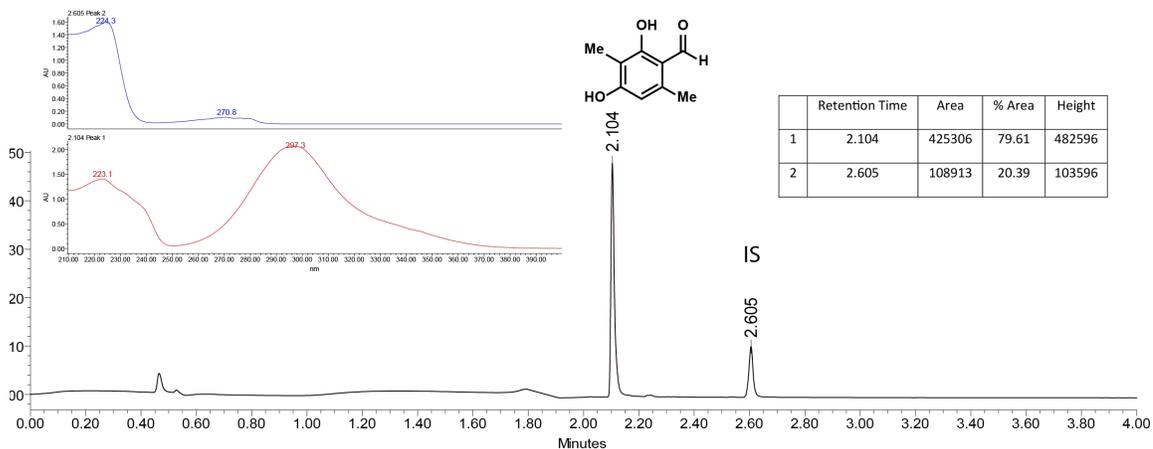
Figure 2.32: Reaction of **2.1** and TropB R206E. PDA traces of enzymatic reaction.

	Retention Time	Area	% Area	Height
1	2.104	425306	79.61	482596
2	2.605	108913	20.39	103596
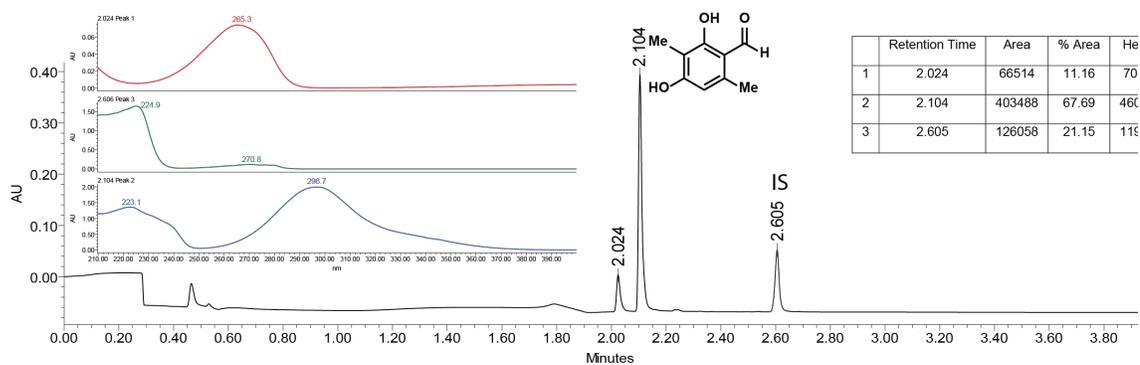
Figure 2.33: Reaction of **2.1** and TropB Y239F. PDA traces of enzymatic reaction.
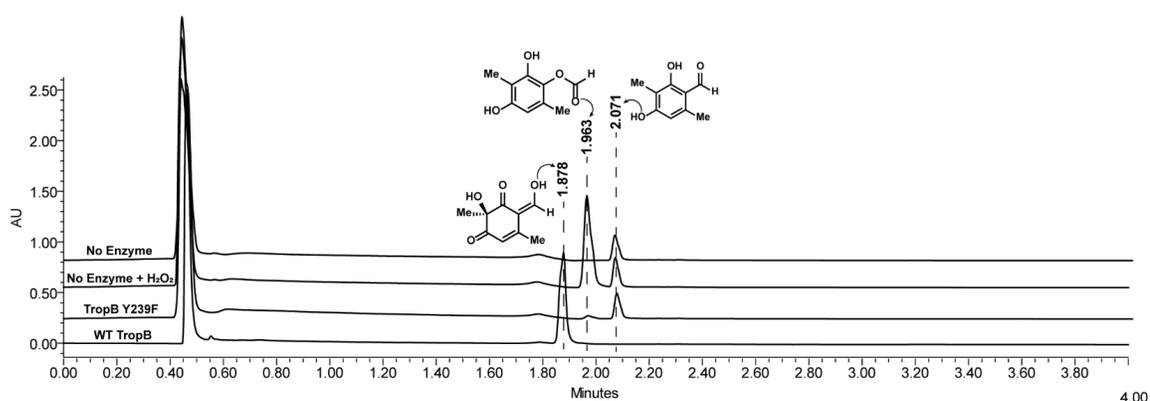
Figure 2.34: Reaction of **2.1** and TropB WT, TropB Y239F, or hydrogen peroxide. PDA traces of enzymatic reaction and control reactions.

Note: In the Fig. 2.33 there is a peak with a retention time of 2.024 min with maximum absorbance of 265 nm. This peak does not have the corresponding absorption spectrum of the dearomatized product- but that of the Baeyer-Villiger product shown below in Fig. 2.34. This product is formed due to a high concentration of hydrogen peroxide generated by this variant. This is further confirmed by a no enzyme control run with 2.5 mM hydrogen peroxide.

Summary of *In vitro* reactions

Enzymes	Average Percent Conversion		
TropB Wild-Type	97.0	±	0.8
No Enzyme	0.0	±	0.0
TropB (Y239F)	0.0	±	0.0
TropB (R206E)	0.0	±	0.0
TropB (R206Q)	0.0	±	0.0
TropB (H331A)	71	±	2
TropB (H235A)	10	±	3
TropB (H330A)	11.1	±	0
TropB (H330A/H235A)	10	±	2
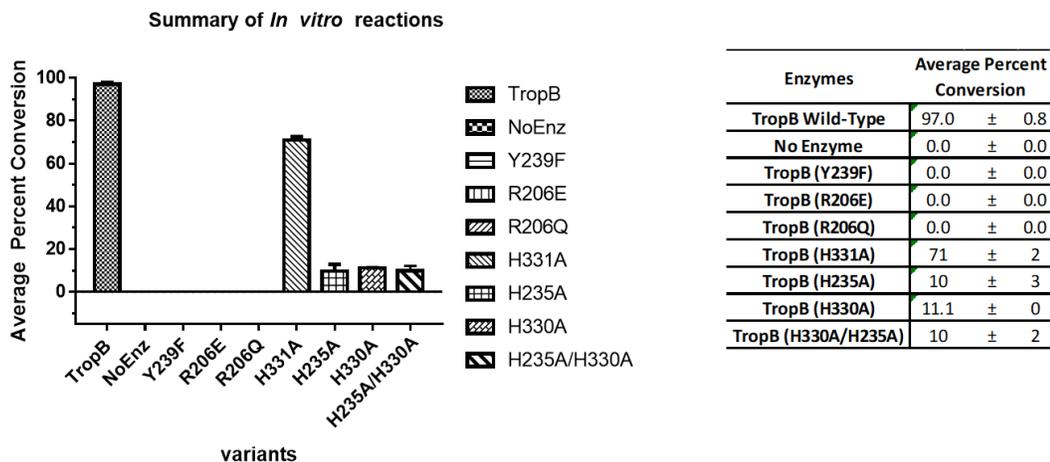
Figure 2.35: Summary of *in-vitro* reactions. Average percent conversion calculated from reactions in triplicate and compared to calibration curve of substrate **2.1**.

IV. Biochemical Characterization of TropB WT and Variants

Substrate binding titrations with TropB WT, TropB (R206Q), TropB (R206E), and TropB (Y239F). All spectral binding measurements were carried out on a UV-2501PC spectropho-tometer (Shimazu). Spectral binding titrations of substrate **2.1** were performed at 4 °C in 50 mM phosphate buffer pH 8.0 using 1 mL volume samples containing 20 μM enzyme in a 1 cm pathlength cuvette. Substrate 1 stock solutions were made at 3-14 mM in 100 mM phosphate buffer pH 8.0. Titrations were performed step wise additions of the substrate **2.1**. After each addition the cuvette was sealed with a Teflon stopper and mixed 5 times slowly. UV-Vis spectra (300-700 nm) were recorded for the ligand-free enzyme and follow-ing each addition of substrate. Substrate additions were continued until no further shifts in the spectrum occurred or 500 μM substrate **2.1** was reached. A turbidity correction was done for the Rayleigh scattering, equation, eq. 2.2.

$$Scatter = \frac{1}{\lambda^4} + Constant \tag{2.2}$$

A difference spectrum was generated at each stage in the titration by subtraction of

the spectrum from the ligand-free enzyme spectrum for each subsequent ligand-bound form of substrate [Substrate 1:FAD-dependent monooxygenase] spectrum produced. The wavelengths of the largest absorbance changes were identified from the difference spectra, and a maximal induced absorbance change Δ_{max} at each point in the titration was determined by subtracting the absorbance at the wavelength minimum from the absorbance maximum, using the same wavelength pair for each titration. Δ_{max} values were plotted against ligand concentration, and data were fit using either a cooperative binding (The Hill equation) function, to determine dissociation constants (K_d values), as described previously. All data processing was done using GraphPad Prism software.
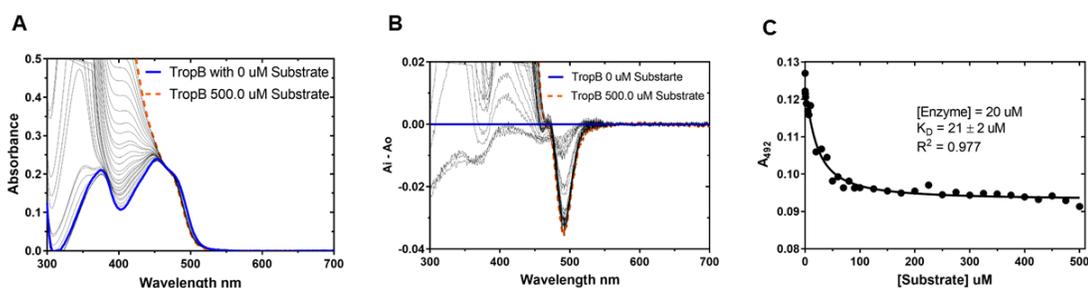
Figure 2.36: Binding titration of TropB WT. (A) Binding titration of TropB WT corrected for turbidity with Rayleigh equation. The substrate-free (20 μM) is shown with a thick solid blue line. Thin solid lines show spectra collected after stepwise additions of substrate **2.1** from a (3 mM-14 mM stock). The final spectrum (collected at 500 μM substrate **2.1**) is near saturated with substrate. (B) Plot of the difference spectra. (C) Substrate-induced absorbance difference against the relevant concentration of substrate 2.1 at 492 nm.
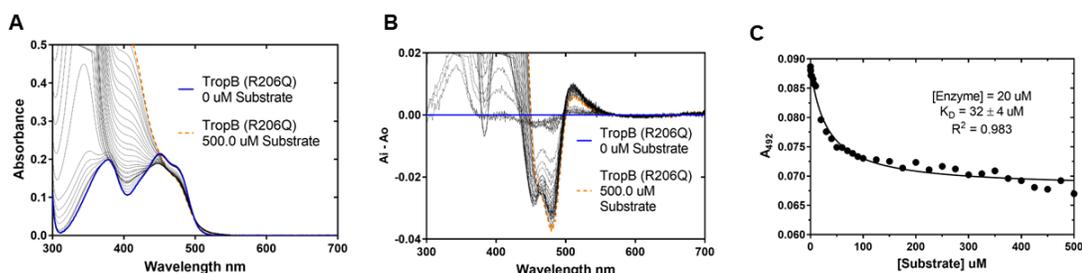
Figure 2.37: Binding titration of TropB R206Q (A) Binding titration of TropB R206Q corrected for turbidity with Rayleigh equation. The substrate-free (20 μM) is shown with a thick solid blue line. Thin solid lines show spectra collected after stepwise additions of substrate **2.1** from a (3 mM–14 mM stock). The final spectrum (collected at 500 μM substrate **2.1**) is near saturated with substrate. (B) Plot of the difference spectra. (C) Substrate-induced absorbance difference against the relevant concentration of substrate **2.1** at 492 nm.
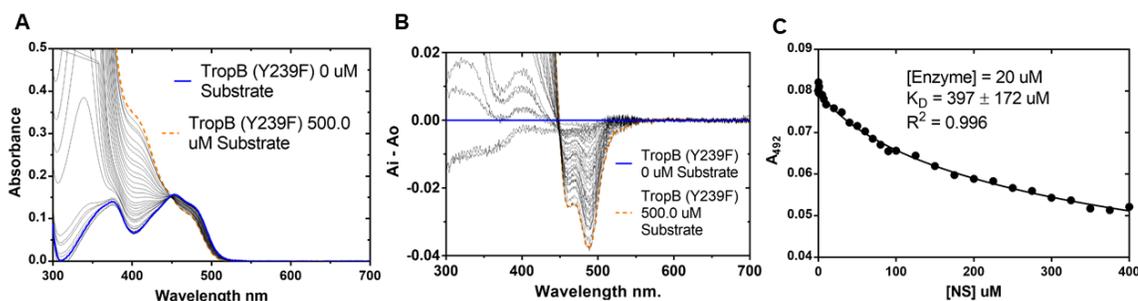
Figure 2.38: Binding titration of TropB Y239F. (A) Binding titration of TropB Y239F corrected for turbidity with Rayleigh equation. The substrate-free (20 μM) is shown with a thick blue solid line. Thin solid lines show spectra collected after step wise additions of substrate **2.1** from a (3 mM–14 mM stock). The final spectrum (collected at 500 μM substrate **2.1**) is near saturated with substrate. (B) Plot of the difference spectra. (C) Substrate-induced absorbance difference against the relevant concentration of substrate **2.1** at 492 nm.
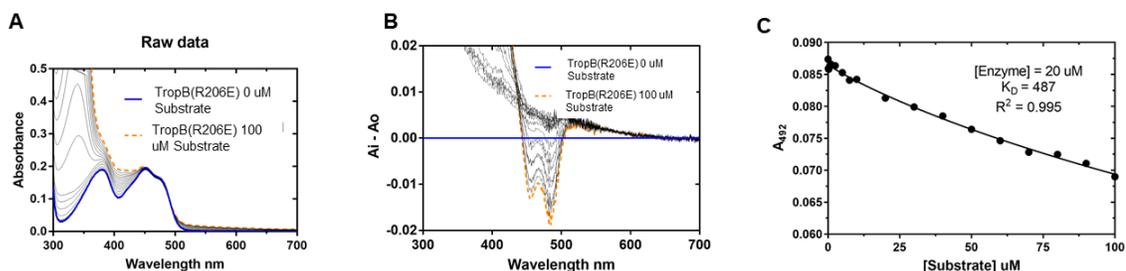
Figure 2.39: Binding titration of TropB R206E. (A) Binding titration of TropB R206E corrected for turbidity with Rayleigh equation. The substrate-free (20 μM) is shown with a thick blue solid line. Thin solid lines show spectra collected after step wise additions of substrate **2.1** from a (3 mM–14 mM stock). The final spectrum (collected at 500 μM substrate **2.1**) is near saturated with substrate. (B) Plot of the difference spectra. (C) Substrate-induced absorbance difference against the relevant concentration of substrate **2.1** at 492 nm

Substrate 2.1 spectra at various pH values: All spectral substrate measurements were carried out on a SpectraMax M5 UV-Vis (Molecular Devices). 2.5 mM of substrate **2.1** (50 mM, DMSO stock) was diluted in 1 mL 50 mM potassium phosphate buffer pH 2.1 at 22 °C. Substrate was titrated to the following pHs (2.6, 3.0, 3.5, 4.3, 5.0, 5.6, 6.0, 6.5, 7.1, 7.8, 8.9, 9.6,10.5). UV-Vis scans were taken at each titration point to generate Fig. 2.40. Scans were taken from 250-700 nm in 2 nm increments. Absorbance of the phenol and the phenolate were plotted against the pH. The inflexion point was calculated to determine that the pKa of the substrate is 7.1 Fig. 2.41.
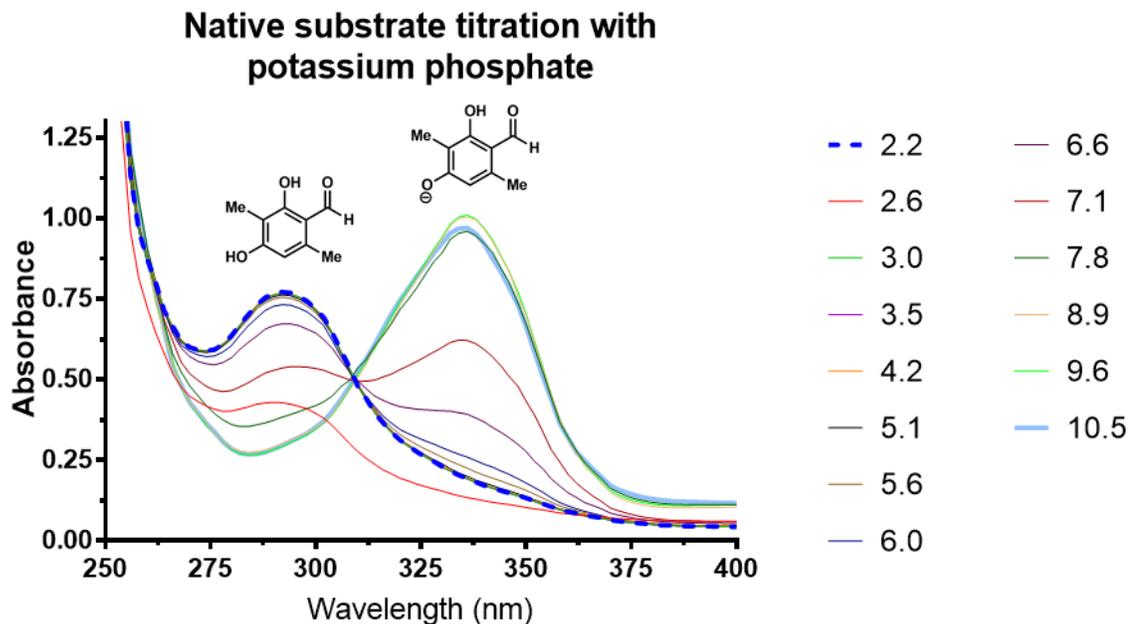
Figure 2.40: Native substrate **2.1** titration with potassium phosphate.
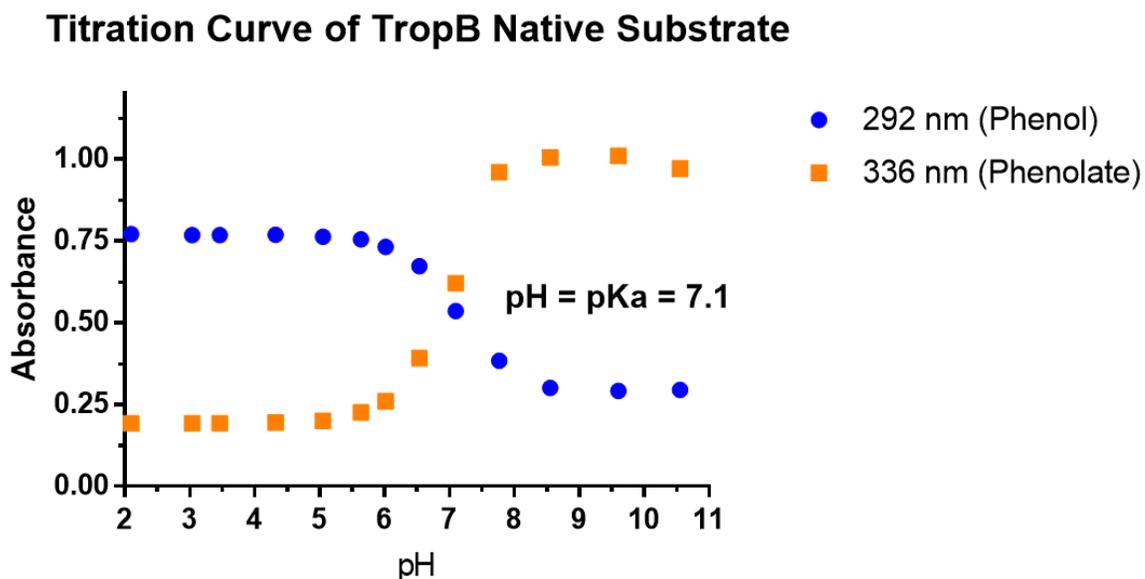
Figure 2.41: Native substrate **2.1** titration with potassium phosphate.

Depleation of NADPH in of TropB WT and variants: All spectral measurements were carried out on a UV-2501PC spectrophotometer (Shimazu). Addition of NADPH was done at 25 °C. Reaction conditions: 50 mM phosphate buffer pH 8.0, 100 nM enzyme, 250 μM NADPH and 125 μM substrate **2.1** was carried out in a 1 cm pathlength cuvette.

Depletion of NADPH was measured at 362 nm for 1000 sec.
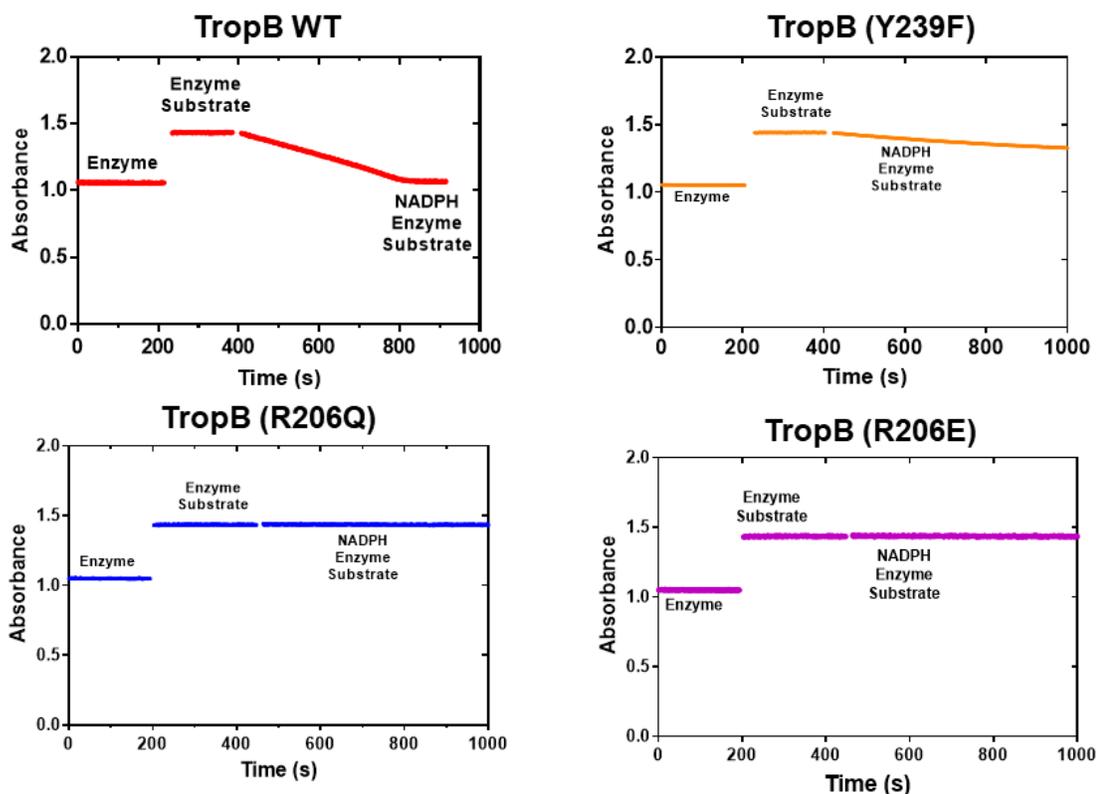
Figure 2.42: NADPH depletion assay monitored at 362 nm.

Reductive half-reaction for TropB WT: The reaction of TropB WT and TropB Y239 was studied in a Hi-Tech Scientific KinetAsyst SF-61 DX2 Stopped-flow spectrophotometer at 4 °C. Enzyme solution was made anaerobic in a tonometer in the presence of 200 μM substrate **2.1** and mixed with anaerobic solutions of NADPH at different concentrations. [74] Kinetic traces were fit to a single exponential using GraphPad PRISM. The observed rate constants for flavin reduction were plotted vs. NADPH concentration and fitted to a square hyperbola to obtain k_{red} and K_d. [75]
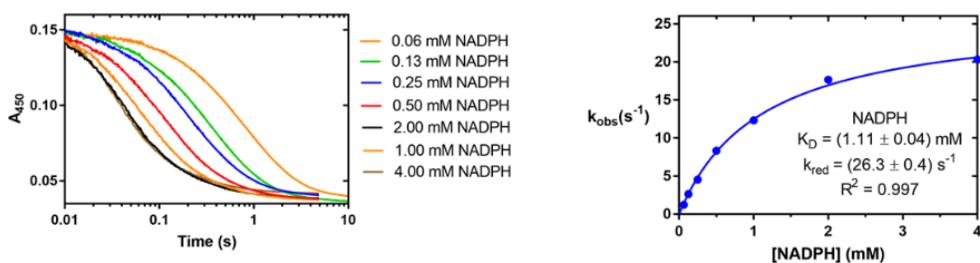
Figure 2.43: TropB wild type reductive half-reaction.

Reductive half-reaction of TropB R206Q monitored over time in the presence of NADH: The reaction of TropB R206Q was studied in UV-2501PC spectrophotometer (Shimazu) UV-Vis spectrophotometer at 25 °C. A 20 μM enzyme solution was prepared anaerobically in a tonometer in the presence of 125 μM substrate **2.1** and mixed with anaerobic solution of 250 μM NADH. [74]UV-Vis scans were taken every 2 min for 3.3 h over 300 nm-700 nm in 1 nm increments. Full reduction of the variant TropB R206Q was achieved at 3.3 h.
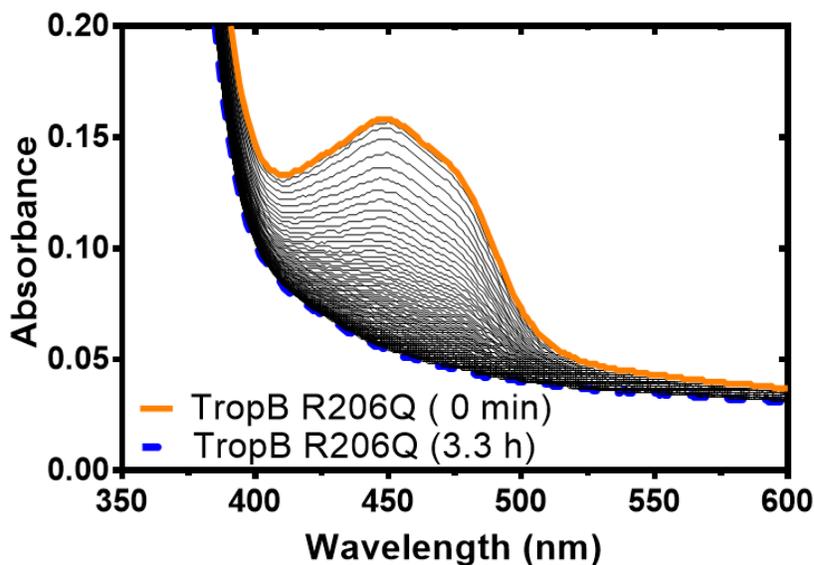
Figure 2.44: TropB R206Q anaerobic reductive half-reaction with NADH monitored over time.

53

Reduction of TropB R206Q: All spectral measurements were carried out on a UV-2501PC spectrophotometer (Shimazu). Addition of NADPH was done at 25 °C. Reaction conditions: 50 mM phosphate buffer pH 8.0, 20 μM enzyme, 250 μM NADPH and 125 μM substrate **2.1** was carried out in a 1 cm pathlength cuvette. Depletion of NADPH was measured at 362 nm for 3.3 h Fig. 2.45.
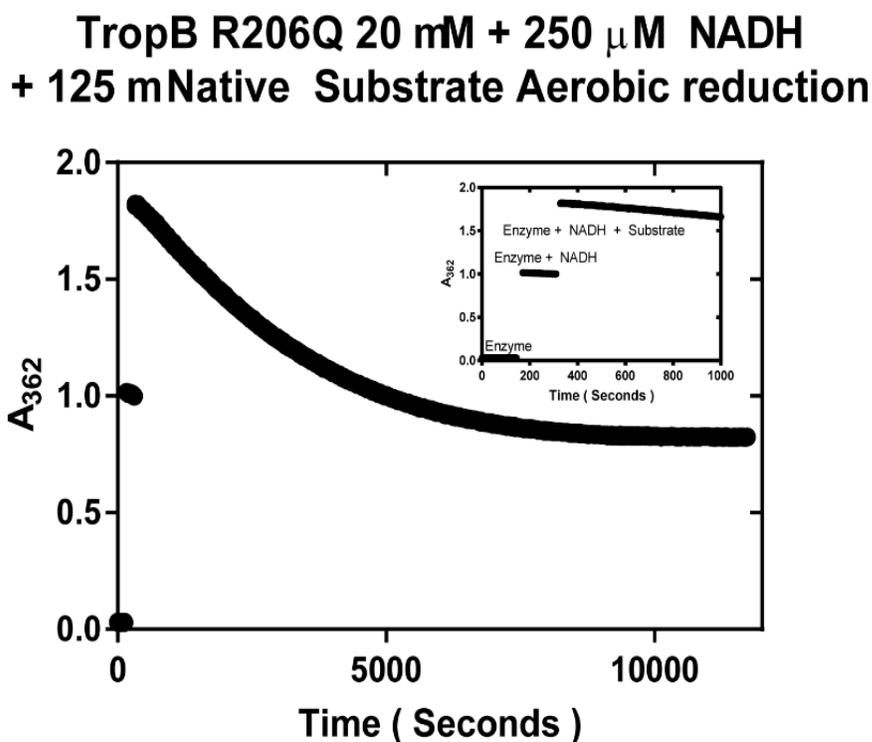
Figure 2.45: NADPH depletion assay monitored at 362 nm with TropB R206Q at 20 μM.

Hydrogen Peroxide Formation Assay: Production of hydrogen peroxide was detected by HRP (2.5 μM) and 4-amino antipyrine (0.75 mM)/2,4,6-tribromo-3-hydroxybenzoic acid dye (0.02% W/V), phosphate buffer pH 8.0 (100 mM). [51] In a 96-well plate, 190 μL of the assay solution was combined with 10 μL of peroxide (16 μM-3 mM) to generate a hydrogen peroxide calibration curve. For reactions, 10 μL of a quenched *in vitro* reaction was added to 190 μL assay solution to measure hydrogen peroxide production measured at 510 nm.
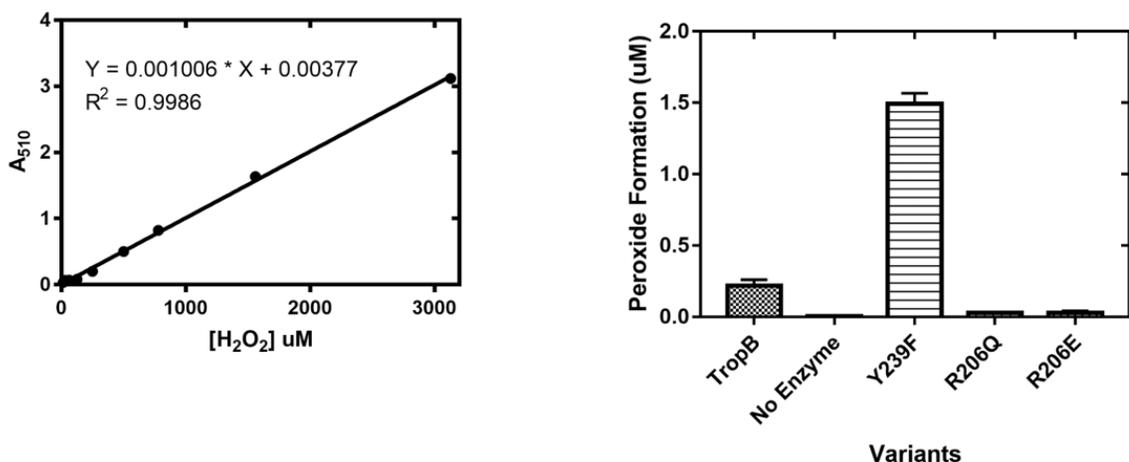
Figure 2.46: NADPH depletion assay monitored at 362 nm with TropB R206Q at 20 μM.

Oxidative half reaction of TropB monitored over time in the presence of O$_2$:
The reaction of 50 μM TropB WT was studied in a Hi-Tech Scientific KinetAsyst SF-61 DX2 Stopped-flow spectrophotometer at 25 °C. An anaerobic enzyme solution was prepared in a tonometer in the presence of 50 μM substrate **2.1** and mixed with buffer, 50 mM Tris:H2SO4 pH 7.8, 25 mM Na$_2$SO$_4$ and 10% glycerol containing 20% or 100% O$_2$. The reaction was monitored at 450 nm to track FAD oxidation. Kinetic traces were plotted using GraphPad PRISM. The blue bold line represents the reaction with 100% O$_2$ and the orange dashed trace represent reaction with 20% O$_2$. The first phase is directly affected by the concentration of oxygen present. Moreover the formation of these two phases indicates the formation of FAD$_{\text{HOOH}}$. [76]
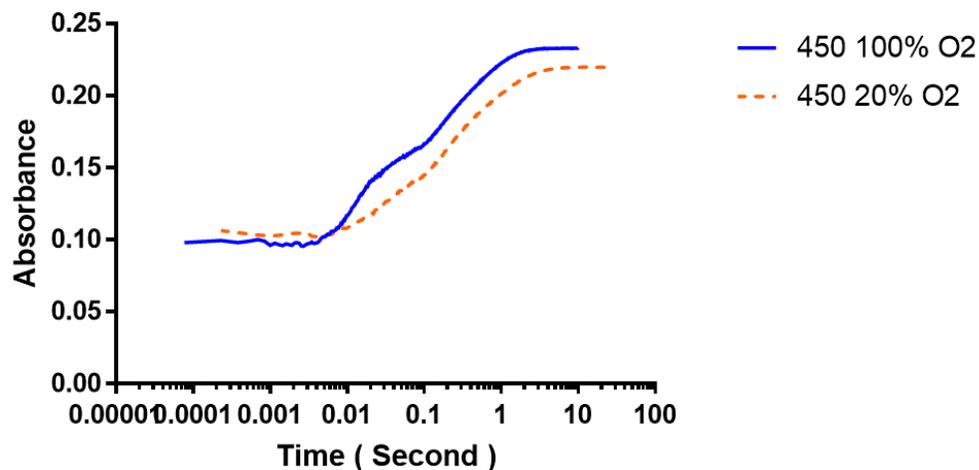
Figure 2.47: TropB WT oxidative half-reaction.

Initial velocity of TropB wild-type, TropB (H331A), TropB (H235A/H330A) with substrate 1: Reaction mixtures were prepared that contained 50 nM enzyme, 5 μM of substrate **2.1**, and 50 mM phosphate buffer pH 8 and Mili-Q water to a final volume of 50 μL. Reactions were initiated by addition 10 μL of a solution of 1 mM NADPH at room temperature. Reactions were quenched with 10 μL of a solution of 10% v/v formic acid in acetonitrile at time points 0, 10, 30, 60, and 300 s. Further, 1.2x dilution with acetonitrile was performed to precipitate protein in solution. The reaction was cleared by centrifugation at 17,000 x g for 12 min. The supernatant was analyzed by LC-MS (TOF). The consumption of the starting material was quantified by comparison to a standard curve of substrate **2.1** and internal standard (1,2,4-trimethylbenzene).
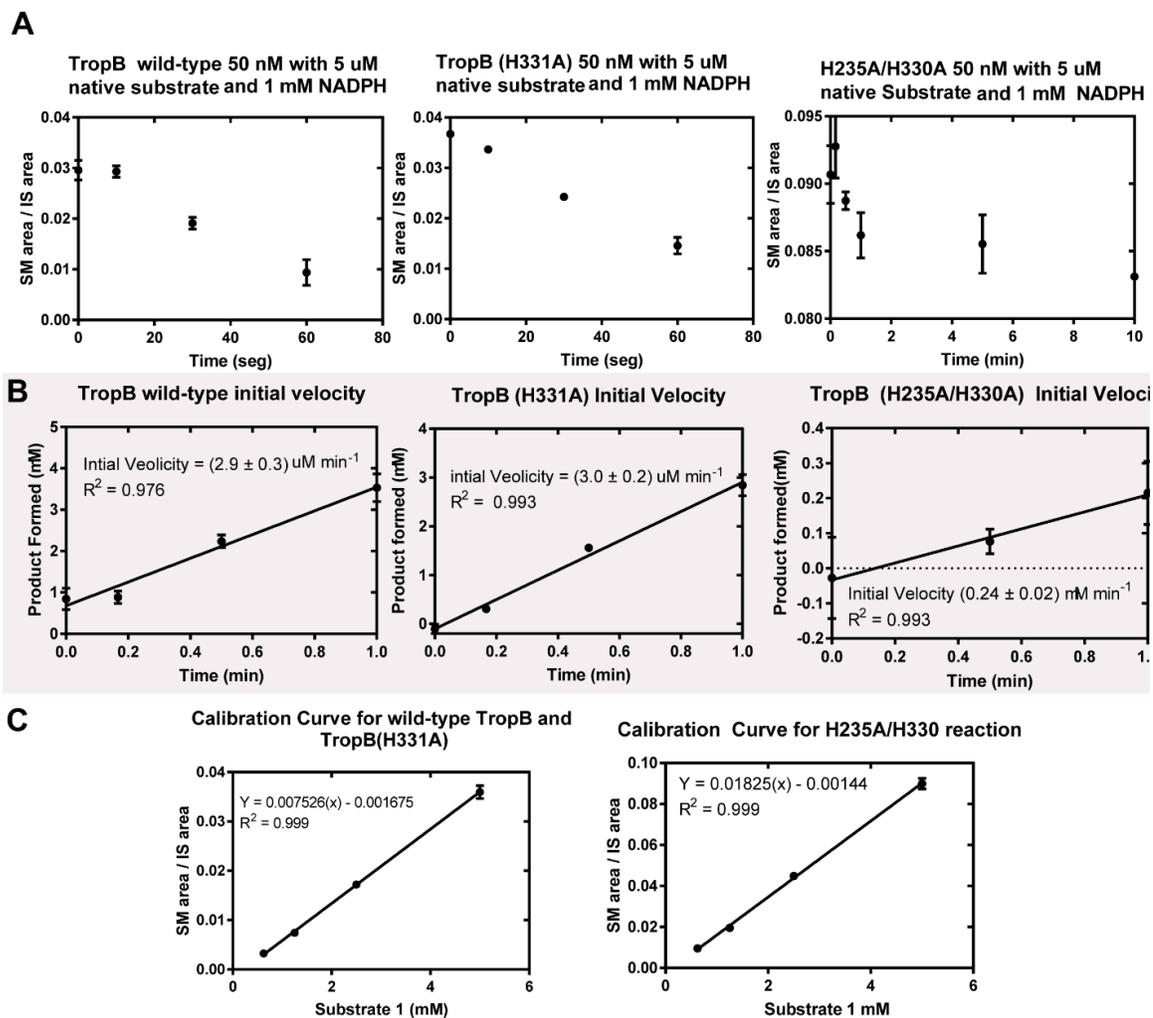
Figure 2.48: (A) Starting material (1) area under the curve divided by internal standard area under the curve plotted over time for enzymatic reactions initiated with NADPH and quenched with 10% v/v formic acid. (B) Initial velocity plots of TropB wild-type, TropB (His331), and TropB (His235A/H330A). (C) Calibration curves used for quantification of starting material determined by LC-MS (TOF).

57

V. Computational modeling of ligand binding

A. Computational Methods

Docking of Substrate 2.1 to TropB

The protocol for docking substrate **2.1** to the holo-form of TropB with the cofactor in the oxidized form was as described by Wu et al. using the CHARMM software package. [77,78] The protocol is also described here due to minor differences from the initial report.

The TropB-cofactor coordinates were obtained from the crystal structure and the protons were added with CHARMM-gui according to their canonical protonation states. [78] The tautomeric state of histidine residues was assigned in order to maximize hydrogen bonds between protein residues. The hydroperoxyflavin coordinates were simply modified from the reduced flavin by adding the hydroperoxy group to C4a on the re face of the flavin. The phenolate form of the substrate Substrate (**2.1**) was used for docking. Grid potentials were first calculated for all fixed elements of the receptor with probe radii corresponding to the latest CGENFF atom types. [79, 80] Flexible side chains were chosen to encompass the broad active site surrounding the flavin, and included Ala55, Thr57, Leu96, Trp98, Phe119, Arg206, Leu226, Met228, Iso237, Tyr239, Asn248, Phe252, and Trp426. A random rotation of the C1-C(aldehyde) bond (the only rotatable bond) was performed with the "obrotamer" algorithm in OpenBabel. [81] The Substrate (**2.1**) structure was then geometry optimized using force field parameters obtained from ParamChem with slight adjustments to the charge distribution in agreement with Mulliken charges from quantum chemical calculations.

The docking cube was centered at the average coordinate of the flexible residues, which were chosen to surround the flavin and active site, and the cube dimensions were defined by the maximum coordinates of the flexible residues in each direction plus 10 Å. The Substrate (**2.1**) position was then translated within the docking cube and randomly rotated and/or further translated. The structure was then geometry optimized with very soft

and soft potentials and then removing any complexes with highly unfavorable poses (>200 kcal/mol). At this stage, the substrate, flavin, and flexible side chains are represented by the CHARMM force field while the rest of the model is represented by the pre-computed grid potentials. A simulated annealing simulation was then performed with very soft and soft potentials (very soft: 4.5 ps of MD with increasing temperature 300 –>700K and 21 ps of MD decreasing temperatures 700 –>300K; soft: 10.5 ps of MD with decreasing temperature 500 –>300K and finally 4.5 ps of MD with decreasing temperature 400 –>50K).

Next, the entire model is represented by the CHARMM force field with a distant dependent dielectric function (ϵ = 3) while geometry optimization is performed with 50 and 100 steps using the Steepest Descent (SD) and Adopted Basis Newton Raphson (ABNR) algorithms, respectively to predict the docked pose. This protocol was repeated 20 times for each of the 25 rotamers docked to the oxidized holo-form TropB for a total of 800 and 500 total docked poses respectively.

Final energies were calculated using generalized Born molecular volume (GBMV) implicit solvent model. [82] Docking poses were ranked by total energy of the complex, and low energy structures consistent with known stereochemical outcome and with a C3 to hydroperoxyflavin O_{distal} distances < 4 Åwere considered for further analysis of reactive positioning.

Additional rigid docking analyses were performed for several substrates (**2.5, 2.6, 2.9, 2.10 and 2.11**) using a slightly modified procedure with no flexible residues. These substrates were docked to the TropB protein structure from the productive binding pose identified during flexible docking with the native substrate **2.1**. This structure used the fully oxidized flavin model (FAD) and had Arg206 and Tyr239 in a position that would allow for hydrogen bonding to the substrate. For two of the bulkier substrates, **2.7** and **2.10**, this rigid docking protocol did not sample favorable productive binding poses, so another round of docking was initiated from the same protein structure, but with residues surrounding the active site made flexible (residues Leu96, Trp98, Phe119, Leu226, Met228,

and Trp426). This flexible protocol produced low-energy, productive poses for the bulkier substrates that were used for analysis. Rigid docking of the native substrate **2.1** was also performed with TropB variants Y239F, R206Q, and R206E based on the corresponding crystal structure coordinates.

Molecular Dynamics Simulations of TropB-peroxy-FADH$_2$-Substrate ternary complex

Select docked structures were subsequently used to construct solvated models. Oxygen atoms identified in the crystal structure as water molecules (TIP3P potential) were retained in the model with protons being added with the HBUILD program in CHARMM. [83] Bulk solvent surrounded the entire ternary complex in a cube with the size being the largest dimension of the protein plus 20 Å. 81 K$^+$ and 78 Cl$^-$ counterions were added to correspond with a 0.15 mol concentration and an overall electrically neutral model. The active-site chloride ion identified in the crystal structure was also represented though this ion immediately translates out of the active site during the simulation. The final model contained 90761 atoms.

MD simulations were performed with the OpenMM/CHARMM interface in version c42b1 and using periodic boundary conditions, particle mesh Ewald for long range electrostatic interactions, and a langevin thermostat (γ = friction coefficient = 5) to maintain the temperature at 298K. [84] Harmonic restraints of 10.0 kcal mol$^-$1
AA$^-$2 were placed on all non-hydrogen atoms of the solute during the first 500 ps of the simulation, which used a 1 femtosecond (fs) timestep. The simulation was continued for another 24 nanoseconds (ns) with a 2 fs timestep. Simulations were also performed on the ternary complex with FAD in reduced form for 100 ns using a 1 fs timestep.

DFTB3/MM Simulations

Snapshots from the MD simulation using a classical force field with the hydroperoxy group aligned for reaction with the oxyanionic substrate were extracted for further simu-

lation using a quantum mechanical/molecular mechanical potential energy function. The DFTB3 method was used for the QM region that included the substrate and the side chains of R206, Y239, and N248, the isoalloxazine ring of the hydroperoxyflavin and seven water molecules proximal to the substrate. [61] The link atom approach was utilized placing the link atom for the cofactor between C25 and C25 so that the first hydroxyl substituent of the R-group was included in the QM region. The total number of DFTB3 atoms was 125. A 150* kcal mol^{-1}

AA^{-2} harmonic restraint was placed on the mass-weighted distance difference between the peroxy Od-Op bond and the forming Od-C3 bond (Reaction coordinate = 0.571 * (Od-Op) − 0.429 * (C3-Od)). The Michaelis complex (where the reactants are lowest in free energy) is represented by a reaction coordinate value of -0.75 Å(full simulation results to be presented in a future manuscript). The complex was simulated with the DFTB3/MM method for 50 ps using a 1 fs timestep. To obtain reliable analysis of the partial charge distribution, the CHELPG application within ORCA was applied to the minimized Michaelis complex structure. [85, 86] The calculations were performed using the B3LYP functional with D3 correction and def2-TZVPP basis set. [87–95] The resulting partial charges are provided in supporting text files.

Native Substrate Docking Pose vs. Crystal Structure Density

Protein/Substrate
TropB Crystal Structure
TropB Docking Model
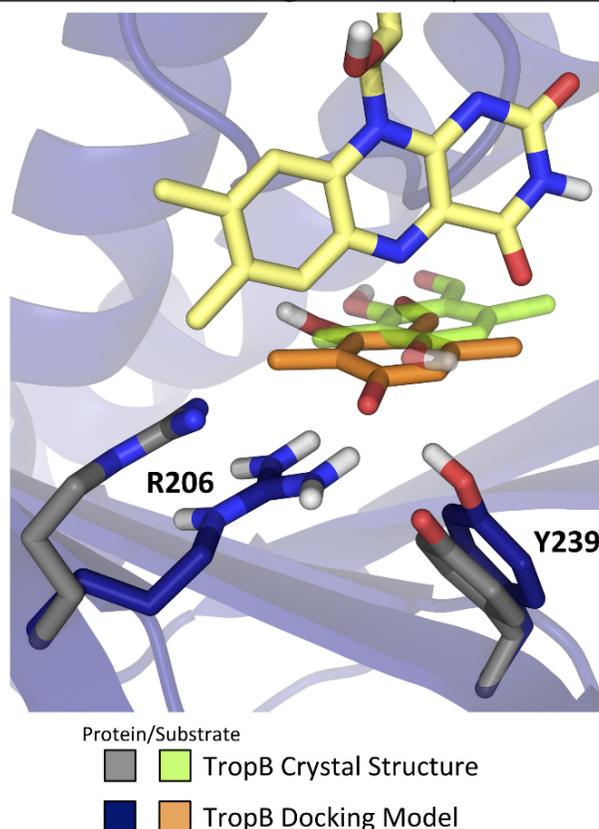
Figure 2.49: Flexible Docking Result for Native Substrate Compared with Crystal Structure Substrate Density. The favorable productive docking pose is close to the crystal structure density presumed to be the substrate. The substrate was modeled into the crystal structure density (transparent green sticks) and shows a 2.1 Åroot-mean-square deviation (RMSD) from the docking pose (solid orange sticks). The protein is shown in the navy cartoon, the crystal structure side chains are shown in gray sticks, the docking model side chains are shown in navy sticks, and the flavin is shown in yellow sticks. In contrast, docking studies with the substrate in the neutral form failed to generate interactions that could activate the substrate through deprotonation or to orient the substrate in a pose that would allow for a reaction with the site- and stereoselectivity necessary to generate the observed product **2.2**. Performed by Sara Tweedy.
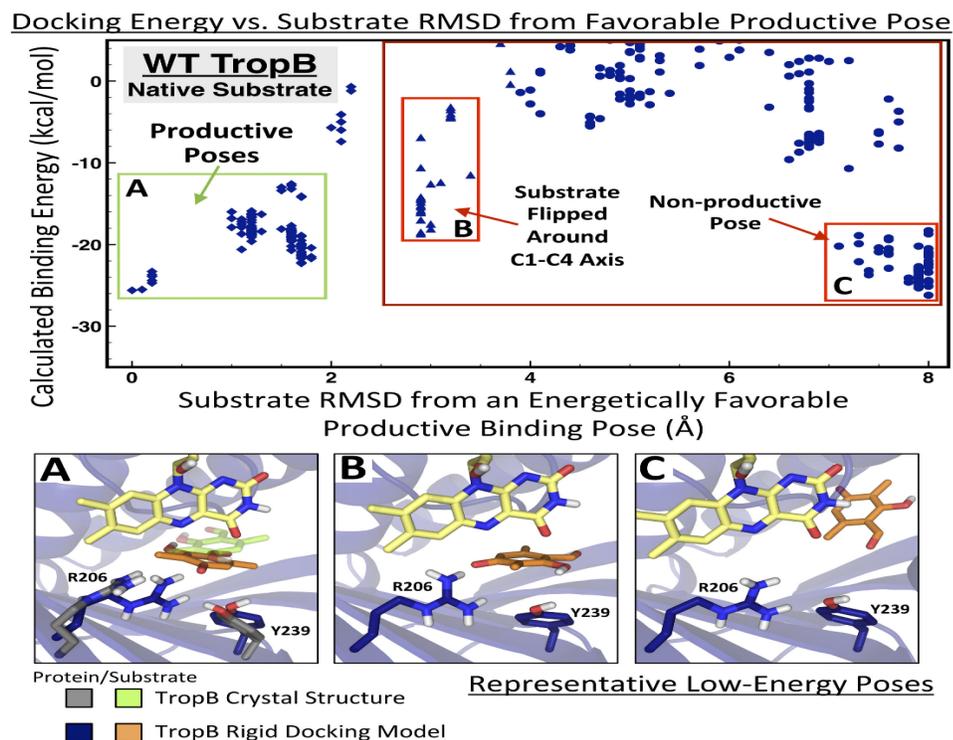
Figure 2.50: Rigid Docking Results for the Native Substrate. Graph of calculated binding energy vs. substrate RMSD from an energetically favorable productive pose (top) and representative low-energy poses (bottom) illustrating A) the productive docking pose compared with the substrate modeled from crystal structure density; B) the pose with the substrate flipped around the C1-C4 axis and presenting the opposite face to the flavin; and C) another energetically favorable pose with the substrate positioned far from the flavin such that hydroxylation with the substrate in this pose would not be productive. A pose was considered productive if the substrate presented the face consistent with the known stereochemical outcome and was positioned such that a c4a-hydroperoxy group on the flavin could attack the substrate (i.e. the substrate C3 carbon was 4 - 8 Åfrom the flavin c4a carbon). The flavin is represented in yellow sticks and the protein in the navy cartoon. The crystal structure side chains are represented in grey sticks while the substrate modeled into the crystal structure density is shown in panel A and represented in transparent lime-green sticks. The substrate poses generated by rigid docking are represented in orange sticks and the side-chain positions are shown in navy sticks. The productive docking poses generated by rigid docking of the native substrate also overlap well with the crystal structure density. The poses consistent with the known stereochemistry (A) are more energetically favorable than the poses where the substrate presents the opposite face (B). The other energetically favorable poses generated are non-productive poses with the substrate far from the flavin (C). Performed by Sara Tweedy.

Representative Low-Energy Poses For Additional Substrates

A R206 Y239 Sub 2.5 Native Substrate

B R206 Y239 Sub 2.6 Native Substrate

C R206 Y239 Sub 2.9 Native Substrate

D R206 Y239 Sub 2.11 Native Substrate
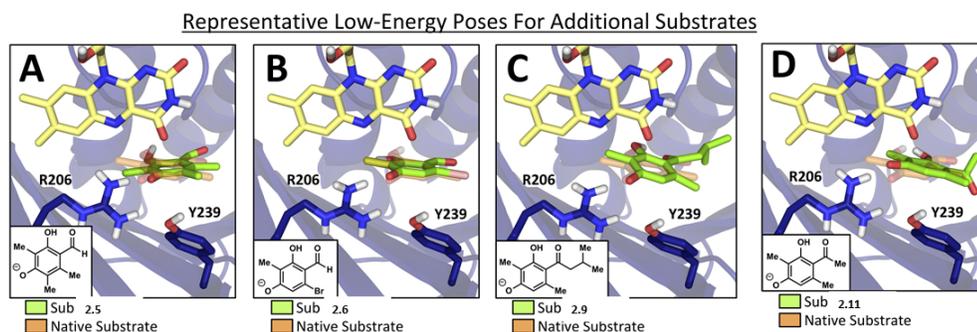
Figure 2.51: Rigid Docking of Other TropB Substrates with WT TropB. Docking of other TropB substrates reveals that they behave similarly to the native substrate and substrates presented in the main text. They form favorable, low energy poses that are consistent with the stereochemical outcome and preserve phenolate group hydrogen bonding with R206 with Y239 nearby. Representative, favorable poses are illustrated for substrates A). **2.5**; B). **2.6**; C). **2.9**; and D). **2.11**. The pose for each substrate (shown in solid green sticks) is overlaid against the pose for the native substrate (shown in transparent, orange sticks). The protein is shown in the navy cartoon, protein residues R206/Y239 are shown in navy sticks, and the flavin is shown in yellow sticks. Performed by Sara Tweedy

A). Substrate docking to the Y239F variant did not reveal the basis for reduced binding affinity as the results closely correspond to those for the native enzyme. Also, the crystal structure does not indicate that this substitution causes any significant change in the active site or enzyme in general. Thus, our results suggest that some other factor such as a change in the distribution of productive binding poses corresponding to a conformational selection model of substrate binding may play a role. In addition, charge transfer from the substrate to the protein surroundings observed in QM/MM calculations of the native (Y239) form and that contribute to substrate binding would be reduced in the F239 variant. Charge transfer effects are not accounted for in the docking calculations.

B). Substrate docking to the R206Q variant resulted in poses with favorable binding energies of similar magnitude as to the native enzyme. Yet, the poses revealed that the substrate was not in a position to be hydroxylated by the reactive hydroperoxyflavin intermediate. This result corresponds with the experimental results that indicate that the native substrate binds to this variant but that the substrate is not hydroxylated. Thus, a hydrogen bond donor is important for binding but alone insufficient for catalysis.

C). Substrate docking to the R206E variant resulted in only a small subset of poses with favorable binding energies of similar magnitude to the native enzyme. In addition, the few poses with the substrate positioned for hydroxylation were unfavorable compared to the native enzyme and the R206Q variant. The docking results suggest that binding to the active site with favorable energetics is less likely than in the native or R206Q and thus perhaps not sufficient to register a change in the spectra that can be observed.
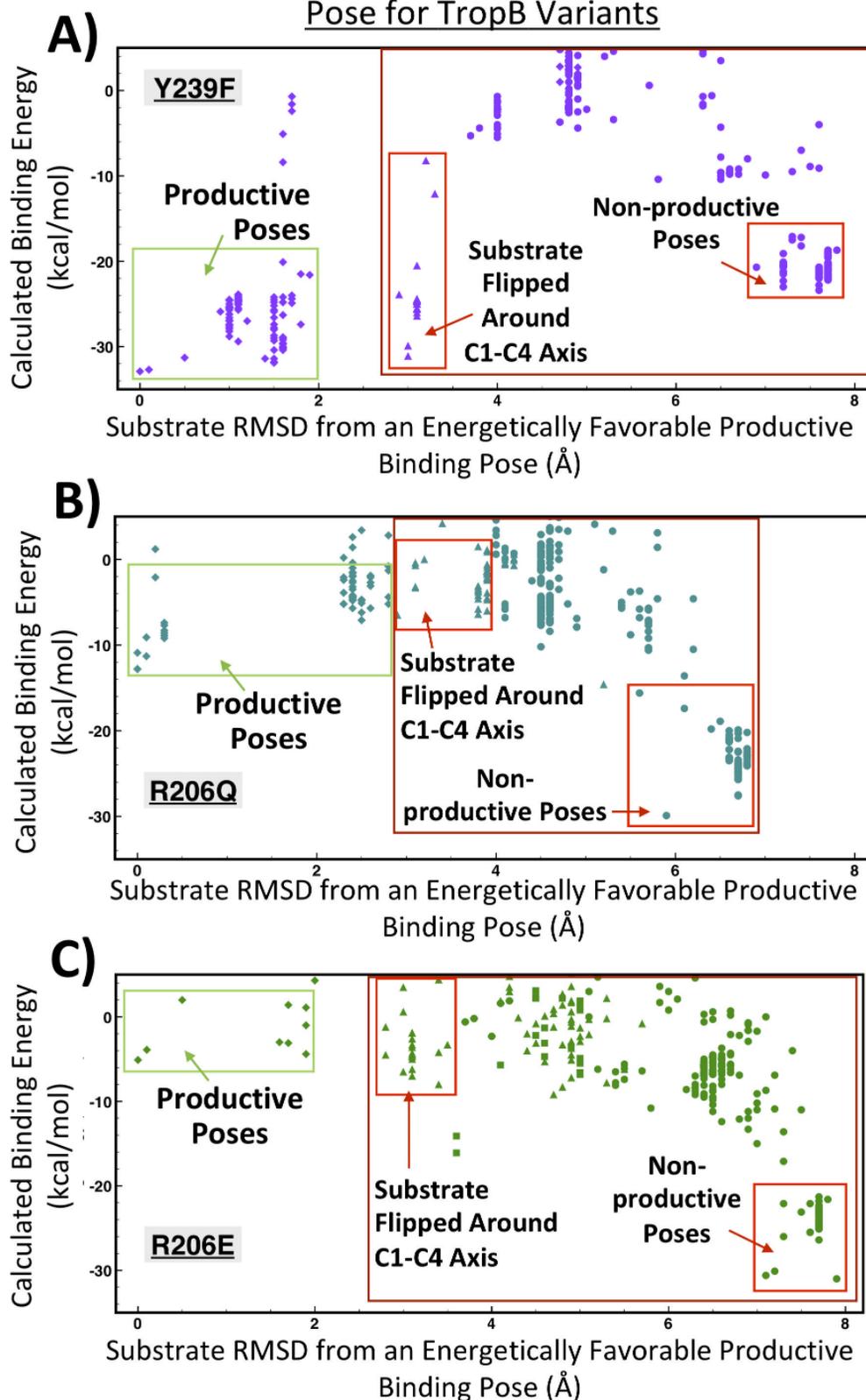
Figure 2.52: Rigid Docking of TropB Native Substrate with TropB Variants A) Y239F, B) R206Q, & C) R206E. Performed by Sara Tweedy.

66
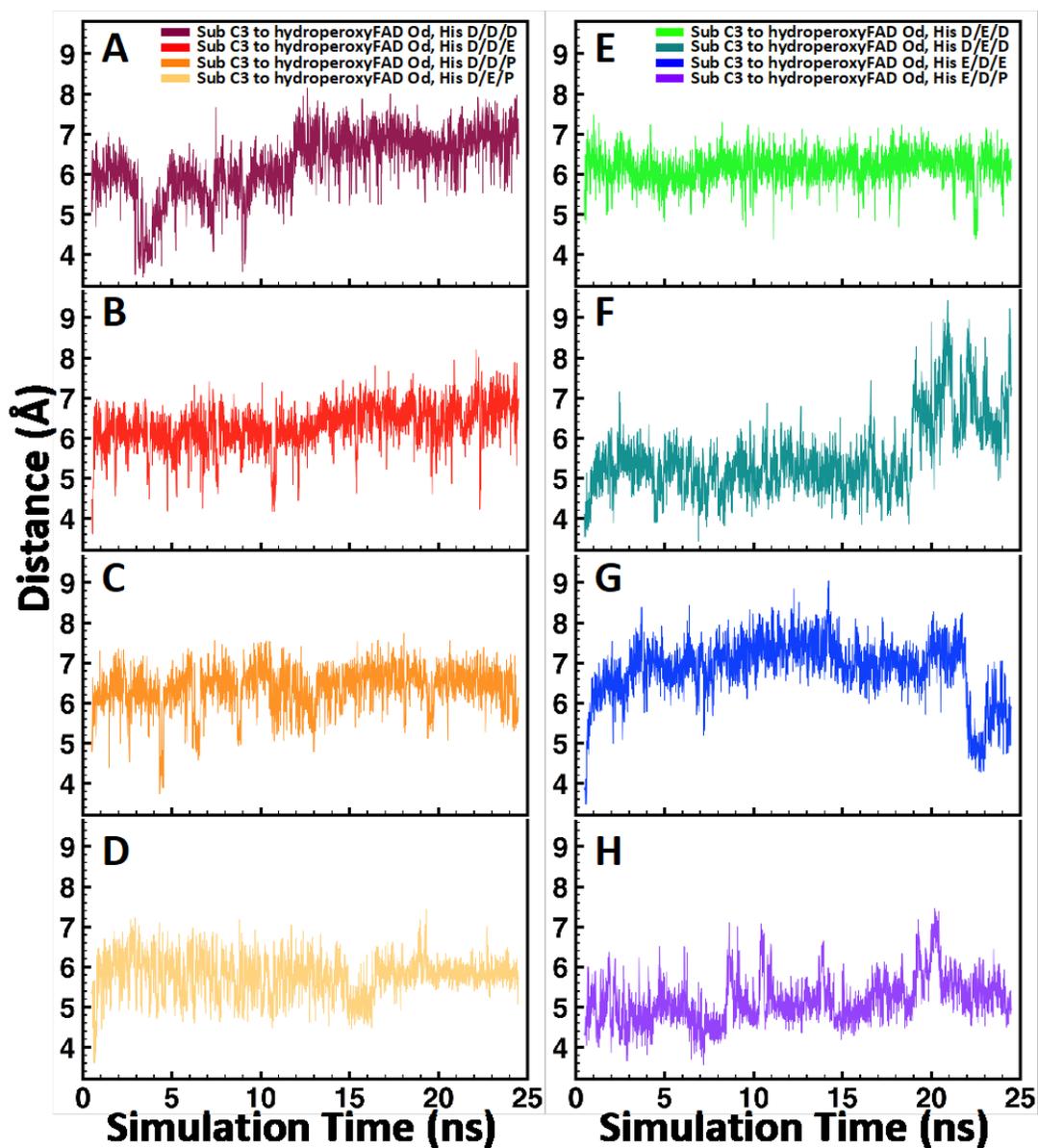
Figure 2.53: Substrate **2.1** C3-Odistal Distance in MD Simulations. The distance between the hydroperoxyflavin distal oxygen and substrate C3 is plotted versus the simulations time for 8 molecular dynamics (MD) simulations. Each simulation was initiated from the same coordinates and docking pose, but had different histidine tautomers for histidines 235, 330 and 331 (HSE or 'E' with the hydrogen on Nϵ, HSD or 'D' with hydrogen on Nσ, or the doubly protonated HSP or 'P'). The histidine tautomers used in each simulation (for His235/His330/His331 respectively) were: A) HSD/HSD/HSD – maroon; B) HSD/HSD/HSE – red; C) HSD/HSD/HSP – orange; D) HSD/HSE/HSP – yellow; E) HSD/HSE/HSE – green; F) HSD/HSE/HSE – teal; G) HSE/HSD/HSE – blue; H) HSE/HSD/HSP – purple. Performed by Sara Tweedy

	D/D/D	D/D/E	D/D/P	D/E/P	D/E/E	D/E/D	E/D/E	E/D/P
Average O$_{distal}$-C3 Distance (Å)	6.2 ± 0.8	6.2 ± 0.6	6.4 ± 0.5	5.8 ± 0.5	6.2 ± 0.4	5.5 ± 0.9	6.8 ± 0.6	5.2 ± 0.6
Average Substrate RMSD from Average Structure (Å)	0.54	0.52	0.65	0.55	0.40	0.40	0.60	0.24

Table 2.3: Average Odistal-C3 Distance and Substrate RMSD for MD Simulations. The average distance between the hydroperoxyflavin distal oxygen and substrate C3 shown for each MD simulation. The average root-mean-square coordinate deviation (rmsd) of the substrate from the average structure is also shown for each simulation. The simulations are labeled by the tautomers used for histidines 235/330/331 respectively, using D for proton on Nσ, E proton on Nϵ, and P for the positively charged His. Performed by Sara Tweedy
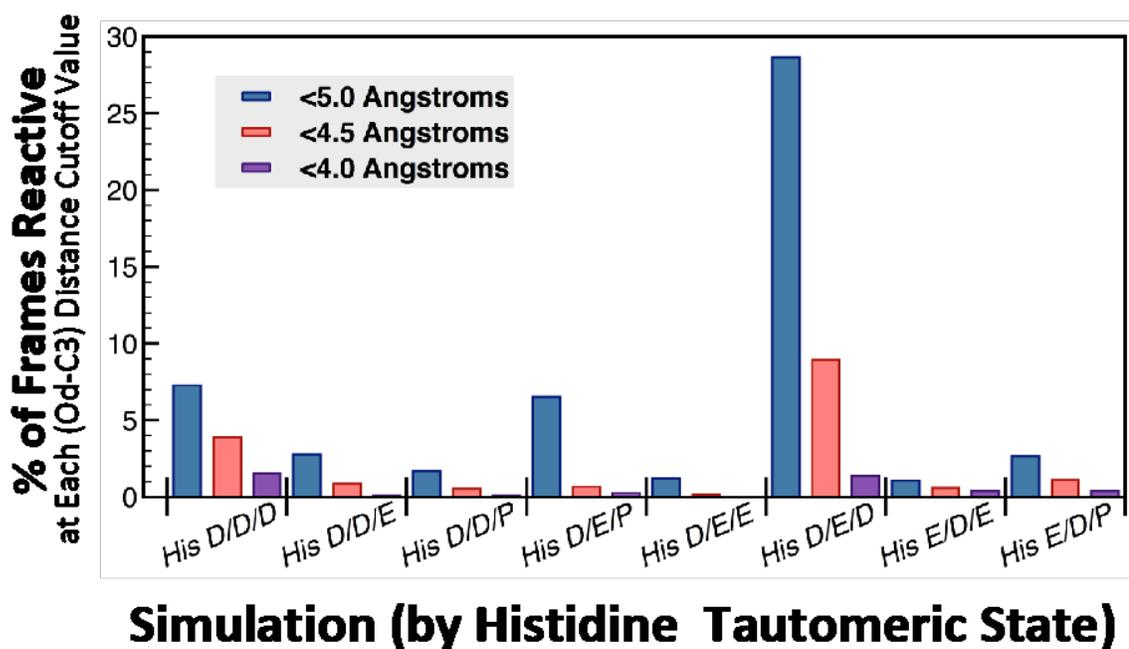
Figure 2.54: Fraction of MD Simulations Considered Reactive with Different Odistal-C3 Distance Cutoffs. Snapshots were considered reactive if the substrate presented the correct face to the hydroperoxyflavin, angle of approach (O$_{distal}$-C3-C3$_{methyl}$ angle between 60° and 115°), and a short distance between reactive entities (hydroperoxyflavin O$_{distal}$ to substrate C3 distance less than 5.0 Å– blue, 4.5 Å– red, or 4.0 Å– purple). The simulation with His235 D, His330 E, and His331 D (His D/E/D) was chosen for further analysis based on the high proportion of reactive frames. Performed by Sara Tweedy
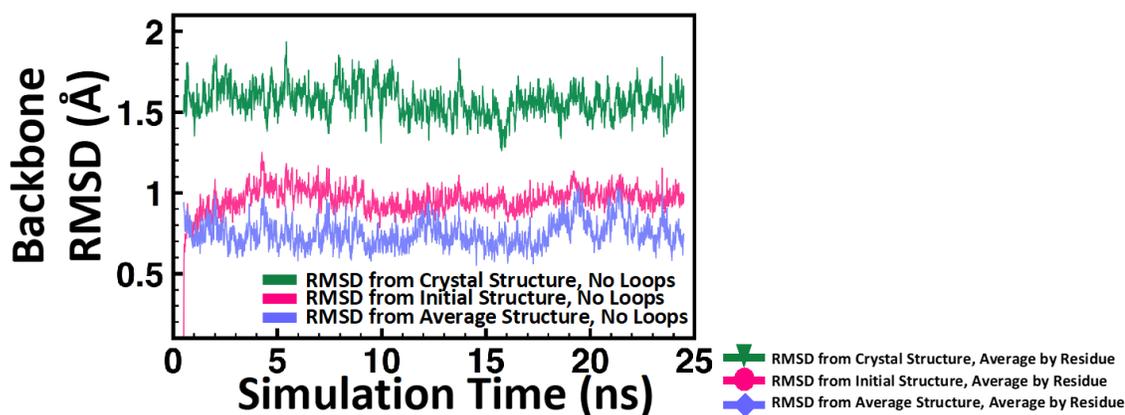
Figure 2.55: Backbone RMSD of the "D/E/D" MD Simulation. The root-mean-square co-ordinate deviation (RMSD) of the TropB backbone atoms, not including highly mobile loop regions (residues 45-55, 80-92, 153-154,190-191, and 255-267), is shown for the simulation. The RMSD was calculated after the structures were aligned for each frame with respect to 1) the crystal structure (green line), 2) the initial pose (pink line), and 3) the average structure (periwinkle line). The average RMSD over the simulation was 1.57 Åfrom the crystal structure, 0.96 Åfrom the initial pose, and 0.75 Åfrom the average structure. Performed by Sara Tweedy
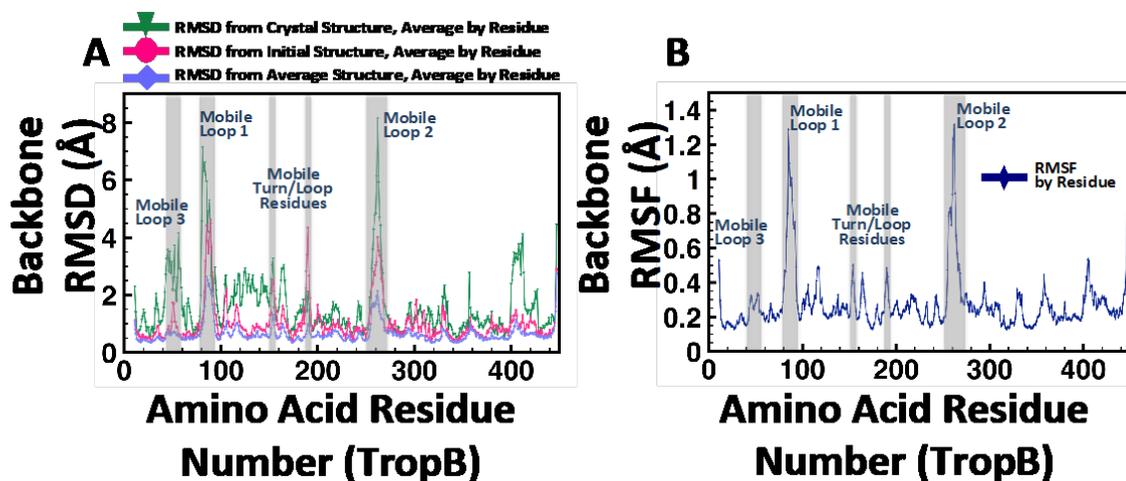
Figure 2.56: Fluctuation of TropB Residues Over "D/E/D" MD Simulation: (Left) The average root-mean-square coordinate deviation (RMSD) of the backbone atoms for each amino acid are shown. The RMSD was calculated after the structures were aligned with respect to 1) the crystal structure (green triangles), 2) the initial pose (pink circles), and 3) the average structure (periwinkle squares). (Right) The average root-mean-square coordinate fluctuation (RMSF) of the backbone atoms for each amino acid are shown (blue diamonds). Three mobile loop regions, two mobile loop turns, and the terminal amino acids show the greatest mobility and deviation from the crystal structure. Mobile loop 1 (residues 80-92) includes a loop region that has no density in the apo crystal structure and mobile loop 3 (residues 45-55) includes the region that becomes disordered (and loses electron density) upon substrate addition. Performed by Sara tweedy
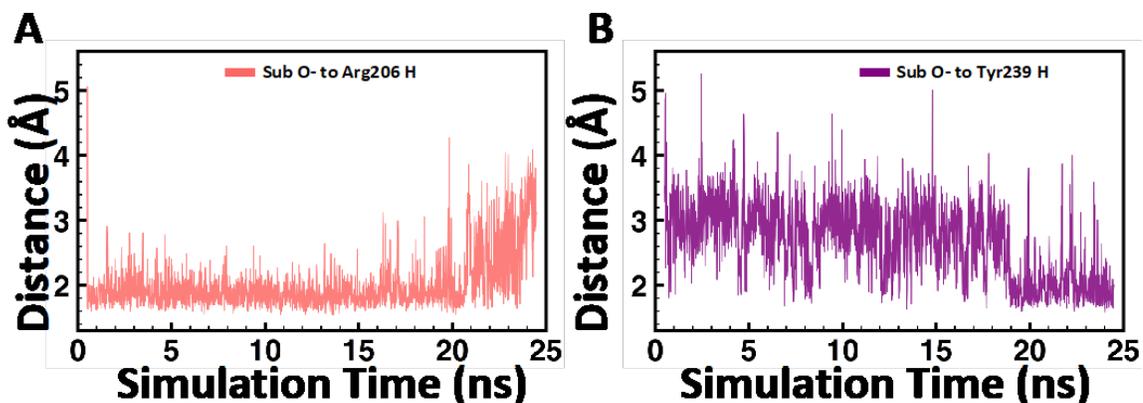
Figure 2.57: Hydrogen Bonding Between Substrate Phenolate Group and Arg206/Tyr239 for "D/E/D" MD Simulation: Distances are shown between the substrate phenolate group and A) Arg206 amine hydrogen (salmon lines), and B) Tyr239 hydroxyl-group hydrogen (purple lines). Performed by Sara Tweedy
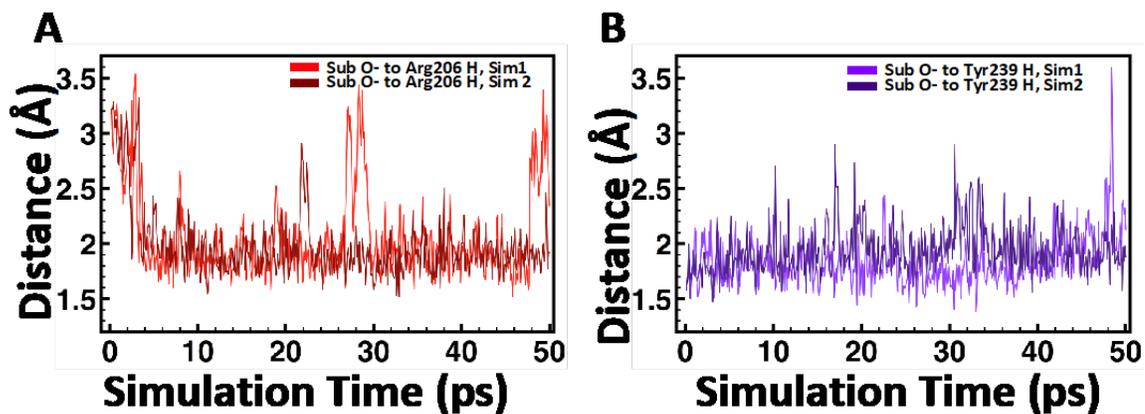
Figure 2.58: Hydrogen Bonding Between Substrate Phenolate Group and Arg206/Tyr239 for DFTB3/CHARMM Simulations of the Michaelis Complex: For the two analyzed simulations, distances are shown between the substrate phenolate group and A) Arg206 amine hydrogen (red lines), and B) Tyr239 hydroxyl-group hydrogen (purple lines). Performed by Sara Tweedy
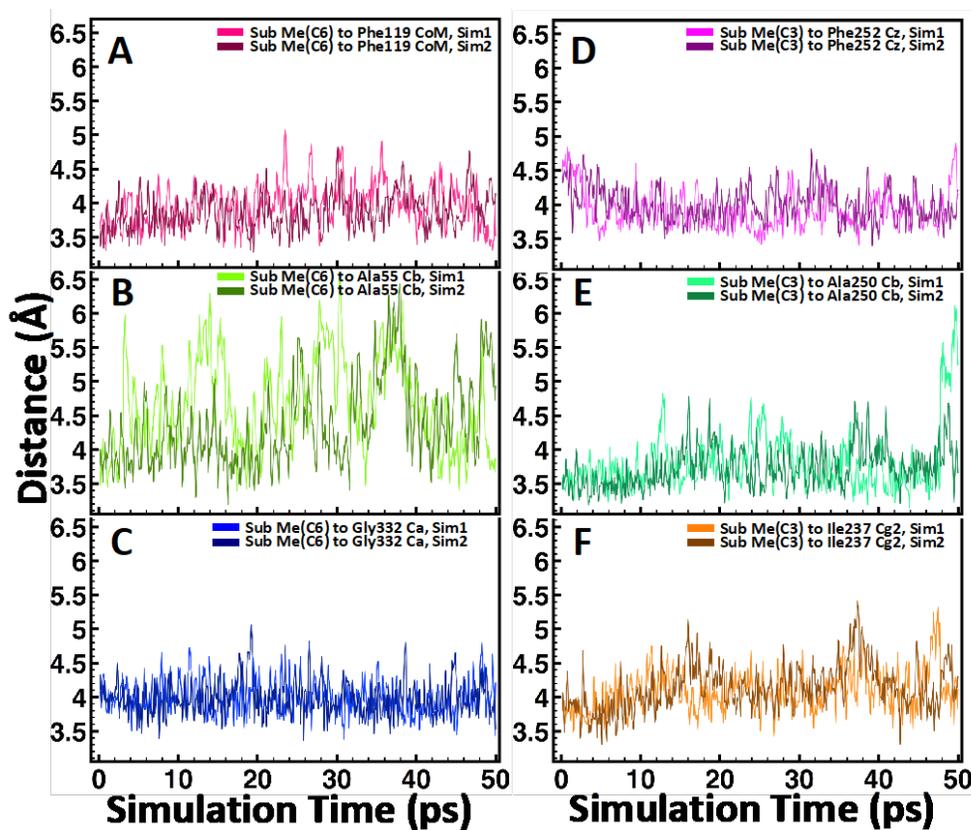
Figure 2.59: Nonpolar Contacts of Substrate Methyl Groups in DFTB3/CHARMM Simulations of the Michaelis Complex: Carbon-carbon distances are shown for the two analyzed simulations over the entire 50ps simulations. In the first column, distances are shown between the substrate C6-methyl group carbon and A) center-of-mass of Phe119 ring carbons (pink lines); B) Cβ of Ala55 (green lines); and C) Cα of Gly332 (blue lines). In the second column, distances are shown between the substrate C3-methyl group carbon and D) Cϵ of Phe252 (purple lines); E) Cβ of Ala250 (green lines); and F) Cγ2 of Ile237 (orange/brown lines). Performed by Sara Tweedy
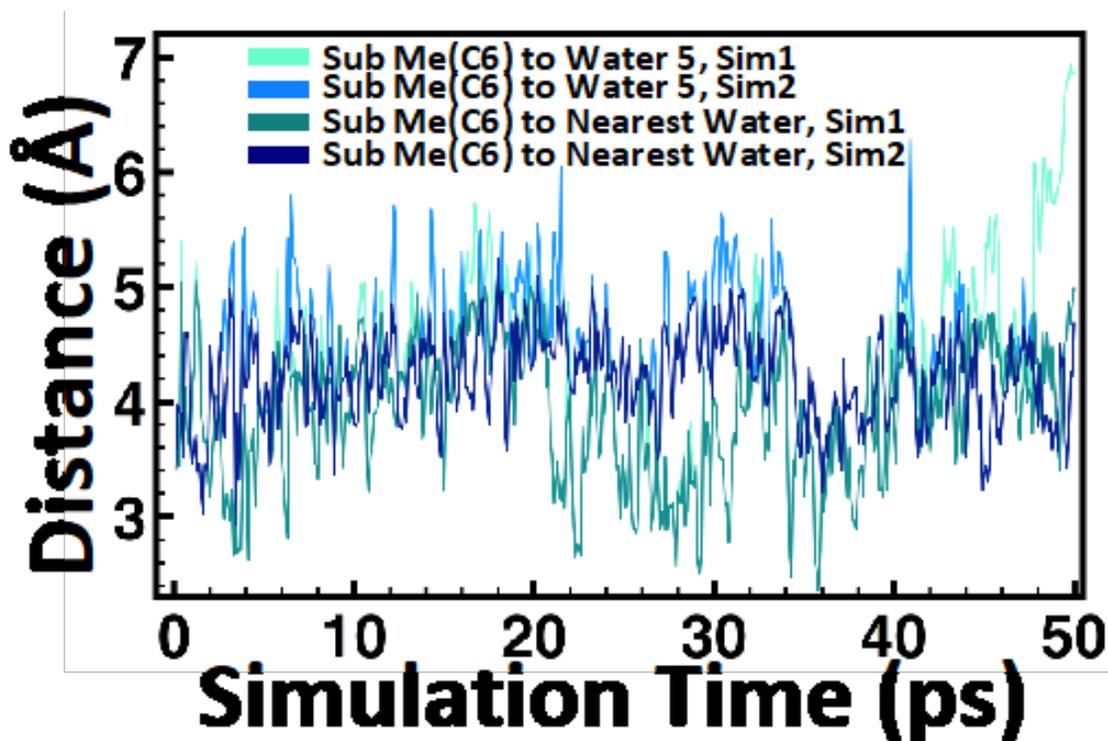
Figure 2.60: Solvation of Substrate C6-Methyl Group Position in DFTB3/CHARMM Simulations of the Michaelis Complex: The distance between the carbon of the substrate C6 methyl group and the nearest water molecule (dark blue/teal lines) is shown alongside the distance to one particular water molecule that is predominantly the closest or second closest water through both analyzed simulations (bright blue/turquoise lines). The nearest water molecule is more than 2Åaway throughout the simulation, and only one water molecule is ever within 3.5Å(the second closest is predominantly more than 4.5-5Åaway). Performed by Sara Tweedy
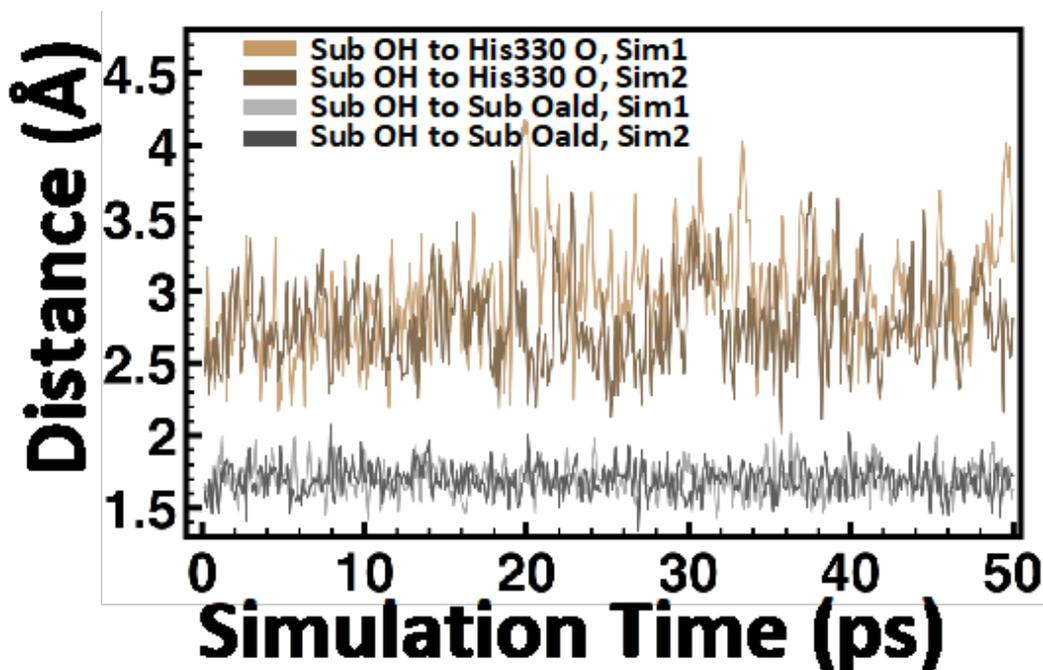
Figure 2.61: Intramolecular Hydrogen Bonding of Substrate C2-Hydroxyl Group in DFTB3/CHARMM Simulations of the Michaelis Complex: The distance between the substrate aldehydic oxygen and hydroxyl hydrogen is plotted alongside the distance between the hydroxyl hydrogen and the backbone carbonyl oxygen of His330. The C2-hydroxyl group is shown to be in close contact with the backbone carbonyl oxygen of His330 (tan/brown lines), but remains exclusively hydrogen bonded to the adjacent aldehydic oxygen (grey/dark grey lines). Performed by Sara Tweedy
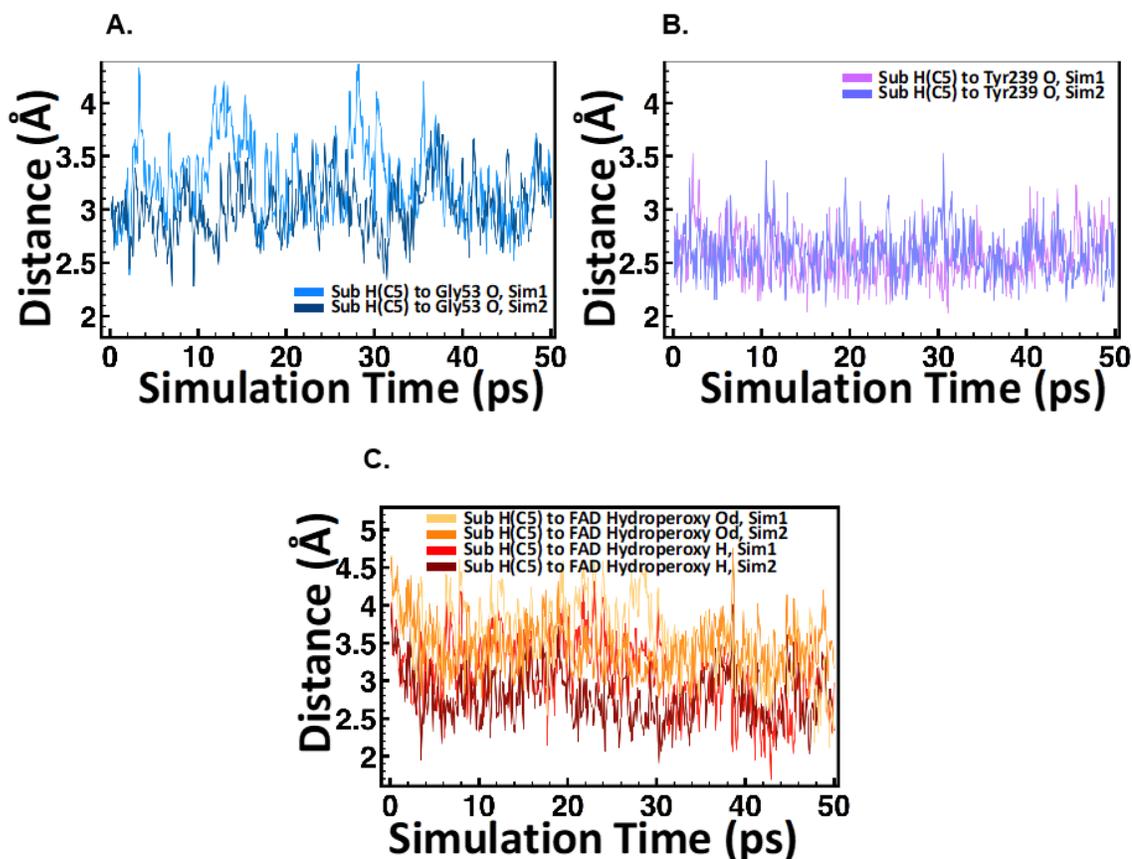
Figure 2.62: The Proximity of the Small C5-Hydrogen to A) Gly53, B) Tyr239, and C) theReactive Hydroperoxyflavin in DFTB3/CHARMM Simulations of the Michaelis Complex: The distance between the C5-hydrogen and the backbone carbonyl of Gly53 (A, top left) is shown in blue next to the distance between C5-H and the oxygen of the Tyr239 phenol group (B, top right; purple/pink) for both analyzed simulations. Below (C, red/orange lines) the distance between the C5-hydrogen and the distal oxygen of the hydroperoxy group is shown alongside the distance to the hydrogen on the hydroperoxy group. The C5-hydrogen is typically within 3 Åof one of those atoms. A larger and more hydrophobic substituent at C5, such as a methyl group, would be placed in contact with the hydrophilic backbone carbonyl, hydroxyl group, and hydroperoxy group. Performed by Sara Tweedy
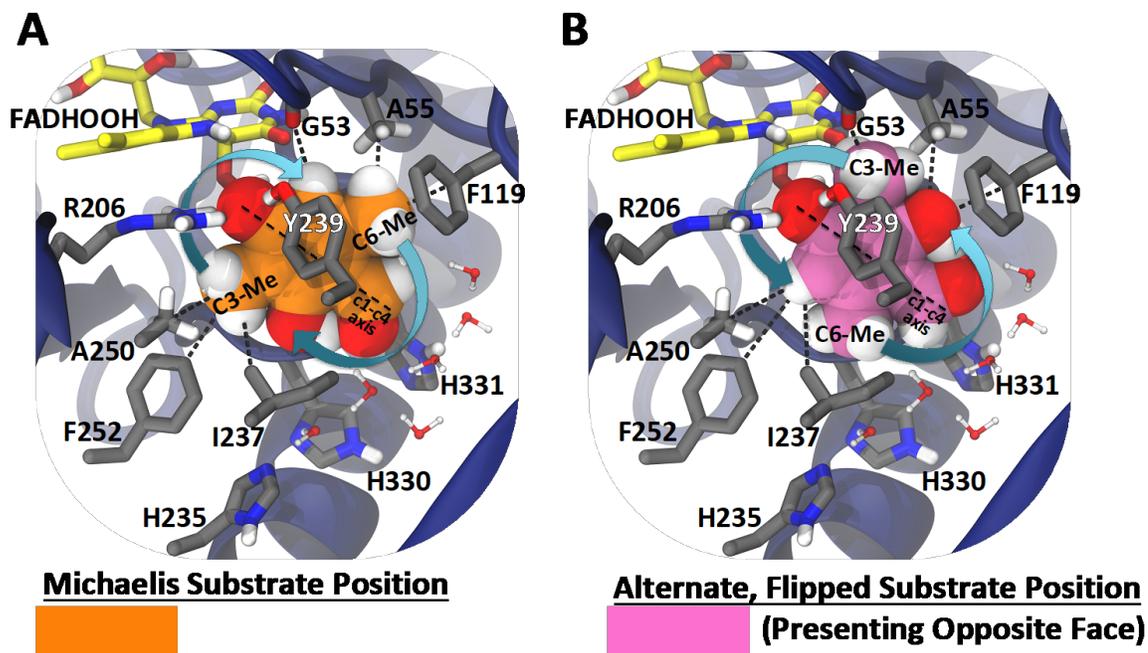
Figure 2.63: Interactions between Substrate Substituents and TropB Demonstrating Basis of Facial Selectivity: (left, A) A DFTB3/CHARMM geometry optimized structure of the enzyme-flavin Michaelis complex is illustrated with the flavin represented in yellow sticks, active-site amino acids represented with grey sticks, water molecules in CPK representation, and the enzyme mainchain as a blue ribbon. The substrate is depicted as spheres; carbons are orange. (right, B) positioning with the substrate flipped around the C1-C4 axis. Coloring as before but with carbons of substrate in pink. In the Michaelis complex, hydrophobic contacts between substrate methyl groups and Ala55, Ala250, Phe119, Phe252, and Ile237 are maintained while the hydroxyl group is positioned in a hydrophilic, solvated region. In contrast, when the substrate is flipped, the hydroxyl group is placed in the hydrophobic region near Ala55 and Phe119 while the C6 methyl group points toward the hydrophilic solvated region. In addition, the bulkier C3 methyl group clashes with nearby hydrophilic groups (Gly53 backbone carbonyl and Tyr239 hydroxyl group). Performed by Sara Tweedy

Profiling Function Across The Sequence Space of Flavin-Dependent Monooxygenase

Summary

Selective access to a targeted stereoisomer is often critical to the synthesis of biologically active molecules. Small-molecule reagents and biocatalysts are routinely available to access either enantiomer of a molecule. However, having a biocatalyst that can access the mirror-image of the same molecule is not readily accessible creating a challenge in the application of biocatalysis in synthesis. Here, I report an approach for accessing biocatalysts with complementary enantioselectivity. An orthogonal approach to protein engineering was employed using a sequence similarity network (SSN). Through this strategy, I iden-

tified enantiocomplementary biocatalysts, like AfoD, and enabled the enantiodivergent synthesis of natural products and analogs. I solved the structure of AfoD and identified the residues that have a potential role in determining the stereoselectivity of the product. With the sequence profiling strategy coupled with experimental data and the crystal structure of AfoD, I was able to select an FAD-dependent biocatalyst library for further testing across a panel of substrates.

3.1 Introduction

Isomers can have stark differences in their biological activities. For example, one enantiomer of the drug thalidomide is sold as a treatment for a wide array of cancers like myeloma and skin conditions like leprosy, while the other enantiomer causes severe teratogenic effects. [96] This example underscores the importance of developing methods that can selectively generate single isomers. Oxidative dearomatization is a reaction that can build complexity and chirality in natural product synthesis by introducing stereocenters in resorcinol scaffolds. [22] An example of a natural product that can arise from this reaction is the azaphilone family. These compounds are characterized by an oxygenated pyranoquinone bicyclic core bearing a single tetrasubstituted carbon Fig. 3.1. [97–101] Isolated from fungal sources, this large family of natural products is known to contain a diverse array of structural features that impart a wide range of biological properties including anticancer, [102–104] antiviral, [105] and anti-inflammatory activities. [104, 106] For example, luteusin A (**3.6**) and rubrotiorin (**3.5**) were found to inhibit the binding of the HIV surface glycoprotein gp120 to the human CD4 protein, making these natural products potential starting points for the development of therapeutic agents against the virus. [105] In addition, monascin (**3.7**), a tricyclic azaphilone natural product, has been shown to down-regulate steatohepatitis in a mouse model, indicating that these secondary metabolites have potential as therapeutics for nonalcoholic fatty liver disease. [107] As illustrated in Fig. 3.1, azaphilone natural products can contain either the *R*- or the *S*-configuration
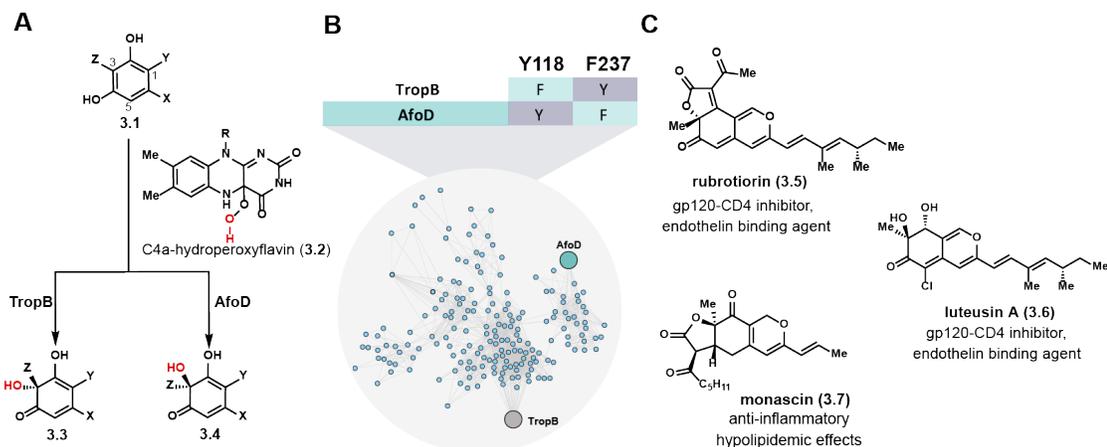
Figure 3.1: A. Reaction scheme for the oxidative dearomatization reaction performed by AfoD and TropB using the C4a-hydroperoxyflavin as the oxidant. B. Sequence similarity network containing with TropB highlighted in gray and AfoD in teal highlighting the change in residues from AfoD to TropB C. Biologically active azaphilone natural products feature both the *R*- and the *S*-configurations at the C7 stereocenter

at the C7-position. Several cases of epimeric azaphilones are known, in which each C7-epimer is produced by a distinct fungal source. [108, 109] The pharmaceutical potential of these molecules has been demonstrated through initial in vitro and cell-based assays of isolated natural products. [110–112]

Furthermore, the application of asymmetric oxidative dearomatization to the synthesis of some tricyclic members of the azaphilone family is yet to be reported. In contrast, Nature has evolved a highly efficient catalyst suite for oxidative dearomatization, a family of flavin-dependent monooxygenases (FDMOs) capable of installing a hydroxyl group with perfect site- and stereoselectivity and with low catalyst loading, mild reaction conditions, with molecular oxygen as the stoichiometric oxidant Fig. 3.1. [11] Selectivity in this class of enzymes is dictated by the substrate binding mode within the active site, often resulting in a generation of a single enantiomer of the product. However, access to both enantiomers is often required to produce scaffolds for natural products, requiring protein engineering and rescreening of enzyme variants to access the desired enantiomer. A structural and mechanistic understanding of FDMO enzymes can create a blueprint for

a unified approach to protein engineering strategies and tight biocatalytic control over stereoselectivity.

Toward this goal, I identified catalysts with complementary selectivity by canvassing the sequence space surrounding TropB in a sequence similarity network (SSN) of putative FDMOs. The sequence space around TropB will provide biocatalysts that can perform similar reactions but with structural differences that can result in complementary selectivity outcomes. This method of selection can be leverage for selecting a library that works across a panel of substrates.

3.2 Sequence Profiling of Flavin-Dependent Monooxygenases

To survey how the selectivity of flavin-dependent monooxygenases is related by sequence, I constructed a sequence similarity network (SSN) of putative flavin-dependent monooxygenases Fig. 3.2A. [25, 29, 113] SSNs are visual representations of the relatedness of protein sequences, which cluster on the basis of similarity thresholds. [25] This approach can enable the rapid and logical investigation of proteins likely to demonstrate similar reactivity and/or selectivity, dramatically reducing the time and search space required during screening, considerations critical to improving the compatibility of biocatalytic approaches and traditional synthetic strategies. The protein family (Pfam01494) of flavin-dependent monooxygenases includes over 45,000 enzyme sequences, which are included in the full SSN displayed in Fig. 3.33. This data set was truncated by limiting the search to edges possessing an alignment score greater than 110, which returned 1211 sequences Fig. 3.2A. We noted that the enzymes we had previously investigated, AzaH, TropB, and SorbC, were each located in distinct groups or "clusters" within the network, and that a fourth tight clustering of sequences formed between the TropB and SorbC clusters containing an enzyme associated with asperfuranone biosynthesis, AfoD. [19, 114, 115] Previous in vivo studies [115] and bioinformatic analysis [114] indicated that AfoD is responsible for oxidative dearomatization of an asperfuranone precursor with the same site-selectivity

as TropB and AzaH; [114] however, the absolute configuration of asperfuranone suggests that AfoD carries out this transformation with the opposite facial selectivity. [19]
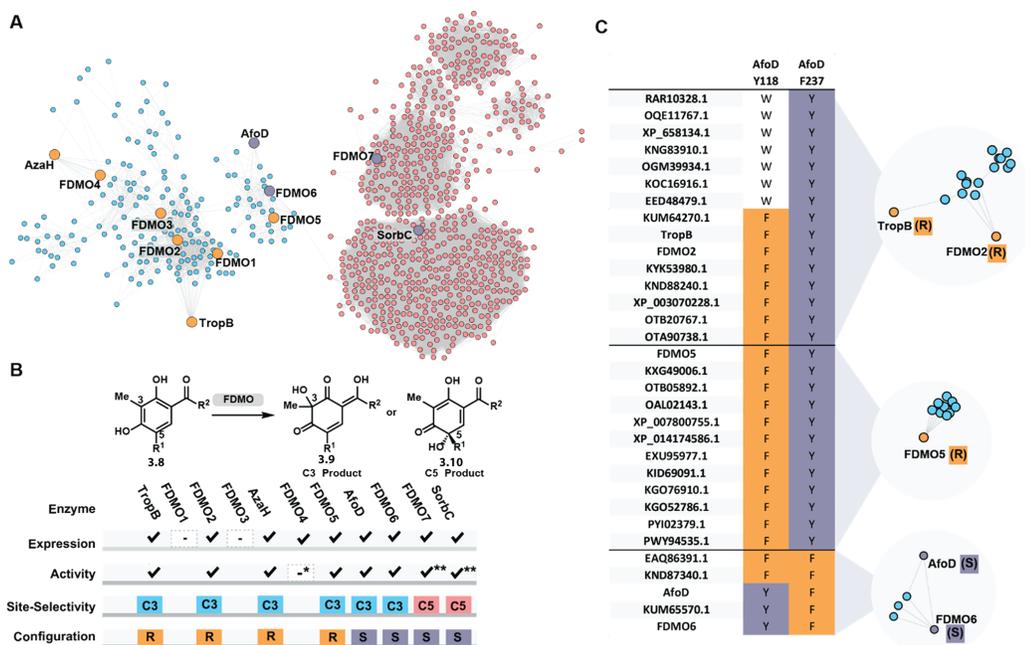
Figure 3.2: (A) Sequence similarity network (SSN) of FAD-dependent monooxygenases (Pfam01494) using a sequence alignment score of 110. (B) Results of expression, activity with model substrate **3.11** or **3.14**** with R$_1$ = H, and site- and stereoselectivity of enzymes chosen from the SSN in panel A. *FDMO4 demonstrated <10% conversion by UPLC with substrate **3.11**. (C) Selected clusters from a more stringent SSN generated with an alignment score of 150 and corresponding analysis of the multisequence alignment of each cluster.

Analysis of sequence alignments between each of these clusters revealed conserved residues at positions 118 and 237 (AfoD numbering), Fig. 3.1B. In general, catalysts that generate products with the *R*-configuration possess a tyrosine at position 237 and an aromatic residue such as phenyalanine or tryptophan at position 118. In contrast, these conserved residues switch positions in biocatalysts that afford products with the *S*-configuration. We previously demonstrated that a two-point coordination of the phenolate substrate by Tyr237 and an arginine residue is critical for positioning the substrate within the active

site in the R-selective enzyme, TropB. [31, 116].

3.3 AfoD Crystal Structure

A sequence alignment of AfoD and TropB revealed that many positions in the substrate-binding pocket have different amino acid side chains. Notably, AfoD possesses a Phe237 analogous to the critical substrate-orienting Tyr239 in TropB. Therefore, to elucidate the basis for selectivity, we solved a crystal structure of AfoD to a 2.0 Å resolution with bound flavin adenine dinucleotide (FAD) cofactor Fig. 3.3B and 3.2. In contrast to dimeric TropB, AfoD was monomeric in crystals and in solution Fig. 3.4. Typical of class A FDMOs, AfoD is composed of a FAD-binding domain with a site for the adenosine diphosphate (ADP) and a catalytic domain. As expected for proteins with 37 % sequence identity, the overall structures of AfoD and TropB are similar (RSMD 1.30 Å for 334 Cα atoms).
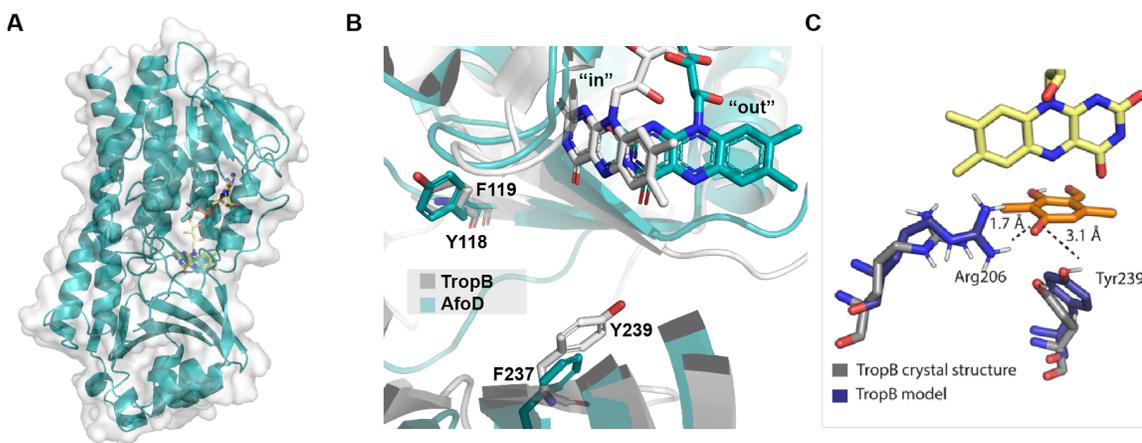
Figure 3.3: AfoD monomer with the flavin cofactor highlighted in yellow. Zoom-in of the active site of AfoD shown with surface contour and overlaid with TropB co-crystallized with its native substrate PDB: 6NET B. AfoD active site in teal overlaid with TropB (PDB:6NET) in light grey. C. TropB crystal structure (PDB:) overlaid with computational model of TropB showcasing the two point binding mode to the substrate engaging Tyr239 and Arg206.
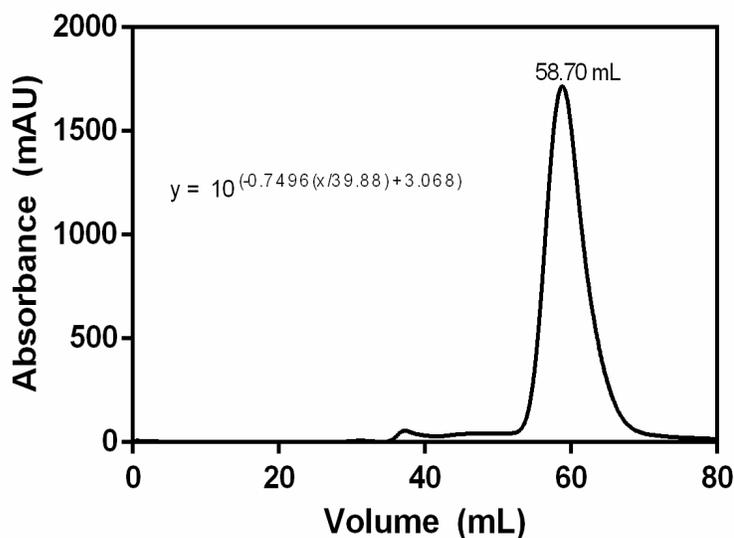
Figure 3.4: Size-exclusion elution profile of AfoD.

The class A FDMOs employ a "wavin' flavin" mechanism for cofactor reduction in which the FAD isoalloxazine ring rotates between an "in" position for substrate engagement and an "out" position for reduction by NAD(P)H. [14] To determine which class the AfoD enzyme belongs to we performed a reaction in the presence and absence of substrate while monitoring the consumption of NADPH. This experiment was to determine if the addition of substrate is required for cofactor reduction. The NADPH depletion was monitored at 362 nm, the isosbestic point between the product and starting material. The reaction was carried out in the presence and absence of substrate. Upon the addition of substrate to AfoD, the rate of oxidation of NADPH was enhanced suggesting that AfoD is a class A FDMO Fig. 3.5. In the AfoD crystal structure, the FAD isoalloxazine ring is in the "out" position where it π-π stacks with Phe297. This is a striking contrast to the "in" flavin position observed in the substrate-free TropB crystal structure. However, substrate is required in both AfoD and TropB to enhance the rate of FAD reduction during catalysis Fig. 3.5. We hypothesize that Phe297 is less bulky than Trp299 in TropB in an analogous position which allows freer movement of the FAD cofactor providing a stable conformation in the AfoD crystals.
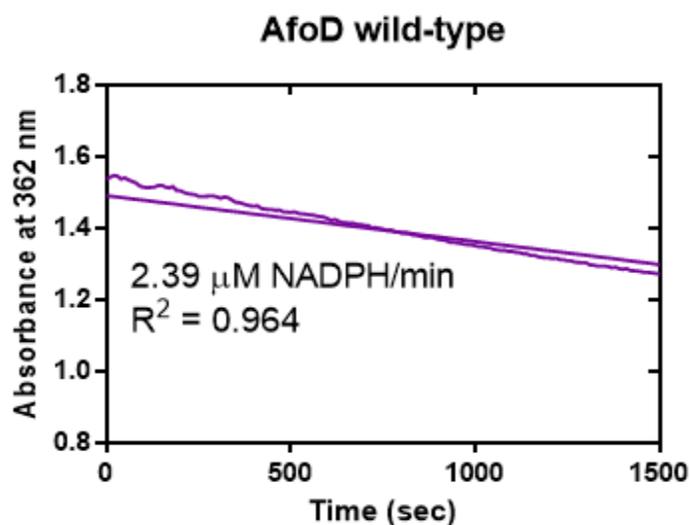
Figure 3.5: NADPH Depletion Assay monitored at 362 nm. Reaction of AfoD without the presence of substrate. Data is fitted to tangent of the curve to determine the rate of consumption of NADPH.

The sequence identity between TropB and AfoD is 37 percent, therefore the overall fold is the same with some amino acid changes that might give rise to their distinctive enantioselectivity. In the active site cavity, Phe237 occupies the position of TropB Tyr239, whose phenolic group is essential for catalysis. [31] On the opposite side of the AfoD active site, Tyr118 (Phe119 in TropB) is near the FAD isoalloxazine ring where it could fulfill a similar role that of Tyr239 in TropB Fig. 3.3B. The opposite positions of these phenolic functional groups render the active sites mirror images of one another and could be a simple way for Nature to control selectivity.

3.4   Mutagenesis and Catalysis of AfoD

Based on the crystal structures and sequence alignment of AfoD and TropB, AfoD Tyr118 and Phe237 are potential determinants of substrate positioning in the active site. To probe the role of these residues in catalysis and in determining facial selectivity, we generated AfoD variants Y118F and F237Y Fig. 3.6.
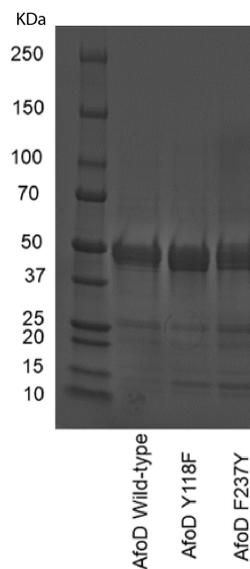
Figure 3.6: AfoD Wild-type, AfoD(Y118F), AfoD(F237Y), SDS-PAGE gel. The gel was stained with Quick Coomassie stain (Anatrace) and visualized with the Azure Gel Imaging System. Units in the left side of the image are in kDa

To measure activity, we used the model substrate (**3.11**) Fig. 3.12, which mimics AfoD native substrate key moieties but with a shorter C6 aliphatic side chain. AfoD variants were assayed though analytical in vitro reaction with substrate (**3.11**) by monitoring substrate consumption by ultra high-performance liquid chromatography coupled to a photodiode array ultra high-performance liquid chromatography coupled to a photodiode array (UPLC-DAD) Fig. 3.7A and B. AfoD Y118F and F237Y both possessed catalytic activity generating product (**3.12**) that spontaneously cyclizes to (**3.13**) Fig. 3.7 B and C.
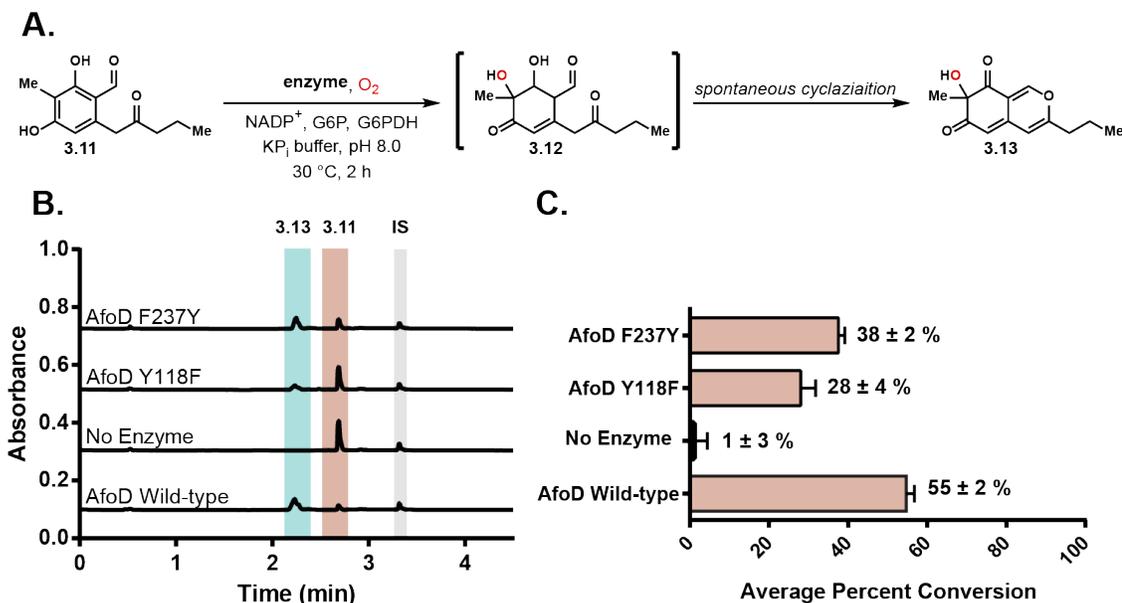
Figure 3.7: A. Reaction scheme for in vitro reaction with substrate (**3.11**). B. UPLC- Photodiode Array PhotoDiode Array (PDA) traces highlight the starting material (**3.11**) and product (**3.13**). C. Average percent conversion from in vitro reactions based on the substrate (**3.11**) calibration curve Fig. 3.8. Reaction conditions: 2.5 mM substrate 3.11, 20 μM enzyme, 1 mM NADP+ , 5 mM glucose-6-phosphate (G6P), 1 μmL-1 glucose-6-phosphate dehydrogenase (G6PDH), 50 mM potassium phosphate buffer pH 8.0, 30 °C, 1 h. C. Average percent conversion from in vitro reactions based on the substrate (**3.11**) calibration curve. *Reaction conditions: 2.5 mM substrate **3.11**, 20 μM enzyme, 1 mM NADP+ , 5 mM glucose-6-phosphate (G6P), 1 μmL-1 glucose-6-phosphate dehydrogenase (G6PDH), 50 mM potassium phosphate buffer pH 8.0, 30 °C, 1 h. IS = internal standard, pentamethylbenzene

[SM] mM	Average	SD
2.50	13.7	1.2
1.25	6.30	0.02
0.63	2.916	0.009
0.31	1.417	0.005
0.16	0.66	0.03
0.08	0.306	0.006
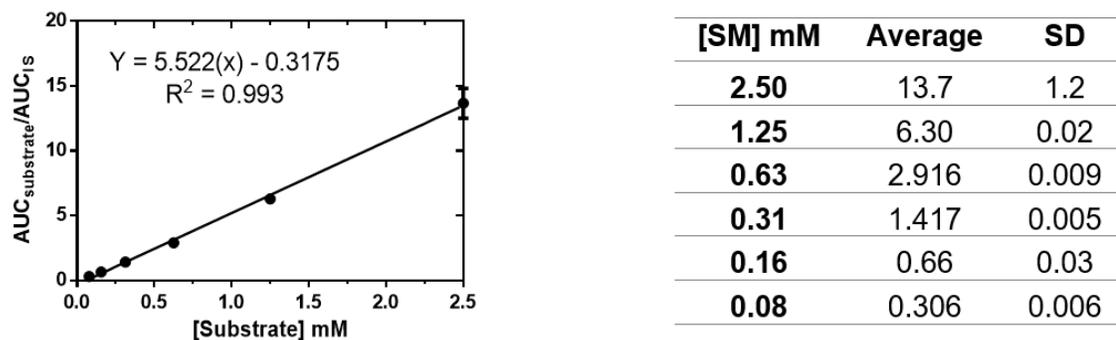
Figure 3.8: Calibration Curve of substrate **3.11**

AfoD wild type has an average conversion of 55% with substrate (**3.11**) and AfoD variants Y118F and F237Y 28% and 38% conversion, respectively, Fig. 3.7 C and Fig. 3.10.

In contrast to TropB, where Phe substitution to the catalytic Tyr239 resulted in no detectable product and hydrogen peroxide generation, the AfoD substitutions generated equal amounts of hydrogen peroxide compared to AfoD wild type, allowing us to determine the effect of amino acid substitutions on enantioselectivity Fig. 3.30.

AfoD Y118F

Instrument Method	Inj. Vol.	Solvent	Column	Sample	Well Location	Temp. (C)	Flow (g/min)	% Modifier	Pressure (Bar)
AD-H_30%_300-330	6 uL	Isopropanol	AD-H	ARB-V-071 AfoD_1	12A	40	3.5	30	120

Peak No	% Area	Area	Ret. Time	Height	Cap. Factor
1	46.8333	12090.3845	2.92 min	1322.9644	0
2	53.1667	13725.3924	4.16 min	1541.5059	0

AfoD F237Y

Instrument Method	Inj. Vol.	Solvent	Column	Sample	Well Location	Temp. (C)	Flow (g/min)	% Modifier	Pressure (Bar)
AD-H_30%_300-330	4 uL	Isopropanol	AD-H	AfoD_2_prep2 ARB-V-071	11A	40	3.5	30	120

Peak No	% Area	Area	Ret. Time	Height	Cap. Factor
1	28.8486	2149.7018	2.97 min	332.0461	0
2	71.1514	5301.9567	4.02 min	631.6176	0

Figure 3.9: PDA traces of AfoD Y118F and AfoD F237Y mediated oxidative dearomatization to yield 3.13
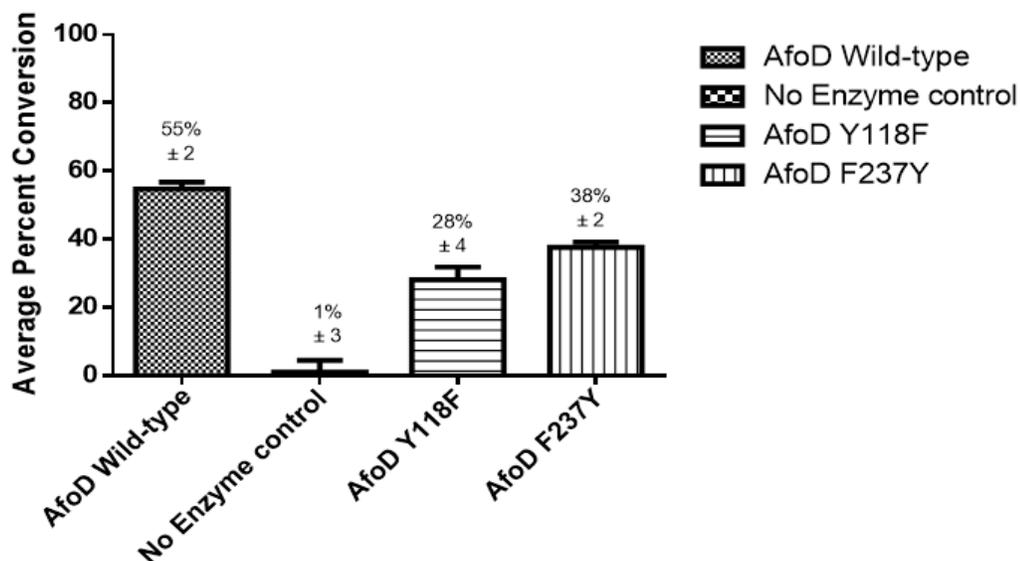
Figure 3.10: Summary of *in vitro* reactions with substrate **3.11**. Reactions were done in triplicate.

3.5 Evaluation of AfoD variants stereoselectivity

Analysis of sequence alignments between each of these clusters revealed conserved residues at positions 118 and 237 (AfoD numbering). In general, catalysts that generate products with the *R*-configuration possess a tyrosine at position 237 and an aromatic residue such as phenylalanine or tryptophan at position 118. In contrast, these conserved residues switch positions in biocatalysts that afford products with the *S*-configuration Fig. 3.2 C. We previously demonstrated that a two-point coordination of the phenolate substrate by Tyr237 and an arginine residue is critical for positioning the substrate within the active site in the *R*-selective enzyme, TropB. [31, 116] To probe the role of Tyr118 in the stereoselectivity provided by AfoD, the AfoD Y118F and F237Y variants were generated. Both AfoD single variants were reacted with resorcinol **3.11** in preparative-scale reactions to generate isolable quantities of product **10**. I previously reported that the AfoD wild-type and Y118F yielded the dearomatized product in 1:99 and 53:47 [30] (S:R) enantiomeric ratio enantiomeric ratio (er), respectively, (Fig. 3.11 and Fig. 3.30), highlighting the importance

of Tyr118 in controlling facial selectivity in the binding site of AfoD. However, with AfoD F237Y, the dearomatized product was formed in 29:71 (S:R) er (Fig. 3.11). AfoD F237Y has a phenolic functional group in the analogous position to the catalytic TropB Tyr239, but also has a second phenolic group at Y118 on the opposite side of the active site Fig. 3.3 B. I hypothesize that this led to the formation of some *R* enantiomer, as formed by TropB.
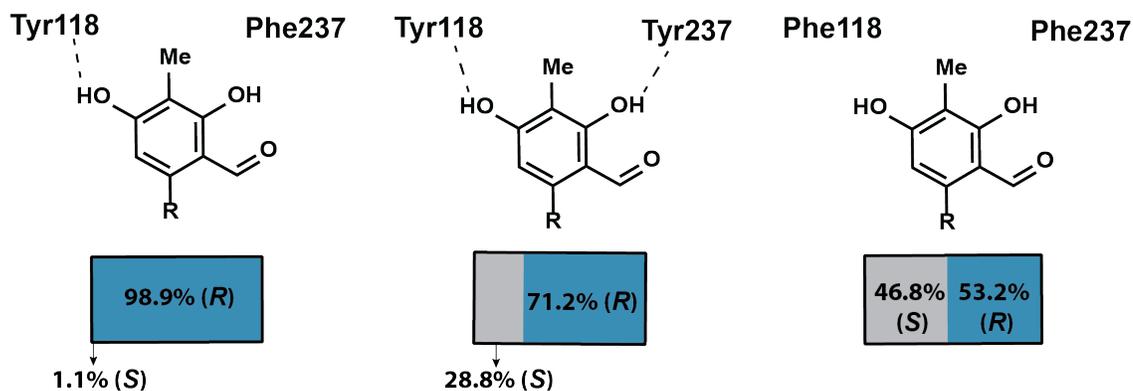
Figure 3.11: Representation of substrate interaction with residues in the active site of the AfoD wild-type and single variants with the enantiomeric ratio across different AfoD variants. Preparative–scale enzymatic reactions were conducted on 10 mg of substrate (3.11) under the following conditions: 20 μM enzyme, 1 mM NADP+ , 5 mM glucose-6-phosphate (G6P), 1 μmL-1 glucose-6-phosphate dehydrogenase (G6PDH), 50 mM potassium phosphate buffer pH 8.0, 30 °C.

3.6 Flavin-Dependent Monooxygenase Library

We hypothesized that sequences with high similarity to TropB, AzaH, AfoD, and SorbC would encode enzymes that possess the ability to carry out oxidative dearomatization reactions, and further questioned if trends in site- and stereoselectivity could be predicted on the basis of sequence. To test this hypothesis, seven sequences were selected on the basis of proximity to either TropB, AzaH, AfoD, or SorbC in the SSN labeled FDMO1 to 7 in Fig. 3.2A. Synthetic genes corresponding to the four known enzymes and FDMO1 to 7 were transformed into *E. coli* BL21(DE3) cells. Under standard expression conditions, nine proteins were successfully obtained, whereas FDMO1 and FDMO3 proved insoluble un-

der these conditions (Fig. 3.2 B). The reactivity of the nine soluble enzymes was evaluated with two substrates, **3.11** and **3.14** (Fig. 3.12).
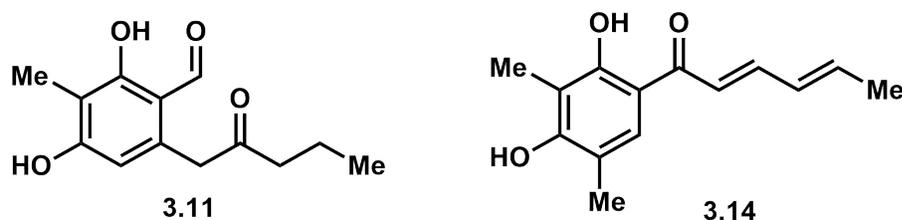
Figure 3.12: Model Substrates tested with the FDMO library

Gratifyingly, all nine enzymes showed activity with one of the two model substrates, eight of which displayed more than 40 percent activity, for product isolation from preparative-scale reactions (Fig. 3.2 B). A single product was obtained from each biocatalytic reaction, and a strong trend in site-selectivity was clear from these data: FDMO7, the enzyme most similar to SorbC, produced C5-hydroxylated product, whereas the remaining enzymes more similar to TropB, AzaH, and AfoD delivered C3-hydroxylated products. However, the trend between location on the SSN (Fig. 3.2 A) and stereoselectivity was not apparent. Increasing the E value from 110 to 150 produced an SSN that provided greater insight into the relationships between sequences in this family (Fig. 3.34). Notably, previously clustered sequences associated with divergent selectivity, such as AfoD and FDMO5, now clearly separated into distinct clusters (Fig. 3.2 C). Analysis of sequence alignments between each of these clusters revealed conserved residues at positions 118 and 237 (AfoD numbering).

Beyond the importance of the Tyr position in the active site, it is clear that other mechanisms for stereocontrol have evolved in this class of enzymes. For example, catalysts AzaH and SorbC break from this Tyr control mechanism, and, in the case of SorbC, I have proposed an alternative mechanism for control of substrate position in the active site. [117] I uncovered that the SorbC enzyme employs hydrophobic pocket to dock the substrate in the active site using the C1 aliphatic sidechain of the substrate. With AzaH I have observed it works with more electron deficient substrates, therefore basic or acidic

residues might be involved in the mechanism. Ultimately, this survey of sequence space surrounding known enzymes provides a greater understanding of the sequence features that can predict site- and stereoselectivity and has increased the number of biocatalysts vetted for this transformation.

With this goal in mind I have set to explore 500 putative FMDOs as biocatalysts employing a sequence and structure-guided approach. This will potentially generate a library of hydroxylated products across a library of substrates. This approach could lead to curating more promiscuous enzymes and with complementary activity, but it could also lead to insoluble biocatalysts or inactive biocatalysts towards or selected substrates. To overcome this challenge we aim to test challenging enzymes for solubility under our conditions and test them against a chemically diverse library compared to our traditional substrates.

3.7   Conclusion

Through structural insights gained from the AfoD crystal structure and sequence analysis, we have uncovered the origins of AfoD stereoselectivity. Facial selectivity is abolished in AfoD Y118F, demonstrating that hydrogen bonding with this residue may be critical for stereoselectivity. Moreover, AfoD F237Y introduced an additional hydrogen bonding partner in the active site, leading to the formation of the *S* enantiomer. These findings illustrate the structural features that control stereoselectivity in the reaction mediated by AfoD. These are critical for future protein engineering and the search for catalysts with complementary stereoselectivity.

A survey of the sequence space of TropB coupled with structural and catalytic findings provides a greater understanding of the features that can potentially predict the site- and steoreoselectivity of a number of biocatalysts. This has informed current work in the Narayan lab to expand the FDMO sequence library implementing a machine learning approach. This technology informed by experimental data will have the potential to predict the function of biocatalysts based on their sequence.

3.8 Experimental

I. Cloning, Protein Expression, and Purification

A. General Considerations

E. coli cloning strains DH5α (Invitrogen) were used for DNA propagation. Phusion HF polymerase was purchased from New England BioLabs. All primers were purchased from Integrated DNA Technologies (IDT) ddH2O was sourced from a MilliQ Biocel water purification unit from Millipore.

B. Sequence Information

A gene encoding *afoD* (Q5BEJ7.1) was codon-optimized for overexpression in *E. coli* synthesized by GeneArt (Thermofisher). GeneArt cloned the synthesized sequence into pET151 vector conferring the T7 expression system, ampicillin resistance, and an N-terminal 6x His tag encoded upstream from the inserted gene. No further modification to the construct was necessary. The plasmid encoding AzaH (G3XMC2.1), in a modified pET28 vector to afford protein with both C- and N-terminal 6 x His-tags, was a generous gift from Professor Yi Tang at the University of California, Los Angeles. The plasmids encoding for FMDO1-7 were synthesized by Twist Bioscience and cloned into a pET28a vector.

Codon-Optimized *afoD* Sequence

```
ATGAGTACAGACTCGATCGAAGTTGCCATTATAGGCGCCGGGATCACGGGAATCACCCTGGCCCTGGGCCTCCTGTCTCGC
GGCATTCCCGTCCGCGTCTACGAGCGAGCCCGCGACTTTCACGAAATTGGAGCCGGTATCGGTTTCACCCCCAACGCCGAA
TGGGCGATGAAAGTCGTCGACCCGCGCATTCAAGCTGCTTTCAAACGCGTCGCTACCCCCAATGCCTCCGACTGGTTCCAG
TGGGTGGACGGATTCAACGAGTCCGGTACCGACCCGCGCGAGACCGAGGAACAGCTACTCTTCAAGATCTACCTCGGCGAG
CGTGGATTTGAGGGCTGCCACCGTGCCGACTTCCTAGGTGAGCTGGCACGTCTACTACCGGAAGGTGTGGTGACATTCCAG
AAGGCGCTGGATACCGTGGAGCCTGCAGCAGATAATAGCCTCGGCCAGCTTCTTCGATTCCAAGATGGCACGACAGCTACC
GCCCACGCGGTGATCGGCTGCGATGGCATTCGGTCGCGCGTTCGTCAGATCCTCCTAGGTGAAGACCATCCGACAGCATCA
GCCCATTACAGTCATAAATATGCAGCACGCGGCCTTATTCCCATGGACCGCGCCCGGGAGGCGCTGGGCGAAGATAAAGTG
GCGACACGCTTCATGCATCTCGGTCCGGATGCCCATGCCCTGACCTTCCCCGTTAGCCATGGGTCCTTGTTGAACGTCGTC
GCCTTCGTCACGGACCCTAACCCTTGGCCATATGCTGATCGCTGGACGGCGCAGGGCCCAAGAAAGACGTGACGGCTGCC
TTTTCCCGCTTTGGTCCGACCATGCGCACCATAATTGACCTCTTGCCTGATCCTATTGATCAATGGGCCGTTTTTGATACA
TACGACCATCCCCCAAATACGTATTCCCGGGGAGCTGTCTGTATAGCAGGGGATGCTGCTCATGCCGCGGCTCCGCATCAC
GGTGCAGGTGCAGGTTGTGGTGTGGAAGACGCGGCTGTGCTGTGCGCTGTGCTTCATATGGCTGCGAAAAAAGTTAACACC
GCAAAAACTGGTTCTGAGGGGAAAGCCGCTCTTATCACGGCCGCATTCGAAACCTATGATTCGGTTTGTCGCGAGCGTGCG
CAGTGGCTGGTGGAAAGTAGTCGCGTTATCGGTAATCTGTATGAGTGGCAGGATAAGGAGGTAGGGTCGGATGCTTCCAGG
TGCCACGATGAGGTGTATTGGCGCTCTCATCGCATTTGGGACTATGATATTGATGCGATGATGAGAGAGACAGCTGAGGTG
TTTGAGGCGCAGGTAGCTGGGGTGGCGAGAAAT
```

AfoD Protein Sequence

```
MADHEQEQEPLSIAIIGGGIIGLMTALGLLHRNIGKVTIYERASAWPDIGAAFAFTGIARECMQRLDPAILSALSKVAQRN
PHDKVRYWDGFHPKSKEEAQDPEKSVLFEIEEKNMAYWACLRGVFHAEMARLLPERVVRFGKRLVAYEDGGDQKVVLRFED
GEVEEADIVIACDGVHSTARRVLLGAEHPAANARYSRKAVYRALVPMPAAIDALGTEKAHVQIAHCGPDAHIVSFPVNNAQ
IYNVFLFTHDSNEWTHGHTMTVPSSKEEILSAVENWGPHIKELASLFPEQLSKYAIFDQADHPLPYYAAGRVALAGDAAHA
SSPFHGAGACMGVEDALVLAELLEKVQNGSAFKEKKSNIELALKTYSDVRIERSQWLVKSSREMGDLYEWRYEDIGGDGVK
CKAEW ERRSRVIWDFDVQGMVDQAREAYERAVVKV
```

Figure 3.13: For protein expression we used a codon-optimized AfoD gene that was or-
dered from GeneArt. [30]

Non-Optimized *azaH* Sequence

```
ATGAGTACAGACTCGATCGAAGTTGCCATTATAGGCGCCGGGATCACGGGAATCACCCTGGCCCTGGGCCTCCTGTCTCGC
GGCATTCCCGTCCGCGTCTACGAGCGAGCCCGCGACTTTCACGAAATTGGAGCCGGTATCGGTTTCACCCCCAACGCCGAA
TGGGCGATGAAAGTCGTCGACCCGCGCATTCAAGCTGCTTTCAAACGCGTCGCTACCCCCAATGCCTCCGACTGGTTCCAG
TGGGTGGACGGATTCAACGAGTCCGGTACCGACCCGCGCGAGACCGAGGAACAGCTACTCTTCAAGATCTACCTCGGCGAG
CGTGGATTTGAGGGCTGCCACCGTGCCGACTTCCTAGGTGAGCTGGCACGTCTACTACCGGAAGGTGTGGTGACATTCCAG
AAGGCGCTGGATACCGTGGAGCCTGCAGCAGATAATAGCCTCGGCCAGCTTCTTCGATTCCAAGATGGCACGACAGCTACC
GCCCACGCGGTGATCGGCTGCGATGGCATTCGGTCGCGCGTTCGTCAGATCCTCCTAGGTGAAGACCATCCGACAGCATCA
GCCCATTACAGTCATAAATATGCAGCACGCGGCCTTATTCCCATGGACCGCGCCCGGGAGGCGCTGGGCGAAGATAAAGTG
GCGACACGCTTCATGCATCTCGGTCCGGATGCCCATGCCCTGACCTTCCCCGTTAGCCATGGGTCCTTGTTGAACGTCGTC
GCCTTCGTCACGGACCCTAACCCTTGGCCATATGCTGATCGCTGGACGGCGCAGGGGCCCAAGAAAGACGTGACGGCTGCC
TTTTCCCGCTTTGGTCCGACCATGCGCACCATAATTGACCTCTTGCCTGATCCTATTGATCAATGGGCCGTTTTTGATACA
TACGACCATCCCCCAAATACGTATTCCCGGGGAGCTGTCTGTATAGCAGGGGATGCTGCTCATGCCGCGGCTCCGCATCAC
GGTGCAGGTGCAGGTTGTGGTGTGGAAGACGCGGCTGTGCTGTGCGCTGTGCTTCATATGGCTGCGAAAAAAGTTAACACC
GCAAAAACTGGTTCTGAGGGGAAAGCCGCTCTTATCACGGCCGCATTCGAAACCTATGATTCGGTTTGTCGCGAGCGTGCG
CAGTGGCTGGTGGAAAGTAGTCGCGTTATCGGTAATCTGTATGAGTGGCAGGATAAGGAGGTAGGGTCGGATGCTTCCAGG
TGCCACGATGAGGTGTATTGGCGCTCTCATCGCATTTGGGACTATGATATTGATGCGATGATGAGAGAGACAGCTGAGGTG
TTTGAGGCGCAGGTAGCTGGGGTGGCGAGAAAT
```

AzaH Protein Sequence

```
MSTDSIEVAIIGAGITGITLALGLLSRGIPVRVYERARDFHEIGAGIGFTPNAEWAMKVVDPRIQAAFKRVATPNASDWFQ
WVDGFNESGTDPRETEEQLLFKIYLGERGFEGCHRADFLGELARLLPEGVVTFQKALDTVEPAADNSLGQLLRFQDGTTAT
AHAVIGCDGIRSRVRQILLGEDHPTASAHYSHKYAARGLIPMDRAREALGEDKVATRFMHLGPDAHALTFPVSHGSLLNVV
AFVTDPNPWPYADRWTAQGPKKDVTAAFSRFGPTMRTIIDLLPDPIDQWAVFDTYDHPPNTYSRGAVCIAGDAAHAAAPHH
GAGAGCGVEDAAVLCAVLHMAAKKVNTAKTGSEGKAALITAAFETYDSVCRERAQWLVESSRVIGNLCHDEVYWRSHRIWD
YDIDAMMRETAEVFEAQVAGVARN
```

Figure 3.14: Sequence for codon-optimized azaH that was gifted by Professor Yi Tang at the University of California, Los Angeles. [30]

Codon-Optimized *fdmo1* Sequence

```
ATGCCGAGCTATAACAAAGATACCGAAAGCGTGGAAGTGGCGGTGATTGGCGGCGGCATTGTGGGCCTGGTGCTGGCGGCG
GGCCTGACCCGCCGCCAGATTAAAGTGAAAGTGTATGAACAGAGCCAGGGCTTTCGCGATATTGGCGCGGGCATTGGCTTT
AACGGCGCGGCGAAAGCGTGCATGCAGATGATTGATCCGGGCGTGATTACCGCGCTGCATCGCGGCGGCGGCGTGGCGGTG
AGCGCGGCGGATGAAGATGATCCGCATGATTATCTGCGCTGGATTGATGGCTTTGATCGCGGCAACGTGCAGCATCTGCAT
GATCAGAAACTGTATTGCAAAGTGGATGCGGGCTATAAAAAGCATTGAAGGCACCCGCCGCGATCGCTTTCTGGAAGAACTG
GCGAAAGATCTGCCGGAAGGCATGGTGGAATTTAAAAAACGCCTGCGCACCGTGGAAGAAGGCGGCGATGATTGCAAACTG
CAGCTGCATTTTGAAGATGGCACCATTGCGGAAGCGGATGCGCGCTGCGATGGCATTAAAAGCCGCATTCGCGAAATTGTG
CTGAGCGAAGCGAGCGTGGCGAGCAAACCGAGCTATACCCATGTGAACTTTTATAGCAGCCTGATTCCGATGAACAAAGCG
GTGGATATTCTGGGCAAATTTAAAGCGAGCGTGTTTCATAACCATATTGGCCCGGGCGCGAACGTGCTGCATTATCCGGTG
GCGAACGGCACCCTGTGCAACGTGAGCGCGTTTGTGCATGATGCGAACGAATGGCCGCCGCGGAAAAAAGCCCGACCAGCATT
GGCTTTCGCAAACATATTCAGGAAAAACTGGTGGGCTGGAGCCCGGTGGTGCGCGGCCTGATTGATCTGTTTCCGGATACC
CTGCCGGTGTGGGCGGTGTTTGATCTGTGGGAACATCCGATGCCGTATTATAACCGCGGCCGCATTTGCGTGGCGGGCGAT
GCGGCGCATGCGAGCAGCCCGCATCATGGCGCGGGCGCGGGCATGGGCATTGAAGATGCGCTGTGCCTGAGCGTGCTGCTG
GATGAAGTGAGCAGCAGCATTCGCCTGGAAGGCGCGAGCCGCCGCGATGCGATTCCGGTGGCGTTTCAGGTGTATGATAGC
ATTCGCCGCCGCCGCAGCCAGTGGCTGGTGAACAGCAGCCGCCGCCTGTGCGATCTGCAGCAGCATCATGATTGGGCGGAT
CCGGCGAAACTGGTGAAAGCGGAAACCTGCTTTGAAGAAATTACCGATCGCACCTATAAAATTTGGAACTTTGATAGCAAC
GGCATGATTAAAGAAAGCATTGAAAAATATGGCCGCGCGATTAACAGCCTGCGCCGCAACGGCCTGGCGACCAACACCGAT
TGCAAAGGCAACGGCCATATGAACGGCGTGCGCGCG
```

Figure 3.15: Codon-optimized FDMO1 sequence. [30]

Codon-Optimized *fdmo2* Sequence

```
ATGGCGAGCACCGAACCGCAGGCGGATAGCGTGGATGTGGTGATTGTGGGCGGCGGCATTATTGGCCTGGTGCTGACCGTG
GGCCTGCTGCGCGTGGGCGTGAAAGTGAAAGTGTATGAACAGGCGCAGGGCTTTCGCGAAATTGGCGCGGGCATTGCGTTT
ACCGCGAACGCGATTCGCTGCATGAACCTGATTGATCCGGCGATTCCGGTGGCGCTGCGCAGCAGCGGCAGCGTGGCGACC
AGCAACGGCGGCGATGAAGATCCGAACGATTATCTGCGCTGGATTGATGGCTATGATCGCCAGCGCGATGATCCGAGCCTG
CAGCAGCTGTTTTTTAAACTGAACGCGGGCTATCGCGGCTTTGAAGGCTGCCGCCGCGATCAGTTTCTGGAAGCGCTGGTG
AAAGTGATTCCGCCGGGCGTGATTGAACTGAAAAAACGCCTGGAAACCGTGCATGATAACGGCAGCGAAAACAAACTGCTG
CTGACCTTTCAGGATGGCCACCACCGCGGAAGCGGATGCGGTGATTGGCTGCGATGGCGTGAAAAGCACCCTGCGCCCGCATT
ATGTTTGGCGATGATCATCCGGCGAGCCGCCCGCGCTATAGCCATTGCGTGGCGTATCGCACCCTGATTCCGATGGATAAA
GCGGTGAGCGCGCTGGGCGCGTATAAAGCGACCAACCAGCATAACCATGTGGGCCCGAACGCGAACATTCTGCATTATCCG
GTGGCGAACAACACCATGATTAACGCGGTGGCGTTTATTCGCGATCCGAACGAATGGACCGATGAAAAAACCGTGGCGGAA
GGCACCCGCGATGATGTGAAAGCGGCGGTGCGCGGCTGGAGCCAGCCGGTGCTGAACCTGGTGGATTGCTTTCCGGATACC
CTGAGCAAATGGGGCATTTTTGATCTGTGGGAATTTCCGGTGCCGAGCTATAACGTGGGCCGCCTGAGCCTGGCGGGCGAT
GCGGCGCATGCGAGCAGCCCGCATCATGGCGCGGGCGCGTGCATGGGCATTGAAGATGCGCTGTGCCTGACCACCCTGATG
GAACAGGTGGTGGTGGAAGCGCAGAAAGCCCGGGCGATAAAGGCCGCGCGCTGATTGCGGCGCTGGATACCTATAGCGCG
GTGCGCCAGACCCGCAGCCAGTGGGTGGTGAACAGCAGCCGCCGCGTGTGCGATCTGCATCAGCAGCAGGAATGGGCGGAT
GCGACCAAACTGATTAAAGCGCAGACCTGCTTTGAAGAAGTGAAAGATCGCAGCCTGAAAATTTGGCATTTTGATTATGAA
CGCATGGTGCGCGATAGCCTGCAGGCTATAAACAGCGCCGCGCGCCGATTAACGGCGCGACCAAAGATAAAAACCTGTAT
```

Figure 3.16: Codon-optimized FDMO2 sequence. [30]

Codon-Optimized *fdmo3* Sequence

```
ATGATTGAAGCGCGCGCGATTGAAGTGGCGATTATTGGCGGCGGCATTACCGGCCTGACCCTGGCGCTGG
GCCTGCAGAAACGCAACACCAACTTTCATATTTATGAACGCGCGCAGAGCCTGCGCGAAATTGGCGCGGGG
CATTGGCTTTACCCCGAACGCGGAACGCGCGATGCTGGCGCTGGATCCGCGCATTCATGAAGCGTTTAAA
AGCGTGGCGAGCAAAAACGCGAGCGATTGGTTTCAGTGGGTGGATGGCTTTAGCGGCGTGAACAACGATA
AAGATACCGTGAAAGAAGATCTGCTGTTTAACATGTATCTGGGCGAACGCGGCTTTGAAGGCTGCCATCG
CGCGCAGTTTCTGAAAGAACTGGTGAACCATCTGCCGCAGGGCTGCGTGACCTTTGGCGCGTGCCTGGAT
ACCATTATTGATCAGGGCGAAAACGAACGCATTCTGCTGAAATTTCATAACGGCACCATTGCGGAAGCGG
ATCTGGTGATTGGCTGCGATGGCATTCGCAGCCGCGTGCGCCAGCTGATTCTGGGCGAAAACAACCCGGC
GAGCTATCCGGCGTATACCCATAAAAAAGCGTATCGCGGCCTGATTCCGATGGAAAAAGCGCTGCCGGCG
CTGGGCGAAAGCAAAGTGAACACCCGCCTGATGCATCTGGGCCCGGATGCGCATACCCTGACCTTTCCGG
TGGCGGGCGGCAAACTGATGAACGTGGTGGCGTTTGTGACCGATCCGGGCGAATGGCCGTATACCGAAAA
ACTGAGCGCGCCGGCGGAAAAAAAAGCGCGATTGAAGGCTTTAGCAAATTTGGCGGCGCGGTGCGCACC
ATTATGAACCTGCTGCCGGAAGATCTGGATGAATGGGCGATTTTTGATACCTATGATCATCCGGCGAGCA
CCTATTATCATGGCCGCATTTGCATTGCGGGCGATGCGGCGCATGCGAGCAGCCCGCATCATGGCGCGGG
CGCGGGCGCGGGCATTGAAGATGTGACCGTGCTGGCGACCGTGATTGAAGTGGCGCAGACCACCCTGCTG
GAAAGCCCGGATAAAGCCGCAGCGGCGTGCTGAACGCGGCGCTGGCGACCTATAACGCGGTGCGCCTGG
AACGCAGCCAGTGGCTGGTGGAAAGCAGCCGCATTCTGGGCGAAATTTATGAATGGCAGTATAAACCGAC
CGGCCGCGATAAAAAAAAATGCGAAGAAGAAGTGTATTGGCGCAGCCATAAAATTTGGGATTATGATATT
GGCCAGATGCTGCAGGAAACCACCGAATATTATAAACAGCGCGTGGGCGCG
```

Figure 3.17: Codon-optimized FDMO3 sequence. [30]

Codon-Optimized *fdmo4* Sequence

```
ATGGATACCAACAAATTTGAAATTGCGATTATTGGCGCGGGCATTACCGGCATTACCCTGGCGCTGGGCCTGCTGAGCCGC
GGCATTCCGCCGCGCGATTTTCATGAAATTGGCGCGGGCATTGGCTTTACCCCGAACGCGGAATGGGCGATGAAAGTGGTG
GATCCGCGCATTCATGCGGCGTTTAAACGCGTGGCGACCCCGAACGCGAGCGATTGGTTTCAGTGGGTGGATGCGTTTAAC
GAAACCGGCGAACGCGGCTTTGAAGGCTGCCATCGCGCGCAGCTGCTGGGCGAACTGGCGCGCCTGCTGCCGGAAGGCATT
GTGACCTTTTATAAAGCGCTGGATACCCTGGAACCGGCGGCGGATAACCGCCTGGGCCAGCTGCTGCGCTTTCAGGATGGC
ACCACCGTGACCGCGCATGCGGTGATTGGCTGCGATGGCATTCGCAGCCGCGTGCGCCAGATTCTGTTTGGCGAAGATCAT
CCGGCGGCGAGCGCGCATTATAGCCATAAATATGCGGCGCGCGGCCTGATTCCGATGGATCGCGCGCGCGAAGCGCTGGGC
GATGCGAAAGTGGCGACCCGCTTTATGCATCTGGGCCCGGATGCGCATGCGCTGACCTTTCCGATTGCGCATGGCAGCCTG
CTGAACGTGGTGGCGTTTGTGACCGATCCGAACCCGTGGCCGTATGCGGATCGCTGGACCGCGCAGCGCAACGAAACCGAT
GTGGCGGCGGCCGTTTAGCCGCTTTGGCCCGACCATGCGCACCATTATTGATCTGCTGCCGGATCCGATTGATCAGTGGGCG
GTGTTTGATACCTATGATCATCCGCCGAACACCTATAGCCGCGGCCCGGTGTGCATTGCGGGCGATGCGGCGCATGCGGCG
GCGCCGCATCATGGCGCGGGCGCGGGCTGCGCGGTGGAAGATGTGGCGGTGCTGTGCGCGGTGCTGGATCTGGCGGCGAAA
CGCGTGGATGCGACCAAATGCGATCCGAAAGGCAAAGCGGCGCTGATTACCACCGCGTTTGAAACCTATGATGCGGTGCGC
CGCGAACGCGCGCAGTGGCTGGTGGAAACCAGCCGCATTATTGGCAACTTTTATGAATGGCAGGATAACGAAGTGGGCCCG
GATGCGAGCATTTGCCATGATGAAGTGTATTGGCGCAGCCATCGCATTTGGGATTATGATATTGATACCATGATGCGCGAA
ACCGCGAAAGTGTTTGAAGTGCGCGTGGCGGAACTGACCAAAAAC
```

Figure 3.18: Codon-optimized FDMO4 sequence. [30]

Codon-Optimized *fdmo5* Sequence

```
ATGGCGAGCAACAACAAAACCACCAACCCGAGCATTGAAGTGGCGGTGGTGGGCGGCGGCGTGATTGGCGTGATGACCGCG
CTGGGCCTGATTCGCCGCGGCATTAAAGTGACCATTTATGAACGCAGCAGCAACTGGCATGAAATTAGCGCGGGCTTTGCG
TTTACCGGCGTGGCGCGCGAATGCATGCAGCGCCTGGATCCGGGCATTCTGGATGTGCTGAGCCGCATTAGCCAGAAAACC
GATCCGAACGATAGCAGCACCACCTATTGGAACGCGTATCATCCGCAGACCAAACAGGATGCGGAAGATGAAAGCACCAGC
CTGCTGTTTCAGCTGCCGGGCAACAAACTGGCGTTTTGGGGCTGCGTGCGCAGCCAGTTTCTGCTGGGCATGGTGGCGCTG
CTGCCGGATGATGTGGCGCGCTTTGGCAAACAGCTGGTGAGCTATGATGATGGCGATGCGAACGATAAAGTGGTGCTGCAT
TTTGCGGATGGCAGCACCGCGGAAGCGGATGTGGTGCTGGGCTGCGATGGCATTCATAGCACCACCCGCAAAACCCTGCTG
GGCGCGCATCATCCGGCGACCCGCCCGAGCTATACCCATACCGTGGCGTATCGCACCATGGTGCCGATTGATGCGGGCATT
GCGGCGCTGGGCGAAGATAAAGCGCGCCGCGCGTGCATGCATTGCGGCCCGAACGCGAACATGATGAGCTATCCGGTGATG
AACGGCACCCTGCTGAACGTGGCGTTTTTTGCGCATGAAAGCAGCGAATTTCCGGATCCGGAAAAAATGACCGCGCCGGGC
ACCCGCGAAGAACTGGAACGCGTGGTGGTGGGCTGGGGCCCGCATCTGGTGGAACTGACCAAACTGTTTCCGGATAACATG
GTGAAATGGGGCATTTTTGATATGGATGAAAACCCGGCGCCGACCTATGCGCGCGGCTGCGTGTGCCTGGCGGGCGATGCG
GCGCATGCGAGCAGCCCGTTTCAGGGCGTGGGCGCGTGCATTGGCGTGGAAGATGCGCTGGTGCTGTGCGAAGCGCTGGCG
ACCGTGCAGGCGGGCGGCAACAGCGGCAGCGATGATGGCAACCATACCCATAGCCAGCGCGAAGTGATTGAACAGGCGCTG
CAGGCGTATAGCCAGGCGCGCATTGATCGCGGCCAGTGGGTGGTGCGCAGCAGCCGCGAACTGGGCCAGATTTATCAGTGG
CGCTATGGCCCGACCGGCCGCGATGCGGAACGCAGCAAACTGAAACTGGAACGCGCGCGAGCCGCACCGTGTGGGATTATGAT
GTGGATAAAATTGTGACCGAAATTCGCGCGGTGGTGGCG
```

Figure 3.19: Codon-optimized FDMO5 sequence. [30]

97

Codon-Optimized *fdmo6* Sequence

```
ATGACCGTGGCGGATCGCGCGCCGCTGGATGTGGCGATTATTGGCGGCGGCATTATTGGCATTATGACCGCGCTGGGCCTG
CTGCATCGCGGCTTTCGCGTGACCGTGTATGAACGCGCGGCGAGCTGGCCGGAAATTGGCGCGGCGTTTGCGTTTACCGGC
GTGGCGCGCCAGTGCATGGAACGCCTGGATCCGCGCGTGCTGGAAAGCCTGGCGCGCGTGGCGCAGCGCAGCCGCATGAA
AAAGTGCGCTATTGGGATGGCTTTCATCCGCGCACCAAAGAAGCGGCGCAGGAAGAAAGCGCGGTGCTGTTTGAAATTCTG
GAAAACATATGGCGTATTGGGCGTGCATTCGCGGCCATTTTCTGCTGGATATGGCGGCGCAGCTGCCGGATGGCGTGGTG
CAGTTTGGCAAACGCCTGGTGGATTATAACGATGATGAAGCGAACGAAAAGTGGTGCTGTGCTTTGCGGATGGCAGCACC
GCGGAAAGCGATGTGGTGATTGCGTGCGATGGCATTCATAGCGCGACCCGCAAAGTGCTGCTGGGCGTGGATCATCCGGCG
GCGAACGCGAGCTATAGCCGCAAAAGCATGTATCGCGCGATGGTGCCGATGGCGGATGCGGTGAGCGCGCTGGGCACCGAA
AAAGCGCATGTGCAGATTGCGCATCTGGGCCCGGATGCGCATGTGGTGAGCTTTCCGGTGAACAACGGCCAGGTGTATAAC
GTGTTTCTGTTTCTGCATGATCCGAACGAATGGGATCATGGCCATACCATGACCGTGCCGAGCAGCCGCAGCGAAGTGATG
GATGCGATTCAGGGCTGGGGCCCGCATATTAAAGAAATTGTGAGCTGCTTTCCGGAAACCGTGAGCAAATATGCGATTTTT
GATCAGGCGGATAACCCGCTGCCGTATTATGCGAGCGGCCGCGTGTGCCTGGCGGGCGATGCGGCGCATGCGAGCAGCCCG
TTTCATGGCGCGGGCGCGTGCATGGGCGTGGAAGATGCGCTGGTGCTGGCGGAACTGCTGGGCCTGGTGGATGCGGGCCCG
GTGGCGGCGCGCCAGCGCAACATTAAAGCGGCGCTGCAGACCTATAGCAGCGTGCGCATTGAACGCAGCCAGTGGCTGGTG
CAGAGCAGCCGCGATATGGGCGATCTGTATGAATGGCGCTATCCGCCGACCGGCGAAGATGGCGCGAAATGCAAAGCGGAA
TTTGAACGCCGCAGCAAAGTGATTTGGGATTTTGATGTGGATGGCATGGTGGCGGGCGCGAAAAAAAAATATGAACATAGC
ATGGAAGCG
```

Figure 3.20: Codon-optimized FDMO6 sequence. [30]

Codon-Optimized *fdmo7* Sequence

```
ATGGAAGCGCCGAACAACCATCCGAACGGCATTAACGTGATTAACGGCCATAAAGCGAAAAGCCTGGAAGTGGCGATTGTG
GGCGGCGGCCTGACCGGCCTGGCGCTGGCGGTGGGCCTGCTGCGCCGCAACATTAACTTTACCATTTATGAACGCGCGGCG
AGCTTTGGCGAACTGGGCGTGGGCATTCATTTTACCCCGAACGCGGAACGCGCGATGGAAGCGCTGGATCCGCGCGTGCTG
CAGAGCTATGTGGATGTGGCGGACCAACGCGGAAGGCGGCTTTCTGAGCTTTGTGGATGGCGCGAGCGGCGATGATGGCCTG
CTGTTTCAGCTGCGCATGGGCAAAGGCTATAAAGCGGCGCGCCGCTGCGATTTTGTGAGCCAGCTGGTGAAACATATTCCG
CAGGAACGCGTGCAGCATCTGAAATGGCTGCAGAGCGTGGAAGAAGATGGCGAAGGCCGCGCGGTGCTGACCTTTCGCGAT
GGCAGCACCGCGGAAGCGGATGTGGTGGTGGGCTGCGATGGCATTCGCAGCCAGGTGCGCAGCGCGATGTTTGGCAGCGGC
CCGAGCGCGCCGCGCGCGCAGTATGCGCATCAGCTGGCGTTTCGCGGCCTGGTGCCGATGGCGAAAGTGGAAGAAGCGCTG
GGCAGCGGCAAAACCAGCCGCGCGATTGGCTATCTGGGCCCGGGCGGCTTTGTGCTGAGCGTGCCGCTGGCGGGCATTAAC
ATGATGCATCTGGAAGTGTTTGTGATGGATCCGCTGGATTGGAGCGATACCCGCAGCAAAAGCGAAAAAGGCAACGATGAA
GATGATGTGAAACGCTATGTGCTGCCGGCGACCCGCGCGGAAGCGGAAAAAGCGTTTACCGAATTTAACCCGACCGTGCGC
AGCCTGATTAGCCTGCTGCCGGAAACCCTGGGCAAATGGGCGATTTTTGATATGCTGGATAGCCCGGCGCCGAGCTATGCG
CTGGGCCGCATGTGCCTGGCGGGCGATGCGGCGCATGCGAGCACCCCGAACCAGGCGGCGGCGCGGGCGCGGGCATGGAA
GATAGCCTGGTGCTGGCGGAAATTCTGGCGGCGCTGGCGGATCGCGAAAACAGCGGCGCGCCGGTGGGCCTGAGCGAAATT
AGCGAAGGCCTGAAAGTGTATAGCGAAGCGCGCTATGAACGCGCGCAGTGGCTGGTGCAGAGCAGCCGCCGCGTGGCGCAG
CTGTTTACCCGCAAAAGCGCGGAACAGGAAGAACCGATTAGCCGCGAAATTCTGGAACGCAGCCATCAGCTGTGGGATCAT
GATGTGGATGCGATGGTGGCGGATGCGCTGGGCAAACTGAAAGCGAAACTGAGCGAAAAAAAA
```

Figure 3.21: Codon-optimized FDMO7 sequence. [30]

C. Site-directed mutagenesis

Variant	Plasmid ID	Oligo sequence
AfoD Y118F	AfoD_Y118F_P1	GAAAAGAACATGGCATTTTGGGCATGTC
AfoD F237Y	AfoD_F237Y_P2	CATATTGTTAGCTATCCGGTTAATAAC

Note: All sequences are 5'-3'.

Table 3.1: Primer Sequences.

98

AfoD(Y118F), AfoD(F237Y). Substitutions were generated by site-directed mutagenesis on pET151-afoD(WT). 25 μL PCR reaction mixtures contained 5 μL of 5X Phusion HF buffer, 1 ng/μL WT parent plasmid, 0.5 μM of primer, 200 μM dNTPs, 0.5 U μL-1 Phusion HF. Amplification was accomplished with the following PCR protocol: 95 °C for 30 s, (95 °C 30s 95 °C for 30 s (-0.5 °C/cycle), 72 °C 0:30/kb) for 12 cycles, (95 °C for 30 s, 65 °C for 30 s, 72 °C 0:30/kb) for 20 cycles with a final extension of 72 °C for 10 min. This was followed by a 10 μL digestion containing 1 μL of NEB CutSmart buffer, 8 μL of PCR mixture and 20 units of DpnI. The reaction was incubated at 37 °C for 3 h and transformed into chemically competent *E. coli* DH5α cells.

D. Gene Expression and Protein Purification

Cell Growth:

AfoD Wild-type, AfoD(Y118F) and AfoD(F237Y), Plasmids were transformed into *E. coli* strain BL21(DE3). 500 mL of Terrific Broth (TB) containing 100 μg mL-1 ampicillin was inoculated with 5 mL overnight culture prepared from a single colony in Luria Broth Luria Broth (LB) and 100 μg mL-1 ampicillin. The culture was grown at 37 °C and 250 rpm for 4 h. The culture was then cooled to 20 °C for 1 h at 200 rpm, and expression was induced with 0.1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) and growth continued at 20 °C for 18 h at 200 rpm. After overnight expression, cultures were centrifuged at 13,881 x g for 30 min. Cell pellets from overexpression were stored at -80 °C for long-term storage.

Protein Purification for Crystallography:

The cell pellet from a 3 L culture was resuspended in 40 mL of lysis buffer (50 mM Tris:HCl pH 7.8, 300 NaCl, 10 mM imidazole, and 10% (v/v) glycerol) with 0.1 mg mL-1 lysozyme, 0.05 mg mL-1 DNase, and 0.1 mM flavin adenine dinucleotide (FAD), incubated on a rocker at 4 °C for 2 h, and lysed by sonication. Insoluble material was removed by centrifugation (46,413 x g for 30 min).

The soluble fraction was sterile-filtered through a 0.45 μM syringe filter. The clarified lysate was loaded onto a 5 mL HisTrap column (GE Healthcare) on a Healthcare ÄKTA FPLC. The column was washed with 10 column volumes (CV) of lysis buffer, and the protein was eluted with an 8-100% gradient of elution buffer (50 mM Tris:HCl pH 7.8, 300 mM NaCl, 400 mM imidazole, and 10% (v/v) glycerol). Fractions containing protein, which were visibly yellow, were concentrated to 2 mL using a 30 kDa cutoff filter by centrifugation at 4,000 x g, 4 °C. Concentrated proteins were further purified by size exclusion chromatography (HiLoad 16/60 Superdex 00, GE Healthcare) in gel filtration buffer (50 mM Tris:HCl pH 7.8, 100 mM NaCl, 10% (v/v) glycerol). Estimated molecular weight including the 6x His tags for AfoD wild-type and variants, 48.5 kDa, was confirmed by SDS-PAGE analysis Fig. 3.6. AfoD wild-type eluted as a monomer with an apparent molecular weight of 48 kDa Fig. 3.4. FAD content was measured by the flavin incorporation assay. [62]

Protein Purification for Assay:

The cell pellet was resuspended in 40 mL lysis buffer (50 mM Tris:HCl pH 7.8, 300 NaCl, 10 mM imidazole, 10% (v/v) glycerol) with 0.1 mg mL^{-1} lysozyme, 0.05 mg mL^{-1} DNase, and 0.1 mM FAD, incubated on a rocker at 4 °C for 45 min, lysed by sonication and cleared by centrifugation (46,413 x g for 30 min). The supernatant was incubated with Ni-NTA on a rocker for 2 h at 4 °C, followed by purification by gravity using a 25-50 mM gradient with increments of 5 mM imidazole. Protein was eluted with 100% elution buffer (50 Mm Tris:HCl pH 7.8, 300 mM NaCl, 400 mM imidazole, 10% (v/v) glycerol). Concentrated protein was desalted over a PD-10 desalting column (GE Healthcare). The protein was concentrated further using a 30 kDa cutoff filter by centrifugation at 4,000 x g, 4 °C. Concentrated protein was divided into 100 μL aliquots, frozen in liquid nitrogen and stored at -80 °C.

Oligomeric state determination:

Separations were conducted as described in respective purification procedures using a Sephacryl S-200 HR gel filtration column. Approximate sizes were determined relative to a calibration curve generated using the Gel Filtration Standards Kit (Sigma). According to the calibration curve where y =10(-0.7496(x/39.88)+3.068), the apparent molecular weight of AfoD was 48 kDa. Based on this analysis, it can be concluded that AfoD forms a monomer in solution.

Determination of flavin incorporation:

A 100 μL sample of enzyme stock was boiled at 100 °C for 10 min. Solution was clarified by centrifugation. Supernatant was then transferred to a 96-well plate and the absorbance was recorded from 300-700 nm in 1 nm increments. Absorbance at 450 nm and the extinction coefficient (11,300 M^{-1} cm^{-1}) for free FAD were used to determine the concentration of FAD in each protein sample using Beer's law. Due to evaporation, the pathlength was measured for each well and used in the calculation of FAD concentration. The FAD incorporation for each protein was AfoD wild-type 64%, AfoD Y118F 58%, AfoD F237Y 65%.

FAD incorporation Assay for AfoD Wild-type

[FAD] = 165.2 μM
FAD incorporation = 64%
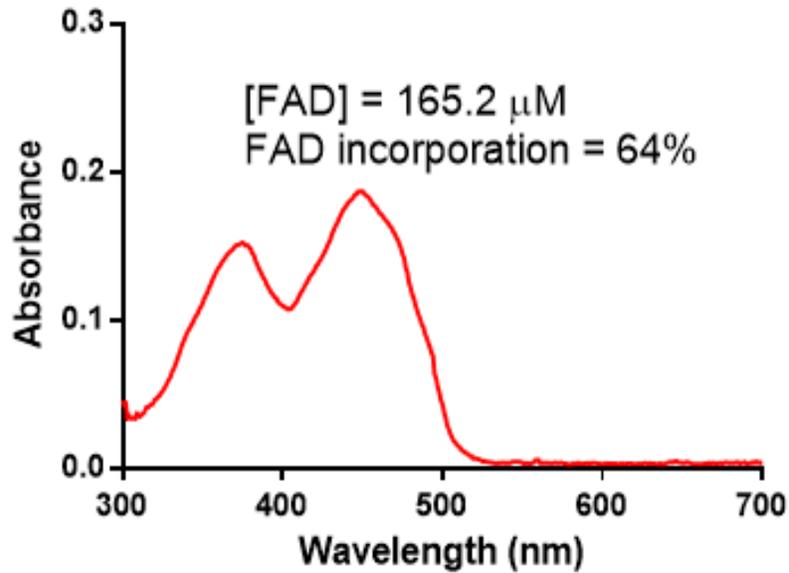
Figure 3.22: UV-Vis of denatured AfoD wild-type.

FAD incorporation Assay for AfoD Y118F

[FAD] = 54 μM
FAD incorporation = 58%
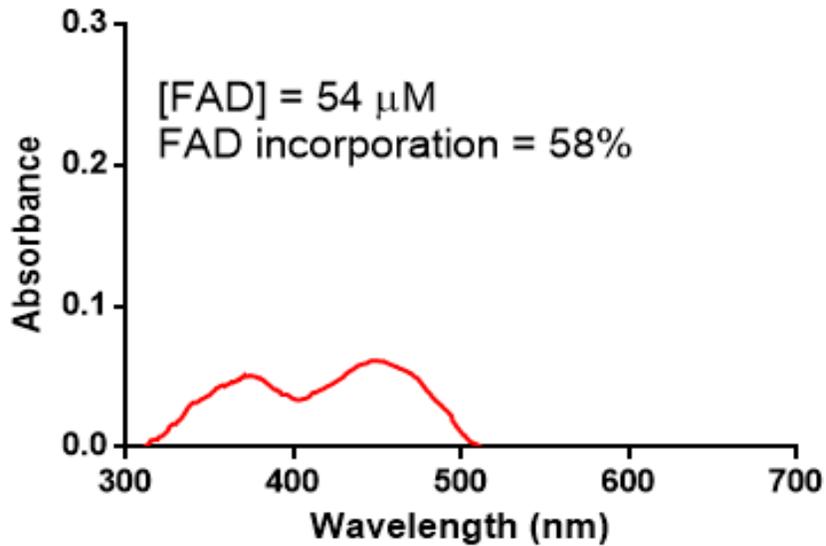
Figure 3.23: UV-Vis of denatured AfoD Y118F.
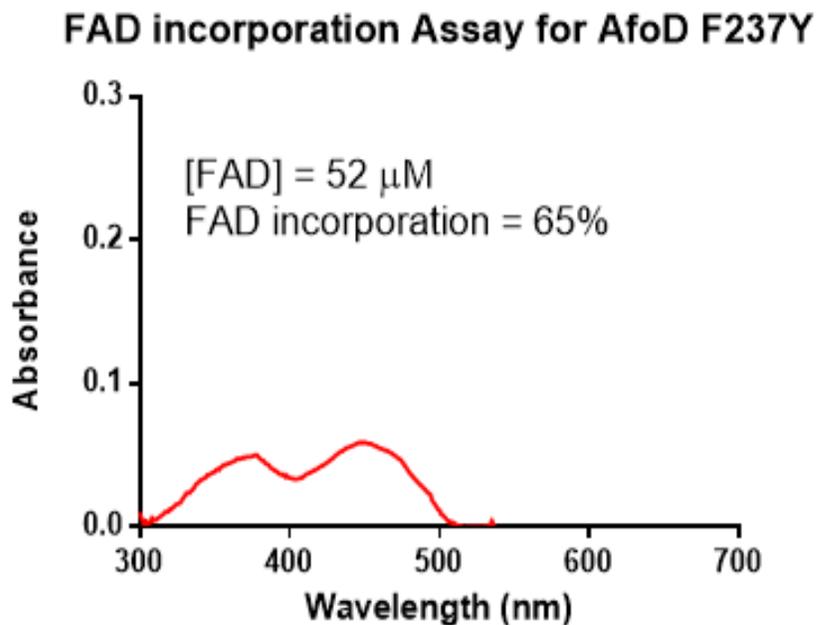
Figure 3.24: UV-Vis of denatured AfoD F237Y.

II. Protein crystallization and structure determination

Crystallization:

Diffraction-quality crystals of wild type AfoD were grown by sitting-drop vapor diffusion at 20 °C from a solution containing 1 μL protein stock and 1 μL reservoir solution (0.2 M $(NH_4)_2SO_4$, 0.1 M MES:NaOH pH 6.5, 30% w/v PEG 5K MME). All crystals were flash cooled in liquid nitrogen without additional cryoprotection. Crystals were sensitive to radiation.

Structure determination and refinement of AfoD wild-type:

Diffraction data were recorded at GM/CA beamline 23ID-B at APS, integrated with XDS [63] and scaled in Aimless [64] in the CCP4 suite [64] table 3.2. The structure was solved by molecular replacement using Phaser [65] with flavin-dependent TropB (PDB code 6NES) [31] as a search model. Further model building was carried out in Coot and

refinement in Phenix AutoBuild. [66, 67] The final model includes amino acids 10 - 352 in AfoD monomer. The monomer model also includes 1 FAD cofactor and 279 water molecules. Figures were made with Pymol. [68] The stereochemical quality of the structure was validated with MolProbity. [71]

Data Collection	AfoD Wild Type
Space Group	$P2_12_12_1$
Cell dimensions a, b, c (Å)	83.54 96.97 116.34
α, β, γ (°)	90 90 90
X-ray Source	APS 23 ID-B
Wavelength (Å)	1.033
d_{min} (Å)	2.09 (2.165 - 2.09)[a]
R_{merge}	0.119 (1.838)[a]
R_{merge} inner shell	0.063
Wilson B-factor ($Å^2$)	47
Mean I/σ (I)	13.9 (1.6)[a]
Completeness (%)	99.9 (99.9)[a]
Multiplicity	13.6 (13.1)[a]
Total Reflections	762,543 (73,275)[a]
$CC_{1/2}$	0.997 (0.721)[a]
Refinement	
Data Range (Å)	42.83 - 2.09
Reflections used in refinement (#)	56,614
R_{work}/R_{free}	0.196/0.233
Non-hydrogen atoms (#)	7,032
protein	6,647
ligand	106
water	279
Amino Acid Residues	845
Deviation from ideality	
RMSD(bonds) (Å)	0.010
RMSD(angles) (°)	1.74
Average B-factor ($Å^2$)	47.4
protein	47.5
ligand	43.2
solvent	47.7
Ramachandran Plot	
Favored (%)	96.3
Allowed (%)	3.7
Outliers (%)	0
PDB	

Table 3.2: AfoD Crystallographic Summary. a Values in parentheses pertain to the outermostshell data

III. Biocatalytic Reactions

General Considerations

A stock solution of the substrate (50 mM) was prepared by dissolving the substrate in DMSO (analytical grade). Stock solutions of NADP⁺ (100 mM) and glucose-6-phosphate (G6P, 500 mM) in water were stored at -20 °C. Aliquots of glucose-6-phosphate dehydrogenase (G6P-DH, 100 U mL^{-1}), All enzymes in this chapter were stored at -80 °C.

In vitro analytical-scale reactions

Each reaction contained 20 μM enzyme, 50 mM of potassium phosphate buffer pH 8.0 (25 μL, 100 mM stock), 2.5 mM substrate (2.5 μL, 50 mM stock), 1 mM NADP+ (0.5 μL, 100 mM stock), 1 mM G6P-DH (0.5 μL, 100 μ mL^{-1}), 5 mM G6P (5 μL, 5 mM stock), and Milli-Q water to a final volume of 50 μL. Reactions were carried out at 30 °C for 2 h and quenched with the addition of 70 μL acetonitrile (HPLC grade) containing 2.5 mM pentamethylbenzene as an internal standard. Reactions were cleared by centrifugation at 16,000 x g, 12 min) and the supernatant was analyzed by UPLC-diode Array Detector (DAD). Consumption of the starting material was quantified by comparison to standard curves of each substrate.

UPLC traces of Biocatalytic Reactions
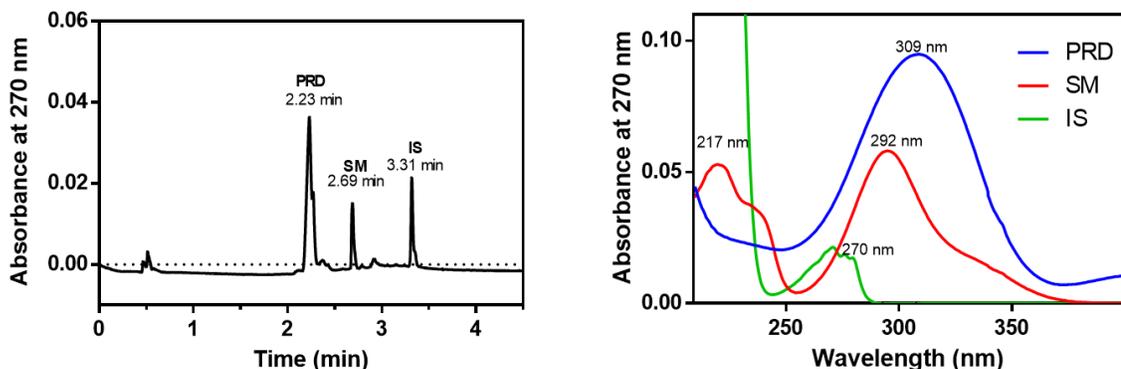
Figure 3.25: Oxidative dearomatization of S1 by AfoD wild-type. PDA traces of enzymatic reaction. IS = internal standard, SM = Starting Material **3.11**, and PRD = Product **3.13**.
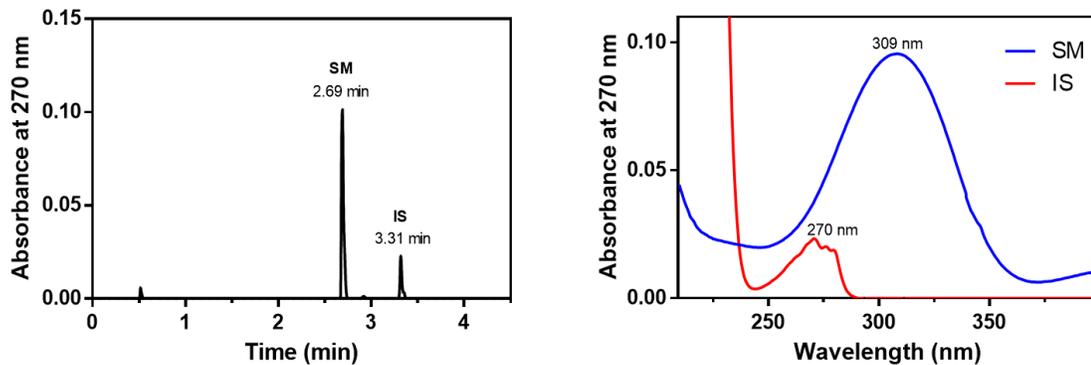
Figure 3.26: No enzyme control of the reaction of S1. PDA traces of enzymatic reaction. IS = internal standard, SM = Starting Material **3.11**.
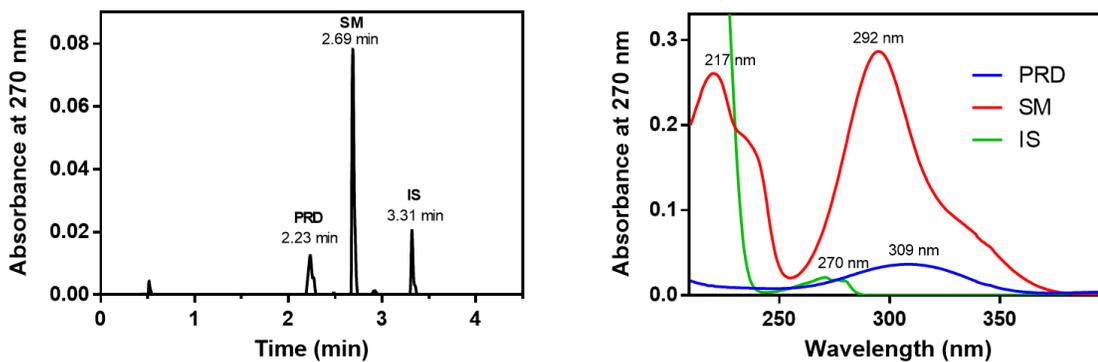
Figure 3.27: The reaction of S1 and AfoD Y118F. PDA traces of enzymatic reaction. IS = internal standard, SM = Starting Material **3.11**, and PRD = Product **3.13**.
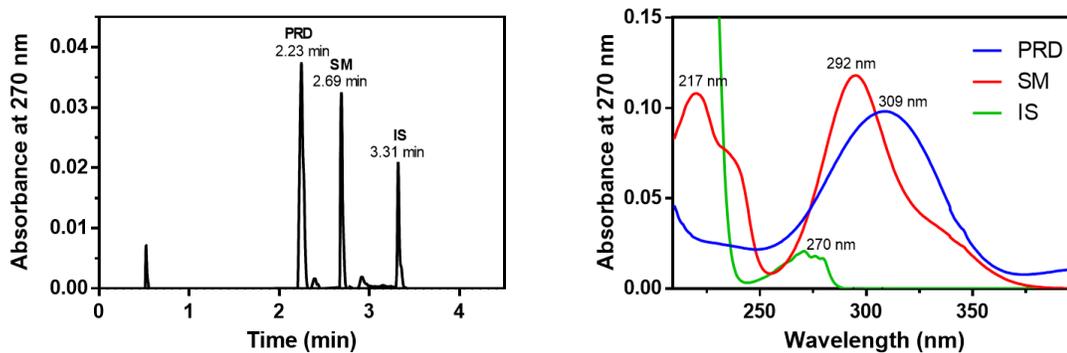
Figure 3.28: The reaction of S1 and AfoD F237Y. PDA traces of enzymatic reaction. IS = internal standard, SM = Starting Material **3.11**, and PRD = Product **3.13**.

IV. Determination of Enantiomeric Excess

General procedure for in vitro preparative-scale reactions:

Preparative–scale enzymatic reactions were conducted on 5 mg of substrate **3.11** under the following conditions: 20 μM flavin-dependent monooxygenase, 2.5 mM substrate, 1 mM NADP+, 1 U/mL G6PDH, and 5 mM G6P for NADPH generation in reaction buffer (50 mM potassium phosphate buffer, pH 8.0). The reaction mixture was added to a 50 mL Erlenmeyer flask and incubated at 30 °C with 100 rpm shaking. After 18 h, the reaction mixture was diluted with acetone (2 x total reaction volume). Precipitated biomolecules were pelleted by centrifugation (4,000 x g, 12 min). Isolation procedure: The supernatant was concentrated under reduced pressure to a final volume of approximately 1 mL. The resulting mixture was filtered through a 0.22 μm filter and purified by preparative HPLC using a Phenomenex Kinetex 5 μm C18, 150 x 21.2 mm column under the following conditions: mobile phase A = deionized water + 0.1% formic acid and B = acetonitrile + 0.1% formic acid; method = 5% to 100% B over 13 min, 100% B for 4 min; flow rate, 15 mL/min.

Determination of Enantiomeric Excess

Racemic Standard
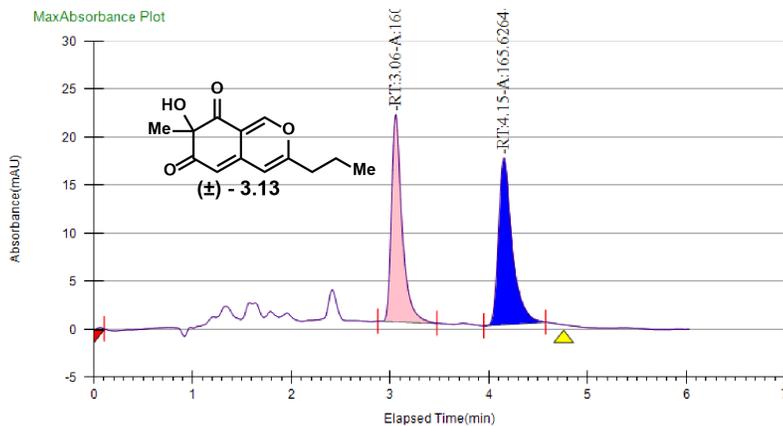
Instrument Method	Inj. Vol.	Solvent	Column	Sample	Well Location	Temp. (C)	Flow (g/min)	% Modifier	Pressure (Bar)
AD-H_30%_300-330	4 uL	Isopropanol	AD-H	rac	11A	40	3.5	30	120

Peak No	% Area	Area	Ret. Time	Height	Cap. Factor
1	49.2722	160.8739	3.06 min	21.5858	0
2	50.7278	165.6264	4.15 min	17.3272	0

AfoD Reaction
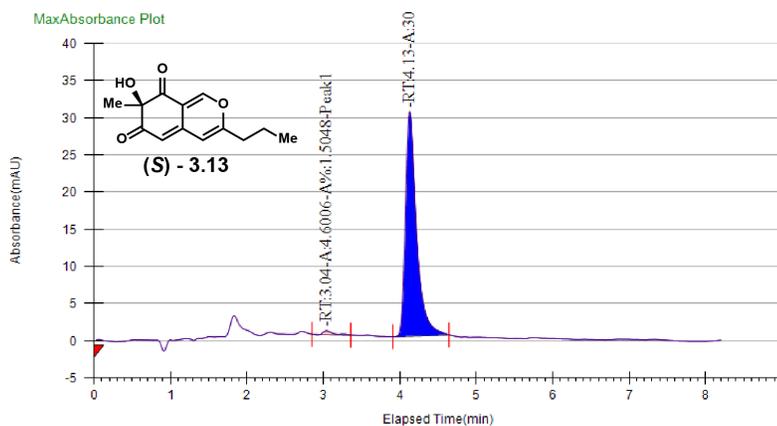
Instrument Method	Inj. Vol.	Solvent	Column	Sample	Well Location	Temp. (C)	Flow (g/min)	% Modifier	Pressure (Bar)
AD-H_30%_300-330	10 uL	iPrOH	AD-H Chiral Analytical	JBP-3-169-AfoD	13A	40	3.5	30	120

Peak No	% Area	Area	Ret. Time	Height	Cap. Factor
1	1.5048	4.6006	3.04 min	0.51	3040.6333
2	98.4952	301.1399	4.13 min	30.1492	4132.2833

Figure 3.29: PDA traces of racemic 3.13 obtained from an IBX-mediated oxidative dearomatization, 3.13 obtained from AfoD-mediated oxidative dearomatization CHIRAL-PAK® ADH, 30%, CO2, 3.5 mL/min.

AfoD Y118F

Instrument Method	Inj. Vol.	Solvent	Column	Sample	Well Location	Temp. (C)	Flow (g/min)	% Modifier	Pressure (Bar)
AD-H_30%_300-330	6 uL	Isopropanol	AD-H	ARB-V-071 AfoD_1	12A	40	3.5	30	120

Peak No	% Area	Area	Ret. Time	Height	Cap. Factor
1	46.8333	12090.3845	2.92 min	1322.9644	0
2	53.1667	13725.3924	4.16 min	1541.5059	0

AfoD F237Y

Instrument Method	Inj. Vol.	Solvent	Column	Sample	Well Location	Temp. (C)	Flow (g/min)	% Modifier	Pressure (Bar)
AD-H_30%_300-330	4 uL	Isopropanol	AD-H	AfoD_2_pre p2 ARB-V-071	11A	40	3.5	30	120

Peak No	% Area	Area	Ret. Time	Height	Cap. Factor
1	28.8486	2149.7018	2.97 min	332.0461	0
2	71.1514	5301.9567	4.02 min	631.6176	0

Figure 3.30: PDA traces of AfoD Y118F and AfoD F237Y mediated oxidative dearomatization to yield 3.13

Biochemical Characterization of AfoD wild-type and Variants

pH Profile of AfoD wild-type

Substrate consumption of AfoD wild-type and variants was monitored across a range of pHs. The pH at which AfoD was optimally active was pH 8.0. Each reaction contained 20 μM enzyme, 50 mM of potassium phosphate buffer pH 8.0 (25 μL, 100 mM stock), 2.5 mM substrate (2.5 μL, 50 mM stock), 1 mM NADP+ (0.5 μL, 100 mM stock), 1 mM G6P-DH (0.5 μL, 100 μ L^{-1}), 5 mM G6P (5 μL, 5 mM stock), and Milli-Q water to a final volume of 50 μL. Reactions were carried out at 30 °C for 2 h and quenched with the addition of 70 μL acetonitrile (HPLC grade) containing 2.5 mM pentamethylbenzene as an internal standard. Reactions were cleared by centrifugation at 16,000 x g, 12 min and the supernatant was analyzed by UPLC-DAD. Consumption of the starting material was quantified by comparison to standard curves of each substrate.

**Figure 3.31:** The relative effect of pH on AfoD activity with **3.11**. Measurements were done in triplicate Note: Above pH 9.7 protein precipitates cannot determine product formation above that.

Reduction of AfoD WT by NADPH

All spectral measurements were carried out on a UV2501PC spectrophotometer (Shimazu). Addition of NADPH was done at 25 °C. Reaction conditions: 50 mM phosphate buffer pH 8.0, 800 nM enzyme, 250 μM NADPH, and 125 μM substrate **3.11** was carried out in a 1 cm pathlength cuvette. Depletion of NADPH was measured at 362 nm.

Hydrogen Peroxide Formation Assay:

Production of hydrogen peroxide was detected by horseradish peroxidase horseradish peroxidase (HRP) (2.5 μM) and 4-amino antipyrine (0.75 mM)/2,4,6-tribromo-3-hydroxybenzoic acid dye (0.02% W/V), phosphate buffer pH 8.0 (100 mM). In a 96-well plate, 190 μL of the assay solution was combined with 10 μL of peroxide (16 μM-3 mM) to generate a peroxide calibration curve. For reactions, 10 μL of a quenched in vitro reaction was added to 190 μL assay solution to measure peroxide production measured at 510 nm.
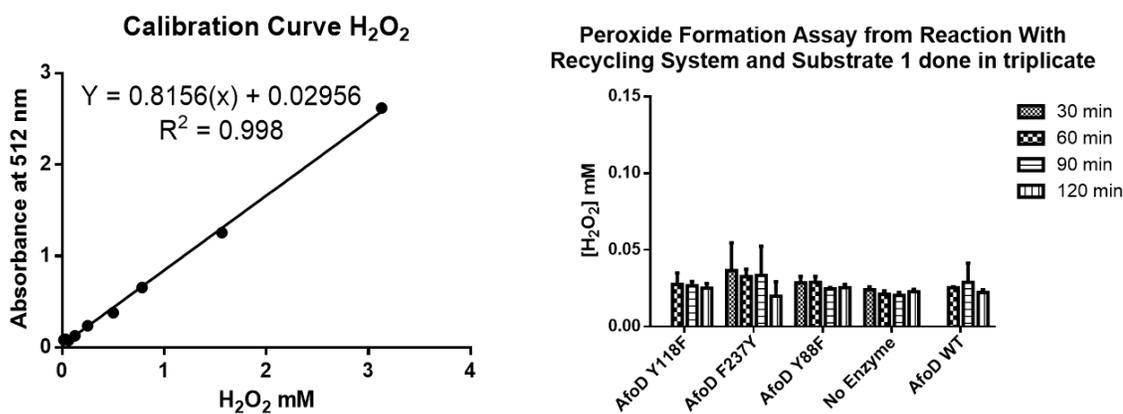
Figure 3.32: Peroxide Formation quantification across different time points from 30 min to 120 min after *in vitro* reaction with substrate 3.11 with AfoD wild-type and variants.
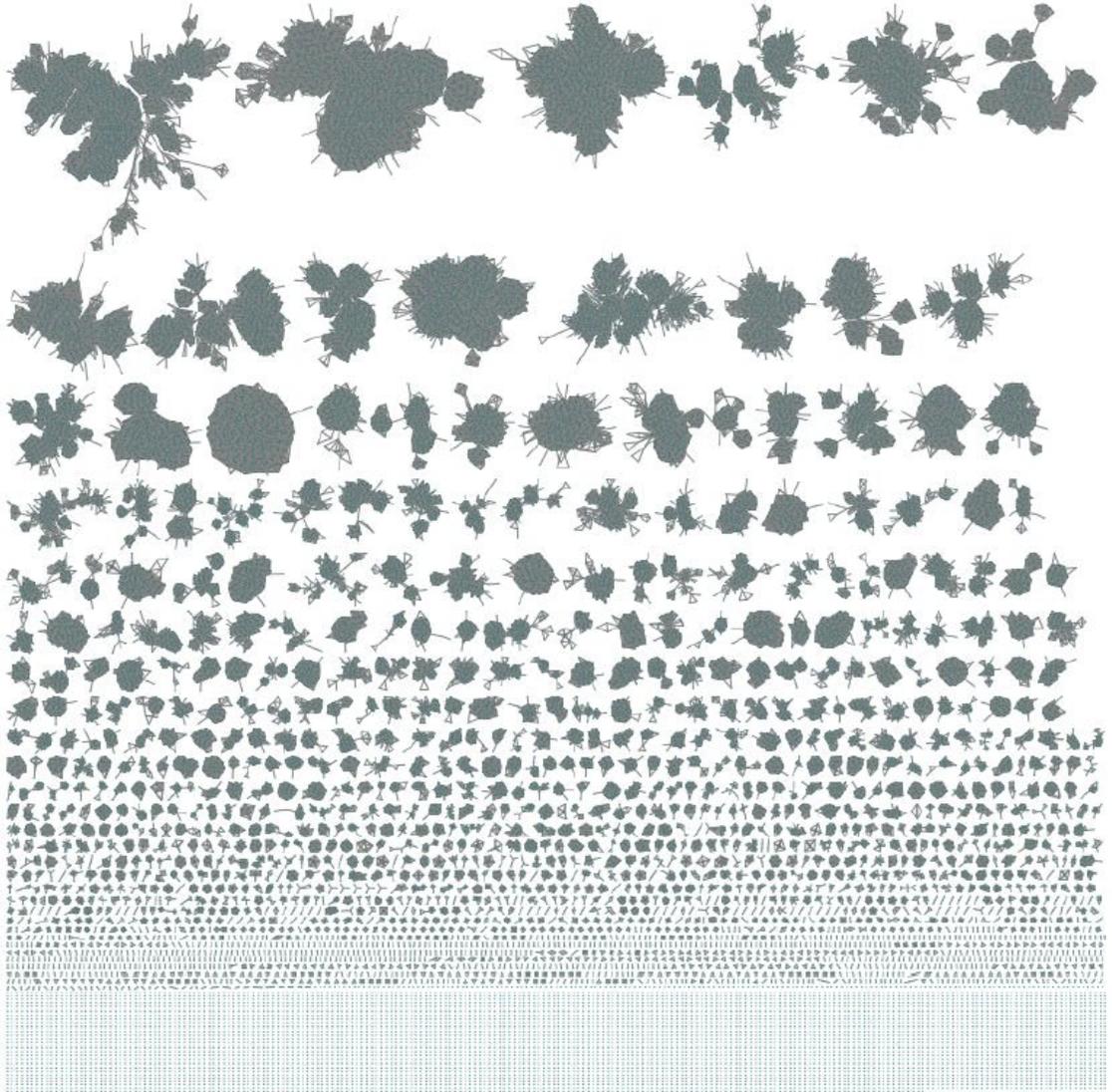
Figure 3.33: SSN of flavin-dependent monooxygenases created using web tools originating from the Enzyme Function Initiative (EFI).This SSN encompasses 45,000 non-redundant FAD-dependent enzyme sequences clustered by sequence similarity. Edge score between 150 and 1,440 inclusive.
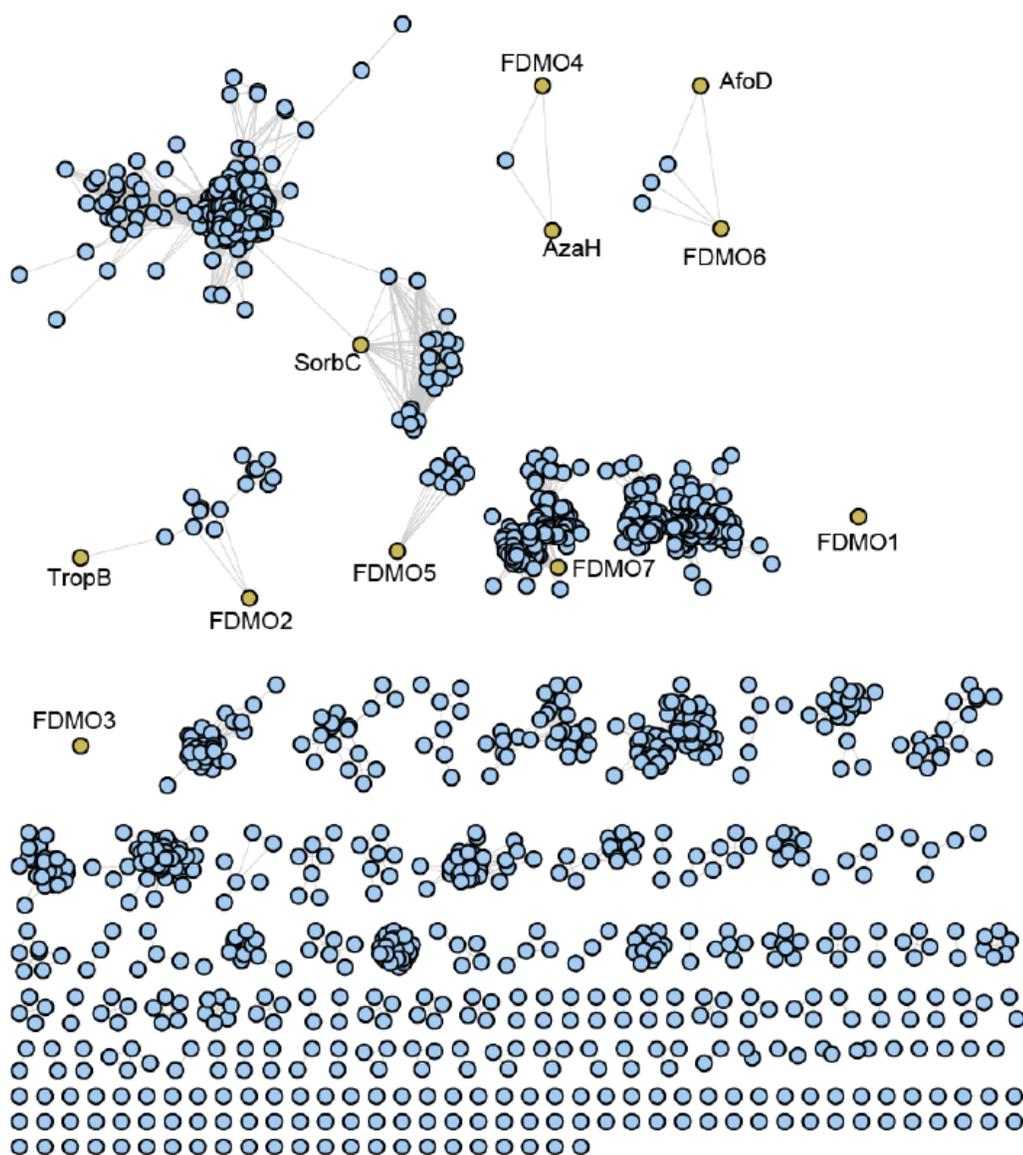
Figure 3.34: SSN of flavin-dependent monooxygenases created using web tools originating from the Enzyme Function Initiative (EFI). Zoom in to cluster containing TropB, AzaH, AfoD, and SorbC with and organic layout applied to rearrange the sequences. Edge score 150.

VI. FDMO Sequence Alignment

Figure 3.35: AfoD cluster sequence alignment with other sequences within its cluster Y118 and F237 are highlighted. Alignment is colored by conservation.

Figure 3.36: FDMO5 cluster sequence alignment with other sequences within its cluster Y118 and F237 are highlighted (From AfoD). Alignment is colored by conservation.

Figure 3.37: TropB cluster sequence alignment with other sequences within its cluster Y118 and F237 are highlighted (From AfoD). Alignment is colored by conservation.

CHAPTER IV

Conclusions and Future Directions

There has been an uprising need to access molecules with complex structures and a desire for chemical reactions that are operationally simple, robust, and environmentally friendly. In the field of biocatalysis, we aim to leverage microorganisms and enzymes to catalyze challenging chemical transformations that meet these requirements for multiple chemical transformations. [118, 119] However, there has been a gap between identifying suitable biocatalysts and integrating them in synthetic chemists' toolbox. [9, 120, 121] The body of work in this thesis seeks to identify and characterize new biocatalysts for the synthesis of complex natural products to bridge the gap between discovery and implementation. We narrowed our scope for the identification of biocatalysts to FAD-dependent monooxygenases (FDMO), which only require molecular oxygen and a nicotinamide cofactor for the oxidative dearomatization reaction. Oxidative dearomatization is a powerful chemical transformation that builds complexity in phenolic compounds. FMDOs can enable this reaction under mild reaction conditions and high site- and stereoselectivity. [13, 22] This Chapter summarizes the results of the preceding two chapters describing the site- and enantioselectivity of a selection of FMDOs through structural, sequence, and modeling insights.

We anticipated that FMDOs can provide an excellent platform due to their exceptional selectivity. We previously explored the substrate scope and scalability of a panel

of FDMOs the mediate the oxidative dearomatization reaction. [22] In this initial study, we demonstrated that three biocatalysts, TropB, AzaH, and SorbC, had high levels of site- and enantioselectivity across a panel of substrates. Given the excellent reaction profile of TropB we wanted to enhance its synthetic utility by expanding its substrate scope and accessing different reaction pathways that are not native to the catalyst. We aimed to gain a structural and mechanistic understanding of this highly selective FDMO, TropB, to achieve our goal. I solved the TropB crystal structure and upon close inspection of the active site, we concluded that it was sufficiently spacious to accommodate a wide range of substrates. Structural studies coupled with molecular dynamics simulations and mutagenesis studies show that the residues Arg206 and Tyr239 are involved in a two-point binding of the substrate. Interactions with these residues allow for excellent selectivity. These findings will aid with future protein engineering and reaction development efforts with TropB and related FDMOs.

We wanted to explore how the mechanistic findings of TropB are translated to other FMDOs systems. Using the TropB sequence as a template, we employed a sequence similarity network (SSN). This approach enabled a rapid and logical investigation of proteins that are likely to demonstrate similar reactivity or selectivity reducing the search space for screening. Through this approach, we identify the FDMO AfoD located between the clusters of SorbC and TropB, which have complementary enantioselectivity. AfoD yielded the C3-hydroxylated product similarly to TropB but with complementary stereoselectivity. Upon close inspection of the sequence alignment of TropB and AfoD, we identified that AfoD has complementary residues that in the TropB system were critical for selectivity. Taking this sequence fingerprint into consideration, we designed a focused library that yielded TropB-like products when an equivalent to Tyr239 was present in the active site and AfoD-like products when an equivalent to Tyr118 was present in the active site. Utilizing these modern tools in bioinformatics allowed us to rapidly identify FDMO homologs with desired properties, an orthogonal approach to protein engineering.

4.1 Future Directions

FMDO expanded library

Through the bioinformatic tools we have exploited and structural and sequence insights from the FMDOs we have studied we aim to identify new catalysts with complementary activity. Leveraging the SSN we have built we can canvas and identify biocatalysts that are similar to TropB and AfoD, but with a more promiscuous substrate scope. This will allow for selecting enzymes that are similar in reactivity but with perhaps a different substrate scope or selectivity depending on how th enzyme selected is clustered. This approach can be coupled with machine learning and the design of ancestral sequences between AfoD and TropB. Enzymes have evolved to react with specific substrates narrowing their scope. However, through machine learning, we can take a protein sequence and computationally predict what the potential ancestral sequence was. We anticipate that this approach will yield enzymes with a wide range of site-, stereoselectivity and substrate scope that can be leveraged in the synthesis of complex natural products.

The different approaches presented in this dissertation demonstrate the potential that FMDOs can have in the synthesis of natural products. However, when designing a library of large scale enzymes they have to be soluble under similar conditions which was a problem with some of the enzymes within our planned library and might need to be address in the future when selecting enzymes. The substrate scope of certain enzymes can also be challenging to predict and relies to be tested experimentally to determine the enzyme promiscuity across a panel of substrates.

This work highlights the identification of new biocatalysts, structural insights, and techniques to better canvas the sequence space for a biocatalyst with the desired activity. I anticipate that as our understanding of the reactivity of different FDMOs grows we can design biocatalysts that fit our needs in the future.

Going beyond the Class A FAD dependent monooxygenases

Currently, there are more than 216 million annotated protein sequences available in public databases, a number that doubles every 28 months, and just like the deep-sea floor, only a minuscule portion of this territory has been explored. Combing through these databases for a well-suited catalyst for the desired reaction with a specific substrate scope and enantioselectivity can be challenging. For the most part, these sequences are annotated with putative functions. Profiling function across a sequence space can be possible if a subset sample is characterized against a panel of substrates. This approach will allow researchers to identify trends in reactivity across a family of enzymes and ultimately identifying enzymes that are capable of the desired reaction. With the evolution of tools available for canvassing and identifying sequences with untapped synthetic potential, there can be a leap into determining the function based on the sequences. Some of these tools include sequence similarity networks (SSNs) and the variational autoencoder latent space model (VAE), which allow scientists to hone in on motifs that can potentially be used to predict enzyme function or pinpoint residues that might be important for a specific enzyme function. The example in this thesis showcases the power of enzymes that hide in plain sight. The sequences are known, but their reactivity will remain a mystery without dedicated experimental work toward family-wide reactivity profiling. These efforts are guided by tools for visualizing sequence space and have the potential to bring light to the deep-sea floor of unexplored enzymes.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] M. O. Nafiu, A. A. Hamid, H. F. Muritala, S. B. Adeyemi, *Medicinal Spices and Vegetables from Africa: Therapeutic Potential Against Metabolic, Inflammatory, Infectious and Systemic Diseases* (Elsevier Inc., 2017), pp. 171–204.

[2] E. Patridge, P. Gareiss, M. S. Kinch, D. Hoyer, "An analysis of FDA-approved drugs: Natural products and their derivatives" (2016). 2013 one third of the FAD-approved new molecular entities were based or derived from natural products.

[3] M. E. Welsch, S. A. Snyder, B. R. Stockwell, "Privileged scaffolds for library design and drug discovery" (2010).

[4] R. Mullin, "Pfizer offers a 'green' progesterone", *Chemistry and Engineering News* (2018).

[5] M. A. Huffman, *et al.*, "Design of an in vitro biocatalytic cascade for the manufacture of islatravir", *Science* **366**, 1255 (2019). DOI: 10.1126/science.aay8484.

[6] Y. Ni, D. Holtmann, F. Hollmann, "How green is biocatalysis? to calculate is to know" (2014).

[7] R. A. Sheldon, J. M. Woodley, "Role of Biocatalysis in Sustainable Chemistry" (2018).

[8] "GenBank and WGS Statistics", (Accessed: 2020-09-18).

[9] U. T. Bornscheuer, *et al.*, "Engineering the third wave of biocatalysis", *Nature* **485**, 185 (2012). DOI: 10.1038/nature11117.

[10] M. D. Truppo, "Biocatalysis in the Pharmaceutical Industry: The Need for Speed", *ACS Medicinal Chemistry Letters* **8**, 476 (2017). DOI: 10.1021/acsmedchemlett.7b00114.

[11] M. M. Huijbers, S. Montersino, A. H. Westphal, D. Tischler, W. J. Van Berkel, "Flavin dependent monooxygenases", *Archives of Biochemistry and Biophysics* **544**, 2 (2014). DOI: 10.1016/j.abb.2013.12.005.

[12] M. H. Eppink, H. A. Schreuder, W. J. Van Berkel, "Identification of a novel conserved sequence motif in flavoprotein hydroxylases with a putative dual function in FAD/-NAD(P)H binding", *Protein Science* **6**, 2454 (1997). DOI: 10.1002/pro.5560061119.

[13] W. J. van Berkel, N. M. Kamerbeek, M. W. Fraaije, "Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts", *Journal of Biotechnology* **124**, 670 (2006). DOI: 10.1016/j.jbiotec.2006.03.044.

[14] B. A. Palfey, C. A. McDonald, "Control of catalysis in flavin-dependent monooxygenases", *Archives of Biochemistry and Biophysics* **493**, 26 (2010). DOI: 10.1016/j.abb.2009.11.028.

[15] D. L. Gatti, *et al.*, "The mobile flavin of 4-OH benzoate hydroxylase.", *Science* **266**, 110 (1994). DOI: 10.1126/science.7939628.

[16] D. P. Ballou, B. Entsch, L. J. Cole, "Dynamics involved in catalysis by single-component and two-component flavin-dependent aromatic hydroxylases", *Biochemical and Biophysical Research Communications* **338**, 590 (2005). DOI: 10.1016/j.bbrc.2005.09.081.

[17] R. A. Sheldon, J. M. Woodley, "Role of Biocatalysis in Sustainable Chemistry", *Chemical Reviews* **118**, 801 (2018). DOI: 10.1021/acs.chemrev.7b00203.

[18] R. D. Ceccoli, D. A. Bianchi, D. V. Rial, "Flavoprotein monooxygenases for oxidative biocatalysis: Recombinant expression in microbial hosts and applications", *Frontiers in Microbiology* **5**, 1 (2014). DOI: 10.3389/fmicb.2014.00025.

[19] Y. M. Chiang, *et al.*, "A gene cluster containing two fungal polyketide synthases encodes the biosynthetic pathway for a polyketide, asperfuranone, in aspergillus nidulans", *Journal of the American Chemical Society* **131**, 2965 (2009). DOI: 10.1021/ja8088185.

[20] A. O. Zabala, W. Xu, Y. H. Chooi, Y. Tang, "Characterization of a silent azaphilone gene cluster from Aspergillus niger ATCC 1015 reveals a hydroxylation-mediated pyran-ring formation", *Chemistry and Biology* **19**, 1049 (2012). DOI: 10.1016/j.chembiol.2012.07.004.

[21] A. A. Fahad, *et al.*, "Oxidative dearomatisation: the key step of sorbicillinoid biosynthesis", *Chemical science* **5**, 523 (2014). DOI: 10.1039/c3sc52911h.

[22] S. A. Baker Dockrey, A. L. Lukowski, M. R. Becker, A. R. H. Narayan, "Biocatalytic site- and enantioselective oxidative dearomatization of phenols", *Nature Chemistry* **10**, 119 (2017). DOI: 10.1038/nchem.2879.

[23] M. Achard, A. B. Beeler, J. A. Porco, "Synthesis of azaphilone-based chemical libraries", *ACS Combinatorial Science* **14**, 236 (2012). DOI: 10.1021/co300002x.

[24] B. F. Fisher, H. M. Snodgrass, K. A. Jones, M. C. Andorfer, J. C. Lewis, "Site-Selective C–H Halogenation Using Flavin-Dependent Halogenases Identified via Family-Wide Activity Profiling", *ACS Central Science* (2019). DOI: 10.1021/acscentsci.9b00835.

[25] J. A. Gerlt, *et al.*, "Enzyme function initiative-enzyme similarity tool (EFI-EST): A web tool for generating protein sequence similarity networks" (2015).

[26] F. Madeira, *et al.*, "The EMBL-EBI search and sequence analysis tools APIs in 2019", *Nucleic Acids Research* **47**, W636 (2019). DOI: 10.1093/nar/gkz268.

[27] G. A. Pavlopoulos, T. G. Soldatos, A. Barbosa-Silva, R. Schneider, "A reference guide for tree analysis and visualization" (2010).

[28] X. Ding, "Methodological Advances for Drug Discovery and Protein Engineering", *Tech. rep.* (2018).

[29] J. A. Gerlt, "Genomic Enzymology: Web Tools for Leveraging Protein Family Sequence-Function Space and Genome Context to Discover Novel Functions", *Biochemistry* (2017). DOI: 10.1021/acs.biochem.7b00614.

[30] J. B. Pyser, *et al.*, "Stereodivergent, Chemoenzymatic Synthesis of Azaphilone Natural Products", *Journal of the American Chemical Society* **141**, 18551 (2019). DOI: 10.1021/jacs.9b09385.

[31] A. Rodríguez Benítez, *et al.*, "Structural Basis for Selectivity in Flavin-Dependent Monooxygenase-Catalyzed Oxidative Dearomatization", *ACS Catalysis* **9**, 3633 (2019). DOI: 10.1021/acscatal.8b04575.

[32] S. P. Roche, J. A. Porco, "Dearomatization strategies in the synthesis of complex natural products.", *Angewandte Chemie International Edition* **50**, 4068 (2011). DOI: 10.1002/anie.201006017.

[33] J. Zhu, N. P. Grigoriadis, J. P. Lee, J. A. Porco, "Synthesis of the azaphilones using copper-mediated enantioselective oxidative dearomatization", *Journal of the American Chemical Society* **127**, 9342 (2005). DOI: 10.1021/ja052049g.

[34] W. X. Wang, *et al.*, "Antibacterial Azaphilones from an Endophytic Fungus, Colletotrichum sp. BS4", *Journal of Natural Products* **79**, 704 (2016). DOI: 10.1021/acs.jnatprod.5b00436.

[35] Q. Yang, *et al.*, "Evolution of an oxidative dearomatization enabled total synthesis of vinigrol", *Organic and Biomolecular Chemistry* **12**, 330 (2014). DOI: 10.1039/c3ob42191k.

[36] H. Y. Shiao, H. P. Hsieh, C. C. Liao, "First total syntheses of (±)-annuionone B and (±)- tanarifuranonol", *Organic Letters* **10**, 449 (2008). DOI: 10.1021/ol7028178.

[37] W. Sun, G. Li, L. Hong, R. Wang, "Asymmetric dearomatization of phenols", *Organic and Biomolecular Chemistry* **14**, 2164 (2016). DOI: 10.1039/c5ob02526e.

[38] M. Uyanik, T. Yasui, K. Ishihara, "Hydrogen bonding and alcohol effects in asymmetric hypervalent iodine catalysis: Enantioselective oxidative dearomatization of phenols", *Angewandte Chemie - International Edition* **52**, 9215 (2013). DOI: 10.1002/anie.201303559.

[39] C. Bosset, *et al.*, "Asymmetric hydroxylative phenol dearomatization promoted by chiral binaphthylic and biphenylic iodanes", *Angewandte Chemie - International Edition* **53**, 9860 (2014). DOI: 10.1002/anie.201403571.

[40] G. Billek, J. SWOBODA, F. WESSELY, "Die Dienon-phenol-umlgerung des 2,2-diacetoxy-cyclohexa-3,5-dienons-1", *Tetrahedron* **18**, 909 (1962).

[41] K. A. Volp, A. M. Harned, "Chiral aryl iodide catalysts for the enantioselective synthesis of para-quinols", *Chemical Communications* **49**, 3001 (2013). DOI: 10.1039/c3cc00013c.

[42] J. K. Boppisetti, V. B. Birman, "Asymmetric Oxidation of o- Alkylphenols with Chiral 2- (o- Iodoxyphenyl) -oxazolines", *Org. Lett* **11**, 1221 (2009). DOI: 10.1021/ol8029092.

[43] T. Hashimoto, Y. Shimazaki, Y. Omatsu, K. Maruoka, "Indanol-Based Chiral Organoiodine Catalysts for Enantioselective Hydrative Dearomatization", *Angewandte Chemie International Edition* **57**, 7200 (2018). DOI: 10.1002/anie.201803889.

[44] T. Uemura, *et al.*, "The catalytic mechanism of decarboxylative hydroxylation of salicylate hydroxylase revealed by crystal structure analysis at 2.5 Å resolution", *Biochemical and Biophysical Research Communications* **469**, 158 (2016). DOI: 10.1016/J.BBRC.2015.11.087.

[45] S. Montersino, *et al.*, "Crystal structure of 3-hydroxybenzoate 6-hydroxylase uncovers lipid-assisted flavoprotein strategy for regioselective aromatic hydroxylation", *Journal of Biological Chemistry* **288**, 26235 (2013). DOI: 10.1074/jbc.M113.479303.

[46] K. S. Ryan, *et al.*, "Crystallographic trapping in the rebeccamycin biosynthetic enzyme RebC", *Proc Natl Acad Sci U S A* **104**, 15311 (2007). DOI: 10.1073/pnas.0707190104.

[47] G. R. Moran, B. Entsch, B. A. Palfey, D. P. Ballou, "Evidence for flavin movement in the function of p-hydroxybenzoate hydroxylase from studies of the mutant Arg220Lys", *Biochemistry* **35**, 9278 (1996). DOI: 10.1021/bi960360s.

[48] H. A. Schreuder, *et al.*, "Crystal Structures of Wild-Type p-Hydroxybenzoate Hydroxylase Complexed with 4-Aminobenzoate, 2,4-Dihydroxybenzoate, and 2-Hydroxy-4-aminobenzoate and of the Tyr222Ala Mutant Complexed with 2-Hydroxy-4-aminobenzoate. Evidence for a Proton Channel and a Ne", *Biochemistry* **33**, 10161 (1994). DOI: 10.1021/bi00199a044.

[49] B. A. Palfey, D. P. Ballou, V. Massey, "Flavin conformational changes in the catalytic cycle of p- hydroxybenzoate hydroxylase substituted with 6-azido- and 6-aminoflavin adenine dinucleotide", *Biochemistry* **36**, 15713 (1997). DOI: 10.1021/bi971427u.

[50] K. K. Frederick, B. A. Palfey, "Kinetics of proton-linked flavin conformational changes in p-hydroxybenzoate hydroxylase", *Biochemistry* **44**, 13304 (2005). DOI: 10.1021/bi051119t.

[51] R. S. Heath, M. Pontini, B. Bechi, N. J. Turner, "Development of an R-selective amine oxidase with broad substrate specificity and high enantioselectivity", *ChemCatChem* **6**, 996 (2014). DOI: 10.1002/cctc.201301008.

[52] C. Enroth, IUCr, "High-resolution structure of phenol hydroxylase and correction of sequence errors", *Acta Crystallographica Section D Biological Crystallography* **59**, 1597 (2003). DOI: 10.1107/S0907444903014902.

[53] T. Hiromoto, S. Fujiwara, K. Hosokawa, H. Yamaguchi, "Crystal Structure of 3-Hydroxybenzoate Hydroxylase from Comamonas testosteroni Has a Large Tunnel for Substrate and Oxygen Access to the Active Site", *Journal of Molecular Biology* **364**, 878 (2006). DOI: 10.1016/J.JMB.2006.09.031.

[54] J. Kobayashi, *et al.*, "Role of the Tyr270 residue in 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase from Mesorhizobium loti", *Journal of Bioscience and Bioengineering* **123**, 154 (2017). DOI: 10.1016/j.jbiosc.2016.07.022.

[55] A. Abood, *et al.*, "Kinetic characterisation of the FAD dependent monooxygenase TropB and investigation of its biotransformation potential", *RSC Adv.* **5**, 49987 (2015). DOI: 10.1039/C5RA06693J.

[56] B. A. Palfey, B. Entsch, D. P. Ballou, V. Massey, "Changes in the Catalytic Properties of p-Hydroxybenzoate Hydroxylase Caused by the Mutation Asn300Asp* *", *Tech. rep.* (1994).

[57] B. Entsch, B. A. Palfey, D. P. Ballou, V. Massey, "Catalytic function of tyrosine residues in para-hydroxybenzoate hydroxylase as determined by the study of site-directed mutants", *Journal of Biological Chemistry* **266**, 17341 (1991). DOI: 10.1016/S0021-9258(19)47379-1.

[58] J. K. Gagnon, S. M. Law, C. L. Brooks, "Flexible CDOCKER: Development and application of a pseudo-explicit structure-based docking method within CHARMM", *Journal of Computational Chemistry* **37**, 753 (2016). DOI: 10.1002/jcc.24259.

[59] T. Wymore, *et al.*, "Molecular recognition of aldehydes by aldehyde dehydrogenase and mechanism of nucleophile activation", *Proteins: Structure, Function, and Bioinformatics* **57**, 758 (2004). DOI: 10.1002/prot.20256.

[60] Kyriacos C. Nicolaou and Scott A. Snyder, *Classics in Total Synthesis II: More Targets, Strategies, Methods* (Wiley, 2003).

[61] M. Gaus, Q. Cui, M. Elstner, "DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB)", *Journal of Chemical Theory and Computation* **7**, 931 (2011). DOI: 10.1021/ct100684s.

[62] A. Aliverti, B. Curti, M. A. Vanoni, *Flavoprotein Protocols* (Humana Press, New Jersey, 1999), pp. 9–24.

[63] W. Kabsch, IUCr, "XDS", *Acta Crystallographica Section D Biological Crystallography* **66**, 125 (2010). DOI: 10.1107/S0907444909047337.

[64] M. D. Winn, *et al.*, "Overview of the CCP4 suite and current developments", *Acta Crystallographica Section D Biological Crystallography* **67**, 235 (2011). DOI: 10.1107/S0907444910045749.

[65] A. J. McCoy, *et al.*, "Phaser crystallographic software", *Journal of Applied Crystallography* **40**, 658 (2007). DOI: 10.1107/S0021889807021206.

[66] T. C. Terwilliger, *et al.*, "Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard", *Acta Crystallographica Section D Biological Crystallography* **64**, 61 (2008). DOI: 10.1107/S090744490705024X.

[67] P. Emsley, K. Cowtan, IUCr, "Coot : model-building tools for molecular graphics", *Acta Crystallographica Section D Biological Crystallography* **60**, 2126 (2004). DOI: 10.1107/S0907444904019158.

[68] L. Schrodinger, "The PyMOL Molecular Graphics System" (2015).

[69] E. Krissinel, K. Henrick, "Inference of Macromolecular Assemblies from Crystalline State", *Journal of Molecular Biology* **372**, 774 (2007). DOI: 10.1016/J.JMB.2007.05.022.

[70] E. Jurrus, *et al.*, "Improvements to the APBS biomolecular solvation software suite", *Protein Science* **27**, 112 (2018). DOI: 10.1002/pro.3280.

[71] V. B. Chen, *et al.*, "MolProbity: all-atom structure validation for macromolecular crystallography", *Acta Crystallographica Section D Biological Crystallography* **66**, 12 (2010). DOI: 10.1107/S0907444909042073.

[72] W. J. H. BERKEL, W. J. J. TWEEL, "Purification and characterisation of 3-hydroxyphenylacetate 6-hydroxylase: a novel FAD-dependent monooxygenase from a Flavobacterium species", *European Journal of Biochemistry* **201**, 585 (1991). DOI: 10.1111/j.1432-1033.1991.tb16318.x.

[73] P. J. Steennis, M. M. Cordes, J. G. H. Hilkens, F. Miller, "On the interaction of para-hydroxybenoazte hydroxylase from pseudomonas fluorescens with halogen ions" **36**, 177 (1973).

[74] B. A. Palfey, *Time Resolved Spectral Analysis, in Kinetic Analysis of Macromolecules* (Oxford University Press, New York, 2003).

[75] S. Strickland, G. Palmer, V. Massey, "Determination of Dissociation Constants and Specific Constants of Enzyme-Substrate (or Protein-Ligand) Interactions from Rapid Reaction Kinetic Data", *The Journal of biological chemistry* **250**, 4048 (1975).

[76] L. Gonzalez-Osorio, K. Luong, S. Jirde, B. A. Palfey, J. L. Vey, "Initial investigations of C4a-(hydro)peroxyflavin intermediate formation by dibenzothiophene monooxygenase", *Biochemical and Biophysical Research Communications* **481**, 189 (2016). DOI: 10.1016/J.BBRC.2016.10.145.

[77] B. R. Brooks, *et al.*, "CHARMM: The biomolecular simulation program", *Journal of Computational Chemistry* **30**, 1545 (2009). DOI: 10.1002/jcc.21287.

[78] S. Jo, *et al.*, "CHARMM-GUI 10 years for biomolecular modeling and simulation", *Journal of Computational Chemistry* **38**, 1114 (2017). DOI: 10.1002/jcc.24660.

[79] K. Vanommeslaeghe, A. D. Mackerell, "Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing", *J. Chem. Inf. Model.* **52**, 3144 (2012). DOI: 10.1021/ci300363c.

[80] K. Vanommeslaeghe, E. P. Raman, A. D. Mackerell, "Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges", *J. Chem. Inf. Model.* **52**, 3155 (2012). DOI: 10.1021/ci3003649.

[81] N. M. O'Boyle, *et al.*, "Open Babel: An open chemical toolbox", *Journal of Cheminformatics* **3**, 33 (2011). DOI: 10.1186/1758-2946-3-33.

[82] W. Im, M. S. Lee, C. L. Brooks, "Generalized born model with a simple smoothing function", *Journal of Computational Chemistry* **24**, 1691 (2003). DOI: 10.1002/jcc.10321.

[83] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, "Comparison of simple potential functions for simulating liquid water", *The Journal of Chemical Physics* **79**, 926 (1983). DOI: 10.1063/1.445869.

[84] P. Eastman, *et al.*, "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics", *PLOS Computational Biology* **13**, e1005659 (2017). DOI: 10.1371/journal.pcbi.1005659.

[85] C. M. Breneman, K. B. Wiberg, "Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis", *Journal of Computational Chemistry* **11**, 361 (1990). DOI: 10.1002/jcc.540110311.

[86] F. Neese, "The ORCA program system", *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2**, 73 (2012). DOI: 10.1002/wcms.81.

[87] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, "Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields", *J. Phys. Chem* **98**, 12 (1994).

[88] S. H. Vosko, L. Wilk, M. Nusair, "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis", *Canadian Journal of Physics* **58**, 1200 (1980). DOI: 10.1139/p80-159.

[89] C. Lee, W. Yang, R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density", *Physical Review B* **37**, 785 (1988). DOI: 10.1103/PhysRevB.37.785.

[90] A. D. Becke, "Density-functional thermochemistry. III. The role of exact exchange", *The Journal of Chemical Physics* **98**, 5648 (1993). DOI: 10.1063/1.464913.

[91] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, "A consistent and accurate <i>ab initio</i> parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu", *The Journal of Chemical Physics* **132**, 154104 (2010). DOI: 10.1063/1.3382344.

[92] S. Grimme, S. Ehrlich, L. Goerigk, "Effect of the damping function in dispersion corrected density functional theory", *Journal of Computational Chemistry* **32**, 1456 (2011). DOI: 10.1002/jcc.21759.

[93] F. Weigend, R. Ahlrichs, "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy", *Physical Chemistry Chemical Physics* **7**, 3297 (2005). DOI: 10.1039/b508541a.

[94] F. Weigend, "Accurate Coulomb-fitting basis sets for H to Rn", *Physical Chemistry Chemical Physics* **8**, 1057 (2006). DOI: 10.1039/b515623h.

[95] A. Hellweg, C. Hättig, S. Höfener, W. Klopper, "Optimized accurate auxiliary basis sets for RI-MP2 and RI-CC2 calculations for the atoms Rb to Rn", *Theoretical Chemistry Accounts* **117**, 587 (2007). DOI: 10.1007/s00214-007-0250-5.

[96] L. A. Nguyen, H. He, C. Pham-Huy, "Chiral drugs: an overview.", *International journal of biomedical science : IJBS* **2**, 85 (2006).

[97] F. Giffhorn, "Fungal pyranose oxidases: Occurrence, properties and biotechnical applications in carbohydrate chemistry", *Applied Microbiology and Biotechnology* **54**, 727 (2000). DOI: 10.1007/s002530000446.

[98] J. A. Laakso, *et al.*, "CT2108A and B: New fatty acid synthase inhibitors as antifungal agents", *Journal of Natural Products* **66**, 1041 (2003). DOI: 10.1021/np030046g.

[99] D. N. Quang, M. Stadler, J. Fournier, A. Tomita, T. Hashimoto, "Cohaerins C-F, four azaphilones from the xylariaceous fungus Annulohypoxylon cohaerens", *Tetrahedron* **62**, 6349 (2006). DOI: 10.1016/j.tet.2006.04.040.

[100] E. YOSHIDA, H. FUJIMOTO, M. BABA, M. YAMAZAKI, "Four New Chlorinated Azaphilones, Helicusins A-D, Closely Related to 7-epi-Sclerotiorin, from an Ascomycetous Fungus, Talaromyces helicus.", *CHEMICAL & PHARMACEUTICAL BULLETIN* **43**, 1307 (1995). DOI: 10.1248/cpb.43.1307.

[101] A. A. Stierle, D. B. Stierle, T. Bugni, "Sequoiatones A and B: Novel antitumor metabolites isolated from a redwood endophyte", *Journal of Organic Chemistry* **64**, 5479 (1999). DOI: 10.1021/jo990277l.

[102] K. Kaur, *et al.*, "The fungal natural product azaphilone-9 binds to HuR and inhibits HuR-RNA interaction in vitro", *PLOS ONE* **12**, e0175471 (2017). DOI: 10.1371/journal.pone.0175471.

[103] W. Wang, *et al.*, "Chlorinated Azaphilone Pigments with Antimicrobial and Cytotoxic Activities Isolated from the Deep Sea Derived Fungus Chaetomium sp. NA-S01-R1", *Marine Drugs* **16**, 61 (2018). DOI: 10.3390/md16020061.

[104] K. Yasukawa, *et al.*, "Azaphilones Inhibit Tumor Promotion by 12-O-Tetradecanoylphorbol-13-Acetate in Two-Stage Carcinogenesis in Mice", *Oncology* **51**, 108 (1994). DOI: 10.1159/000227320.

[105] K. Matsuzaki, *et al.*, "New brominated and halogen-less derivatives and structure-activity relationship of azaphilones inhibiting gp120-CD4 binding.", *The Journal of antibiotics* **51**, 1004 (1998). DOI: 10.7164/antibiotics.51.1004.

[106] J. L. Tang, *et al.*, "Azaphilone Alkaloids with Anti-inflammatory Activity from Fungus Penicillium sclerotiorum cib-411", *Journal of Agricultural and Food Chemistry* **67**, 2175 (2019). DOI: 10.1021/acs.jafc.8b05628.

[107] W. H. Hsu, T. H. Chen, B. H. Lee, Y. W. Hsu, T. M. Pan, "Monascin and ankaflavin act as natural AMPK activators with PPARα agonist activity to down-regulate nonalcoholic steatohepatitis in high-fat diet-fed C57BL/6 mice", *Food and Chemical Toxicology* **64**, 94 (2014). DOI: 10.1016/j.fct.2013.11.015.

[108] W. B. Whalley, G. Ferguson, W. C. Marsh, R. J. Restivo, "The chemistry of fungi. Part LXVIII. The absolute configuration of (+)-sclerotiorin and of the azaphilones", *Journal of the Chemical Society, Perkin Transactions 1* **0**, 1366 (1976). DOI: 10.1039/p19760001366.

[109] S. Udagawa, "(-)-Sclerotiorin, A Major Metabolite of Penicillium hirayamae UDA-GAWA.", *CHEMICAL & PHARMACEUTICAL BULLETIN* **11**, 366 (1963). DOI: 10.1248/cpb.11.366.

[110] J.-H. Park, *et al.*, "Antifungal activity against plant pathogenic fungi of chaetoviridins isolated from <Chaetomium globosum<", *FEMS Microbiology Letters* **252**, 309 (2005). DOI: 10.1016/j.femsle.2005.09.013.

[111] E. Thines, H. Anke, O. Sterner, "Trichoflectin, a bioactive azaphilone from the ascomycete Trichopezizella nidulus", *Journal of Natural Products* **61**, 306 (1998). DOI: 10.1021/np970469g.

[112] M. Nukina, S. Marumo, "Lunatoic acid A and B, aversion factor and its related metabolite of cochliobolus lunata", *Tetrahedron Letters* **18**, 2603 (1977). DOI: 10.1016/S0040-4039(01)83831-4.

[113] R. Zallot, N. O. Oberg, J. A. Gerlt, "'Democratized' genomic enzymology web tools for functional assignment" (2018).

[114] J. Davison, *et al.*, "Genetic, molecular, and biochemical basis of fungal tropolone biosynthesis", *Proc Natl Acad Sci U S A.* **109**, 7642 (2012).

[115] A. D. Somoza, K. H. Lee, Y. M. Chiang, B. R. Oakley, C. C. Wang, "Reengineering an azaphilone biosynthesis pathway in Aspergillus nidulans to create lipoxygenase inhibitors", *Organic Letters* (2012). DOI: 10.1021/ol203094k.

[116] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. Sternberg, "The Phyre2 web portal for protein modeling, prediction and analysis", *Nature Protocols* (2015). DOI: 10.1038/nprot.2015.053.

[117] N. A. Baker Dockrey S.A, Suh C.E., Rodríguez Benítez A., Wymore T., Brooks C.L. III, "Positioning-Group-Enabled Biocatalytic Oxidative Dearomatization", *ACS Cent. Sci* **5**, 1010 (2019). DOI: 10.1021/acscentsci.9b00163.

[118] A. R. Alcántara, "Biocatalysis and Pharmaceuticals: A Smart Tool for Sustainable Development", *Catalysts* **9**, 792 (2019). DOI: 10.3390/catal9100792.

[119] N. C. Goodwin, J. P. Morrison, D. E. Fuerst, T. Hadi, "Biocatalysis in Medicinal Chemistry: Challenges to Access and Drivers for Adoption" (2019). DOI: 10.1021/acsmedchemlett.9b00410.

[120] U. T. Bornscheuer, "The fourth wave of biocatalysis is approaching", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **376**, 20170063 (2018). DOI: 10.1098/rsta.2017.0063.

[121] A. S. Bommarius, "Biocatalysis: A Status Report", *Annual Review of Chemical and Biomolecular Engineering* **6**, 319 (2015). DOI: 10.1146/annurev-chembioeng-061114-123415.