

# RATIONAL ENGINEERING OF CYTOCHROME P450 ENZYMES

Joshua Gregory Gober

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry.

Chapel Hill  
2017

Approved by:

Eric Brustad

Marcey Waters

Jeffrey Johnson

Bo Li

Gary Pielak

© 2017  
Joshua Gregory Gober  
ALL RIGHTS RESERVED

## **ABSTRACT**

Joshua Gregory Gober: Rational Engineering of Cytochrome P450 Enzymes  
(Under the direction of Eric Brustad)

Cytochrome P450s, which typically catalyze oxidation reactions via an iron-oxo species, have recently been reported to cyclopropanate alkenes in the presence of diazoacetate reagents through formation of an iron carbenoid. Stereoselective enzymes for carbenoid insertion into aryl olefins have been reported; however, engineering selective variants for all of the possible isomers of the reaction remains a challenge. Earlier work with a model P450 (P450 BM3) reported a highly activating mutation of a conserved active site threonine to alanine that resulted in dramatic improvements in enantioselectivity and diastereoselectivity for the model reaction of styrene with ethyl diazoacetate. This work demonstrates that by incorporating this single mutation into a diverse panel of a dozen P450s from various microorganisms, enantioselective and diastereoselective catalysts can be quickly identified for all isomers in the reaction of styrene with ethyl diazoacetate.

This work also demonstrates the utility of intermolecular, P450-mediated olefin cyclopropanation for selective, late-stage modification of complex natural products. In this study, a diverse set of engineered P450s were found to catalyze cyclopropanation of dehydroalanines (Dhas), which are commonly found in natural products. P450s involved in the biosynthesis of a pyridine-containing thiopeptide, thiomuracin GZ, were found to catalyze cyclopropanation of thiomuracin derivatives engineered to display Dhas at key residues. This work presents a strategy

for chemoselective and stereoselective modification of complex natural products using a rational engineering approach.



## **ACKNOWLEDGMENTS**

I would like to thank my advisor Dr. Eric Brustad for his guidance and mentorship. His enthusiasm for teaching and scientific discovery as well as his contagious optimism have inspired me throughout my graduate school career.

I would like to thank my lab members, colleagues, and collaborators for their instrumental roles in my scientific training. I would like to thank Stephanie Baril and Adrienne Huntress for training me in molecular biology techniques. I would also like to thank all of the current and past members of the Brustad lab for helpful discussions and advice for troubleshooting experimental issues. I would like to thank Dr. Albert Bowers for the opportunity to work on a collaborative project with members of his lab, and I would especially like to thank Dr. Swapnil Ghodge of the Bowers lab for his help with synthesis and preparative purifications.

Lastly, I would like to thank my family for their continuous love and support. None of this would have been possible without you. I love you.

## TABLE OF CONTENTS

LIST OF TABLES .....	viii
LIST OF FIGURES .....	ix
LIST OF ABBREVIATIONS .....	xii
CHAPTER 1: NON-NATURAL CARBENOID AND NITRENOID INSERTION REACTIONS CATALYZED BY HEME PROTEINS .....	15
Introduction .....	15
Carbenoid transfer reactions .....	17
Nitrenoid transfer reactions .....	27
Conclusion .....	31
REFERENCES .....	32
CHAPTER 2: MUTATING A HIGHLY CONSERVED RESIDUE IN DIVERSE CYTOCHROME P450S FACILITATES DIASTEREOSELECTIVE OLEFIN CYCLOPROPANATION .....	34
Introduction .....	34
Results .....	37
Discussion .....	41
Conclusion .....	42
Methods .....	43
Supporting Information .....	59
REFERENCES .....	98

CHAPTER 3: P450-MEDIATED NON-NATURAL CYCLOPROPANATION OF  
DEHYDROALANINE-CONTAINING THIOPEPTIDES

Introduction.....	101
Results and Discussion .....	102
Conclusion .....	112
Methods.....	113
Supporting Information.....	148
REFERENCES .....	199

## LIST OF TABLES

Table 1.1. Summary of carbenoid- and nitrenoid-mediated transformations catalyzed by heme-containing enzymes .....	20
Table 2.1. Activities and stereoselectivities of P450 variants for the reaction of styrene with ethyl diazoacetate.....	38
Table 2.2. Substrate scope for most selective variants from the P450 library .....	39
Table S2.1. % Identity to P450 <sub>BM3</sub> , representative PDB accession numbers (PDB#), and expression level in <i>E. coli</i> of each P450 library member .....	69
Table S2.2. Representative substrates and reactions for P450 library members .....	70
Table 3.1. Dha cyclopropanation with P450 variants .....	103
Table 3.2. Dha cyclopropanation with P450 <sub>TbtJ1</sub> and P450 <sub>TbtJ2</sub> variants.....	105
Table 3.3. Dha cyclopropanation on linear thiopeptide cores .....	109
Table S3.1. Conversions and turnovers for EDA dimer formation during cyclopropanation reactions with <b>1</b> .....	162
Table S3.2. Data collection and refinement statistics for TbtJ1 crystallography .....	163

## LIST OF FIGURES

Figure 1.1. Reactive intermediates involved in metalloporphyrin catalyzed transformations .....	18
Figure 1.2. Selectivity, structural insight, and application of P450-catalyzed cyclopropanation.....	19
Figure 1.3. Proposed mechanisms for some carbenoid- and nitrenoid-mediated transformations catalyzed by heme-containing enzymes .....	25
Figure 2.1. Styrene cyclopropanation with ethyl diazoacetate catalyzed by P450 variants.....	35
Figure 2.2. Active site of substrate-free P450 <sub>BM3</sub> (PDB: 2IJ2).....	36
Figure S2.1. ClustalW2 protein sequence alignment of P450 library.....	59
Figure S2.2. Percent identity matrix for P450 library sequence alignment .....	63
Figure S2.3. Active site structures of P450s from library.....	64
Figure S2.4. SDS-PAGE analysis of his-tag purified P450 variants .....	67
Figure S2.5. Chiral GC traces from enzymatic cyclopropanations of styrene with EDA; carbon-monoxide binding spectra.....	73
Figure S2.6. Chiral GC traces from enzymatic cyclopropanations of $\alpha$ -methylstyrene with EDA.....	86
Figure S2.7. Chiral GC traces from enzymatic cyclopropanations of 4-methoxystyrene with EDA.....	90
Figure S2.8. Chiral GC traces from enzymatic cyclopropanations of 4-(trifluoromethyl)styrene with EDA.....	94
Figure 3.1. Key post-translational modifications of thiomuracin GZ core.....	105
Figure 3.2. P450-mediated cyclopropanation of linear I8Dha thiopeptide core substrates bearing either one ( <b>7</b> , panel A) or three ( <b>9</b> , panel F) Dha olefins .....	107
Figure 3.3. Non-natural enzymatic modification of a cyclic thiopeptide .....	110
Figure S3.1. HPLC traces from enzymatic cyclopropanations of methyl 2-benzamidoacrylate .....	148

Figure S3.2. HPLC traces from cyclopropanations of ethyl 2-(1-acetamidovinyl)thiazole-4-carboxylate .....	155
Figure S3.3. Chiral GC traces from cyclopropanations of methyl 2-benzamidoacrylate with EDA .....	160
Figure S3.4. TbtJ1 structure determination.....	164
Figure S3.5. X-ray crystal structure of ligand-free TbtJ1 .....	165
Figure S3.6. TbtJ1 alignment with the open and closed forms of PikC .....	166
Figure S3.7. Loop modeling of TbtJ1 based on the closed form of PikC.....	167
Figure S3.8. Comparison of P450 binding pockets for highly active cyclopropanation variants.....	168
Figure S3.9. ClustalOmega alignment of P450 protein sequences.....	169
Figure S3.10. Percent identity matrix for TbtJ1, TbtJ2, and P450 library.....	170
Figure S3.11. LC-MS/MS of linear TbtA I8-1Dha core.....	171
Figure S3.12. HPLC and EIC traces of enzymatic cyclopropanations of linear TbtA I8-1Dha core .....	172
Figure S3.13. LC-MS/MS of cyclopropanated linear TbtA I8-1Dha mono core .....	174
Figure S3.14. LC-MS/MS of linear TbtA core .....	175
Figure S3.15. LC-MS/MS of linear TbtA F5-3Dha core.....	176
Figure S3.16. LC-MS/MS of linear TbtA I8-3Dha core.....	177
Figure S3.17. HPLC and EIC traces of enzymatic cyclopropanation of linear TbtA core .....	178
Figure S3.18. HPLC and EIC traces of enzymatic cyclopropanation of linear TbtA F5-3Dha core .....	181
Figure S3.19. HPLC and EIC traces of enzymatic cyclopropanation of linear TbtA I8-3Dha core .....	184
Figure S3.20. LC-MS/MS of cyclopropanated linear TbtA core.....	187
Figure S3.21. LC-MS/MS of cyclopropanated linear TbtA F5-3Dha core .....	188

Figure S3.22. LC-MS/MS of cyclopropanated linear TbtA I8-3Dha core .....	189
Figure S3.23. HPLC traces for time course experiments with TbtJ2-T247A and I8-3Dha core .....	190
Figure S3.24. HPLC traces for catalyst loading experiments with TbtJ2-T247A and I8-3Dha core .....	191
Figure S3.25. HPLC and EIC traces of cyclopropanations of TbtA I8-3Dha core with traditional catalysts .....	192
Figure S3.26. LC-MS/MS of 15LP-TbtA F5Dha .....	194
Figure S3.27. LC-MS/MS of 15LP-TbtA I8Dha .....	195
Figure S3.28. LC-MS/MS of cyclized TbtA F5Dha .....	196
Figure S3.29. LC-MS/MS of cyclopropanated cyclized TbtA F5Dha .....	197
Figure S3.30. LC-MS traces of cyclic peptide cyclopropanation experiments .....	198

## LIST OF ABBREVIATIONS

°C	Degrees Celsius
Å	Angstrom
ACCA	1-Aminocyclopropane-1-carboxylic acid
ALA	5-Aminolevulinic acid
Ala	Alanine
BM3	Bacillus megaterium-3
CDCl <sub>3</sub>	Deuterated chloroform
CE	Collision energy
CID	Collision induced dissociation
CYP	Cytochrome P450
Cys	Cysteine
d.e.	Diastereomeric excess
d.r.	Diastereomeric ratio
DASP2	Deacon Active Site Profiler 2
Dha	Dehydroalanine
DIEA	N,N-diisopropylethylamine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
e.e.	Enantiomeric excess
EDA	Ethyl diazoacetate
EDT	Ethanedithiol



EtOH	Ethanol
FeTPP	Iron(III) tetraphenylporphyrin
FID	Flame ionization detector
g	Grams
GC	Gas chromatography
h	Hour(s)
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HOBt	Hydroxybenzotriazole
HPLC	High performance liquid chromatography
hr	Hour(s)
Ile	Isoleucine
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
kDA	Kilodaltons
KPi	Potassium phosphate
LCMS	Liquid chromatography mass spectrometry
M	Molar
Mb	Myoglobin
MeOH	Methanol
Min	Minute(s)
Mol	Mol
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate

NHC	N-heterocyclic carbene
nm	Nanometer
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethyleneglycol
pH	Decimal logarithm of hydronium ion concentration
Phe	Phenylalanine
R <sub>f</sub>	Retardation factor
RMSD	Root-mean-square deviation
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sec	Second(s)
Ser	Serine
TB	Terrific broth
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
Thr	Threonine
TIPS	Triisopropylsilane
TTN	Total turnover number
UV	Ultraviolet

## CHAPTER 1: NON-NATURAL CARBENOID AND NITRENOID INSERTION REACTIONS CATALYZED BY HEME PROTEINS

### Introduction

In recent years, the use of biocatalysts to synthesize optically pure organic molecules has garnered considerable interest for use in pharmaceutical and industrial applications. Enzymes are especially attractive because they offer advantages of high selectivity and mild, sustainable reaction conditions. Traditional enzyme engineering efforts have focused on adapting natural biocatalysis. For example, directed evolution has proven to be a valuable tool to enhance the organic solvent tolerance, pH stability, and thermostability of enzymes for use under industrially relevant reaction conditions.<sup>1</sup> In addition, directed evolution as well as rational and computational design are now widely used to expand the scope of native enzymes to include synthetic substrates of medicinal and fine chemical importance. Despite the increasing interest in using enzymes as tools for synthesis, adapting reactions discovered through chemical ingenuity remains challenging, and many synthetic transformations have no biological counterpart. This vacancy in the field has compelled researchers to develop strategies to adapt protein scaffolds for new reactivity. Historically, a number of approaches have been applied to engineer protein pockets to carry out non-native chemical transformations. For example, protein cavities have been engineered to bind non-natural transition states using *in silico* tools such as Rosetta, or

---

<sup>1</sup>Edited with permission from: Gober, J. G.; Brustad, E. M. "Non-natural carbenoid and nitrenoid insertion reactions catalyzed by heme proteins." *Curr. Opin. Chem. Biol.* **2016**, 35, 124-132. Copyright © 2016 Elsevier Ltd.

through immunization with transition state analogs to generate catalytic antibodies.<sup>2,3</sup>

Alternatively, natural active sites may show weak but fortuitous activity for side reactions that share mechanistic features with the enzyme's native chemistry.<sup>4</sup> These promiscuous functions can often be optimized by evolution in both nature and in the lab when enzyme variants are interrogated against reaction conditions where new reactivity may provide a selective advantage.

In addition to promiscuous reactivity driven directly by protein scaffolds, recent efforts have sought to take advantage of alternative reaction landscapes mediated by enzyme cofactors. (Note: The term cofactor is used to broadly encompass protein-bound accessory molecules including cofactors, coenzymes, prosthetic groups, and metal ions). In nature, cofactors serve to augment the reactivity of proteins beyond limitations imposed by the canonical amino acids. In many cases, cofactors have served as inspiration for the design of synthetic small molecule catalysts that mimic or expand on cofactor function. For example, about 60 years ago, in research aimed at elucidating the mechanistic basis of thiamine (vitamin B1) catalysis, Breslow correctly identified that the central thiazole moiety of thiamine forms a stabilized N-heterocyclic carbene (NHC) that is responsible for catalysis.<sup>5</sup> This seminal discovery proved to be a launching point for a new field of chemistry aimed at developing synthetic NHCs for organocatalysis, among other applications. In a complementary fashion, taking inspiration from this resulting synthetic creativity, Müller and coworkers have gone on to show that thiamine-dependent enzymes catalyze a number of non-natural, NHC-dependent reactions including asymmetric benzoin condensation and Stetter reactions.<sup>6</sup>

Transition metals offer intriguing targets for expanding protein function due to their diverse range of reactivity and their ability to function on a wide range of substrates. Indeed, since the 1970s, a variety of techniques have been developed to introduce non-natural

metallocofactors into protein cavities to generate novel biocatalysts.<sup>7,8</sup> Remarkably, while approximately one third of all proteins are metalloproteins, little effort has focused on determining whether or not natural protein-bound transition metal centers possess the diverse reactivity observed with similar synthetic transition metal complexes.

Heme proteins have served as sources of inspiration for synthetic chemists due to the diverse range of activities exhibited by these proteins. For example, cytochrome P450s, many of which use a high-valent iron-oxo species to oxidize unactivated C-H bonds, have motivated many research groups to design small molecule catalysts that mimic features of these powerful catalysts. Conversely, biological engineers are now taking lessons learned from synthetic metalloporphyrin chemistry to augment the chemical capabilities of protein-bound heme molecules. Small molecule metalloporphyrins have been shown to catalyze P450-like oxidation reactions, and in addition, mechanistically related carbenoid and nitrenoid transfer reactions using reactive intermediates that bear significant structural and electronic similarity to the native P450-oxenoid (Figure 1.1). Here, I highlight recent work demonstrating that in the presence of synthetic reagents, numerous heme proteins, including P450s and globins, catalyze a broad array of abiological carbenoid and nitrenoid transfer transformations, opening new avenues for the construction of complex carbon-carbon, carbon-nitrogen, and other bonds using chemistry that has not previously been explored by Nature.

## **Carbenoid transfer reactions**

### *Enzymatic olefin cyclopropanation*

Cyclopropanation via carbenoid transfer to olefins is a hallmark reaction catalyzed by synthetic metalloporphyrins and other organometallic complexes. In 2013, Arnold and colleagues reported that the iron porphyrin of heme proteins is able to carry out similar non-

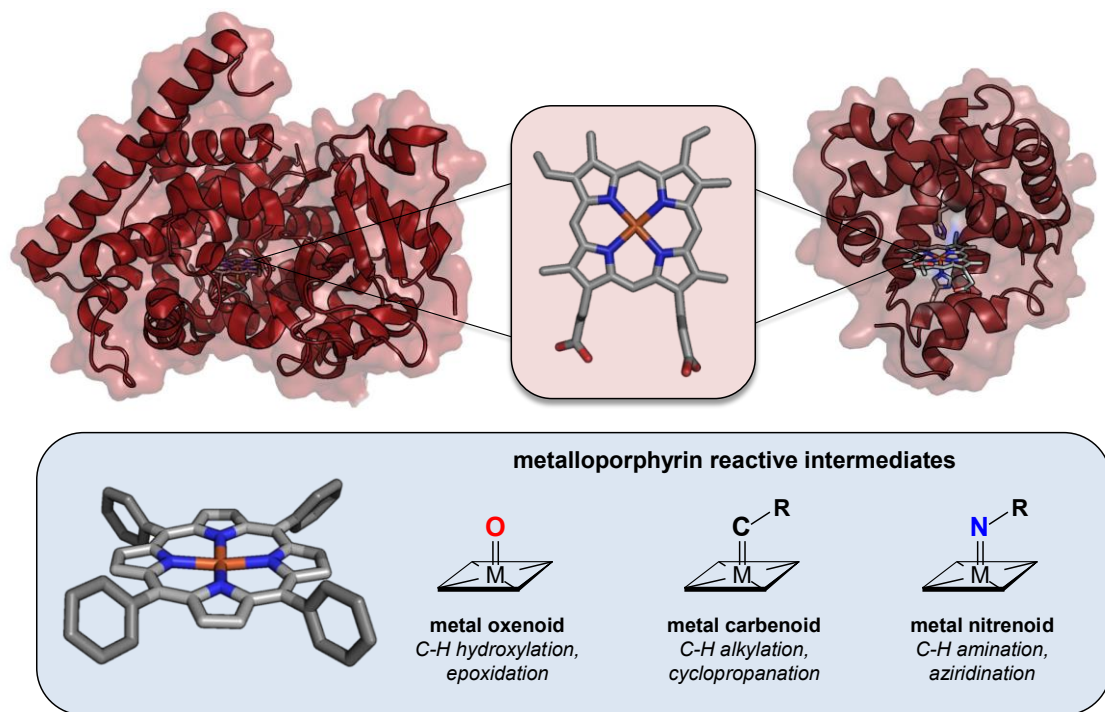


Figure 1.1. Reactive intermediates involved in metalloporphyrin catalyzed transformations. Cytochrome P450<sub>BM3</sub> (PDB 2IJ2, top left) and myoglobin (PDB 1MBN, top right) contain the cofactor iron protoporphyrin IX (heme). Non-natural reactions catalyzed by P450<sub>BM3</sub> and myoglobin were inspired by biomimetic small molecule catalysts such as metal tetraphenylporphyrins that have been shown to catalyze transformations that proceed through carbenoid and nitrenoid intermediates similar in structure to the reactive oxenoid intermediate that drives native cytochrome P450 catalysis.

natural metallocarbenoid chemistry.<sup>9</sup> Diverse heme scaffolds including free hemin, horseradish peroxidase, myoglobin, cytochrome c, and cytochrome P450<sub>BM3</sub> (BM3), all catalyzed the model cyclopropanation of styrene using ethyl diazoacetate (EDA) as a carbene precursor, producing four diastereomeric products (two *cis* and two *trans*, Figure 1.2A and Table 1.1, Entry 1) albeit with low yield. Intriguingly, BM3, a cytochrome P450 (P450) that has proven robust for enzyme engineering applications, showed altered diastereoselectivity (37:63; *cis:trans*) when compared to free hemin (6:94), providing evidence that product distribution can be controlled by the architecture of the heme-binding pocket.

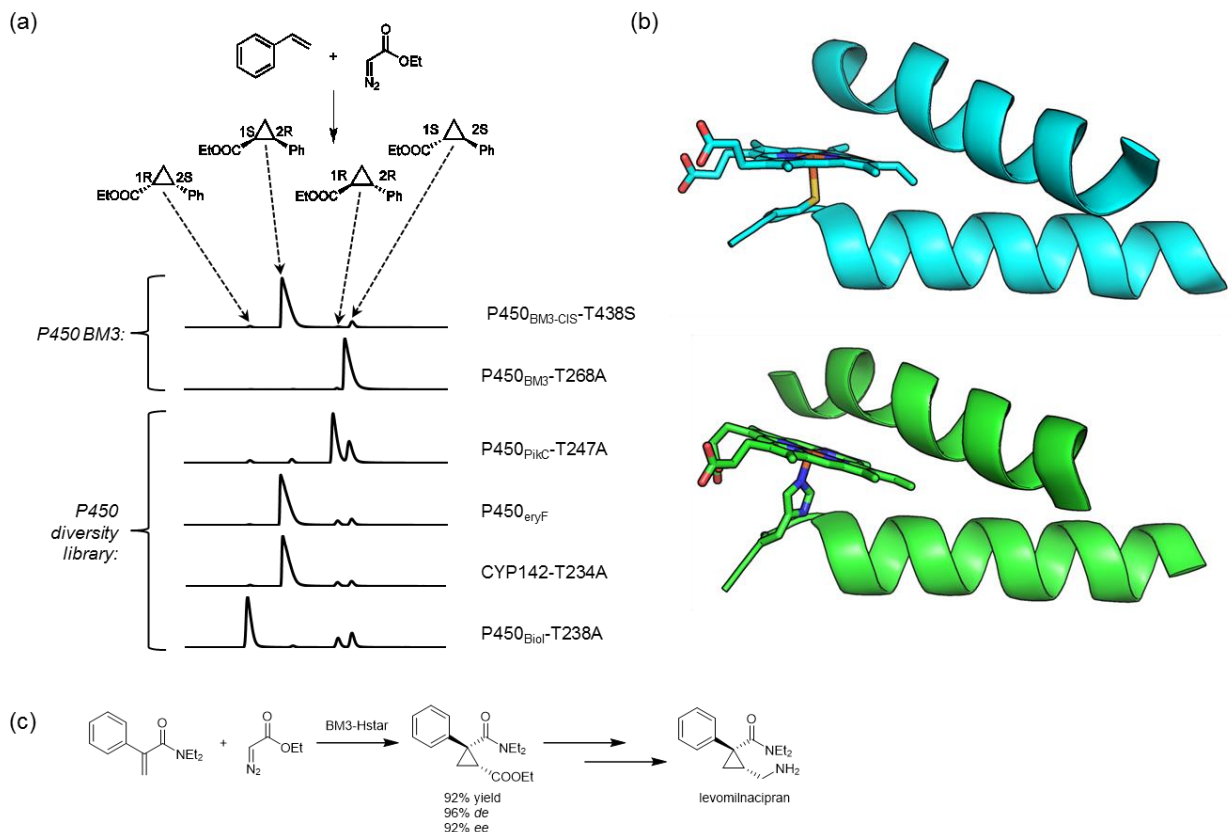
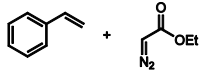

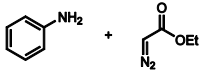
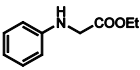
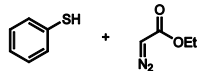
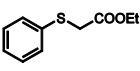
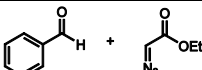
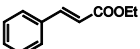
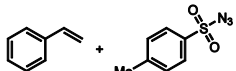
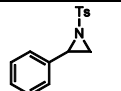
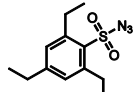
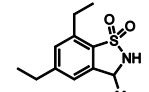
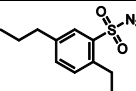
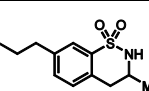
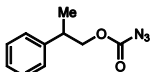
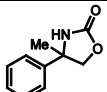
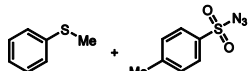
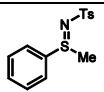


Figure 1.2. Selectivity, structural insight, and application of P450-catalyzed cyclopropanation. **(a)** Chiral gas chromatography traces show the stereoselectivities obtained in the reaction of styrene and ethyl diazoacetate using variants of the model P450<sub>BM3</sub> and variants generated in a diversity-oriented P450 library. **(b)** Structural comparison of wild type CYP119 (cyan, PDB 1F4T) and CYP119-T213A/C317H (green, PDB 5bv5) shows that mutation of the axially-ligating cysteine to histidine causes considerable changes to the orientation of the heme cofactor within the active site. **(c)** An engineered P450<sub>BM3</sub> variant BM3-Hstar allows for the synthesis of a levomilnacipran precursor in high yield and excellent stereoselectivity.

Enzymatic cyclopropanation was improved by screening a panel of 92 pre-engineered BM3 variants that had been isolated over the course of past directed evolution efforts aimed at expanding native P450 biocatalysis. Of the ~ 10 active variants that were identified, most displayed hemin-like product distributions favoring formation of *trans*-cyclopropanes; however, a few variants displayed altered preference for *cis*-cyclopropanes.

Table 1.1. Summary of carbenoid- and nitrenoid-mediated transformations catalyzed by heme-containing enzymes				
Carbenoid-mediated				
Entry	Substrate(s)	Product	Protein(s)	Reference(s)
1			cytochrome P450; myoglobin	9-14, 16
2			cytochrome P450; myoglobin	17-18
3			myoglobin	19
4			myoglobin; YfeX	20-21
Nitrenoid-mediated				
Entry	Substrate(s)	Product	Protein(s)	Reference(s)
5			cytochrome P450	29
6			cytochrome P450; myoglobin	22, 24-25
7			cytochrome P450	23
8			cytochrome P450	26
9			cytochrome P450	27-28

Sequence analysis of highly active BM3 variants revealed that, in addition to other mutations, each contained the same Ala mutation in place of a highly conserved active site Thr (T268A). Remarkably, grafting the T268A mutation alone into wild type BM3 transformed the enzyme into a highly active and *trans*-selective cyclopropanation catalyst (Figure 1.2A, 323 total turnover (TTN), 1:99 *dr*, 96 % *ee*). Mutations were also capable of tuning product distribution. Site saturation mutagenesis performed individually on active site residues of a moderately *cis*-selective variant, BM3-CIS (13 mutations from wild type; 71:29 *dr* and 94% *ee*), led to the isolation of a variant with a single point mutation (BM3-CIS-T438S) that enriched production



(Figure 1.2A, 293 TTN, 92:8 *dr*, 97% *ee*) of a *cis*-cyclopropane diastereomer that is challenging to make using traditional synthetic cyclopropanation catalysts.

Since this initial report, metallocarbenoid-mediated cyclopropanation has been shown to be a general feature of the P450 family of enzymes.<sup>10,11</sup> Brustad and coworkers recently explored the generality of enzyme-catalyzed cyclopropanation by tapping into natural P450 diversity (this work is described in more detail in Chapter 2).<sup>11</sup> A library of over one dozen wild type P450s (24% average sequence identity to BM3) and their corresponding T268A mutation (found to be activating in BM3) was constructed and screened in the reaction of styrene with EDA. All of the wild type P450s were active, and over half of the enzymes from the library experienced an increase in activity and/or selectivity in the presence of the Ala mutation. Notably, from this small, diversity-oriented library, individual protein scaffolds were identified that provided selective enrichment for each of the four possible cyclopropane stereoisomers (Figure 1.2A).

Under biological conditions, carbenoid insertion is outcompeted by native P450-mediated monooxygenation (providing epoxides from olefins). Similar to monooxygenation, carbene precursor activation requires initial single-electron reduction of the resting state Fe<sup>3+</sup>-heme. During native P450 catalysis, substrate binding leads to an increase in the heme iron reduction potential, facilitating electron transfer from biological reductants such as NAD(P)H. Non-natural substrates such as styrene, however, may not efficiently induce the conformational changes necessary to drive this process. Accordingly, an anaerobic environment and strong reductants, such as sodium dithionite, are required to provide useful levels of *in vitro* enzymatic cyclopropanation. Several groups have shown that the activity of P450s for non-native chemistry can be significantly improved by tuning the inner coordination sphere of the iron center. Arnold and Brustad demonstrated that mutation of the conserved axial Cys ligand of P450s to Ser

provided a new variant (denoted cytochrome P411 due to a distinctive shift in the carbon monoxide-bound heme Soret band from 450 nm to 411 nm) that was incapable of carrying out monooxygenation chemistry and fortuitously showed enhanced activity towards enzymatic carbenoid-insertion reactions.<sup>12</sup> The mutation increased the resting state heme reduction potential from  $\sim -420$  mV to  $\sim -290$  mV (in the wild type BM3 backbone, all potentials versus standard hydrogen electrode) resulting in thermodynamically favorable P411 reduction by NAD(P)H ( $-320$  mV). Enhanced activity by P411s enabled whole-cell cyclopropanation reactions for the preparative-scale synthesis of phenyl cyclopropanes (27 g/L; 48,800 TTN), and lyophilized cells performed well without major loss of stereoselectivity, providing an operationally facile biocatalytic platform for cyclopropane synthesis.

Wang et al. have gone on to show that a number of altered proximal ligation states are accommodated within expressed P450s.<sup>13</sup> Notably, a Cys-to-His mutation in BM3-T268A greatly increased cyclopropanation activity and expanded substrate scope to include an electron-deficient olefin that can be used as an intermediate in the synthesis of the antidepressant levomilnacipran (Figure 1.2B and C). Further directed evolution via iterative site saturation mutagenesis led to the isolation of a new catalyst, BM3-Hstar, which catalyzed the reaction in 92% yield, 96% *de*, and 92% *ee*. BM3-Hstar exhibits an impressive rate of cyclopropanation, achieving 80% yield in 10 mins, and remains active and stereoselective in the presence of oxygen. BM3-Hstar was used to prepare, in whole cells, a small library of levomilnacipran derivative precursors in high yields and moderate to very good selectivities.<sup>14</sup> Analysis of an X-ray crystal structure of a thermostable P450, CYP119, bearing the Cys-to-His mutation showed that substantial structural rearrangements were required to accommodate the altered heme-

ligation state (Figure 1.2B).<sup>15</sup> Accordingly, these destabilizing axial mutations may not be accessible in all P450 scaffolds.

Expanding beyond cytochrome P450s, Fasan and coworkers have shown that myoglobin (Mb), which natively contains a His-ligated heme, can also be engineered to produce highly functional cyclopropanation catalysts.<sup>16</sup> Using a limited diversity active site library in which residues in the distal pocket of Mb were mutated to Ala, Val, Phe, or Trp, beneficial mutations were identified that when combined produced a remarkable variant (Mb-H64V/V68A) that catalyzed the cyclopropanation of styrene with EDA in >99% *de*, >99% *ee*, and > 10,000 TTN, indicating that protein-based, carbenoid-insertion catalysts can achieve synthetically relevant activities and selectivities. The engineered Mb variant maintained high diastereo- and enantioselectivity on a variety of substituted styrenes, and mechanistic investigations performed on electronically tuned styrenes provided evidence that an electrophilic heme-centered carbenoid is generated during the catalytic cycle.

#### *Enzymatic X-H carbenoid insertion reactions*

In addition to carbenoid-mediated olefin cyclopropanation, iron porphyrins are also able to catalyze carbenoid insertion into O-H, N-H, and S-H bonds. By screening a small panel of BM3 cyclopropanation catalysts, Wang et al. identified one variant, H2-5-F10, that was able to catalyze the N-H insertion reaction of EDA and aniline in 47% yield (Table 1.1, Entry 2).<sup>17</sup> Using H2-5-F10 as a catalyst, electron-rich anilines gave moderate to high yields, but alkyl amines were poor substrates for the reaction. Substitutions to the para- position of aniline were well tolerated; however ortho- substitutions and increased steric bulk with respect to the carbene precursor resulted in decreased yields.

Fasan and coworkers have demonstrated that myoglobin variants can also catalyze N-H and S-H insertion reactions (Table 1.1, Entries 2 and 3).<sup>18,19</sup> For example, Mb-H64V/V68A, previously engineered to be a highly active and selective cyclopropanation catalyst, was also a very robust catalyst for carbenoid insertion into anilines. Substitutions at the *ortho*-, *meta*-, and *para*- positions were well tolerated, as were both electron withdrawing and electron donating groups. In the reaction of aniline with ethyl diazoacetate, Mb-H64V/V68A catalyzed formation of the single N-H insertion product in > 99% yield and ~3000 TTN at high substrate loading. Mb variants were also screened in the S-H insertion reaction of thiophenol with ethyl diazoacetate (Table 1.1, Entry 3).<sup>19</sup> Mb-L29A and Mb-L29A/H64V showed a greater than two-fold increase in activity for the model reaction compared to wild type Mb. Thiophenols containing electron donating and electron withdrawing groups at the *ortho*-, *meta*-, and *para*- positions as well as benzyl, alkyl, and cycloalkyl mercaptans worked well in the reaction. Yields of 60% to > 99% could be achieved on select substrates using low catalyst loading (0.2 mol%). The reaction of ally(phenyl)sulfane and ethyl diazoacetate was used to probe the mechanism of S-H insertion. A [2,3]-sigmatropic rearrangement was observed, which is consistent with a stepwise mechanism that proceeds via a sulfonium ylide intermediate (Figure 1.3A). The observed enantioselectivity suggests that the sulfonium ylide is associated with the heme or the active site prior to protonation.

#### *Enzymatic olefination of carbonyl compounds*

Phosphonium ylides are known to react with aldehydes to generate *E* or *Z* olefins. Fasan and associates have shown that both hemin and Mb variants are active olefination catalysts in the reaction of benzaldehyde with ethyl diazoacetate in the presence of either triphenylphosphine or triphenylarsine oxophiles (Table 1.1, Entry 4).<sup>20</sup> The choice of triphenylphosphine and its

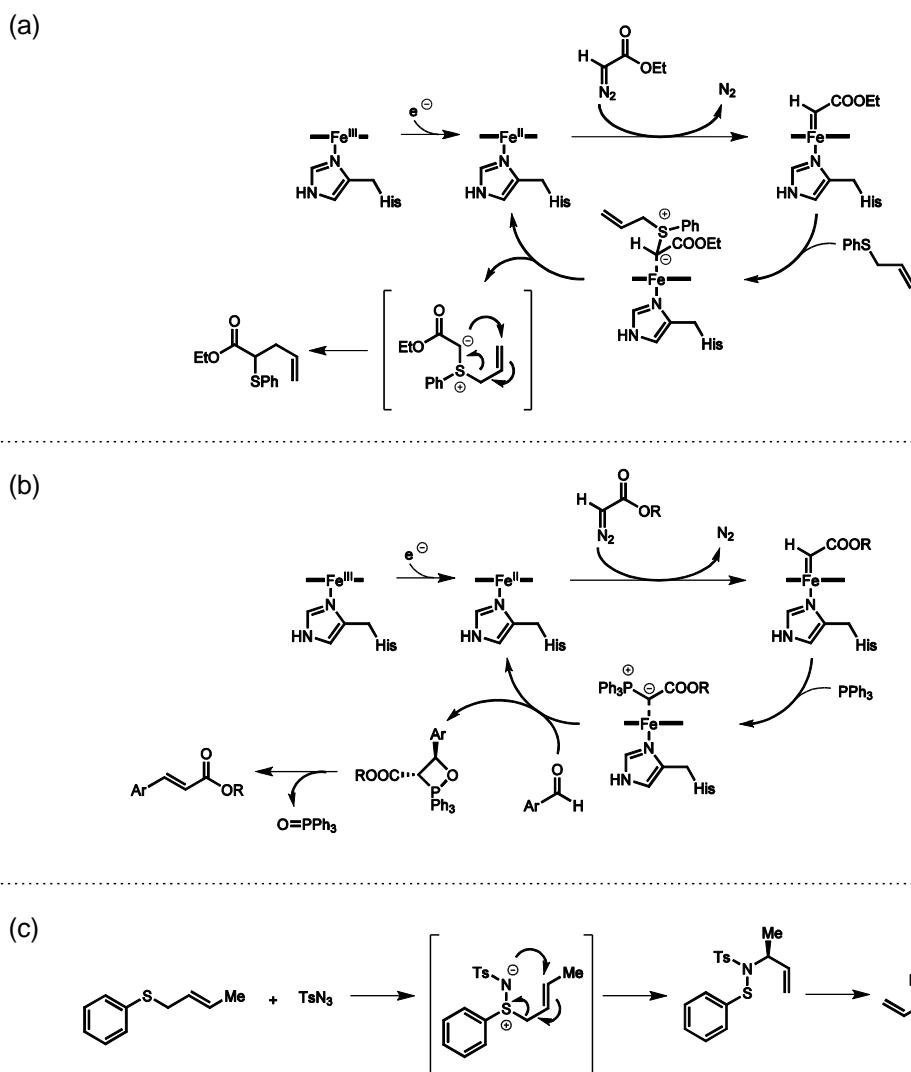


Figure 1.3. Proposed mechanisms for some carbenoid- and nitrenoid-mediated transformations catalyzed by heme-containing enzymes. **(a)** The S-H insertion reaction of ally(phenyl)sulfane and ethyl diazoacetate catalyzed by myoglobin variants is proposed to involve a sulfonium ylide intermediate that undergoes a [2,3]-sigmatropic rearrangement to furnish the final product. **(b)** The olefination of aryl aldehydes catalyzed by myoglobin variants is proposed to involve a phosphonium ylide that reacts in a Wittig-like process to yield  $\alpha,\beta$ -unsaturated esters. **(c)** Allylic amination catalyzed by engineered P450<sub>BM3</sub> variants is proposed to proceed through initial sulfimination of an allylic sulfide using p-toluenesulfonyl azide followed by a [2,3]-sigmatropic rearrangement and reductive cleavage to give the chiral protected allylic amine.

congeners in addition to mutations within the protein scaffold had an impact on conversion and diastereoselectivity. Aromatic aldehydes with electronic and steric substitutions could be used in the reaction without compromising selectivity; however, substitutions to the  $\alpha$ -position of the carbene precursor were not tolerated. Similar to sulfonium ylide formation in S-H insertion reactions, the proposed mechanism for aldehyde olefination involves attack of an electrophilic iron carbenoid by triphenylphosphine to generate a phosphonium ylide (Figure 1.3B). The ylide is then proposed to react with an aldehyde to form an oxaphosphetane intermediate that rearranges to yield the olefin and triphenylphosphine oxide. The selectivity-determining step likely occurs during formation of the oxaphosphetane. The phosphine oxide byproduct however, was found to be inhibitory, limiting total conversion to < 50%.

Hauer and coworkers have gone on to show that some heme-containing enzymes also catalyze carbonyl olefination in the absence of an oxophile (Table 1.1, Entry 4).<sup>21</sup> From a screen of concentrated and fractionated *E. coli* lysate, four heme-containing proteins were found to catalyze the olefination of benzaldehyde with ethyl diazoacetate in the absence of triphenylphosphine. YfeX, an *E. coli* protein of unknown function, was found to be the most active of the screened enzymes. Overexpression of YfeX in *E. coli* resulted in a > 2 fold increase in product formation in whole cells (440 mgL<sup>-1</sup>hr<sup>-1</sup>). Distinct from Mb-catalyzed olefination, in the absence of an oxophile, a nucleophilic carbene has been hypothesized to attack the aldehyde directly, forming an iron oxetane intermediate which rearranges to form the alkene and the native P450 monooxygenation intermediate, Compound I. The proposed formation of Compound I is supported by the observation of non-productive alkene epoxidation byproducts.

## Nitrenoid transfer reactions

### *Enzymatic intramolecular C-H amination*

Shortly after reporting P450-mediated cyclopropanation, Arnold and coworkers went on to show that many of the same P450-scaffolds also catalyzed metallonitrenoid insertions, including the intramolecular C-H amination of benzylic carbons.<sup>22</sup> In this work, a panel of Cys- and Ser-ligated BM3 variants was screened for C-H amination of 2,4,6-triethylbenzene-1-sulfonyl azide (Table 1.1, Entry 6). While the wild type enzyme was poorly active, mutations previously identified as activating for P450-mediated cyclopropanation (T268A and C400S) substantially enhanced enzymatic C-H amination. Substoichiometric quantities of NADPH or sodium dithionite as reductants supported activity, providing evidence that the active ferrous-heme is regenerated during the catalytic cycle. Activity was inversely correlated with C-H bond strength when free hemin was used as a catalyst, yielding a racemic mixture of the benzosultam products. Intriguingly, highly active P411-based catalysts did not show the same dependence on bond strength and catalyzed the reactions with considerable asymmetric induction. These data suggest that the enzyme active site influences both stereoselectivity and reactivity in intramolecular C-H aminations. In an exciting follow up, Hyster et al. have shown that P450 active sites can also tune the regioselectivity of C-H aminations.<sup>23</sup> Using 2,5-diisopropylbenzene-1-sulfonyl azide as a model substrate, P411 screening and subsequent active site directed evolution led to the isolation of variants enriched for C-H amination at the homobenzylic carbon (Table 1.1, Entry 7), providing evidence that enzymatic nitrenoid insertion reactions can also target less activated C-H bonds.

Intramolecular C-H amination via Cys-ligated BM3 variants has been further confirmed by Fasan and colleagues, who examined a series of benzenesulfonyl azide substrates in order to

better understand the requirements for C-H activation.<sup>24</sup> Altering electronics on the benzene ring did not appreciably affect catalysis; however, increased steric bulk was shown to be an important factor to achieve high activity. It has been suggested that larger substituents may help orient the C-H bond in close proximity to the metallonitrenoid intermediate in a manner that favors C-H insertion. The Fasan lab also showed that a number of heme proteins, including catalase, horseradish peroxidase, myoglobin, and hemoglobin, all catalyze intramolecular benzylic C-H amination using benzenesulfonyl azide substrates.<sup>25</sup> In addition, these authors have demonstrated that alternative organoazide nitrene precursors, including carbonazidates, also function in P450-mediated C-H amination reactions, providing an enzymatic source to synthetic precursors for valuable 1,2-amino alcohols (Table 1.1, Entry 8).<sup>26</sup>

One limiting feature in P450-mediated nitrenoid insertions is the non-productive over-reduction of the sulfonyl azide precursor to yield the corresponding sulfonamide. Through their investigations, Singh et al. observed that the T268A mutation in BM3 (generally activating in carbenoid and nitrenoid insertions), also substantially reduced formation of this byproduct.<sup>24</sup> In native P450 catalysis, this conserved active site threonine plays a pivotal role in proton transfer events that are important for monooxygenation. Mutation to Ala likely favors productive C-H amination by decreasing the rate of off-path, non-productive protonation of the metallonitrenoid intermediate, thereby improving the efficiency of the desired nitrenoid insertion.

#### *Enzymatic intermolecular sulfimidation*

It is well known that the active oxenoid intermediate (Compound I) responsible for P450-mediated monooxygenations, can also react with aryl sulfides to generate sulfoxides. Analogously, Farwell et al. have recently shown that P450-centered metallonitrenoids can mimic this process to generate sulfimides in an asymmetric fashion.<sup>27</sup> A number of BM3 variants were



shown to be active in the reaction of tosyl azide (as a nitrene precursor) with thioanisole (Table 1.1, Entry 9). While Cys-ligated enzymes were poorly active, substantial product formation (>300 TTN) was observed for Ser-substituted P411<sub>BM3</sub> variants. Importantly, similar to other carbenoid and nitrenoid insertion chemistries described here, the composition of the heme-binding pocket plays an important role in substrate orientation as selective scaffolds were identified for the enriched production of either sulfimide enantiomer.

To date, substrate scope in enzymatic sulfimidation reactions remains quite limited. Hammett analysis comparing rates of enzymatic sulfimidation on electronically tuned thioanisoles produced a linear free energy relationship indicative of a buildup of partial positive charge on the sulfide during the rate-limiting step.<sup>27</sup> These data suggest that the electronics of the aryl sulfide also play a key role in determining the overall yields of sulfimidation products. In addition, while a number of sulfonyl azide nitrene precursors of varied size were examined in this reaction, only tosyl azide provided productive sulfimidation. Limitations in nitrene precursor scope in this and other enzymatic nitrenoid insertion reactions remains poorly understood.

In a clever follow up to this initial work, Prier et al. have adapted enzymatic sulfimidation for the selective synthesis of chiral amines.<sup>28</sup> By using a substrate walking approach in combination with directed evolution, the authors were able to isolate a P411 variant that demonstrated sulfimidation on phenyl crotyl sulfide (Figure 1.3C). Allylic sulfimides are known to undergo stereospecific [2,3]-sigmatropic rearrangements to yield protected allylic amines. Whole-cell sulfimidation of phenyl crotyl sulfide in the presence of tosyl azide using an optimized P411 scaffold led to the formation of the corresponding allylic amine in 77% yield, 2200 TTN, and 68% *ee* upon reductive workup. Enantioselectivity for the resulting allylic amines, while moderate, were attenuated in comparison to stereoselectivities achieved during

sulfimination of alkyl sulfides. The difference in observed selectivities using alkyl sulfides and allylic sulfides of comparable size suggests that enantiopurity may erode during the sigmatropic rearrangement.

#### *Enzymatic intermolecular aziridination*

Nitrenoid-mediated aziridination reactions of olefins bear mechanistic similarity to both carbenoid-mediated cyclopropanation and oxenoid-mediated epoxidation reactions. In nature, aziridine biosynthesis is accomplished through nucleophilic substitution reactions. Accordingly, enzymatic aziridination via nitrene transfer would provide a novel strategy for synthesizing these synthetically relevant functionalities in a manner not previously observed in biology. Based on their earlier success, Arnold and coworkers screened a variety of P411 enzymes that previously demonstrated C-H amination and/or sulfimination activity for the nitrenoid-mediated aziridination of substituted styrenes.<sup>29</sup> Using tosyl azide as a nitrene precursor, one variant, denoted P-I263F, demonstrated considerable activity (>150 TTN) on a number of *p*-substituted styrenes. Electron-rich styrenes provided ring-opened products due to favorable hydrolysis of the aziridine product, but less electron-rich substrates including styrene and *p*-methylstyrene provided the desired aziridine product in moderate yield. Initial yields were limited by non-productive reduction of the metallonitrenoid intermediate yielding sulfonamide byproducts; however, aziridination activity and selectivity were further improved via directed evolution using site saturation mutagenesis targeted to the substrate-binding pocket. One variant, P-I263F-A328V-L437V, showed excellent enantioselectivity (99% *ee*) on *p*-methylstyrene and improved yields due to reduced formation of the sulfonamide byproduct. Impressively, good to excellent selectivity was maintained on a variety of substituted styrenes, although complete erosion of

stereochemistry was observed in reactions with  $\alpha$ -methyl styrene and *p*-methoxystyrene due to ring opening of the aziridine with water.

## Conclusion

Engineering enzyme scaffolds that carry out non-natural chemical transformations has been a longstanding goal of protein chemists in an effort to combine lessons learned from synthetic chemistry with the selectivity and evolvability imparted by genetically encoded biocatalysts. Since the initial report of protein-mediated carbenoid insertion in 2013, rapid progress has been made with regards to increasing the synthetic repertoire of heme proteins to include a variety of metallocarbenoid and nitrenoid transfer reactions. Importantly, researchers have shown that the heme cofactor provides reactivity that can be tuned by the protein scaffold using well-established protein engineering and directed evolution techniques. It is reasonable to suspect that, under the right conditions, other synthetic transformations that exploit these reactive intermediates as well as alternative chemistries driven by synthetic iron porphyrins, may similarly be accessible using heme proteins as catalysts. For example, expanding heme protein catalysis to include intermolecular carbenoid and nitrenoid insertion into unactivated C-H bonds – advancing on the early intramolecular C-H amination work highlighted here – would mark a significant achievement and would accelerate the late-stage diversification of small molecules. The outlook for the continued expansion of synthetically useful metallo- and organocofactor-catalyzed reactions is promising given the wealth of transformations mediated by synthetic transition metal complexes and organocatalysts. Using chemistry as a guide, the identification of promiscuous reactivity in heme and other natural, protein-bound catalysts may provide a rapid means of expanding the reactive landscape available to protein engineers.

## REFERENCES

1. Jemli, S.; Ayadi-Zouari, D.; Hlima, H. Ben; Bejar, S. *Crit. Rev. Biotechnol.* **2016**, 36 (2), 246.
2. Hilvert, D. *Annu. Rev. Biochem.* **2013**, 82 (1), 447.
3. Wentworth, P.; Janda, K. D. *Cell Biochem. Biophys.* **2001**, 35 (1), 63.
4. Khersonsky, O.; Tawfik, D. S. *Annu. Rev. Biochem.* **2010**, 79 (1), 471.
5. Breslow, R. *J. Am. Chem. Soc.* **1958**, 80 (14), 3719.
6. Müller, M.; Gocke, D.; Pohl, M. *FEBS J.* **2009**, 276 (11), 2894.
7. Heinisch, T.; Ward, T. R. *Curr. Opin. Chem. Biol.* **2010**, 14 (2), 184.
8. Lewis, J. C. *ACS Catal.* **2013**, 3 (12), 2954.
9. Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. *Science (80-. )*. **2013**, 339 (6117), 307.
10. Heel, T.; McIntosh, J. A.; Dodani, S. C.; Meyerowitz, J. T.; Arnold, F. H. *ChemBioChem* **2014**, 15 (17), 2556.
11. Gober, J. G.; Rydeen, A. E.; Gibson-O'Grady, E. J.; Leuthaeuser, J. B.; Fetrow, J. S.; Brustad, E. M. *ChemBioChem* **2016**, 17 (5), 394.
12. Coelho, P. S.; Wang, Z. J.; Ener, M. E.; Baril, S. A.; Kannan, A.; Arnold, F. H.; Brustad, E. M. *Nat. Chem. Biol.* **2013**, 9 (8), 485.
13. Wang, Z. J.; Renata, H.; Peck, N. E.; Farwell, C. C.; Coelho, P. S.; Arnold, F. H. *Angew Chem Int Ed* **2014**, 126 (26), 6928.
14. Renata, H.; Wang, Z. J.; Kitto, R. Z.; Arnold, F. H. *Catal Sci Technol* **2014**, 4 (10), 3640.
15. McIntosh, J. A.; Heel, T.; Buller, A. R.; Chio, L.; Arnold, F. H. *J Am Chem Soc* **2015**, 137 (43), 13861.
16. Bordeaux, M.; Tyagi, V.; Fasan, R. *Angew Chem Int Ed* **2015**, 54 (6), 1744.
17. Wang, Z. J.; Peck, N. E.; Renata, H.; Arnold, F. H. *Chem. Sci.* **2014**, 5 (2), 598.
18. Sreenilayam, G.; Fasan, R. *Chem Commun* **2015**, 51 (8), 1532.
19. Tyagi, V.; Bonn, R. B.; Fasan, R. *Chem. Sci.* **2015**, 6 (4), 2488.

20. Tyagi, V.; Fasan, R. *Angew Chem Int Ed* **2016**, 55 (7), 2512.
21. Weissenborn, M. J.; Löw, S. A.; Borlinghaus, N.; Kuhn, M.; Kummer, S.; Rami, F.; Plietker, B.; Hauer, B. *ChemCatChem* **2016**, 8, 1636.
22. McIntosh, J. A.; Coelho, P. S.; Farwell, C. C.; Wang, Z. J.; Lewis, J. C.; Brown, T. R.; Arnold, F. H. *Angew. Chemie Int. Ed.* **2013**, 52 (35), 9309.
23. Hyster, T. K.; Farwell, C. C.; Buller, A. R.; McIntosh, J. A.; Arnold, F. H. *J. Am. Chem. Soc.* **2014**, 136 (44), 15505.
24. Singh, R.; Bordeaux, M.; Fasan, R. *ACS Catal.* **2014**, 4 (2), 546.
25. Bordeaux, M.; Singh, R.; Fasan, R. *Bioorg Med Chem* **2014**, 22 (20), 5697.
26. Singh, R.; Kolev, J. N.; Sutura, P. A.; Fasan, R. *ACS Catal* **2015**, 5 (3), 1685.
27. Farwell, C. C.; McIntosh, J. A.; Hyster, T. K.; Wang, Z. J.; Arnold, F. H. *J Am Chem Soc* **2014**, 136 (24), 8766.
28. Prier, C. K.; Hyster, T. K.; Farwell, C. C.; Huang, A.; Arnold, F. H. *Angew Chem Int Ed* **2016**, 55 (15), 4711.
29. Farwell, C. C.; Zhang, R. K.; McIntosh, J. A.; Hyster, T. K.; Arnold, F. H. *ACS Cent Sci* **2015**, 1 (2), 89.

## CHAPTER 2: MUTATING A HIGHLY CONSERVED RESIDUE IN DIVERSE CYTOCHROME P450S FACILITATES DIASTEREOSELECTIVE OLEFIN CYCLOPROPANATION<sup>2</sup>

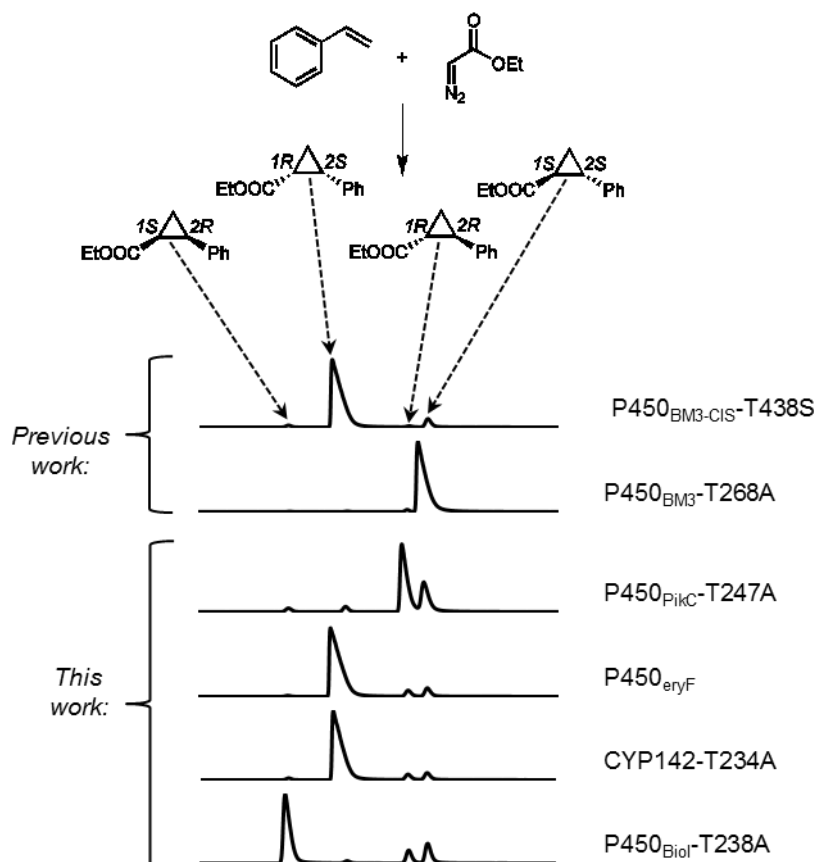
### Introduction

One approach to uncovering new modes of enzyme catalysis is to use man-made reagents that provide access to reactive intermediates not typically found in nature.<sup>1</sup> For example, intermolecular metal-catalyzed cyclopropanation is a well-characterized reaction that allows functionalization of olefins with a variety of synthetic carbene precursors. Cyclopropanes are valuable synthetic targets due to their presence in natural products and pharmaceuticals as well as their use as synthetic intermediates that undergo stereoselective ring-opening transformations.<sup>2-5</sup> A variety of transition metal complexes, including metalloporphyrins similar to the native prosthetic group heme, have been applied for this transformation, but the design of competent catalysts that demonstrate high diastereo- and enantioselectivity has remained challenging.<sup>6-14</sup>

Recent work by Arnold and Fasan has shown that heme-containing proteins, including members of the cytochrome P450 family of enzymes (P450s) and myoglobin, promote the promiscuous cyclopropanation of styrenes in the presence of diazoacetates.<sup>15-17</sup> In contrast to free heme, which produces a racemic mixture of predominantly *trans*-cyclopropanes with low total

---

<sup>2</sup>Edited with permission from: Gober, J. G.; Rydeen, A. E.; Gibson-O'Grady, E. J.; Leuthaeuser, J. B.; Fetrow, J. S.; Brustad, E. M. "Mutating a Highly Conserved Residue in Diverse Cytochrome P450s Facilitates Diastereoselective Olefin Cyclopropanation." *ChemBioChem* **2016**, 17 (5), 394-397. Copyright © 2016 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.



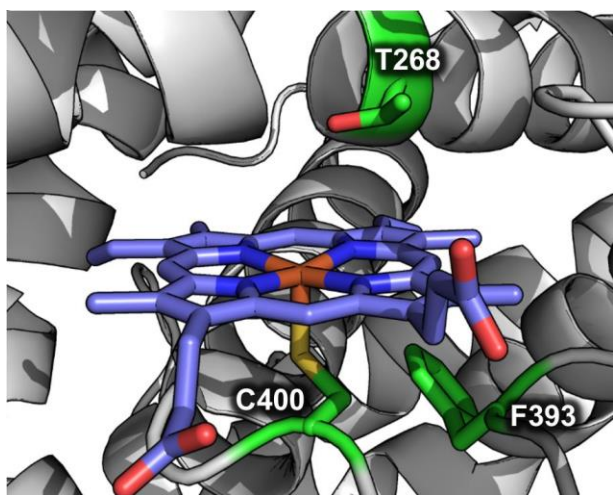
**Figure 2.1.** Styrene cyclopropanation with ethyl diazoacetate catalyzed by P450 variants. Chiral GC traces are aligned to provide a visual representation of product distributions.

turnover number (TTN), several native heme proteins exhibit weak to moderate stereoselection and modest catalytic efficiency.

Engineering efforts on cyclopropanation catalysts derived from myoglobin and the bacterial P450<sub>BM3</sub> have resulted in increased activity and have led to the facile isolation of highly *trans*-selective enzymes. For example, an engineered myoglobin with two mutations from wild type showed near-perfect selectivity in the cyclopropanation of styrene with ethyl diazoacetate (EDA) producing the 1*S*,2*S* isomer with >99% conversion.<sup>15</sup> Mutations have also been shown to affect activity. While wild type P450<sub>BM3</sub> is a weak cyclopropanation catalyst (< 5 TTN), a single active site mutation, T268A, improved TTN ~ 65 fold and produced a highly selective variant for the 1*S*,2*S* isomer (99:1 *dr*, 97% *ee*; Figures 2.1 and 2.2 and Table 2.1).<sup>16</sup> Together these studies

demonstrate that the architecture of heme-binding pockets can be leveraged to improve selectivity in intermolecular cyclopropanation reactions.

Despite these efforts, engineering biocatalysts selective for thermodynamically unfavorable *cis*-diastereomers remains a challenge. *Cis*-selective catalysts for the cyclopropanation of styrene with EDA have been identified by screening a library of P450<sub>BM3</sub> variants possessing diverse active sites (see P450<sub>BM3</sub>-CIS-T438S, Figure 2.1 and Table 2.1);<sup>16,18</sup> however, these catalysts contained a large number of mutations (> 10), most of which were acquired over a decade of directed evolution. Identifying similar trajectories in other, non-*cis*-selective scaffolds would require considerable engineering and screening efforts. In addition, minor substitutions on styrene lead to attenuated or even reversed selectivity (*vide infra*).<sup>16,18</sup> The difficulty in engineering *cis*-selective enzymes illustrates the challenge of using a single protein scaffold to alter both diastereo- and enantioselectivity.



**Figure 2.2.** Active site of substrate-free P450<sub>BM3</sub> (PDB: 2IJ2). Conserved residues T268, F393, and C400 are colored green, and the heme cofactor is colored purple.

We hypothesized that we could access stereoselective cyclopropanation catalysts by sampling a diverse library of natural P450 active sites and by introducing strategic mutations informed by previous P450<sub>BM3</sub>

engineering efforts. In P450<sub>BM3</sub>, a single mutation, T268A, significantly improves activity and selectivity, and the majority of highly active P450<sub>BM3</sub> variants contain this mutation (Figures 2.1 and 2.2 and Table 2.1).<sup>16</sup> Because this threonine is highly conserved, we reasoned that this mutation might affect cyclopropanation activity and selectivity in other scaffolds. Here we show



that a small library composed of thirteen diverse P450s and their corresponding alanine variants allows for the rapid identification of selective biocatalysts for all four diastereomers of the cyclopropanation reaction.

## Results

Given the ubiquity of P450s in nature, numerous gene sequences are known that encode scaffolds that catalyze a range of oxidative and reductive transformations on a wide variety of substrates. The vast number of P450s that have been characterized or hypothesized based on genome mining makes the selection of viable sequences a non-trivial endeavor. To guide our search, we used the bioinformatics tool Deacon Active Site Profiler 2 (DASP2), which allowed us to maximize diversity by identifying scaffolds that share motifs necessary for function (*e.g.* the heme ligating cysteine, Figure 2.2) among sequences that share little overall sequence identity (See Supporting Information).<sup>19-21</sup> From the DASP2 search we identified hundreds of unique P450 sequences as potential candidates for library design. The library was further narrowed down to enzymes that fit at least two of the following criteria: 1) P450s derived from bacterial origin or previously produced in *Escherichia coli* (to facilitate heterologous expression), 2) P450s of known structure (to aid future protein engineering), and 3) sequences that shared <20% sequence identity to P450<sub>BM3</sub> (to maximize active site structural variation). Our final library comprised sixteen P450s, including P450<sub>BM3</sub>, with diverse activities and substrate scopes (Tables S2.1 and S2.2 and Supporting Information text). The average sequence identity across the entire library is 24%, and the maximum identity is 46% (P450<sub>PikC</sub> and P450<sub>eryF</sub>; Figures S2.1 and S2.2). Each enzyme contains the conserved active site threonine, with the exceptions of P450<sub>eryF</sub> and CYP122A2, which contain an alanine and a serine respectively at this position (Figures S2.1 and S2.3). The genes of the wild type P450s and their corresponding

alanine variants were synthesized by Gen9 Inc. Of the thirty-one wild type and Thr→Ala constructs under investigation, 25 proteins were successfully purified in sufficient quantity for screening (Figure S2.4 and Table S2.1).

**Table 2.1.** Activities and stereoselectivities of P450 variants for the reaction of styrene with ethyl diazoacetate.

catalyst	yield	TTN <sup>a</sup>	dr (cis:trans)	ee cis [%] <sup>b</sup>	ee trans [%] <sup>c</sup>
Hemin	16	79	13:87	-2	-4
P450 <sub>BM3</sub>	1	7	12:88	0	-2
P450 <sub>BM3</sub> -T268A	67	338	1:99	-18	-97
P450 <sub>BM3</sub> -CIS-T438S	62	311	93:7	-97	-79
P450 <sub>Biol</sub>	27	135	12:88	8	13
P450 <sub>Biol</sub> -T238A	48	241	71:29	95	-24
P450 <sub>cam</sub>	41	207	88:12	-43	9
P450 <sub>cam</sub> -T252A	30	151	71:29	-86	-5
P450 <sub>eryF</sub> (A245)	70	349	89:11	-99	-19
CYP142	49	246	44:56	-84	-6
CYP142-T234A	54	272	90:10	-97	-14
CYP164A2	7	34	14:85	5	2
CYP164A2-T260A	70	350	18:82	-82	-9
CYP107N1	7	36	9:91	0	-4
CYP107N1-T251A	48	238	9:91	-3	-3
P450 <sub>nor</sub>	5	27	12:88	-6	-3
P450 <sub>nor</sub> -T243A	13	66	10:90	-3	-2
P450 <sub>EpoK</sub>	49	247	11:89	-15	-22
P450 <sub>EpoK</sub> -T258A	44	219	18:82	-25	-14
P450 <sub>PikC</sub>	50	249	8:92	-2	3
P450 <sub>PikC</sub> -T247A	46	231	7:93	-15	32
P450 <sub>RhF</sub>	52	258	9:91	-3	-2
P450 <sub>RhF</sub> -T275A	34	171	12:88	-11	-2
P450 <sub>TxtE</sub>	37	187	10:90	6	-2
P450 <sub>TxtE</sub> -T250A	45	225	10:90	-2	-2
P450 <sub>TylH1</sub>	48	242	13:87	16	8
P450 <sub>TylH1</sub> -T279A	50	251	10:90	1	-28

<sup>a</sup> TTN = total turnover number. <sup>b</sup> (1S,2R) – (1R,2S). <sup>c</sup> (1R,2R) – (1S,2S). TTNs and stereoselectivities determined by chiral GC analysis.

**Table 2.2.** Substrate scope for most selective variants from the P450 library

$\text{R}_1$ - $\text{C}_6\text{H}_4$ -CH=CH-R<sub>2</sub> +  $\text{CH}_2\text{N}_2\text{COOEt} \xrightarrow[0.1 \text{ M KPi, pH 8.0, argon, RT, 16 hrs}]{20 \text{ } \mu\text{M P450, 5\% MeOH}}$   $\text{R}_1$ - $\text{C}_6\text{H}_4$ -CH<sub>2</sub>-CH(R<sub>2</sub>)-COOEt

**1b** - R<sub>1</sub> = H, R<sub>2</sub> = CH<sub>3</sub>  
**1c** - R<sub>1</sub> = OCH<sub>3</sub>, R<sub>2</sub> = H  
**1d** - R<sub>1</sub> = CF<sub>3</sub>, R<sub>2</sub> = H

**4a** - R<sub>1</sub> = H, R<sub>2</sub> = CH<sub>3</sub>  
**4b** - R<sub>1</sub> = OCH<sub>3</sub>, R<sub>2</sub> = H  
**4c** - R<sub>1</sub> = CF<sub>3</sub>, R<sub>2</sub> = H

	yield	TTN <sup>a</sup>	dr (cis:trans)	ee cis [%] <sup>b</sup>	ee trans [%] <sup>b</sup>
<b>4a</b>					
hemin	16	79	22:78	0	-3
P450 <sub>BM3</sub> -T268A	68	342	1:99	33	93
P450 <sub>PikC</sub> -T247A	50	248	25:75	11	-49
P450 <sub>BM3</sub> -CIS-T438S	19	95	8:92	8	90
P450 <sub>eryF</sub> (A245)	30	149	42:58	-5	-37
P450 <sub>Biol</sub> -T238A	54	271	87:13	-96	8
CYP142-T234A	47	234	89:11	94	0
<b>4b</b>					
hemin	18	90	17:83	0	n.d. <sup>c</sup>
P450 <sub>BM3</sub> -T268A	65	326	1:99	-40	n.d. <sup>c</sup>
P450 <sub>PikC</sub> -T247A	55	276	11:89	-4	n.d. <sup>c</sup>
P450 <sub>BM3</sub> -CIS-T438S	61	305	75:25	-79	n.d. <sup>c</sup>
P450 <sub>eryF</sub> (A245)	41	204	73:27	-93	n.d. <sup>c</sup>
P450 <sub>Biol</sub> -T238A	43	214	66:34	88	n.d. <sup>c</sup>
CYP142-T234A	48	241	91:9	-96	n.d. <sup>c</sup>
<b>4c</b>					
hemin	6	30	13:87	-1	n.d. <sup>c</sup>
P450 <sub>BM3</sub> -T268A	39	195	13:87	-36	n.d. <sup>c</sup>
P450 <sub>PikC</sub> -T247A	41	205	9:91	0	n.d. <sup>c</sup>
P450 <sub>BM3</sub> -CIS-T438S	59	297	35:65	71	n.d. <sup>c</sup>
P450 <sub>eryF</sub> (A245)	17	85	73:27	-81	n.d. <sup>c</sup>
P450 <sub>Biol</sub> -T238A	14	70	30:70	59	n.d. <sup>c</sup>
CYP142-T234A	34	171	90:10	-96	n.d. <sup>c</sup>

<sup>a</sup>TTN = total turnover numbers. <sup>b</sup>stereochemistry unassigned; negative sign indicates opposite enantiomer is formed. <sup>c</sup>not determined; baseline separation could not be achieved. TTNs and stereoselectivities determined by chiral GC analysis.

All of the wild type P450s screened were active catalysts for the cyclopropanation of styrene with EDA, though to varying extents (Table 2.1). The threonine to alanine mutation, which was previously found to strongly activate P450<sub>BM3</sub> variants,<sup>16</sup> induced similar enhancements in three scaffolds (P450<sub>nor</sub>, CYP107N1, and CYP164A2) that showed weak native activity (< 50 TTN) but experienced three to ten-fold increases in TTN upon mutation (Table 2.1). For example, wild type CYP164A2 displayed negligible activity (34 TTN), whereas CYP164A2-T260A produced cyclopropanes in 70% yield (350 TTN).

Nine of the wild type P450s showed considerable activity (> 200 TTN) even in the absence of the mutation, indicating that it is not absolutely required for high activity (Table 2.1). In some

scaffolds, a small but significant decrease in activity was observed upon mutation, with P450<sub>cam</sub>-T252A and P450<sub>RhF</sub>-T274A showing 27% and 33% decreases, respectively. Interestingly, wild type P450<sub>eryF</sub>, which contains an alanine in place of the highly conserved threonine, was the most active wild type P450 and one of the most active enzymes screened (349 TTN).

Most of the variants were *trans*-selective and showed hemin-like product profiles. A subset, however, possessed notable stereoselectivity (Table 2.1). P450<sub>PikC</sub>-T247A and P450<sub>BM3</sub>-T268A showed impressive *trans*-diastereoselectivity (93:7 and 99:1 *dr*, respectively) and moderate to excellent enantioselectivity (32% and 97% *ee*) for the 1*R*,2*R* and 1*S*,2*S* isomers, respectively. The selectivity of P450<sub>BM3</sub>-T268A is consistent with previous reports.<sup>16</sup> In both variants, the threonine to alanine mutation improved enantioselectivity. Interestingly, five P450s in the library were *cis*-selective, including a few that were highly enantioselective for the 1*R*,2*S* and 1*S*,2*R* isomers. Wild type P450<sub>cam</sub> produced the 1*R*,2*S* enantiomer (88:12 *dr*) with moderate enantioselectivity (43%), which improved to 86% *ee* (71:29 *dr*) in its T252A variant. P450<sub>eryF</sub>, which natively contains the active site alanine, and CYP142-T234A showed strong preference for the 1*R*,2*S* isomer, catalyzing the reaction with ~90:10 *dr* and ≥97% *ee*. Substituting the conserved threonine with alanine had a drastic effect on P450<sub>Biol</sub>, transforming the *trans*-selective scaffold into a highly *cis*-selective catalyst with unprecedented enantioselectivity among biocatalysts for the 1*S*,2*R* isomer (95% *ee*, 71:29 *dr*). Notably this simple, limited diversity library produced selective variants for all four possible ethyl-2-phenylcyclopropanecarboxylate diastereomers (Figure 2.1, Table 2.1).

To assess how selectivity translates to other substrates, we briefly explored the substrate scope of our most selective variants compared to previously engineered *trans*- and *cis*-selective P450<sub>BM3</sub> variants (P450<sub>BM3</sub>-T268A and P450<sub>BM3-CIS</sub>-T438S, respectively).<sup>16</sup> We screened variants

against styrenes with substituents at the  $\alpha$ -vinyl position (**1b**) and with electron donating or withdrawing substituents on the aromatic ring (**1c** and **1d**, respectively.) Tolerance to substitutions varies widely depending on the protein scaffold (Table 2.2). *Trans*-selective scaffolds, P450<sub>BM3</sub>-T268A and P450<sub>PikC</sub>-T247A, identified using styrene as a model substrate, retained moderate to high activity and *trans*-diastereoselectivity on all substrates tested. Conversely, the *cis*-selective enzymes demonstrated more variability. For example, P450<sub>BM3</sub>-CIS-T438S, retained *cis*-selectivity against methoxy-substituted styrene **1c**, but became a *trans*-selective enzyme when presented with electron withdrawing substituents and increased branching on the olefin. P450<sub>Biol</sub>-T238A showed improved *cis*-selectivity in the presence of  $\alpha$ -methyl styrene (**1b**); however, substitutions on the styrene ring lead to diminished or reversed diastereoselectivity. CYP142-T234A, however, remains highly *cis*-selective and displays high enantioselectivity (>90% *ee*) for all substrates tested. Despite the difficulties in predicting trends, using only a small P450 library, moderate to high selectivity for encumbered *cis*-diastereomers was observed for each model substrate.

## Discussion

Interestingly, the most active and stereoselective variants presented in this study contained the mutation of the active site threonine to alanine, and the mutation tended to increase activity and/or selectivity in over half of the enzymes screened. This highly conserved residue is located in the kinked region of the I-helix (Figure S2.3) and has been proposed to facilitate proton delivery, activation of molecular oxygen, and the stabilization of other catalytic intermediates.<sup>22</sup> In cysteine-ligated BM3 variants the T268A mutation is required for cyclopropanation activity, which initially led us to hypothesize that the mutation to a smaller residue may relieve steric clash that prevents favorable binding of reactants. However, mutating the active site threonine to

valine, a residue that is similar in size, also produced a highly active variant with similar selectivity to P450<sub>BM3</sub>-T268A (Figure S2.5). Alternatively, enzyme inactivation may result from direct carbenoid insertion into the protein scaffold (e.g. O-H insertion into the threonine side chain). However, no change in protein mass was observed after incubating the enzyme with styrene and ethyl diazoacetate. While the mutation does not appear to cause significant changes to secondary and tertiary structure in P450<sub>BM3</sub> (RMSD of 0.5 Å between the wild type and T268A structures, 2IJ2 and 1YQO, respectively), backbone rearrangements in this region have been observed in other P450 crystal structures.<sup>23-24</sup> The mutation may induce subtle structural changes or alter hydrogen bonding networks, active site water composition, or protein dynamics that affect enzymatic cyclopropanation in a manner that has yet to be elucidated.

## Conclusion

In summary, we have shown that natural P450 diversity provides a rich and rapid means for identifying biocatalysts with moderate to high selectivity for most cyclopropane diastereomers including *cis*-diastereomers that are traditionally difficult to produce. In addition, mutations identified in previous engineering experiments can guide library design, increasing the likelihood of isolating robust and stereoselective catalysts. Although the effects of the conserved threonine to alanine mutation are not universal, it provides an important target for engineering P450-based cyclopropanation catalysts. Importantly, incorporation of this mutation allowed the discovery of *cis*-selective catalysts that would not have been discovered by screening only the wild type enzymes. Although there remains room for improvements, this work shows that a diversity-based strategy incorporating key mutations can help create small, focused libraries that provide rapid access to selective starting points for further engineering and laboratory evolution.

## Methods

*General.* All chemicals and reagents were purchased from commercial suppliers (Acros, Fisher, Sigma). Silica gel chromatography was performed on an automated Biotage Isolera One using 10 g SNAP columns. Proton and carbon magnetic resonance spectra ( $^1\text{H}$  NMR at 600 MHz and  $^{13}\text{C}$  NMR at 150 MHz) were recorded on a Bruker model DRX 600 spectrometer with solvent resonance as the internal standard ( $^1\text{H}$  NMR:  $\text{CDCl}_3$  at 7.27 ppm;  $^{13}\text{C}$  NMR:  $\text{CDCl}_3$  at 77.0 ppm). Synthetic reactions were monitored by thin layer chromatography (EMD Millipore TLC silica gel 60 F254) using a UV-lamp for visualization. Gas chromatography analyses were carried out using an Agilent 7820A gas chromatograph, FID detector, and a J&W scientific cyclosil-B column (30m x 0.25 mm x 0.25  $\mu\text{m}$  film). Cyclopropane product standards for the reaction of ethyl diazoacetate (EDA) with styrene (ethyl 2-phenylcyclopropane-1-carboxylate),  $\alpha$ -methylstyrene (ethyl 2-methyl-2-phenylcyclopropane-1-carboxylate), 4-methoxystyrene (ethyl 2-(4-methoxyphenyl)cyclopropane-1-carboxylate), and 4-trifluoromethylstyrene (ethyl 2-(4-(trifluoromethyl)phenyl)cyclopropane-1-carboxylate) were prepared using purified P450<sub>BM3</sub>-T268A and P450<sub>BM3-CIS</sub>-T438S as previously described.<sup>16</sup>

*Library construction by functional site profiling.* To guide library design we used the bioinformatics tool Deacon Active Site Profiler 2 (DASP2). This program allows the user to create a signature active site profile for protein(s) of known structure and then search for sequences in existing databases that contain similar features.<sup>19-21,25</sup> We used the substrate-free and N-palmitoylglycine-bound structures of P450<sub>BM3</sub> as well as a substrate-free structure of P450<sub>cam</sub> to create an active site profile (PDB: 2IJ2, 1JPZ, and 1PHC respectively).<sup>23,26-27</sup> We then chose 3 key residues that are highly conserved among P450s: T268, F393, and C400 in P450<sub>BM3</sub> (Figure 2.2); and T252, F350, and C357 in P450<sub>cam</sub>. Using these structures as input, DASP2 extracted an active site signature for each protein consisting of the residues in the immediate

structural vicinity of each of the defined residues. These signatures were aligned by DASP2 to create an active site profile. The program then split the profile into small motifs based on continuous fragments in a protein sequence and searched the PDB and GenBank NR databases for proteins that share these characteristic motifs.

From the DASP2 PDB and GenBank searches, hundreds of unique P450 sequences were identified with significant DASP search scores that could be potential candidates for library design. Sequences with DASP search scores  $\leq 1\text{E-}6$  (PDB) or  $\leq 1\text{E-}9$  (GenBank) were selected for further evaluation.<sup>19</sup> From these sequences the library was narrowed down to enzymes that fit into at least two of the following categories: 1) P450s derived from bacterial origin or previously produced in *E. coli* to facilitate heterologous expression, 2) P450s with available crystal structures to aid future protein engineering efforts, and 3) P450 sequences that shared  $< 20\%$  sequence identity to P450<sub>BM3</sub> to maximize structural variation among P450 active sites.

*P450 gene sequences and cloning.* Wild type and mutant variants of P450<sub>BM3</sub> were obtained from the Arnold lab (Caltech). Wild type and threonine-to-alanine mutant genes for the P450 scaffolds used in this study were obtained via commercial gene synthesis from Gen9 Inc. (Cambridge, MA). Gen9 Inc. gene sequences were codon optimized to facilitate gene synthesis. pET-21c(+) was used as a cloning and expression vector for all enzymes described in this study. P450 genes were cloned into pET-21c(+) using NdeI and XhoI restriction sites to provide constructs with a C-terminal 6xHis tag for protein purification. The P450 genes in pET-21c(+) were transformed into *E. coli* strain BL21(DE3) for subsequent expression. In addition to P450s identified by the DASP2 algorithm (above), the library was supplemented with two additional enzymes, P450<sub>eryF</sub> and CYP122A2, which are known in the literature and contain a native alanine and serine respectively in place of the active site threonine (Figure S2.1).



*Gene sequence information.* Codon optimized gene sequences purchased from Gen9 are listed below. The conserved active site threonine is in bold and underlined. For P450<sub>eryF</sub> and CYP122A2 the corresponding alanine and serine residues respectively are bolded and underlined. The optimized gene sequences for the threonine-to-alanine mutants are identical to wild type except for the codon encoding the active site threonine, which was changed to GCG for alanine.

P450<sub>cam</sub>: *Pseudomas putida*

ATGACGACTGAAACCATACAAAGCAACGCCAATCTTGCCCCTCTGCCACCCCATGTG  
CCAGAGCACCTGGTATTTCGACTTCGACATGTACAATCCGTCGAATCTGTCTGCCGGC  
GTGCAGGAGGCCTGGGCAGTTCTGCAAGAATCAAACGTACCGGATCTGGTGTGGAC  
TCGCTGCAACGGCGGACACTGGATCGCCACTCGCGGCCAACTGATCCGTGAGGCCT  
ATGAAGATTACCGCCACTTTTCCAGCGAGTGCCCGTTTCATCCCTCGTGAAGCCGGCG  
AAGCCTACGACTTCATTCCCACCTCGATGGATCCGCCCCGAGCAGCGCCAGTTTCGTG  
CGCTGGCCAACCAAGTGGTTGGCATGCCGGTGGTGGATAAGCTGGAGAACCGGATC  
CAGGAGCTGGCCTGCTCGCTGATCGAGAGCCTGCGCCCCGCAAGGACAGTGCAACTT  
CACCGAGGACTACGCCGAACCCCTTCCCGATACGCATCTTCATGCTGCTCGCAGGTCT  
ACCGGAAGAAGATATCCCGCACTTGAAATACCTAACGGATCAGATGACCCGTCCGG  
ATGGCAGCATGACCTTCGCAGAGGCCAAGGAGGCGCTCTACGACTATCTGATACCG  
ATCATCGAGCAACGCAGGCAGAAGCCGGGAACCGACGCTATCAGCATCGTTGCCAA  
CGGCCAGGTCAATGGGCGACCGATCACCAGTGACGAAGCCAAGAGGATGTGTGGCC  
TGTTACTGGTCGGCGGCCTGGAT**ACCG**GTGGTCAATTTCTCTCAGCTTCAGCATGGAGT  
TCCTGGCCAAAAGCCCGGAGCATCGCCAGGAGCTGATCGAGCGTCCCGAGCGTATT  
CCAGCCGCTTGCGAGGAAGTACTCCGGCGCTTCTCGCTGGTTGCCGATGGCCGCATC  
CTCACCTCCGATTACGAGTTTCATGGCGTGCAACTGAAGAAAGGTGACCAGATCCTG  
CTACCGCAGATGCTGTCTGGCCTGGATGAGCGCGAAAACGCCTGCCCGATGCACGT  
CGACTTCAGTCGCCAAAAGGTTTACACACACCACTTTGGCCACGGCAGCCATCTGTG  
CCTTGGCCAGCACCTGGCCCCGCCGGGAATCATCGTCACCCTCAAGGAATGGCTGA  
CCAGGATTCTGACTTCTCCATTGCCCCGGGTGCCAGATTTCAGCACAAGAGCGGCA  
TCGTCAGCGGCGTGCAGGCACTCCCTCTGGTCTGGGATCCGGCGACTACCAAAGCG  
GTA

P450<sub>eryF</sub>: *Saccharopolyspora erythraea*

ATGACGACCGTTCCCGATCTCGAAAGCGACTCCTTCCACGTCGACTGGTACCGCACC  
TACGCCGAGCTGCGCGAGACCGCGCCGGTGACGCCGGTGCGCTTCTCGGCCAGGA  
CGCGTGGCTGGTCACCGGCTACGACGAGGCGAAGGCCGCGCTGAGCGACCTGCGCC  
TGAGCAGCGACCCGAAGAAGAAGTACCCGGGCGTGAGAGTTCAGTTCCCGGCATAC  
CTCGGTTTCCCCGAGGACGTGCGGAACTACTTCGCCACCAACATGGGCACCAGCGA  
CCCGCCGACCCACACCCGGCTGCGCAAGCTGGTGTGCGCAGGAGTTCACCGTCCGCC  
GCGTGGAGGCGATGCGGCCCCGCGTCGAGCAGATCACCGCGGAGCTGCTCGACGAG  
GTGGGCGACTCCGGCGTGGTTCGACATCGTCGACCGCTTCGCCCACCCGCTGCCCATC

AAGGTCATCTGCGAGCTGCTCGGCGTCGACGAGAAGTACCGCGGGGAGTTCGGGCG  
 GTGGAGCTCGGAGATCCTGGTCATGGACCCGGAGCGGGCCGAACAGCGCGGGCAGG  
 CGGCCAGGGAGGTCGTCAACTTCATCCTCGACCTGGTCGAGCGCCGCCGACCGAG  
 CCCGGCGACGACCTGCTGTCCGCGCTGATCAGGGTCCAGGACGACGATGACGGTCG  
 GCTCAGCGCCGACGAGCTGACCTCCATCGCGCTGGTGCTGCTGCTGGCCGGTTTCGA  
GCGGTCGGTGAGCCTCATCGGGATCGGCACCTACCTGCTGCTCACCCACCCGGACC  
 AGCTCGCGCTGGTGCGGCGGGACCCGTCGGCGCTGCCCAACGCCGTCGAGGAGATC  
 CTGCGCTACATCGCTCCGCCGGAGACCACCGCGCTTCGCCGCGGAGGAGGTGGA  
 GATCGGCGGTGTCGCGATCCCCAGTACAGCACGGTGCTGGTCGCGAACGGCGCGG  
 CCAACCGCGACCCGAAGCAGTTCCCGGACCCCCACCGCTTCGACGTCACCCGCGAC  
 ACCCGCGGCCACCTGTCGTTCCGGGCAGGGCATCCACTTCTGCATGGGCCGGCCGCTG  
 GCCAAGCTGGAGGGCGAGGTGGCGCTGCGGGCGCTGTTCCGGCCGCTTCCCCGCTCT  
 GTCGCTGGGAATCGACGCCGACGACGTGGTGTGGCGGCGTTCGCTGCTGCTGCGGG  
 GCATCGACCACCTACCGGTGCGGCTCGACGGA

P450<sub>nor</sub>: *Fusarium oxysporum*

ATGGCATCTGGCGCACCGAGCTTCCCTTTCTCGCGCGCCTCTGGTCCTGAGCCACCC  
 GCCGAGTTCGCCAACTTCGAGCTACGAATCCCGTTTCCCAAGTCAAGCTTTTCGAT  
 GGCAGTCTCGCCTGGCTCGTCACTAAGCACAAAGATGTCTGCTTCGTAGCTACTTCT  
 GAAAAGCTCTCCAAGGTCCGCACTCGCCAAGGCTTCCCTGAGCTTAGCGCCAGTGG  
 AAAGCAAGCAGCCAAGGCAAAGCCGACATTTGTCGACATGGATCCCCCAGAGCACA  
 TGCATCAGAGGAGCATGGTGAGCCGACCTTTACCCCCGAAGCTGTCAAGAATTTGC  
 AGCCTTACATCCAGAGGACTGTCGATGATCTACTGGAGCAGATGAAGCAGAAGGGG  
 TGTGCCAATGGTCCAGTTGACCTTGTCAAAGAGTTTGCTCTCCCTGTCCCCTCCTATA  
 TCATCTACACACTTCTCGGAGTTCCCTTCAATGATCTCGAATATCTTACGCAGCAGA  
 ACGCCATTTCGGACAAATGGTAGCTCCACTGCCCCGAGAGGCCTCTGCCGCTAACCAG  
 GAACCTCTTGATTACCTTGCAATTCTTGTCGAGCAGCGTCTCGTAGAGCCAAAGGAT  
 GATATCATCAGCAAGCTCTGCACTGAGCAAGTCAAGCCTGGAAATATCGACAAGTC  
 CGATGCTGTCCAGATTGCTTTCTTGCTCCTTGTCGCTGGCAACGCAACCATGGTAAA  
 CATGATTGCTCTGGGCGTCGCCACTCTGGCTCAGCACCCCTGATCAGCTGGCGCAACT  
 CAAGGCCAACCCATCCCTTGCGCCCCAGTTTGTCGAGGAAGTCTGTGCTACCATAC  
 TGCTTCTGCACTAGCTATCAAGCGTACTGCCAAGGAGGATGTCATGATCGGTGACAA  
 GCTGGTCCGGGCGAATGAAGGCATTATCGCATCCAACCAGTCAGCCAACAGAGATG  
 AAGAAGTCTTCGAGAATCCCGATGAGTTCAACATGAACCGCAAGTGGCCTCCTCAA  
 GATCCTCTTGGCTTTGGCTTTGGTGATCACAGATGTATCGCTGAGCATCTTGCAAAG  
 GCTGAACTCACAACCTGTGTTCTCAACACTGTATCAGAAGTTCCCAGATCTCAAGGTT  
 GCGGTTCCATTGGGAAAGATCAACTACACGCCTCTGAACCGAGATGTCGGAATCGT  
 GGATCTTCCTGTCATTTTT

P450<sub>TxE</sub>: *Streptomyces scabiei* 87.22

ATGACCGTCCCCTCGCCGCTCGCCGACCCGTCCATCGTGCCCGACCCCTACCCTGTC  
 TACGCCGACCTGGCCCAGCGCCGCCCGTCCACTGGGTCGAGCGCCTGAACGCCTG  
 GGCGGTCTTGACGTACGCCGACTGCGCCGCCGGGCTGAAGGATCCCCGGCTCACCG  
 CCGACCGGGGGACGGAAGTGCTGGCCGCGAAGTTCCCCGGACAGCCGCTGCCGCCG  
 GACAACATCTTCCACCGCTGGACCAAGAACGTGGTGATGTACACGGACCCGCCGCT  
 CCACGACGCGCTACGCCGGTCCGTCCGCGCAGGCTTCACCCGTGCCGCGCACCAGC

ACTACGACCAAGTCCTCCAGAAGGTCGCGCACGACCTGGTCGCTTCCATCCCCGGCCG  
GTGCCACCGAGATCGACGCCGTCCCCGCCCTGGCTGCCGAACCTCCCCGTACGCTCCG  
CCGTGCACGCCTTCGGGGTCCCCGAGGAGGACCTCGGATTCCTCATCCCCGCGCGTGA  
ATACGATCATGACGTACCACTCCGGTCCGAAGGATCAGCCGGTGACGCAGGAGATA  
ATCCTGGAAAAGCTCACCGACCTGCACACGTACGCCTCCGAACCTCTCCAGGGCATG  
CGGGGCAAGGTCCTGCCGGACACCGTCATCGCCCCGCTGGCAGCCGCCAGGACGG  
CCTGACCGAGACCACGCCGGAACAGACCGTGCACCAGCTGGCGCTGGTGTTCATCG  
CGTTGTTTCGCGCCCACGACGCCGGGCTCTCTCAGCAGCGGCACGCTCGCGTTTCGCCC  
GCAACCCGCGGCAGGTCGAACGCTTCCTGGCGGACCAGGCGTGCGTGGAACAACAG  
GCGAACGAGGTCTCCGCTACAACGCCTCGAACCAAGTTACCTGGCGCGTCGCGGC  
CAAGGACGTCGAGATGGGCGGCGTACGGATCGAGGCCGGGCAGACTCTCGCCCTGT  
TCCTGGGCTCGGCCAACCGGGACGCCAACATGTTTCGAGCGACCGAACGACTTCGAC  
CTCGACCGTCCCAACAGCGCTCGGCACCTGTCTGTTTCGGCCAAGGGGTGCACGCCTGT  
CTCGCCGCGCAGCTCATCTCCCTGCAGCTGAAGTGTTCTACGTCGCCCTGCTGAAC  
CGCTTCCCGGGCATCCGGACGGCGGGCGAGCCGATCTGGAACGAGAACCTCGAATT  
CCGCTCCCTTCGCTCCCTGCCGCTCAGCCTCCGC

P450<sub>RhF</sub>: *Rhodococcus* sp.NCIMB9784

ATGAGTGCATCAGTTCCGGCGTCGGCGCCGGCGTGTCCCGTCGACCACGCGGCCCTG  
GCGGGCGGCTGCCCCGTGTGGCGAACGCCGCGGCGTTCGATCCGTTCCGTTCCGCG  
TACCAGACCGATCCGGCCGAGTCGCTGCGCTGGTCCCGCGACGAGGAGCCGGTGT  
CTACAGCCCCGAACCTCGGCTACTGGGTCTGCACCCGGTACGAGGATGTGAAGGCGG  
TGTTCCGCGACAACATCCTGTTCTCGCCGGCGATCGCGCTGGAGAAGATCACTCCCG  
TCTCGGCGGAGGCCACCGCCACCTCGCCCCGTACGACTACGCCATGGCCCCGACC  
CTCGTGAACGAGGACGAGCCCCGCCACATGCCGCGCCGCCGCGCGCTCATGGATCC  
GTTACCCCCGAAGGAACCTGGCGCACACGAGGCGATGGTGCGACGGCTCACGCGCG  
AATACGTCGACCGCTTCGTCTGAATCCGGCAAGGCCGACCTGGTGGACGAGATGCTG  
TGGGAGGTTCCGCTCACCGTCGCCCTGCACTTCCTCGGCGTGCCGGAGGAGGACATG  
GCGACGATGCGCAAGTACTCGATCGCGCACACCGTGAACACCTGGGGCCGCCCCGC  
GCCCCGAGGAGCAGGTGGCCGTGCGCCGAGGCGGTTCGGCAGTTCTGGCAGTACGCGG  
GCACGGTGCTCGAGAAGATGCGGCAGGACCCGTGCGGACACGGCTGGATGCCCTAC  
GGGATCCGCAAGCAGCGGGAGATGCCGGACGTCGTCACCGACTCCTACCTGCACTC  
GATGATGATGGCCGGCATCGTCGCCGCGCACGAGACCACGGCCAACGCGTCCGCGA  
ACGCGTTCAAGCTGCTGCTCGAGAACC GCGCGGTGTGGGAGGAGATCTGCGCGGAT  
CCGTCGCTGATCCCCAACGCCGTGAGGAGTGCTGCGCCACTCCGGGTCCGTGGCG  
GCGTGGCGACGGGTGGCCACCGCCGACACCCGCATCGGCGACGTCGACATCCCCGC  
CGGCGCCAAGCTGCTCGTCGTCAACGCGTCCGCCAACCACGACGAGCGCCACTTCG  
AGCGCCCCGACGAGTTCGACATCCGGCGCCCCGAACCTCGAGCGACCATCTCACCTTCG  
GGTACGGCAGCCACCAAGTGCATGGGCAAGAACCTGGCCCCGCATGGAGATGCAGATC  
TTCCTCGAGGAACCTACCACGCGGCTTCCCCACATGGAACCTCGTACCCGATCAGGAG  
TTCACCTACCTGCCGAATACGTCCTTCCGCGGACCCGACCACGTGTGGGTGCAGTGG  
GATCCGCGAGGCGAATCCCGAGCGCACCGATCCTGCTGTGCTGCACCGGCATCAACC  
GGTCACCATCGGAGAACCCGCCGCCCGGGCGGTGTCCCGCACCGTCACCGTCGAGC  
GCCTGGACCGGATCGCCGACGACGTGCTGCGCCTCGTCCTGCGCGACGCCGGCGGA  
AAGACATTACCCACGTGGACTCCCGGCGGCCATATCGACCTCGACCTCGGCGCGCTG  
TCGCGCCAGTACTCCCTGTGCGGCGCGCCCCGATGCGCCGAGCTACGAGATTGCCGTG

CACCTGGATCCCGAGAGCCGCGGGCGGTTCGCGCTACATCCACGAACAGCTCGAGGT  
GGGAAGCCCGCTCCGGATGCGCGGGCCCTCGGAACCATTTTCGCGCTCGACCCCGGCG  
CCGAGCACTACGTGTTTCGTCGCCGGCGGCATCGGCATACCCCAAGTCCTGGCCATGG  
CCGACCACGCCCCGCGCCCGGGGGTGGAGCTACGAACTGCACTACTGCGGGCCGAAAC  
CGTTCCGGCATGGCCTATCTCGAGCGTGTGCGCCGGGCACGGTGACCGGGCCGCCCTG  
CACGTGTCCGAGGAAGGCACCCGGATCGACCTCGCCGCCCTCCTCGCCGAGCCCCG  
CCCCGGCGTCCAGATCTACGCGTGCGGGGCCCGGGCGGCTGCTCGCCGGACTCGAGG  
ACGCGAGCCGGAACCTGGCCCGACGGGGCGCTGCACGTCGAGCACTTCACCTCGTCC  
CTCGCGGGCGCTCGATCCGGACGTCGAGCACGCCTTCGACCTCGAACTGCGTGACTCG  
GGGCTGACCGTGCGGGTTCGAACCCACCCAGACCGTCCTCGACGCGTTGCGCGCCAA  
CAACATCGACGTGCCAGCGACTGCGAGGAAGGCCTCTGCGGCTCGTGCGAGGTCG  
CCGTCTCGACGGCGAGGTCGACCATCGCGACACGGTGCTGACCAAGGCCGAGCGG  
GCGGCGAACCGGCAGATGATGACCTGCTGCTCGCGTGCCTGTGGCGACCGGCTGGC  
CCTGCGACTC

P450<sub>PikC</sub>: *Streptomyces venezuelae*

ATGCGCCGTACCCAGCAGGGAACGACCGCTTCTCCCCCGGTACTCGACCTCGGGGGCC  
CTGGGGGCAGGATTTTCGCGGGCCGATCCGTATCCGACGTACGCGAGACTGCGTGCCGA  
GGGTCCGGCCCCACCGGGTGCGCACCCCCGAGGGGGACGAGGTGTGGCTGGTCGTCG  
GCTACGACCGGGCGCGGGCGGTCTCTGCCGATCCCCGGTTCAGCAAGGACTGGCGC  
AACTCCACGACTCCCCTGACCGAGGCCGAGGCCGCGCTCAACCACAACATGCTGGA  
GTCCGACCCGCCGCGGCACACCCGGCTGCGCAAGCTGGTGGCCCCGTGAGTTCACCA  
TGCGCCGGGTTCGAGTTGCTGCGGCCCCGGGTCCAGGAGATCGTCGACGGGCTCGTG  
GACGCCATGCTGGCGGGCGCCCGACGGCCGCGCCGATCTGATGGAGTCCCTGGCCTG  
GCCGCTGCCGATCACCGTGATCTCCGAACCTCCTCGGCGTGCCCCGAGCCGGACCGCGC  
CGCCTTCCGCGTCTGGACCGACGCCTTCGTCTTCCCGGACGATCCCCGCCAGGCCCA  
GACCGCCATGGCCGAGATGAGCGGCTATCTCTCCCGGCTCATCGACTCCAAGCGCG  
GGCAGGACGGCGAGGACCTGCTCAGCGCGCTCGTGCGGACCAGCGACGAGGACGG  
CTCCCGGCTGACCTCCGAGGAGCTGCTCGGTATGGCCACATCCTGCTCGTCGCGGG  
GCACGAGACCACGGTCAATCTGATCGCCAACGGCATGTACGCGCTGCTCTCGCACC  
CCGACCAGCTGGCCGCCCTGCGGGCCGACATGACGCTCTTGAGCGGCGCGGTGGAG  
GAGATGTTGCGCTACGAGGGCCCCGGTGAATCCGCGACCTACCGCTTCCCGGTCTGA  
GCCCCGTCGACCTGGACGGCACGGTCATCCCGGGCCGGTGACACGGTCCTCGTCGTCCT  
GGCCGACGCCCACCGCACCCCCGAGCGCTTCCCGGACCCGCACCGCTTCGACATCCG  
CCGGGACACCGCCGGCCATCTCGCCTTCGGCCACGGCATCCACTTCTGCATCGGCGC  
CCCCTTGGCCCCGGTTGGAGGCCCGGATCGCCGTCCGCGCCCTTCTCGAACGCTGCCC  
GGACCTCGCCCTGGACGTCTCCCCCGGCGAACTCGTGTGGTATCCGAACCCGATGAT  
TCGCGGGCTCAAGGCCCTGCCGATCCGCTGGCGGCGAGGACGGGAGGCGGGCCGCC  
GTACCGGT

P450<sub>Biol</sub>: *Bacillus subtilis*

ATGACAATTGCATCGTCAACTGCATCTTCTGAGTTTTTTGAAAAACCCATATTCTTTTT  
ACGACACATTGCGAGCTGTTTCATCCTATCTATAAAGGGAGTTTCTTAAATACCCGG  
GCTGGTATGTACAGGATATGAAGAAACGGCTGCTATTTTGAAAGATGCGAGATTC  
AAAGTCCGCACCCCGCTGCCTGAGAGCTCAACCAAATATCAGGACCTTTCACATGTG  
CAAAATCAAATGATGCTGTTTCAGAACCAGCCTGATCATAGACGATTGCGGACGCTT

GCCAGCGGAGCGTTTACGCCGAGAACGACAGAGAGTTATCAGCCGTATATCATTGA  
AACTGTCCATCATTTGCTTGATCAAGTGCAAGGTAAAAAAGATGGAGGTCATTTG  
GACTTTGCTTTTCCTTTAGCAAGTTTTGTCATAGCTAACATTATAGGTGTACCGGAG  
GAAGATAGGGAGCAATTAAGGAGTGGGCTGCGAGTCTCATTCAAACGATTGATTT  
TACCCGCTCAAGAAAGGCATTAACAGAGGGCAATATTATGGCTGTGCAGGCTATGG  
CATATTTCAAAGAGCTGATTCAAAAGAGAAAACGCCACCCTCAACAGGATATGATC  
AGCATGCTCTTGAAGGGGAGAGAAAAGGATAAGCTGACGGAAGAGGAGGCGGCAT  
CTACGTGCATATTGCTGGCGATCGCCGGACATGAGACAACGGTCAATCTCATCAGCA  
ATTCAGTCCTTTGTCTGCTGCAGCATCCAGAACAGCTTTTGAAACTGAGAGAAAATC  
CAGATCTTATTGGTACCGCAGTCGAGGAATGTTTACGCTATGAAAGCCCCACGCAAA  
TGACAGCCAGAGTTGCGTCAGAGGATATTGACATCTGCGGGGTGACGATCCGTCAA  
GGAGAACAAGTCTATCTTTTGTTAGGAGCGGCTAATCGAGACCCTAGCATATTCACG  
AACCCCGATGTCTTCGATATTACGAGAAGTCCTAATCCGCATCTTTCATTTCGGGCAT  
GGCCATCATGTTTGCTTAGGGTCCTCGCTGGCACGATTAGAAGCGCAAATTGCGATT  
AACACTCTTCTGCAGCGAATGCCAGCCTTAATCTTGCGGATTTTGAATGGCGGTAT  
CGGCCGCTTTTTGGATTTCGGGCGCTTGAGGAGCTGCCGGTGACTTTTGAA

CYP164A2: *Mycobacterium Smegmatis* str.MC2-155

ATGCACAATGGGTGGATGTCGACCGCCGCGACAGCCCAAGAGGCGCAGGGACTGCT  
GCTCCAACCTGCTCGACCCTGCCACCCGTGCCGACCCGTACCCGATCTATGACCGCAT  
CCGCCGCGGGCGGACCGCTTGCGCTCCCCGAGGCGAATCTCGCGGTGTTCTCCAGCTT  
TTCCGATTGCGACGACGTTCTGCGGCACCCCTCGTCGTGCAGCGACCGCACGAAATC  
GACCATCTTTCAACGGCAACTCGCCGCCGAGACGACGCGCGACCGCAGGGGCCCCG  
CCAGCTTCCTGTTCTCGACCCGCCCGATCACACGAGGCTGCGCGGCCTGGTCAGCA  
AGGCGTTCGCGCCGCGCGTGATCAAGCGGCTCGAACCCGAGATCACCGCGCTGGTC  
GACCAACTGCTCGACGCTGTGACGGGCCGGAGTTCAACCTCATCGACAACCTCGCC  
TACCCGCTGCCGGTGCGGGTGATCTGCCGACTGCTCGGCGTGCCGATCGAGGACGA  
ACCGAAGTTCAGCCGCGCGTGCGGCGCTGCTGGCCGCGGCCCTGGACCCGTTCTGGC  
GCTCACCGGCGAGACGTCGGATCTGTTGACGAACAGATGAAGGCCGGGATGTGGC  
TGCGCGACTATCTGCGAGCCCTCATCGACGAGCGGCGCCGCACACCGGGTGAGGAC  
CTCATGTCCGGTTTGCTGCGGGTCGAGGAGTCGGGCGATCAGCTCACCGAGGACGA  
GATCATCGCGACCTGCAACCTGCTGCTGATCGCGGGCCACGAAACGACGGTGAACC  
TGATCGCCAACGCGGCCCTGGCCATGTTGCGCACCCCGGACAGTGGGCGCGCTG  
GCCGCCGACGGATCCCGGGCCTCCGCGGTATCGAGGAGACCATGCGGTACGACCC  
GCCCGTGCAAGTTGGTGTCGCGGTACGCAGGCGACGACCTGACCATCGGGACGCACA  
CGGTGCCCCAAGGGCGACACCATGTTGCTGCTGCTCGCCGCGGCGCACCGCGACCCG  
ACGATCGTCGCGCCCCCGACCGGTTTCGATCCGGACCGCGCTCAGATCCGGCACCTC  
GGCTTCGGCAAGGGGGCGCACTTCTGTCTGGGGGCACCTCTGGCCCCGACTCGAGGC  
GACCGTGCGGTTGCCCGCGCTCGCGGCCCGGTTCCCCGAGGCCCGGCTCAGCGGCG  
AACCGGAGTACAAGCGGAATTTGACGCTGCGCGGCATGTCTACGTTATCCATCGCGG  
TA

CYP107N1: *Streptomyces lavendulae*

ATGACCGGCCCGAGGCCGCGGTGCGCGGGTGCCCCCTTCGGCGCCGGCGAGGCGCC  
CGCGTACCCCTTCCACGCCCCCGACCGGCTGGAGCCCGACCCGTACTGGGAGCCGCT  
GCGCCGCGAGCGGCCGCTGCAACGCGTCACGCTGCCGTACGGCGGCGAGGCGTGCC

TCGCCACCCGCTATCAGGACGTGCGCGCGGTCTTCGCCGACCGCAGGTTCTCCCGGC  
AGCTCGCCGTCGCGCCCCGGCGCTCCGCGCTTCTCCCGCACCAGCCGCCGCCGGACG  
CCGTCCTGAGCGTCGAGGGCCCCGACCACGCGCGGCTGCGCCGGCTGGTCGGGAAG  
GTCTTCACGCCGCGCCGCGTGGAGGACATGCGTCCGCTCATCCAGCGCACCGCCGAC  
GGACTCCTCGACGCGATGGAGGAGATGGGGCCGCCCGCGGACCTGGTTCGAGGACTT  
CTCCCTGCCCTTCGCCGTGTCCATGATCTGCGAGCTGCTCGGGCGTGCCGCCCGAGGA  
CCGCAAGCGGTTCTGCGTCTGGTCGGACGCGCTGCTGACGACCACCGCGCACACCCC  
CGCCCAGGTGCGCGACTACATGATGCAGATGCACGACTACCTCGGCGGGCTCGTCG  
CGCAGCGCCGGGTGCGGCCACCGCGGACCTGATCGGCTCCCTCGTGACCGCGCGC  
GACGAGGAGGACAAGCTCACCGAGGGCGAGCTGGTGCGGCTGGCCGAGGCCATCCT  
CATCGCCGGCTACGAGACCTCGGCGAGCCAGATCCCCAACTTCTCTACGTCCTCTT  
CCGCCACCCGCAGCTGCTGGAGCGGATCAGGAACGACCACGACCTCATCCCCGACG  
CCGTCGAGGAACTGCTGCGCTTCGTGCCCATCGGCACCGTGACGGCTTTCCCCGTA  
CGGCCACCGAGGACGTCGAGCTCGGGGGAGTCCTGGTCAGGGCCGGGGAGACGGTC  
GTGCCGTCGATGGGCGCCGCCAACC GCGACCCCGAGCTGTTACGGACCCCGACGA  
GCTGGACCTCGCGCGGGCGGCCGAATCCGCACCTGGGCTTCGGCGCGGGACCGCACC  
ACTGCCTGGGCGCCCAACTGGCCCGGGTGAGCTCCAGATCACGCTCACGACGCTG  
TTCCGCAGATACCCCCGCCTGCGGCTGGCCGTGCCGGAGGAGAGCCTCTCGTGGA  
GGAGGGGCTGATGGTCCGCGGCATGCACACCATGCCGGTCACCTGG

CYP142: *Mycobacterium tuberculosis* UT205

ATGACTGAAGCTCCGGACGTGGATCTGGCCGACGGCAACTTCTACGCCAGCCGCGA  
GGCGCGGGCCGCGTACCGGTGGATGCGGGCCAACCAACCGGTGTTCCGCGATCGCA  
ACGGCCTGGCGGCCGCGTCGACGTACCAGGCGGTGATCGACGCCGAACGTCAACCC  
GAGCTGTTCTCCAACGCCGGCGGCATCCGCCCCGACCAGCCCGCCCTGCCGATGATG  
ATCGACATGGACGATCCCGCACATCTGTTGCGGCGCAAGCTGGTTAACGCCGGCTTC  
ACCCGCAAGCGGGTGAAGGACAAGGAGGCGTCGATTGCCGCGCTGTGTGACACCCT  
GATCGACGCCGTGTGCGAACGCGGGCAGTGTGACTTCGTGCGGGACCTGGCCGCGC  
CGCTACCGATGGCGGTGATCGGCGACATGCTCGGGGTGCGTCCAGAGCAGCGGGAC  
ATGTTCTTGCGGTGGTCCGACGATCTGGTGACATTCTCAGTTCGCATGTGTCTCAAG  
AGGATTTCCAGATCACCATGGACGCCTTCGCGGCCTACAACGACTTCACCCGGGCCA  
CCATTGCGGCACGGCGAGCGGACCCACCGACGACCTGGTCAGCGTGCTGGTGAGT  
TCCGAAGTTGACGGCGAGCGGCTAAGCGACGACGAGCTGGTCATGGAGACGCTGCT  
GATCCTGATCGGCGGCGACGAGACCACGCGGCATACCTTGAGCGGTGGTACCGAGC  
AGCTGCTGCGCAACCGTGACCAGTGGGACCTGCTGCAGCGCGACCCGTGTTGCTGC  
CCGGGGCCATCGAGGAGATGCTACGTTGGACCGCCCCGGTAAAGAACATGTGCCGG  
GTGTTGACCGCGGATACCGAGTTTCACGGCACGGCGTTGTGTGCCGGCGAGAAGAT  
GATGCTGCTCTTCGAGTCGGCGAACTTCGACGAGGCGGTTTTCTGTGAACCGGAAAA  
GTTTGATGTTTCAGCGAAATCCAAACAGCCACTTGGCGTTTTGGCTTCGGCACGCATTT  
CTGCCTGGGCAATCAGCTGGCCCGGTTGGAGCTGTCGTTGATGACGGAACGGGTGTT  
GCGGCGGCTACCCGACCTGCGGTTGGTCGCCGATGACTCCGTGTTGCCGCTGCGGCC  
GGCGAACTTTGTCAGCGGCCTGGAATCCATGCCGGTGGTGTTACGCCGAGCCCGCC  
GCTGGGC

P450<sub>TyIH1</sub>: *Streptomyces fradiae*

ATGGCCTGGGCACCGGACACGGTGTTCGTTGAGGCCGCAGCGAGTCGTGTCCTCG  
TCCGGGGACGCCCCGCGTCACAAAAGGGGATTCTGTTGCCCGCAGCACGCGCCAA  
CGACACCGACGAGGCCGCAGGCCGGCGCAGCATCGCCTGGCCGGTCGCCCCGACCT  
GTCCCTTCAGCCCTCCCGAGCAGTACGCCGCCCTCCGGGCGGAGGAGCCGATCGCCC  
GGGCCGAGCTGTGGGACGGGGCGCCGGTCTGGCTGATCTCCCGCCAGGATCACGTC  
CGGGCGCTGCTGGCCGACCCCCGGGTACAGCATCCATCCCGCGAAGCTCCCGCGGCTC  
TCACCCTCCGACGGTGAGGCCGAGGCGTCCCGTTCGCTGCTGACACTGGACCCGCCC  
GACCACGGAGCGCTCCGCGGCCACTTCATCCCCGAGTTCGGCCTGCGGCGGGGTGCG  
GGAGCTGCGCCCCCTCCGTCGAACAGATCGTACCGGCCTGCTGGATGACCTACCGC  
CCGCGGCGACGAGGCGGATCTGCTGGCCGACTTCGCGCTCCCCATGGCGACCCAGG  
TGATCTGCCGGCTGCTCGACATCCCCTACGAGGACCGGGACTACTTCCAGGAGCGCA  
CCGAACAGGCCACCCGCCCCGCGGCCGGCGAGGAGGCGCTGGAGGCGCTGCTGGAG  
CTGCGCGACTACCTCGACCGGCTGATCAGCGGCAAGACCGGCCGGGAATCCGGCGA  
CGGGATGCTCGGCAGCATGGTGGCGCAGGCCCGCGGTGGCGGGCTGTGCGACGCCG  
ACGTCCTGGACAACGCGGTGCTCCTGCTGGCCGCCGGGCACGAGACCACGGCCAGC  
ATGGTCACGATGAGCGTGCTCGTTCTGCTACAGCACCCACGGCCTGGCGCGAACTC  
ACCGTGAATCCCGGCCTGTTGCCGGGCGCGGTGGACGAACTGCTGCGCTATCTGTGC  
ATCGCCGACGGGCTGCGCCGCTCGGCCACCGCGGACATCGAGATCGACGGCCACAC  
CATCCGCGCCGGCGACGGCCTGGTCTTCCTGCTGGCCGCGGCCAACCGGGACGAGG  
CCGTCTTCTCCGAACCGGAGGCCTTCGACATCCACCGGTCCGCGAGGCGGCATGTGC  
CCTTCGGCTACGGACCCACCAGTGCTGGGACAGAACCTCGCCCCGGATGGAAGT  
GAAGTGGCCCTCGGCGCCGTGCTGGAGAGGCTGCCCCGACTGCGGCCGACCACGGA  
CGTCGCCGGGCTGCGGCTCAAGAGCGACTCCGCGGTCTTCGGGGTGTACGAGCTGC  
CCGTCGCTGG

P450<sub>NovI</sub>: *Streptomyces niveus*

ATGAGCACCCGTCCCACGGTGTCCCCAGCGAACTCGAACAGATCGACCTGGCATC  
ACCGGTCCTGCACGCCGAGTACGAACTGGACGAGATCTTCCGCCACCTGCGCGCCG  
ATGAGCCGGTGTAAGTGGCAGCAGCCACGGAACGAGCAGCCCGGCTTCTGGGTGATC  
AGTCGCCACGCCGACGTGAACGAGGTGTACAAGGACAAGGAGCACTTCACGACGGA  
GCACGGCAACGCGCTGGCCACTTTGCTGACCGGCGGTGACTCTGCCTCGGGCGCCAT  
GCTCGCCGTACCGACGGAGTGCGCCACCATCAGGTGCGCAACGTGTTGTCCAGGG  
GCTTCTCGGCGCGGATGCTCGACCTCATCGCCACACGCTGCAGGAAACCGTGGAC  
GGTCTGCTGCTGGCGGCACTGGAGCGAGGCGAATGCGACGCCGCGCAGGACATCGC  
GGCGGACGTGCCACTCGGGGCGATCTGCGACCTGCTGGAGATTCCCCACGCGGACC  
GGAAGTACCTGCTCGGTCTGACCTCGCACGCGTGGAGCACGGACTACGCGGACGAG  
CCTCCCGAGGAGAGCTGGGTGCGCAAGAACGAGATCCTGCTGTACTTCAGCAAGCT  
GCTCAAGGAGCGTCGCGGTGGAGTCCGGGAGGACATGGTCAGCCTGCTGGCGAACT  
GCCGGATCGACGGCGATCCGCTCAAGGCGGCCGAGCAGATGGCCAACTGCTACGGG  
CTGATGATCGGCGGCGACGAGACCGGCGAGGCACGCCATCACCGGCACGATCCTCGC  
GTTGATCCAGAACCCCGACCACTGGCGGTGCGCTGAAGAACGGCGACGTCGATCTGA  
ACACGGCGACCGAGGAGGCCCTGCGCTGGACCGTGCCGTCGCTGCACGGAGGCCGG  
AAGGCGACCGGAGACGTCGTGATCAACGGCCGGCGGATCAACGCCGGCGATGTGGT  
CAGTGTGTGGATCTCCTCGGCCAACCGTGACGAGACCGTCTTCGACGCGCCGGACG  
AGTTCAACCTCGCCCGCACCCCGAACAAGCACTTCACCTTCGCGTACGGCTCGCACT

ACTGCCTCGGCCACTACCTCGGCCGGATGGAGGTCTACGCCGTGCTCGACGGGCTGC  
GCCGGCTGGTGGGCGACCTGGAGCAGATCGGGCAGGAGCGATGGATCTACTCCAGC  
ATCCTGCACGGGATGAGCTCACTGCCGATCAGGATCACGGGC

P450<sub>EpoK</sub>: *Sorangium cellulosum*

ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTTCGACTTCAAGCC  
GTTCGCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGG  
CAACCCCCATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCTCACC CGATAACCACG  
ACGTGTCGGCGGTGTTCCGCGACGAACGCTTCGCGGTTCAGTCGAGAAGAGTGGGAA  
TCGAGCGCGGAGTACTCGTCGGCCATTCCCAGGCTCAGCGATATGAAGAAGTACGG  
ATTGTTTCGGGCTGCCGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTC  
GTTTACGTCACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCACCCGTCGACCA  
GCTGCTCGATGCTCGCTCCGGACAAGAGGAGTTTCGACGTTGTGCGGGATTACGCGG  
AGGGAATCCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGAC  
GAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTG  
CCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGTCACCGAGGGGCTCGC  
GCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGAAAATGACGTCTT  
GACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTGG  
TCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCTACCTTATCG  
CGTTCGCTGTGCTCAACCTGCTGCGGTGCCCCGAGGCGCTCGAGCTGGTGAAGGCCG  
AGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCCTCA  
GAATAGGAAGTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC  
AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGT  
ATTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGT  
ACGGTAGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGA  
TCGCCGTGGGCACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAACTCCCC  
TGTTTGGATAACACCCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGC  
CCTCCAAAGCTGGA

*Protein expression.* BL21(DE3) *E. coli* cells harboring a pET-21c(+) plasmid encoding the P450 of interest were grown from glycerol stocks overnight (37 °C, 225 rpm) in 5 mL of LB media containing 100 ug/mL ampicillin. The saturated overnight culture was used to inoculate 1 L of terrific broth supplemented with trace metal mix (1000 × = 50 mM FeCl<sub>3</sub>, 20 mM CaCl<sub>2</sub>, 10 mM MnSO<sub>4</sub>, 10 mM ZnSO<sub>4</sub>, 2 mM CoSO<sub>4</sub>, 2 mM CuCl<sub>2</sub>, 2 mM NiCl<sub>2</sub>, 2 mM Na<sub>2</sub>MoO<sub>4</sub>, and 2 mM H<sub>3</sub>BO<sub>3</sub>) and 100 µg/mL ampicillin. The culture was grown at 37 °C and 225 rpm to an O.D. of 1.2 – 1.8 (3.5 – 4 hours). The temperature was lowered to 25 °C, and the culture was shaken for 20 min at the reduced temperature. Isopropyl β-D-1-thiogalactopyranoside (IPTG) and 5-



aminolevulinic acid (ALA) were then added to a final concentration of 0.5 mM and the cells were allowed to shake at 25 °C and 225 rpm for 16 – 20 hours. The cells were harvested by centrifugation (4 °C, 10 min, 4400 xg) and subsequently frozen at -20 °C for at least two hours.

*Protein purification.* Frozen cell pellets were thawed on ice, resuspended in 25 mL of lysis solution containing 0.5 mg/mL lysozyme in buffer A (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 10 mM imidazole, 0.3 M NaCl, pH 8.0), incubated for 30 min on ice, and sonicated (8 min, 20% amplitude, pulse on 0.5s, pulse off 0.8s; Sonic Dismembrator Model 500, Fisher Scientific). Cell debris was removed by centrifugation for 45 min at 27000 xg, and the supernatant was filtered through a 0.45 µm syringe filter before loading onto a pre-equilibrated Ni-NTA HisTrap HP column (5 mL HisTrap HP, GE Healthcare, Piscataway, NJ) using an AKTA Purifier or AKTApurify (GE Healthcare). The column was washed with buffer A until the signal from the detector reached ≤ 60 mAU. P450 was then eluted with 55% buffer B (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 0.3 M imidazole, 0.3 M NaCl, pH 8.0). The combined P450 fractions were concentrated and exchanged into 0.1 M potassium phosphate buffer (pH 8.0) using a 30 kDa molecular weight cut-off Amicon centrifugal filter (EMD Millipore). P450 concentrations were determined by carbon monoxide binding assay (*vide infra*), portioned into 50-100 µL aliquots, and frozen at -80 °C until further use.

*P450 concentration determination.* The concentration of cytochrome P450s was determined using a standard carbon monoxide binding assay. In a polystyrene 96-well plate (Evergreen Scientific) 50 µL of 0.4 M sodium dithionite in 1 M potassium phosphate buffer (pH 8.0) was added to a 200 µL solution of cytochrome P450 (0.1 M potassium phosphate, pH 8.0). The absorbance at 450 nm and 490 nm was recorded using a Tecan M1000 PRO UV/Vis plate reader. The 96-well plate was then placed in a vacuum chamber in a fume hood. The chamber

was evacuated to approximately -15 mmHg, backfilled with carbon monoxide gas, and incubated for 25 min. The plate was removed, and the absorbance at 450 nm and 490 nm was measured again using the plate reader. The P450 concentration was determined from the ferrous carbon monoxide binding difference spectrum by using previously reported extinction coefficient for P450<sub>cam</sub> ( $\epsilon_{450-490}=91,000 \text{ M}^{-1}\text{cm}^{-1}$ ).<sup>28</sup>

*Typical procedure for small-scale cyclopropanation bioconversions under anaerobic conditions.* Milli-Q water and 0.1 M Na<sub>2</sub>S<sub>2</sub>O<sub>4</sub> were placed into a separate sealed vial and degassed with argon for at least 10 minutes prior to use. Small-scale reactions (400  $\mu\text{L}$ ) were conducted in 2-mL crimp vials (Agilent Technologies, San Diego, CA). The headspace of a crimp vial containing P450 and a stir bar was similarly flushed with argon to remove oxygen, however with no bubbling to minimize protein denaturation. Using glass syringes, water and P450 were added under argon to a final volume of 340  $\mu\text{L}$  followed by 40  $\mu\text{L}$  of 0.1 M Na<sub>2</sub>S<sub>2</sub>O<sub>4</sub>. 10  $\mu\text{L}$  of olefin solution (1.2 M in MeOH) was added and allowed to mix for at least 30 sec before adding 10  $\mu\text{L}$  of EDA (0.4 M in MeOH). The argon lines were subsequently removed, and the reactions were left to stir overnight. The final concentration of reagents in the reactions were as follows: 30 mM olefin, 10 mM EDA, 10 mM Na<sub>2</sub>S<sub>2</sub>O<sub>4</sub>, and 20  $\mu\text{M}$  P450. After 12 hours, 15  $\mu\text{L}$  of 6 M HCl was added by syringe to the crimp vials to quench the reaction. The vials were opened and supplemented with 20  $\mu\text{L}$  of internal standard (20 mM 2-phenylethanol in MeOH). The reaction was extracted with 1 mL of ethyl acetate, transferred to a 1.7 mL eppendorf tube, and centrifuged for 5 min at 18400 xg to remove precipitated protein. The organic layer was transferred to a separate eppendorf tube containing anhydrous sodium sulfate for drying. After 10 minutes, the organic layer was filtered through a Pasteur pipette packed with glass wool into a glass vial for chiral GC analysis.

*Synthesis of cyclopropane product standards.* Product standards for the cyclopropanation of styrene,  $\alpha$ -methylstyrene, 4-methoxystyrene, and 4-(trifluoromethyl)styrene were synthesized on a preparative scale as previously reported using P450<sub>BM3</sub>-T268A and P450<sub>BM3</sub>-CIS-T438S.<sup>16</sup>

**Ethyl 2-phenylcyclopropane-1-carboxylate (3a-d)** was synthesized from styrene with EDA and isolated as predominately trans isomers. Diagnostic data for the *E* cyclopropanes: <sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>):  $\delta$  7.30 (m, 2H), 7.22 (m, 1H), 7.11 (m, 2H), 4.18 (q, 7.2 Hz, 2H), 2.53 (m, 1H), 1.92 (m, 1H), 1.62 (m, 1H), 1.33 (m, 1H), 1.30 (t, *J* = 7.2 Hz, 3H); <sup>13</sup>C NMR (150 MHz, CDCl<sub>3</sub>):  $\delta$  173.4, 140.1, 128.4, 126.4, 126.1, 60.7, 26.2, 24.2, 17.1, 14.2.

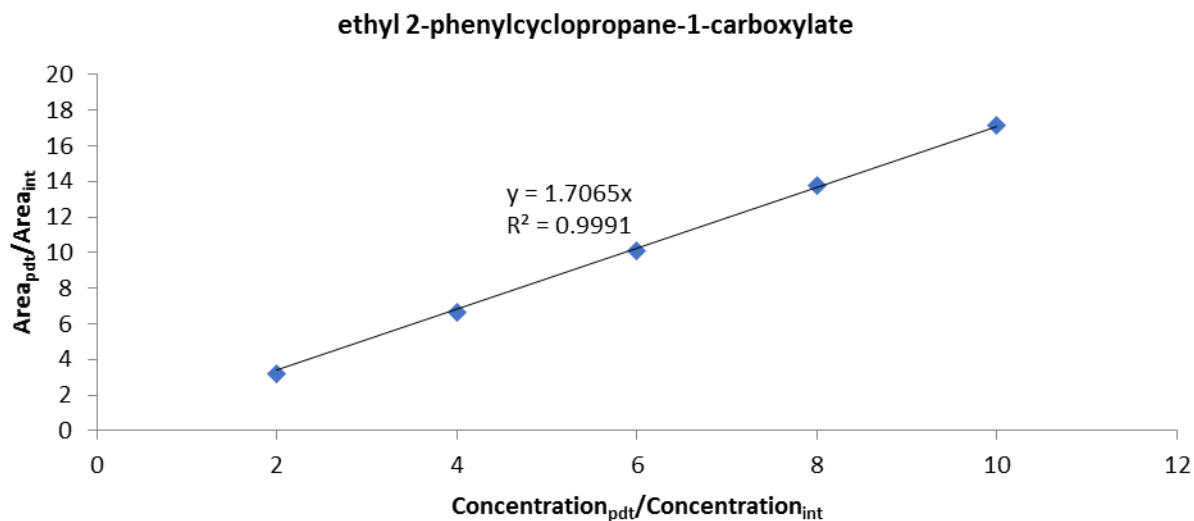
**Ethyl 2-methyl-2-phenylcyclopropane-1-carboxylate (4a)** was synthesized from  $\alpha$ -methylstyrene with EDA and isolated as predominately trans isomers. Diagnostic data for the *E* cyclopropanes: <sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>):  $\delta$  7.31 (m, 4H), 7.22 (m, 1H), 4.20 (q, *J* = 7.1 Hz, 2H), 1.98 (dd, *J* = 8.3, 6.0 Hz, 1H), 1.54 (s, 3H), 1.44 (m, 2H), 1.31 (t, *J* = 7.1 Hz, 3H); <sup>13</sup>C NMR (150 MHz, CDCl<sub>3</sub>):  $\delta$  172.2, 145.8, 128.4, 127.3, 126.4, 60.5, 30.6, 27.9, 20.8, 19.9, 14.4.

**Ethyl 2-(4-methoxyphenyl)cyclopropane-1-carboxylate (4b)** was synthesized from 4-methoxystyrene with EDA and isolated as exclusively trans isomers. Diagnostic data for the *E* cyclopropanes: <sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>):  $\delta$  7.04 (d, *J* = 8.9 Hz, 2H), 6.83 (d, *J* = 8.6 Hz, 2H), 4.18 (q, *J* = 7.2 Hz, 2H), 3.79 (s, 3H), 2.49 (m, 1H), 1.83 (m, 1H), 1.57 (m, 1H), 1.29 (t, *J* = 7.2 Hz, 3H), 1.26 (m, 1H); <sup>13</sup>C NMR (150 MHz, CDCl<sub>3</sub>):  $\delta$  173.6, 158.3, 132.0, 127.3, 113.8, 60.7, 55.3, 25.7, 23.9, 16.8, 14.3.

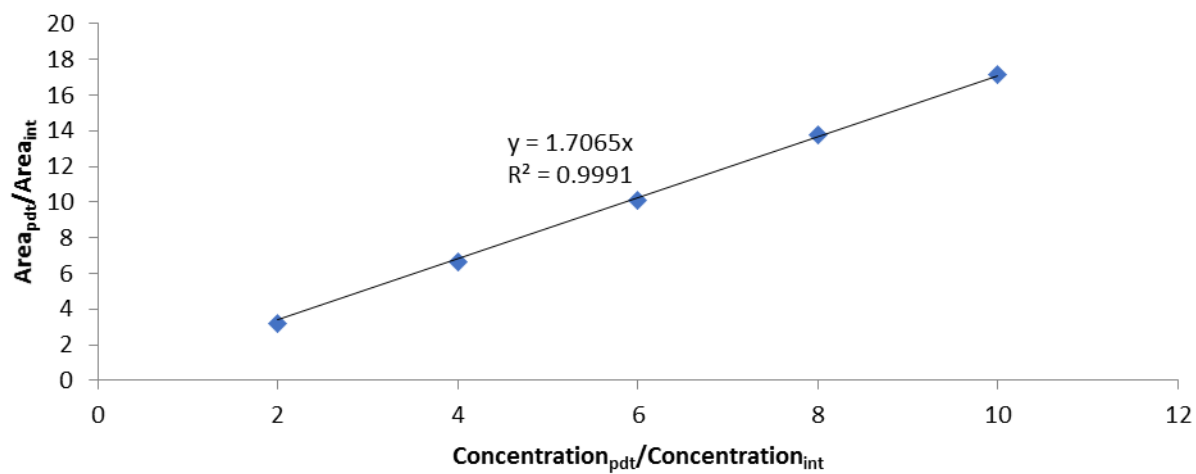
**Ethyl 2-(4-trifluoromethylphenyl)cyclopropane-1-carboxylate (4c)** was synthesized from 4-(trifluoromethyl)styrene with EDA and isolated as a mixture of cis and trans isomers. Diagnostic data for the *E* cyclopropanes: <sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>):  $\delta$  7.53 (d, *J* = 7.8 Hz, 2H), 7.20 (d, *J* = 8.1 Hz, 2H), 4.19 (q, *J* = 6.9 Hz, 2H), 2.56 (m, 1H), 1.95 (m, 1H), 1.67 (m, 1H), 1.35

(m, 1H), 1.30 (t,  $J = 7.1$  Hz, 3H).  $^{13}\text{C}$  NMR (150 MHz,  $\text{CDCl}_3$ ):  $\delta$  172.9, 144.4, 126.3, 125.3, 60.9, 25.7, 24.5, 17.3, 14.2. Diagnostic data for the Z cyclopropanes:  $^1\text{H}$  NMR (600 MHz,  $\text{CDCl}_3$ ):  $\delta$  7.53 (d,  $J = 7.8$  Hz, 2H), 7.39 (d,  $J = 8.2$  Hz, 2H), 3.90 (q,  $J = 7.1$  Hz, 2H), 2.58 (m, 1H), 2.15 (m, 1H), 1.74 (m, 1H), 1.40 (m, 1H), 1.00 (t,  $J = 7.2$  Hz, 3H).

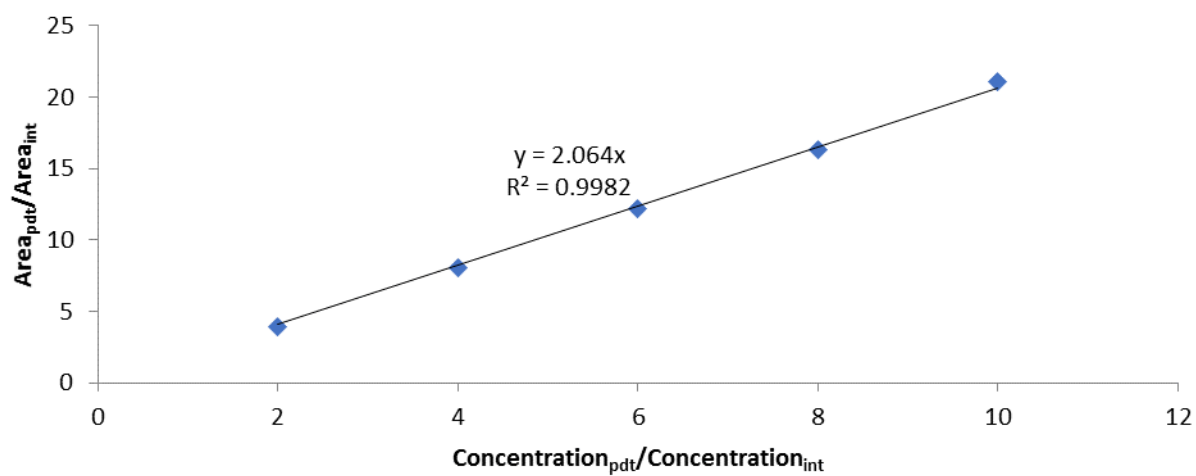
*Calibration curves.* Yields of cyclopropane products were determined using calibration curves made with independently synthesized product standards. A 0.2 M stock solution of product standard was used to prepare samples of product standard at 2 mM, 4 mM, 6 mM, 8 mM, and 10 mM in 0.1 M potassium phosphate (pH 8.0) containing 0.4 mM internal standard (2-phenylethanol) and 5% methanol. The samples were extracted with 1 mL of ethyl acetate, transferred to a 1.7 mL eppendorf tube, and centrifuged for 5 min at 18400 xg. The organic layer was transferred to a separate eppendorf tube containing anhydrous sodium sulfate for drying. After 10 minutes, the organic layer was filtered through a Pasteur pipette packed with glass wool into a glass vial for GC analysis.



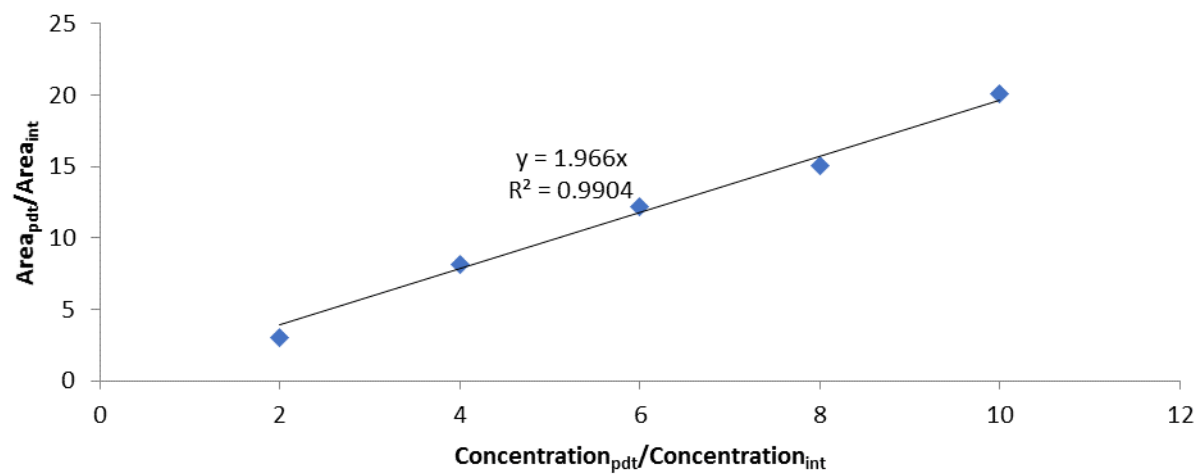
ethyl 2-phenylcyclopropane-1-carboxylate



ethyl 2-methyl-2-phenylcyclopropane-1-carboxylate



ethyl 2-(4-(trifluoromethyl)phenyl)cyclopropane-1-carboxylate



## Supporting Information

**Figure S2.1.** ClustalOmega alignment of the P450 library protein sequences. The conserved active-site threonine is highlighted in yellow, and amino acid substitutions that differ from the highly conserved threonine are highlighted in red.

BM3h	-----MTIKEMPQPKTFG	13
cam	MTT-----ETIQSNANLAPL-----PPHVPEHLVDFDFMYNPSNLSAG	38
TxtE	-----MTVPSPLADPSIVPD	15
NovI	MST-----RPTVSPSELEQIDLASPVLHAEYE	27
RhF	MSA-----SVPASAPACPVD-----HAALAGGCPVSANAAAFDPFGSAYQTD	42
terp	-----MDARATIPHEHIARTVILPQGYADDEV	26
CYP122A2	-----MSTEAQQESTPTARCPFS	18
EpoK	MTQ-----EQANQSETKPAFDKPFAPGYAED	27
CYP142	-----MTEAPDVDLADGNFYASRE	19
nor	-----MASGAPSFPPSRASGPE	17
TylH1	MAWAPDTVFSLRPQRVSVSSGDARPSQKGILLPAARANDTDEAAGRRSIAWPVARTCPFS	60
BioI	-----MTIASSTASSEFLKN	15
CYP107N1	MTG-----PEAAVRGCPFGAGEAPAYPFHAPDRLE	30
CYP164A2	MHN-----GWMSTAATAQEAQGLLLQLLDPATRAD	30
PikC	MRR-----T---QQGTTASPPVLDL GALGQDFAAD	27
eryF	-----MTTVPDLESDFSFHD	15
BM3h	ELKNLPLLNTDKPVQALMKIADDELGEIFKFEAPGRVTRYLSSQRLIKEACDESFRDKNLS	73
cam	VQEAWAVLQESNVDPDLVWTRC-N-----GGHWIATRGQLIRE---AYEDYRHSSEC	86
TxtE	PYPVYADLAQRRPVHWVE---R-----LNAWAVLTYADCAA---GLKDPRLTADRG	60
NovI	LDEIFRHLRADEFPVYWQQPRNEQ-----PGFWVISRHADVNE---VYKDKEHFTTEH	76
RhF	PAESLRWSRDEEPVFYSP---E-----LGYWVVTRYEDVKA---VFRDNILFSPA	87
terp	IYPAFKWLRDEQPLAMAHIEG-Y-----DPMWIATKHADVMO---IGKQPGLFSNAE	74
CYP122A2	IQDGHRTILETGTVGAHELFG---G-----VKQWLVAEAEDVKL---VTNDPRFSSAAP	65
EpoK	PFPATIERLREATPIFY---WD-E-----GRSWVLTRYHDVSA---VFRDERFAVSRE	72
CYP142	ARAAAYRWMRANQPVFRD---R-----NGLAAASTYQAVID---AERQPELFSNAG	63
nor	PPAEFAKL RATNPVSQVKLFD-G-----SLAWLVTKHKDVCF---VATSEKLSKVRT	65
TylH1	PPEQYAALRAEEPIARAELWD-G-----APVWLISRQDHVRA---LLADPRVSIHPA	108
BioI	PYSFYDTLRAVHPIYKGSFL---K-----YPGWYVTGYEETA---ILKDARFKVRTP	62
CYP107N1	PDPYWEPLRREPLQRVTLPLY-G-----GEAWLATRYQDVRA---VFADRRFSRQLA	78
CYP164A2	PYPIYDRIRRGGPLALP---E-----ANLAVFSSFSDCDD---VLRHPSSCSDRT	74
PikC	PYPTYARLRAEGPAHRVRTPE-G-----DEVWLVVGYDRARA---VLADPRFSKDW	75
eryF	WYRTYAELRETAPVTPVRFL-G-----QDAWLVTGYDEAKA---ALSDLRLSSDPK	62
BM3h	QALKF-----VRDFAGDGLFTSWTHEKNWKKAHNILLPSFSQQAMKGYHAMM	120
cam	PFIPR-----EAGEAYDF-IPTSMDPPEQRQFRALANQVVGMPVVDKLENRI	132
TxtE	TEVLA-AKFPQQPLPPDNIFHRWTKN-VVMYTDPLHDALRRSVRAGFTRAAHQHYDQVL	118
NovI	GNALA-----TLLTGGSASGAMLAVTDGVRHHQVRNVLSRGFSARMLDLIAHTL	126
RhF	ALEKITPVSAEATATLARYDYAMAR-TLVNEDEPAHMPRRRALMDPFTPKELAHHEAMV	145
terp	GSEILYDQNNAEAFMRSISGGCPHVID-SLTSMDPPTHTAYRGLTLNWFQPASIRKLEENI	133
CYP122A2	SG-----ILGDRRPG-WFSGMDSPEHNRYRQKIARDFTLRAARKQEEFI	108
EpoK	EWESS-----AEYSSAIPELSDMKKY-GLFGLPPEDHARVRKLVNPSFTSRAIDLRAEI	126
CYP142	-----GIRPDQPALP-MMIDMDPAHLLRRKLVNAGFTRKRVKDKEASI	106
nor	RQGFPG-----ELSASGKQAAKAKP-TFVDMDPPEHMHQSRMVEPTFTPEAVKNLQPYI	117
TylH1	KLPRLL-----SPSDGEAEASR-SLLTLDPPDHGALRGHFIPFGLRRVREL RPSV	157
BioI	LPE-----SSTKYQDL SHVQNQ-MMLFQNPQPDHRLRLTLASGAFTPRTTESYQPYI	112
CYP107N1	VAPGA-----P-----RFLPHQPPPD-AVLSVEGPDHARLRLVGVKVFPPRRVEDMRPLI	127
CYP164A2	KSTI-----FQRQLAAETQPRPQGPA-SFLFLDPPDHTRLRGLVSKAFAPRVIKRLEPEI	128
PikC	NSTTP-----LTEAEALNH-NMLESPPRHTRLRKLVAAREFTMRREVLLRPRV	123
eryF	KKYPGVEVEFPAYLGFPEDVRNYFAT-NMGTSDPPTHTRLRKLVSQEFTVRRVEAMRPRV	121
BM3h	VDIAVQLVQKWERL-NADEHIEVPEDMTRLTLDTIGLCGFNYRFNSFYRDQPHPFITSMV	179

cam	QELACSLIESLRPQ----	GQCNF-----	TEDYAEPFPIRIFMLLA	168
TxtE	QKVAHDLVASIPAGA---	TEIDA-----	VPALAAELPVRSAVHAF	155
NovI	QETVDGLLLAALER---	GECD-----	AQDIAADVPLGAICDLL	162
RhF	RRLTREYVDRFVES---	GKADL-----	VDEMLWEVPLTVALHFL	181
terp	RRIAQASVQRLLDFFD---	GECDF-----	MTDCALYYPLHVMTAL	170
CYP122A2	VRAADSCLDDEASG---	PGTDL-----	VPGYAKRLASLAHDL	145
EpoK	QRTVDQLLDARSGQ---	EEFDV-----	VRDYAEGIPMRAISALL	162
CYP142	AALCDTLIDAVCER---	GECDF-----	VRDLAAPLPMAVIGDML	142
nor	QRTVDLLLEQMKQKGCANGPVDL-----		VKEFALPVPSYIIYTLL	157
TylH1	EQIVTGLLDDLTARG---	DEADL-----	LADFALPMATQVICRLL	194
BioI	IETVHHLLDQVQGK---	KKMEV-----	ISDFAFPLASFVIANII	148
CYP107N1	QRTADGLLDAMEEMG---	PPADL-----	VEDFSLPFAVSMICELL	164
CYP164A2	TALVDQLLDVAVDG---	PEFNL-----	IDNLAYPLPVAVICRLL	163
PikC	QEIVDGLVDAMLAAAP--	DGRADL-----	MESLAWPLPITVISELL	161
eryF	EQITAEELLDEVGDS---	GVVDI-----	VDRFAHPLPIKVICELL	157

: :

BM3h	RALDEAMNKLQRANPDDPAY-----	DENKRQFQEDIKVMNDLVDKIIADR	224
cam	GLPEEDIPHLKYLTDMQTR-----	PDGSMTFAEAKEALYDYLIPITIEQR	212
TxtE	GVPEEDLGFLIPRVNTIMTYHSGPK-----	DQPVQTQEIILEKLTDLHTYASELLQGM	207
NovI	EIPHADRKYLLGLTSHAWST-----	DYADEPPEESWVAKNEILLYFSKLLKER	210
RhF	GVPEEDMATMRKYSIAHTVNTWG-----	RPAPEEQVAVAEAVGRFWQYAGTVLEKM	232
terp	GVPEDEDEPLMLKLTQDFFGVHEPDEQAVAAPRQSADEAARRFHETIATFYDYFNGFTVDR		230
CYP122A2	GLNEEGPVLEGQMRAM-----	EGGTMESIKRLTDEFGHVLALV	184
EpoK	KVPAECDEKFRFRFGSATARALGVGLV-----	PRVDEETKTLVASVTEGLALLHGVLDER	216
CYP142	GVRPEQRDMFLRWSDDLVTFLSSHVS-----	QEDFQITMDAFAAYNDFTRATIAAR	193
nor	GVPFNDLEYLTQQNAIRTNG-----	SSTAREASAAANQELLDYLAILVEQR	202
TylH1	DIPYEDRDYFQERTEQATR-----	PAAGEEALQELLELDYLDRLISGK	238
BioI	GVPEEDREQLKEWAASLIQITIDFTRS----	R-----KALTEGNIMAVQAMAYFKELIQKR	199
CYP107N1	GVPPEDRKRFCVWSDALLTT-----	TAHTPAQVRDYMMQMHYDLGGGLVAQR	210
CYP164A2	GVPIEDEPKFSRASALLAAALDPFLA----	LTGETSDLFDEQMKAGMWLRDYLRLALIDER	219
PikC	GVPEPDRAAFRVWTDAFVF-----	PDDPAQAQTAMAEMSGYLSRLIDSK	205
eryF	GVDEKYRGEFGRWSSEILVM-----	DPERAEQRGQAAREVVNFILDLVERR	203

BM3h	KASGEQSD-DLLTHMLNGKDPETGEPLDDENIRYQIITFLIAGHETTSGLLSFALYFLVK	283
cam	R--QKPGT-DAISIVANGQ--VNGRPITSDEAKRMCGLLLVGGLDTVVNFLSFSMEFLAK	267
TxtE	R--GKVLPTVIARLAAAQD-GLTETTPEQTVHQALVFIALFAPTTPGSLSSSGTALAFAR	264
NovI	R--GGVRE-DMVSLLANCR--IDGDPKAAEQMANCYGLMIGGDETGRHAITGTILALIQ	265
RhF	R--QDPSGHGWMPYGIKQQR-EMPDPVVTDSYLHSMMAAGIVAAHETTANASANAFKLLLE	289
terp	R--SCPKD-DVMSLLANSK--LDGNYIDDKYINAYYVAIATAGHDTTSSSSGGAIIGLSR	285
CYP122A2	RAKRDEAGDRLLHRLAESG--EDEILLSDEATGVFATLLFAGHDSMQQMVGYSLYALLS	242
EpoK	R--RNPLENDVLTMLLQAE--ADGSRLSTKELVALVGAIIAAGTDTTIYILIAFAVLNLLR	272
CYP142	R--ADPTD-DLVSVLVSS--VDGERLSDELVMETLLILIGGDETRHRTLSSGGTEQLLR	248
nor	L--VEPKD-DIISKLTEQ--VKPGNIDKSDAVQIAFLLLVAGNATMVNMIALGVATLAQ	257
TylH1	T--GRESGDGMLGSMVAQA--RGGGLSHADVLDNAVLLLAAGHETTASMTMSVLVLVLLQ	293
BioI	K--RHPQQ-DMISMLLKGR---EKDKLTEEEAASTCILLAIAGHETTIVNLISNSVLCLLQ	253
CYP107N1	R--VRPTA-DLIGSLVTAR--DEEDKLTEGELVRLAEAILIAGYETSASQIPNFLYVLFR	265
CYP164A2	R--RTPGE-DLMSGVLVAVE--ESGDQLTEDEIIATCNLLLIAGHETTIVNLIANAALAMLR	274
PikC	R--GQDGE-DLLSALVRTSD-EDGSRLTSEELLGMAHILLVAGHETTIVNLIANGMYALLS	261
eryF	R--TEPGD-DLLSALIRVQD-DDDGRLSADELTSIALVLLLAGFESVSLIGIGTYLLLT	259

: : :

BM3h	NPHVLQKAAEEAARVLVDPVPSYKQVKQLKYVGMVLNEALRLWPTA--PAFSLYAKEDTV	341	
cam	SPEHRQELIERPER-----	IPAACEEELRRFSLV---ADGRILTSDYE	307
TxtE	NPRQVERFLADQAC-----	VDNTANEVLRYNASN--QFTWRVAAKDVE	305
NovI	NPDQWRALKNGDVD-----	LNTATEEALRWTVPS--LHGGRKATGDV	306
RhF	NRAVWEEICADPSL-----	IPNAVEECLRHSGSV--AAWRRVATADTR	330
terp	NPEQLALAKSDPAL-----	IPRLVDEAVRWTAAPV--KSFMRALADTE	326
CYP122A2	HPEQRAALRENPD-----	IDGAVEELLRFLPLNQ-LGVPRVCVEDVE	284
EpoK	SPEALELVKAEPGL-----	MRNALDEVLRFDNILR-IGTVRFARQDLE	314
CYP142	NRDQWDLQLQRDPSL-----	LPGAIEEMLRWTAAPV--KNMCRVLTADTE	289
nor	HPDQLAQLKANPSL-----	APQFVEELCRYHTASA-LAIKRTAKEDVM	299



TylH1	HPTAWRELTVPNGL-----LPGAVDELLRYLSIA--DGLRRSATADIE	334
BioI	HPEQLLKLRENPD-----IGTAVEECLRYESPT--QMTARVASEDID	294
CYP107N1	HPQLLERIRNDHDL-----IPDAVEELLRFVPIGTVDGFPRTATEDVE	308
CYP164A2	TPGQWAALAADGSR-----ASAVIEETMRYDPPV--QLVSRYAGDDLT	315
PikC	HPDQLAALRADMTL-----LDGAVEEMLRYEGPVE-SATYRFPVEPVD	303
eryF	HPDQLALVRRDPSA-----LPNAVEEILRYIAPP--ETTRFAAAEEVE	300
	: * *	
BM3h	LGGEYPLEKGDDELMVLIPQLHRDKTIWGDDVEEFRPERFENPSAIPQHAFKPFNGQRAC	401
cam	FHGV-QLKKGDAQILLPQMLSGLDERENA-----CPMHVDFSRQ--KVSHTTFGHGSHLC	358
TxtE	MGGV-RIEAGQTLALFLGSANRDANMFE-----RPNDFDLDRP-NSARHLSFGQGVHAC	357
NovI	INGR-RINAGDVVS VWISSANRDET VFD-----APDEFNLART--PNKHFTFAYGSHYC	357
RhF	IGDV-DIPAGAKLLVNASANHDERHFE-----RPDEFDIRRP-NSSDHLTFGYGSHQC	382
terp	VRGQ-NIKRGDRIMLSYPSANRDEEVFS-----NPDEFDITRF--PNRHLGFGWGAHMC	377
CYP122A2	LHGQ-TISAGDNVIPLYSTANRDGPVFA-----DPDTFDITRK--PEHNFAFGYGIHGC	335
EpoK	YCGA-SIKKGE MVFL LIP SALRDGT VFS-----RPDVFDVRRD--TSASLAYGRGPHVC	365
CYP142	FHGT-ALCAGEKMMLLFESANFDEAVFC-----EPEKFDVQRN--PNSHLAFGFGTHFC	340
nor	IGDK-LVRANEGIIASNQSANRDEEVFE-----NPDEFNMNRKWPPQDPLGFGGDHRC	352
TylH1	IDGH-TIRAGDGLVFL LAAANRDEAVFS-----EPEAFDIHRS--ARRHVAFGYGPHQC	385
BioI	ICGV-TIRQGEQVYLL LGAANRDP SIFT-----NPDVFDITRS--PNPHLSFGHGHVVC	345
CYP107N1	LGGV-LVRAGETVVP S MGAANRDEPELFT-----DPDELDLARR--PNPHLGFGAGPHHC	359
CYP164A2	IGTH-TVPKGD TML LLLAAHRDPTIVG-----APDRFDPDRA--QIRHLGFGKGAHFC	366
PikC	LDGT-VIPAGDTVLVVLADAHRTPERFP-----DPHRFDIRRD--TAGHLAFGHHIHC	354
eryF	IGGV-AIPQYSTVLVANGAANRDPKQFP-----DPHRFDVTRD--TRGHLSFGQGIHFC	351
	: : * .: .: *	
BM3h	IGQQFALHEATLV LGMMLKHDFEDHTNYELDIKETLTLKPEGFVVKAKSKKIPLGGIPS	461
cam	LGQHLARREIIVTLKEWLTRIPDFSIA-----GAQIQHKSGIVSGVQ-----ALPL	405
TxtE	LAAQLISLQLKWIFYALLNRFPGIRTAGE-----PIWNNLEFRSLR-----SLPL	403
NovI	LGHYLG RMEVYAVLDGLRRLVGDL EQIGE-----ERWIYSSILHGMS-----SLPI	403
RhF	MGKNLARMEMQIFLEELTTRLPHEMELVPD-----QEFTYLPNTSFRGPD-----HVWV	430
terp	LGQHLAKLEMKIFFEELLPKLKSVELSGP-----PRLVATNFVGGPK-----NVPI	423
CYP122A2	PGQHLARVLIKVATVRLFERFPDVRLAGD-----VPMNEGLGLFSPA-----ELRV	381
EpoK	PGVSLARLEAEIAVGTIFRRFPKMLKET-----PVFGYHPAFRNIE-----SLNV	411
CYP142	LGNQLARLELSLMTERVLRRLPDLRLVADD----SVLPLRPANFVSGLE-----SMPV	389
nor	IAEHLAKAELTTVFSTLYQKFDPDKVAVPL----GKINYTPLNRDVGIV-----DLPV	401
TylH1	LGQNLARMELEVALGAVLERLPALRPPTDV---AGLRLKSDSAVFGVY-----ELPV	434
BioI	LGSSLARLEAQIAINTLLQRMPSLNAD-----FEWRYRPLFGFRALE-----ELPV	392
CYP107N1	LGAQLARVELQITLTTLFRYPRLRLAVPE----ESLSWKEGLMVRGMH-----TMPV	408
CYP164A2	LGAPLARLEATVALPALAARFPEARLSGE-----PEYKRNLTLRGMS-----TL SI	412
PikC	IGAPLARLEARIAVRALLERCPLDALDVSP----GELVWYPNPMIRGLK-----ALPI	403
eryF	MGRPLAKLEGEVALRALFGRFPALSLGIDA----DDVVWRRSLLLRGID-----HLPV	400
	. : :	
BM3h	PST-----	464
cam	VWDPATTKAV-----	415
TxtE	SLR-----	406
NovI	RITG-----	407
RhF	QWDPQANPERTDPAVLHRHQVPTIGEPAAARAVSRTVTVERLDRIADDVLRLVLRDAGGKT	490
terp	RFTKA-----	428
CYP122A2	TWGAE-----	386
EpoK	ILKPSKAG-----	419
CYP142	VFTSPPLG-----	398
nor	IF-----	403
TylH1	AW-----	436
BioI	TFE-----	395
CYP107N1	TW-----	410
CYP164A2	AV-----	414
PikC	RWRRGREAGRRTG-----	416
eryF	RLDG-----	404
BM3h	-----	464

cam	-----	415
TxtE	-----	406
NovI	-----	407
RhF	LPTWTPGAHIDLDLGALSRQYSLCGAPDAPSYEIAVHLDPESRGGSRYIHEQLEVGSPRLR	550
terp	-----	428
CYP122A2	-----	386
EpoK	-----	419
CYP142	-----	398
nor	-----	403
TylH1	-----	436
BioI	-----	395
CYP107N1	-----	410
CYP164A2	-----	414
PikC	-----	416
eryF	-----	404

BM3h	-----	464
cam	-----	415
TxtE	-----	406
NovI	-----	407
RhF	MRGPRNHFALDPGAEHYVFVAGGIGITPVLAMADHARARGWSYELHYCGRNRSGMAYLER	610
terp	-----	428
CYP122A2	-----	386
EpoK	-----	419
CYP142	-----	398
nor	-----	403
TylH1	-----	436
BioI	-----	395
CYP107N1	-----	410
CYP164A2	-----	414
PikC	-----	416
eryF	-----	404

BM3h	-----	464
cam	-----	415
TxtE	-----	406
NovI	-----	407
RhF	VAGHGDRAALHVSEEGTRIDLAALLAEPAPGVQIYACGPGRLLAGLEDASRNWPDGALHV	670
terp	-----	428
CYP122A2	-----	386
EpoK	-----	419
CYP142	-----	398
nor	-----	403
TylH1	-----	436
BioI	-----	395
CYP107N1	-----	410
CYP164A2	-----	414
PikC	-----	416
eryF	-----	404

BM3h	-----	464
cam	-----	415
TxtE	-----	406
NovI	-----	407
RhF	EHFTSSLAALDPDVEHAFDLELRDGLTVRVEPTQTVLDALRANNIDVPSDCEEGLCGSC	730
terp	-----	428
CYP122A2	-----	386
EpoK	-----	419
CYP142	-----	398
nor	-----	403

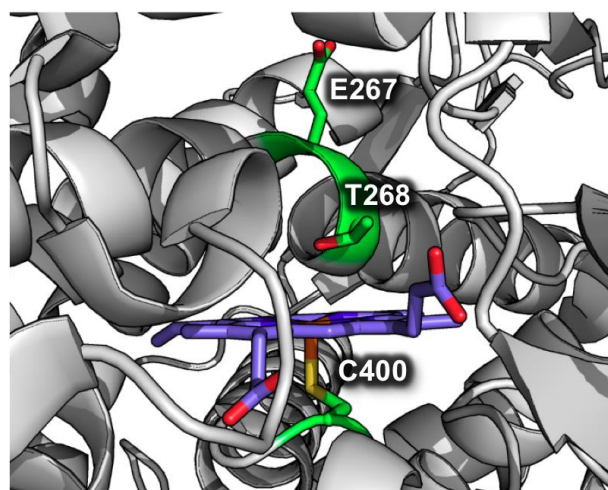
TylH1	-----	436
BioI	-----	395
CYP107N1	-----	410
CYP164A2	-----	414
PikC	-----	416
eryF	-----	404

BM3h	-----	464
cam	-----	415
TxtE	-----	406
NovI	-----	407
RhF	EVAVLDGEVDHRDTVLTKAERAANRQMMTCCSRACGDRLALRL	773
terp	-----	428
CYP122A2	-----	386
EpoK	-----	419
CYP142	-----	398
nor	-----	403
TylH1	-----	436
BioI	-----	395
CYP107N1	-----	410
CYP164A2	-----	414
PikC	-----	416
eryF	-----	404

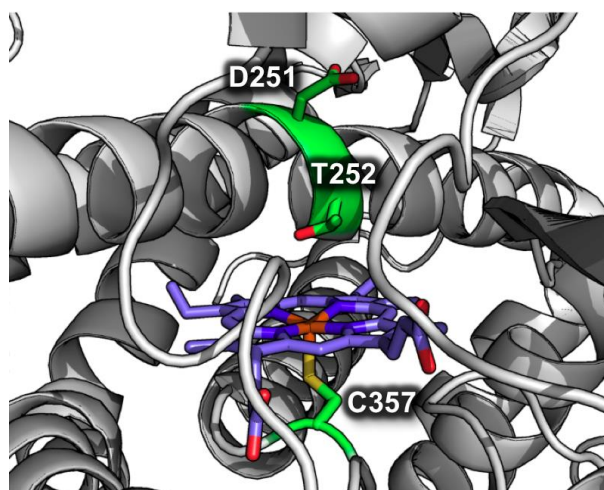
**Figure S2.2.** Percent identity matrix generated by Clustal 12.1.

BM3h	100.00	18.96	17.16	15.68	20.00	19.57	18.77	15.96	18.35	14.44	13.71	17.33	17.57	15.72	15.90	18.90
BioI	18.96	100.00	34.13	29.04	27.51	33.68	30.69	25.07	35.32	24.62	25.65	32.01	23.53	33.76	29.03	24.55
CYP107N1	17.16	34.13	100.00	30.40	31.84	32.30	31.27	26.36	37.44	28.78	27.08	34.24	24.87	33.77	31.95	25.25
CYP122A2	15.68	29.04	30.40	100.00	25.14	25.41	25.41	22.99	30.40	21.26	22.99	28.99	24.07	29.89	27.23	23.70
CYP142	20.00	27.51	31.84	25.14	100.00	31.48	27.03	28.53	30.37	27.16	24.15	30.00	26.55	31.65	24.93	29.61
CYP164A2	19.57	33.68	32.30	25.41	31.48	100.00	29.48	25.70	38.93	24.14	24.23	29.74	23.42	32.56	27.30	26.11
EpoK	18.77	30.69	31.27	25.41	27.03	29.48	100.00	23.41	30.85	25.55	23.14	27.76	20.94	28.83	24.15	25.00
NovI	15.96	25.07	26.36	22.99	28.53	25.70	23.41	100.00	27.11	22.75	22.05	29.20	23.64	23.12	24.80	25.94
PikC	18.35	35.32	37.44	30.40	30.37	38.93	30.85	27.11	100.00	27.94	28.87	35.97	25.26	46.80	29.43	25.00
RhF	14.44	24.62	28.78	21.26	27.16	24.14	25.55	22.75	27.94	100.00	24.44	26.15	21.76	25.76	24.74	27.49
TxtE	13.71	25.65	27.08	22.99	24.15	24.23	23.14	22.05	28.87	24.44	100.00	23.30	18.73	27.39	22.45	22.31
TylH1	17.33	32.01	34.24	28.99	30.00	29.74	27.76	29.20	35.97	26.15	23.30	100.00	24.62	32.02	29.02	26.26
cam	17.57	23.53	24.87	24.07	26.55	23.42	20.94	23.64	25.26	21.76	18.73	24.62	100.00	22.13	23.16	24.36
eryF	15.72	33.76	33.77	29.89	31.65	32.56	28.83	23.12	46.80	25.76	27.39	32.02	22.13	100.00	29.47	22.34
nor	15.90	29.03	31.95	27.23	24.93	27.30	24.15	24.80	29.43	24.74	22.45	29.02	23.16	29.47	100.00	27.23
terp	18.90	24.55	25.25	23.70	29.61	26.11	25.00	25.94	25.00	27.49	22.31	26.26	24.36	22.34	27.23	100.00

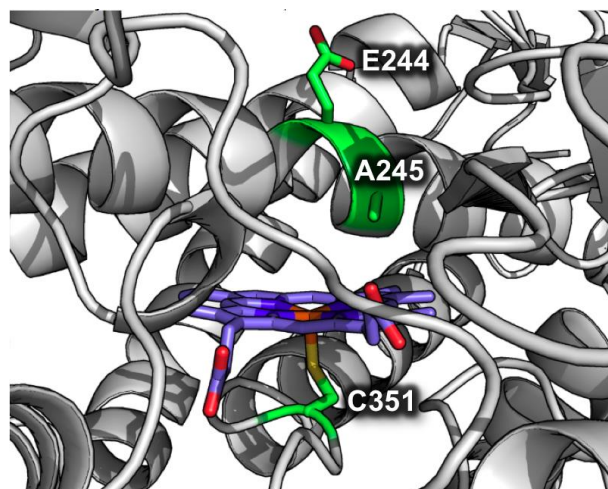
**Figure S2.3.** Active sites of P450s from the library with known structure. The position of the conserved threonine is shown (except for in eryF where this residue is alanine). In the structures of the P450s that perform monooxygenation, a conserved glutamate/aspartate precedes the threonine. P450<sub>TxtE</sub> and P450<sub>nor</sub>, which do not perform monooxygenation, have unusual residues in this position.



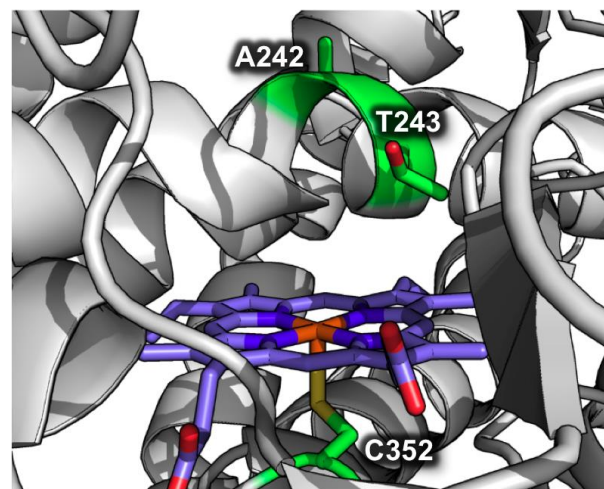
P450<sub>BM3</sub>, PDB: 2IJ2



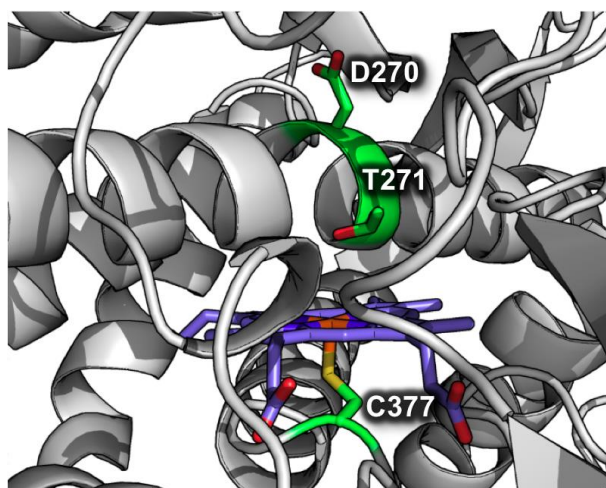
P450<sub>cam</sub>, PDB: 1PHC



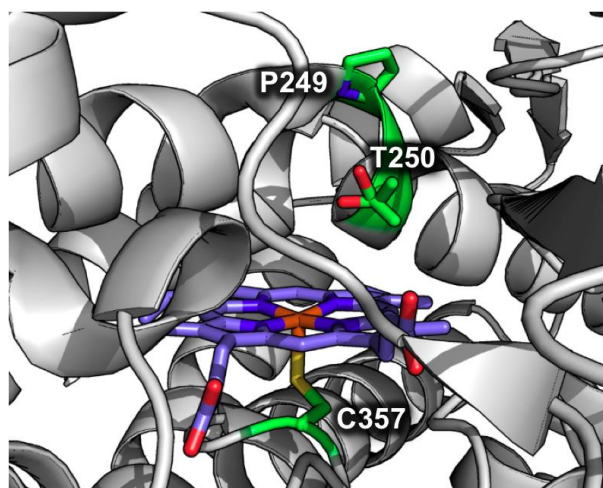
P450<sub>eryF</sub>, PDB: 1OXA



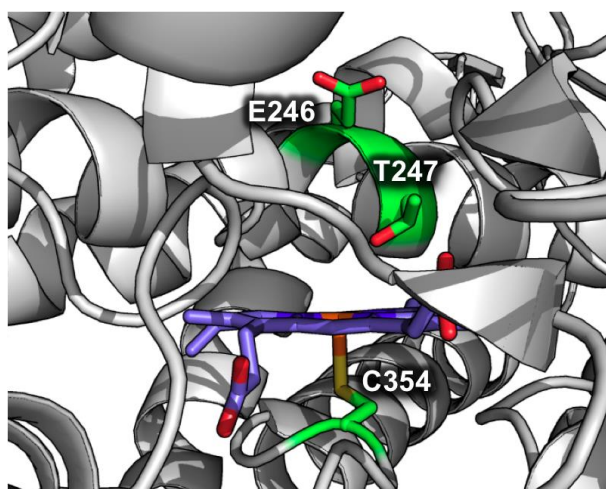
P450<sub>nor</sub>, PDB: 1XQD



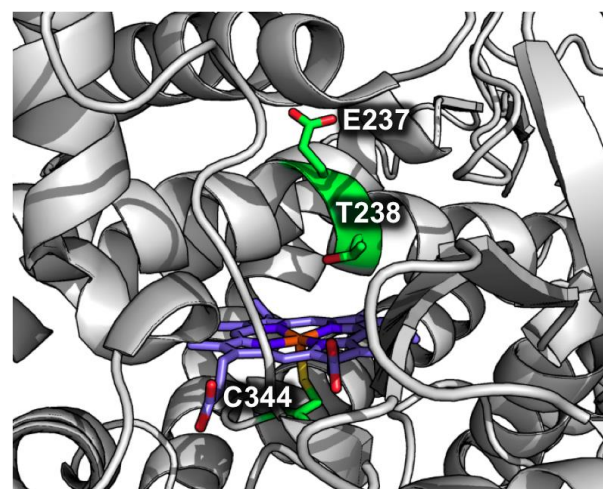
P450<sub>terr</sub>, PDB: 1CPT



P450<sub>TxtE</sub>, PDB: 4TPN

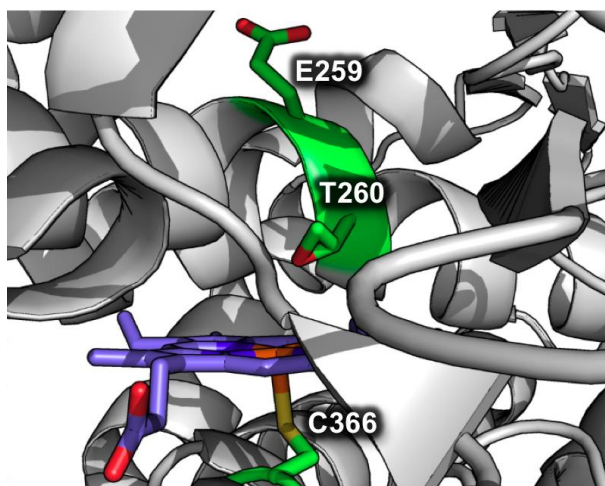


P450<sub>PikC</sub>, PDB: 2BVJ

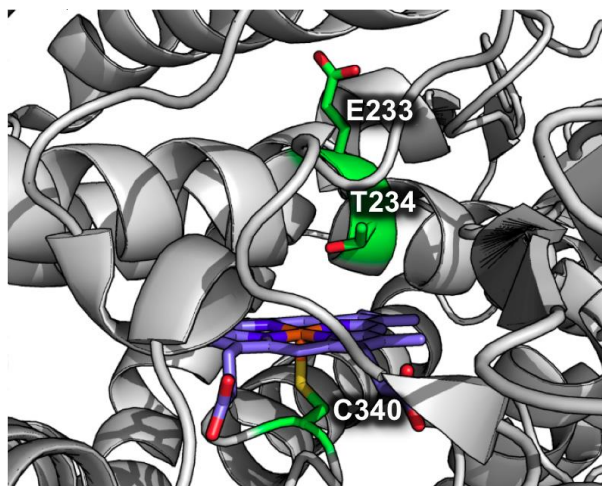


P450<sub>Biol</sub>, PDB: 3EJB

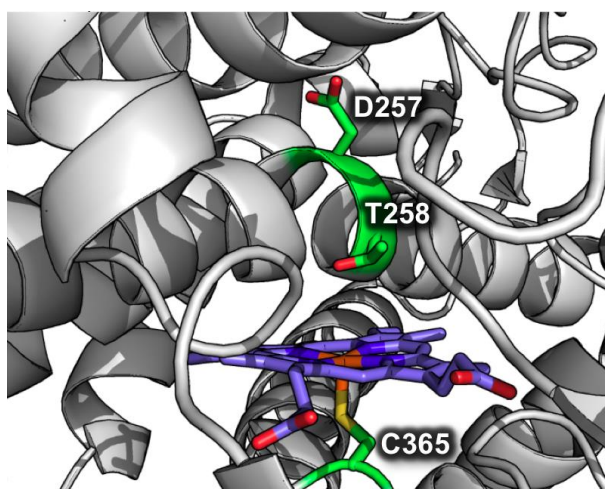




CYP164A2, PDB: 3R9B

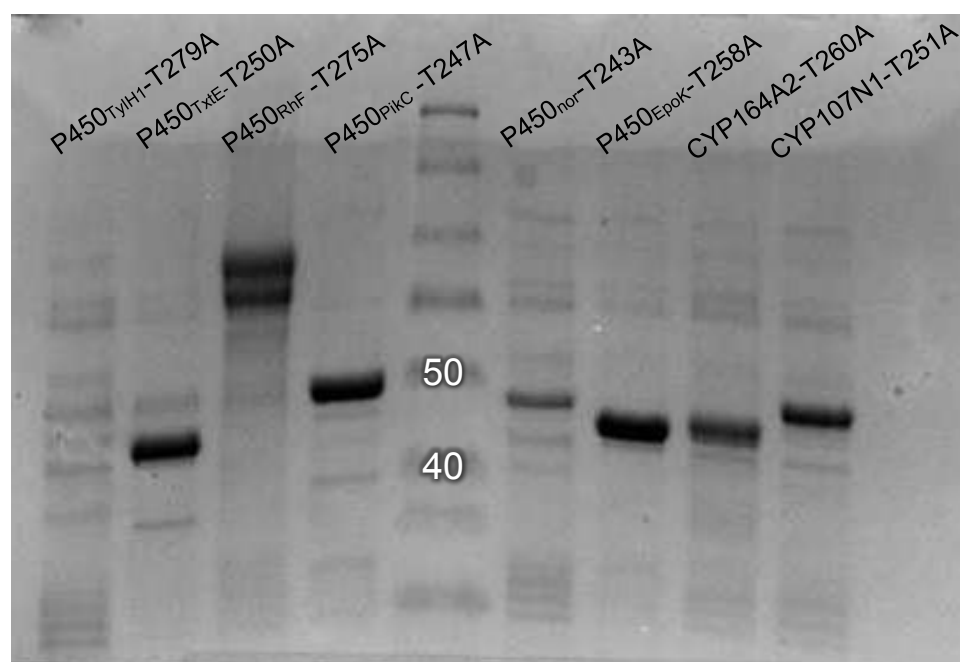
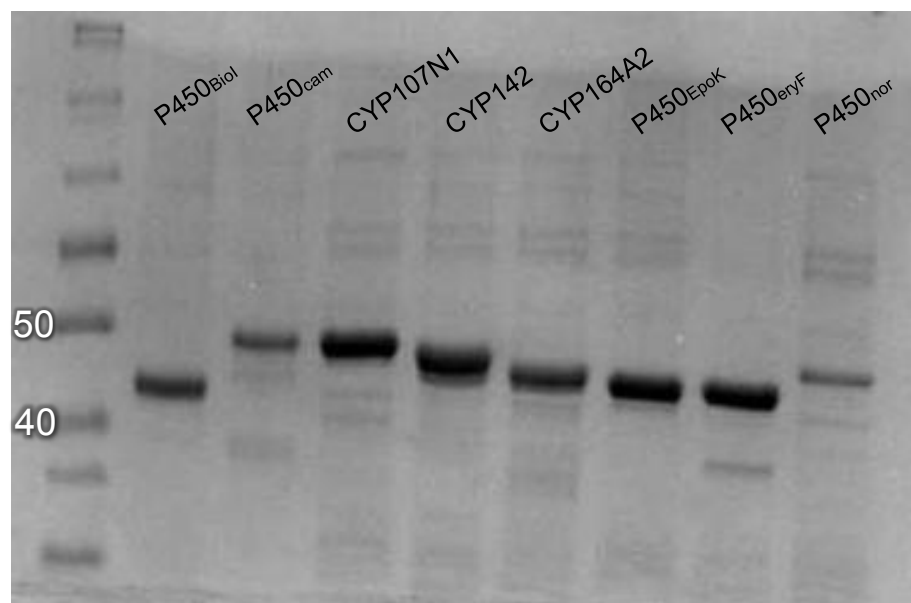


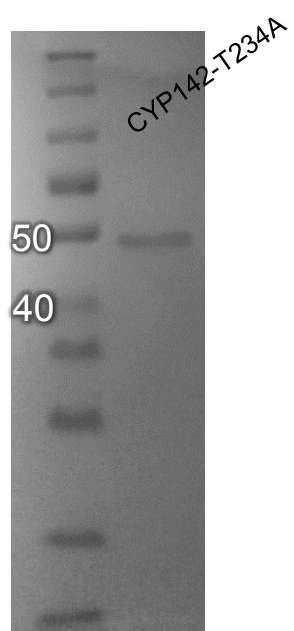
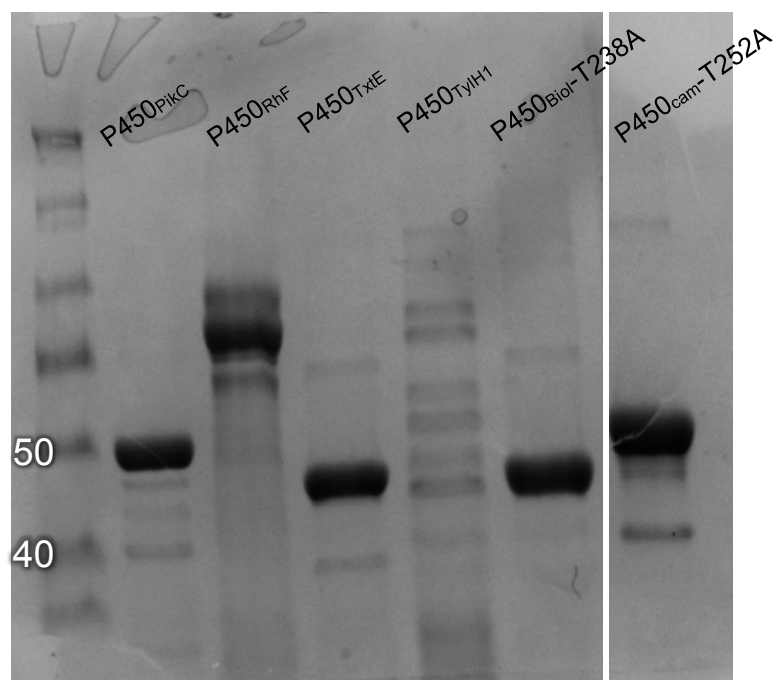
CYP142, PDB: 2XKR



P450<sub>EpoK</sub>, PDB: 1Q5E

**Figure S2.4.** SDS-polyacrylamide gel electrophoresis analysis of his-tag purified P450 variants. The 40 kDa and 50 kDa protein markers are indicated with white text in each gel image.



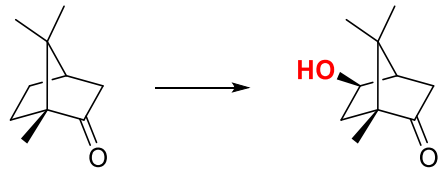
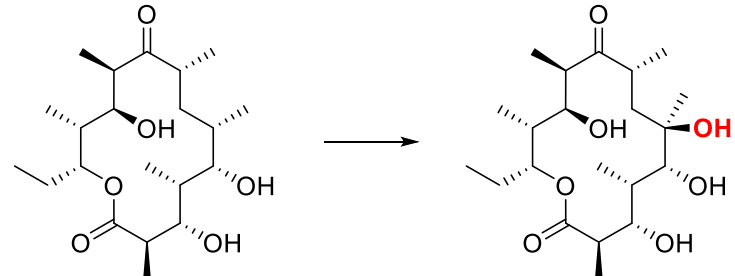
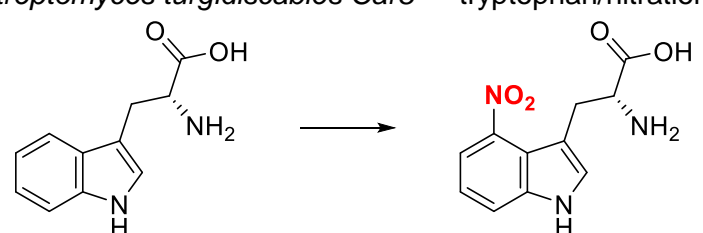
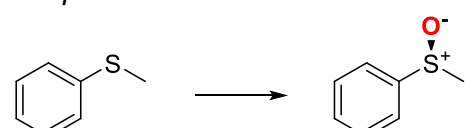




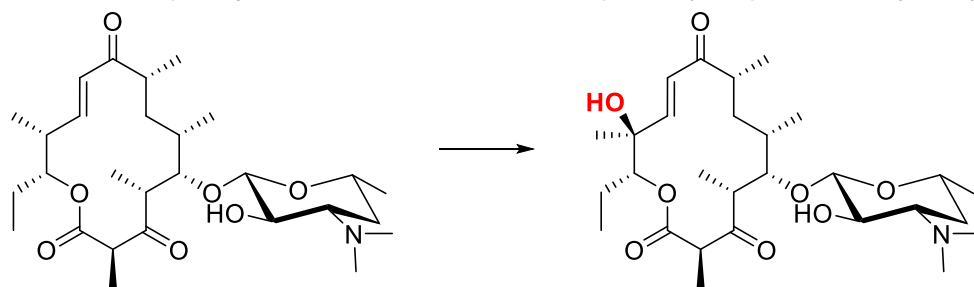
**Table S2.1.** % Identity to P450<sub>BM3</sub>, representative PDB accession numbers (PDB#), and expression level in *E. coli* of each P450 library member.<sup>27,29-44</sup>

Name	% Identity	PDB#	Expression level in <i>E. coli</i>
P450 <sub>cam</sub>	13.25	1PHC	High, mostly folded P450
P450 <sub>eryF</sub>	12.87	1OXA	High, mostly folded P450
P450 <sub>nor</sub>	14.39	1XQD	Low, mostly folded P450
P450 <sub>terp</sub>	15.65	1CPT	Low, mostly misligated P420**
P450 <sub>TxtE</sub>	11.33	4TPN	High, mostly folded P450
P450 <sub>RhF</sub> *	12.42	N/A	High, some misligated P420**
P450 <sub>PikC</sub>	17.79	2BVJ	High, mostly folded P450
P450 <sub>Biol</sub>	16.46	3EJB	High, mostly folded P450
CYP164A2	16.51	3R9B	High, mostly folded P450
CYP107N1	14.39	N/A	High, mostly folded P450
CYP122A2	12.44	N/A	None
CYP142	16.33	2XKR	High, mostly folded P450
P450 <sub>TylH1</sub>	12.39	N/A	Low, some misligated P420
P450 <sub>NovI</sub>	11.30	N/A	None
P450 <sub>EpoK</sub>	16.47	1Q5E	High, mostly misligated P420**
*P450-reductase fusion			
**P420, shown by the presence of an absorbance maximum at 420 nm for the carbon monoxide-bound ferrous enzyme, indicates a misligated heme state and possible protein misfolding. See Figure S1 for CO-bound absorbance traces. <sup>45</sup>			

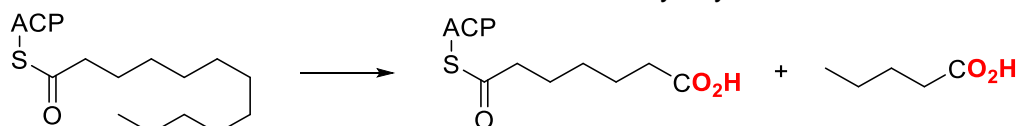
**Table S2.2.** Representative substrates and reactions for P450 library members.<sup>29,33,35-44</sup>

Name	Organism	Substrate/ Reaction
P450 <sub>cam</sub>	<i>Pseudomonas putida</i>	camphor/hydroxylation
		
P450 <sub>eryF</sub>	<i>Saccharopolyspora erythraea</i>	6-deoxyerythronolide B/hydroxylation
		
P450 <sub>nor</sub>	<i>Fusarium oxysporum</i>	nitric oxide and NADH/reduction
	$2 \text{ NO} + \text{NADH} + \text{H}^+ \longrightarrow \text{N}_2\text{O} + \text{H}_2\text{O} + \text{NAD}^+$	
P450 <sub>TxtE</sub>	<i>Streptomyces turgidiscabies</i> Car8	tryptophan/nitration
		
P450 <sub>RhF</sub>	<i>Rhodococcus</i> sp. NCIMB 9784	unknown/various oxidations
		

P450<sub>PikC</sub>      *Streptomyces venezuelae*      pikromycin precursor/hydroxylation



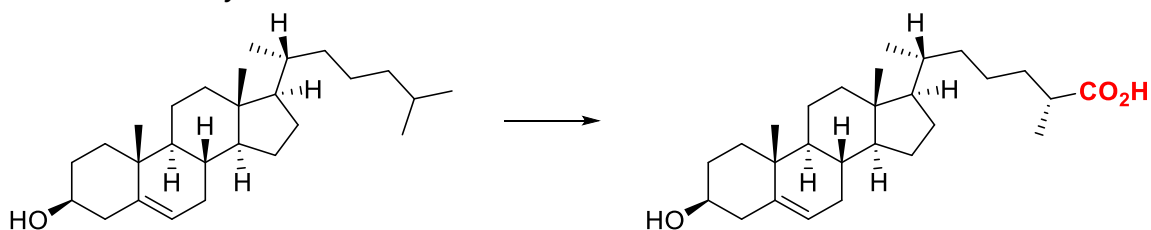
P450<sub>Biol</sub>      *Bacillus subtilis*      fatty acyl-ACP/oxidative cleavage



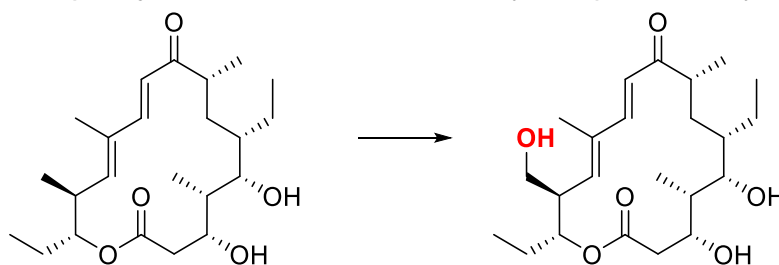
CYP164A2      *Mycobacterium smegmatis*      fatty acid oxidation (hypothesized)

CYP107N1      *Streptomyces lavendulae*      mitomycin precursor/oxidation  
(not fully characterized)

CYP142      *Mycobacterium tuberculosis*      cholesterol/oxidation



P450<sub>TylH1</sub>      *Streptomyces fradiae*      tylsoin precursor/hydroxylation

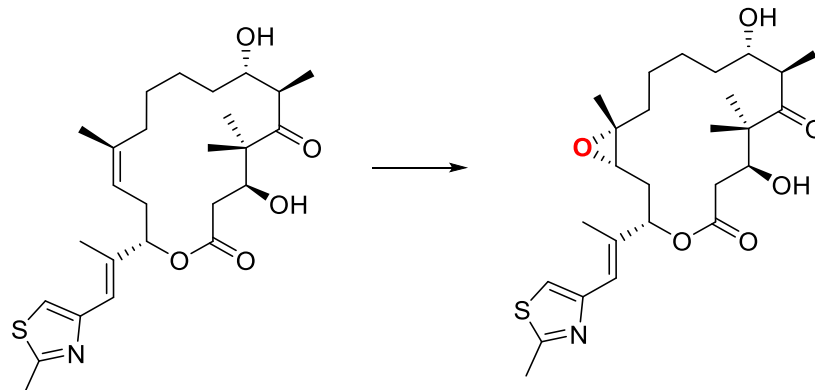


---

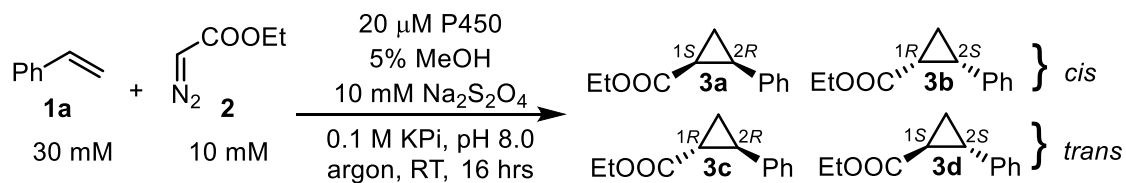
P450<sub>EpoK</sub>

*Sorangium cellulosum*

epothilone precursor/epoxidation

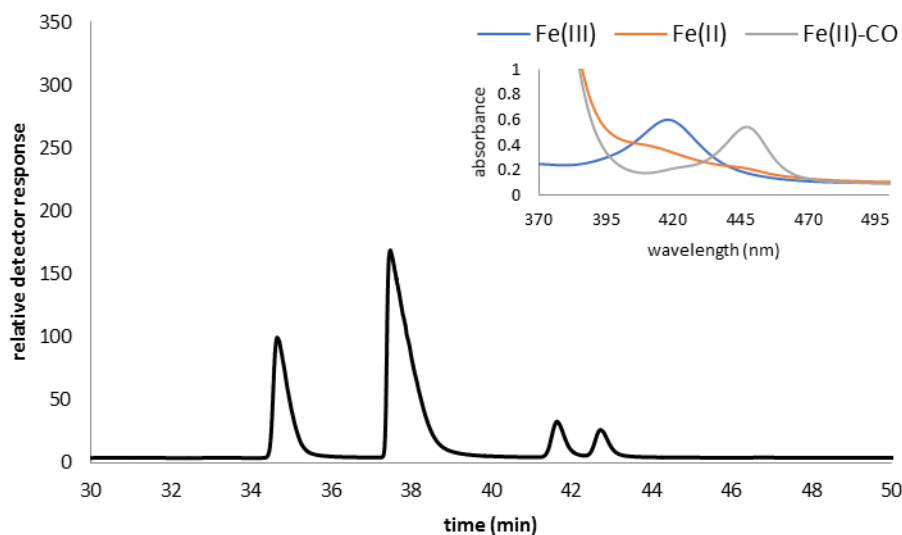


**Figure S2.5.** Chiral GC traces from enzymatic cyclopropanations of styrene with EDA.

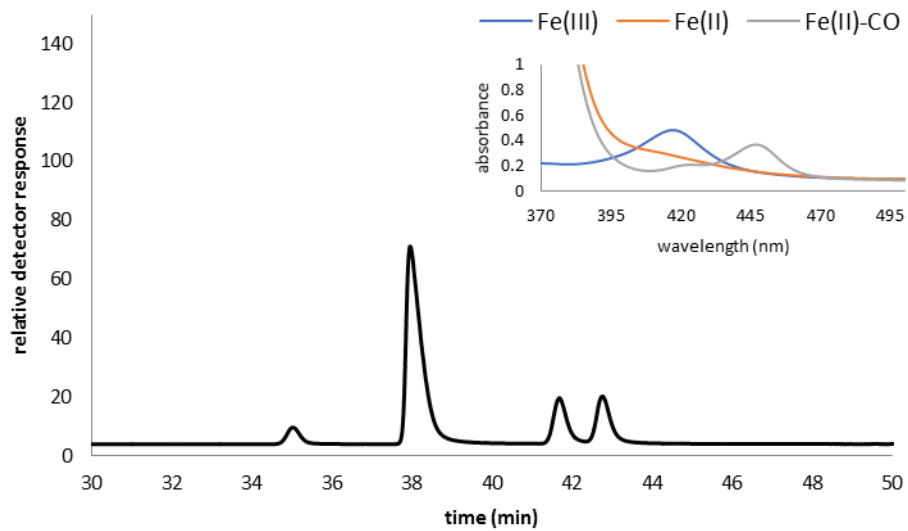


Constant pressure; GC Oven temperature = 100 °C for 30 min, 1 °C/min to 135 °C, 135 for 10 min, 10 °C/min to 200 °C, 200 °C for 5 min. Inset: UV/vis spectra for ferric P450 (blue), ferrous P450 (orange), and ferrous-carbon monoxide-bound P450 (gray). The ferrous-carbon monoxide-bound spectra show qualitative compositions of folded (450 nm peak) and misligated (420 nm peak) protein. Elution times: **3a** (35.0min), **3b** (37.5 min), **3c** (41.6 min), and **3d** (42.7 min). Variable dilutions and injection volumes were used to attempt separation of cyclopropane isomers; product peaks can be compared within the same chromatogram but not between chromatograms.

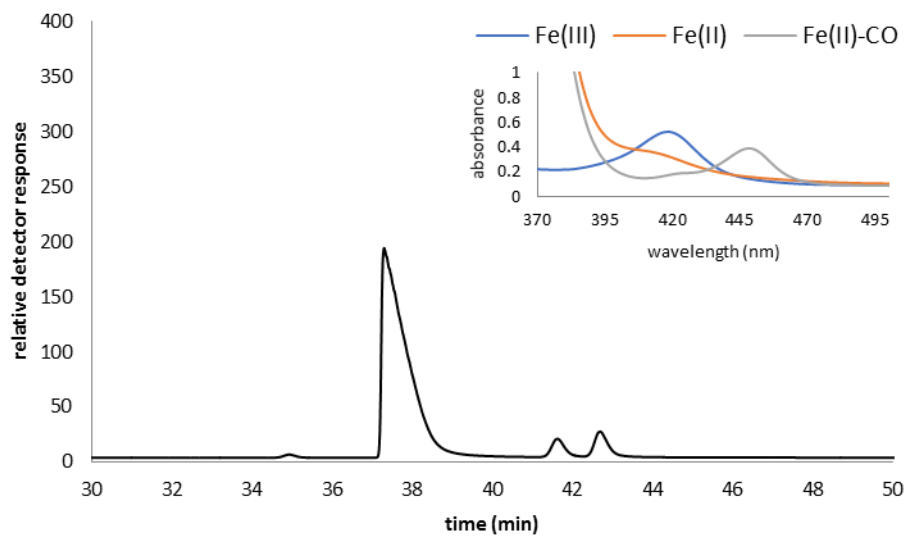
#### P450<sub>cam</sub>



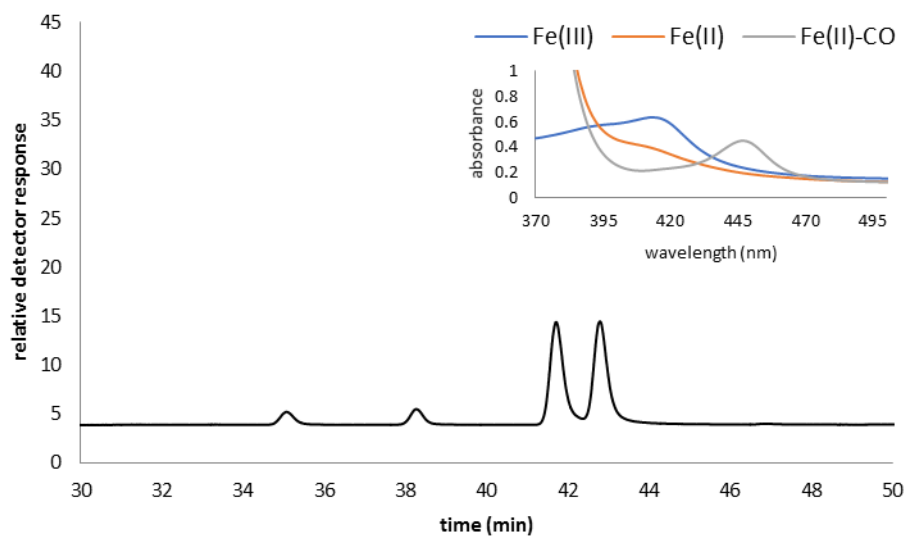
### P450<sub>cam</sub>-T252A



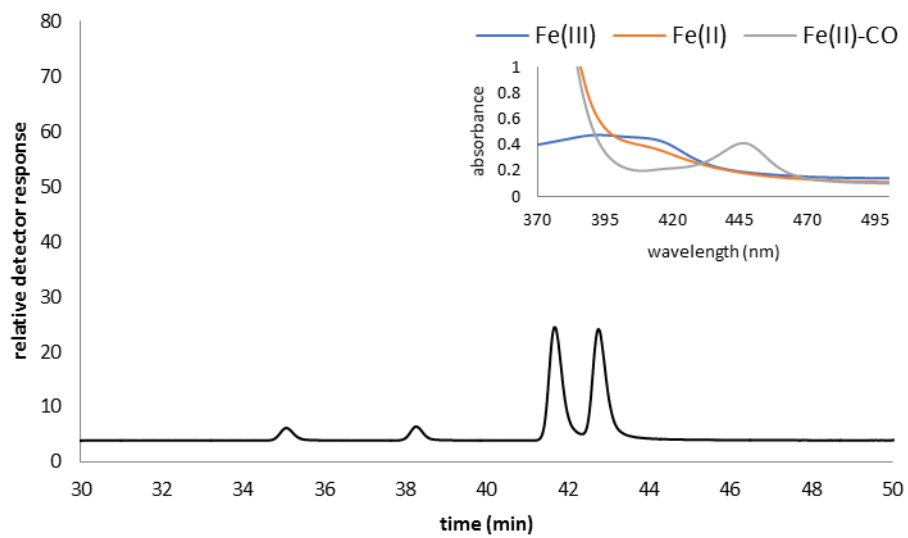
### P450<sub>eryF</sub>



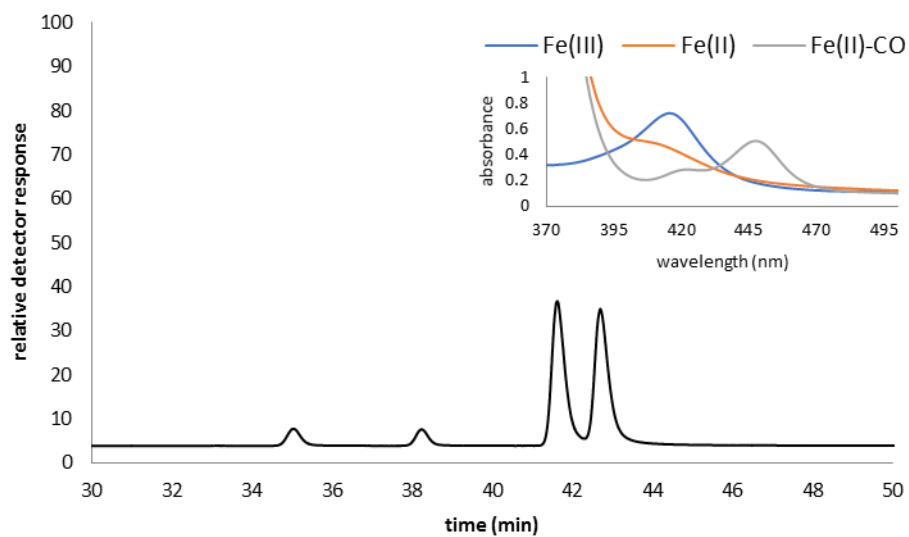
### P450<sub>nor</sub>



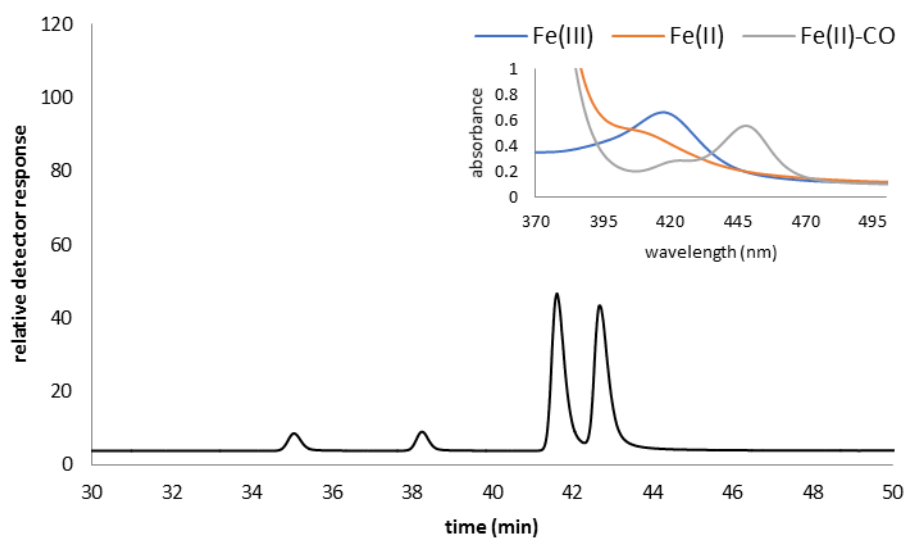
### P450<sub>nor</sub>-T243A



### P450<sub>TxtE</sub>

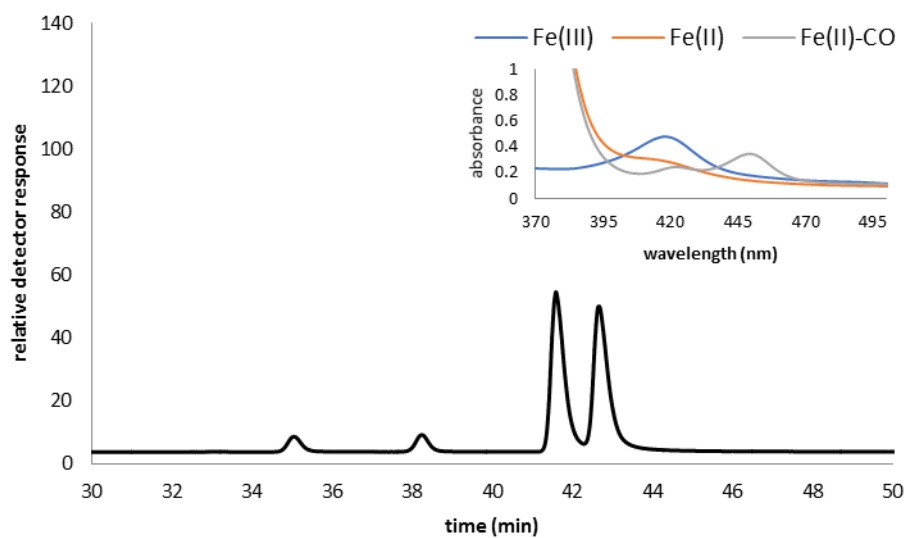


### P450<sub>TxtE</sub>-T250A

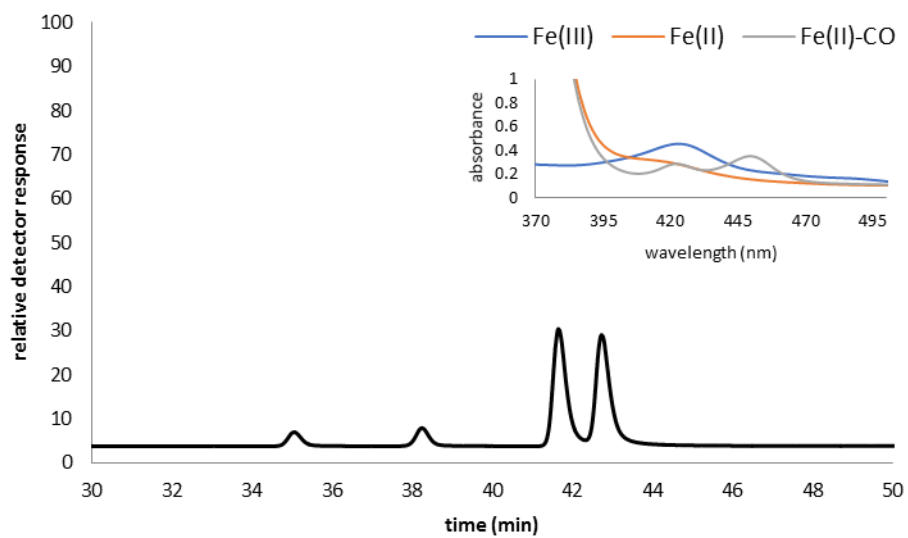


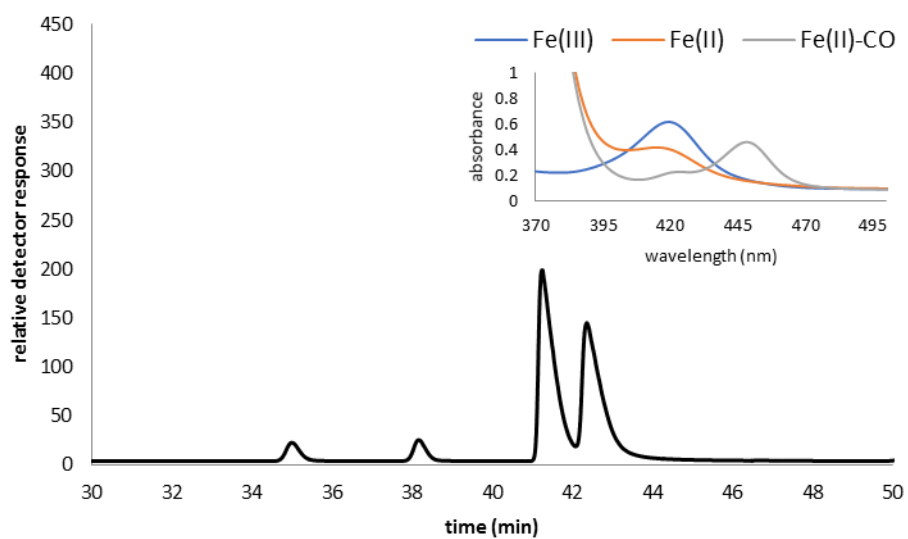
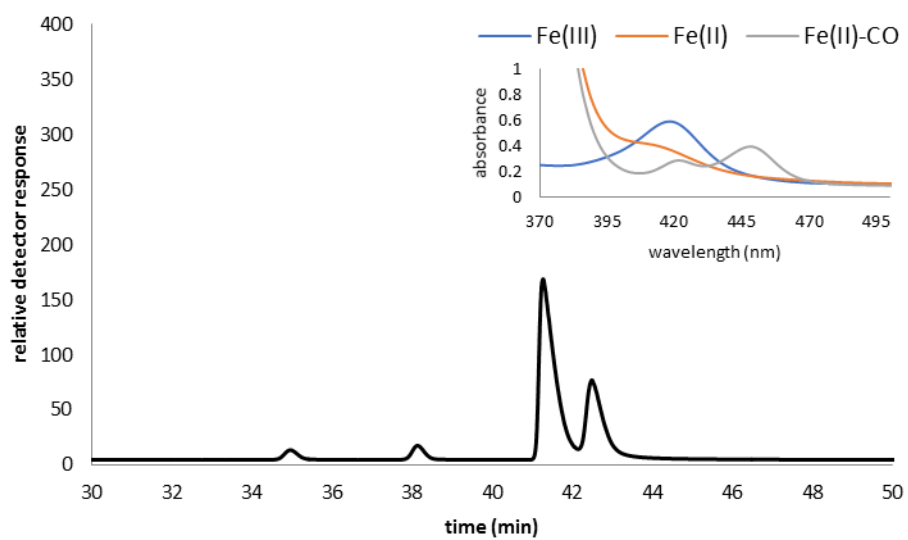


### P450<sub>RhF</sub>

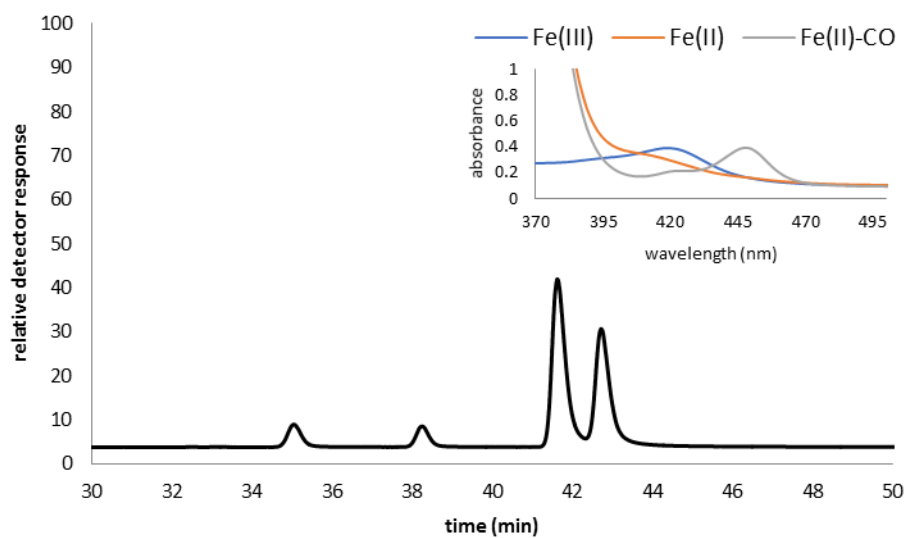


### P450<sub>RhF</sub>-T275A

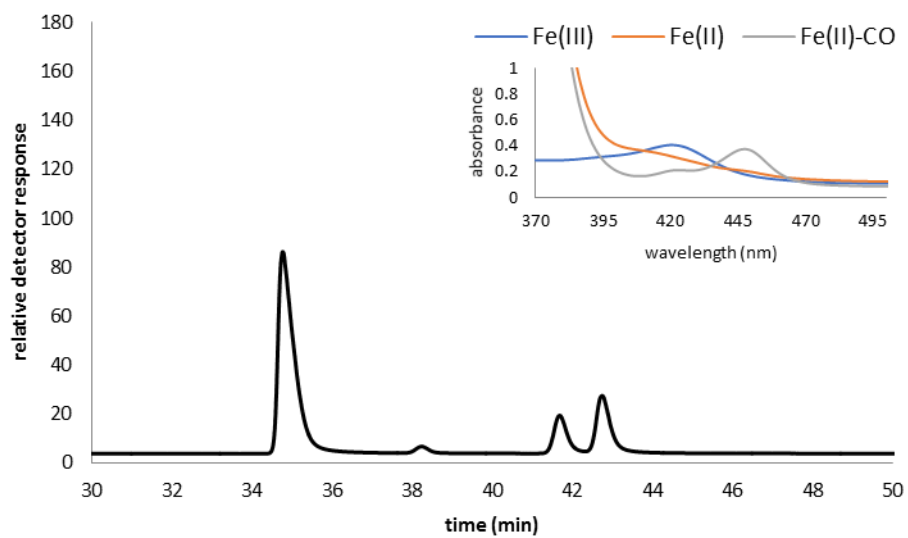


**P450<sub>PikC</sub>****P450<sub>PikC</sub>-T247A**

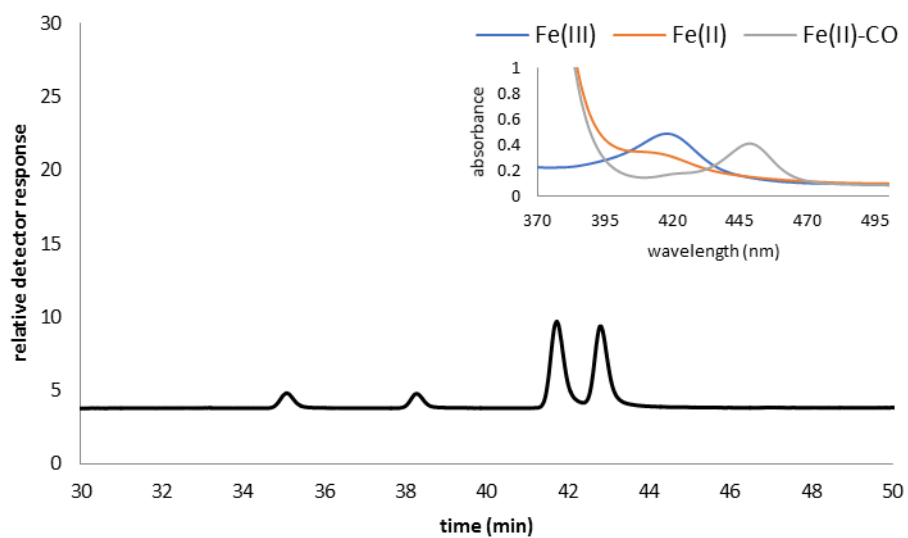
### P450<sub>Biol</sub>



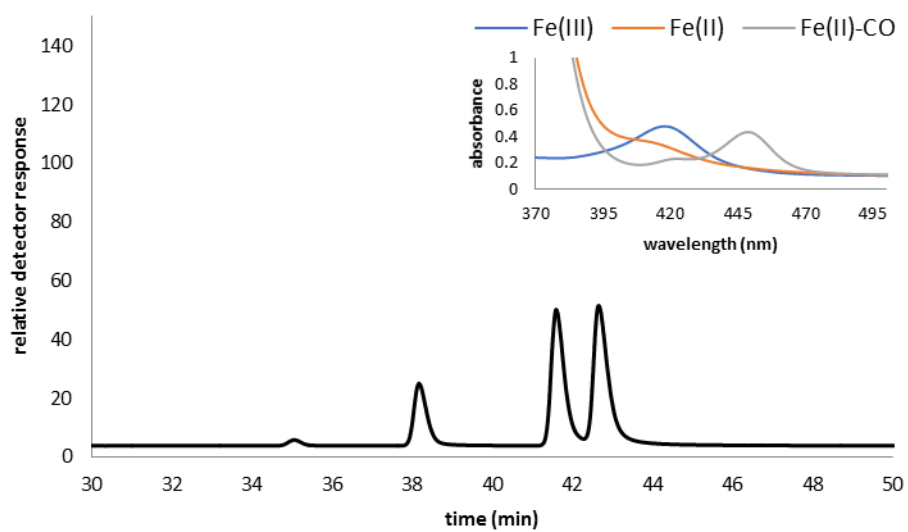
### P450<sub>Biol</sub>-T238A



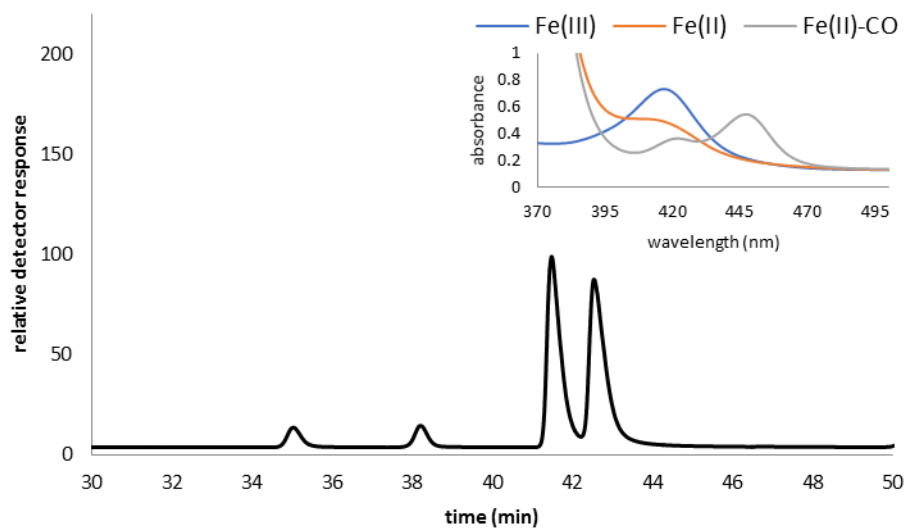
## CYP164A2



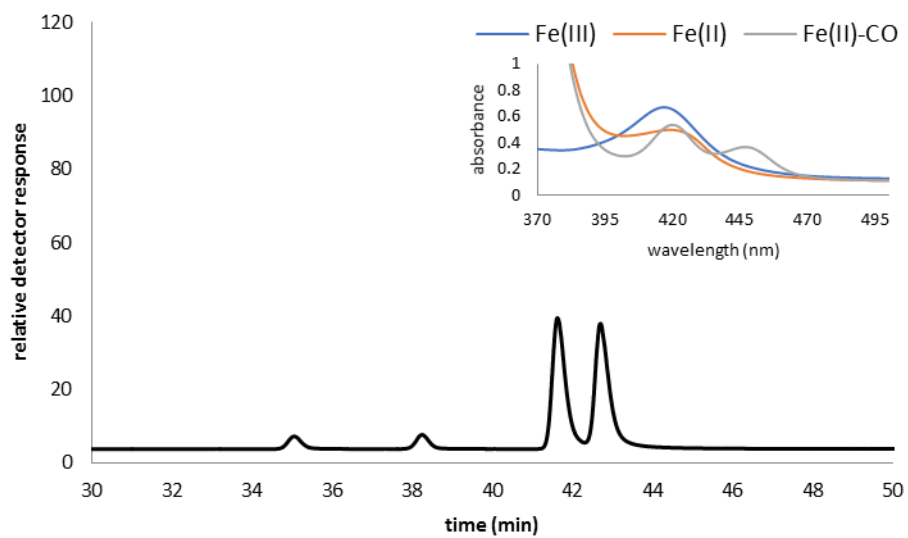
## CYP164A2-T260A



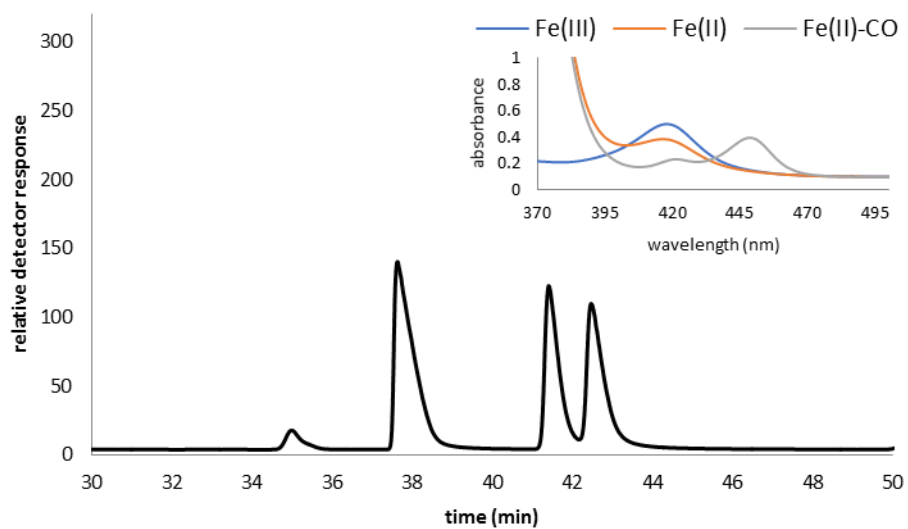
## CYP107N1



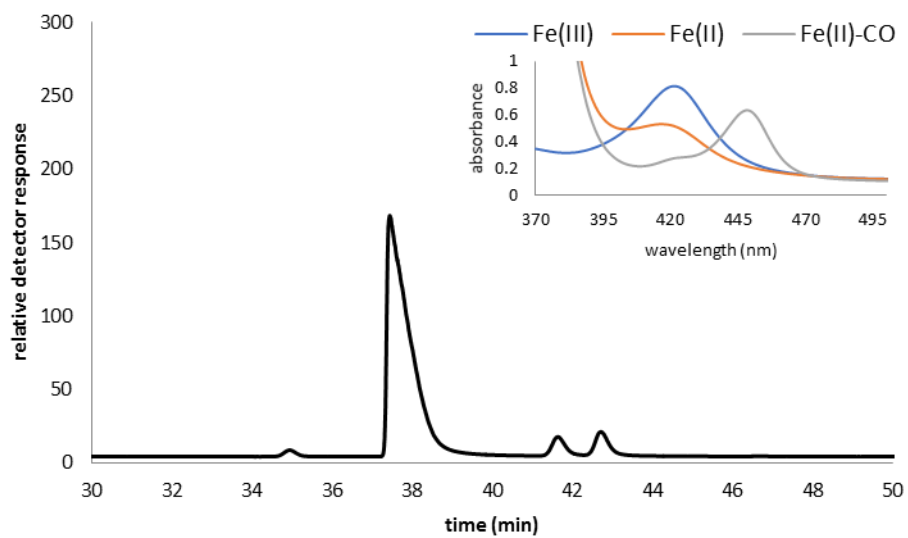
## CYP107N1-T251A



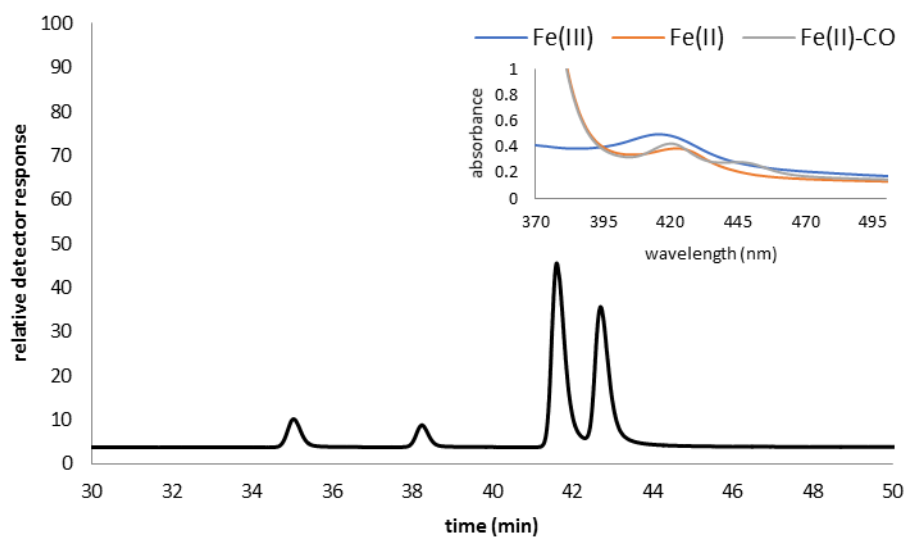
## CYP142



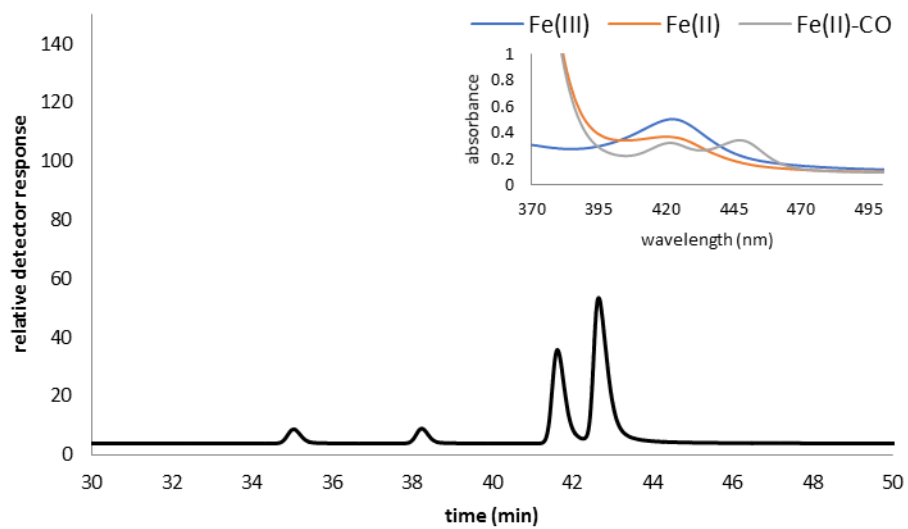
## CYP142-T234A



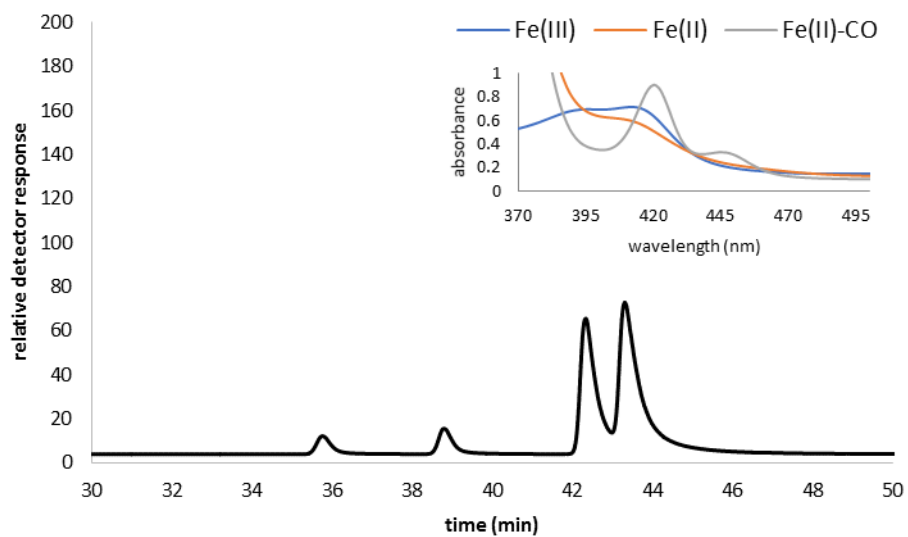
**P450<sub>TyH1</sub>**



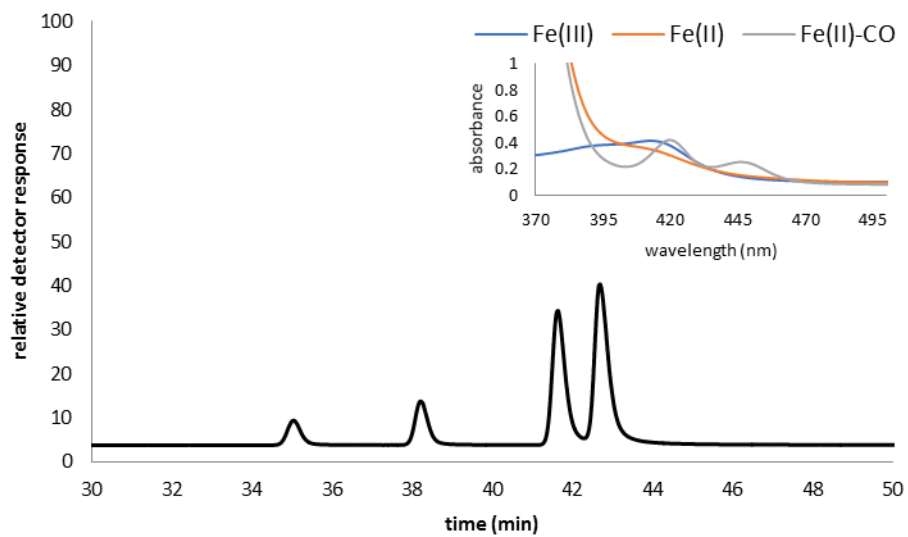
**P450<sub>TyH1</sub>-T279A**



### P450<sub>EpoK</sub>

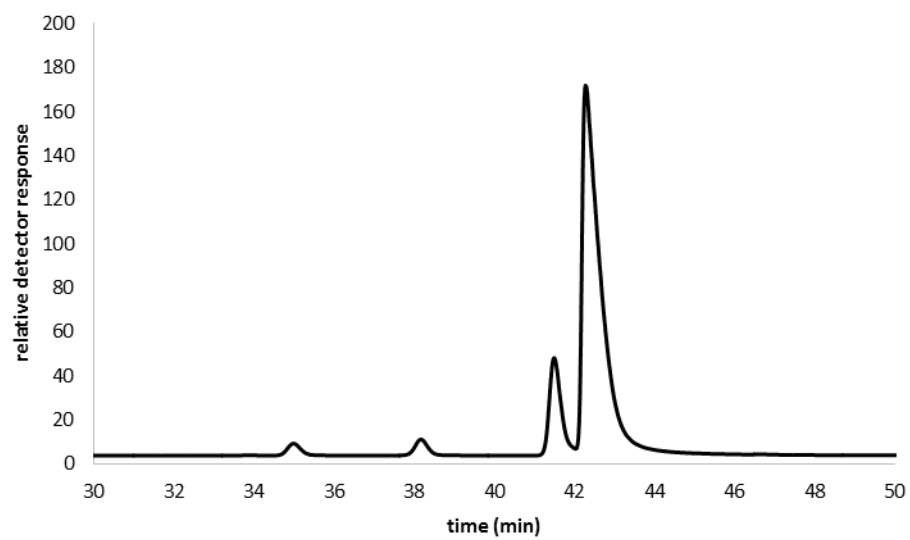


### P450<sub>EpoK</sub>-T258A

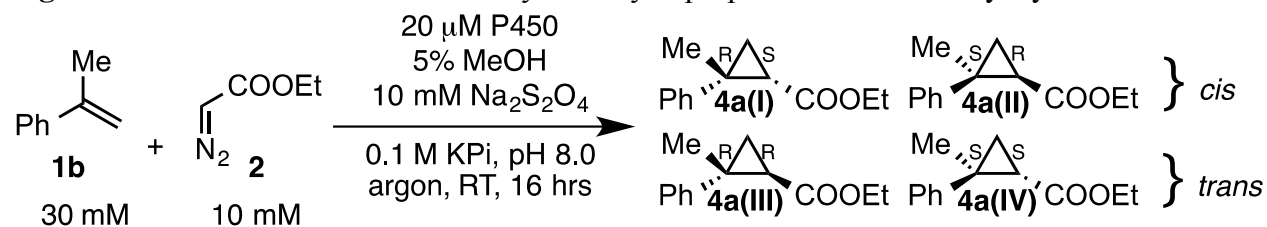




**P450<sub>BM3</sub>-T268V**

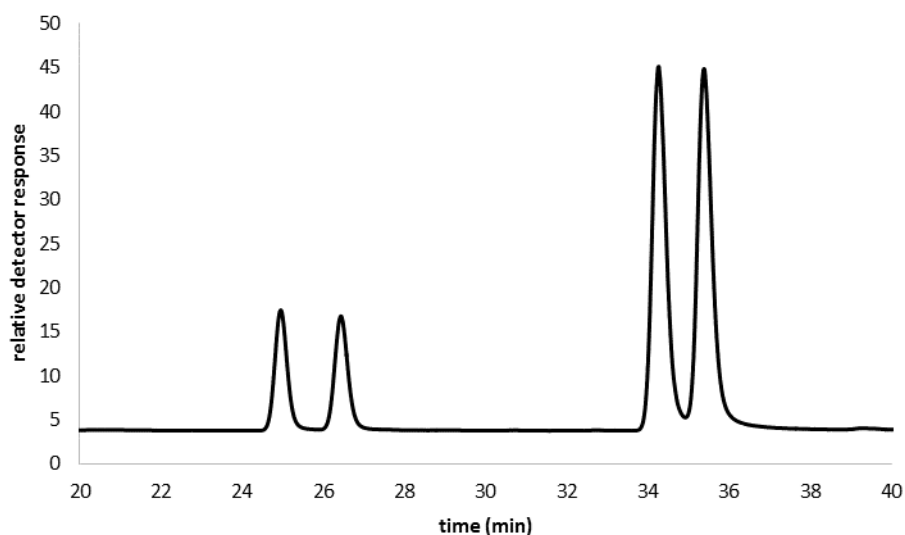


**Figure S2.6.** Chiral GC traces from enzymatic cyclopropanations of  $\alpha$ -methylstyrene with EDA.

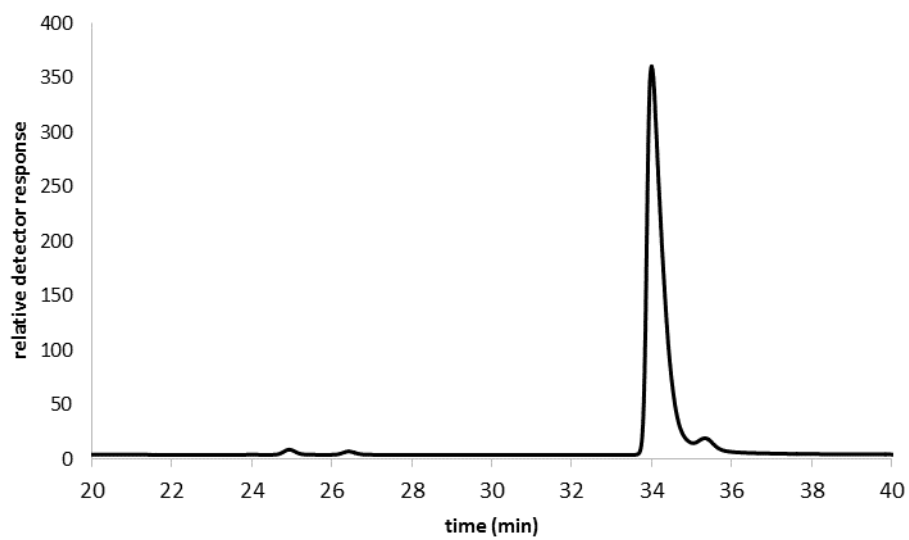


Constant pressure; GC Oven temperature = 100 °C for 30 min, 1 °C/min to 135 °C, 135 for 10 min, 10 °C/min to 200 °C, 200 °C for 5 min. Elution times: *Z* isomers (25.0 min, 26.5 min); *E* isomers (34.3 min, 35.4 min). Variable dilutions and injection volumes were used to attempt separation of cyclopropane isomers; product peaks can be compared within the same chromatogram but not between chromatograms.

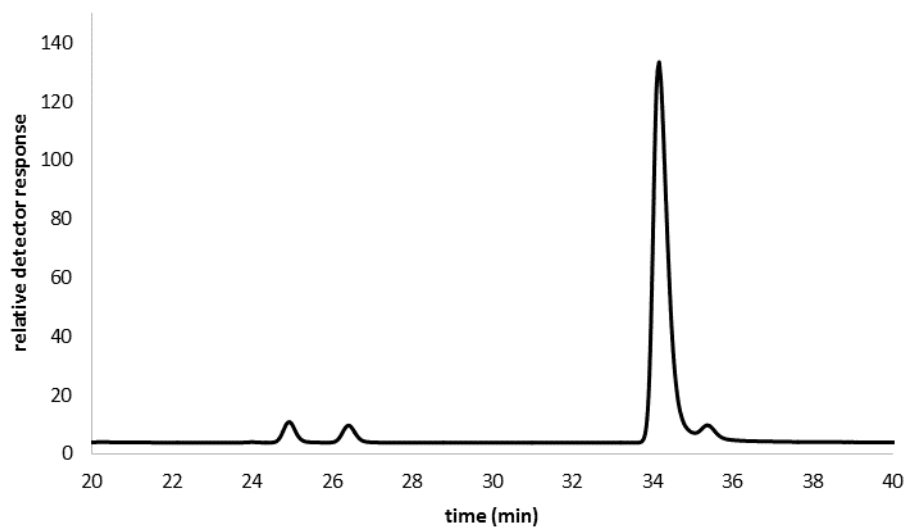
#### FePP



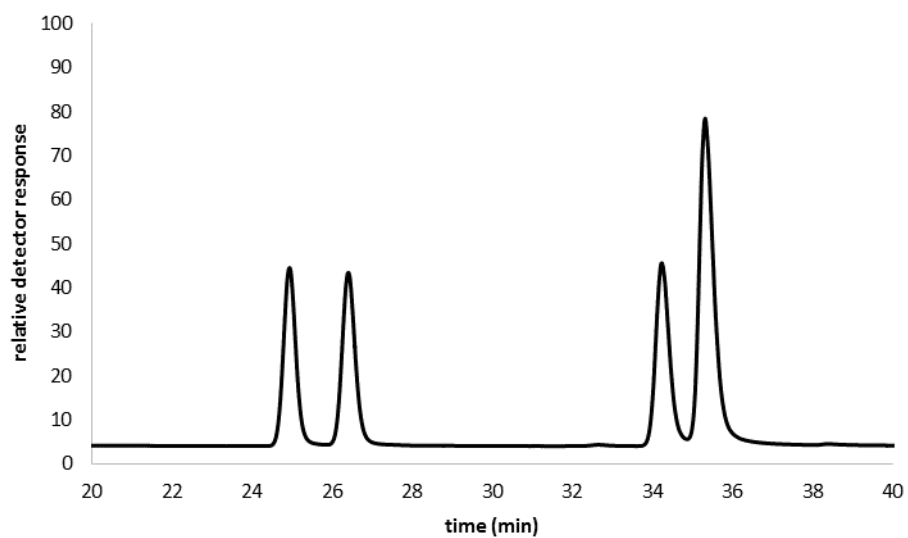
**P450<sub>BM3</sub>-T268A**



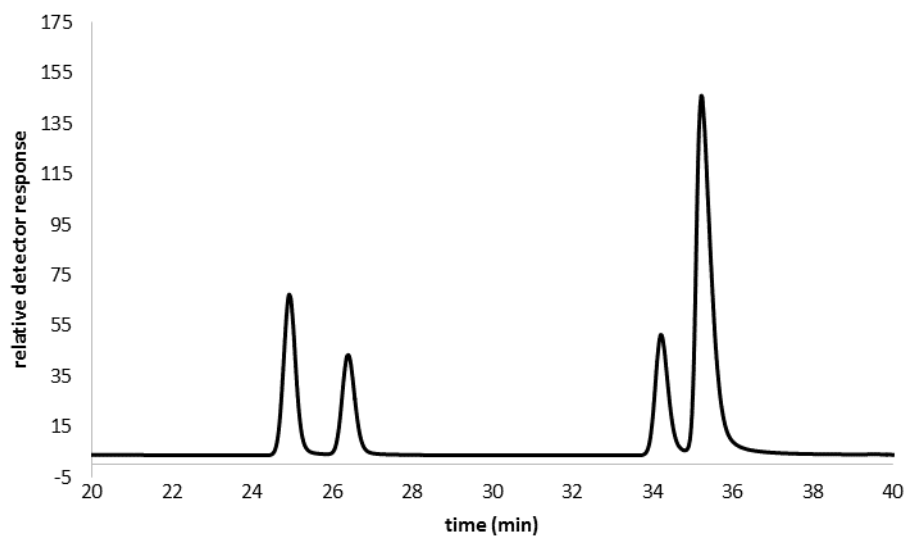
**P450<sub>BM3</sub>-CIS-T438S**



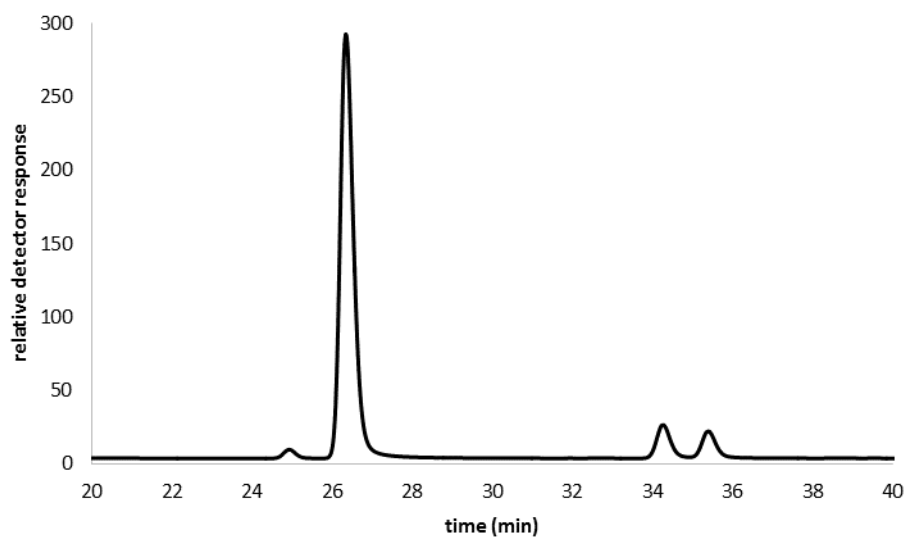
**P450<sub>eryF</sub>**



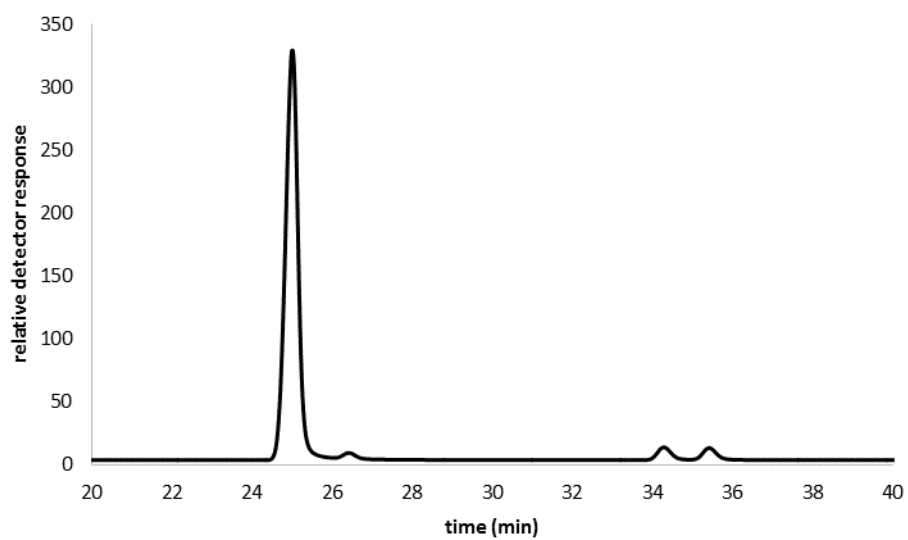
**P450<sub>PikC</sub>-T247A**



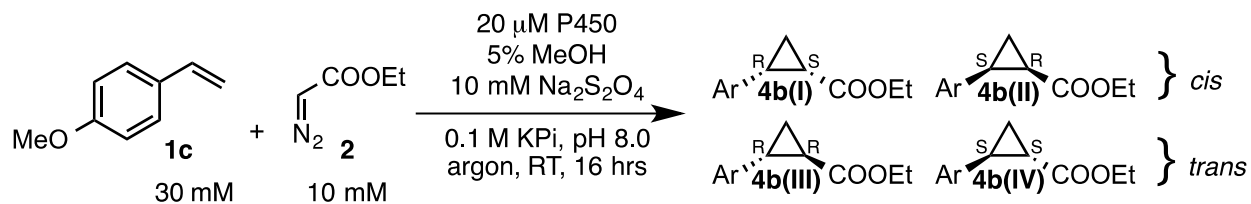
**P450<sub>Biol</sub>-T238A**



**CYP142-T234A**

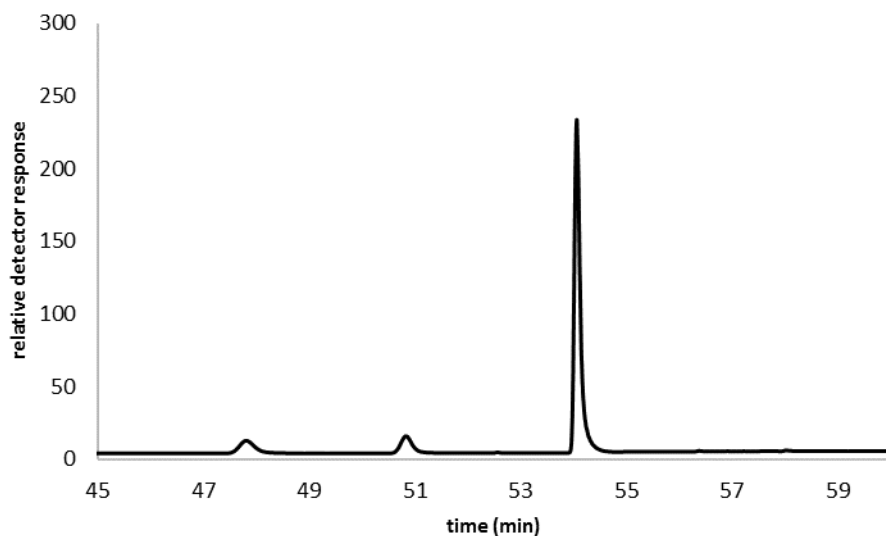


**Figure S2.7.** Chiral GC traces from enzymatic cyclopropanations of 4-methoxystyrene with EDA.

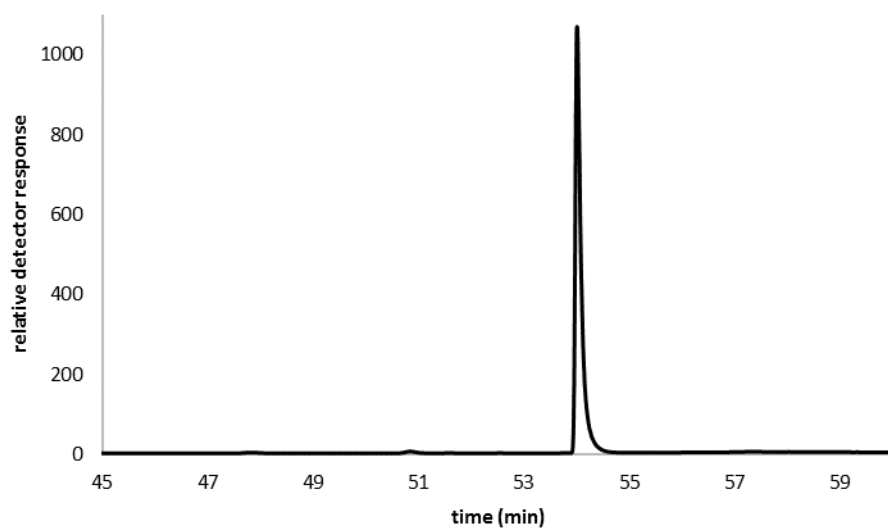


Constant pressure; GC Oven temperature = 100 °C for 5 min, 1 °C/min to 135 °C, 135 for 10 min, 10 °C/min to 200 °C, 200 °C for 5 min. Elution times: *Z* isomers (47.9 min, 50.9 min); *E* isomers (54.1 min). Variable dilutions and injection volumes were used to attempt separation of cyclopropane isomers; product peaks can be compared within the same chromatogram but not between chromatograms.

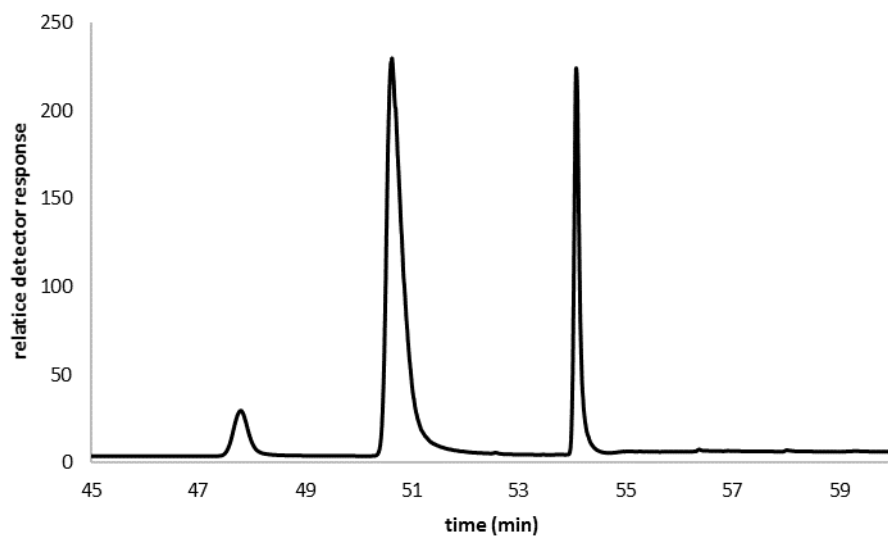
#### FePP



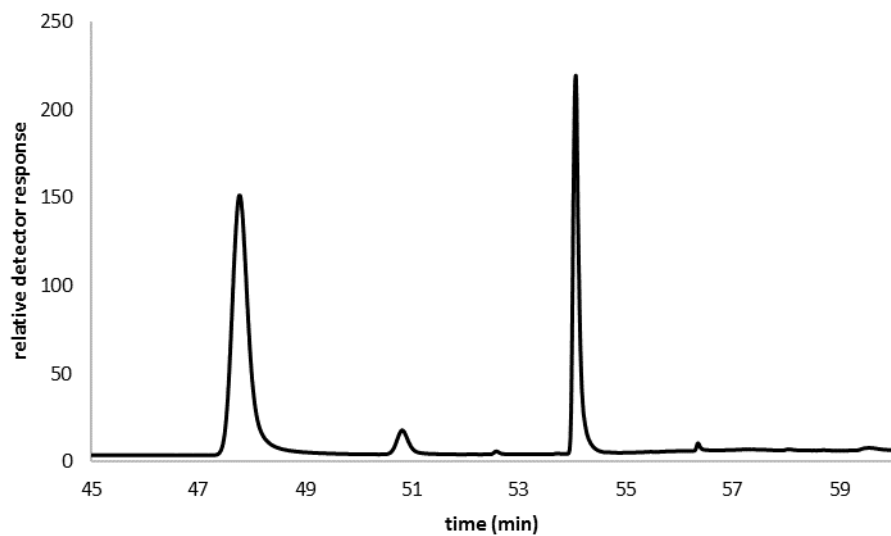
**P450<sub>BM3</sub>-T268A**



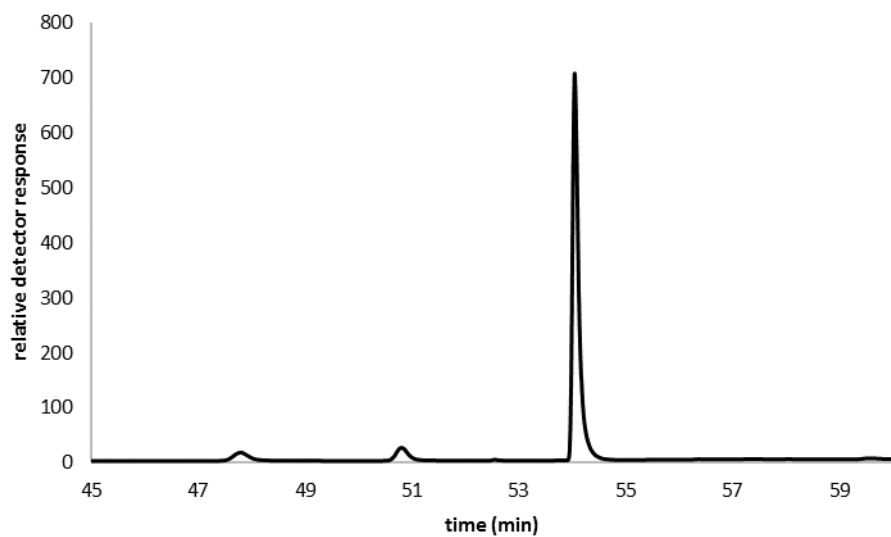
**P450<sub>BM3-CIS</sub>-T438S**



**P450<sub>eryF</sub>**

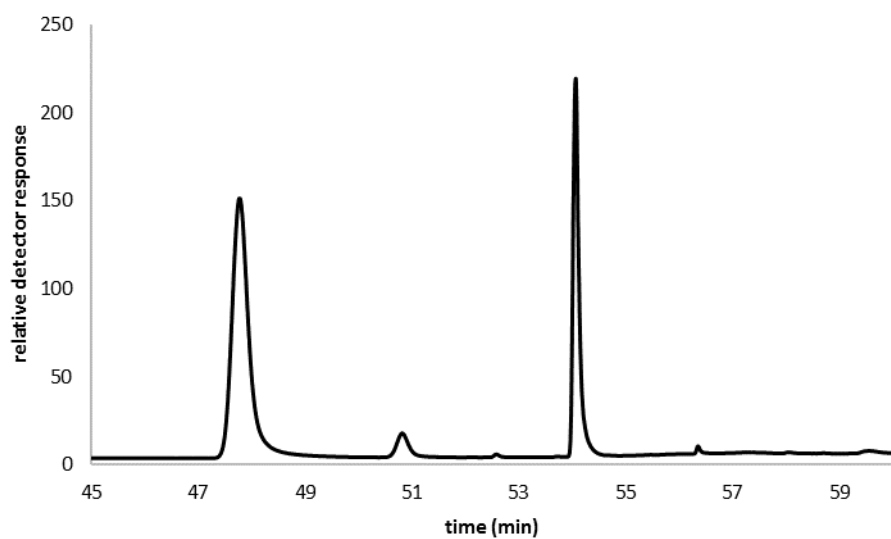


**P450<sub>PikC</sub>-T247A**

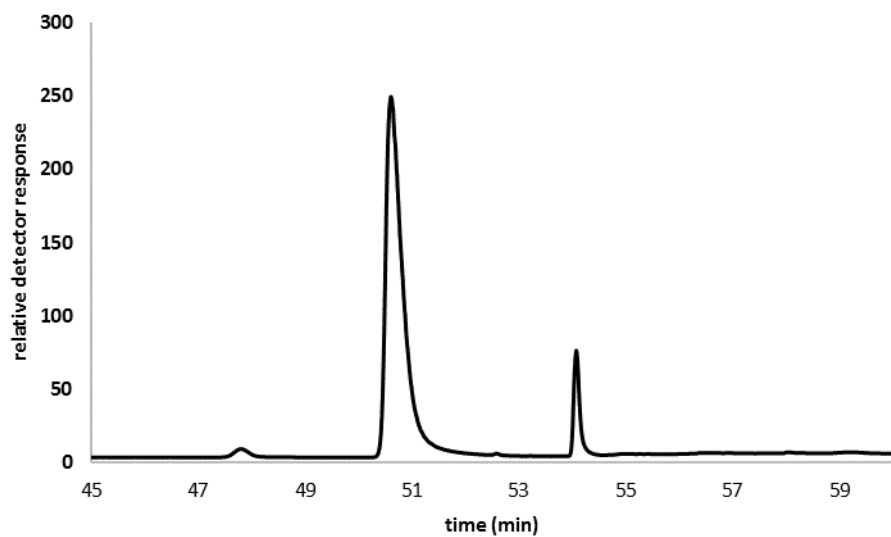




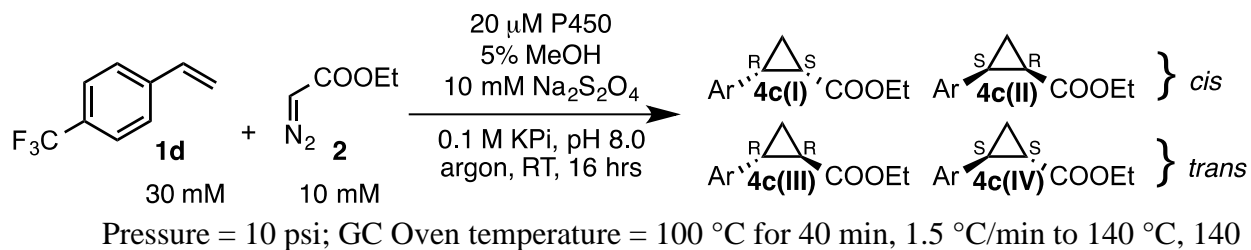
**P450<sub>Biol</sub>-T238A**



**CYP142-T234A**

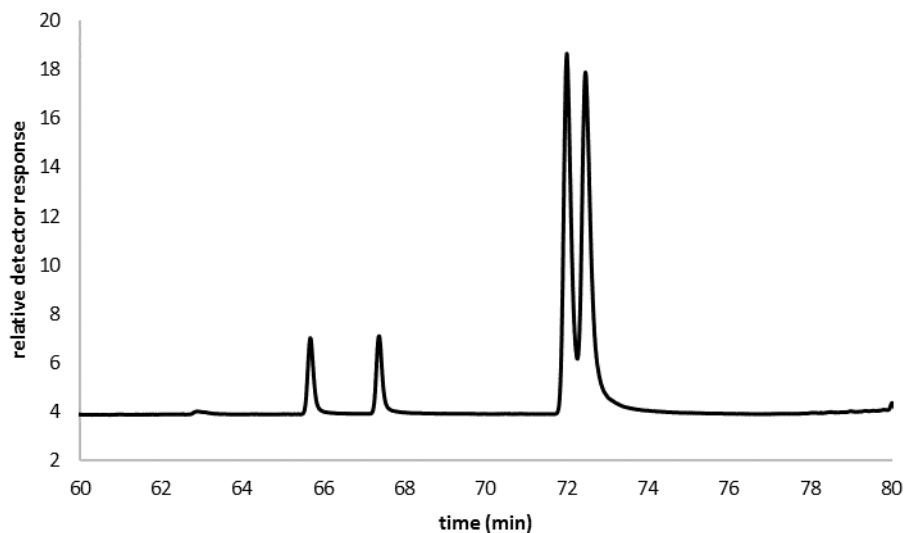


**Figure S2.8.** Chiral GC traces from enzymatic cyclopropanations of 4-(trifluoromethyl)styrene with EDA.

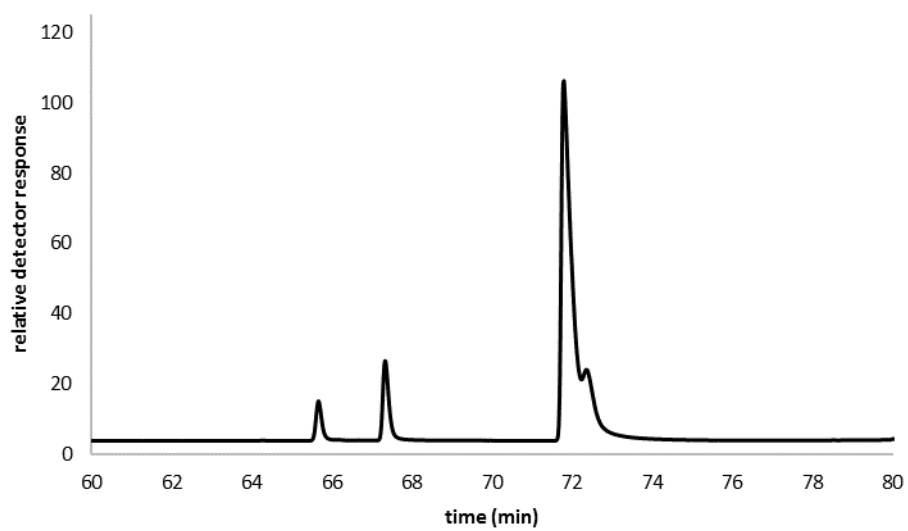


for 10 min, 10 °C/min to 200 °C, 200 °C for 5 min. Elution times: *Z* isomers (65.7 min, 67.4 min); *E* isomers (72.0 min, 72.4 min). Variable dilutions and injection volumes were used to attempt separation of cyclopropane isomers; product peaks can be compared within the same chromatogram but not between chromatograms.

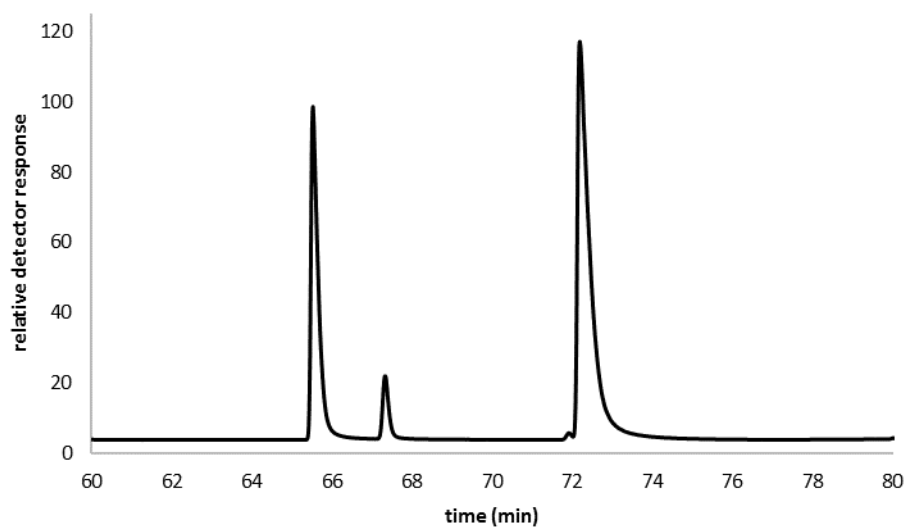
#### FePP



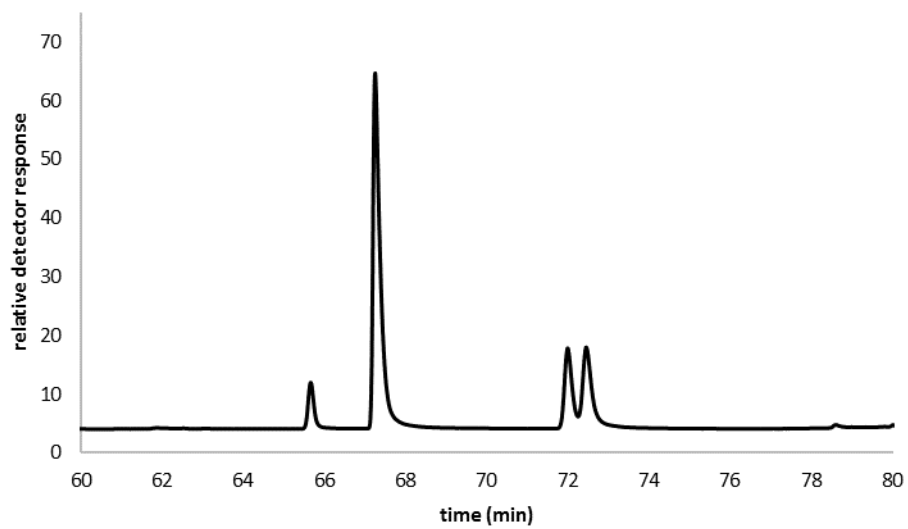
**P450<sub>BM3</sub>-T268A**



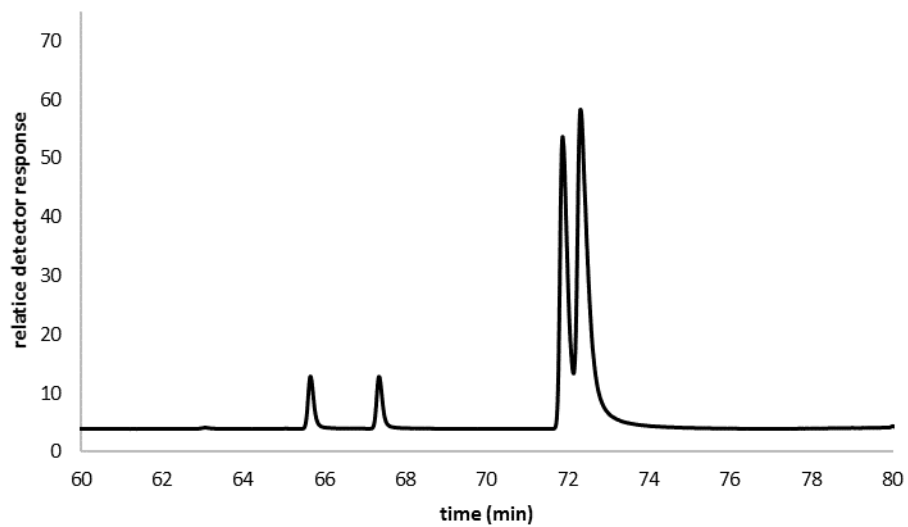
**P450<sub>BM3</sub>-CIS-T438S**



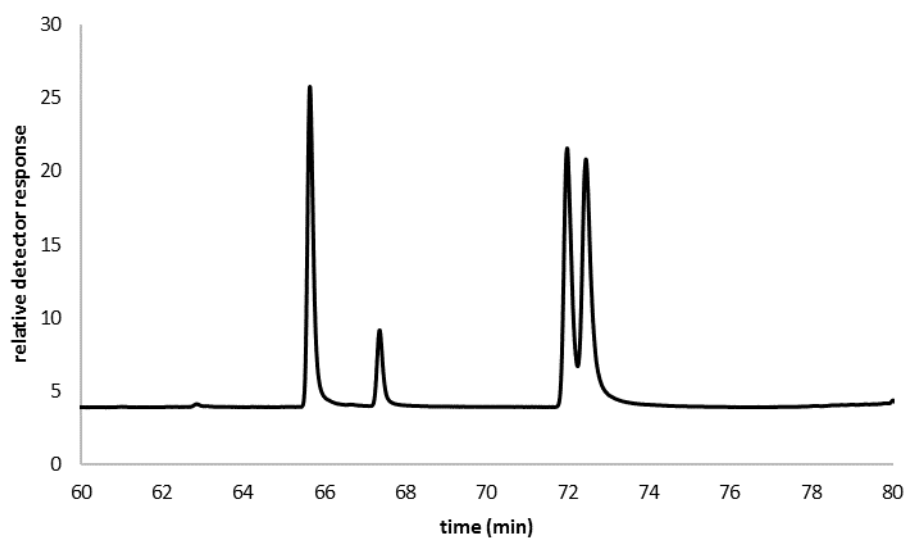
**P450<sub>eryF</sub>**



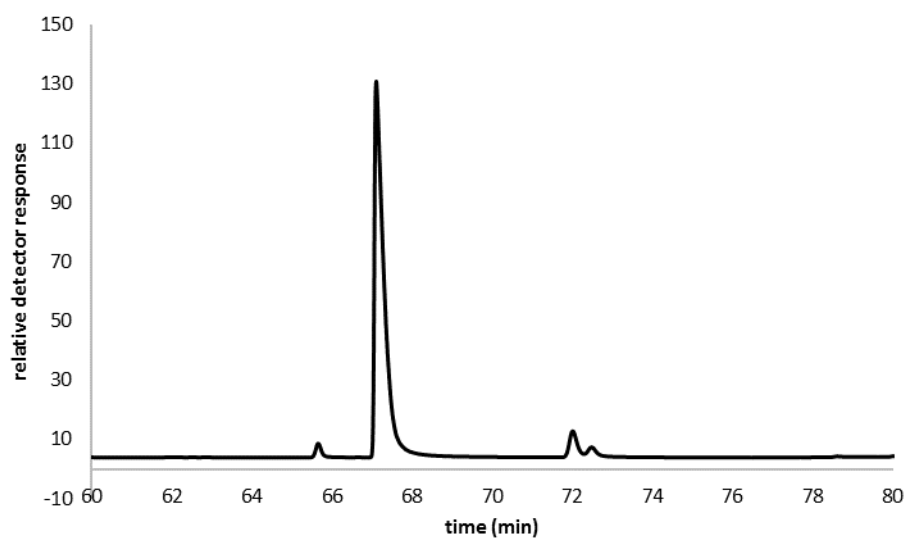
**P450<sub>PikC</sub>-T247A**



**P450<sub>Biol</sub>-T238A**



**CYP142-T234A**



## REFERENCES

1. Arnold, F.H. *Q. Rev. Biophys.* **2015**, FirstView, 1.
2. Alliot, J.; Gravel, E.; Pillon, F.; Buisson, D.-A.; Nicolas, M.; Doris, E. *Chem. Commun.* **2012**, 48, 8111.
3. Chen, D. Y. K.; Pouwer, R. H.; Richard, J.-A. *Chem. Soc. Rev.* **2012**, 41, 4631.
4. Shuto, S.; Ono, S.; Hase, Y.; Kamiyama, N.; Takada, H.; Yamasihita, K.; Matsuda, A. *J. Org. Chem.* **1996**, 61, 915.
5. Wong, H. N. C.; Hon, M. Y.; Tse, C. W.; Yip, Y. C.; Tanko, J.; Hudlicky, T. *Chem. Rev.* **1989**, 89, 165.
6. Morandi, B.; Carreira, E. M. *Science* **2012**, 335, 1471.
7. Wolf, J. R.; Hamaker, C. G.; Djukic, J.-P.; Kodadek, T.; Woo, L. K. *J. Am. Chem. Soc.* **1995**, 117, 9194.
8. Davies, H. M. L.; Beckwith, R. E. J. *Chem. Rev.* **2003**, 103, 2861.
9. Doyle, M. P.; Forbes, D. C. *Chem. Rev.* **1998**, 98, 911.
10. Ho, C.-M.; Zhang, J.-L.; Zhou, C.-Y.; Chan, O.-Y.; Yan, J. J.; Zhang, F.-Y.; Huang, J.-S.; Che, C.-M. *J. Am. Chem. Soc.* **2010**, 132, 1886.
11. Lai, T. S.; Chan, F. Y.; So, P. K.; Ma, D. L.; Wong, K. Y.; Che, C. M. *Dalton Trans.* **2006**, 4845.
12. Mazet, C.; Kohler, V.; Pfaltz, A. *Angew. Chem. Int. Ed.* **2005**, 44, 4888.
13. Zhu, S.; Xu, X.; Perman, J. A.; Zhang, X. P. *J. Am. Chem. Soc.* **2010**, 132, 12796.
14. Lindsay, V. N. G.; Lin, W.; Charette, A. B. *J. Am. Chem. Soc.* **2009**, 131, 16383.
15. Bordeaux, M.; Tyagi, V.; Fasan, R. *Angew. Chem. Int. Ed.* **2015**, 54, 1744.
16. Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. *Science* **2013**, 339, 307.
17. Heel, T.; McIntosh, J. A.; Dodani, S. C.; Meyerowitz, J. T.; Arnold, F. H. *Chembiochem* **2014**, 15, 2556.
18. Coelho, P. S.; Wang, Z. J.; Ener, M. E.; Baril, S. A.; Kannan, A.; Arnold, F. H.; Brustad, E. M. *Nat. Chem. Biol.* **2013**, 9, 485.

19. Cammer, S. A.; Hoffman, B. T.; Speir, J. A.; Canady, M. A.; Nelson, M. R.; Knutson, S.; Gallina, M.; Baxter, S. M.; Fetrow, J. S. *J. Mol. Biol.* **2003**, *334*, 387.
20. Fetrow, J. S. in *Curr. Protoc. Bioinformatics*, John Wiley & Sons, Inc., **2006**.
21. Huff, R. G.; Bayram, E.; Tan, H.; Knutson, S. T.; Knaggs, M. H.; Richon, A. B.; Santago, P.; Fetrow, J. S. *Chem. Biodivers.* **2005**, *2*, 1533.
22. Whitehouse, C. J. C.; Bell, S. G.; Wong, L.-L. *Chem. Soc. Rev.* **2012**, *41*, 1218.
23. Girvan, H. M.; Seward, H. E.; Toogood, H. S.; Cheesman, M. R.; Leys, D.; Munro, A. W. *J. Biol. Chem.* **2007**, *282*, 564.
24. McIntosh, J. A.; Heel, T.; Buller, A. R.; Chio, L.; Arnold, F. H. *J. Am. Chem. Soc.* **2015**, *137*, 13861.
25. Leuthaeuser, J.B.; Morris, J. H.; Harper, A.F.; Ferrin; T. E.; Babbitt; P. C.; Fetrow, J. S. *BMC Bioinformatics* **2016**, *17* (1), 458.
26. Haines, D. C.; Tomchick, D. R.; Machius, M.; Peterson, J. A. *Biochemistry* **2001**, *40*, 13456.
27. Poulos, T. L.; Finzel, B. C.; Howard, A. J. *Biochemistry* **1986**, *25*, 5314.
28. Otey, C. R. In *Directed Enzyme Evolution*; Arnold, F. H., Georgiou, G., Eds.; Humana Press: **2003**; Vol. 230, 137.
29. Barry, S. M.; Kers, J. A.; Johnson, E. G.; Song, L.; Aston, P. R.; Patel, B.; Krasnoff, S. B.; Crane, B. R.; Gibson, D. M.; Loria, R.; Challis, G. L. *Nat. Chem. Biol.* **2012**, *8*, 814.
30. Chen, H.; Walsh, C. T. *Chem. Biol.* **2001**, *8*, 301.
31. Chung, L. L., L; Patel, S.; Carney, J. R. *J. Antibiot.* **2001**, *54*, 250.
32. Dodani, S. C.; Cahn, J. K.; Heinisch, T.; Brinkmann-Chen, S.; McIntosh, J. A.; Arnold, F. H. *Chembiochem* **2014**, *15*, 2259.
33. Ogura, H.; Nishida, C. R.; Hoch, U. R.; Perera, R.; Dawson, J. H.; Ortiz de Montellano, P. R. *Biochemistry* **2004**, *43*, 14712.
34. Peterson, J. A.; Lu, J. Y.; Geisselsoder, J.; Graham-Lorence, S.; Carmona, C.; Witney, F.; Lorence, M. C. *J. Biol. Chem.* **1992**, *267*, 14193.
35. Roberts, G. A.; Celik, A.; Hunter, D. J.; Ost, T. W.; White, J. H.; Chapman, S. K.; Turner, N. J.; Flitsch, S. L. *J. Biol. Chem.* **2003**, *278*, 48914.
36. Stok, J. E.; De Voss, J. *J. Arch. Biochem. Biophys.* **2000**, *384*, 351.

37. Warrilow, A. G.; Jackson, C. J.; Parker, J. E.; Marczylo, T. H.; Kelly, D. E.; Lamb, D. C.; Kelly, S. L. *Antimicrob. Agents Chemother.* **2009**, *53*, 1157.
38. Xue, Y.; Wilson, D.; Zhao, L.; Liu, H.-W.; Sherman, D. H. *Chem. Biol.* **1998**, *5*, 661.
39. Tomura, D.; Obika, E.; Fukamizu, K.; Shoun, I. *J. Biochem.* **1994**, *116*, 88.
40. Keller, R. M.; Wüthrich, K.; Debrunner, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **1972**, *69*, 2073.
41. Driscoll, M. D.; McLean, K. J.; Levy, C.; Mast, N.; Pikuleva, I. A.; Lafite, P.; Rigby, S. E. J.; Leys, D.; Munro, A. W. *J. Biol. Chem.* **2010**, *285*, 38270.
42. Fouces, R.; Mellado, E.; Diez, B.; Barredo, J. L. *Microbiology* **1999**, *145*, 855.
43. Andersen, J. F.; Tatsuta, K.; Gunji, H.; Ishiyama, T.; Hutchinson, C. R. *Biochemistry* **1993**, *32*, 1905.
44. Kelly, S. L.; Kelly, D. E. *Philosophical Transactions of the Royal Society B: Biol. Sci.* **2013**, *368*, 20120476.
45. Egawa, T.; Hishiki, T.; Ichikawa, Y.; Kanamori, Y.; Shimada, H.; Takahashi, S.; Kitagawa, T.; Ishimura, Y. *J. Biol. Chem.* **2004**, *279*, 32008.



## CHAPTER 3: P450-MEDIATED NON-NATURAL CYCLOPROPANATION OF DEHYDROALANINE-CONTAINING THIOPEPTIDES<sup>3</sup>

### Introduction

Cytochrome P450s are versatile biocatalysts that are of synthetic importance due to their ability to catalyze late-stage modifications of small molecules and natural products.<sup>1-4</sup> For example, P450s identified from *S. bungoensis* A-1544 have been instrumental in preparing C16-hydroxylated pladienolide analogs for clinical trials.<sup>5,6</sup> More recently, Arnold, Fasan, and others have shown that P450s and other heme proteins catalyze non-native metallocarbenoid and metallonitrenoid insertion reactions, providing novel biosynthetic routes to the formation of C-C, C-N, and other bonds.<sup>4,7-14</sup> P450-mediated cyclopropanation of aryl olefins using diazoacetate reagents has garnered attention due to the high turnover and diastereoselectivities displayed by native and engineered P450s.<sup>7,15</sup> This reaction has been adapted for the enzymatic synthesis of precursors to pharmaceuticals such as levomilnacipran and ticagrelor.<sup>16,17</sup> In addition, tuning P450 reactivity through mutations of the proximal Cys heme-ligand to Ser or His has led to efficient whole cell cyclopropanation catalysts.<sup>15,16</sup> Due to the prevalent role of P450s in natural product biosynthesis, non-natural P450-mediated reactions may be adaptable to the diversification of complex molecules. Notably, Fasan and colleagues have demonstrated that engineered P450s catalyze intramolecular C-H amination of modified-terpenes covalently

---

<sup>3</sup>Edited with permission from: Gober, J. G.; Ghodge, S. V.; Bogart, J. W.; Wever, W. J.; Watkins, R. R.; Brustad, E. M.; and Bowers, A. A. Submitted, 2017. Under revision at *J. Am. Chem. Soc.*

decorated with carbonazidate metallonitrenoid precursors.<sup>4</sup>

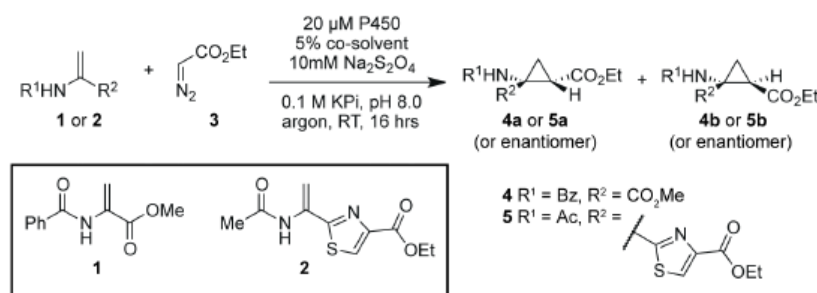
We hypothesized that P450s may be capable of intermolecular carbenoid-mediated cyclopropanations of natural products bearing native or engineered olefins. In particular, dehydroalanines (Dhas) and related olefins, are common features of ribosomally-produced cyclic peptides, including lantipeptide and thiopeptide antibiotics.<sup>18</sup> Despite the fact that hemoprotein-catalyzed cyclopropanation reactions have typically been limited to aryl olefins (e.g. styrenes), we speculated that the geminal push-pull olefin present in Dhas would be sufficiently electron rich to permit modification by P450-metallocarbenoids. On small molecule Dhas, this chemistry would allow access to 1-amino-2-cyclopropane carboxylic acids (ACCAs), important motifs in natural products and pharmaceuticals.<sup>19-21</sup> Alternatively, Dhas might provide reactive handles for ACCA incorporation into natural products.

## Results and Discussion

We began our investigations by targeting cyclopropanation of two model compounds representative of Dhas observed in lantipeptides (**1**) and thiopeptides (**2**). For our first survey of this chemistry, we employed a panel of diverse P450 variants which we previously identified as highly active and/or stereoselective for cyclopropanation of styrene with ethyl diazoacetate (EDA, **3**).<sup>22</sup> To enrich activity, we also added an active site Thr→Ala mutation that has been found to be generally activating with respect to carbenoid transfer. As summarized in Table 3.1 (and Figure S3.1, Figure S3.2, and Table S3.1), each variant screened was active for cyclopropanation against **1** and **2**, to yield cyclopropane products **4** and **5**, respectively. Conversions and diastereoselectivities were determined based on UV absorbance compared to authentic standards that were isolated and fully characterized by NMR. Diastereomers of **4** have

been previously described in the literature;<sup>23</sup> key NOEs were used to discern diastereomers of compound **5** (see Supporting Methods). A number of trends can be observed in these results. Most notably, thiazole Dha, **2**, was cyclopropanated more readily when compared to the more electron-deficient Dha ester (**1**); N-H insertion, a preferred target for P450-generated carbenoids with olefin-substituted anilines,<sup>11</sup> was not observed for either substrate. While conversions in excess of 60% could be obtained with **2** (Table 3.1, Figures S3.1 and S3.2), cyclopropanation of **1** gave low conversion (2 – 29 %) even in the presence of excess olefin. In reactions with compound **1**, reduced cyclopropanation is accompanied by competing EDA dimerization, which proceeds in 14 – 71 % conversion (Figures S3.3 and Table S3.1). This result suggests that slow insertion into more electron-deficient Dha olefins may enhance off-path dimerization products.

**Table 3.1.** Dha cyclopropanation with P450 variants.



catalyst	conversion ( <b>4</b> ) <sup>a</sup>	TTN( <b>4</b> )	dr ( <b>4</b> )	conversion ( <b>5</b> ) <sup>b</sup>	TTN( <b>5</b> )	dr ( <b>5</b> )
hemin	10	50	66:34	67	337	82:18
P450 <sub>BM3</sub> -T268A	2	8	63:37	1	3	94:06
P450 <sub>BM3</sub> -T268A/C400S	6	29	68:32	13	67	82:18
P450 <sub>BM3-CIS</sub> -C400S	6	32	69:31	11	56	66:34
P450 <sub>Biol</sub> -T238A	4	21	49:51	13	65	46:54
P450 <sub>Biol</sub> -T238A/C344S	18	90	55:45	61	304	72:28
P450 <sub>EryF</sub>	12	60	76:24	18	91	72:28
CYP142-T234A	2	12	74:26	3	16	84:16
P450 <sub>PikC</sub> -T247A	29	145	83:17	64	321	91:09
CYP119-T213A	19	94	85:15	11	57	86:14
CYP119-T213A/C317S	6	28	57:43	44	221	10:90

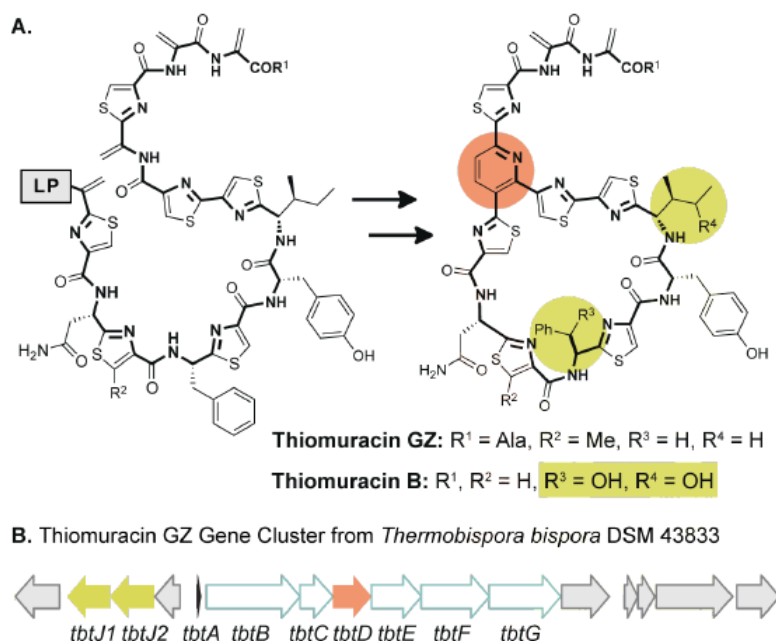
TTN = total turnover number. TTNs and diastereomeric ratios determined by HPLC analysis.

<sup>a</sup>Conditions: 30 mM **1**, 10 mM **3**. <sup>b</sup>Conditions: 10 mM **2**, 10 mM **3**.

As anticipated, the Thr→Ala mutation proved generally helpful. However, addition of an axial Cys→Ser mutation, which is activating in aryl olefin cyclopropanation,<sup>15,22</sup> led to variable improvements across both substrates. Interestingly, the Cys→Ser mutation had a profound effect on stereoselectivity in CYP119. In general, the C-C *cys* diastereomer comprised the preferred product for either substrate **2** (see Supporting Methods), with moderate *dr* observed for **1** and very good *dr* for **2** (Table 3.1). With CYP119-T213A-C317S, a loss of stereoselectivity was observed for compound **1**, whereas, against compound **2**, the catalyst showed a significant inversion of diastereoselectivity (*dr*: 84:16 → 10:90).

We next expanded our P450 library in order to examine enzymes that evolved to oxidize natural products containing motifs analogous to compounds **1** and **2**. Many thiopeptides undergo post-translational monooxygenation by P450s. For example, thiomuracin and related thiopeptides have hydroxyls and/or epoxides installed on side chains (Figure 3.1).<sup>24-27</sup> Mutagenesis experiments have tied these modifications to P450s, but there have been no reports of *in vitro* reconstitution to date.<sup>27,28</sup> We chose to pursue two such P450s, P450<sub>TbJ1</sub> and P450<sub>TbJ2</sub>, identified from the recently reported thiomuracin GZ pathway from *T. bispora* DSM 43833 (Figure 3.1).<sup>29</sup> P450<sub>TbJ1</sub> and P450<sub>TbJ2</sub> variants bearing conserved activating mutations were screened for cyclopropanation of **1** and **2** with EDA. P450<sub>TbJ1</sub> and P450<sub>TbJ2</sub> variants gave similar efficiencies to library constituents on **1** (Table 3.2, and Figure S3.1 and S3.2) and showed enhanced activity against Dha-thiazole, **2**. Notably, P450<sub>TbJ2</sub>-T247A provided the best conversion (79 %), turnovers, and *dr* of any of the biocatalysts surveyed against **2**. These results suggested that P450<sub>TbJ1</sub> and P450<sub>TbJ2</sub> variants might be suitable candidates to carry out similar chemistry on Dha-containing analogues of native substrates. In effort to better understand the differences in activities observed among P450 catalysts for **2**, we determined the crystal structure

of substrate-free P450<sub>TbtJ1</sub> (Table S3.2, Figure S3.4, 2.4 Å resolution, PDB: t.b.d.). Initial efforts to obtain structures for P450<sub>TbtJ2</sub> did not yield diffraction-quality crystals. Nevertheless, the P450<sub>TbtJ1</sub> structure allows a visual comparison of X-ray crystal structures for P450 biocatalysts tested herein. P450<sub>TbtJ1</sub> exhibits the typical triangular prism-shaped P450 fold with the heme prosthetic group embedded adjacent to the central I helix (Figure S3.5). The surface of the protein shows a deep but narrow channel that may accommodate its preferred linear peptide



**Figure 3.1.** Key post-translational modifications of thiomuracin GZ core. A. Cyclization of the TbtA precursor by TbtD leads to the pyridine core (orange) of Thiomuracin GZ. LP = leader peptide B. Two monooxygenations, highlighted in yellow, are predicted to be carried out by P450 gene products of *tbtJ1* and *tbtJ2*.

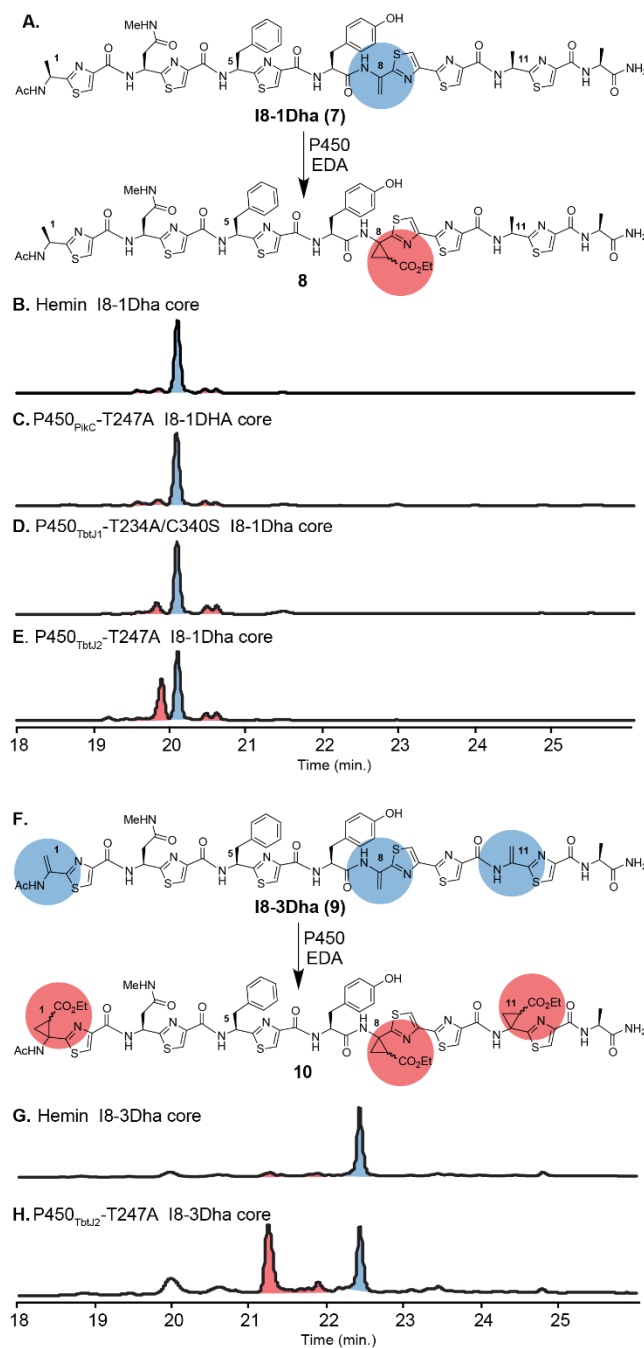
**Table 3.2.** Dha cyclopropanation with P450<sub>TbtJ1</sub> and P450<sub>TbtJ2</sub> variants

catalyst	conversion (4) <sup>a</sup>	TTN(4)	dr (4)	conversion (5) <sup>b</sup>	TTN(5)	dr (5)
P450 <sub>TbtJ1</sub>	2	12	80:20	nt	--	--
P450 <sub>TbtJ1</sub> -T234A	10	50	74:26	54	267	78:22
P450 <sub>TbtJ1</sub> -C340S	12	59	65:35	nt	--	--
P450 <sub>TbtJ1</sub> -T234A/C340S	16	81	54:46	55	277	81:19
P450 <sub>TbtJ2</sub>	4	20	71:29	nt	--	--
P450 <sub>TbtJ2</sub> -T247A	20	101	90:10	79	393	95:05
P450 <sub>TbtJ2</sub> -C353S	6	29	49:51	nt	--	--
P450 <sub>TbtJ2</sub> -T247A/C353S	9	44	59:41	38	192	69:31

TTN = total turnover number. TTNs and diastereomeric ratios determined by HPLC analysis. nt = not tested. <sup>a</sup>Conditions: 30 mM **1**, 10 mM **3**. <sup>b</sup>Conditions: 10 mM **2**, 10 mM **3**.

substrate. This large substrate-binding cavity is flanked by the distinctive B-C and F-G loops shared by many P450s.<sup>1</sup> Common to substrate-free crystal structures of other P450s, the B-C loop appears partially disordered (Figure S3.6 and S3.7). Dynamic interactions with the B-C loop play an important role in P450 substrate recognition, and substrate binding typically results in organization of the B-C loop in other P450 crystal structures.<sup>30</sup>

Qualitatively, P450s that show high activity against substrate **2** (i.e. P450<sub>PikC</sub> (PDB: 2C7Z), P450<sub>TbU1</sub> (PDB: t.b.d.), and CYP119 (PDB:1IO7)) are marked by large, solvent-accessible heme binding pockets in their ligand-free form (Figure S3.8). Alternatively, the heme of poorly active P450s, including P450<sub>BM3</sub> and CYP142 (PDB: 1YQO and 3EJB, respectively), is buried within the protein, with only narrow solvent channels leading to the cofactor. This observation is supported by calculations of accessible channel volumes (see Figure S3.8 for details), which suggest highly active P450s possess large channels that facilitate substrate access to the metallocarbenoid intermediate. A notable structural difference for P450<sub>TbU1</sub> is the presence of several phenylalanines (F70, F277, and F300, Figure S5) near the active site. These aromatic residues may be involved in pi-pi interactions that bind and/or orient native polythiazole substrates and may contribute to the enriched activity observed for P450<sub>TbU1</sub> variants on thiazole Dha, **2**. A similar motif has been proposed for thiazole recognition by EpoK, a P450 involved in epothilone biosynthesis, and similar residues can be found in alignments with P450<sub>TbU2</sub> (Figure S9 and S10).<sup>31</sup> Further structural and biochemical studies to confirm the importance of these features for substrate recognition, as well as P450<sub>TbU2</sub> structural investigations, are ongoing; however, we anticipate that the substrate-free structure will facilitate further engineering and applications of this chemistry on thiomuracin and other thiopeptides.



**Figure 3.2.** P450-mediated cyclopropanation of linear I8Dha thiopeptide core substrates bearing either one (7, panel A) or three (9, panel F) Dha olefins. B-E. UV traces of cyclopropanation reactions on 7 show starting material (blue) and monosubstituted cyclopropane products (8, red). G-H. UV traces of reactions on 9 show starting material (blue) and tri-substituted cyclopropane products (10, red). Reaction conditions: 20  $\mu$ M catalyst, 100  $\mu$ M thiomuracin derivative, 10 mM EDA, 10 mM  $\text{Na}_2\text{S}_2\text{O}_4$ , 0.1 M KPi pH 8, 7.5% co-solvent, argon, 16 h.

With thiomuracin P450 variants in hand, we turned to the synthesis of putative substrates. Based on prior mutagenesis work, it seemed likely that these enzymes would function on linear thiopeptide cores prior to installation of the central pyridine core (Figure 3.1).<sup>28</sup> We exploited a recently disclosed solid-phase route to thiopeptide cores to prepare a substrate analog in which Ile8, which is monooxygenated in a number of thiomuracin variants, was replaced with a Cys residue that could be eliminated to yield a Dha at this position; all other DhAs normally present in this core were replaced with alanine to reduce potential complexity (Figure S3.11).<sup>32,33</sup> Compound **7** (I8-1Dha, Figure 3.2A and Table 3.3) was separately incubated with three of the best biocatalysts for cyclopropanation of **1** and **2** (P450<sub>PikC</sub>-T247A, P450<sub>TbJ1</sub>-T234A/C340S and P450<sub>TbJ2</sub>-T247A), as well as free hemin. Due to the limited solubility of **7**, the maximum peptide concentration investigated was 100  $\mu$ M. Catalyst concentrations were maintained at 20  $\mu$ M, consistent with small molecule reactions. P450<sub>PikC</sub>-T247A, P450<sub>TbJ1</sub>-T234A/C340S, and hemin all gave very low levels of modified products exhibiting different retention times (Figure 3.2B-D, Table 3.3). These peaks likely correspond to multiple diastereomers of the expected cyclopropanation product, which are generated in low yields by these catalysts. In contrast, P450<sub>TbJ2</sub>-T247A provides robust conversion to a single major peak in the LC trace (**8**, Figures 3.2A and 3.2E, Figures S3.12 and S3.13); a few minor peaks can also be observed near the baseline. To confirm that this product was the expected cyclopropane and to investigate diastereoselectivity, we repeated the reaction on scale using <sup>13</sup>C-labeled EDA in order to facilitate subsequent characterization by NMR. <sup>1</sup>H and <sup>13</sup>C NMRs of the purified major product confirm that it is a single diastereomer of the cyclopropanation product (See Supporting Methods). A trace amount (<5% based on integration) of an unidentified compound with resonance at ~40 ppm can also be observed in the <sup>13</sup>C NMR. Together, these results show that



P450<sub>TbIJ2</sub>-T247A catalyzes robust and diastereoselective cyclopropanation on an engineered variant of its natural substrate. As P450<sub>TbIJ2</sub>-T247A exhibits proclivity towards modification of an I8-1Dha core (**7**), this may indicate P450<sub>TbIJ2</sub>'s role in native modifications at the I8 position.

We further investigated the potential for site-specific cyclopropanation on thiopeptide substrates exhibiting multiple Dhas. Substrates F5-3Dha (Figure S3.14 and S3.15) and I8-3Dha (**9**, Figure 3.2F and S3.16) display three Dhas: one Dha at one of the two different positions that

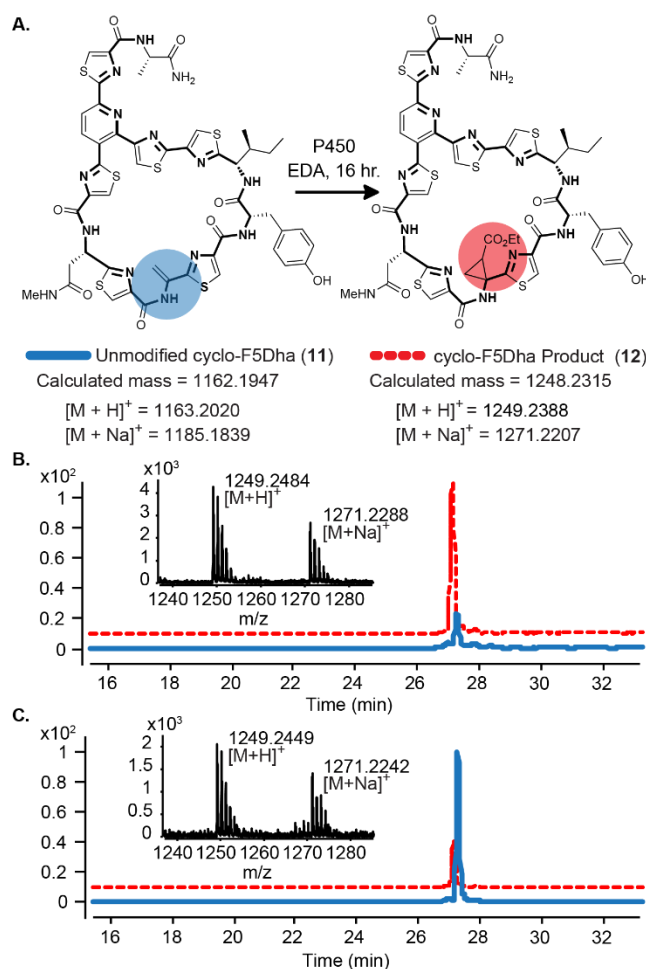
**Table 3.3.** Dha cyclopropanation on linear thiopeptide cores

catalyst	core	% conversion <sup>a</sup>
Hemin	I8-1Dha ( <b>7</b> )	14
P450 <sub>PikC</sub> -T247A	I8-1Dha ( <b>7</b> )	9
P450 <sub>TbIJ1</sub> -T234A/C340S	I8-1Dha ( <b>7</b> )	29
P450 <sub>TbIJ2</sub> -T247A	I8-1Dha ( <b>7</b> )	46
Hemin	F5-3Dha	11
P450 <sub>PikC</sub> -T247A	F5-3Dha	12
P450 <sub>TbIJ1</sub> -T234A/C340S	F5-3Dha	44
P450 <sub>TbIJ2</sub> -T247A	F5-3Dha	30
Hemin	I8-3Dha ( <b>9</b> )	14
P450 <sub>PikC</sub> -T247A	I8-3Dha ( <b>9</b> )	25
P450 <sub>TbIJ1</sub> -T234A/C340S	I8-3Dha ( <b>9</b> )	42
P450 <sub>TbIJ2</sub> -T247A	I8-3Dha ( <b>9</b> )	68

<sup>a</sup>% Conversion based on UV<sub>A254</sub> integration versus that of starting material.

normally undergo P450-mediated monooxygenation (Phe5 or Ile8) and two Dhas at positions 1 and 11, which are natively used as handles for pyridine formation en route to the macrocyclic thiopeptide. Variants of P450<sub>TbIJ1</sub> and P450<sub>TbIJ2</sub> provided cyclopropane products (i.e. **10** in Figure 3.2F) with improved conversion relative to P450<sub>PikC</sub>-T247A or hemin alone (Table 3.3, Figure 3.2G and 3.2H; Figure S3.17-S3.22). Somewhat surprisingly, the major products exhibit modifications at all three olefins (Figures 3.2, S3.21 and S3.22); mono- and di-cyclopropanation products are only detectable in the baseline of the mass spectra. Again, P450<sub>TbIJ2</sub>-T247A shows particularly strong conversion (~68%, Figure 3.2H and Table 3.3) from substrate **9** and yields primarily **4a** single peak in the UV trace.

In order to investigate site selectivity for enzymatic cyclopropanations by P450<sub>TbJ2</sub>-T247A, product formation in the presence of **9** was monitored over time. Notably, the reaction nears completion in under an hour, and the tri-cyclopropanated species was the major product observed even at early time points (Figures S3.23 and S3.24). Accumulation of mono- and di-modified product was not observed. These results suggest that the native binding affinity of P450<sub>TbJ1</sub> and P450<sub>TbJ2</sub> may permit processive scanning of the substrate, transiently orienting



**Figure 3.3** Non-natural enzymatic modification of a cyclic thiopeptide. A. Cyclic F5Dha peptide precursor (**11**) and cyclopropane product (**12**). B. Extracted ion chromatograms from P450<sub>TbJ1</sub>-T234A/C340S cyclopropanation reaction with cyclo-F5Dha C. Extracted ion chromatograms from P450<sub>TbJ2</sub>-T247A cyclopropanation reaction with cyclo-F5Dha. Cyclo-F5Dha is shown in blue, and the cyclopropane product is shown in red. Reaction conditions: 20  $\mu$ M P450, 100  $\mu$ M thiomuracin derivative, 10 mM EDA, 10 mM Na<sub>2</sub>S<sub>2</sub>O<sub>4</sub>, 0.1 M KPi pH 8, 7.5% co-solvent, argon, 16 h.

each olefin within the active site for modification. A broader analysis of substrate promiscuity with additional analogs will be necessary to confirm this mechanism.

For comparison, we investigated modification of Dha modified peptides using standard metal catalysts. Rhodium catalysts have been used by Romo and colleagues to catalyze microscale intermolecular C-H amination and olefin cyclopropanation of natural products.<sup>34,35</sup> Attempts to cyclopropanate I8-3Dha (**9**) using a panel of Rh catalysts provided only trace amounts of product (Figure S25). This data may suggest an inhibitory interaction between the peptide and catalyst. In addition, poor peptide solubility prevents high substrate concentrations typically used for Rh-mediated cyclopropanations. Alternatively, catalysis by Fe<sup>III</sup>-tetraphenylporphyrin provided a complex mixture of cyclopropanation products enriched for one or two cyclopropane modifications (Figure S3.25). Finally, in order to access cyclopropanated thiopeptides, we sought to determine whether P450<sub>TbJ1</sub> or P450<sub>TbJ2</sub> variants can modify a macrocyclic thiopeptide. This would provide a unique enzymatic opportunity for late-stage introduction of new C-C bonds, functional handles, and structural diversity into a natural product or natural product-like molecule. Accordingly, both the F5-3Dha and I8-3Dha core substrates were elaborated with a 15-residue fragment of the thiomuracin leader peptide. In previous work, we have reported that a pyridine synthase recognition sequence is located in this region of the thiomuracin leader peptide (Figure S3.26 and S3.27).<sup>33,36</sup> Treatment of the F5-3Dha core with the thiomuracin pyridine synthase, TbtD (Figure 1), yielded the pyridine macrocycle (**11**, Figure 3.3 and S3.28). In contrast, no cyclization was observed with the I8-3Dha core, suggesting that TbtD is intolerant to this modification. Incubation of the cyclic F5-Dha thiomuracin analog (**11**, Figure 3.3A) and EDA with P450<sub>TbJ1</sub>-T234A/C340S or P450<sub>TbJ2</sub>-T247A afforded the cyclopropanated thiopeptide product (**12**, Figure 3B and 3C; confirmation by LCMS/MS in Figure S3.29 and

S3.30). We additionally screened P450<sub>TbJ1</sub> and P450<sub>TbJ2</sub> variants against two olefin-containing thiopeptides, nosiheptide and thiostrepton, derived from alternative biosynthetic pathways. Only baseline detection of cyclopropanated nosiheptide was observed, whereas thiostrepton showed no modification (data not shown). These results suggest that P450<sub>TbJ1</sub> and P450<sub>TbJ2</sub> variants show preference for linear and cyclic thiopeptide analogs of their natural substrate for cyclopropanation.

## Conclusion

In conclusion, we have demonstrated that P450 mediated cyclopropanation can be extended to modification of Dhas, resulting in enzymatic synthesis of ACCAs. Both electron-rich and electron-deficient Dhas can be transformed in this way, although further catalyst tuning will be necessary to obtain appreciable yields from the latter. Additionally, we characterized two new P450s, P450<sub>TbJ1</sub> and P450<sub>TbJ2</sub> from the thiomuracin pathway in *T. bispora* DSM 43833. Both P450s were able to affect cyclopropanation of Dhas at multiple sites on thiopeptide precursors. Significantly, engineered variants of P450<sub>TbJ1</sub> and P450<sub>TbJ2</sub>, which displayed weak activity on model Dha **1**, were more efficient catalysts on thiopeptides resembling their native substrates. In addition, these thiopeptide substrates were poorly tolerated by hemin and other P450s (i.e. PikC). These results suggest that innate substrate recognition motifs can be exploited to enhance non-natural enzymatic modifications on natural products and their analogues. Furthermore P450<sub>TbJ1</sub>, which we structurally characterized, is able to carry out ACCA incorporation on a cyclic thiomuracin analog, thereby demonstrating that P450s can carry out non-natural, late-stage modification of complex molecules and natural products, in addition to monooxygenation. Beyond cyclopropanation, a diverse array of metallocarbenoid and metallonitrenoid chemistries have been demonstrated in P450s<sup>4,7,8,10-14</sup> and Ir-substituted heme proteins,<sup>37,38</sup> suggesting that

P450-mediated modification of natural products can be extended to other non-natural C-C, C-N, and C-X modifications. Future efforts in our labs will focus on investigating the substrate scope for P450<sub>TbU1</sub> and P450<sub>TbU2</sub>, engineering enzyme variants for improved activity and selectivity on small molecule or peptide Dhas, and assessing the biological activity of analogs prepared by this method.

## Methods

### *General*

All reactions were carried out in an oven-dried round-bottomed-flask under an inert nitrogen atmosphere with stirring. Solvents, reagents, and chemicals were purchased from commercial suppliers (Acros, Fisher, Sigma, ChemPep) and used as received unless otherwise noted. Spectra for <sup>1</sup>H and <sup>13</sup>C NMR were recorded at room temperature with a Varian Inova 400 spectrometer (400 MHz and 100 MHz, respectively), Varian Inova 500 spectrometer (500 MHz and 125 MHz, respectively), a Bruker model DRX 600 spectrometer (600 MHz and 151 MHz respectively) and a Bruker 700 Avance3E spectrometer (700 MHz and 176 MHz respectively). Chemical shifts are reported in  $\delta$  (ppm) relative units to residual solvent peaks CDCl<sub>3</sub> (7.26 ppm for <sup>1</sup>H and 77.0 ppm for <sup>13</sup>C) and CD<sub>3</sub>OD (3.35 ppm and 4.78 ppm for <sup>1</sup>H; 49.32 ppm for <sup>13</sup>C). Splitting patterns are assigned as s (singlet), d (doublet), t (triplet), q (quartet), quint (quintet), multiplet (m), dd (doublet of doublets), and td (triplet of doublets). LC-MS and LC-MS/MS measurements were recorded using an Agilent 6520 Accurate-Mass Q-TOF ESI positive in high resolution mode (350 °C, 250 V fragmentor for peptides; 310 °C, 130 V fragmentor for small molecules). Analytical columns used in LC-MS analysis were a Phenomenex Kinetex® 2.6  $\mu$  C18 100 Å column (50 x 2.1 mm), a Thermo Scientific™ 2.6  $\mu$  Accucore™ RP-MS LC Column (50 x 3 mm), and a Restek 5  $\mu$  Ultra Biphenyl Column (50 x 2.1 mm). MS-MS was performed with collision energy of 45 V or 55 V. Predicted masses were extracted to  $\pm 5$  ppm. Semi-

preparatory HPLC was carried out using a Shimadzu UFLC CBM-20A with a dual channel wavelength detector at 220 nm and 254 nm with a Phenomenex LUNA 10  $\mu$  C18(2) 100 Å, AXIA semi-preparatory column or a Restek 5  $\mu$  Ultra Biphenyl Prep Column (50 x 21.2 mm). Gas chromatography analyses were carried out using an Agilent 7820A gas chromatograph, FID detector, and a J&W scientific cyclosil-B column (30m x 0.25 mm x 0.25  $\mu$ m film).

LC-MS Method A		
Solvent A: 0.1% formic acid in water Solvent B: 0.1% formic acid in acetonitrile Column: Accucore RP-MS		
Time (min)	%B	Flow (mL/min)
0.00	5	0.5
6.00	5	0.5
24.00	50	0.5
25.00	50	0.5
26.00	95	0.5
29.00	95	0.5
29.01	5	0.5
32.00	5	0.5

LC-MS Method B		
Solvent A: 5 mM ammonium formate in water, pH 8.3 Solvent B: 5 mM ammonium formate in methanol Column: Accucore RP-MS		
Time (min)	%B	Flow (mL/min)
0.00	5	0.5
6.00	5	0.5
24.00	50	0.5
25.00	50	0.5
26.00	95	0.5
35.00	95	0.5
35.01	5	0.5
38.00	5	0.5

LC-MS Method C		
Solvent A: 0.1% formic acid in water Solvent B: 0.1% formic acid in acetonitrile Column: Kinetex C18		
Time (min)	%B	Flow (mL/min)
0.00	2	0.5
2.00	2	0.5
13.00	95	0.5
15.00	95	0.5
15.01	2	0.5

Chiral GC Method			
	Rate (°C/min)	Value (°C)	Hold time (min)
Initial		80	5
Ramp 1	1	115	10
Ramp 2	17	200	5

LC-MS Method D		
Solvent A: 0.1% formic acid in water Solvent B: 0.1% formic acid in acetonitrile Column: Restek Biphenyl (analytical)		
Time (min)	%B	Flow (mL/min)
0.00	5	0.5
6.00	5	0.5
20.00	15	0.5
24.00	40	0.5
28.00	40	0.5
32.00	95	0.5
35.00	95	0.5
35.01	5	0.5
40.00	5	0.5

Semi-Prep HPLC Method A (for small molecules)		
Solvent A: water Solvent B: acetonitrile Column: Restek Biphenyl (prep)		
Time (min)	%B	Flow (mL/min)
0.00	5	21
6.00	5	21
20.00	15	21
24.00	40	21
28.00	40	21
32.00	95	21
35.00	95	21
35.01	5	21
40.00	5	21

Semi-Prep HPLC Method B (for core peptides)		
Solvent A: 0.1% trifluoroacetic acid in water Solvent B: 0.1% trifluoroacetic acid in acetonitrile Column: LUNA C18		
Time (min)	%B	Flow (mL/min)
0.00	5	15
5.00	5	15
8.00	55	15
22.00	85	15
24.00	100	15
27.00	100	15
29.00	5	15
31.00	5	15

<b>Semi-Prep HPLC Method C (for leader peptide-containing peptides)</b>		
Solvent A: 0.1% trifluoroacetic acid in water Solvent B: 0.1% trifluoroacetic acid in acetonitrile Column: LUNA C18		
Time (min)	%B	Flow (mL/min)
0.00	5	15
8.00	5	15
10.00	45	15
23.00	60	15
24.00	100	15
27.00	100	15
28.00	5	15
30.00	5	15

<b>Semi-Prep HPLC Method D (for small molecules)</b>		
Solvent A: water Solvent B: acetonitrile Column: LUNA C18		
Time (min)	%B	Flow (mL/min)
0.00	5	15
6.00	5	15
8.00	20	15
26.00	35	15
29.00	100	15
30.00	100	15
31.00	5	15
33.00	5	15



## Cloning

TbtJ1 and TbtJ2 were cloned into pMCSG7 using a standard ligation independent cloning (LIC) protocol. The codon optimized genes for expression in *E.coli* were obtained from Integrated DNA Technologies (IDT) and amplified by PCR using Q5 High-Fidelity DNA polymerase (NEB) according to the manufacturer's instructions. The codon optimized gene sequences and the primer sequences used for gene amplification are below:

### tbtJ1

```
ATGATGGAATTGCCCGGACAACCTTCCTTAACGGATGGAGGAGCAGCTCTGTTTGCT
TGGTTACGCACAATGCGTAATGAACATCCGGTGTGGCGCGACCAGTTCGGAATTTAC
CATGTTTTTCGTTATGACGATGTACGCCAGATCCTGGGGGACTATCAGACCTTCAGT
TCTGACCGCACGCGCCTGATGCCGACAGCCCAGGGCTTCGGAAAAGGGGGGTATCAC
AATGATTGATCCCCCGGAGCACC GCCATCAACGCCGTCTGATCACACATGCGTTCAC
ACCCAGTCAATCTCAGCCATGGAACCACGCATCCGTCAAATCGCCGATCATCTTCT
TGATGAGCTTCCGGGACCAGAGTTCGACTTAGTAGAACATTTTGCATACCCTTTGCC
AGTCATTGTAATTGCTGAGTTGCTTGGCGTGCCCCCAGGAGACCGCCACTTATTCCG
CACTTGGTCAGACCGCCTTATGAGCTTACAAGTTGAGAACTACGCTGATCCTGAACT
GGCGCGCACCGTCGCTGCAGCGATGACGGAGATGAACGATTACCTTCGTGAACATT
GTCGTAGTCGCCGTACCCATCCCCGCGACGACTTGTTAACTCGCTTAGTCCAGGCAG
AAGTCGAAGGCAAGCGCTTAGATCTGGAAGAGGTGGTCAATACAGCTAGTTTACTTT
TGCTGGCCGGCCATTTAACCCTACCGTTCTGATCGGCAATACCATGTTGTGCTTGTG
GGACCATCCAGAAGCAGAAAAAGCGGTGCGTGCTGACCCCTCCTTAATCCCCGCAG
CATTGGAGGAGTCGCTGCGTTTACGTTCCGCAATTCCTGCAGGCTGGACGTGTTACTA
CCCGCGATGTAACAATCGCCGGTGAGACAATTCCAGCAAATCGCTTCGTGATGGCCT
GGATTTTAAGCGCAAACCACGACGACCGTCGTTTTCCCGATCCGGAGCGCTTTGACC
TGCACCGTCAAACCACTGGTCACATCGCATTCCGGACACGGGGTCCATTTTGTGTTAG
GAGCTCAGCTGGGGCGTCTTGAGGGTTCGCATCGCGCTGGAACGCTTGTTAGGTCGTT
TACTGAAATTCACCCCTGGCCCCGTGAAGGCATTTCTTCTACCAATCTGCTATCTT
CGGGGCCAGTCGTATGCCTGTTCGTTGCGGATGA
```

### tbtJ2

```
ATGAATGAACCCCTGACCGCGCAGGGCACGCCTACAATGACGATGCCTGCCGCTCA
ACGTCCACGCATCTCTCACGGGGGGCAGGCTTTGCTTAAGTGGCTGGATGAGATGCG
TGAATCCCAACCCGTATGGCGTGATGGATTCCGCATCTTCCATGTTTTCCGTCATGCG
GATGTTTACGCGCGTAATGGCAGATTATGCAACTTTCTCTTCCGACATCAATCGCTTG
CGCCAGGGGGTGATCCGTTTACGCGCAGGATCATTAATGCTGACTGATCCTCCGGAG
CACCGTAAACTTCGCCGTCTGATCTCGCAAGCCTTCAGTCCAAAGATGAGTGCAGAC
ATGAAGCCGCGCATCGCAGAGCTGACAGAGGAACCTTTTAGATGATATTGAGGAAGA
CGAGTTCGATCTGGTCGAAAAATTCGCTCACCCCTTACCCATCATGGTTATCGCTGA
```

GTTACTGGGTATTCCTATTCATGATCGTGGTTTGTTCGTACTTGGGCTGATCGTTTG  
ATTGCGTTGCACGTGGATGACCCTACGGATGTAGAGATTGGACGCATGGTCGGCGA  
AGCAATGCGTGAGATGGGAGAATACGTGCAAACCCACGTGCGCAAACGCCGCGCAG  
ATCCTCAAGATGACCTTGTAAGCAAATTAATCGCAGCAGAAGTAGACGGCGAGCGT  
CTGACGGATGCAGAGATCGTCAACACTTCTTGTCTGCTGCTTTTAGCGGGACAGATT  
ACCTCTACAATGGCGCTTGGAACACCTTCCTTTGCTTTCGTGACGCCCCGGATGCA  
GAACGTGCGGTCCGCGCCGACTTTAGTTTGTAGGCCCGGCCTTCGAAGAAGTGCTG  
CGTTTGCGCCCTCCACTGACACAGGCCGACGTTTGACCACAACAGATGTCGAGGTG  
GCAGGGACGTTGATTCCAGCAGGGAGTCTTGTGATTAAGCGCTAAGCGCTAAC  
TACGATGAGCGCCAGTTCCCGAACCCGTATCGCTTTGATTAAACACGTTTCGCCAAC  
CGCCACTTCGCCTTTGGTCATGGTATTCACCATTTGCTTGGTGCACCTTTGGCGCGTG  
TAGAGGGCCGTGTTGCTTTAGAGCTTTTGTACGCCGTTTCTCCGAGATTACTATTGA  
TCCCGATGCCGAGTTGAGCTACTATGAGGACCCGATGTTTGGAGTCAAATCCCTGCC  
AGTCCGTGTTTCGCCGTGCAGCGTAG

TbtJ1\_F: 5'-TACTTCCAATCCAATGCGATGATGGAATTGCCCGGACAACCTTCC-3'

TbtJ1\_R: 5'-TTATCCACTTCCAATGCGCTATCATCCGCAACGAACAGGCATACG-3'

TbtJ2\_F: 5'-TACTTCCAATCCAATGCGATGAATGAACCCCTGACCGCGCAG-3'

TbtJ2\_R: 5'-TTATCCACTTCCAATGCGCTACTACGCTGCACGGCGAACACGGAC-3'

The purified PCR product was phosphorylated with T4-PNK and then treated with T4-DNA polymerase to create LIC overhangs. pMCSG7 was linearized with SspI, dephosphorylated with Antarctic Phosphatase, and then treated with T4-DNA polymerase to create LIC overhangs. Treated PCR product and vector were combined in 1:1 ratio by volume and transformed into One-Shot Top10™ competent cells (Agilent) and plated onto LB-agar plates containing 100 µg/mL ampicillin. Individual colonies were picked and grown overnight in 5 mL LB-medium containing 100 µg/mL ampicillin. Cells were spun down and DNA was extracted using plasmid purification (Life Technologies). The plasmid was transformed into BL21-DE3 competent cells (Agilent).

Site-directed mutagenesis of TbtJ1 and TbtJ2 was carried out by the QuikChange mutagenesis protocol using PfuTurbo DNA polymerase. The template DNA was digested after PCR with the restriction enzyme DpnI. The PCR product was then transformed into One-Shot

Top10™ Competent cells and grown on LB-agar plates containing ampicillin. Primer sequences used in the mutagenesis PCR reactions are given below:

TbtJ1\_T234A\_F: 5'-CTGGCCGGCCATTTA**GCC**ACTACCGTTCTGATC-3'

TbtJ1\_T234A\_R: 5'-GATCAGAACGGTAGT**GGCT**AATGGCCGGCCAG-3'

TbtJ1\_C340S\_F: 5'-CACGGGGTCCATTTT**AGT**TTAGGAGCTCAGCTG-3'

TbtJ1\_C340S\_R: 5'-CAGCTGAGCTCCTAA**ACT**AAAATGGACCCCGTG-3'

TbtJ2\_T247A\_F: 5'-TTAGCGGGACAGATT**GCCT**CTACAATGGCGCTT-3'

TbtJ2\_T247A\_R: 5'-AAGCGCCATTGTAGA**GGCA**ATCTGTCCCGCTAA-3'

TbtJ2\_C353S\_F: 5'-CATGGTATTCACCAT**AGT**CTTGGTGCACCTTTG-3'

TbtJ2\_C353S\_R: 5'-CAAAGGTGCACCAAG**ACT**ATGGTGAATACCATG-3'

pCYP119 was a gift from Stephen Sligar (Addgene plasmid # 50771). The CYP119 gene contains the recognition sequence for the restriction enzyme NdeI (CATATG). A single point mutation was introduced to remove the NdeI recognition sequence by a Golden Gate assembly protocol. Whole plasmid PCR with Phusion polymerase (NEB) was performed using the primers below (BsaI recognition sequence is underlined):

CYP119\_NdeI\_OUT\_F: 5'-

GATGTT**GGTCTCT**TGAAACA**C**ATGGCAGGTGTTTTCCTATAG-3'

CYP119\_NdeI\_OUT\_R: 5'-GATGTT**GGTCTCT**TTTCCGTCATAATACTGG-3'

The PCR product was digested with the restriction enzyme DpnI (NEB) overnight, column purified, and digested with the restriction enzyme BsaI HF (NEB) for 6 hours. The digested PCR product was then column purified and ligated with T4 DNA ligase (NEB). The ligation mixture was column purified and transformed into *E. coli* strain BL21-GOLD (DE3) competent cells (Agilent).

The modified CYP119 gene was amplified from purified plasmid DNA using Phusion polymerase and the following primers (restriction enzyme recognition sequences are underlined):

CYP119\_NdeI\_F: 5'-GATGTTCATATGTATGACTGGTTTAGTGAGATG-3'

CYP119\_XhoI\_R: 5'-GATGTTCTCGAGTCATTACTCTTCAACCTGACCAC-3'

The PCR product was digested with NdeI, XhoI, and DpnI and then ligated into pET21c+ (digested with NdeI and XhoI and treated with Calf Intestinal Alkaline Phosphatase) using T4 DNA ligase. The ligation mixture was column purified and then transformed into *E. coli* strain BL21-GOLD (DE3) competent cells.

Site-directed mutagenesis to introduce mutations to the conserved threonine and cysteine residues was performed using the whole plasmid Golden Gate assembly method described above. Primer sequences used in the mutagenesis PCR reactions are given below (BsaI recognition sequence is underlined):

CYP119\_T213A\_BsaI\_F: 5'-GATGTTGGTCTCTTGAGGCTACAACCTAATATCAAACCTCTG-3'

CYP119\_T213A\_BsaI\_R: 5'-GATGTTGGTCTCTCTCATTACCCGCTATGAGAAG-3'

CYP119\_C317S\_BsaI\_F: 5'-GATGTTGGTCTCTTCTGTCTTTAGGTGCTCCTTTGGCTAG-3'

CYP119\_C317S\_BsaI\_R: 5'-GATGTTGGTCTCTGATGTATTCCAGACCCAAAG-3'

The sequences and cloning for all other P450s under investigation in this study were reported previously.<sup>1</sup>

#### *Protein expression and purification*

The cytochrome P450s used in this study, as well as TbtD, were expressed and purified as previously described.<sup>22,33</sup> SDS-PAGE analysis of purified P450s is shown in Figure S30.

TbtJ1 used for crystallography was purified using a HisTrap HP column (GE Healthcare Life Sciences) and then purified to homogeneity using an S200 gel filtration column on an

ÄKTAexpress (GE Healthcare Life Sciences) after removal of the N-terminal his-tag by TEV protease. Purified TbtJ1 was stored in 10 mM Tris pH 8, 10 mM NaCl at 4 °C prior to crystallography.

#### *Crystallization conditions for TbtJ1*

TbtJ1 was crystallized by vapor diffusion in 96-well sitting drop plates (Hampton Research) using an automated Phoenix liquid handling robot (Art Robbins Instruments) and sampling conditions from the commercial MCSG crystallization suite screen (Anatrace). A 0.2  $\mu$ L:0.1  $\mu$ L mixture of protein stock (15 mg/mL in 10 mM Tris pH 8, 10 mM NaCl) to mother liquor was combined in each drop. Optimal crystallization conditions for TbtJ1 were 1 M  $(\text{NH}_4)_2\text{SO}_4$ , 1 % w/v PEG 3350, and 0.1 M Bis-tris pH 5.5. Crystals typically grew over a span of 1 – 3 days at room temperature under these conditions.

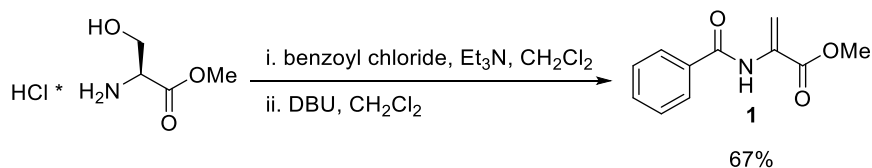
#### *X-ray data collection and protein structure determination*

X-ray diffraction data were collected at Southeast Regional Collaborative Access Team (SER-CAT) at the Advanced Photon Source (Argonne National Laboratory) using beamline 22-ID and a MAR300HS CCD detector. Data were collected at 100 K. Statistics for data collection and refinement are listed in Table S3.2. Diffraction data sets were integrated and scaled with the automated data processing software KYLIN provided by SER-CAT.<sup>39</sup> Initial phases for TbtJ1 were determined by molecular replacement using Phenix Phaser.<sup>40</sup> The ligand free form of P450 EryK from *Saccharopolyspora erythraea* (54% identity to TbtJ1, deposited under PDB code 2WIO) was used as a molecular replacement model.<sup>41</sup> Refinement was accomplished by iterative cycles of manual model building with *Coot* and automated refinement using Phenix Refine.<sup>40,42</sup> Model quality was assessed with the Phenix Validation tool.<sup>40</sup> All protein structure figures were

generated using PyMOL software (The PyMOL Molecular Graphics System, Version 1.6, Schrödinger LLC.).

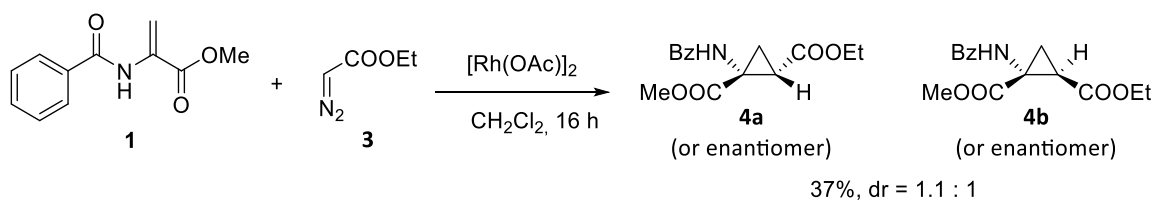
### Synthetic Procedures

#### Synthesis and characterization of methyl 2-benzamidoacrylate (**1**)



Compound **1** was prepared according to a literature procedure.<sup>43</sup> (800 mg, 67% yield, 2 steps). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ 8.52 (broad, s, 1H), 7.80 – 7.83 (m, 2H), 7.42 – 7.54 (m, 3H), 6.77 (s, 1H), 5.97 (d, *J* = 1.6 Hz, 1H), 3.85 (s, 3H).

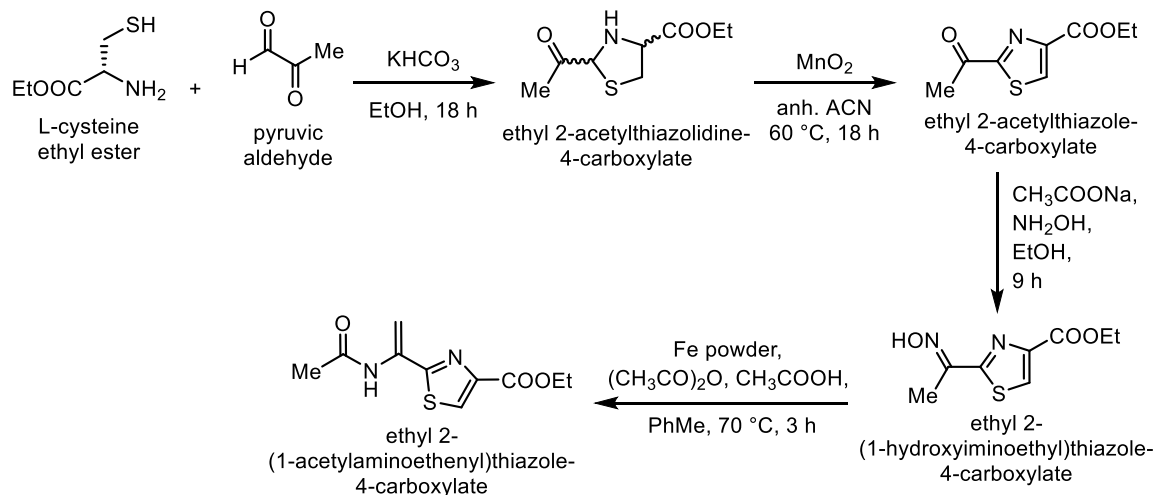
#### Synthesis and characterization of 2-ethyl 1-methyl 1-benzamidocyclopropane-1,2-dicarboxylate



Inside a glove box under nitrogen atmosphere, compound **1** (98.8 mg, 0.482 mmol) was dissolved in CH<sub>2</sub>Cl<sub>2</sub> (2 mL). Ethyl diazoacetate (87%, 87.4 μL, 0.723 mmol) was added dropwise followed by addition of [Rh(OAc)<sub>3</sub>]<sub>2</sub> (4.26 mg, 9.64 μmol). The reaction was stirred under N<sub>2</sub> for 12 h, after which an additional 87.4 μL of ethyl diazoacetate was added and stirred for an additional 4 h. The reaction was concentrated under reduced pressure and purified by silica gel (~60 mesh) using a hexanes : ethyl acetate system (1:1, *R<sub>f</sub>* = 0.45) to afford compound **4** (52.3 mg) in 37% yield as a mixture of diastereomers (dr = 1.1 : 1, major : minor). The diastereomers

were separated using Semi-Prep HPLC Method A. The structure of the diastereomers was determined based on comparison to previously reported  $^1\text{H}$  NMR spectra for diethyl 1-benzamidocyclopropane-1,2-dicarboxylate diastereomers.<sup>23</sup>  $^1\text{H}$  NMR (600 MHz,  $\text{CDCl}_3$ , **4a**):  $\delta$  7.77-7.79 (m, 2H), 7.51-7.54 (m, 1H), 7.43-7.46 (m, 2H), 6.80 (broad, s, 1H), 4.13-4.19 (dq,  $J = 1.6, 7.2$  Hz, 2H), 3.76 (s, 3H), 2.79 (dd,  $J = 7.2, 9.0$  Hz, 1H), 2.04 (dd,  $J = 5.2, 9.4$  Hz, 1H), 1.79 (dd,  $J = 5.2, 8.6$  Hz, 1H), 1.26 (t,  $J = 7.2$  Hz, 3H).  $^1\text{H}$  NMR (600 MHz,  $\text{CDCl}_3$ , **4b**):  $\delta$  7.78-7.80 (m, 2H), 7.53-7.56 (m, 1H), 7.45-7.47 (m, 2H), 6.85 (broad, s, 1H), 4.21 (q,  $J = 7.2$  Hz, 2H), 3.74 (s, 3H), 2.43 (dd,  $J = 8.1, 9.6$  Hz, 1H), 2.35 (dd,  $J = 6, 8.2$  Hz, 1H), 1.77 (dd,  $J = 5.9, 9.6$  Hz, 1H), 1.30 (t,  $J = 7.2$  Hz, 3H).  $^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ , mixture of diastereomers):  $\delta$  14.00, 14.06, 20.67, 21.81, 28.53, 31.27, 39.80, 40.03, 52.92, 53.18, 61.62, 61.67, 127.16, 127.19, 128.64, 128.67, 132.14, 132.31, 132.89, 133.22, 168.40, 168.63, 169.33, 169.37, 169.56, 170.13. Expected  $[\text{M}+\text{H}]^+ = 292.1179$ ; Observed  $[\text{M}+\text{H}]^+ = 292.1198$ .

## Synthesis of and characterization of ethyl 2-(1-acetylaminoethenyl)thiazole-4-carboxylate (2).



Ethyl 2-(1-acetylaminoethenyl)thiazole-4-carboxylate was synthesized according to the procedure published by Hughes et al.<sup>44</sup> A brief description of the adopted procedure is provided here. Potassium hydrogen carbonate (2.16 g, 21.6 mmol) and L-cysteine ethyl ester (4.0 g, 21.6 mmol) were mixed in water (200 mL) and ethanol (200 mL) in a 1000 mL flask. The reaction mixture was stirred under nitrogen at room temperature for 18 hours. The reaction mixture was concentrated to half its original volume, sodium chloride was added to saturate the aqueous mixture, which was then extracted with chloroform (4 x 100 mL). The combined organic phase was dried over magnesium sulfate and concentrated in vacuo using a rotary evaporator to yield ethyl 2-acetylthiazolidine-4-carboxylate (4.0 g) as a red oil, which was used in the next step without purification.

Manganese dioxide (31.2 g, 0.36 mol) was added to ethyl 2-acetylthiazolidine-4-carboxylate (4.0 g) in dry acetonitrile (160 mL). The reaction mixture was heated to  $60^\circ\text{C}$  under nitrogen for 18 hours. The reaction mixture was filtered through Celite and the filter cake was washed thoroughly with acetonitrile (3 x 50 mL). The combined organic filtrate was



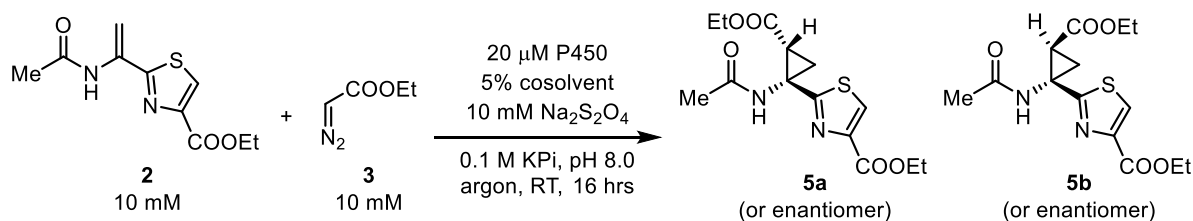
concentrated *in vacuo* to give a brown oil. Purification by column chromatography gave ethyl 2-acetylthiazole-4-carboxylate as a yellow solid (0.35 g, 8% over two steps).

Sodium acetate (0.139 g, 1.708 mmol) was added to a suspension of hydroxylamine hydrochloride (0.118 g, 1.708 mmol) in ethanol (2 mL). After stirring for 30 minutes, a solution of ethyl 2-acetylthiazole-4-carboxylate (0.34 mg, 1.708 mmol) in ethanol (4 mL) was added dropwise over an hour using dropping funnel, and then the reaction mixture was stirred at room temperature under nitrogen for 12 hours. Water (15 mL) was added and the mixture was extracted with dichloromethane (3 x 15 mL). The combined organic layer was dried over magnesium sulfate and concentrated *in vacuo* to give ethyl 2-(1-hydroxyiminoethyl)thiazole-4-carboxylate, which was used in the next step without purification.

Acetic anhydride (0.445 mL, 4.74 mmol) was added in portions to a solution of ethyl 2-(1-hydroxyiminoethyl)thiazole-4-carboxylate (0.34 g, 1.58 mmol) in toluene (4 mL) under nitrogen. Glacial acetic acid (0.27 mL, 4.74 mmol) was added followed by iron powder (0.177 g, 3.16 mmol), and the reaction mixture was heated to 70 °C for three hours. The reaction mixture was allowed to cool to room temperature and filtered through Celite. The filter cake was washed with toluene (3 x 5 mL). The combined filtrates were cooled in an ice bath and washed with aqueous sodium hydroxide solution (2 M, 3 x 15 mL). The organic phase was separated, dried over magnesium sulfate, and the solvents were removed using rotary evaporator. The product was then dissolved in dimethyl sulfoxide and diluted with an equal volume of water, and purified using semi-preparatory reverse-phase HPLC using Phenomenex C18 column as described above. Water was used as solvent A and acetonitrile was used as solvent B. Eluent fractions were analyzed by LC-MS and the solvent was removed using freeze-drying. Ethyl 2-(1-acetylaminoethenyl)thiazole-4-carboxylate (0.094 g, 24% over two steps) was obtained as a

white solid.  $^1\text{H}$  NMR (400 MHz;  $\text{CDCl}_3$ , 298 K)  $\delta$  8.38 (1H, br s, NH), 8.07 (1H, s, 5-H), 6.51 (1H, s, =CHH), 5.47 (1H, s, =CHH), 4.40 (2H, q,  $J$  7.1,  $\text{OCH}_2\text{Me}$ ), 2.20 (3H, s, COMe), 1.40 (3H, t,  $J$  7.1,  $\text{OCH}_2\text{Me}$ );  $^{13}\text{C}$  NMR (400 MHz,  $\text{CDCl}_3$ , 298 K)  $\delta$  169.2, 166.0, 160.8, 146.6, 133.7, 127.7, 103.2, 61.6, 24.9, 14.3. Expected  $[\text{M}+\text{H}]^+$ : 241.0641; Observed  $[\text{M}+\text{H}]^+$ : 241.0655.

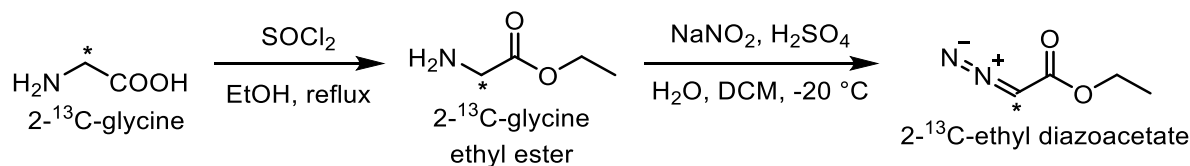
Synthesis and characterization of ethyl 2-(1-acetamido-2-(ethoxycarbonyl)cyclopropyl)thiazole-4-carboxylate (5).



Compound **5a** was synthesized on a preparative scale according to a previously reported procedure using P450<sub>TbJ2</sub>-T247A and purified using Semi-Prep HPLC Method D.<sup>7</sup>  $^1\text{H}$  NMR (600 MHz,  $\text{CDCl}_3$ )  $\delta$  1.28 (t,  $J$ =7.2 Hz, 3 H), 1.38 (t,  $J$ =7.2 Hz, 4 H), 1.88 (dd,  $J$ =7.2, 5.3 Hz, 1 H), 2.06 (s, 3 H), 2.24 (q,  $J$ =5.2 Hz, 1 H), 2.98 (dd,  $J$ =9.2, 7.3 Hz, 1 H), 4.11 - 4.18 (m, 2 H), 4.35 - 4.41 (m, 2 H), 6.51 (s, 1 H), 8.02 (s, 1 H);  $^{13}\text{C}$  NMR (151 MHz,  $\text{CDCl}_3$ )  $\delta$  14.05, 14.30, 23.05, 24.71, 31.38, 41.22, 61.35 (2 C), 127.27, 146.87, 161.17, 169.72, 171.3, 172.36. Expected  $[\text{M}+\text{H}]^+$  = 327.1009; Observed  $[\text{M}+\text{H}]^+$  = 327.1057.

Compound **5b** was synthesized on a preparative scale according to a previously reported procedure using CYP119-T213A-C317S and purified using Semi-Prep HPLC Method D.<sup>7</sup>  $^1\text{H}$  NMR (600 MHz,  $\text{CDCl}_3$ )  $\delta$  1.14 (t,  $J$ =7.2 Hz, 3 H), 1.38 (t,  $J$ =7.2 Hz, 4 H), 1.83 (dd,  $J$ =8.8, 5.9 Hz, 1 H), 1.94, (s, 3 H), 2.29 (t,  $J$ =6.3 Hz, 1 H), 2.35 (dd,  $J$ =8.9, 7.2 Hz, 1 H), 4.00 (q,  $J$ =7.2 Hz, 2 H), 4.39 (q,  $J$ =7.2 Hz, 2 H), 7.19 (br. s, 1 H), 8.12 (s, 1 H);  $^{13}\text{C}$  NMR (151 MHz,  $\text{CDCl}_3$ )  $\delta$  14.07, 14.39, 21.79, 23.07, 30.16, 39.59, 61.19, 61.37, 128.79, 145.91, 161.19, 167.78, 169.22, 170.59. Expected  $[\text{M}+\text{H}]^+$  = 327.1009; Observed  $[\text{M}+\text{H}]^+$  = 327.1068.

### Synthesis and characterization of $^{13}\text{C}$ -labelled EDA.



Ethyl 2- $^{13}\text{C}$ -diazoacetate was synthesized by adapting the procedure described by Ranocchiari and Mezzeti.<sup>45</sup> 2- $^{13}\text{C}$ -glycine (Sigma, 0.30 g, 4.21 mmol) was suspended in 30 mL ethanol under nitrogen, and the suspension was cooled in an ice-water-salt bath. Thionyl chloride (5.47 mmol, 0.397 mL) was added dropwise using a syringe and a needle. The flask was allowed to warm to room temperature. The mixture was then refluxed for 2 hours. After cooling the reaction mixture to room temperature, the solvent was removed under reduced pressure. The resulting white solid (ethyl 2- $^{13}\text{C}$ -glycine hydrochloride) was dried *in vacuo* and recrystallized from ethanol. Yield: 0.22 g (40%).

Ethyl 2- $^{13}\text{C}$ -glycine hydrochloride (0.21 g, 1.49 mmol) was mixed with water (2 mL) and dichloromethane (3 mL) in a 10-mL flask with a magnetic stirrer under nitrogen. The flask was placed in a freezing mixture of ice, water and salt. Sodium nitrite (0.124 g, 1.8 mmol) was dissolved in 1 mL water and the solution was cooled to 0 °C separately, and then added to the flask dropwise. After stirring the mixture for ~10 minutes after complete addition of sodium nitrite, 5%  $\text{H}_2\text{SO}_4$  solution (0.143 mL, 0.073 mmol) was added slowly in a dropwise manner. The reaction mixture was allowed to stir for 30 minutes, during which it turns progressively yellow, while ensuring that the temperature of the bath was maintained below 0 °C throughout. The reaction mixture was then poured into an ice-cold separating funnel. The yellow organic phase was recovered, and the aqueous phase was extracted with ice-cold dichloromethane (2 x 4 mL). The combined organic phase was then washed with 5% aqueous solution of sodium carbonate (3 x 4 mL). The organic phase was set aside, and the aqueous phase was washed with

dichloromethane (2 x 4 mL). Organic phases were combined and dried over Na<sub>2</sub>SO<sub>4</sub>. The solvent was carefully removed under mildly reduced pressure in a rotary evaporator. The synthesized EDA was further purified using silica gel flash chromatography and eluted using 1:9 ethyl acetate:hexane solvent mixture. Purified EDA was carefully concentrated *in vacuo* by repeated dilutions with dichloromethane. Yield ~ 0.178 g of 50% EDA solution in CH<sub>2</sub>Cl<sub>2</sub> (<sup>1</sup>H-NMR) (~52% yield). <sup>1</sup>H-NMR (400 MHz; CDCl<sub>3</sub>) δ 4.72 (1H, d, <sup>1</sup>J<sub>C-H</sub> 204 Hz, N<sub>2</sub><sup>13</sup>CH), 4.20 (2H, q, <sup>2</sup>J<sub>H,H'</sub> 7.0 Hz, -OCH<sub>2</sub>Me), 1.26 (3H, t, <sup>2</sup>J<sub>H,H'</sub> 7.1 Hz, -OCH<sub>2</sub>CH<sub>3</sub>); <sup>13</sup>C NMR (400 MHz, CDCl<sub>3</sub>, 298 K) δ 183.9, 60.82, 46.10 (s, N<sub>2</sub><sup>13</sup>CH), 14.39.

#### Solid phase synthesis of linear thiopeptides

Peptide synthesis was carried out using solid-phase peptide synthesis on a Biotage Initiator+ Alstra microwave peptide synthesizer using Rinkamide Chemmatrix resin (Biotage). Peptidic scaffolds were synthesized from the C-terminus to the N-terminus with a C-terminal amide functional group using standard fluorenylmethoxycarbonyl (Fmoc) chemistry. The building blocks comprised of commercially available N-α-Fmoc-L-amino acids, and chemically synthesized N-Fmoc-L-amino acid-thiazole building blocks. The detailed synthetic procedures for N-Fmoc-L-amino acid-thiazole building blocks are described elsewhere.<sup>33</sup> The solid-phase synthesis involved swelling the resin (15 mg) in 1.5 mL DMF at 70 °C for 20 minutes. (i) Coupling: Fmoc-building block (5.0 equiv., 0.4 M in DMF), HATU (5.0 equiv., 0.2 M in DMF), and DIEA (10.0 equiv., 0.2 M in DMF) were added to the swollen resin in that order. The resulting suspension was heated under microwave irradiation: 5 min at 75 °C. The reaction vessel was then drained, and the resin was thoroughly washed with DMF four times. Acetic anhydride was used for end-capping the N-terminus of core peptides. Acetic anhydride (10 equiv., 5 M in DMF) and DIEA (10 equiv., 0.5 M in DMF) were added to the reaction vial, and

the suspended resin was incubated at room temperature for 10 minutes followed by draining and washing as above. The bisthiazole-containing building block was coupled using TBTU (5 equiv., 0.2 M in DMF) and HOBt (5 equiv., 0.2 M in DMF) instead of HATU. (ii) *Fmoc removal*: Following coupling at each step, excess 20% piperidine was added to the reaction vessel and allowed to incubate for 3 min while stirring at room temperature. The reaction vessel was drained, washed with DMF, and excess 20% piperidine was introduced again and allowed to stir at room temperature for 10 min. The reaction vessel was drained, and the resin was thoroughly washed with DMF four times. After the final deprotection to obtain a primary amine at the N-terminus of the synthesized leader-peptide containing peptide or the end-capping of the core peptides using acetic anhydride, the resin was washed with dichloromethane and dried. The final peptide product was cleaved using a cleavage cocktail containing 94 - 95% (v/v) TFA, 2.5% (v/v) water, 2.5% (v/v) triisopropylsilane (TIPS), and 1 or 0% (v/v) ethanedithiol (EDT). EDT was added when the peptide sequence contained at least one methionine. Cleavage reactions were carried out by shaking the resin with the cleavage cocktail at 37 °C for one hour followed by room temperature incubation for about 30 minutes in syringes fitted with frits. The cleaved peptide was precipitated by the dropwise addition of the cleavage reaction mixture into cold diethyl ether. The diethyl ether was evaporated under vacuum, and the dried solid was dissolved in DMSO and diluted with a mixture of 50% v/v acetonitrile-water. This crude product was then purified using semi-preparatory scale HPLC, and the fractions were analyzed using the UV trace at 220 nm, 254 nm, and LC-MS. In general, the synthesized peptides eluted in a range between 50%-80% buffer B. Fractions containing the relatively pure product were dried using lyophilization. The dried product was then weighed and stored at -80 °C. The peptide products were dissolved in DMSO to give a 2 mM stock solution for further use.

## Conversion of thiopeptide cysteines to dehydroalanines

The elimination of hydrogen sulfide from cysteine side chains to obtain dehydroalanines was adapted from a previously published procedure.<sup>46</sup> In brief, cysteine-containing linear thiopeptide (1 mM) was incubated at 37 °C in an aqueous solution containing 50% DMSO, 4 mM TCEP, and 100 mM K<sub>2</sub>CO<sub>3</sub> for 15 minutes. 100 mM of 2,5-dibromopentanoic acid was then added to a final concentration of 100 mM, and the reaction mixture was incubated at 37 °C for 3 hours with continuous shaking. Excess unreacted dibromopentanoic acid was removed by washing the reaction with diethyl ether (6:1 ratio of volumes of diethyl ether to the reaction mixture, at least three washes per reaction), which included vortexing, centrifugation, and careful removal of the top diethyl ether layer. Residual amounts of diethyl ether after the final wash were removed by keeping the reaction vial open at 37 °C for 10-15 minutes. The remaining reaction mixture was then lyophilized overnight. DMSO was added to the dried solid residue to yield a 2 mM stock solution of the peptide dehydroalanine product (based on the amount of peptide originally taken) followed by centrifugation to separate undissolved solids. This stock solution was used as is for peptide characterization and enzymatic reactions.

### *Enzymatic Reaction Protocols*

#### Enzymatic heterocyclization using pyridine synthase TbtD

Dehydroalanine-containing linear substrate with the 15-residue leader peptide (20 μM) was incubated with 10 μM TbtD enzyme at room temperature for 16 hours in a solution containing 50 mM HEPES pH 7.5, and 250 mM NaCl. The reaction mixture was frozen and lyophilized. The solid residue was treated with methanol washes (250 μL x 2) with vortexing, followed by centrifugation to separate undissolved solids. The methanolic solution was dried *in vacuo*, and the solid residue obtained was redissolved in DMSO to give a stock solution of the

dehydroalanine-containing analog of thiomuracin. The aqueous reaction mixture, and the methanol and DMSO solutions were analyzed by LC-MS. Impurities included the unreacted linear substrate as well as the cleaved leader peptide carboxamide, a known byproduct of the TbtD reaction.

#### *Small-scale cyclopropanation reaction setup*

##### Methyl 2-benzamidoacrylate

Small-scale 400  $\mu$ L reactions with methyl 2-benzamidoacrylate and EDA were setup following conditions previously reported for the cyclopropanation of styrene with EDA.<sup>22</sup> A 1.2 M stock solution of methyl 2-benzamidoacrylate was prepared in DMSO, and a 0.4 M stock solution of ethyl diazoacetate was prepared in ethanol. The final concentrations in 400  $\mu$ L were as follows: 20  $\mu$ M enzyme, 10 mM ethyl diazoacetate, 30 mM methyl 2-benzamidoacrylate, 10 mM sodium dithionite, and 0.1 M KPi pH 8.0. After 16 h, 20  $\mu$ L of 100 mM diethyl benzamidomalonate (internal standard) was added. The reaction mixture was extracted with 1 mL of ethyl acetate, transferred to a 1.7 mL microcentrifuge tube, and centrifuged for 5 min at 18400 xg. The organic layer was diluted 1:1 with 50% ethanol, and 20  $\mu$ L of the solution was analyzed using LC-MS Method A.

##### Ethyl 2-(1-acetylaminoethenyl)thiazole-4-carboxylate

Small-scale 200  $\mu$ L reactions with ethyl 2-(1-acetylaminoethenyl)thiazole-4-carboxylate and EDA were setup following conditions previously reported for the cyclopropanation of styrene with EDA.<sup>22</sup> A 0.4 M stock solution of ethyl 2-(1-acetylaminoethenyl)thiazole-4-carboxylate was prepared in DMSO, and a 0.4 M stock solution of ethyl diazoacetate was prepared in ethanol. The final concentrations in 200  $\mu$ L were as follows: 20  $\mu$ M enzyme, 10 mM ethyl diazoacetate, 10 mM ethyl 2-(1-acetylaminoethenyl)thiazole-4-carboxylate, 10 mM sodium

dithionite, and 0.1 M KPi pH 8.0. After 16 h, 10  $\mu$ L of 100 mM diethyl benzamidomalonate (internal standard) was added. The reaction mixture was extracted with 1 mL of ethyl acetate, transferred to a 1.7 mL microcentrifuge tube, and centrifuged for 5 min at 18400 xg. The organic layer was collected in a separate microcentrifuge tube and evaporated to dryness. The dry material was dissolved in 1 mL of ethanol, and 1  $\mu$ L of the solution was analyzed using LC-MS Method D.

#### Thiomuracin derivatives

Milli-Q water and 0.1 M  $\text{Na}_2\text{S}_2\text{O}_4$  in 1 M KPi (pH 8) were placed into a separate sealed vial and degassed with argon for at least 10 minutes prior to use. Small-scale reactions (200  $\mu$ L) were conducted in 2-mL crimp vials (Agilent Technologies, San Diego, CA). The headspace of a crimp vial containing P450 and a stir bar was flushed with argon (no bubbling). Using glass syringes, water and P450 were added under argon to a final volume of 166  $\mu$ L followed by 20  $\mu$ L of 0.1 M  $\text{Na}_2\text{S}_2\text{O}_4$ . 10  $\mu$ L of a solution of thiomuracin derivative (2 mM in DMSO) was added and allowed to mix for at least 30 seconds before adding 5  $\mu$ L of EDA (0.4 M in EtOH). The argon lines were subsequently removed, and the reactions were left to stir for at least 16 hours. The final concentration of reagents in the reactions were as follows: 20  $\mu$ M P450, 100  $\mu$ M thiomuracin derivative, 10 mM EDA, 10 mM  $\text{Na}_2\text{S}_2\text{O}_4$ , 0.1 M KPi pH 8, and 7.5% cosolvent. The reactions with linear thiomuracin substrates were extracted with 1 mL of ethyl acetate, transferred to a 1.7 mL microcentrifuge tube, and centrifuged for 5 min at 18400 xg to remove precipitated protein. The organic layer was transferred to a separate microcentrifuge tube, evaporated to dryness, and dissolved in 200  $\mu$ L of DMSO prior to analysis using LC-MS Method A.

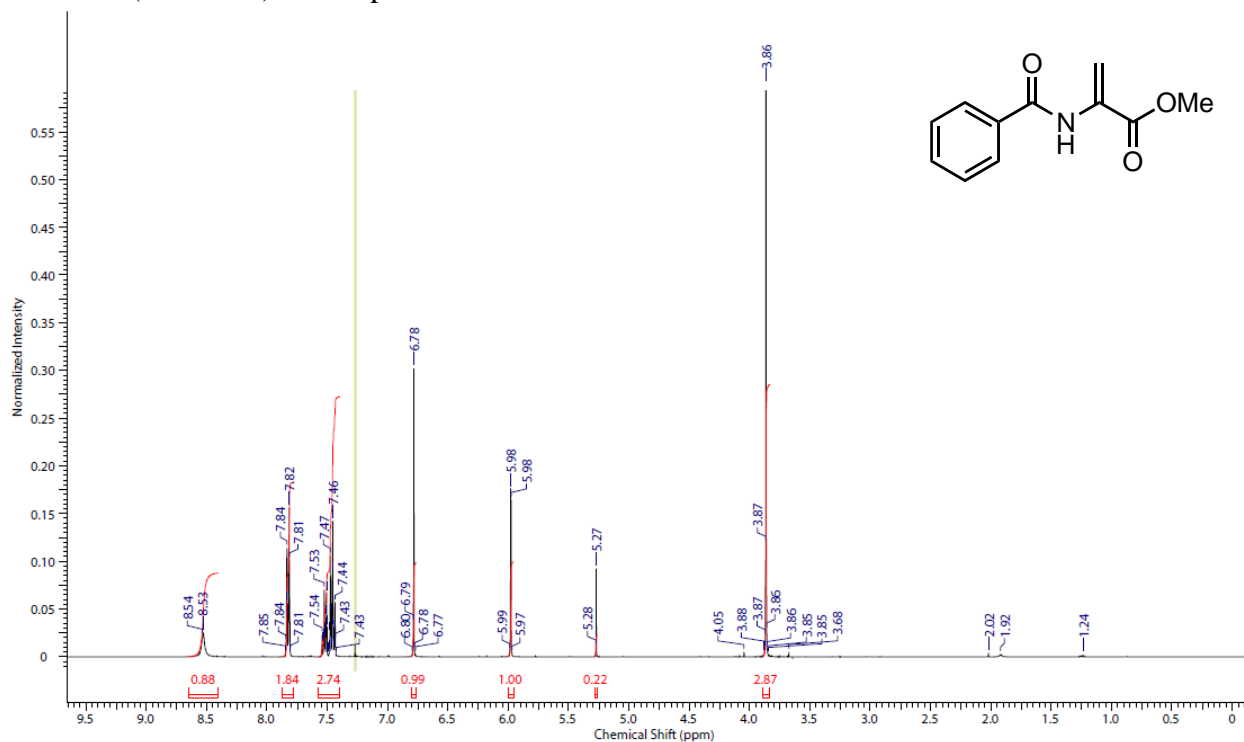


Small-scale 200  $\mu$ L reactions with traditional transition metal catalysts and I8Dha core were setup following a similar procedure to the one described above. The reaction with  $\text{Rh}_2(\text{OAc})_2$  in water was setup using previously reported reaction conditions.<sup>13</sup> Reactions with iron(III) tetraphenylporphyrin (FeTPP),  $\text{Rh}_2(\text{OAc})_2$ ,  $\text{Rh}_2(\text{Oct})_2$ , and  $\text{Rh}_2(\text{esp})_2$  were setup under argon in dichloromethane with the following final concentrations: 100  $\mu$ M rhodium (50  $\mu$ M FeTPP), 100  $\mu$ M thiomuracin derivative, 10 mM EDA, and 2.5% ethanol. For the reactions with rhodium catalysts, the thiomuracin derivative was added to the crimp vial in DMSO and then lyophilized prior to addition of catalyst and solvent (DMSO was found to be inhibitory). The aqueous reaction was worked up using the procedure described above for thiomuracin derivatives. For the reactions performed in dichloromethane, the reaction mixture was transferred to a microcentrifuge tube and evaporated to dryness. The remaining solid was dissolved in 200  $\mu$ L of DMSO prior to analysis using LC-MS Method A.

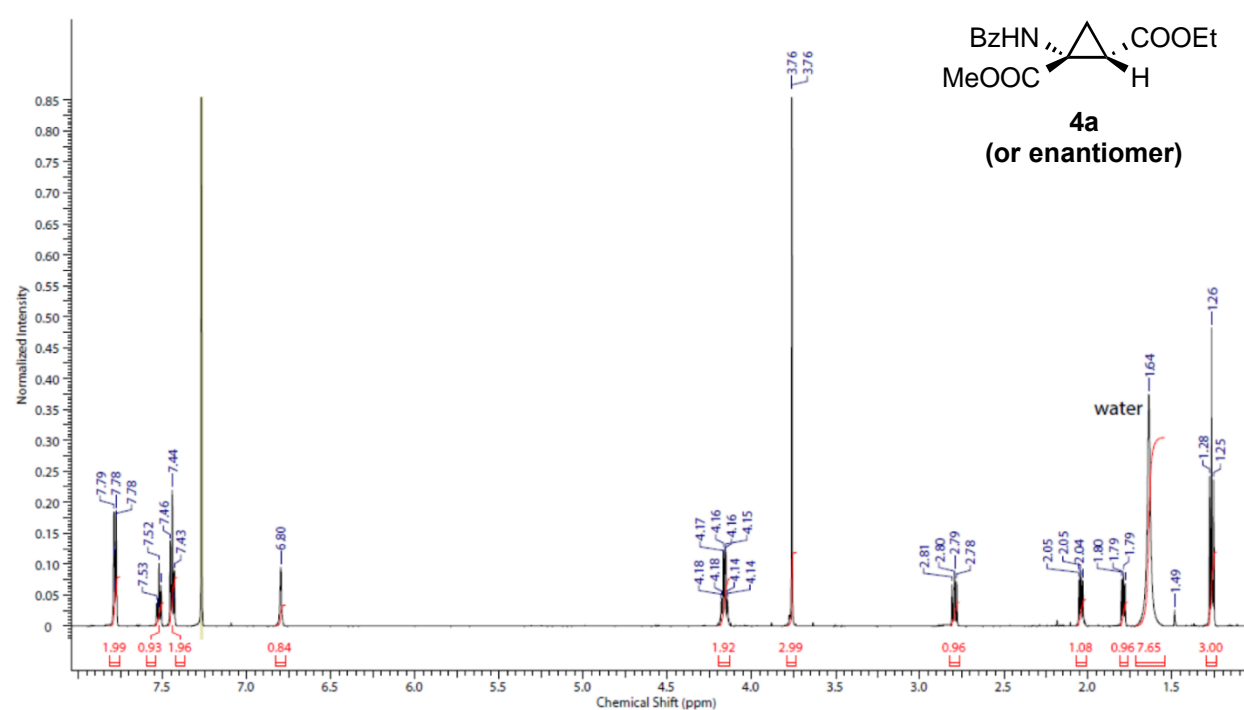
For the reactions with cyclic thiomuracin substrates, an equal volume (200  $\mu$ L) of methanol was added to precipitate protein. The mixture was centrifuged at 18400  $\times g$  for 5 min and then filtered through glass wool prior to analysis using LC-MS Method B.

## NMR Spectra

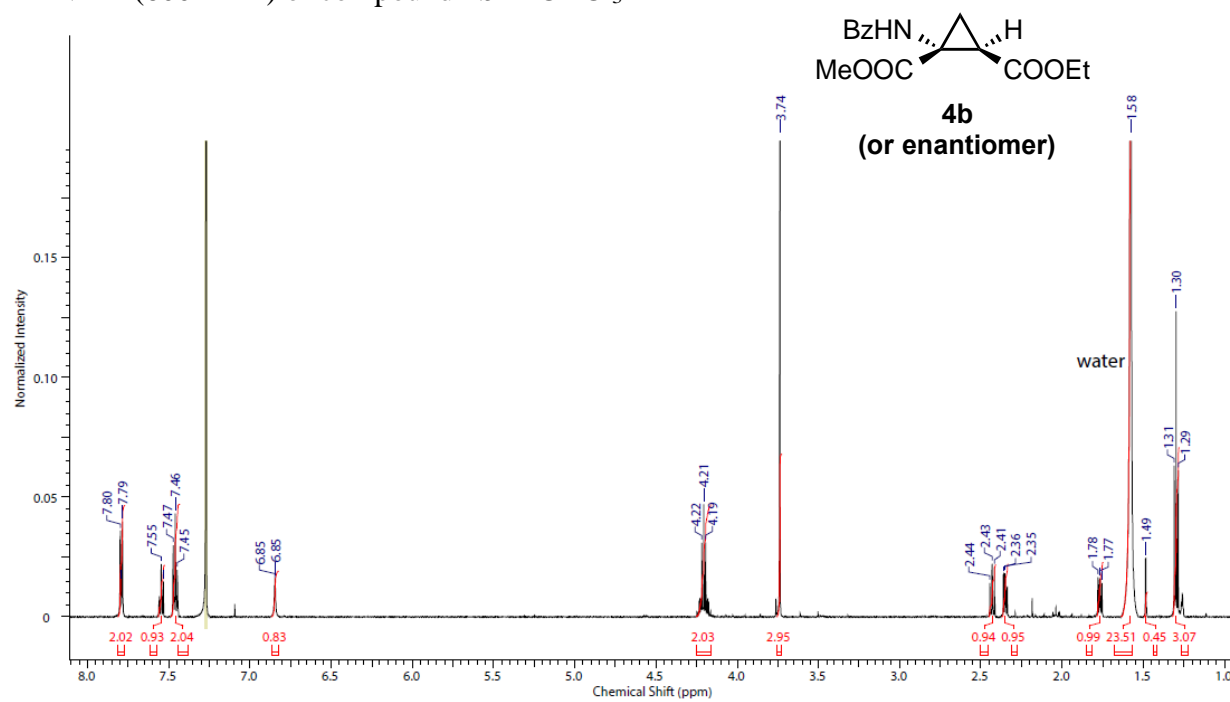
$^1\text{H}$  NMR (400 MHz) of compound **1** in  $\text{CDCl}_3$



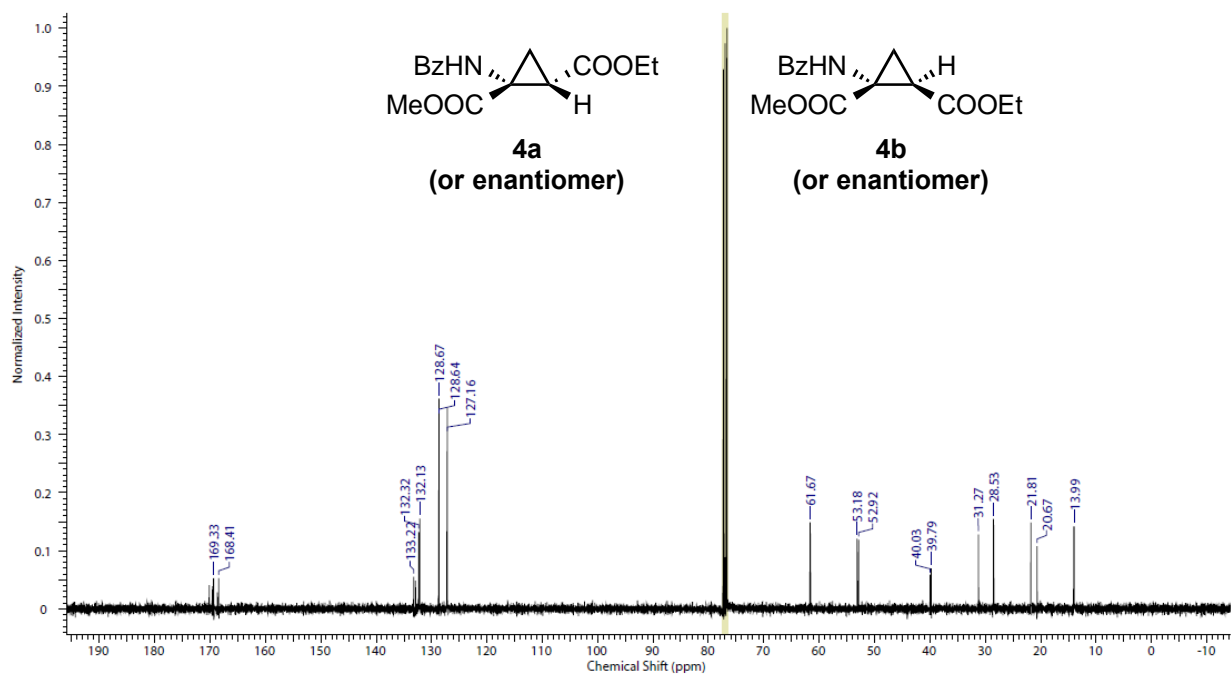
$^1\text{H}$  NMR (600 MHz) of compound **4a** in  $\text{CDCl}_3$



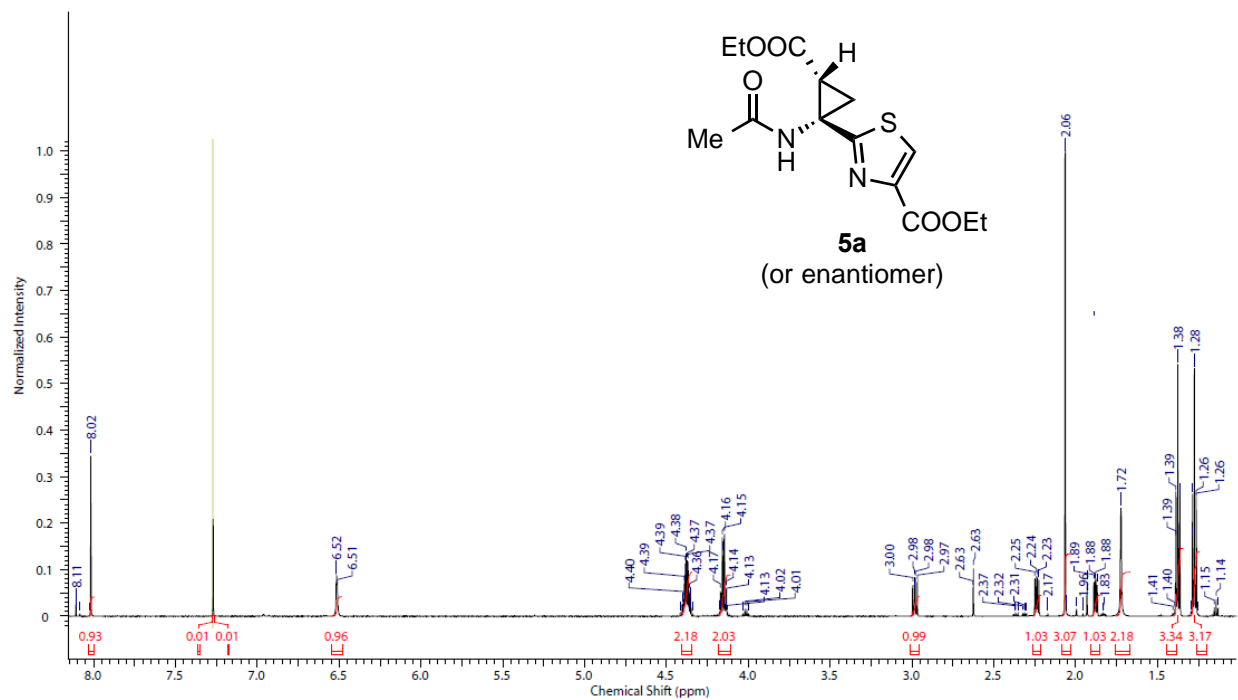
$^1\text{H}$  NMR (600 MHz) of compound **4b** in  $\text{CDCl}_3$



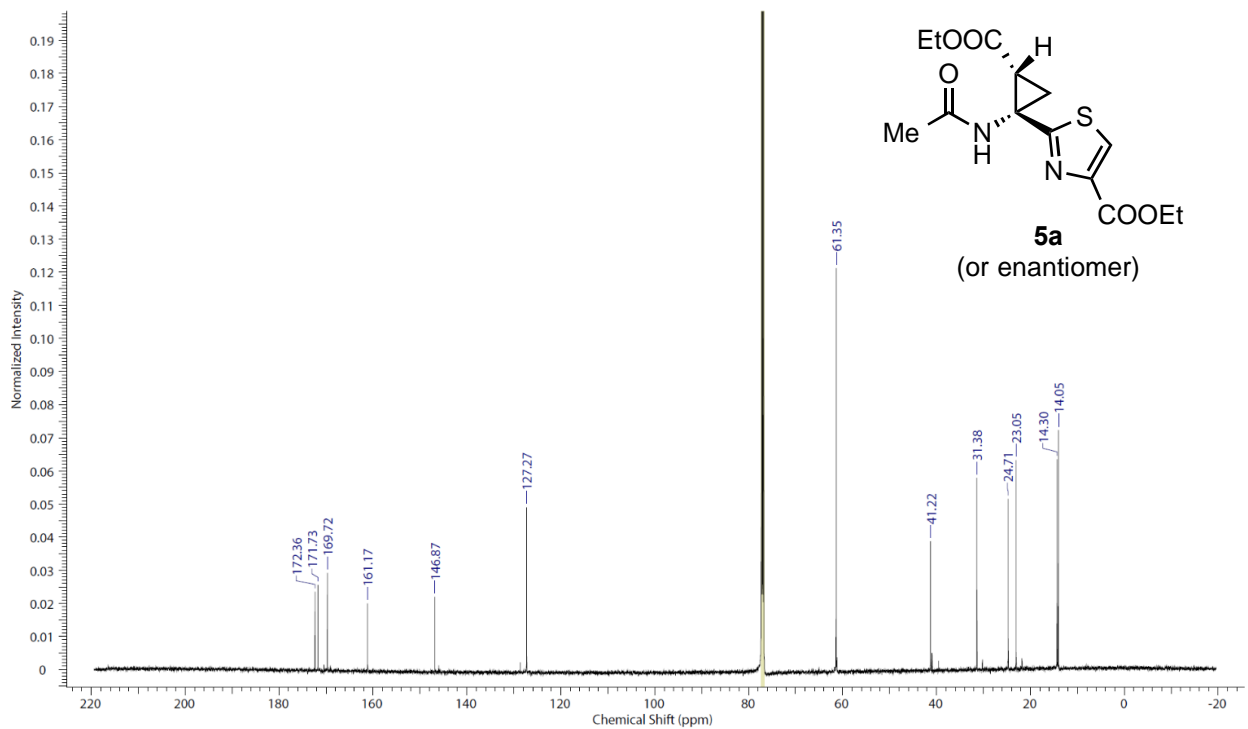
$^{13}\text{C}$  NMR (125 MHz) of compound **4** (mixture of diastereomers) in  $\text{CDCl}_3$



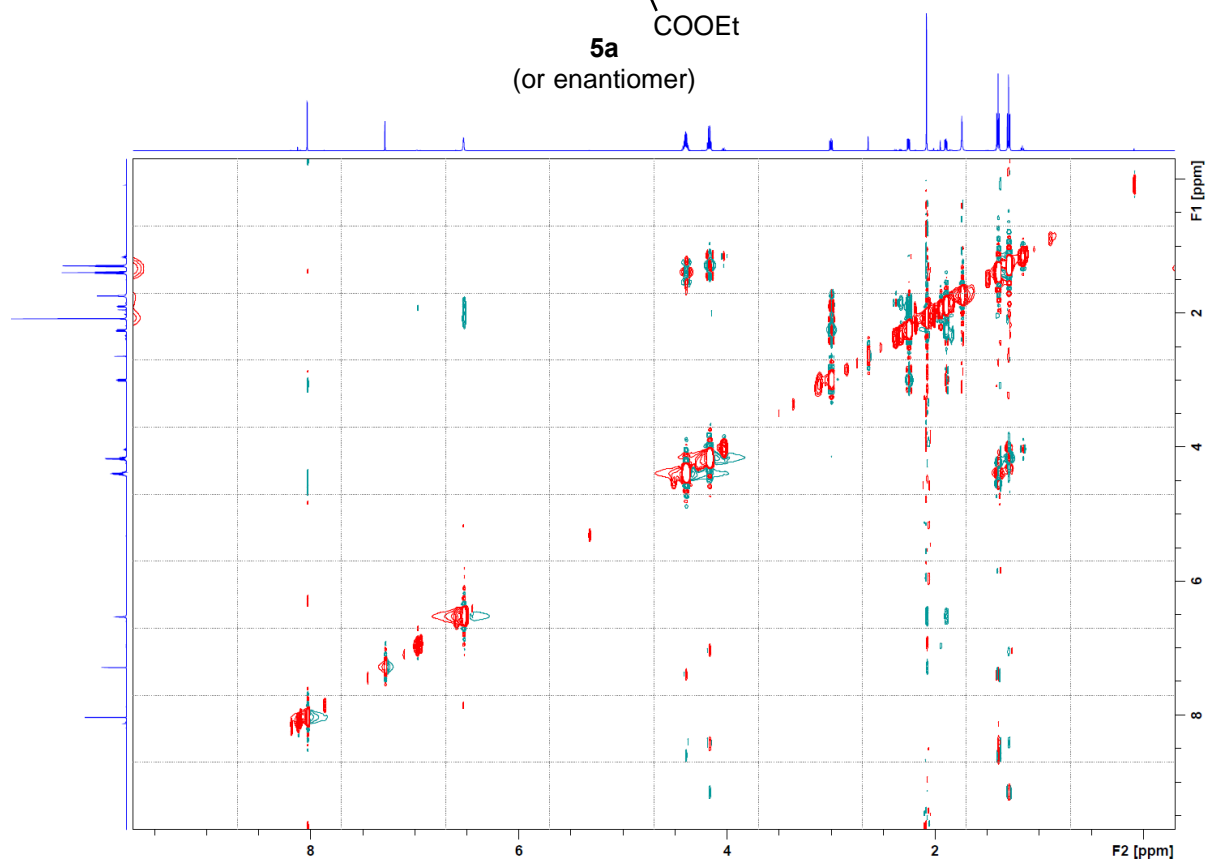
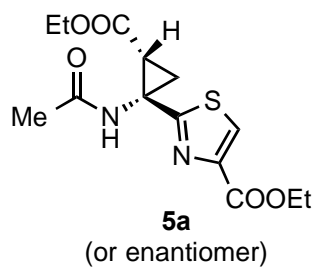
$^1\text{H}$  NMR (600 MHz) of **5a** in  $\text{CDCl}_3$ .



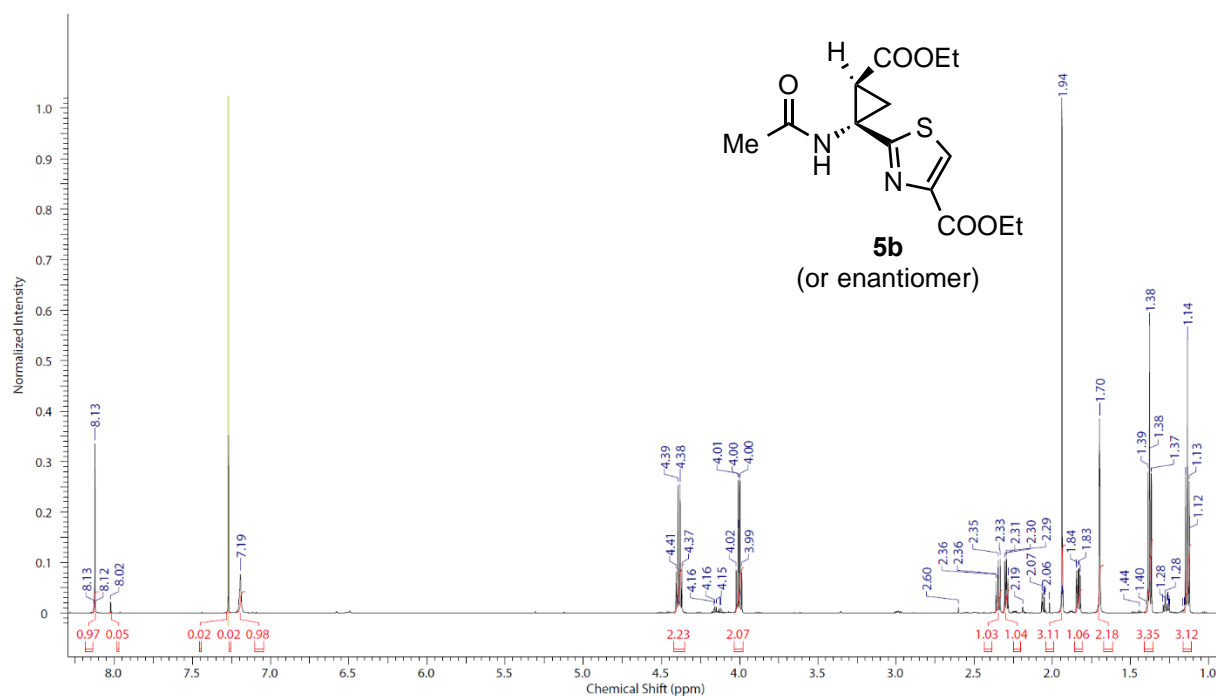
$^{13}\text{C}$  NMR (151 MHz) of **5a** in  $\text{CDCl}_3$ .



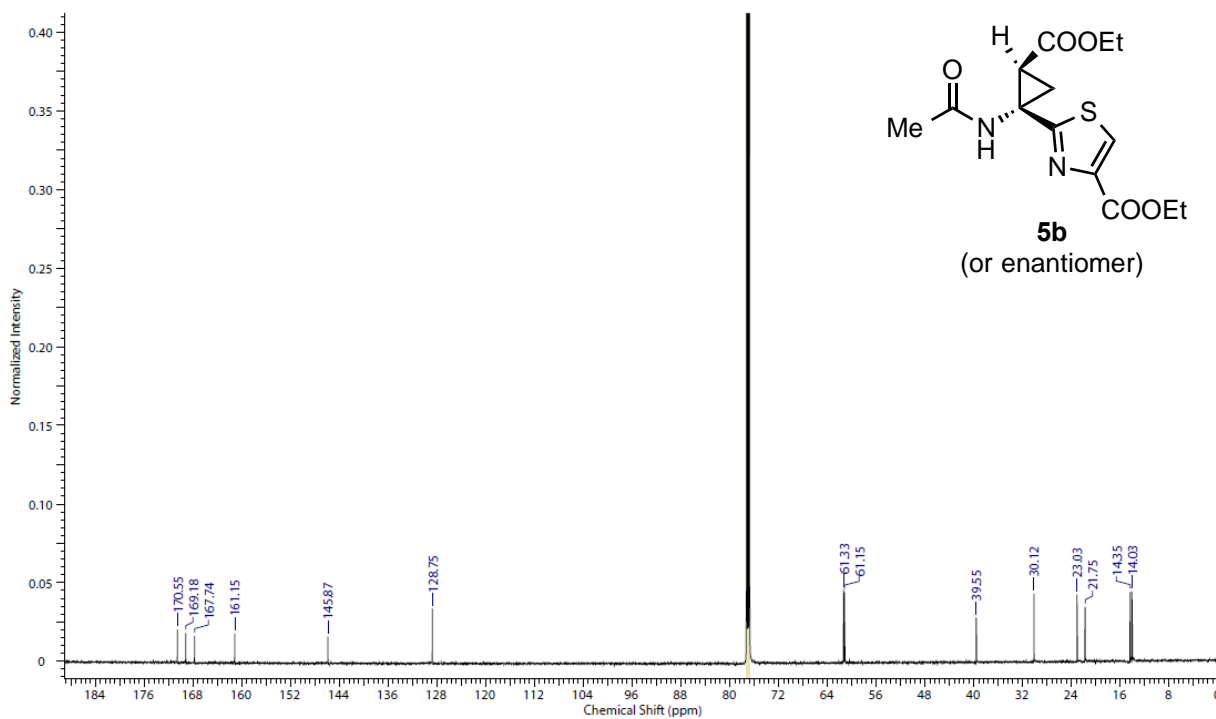
$^1\text{H}$ - $^1\text{H}$  NOESY (600 MHz) of **5a** in  $\text{CDCl}_3$ .



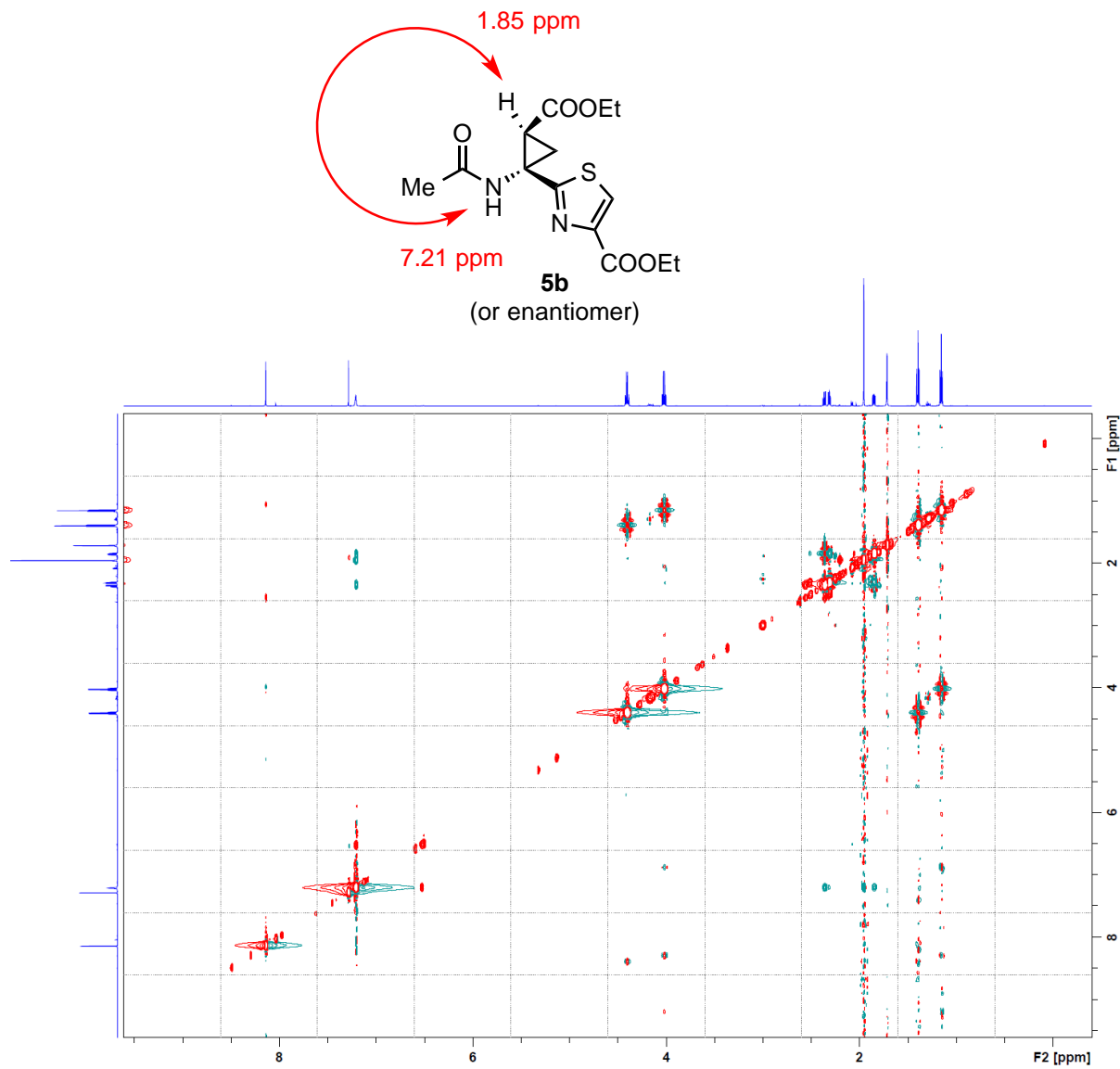
$^1\text{H}$  NMR (600 MHz) of **5b** in  $\text{CDCl}_3$ .



$^{13}\text{C}$  NMR (151 MHz) of **5b** in  $\text{CDCl}_3$ .



$^1\text{H}$ - $^1\text{H}$  NOESY (600 MHz) of **5b** in  $\text{CDCl}_3$ .



*Characterization of  $^{13}\text{C}$ -labelled cyclopropanated I8Dha mono core*

$^{13}\text{C}$ -labelled cyclopropanated I8-1Dha mono core was synthesized on a preparative scale according to a previously reported procedure using P450<sub>TbU2</sub>-T247A and 2- $^{13}\text{C}$ -EDA.<sup>10</sup> The isotopically labeled product was purified using Semi-Prep HPLC Method D.

**$^1\text{H}$ -NMR chemical shifts for  $^{13}\text{C}$ -labelled cyclopropanated I8Dha mono core**

Sr. no.	Chemical shift ( $\delta$ )	Multiplicity	Coupling constant ( $J$ , Hz)*
1	9.33	s	-
2	9.19	d	9.0
3	8.85	m	3.0
4	8.845	d	7.5
5	8.59	d	9.0
6	8.52	d	8.0
7	8	s	-
8	7.97	s	-
9	7.96	s	-
10	7.95	s	-
11	7.94	s	-
12	7.92	s	-
13	7.91	s	-
14	7.87	s	-
15	7.01	t	7.5
16	7	d	7.3
17	6.99	d	7.3
18	6.98	t	7.5
19	6.96	t	7.5
20	6.74	d	8.5
21	6.37	d	9.0
22	5.56	m	ND
23	5.48	q	7.9
24	5.35	m	ND
25	5.08	m	ND
26	4.58	s	-
27	3.99	q	7.3
28	3.74	m	ND
29	3.3	m	ND

\*ND: Not determined

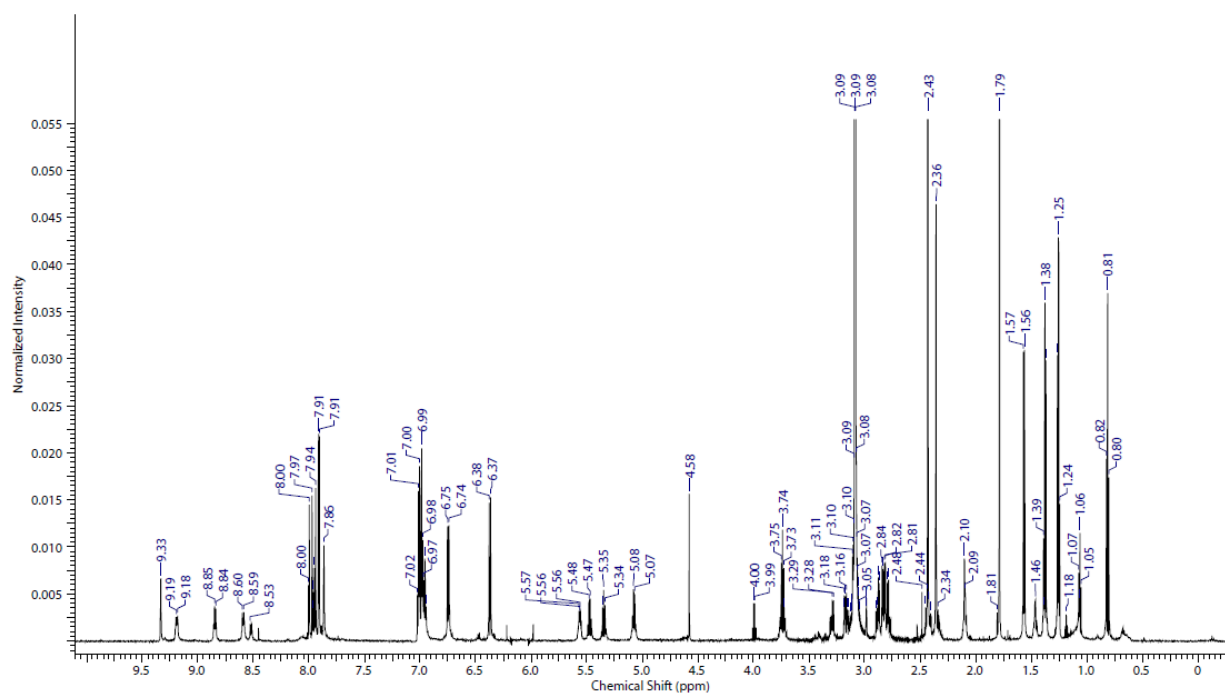
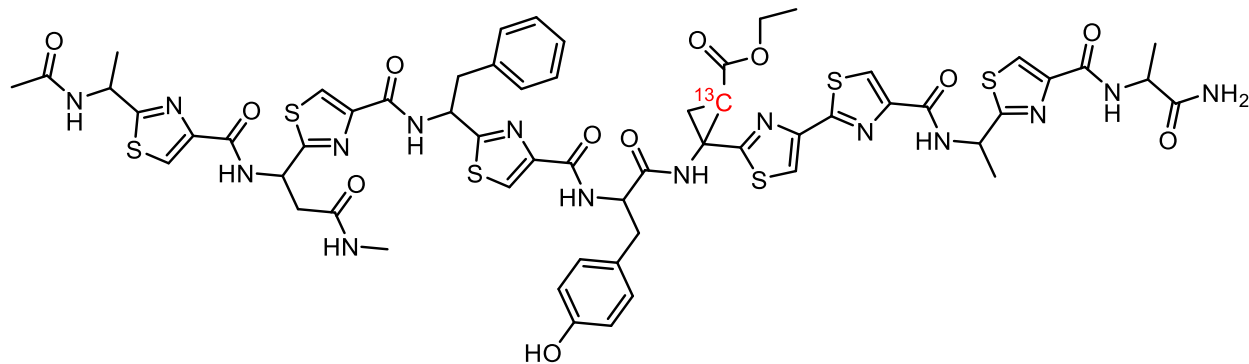


**<sup>1</sup>H-NMR chemical shifts for <sup>13</sup>C-labelled cyclopropanated l8Dha mono core (continued)**

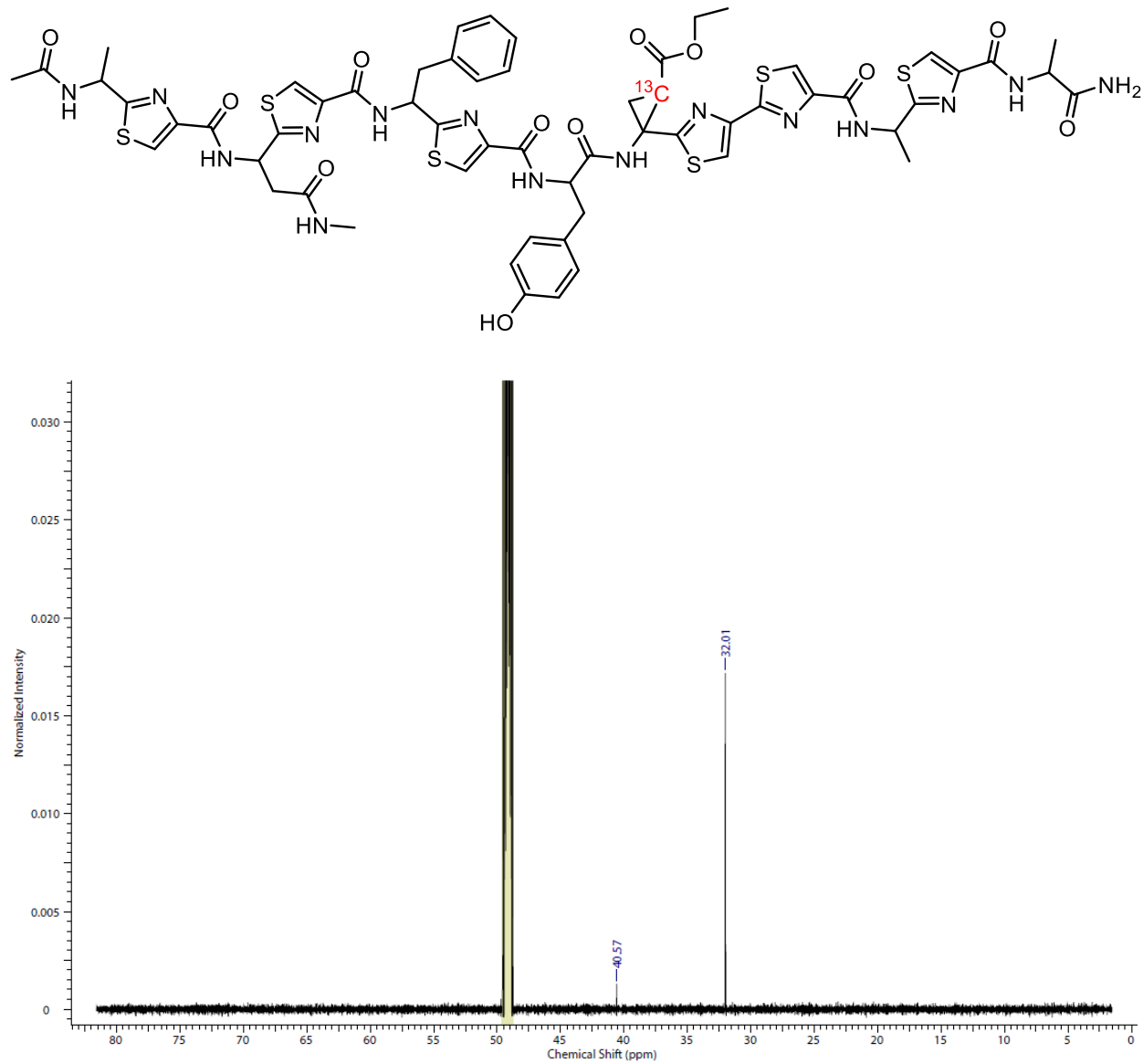
Sr. no.	Chemical shift (δ)	Multiplicity	Coupling constant (J, Hz)*
30	3.3	d	7.0
31	3.28	d	7.0
34	3.16	m	ND
35	3.15	m	ND
36	3.14	m	ND
37	2.98	quin	1.6
38	2.89	d	6.8
46	2.45	m	ND
47	2.41	m	ND
48	2.4	m	ND
49	2.34	m	ND
50	2.33	s	-
51	2.1	m	ND
52	1.79	s	-

\*ND: Not determined

$^1\text{H}$  NMR (700 MHz, 298.2 K) of  $^{13}\text{C}$ -labelled cyclopropanated I8-1Dha mono core in  $\text{CD}_3\text{OD}$ .

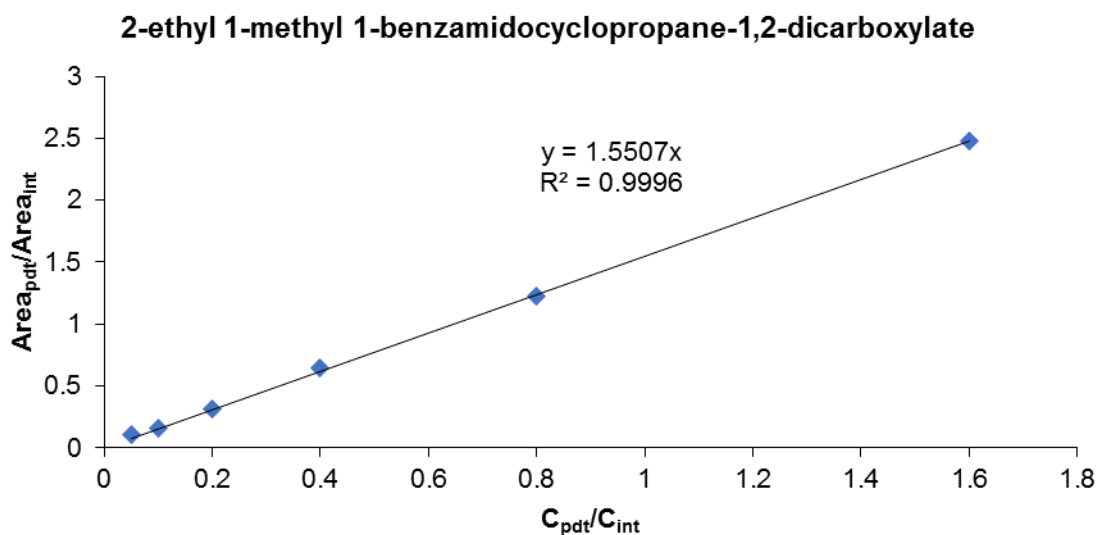


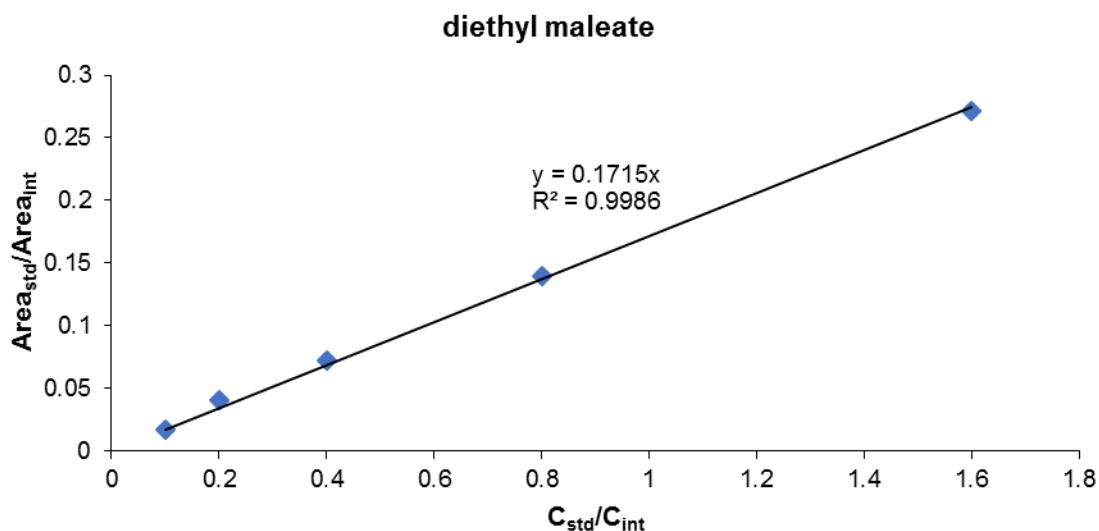
$^{13}\text{C}$  NMR (176 MHz, 298.2 K) of  $^{13}\text{C}$ -labelled cyclopropanated I8-1Dha mono core in  $\text{CD}_3\text{OD}$ .



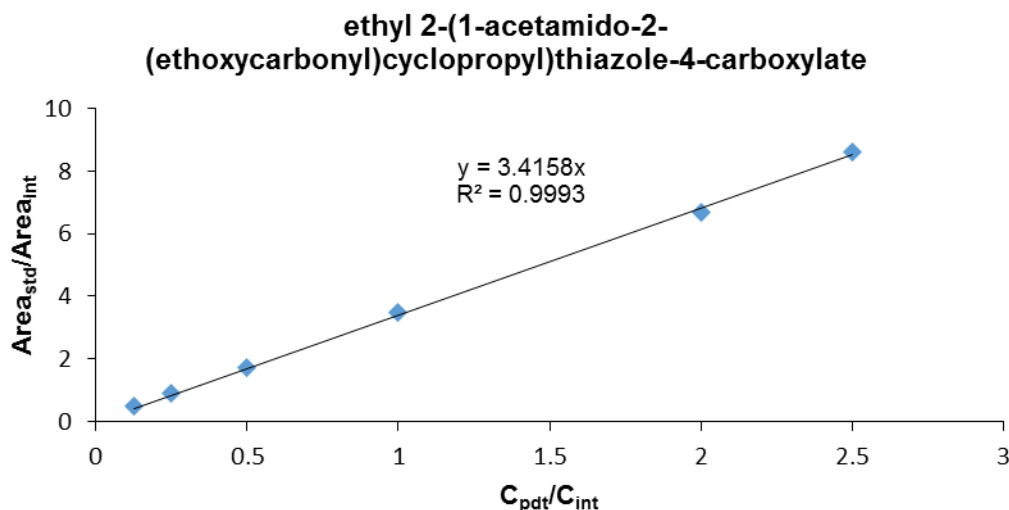
### Calibration curves

Calibration curves for 2-ethyl 1-methyl 1-benzamidocyclopropane-1,2-dicarboxylate and diethyl maleate were prepared according to the following procedure: A 0.2 M stock solution of product standard was used to prepare 400  $\mu$ L samples of product standard at 0.1 mM, 0.2 mM, 0.4 mM, 0.8 mM, 1.6 mM and 2 mM in 0.1 M potassium phosphate (pH 8.0) containing 2 mM internal standard (diethyl 2-benzamidomalonate) and 5% ethanol. The samples were extracted with 1 mL of ethyl acetate, transferred to a 1.7 mL microcentrifuge tube, and centrifuged for 5 min at 18400 xg. The organic layer was diluted 1:1 with 50% ethanol, and 20  $\mu$ L of the solution was analyzed using LC-MS Method A. The absorbance at 254 nm was used for integration.



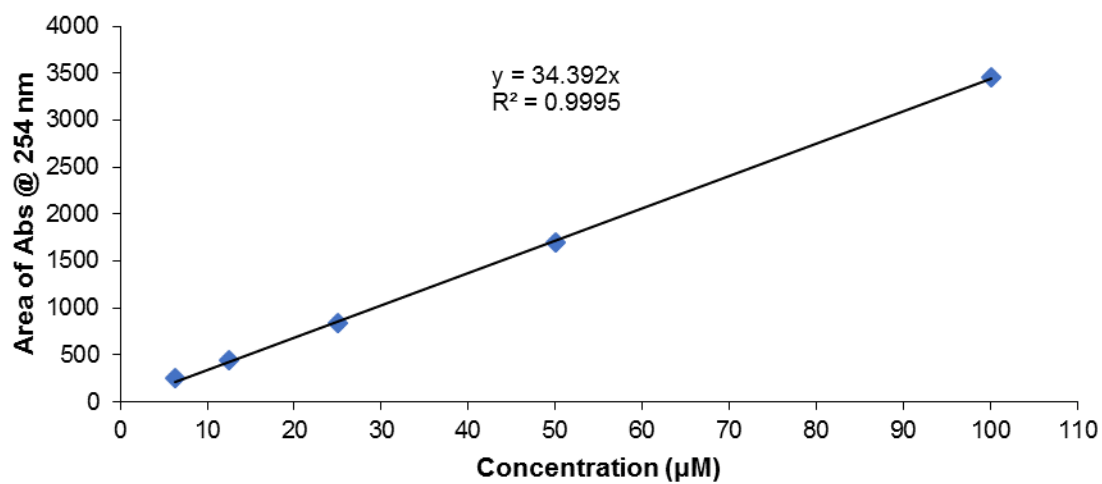


The calibration curve for ethyl 2-(1-acetamido-2-(ethoxycarbonyl)cyclopropyl)thiazole-4-carboxylate was prepared according to the following procedure: A 0.2 M stock solution of product standard was used to prepare 200  $\mu$ L samples of product standard at 0.1 mM, 0.2 mM, 0.4 mM, 0.8 mM, 1.6 mM and 2 mM in 0.1 M potassium phosphate (pH 8.0) containing 1 mM internal standard (diethyl 2-benzamidomalonate) and 5% DMSO. The samples were extracted with 1 mL of ethyl acetate, transferred to a 1.7 mL microcentrifuge tube, and centrifuged for 5 min at 18400 xg. The organic layer was collected in a separate microcentrifuge tube and evaporated to dryness. The dry material was dissolved in 1 mL of ethanol, and 1  $\mu$ L of the solution was analyzed using LC-MS Method D. The absorbance at 254 nm was used for integration.

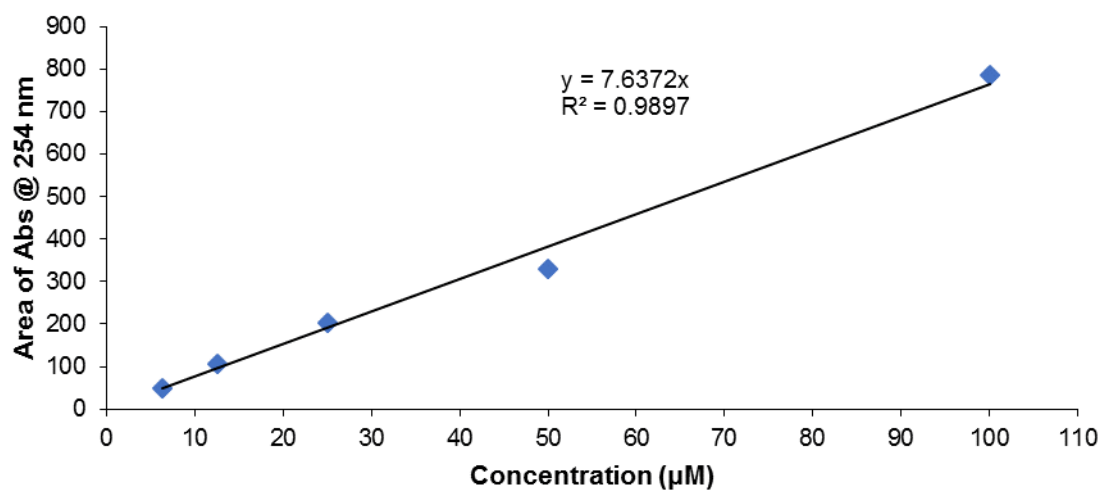


Calibration curves for TbtA F5Dha core and TbtA I8Dha core were prepared according to the following procedure: A 2 mM stock solution of thiomuracin derivative was used to prepare 200  $\mu\text{L}$  samples at 100  $\mu\text{M}$ , 50  $\mu\text{M}$ , 25  $\mu\text{M}$ , 12.5  $\mu\text{M}$ , and 6.25  $\mu\text{M}$  in 0.1 M potassium phosphate (pH 8.0) containing 5% ethanol and 20  $\mu\text{M}$  hemin. The samples were extracted with 1 mL of ethyl acetate, transferred to a 1.7 mL microcentrifuge tube, and centrifuged for 5 min at 18400 xg. The organic layer was then completely evaporated. The remaining pellet was dissolved in 200  $\mu\text{L}$  of DMSO, and 20  $\mu\text{L}$  of the resulting solution was analyzed using LC-MS Method A. The absorbance at 254 nm was used for integration.

**TbtA F5Dha core calibration curve**



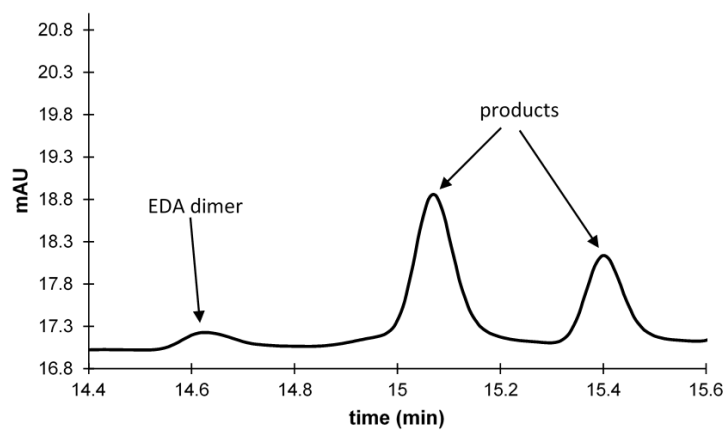
**TbtA I8Dha core calibration curve**



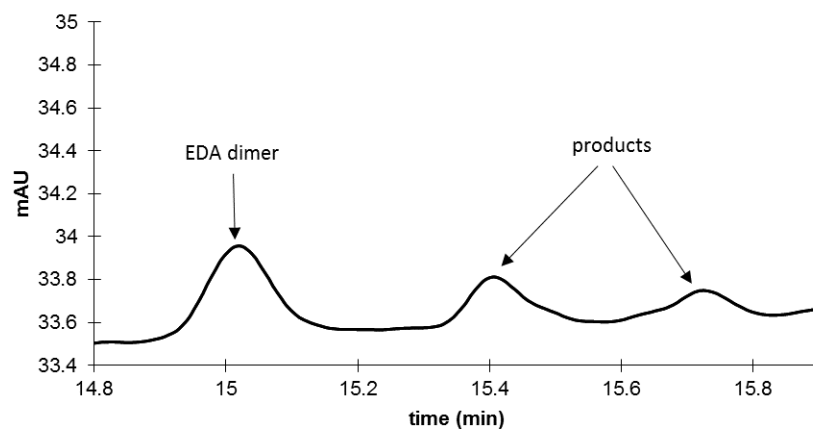
## Supporting Information

**Figure S3.1.** HPLC traces from enzymatic cyclopropanations of methyl 2-benzamidoacrylate (**1**) (N-Bz-Dha-COOMe).

### Hemin - N-Bz-Dha-COOMe

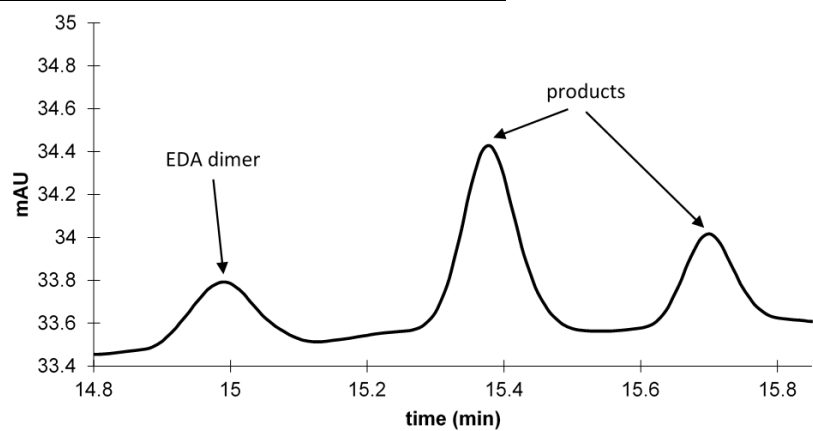


### P450<sub>BM3</sub>-T268A - N-Bz-Dha-COOMe

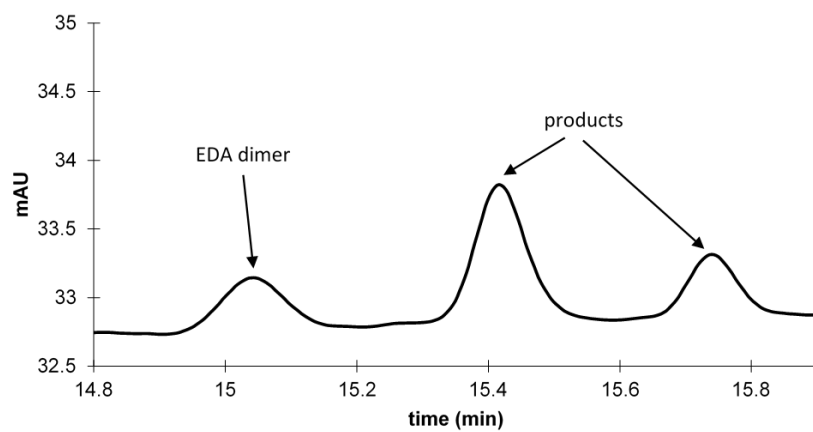




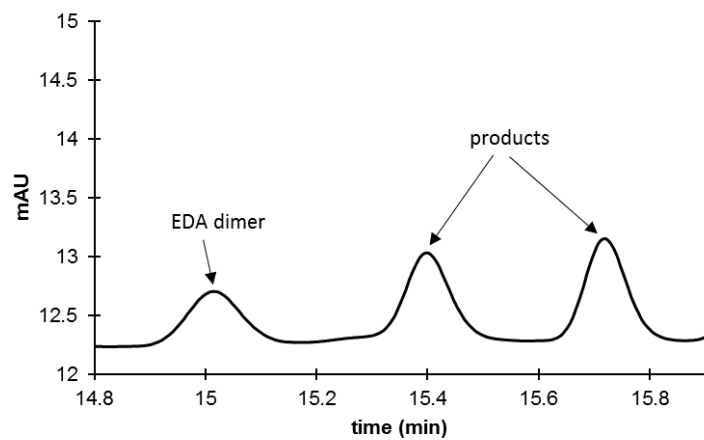
P450<sub>BM3</sub>-T268A/C400S - N-Bz-Dha-COOMe



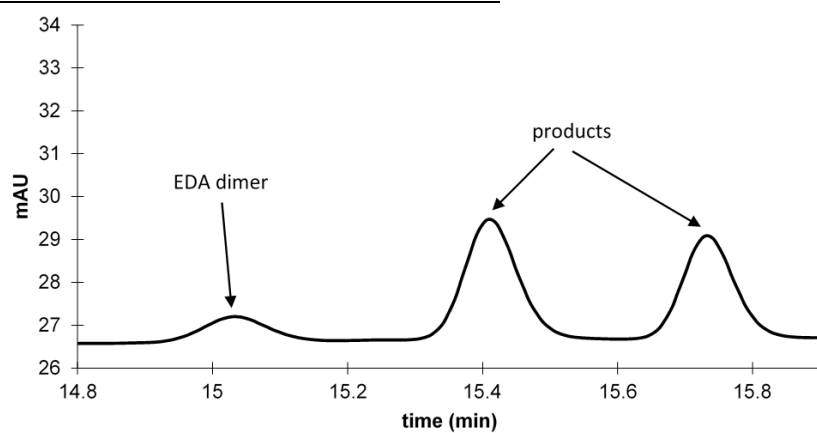
P450<sub>BM3-CIS</sub>-C400S - N-Bz-Dha-COOMe



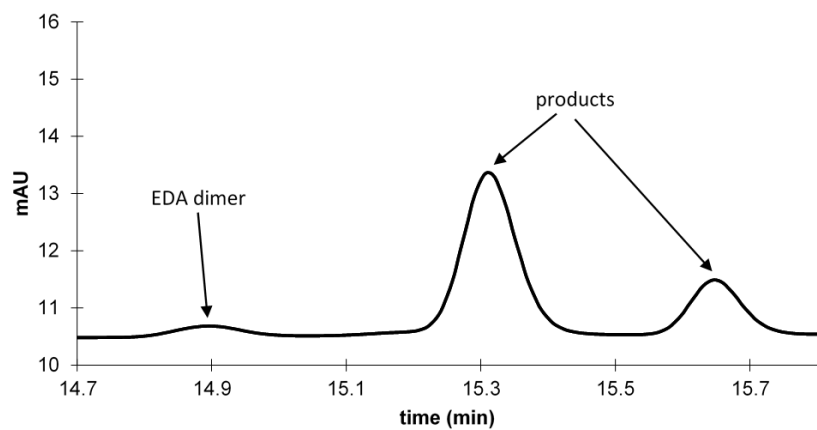
P450<sub>Biol</sub>-T238A - N-Bz-Dha-COOMe



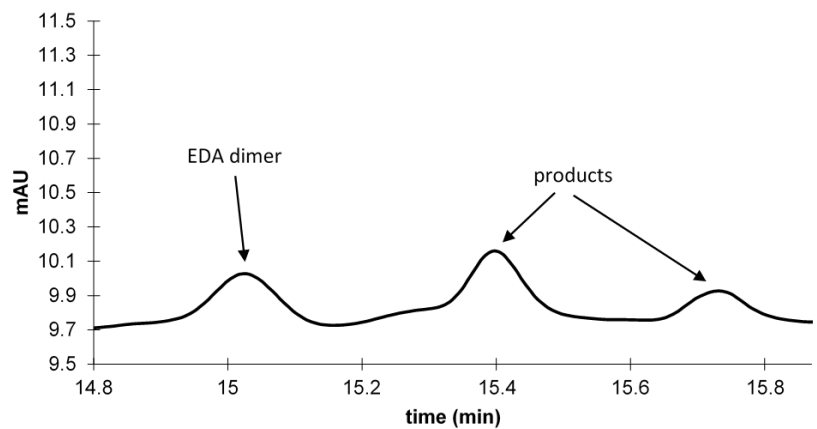
P450<sub>Biol</sub>-T238A/C344S - N-Bz-Dha-COOMe



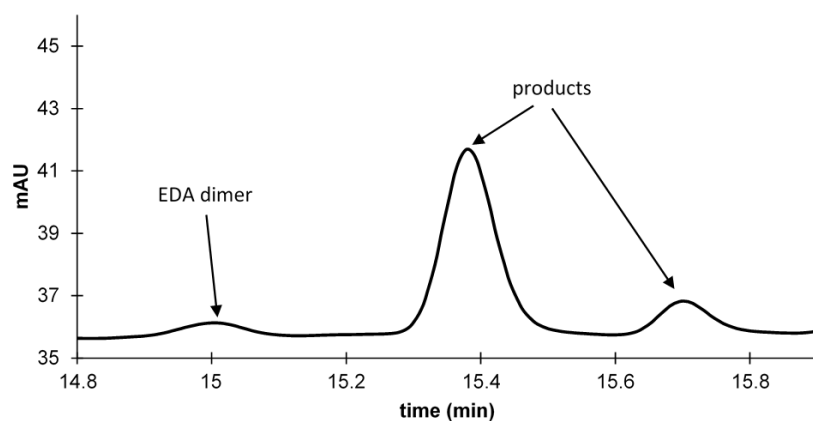
P450<sub>EryF</sub> - N-Bz-Dha-COOMe



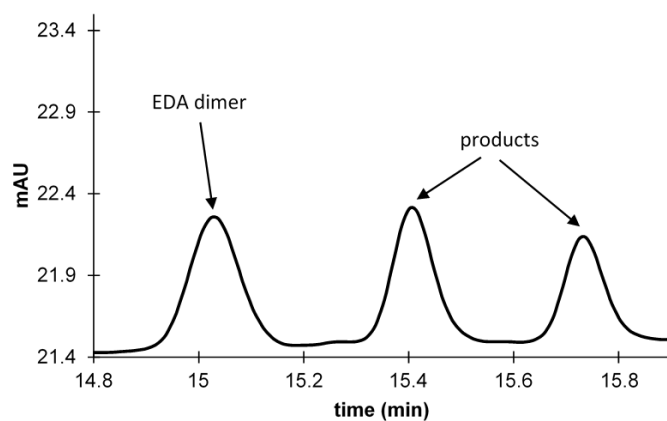
CYP142-T234A - N-Bz-Dha-COOMe



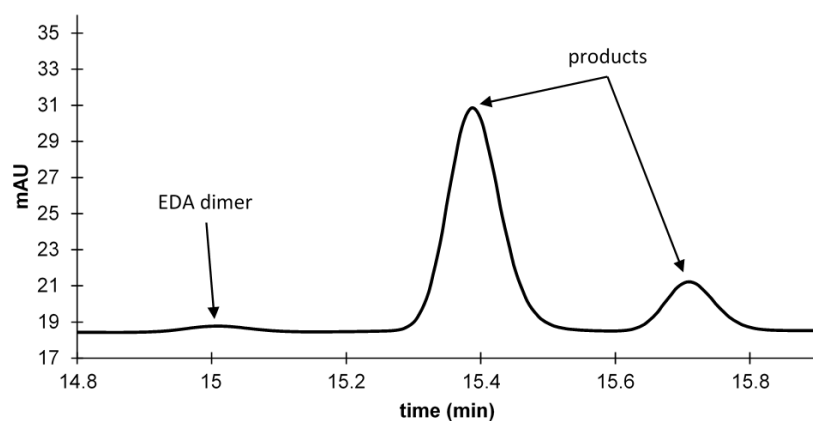
CYP119-T213A - N-Bz-Dha-COOMe



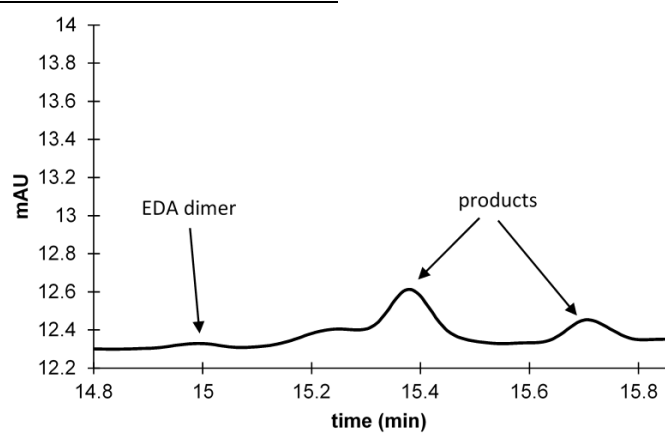
CYP119-T213A/C317S - N-Bz-Dha-COOMe



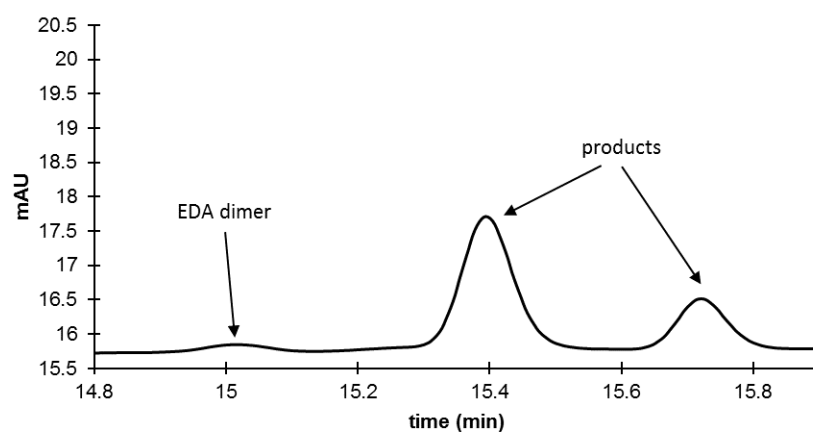
P450<sub>PikC</sub>-T247A - N-Bz-Dha-COOMe



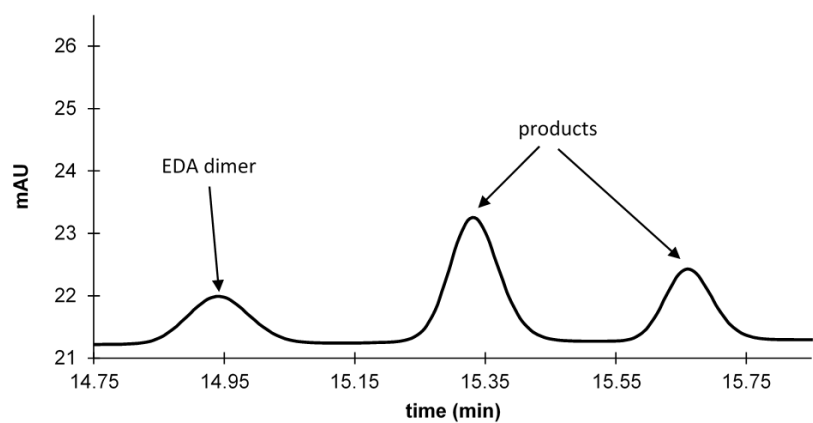
P450<sub>TbIJ1</sub> - N-Bz-Dha-COOMe



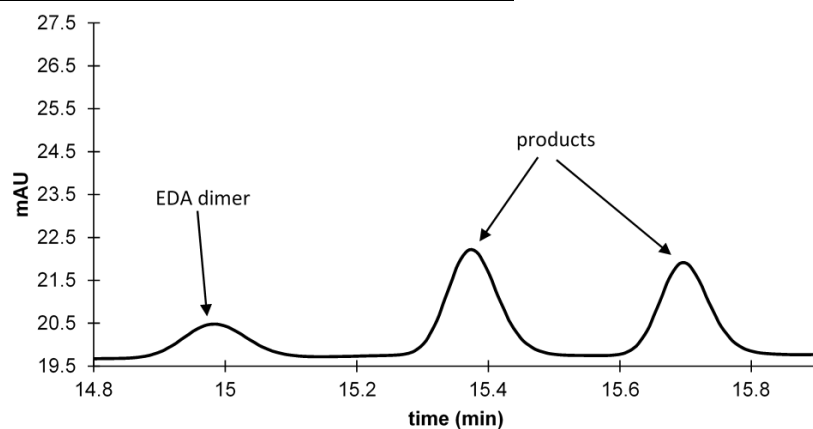
P450<sub>TbIJ1</sub>-T234A - N-Bz-Dha-COOMe



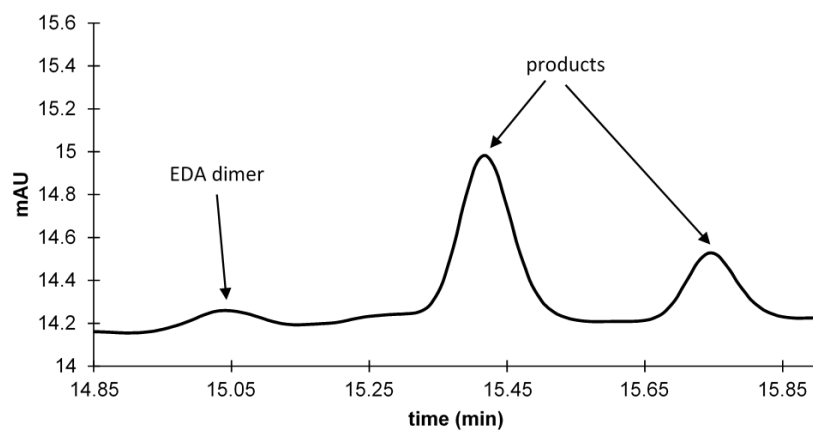
P450<sub>TbIJ1</sub>-C340S - N-Bz-Dha-COOMe



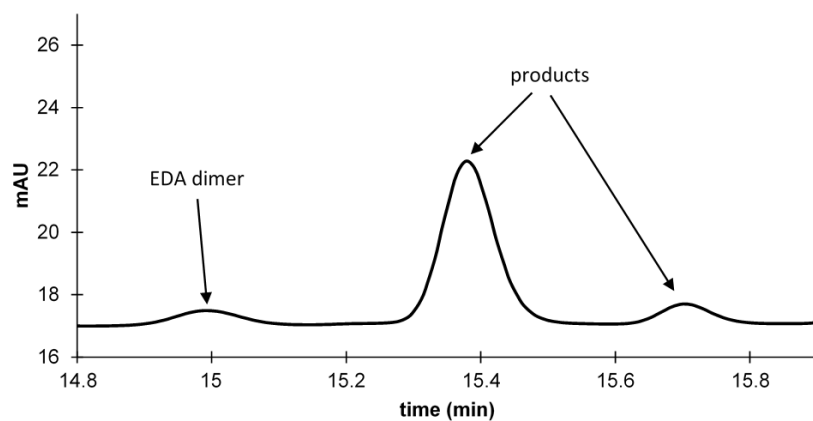
P450<sub>TbJ1</sub>-T234A/C340S - N-Bz-Dha-COOMe



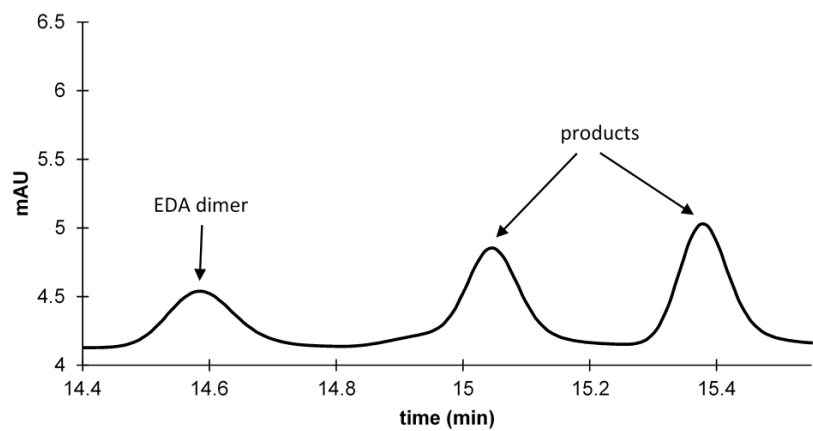
P450<sub>TbJ2</sub> - N-Bz-Dha-COOMe



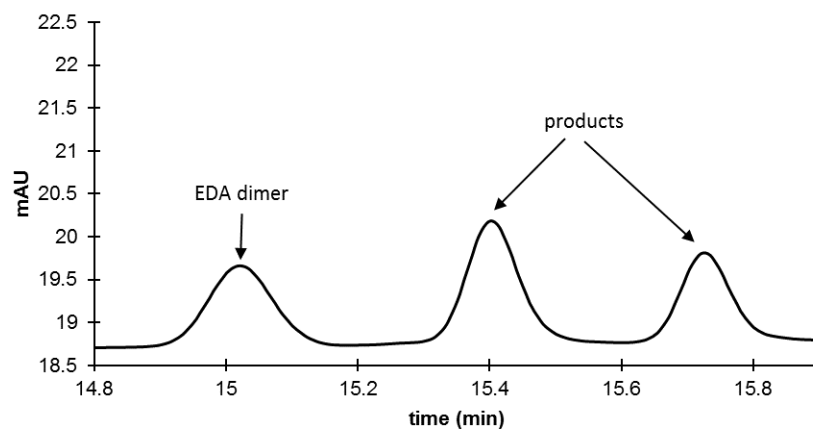
P450<sub>TbJ2</sub>-T247A - N-Bz-Dha-COOMe



P450<sub>TbIJ2</sub>-C340S - N-Bz-Dha-COOMe

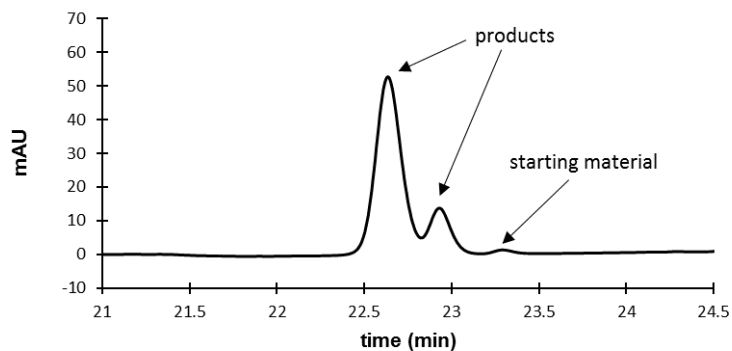


P450<sub>TbIJ2</sub>-T247A/C353S - N-Bz-Dha-COOMe

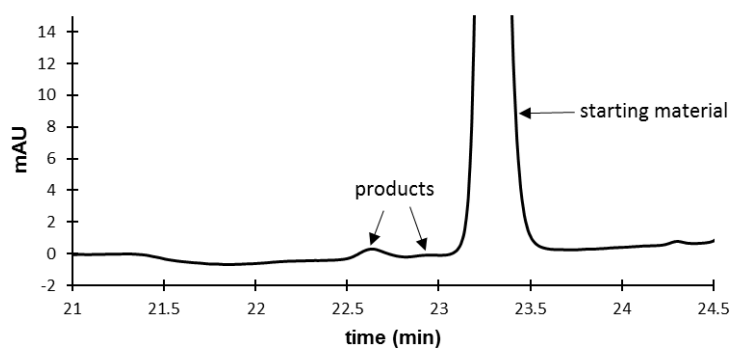


**Figure S3.2.** HPLC traces from enzymatic cyclopropanations of ethyl 2-(1-acetamidovinyl)thiazole-4-carboxylate (**2**) (N-Ac-ThzDha-COOEt).

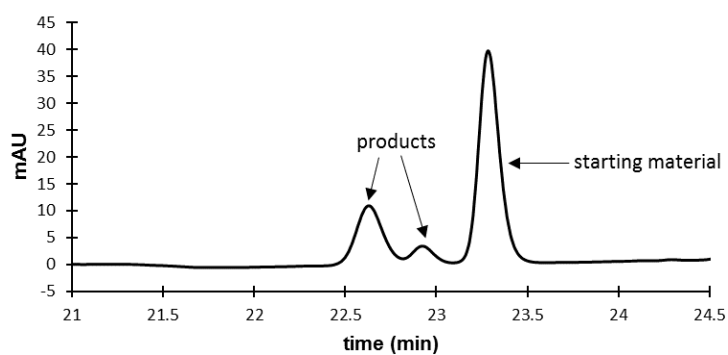
Hemin - N-Ac-ThzDha-COOEt



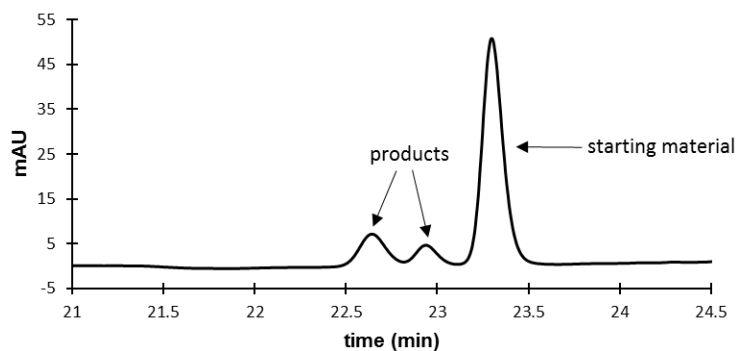
P450<sub>BM3</sub>-T268A - N-Ac-ThzDha-COOEt



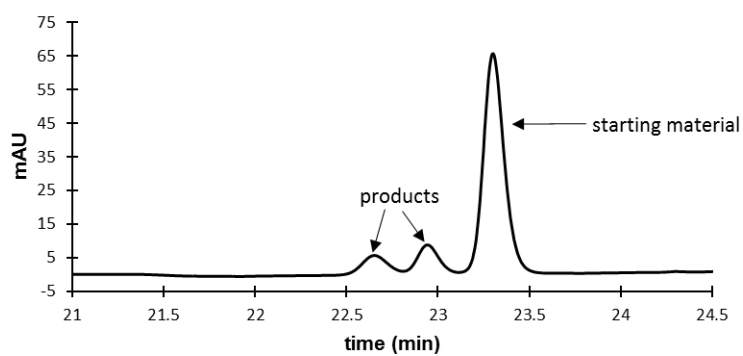
P450<sub>BM3</sub>-T268A/C400S - N-Ac-ThzDha-COOEt



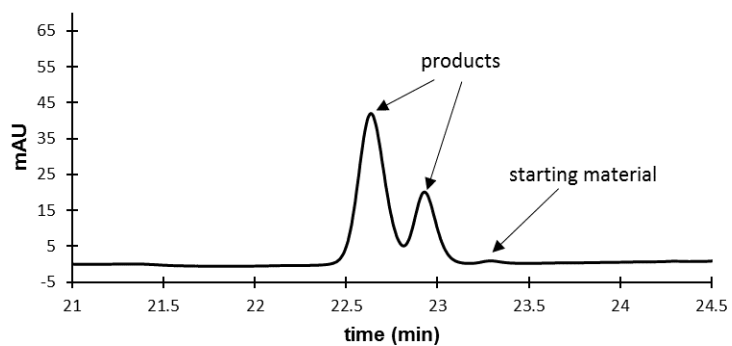
P450<sub>BM3</sub>-CIS-C400S - N-Ac-ThzDha-COOEt



P450<sub>Biol</sub>-T238A - N-Ac-ThzDha-COOEt

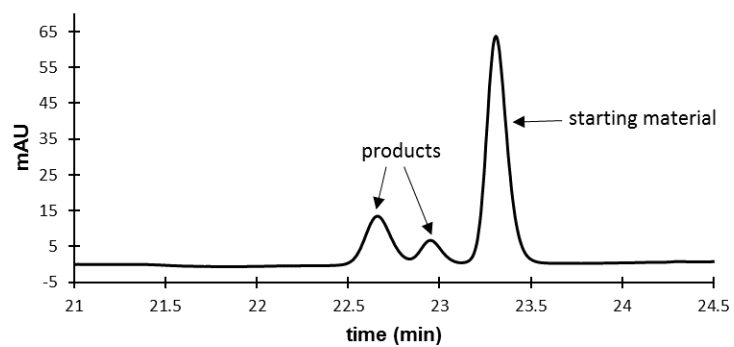


P450<sub>Biol</sub>-T238A - N-Ac-ThzDha-COOEt

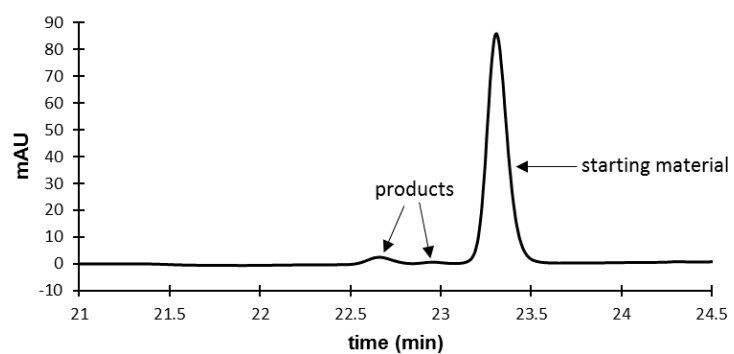




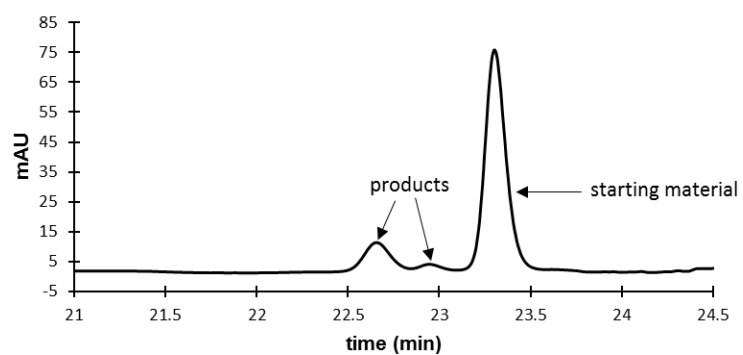
P450<sub>EryF</sub> - N-Ac-ThzDha-COOEt



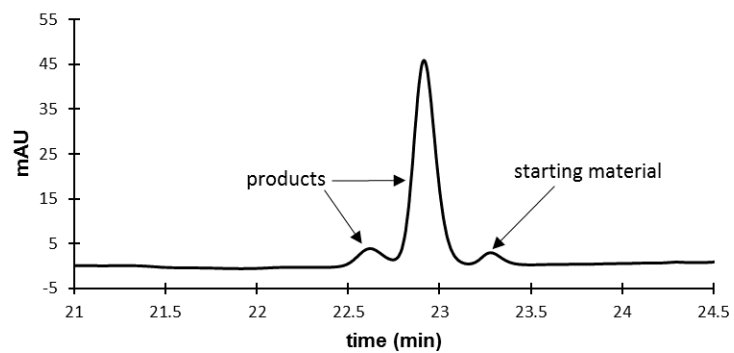
CYP142-T234A - N-Ac-ThzDha-COOEt



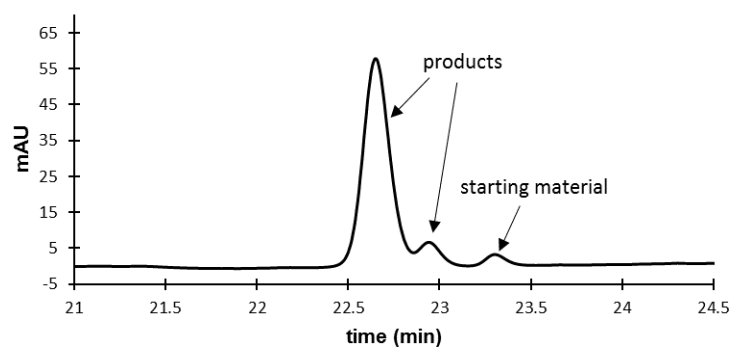
CYP119-T213A - N-Ac-ThzDha-COOEt



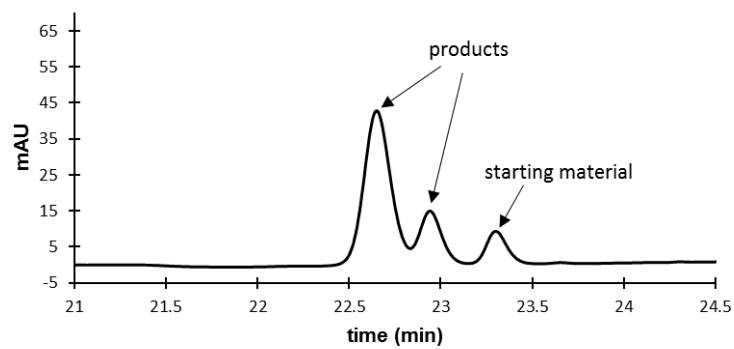
CYP119-T213A/C317S - N-Ac-ThzDha-COOEt



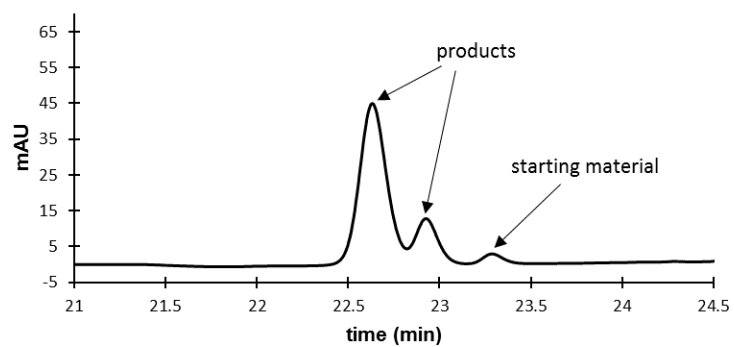
P450<sub>PikC</sub>-T247A - N-Ac-ThzDha-COOEt



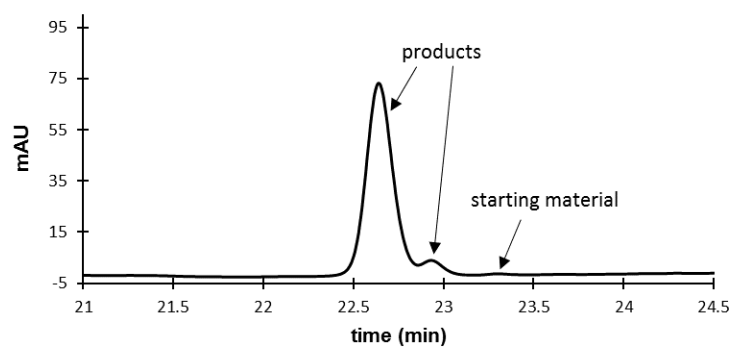
P450<sub>TbtJ1</sub>-T234A - N-Ac-ThzDha-COOEt



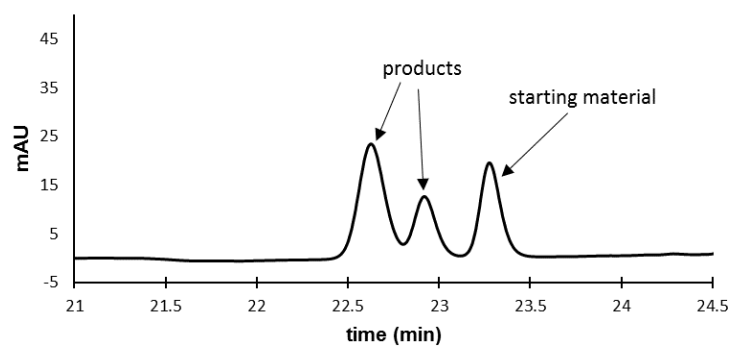
P450<sub>TbIJ1</sub>-T234A/C340S - N-Ac-ThzDha-COOEt



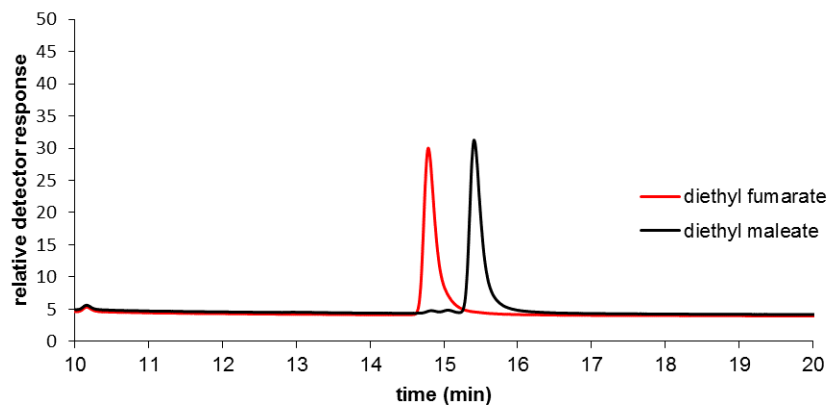
P450<sub>TbIJ2</sub>-T247A - N-Ac-ThzDha-COOEt



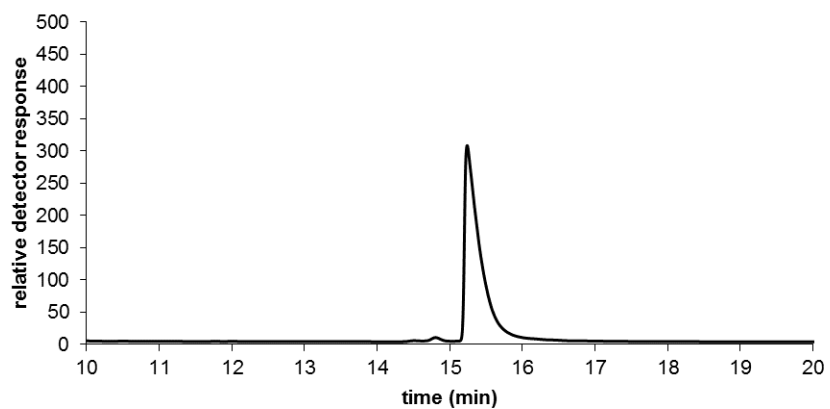
P450<sub>TbIJ2</sub>-T247A/C353S - N-Ac-ThzDha-COOEt



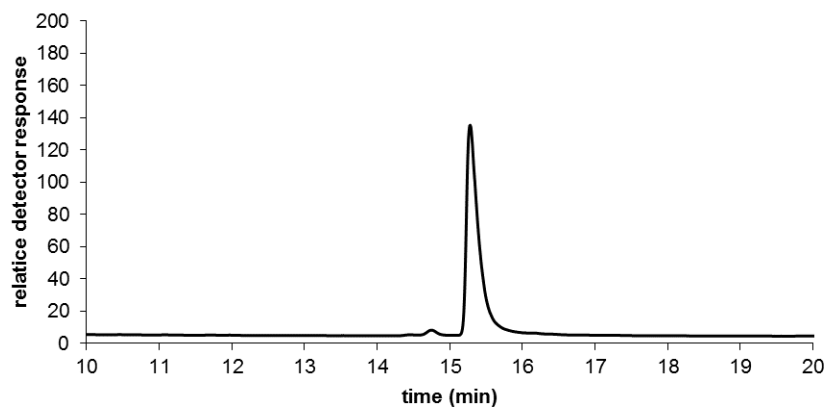
**Figure S3.3.** Chiral GC traces from cyclopropanations of methyl 2-benzamidoacrylate with EDA.



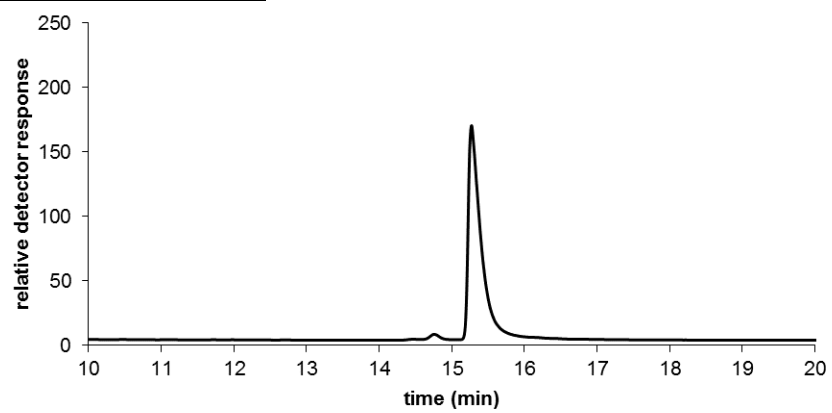
CYP119-T213A/C317S



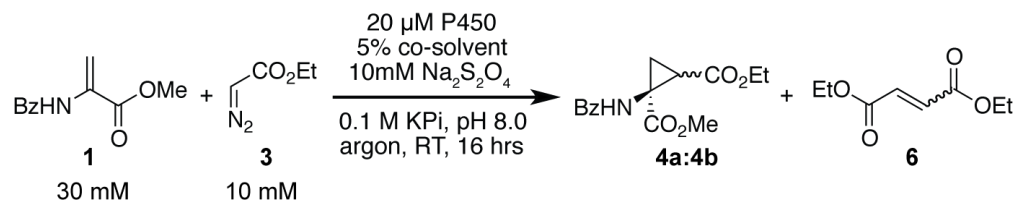
P450<sub>TbJ1</sub>-T234A/C340S



P450<sub>TbIJ2</sub>-T247A/C353S



**Table S3.1.** Conversions and turnovers for EDA dimer formation during cyclopropanation reactions with **1**.



catalyst	conversion ( <b>4</b> )	TTN( <b>4</b> ) <sup>a</sup>	dr ( <b>4</b> )	conversion ( <b>6</b> )	TTN( <b>6</b> ) <sup>a</sup>
hemin	10	50	66:34	15	38
P450 <sub>BM3</sub> -T268A	2	8	63:37	40	101
P450 <sub>BM3</sub> -T268A/C400S	6	29	68:32	31	78
P450 <sub>BM3-CIS-3</sub> -C400S	6	32	69:31	38	96
P450 <sub>Biol</sub> -T238A	4	21	49:51	26	65
P450 <sub>Biol</sub> -T238A/C344S	18	90	55:45	45	113
P450 <sub>EryF</sub>	12	60	76:24	13	32
CYP142-T234A	2	12	74:26	25	62
CYP119-T213A	19	94	85:15	27	67
CYP119-T213A/C317S	6	28	57:43	71	176
P450 <sub>PkC</sub> -T247A	29	145	83:17	14	36
P450 <sub>TblJ1</sub>	2	12	80:20	3	6
P450 <sub>TblJ1</sub> -T234A	10	50	74:26	8	21
P450 <sub>TblJ1</sub> -C340S	12	59	65:35	64	160
P450 <sub>TblJ1</sub> -T234A/C340S	16	81	54:46	63	156
P450 <sub>TblJ2</sub>	4	20	71:29	7	16
P450 <sub>TblJ2</sub> -T247A	20	101	90:10	36	91
P450 <sub>TblJ2</sub> -C353S	6	29	49:51	33	83
P450 <sub>TblJ2</sub> -T247A/C353S	9	44	59:41	75	187

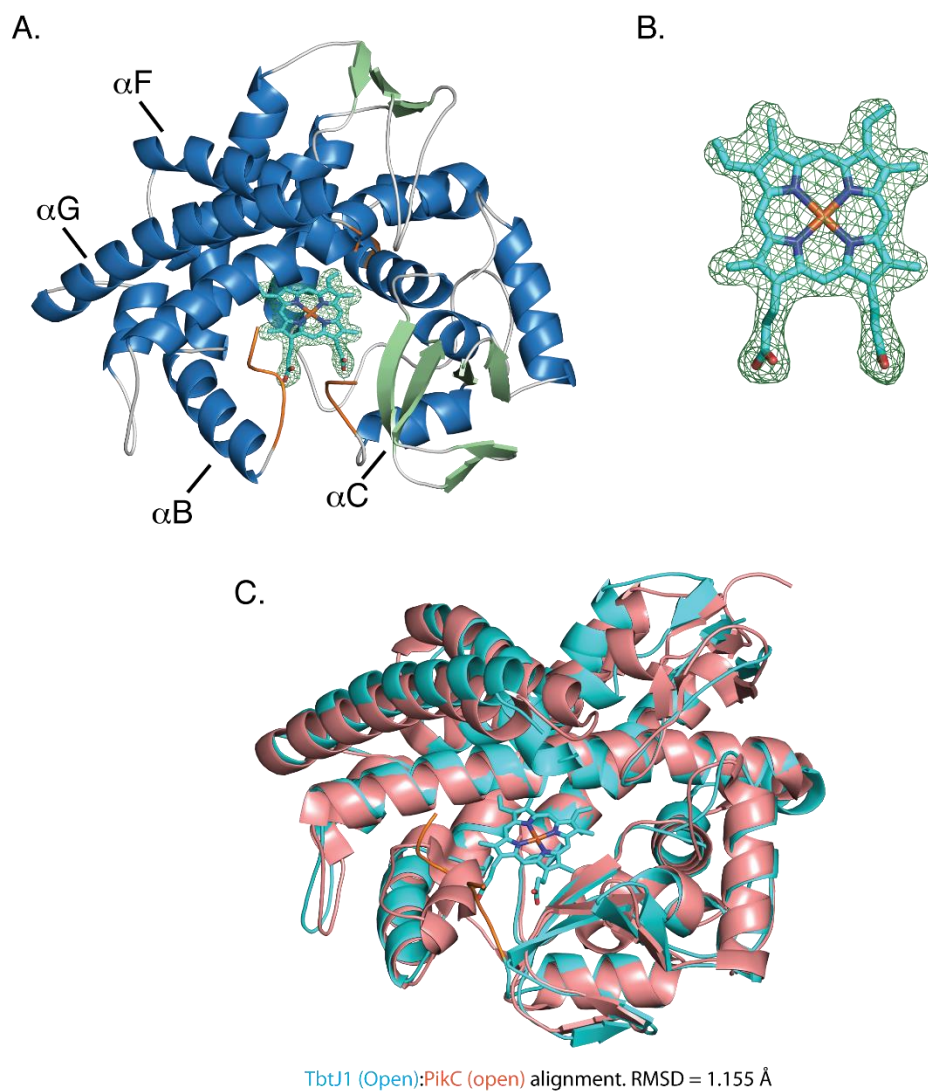
<sup>a</sup>TTN = total turnover number. TTNs and diastereomeric ratios determined by HPLC analysis.

**Table S3.2.** Data collection and refinement statistics for TbtJ1 crystallography.

	TbtJ1
<b>pdb accession #</b>	t.b.d.
<b>Data collection</b>	
Space group	<b>P 41 21 2</b>
Wavelength	<b>1.000</b>
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	<b>95.316, 95.316, 126.782</b>
<i>a</i> , <i>b</i> , <i>g</i> (°)	<b>90</b>
Resolution (Å)	<b>26.44 – 2.41 (2.50 – 2.41) *</b>
<i>R</i> <sub>merge</sub>	<b>10.0 (47.9)</b>
<i>I</i> / <i>σI</i>	<b>6.0 (2.0)</b>
Completeness (%)	<b>91.2 (67.5)</b>
Redundancy	<b>3.7 (2.4)</b>
<b>Refinement</b>	
Resolution (Å)	<b>26.44 – 2.41</b>
No. reflections	<b>21195</b>
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	<b>0.21/ 0.28</b>
No. atoms	
Protein	<b>3003</b>
Ligand/ion	<b>43</b>
Water	<b>96</b>
<i>B</i> -factors	
Protein	<b>23.3</b>
Ligand/ion	<b>12.8</b>
Water	<b>20.3</b>
R.m.s. deviations	
Bond lengths (Å)	<b>0.009</b>
Bond angles (°)	<b>1.03</b>
Ramachandran outliers	<b>1.6 %</b>

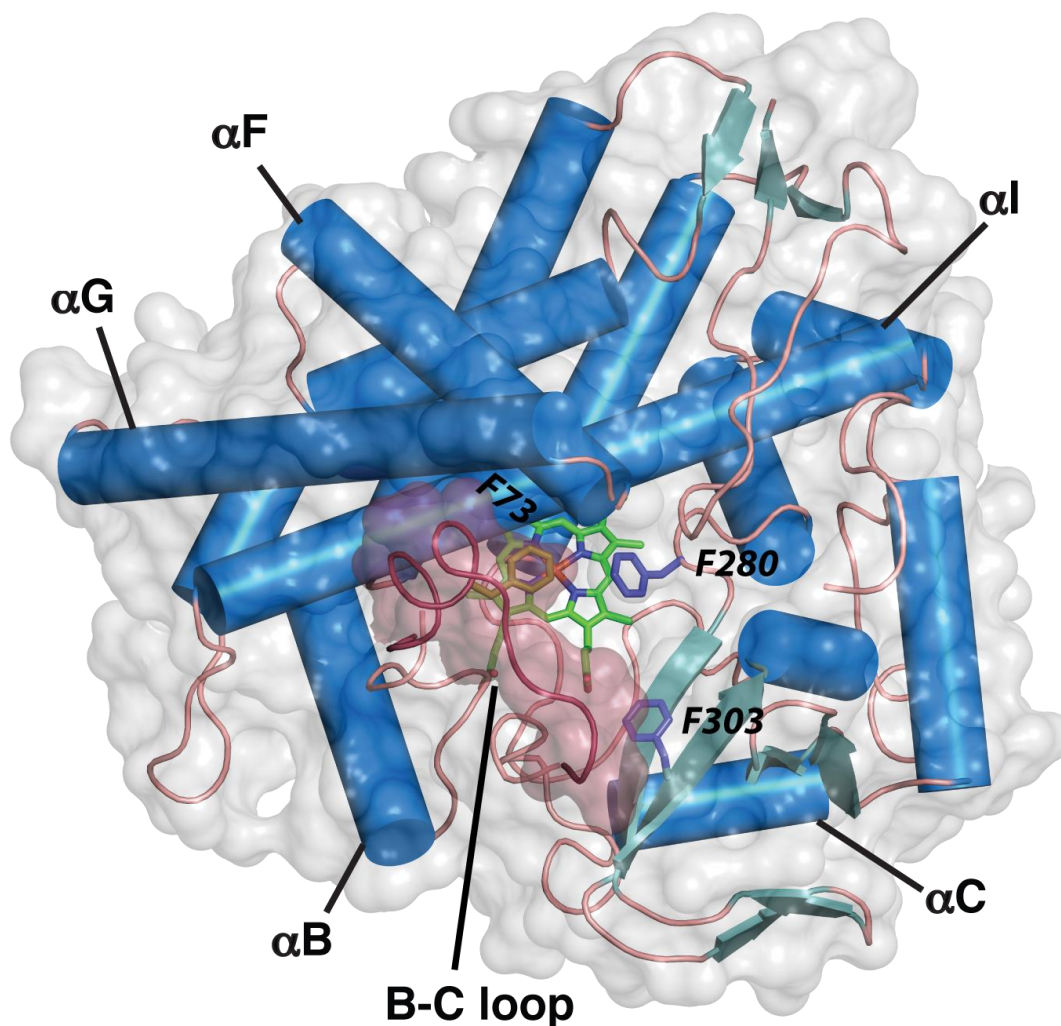
\*All data sets were collected from single crystals. Highest-resolution shell is shown in parentheses.

**Figure S3.4.** TbtJ1 structure determination. (A) The structure of TbtJ1 was determined by X-ray crystallography to 2.4 Å via molecular replacement using P450 EryK (54% identity, PDB: 2WIO) as a model template. The heme is shown as cyan sticks. Two regions of TbtJ1 were unstructured in the final structure including residues 64-70 (MPTAQGF) in the B-C loop and residues 165 and 166 (NY) in the F-G loop. Flanking residues for these missing regions are highlighted in orange. (B) Close up of the modeled heme from TbtJ1 overlaid with the 2Fo – Fc maximum likelihood weighted electron density map. (C) Structural alignments of TbtJ1 with the EryK structure used as a molecule replacement model.



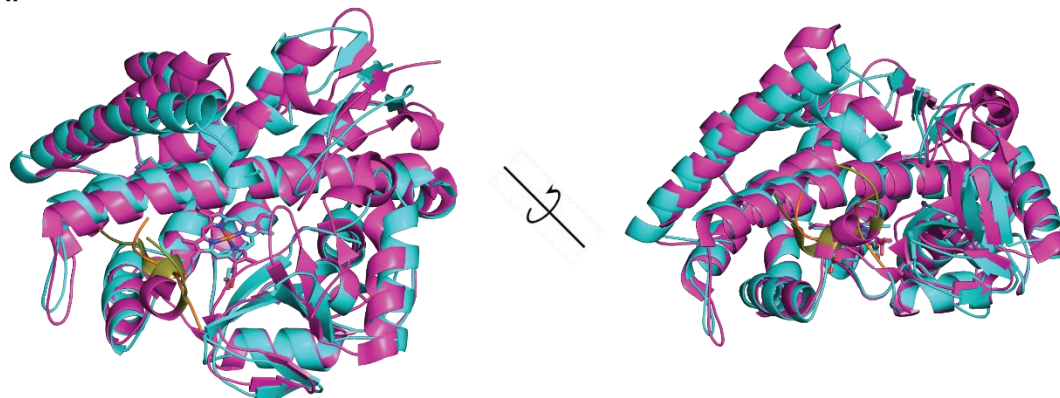


**Figure S3.5.** X-ray crystal structure of ligand-free TbtJ1. The heme is shown as green sticks. The B-C loop that caps the active site is disordered in this structure. The B-C loop from ligand-bound PikC has been superimposed in red, to provide a rough estimate of this capping interaction. Three Phe residues flank the active site. F280 and F303 (purple) are well resolved in the crystal structure. The side chain of F73 (olive) is poorly structure in the crystal structure, but modeled here (using coot) to show the approximate orientation with respect to the active site.



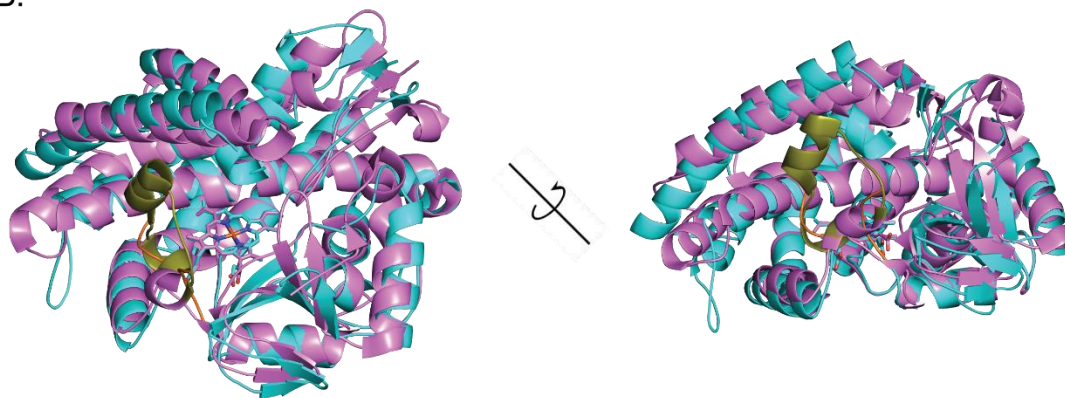
**Figure S3.6.** TbtJ1 alignment with the open and closed forms of PikC. (A) TbtJ1 alignment with the ligand-free, open form of PikC (PDB: 2BVJ). Residues flanking the missing, disordered regions of TbtJ1 are highlighted in orange. Similar disordered regions (highlighted in olive) are observed in the ligand-free form of PikC. (B) TbtJ1 alignment with the ligand-bound, closed form of PikC (PDB:2C7X). The disordered region of PikC is now well structured (olive) and provides a rough estimate of where missing residues in TbtJ1 may be located when bound to thiopeptide substrates.

A.



TbtJ1 (Open):PikC (open) alignment. RMSD = 1.155 Å

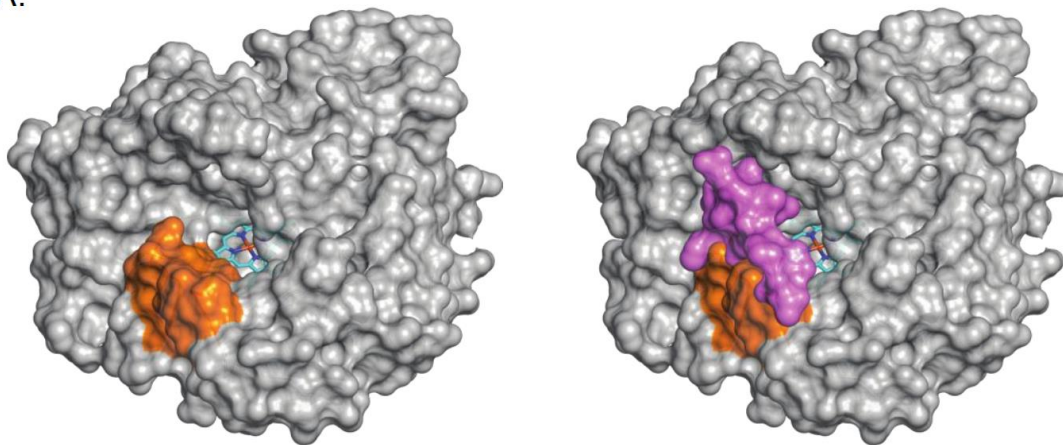
B.



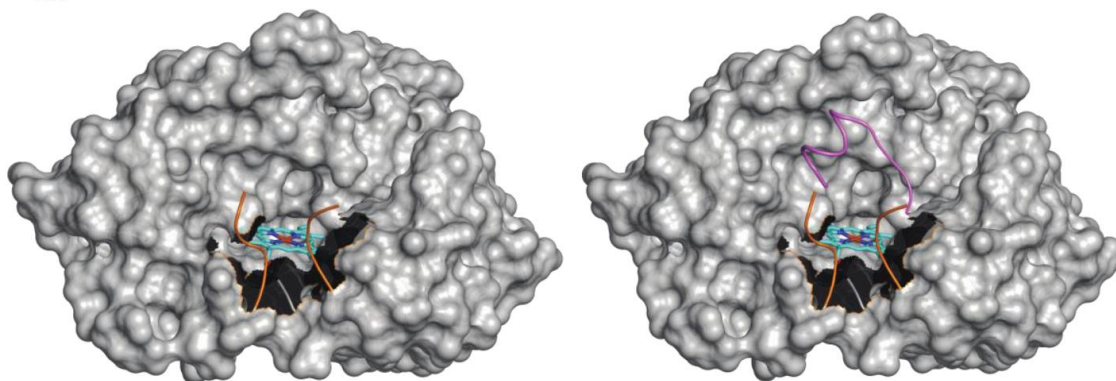
TbtJ1 (Open):PikC (closed) alignment. RMSD = 2.424 Å

**Figure S3.7.** Loop modeling of TbtJ1 based on the closed form of PikC. Structural alignment with the closed form of PikC provides a crude estimate of the location of unstructured residues in TbtJ1. These regions are shown as surface representations in panel A and, for clarity, as cartoons in panel B. Residues flanking this unstructured region in TbtJ1 are highlighted in orange. The ordered loop from the ligand-bound form of PikC (PDBL2C7X) is highlighted in pink and shows that missing residues in TbtJ1 likely form a cap over the deep heme-bound active site.

A.

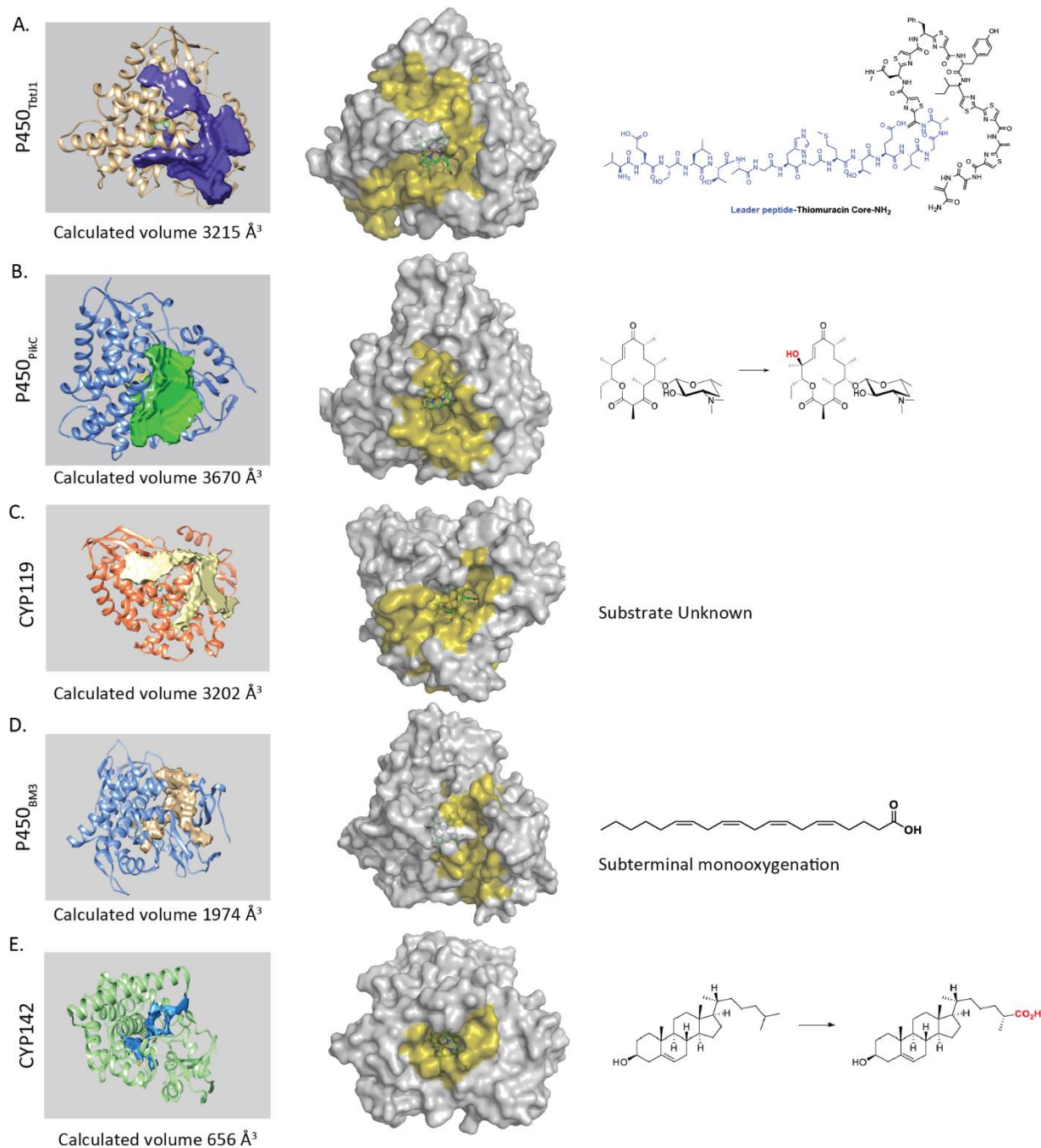


B.





**Figure S3.8.** Comparison of P450 binding pockets for highly active cyclopropanation variants (A) P450<sub>TbtJ1</sub> (PDB: t.b.d.), (B) P450<sub>PikC</sub> (PDB: 2BVJ), and (C) CYP119 (PDB: 1IO7) as well as poorly active variants (D) P450<sub>BM3</sub> (PDB: 1YQO) and (E) CYP142 (PDB: 2XKR). Cavity volumes calculated using the 3V, volume calculator (<http://3vee.molmovdb.org>),<sup>14</sup> are shown on the left. Surface representations of each crystal structure are shown in the middle. Regions identified in 3V are highlighted in yellow in the corresponding surface structures. Highly active variants show considerably larger cavities providing access to the heme cofactor. Structures of representative substrates for each P450 are shown to the right.



**Figure S3.9.** ClustalOmega alignment of P450 protein sequences. The conserved active-site threonine and ligating cysteine are highlighted in yellow, and amino acid substitutions that differ from the highly conserved threonine are highlighted in red. Active site phenylalanine residues are highlighted in green.

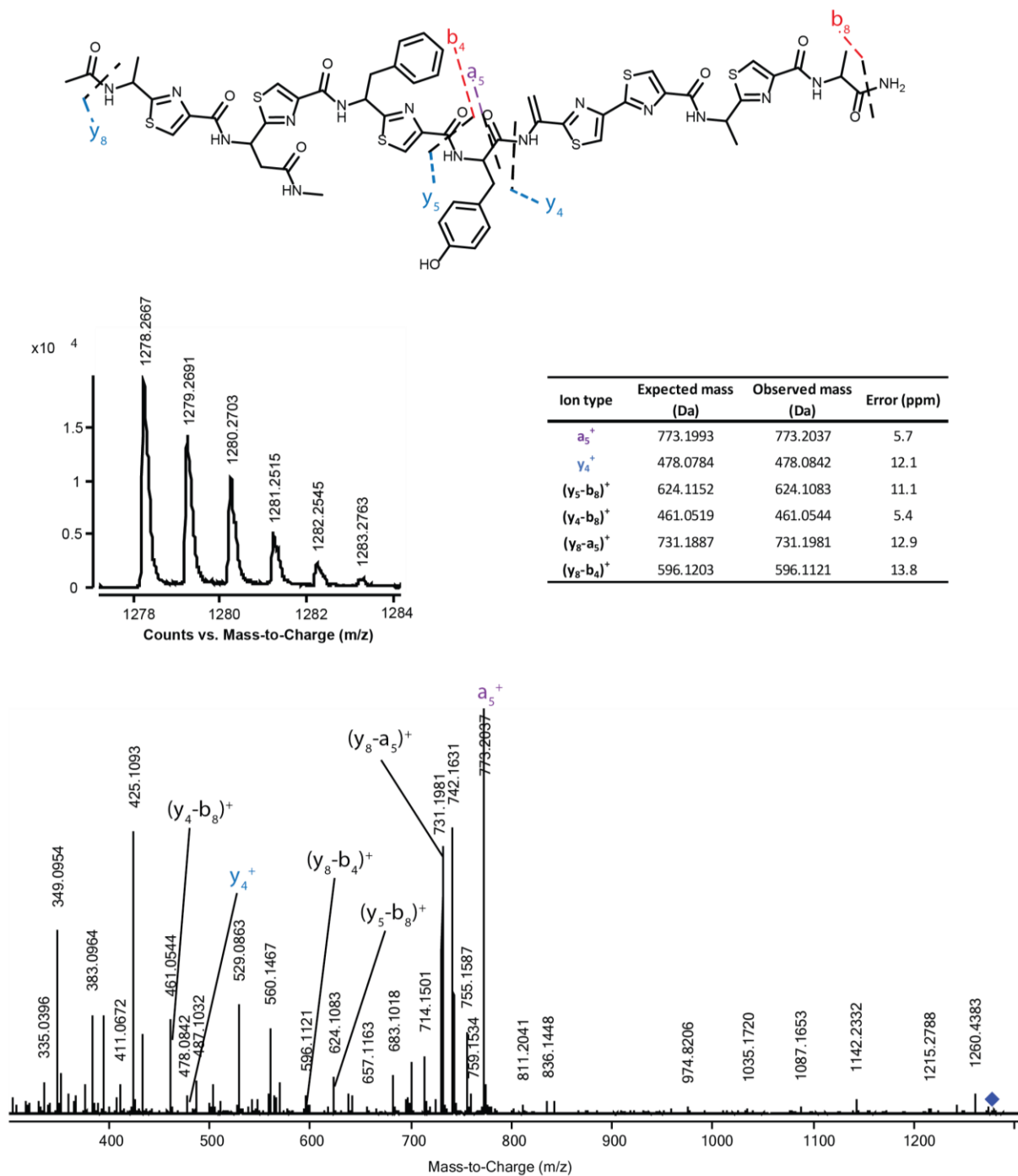
BM3	MTIKEMPQPKT-FGELKNLPLLN---TD---KPVQALMKIADL---GEIFKFEA----	45
CYP142	-----MTEAPDVLADGNFYASREARA---AYRWMRANQPVFRDRN----	38
CYP119	-----MYDWFSEMRKKDPVYYDG-----	18
TbtJ1	-----MMELP---GQPSLTDGGAALFAWLRTMRNEHPVWRDQF----	35
TbtJ2	-----MNEPLTAQGTPTMTMPAA---QRPRISHGGQALLKWLDEMRESQPVWRDGF----	48
BioI	-----MTIA---SSTASSEFLKNPYSFYDTLRAVHPITYKGSF-LKY	37
PikC	-----MRRTQQGTTASPPVLD---LGALGQDFAADPYPTYARLRAEGPAHRVRTPEGD	50
EryF	-----MTTVP-----DLESDFHVDWYRTYAELERETAPVTPVRF-LGQ	37
:		
BM3	-PGRVTRYLSSQRLIKEACDESFRDKNLS-----QALKFVRDFAG---DGLFTS	90
CYP142	GLAAASTYQAVIDAERQP---ELFSN-AGGIR-----PDQPALPMMIDM	78
CYP119	NIWQVFSYRYTKEVLNNF---SKFSSDLTGYHERLEDLRNGKIR----FDIPTRYTMLTS	71
TbtJ1	GIYHVFRYDDVRQILGDY---QTFSDDRTRLMP-----AQGGFGKGGITMI	78
TbtJ2	GIFHVFRHADVQRVMADY---ATFSSDINRLRPG-----GDPFGSAGSLMLT	91
BioI	PGWYVTGYEETAAILKDA---R-FKVRTPLPE-----SSTK---YQDLSHVQNQMMLFQ	84
PikC	EVVLVVGVDYDRARAVLADP---R-FSKDWRNSTTPL-----TEAEAALNHNLMLES	95
EryF	DAWLVTGYDEAKAALS DL---R-LSSDPKKKYPGVEVEFPAYLGFPEDVRNYFATNMGTS	93
. : :		
BM3	WTHEKNWKKAHNILLPSFSQAMKGYHAMMV DIAVLVQKWERLNADEHIEVPEDMTRLT	150
CYP142	--DDPAHLLRRKLVNAGFTRKRVKDKEASIAALCDTLIDAV--CE-----	119
CYP119	--DPPLHDELRSMSADIFSPQKLQTLTFIRETTRSLSDSI---D-----	111
TbtJ1	--DPPEHRHQRLITHAFTPQSI SAMEPRIRQIADHLLDEL---P-----	118
TbtJ2	--DPPEHRKLRLISQAFSPKMSADMKPRIAE LTEELLDDI---E-----	131
BioI	--NQPDHRRRLRTLASGAFTPRTTESYQPYIIETVHHLLDQV--QG-----	125
PikC	--DPPRHTRLRKLVA REFTMRRLVLLRPRVQEIVDGLVDAMLAAP-----	138
EryF	--DPPTHTRLRKLVSQEFTVRRVEAMRPRVEQITAE LLDDEV--GD-----	134
. : * : . : * :		
BM3	LDTIGLCGFNYRFNSFYRDQPHPFITSMVRALDEAMNK-----LQRANPDDPAYDENK	203
CYP142	---RGECDFV---RDLAAPLPMAVIGDMLGVRPEQRDMFLRWSDDLVTFLSSHVSQ-EDF	172
CYP119	---PREDDIV---KKLAVPLPIIVISKILGLPIEDKEKFKEWSDLVAFRLGK----PGEI	161
TbtJ1	---GPEFDLV---EHFAYPLPVIVIAELLGVPPGDRHLFRTWSDRMLSQVENYADPELA	172
TbtJ2	---EDEFDLV---EKFAHPLPIMVIAELLGIPIHDRGLFRTWADRLIALHVDDPTDVEIG	185
BioI	---KKKMEVI---SDFAFPLASFVIANIIGVPEEDREQLKEWAASLIQTI DFTRSR-KAL	178
PikC	---DGRADLM---ESLAWPLPITVISELLGVPEPDRAAFRVWTDADFVFPDD--P-----A	185
EryF	---SGVVDIV---DRFAHPLPIKVICELLGVDEKYRGEFGRWSSEILVMDPERA-----E	183
. : . * : :		
BM3	RQFQEDIKVMNDLVDKIIADRKASGEQSDDLLTHMLNGKDPETGEPLDDENIRYQIITFL	263
CYP142	QITMDAFAAYNDFTRATIAARR---ADPTDDLVSVLVSS-EV-DGERLSDDDELVMETLLIL	228
CYP119	FELGKKYLELIGYVKDHLNS-----GTEVVS RVV-----NSNLSDIEKLG YI ILLL	207
TbtJ1	RTVAAAMTEMNDYLREHCRSR--THPRDDLTRLVQA-EV-EGKRLDLEEVNTASLLL	228
TbtJ2	RMVGEAMREMGEYVQTHVRKR--ADPQDDLVS KLIAA-EV-DGERLTD AEIVNTSCLLL	241
BioI	TEGNIMAVQAMAYFKELIQKR--RHPQQDMI SMLLKGREK---DKLTEEEEA STCILLA	233
PikC	Q-AQTAMAEMSGYLSRLIDSKR--GQDGEDLLSALVRTSDE-DGSRLTSEELLGMAHILL	241
EryF	Q-RGQAAREVVNFILDLVERR--TEPGD LLSALIRVQDD-DDGRLSADELTSIALVLL	239
::: : * :		
BM3	IAGHETTSGLLSFALYFLVKNPHVLQKAAEEAARVLVDPVPSYKQVKQLKYVGMVLNEAL	323
CYP142	IGGDETTRHTLSGGTEQLLRNRDQWDL LQRDP-----SLLPGAIEEML	271
CYP119	IAGNETTTNLI NSVIDFTRFN LW--QRIREE-----NLYLKAIEEAL	248
TbtJ1	LAGHLTTTVLIGNTMLCLWDHPEAEKAVRADP-----SLIPAALEESL	271
TbtJ2	LAGQITSTMALGNTFLCFRDAPDAERAVRADF-----SLLGPAFEEVL	284
BioI	IAGHETTVNLI NSVLC LLQHPEQLLKLREN-----DLIGTAVEECL	276

PikC	VAGHETTVNLIANGMYALLSHPDQLAALRADM-----TLLDGAVEEML	284
EryF	LAGFEASVSLIGIGTYLLLTHPDLQALVRDP-----SALPNAVEEIL	282
	:.* :: : :	:::* *
BM3	RLWPTAPA-FSLYAKEDTVLGGEYPLEKGDLMVLIPQLHRDKTIWGDDVEEFRPERFEN	382
CYP142	RWTAPVK-NMCRVLTADEFHGHT-ALCAGEKMMLLFESANFDEAVFCE-----PEKFDF	323
CYP119	RYSPVMR-TVRRKTKEVKLGDQ-TIEEGEYRVVWIASANRDDEEVFHD-----GKGFIP	300
TbtJ1	RLRSPLQ-AGRVTTRDVTIAGE-TIPANRFVMAWILSANHDDRFPD-----PERFDL	323
TbtJ2	RLRPPLIQ-AARLTITTDVEVAGT-LIPAGSLVINWLLSANYDERQFPN-----PYRFDL	336
BioI	RYESPTQM-TARVASEDIIDCGV-TIRQGEQVYLLGLGAANDPSIFTN-----PDVFDI	328
PikC	RYEGPVESATYRFFVPVVDLDGT-VIPAGDTVLVVLAHAHTPERFPD-----PHRFDI	337
EryF	RYIAPPET-TTRFAAAEEVEIGGV-AIPQYSTVLVANGAANDPKQFPD-----PHRFDV	334
	* . . :	:
BM3	PSAIPQHAFKPFGNGQRACIGQQFALHEATLVLGMMCLKHFDFEDHTNYELDIKE-TLTLK	441
CYP142	QRN--PNSHLAFGGFTHFCLGNQLARLELSLMTERVLRRLPDLR----LVADDSVLPLR	376
CYP119	DRN--PNPHLSFGSGIHLC LGAPLARLEARIAIEEF SKRFRHIE----ILDTE----KV	349
TbtJ1	HRQ--TTGHIAFGHGVHFC LGAQLGRLEGRIALERLLGRFTEIH----PWPRE-GISFY	375
TbtJ2	TRS--PNRHFAFGHGIHHCL GAPLARVEGRALELLRRSEIT----IDPDA-ELSYY	388
BioI	TRS--PNPHLSFGFHGHVCL GSSSLARLEAQIAINTLLQRMPNLN----LA---DFEWR	377
PikC	RRD--TAGHLAFGHHGHIFC I GAPLARLEARIAVRALLERCPCDLA----LDVSPGELVWY	390
EryF	TRD--TRGHL SFGQGIIHFC MGRPLAKLEGEVALRALFGRFPALS-----LGIDADDVVWR	387
	. ** * : *. * : . :	.
BM3	PEGFVVKAKSKKIPLGGIPSPST-----	464
CYP142	PA--NFVSGL SMPVVFTP SPPLG----	398
CYP119	PN--EVLNGYKRLVVRLKSNE-----	368
TbtJ1	QSAIF---GASRM PVRCG-----	390
TbtJ2	EDPMF---GVKSLPVR VRRRA-----	406
BioI	YRPLFGFR AEELPVTTFE-----	395
PikC	PNP--MIRGLKALPI RWRRGREAGRRTG	416
EryF	RSL---LLRGIDHLP VR LDG-----	404

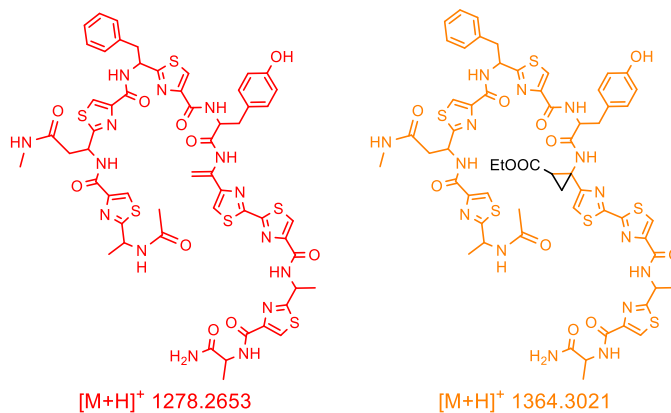
1: BM3	100.00	19.35	18.66	19.35	19.37	18.21	19.47	16.98
2: CYP142	19.35	100.00	28.77	32.02	33.07	27.25	31.87	31.66
3: CYP119	18.66	28.77	100.00	33.24	34.65	30.77	34.38	29.72
4: TbtJ1	19.35	32.02	33.24	100.00	53.33	30.79	37.47	34.75
5: TbtJ2	19.37	33.07	34.65	53.33	100.00	30.26	34.61	35.70
6: BioI	18.21	27.25	30.77	30.79	30.26	100.00	35.43	33.85
7: PikC	19.47	31.87	34.38	37.47	34.61	35.43	100.00	46.04
8: EryF	16.98	31.66	29.72	34.75	35.70	33.85	46.04	100.00

**Figure S3.11.** LC-MS/MS of linear TbtA I8-1Dha core.

LC-MS Method A; CID = 55 eV. Expected  $[M+H]^+$ : 1278.2653; Observed  $[M+H]^+$ : 1278.2667.

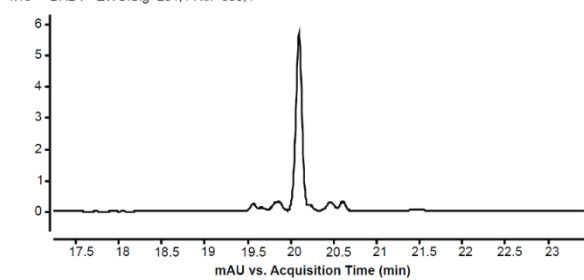


**Figure S3.12.** HPLC and extracted ion chromatogram (EIC) traces of enzymatic cyclopropanations of linear TbtA I8-I Dha core (analyzed by LC-MS Method A).

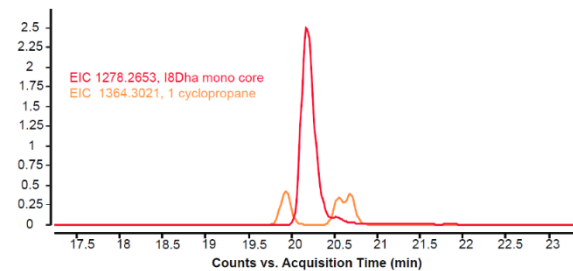


### Hemin

x10<sup>2</sup> DAD1 - EWC:Sig=254,4 Ref=600,4

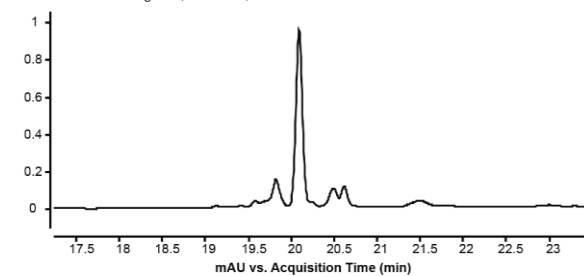


x10<sup>6</sup>

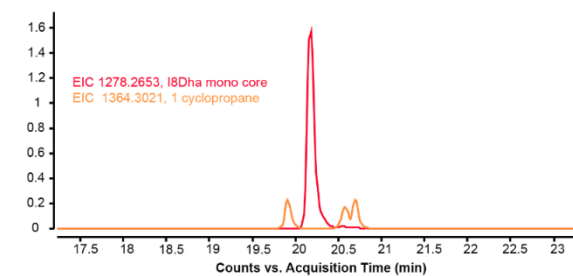


### P450<sub>TbtJ1</sub>-T234A/C340S

x10<sup>2</sup> DAD1 - EWC:Sig=254,4 Ref=600,4

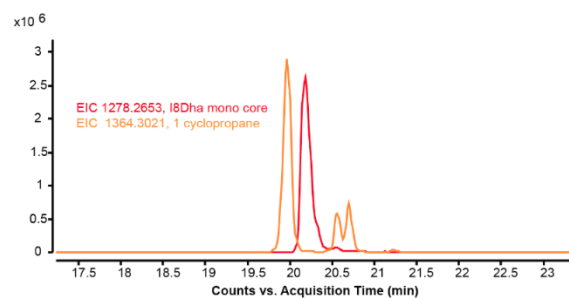
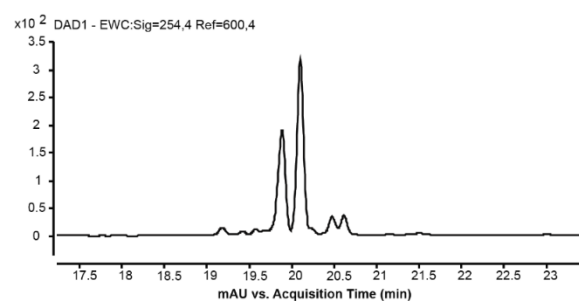


x10<sup>6</sup>

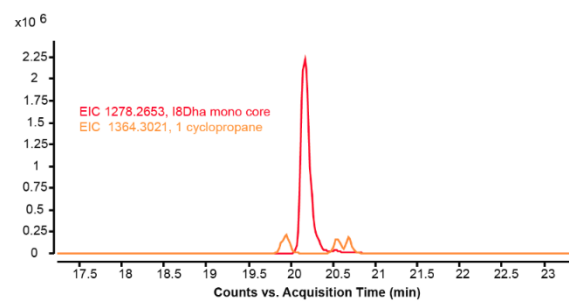
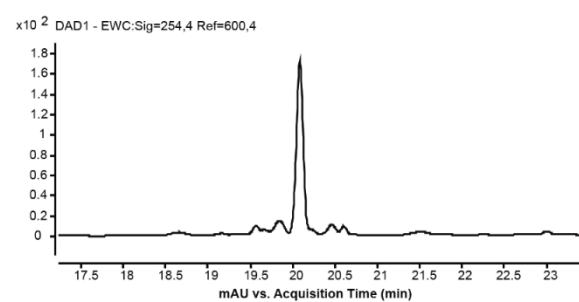




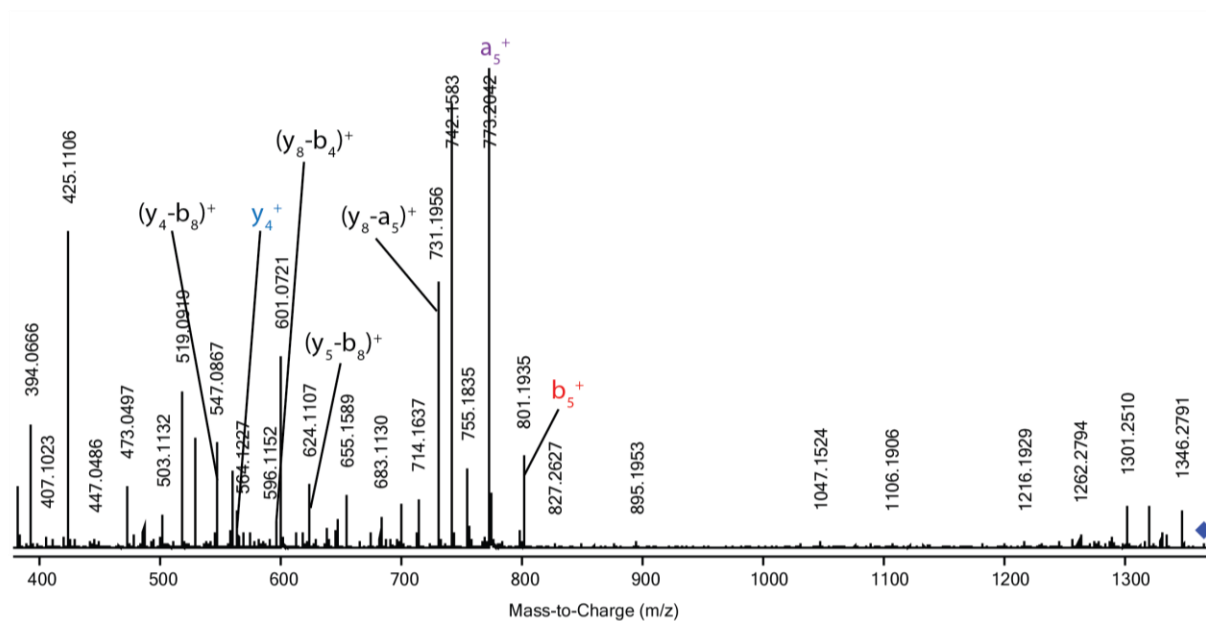
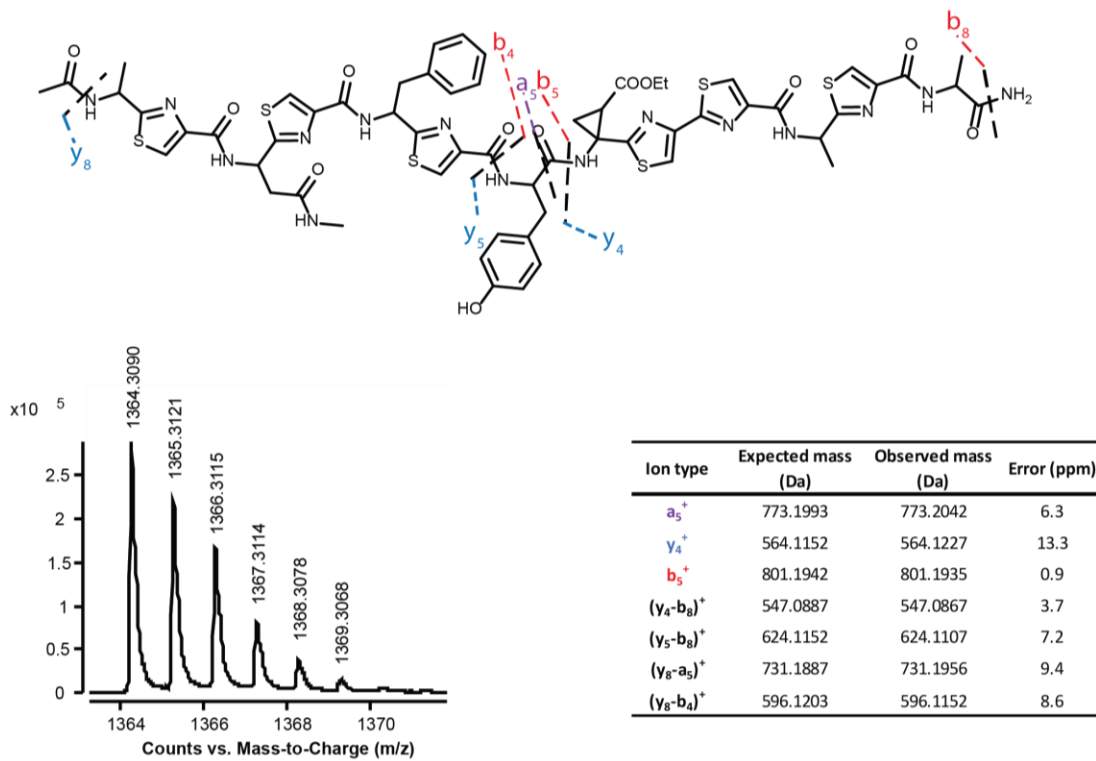
### P450<sub>TbtJ2</sub>-T247A



### P450<sub>PikC</sub>-T247A

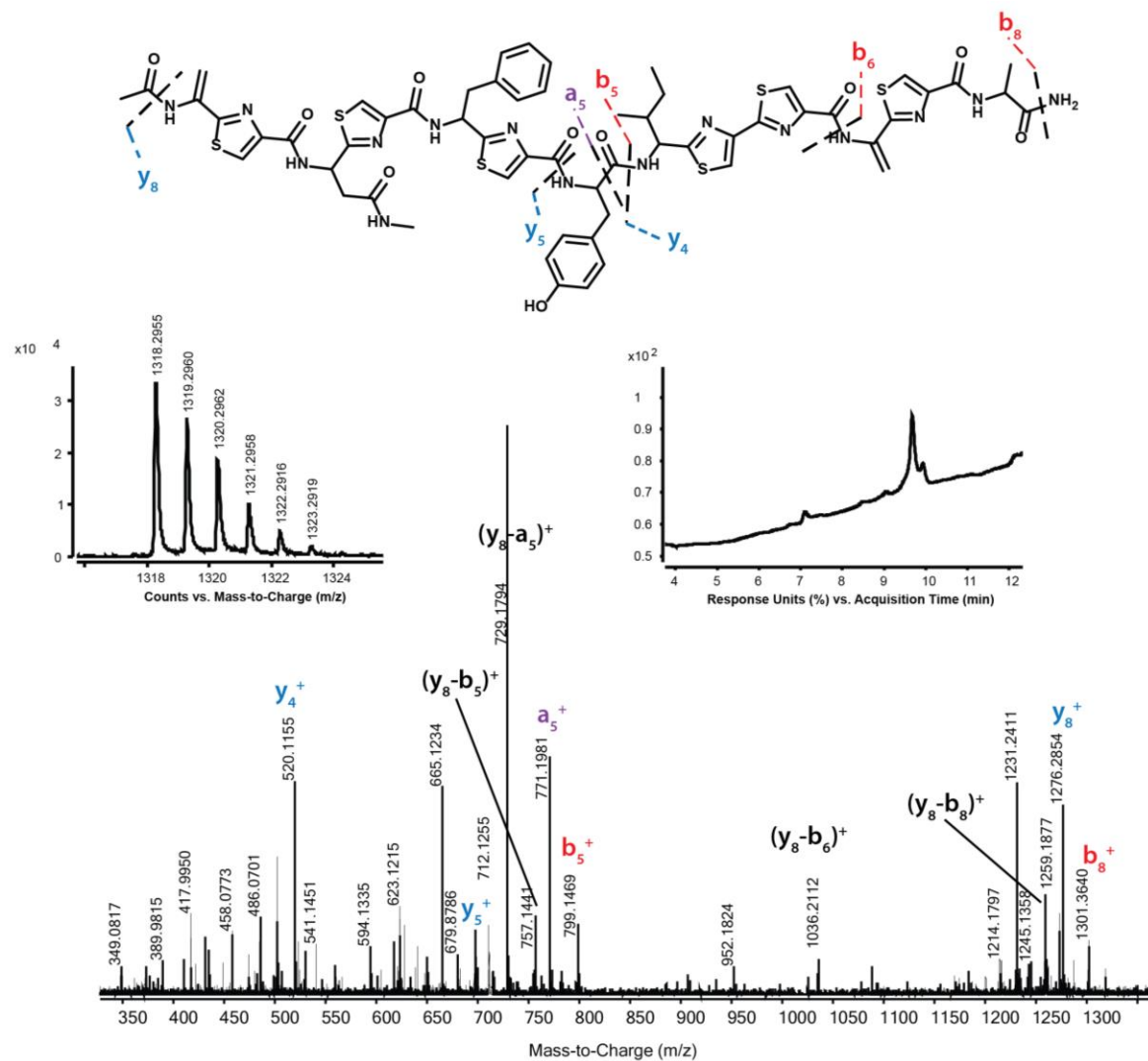


**Figure S3.13.** LC-MS/MS of cyclopropanated linear TbtA I8-1Dha core.  
LC-MS Method A; CID = 55 eV. Expected  $[M+H]^+$ : 1364.3021; Observed  $[M+H]^+$ : 1364.3090.



**Figure S3.14.** LC-MS/MS of linear TbtA core.

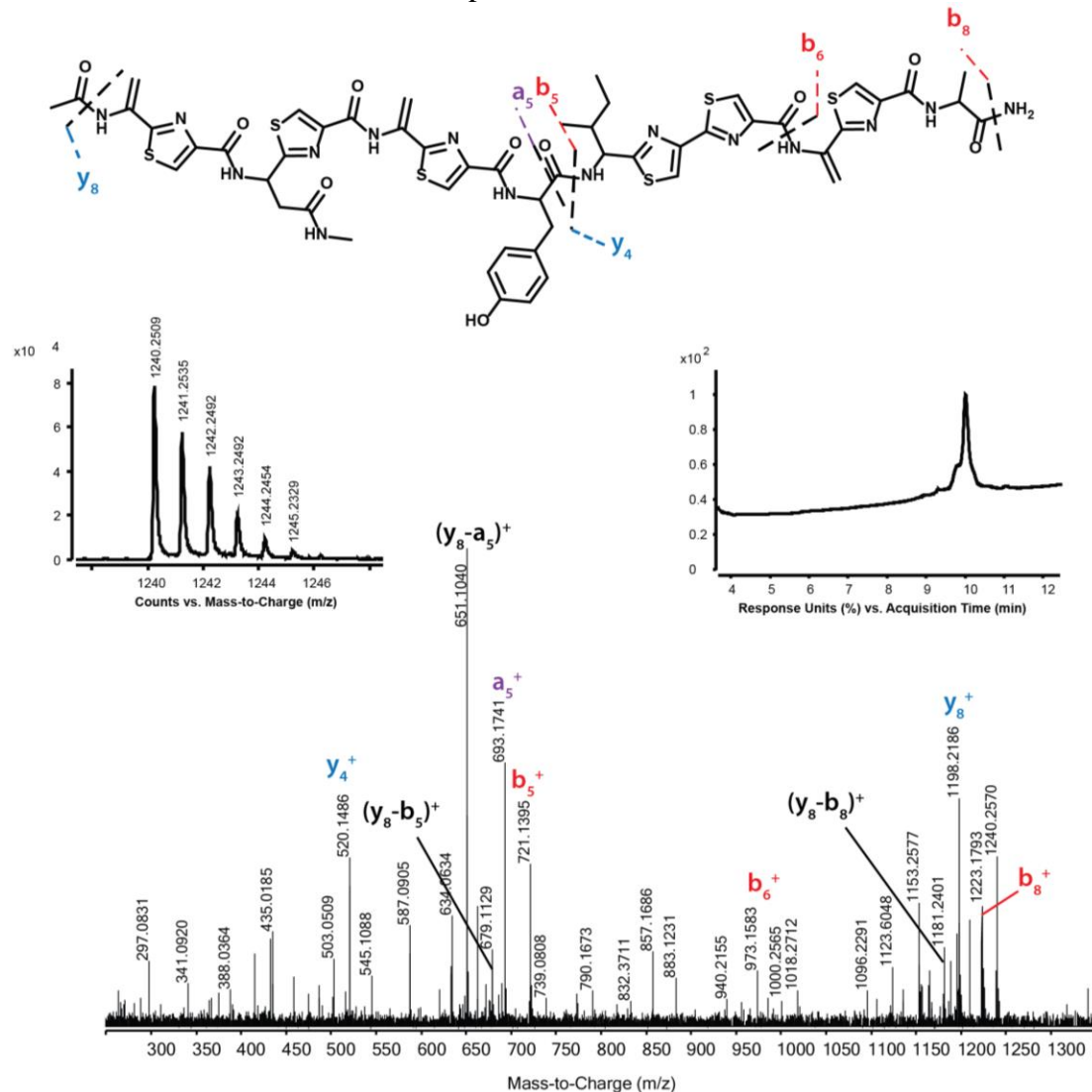
LC-MS Method C; CID = 45.0 eV. Expected  $[M+H]^+$ : 1318.2966, Observed  $[M+H]^+$ : 1318.2955.



Ion type	Expected mass (Da)	Observed mass (Da)	Error (ppm)
$a_5^+$	771.1836	771.1981	18.8
$b_5^+$	799.1785	799.1469	39.5
$b_8^+$	1301.2701	1301.364	72.2
$y_4^+$	520.1254	520.1155	19.0
$y_5^+$	683.1887	683.1771	17.0
$y_8^+$	1276.2861	1276.2854	0.5
$(y_8-a_5)^+$	729.1731	729.1794	8.6
$(y_8-b_5)^+$	757.168	757.1441	31.6
$(y_8-b_6)^+$	1036.218	1036.2112	6.6
$(y_8-b_8)^+$	1259.2595	1259.1877	57.0

**Figure S3.15.** LC-MS/MS of linear TbtA F5-3Dha core.

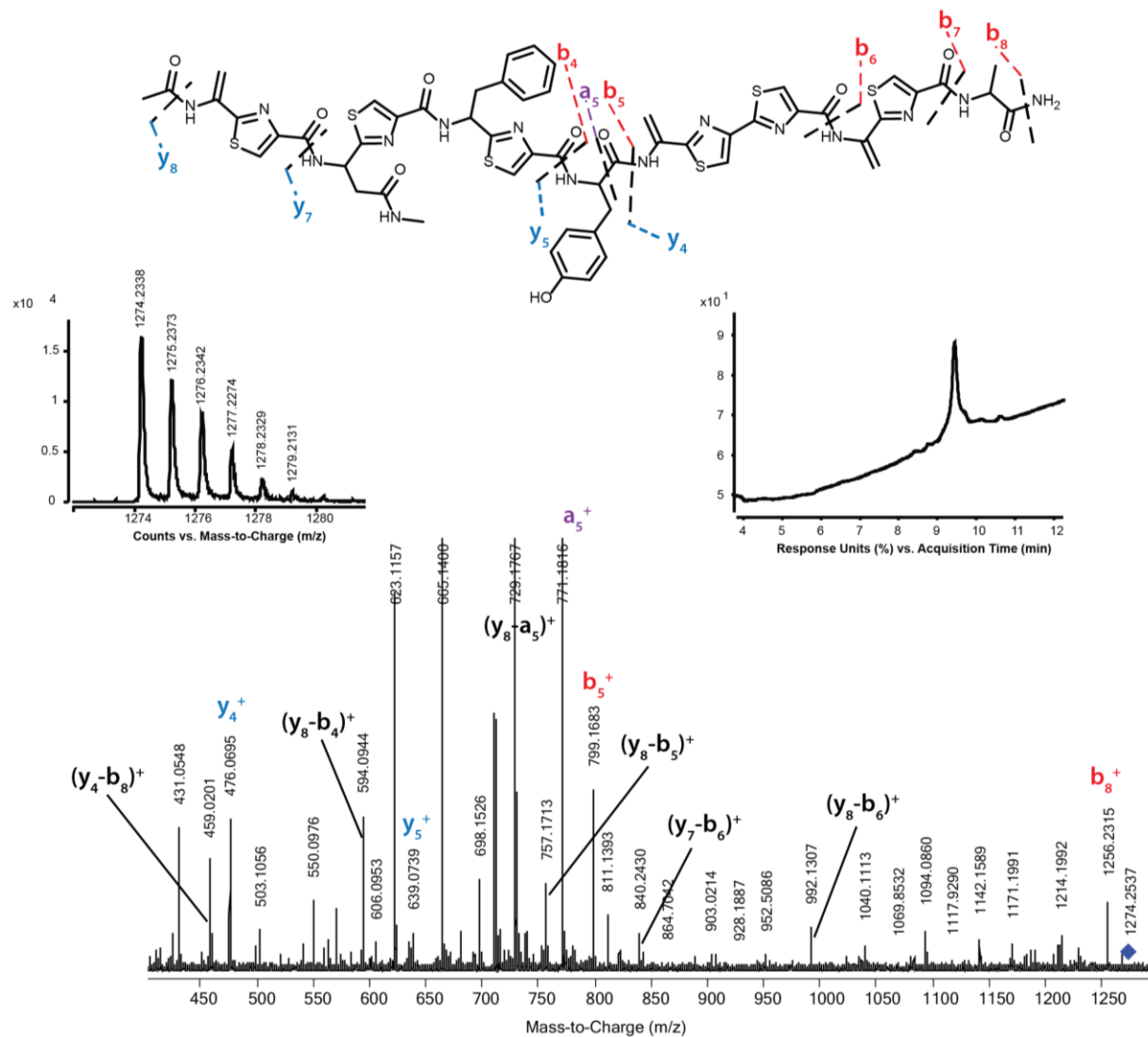
LC-MS Method C; CID = 45 eV. Expected  $[M+H]^+$ : 1240.2497; Observed  $[M+H]^+$ : 1240.2509.



Ion type	Expected mass (Da)	Observed mass (Da)	Error (ppm)
$a_5^+$	693.1367	693.1741	54.0
$b_5^+$	721.1316	721.1395	11.0
$b_6^+$	1000.1816	1000.2565	74.9
$b_8^+$	1223.2231	1223.1793	35.8
$y_4^+$	520.1254	520.1486	44.6
$y_8^+$	1198.2391	1198.2186	17.1
$(y_8 - a_5)^+$	651.1261	651.1040	33.9
$(y_8 - b_5)^+$	679.121	679.1129	11.9
$(y_8 - b_8)^+$	1181.2126	1181.2401	23.3

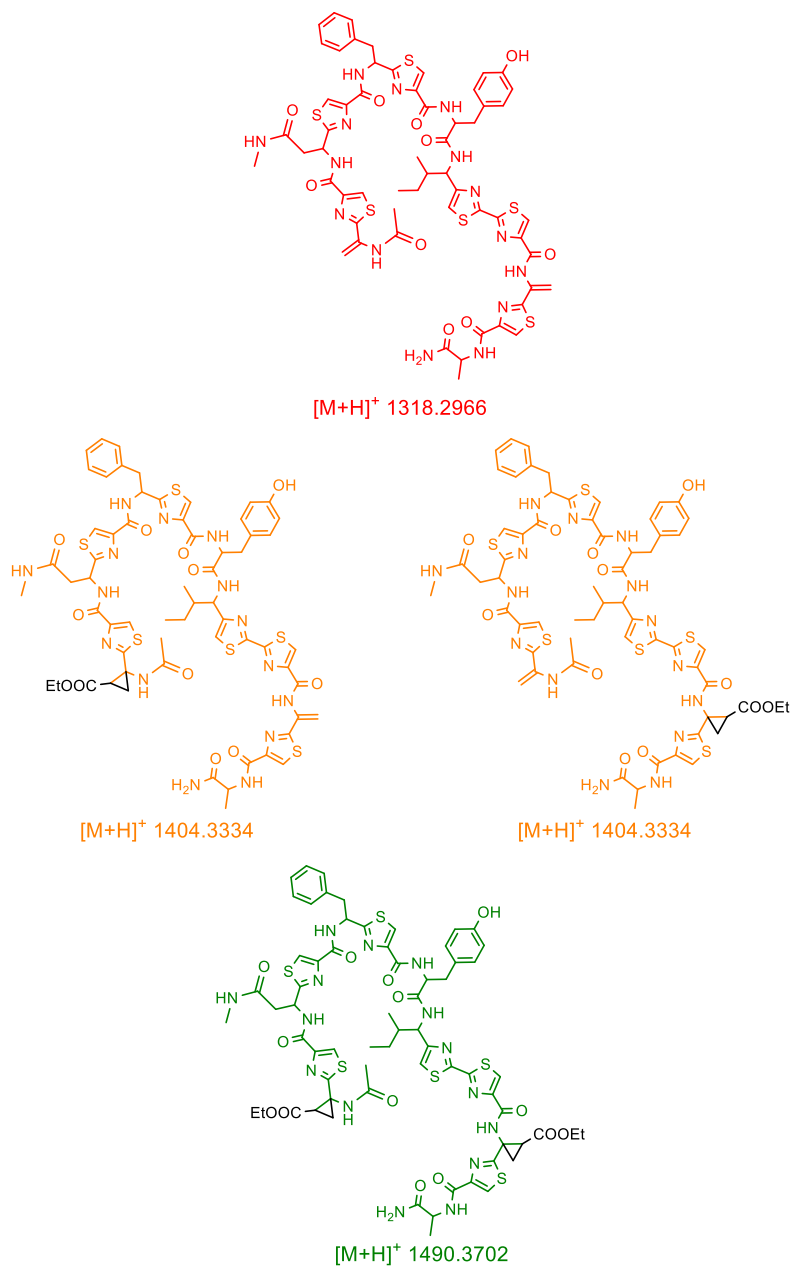
**Figure S3.16.** LC-MS/MS of linear TbtA I8-3Dha core (**9**).

LC-MS Method C; CID = 45 eV. Expected  $[M+H]^+$ : 1274.2340; Observed  $[M+H]^+$ : 1274.2338.



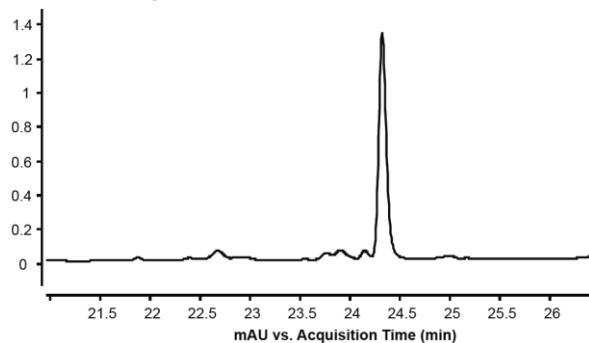
Ion type	Expected mass (Da)	Observed mass (Da)	Error (ppm)
$a_5^+$	771.1836	771.1816	2.6
$b_5^+$	799.1785	799.1683	12.8
$y_4^+$	476.0628	476.0695	14.1
$y_5^+$	639.1261	639.0739	81.7
$(y_8-a_5)^+$	729.1731	729.1767	4.9
$(y_4-b_8)^+$	459.0362	459.0201	35.1
$(y_8-b_4)^+$	594.1046	594.0944	17.2
$(y_8-b_5)^+$	757.1680	757.1713	4.4
$(y_8-b_6)^+$	992.1554	992.1307	24.9
$(y_7-b_6)^+$	840.1509	840.2430	109.6

**Figure S3.17.** HPLC and extracted ion chromatogram (EIC) traces of enzymatic cyclopropanations of linear TbtA core (analyzed by LC-MS Method A).

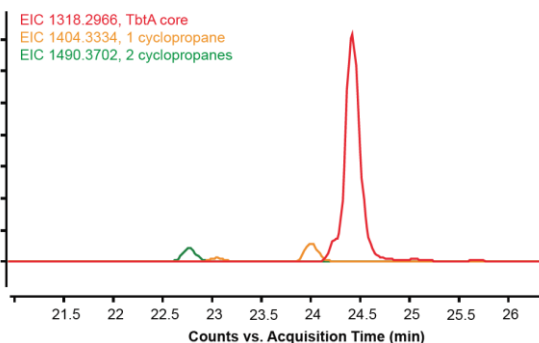


### Hemin

x10<sup>-2</sup> DAD1 - EWC:Sig=254,4 Ref=600,4

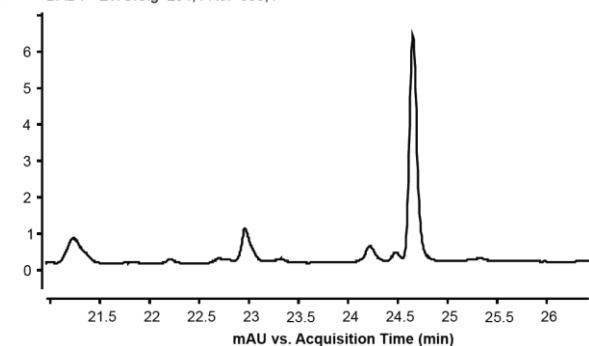


x10<sup>-5</sup>

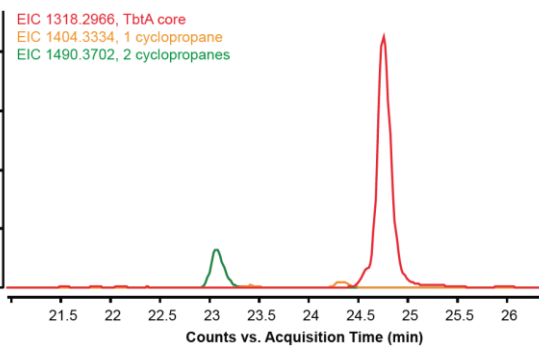


### P450<sub>TbtJ1</sub>-T234A

x10<sup>-1</sup> DAD1 - EWC:Sig=254,4 Ref=600,4

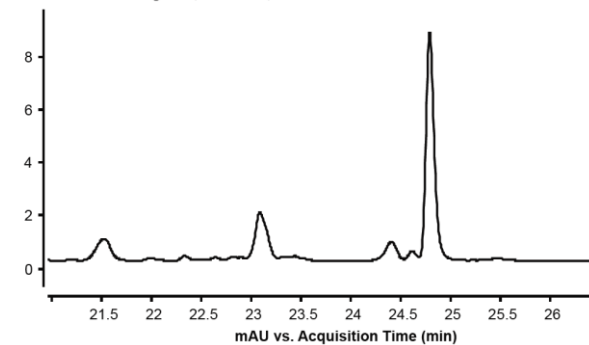


x10<sup>-5</sup>

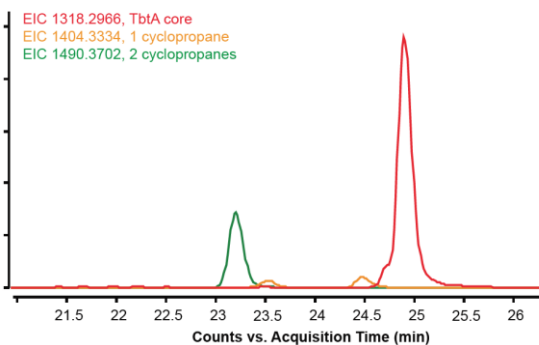


### P450<sub>TbtJ2</sub>-T247A

x10<sup>-1</sup> DAD1 - EWC:Sig=254,4 Ref=600,4

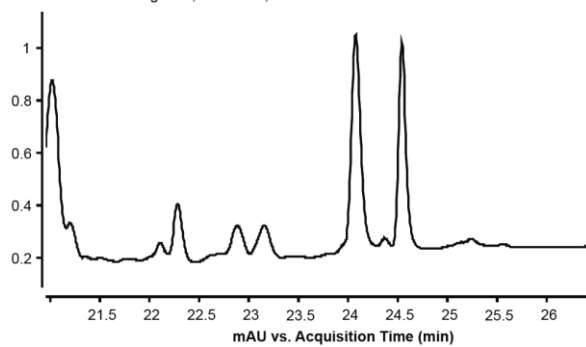


x10<sup>-5</sup>

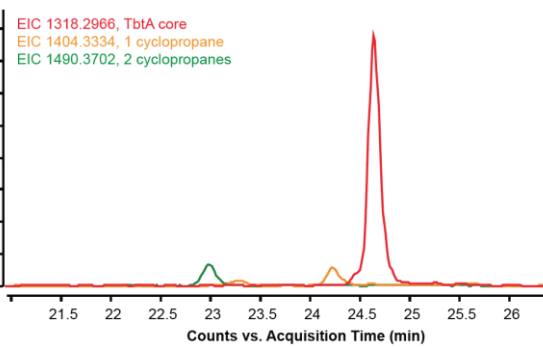


# P450<sub>PikC</sub>-T247A

x10<sup>-1</sup> DAD1 - EWC:Sig=254,4 Ref=600,4

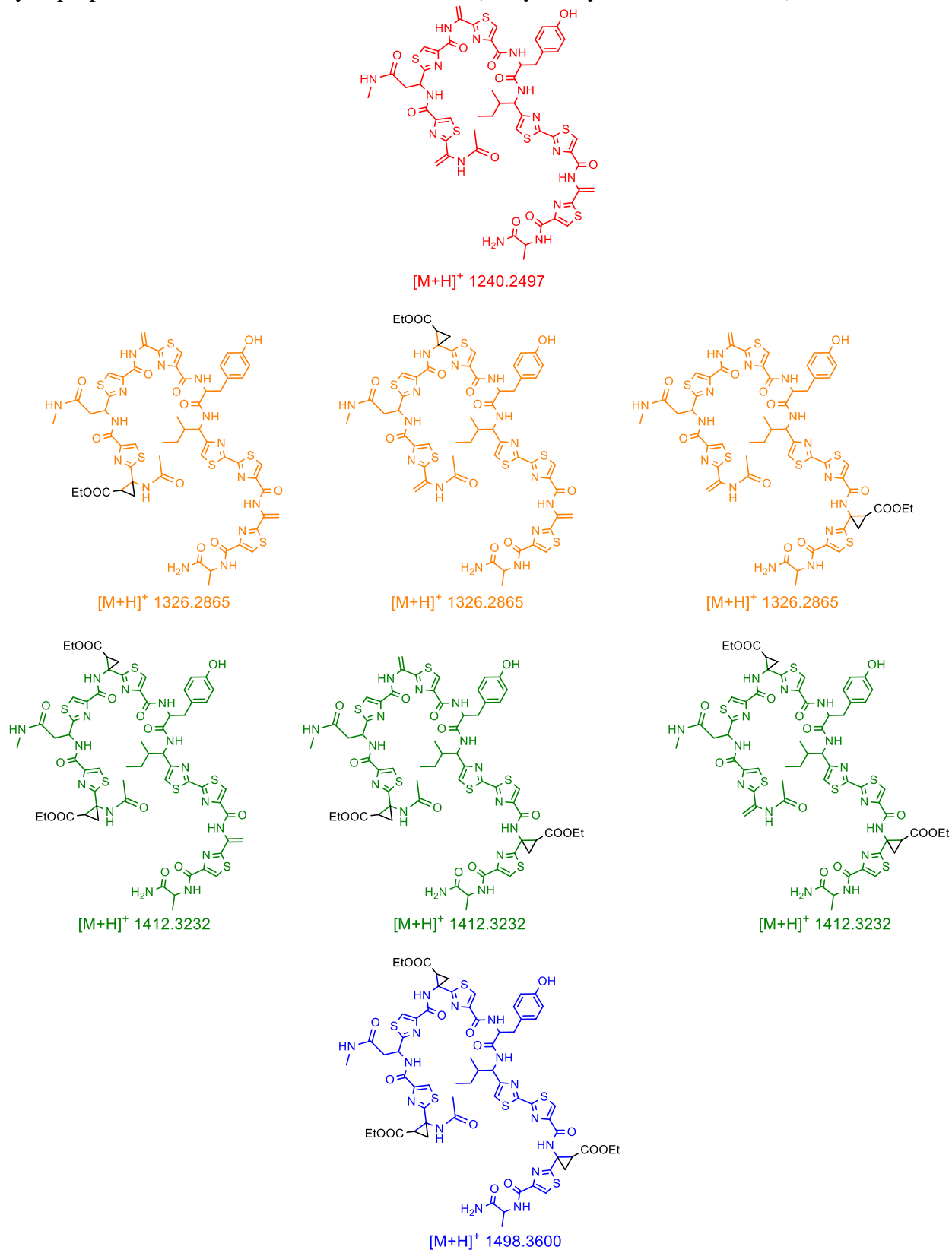


x10<sup>-4</sup>

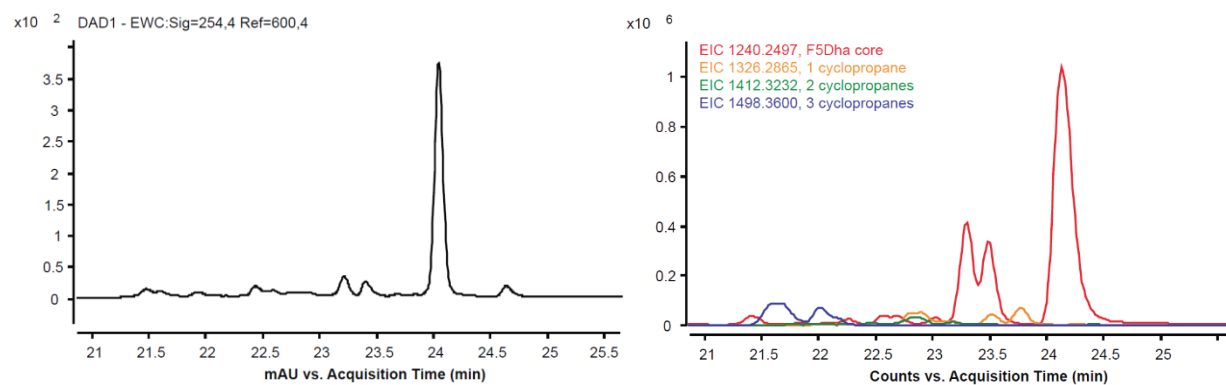




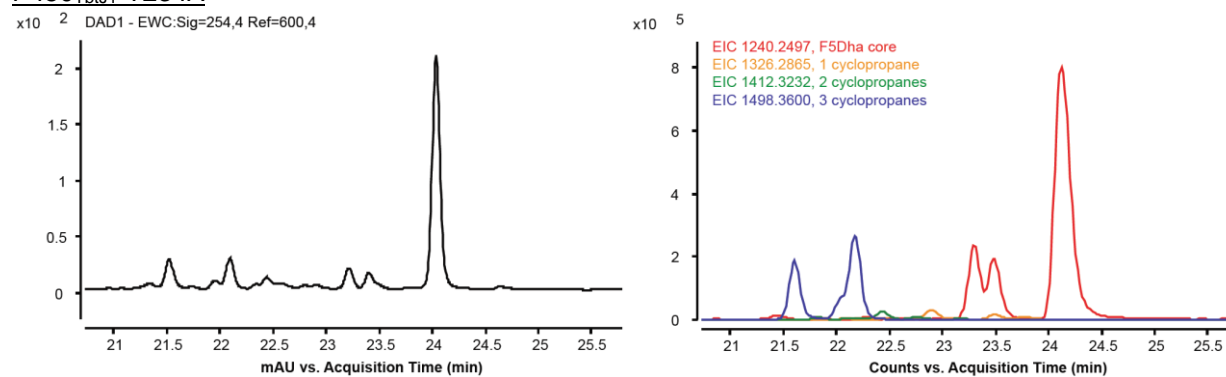
**Figure S3.18.** HPLC and extracted ion chromatogram (EIC) traces of enzymatic cyclopropanations of linear TbtA F5-3Dha core (analyzed by LC-MS Method A).



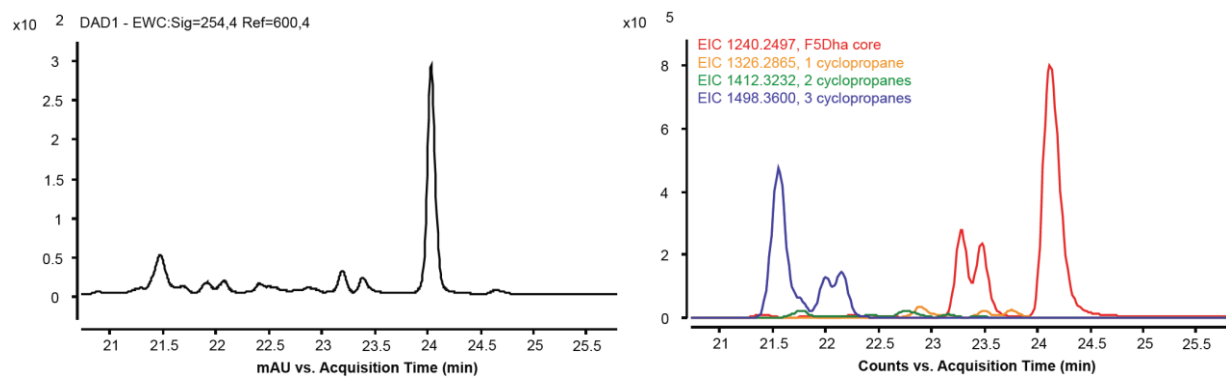
## Hemin



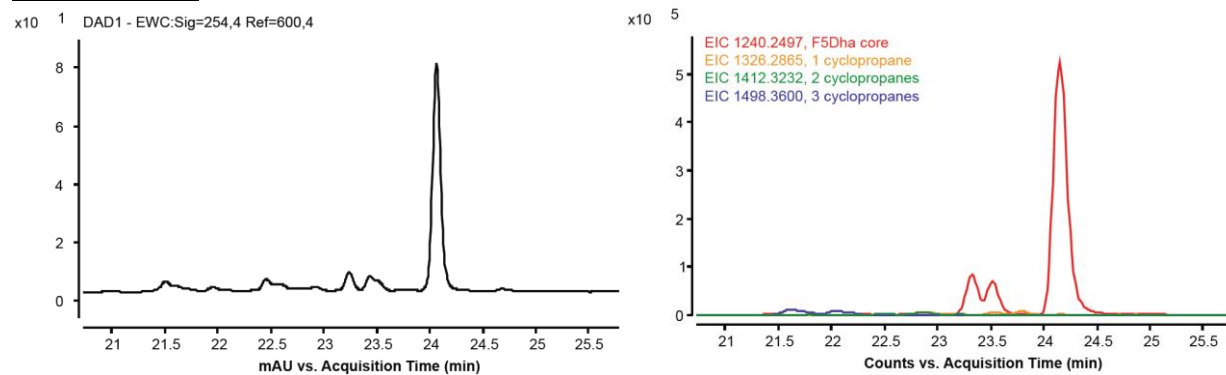
## P450<sub>TbtJ1</sub>-T234A



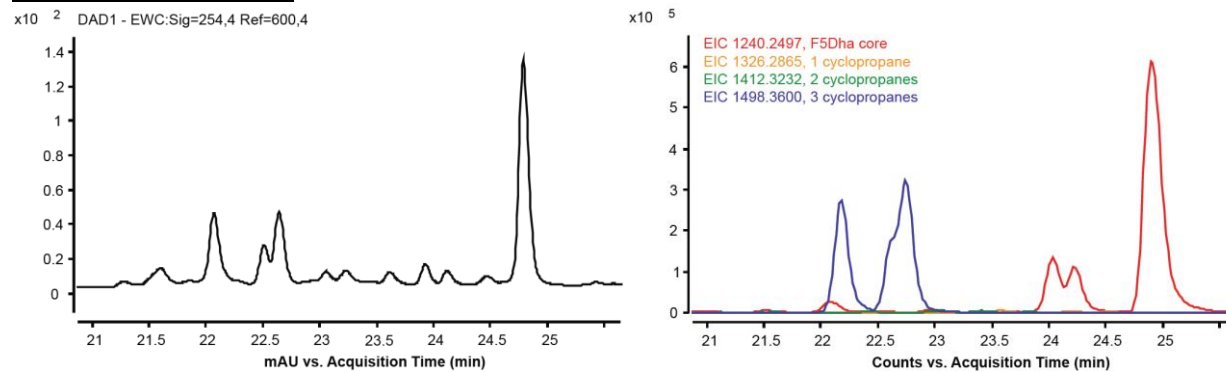
## P450<sub>TbtJ2</sub>-T247A



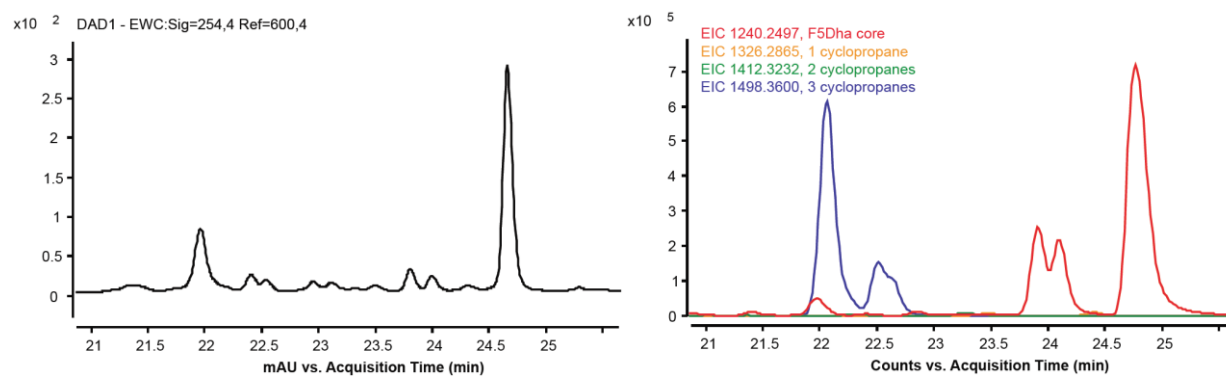
### P450<sub>PikC</sub>-T247A



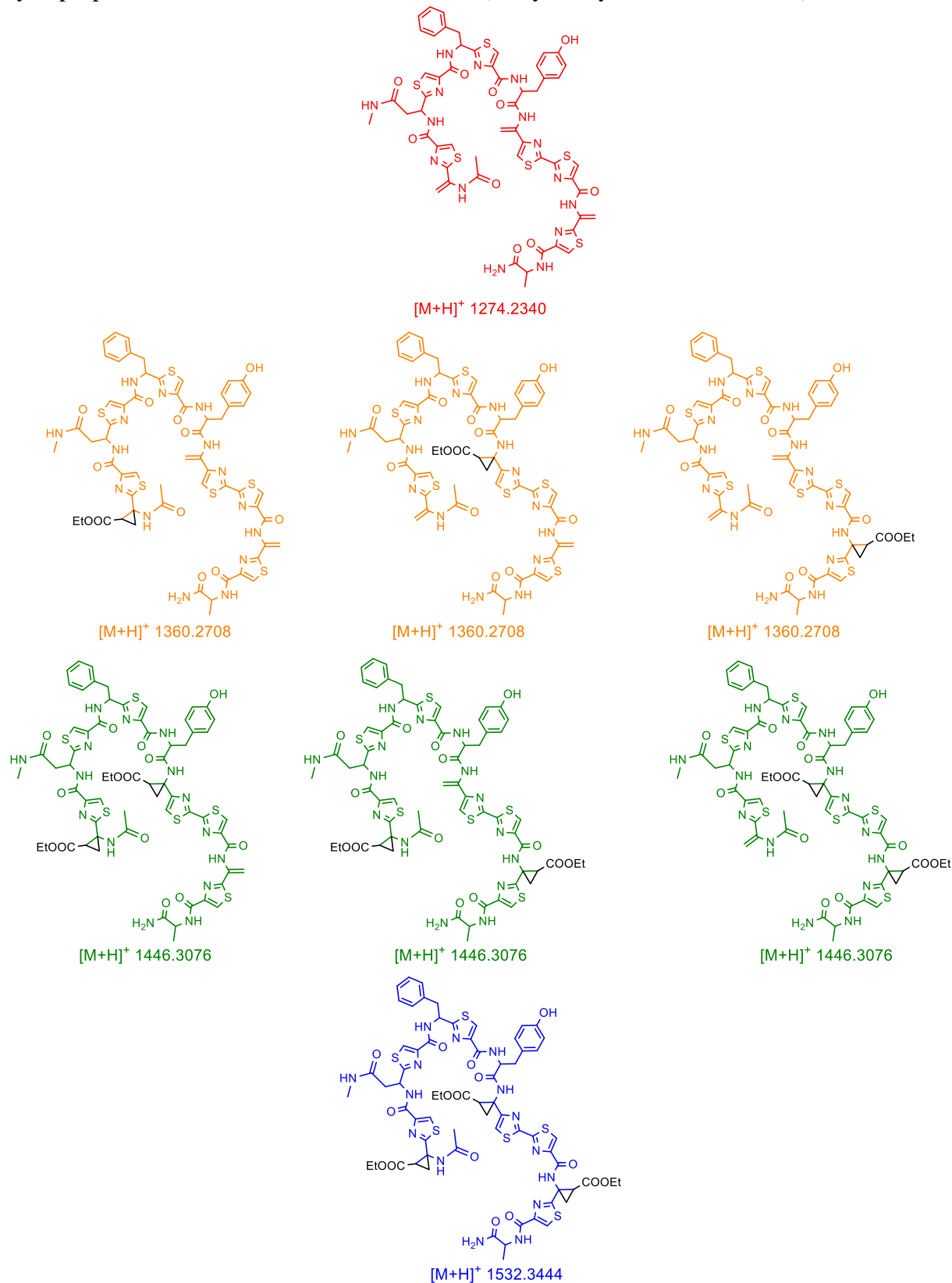
### P450<sub>TbtJ1</sub>-T234A/C340S



### P450<sub>TbtJ2</sub>-T247A/C353S

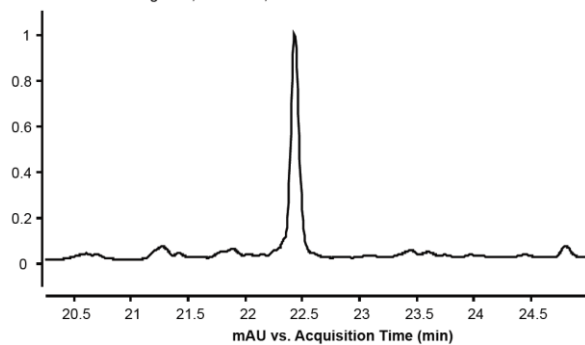


**Figure S3.19.** HPLC and extracted ion chromatogram (EIC) traces of enzymatic cyclopropanations of linear TbtA I8-3Dha core (analyzed by LC-MS Method A).

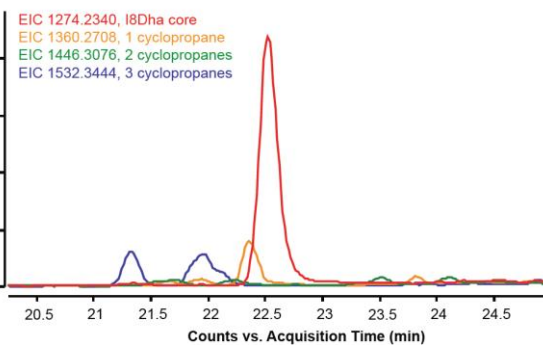


## Hemin

x10<sup>-2</sup> DAD1 - EWC:Sig=254,4 Ref=600,4

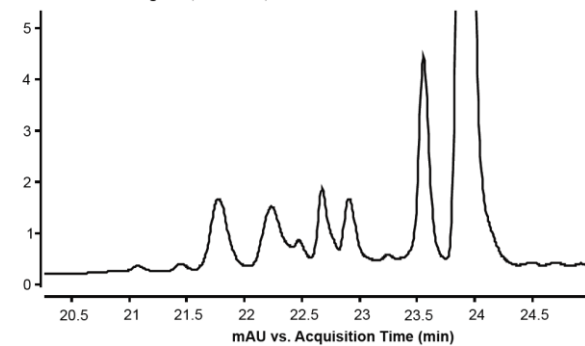


x10<sup>-5</sup>

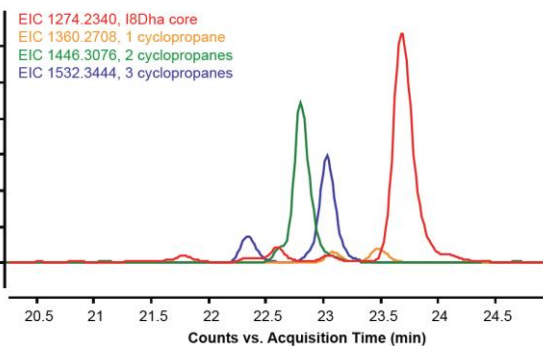


## P450<sub>TbtJ1</sub>-T234A

x10<sup>-1</sup> DAD1 - EWC:Sig=254,4 Ref=600,4

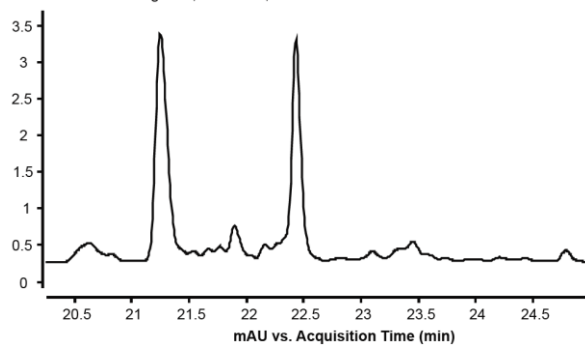


x10<sup>-5</sup>

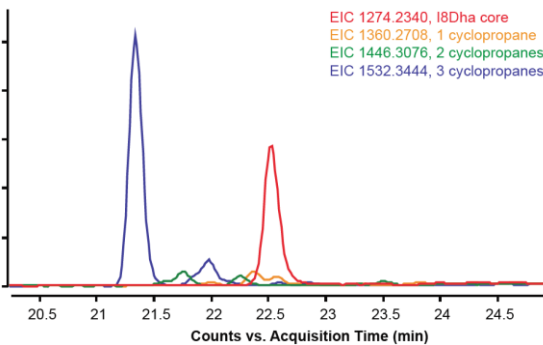


## P450<sub>TbtJ2</sub>-T247A

x10<sup>-1</sup> DAD1 - EWC:Sig=254,4 Ref=600,4

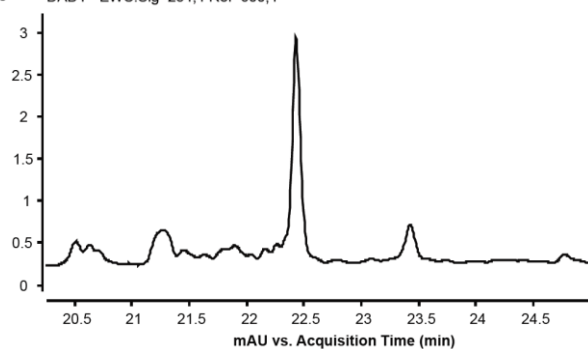


x10<sup>-5</sup>

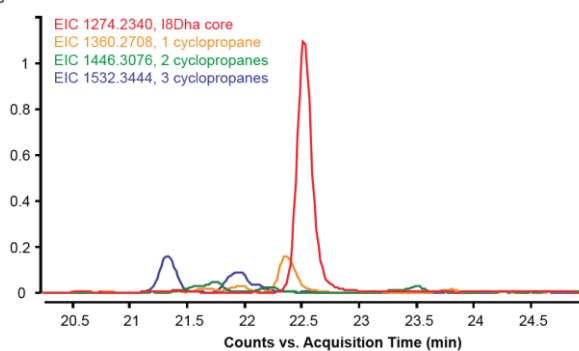


### P450<sub>PikC</sub>-T247A

x10<sup>-1</sup> DAD1 - EWC:Sig=254,4 Ref=600,4

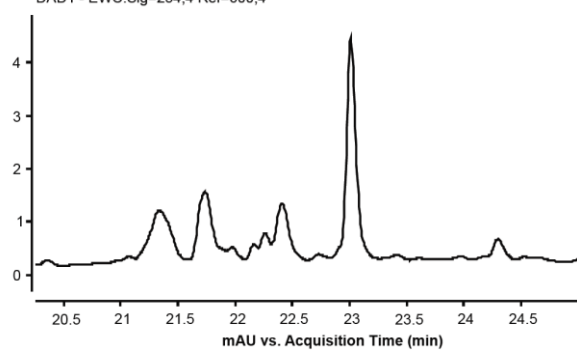


x10<sup>-5</sup>

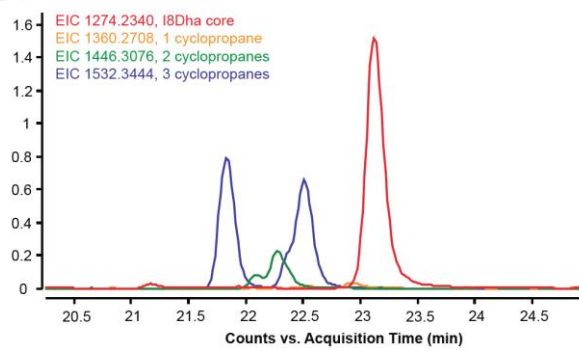


### P450<sub>TbtJ1</sub>-T234A/C340S

x10<sup>-1</sup> DAD1 - EWC:Sig=254,4 Ref=600,4

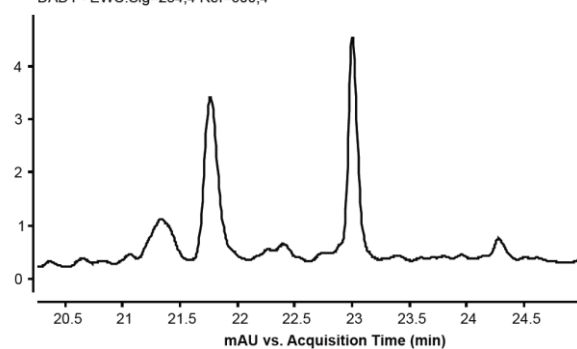


x10<sup>-5</sup>

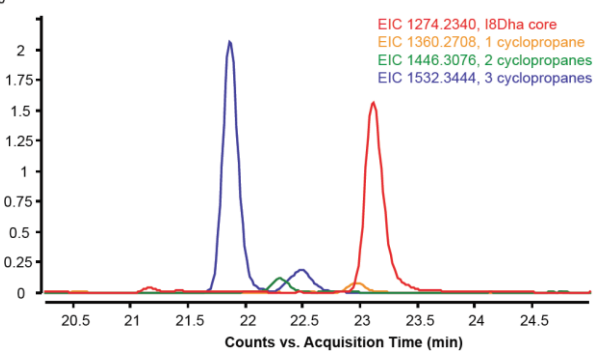


### P450<sub>TbtJ2</sub>-T247A/C353S

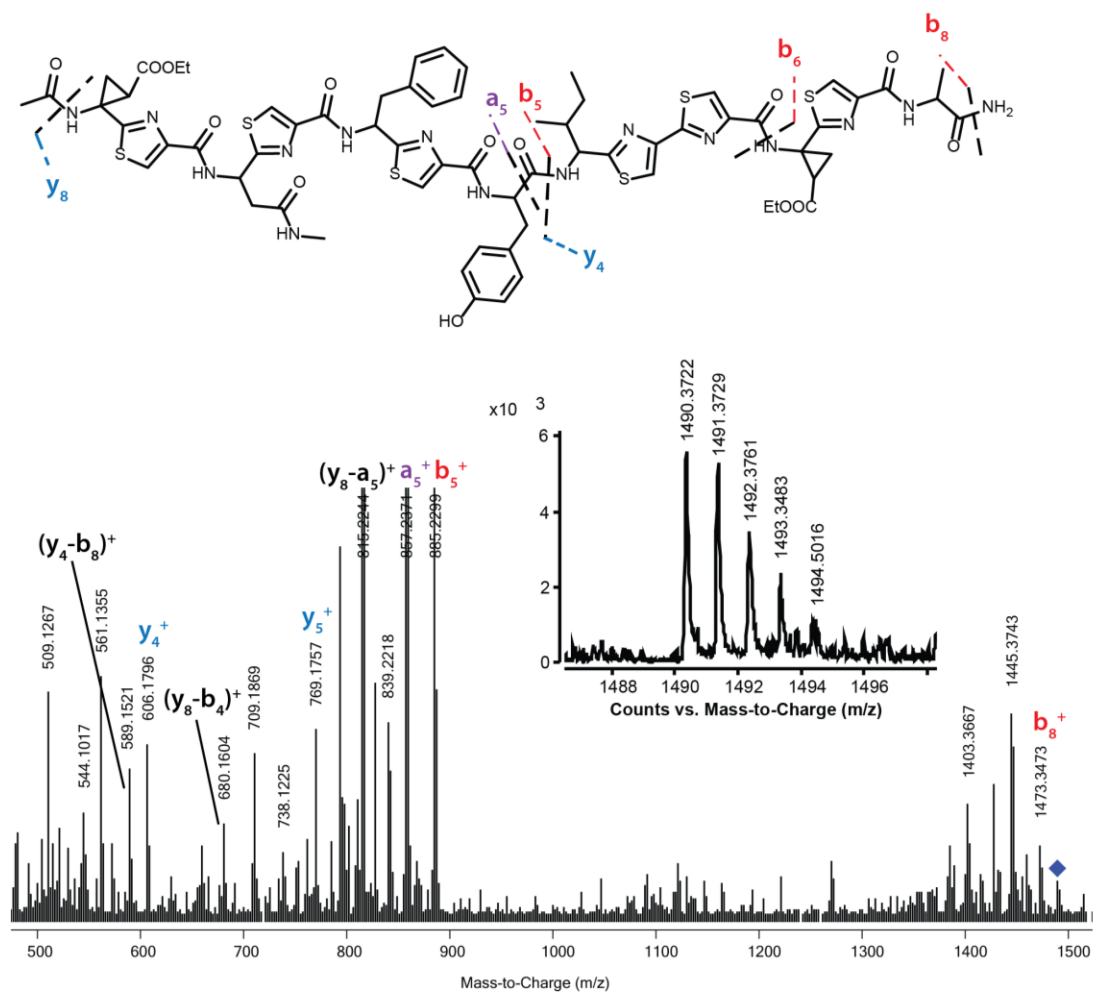
x10<sup>-1</sup> DAD1 - EWC:Sig=254,4 Ref=600,4



x10<sup>-5</sup>

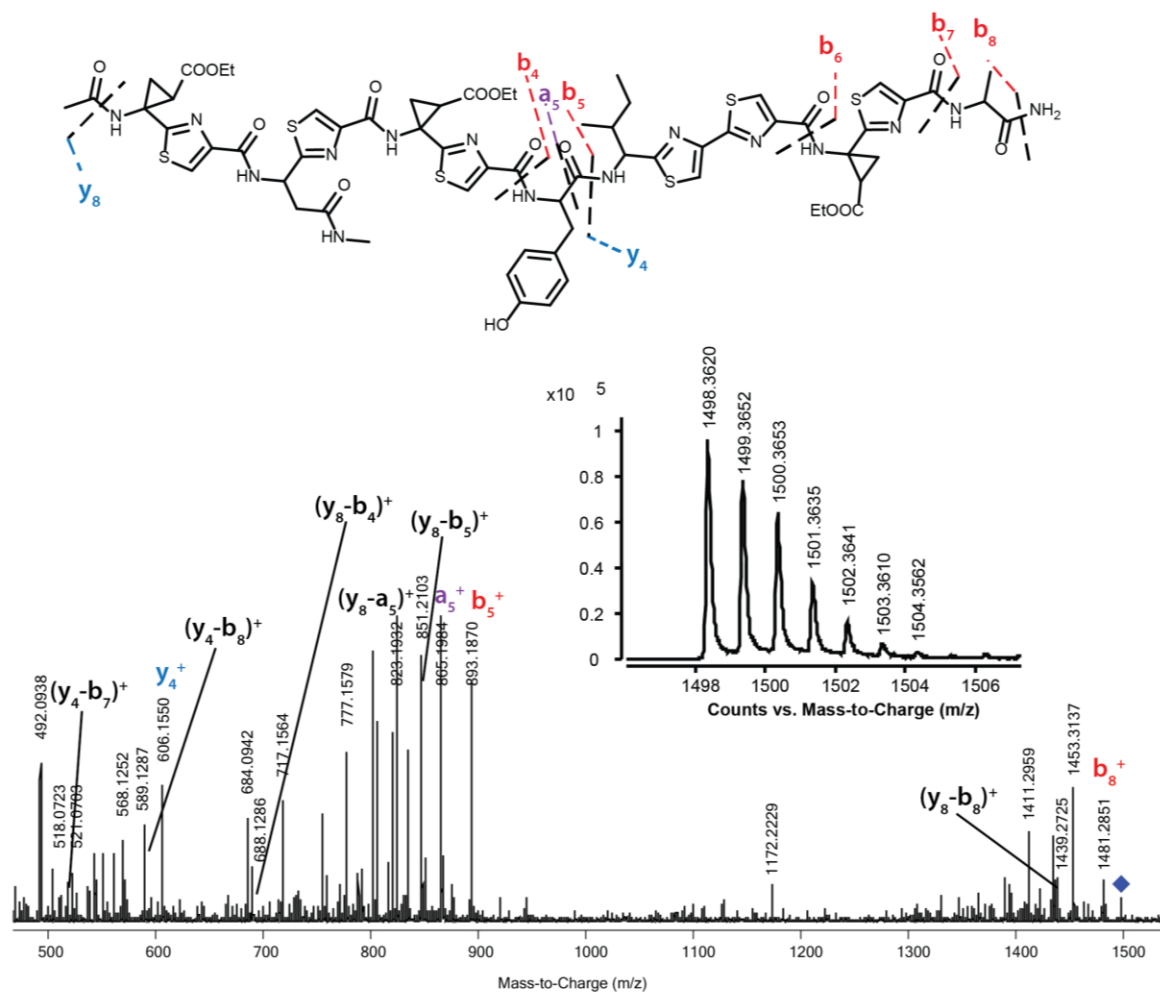


**Figure S3.20.** LC-MS/MS of cyclopropanated linear TbtA core.  
LC-MS Method A; CID = 55 eV. Expected  $[M+H]^+$ : 1490.3702; Observed  $[M+H]^+$ : 1490.3722.



Ion type	Expected mass (Da)	Observed mass (Da)	Error (ppm)
$a_5^+$	857.2204	857.2371	19.5
$b_5^+$	885.2153	885.2299	16.5
$b_8^+$	1473.3436	1473.3473	2.5
$y_4^+$	606.1622	606.1796	28.7
$y_5^+$	769.2255	769.1757	64.7
$(y_4-b_8)^+$	589.1356	589.1521	28.0
$(y_8-a_5)^+$	815.2098	815.2244	17.9
$(y_8-b_4)^+$	680.1414	680.1604	27.9

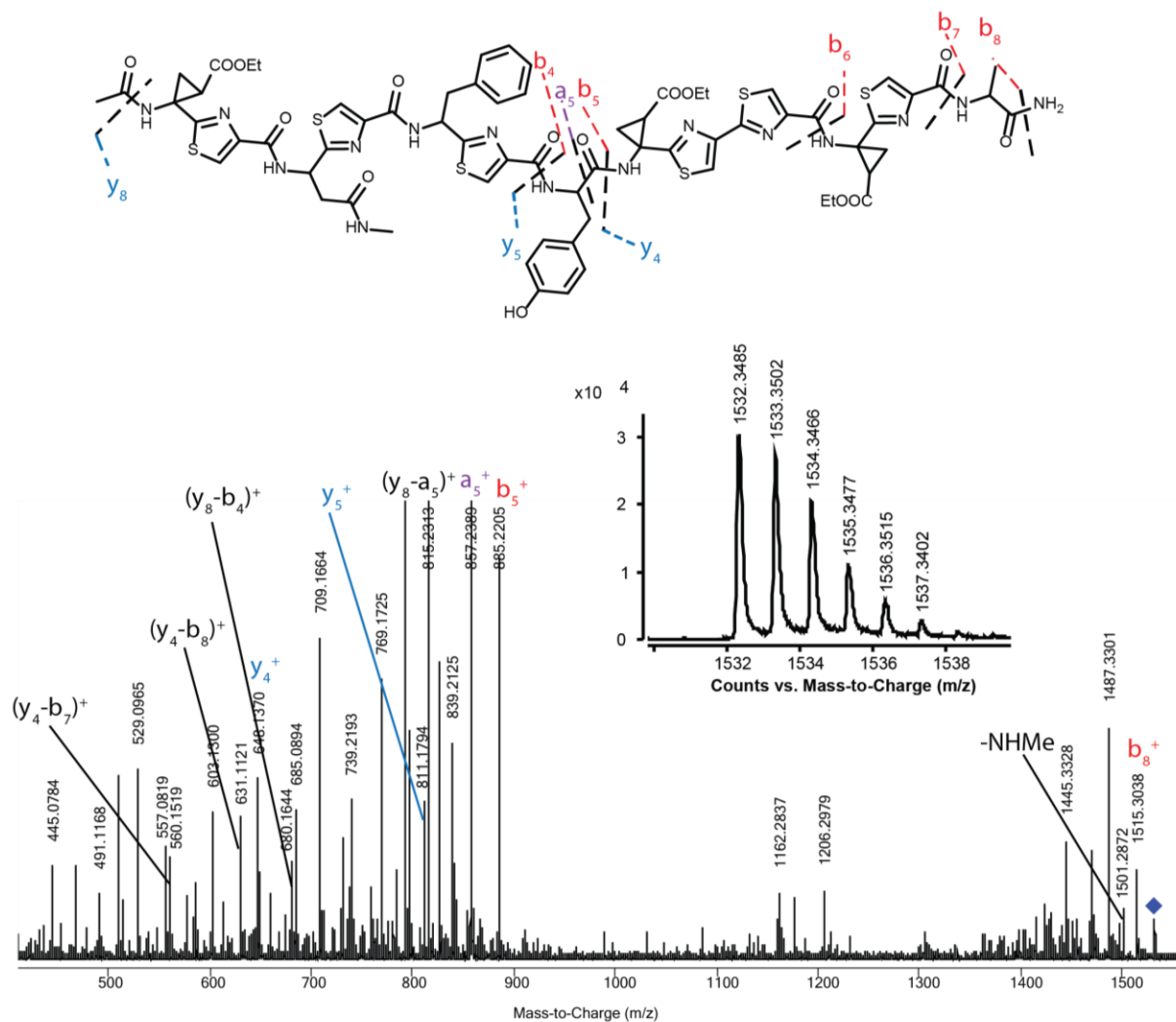
**Figure S3.21.** LC-MS/MS of cyclopropanated linear TbtA F5-3Dha core.  
LC-MS Method A; CID = 55 eV. Expected  $[M+H]^+$ : 1498.3600; Observed  $[M+H]^+$ : 1498.3620.



Ion type	Expected mass (Da)	Observed mass (Da)	Error (ppm)
$a_5^+$	865.2102	865.1984	13.6
$b_5^+$	893.2051	893.187	20.3
$b_8^+$	1481.3335	1481.2851	32.7
$y_4^+$	606.1622	606.155	11.9
$(y_4-b_7)^+$	518.0985	518.0723	50.6
$(y_4-b_8)^+$	589.1356	589.1287	11.7
$(y_8-a_5)^+$	823.1997	823.1932	7.9
$(y_8-b_4)^+$	688.1312	688.1286	3.8
$(y_8-b_5)^+$	851.1946	851.2103	18.4
$(y_8-b_8)^+$	1439.3229	1439.2725	35.0

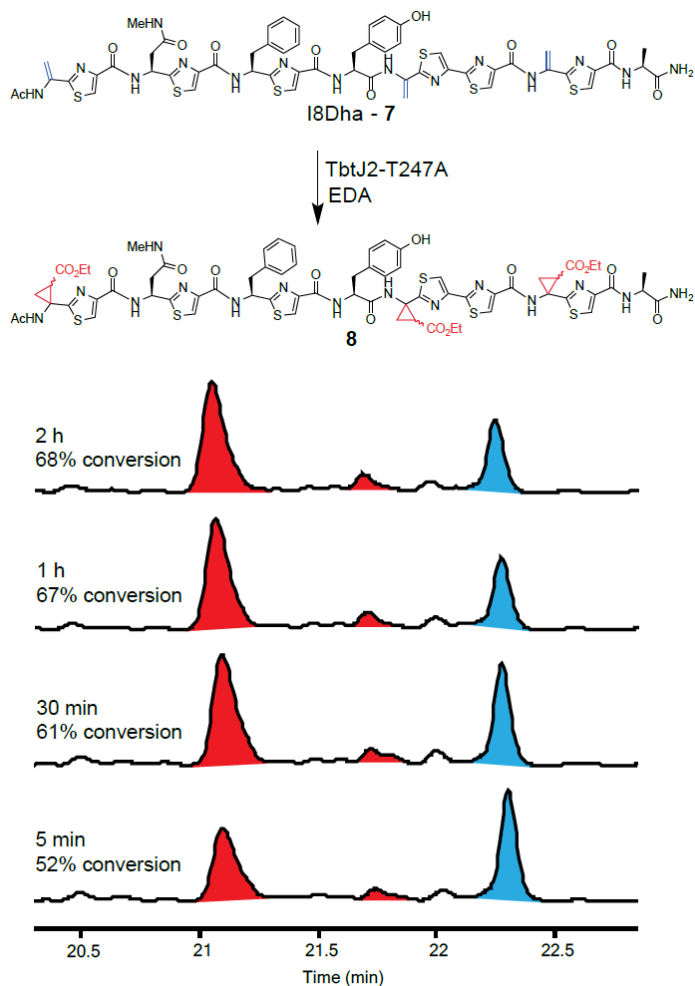


**Figure S3.22.** LC-MS/MS of cyclopropanated linear TbtA I8-3Dha core.  
LC-MS Method A; CID = 55 eV. Expected  $[M+H]^+$ : 1532.3444; Observed  $[M+H]^+$ : 1532.3485.

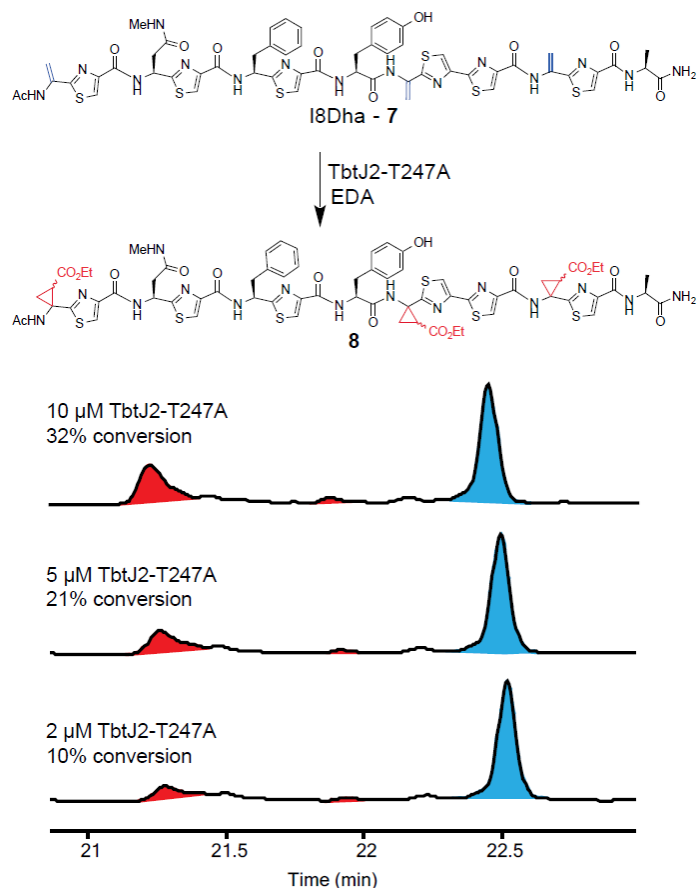


Ion type	Expected mass (Da)	Observed mass (Da)	Error (ppm)
$a_5^+$	857.2204	857.2389	21.6
$b_5^+$	885.2153	885.2205	5.9
$b_8^+$	1515.3178	1515.3038	9.2
$y_4^+$	648.1363	648.137	1.1
$y_5^+$	811.1997	811.1794	25.0
$(y_4-b_7)^+$	560.0727	560.1519	141.4
$(y_4-b_8)^+$	631.1098	631.1121	3.6
$(y_8-a_5)^+$	815.2098	815.2313	26.4
$(y_8-b_4)^+$	680.1414	680.1644	33.8
-NHMe	1501.3022	1501.2872	10.0

**Figure S3.23.** HPLC traces for time course experiments with TbtJ2-T247A and I8-3Dha core. Reactions were setup following the procedure for small-scale cyclopropanation of thiomuracin derivatives described above. Percent conversion was measured after 5 min, 30 min, 1 h, and 2 h. The starting material peak (I8Dha) is colored blue, and the 3 cyclopropane product peaks are colored red.

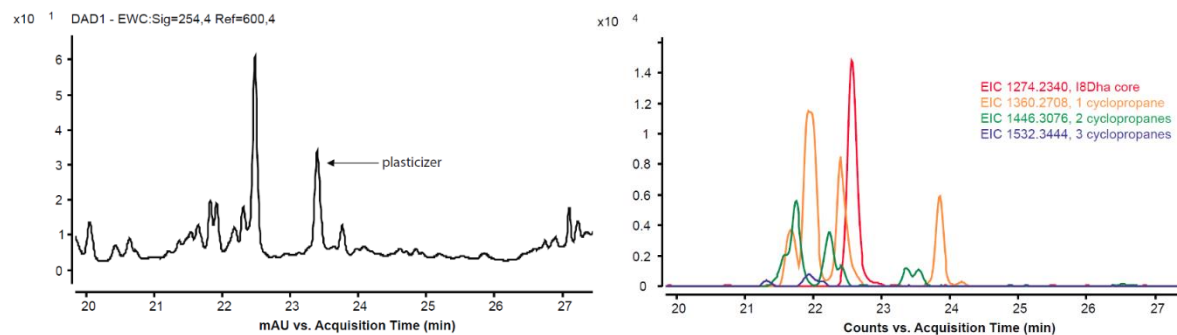


**Figure S3.24.** HPLC traces for catalyst loading experiments with TbtJ2-T247A and I8-3Dha core. Reactions were setup following the procedure for small-scale cyclopropanation of thiomuracin derivatives described above. Percent conversion was measured at enzyme loadings of 2 mol% (2  $\mu$ M), 5 mol% (5  $\mu$ M), and 10 mol% (10  $\mu$ M) after 16 hours. The starting material peak (I8Dha) is colored blue, and the 3 cyclopropane product peaks are colored red.

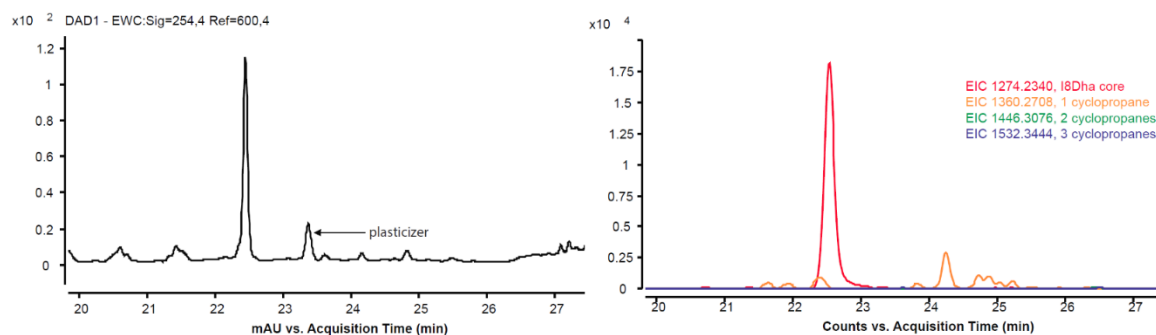


**Figure S3.25.** HPLC and extracted ion chromatogram (EIC) traces of cyclopropanations of linear TbtA I8-3Dha core with traditional catalysts (analyzed by LC-MS Method A).

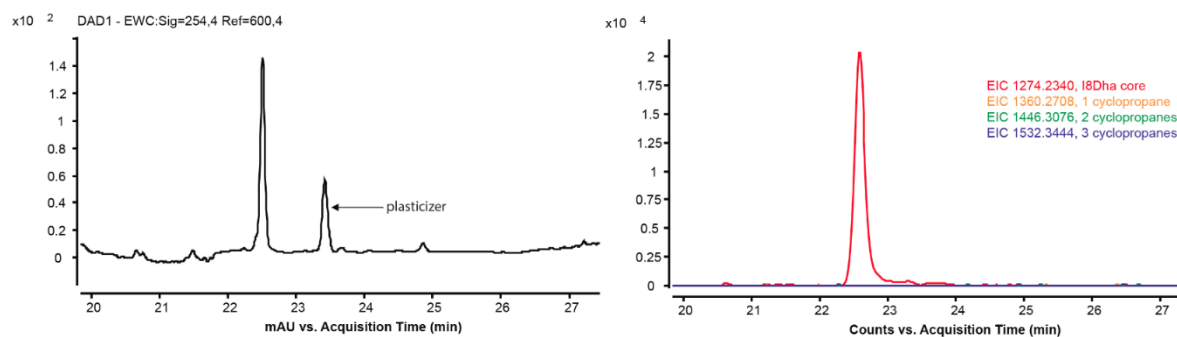
Iron (III) tetraphenylporphyrin in dichloromethane



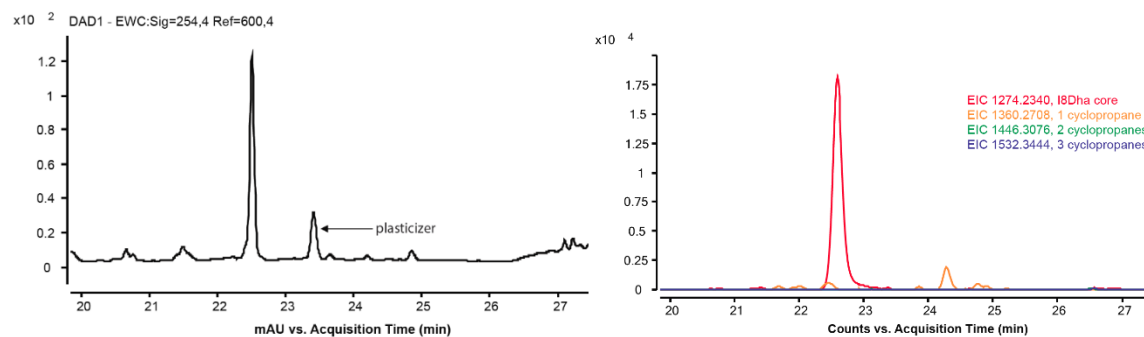
Rhodium (II) acetate in dichloromethane



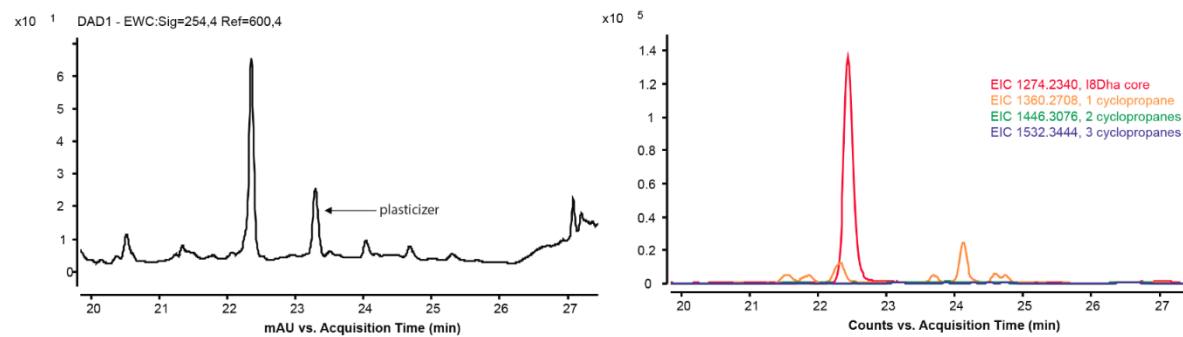
Rhodium (II) acetate in water



### Rhodium (II) octanoate in dichloromethane

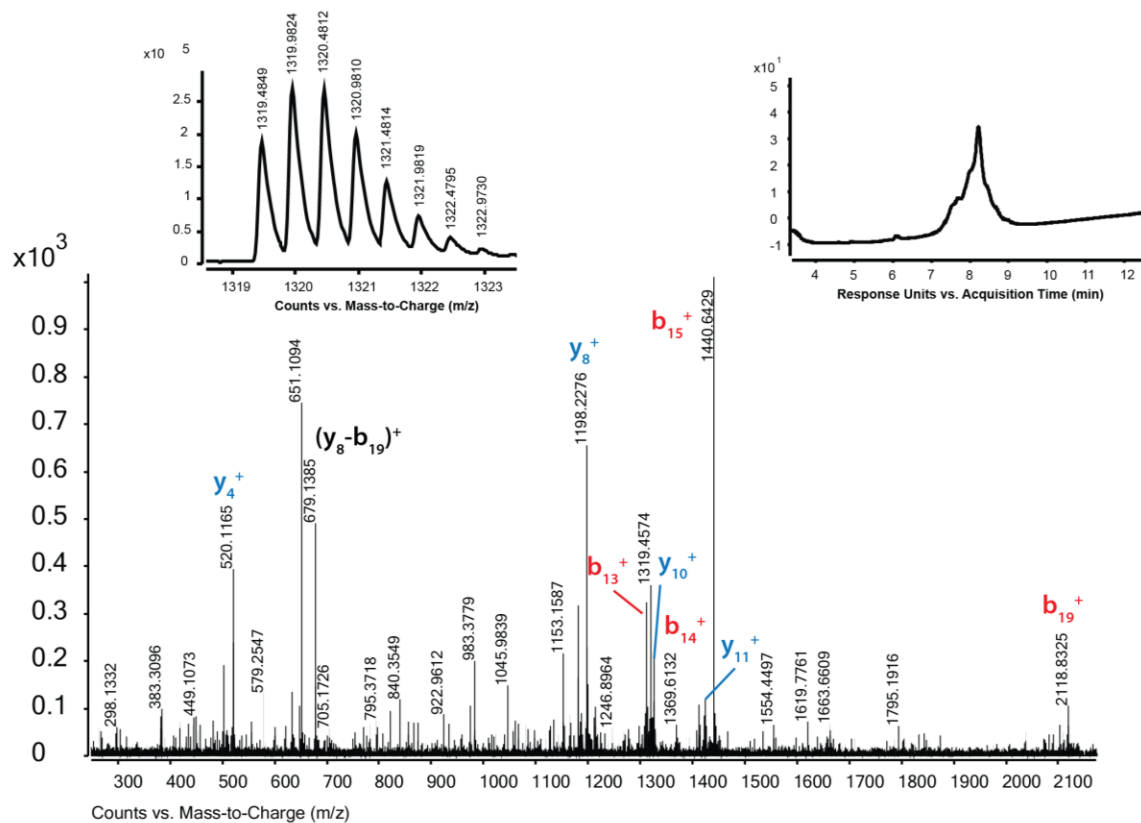
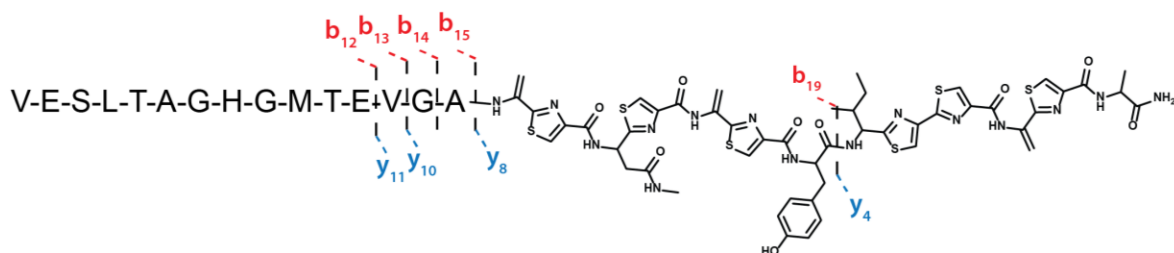


### Bis[rhodium( $\alpha,\alpha,\alpha',\alpha'$ -tetramethyl-1,3-benzenedipropionic acid)] ( $\text{Rh}_2(\text{esp})_2$ ) in dichloromethane



**Figure S3.26.** LC-MS/MS of 15LP-TbtA F5Dha.

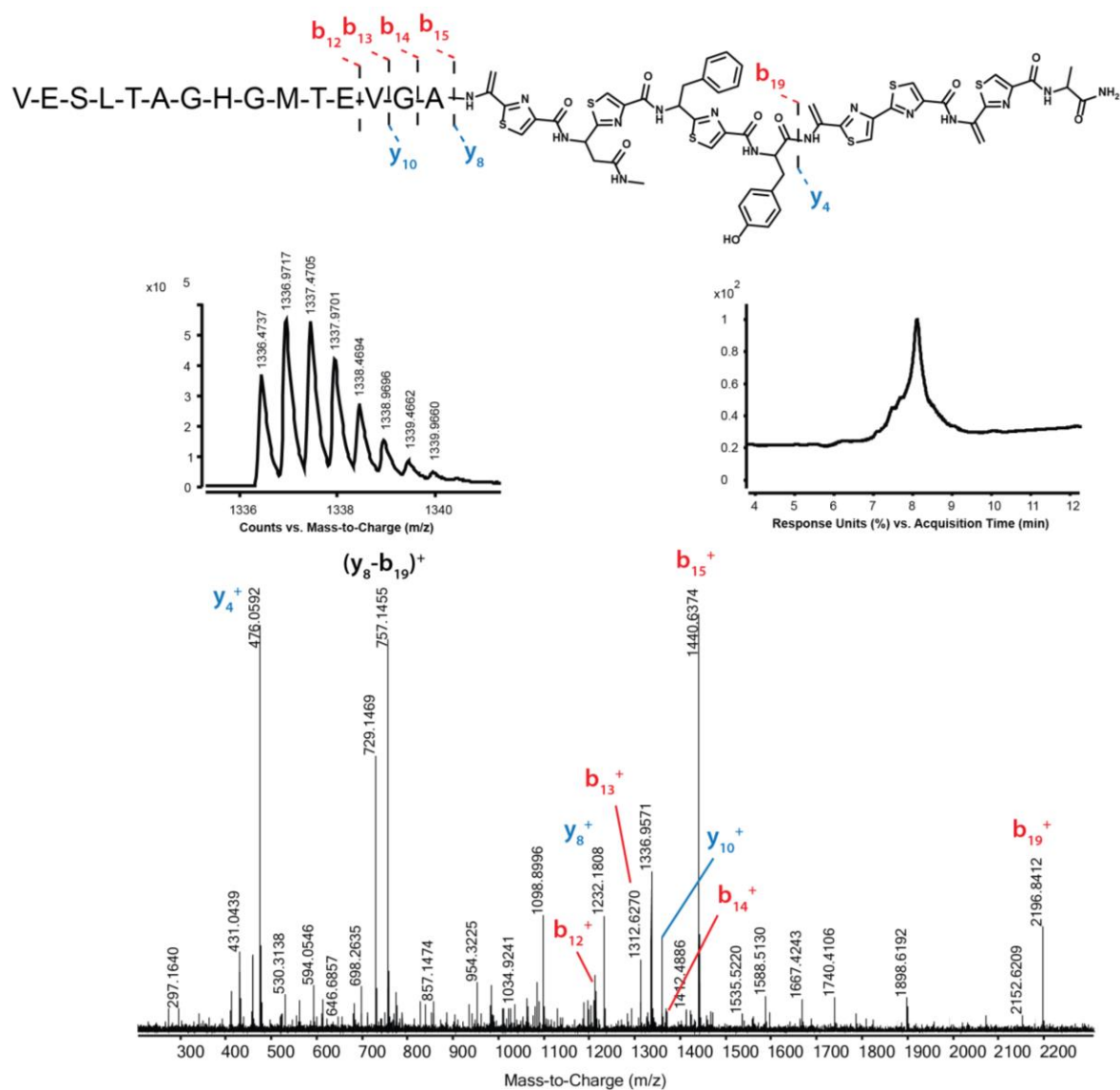
LC-MS Method C; CID = 45 eV. Expected  $[M+2H]^{+2}$ : 1319.4589; Observed  $[M+2H]^{+2}$ : 1319.4849.



Ion type	Expected mass (Da)	Observed mass (Da)	Error (ppm)
$b_{12}^{+}$	1213.5518	1213.6086	46.8
$b_{13}^{+}$	1312.6202	1312.6083	9.06
$b_{14}^{+}$	1369.6416	1369.6132	20.7
$b_{15}^{+}$	1440.6788	1440.6429	24.9
$b_{19}^{+}$	2118.7925	2118.8325	18.8
$y_4^{+}$	520.1254	520.1165	17.1
$y_8^{+}$	1198.2391	1198.2276	9.60
$y_{10}^{+}$	1326.2977	1326.3181	15.4
$y_{11}^{+}$	1425.3661	1425.3109	38.7
$(y_8 - b_{19})^{+}$	679.1210	679.1385	25.8

**Figure S3.27.** LC-MS/MS of 15LP-TbtA I8Dha.

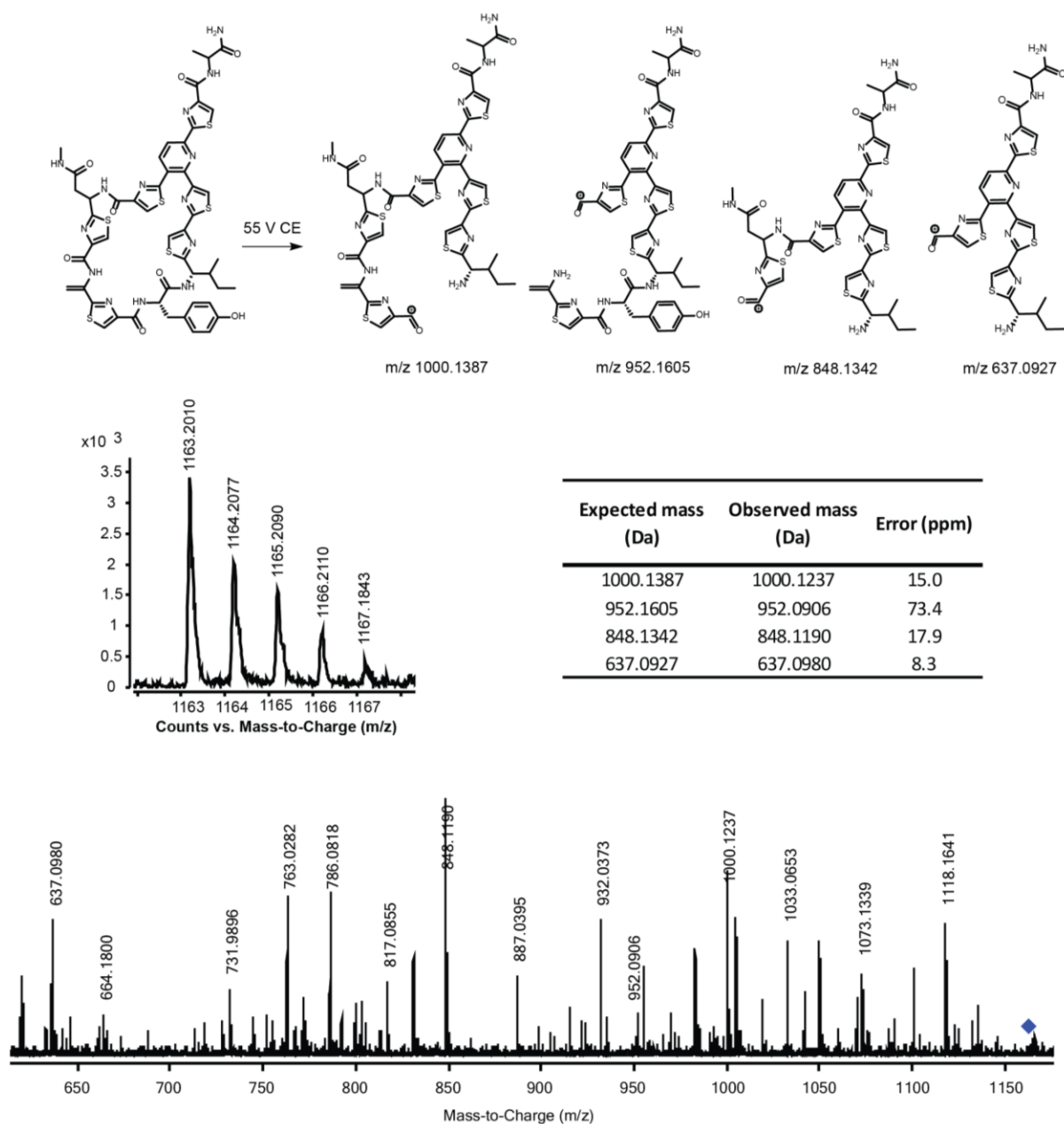
LC-MS Method C; CID = 45 eV. Expected  $[M+2H]^{+2}$ : 1336.4511; Observed  $[M+2H]^{+2}$ : 1336.4737.



Ion type	Expected mass (Da)	Observed mass (Da)	Error (ppm)
$b_{12}^+$	1213.5518	1213.4622	73.8
$b_{13}^+$	1312.6202	1312.6270	5.18
$b_{14}^+$	1369.6416	1369.5631	57.3
$b_{15}^+$	1440.6788	1440.6374	28.7
$b_{19}^+$	2196.8394	2196.8412	0.819
$y_4^+$	476.0628	476.0592	7.56
$y_8^+$	1232.2235	1232.1808	34.6
$y_{10}^+$	1360.2820	1360.2706	8.38
$(y_8 - b_{19})^+$	757.1680	757.1455	29.7

**Figure S3.28.** LC-MS/MS of cyclized TbtA F5Dha.

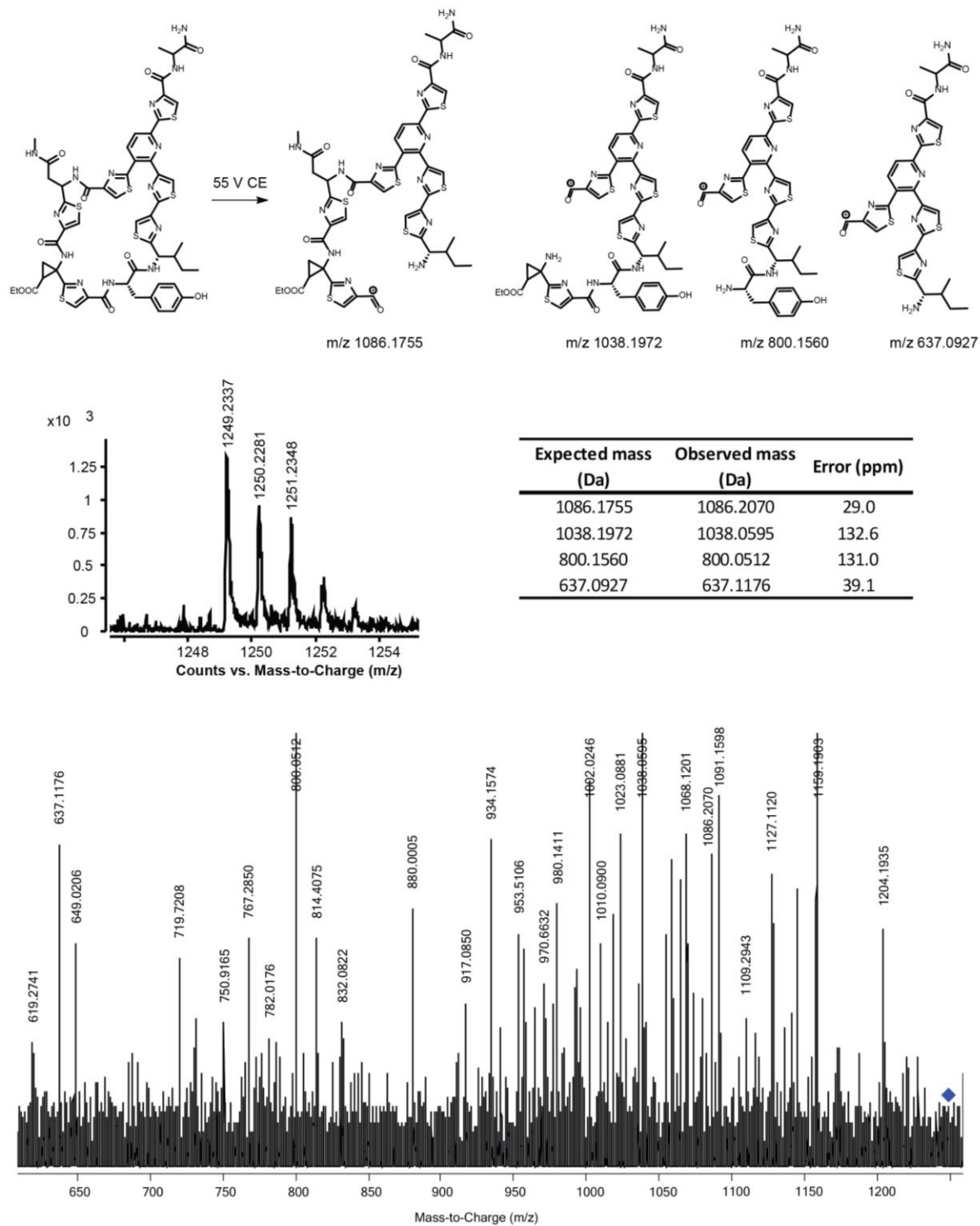
LC-MS Method B; CID = 55 eV. Expected  $[M+H]^+$ : 1163.2020; Observed  $[M+H]^+$ : 1163.2010.



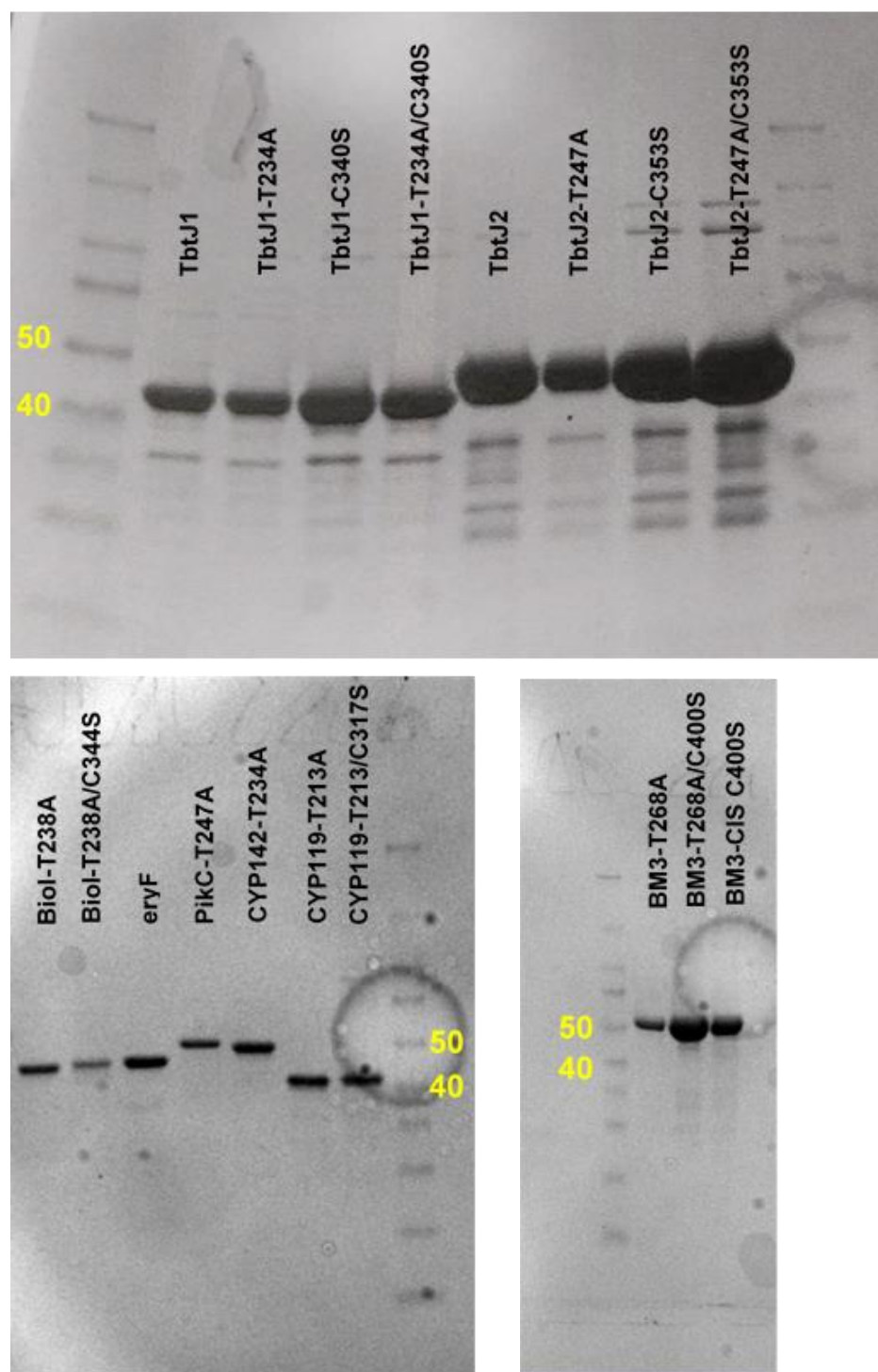


**Figure S3.29.** LC-MS/MS of cyclopropanated cyclized TbtA F5Dha.

LC-MS Method B; CID = 55 eV. Expected  $[M+H]^+$ : 1249.2388; Observed  $[M+H]^+$ : 1249.2337.



**Figure S3.30.** SDS-polyacrylamide gel electrophoresis analysis of his-tag purified P450 variants. The 40 kDa and 50 kDa protein markers are indicated with white text in each gel image.



## REFERENCES

1. Podust, L. M.; Sherman, D. H. *Nat. Prod. Rep.* **2012**, 29 (10), 1251.
2. Zhang, K.; Shafer, B. M.; Demars, M. D.; Stern, H. A.; Fasan, R. *J. Am. Chem. Soc.* **2012**, 134 (45), 18695.
3. Dietrich, J. A.; Yoshikuni, Y.; Fisher, K. J.; Woolard, F. X.; Ockey, D.; McPhee, D. J.; Renninger, N. S.; Chang, M. C. Y.; Baker, D.; Keasling, J. D. *ACS Chem. Biol.* **2009**, 4 (4), 261.
4. Singh, R.; Kolev, J. N.; Suter, P. A.; Fasan, R. *ACS Catal.* **2015**, 5 (3), 1685.
5. Machida, K.; Aritoku, Y.; Tsuchida, T. *J. Biosci. Bioeng.* **2009**, 107 (6), 596.
6. Machida, K.; Aritoku, Y.; Nakashima, T.; Arisawa, A.; Tsuchida, T. *J. Biosci. Bioeng.* **2008**, 105 (6), 649.
7. Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. *Science* **2013**, 339 (6117), 307.
8. Sreenilayam, G.; Fasan, R. *Chem. Commun.* **2015**, 51 (8), 1532.
9. Kan, S. B. J.; Lewis, R. D.; Chen, K.; Arnold, F. H. *Science* **2016**, 354 (6315), 1048.
10. Tyagi, V.; Bonn, R. B.; Fasan, R. *Chem. Sci.* **2015**, 6 (4), 2488.
11. Wang, Z. J.; Peck, N. E.; Renata, H.; Arnold, F. H. *Chem. Sci.* **2014**, 5 (2), 598.
12. McIntosh, J. A.; Coelho, P. S.; Farwell, C. C.; Wang, Z. J.; Lewis, J. C.; Brown, T. R.; Arnold, F. H. *Angew. Chem. Int. Ed.* **2013**, 52 (35), 9309.
13. Singh, R.; Bordeaux, M.; Fasan, R. *ACS Catal.* **2014**, 4 (2), 546.
14. Gober, J. G.; Brustad, E. M. *Curr. Opin. Chem. Biol.* **2016**, 35, 124.
15. Coelho, P. S.; Wang, Z. J.; Ener, M. E.; Baril, S. A.; Kannan, A.; Arnold, F. H.; Brustad, E. M. *Nat. Chem. Biol.* **2013**, 9 (8), 485.
16. Wang, Z. J.; Renata, H.; Peck, N. E.; Farwell, C. C.; Coelho, P. S.; Arnold, F. H. *Angew. Chem. Int. Ed.* **2014**, 53 (26), 6810.
17. Hernandez, K. E.; Renata, H.; Lewis, R. D.; Kan, S. *ACS Catal.* **2016**, 6, 7810.
18. Ortega, M. A.; van der Donk, W. A. *Cell Chem. Biol.* **2016**, 23 (1), 31.
19. Sato, T.; Izawa, K.; Aceña, J. L.; Liu, H. *Eur. J. Org. Chem.* **2016**, 2016 (16), 2757.

20. Cativiela, C.; Ordóñez, M. *Tetrahedron Asymmetry* **2009**, *20* (1), 1.
21. Brackmann, F.; de Meijere, A. *Chem. Rev.* **2007**, *107* (11), 4493.
22. Gober, J. G.; Rydeen, A. E.; Gibson-O'Grady, E. J.; Leuthaeuser, J. B.; Fetrow, J. S.; Brustad, E. M. *ChemBioChem* **2016**, *17* (5), 394.
23. Pyne, S. G.; Schafer, K.; Skelton, B. W.; White, A. H. *Aust. J. Chem.* **1998**, *51* (2), 127.
24. Bagley, M. C.; Dale, J. W.; Merritt, E. A.; Xiong, X. *Chem. Rev.* **2005**, *105* (2), 685.
25. Liu, W.; Xue, Y.; Ma, M.; Wang, S.; Liu, N.; Chen, Y. *ChemBioChem* **2013**, *14* (13), 1544.
26. Zheng, Q.; Wang, S.; Liao, R.; Liu, W. *ACS Chem. Biol.* **2016**, *11* (10), 2673.
27. Morris, R. P.; Leeds, J. A.; Naegeli, H. U.; Oberer, L.; Memmert, K.; Weber, E.; LaMarche, M. J.; Parker, C. N.; Burrer, N.; Esterow, S.; Hein, A. E.; Schmitt, E. K.; Krastel, P. *J. Am. Chem. Soc.* **2009**, *131* (16), 5946.
28. Young, T. S.; Dorrestein, P. C.; Walsh, C. T. *Chem. Biol.* **2012**, *19* (12), 1600.
29. Hudson, G. A.; Zhang, Z.; Tietz, J. I.; Mitchell, D. A.; van der Donk, W. A. *J. Am. Chem. Soc.* **2015**, *137* (51), 16012.
30. Sherman, D. H.; Li, S.; Yermalitskaya, L. V.; Kim, Y.; Smith, J. A.; Waterman, M. R.; Podust, L. M. *J. Biol. Chem.* **2006**, *281* (36), 26289.
31. Nagano, S.; Li, H.; Shimizu, H.; Nishida, C.; Ogura, H.; Ortiz de Montellano, P. R.; Poulos, T. L. *J. Biol. Chem.* **2003**, *278* (45), 44886.
32. Wever, W. J.; Bogart, J. W.; Baccile, J. A.; Chan, A. N.; Schroeder, F. C.; Bowers, A. A. *J. Am. Chem. Soc.* **2015**, *137* (10), 3494.
33. Wever, W. J.; Bogart, J. W.; Bowers, A. A. *J. Am. Chem. Soc.* **2016**, *138* (41), 13461.
34. Li, J.; Cisar, J. S.; Zhou, C.-Y.; Vera, B.; Williams, H.; Rodríguez, A. D.; Cravatt, B. F.; Romo, D. *Nat. Chem.* **2013**, *5* (6), 510.
35. Robles, O.; Serna-Saldívar, S. O.; Gutiérrez-Urbe, J. A.; Romo, D. *Org. Lett.* **2012**, *14* (6), 1394.
36. Zhang, Z.; Hudson, G. A.; Mahanta, N.; Tietz, J. I.; van der Donk, W. A.; Mitchell, D. A. *J. Am. Chem. Soc.* **2016**, *138* (48), 15507.
37. Key, H. M.; Dydio, P.; Clark, D. S.; Hartwig, J. F. *Nature* **2016**, *534*, 534.

38. Dydio, P.; Key, H. M.; Nazarenko, A.; Rha, J. Y.-E.; Seyedkazemi, V.; Clark, D. S.; Hartwig, J. F. *Science* **2016**, *354* (6308), 102.
39. Fu, Z.-Q. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2005**, *61* (12), 1643.
40. Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66* (2), 213.
41. Savino, C.; Montemiglio, L. C.; Sciara, G.; Miele, A. E.; Kendrew, S. G.; Jemth, P.; Gianni, S.; Vallone, B. *J. Biol. Chem.* **2009**, *284* (42), 29170.
42. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66* (4), 486.
43. Wang, S.; Otani, Y.; Liu, X.; Kawahata, M.; Yamaguchi, K.; Ohwada, T. *J. Org. Chem.* **2014**, *79* (11), 5287.
44. Hughes, R. A.; Thompson, S. P.; Alcaraz, L.; Moody, C. J. *J. Am. Chem. Soc.* **2005**, *127* (44), 15644.
45. Ranocchiari, M.; Mezzetti, A. *Organometallics* **2009**, *28* (13), 3611.
46. Morrison, P. M.; Foley, P. J.; Warriner, S. L.; Webb, M. E. *Chem. Commun.* **2015**, *51* (70), 13470.
47. Antos, J. M.; Francis, M. B. *J. Am. Chem. Soc.* **2004**, *126* (33), 10256.
48. Voss, N. R.; Gerstein, M. *Nucleic Acids Res.* **2010**, *38* (Web Server issue), W555.