

EXPANDING THE SCOPE OF METALLOPROTEIN  
FAMILIES AND SUBSTRATE CLASSES IN NEW-TO-  
NATURE REACTIONS

Thesis by  
Anders Matthew Knight

In Partial Fulfillment of the Requirements for the Degree of  
Doctor of Philosophy

The logo for the California Institute of Technology (Caltech), featuring the word "Caltech" in a bold, orange, sans-serif font.

CALIFORNIA INSTITUTE OF TECHNOLOGY  
Pasadena, California

2020  
(Defended June 9, 2020)

© 2020

Anders Matthew Knight

ORCID: 0000-0001-9665-8197

## ACKNOWLEDGEMENTS

I would first like to thank my advisor, Prof. Frances Arnold, for her help and guidance during my thesis work. When I spoke with Frances at my Caltech interview, she told me that if I wanted to learn how to engineer proteins, the Arnold lab was the best place in the world to do it. As usual, Frances was correct. Thank you for challenging me to come up with interesting, novel, and useful research questions, and for giving me the freedom to apply myself in answering those questions. I am thankful to my committee members, Prof. Sarah Reisman, Prof. William Clemons, Prof. Mikhail Shapiro, and Prof. Justin Bois for their insights and guidance on my PhD research. I would also like to thank Justin for his pivotal role in my graduate education, having taught most of the courses I took at Caltech and teaching me most of the programming I know. His skill and enthusiasm for educating junior scientists is immediately apparent.

The Arnold laboratory is full of incredibly talented scientists. Dr. Sabine Brinkmann-Chen has been there from the beginning for research, writing, and life advice. I am indebted to Dr. Andrew Buller and Dr. David Romney for mentoring me during my rotation and beyond. Dr. Rusty Lewis, Dr. Stephan Hammer, Dr. Oliver Brandenburg, Dr. Tina Boviile, Dr. Nicholas Porter, and Dr. Xiongyi Huang all provided excellent feedback on my new project ideas and helped me decipher data. I would like to thank Dr. Jennifer Kan in particular for her constant mentorship, advice, and support throughout my PhD – my time at Caltech would have been very different without your guidance. I have learned so much from collaborations and discussions with Lucas Schaus, Ruijie Zhang, Patrick Almhjell, Ella Watkins, Bruce Wittmann, Nicholas Porter, and Nat Goldberg. Thank you all for making me a better scientist, and, together with Austin Dulaney and Brad Boviile, for making my time outside of work enjoyable. I am richer for having known you, friends.

I would not have gotten to Caltech in the first place if it were not for the many role models and advisors on my way. My high-school chemistry teacher, Michael Smits, inspired me to study chemistry and chemical engineering in college. My undergraduate research advisor, Prof. Silvia Cavagnero, took a chance in hiring an enthusiastic but inexperienced freshman with no background on proteins to do research in her group. At the time I thought I would be a “traditional” chemical engineer, but I was quickly hooked on researching proteins. The foundation of both technical skills

and the research process that I learned in your laboratory have been invaluable. Prof. Randall Goldsmith, on top of being a fantastic teacher, helped me navigate the process of deciding to pursue a PhD and which program to choose. Prof. Uwe Bornscheuer guided me through my first forays into protein engineering and helped lay the groundwork for me to hit the ground running when I started at Caltech.

Mental health is a common topic when speaking about graduate school, and I am not alone in my struggles over the last five years. I am grateful to Dr. Lee Coleman for helping me to persevere, and to be kind to myself along the way.

I am thankful to have an incredibly supportive family. I know that my parents, Matthew and Carolyn Knight, and sister, Freya Ludeman, are always there for me. I look forward to celebrating in person when we can. Lastly, thank you to my partner, Silken Jones, for being with me through the ups and downs of grad school over the last four years. I cannot wait to start the next chapter of our adventure together.

## ABSTRACT

Heme proteins, in particular cytochromes P450, have been extensively used in biocatalytic applications due to their high degree of regio-, chemo-, and stereoselectivity in oxene-transfer reactions. In 2013, it was shown for the first time that engineered heme proteins can also catalyze analogous carbene- and nitrene-transfer reactions. Research in this field has since grown dramatically, with emphasis on developing new heme protein variants to increase the scope of biotransformations accessible through these new transfer reactions. This thesis details the expansion of these new-to-nature carbene and nitrene-transfer reactions to include new substrate classes previously unexplored with iron-porphyrin proteins, the use of non-heme metalloproteins for these transformations, and steps toward improving the robustness of the new-to-nature biocatalytic platform. Chapter 1 introduces the steps the field of biocatalysis has taken toward engineering enzymes with new catalytic functions and the process by which these activities are discovered and enhanced. Chapter 2 details the discovery and engineering of heme proteins which catalyze the stereodivergent cyclopropanation of unactivated and electron-deficient alkenes via carbene transfer, expanding the substrate classes beyond styrenyl alkenes. Chapter 3 shows the development of engineered variants of a heme protein (*Rhodothermus marinus* nitric oxide dioxygenase) for the diastereodivergent synthesis of cyclopropanes functionalized with a pinacolborane moiety, enabling product diversification through standard cross-coupling reactions. In Chapter 4, a collection of non-heme metalloproteins is curated, and a non-heme iron enzyme (*Pseudomonas savastanoi* ethylene-forming enzyme) is shown to be both amenable to directed evolution and non-native ligand substitution to enhance its nitrene-transfer activity. Chapter 5 describes the expansion of sequence space targeted for screening in the serine-ligated cytochrome P411 from *Bacillus megaterium* (P411<sub>BM3</sub>) biocatalytic platform to enhance the mutational robustness of these remarkable enzymes. Overall, this work provides a framework for bringing model new-to-nature reactions to their full potential in synthetic biocatalytic reactions.

## PUBLISHED CONTENT AND CONTRIBUTIONS

† denotes equal contribution

1. Hammer, S. C.†; **Knight, A. M.**† Arnold, F. H. Design and Evolution of Enzymes for Non-Natural Chemistry. *Curr. Opin. Green Sustain. Chem.* **2017**, *7*, 23–30. DOI: 10.1016/j.cogsc.2017.06.002.

A.M.K wrote the first draft of the *Introduction* and *Non-natural activities are readily accessed from diverse starting points* sections and participated in the writing and editing of the current opinion.

2. **Knight, A. M.**; Kan, S. B. J.; Lewis, R. D.; Brandenburg, O. F.; Chen, K.; Arnold, F. H. Diverse Engineered Heme Proteins Enable Stereodivergent Cyclopropanation of Unactivated Alkenes. *ACS Cent. Sci.* **2018**, *4*, 372–377. DOI: 10.1021/acscentsci.7b00548. Previously deposited on ChemRxiv, DOI: 10.26434/chemrxiv.5718076.v1.

A.M.K. participated in the conception of the project, discovered initial enzymatic activity, performed the majority of the mutagenesis and screening, synthesized authentic standards, developed analytical methods to determine enantioselectivity, and wrote the manuscript with input from all authors. A.M.K. generated the crystallography constructs, purified, and carried out crystallography experiments. A.M.K. collected data for and refined the tagless *RmaNOD Q52V* crystal structure.

3. Brandenburg, O. F.; Prier, C. K.; Chen, K.; **Knight, A. M.**; Wu, Z.; Arnold, F. H. Stereoselective Enzymatic Synthesis of Heteroatom-Substituted Cyclopropanes. *ACS Catal.* **2018**, *8*, 2629–2634. DOI: 10.1021/acscatal.7b04423.

A.M.K. participated in preparation of P411<sub>BM3</sub> variants for enhanced stereoselectivity on substrate scope targets and participated in characterizing the stereoselectivity of enzymatic products. A.M.K. participated in editing the manuscript.

4. Wittmann, B. J.†; **Knight, A. M.**†; Hofstra, J. L.; Reisman, S. E.; Kan, S. B. J.; Arnold, F. H. Diversity-Oriented Synthesis of Cyclopropane Building Blocks. *ACS Catal.* **2020**, *10*, 7112–7116. DOI: 10.1021/acscatal.0c01888.

A.M.K. and S.B.J.K. conceived the project. A.M.K. prepared compilation plates for the initial activity screen and determined enzymes with initial activity. A.M.K. designed, constructed, and screened the libraries for the *cis*-selective lineage. A.M.K. ran multi-gram scale enzymatic reactions. A.M.K. developed the column-free purification procedure via trifluoroboration methods. A.M.K. and B.J.W wrote the manuscript with input from all authors.

5. Goldberg, N. W.†; **Knight, A. M.**†; Zhang, R. K.; Arnold, F. H. Nitrene Transfer Catalyzed by a Non-Heme Iron Enzyme and Enhanced by Non-Native Small-Molecule Ligands. *J. Am.*

*Chem. Soc.* **2019**, *141*, 19585–19588. DOI: 10.1021/jacs.9b11608. Previously deposited on ChemRxiv, DOI: 10.26434/chemrxiv.10062044.v1.

A.M.K. and N.W.G. ran controls to confirm the activity to be enzymatic. A.M.K. carried out alternative ligand experiments. A.M.K. designed site-saturation mutagenesis libraries. A.M.K. and N.W.G. constructed the libraries, A.M.K. screened and analyzed the libraries, and N.W.G. performed shake-flask validation of potentially improved variants. A.M.K. ran alternative ligand screens, thermostability assays, and time courses for evolved variants. A. M. K. and N. W. G. wrote the manuscript with input from all authors.

## TABLE OF CONTENTS

Acknowledgements .....	iii
Abstract.....	v
Published content and contributions .....	vi
Table of contents .....	viii
List of figures .....	x
List of tables .....	xiii
Abbreviations .....	xiv
1. Design and evolution of enzymes for non-natural chemistry .....	1
Abstract .....	2
1.1 Introduction .....	3
1.2 Non-natural activities are readily accessed from diverse starting points .....	5
1.3 Engineering enzymes for new-to-nature chemistry .....	7
1.4 Engineered proteins offer exquisite control over non-natural catalytic functions .....	12
1.5 Conclusions .....	14
1.6 References for Chapter 1 .....	15
2. Diverse engineered heme proteins enable stereodivergent cyclopropanation of unactivated and electron-deficient alkenes.....	21
Abstract .....	22
2.1 Introduction .....	23
2.1 Results and Discussion .....	28
2.2 Structural characterization of engineered hemoproteins .....	38
2.4 Expanding enzymatic substrate scope to transfer trifluoromethylcarbenes to activated, unactivated, and electron-deficient alkenes.....	48
2.5 Conclusions .....	53
2.6 Supplementary information for Chapter 2 .....	53
2.7 References for Chapter 2 .....	79
3. Diversity-oriented enzymatic synthesis of cyclopropane building blocks .....	87
Abstract .....	88
3.1 Leveraging biocatalysis in the drug development pipeline .....	89
3.2 Initial activity determination and enhancement via directed evolution .....	92
3.3 Derivatization of cyclopropylboronates via cross-coupling reactions .....	98
3.4 Reaction engineering for preparative-scale production of cyclopropylboronates .....	102
3.5 Conclusions .....	106
3.6 Supplementary information for Chapter 3 .....	107
3.7 References for Chapter 3 .....	124
4. Expanding new-to-nature reactions beyond heme proteins .....	128
Abstract .....	129
4.1 Expanding new-to-nature metalloprotein chemistry beyond heme proteins .....	130

4.2 Curating a collection of non-heme metalloproteins for discovering new-to-nature activities .....	132
4.3 Toward enzymatic hydrosilylation with non-heme metalloproteins.....	136
4.4 Nitrene-transfer activity with $\alpha$ KG enzymes .....	146
4.5 Evolution of enzymes with enhanced nitrene-transfer activity .....	153
4.6 Structural characterization of <i>PsEFE</i> variants .....	164
4.7 Adaptation of <i>PsEFE</i> for high-throughput screening .....	169
4.8 Conclusions .....	173
4.9 Supplementary information for Chapter 4. ....	175
4.10 References for Chapter 4 .....	189
5. Increasing robustness of cytochrome P411 <sub>BM3</sub> to facilitate engineering for new-to-nature reactions.....	195
Abstract .....	196
5.1 Effect of protein stability in directed evolution .....	197
5.2 Cytochrome P411 <sub>BM3</sub> as a platform for evolving new-to-nature biocatalysts .....	197
5.3 High-throughput techniques to assay protein fitness .....	198
5.4 Developing a screen to discover stabilizing mutations in P411 <sub>BM3</sub> variants.....	200
5.5 Design of tile-based comprehensive site-saturation mutagenesis libraries .....	204
5.6 Identification and recombination of stabilizing mutations .....	208
5.7 Conclusions .....	211
5.8 Supplementary information for Chapter 5 .....	212
5.9 References for Chapter 5 .....	216
Appendix A: Sequence information .....	220

## LIST OF FIGURES

Figure 1-1. Directed evolution workflow.....	4
Figure 1-2. Discovery of starting points and engineering for non-natural enzymatic function. ....	6
Figure 1-3. New catalytic activation modes engineered in proteins and their putative catalytic intermediates. ....	9
Figure 1-4. Examples of selective syntheses catalyzed by non-natural enzymes. ....	14
Figure 2-1. Cyclopropane-containing pharmaceutical and agrochemical compounds. ....	24
Figure 2-2. Homology models of proteins engineered in the study. ....	30
Figure 2-3. Stereodivergent enzymatic cyclopropanation of an unactivated alkene.....	32
Figure 2-4. Cyclopropanation substrate scope. ....	34
Figure 2-5. Selective, preparative-scale cyclopropane syntheses from various aliphatic alkenes and dienes. ....	36
Figure 2-6. Differences in structure between the prepared homology model and the solved crystal structure. ....	41
Figure 2-7. Homology model of <i>ApePgb</i> based on <i>MacPgb</i> . ....	45
Figure 2-8. Results of random matrix microseeding in <i>ApePgb</i> AGW. ....	46
Figure 2-9. Substrate scope of <i>ApePgb</i> LQ (red) and GW (blue) variants against activated and electron-deficient alkenes. ....	49
Figure 2-10. Protein identity matrix showing the amino-acid sequence identity correlations for the various <i>XxxPgb</i> s. ....	51
Figure 2-11. Synthesis of Weinreb amide-substituted cyclopropanes.....	52
Figure 2-12. Compound nomenclature used in this work. ....	54
Figure 3-1. Drug development pipeline.....	90
Figure 3-2. Expanding cyclopropane diversity through selective installation of derivatizable functional groups. ....	91
Figure 3-3. Proposed enzymatic reaction of vinylboronic acid pinacol ester ( <b>1</b> ) and ethyl diazoacetate ( <b>2</b> ) to form the cyclopropylboronate ester ( <b>3</b> ).....	92
Figure 3-4. Location of residues 32 and 52 in the <i>RmaNOD</i> scaffold, based on the <i>RmaNOD</i> Q52V crystal structure (PDB ID: 6WK3). ....	94
Figure 3-5. <i>RmaNOD</i> scaffold (PDB ID: 6WK3) displaying residues targeted in the <i>cis</i> -selective lineage. ....	95
Figure 3-6. Activity and diastereoselectivity of the <i>cis</i> - and <i>trans</i> - specific lineages for the formation of <b>3</b> . ....	96
Figure 3-7. <i>RmaNOD</i> scaffold (PDB ID: 6WK3) displaying residues targeted in the <i>trans</i> -selective lineage. ....	97
Figure 3-8. Proposed access to all four stereoisomers of derivatized cyclopropane esters. ....	99
Figure 3-9. Derivatizations of <b>3</b> . ....	101
Figure 3-10. Nonlinear response behavior of the cyclopropylboronate compounds. ....	103
Figure 3-11. Representative 96-well plate layout for screening two libraries per plate. ...	109
Figure 4-1. Potential primary coordination sphere plasticity in heme proteins and non-heme metalloproteins.....	131

Figure 4-2. Initial hydrosilylation reaction tests.....	137
Figure 4-3. Copper-binding protein metallation.....	138
Figure 4-4. Pooled substrate sets.. ..	140
Figure 4-5. Final hydrosilylation substrate matrix. ....	141
Figure 4-6. Background reactivity caused by a silicone sealing mat.. ..	145
Figure 4-7. Typical oxene-transfer reactions carried out by non-heme metalloenzymes..	147
Figure 4-8. Nitrene-transfer reaction catalyzed by non-heme iron proteins. ....	147
Figure 4-9. Reactions catalyzed natively by <i>Pseudomonas savastanoi</i> ethylene-forming enzyme ( <i>PsEFE</i> ). ....	148
Figure 4-10. Aziridination of unactivated alkenes. ....	150
Figure 4-11. Attempted carbene transfer reactions.. ..	151
Figure 4-12. $\alpha$ KG and analogs tested for their effect on <i>PsEFE</i> aziridination activity.. ...	152
Figure 4-13. Amino-acid residues targeted for site-saturation mutagenesis in the first round of <i>PsEFE</i> engineering. ....	154
Figure 4-14. Amino-acid residues targeted for site-saturation mutagenesis and screened via LC-MS in the first round of <i>PsEFE</i> engineering. ....	155
Figure 4-15. Aziridination activity and selectivity of <i>PsEFE</i> wild type and engineered variants. ....	156
Figure 4-16. Amino-acid residues targeted in the second round of mutagenesis.....	157
Figure 4-17. <i>PsEFE</i> recombination library primer design.....	158
Figure 4-18. Position of R277 relative to $\alpha$ KG. ....	160
Figure 4-19. Aziridination activity with <i>PsEFE</i> MLHMM under different reaction conditions. ....	160
Figure 4-20. Intramolecular C–H insertion reaction.. ..	161
Figure 4-21. Intramolecular C–H insertion with variants <i>PsEFE</i> VMM and <i>PsEFE</i> VHMM.....	162
Figure 4-22. Aziridination time course.....	163
Figure 4-23. Thermostability of wild-type and evolved <i>PsEFE</i> variants for aziridination and intramolecular C-H insertion. ....	164
Figure 4-24. <i>PsEFE</i> VMM crystals. ....	166
Figure 4-25. Structural comparison of Mn- and NOG-bound <i>PsEFE</i> VMM and Mn- and $\alpha$ KG-bound <i>PsEFE</i> WT (PDB ID: 5V2Y). ....	167
Figure 4-26. Electron density of NOG-bound <i>PsEFE</i> VMM matches NOG in two conformations. ....	168
Figure 4-27. <i>PsEFE</i> VMM 127-site-saturation library. ....	171
Figure 4-28. Split-fluorescent protein labeling enables measurement of <i>PsEFE</i> folding and aggregation. ....	173
Figure 4-29. Expression tests of the non-heme metalloprotein collection. ....	179
Figure 4-30. SDS-PAGE showing purification of the cleavable <i>N</i> -His <sub>6</sub> -SUMO- <i>PsEFE</i> variant constructs. ....	187
Figure 4-31. C-H insertion activity for <i>PsEFE</i> VHMM and <i>PsEFE</i> VMM with glycolic acid as an alternative ligand. ....	188
Figure 4-32 SDS-PAGE of <i>PsEFE</i> aziridination variant T97M R171L R277H F314M C317M ( <i>PsEFE</i> MLHMM). ....	189

Figure 5-1. Colony-based detection of P411 <sub>BM3</sub> expression via split-fluorescent protein tagging.....	202
Figure 5-2. Supernate fluorescence of P450 <sub>BM3</sub> -C10 tagged with mNG2(11) and co-expressed with the complementary mNG2(1-10) fragment.....	203
Figure 5-3. Overview of sites targeted in cytochrome P450 <sub>BM3</sub> heme and FMN domains (PDB ID: 1BVY).....	206
Figure 5-4. Example tile in P411 <sub>BM3</sub> -C10. ....	207
Figure 5-5. Potential recombination methods to generate combinatorial diversity from beneficial mutations found in tile libraries.....	210
Figure 5-6. Agarose gel of PCR-amplified tile regions of P411 <sub>BM3</sub> .....	213

## LIST OF TABLES

Table 2-1. Literature precedent for stereocomplementary cyclopropanation of styrenyl alkenes via enzymatic carbene transfer.....	25
Table 2-2. State-of-the-art methods for the intermolecular cyclopropanation of unactivated alkenes prior to our report of the stereodivergent cyclopropanation of unactivated alkenes with heme proteins. ....	25
Table 2-3. GC yields for analytical-scale reactions reported in Figure 2-4. ....	33
Table 2-4. Activity and selectivity of final protein variants as whole-cell, sonicated cell lysate, and purified protein for <b>3a</b> product formation.....	37
Table 2-5. List of protoglobin proteins selected for the transfer of mutation experiment. ..	51
Table 2-6. Activity and diastereoselectivity of <i>TarPgb</i> and <i>TdaPgb</i> wild type and variants in the synthesis of N-methoxy-N-methyl-2-(trifluoromethyl)cyclopropane-1-carboxamide. ....	52
Table 2-7. Heme-binding proteins tested for unactivated alkene cyclopropanation activity using 1,7-octadiene and EDA as substrates. ....	55
Table 2-8. Activity and selectivity of hemin controls, wild-type protein, and engineering lineage intermediate proteins in <b>3a</b> product formation. ....	60
Table 2-9. Activity and selectivity of wild-type and engineering lineage intermediate proteins in <b>3b</b> product formation.....	60
Table 2-10. Substrate scope and diastereoselectivity of the four final variants. ....	61
Table 2-11. Conditions for the preparative-scale reactions reported in the chapter.....	62
Table 2-12. Crystal data and refinement statistics for <i>RmaNOD Q52V</i> . ....	65
Table 2-13. P450 <sub>BM3</sub> “superscreen” crystallographic conditions. ....	66
Table 2-14. Chromatographic separation conditions. ....	77
Table 2-15. Information on engineered proteins in the chapter.....	79
Table 3-1. Protein scaffolds tested for formation of <b>3</b> in the initial-activity screening. ....	113
Table 3-2. Activity of top variants from initial screening.....	114
Table 3-3. Achiral separation conditions for compound <b>3</b> .....	123
Table 3-4. Chiral separation conditions.....	123
Table 4-1. Activities of $\alpha$ -KG-dependent iron enzymes toward aziridination to form <b>3</b> ..	148
Table 4-2. Control and ligand-substitution reactions for <i>PsEFE</i> WT aziridination to form <b>3</b> .....	149
Table 4-3. Intramolecular C–H insertion activity with ligand substitutions in <i>PsEFE</i> VMM.....	162
Table 4-4. Collection of non-heme metalloproteins acquired as synthetic gene fragments from IDT DNA or as clonal constructs from other research groups. ....	175
Table 4-5. Collection of non-heme metalloproteins acquired as clonal constructs from Twist Biosciences. ....	176
Table 5-1. P411 <sub>BM3</sub> -C10 tile locations.....	215
Table 5-2. Oligonucleotide sequences to amplify the 12 tile-based mutagenesis libraries for P411 <sub>BM3</sub> -C10. ....	216
Table A-1. Proteins (and expression vectors) for which sequences are given in the worksheet deposited to CaltechDATA (DOI: 10.22002/D1.1437).....	220

## ABBREVIATIONS

5-ALA	5-aminolevulinic acid
Å	Ångstrom
amp	ampicillin
<i>ApePgb</i>	protoglobin from <i>Aeropyrum pernix</i>
Ar	aryl group
Bpin	pinacolborane
carb	carbenicillin
chlor	chloramphenicol
CO	carbon monoxide
cyt <i>c</i>	cytochrome <i>c</i>
dr	diastereomeric ratio
DTFE	2-diazo-1,1,1-trifluoroethane
EDA	ethyl diazoacetate
ee	enantiomeric excess
EtOAc	Ethyl acetate
EtOH	ethanol
FID	flame ionization detection
FMN	flavin mononucleotide
FPLC	fast protein liquid chromatography
GC	gas chromatography
GC-MS	gas chromatography with mass spectrometry
HB	Hyperbroth medium (AthenaES)
HPLC	high performance liquid chromatography
HRMS	high-resolution mass spectrometry
IPC	iron-porphyrin carbene
IPTG	isopropyl $\beta$ -D-1-thiogalactopyranoside
LB	Lysogeny Broth (Luria-Bertani medium)
LC-MS	liquid chromatography with mass spectrometry
M9-N	M9 minimal medium with no nitrogen added
Me	methyl
Me-EDA	ethyl 2-diazopropanoate (methyl-EDA)
mNG2	monomeric Neon Green 2 fluorescent protein
mNG2(1-10)	mNG2 split-fluorescent protein fragment containing $\beta$ -strands 1-10
mNG2(11)	mNG2 split-fluorescent protein fragment containing $\beta$ -strand 11
MOPS	3-(N-morpholino)propanesulfonic acid
NOG	N-oxalylglycine
OD <sub>600</sub>	optical density at 600 nm
P411 <sub>BM3</sub>	cytochrome P411 (P450 <sub>BM3</sub> containing C400S mutation)
P450 <sub>BM3</sub>	cytochrome P450 from <i>Bacillus megaterium</i>
PCR	polymerase chain reaction

PDB	Protein Data Bank
<i>Rma</i>	<i>Rhodothermus marinus</i>
<i>Rma</i> NOD	nitric oxide dioxygenase from <i>Rhodothermus marinus</i>
RT	room temperature
SOC	Super Optimal broth with Catabolite repression
TB	Terrific Broth
TM	Melting temperature
Ts	para-toluenesulfonyl (tosyl)
TTN	total turnover number
WT	wild type
$\alpha$ KG	$\alpha$ -ketoglutarate

## DESIGN AND EVOLUTION OF ENZYMES FOR NON-NATURAL CHEMISTRY

Content in this chapter is adapted from published work:

† denotes equal contribution

Hammer, S.C.<sup>†</sup>; **Knight, A.M.**<sup>†</sup>; Arnold, F.H. Design and Evolution of Enzymes for Non-Natural Chemistry. *Curr. Opin. Green Sustain. Chem.* **2017**, 7, 23–30.

A.M.K wrote the first draft of the *Introduction* and *Non-natural activities are readily accessed from diverse starting points* sections and participated in the writing and editing of the current opinion.

## **Abstract**

Enzymes are used in biocatalytic processes for the efficient and sustainable production of pharmaceuticals, fragrances, fine chemicals, and other products. Most bioprocesses exploit chemistry found in nature, but we have now entered a realm of biocatalysis that goes well beyond natural reactions. Enzymes have been engineered to catalyze reactions previously only accessible with synthetic catalysts. Because they can be tuned by directed evolution, many of these new biocatalysts have been shown to perform abiological reactions with high activity and selectivity. This chapter discusses examples which showcase catalyst improvements achieved using directed evolution, and commentary on current and future implications of non-natural enzyme evolution for sustainable chemical synthesis.

## 1.1 Introduction

Chemical synthesis is integral to modern life: from the Haber-Bosch process used to fix nitrogen for fertilizer and feed the Earth's growing population to ubiquitous synthetic fabrics to life-saving medicines, we are surrounded by the products of chemical transformations. Many synthetic routes, especially those for more complex molecules, are far from ideal: poor atom economy, harsh reaction conditions, and multiple steps that require purification of intermediates all lead to lower yield and increased waste. Rare metal catalysts are increasingly costly to obtain, and their mining and refinement has significant and negative environmental impacts.<sup>1</sup> The recovery or removal of toxic catalysts from the products can also be a challenge. Thus both economic and environmental factors favor chemistry that is inherently "green."<sup>2,3</sup> Discussions of "green" chemistry often include biocatalysis, where enzymes stand out for their ability to selectively convert inexpensive starting materials into complex molecules under mild aqueous conditions. Enzymes have been used for organic synthesis for well over a century.<sup>4</sup> However, a biocatalytic process was not possible if the desired enzyme could not be found in nature and made available at low cost. This has changed with the development of recombinant protein production technology, which has delivered many biocatalysts at low cost. Furthermore, enzymes have been engineered, at the level of their amino acid sequences, for improved performance under process conditions. The growing need for sustainable synthetic methods, coupled with advances in enzyme engineering and related process technologies, have helped drive adoption of biocatalytic processes in the pharmaceutical and chemical industries.<sup>5,6</sup>

A key innovation that underlies growing interest in biocatalysis is directed evolution, a protein engineering method that mimics Darwinian evolution, using iterations of genetic variation and screening in order to accumulate performance-enhancing mutations (Figure 1-1). Researchers have used directed evolution to increase enzyme performance under reaction conditions, including higher temperatures, high acidity or basicity, and high concentrations of organic solvents. The ranges of substrates that enzymes accept have been expanded, and product selectivity has been altered and optimized.<sup>7</sup> Rather than having to design a process to fit an enzyme's requirements, directed evolution has made it possible to fit the catalyst to

optimal manufacturing conditions.<sup>6</sup> A beautiful example of this ability to tailor a biocatalyst to fit the process is Merck's engineered transaminase that catalyzes a key step in the production of the antidiabetic drug sitagliptin. Through directed evolution, a transaminase was engineered for near-perfect enantioselectivity and high yields in 50% DMSO, 200 g L<sup>-1</sup> starting material, and elevated temperatures.<sup>8</sup> Replacing the precious metal catalyst previously used for multi-ton-scale synthesis with the engineered enzyme dramatically reduced solvent waste and increased process yield. Many other industrial syntheses make use of engineered enzymes with enhanced stability, activity, and selectivity.<sup>6,9-13</sup>

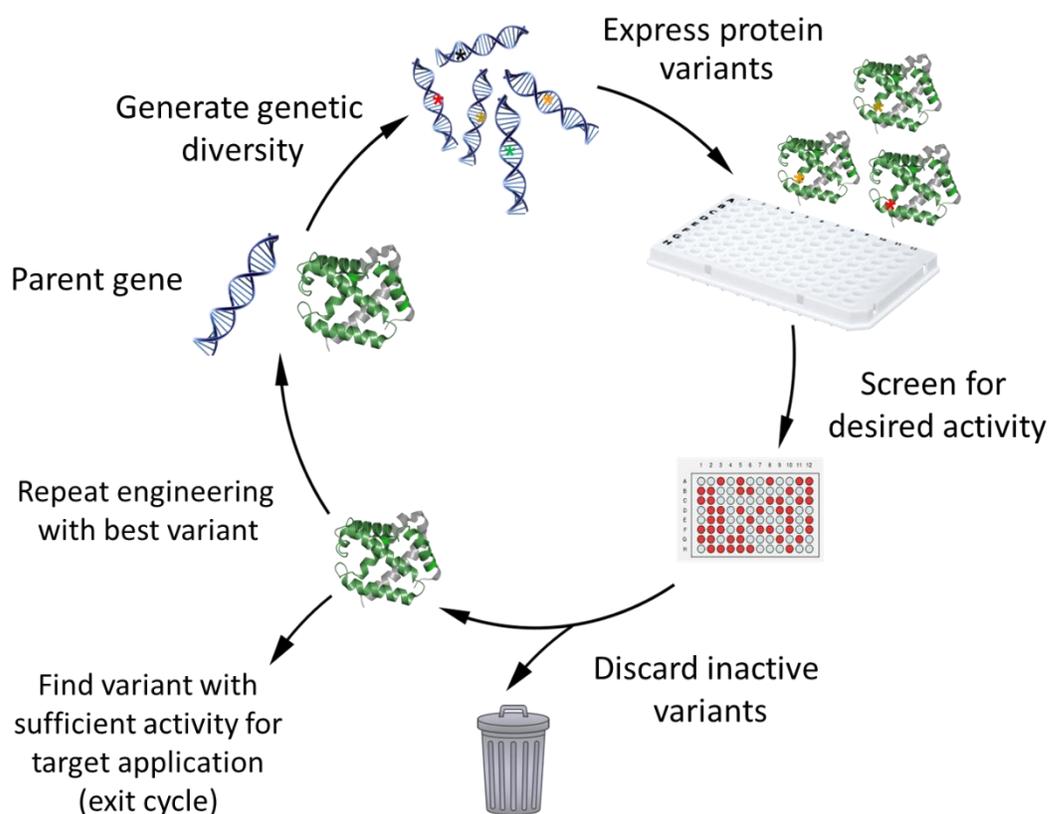


Figure 1-1. Directed evolution workflow. Variants of the gene of interest are generated and the corresponding protein variants are expressed in (typically bacterial) cells. The proteins are screened for the desired activity. Proteins with enhanced activity are used as the parent for subsequent rounds until a variant with sufficient activity is found, at which point the directed evolution cycle is exited.

An exciting development is engineering that goes well beyond enhancing native enzyme activities or broadening the substrates accepted: enzymes have been engineered for new chemical transformations not known in the biological world, and new catalytic capabilities have even been engineered into proteins that do not have any native catalytic activity. These advances make it possible to take the concept of engineering a protein to fit a process to the next level, where biochemical syntheses are no longer limited to reaction classes known in nature, but can expand to include whole new transformations and entirely new classes of products. In this Chapter, I discuss some of the non-natural activities that have been engineered into proteins, the ease with which such activities have been enhanced by directed evolution, and the current and future implications of such non-natural enzymes for green chemistry.

## **1.2 Non-natural activities are readily accessed from diverse starting points**

Directed evolution of a non-natural activity requires a starting point – it can enhance an existing activity, but cannot create it entirely from scratch (Figure 1-2). Although enzymes are often thought to catalyze reactions with great selectivity, many enzymes are in fact catalytically promiscuous – that is, they are able to catalyze reactions other than their native, biologically relevant reactions.<sup>6,14,15</sup> It has been recognized for some time that these promiscuous activities are good starting points for directed evolution.<sup>16</sup> In the initial report of enzymatic cyclopropanation, for example, many of the heme proteins tested – including wild-type proteins and previously engineered cytochrome P450 variants – catalyzed non-natural styrene cyclopropanation at some (low) level.<sup>17</sup> Different protein scaffolds could serve as starting points for directed evolution as has been observed for other reactions<sup>18–20</sup> and in a broader survey of P450s from nature.<sup>21</sup> Starting points are not even limited to enzymes: heme-binding proteins with no native enzymatic functions also exhibit non-natural catalytic activities that could be improved by directed evolution. Fasan and coworkers showed that sperm whale myoglobin, an oxygen-binding heme protein, could be engineered to catalyze the cyclopropanation of styrene stereoselectively with 46,800 turnovers,<sup>18</sup> whereas biocatalytic carbon-silicon bond formation was demonstrated with an engineered

cytochrome *c*.<sup>20</sup> In each case, engineering active and selective biocatalysts required just a few rounds of directed evolution. There is no evidence that these non-natural biocatalysts have reached their full potential either: further evolution could continue to increase these activities, for which the upper limits are unknown.

An initial activity can sometimes be found by screening wild-type and/or engineered proteins. Another approach is to install an artificial cofactor<sup>22</sup> to generate activity that can be improved by directed evolution,<sup>23–27</sup> albeit with low throughput. Although directed evolution with *in vitro* assembly of an artificial metalloenzyme is more work than with fully genetically encoded catalysts, the evolution of an artificial metalloenzyme as shown by Hartwig and coworkers<sup>23</sup> is testament to the efficiency and reliability of this catalyst improvement method.

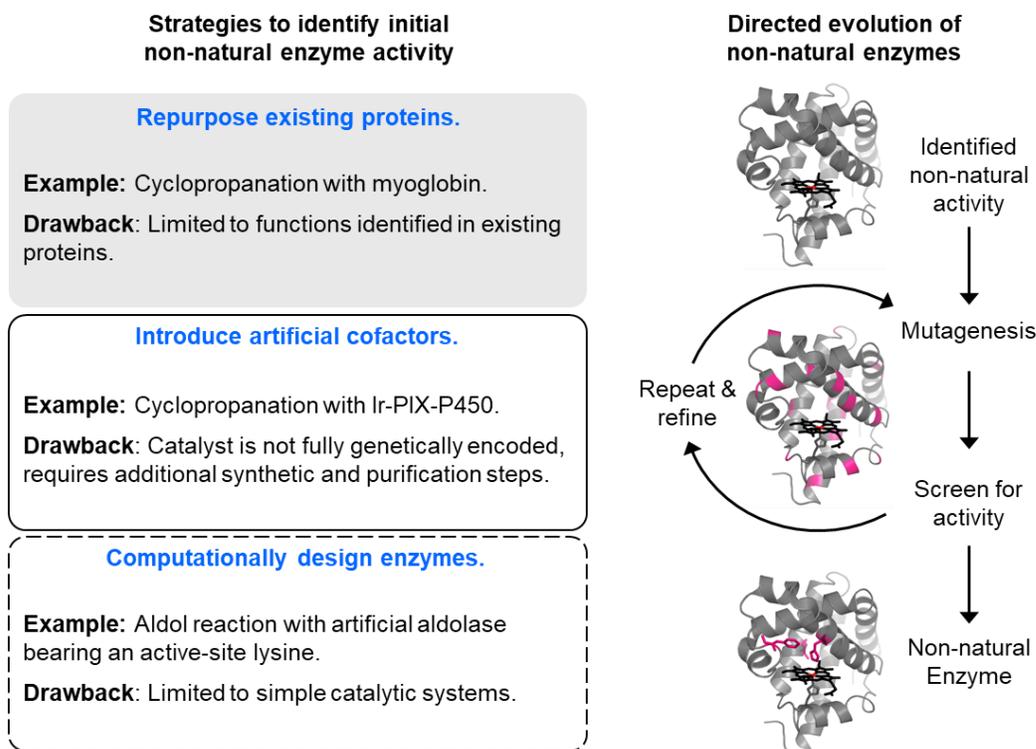


Figure 1-2. Discovery of starting points and engineering for non-natural enzymatic function. Left: Different strategies for finding initial enzymatic activity. Right: proteins with initial activity are engineered through an iterative process of mutagenesis and screening (directed evolution) to accumulate mutations that enhance the non-natural function.

For some model reactions, one can even rationally design a protein with initial non-natural enzymatic activity.<sup>28</sup> A recent example of pairing a rationally designed starting point with directed evolution comes from the Baker and Hilvert labs, who designed an active site into a natural TIM-barrel protein to give it trace activity for retroaldol cleavage.<sup>29</sup> The Hilvert lab then used high-throughput directed evolution with detection of the fluorescent product to obtain an enzyme with reaction kinetics comparable to native aldolases.<sup>30,31</sup> The computational design of non-natural enzymes is currently limited to relatively simple reactions. Should rational design extend to more difficult and especially to cofactor-dependent modes of catalysis, this approach could provide valuable new starting points for future biocatalysts.

The ease with which the non-natural activities were improved is a common theme across these new biocatalysts. Just a few rounds of directed evolution (each taking 1–4 weeks) increased the activities and/or selectivities well beyond the starting catalysts and even the activities of their small-molecule catalyst analogs.<sup>32</sup> Given measurable starting activity and efficient analytical methods (e.g. spectroscopic or chromatographic analyses), a protein's catalytic activity can be improved with no additional knowledge of the catalyst's mechanism or structure beyond the DNA sequence encoding the protein. This engineering of non-natural activity proceeds using the same methods and with the same efficacy as the directed evolution of native activities. In fact, given that nature never optimized these proteins for such non-natural catalytic activities, there is plenty of room for improvement and activating mutations are relatively easy to find.

### **1.3 Engineering enzymes for new-to-nature chemistry**

Many of the concepts underlying creation of non-natural enzymes, such as exploiting natural catalytic promiscuity,<sup>14,15,33</sup> chemomimetic biocatalysis,<sup>34,35</sup> computational enzyme design,<sup>28</sup> or introduction of artificial cofactors,<sup>22</sup> have been part of the biocatalysis community for a decade or more. These approaches have enabled researchers to identify (usually low levels of) enzymatic activity for numerous new-to-nature chemical transformations. What has radically advanced the field in recent years is the combination of the discovery of new

enzymatic activities with directed evolution to enhance and tune those activities and selectivities.<sup>16,36,37</sup>

Multiple important modes of catalysis have been implemented in protein scaffolds by exploiting the reactivities of both canonical<sup>30,38,39</sup> and non-canonical amino acids,<sup>40</sup> as well as natural<sup>35</sup> and artificial<sup>22,41</sup> cofactors. Hydrolases have been particularly popular enzymes to explore for new chemistry due to their availability and the versatility of their hydrogen bonding activation modes and acyl-transfer activity.<sup>38</sup> More than 20 years ago, visionary contributions showed that the acyl-enzyme intermediate can be trapped with various nucleophiles such as amines.<sup>42</sup> Subsequent optimization of the acyl donor improved this non-natural activity<sup>43</sup> and enabled chiral synthesis of various amines on a multi-ton scale.<sup>44</sup> Similarly, hydrogen bonding in a halohydrin dehalogenase was harnessed for selective epoxide opening with cyanide. Directed evolution of a weakly active wild-type enzyme yielded a biocatalyst that was stable and highly efficient under process conditions.<sup>45</sup>

Various activation modes from organocatalysis<sup>46</sup> have been explored in proteins.<sup>30,39,47–50</sup> Inspired by aminocatalysis, active-site proline and lysine residues were recently exploited for biocatalytic C–C bond formations: Michael(-type) additions,<sup>47,49</sup> aldol reactions,<sup>30</sup> and Knoevenagel condensation reactions.<sup>50</sup> In an outstanding report, Hilvert and coworkers demonstrated that directed evolution can convert proof-of-concept low level activity into an impressive biocatalyst.<sup>30</sup> An artificial aldolase (class I aldolase mimic) performing enamine-catalyzed aldol reactions via an active-site lysine was evolved to perform aldol reactions with high yields, high enantioselectivities, and up to 28,500 total turnovers under preparative-scale conditions (Figure 1-3a). In another example, the protonation machinery of a squalene hopene cyclase was harnessed for Brønsted acid-catalyzed activation of carbonyls, epoxides, and alkenes at neutral pH in water (Figure 1-3b). Several valuable cyclohexanoids have been synthesized with very high stereoselectivity, and active-site engineering improved the activity and selectivity of the biocatalyst. Further enzyme engineering might provide a broad biocatalytic platform for Brønsted acid catalysis.<sup>39</sup>

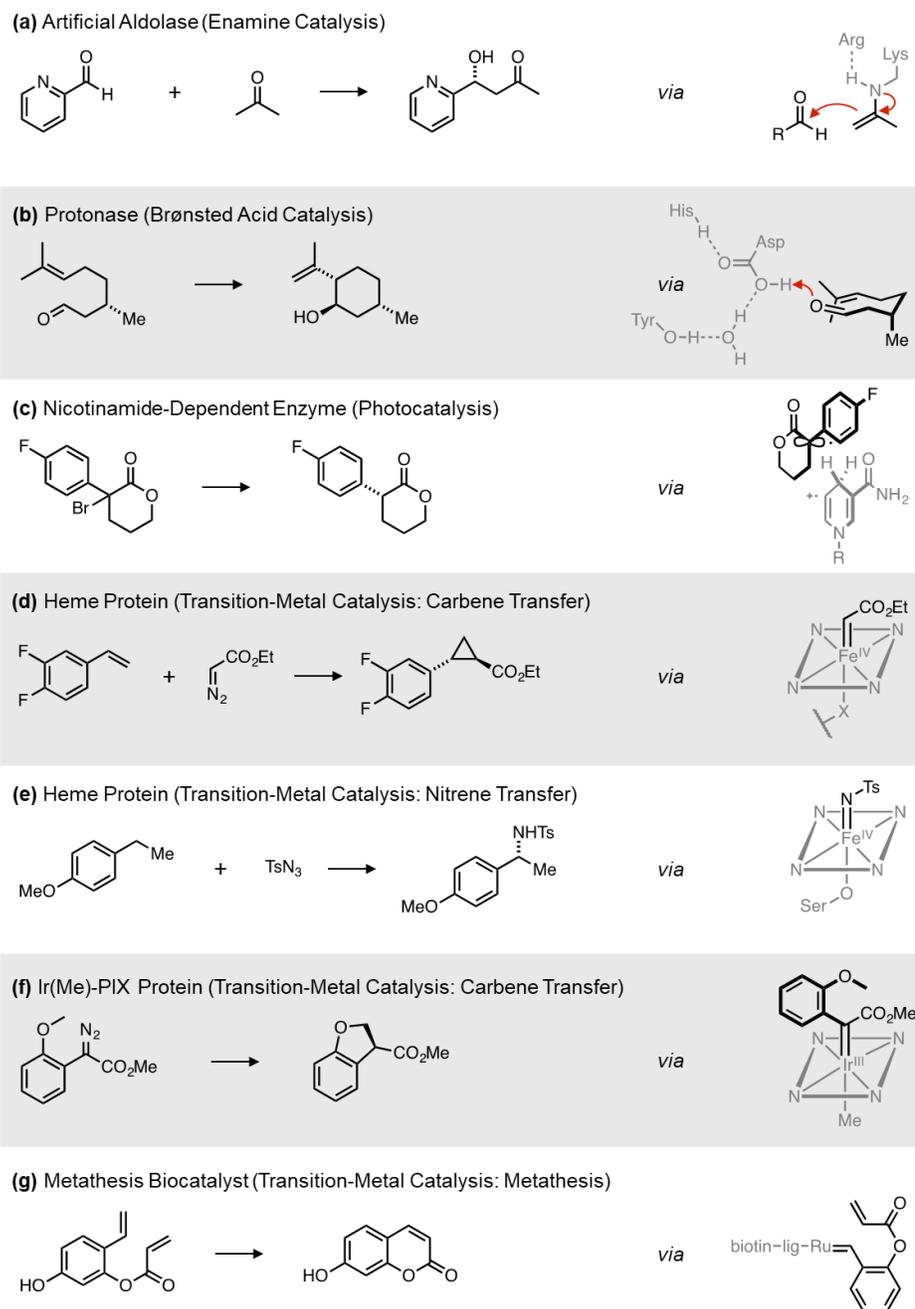


Figure 1-3. New catalytic activation modes engineered in proteins and their putative catalytic intermediates. Examples include **(a)** enamine-catalyzed aldol reaction, **(b)** Brønsted acid-catalyzed Prins/Ene-reaction, **(c)** photo-catalyzed radical dehalogenation of halolactones, **(d)** heme-catalyzed cyclopropanation of olefins, **(e)** heme-catalyzed intermolecular C-H amination, **(f)** Ir(Me)-protoporphyrin IX-catalyzed insertion of carbenes into C-H bonds and **(g)** olefin metathesis using a Hoveyda-Grubbs second-generation catalyst embedded in a protein.

Hyster and coworkers shone a whole new light on non-natural biocatalysis by using photoexcitation of nicotinamide-dependent enzymes in enantioselective radical dehalogenations of halolactones (Figure 1-3c).<sup>51</sup> The nicotinamide cofactor of ketoreductases, whose native function is carbonyl reduction, was utilized as photoreductant and hydrogen source. Cofactor photoexcitation leads to electron transfer to the substrate which forms a prochiral radical intermediate upon dehalogenation. The intermediate generates a chiral lactone upon hydrogen-atom transfer from the photoexcited cofactor. Accessing radical intermediates via photoredox processes within an evolvable protein scaffold provides a new level of control over challenging radical-mediated reactions, as has been shown in novel “ene” reductase-photoredox catalyst systems for cyclization<sup>52</sup> and hydrogenation<sup>53</sup> reactions.

New-to-nature biocatalysis involving transition metals has flourished in the years following the seminal contribution of Coelho *et al.*,<sup>17</sup> who demonstrated that a heme protein can be engineered to generate and selectively transfer a reactive carbene intermediate to an olefin to make cyclopropanes. Efficient carbene-transfer enzymes have been evolved in the laboratory with astonishing ease.<sup>32</sup> Since 2013, the scope of non-natural reactions of heme proteins has expanded dramatically to include an array of carbene- and nitrene-transfer reactions, such as aldehyde olefination, Doyle-Kirmse reaction, olefin aziridination, and C–H amination.<sup>32</sup> Further work has demonstrated that enzyme-catalyzed carbene-transfer reactions are not limited to  $\alpha$ -diazoester carbene precursors: Fasan and coworkers described synthesis of trifluoromethyl-substituted cyclopropanes by selectively transferring trifluoromethylcarbene to olefins,<sup>54</sup> and the Arnold group has reported the use of  $\alpha$ -trifluoromethylated diazo compounds in biocatalytic organoborane<sup>55,56</sup> and fluoroalkane<sup>57</sup> syntheses. Biocatalytic carbene-transfer chemistry has been applied in the synthesis of several important cyclopropane drug precursors, including ticagrelor (Figure 1-3d),<sup>19,58</sup> levomilnacipran,<sup>59</sup> and grazoprevir.<sup>60</sup> In addition to cyclopropanation, engineered cytochromes P411 have been shown to perform truly remarkable cyclopropanation and bicyclobutanation reactions;<sup>61</sup> these highly strained carbocycles were first reported to be accessible from a wide range of

terminal alkynes and, with further protein engineering, P411<sub>BM3</sub> variants could also cyclopropanate internal alkynes.<sup>62</sup> Direct alkylation of *sp*<sup>3</sup> C–H bonds is limited primarily to methylation in nature, and biocatalytic approaches have required the use of enzymes binding noble-metal-substituted porphyrins.<sup>24</sup> Arnold and coworkers reported an iron-porphyrin protein that is up for the challenge: through 13 rounds of directed evolution, a P411<sub>BM3</sub> variant was improved from a starting point of just 13 TTN for the *sp*<sup>3</sup> C–H alkylation of *p*-methoxybenzyl methyl ether to a biocatalyst capable of C–H functionalization of a broad scope of benzylic, allylic, and  $\alpha$ -amino C–H bonds.<sup>63</sup>

The first enzyme-catalyzed carbon–silicon bond formation was realized using a heme protein for carbene insertion into Si–H bonds.<sup>20</sup> Achieved by directed evolution of a cytochrome *c*, an electron-transfer protein with no natural enzymatic activity, this novel reactivity could be the first step toward biological production of organosilicon compounds. Nitrene-transfer chemistry has also enjoyed significant success: Prier *et al.* reported the first example of enzymatic intermolecular C–H amination, which they accomplished using an engineered cytochrome P450.<sup>64</sup> Efficient and highly enantioselective intermolecular amination of *sp*<sup>3</sup>-hybridized C–H bonds has long been a challenge in catalysis. Using directed evolution, a biocatalyst for the amination of benzylic C–H bonds could be engineered to catalyze up to 1,300 turnovers with excellent enantioselectivities (>99% *ee*) (Figure 1-3e). The protein framework tuned the earth-abundant iron to catalyze this reaction, which contrasts with the C–H functionalization field’s heavy reliance on precious metals. While these “new-to-nature” reactions had not been found in natural metabolism at the time of this report, Nature was already using nitrene chemistry: shortly after this nitrene-transfer reaction was reported, researchers studying the benzastatin biosynthesis pathway determined that a cytochrome P450, BezE, formed an aziridine via an iron-nitrenoid intermediate.<sup>65</sup>

Contributions from Hartwig, Clark, and coworkers have further merged transition-metal and enzyme catalysis by replacing native iron-porphyrin cofactors with a synthetic iridium porphyrin to access carbene- and nitrene-transfer activities. Directed evolution of iridium-substituted myoglobin and P450 yielded enzymes with impressive catalytic performance for

non-natural reactions, such as intra- and intermolecular carbene insertion into C–H bonds (Figure 1-3f),<sup>23,24</sup> highly chemoselective, intramolecular C–H bond amination,<sup>25</sup> and cyclopropanation of various internal vinylarenes and unactivated alkenes.<sup>26</sup> Lewis and coworkers took a different approach to obtaining non-natural reactivities by covalently linking an alkyne-substituted dirhodium catalyst with a protein scaffold containing a genetically encoded azidophenylalanine. Mutations improved enantioselectivity and reduced side reactions in the cyclopropanation of styrenes using donor-acceptor carbene precursors.<sup>27</sup> These artificial metalloenzyme systems provide access to activities not yet reported in genetically encoded biocatalysts. But there are drawbacks to this approach. The artificial cofactor must be synthesized separately, and the *in vitro* catalyst assembly initially required protein purification, lowering the throughput of protein engineering and raising the cost of the prepared biocatalyst. Fasan and Brustad have shown methods for *in vivo* incorporation of non-native metalloporphyrins into heme proteins, improving the non-natural biocatalyst preparation in ways that could be translated to other processes requiring cofactor substitutions.<sup>66,67</sup> The Ward group solved the issue of low-throughput biocatalyst preparation through *in vivo* assembly of an artificial, ruthenium-based olefin metathesis biocatalyst. This artificial metalloenzyme, containing a biotinylated Hoveyda-Grubbs second-generation catalyst, was assembled in the cell's periplasm, where the concentration of metathesis inhibitors is reduced.<sup>68</sup> The ability to assemble this artificial metalloenzyme *in vivo* enabled the use of directed evolution to improve its activity (Figure 1-3g). In a complementary approach, the Lewis group developed and optimized a high-throughput screen of a dirhodium catalyst bioconjugated to a protein scaffold in heat-treated cell lysate, enabling combinatorial and random mutagenesis approaches.<sup>69</sup>

## **1.4 Engineered proteins offer exquisite control over non-natural catalytic functions**

Not only can these novel, engineered enzymes catalyze reactions that previously only synthetic catalysts could perform, they can even outcompete their small-molecule counterparts in terms of efficiency and selectivity because they can be tuned by directed

evolution. Near-perfect stereoselectivities have been reported with non-natural biocatalysis, such as in the enantiodivergent cyclopropanation of unactivated alkenes (Figure 1-4a).<sup>26</sup> The macromolecular protein structure further enables outstanding control over product selectivity that is very difficult to attain with other catalysts. This is demonstrated by the challenging catalyst-controlled, regiodivergent C–H amination achieved using heme proteins (Figure 1-4b)<sup>70</sup> and the remarkable control of carbene insertion into an Si–H bond rather than an N–H bond in the same substrate (Figure 1-4c).<sup>20</sup> The protein matrix can also dramatically change the reactivity of a metal catalyst, with profound consequences for catalytic function.<sup>71</sup> As shown in enzymatic nitrene-transfer reactions,<sup>64</sup> less reactive, earth-abundant iron can be activated within a heme protein to outperform catalysts that rely on precious metals (Figure 1-3). In this context, substituting iron in natural metalloproteins with noble metals such as iridium<sup>23–26,72</sup> is an exciting strategy to access new chemical space in biology, but is not required in all cases; further reports have shown that engineered enzymes binding iron porphyrin can catalyze challenging new-to-nature reactions<sup>63,73</sup> previously only reported with noble-metal substituted porphyrins.<sup>41</sup> Enzymes control substrate conformations with remarkable precision and offer multiple secondary interaction sites to stabilize and guide high-energy intermediates during complex catalytic capabilities.<sup>74,75</sup> We expect that this control can be used to enable whole new catalytic capabilities.

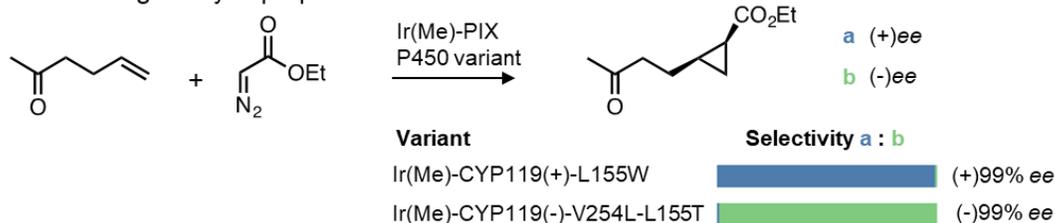
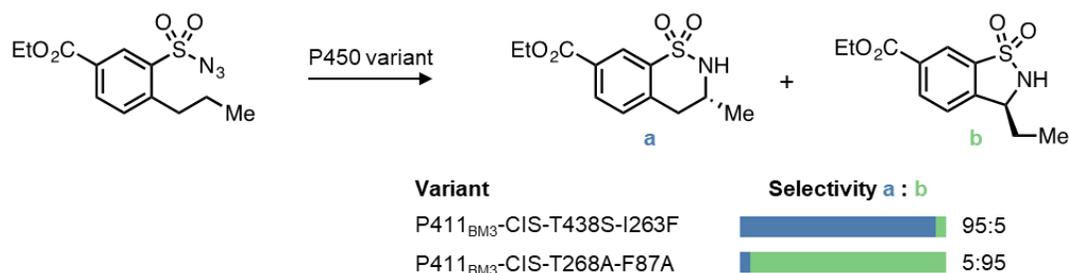
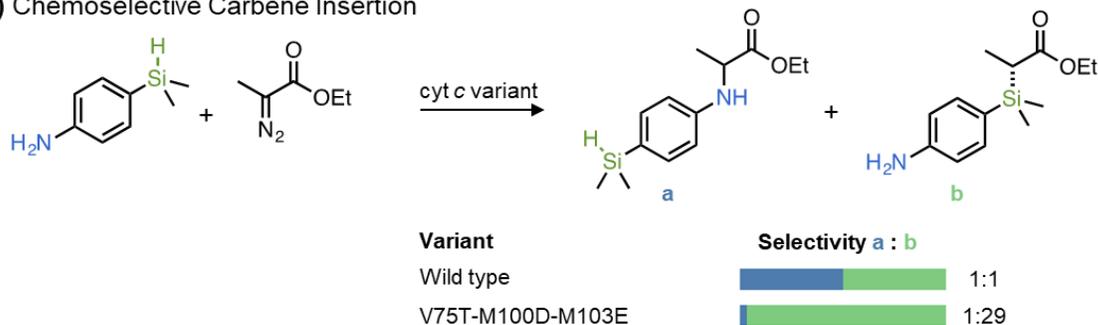
**(a) Enantiodivergent Cyclopropanation****(b) Regiodivergent C-H Amination****(c) Chemoselective Carbene Insertion**

Figure 1-4. Examples of selective syntheses catalyzed by non-natural enzymes. **(a)** Enantiodivergent cyclopropanation of unactivated alkenes. **(b)** Engineered P450s for challenging catalyst-controlled, regiodivergent C–H amination. **(c)** Wild-type and engineered cytochrome *c* for catalyst-controlled chemoselective carbene Si–H insertion over N–H insertion.

## 1.5 Conclusions

Whether the starting point is a natural enzyme, a designed enzyme, a protein without a native catalytic function, or a protein-bound artificial cofactor, directed evolution has been shown to improve activity and selectivity of non-natural catalytic function. Because we are just beginning to explore new functional spaces for biocatalysis, we do not yet know what starting

points will lead to the most efficient enzymes, or even whether all poor enzymes can become good ones. What is certain, however, is that the field has entered a vast goldmine of new biocatalytic possibilities, and that chemical creativity will play a major role in future discovery and exploitation of these remarkable biological resources. Furthermore, advances in directed evolution methods will continue to improve the ease of engineering these new capabilities. We expect that this new era of non-natural biocatalysis will highlight the capabilities of enzymes to activate earth-abundant metals and catalyze new-to-nature transformations under mild conditions. Biocatalysis will offer competitive, sustainable routes to making the molecules and materials we need.

## 1.6 References for Chapter 1

- (1) Nuss, P.; Eckelman, M. J. Life Cycle Assessment of Metals: A Scientific Synthesis. *PLoS One* **2014**, *9*, e101298. <https://doi.org/10.1371/journal.pone.0101298>.
- (2) Li, C.-J.; Trost, B. M. Green Chemistry for Chemical Synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 13197–13202. <https://doi.org/10.1073/pnas.0804348105>.
- (3) Tucker, J. L.; Faul, M. M. Industrial Research: Drug Companies Must Adopt Green Chemistry. *Nature* **2016**, *534*, 27–29. <https://doi.org/10.1038/534027a>.
- (4) Rosenthaler, L. Durch Enzyme Bewirkte Asymmetrische Synthese. *Biochem. Z.* **1908**, *14*, 238–253.
- (5) Wohlgemuth, R. Biocatalysis—Key to Sustainable Industrial Chemistry. *Curr. Opin. Biotechnol.* **2010**, *21*, 713–724. <https://doi.org/10.1016/j.copbio.2010.09.016>.
- (6) Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. Engineering the Third Wave of Biocatalysis. *Nature* **2012**, *485*, 185–194. <https://doi.org/10.1038/nature11117>.
- (7) Reetz, M. T. Biocatalysis in Organic Chemistry and Biotechnology: Past, Present, and Future. *J. Am. Chem. Soc.* **2013**, *135*, 12480–12496. <https://doi.org/10.1021/ja405051f>.
- (8) Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J. Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science* **2010**, *329*, 305–309. <https://doi.org/10.1126/science.1188934>.
- (9) Hughes, D. L. Biocatalysis in Drug Development - Highlights of the Recent Patent Literature. *Org. Process Res. Dev.* **2018**, *22*, 1063–1080. <https://doi.org/10.1021/acs.oprd.8b00232>.
- (10) Sheldon, R. A.; Brady, D.; Bode, M. L. The Hitchhiker's Guide to Biocatalysis: Recent Advances in the Use of Enzymes in Organic Synthesis. *Chem. Sci.* **2020**, *11*, 2587–2605. <https://doi.org/10.1039/C9SC05746C>.
- (11) Huffman, M. A.; Fryszkowska, A.; Alvizo, O.; Borra-Garske, M.; Campos, K. R.;

- Canada, K. A.; Devine, P. N.; Duan, D.; Forstater, J. H.; Grosser, S. T.; Halsey, H. M.; Hughes, G. J.; Jo, J.; Joyce, L. A.; Kolev, J. N.; Liang, J.; Maloney, K. M.; Mann, B. F.; Marshall, N. M.; McLaughlin, M.; Moore, J. C.; Murphy, G. S.; Nawrat, C. C.; Nazor, J.; Novick, S.; Patel, N. R.; Rodriguez-Granillo, A.; Robaire, S. A.; Sherer, E. C.; Truppo, M. D.; Whittaker, A. M.; Verma, D.; Xiao, L.; Xu, Y.; Yang, H. Design of an in Vitro Biocatalytic Cascade for the Manufacture of Islatravir. *Science* **2019**, *366*, 1255–1259. <https://doi.org/10.1126/science.aay8484>.
- (12) Woodley, J. M. Protein Engineering of Enzymes for Process Applications. *Curr. Opin. Chem. Biol.* **2013**, *17*, 310–316. <https://doi.org/10.1016/j.cbpa.2013.03.017>.
- (13) Huisman, G. W.; Collier, S. J. On the Development of New Biocatalytic Processes for Practical Pharmaceutical Synthesis. *Curr. Opin. Chem. Biol.* **2013**, *17*, 284–292. <https://doi.org/10.1016/j.cbpa.2013.01.017>.
- (14) Hult, K.; Berglund, P. Enzyme Promiscuity: Mechanism and Applications. *Trends Biotechnol.* **2007**, *25*, 231–238. <https://doi.org/10.1016/j.tibtech.2007.03.002>.
- (15) Tawfik, O. K. and D. S. Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. *Annu. Rev. Biochem.* **2010**, *79*, 471–505. <https://doi.org/10.1146/annurev-biochem-030409-143718>.
- (16) Renata, H.; Wang, Z. J.; Arnold, F. H. Expanding the Enzyme Universe: Accessing Non-Natural Reactions by Mechanism-Guided Directed Evolution. *Angew. Chem., Int. Ed.* **2015**, *54*, 3351–3367. <https://doi.org/10.1002/anie.201409470>.
- (17) Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. Olefin Cyclopropanation via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes. *Science* **2013**, *339*, 307–310. <https://doi.org/10.1126/science.1231434>.
- (18) Bordeaux, M.; Tyagi, V.; Fasan, R. Highly Diastereoselective and Enantioselective Olefin Cyclopropanation Using Engineered Myoglobin-Based Catalysts. *Angew. Chem., Int. Ed.* **2015**, *54*, 1744–1748. <https://doi.org/10.1002/anie.201409928>.
- (19) Hernandez, K. E.; Renata, H.; Lewis, R. D.; Kan, S. B. J. J.; Zhang, C.; Forte, J.; Rozzell, D.; McIntosh, J. A.; Arnold, F. H. Highly Stereoselective Biocatalytic Synthesis of Key Cyclopropane Intermediate to Ticagrelor. *ACS Catal.* **2016**, *6*, 7810–7813. <https://doi.org/10.1021/acscatal.6b02550>.
- (20) Kan, S. B. J. J.; Lewis, R. D.; Chen, K.; Arnold, F. H. Directed Evolution of Cytochrome c for Carbon–Silicon Bond Formation: Bringing Silicon to Life. *Science* **2016**, *354*, 1048–1051. <https://doi.org/10.1126/science.aah6219>.
- (21) 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. <https://doi.org/10.1016/j.cbpa.2016.09.004>.
- (22) Hyster, T. K.; Ward, T. R. Genetic Optimization of Metalloenzymes: Enhancing Enzymes for Non-Natural Reactions. *Angew. Chem., Int. Ed.* **2016**, *55*, 7344–7357. <https://doi.org/10.1002/anie.201508816>.
- (23) Dydio, P.; Key, H. M.; Nazarenko, A.; Rha, J. Y. E.-E.; Seyedkazemi, V.; Clark, D. S.; Hartwig, J. F. An Artificial Metalloenzyme with the Kinetics of Native Enzymes. *Science* **2016**, *354*, 102–106. <https://doi.org/10.1126/science.aah4427>.
- (24) Key, H. M.; Dydio, P.; Clark, D. S.; Hartwig, J. F. Abiological Catalysis by Artificial Haem Proteins Containing Noble Metals in Place of Iron. *Nature* **2016**, *534*, 534–537. <https://doi.org/10.1038/nature17968>.

- (25) Dydio, P.; Key, H. M.; Hayashi, H.; Clark, D. S.; Hartwig, J. F. Chemoselective, Enzymatic C-H Bond Amination Catalyzed by a Cytochrome P450 Containing an Ir(Me)-PIX Cofactor. *J. Am. Chem. Soc.* **2017**, *139*, 1750–1753. <https://doi.org/10.1021/jacs.6b11410>.
- (26) Key, H. M.; Dydio, P.; Liu, Z.; Rha, J. Y. E.; Nazarenko, A.; Seyedkazemi, V.; Clark, D. S.; Hartwig, J. F. Beyond Iron: Iridium-Containing P450 Enzymes for Selective Cyclopropanations of Structurally Diverse Alkenes. *ACS Cent. Sci.* **2017**, *3*, 302–308. <https://doi.org/10.1021/acscentsci.6b00391>.
- (27) Srivastava, P.; Yang, H.; Ellis-Guardiola, K.; Lewis, J. C. Engineering a Dirhodium Artificial Metalloenzyme for Selective Olefin Cyclopropanation. *Nat. Commun.* **2015**, *6*, 7789. <https://doi.org/10.1038/ncomms8789>.
- (28) Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N. Computational Enzyme Design. *Angew. Chem., Int. Ed.* **2013**, *52*, 5700–5725. <https://doi.org/10.1002/anie.201204077>.
- (29) Althoff, E. A.; Wang, L.; Jiang, L.; Giger, L.; Lassila, J. K.; Wang, Z.; Smith, M.; Hari, S.; Kast, P.; Herschlag, D.; Hilvert, D.; Baker, D. Robust Design and Optimization of Retroaldol Enzymes. *Protein Sci.* **2012**, *21*, 717–726. <https://doi.org/10.1002/pro.2059>.
- (30) Obexer, R.; Godina, A.; Garrabou, X.; Mittl, P. R. E.; Baker, D.; Griffiths, A. D.; Hilvert, D. Emergence of a Catalytic Tetrad during Evolution of a Highly Active Artificial Aldolase. *Nat. Chem.* **2017**, *9*, 50–56. <https://doi.org/10.1038/nchem.2596>.
- (31) Giger, L.; Caner, S.; Obexer, R.; Kast, P.; Baker, D.; Ban, N.; Hilvert, D. Evolution of a Designed Retro-Aldolase Leads to Complete Active Site Remodeling. *Nat. Chem. Biol.* **2013**, *9*, 494–498. <https://doi.org/10.1038/nchembio.1276>.
- (32) Brandenberg, O. F.; Fasan, R.; Arnold, F. H. Exploiting and Engineering Hemoproteins for Abiological Carbene and Nitrene Transfer Reactions. *Curr. Opin. Biotechnol.* **2017**, *47*, 102–111. <https://doi.org/10.1016/j.copbio.2017.06.005>.
- (33) Bornscheuer, U. T.; Kazlauskas, R. J. Catalytic Promiscuity in Biocatalysis: Using Old Enzymes to Form New Bonds and Follow New Pathways. *Angew. Chem., Int. Ed.* **2004**, *43*, 6032–6040. <https://doi.org/10.1002/anie.200460416>.
- (34) Sheldon, R. A. E Factors, Green Chemistry and Catalysis: An Odyssey. *Chem. Commun.* **2008**, No. 29, 3352. <https://doi.org/10.1039/b803584a>.
- (35) Prier, C. K.; Arnold, F. H. Chemomimetic Biocatalysis: Exploiting the Synthetic Potential of Cofactor-Dependent Enzymes To Create New Catalysts. *J. Am. Chem. Soc.* **2015**, *137*, 13992–14006. <https://doi.org/10.1021/jacs.5b09348>.
- (36) Arnold, F. H. The Nature of Chemical Innovation: New Enzymes by Evolution. *Q. Rev. Biophys.* **2015**, *48*, 404–410. <https://doi.org/10.1017/S003358351500013X>.
- (37) Arnold, F. H. Directed Evolution: Bringing New Chemistry to Life. *Angew. Chem., Int. Ed.* **2018**, *57*, 4143–4148. <https://doi.org/10.1002/anie.201708408>.
- (38) Busto, E.; Gotor-Fernández, V.; Gotor, V. Hydrolases: Catalytically Promiscuous Enzymes for Non-Conventional Reactions in Organic Synthesis. *Chem. Soc. Rev.* **2010**, *39*, 4504. <https://doi.org/10.1039/c003811c>.
- (39) Hammer, S. C.; Marjanovic, A.; Dominicus, J. M.; Nestl, B. M.; Hauer, B. Squalene Hopene Cyclases Are Protonases for Stereoselective Brønsted Acid Catalysis. *Nat. Chem. Biol.* **2015**, *11*, 121–126. <https://doi.org/10.1038/nchembio.1719>.

- (40) Drienovská, I.; Roelfes, G. Expanding the Enzyme Universe with Genetically Encoded Unnatural Amino Acids. *Nat. Catal.* **2020**, *3*, 193–202. <https://doi.org/10.1038/s41929-019-0410-8>.
- (41) Natoli, S. N.; Hartwig, J. F. Noble–Metal Substitution in Hemoproteins: An Emerging Strategy for Abiological Catalysis. *Acc. Chem. Res.* **2019**, *52*, 326–335. <https://doi.org/10.1021/acs.accounts.8b00586>.
- (42) de Zoete, M. C.; van Rantwijk, F.; Sheldon, R. A. Lipase-Catalyzed Transformations with Unnatural Acyl Acceptors. *Catal. Today* **1994**, *22*, 563–590. [https://doi.org/10.1016/0920-5861\(94\)80124-X](https://doi.org/10.1016/0920-5861(94)80124-X).
- (43) Balkenhohl, F.; Ditrich, K.; Hauer, B.; Ladner, W. Optisch Aktive Amine Durch Lipase-Katalysierte Methoxyacetylierung. *J. für Prakt. Chemie/Chemiker-Zeitung* **1997**, *339*, 381–384. <https://doi.org/10.1002/prac.19973390166>.
- (44) Breuer, M.; Ditrich, K.; Habicher, T.; Hauer, B.; Keßeler, M.; Stürmer, R.; Zelinski, T. Industrial Methods for the Production of Optically Active Intermediates. *Angew. Chem., Int. Ed.* **2004**, *43*, 788–824. <https://doi.org/10.1002/anie.200300599>.
- (45) Ma, S. K.; Gruber, J.; Davis, C.; Newman, L.; Gray, D.; Wang, A.; Grate, J.; Huisman, G. W.; Sheldon, R. A. A Green-by-Design Biocatalytic Process for Atorvastatin Intermediate. *Green Chem.* **2010**, *12*, 81–86. <https://doi.org/10.1039/B919115C>.
- (46) List, B. *Asymmetric Organocatalysis*; List, B., Ed.; Topics in Current Chemistry; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009; Vol. 291. <https://doi.org/10.1007/978-3-642-02815-1>.
- (47) Miao, Y.; Tepper, P. G.; Geertsema, E. M.; Poelarends, G. J. Stereochemical Control of Enzymatic Carbon-Carbon Bond-Forming Michael-Type Additions by “Substrate Engineering.” *Eur. J. Org. Chem.* **2016**, *2016*, 5350–5354. <https://doi.org/10.1002/ejoc.201601126>.
- (48) Beigi, M.; Gauchenova, E.; Walter, L.; Waltzer, S.; Bonina, F.; Stillger, T.; Rother, D.; Pohl, M.; Müller, M. Regio- and Stereoselective Aliphatic-Aromatic Cross-Benzoin Reaction: Enzymatic Divergent Catalysis. *Chem. Eur. J.* **2016**, *22*, 13999–14005. <https://doi.org/10.1002/chem.201602084>.
- (49) Garrabou, X.; Beck, T.; Hilvert, D. A Promiscuous De Novo Retro-Aldolase Catalyzes Asymmetric Michael Additions via Schiff Base Intermediates. *Angew. Chem., Int. Ed.* **2015**, *54*, 5609–5612. <https://doi.org/10.1002/anie.201500217>.
- (50) Garrabou, X.; Wicky, B. I. M. M.; Hilvert, D. Fast Knoevenagel Condensations Catalyzed by an Artificial Schiff-Base-Forming Enzyme. *J. Am. Chem. Soc.* **2016**, *138*, 6972–6974. <https://doi.org/10.1021/jacs.6b00816>.
- (51) Emmanuel, M. A.; Greenberg, N. R.; Oblinsky, D. G.; Hyster, T. K. Accessing Non-Natural Reactivity by Irradiating Nicotinamide-Dependent Enzymes with Light. *Nature* **2016**, *540*, 414–417. <https://doi.org/10.1038/nature20569>.
- (52) Biegasiewicz, K. F.; Cooper, S. J.; Gao, X.; Oblinsky, D. G.; Kim, J. H.; Garfinkle, S. E.; Joyce, L. A.; Sandoval, B. A.; Scholes, G. D.; Hyster, T. K. Photoexcitation of Flavoenzymes Enables a Stereoselective Radical Cyclization. *Science* **2019**, *364*, 1166–1169. <https://doi.org/10.1126/science.aaw1143>.
- (53) Hyster, T.; Nakano, Y.; Black, M. J.; Meichan, A. J.; Sandoval, B. A.; Chung, M.; Biegasiewicz, K.; Zhu, T. Photoenzymatic Hydrogenation of Heteroaromatic Olefins

- Using 'Ene'-Reductases with Photoredox Catalysts. *Angew. Chem., Int. Ed.* **2020**, anie.202003125. <https://doi.org/10.1002/anie.202003125>.
- (54) Tinoco, A.; Steck, V.; Tyagi, V.; Fasan, R. Highly Diastereo- and Enantioselective Synthesis of Trifluoromethyl-Substituted Cyclopropanes via Myoglobin-Catalyzed Transfer of Trifluoromethylcarbene. *J. Am. Chem. Soc.* **2017**, *139*, 5293–5296. <https://doi.org/10.1021/jacs.7b00768>.
- (55) Kan, S. B. J.; Huang, X.; Gumulya, Y.; Chen, K.; Arnold, F. H. Genetically Programmed Chiral Organoborane Synthesis. *Nature* **2017**, *552*, 132–136. <https://doi.org/10.1038/nature24996>.
- (56) Huang, X.; Garcia-Borràs, M.; Miao, K.; Kan, S. B. J.; Zutshi, A.; Houk, K. N.; Arnold, F. H. A Biocatalytic Platform for Synthesis of Chiral  $\alpha$ -Trifluoromethylated Organoborons. *ACS Cent. Sci.* **2019**, *5*, 270–276. <https://doi.org/10.1021/acscentsci.8b00679>.
- (57) Zhang, J.; Huang, X.; Zhang, R. K.; Arnold, F. H. Enantiodivergent  $\alpha$ -Amino C–H Fluoroalkylation Catalyzed by Engineered Cytochrome P450s. *J. Am. Chem. Soc.* **2019**, *141*, 9798–9802. <https://doi.org/10.1021/jacs.9b04344>.
- (58) Bajaj, P.; Sreenilayam, G.; Tyagi, V.; Fasan, R. Gram-Scale Synthesis of Chiral Cyclopropane-Containing Drugs and Drug Precursors with Engineered Myoglobin Catalysts Featuring Complementary Stereoselectivity. *Angew. Chem., Int. Ed.* **2016**, *55*, 16110–16114. <https://doi.org/10.1002/anie.201608680>.
- (59) Wang, Z. J.; Renata, H.; Peck, N. E.; Farwell, C. C.; Coelho, P. S.; Arnold, F. H. Improved Cyclopropanation Activity of Histidine-Ligated Cytochrome P450 Enables the Enantioselective Formal Synthesis of Levomilnacipran. *Angew. Chem., Int. Ed.* **2014**, *53*, 6810–6813. <https://doi.org/10.1002/anie.201402809>.
- (60) Kim, T.; Kassim, A. M.; Botejue, A.; Zhang, C.; Forte, J.; Rozzell, D.; Huffman, M. A.; Devine, P. N.; McIntosh, J. A. Hemoprotein-Catalyzed Cyclopropanation En Route to the Chiral Cyclopropanol Fragment of Grazoprevir. *ChemBioChem* **2019**, *20*, 1129–1132. <https://doi.org/10.1002/cbic.201800652>.
- (61) Chen, K.; Huang, X.; Kan, S. B. J.; Zhang, R. K.; Arnold, F. H. Enzymatic Construction of Highly Strained Carbocycles. *Science* **2018**, *360*, 71–75. <https://doi.org/10.1126/science.aar4239>.
- (62) Chen, K.; Arnold, F. H. Engineering Cytochrome P450s for Enantioselective Cyclopropanation of Internal Alkynes. *J. Am. Chem. Soc.* **2020**, *142*, 6891–6895. <https://doi.org/10.1021/jacs.0c01313>.
- (63) Zhang, R. K.; Chen, K.; Huang, X.; Wohlschlager, L.; Renata, H.; Arnold, F. H. Enzymatic Assembly of Carbon–Carbon Bonds via Iron-Catalysed Sp<sup>3</sup> C–H Functionalization. *Nature* **2019**, *565*, 67–72. <https://doi.org/10.1038/s41586-018-0808-5>.
- (64) Prier, C. K.; Zhang, R. K.; Buller, A. R.; Brinkmann-Chen, S.; Arnold, F. H. Enantioselective, Intermolecular Benzylic C–H Amination Catalysed by an Engineered Iron-Haem Enzyme. *Nat. Chem.* **2017**, *9*, 629–634. <https://doi.org/10.1038/nchem.2783>.
- (65) Tsutsumi, H.; Katsuyama, Y.; Izumikawa, M.; Takagi, M.; Fujie, M.; Satoh, N.; Shin-ya, K.; Ohnishi, Y. Unprecedented Cyclization Catalyzed by a Cytochrome P450 in Benzastatin Biosynthesis. *J. Am. Chem. Soc.* **2018**, *140*, 6631–6639.

- <https://doi.org/10.1021/jacs.8b02769>.
- (66) Reynolds, E. W.; Schwochert, T. D.; McHenry, M. W.; Watters, J. W.; Brustad, E. M. Orthogonal Expression of an Artificial Metalloenzyme for Abiotic Catalysis. *ChemBioChem* **2017**, *18*, 2380–2384. <https://doi.org/10.1002/cbic.201700397>.
- (67) Sreenilayam, G.; Moore, E. J.; Steck, V.; Fasan, R. Metal Substitution Modulates the Reactivity and Extends the Reaction Scope of Myoglobin Carbene Transfer Catalysts. *Adv. Synth. Catal.* **2017**, *359*, 2076–2089. <https://doi.org/10.1002/adsc.201700202>.
- (68) Jeschek, M.; Reuter, R.; Heinisch, T.; Trindler, C.; Klehr, J.; Panke, S.; Ward, T. R. Directed Evolution of Artificial Metalloenzymes for in Vivo Metathesis. *Nature* **2016**, *537*, 661–665. <https://doi.org/10.1038/nature19114>.
- (69) Yang, H.; Swartz, A. M.; Park, H. J.; Srivastava, P.; Ellis-Guardiola, K.; Upp, D. M.; Lee, G.; Belsare, K.; Gu, Y.; Zhang, C.; Moellering, R. E.; Lewis, J. C. Evolving Artificial Metalloenzymes via Random Mutagenesis. *Nat. Chem.* **2018**, *10*, 318–324. <https://doi.org/10.1038/nchem.2927>.
- (70) Hyster, T. K.; Farwell, C. C.; Buller, A. R.; McIntosh, J. A.; Arnold, F. H. Enzyme-Controlled Nitrogen-Atom Transfer Enables Regiodivergent C–H Amination. *J. Am. Chem. Soc.* **2014**, *136*, 15505–15508. <https://doi.org/10.1021/ja509308v>.
- (71) Reedy, C. J.; Gibney, B. R. Heme Protein Assemblies. *Chem. Rev.* **2004**, *104*, 617–650. <https://doi.org/10.1021/cr0206115>.
- (72) Bordeaux, M.; Singh, R.; Fasan, R. Intramolecular C(Sp<sup>3</sup>)H Amination of Arylsulfonyl Azides with Engineered and Artificial Myoglobin-Based Catalysts. *Bioorg. Med. Chem.* **2014**, *22*, 5697–5704. <https://doi.org/10.1016/j.bmc.2014.05.015>.
- (73) Knight, A. M.; Kan, S. B. J. B. J.; Lewis, R. D.; Brandenburg, O. F.; Chen, K.; Arnold, F. H. Diverse Engineered Heme Proteins Enable Stereodivergent Cyclopropanation of Unactivated Alkenes. *ACS Cent. Sci.* **2018**, *4*, 372–377. <https://doi.org/10.1021/acscentsci.7b00548>.
- (74) Miller, D. J.; Allemann, R. K. Sesquiterpene Synthases: Passive Catalysts or Active Players? *Nat. Prod. Rep.* **2012**, *29*, 60–71. <https://doi.org/10.1039/C1NP00060H>.
- (75) Knowles, R. R.; Jacobsen, E. N. Attractive Noncovalent Interactions in Asymmetric Catalysis: Links between Enzymes and Small Molecule Catalysts. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 20678–20685. <https://doi.org/10.1073/pnas.1006402107>.

DIVERSE ENGINEERED HEME PROTEINS ENABLE  
STEREODIVERGENT CYCLOPROPANATION OF UNACTIVATED  
AND ELECTRON-DEFICIENT ALKENES

Content in this chapter is adapted from published work:

**Knight, A.M.;** Kan, S.B.J.; Lewis, R.D.; Brandenberg, O.F.; Chen, K.; Arnold, F.H. Diverse Engineered Heme Proteins Enable Stereodivergent Cyclopropanation of Unactivated Alkenes. *ACS Cent. Sci.* **2018**, *4*, 372–377.

A.M.K. participated in the conception of the project, discovered initial enzymatic activity, performed the majority of the mutagenesis and screening, synthesized authentic standards, developed analytical methods to determine enantioselectivity, and wrote the manuscript with input from all authors. A.M.K. generated the crystallography constructs, purified, and carried out crystallography experiments. A.M.K. collected data for and refined the tagless RmaNOD Q52V crystal structure.

## Abstract

Developing catalysts that selectively produce each stereoisomer of a desired product is a longstanding synthetic challenge. Biochemists have addressed this challenge by screening nature's diversity to discover enzymes that catalyze the formation of complementary stereoisomers. I show here that the same approach can be applied to a new-to-nature enzymatic reaction: alkene cyclopropanation via carbene transfer. By screening diverse native and engineered heme proteins, we identified globins and serine-ligated P411 variants of cytochromes P450 with promiscuous activity for cyclopropanation of unactivated alkene substrates. I then enhanced their activities and stereoselectivities by directed evolution: just 1–3 rounds of site-saturation mutagenesis and screening generated enzymes that transform unactivated alkenes and electron-deficient alkenes into each of the four stereoisomeric cyclopropanes with up to 5,400 total turnovers and 98% enantiomeric excess. These fully genetically encoded biocatalysts function in whole *Escherichia coli* (*E. coli*) cells in mild, aqueous conditions and provide the first example of enantioselective, intermolecular iron-catalyzed cyclopropanation of unactivated alkenes. I also describe efforts toward structural characterization of these variants and expanding cyclopropanation of unactivated and electron-deficient alkenes beyond  $\alpha$ -diazo ester carbenes.

## 2.1 Introduction

The biological world is a marvelous ensemble of chiral molecules. From the amino acid and nucleoside building blocks that form proteins and DNA to intricate natural products produced by living organisms, chirality dictates how molecules interact with living systems.<sup>1</sup> Modern medicines often draw inspiration from chiral natural products.<sup>2</sup> Because alternate stereoisomers can have very different biological effects,<sup>3</sup> characterization of novel bioactive compounds during drug candidate screening should include testing each stereoisomer.<sup>4</sup> Developing stereodivergent syntheses, where a set of complementary catalysts can generate every possible stereoisomer of the product, is therefore useful and is actively sought after in catalysis.<sup>5</sup> Enzymes are green, sustainable options for stereoselective catalysis, and stereo-complementary enzymes can often be found in nature's diversity: lipases,<sup>6</sup> ketoreductases,<sup>7</sup> and transaminases<sup>8</sup> chosen using genome mining,<sup>9</sup> for example, have all afforded products with different stereoselectivities. We show here that natural protein diversity can be leveraged in a similar fashion to achieve stereodivergence for a new, non-natural enzyme-catalyzed reaction: cyclopropanation of unactivated alkenes via carbene transfer. Cyclopropanes are found in a variety of natural products and have been incorporated as motifs in agrochemical and pharmaceutical compounds (Figure 2-5). They are commonly biosynthesized through intramolecular cyclization of cationic or radical intermediates catalyzed by terpene cyclases, desaturases, or radical S-adenosylmethionine-dependent enzymes. The formation of cyclopropane moieties via carbene transfer to alkenes, widely used in synthetic chemistry, is not known in nature.<sup>10</sup>

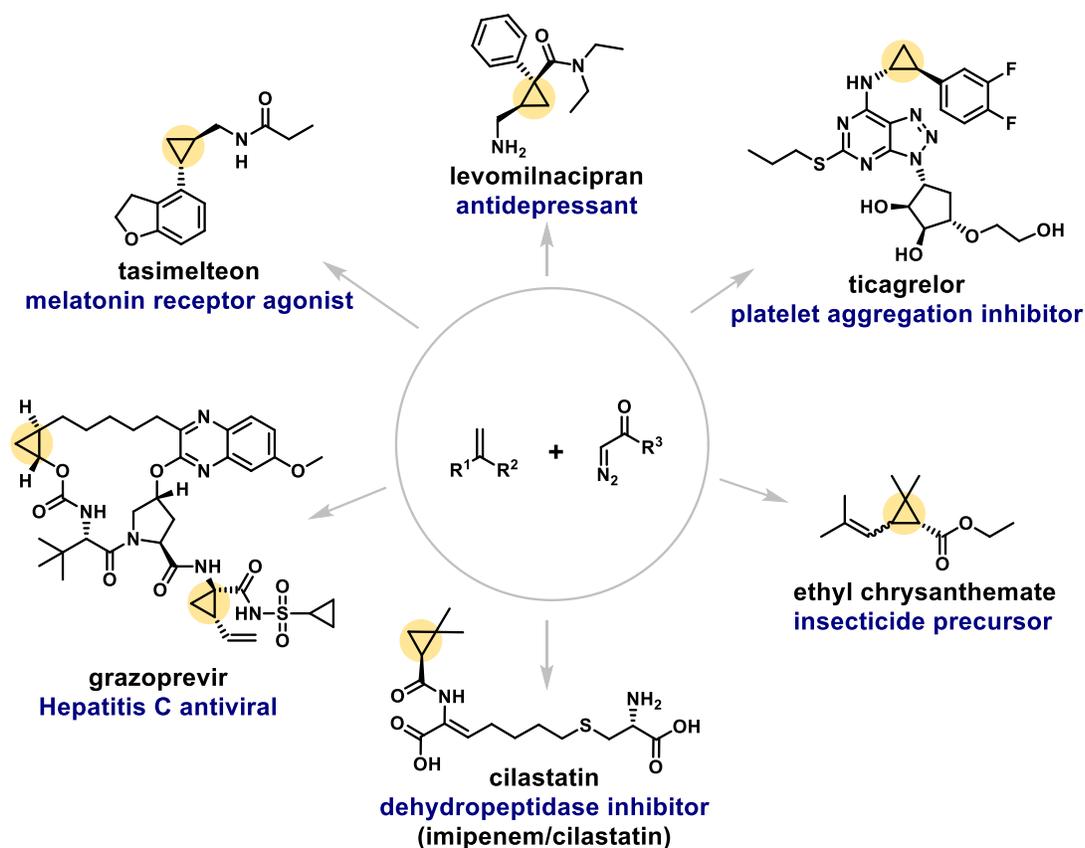


Figure 2-5. Cyclopropane-containing pharmaceutical and agrochemical compounds.

Previous work from this group and others has shown that iron-porphyrin (heme) proteins can be engineered to catalyze the cyclopropanation of styrenyl alkenes with ethyl diazoacetate (EDA, **1**).<sup>11-14</sup> This new-to-nature carbene transfer reaction has been applied in the synthesis of key pharmaceutical intermediates such as levomilnacipran,<sup>15</sup> ticagrelor,<sup>16,17</sup> tasimelteon,<sup>16</sup> and grazoprevir<sup>18</sup> (Figure 2-5). Work from several groups has also shown that cytochromes P450 and myoglobins can be engineered to switch their diastereo- and/or enantioselectivity preference toward styrenyl olefin cyclopropanation,<sup>11,13,16</sup> but these precedents did not yield all possible stereoisomers in high turnover and selectivity (Table 2-1). In addition, alkene cyclopropanation by heme proteins with the native iron cofactor has been limited to styrenyl and other activated alkenes. Unactivated, aliphatic alkenes are attractive feedstocks for chemical synthesis, but their transformation to higher value chiral products is challenging

due to their inert nature, high degree of conformational flexibility, and limited steric and electronic bias to guide stereocontrol.<sup>19</sup>

Table 2-1. Literature precedent for stereocomplementary cyclopropanation of styrenyl alkenes via enzymatic carbene transfer. Reference numbers are given for the main text reference numbers.

Reference	<i>Trans</i> -cyclopropane		<i>Cis</i> -cyclopropane	
	( <i>R,R</i> )	( <i>S,S</i> )	( <i>R,S</i> )	( <i>S,R</i> )
11	<b>Not reported</b>	Up to 98% de, 96% ee	Up to 84% de, 97% ee	<b>Not reported</b>
13	Up to 86% de, <b>32% ee</b>	Up to 98% de, 97% ee	Up to 78% de, 99% ee	Up to <b>42% de</b> , 95% ee
16	Up to 99.9% de, 95% ee	Up to 99.9% de, 99.9% ee	<b>Not reported</b>	<b>Not reported</b>
20	Up to 99.5% de, 92% ee	Up to 99.9% de, 99.9% ee	<b>Not reported</b>	<b>Not reported</b>

State-of-the art methods for enantioselective unactivated alkene cyclopropanation often rely on noble metals,<sup>21–23</sup> with only a few examples of Co-<sup>24</sup> and Cu-based<sup>25</sup> catalysts (Table 2-2); no iron-based catalyst for the enantioselective intermolecular cyclopropanation of unactivated alkenes has been reported at the time of this work. There is precedent that iron-porphyrin catalysts are capable of unactivated alkene cyclopropanation, albeit achirally and with a non-natural iron porphyrin. Two decades ago, Woo and coworkers showed that iron *meso*-tetrakis(pentafluorophenyl)porphyrin chloride (Fe(PFP)Cl) can catalyze the reaction of 2-ethyl-1-butene and EDA with 390 TTN; they reported the formation of cyclopropane products using 1-decene as well.<sup>26</sup>

Table 2-2. State-of-the-art methods for the intermolecular cyclopropanation of unactivated alkenes prior to our report of the stereodivergent cyclopropanation of unactivated alkenes with heme proteins. Because iron-catalyzed asymmetric examples were not known, achiral examples are listed for iron. Enantioselective reactions are reported for other metals.

Entry	Metal Catalyst	Enantioselective?	Ref.	Note
1	Fe	No	1	Iron porphyrin-catalyzed, <u>achiral</u> . Two examples involved unactivated alkenes. Fe(PFP)Cl (0.02-0.05 mol%) catalyzed the cyclopropanation of 2-ethyl-1-butene (100 eq.) with EDA in 390 TON, d.r. not reported. PFP = <i>meso</i> -tetrakis(pentafluorophenyl)porphyrin

Entry	Metal Catalyst	Enantioselective?	Ref.	Note
2	Fe	No	2	Iron porphyrin-catalyzed, <u>achiral</u> . One example involved unactivated alkenes. Fe(TPP)Cl (3 mol%) catalyzed the cyclopropanation of allyl benzene with <i>in situ</i> generated trifluoromethyl diazomethane (1.5 eq.) in water with 3.3 TON, d.r. not reported. TPP = 5,10,15,20-tetraphenyl-21 <i>H</i> ,23 <i>H</i> -porphine
3	Fe	No	3	Iron porphyrin-catalyzed, <u>achiral</u> . One example involved unactivated alkenes. This example is similar to entry 2, but a different <i>in situ</i> method was used to generate the trifluoromethyl diazomethane. No product was detected.
4	Fe	No	4	Iron porphyrin-catalyzed, <u>achiral</u> . Three examples involved unactivated alkenes. Fe(TPP)Cl (10 mol%) catalyzed the cyclopropanation of aliphatic alkenes using <i>in situ</i> generated difluoromethylcarbene (2 eq.) in 6–8 TONs, 7:1 to 13: 1 d.r. ( <i>trans</i> ).
5	Rh	Yes	5-13	All Rh-carboxylate or Rh-carboxamidate-catalyzed (0.2–2 mol% Rh) examples of unactivated alkene cyclopropanation reported are <b><i>trans</i>-selective</b> . The most efficient and selective examples are the cyclopropanation of 1-hexene (72 TTN, >20:1 d.r. ( <i>trans</i> ), 96% <i>e.e.</i> ), and 1-octene (70 TTN, >20:1 d.r. ( <i>trans</i> ), 90 % <i>e.e.</i> ).
7	Ir	Yes	14	Ir-salen (1 mol%) catalyzed the cyclopropanation of >10 unactivated alkenes with up to 93 TTN, 98:2 d.r. ( <i>cis</i> major), 99% <i>e.e.</i> .
8	Ir	Yes	15	Ir(Me)PPIX in engineered myoglobin (0.5 mol%) catalyzed the cyclopropanation of 1-octene in 40 TON, 91:9 d.r., 40:1 d.r. ( <i>trans</i> major). 6 equivalents of EDA were added <i>via</i> syringe pump over 12 h.
9	Ir	Yes	16	Ir(Me)PPIX in engineered CYP119 enzymes catalyzed the cyclopropanation of a variety of unactivated alkenes (terminal, internal and 1,1-disubstituted) with up to 1300 TON and 99% <i>e.e.</i> . 3 equivalents of EDA were added <i>via</i> syringe pump over 3 h.
10	Ru	Yes	17, 18	Ru-(iminophosphoranyl)ferrocene (2 mol%) catalyzed the cyclopropanation of 3 unactivated alkenes with up to 37 TON, 72:28 d.r. ( <i>cis</i> major), 98% <i>e.e.</i> .
11	Co	Yes	19, 20	All Co-catalyzed (1–5 mol% Co) examples of unactivated alkene cyclopropanation reported are <b><i>cis</i>-selective</b> and based on Co-porphyrins. The most efficient and selective example is the cyclopropanation of phenylbutene (90 TTN, >99:1 d.r. ( <i>cis</i> ), 96% <i>e.e.</i> ).
12	Cu	Yes	21, 22	All Cu-catalyzed (1 mol% Cu) examples of unactivated alkene cyclopropanation reported are <b><i>trans</i>-selective</b> . The most efficient and selective example is the cyclopropanation of 1-octene (80 TTN, 93:7 d.r. ( <i>trans</i> ), 90% <i>e.e.</i> ).

Metalloporphyrin catalysts have been used in synthetic chemistry for decades, but nature has used them for millions of years. Present in all forms of life on Earth, heme-binding proteins have diverse functions as well as promiscuous activities for which they were never selected, such as the ability to form reactive carbene intermediates. These heme-binding proteins have also been shown to bind metal-substituted porphyrins for new-to-nature reactions: Hartwig, Clark, and coworkers cited the need for greater metal center reactivity than iron porphyrin could provide in order to perform unactivated alkene cyclopropanation and C–H functionalization reactions when they showed that heme proteins could bind an artificial iridium cofactor in place of iron heme for carbene transfer chemistry.<sup>22</sup> They reported that a protein's active site can confer selectivity to noble-metal, small-molecule catalysts that can already catalyze the reaction.<sup>27</sup> Use of an artificial iridium cofactor (Ir(Me)PIX) required the lysis, purification, and *in vitro* metallation of the apoprotein with the Ir(Me)PIX, all of which add time and cost to catalyst preparation. Though it may be possible to incorporate the new metal *in vivo*,<sup>28</sup> the synthetic, noble-metal cofactor is more expensive than the native heme, which is manufactured by the cell and loaded into the catalyst during protein expression *in vivo*. The use of iridium is also not ideal due to the negative impact mining and refining precious metals has on the environment.<sup>29</sup>

In aqueous buffer with no protein, however, hemin can “catalyze” the formation of **3a**, albeit with only 0.4 TTN. Our group has previously used directed evolution of iron-porphyrin proteins to enhance the intrinsic heme activity, accessing carbene-transfer reactions typically performed with noble-metal catalysts, such as intermolecular C–H amination<sup>30</sup> and carbene insertion into Si–H,<sup>31</sup> B–H,<sup>32</sup> and C–H<sup>33</sup> bonds. In these cases, heme's basal activity is greatly enhanced, and stereoselectivity is enforced by the protein environment, enabling remarkable activity and selectivity enhancements through directed evolution. Different local heme environments can be accessed by screening natural and engineered protein diversity; directed evolution then fine-tunes these features. We therefore set out to create a collection of genetically encoded biocatalysts equipped with the native heme cofactor for

stereodivergent unactivated alkene cyclopropanation, taking advantage of the natural diversity of heme proteins to identify suitable starting activity and stereoselectivity.

## 2.1 Results and Discussion

In order to find initial enzymatic activity for unactivated alkene cyclopropanation, we collected a panel of 11 heme proteins from thermophilic and hyperthermophilic bacteria and archaea (Supplementary Information, Table 2-7), as thermostable proteins can better withstand the potentially destabilizing effects of mutations and are therefore more ‘evolvable.’<sup>34</sup> They are also often easier to work with and better tolerate polar organic solvents used to solubilize substrates. These heme proteins were initially tested with mutations to the putative distal ligand and active-site entrance loop (similar mutations have been found by Fasan and coworkers to enhance myoglobin-catalyzed cyclopropanation of vinylarenes)<sup>12</sup> for the cyclopropanation of 1,7-octadiene using EDA as a carbene precursor (Supplementary Information, Table 2-7). 1,7-Octadiene was chosen as an initial test substrate due to the higher apparent terminal alkene concentration. Protein variants that showed catalytic activity were subsequently tested as wild-type proteins for 1-octene cyclopropanation; 1-octene is a model substrate for unactivated alkene cyclopropanation. Through these experiments, wild-type *Aeropyrum pernix* protoglobin (*ApePgb* WT, UniProt ID: Q9YFF4) and wild-type *Rhodothermus marinus* nitric oxide dioxygenase (*RmaNOD* WT, UniProt ID: D0MGT2) were found to have low but measurable cyclopropanation activity on 1-octene (**2a**), catalyzing the reaction with 18 and 27 total turnovers per enzyme active site (TTN). Notably, *ApePgb* WT and *RmaNOD* WT displayed complementary diastereoselectivity, preferentially producing *cis*–(1*R*,2*S*)-**3a** and *trans*–(1*S*,2*S*)-**3a**, respectively.

In addition to searching globin protein diversity for this novel activity, we also investigated heme proteins obtained in previous directed evolution studies. A panel of 36 variants of serine-ligated *Bacillus megaterium* cytochrome P411 (P411<sub>BM3</sub>)<sup>35</sup> engineered for other non-natural carbene- and nitrene-transfer reactions<sup>14</sup> was tested for the ability to cyclopropanate 1-octene **2a** and 4-phenyl-1-butene (**2b**). **2b** Was chosen as a substrate for library screening,

because the UV-visible phenyl group enables screening by HPLC-UV. The P411<sub>BM3</sub> variant P411<sub>BM3</sub>-CIS L437F T438Q L75Y L181I (P411<sub>BM3</sub>-UA) showed significant activity and selectivity for production of *cis*-(1*S*,2*R*)-**3a**, the third of the four possible isomers. This variant of a serine-ligated “P411” (P411<sub>BM3</sub>-CIS<sup>29</sup>) had been engineered for cyclopropanation activity on electron-rich, non-styrenyl alkenes such as *N*-vinyl amides.<sup>36</sup>

Site-saturation mutagenesis libraries were generated and screened to increase the activities and selectivities of the different enzymes. Because crystal structures of *ApePgb* and *RmaNOD* have not been reported, homology models were built to help us identify residues within the putative distal heme pocket, where carbenoid formation and substrate binding are predicted to take place (Figure 2-6). P411<sub>BM3</sub>-UA residues were selected based on the crystal structure of its P411<sub>BM3</sub>-CIS predecessor (PDB ID: 4H23). Individual site-saturation libraries were screened for increased activity and diastereoselectivity using **2b** and **1** as substrates. Variants with enhanced diastereoselectivity in the production of **3b** were regrown on larger scale, and their activities were tested in the cyclopropanation of **2a** and **2b** with EDA. Enzyme variants with the greatest overall selectivity enhancements against **3a** and **3b** were used as parents in the next rounds of site-saturation mutagenesis and screening.

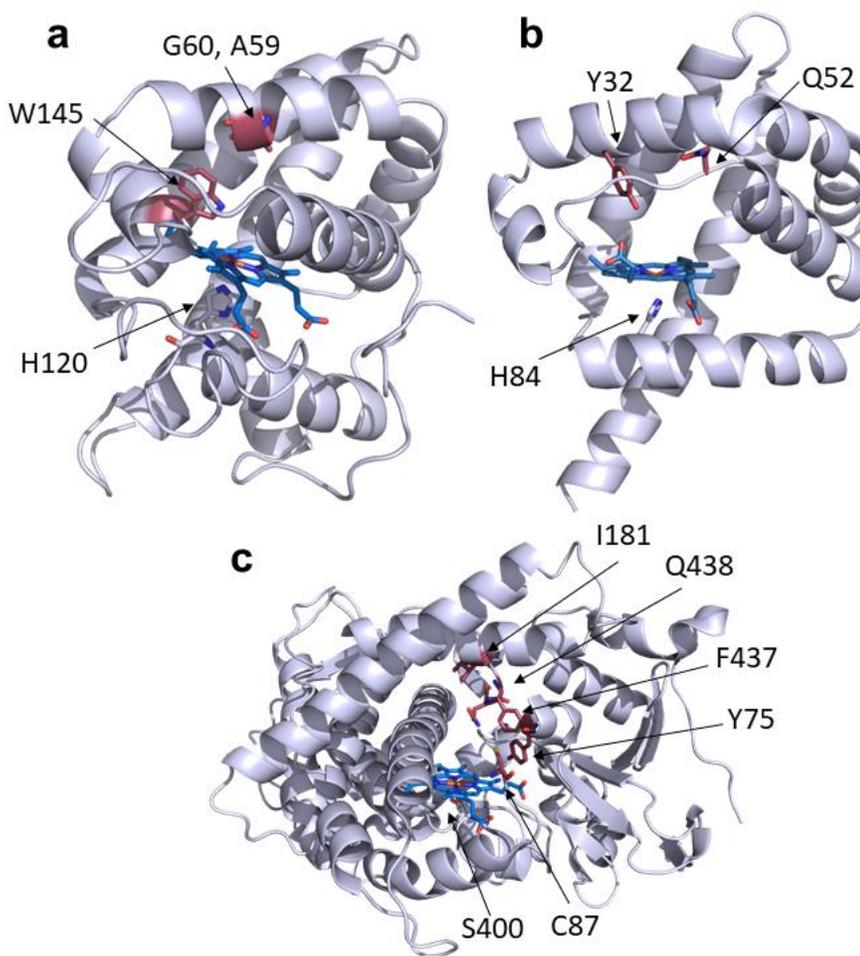


Figure 2-6. Homology models of proteins engineered in the study. The protein and proximal ligand are shown in gray. Red indicates residues at which mutations were made for the engineered variants. The heme cofactor is shown in blue. a) *ApePgb* homology model with W59A, Y60G, and F145W mutations in red. b) *RmaNOD* homology model with Q52 and the putative distal ligand Y32 (where mutations also enhanced activity) in red. c) P411<sub>BM3</sub>-UA homology model with V87C in red, as well as other mutated residues 75, 181, 437, and 438 colored.

*RmaNOD* was mutated at active-site positions Y32 (the putative distal axial ligand), Q52, and V97 using single-site site-saturation mutagenesis. Though the Y32G mutation found in the Y32X site-saturation mutagenesis library showed modest improvements in activity, the most significant increase in activity from this first round was the Q52V mutation. This single mutation (Q52V) gave *RmaNOD* near-perfect stereoselectivity for producing *trans*-(1*S*,2*S*)-**3a**.

*ApePgb* was first mutated at position Y60 (the putative distal axial ligand). This yielded the Y60G variant as the most active catalyst. *ApePgb* Y60G was then subjected to single-site site-saturation mutagenesis at residues W59, F73, F93, and F145 in parallel, all of which are positions known to modulate the gaseous ligand binding properties of the homologous *Methanosarcina acetivorans* protoglobin.<sup>31</sup> The variant that showed the greatest increase in activity while maintaining high selectivity was W59A Y60G. This new variant was used as parent for the next round of single-site site-saturation mutagenesis at positions F73, F93, and F145. After verification in small-scale biocatalytic reactions, *ApePgb* W59A Y60G F145W demonstrated a significantly increased enantioselectivity, making *cis* (*1R,2S*)-**3a** with 89:11 diastereomeric ratio (dr) and 99% enantiomeric excess (ee).

P411<sub>BM3</sub>-UA was first mutated at four positions found to modulate activity for heteroatom-substituted cyclopropanation reactions.<sup>36</sup> A simultaneous NDT double site-saturation mutagenesis library at F437/Q438 and single 22-codon site-saturation mutagenesis libraries<sup>28</sup> at Y75 and I181 were generated and screened for improved diastereoselectivity in the formation of **3b**. In each case, P411<sub>BM3</sub>-UA was among the most active and selective variants. Using P411<sub>BM3</sub>-UA as parent protein, single-site saturation-mutagenesis was performed at additional active-site residues V87, I263, E267, and A328, screening for enhanced diastereoselectivity. The most significant changes in diastereoselectivity were in the V87X site-saturation library, in which an enhanced *cis* diastereoselectivity mutation (V87C, producing the (*1S,2R*) isomer with 89:11 *cis:trans* dr) and a complete inversion in stereochemistry (V87F, producing the (*1R,2R*)-isomer with 4:96 *cis:trans* dr) were found. Residue 87 has often been reported to modulate the stereoselectivity of P450<sub>BM3</sub> for oxygenation of various substrates.<sup>37</sup>

With initial screening of 11 new and 36 previously engineered heme proteins, followed by just one to three rounds of site-saturation mutagenesis, we discovered four protein variants capable of cyclopropanating unactivated alkenes (*RmaNOD* Q52V, *ApePgb* W59A Y60G F145W (= *ApePgb* AGW), P411<sub>BM3</sub>-UA-V87C, and P411<sub>BM3</sub>-UA-V87F), each of which

produced a distinct stereoisomer of the desired product **3a** with 89:11 to <1:99 d.r. and 96 to >99% *e.e.* (Figure 2-7). The enzyme activities against unactivated alkenes are comparable to the state-of-the-art catalysts, with 100–490 TTN for **3a** and as high as 2400 TTN for **3b**, the substrate against which the enzymes were screened. These analytical-scale reactions were prepared under conditions unoptimized for product yield, but GC yields for **3b** and **3c** under these conditions are modest to good (Table 2-3). The system is straightforward and easy to use: the protein-expressing bacterial cells need only be resuspended to the desired concentration, and the alkene and diazo ester is added directly under an anaerobic atmosphere. When the reaction is complete, the product is extracted into organic solvents for analysis or purification.

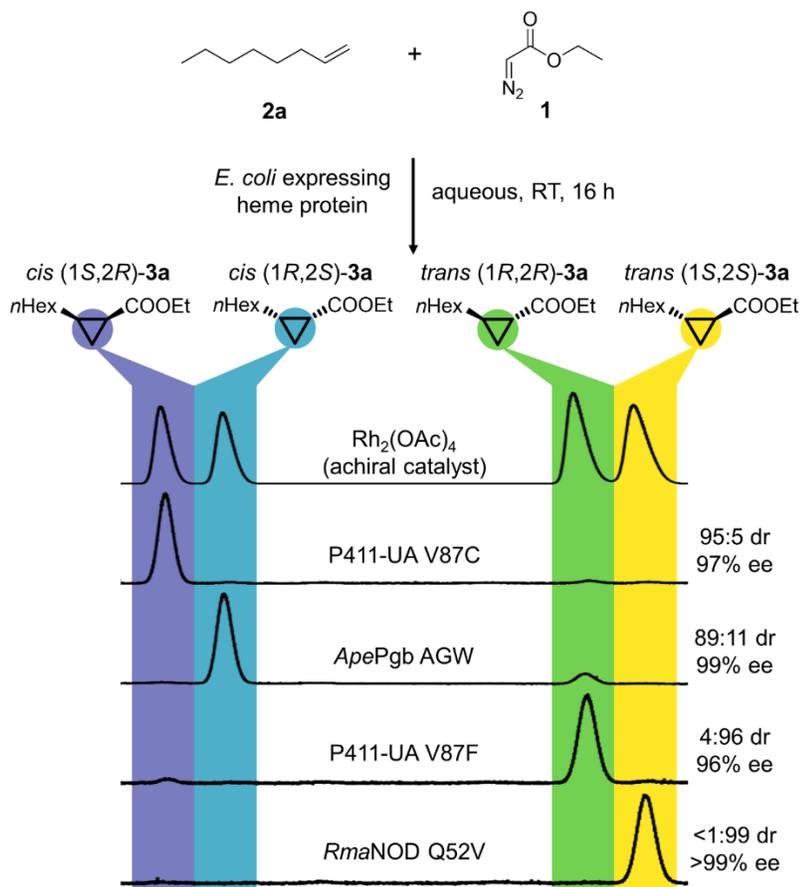


Figure 2-7. Stereodivergent enzymatic cyclopropanation of an unactivated alkene. The reaction of **2a** and **1** with four engineered heme proteins yields each of four stereoisomers of cyclopropane product **3a** with diastereoselectivities from 89:11 to <1:99 d.r. and enantioselectivities from 96% to

>99% ee. The diastereoselectivity ratio (dr) is given as *cis:trans*, and the enantiomeric excess (ee) is given for the major diastereomer.

Table 2-3. GC yields for analytical-scale reactions reported in Figure 2-8. These yields are determined by comparing the GC yield to the calibration curves. These reactions were run under conditions designed to demonstrate the catalysts' potential TTNs, rather than yields.

Alkene, product	P411 <sub>BM3</sub> -UA V87C	ApePgb AGW	P411 <sub>BM3</sub> -UA V87F	RmaNOD Q52V
1-octene ( <b>3a</b> )	4%	8%	7%	5%
4-phenyl-1-butene ( <b>3b</b> )	26%	40%	44%	4%
benzyl acrylate ( <b>3c</b> )	40%	91%	53%	71%
6-bromo-1-hexene ( <b>3d</b> )		19%		
vinyl cyclohexane ( <b>3e</b> )		18%		
methylenecyclohexane ( <b>3f</b> )		10%		
1-penten-3-one ( <b>3g</b> )		41%		
2-vinylpyridine ( <b>3h</b> )				60%

The four engineered biocatalysts were tested on a range of alkenes. Their activities and selectivities were high on unbranched aliphatic alkenes similar to those for which they were engineered, but their substrate scope extends to sterically hindered and electron-deficient alkenes as well (Figure 2-8). Though the activity and stereoselectivity differed on different substrates, each catalyst accepted most of the substrates tested. It is likely that activity on specific substrates can be optimized further, if desired, as has been shown in many other directed evolution studies.<sup>38,39</sup>

		P411-UA-V87C (1S,2R)	ApePgb AGW (1R,2S)	P411-UA-V87F (1R,2R)	RmaNOD Q52V (1S,2S)
<b>a</b>					
<b>3a</b>		270 TTN 95:5 dr 97% ee	490 TTN 89:11 dr 99% ee	310 TTN 4:96 dr 96% ee	100 TTN <1:99 dr >99% ee
<b>3b</b>		1900 TTN 96:4 dr >99% ee	2400 TTN 84:16 dr 95% ee	1900 TTN <1:99 dr >99% ee	210 TTN 3:97 dr 98% ee
<b>3c*</b>		2900 TTN 91:9 dr 96% ee	5400 TTN 71:29 dr 98% ee	2200 TTN 2:98 dr 96% ee	3700 TTN <1:99 dr >99% ee
<b>b</b>					
<b>3d</b>		1100 TTN <sup>1</sup> 92:8 dr 99% ee			
<b>3e</b>			550 TTN <sup>1</sup> 76:24 dr 96% ee		
<b>3f</b>		610 TTN <sup>1</sup> 96% ee			
<b>3g</b>			610 TTN <sup>1</sup> 89:11 dr 99% ee		
<b>3h</b>		3100 TTN <sup>2</sup> <1:99 dr 97% ee			

Figure 2-8. Cyclopropanation substrate scope. (a) Activity and selectivity of each protein variant against **3a**, **3b**, and **3c**. (b) Activity and selectivity against a variety of alkenes. The diastereoselectivity ratio (dr) is given as *cis:trans*, and the enantiomeric excess (ee) is given for the major diastereomer. Enzyme variant used is denoted by superscripts: 1, *ApePgb* AGW; 2, *RmaNOD* Q52V. General reaction conditions: whole *E. coli* cells ( $OD_{600} = 5$  (*ApePgb* AGW, *RmaNOD* Q52V),  $OD_{600} = 20$  (P411<sub>BM3</sub>-UA-V87C, P411<sub>BM3</sub>-UA-V87F)) in 1X-M9-N buffer, 25 mM glucose, 10 mM alkene, direct addition of 20 mM **1** under anaerobic conditions, 5% ethanol cosolvent. Modified reaction conditions: **3g**  $OD_{600} = 20$ ; **3e**  $OD_{600} = 10$ . *RmaNOD* Q52V **3a**,  $OD_{600} = 10$ . Analytical yields for these reactions are given in Table 2-3. The absolute configurations of products **3b** and **3c** are assigned by analogy to the **3a** products. Chiral separation conditions reported in the Supporting Information. \*The benzyl ester of **3c** has IUPAC naming priority, and therefore the stereogenic carbon center numbering is reversed for these compounds.

We have shown that these protein variants function with the commonly used diazo carbene precursor EDA, which is especially useful because the ester moiety can be further derivatized, e.g. to form amides and other esters. It can also be reduced to alcohols and aldehydes, which are versatile synthetic handles for accessing a broad range of functional groups. Reports of trifluoromethylcarbene transfer for alkene cyclopropanation<sup>20</sup> and insertion into B–H and C–H bonds,<sup>40,41</sup> carbene transfer to silanes<sup>31</sup> and boranes,<sup>32</sup> and diazolactone carbene transfer<sup>42,43</sup> have shown that heme proteins can be engineered to use different carbene precursors in new-to-nature reactions.

The small-molecule catalyzed enantioselective preparation of cyclopropyl esters from electron-deficient alkenes has previously been limited to making the *trans*-cyclopropanes,<sup>44</sup> whereas strategies to directly access 1-keto,2-ester or 1,2-diester *cis*-cyclopropanes (or their corresponding carboxylates) via enantioselective cyclopropanation have not been reported. The biocatalysts, in contrast, enable access to the *cis*-1-keto-2-ester and *cis*-1,2-diester products in a single, intermolecular step using an *E. coli*-based platform (*cis*-**3c**, *cis*-**3g**, Figure 2-8). Some of these products are precursors to valuable compounds: cyclopropyl esters of unbranched, aliphatic alkenes are used in fragrances, for example, including the essential odorants in frankincense.<sup>45</sup> Notably, the enzymes catalyze the reaction on 2-vinylpyridine (**2h**), which is a difficult substrate for many catalysts due to pyridine's propensity to coordinate to and inhibit metal centers. This cyclopropanation product is a precursor for an orphan GPR88 agonist.<sup>46</sup> A similar vinylpyridine substrate was used in myoglobin-catalyzed cyclopropanation.<sup>16</sup>

Enzymes are chemoselective and can generate desired products without additional steps to protect and deprotect other reactive functional groups on the same molecule. As shown in Figure 2-9, the enzymes described here can selectively cyclopropanate terminal alkenes in the presence of alcohol and carboxylic acid functional groups which often undergo competitive O–H insertion reactions with small-molecule carbene-transfer catalysts like rhodium acetate dimer.<sup>47</sup> *ApePgb* AGW performed particularly well with unprotected 7-octen-1-ol (**2i**) and 7-octen-1-oic acid (**2j**), yielding products *cis*-**3i** and *cis*-**3j** at 77% and

64% isolated yield, respectively, in preparative-scale reactions. Chemo- and regioselectivity is even more important when functional groups cannot be protected easily. In the cases of 1,3-(*E*)-pentadiene (**2k**) and 1,3-(*Z*)-pentadiene (**2l**), all four engineered proteins cyclopropanate the terminal alkene with perfect regioselectivity, likely due to higher accessibility of this double bond in each enzyme's active site. The diastereoselectivity varied for **3k** and **3l**, though the enantioselectivity for the major isomer remained high. As the electronic properties of **2k** and **2l** are similar, the difference in stereoselectivity likely reflects steric constraints of the enzyme active sites.

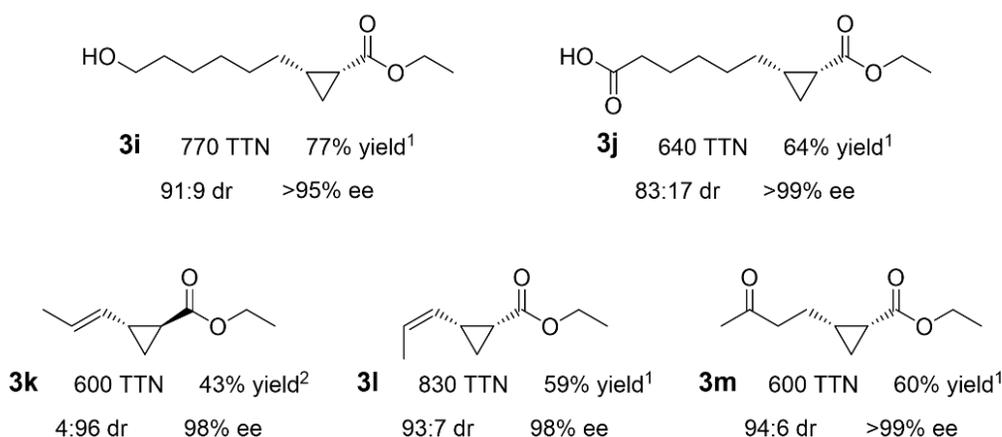


Figure 2-9. Selective, preparative-scale cyclopropane syntheses from various aliphatic alkenes and dienes. Preparative-scale reactions against substrates with free alcohol (7-octen-1-ol, **2i**), free carboxylic acid (7-octen-1-oic acid, **2j**), the two geometric isomers of penta-1,3-diene (**2k**, **2l**), and ketone (5-hexen-2-one, **2m**). The diastereoselectivity ratio (dr) is given as *cis:trans*, and the enantiomeric excess (ee) is given for the major diastereomer. The yields are reported for isolated products. Enzyme variant used is denoted by superscripts: 1, *ApePgb* AGW; 2, *RmaNOD* Q52V.

The four final protein variants were tested under whole-cell, aerobic conditions at analytical scale (400  $\mu$ L) for their ability to form **3a**. The reactions were set up under the same conditions as the anaerobic analytical-scale reactions, with the exception that they were set up outside of the Coy anaerobic chamber. The aerobic conditions resulted in a nearly complete loss in cyclopropanation for the globins, with only traces of **3a** detected. The formation of EDA dimer was also severely attenuated, suggesting that carbene formation was significantly slower under these conditions. The P411<sub>BM3</sub>-UA variants had a substantial

loss in activity, but were still catalytically active (P411<sub>BM3</sub>-UA V87C: 140 ± 5 TTN, 95:5 dr; P411<sub>BM3</sub>-UA V87F: 25 ± 3 TTN, 6:94 dr). The loss of activity for globin variants but not P411<sub>BM3</sub> variants could be due to the globins' high affinity for gaseous ligands like O<sub>2</sub>, CO, and NO,<sup>48</sup> whereas some engineered P411<sub>BM3</sub> variants have previously been shown to function, albeit with attenuated catalytic activity, in aerobic conditions.<sup>15</sup> In addition to being tested as anaerobic and aerobic whole-cell catalysts, the final protein variants were tested as sonicated cell lysates and as purified proteins. In cell lysate, the proteins have decreased in activity relative to whole-cell catalysis (approximately 2-fold for *RmaNOD* Q52V and *ApePgb* AGW, 5-fold for P411<sub>BM3</sub> variants). The enzymes also had decreased activity as purified proteins (2-fold for *RmaNOD* Q52V, 5-fold for *ApePgb* AGW, approximately 30-fold for P411<sub>BM3</sub> variants). In both lysate and purified protein, the catalysts maintained their stereoselectivity. Decreased activity in lysate and purified protein relative to whole-cell reactions has been observed in many enzyme-catalyzed carbene transfer reactions<sup>34,35</sup> and is likely due to reduced carbene transfer-based enzyme inactivation in whole cells<sup>36</sup>. Whole cells could also contribute to the stabilization of the protein (through proper macromolecular crowding effects, chaperones, etc.). As our focus was on developing a straightforward and ready-to-use system for biocatalysis, we opted for whole-cell catalysis which also eliminates additional catalyst purification steps.

Table 2-4. Activity and selectivity of final protein variants as whole-cell, sonicated cell lysate, and purified protein for **3a** product formation. The reactions were performed on analytical scale (10 mM **2a**, direct addition of 20 mM **1**, 16-hour anaerobic reaction). Whole-cell catalyst loading was OD<sub>600</sub>= 5 (*ApePgb* AGW, *RmaNOD* Q52V) and OD<sub>600</sub>=20 (P411<sub>BM3</sub>-UA V87C, P411<sub>BM3</sub>-UA V87F). Cell lysates were diluted to the apparent OD<sub>600</sub>= 5 (*ApePgb* AGW, *RmaNOD* Q52V) and OD<sub>600</sub>=20 (P411<sub>BM3</sub>-UA V87C, P411<sub>BM3</sub>-UA V87F), corresponding to protein concentrations of 0.9 – 2.1 μM. Catalyst loading in purified protein reactions was 5 μM. The diastereoselectivity ratio (dr) is given as *cis:trans*.

Protein	Whole cell		Cell lysate		Purified protein	
	TTN	dr	TTN	dr	TTN	dr
<i>RmaNOD</i> Q52V	100 ± 6	< 1 : 99	43 ± 2	< 1 : 99	53 ± 9	< 1 : 99
<i>ApePgb</i> AGW	490 ± 20	89 : 11	190 ± 9	89 : 11	80 ± 10	87 : 13
P411 <sub>BM3</sub> -UA V87C	270 ± 30	95 : 5	56 ± 4	95 : 5	6.9 ± 0.9	95 : 5
P411 <sub>BM3</sub> -UA V87F	310 ± 20	4 : 96	54 ± 5	5 : 95	13 ± 2	4 : 96

While biocatalysts often possess very high selectivity, this selectivity can be synthetically limiting. A single enzyme may make only a single isomer, but access to other isomers may be equally important; natural diversity can be leveraged effectively for this challenge. A combination of natural diversity and directed evolution enabled the stereodivergent cyclopropanation of unactivated and electron-deficient alkenes in mild, aqueous conditions with a fully genetically encoded heme protein expressed in bacteria. This set of biocatalysts can serve as starting points for green, sustainable synthesis of valuable cyclopropanated products.

## **2.2 Structural characterization of engineered hemoproteins**

Following the engineering of a set of four heme proteins for stereodivergent cyclopropanation of unactivated and electron-deficient alkenes, we were interested in better understanding the structural and mechanistic determinants for their activity and stereoselectivity. One method to investigate structural determinants is through protein crystallography. We set out to obtain crystal structures for the four final biocatalysts, with the goal of seeing substrates, reactive intermediates, and products bound in the distal heme pocket.

### **2.3.1 RmaNOD crystallography**

There were no previous structures for *RmaNOD*, so we started our search for crystallization conditions with sparse-matrix screening.<sup>49</sup> These sparse-matrix screens are sets of 96 diverse conditions which have been designed to find initial crystallization conditions for most proteins of interest. During sparse-matrix screening for *RmaNOD* WT and *RmaNOD* Q52V crystallization, we found two conditions which gave good initial crystallization (0.5 M ammonium sulfate, 0.1 M sodium citrate tribasic dihydrate pH 5.6, 1.0 M lithium sulfate monohydrate and 0.2 M lithium sulfate, 0.1 M phosphate-citrate pH 4.2, 20 %(w/v) PEG 1000). Seeing that both conditions had high concentrations of lithium sulfate and acidic pH as primary components, a refinement tray was set up across which citrate buffer pH and lithium sulfate concentration were varied. Under these conditions, the initial *RmaNOD* WT and Q52V crystals formed within 24 hours into large, hexagonal crystals. Most of these

crystals attached to the plastic surface within the sitting drop crystal trays. As the crystals formed readily, instead of removing crystals from the surface, the subset of crystals which formed in solution were selected for cryoprotection and data collection on the Stanford Synchrotron Radiation Lightsource beamline 12-2 (SSRL 12-2). These crystals for *RmaNOD* WT and Q52V consistently diffracted with low-to-medium resolution (3–3.5 Å). Methods such as seeding, hanging drop (of sitting drop) trays, overlaying sitting drops on Fluorinert (to prevent adsorption to the well surface), and modifying cryoprotectant conditions were used to attempt to generate protein crystals with improved diffraction, but in each case the diffraction was the same or worse.

The presence of the C-terminal polyhistidine tag can have an effect on the packing and crystallization of a target protein, particularly in small proteins or when the tagged terminus would form a crystal contact.<sup>50</sup> Within the Arnold lab it was shown that even a few mutations to the small protein *Rma* cyt *c* led to a variant which was only found to crystallize after changing the location of the His expression tag. I therefore prepared a crystallization construct for *RmaNOD* Q52V where the protein was expressed with an N-terminal His<sub>6</sub>-SUMO tag. The SUMO protein tag is commonly used as an expression tag and, using the Ulp1 protease to cleave the SUMO tag after immobilized metal affinity chromatography (IMAC), the protein of interest is left with a single serine as an N-terminal scar.<sup>51</sup> The His-tagged SUMO protein and Ulp1 protease can then be removed through a second IMAC purification.

The tagless *RmaNOD* Q52V construct did not crystallize under the same conditions in which the C-terminally His-tagged crystallized, but a round of sparse-matrix screening followed by iterative refinement (Supplementary Information) led to crystallization conditions which yielded single crystals. These new crystals formed over several days and remained in solution rather than adsorbed to the well's surface. The diffraction of these proteins did improve relative to the His-tagged constructs, with crystals diffracting to approximately 2.5 Å.

Since no previous structure had been reported for *RmaNOD*, experimental phasing was chosen over potentially biasing the phase solution using a search model with low sequence

identity for molecular replacement. With the heme cofactor containing an iron atom, the structure could be solved using experimental phasing of diffraction data with anomalous signal. The data were therefore collected using single-wavelength anomalous dispersion (SAD) at the iron K edge (1.74 Å). Details on the software used for phase solution and structure refinement are given in the supporting information. The resulting crystal structure (PDB ID 6WK3) has a resolution of 2.45 Å and displays C 1 2 1 symmetry with four monomers in the asymmetric unit.

Initially, water was modeled at iron's sixth coordination site, but through further refinement, it was found that acetate fit the density better. Acetate has been found to bind to heme in crystallization conditions previously,<sup>52</sup> and the concentration of sodium acetate, the primary precipitant, is over 1.2 M in all of the refinement screening conditions used to crystallize *RmaNOD* Q52V. There were also copper ions modeled into the structure; these copper ions are likely crystallographic artifacts, as they form intermolecular contacts and all crystallization conditions for the tagless *RmaNOD* Q52V included a divalent metal ion.

With an experimental protein crystal structure in hand, we were interested in comparing the structure to the homology model. While the homology model originally used for *RmaNOD* is moderately accurate globally (1.568-Å RMSD), the structure of the heme distal pocket is dramatically different (Figure 2-10). The B-helix, which includes residues 21 – 38, is shifted more than 1.5 Å closer to the heme, reducing the active-site volume. Even more striking is the structure for the E helix; while the homology model predicted a short E-helix followed by a 9-residue random coil before the CD helix, the random coil between the E-helix is only two amino acids long. The longer E helix forms a side of the distal heme pocket, with multiple residues (including Q52V, the key mutation to enhance cyclopropanation activity and selectivity in *RmaNOD*) pointing directly into the active site. The residues which were incorrectly predicted to be part of a random coil in the homology model are highly likely to interact with substrates either as they enter the active site or during catalysis. Position Q52 was targeted based on sequence alignment with the mutated residue H64V in the sperm whale myoglobin engineered for carbene transferase activity;<sup>12</sup> this residue would have been

unlikely to be chosen when guided by only the homology model. This deviation between predicted and experimental structures highlights the importance of accurate structural models when selecting residues for site-saturation mutagenesis.

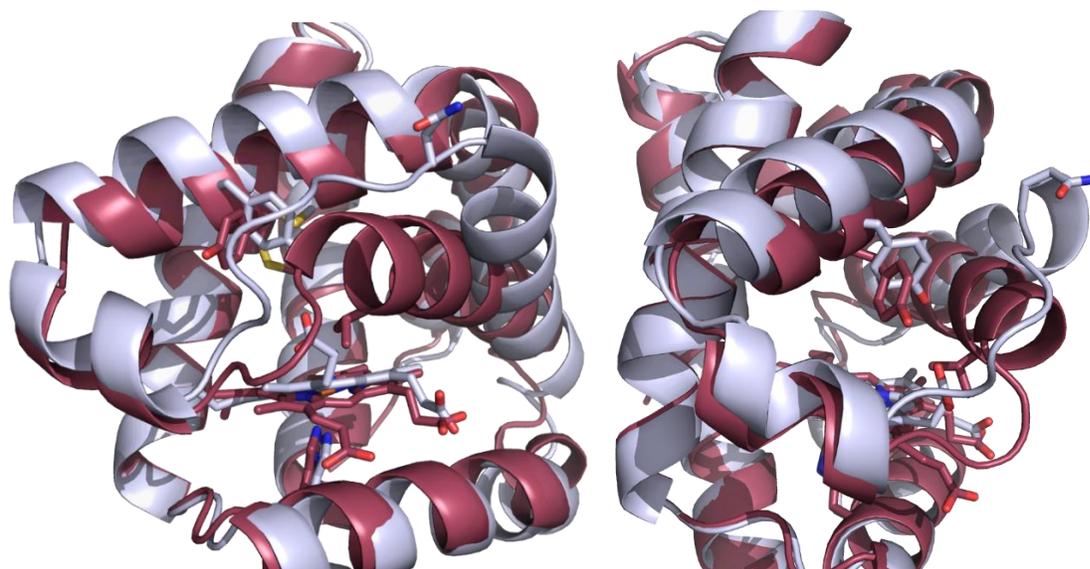


Figure 2-10. Differences in structure between the prepared homology model and the solved crystal structure. The homology model is shown in gray and the crystal structure is shown in red. The foreground displays the differing lengths of the E helix in the homology model and crystal structure.

One of the clearest methods to learn the origin of an enzyme's stereoselectivity is through observing the substrate or product bound within the active site. Getting structures of substrate- or product-bound proteins is done by either soaking, in which the protein is crystallized before being exposed to the substrate (e.g. put in a solution containing the small molecule of interest before cryoprotecting and freezing), or co-crystallization, in which the crystallization condition is prepared including the small molecule of interest. Soaking has the advantage that many different small molecules can be soaked into crystals with known structure and diffraction quality.<sup>53</sup> Soaking experiments are also necessary when the compound to be soaked into the crystal or a resultant intermediate compound are not stable over the time span required for the protein of interest to crystallize. However, if binding the substrate or product causes a substantial conformational shift in the protein, soaking a pre-

formed protein crystal might not capture that biochemically relevant conformational change.<sup>54</sup> Tight crystal packing or crystal contacts at an active-site entrance tunnel can also prevent small molecules from accessing the protein's active site. In contrast, co-crystallization can capture larger conformational changes and provide a snapshot which is more likely to be biochemically accurate, with the caveats that a protein with small-molecule induced conformational changes might not crystallize under the same conditions and that the addition of the small molecule (and solvents used to solubilize it) can disrupt protein crystallization.<sup>55</sup>

The degree of conformational change upon substrate binding to *RmaNOD* Q52V is unknown and the *RmaNOD* Q52V construct crystallized in large quantities without issue; co-crystallization was therefore chosen as the method of choice when attempting to observe substrate-bound *RmaNOD* Q52V. Co-crystallization trays for *RmaNOD* Q52V were prepared with added benzyl acrylate, a model substrate against which *RmaNOD* Q52V displayed 3700 TTN with >99% ee and dr. Benzyl acrylate is also more soluble in aqueous solution than 1-octene and 4-phenyl-1-butene, requiring the addition of less organic cosolvent to remain dissolved in the well solution. *RmaNOD* Q52V readily crystallized in the presence of benzyl acrylate, but data collected on these structures consistently showed an active site with no substrate bound.

The orientation in the active site of the ester moiety in the iron-carbenoid reactive intermediate, together with the alkene's angle-of-attack and substrate-binding mode, would dictate the stereoselectivity observed in the reaction. Iron-porphyrin carbenes (IPCs) have been studied experimentally and computationally,<sup>56,57</sup> and in 2018 there were two reported protein crystal structures containing IPCs. Hilvert and coworkers reported a long-lived reactive IPC derived from EDA using myoglobin with the non-canonical amino acid *N*-methyl-histidine as the heme axial ligand.<sup>58</sup> Arnold and coworkers reported an IPC derived from ethyl 2-diazopropanoate (Me-EDA) in an engineered *Rma* cyt *c* protein variant used in the construction of carbon–silicon bonds.<sup>31,59</sup>

Given the precedent of IPCs in protein crystal structures, we attempted to obtain a carbene-bound structure of *RmaNOD Q52V*. As EDA and the resultant IPC would not be stable for a week-long co-crystallization experiment, soaking experiments were used to add EDA to the crystals. Even at 4 °C, the addition of EDA and sodium dithionite to *RmaNOD Q52V* crystals resulted in the rapid formation of bubbles and the protein crystals were destroyed faster than they could be looped and flash frozen. While the enzyme activity is inhibited under aerobic conditions, sodium dithionite can act as an oxygen scavenger and reduce the dissolved oxygen content.<sup>60</sup> *RmaNOD Q52V* will also readily form EDA dimers (diethyl fumarate and diethyl maleate), consuming the iron-carbenoid intermediate. The carbene formation (and ostensibly EDA dimerization) in the presence of sodium dithionite was enough to preclude the use of this system to trap the EDA carbene intermediate; similar results were observed by Dr. Rusty Lewis when attempting to capture EDA in *Rma cyt c* variants (personal communication).

As a surrogate substrate, crystal soaking with Me-EDA was attempted in place of EDA. Me-EDA did not display activity for cyclopropanation with *RmaNOD Q52V*, but *RmaNOD WT* and variants have been reported to form carbon–silicon bonds with Me-EDA as the carbene precursor.<sup>61</sup> Me-EDA dimerization has not been observed in our hands, and the resultant iron-carbenoid species was stable for over an hour in the *Rma cyt c* enzyme system in which it was studied.<sup>59</sup> Although a structure with Me-EDA-derived carbene is not guaranteed to orient in the active site similar to the EDA-derived IPC would, a well-defined ester binding mode would be a strong indication of the EDA binding modality and provide insight into the origin of alkene cyclopropanation stereoselectivity. *RmaNOD Q52V* crystals were soaked with Me-EDA and sodium dithionite; unlike the EDA-soaked crystals, these crystals did not rapidly break down. There was no noticeable nitrogen evolution, but the lack of Me-EDA dimerization would make nitrogen evolution stoichiometric and limited by the protein concentration (compared to the catalytic EDA dimer formation). Unfortunately, as with *RmaNOD Q52V* crystals co-crystallized with benzyl acrylate, there was no electron density in the active site supporting occupancy of Me-EDA or a carbene derived from Me-EDA.

### 2.3.2 ApePgb AGW crystallography

*ApePgb* AGW had by far the broadest substrate scope of the four enzymes used in the stereodivergent cyclopropanation of unactivated alkenes (Table 2-10), and multiple variants of *ApePgb* have been found to be starting points for additional new-to-nature directed evolution projects in the lab. It was therefore of substantial interest to obtain a crystal structure for this scaffold to inform the design and construction of further *ApePgb* mutagenesis libraries. Like *RmaNOD*, no structures for *ApePgb* had been reported before, although *ApePgb* WT was previously heterologously expressed in *E. coli* and functionally characterized.<sup>62</sup> Only one member of the protoglobin family, *Methanosarcina acetivorans* protoglobin (*MacPgb*), has been crystallized.<sup>48,63</sup>

To begin the search for crystallization conditions of *ApePgb*, sparse-matrix screens were set up for C-terminally polyhistidine-tagged *ApePgb* WT and *ApePgb* AGW. These screens resulted in a small number of conditions in which phase separation or oil droplets were observed; the distinct red coloration of the heme-bound protein made these distinct from precipitant phase separation and protein precipitation. Following multiple refinement screens with no improvement to the crystallization, the crystallization construct was revisited. When reviewing the homology model based on *MacPgb*, we noted that the C-terminus of the protein is in the dimer interface; if *ApePgb* would crystallize with similar crystal packing, the C-terminal polyhistidine tag would disrupt that crystal contact (Figure 2-11). A crystallization construct of *ApePgb* AGW was prepared with the N-terminal polyhistidine-SUMO tag to prevent this potential crystal packing disruption.

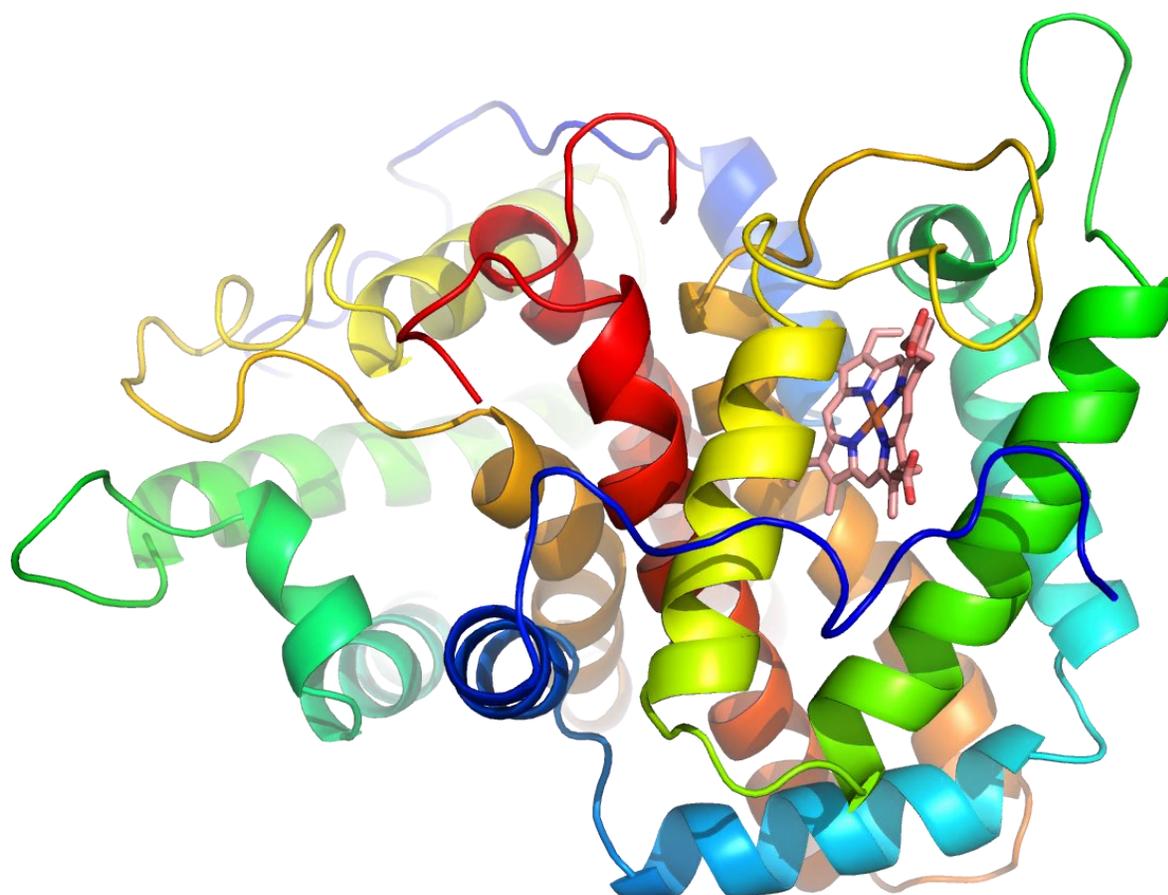


Figure 2-11. Homology model of *ApePgb* based on *MacPgb*. The two monomers in the crystallized dimer are colored as chainbows, with blue signifying the N-terminus and red signifying the C-terminus. The two C termini are in crystal contacts near the dimer interface; C-terminal expression tags are therefore likely to affect the protein's crystallization.

In sparse-matrix screens with the tagless *ApePgb* AGW construct, a set of what were potentially “sea urchin” crystals formed in a condition (Figure 2-12a; condition: 2.8 M 1,6-hexanediol, 0.1 M Tris pH 8.5, 0.2 M  $\text{MgCl}_2$ ). Refinement screening around these conditions had some effect on the urchin number, size, and rate of formation, but the quality of the crystals did not improve with variations in protein concentration, buffer pH, or precipitant concentration. Substrates (4-phenyl-1-butene and benzyl acrylate), substrate mimics (glycine ethyl ester) and heme-binding small-molecules (imidazole and thiocyanate) were also added

to determine whether co-crystallization could improve the crystal quality. In each of these cases there was no improvement in the crystal morphology.

A complementary technique to refinement screening which can be used to find improved protein crystallization conditions is random microseed matrix screening (rMMS).<sup>64,65</sup> In random microseed matrix experiments, like other seeding experiments, a seed stock is prepared by finely crushing crystals, which are then added to new crystallization experiments to provide nucleation sites for crystals to grow. In rMMS, the seed stocks are added to sparse-matrix screens, providing a very different chemical space from the space in which the initial crystals grew. When sparse-matrix screens were set up with *ApePgb* AGW purified protein and seed stock, microcrystals and small needles formed in 0.1 M sodium cacodylate trihydrate pH 6.5, 1.4 M sodium acetate trihydrate (Figure 2-12b).

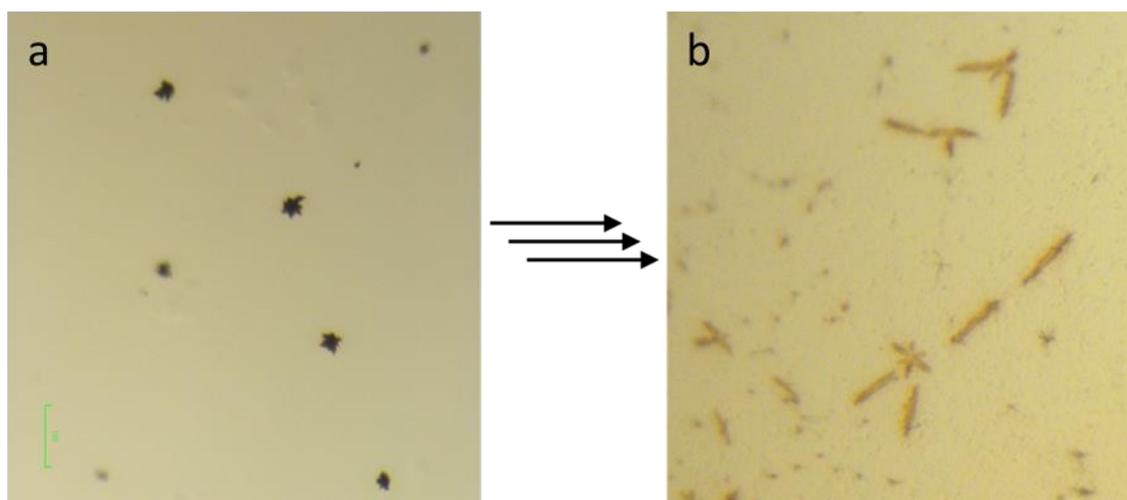


Figure 2-12. Results of random matrix microseeding in *ApePgb* AGW. (a) “sea urchin”-like initial protein crystals found during sparse-matrix screening and refinement. (b) small needles formed in 0.1 M sodium cacodylate trihydrate pH 6.5, 1.4 M sodium acetate trihydrate when microseeded with a crystal seed stock from the sea urchin crystals.

Varying seed stock concentrations, buffer pH, and precipitant concentrations did not lead to improvement in the crystal size or morphology. The Hampton additive screen was also tested under these conditions without observable improvement in crystal size or quality. One potential next step forward would be returning to the crystallization construct and truncating

the termini to remove flexible regions, introducing mutations at positions on the surface to improve the probability of crystallization, or using limited proteolysis techniques to remove flexible regions of the protein.<sup>66,67</sup> As the goal is to investigate structural determinants for the selectivity of particular *ApePgb* variants, truncating or mutating the protein in order to crystallize it would interfere with this investigation. The work on crystallizing *ApePgb* AGW was therefore halted; further attempts to crystallize other protein variants based on the *ApePgb* scaffold are ongoing in the Arnold lab.

### 2.3.3 P411<sub>BM3</sub>-UA crystallography

Although many have attempted to crystallize the full-length P450<sub>BM3</sub>, the closest X-ray crystallographic structure has been of the heme and FMN domain,<sup>68</sup> with the FAD domain crystallized separately.<sup>69</sup> Only recently has cryo-EM enabled the solution of the full-length P450<sub>BM3</sub> structure at low resolution (7.6 Å).<sup>70,71</sup> For structural insight into the effect of mutations on enzyme activity and selectivity, the P450<sub>BM3</sub> variant of interest is typically expressed and purified as a heme-domain only crystallization construct. These heme-domain crystallographic constructs were prepared for P411<sub>BM3</sub>-UA V87C and P411<sub>BM3</sub>-UA V87F. The heme-domain constructs were prepared with the N-terminal polyhistidine-SUMO cleavable tag, and the proteins were expressed and purified as N-His<sub>6</sub>-SUMO-P411<sub>BM3</sub> constructs, the SUMO tag was cleaved with Ulp1 protease to yield purified protein for tagless P411<sub>BM3</sub>-UA V87C and P411<sub>BM3</sub>-UA V87F.

A “superscreen” of conditions commonly found to crystallize P450<sub>BM3</sub> heme domain constructs had been developed by previous lab members (Supplementary Information, Table 2-13); this screen was used to find initial crystallization conditions for the two P411<sub>BM3</sub>-UA variants. Both proteins crystallized as pale green, thin, single needles surrounded by precipitate in 0.1 M Bis-Tris pH 5.5, 20% PEG 3350, 0.2 M Na(HCOO). Varying protein concentration, Bis-Tris buffer pH and PEG 3350 concentration decreased the level of precipitation and increased the crystal formation. Crystals of both P411<sub>BM3</sub>-UA variants were cryoprotected, flash frozen, and sent to SSRL for X-ray diffraction. While crystals from both proteins did diffract, the resolution was very poor (>6 Å).

## 2.4 Expanding enzymatic substrate scope to transfer trifluoromethyl-carbenes to activated, unactivated, and electron-deficient alkenes

The stereodivergent cyclopropanation of unactivated alkenes described in this chapter, like the majority of enzymatic cyclopropanation via carbene transfer,<sup>11–13,36</sup> was limited to the use of  $\alpha$ -diazoester carbene precursors. The use of carbene precursors without  $\alpha$ -ester substitutions would increase the scope of chiral cyclopropanes accessible through these engineered biocatalytic platforms. One alternative acceptor-only diazo compound class are  $\alpha$ -diazoperfluoroalkanes such as 2-diazo-1,1,1-trifluoroethane (DTFE), which can generate trifluoromethyl-substituted cyclopropanes through carbene-transfer reactions. Trifluoromethyl-substituted cyclopropanes have been used extensively in medicinal chemistry applications owing to the conformational rigidity conferred by the cyclopropane ring and the metabolic stability conferred by the trifluoromethyl substituent.<sup>72</sup>

Biocatalytic trifluoromethyl-substituted cyclopropane synthesis has been previously reported: Fasan and coworkers reported an elegant biocatalytic method to prepare trifluoromethyl-substituted cyclopropanes from vinylarenes and *ex situ*-generated DTFE with high enantio- and diastereoselectivity.<sup>20</sup> This example was limited to vinylarenes, however, and only *trans*-substituted cyclopropanes were synthesized. Diastereodivergent synthetic approaches to  $\alpha$ -trifluoromethylcarbene transfer of unactivated and electron-deficient alkenes would be useful for pharmaceutical and agrochemical compound development.

In parallel to my work on the cyclopropanation of unactivated and electron-deficient alkenes with EDA, a postdoctoral fellow in the laboratory, Dr. Xiongyi Huang, was investigating the use of DTFE in enzymatic carbene-transfer reactions as part of his work on  $\alpha$ -trifluoromethylcarbene-transfer reactions.<sup>40,41</sup> In his initial screening, he found that the lineage of *ApePgb* protein variants I had engineered displayed initial activity for cyclopropanation via DTFE carbene transfer and came to me for help with engineering the protein. We worked together on the construction and screening of additional single site-saturation mutagenesis libraries, resulting in the double mutants *ApePgb* W59L Y60Q

(*ApePgb* “LQ”) and *ApePgb* Y60G F73W (*ApePgb* “GW”) which were active against both styrenyl and electron-deficient alkenes (Figure 2-13). Notably, the *ApePgb* LQ variant formed the *cis* diastereomer with electron-deficient alkenes in moderate to good stereoselectivity; these *cis*-trifluoromethyl-substituted cyclopropanes are difficult to access with reported synthetic methods. Cyclopropanation of unactivated alkenes was also tested, but *ApePgb* LQ and GW only show trace activity for the previous model unactivated alkenes, 1-octene and 4-phenyl-1-butene.

	TTN	<i>cis</i> : <i>trans</i> or ( <i>Z</i> : <i>E</i> )		TTN	<i>cis</i> : <i>trans</i> or ( <i>Z</i> : <i>E</i> )
	1620	81 : 19	 methacrylates	3310	97 : 3
	2760	4 : 96			1850
	P/I = 4.2	95 : 5		P/I = 4.0	98 : 2
	P/I = 2.7	45 : 55		P/I = 0.8	71 : 29
	P/I = 1.3	80 : 20		P/I = 1.3	90 : 10
	P/I = 2.7	6 : 94		P/I = 4.7	6 : 94
	P/I = 0.7	86 : 14			
	P/I = 1.1	15 : 85			

Figure 2-13. Substrate scope of *ApePgb* LQ (red) and GW (blue) variants against activated and electron-deficient alkenes. GC-MS data reported were collected by Dr. Xiongyi Huang and Lucas Schaus.

Dr. Huang transferred the mutations from *ApePgb* to the protoglobin from the mesophilic methanogen *Methanosarcina acetivorans* (*MacPgb*), as *MacPgb* has been structurally characterized previously,<sup>48</sup> as he intended to perform molecular dynamics simulations that required an accurate crystal structure. When that initial transfer of mutations to *MacPgb* yielded variants with comparable activity and selectivity to the *ApePgb* variants, I became interested in how general the effect on carbene transfer activity and selectivity is when transferring these mutations to other proteins in the protoglobin fold class. Transferring mutations from one scaffold to another has been quite successful for promiscuous native<sup>73</sup> as

well as new-to-nature<sup>74</sup> reactions – though there is certainly a positive bias in the literature toward successful transfers of mutations. I hypothesized that the carbene-transfer activity observed in *ApePgb* and *MacPgb* protein variants would be seen in protoglobin homologs and that the homologous proteins would provide additional starting points for directed evolution of carbene transferases.

To build a set of homologous proteins, I used Protein-BLAST to search by protein sequence identity to assemble a set of representative proteins from the protoglobin fold class, which has been identified in archaea and bacteria. Of this list, proteins from thermophilic organisms were selected; proteins from thermophilic origins are (usually) more stable than their mesophilic orthologs, and this stability helps counteract the destabilizing nature of most mutations.<sup>75</sup> The amino-acid residues mutated for DTFE cyclopropanation (W59, Y60, and F73) were conserved in all homologous protoglobins found, as were most first-shell active-site residues which would be prime targets for site-saturation mutagenesis. I will refer to these homologous protoglobins in general as *XxxPgb*. I ordered nine additional protoglobin sequences as linear DNA fragments (gBlocks, IDT) codon-optimized for *E. coli* and containing the LQ mutations (Table 2-5), and the oligonucleotides required to generate the GW variants using *XxxPgb* LQ genes as templates. These *XxxPgb*s have between 51% and 83% pairwise amino-acid sequence identity (Figure 2-14).

Table 2-5. List of protoglobin proteins selected for the transfer of mutation experiment. The table lists the protoglobin names, the originating organism, and Uniprot ID or NCBI accession number. ApePgb, MacPgb, and PfePgb were ordered and cloned prior to this gene acquisition.

Protein	Organism	UniProt ID
<i>AauPgb</i>	<i>Acetothermus autotrophicum</i>	H5SUA1
<i>CthPgb</i>	<i>Crenotalea thermophila</i>	A0A1I7NC60
<i>ParPgb</i>	<i>Pyrobaculum arsenaticum</i>	A4WIC7
<i>PmePgb</i>	<i>Pyrinomonas methylaliphatogenes</i>	A0A0B6WXB4
<i>TamPgb</i>	<i>Thermus amyloliquefaciens</i>	WP_038057460.1 (NCBI)
<i>TarPgb</i>	<i>Thermus arciformis</i>	A0A1G7GW55
<i>TdaPgb</i>	<i>Thermanaerotherix daxensis</i>	A0A0P6XZU8
<i>ThuPgb</i>	<i>Thermoflexus hugenholtzii</i>	A0A212QV80
<i>TpePgb</i>	<i>Thermorudis peleae</i>	A0A1E5BNX2
<i>ApePgb</i>	<i>Aeropyrum pernix</i>	Q9YFF4
<i>MacPgb</i>	<i>Methanosarcina acetivorans</i>	Q8TLY9
<i>PfePgb</i>	<i>Pyrobaculum ferrireducens</i>	G7VHJ7

	ApePgb	ParPgb	PfePgb	CthPgb	TpePgb	TdaPgb	MaPgb	PmePgb	ThuPgb	AauPgb	TamPgb	TarPgb
ApePgb		62.83	64.58	51.81	58.25	58.64	56.99	60.31	61.86	60.1	59.9	59.59
ParPgb	62.83		82.72	54.74	51.83	54.21	53.68	54.45	57.07	58.12	58.12	61.26
PfePgb	64.58	82.72		54.97	56.25	59.69	59.16	59.38	60.94	61.46	61.46	61.46
CthPgb	51.81	54.74	54.97		59.9	62.63	64.58	57.51	61.14	60.94	58.12	59.38
TpePgb	58.25	51.83	56.25	59.9		62.69	60.82	64.62	60	61.14	63.4	59.49
TdaPgb	58.64	54.21	59.69	62.63	62.69		70.47	69.43	68.91	71.73	67.36	66.84
MaPgb	56.99	53.68	59.16	64.58	60.82	70.47		61.54	66.15	69.79	66.32	65.46
PmePgb	60.31	54.45	59.38	57.51	64.62	69.43	61.54		68.37	65.28	62.89	62.56
ThuPgb	61.86	57.07	60.94	61.14	60	68.91	66.15	68.37		70.98	65.98	66.15
AauPgb	60.1	58.12	61.46	60.94	61.14	71.73	69.79	65.28	70.98		69.27	72.02
TamPgb	59.9	58.12	61.46	58.12	63.4	67.36	66.32	62.89	65.98	69.27		82.99
TarPgb	59.59	61.26	61.46	59.38	59.49	66.84	65.46	62.56	66.15	72.02	82.99	

Figure 2-14. Protein identity matrix showing the amino-acid sequence identity correlations for the various XxxPgbs. The lowest and highest sequence identities are *CthPgb* – *ApePgb* (51.8%) and *PfePgb* – *ParPgb* (82.7%).

Together with a Master's student mentee, Lucas Schaus, I subcloned these new genes into the pET22b protein expression vector. The resulting XxxPgb LQ variant plasmids were used as PCR templates to generate the XxxPgb GW variants. *E. coli* competent cells were then transformed with the plasmids encoding both sets of protein variants to produce the proteins of interest. The biocatalysts were tested for their activity with DTFE against an array of activated, unactivated, and electron-deficient alkenes. Every successfully cloned XxxPgb

protein variant displayed carbene-transfer activity for most of the alkenyl substrates tested. While *ApePgb* variants showed only trace activity the vinyl Weinreb amide substrate *N*-methoxy-*N*-methylacrylamide, variants of *TarPgb* and *TdaPgb* had high activity and diastereoselectivity against this substrate (Table 2-6). Weinreb amides are useful substrates as the corresponding products can be easily derivatized to generate myriad ketone-substituted cyclopropane products. These *TarPgb* and *TdaPgb* variants are therefore excellent starting points for directed evolution for biocatalytic synthesis and derivatizations to synthesize diverse trifluoromethyl-cyclopropylketones. Expansion of this reaction scope to include 2-substituted *N*-methoxy-*N*-methylacrylamides would enable the synthesis of trifluoromethyl-substituted cyclopropylketones with stereogenic quaternary carbons (Figure 2-15). Following his Master's thesis work, Lucas joined the Arnold laboratory as a graduate student and has continued this work.

Table 2-6. Activity and diastereoselectivity of *TarPgb* and *TdaPgb* wild type and variants in the synthesis of *N*-methoxy-*N*-methyl-2-(trifluoromethyl)cyclopropane-1-carboxamide. Activities were determined via NMR, using fluorobenzene as an internal standard. NMR data were collected by Lucas Schaus.

Protoglobin	Yield	TTN	dr
<i>TarPgb</i> WT	21.5 %	600	36.3 %
<i>TarPgb</i> GW	21.8 %	2000	96.0 %
<i>TdaPgb</i> WT	17.6 %	1200	38.4 %
<i>TdaPgb</i> LQ	20.5 %	1800	93.9 %

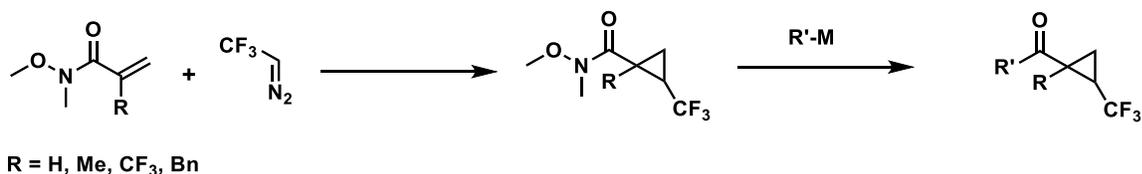


Figure 2-15. Synthesis of Weinreb amide-substituted cyclopropanes. Carbene transfer to substituted *N*-methoxy-*N*-methylacrylamide forms a cyclopropane with a stereogenic quaternary carbon center. This compound can be further derivatized with Weinreb ketone synthesis to generate myriad carbonyl-substituted cyclopropanes.

## 2.5 Conclusions

Prior to the work described in this Chapter, cyclopropanation via carbene transfer using iron-porphyrin-binding proteins had been limited to styrenyl alkenes. Scientists demonstrated that heme proteins that bound a noble-metal substituted porphyrin in place of heme could transfer carbenes to unactivated alkenes,<sup>27</sup> and an iron porphyrin catalyst had been shown to perform the reaction in organic solvent.<sup>26</sup> I found that heme proteins could catalyze the cyclopropanation of unactivated and electron-deficient alkenes, and that their activity and selectivity could be improved through just 1–3 rounds of site-saturation mutagenesis of active-site residues. The four engineered protein variants enabled access to all four stereoisomers of unactivated and electron-deficient alkenes. I solved the crystal structure of one of these variants, *RmaNOD* Q52V; I used this structure to guide library design for *RmaNOD* variants engineered in Chapter 3. The *ApePgb* scaffold was found to catalyze trifluoromethylcarbene transfer to activated and electron-deficient alkenes, and the activating mutations could be successfully transferred into homologous protoglobins. This work and other recent research on enzymatic carbene- and nitrene-transfer reactions showcase the breadth of reactions that iron-porphyrin proteins can catalyze, and I anticipate that the heme protein biocatalytic platform will continue to expand in the near future.

## 2.6 Supplementary information for Chapter 2

### Nomenclature for compound labeling

Alkenyl substrates are named **2x** (where x is a-m). The corresponding cyclopropyl esters (from reaction with ethyl diazoacetate **1**) are named **3x**. Single diastereomer compounds are named *cis*-**3x** or *trans*-**3x**. The 1-octene cyclopropane products, whose absolute configuration is known (see *Compound chiral separation conditions and representative traces*), are labeled (*1(R/S)*, *2(R/S)*)-**3a**. Alkenyl substrates with functional groups protected (e.g. 7-octen-1-ol, 7-octen-1-oic acid) are named **2xa**, and their corresponding cyclopropane products are named **3xa**. These are depicted in Figure 2-16.

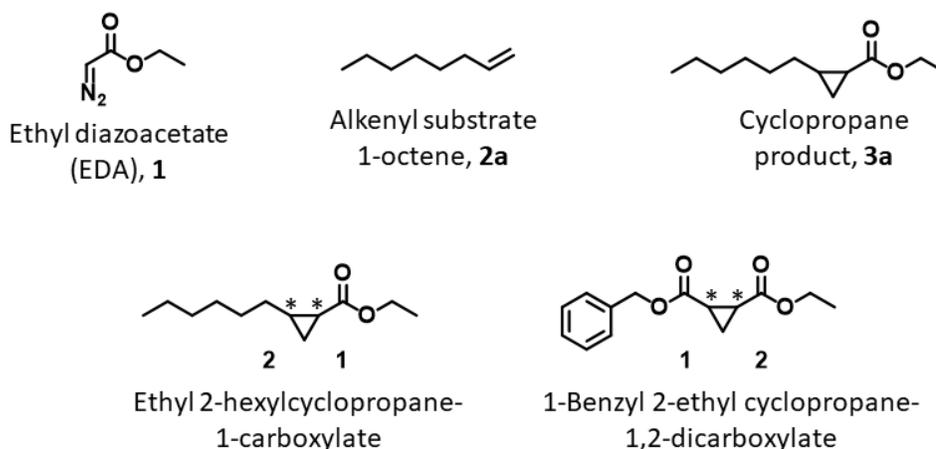


Figure 2-16. Compound nomenclature used in this work.

## Materials and Methods

Solvents and reagents were ordered from Sigma Aldrich, TCI, CombiBlocks, or Alfa Aesar and used without further purification. GC-FID data were collected on a Shimadzu GC-17A, Agilent 6850 GC system, and Agilent 7820A GC system. GC-MS data were collected on a Shimadzu GCMS-QP2010 SE. Screening HPLC-UV data were taken on an Agilent 1200 series HPLC. Normal-phase chiral HPLC data were taken on an Agilent 1100 series HPLC. NMR spectra were recorded on a Bruker Prodigy 400 MHz instrument or Varian 300 MHz instrument with  $\text{CDCl}_3$  as solvent.  $^1\text{H}$  NMR spectra were recorded at 400 MHz and  $^{13}\text{C}$  NMR spectra were recorded at 100 MHz. Chemical shifts were normalized to the chloroform solvent's protio impurity ( $^1\text{H}$  NMR 7.26 ppm,  $^{13}\text{C}$  NMR 77.16 ppm). Optical rotation data were collected on a JASCO P-2000 Polarimeter.

## Proteins tested in enzyme discovery

Genes encoding eleven heme-binding proteins were ordered as codon-optimized gBlocks (Integrated DNA Technologies, Coralville, Iowa) and assembled into pET22b(+) with the pelB leader sequence removed and a C-terminal 6xHis tag. As the putative distal ligand could interfere with the initially low substrate binding affinities, the proteins were ordered with the distal axial ligands mutated to smaller, nonpolar residues, found to be beneficial for

vinylarene cyclopropanation in myoglobin<sup>24</sup>. Some gBlocks were also ordered with a mutation in a putative entrance tunnel residue previously found to enhance styrene cyclopropanation activity in myoglobin.<sup>24</sup> These heme proteins were first tested for cyclopropanation activity using 1,7-octadiene and EDA as substrates (1,7-octadiene was chosen for its higher effective concentration of terminal olefin). The most active and selective proteins identified from these experiments, protoglobin from *Aeropyrum pernix* (ApePgb) and nitric oxide dioxygenase from *Rhodothermus marinus* (RmaNOD), were subsequently tested as wild-type proteins against 1-octene **2a**, a commonly used model substrate for unactivated alkene cyclopropanation studies.

Table 2-7. Heme-binding proteins tested for unactivated alkene cyclopropanation activity using 1,7-octadiene and EDA as substrates.

UniProt ID	Organism	Annotation	Mutation(s) from WT	Cyclopropane product formation detected
Q3IDI7	<i>Pseudoalteromonas haloplanktis</i>	Putative hemoglobin-like oxygen-binding protein	Y42V F69A	No
Q7CX73	<i>Agrobacterium fabrum</i>	Uncharacterized protein	Y26V F53A	No
Q5L1S0	<i>Geobacillus kaustophilus</i>	Hypothetical conserved protein	Y29V Q50A	No
Q9NPG2	<i>Homo sapiens</i>	Neuroglobin	F28V F61I H64A	No
Q0PB48	<i>Campylobacter jejuni</i>	Truncated hemoglobin	none	No
B3DVC3	<i>Methylobacterium inferorum</i>	Hemoglobin IV	H71V L93A	Yes
D0MGT2	<i>Rhodothermus marinus</i>	Nitric oxide dioxygenase	Y32V Q52A	Yes
G7VHJ7	<i>Pyrobaculum ferrireducens</i>	Protoglobin	Y58V	Yes
Q9YFF4	<i>Aeropyrum pernix</i>	Protoglobin	Y60V	Yes
O66586	<i>Aquifex aeolicus</i>	Thermoglobin	Y29V Q50A	Yes

### Screening previously engineered cytochromes P411<sub>BM3</sub> for cyclopropanation activity

A composite plate of 36 P411<sub>BM3</sub>-CIS proteins from lineages engineered for non-natural reactions such as aziridination, sulfimidation, amination, and non-styrenyl, activated alkene cyclopropanation<sup>14</sup> was screened for activity and selectivity in **3a** product formation. Most reactions showed trace activity and moderate diastereoselectivity, but the highest activities by a large margin were found in the lineage engineered for *N*-vinyl amide cyclopropanation.<sup>36</sup> Screening this lineage for stereoselectivity showed that P411<sub>BM3</sub>-CIS L437F T438Q L75Y L181I, referred to in this study as P411<sub>BM3</sub>-UA, had the highest activity and enantioselectivity. It was therefore used as a starting point in unactivated alkene cyclopropanation.

### Homology models

*ApePgb* was modeled with the *Methanosarcina acetivorans* protoglobin (PDB ID: 3ZJL) and *RmaNOD* was modeled with *Alcaligenes eutrophus* flavohemoglobin (PDB ID: 1CQX) using SWISS-MODEL.<sup>76</sup> A homology model of P411<sub>BM3</sub>-UA was generated through side-chain mutations to the P411<sub>BM3</sub>-CIS crystal structure (PDB ID: 4H23). Figures generated from protein homology models and crystal structures were made with PyMOL (Schrödinger, Inc.).

### Subcloning and transformation of genes

Genes for *Aeropyrum pernix* protoglobin (*ApePgb*) and *Rhodothermus marinus* nitric oxide dioxygenase (*RmaNOD*) were ordered as codon-optimized gBlocks (Integrated DNA Technologies, Coralville, Iowa). The gBlocks were amplified via polymerase chain reaction (PCR), and the PCR products were gel extracted and purified with Zymoclean Gel DNA Recovery Kit (Zymo Research Corp, Irvine, CA). The PCR product was subcloned into pET22b(+) via Gibson assembly.<sup>77</sup> Electrocompetent *E. coli* EXPRESS BL21(DE3) cells (Lucigen, Middleton, WI) were transformed with Gibson assembly products using a Gene Pulser Xcell (Bio-Rad, Hercules, CA). Aliquots of SOC medium (750  $\mu$ L) were added, and the cells were incubated at 37 °C and 230 rpm for 45 minutes before being plated on LB-ampicillin (100  $\mu$ g mL<sup>-1</sup>) agar plates. Overnight cultures (5 mL LB-amp in culture tubes)

were grown at 37 °C and 230 rpm for 12–18 hours. Overnight cultures were used to inoculate flask cultures, prepare glycerol stocks, and isolate plasmids. Plasmids were isolated with Qiagen Miniprep kits, and the genes were sequence verified (T7 promoter/terminator sequencing primers, Laragen, Inc.).

### **Medium-scale protein expression**

Cultures of Hyperbroth (HB, AthenaES) with 100  $\mu\text{g mL}^{-1}$  ampicillin in unbaffled Erlenmeyer flasks were inoculated 1% (v/v) with stationary-phase overnight cultures and shaken in an Innova 42 shaker at 230 rpm, 37 °C. At  $\text{OD}_{600} = 0.8$ , cultures were chilled on ice for 20 minutes. Protein expression was induced with 0.5 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) and heme production was enhanced with supplementation of 1 mM 5-aminolevulinic acid (5-ALA). The cultures were shaken at 180 rpm and 22 °C overnight (18–24 hours). Cells were pelleted via centrifugation at 4000 $\times$ g for 10 minutes at 4 °C. The supernate was decanted, and the cells were resuspended in M9-N buffer supplemented with 25 mM glucose.

### **Site-saturation library construction**

Site-saturation mutagenesis was performed using the 22-codon method.<sup>78</sup> Briefly, oligonucleotides were ordered with NDT, VHG, and TGG codons in the coding strand at the amino acid position to be saturated. A reverse primer complementary to all three forward primers was also ordered. Two PCRs were performed for each library, the first containing a mixture of forward primers (12:9:1 NDT:VHG:TGG) and a pET22b(+) internal reverse primer and the second containing the complementary reverse primer and a pET22b(+) internal forward primer. The two PCR products were gel-purified with Zymoclean Gel DNA Recovery Kit (Zymo Research Corp, Irvine, CA) and ligated via Gibson assembly. Electrocompetent *E. coli* EXPRESS BL21(DE3) cells (Lucigen, Middleton, WI) were transformed with the Gibson assembly product. Aliquots of SOC medium (750  $\mu\text{L}$ ) were added, and the cells were incubated at 37 °C for 45 minutes before being plated on LB-ampicillin (100  $\mu\text{g mL}^{-1}$ ) agar plates.

### Site-saturation library expression

Single colonies from the LB-ampicillin agar plates were picked using sterile toothpicks and grown in 300  $\mu$ L LB-ampicillin in 2-mL 96-well deep-well plates at 37 °C, 250 rpm, 80% humidity overnight (12–18 hours). Multichannel pipettes were used to transfer 30  $\mu$ L of starter culture into deep-well plates containing 1 mL HB-amp per well. Glycerol stocks of these plates were prepared in parallel by adding starter culture (100  $\mu$ L) and 50% (v/v) sterile glycerol (100  $\mu$ L) to a 96-well microplate, which was then stored at –80°C. The deep-well expression culture plate was incubated at 37 °C, 250 rpm, 80% humidity for 2.5 hours. The plate was then chilled on ice for 30 minutes. The cultures were induced with 0.5 mM IPTG and supplemented with 1 mM 5-ALA to increase cellular heme production. The plate was incubated at 22 °C and 250 rpm overnight. The plate was centrifuged at 4000 $\times$ g for 10 minutes at 4 °C.

### Site-saturation library reactions and screening

The pellets in the site-saturation library deep-well plates were resuspended in nitrogen-free M9 minimal medium (47.7 mM Na<sub>2</sub>HPO<sub>4</sub>, 22.0 mM KH<sub>2</sub>PO<sub>4</sub>, 8.6 mM NaCl, 2.0 mM MgSO<sub>4</sub>, and 0.1 mM CaCl<sub>2</sub>, abbreviated as M9-N, 400  $\mu$ L). In an anaerobic chamber, 50  $\mu$ L reactant mixture in ethanol (final concentrations in a 450  $\mu$ L reaction: 20 mM EDA **1**, 20 mM 4-phenyl-1-butene **2b**) were added to the reaction plate. The reaction plate was covered with a pierceable foil cover (USA Scientific) and shaken at 500 rpm for 3 hours. To quench the reaction and extract the substrates, 400  $\mu$ L of a mixture of acetonitrile (49 mL) and 3 M HCl (1 mL) were added to each well. The reaction plate was shaken for an additional 30 minutes, followed by centrifugation (4000 $\times$ g, 10 minutes, 4 °C). The supernate was filtered through a 0.2  $\mu$ m PTFE 96-well filter plate into a 96-well microplate (4000 $\times$ g, 1 minute, RT). The microplate was sealed with a pierceable foil cover. The wells were screened for activity and diastereoselectivity of **3b** formation via HPLC using a Kromasil 100-5-C18 column, 4.6 $\times$ 50 mm with a 71% acetonitrile isocratic method (3 minutes). In later screening with higher enzymatic activity, the separation of *cis*- and *trans*-isomers of **3b** was improved with the use of an Eclipse XDB-C18 column, 5  $\mu$ m particle size, 4.6 $\times$ 150 mm and a 6-minute 71% acetonitrile isocratic method. Wells with improved activity relative to the parent protein

were streaked out from the glycerol stock onto LB-amp plates. A single colony was picked and grown in 5 mL LB-amp overnight (230 rpm, 37 °C). These overnight cultures were used in flask protein expression and small-scale biocatalytic reactions to verify enhanced activity and/or selectivity relative to the parent sequence.

### **Sonicated lysate preparation and hemochrome assay**

An aliquot of cells for protein concentration determination was sonicated (QSonica Q500 Sonicator, 1/8 in tip) for 2 minutes, 1 second on, 1 second off at 25% amplitude. Cells expressing *ApePgb* or *RmaNOD* variants had 0.1 volume equivalents Bugbuster 10X protein extraction reagent (EMD Millipore) added prior to sonication. The sonicated lysate was clarified via centrifugation at 4500×g and 4 °C for 10 minutes. The concentration of heme-loaded protein was determined with the pyridine hemochromagen (hemochrome) assay.<sup>79</sup> Briefly, sonicated and clarified lysate (500 μL) was sterile filtered and added to a cuvette. Solution I (500 μL, 0.2 M NaOH, 40% (v/v) pyridine, 500 μM potassium ferricyanide) was added, and the spectrum of this oxidized sample was taken from 350–600 nm. Sodium dithionite (10 μL of 0.5 M solution in 0.5 M NaOH) was added, and the reduced spectrum was taken from 350–600 nm. The pyridine hemochromagen concentration was determined using its Q bands, with extinction coefficient 23.98 mM<sup>-1</sup> cm<sup>-1</sup> for (557 nm<sub>reduced</sub> – 540 nm<sub>oxidized</sub>).<sup>79</sup>

### **Small-scale, whole-cell biocatalytic reaction preparation and work-up**

Small-scale reactions were set up in 2-mL GC crimp vials. *E. coli* expressing the appropriate heme protein catalyst (380 μL, adjusted to the appropriate optical density or protein concentration) were added to the vials, and they were brought into a Coy anaerobic chamber (~ 0-10 ppm O<sub>2</sub>). To each vial was added alkene (final concentration 10 mM) followed by EDA **1** (final concentration 20 mM) with 5% ethanol as a cosolvent. Directly following addition of EDA, the reaction vial was crimped and shaken at 500 rpm at RT for 16 hours. Reactions were worked up by the addition of HCl (16 μL, 3 M stock) and internal standard (16 μL of 40 mM acetophenone in cyclohexane). Cyclohexane (700 μL) was added and the reaction was transferred into 1.7-mL Eppendorf tubes for extraction. The extraction was

carried out with a Retsch MM 301 mixing mill (1 minute, 30 Hz / 1800 rpm). Samples were centrifuged at 20000×g for 5 minutes at RT, and the organic layer was used for chromatographic analysis.

Table 2-8. Activity and selectivity of hemin controls, wild-type protein, and engineering lineage intermediate proteins in **3a** product formation. Mean and standard deviation were determined from two biological replicates of technical duplicates. The reactions were performed on analytical scale (10 mM **2a**, direct addition of 20 mM **1**, 16-hour anaerobic reaction). n.d.: not determined. The diastereoselectivity ratio (dr) is given as *cis:trans* and the enantiomeric excess (ee) is given for the major diastereomer.

Catalyst	TTN	dr	ee (major)
Hemin	0.39 ± 0.04	25 : 75	n.d.
Hemin + BSA	0.41 ± 0.03	23 : 77	n.d.
<i>RmaNOD</i> WT	27 ± 9	3 : 97	89%
<i>ApePgb</i> WT	18 ± 2	81 : 19	69%
<i>ApePgb</i> Y60G	140 ± 40	82 : 18	96%
<i>ApePgb</i> W59A Y60G	360 ± 90	85 : 15	96%
P411 <sub>BM3</sub> -UA	500 ± 130	89 : 11	94%

Table 2-9. Activity and selectivity of wild-type and engineering lineage intermediate proteins in **3b** product formation. Mean and standard deviation were determined from two biological replicates of technical duplicates. The reactions were performed on analytical scale (10 mM **2b**, direct addition of 20 mM **1**, 16-hour anaerobic reaction). The diastereoselectivity ratio (dr) is given as *cis:trans* and the enantiomeric excess (ee) is given for the major diastereomer.

Protein	TTN	dr	ee (major)
<i>RmaNOD</i> WT	34 ± 4	17 : 83	71%
<i>ApePgb</i> WT	85 ± 3	88 : 12	81%
<i>ApePgb</i> Y60G	125 ± 3	83 : 17	94%
<i>ApePgb</i> W59A Y60G	570 ± 50	80 : 20	94%
P411 <sub>BM3</sub> -UA	2500 ± 130	91 : 9	>99%

Table 2-10. Substrate scope and diastereoselectivity of the four final variants. Activities were confirmed via GC-MS. +: activity against the substrate. \*: trace activity detected. -: no detectable activity. n.d.: not determined. Diastereoselectivity ratio is given as *cis:trans*. Variants with high activity were characterized further in analytical-scale (Figure 2-8) and preparative-scale (Figure 2-9) reactions.

Alkene, product	P411 <sub>BM3</sub> -UA V87C	<i>ApePgb</i> AGW	P411 <sub>BM3</sub> -UA V87F	<i>RmaNOD</i> Q52V
1-octene ( <b>3a</b> )	+, 95:5 d.r.	+, 89:11 d.r.	+, 4:96 d.r.	+, <1:99 d.r.
4-phenyl-1-butene ( <b>3b</b> )	+, 96:4 d.r.	+, 84:16 d.r.	+, <1:99 d.r.	+, 3:97 d.r.
benzyl acrylate ( <b>3c</b> )	+, 91:9 d.r.	+, 71:29 d.r.	+, 2:98 d.r.	+, <1:99 d.r.
6-bromo-1-hexene ( <b>3d</b> )	+, 92:8 d.r.	+, 92:8 d.r.	+, 4:96 d.r.	+, <1:99 d.r.
vinyl cyclohexane ( <b>3e</b> )	*	+, 76:24 d.r.	*	*
methylenecyclohexane ( <b>3f</b> )	*	+	+	*
1-penten-3-one ( <b>3g</b> )	*	+, 89:11 d.r.	*	+, <1:99 d.r.
2-vinylpyridine ( <b>3h</b> )	*	*	*	+, <1:99 d.r.
7-octen-1-ol ( <b>3i</b> )	+	+, 91:9 d.r.	*	*
7-octen-1-oic acid ( <b>3j</b> )	n.d.	+, 83:17 d.r.	n.d.	n.d.
( <i>E</i> )-penta-1,3-diene ( <b>3k</b> )	+, 96:4 d.r.	+, 91:9 d.r.	+, 31:69 d.r.	+, 4:96 d.r.
( <i>Z</i> )-penta-1,3-diene ( <b>3l</b> )	+, 55:45 d.r.	+, 93:7 d.r.	+, 35:65 d.r.	+, 35:65 d.r.
5-hexen-2-one ( <b>3m</b> )	-	+, 94:6 d.r.	-	*

### Preparative-scale, whole-cell biocatalytic reaction setup and product purification

Whole cells resuspended in M9-N supplemented with 25 mM glucose were brought into a Coy anaerobic chamber. Whole-cell catalyst was added to unbaffled Erlenmeyer flasks, followed by alkene (5 mM, 1.0 eq.) and EDA (10 mM, 2.0 eq.) diluted in ethanol (5% final ethanol cosolvent). The reactions were sealed and shaken at room temperature at 180 rpm for 16 hours. The product was extracted three times from the aqueous reaction mixture with 1 volume eq. of 2:1 pentane:diethyl ether. The organic layer was dried with sodium sulfate and concentrated via rotary evaporation. The concentrated reaction mixture was then purified via flash chromatography (Biotage, Inc.). Pentane:diethyl ether gradients were generally more effective at separating the cyclopropane products from the EDA dimer byproducts. Fractions containing cyclopropanation products were pooled and concentrated *in vacuo*.

Table 2-11. Conditions for the preparative-scale reactions reported in the chapter. **3k** and **3l** were run at higher reaction volume due to the lower molecular weight of the substrates (and corresponding products).

Product	Protein variant used	Reaction volume (mL)	Whole-cell OD <sub>600</sub>
<b>3i</b>	<i>ApePgb</i> AGW	40	8
<b>3j</b>	<i>ApePgb</i> AGW	40	8
<b>3k</b>	<i>RmaNOD</i> Q52V	80	10
<b>3l</b>	<i>ApePgb</i> AGW	80	10
<b>3m</b>	<i>ApePgb</i> AGW	40	8

### Large-scale protein expression and purification

HB-amp (1 L in 2.8 L unbaffled flask) was inoculated with 1% (v/v) overnight culture and shaken at 37 °C and 160 rpm. At OD<sub>600</sub> = 1.3–1.5 the flasks were chilled on ice for 20 minutes. Protein expression was then induced with a final concentration of 0.5 mM IPTG and 1 mM 5-ALA, and the cultures were grown for 24 hours at 22 °C and 140 rpm. The cultures were pelleted (4000×g, 5 minutes, 4 °C) and frozen at -20 °C. The cells were resuspended in binding buffer (25 mM Tris HCl pH 7.5, 100 mM NaCl, 25 mM imidazole). Hemin (1 mg gwet cells<sup>-1</sup>) was added to increase heme loading of the protein, and DNase I (0.1 mg mL<sup>-1</sup>) was added to reduce lysate viscosity. The cells were sonicated, the lysate was clarified via centrifugation (20000×g, 20 minutes, 4 °C), and the clarified lysate was filtered (0.45 μm sterile filter). The protein was purified via HisTrap (1 mL column) on an ÄktaPurifier (GE Healthcare Life Sciences), using a 25–300 mM imidazole gradient over 10 column volumes. Fractions containing the protein of interest were pooled and buffer exchanged via centrifugal concentration to a 25 mM Tris HCl pH 7.5, 25 mM NaCl. The buffer-exchanged protein was flash-frozen and stored in 25 μL aliquots at -80 °C. The purified protein concentration was determined using the bicinchoninic acid assay (BCA assay, Thermo Scientific) using the standard protocol provided.

### Cloning, expression, and purification for crystallography

DNA encoding *RmaNOD* Q52V, *ApePgb* AGW, P411<sub>BM3</sub>-UA V87C, and P411<sub>BM3</sub>-UA V87F were subcloned into a pET22b vector encoding an N-terminally His-tagged SUMO fusion through PCR amplification and Gibson assembly<sup>77</sup> to generate “pSUMO” constructs of each variant. The P411<sub>BM3</sub>-UA variants were cloned into pSUMO as the heme domain only by truncation to residue 463. BL21(DE3) *E. coli* cells (Lucigen) were

transformed with the pSUMO constructs. Each was then expressed using the following method. A single colony was used to start an overnight culture in LB-amp, which was grown at 37 °C and 230 rpm. HB-amp (1 L in an unbaffled 2.8-L flask) was inoculated with 1% (v/v) overnight culture and shaken at 37 °C and 160 rpm (NFors shaker). At  $OD_{600} = 1.0$ , the flask was put on ice for 30 minutes, induced with final concentration of 0.5 mM IPTG, and supplemented with a final concentration of 1 mM 5-ALA. The culture was shaken overnight at 22 °C and 160 rpm. Following expression, the culture was centrifuged at  $4000 \times g$  for 10 minutes at 4 °C and frozen at -20 °C.

The pellet was thawed on ice and resuspended in resuspension buffer (25 mM Tris pH 7, 100 mM NaCl). The cells were lysed via sonication (QSonica Q500 Sonicator). The lysate was clarified via centrifugation (15 minutes,  $20817 \times g$ ) and filtered through 0.22- $\mu$ m filters. The protein was purified on an Äkta purifier (GE Healthcare) with a 1-mL HisTrap column running a linear gradient between resuspension buffer (25 mM Tris pH 7, 100 mM NaCl, 20 mM imidazole) and elution buffer (25 mM Tris pH 7, 100 mM NaCl, 500 mM imidazole). Fractions containing protein of interest were pooled and buffer exchanged into 25 mM Tris pH 7, 25 mM NaCl, 5 mM DTT. His-tagged Ulp1 protease (a gift from Christopher Bley, Hoelz lab Caltech) was added, and the proteolysis was allowed to proceed overnight. Following proteolysis, the sample was passed through two hand columns with Nickel-NTA agarose bulk resin (Qiagen) to remove the cleaved His-tagged SUMO and His-tagged Ulp1 protease. The flow-through solution containing the protein of interest was buffer exchanged and concentrated in 25 mM Tris pH 7, 25 mM NaCl buffer, and flash-frozen in aliquots.

### **Crystallization of tagless RmaNOD Q52V**

Sparse-matrix screens were set up using 0.2  $\mu$ L protein and 0.2  $\mu$ L well solution with a Gryphon robotic platform (Art Robbins Instruments). Quasicrystals were found in Crystal HT G10 (Hampton Research, 50 mM Cd(II)SO<sub>4</sub>, 0.1 M HEPES pH 7.5, 1.0 M sodium acetate trihydrate). Testing other divalent metal salts resulted in Cu(II)SO<sub>4</sub> as an equivalent replacement for the cadmium salt. Refining from this condition using 24-well sitting-drop well plates (Hampton Research), 2  $\mu$ L protein solution and 2  $\mu$ L well solution yielded

crystals in approximately one week in 40 mM Cu(II)SO<sub>4</sub>, 0.1M HEPES pH 7.6, 1.2 M sodium acetate trihydrate. These crystals were cryoprotected with well solution supplemented with glycerol (increased by steps of 5% sequentially up to 25% final concentration glycerol) and flash-frozen in liquid nitrogen.

### **Data collection and refinement**

Diffraction data were collected on the Stanford Synchrotron Radiation Laboratory Beamline 12-2. As there are no previously reported structures for the *RmaNOD* scaffold, the data were collected as single anomalous diffraction using the iron K edge (1.74 Å). Data were processed using XDS<sup>80</sup>, and three data sets were merged to increase anomalous signal redundancy. The data were used with AutoSol<sup>81</sup> for experimental phasing and model building. Following model building and initial refinements in PHENIX,<sup>82</sup> chain A of this model was used as a search model. The highest quality data set was determined in AIMLESS<sup>83</sup> to have a high-resolution shell limit of 2.45 Å. Molecular replacement was performed using Phaser-MR<sup>84</sup> with the merged data set model as the search model. The model was iteratively refined using PHENIX refine and COOT<sup>85</sup> and validated with MolProbity. Crystallographic and model statistics are described in Table 2-12.

Table 2-12. Crystal data and refinement statistics for *RmaNOD* Q52V.

Structure name	Engineered carbene transferase <i>RmaNOD</i> Q52V, putative nitric oxide dioxygenase from <i>Rhodothermus marinus</i>
PDB ID	6WK3
Wavelength	1.74 Å
Resolution range	43.65 – 2.45 (2.538 – 2.45)
Space group	C 1 2 1
Unit cell	a = 124.269, b = 102.472, c = 84.273 $\alpha = 90, \beta = 98.529, \gamma = 90$
Total reflections	248224 (25486)
Unique reflections	36943 (3600)
Multiplicity	6.7 (7.1)
Completeness (%)	95.98 (94.17)
Mean I/sigma(I)	13.34 (1.53)
Wilson B-factor	58.05
R-merge	0.09393 (1.464)
R-meas	0.1022 (1.581)
R-pim	0.03949 (0.5911)
CC1/2	0.997 (0.647)
Reflections used in refinement	36927 (3600)
Reflections used for R-free	1816 (177)
R-work	0.2077 (0.2991)
R-free	0.2467 (0.3343)
CC(work)	0.966 (0.785)
CC(free)	0.959 (0.723)
Number of non-hydrogen atoms	4680
macromolecules	4481
Ligands	198
Solvent	1
Protein residues	578
RMS(bonds)	0.004
RMS(angles)	0.74
Ramachandran favored (%)	99.30
Ramachandran allowed (%)	0.70
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.00
Clashscore	4.49
Average B-factor	60.86
macromolecules	61.08
Ligands	55.85
Solvent	55.94
Number of TLS groups	1

Table 2-13. P450<sub>BM3</sub> “superscreen” crystallographic conditions. Buffers are added at 0.2 M concentration unless otherwise noted.

Well	Condition
A1	pH 5.5 Bis-Tris, 20% PEG 3350
A2	pH 5.5 Bis-Tris, 10% PEG 4000, 0.2 M MgCl <sub>2</sub>
A3	pH 5.5 Bis-Tris, 14% PEG 4000, 0.2 M MgCl <sub>2</sub>
A4	pH 7.0 HEPES, 1.3 M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , 0.2 M MgCl <sub>2</sub>
A5	pH 7.0 HEPES, 1.6 M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , 0.2 M MgCl <sub>2</sub>
A6	pH 8.5 Tris, 2 M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>
B1	pH 5.5 Bis-Tris, 20% PEG3350, 0.2 M Na(HCOO)
B2	pH 6.5 Bis-Tris, 10% PEG4000, 0.2 M MgCl <sub>2</sub>
B3	pH 6.5 Bis-Tris, 14% PEG4000, 0.2 M MgCl <sub>2</sub>
B4	3 M Na(HCOO)
B5	3.25 M Na(HCOO)
B6	3.5 M Na(HCOO)
C1	pH 8.5 Tris, 25% PEG3350, 0.2 M KNO <sub>3</sub>
C2	pH 7.5 Tris, 10% PEG 4000, 0.2 M MgCl <sub>2</sub>
C3	pH 7.5 Tris, 14% PEG 4000, 0.2 M MgCl <sub>2</sub>
C4	pH 6.5 Bis-Tris, 1.5 M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , 0.1 M NaCl
C5	pH 8.5 Tris, 20% PEG 8000
C6	pH 7.5 HEPES, 15% PEG 20000
D1	pH 8.5 Tris, 25% PEG 3350, 0.2 M MgCl <sub>2</sub>
D2	pH 8.5 Tris, 10% PEG 4000, 0.2 M MgCl <sub>2</sub>
D3	pH 8.5 Tris, 14% PEG 4000, 0.2 M MgCl <sub>2</sub>
D4	pH 7.5 HEPES, 20% PEG 550MME, 0.05 M CaCl <sub>2</sub>
D5	pH 6 Cacodylate, 13% PEG3350, 0.14 M MgCl <sub>2</sub>
D6	pH 6 Cacodylate, 18% PEG3350, 0.14 M MgCl <sub>2</sub>

### Small-scale lysate and purified protein reactions

Small-scale reactions were set up in 2-mL GC crimp vials. Lysate or purified protein was diluted to the desired concentration with M9-N buffer (no glucose added), added to the vials, and brought into a Coy anaerobic chamber (~ 0-10 ppm O<sub>2</sub>). To each vial was added sodium

dithionite (final concentration 2 mM), alkene (final concentration 10 mM), and EDA **1** (final concentration 20 mM) with 5% ethanol as a cosolvent. Directly following addition of EDA, the reaction vial was crimped and shaken at 500 rpm at RT. Reactions were worked up by the addition of HCl (16  $\mu$ L, 3 M stock) and internal standard (16  $\mu$ L of 40 mM acetophenone in cyclohexane). Cyclohexane (700  $\mu$ L) was added, and the reaction was transferred into 1.7-mL Eppendorf tubes for extraction. The extraction was carried out with a Retsch MM 301 mixing mill (1 minute, 30 Hz/1800 rpm). Samples were centrifuged at 20000 $\times$ g for 5 minutes at RT, and the organic layer was used for chromatographic analysis.

### **Determination of hemin-catalyzed cyclopropanation activity**

Hemin-catalyzed cyclopropanation reactions were set up as small-scale reactions above; hemin (50  $\mu$ M final concentration) in M9-N buffer, with or without 1 mg mL<sup>-1</sup> bovine serum albumin (BSA), was brought into the Coy anaerobic chamber in 2-mL glass crimp vials. To each vial were added sodium dithionite (final concentration 2 mM), alkene (final concentration 10 mM), and EDA **1** (final concentration 20 mM) with 5% ethanol as a cosolvent. Directly following addition of EDA, the reaction vial was crimped and shaken at 500 rpm at room temperature for 16 hours.

### **Compound synthesis and characterization**

#### **General procedure A:**

Rhodium acetate dimer (10  $\mu$ mol, 4.4 mg) and a stir bar were added to a 5-mL dram vial, and it was sealed with a septum. The sealed vial was purged with three cycles of vacuum and argon. Neat olefin (8 mmol) was added to the vial. EDA (2 mmol) was added in a 2-hour slow addition on ice and reacted overnight at room temperature. The crude reaction mixture was concentrated *in vacuo* and loaded on a SNAP Ultra silica flash cartridge. The reaction mix was separated on an Isolera flash purification system (Biotage, Charlotte, NC) with a hexane/ethyl acetate gradient. Fractions containing the desired product were pooled and concentrated *in vacuo*. Yields were approximately 5–30%, due in part to both significant EDA dimer formation and low conversion of the unactivated alkenes.

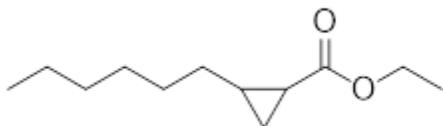
**General procedure B:**

Rhodium acetate dimer (40  $\mu$ mol, 17 mg) and a stir bar were added to a scintillation vial and sealed with a septum. The sealed vial was purged with three cycles of vacuum and argon. The vial was charged with alkene (5 mmol) and dichloromethane (5 mL). EDA **1** (2.0 eq. diluted in 4 mL dichloromethane) was added at room temperature by slow addition over 2 hours and reacted overnight at room temperature. The crude reaction mixture was concentrated *in vacuo* and loaded on a SNAP Ultra silica flash cartridge. Using pentane/diethyl ether as eluents, the reaction mix was separated on an Isolera flash purification system (Biotage, Charlotte, NC). Fractions containing the desired product were pooled and concentrated *in vacuo*. Yields of pure fractions were approximately 5–30%, due in part to both significant EDA dimer formation and low conversion of the unactivated alkenes. This method is preferable for more volatile alkenes, and the pentane/diethyl ether gradients appeared to give better separation from the EDA dimer byproducts compared to the hexanes/ethyl acetate gradient.

**Determination of absolute configurations of the cyclopropane products**

Absolute configurations of **3a** were confirmed by comparison to literature chiral GC.<sup>15</sup> Cyclosil-B column, 90 °C isothermal, absolute configuration of products elute in order: *cis* (1*S*, 2*R*)–**3a**, (1*R*, 2*S*)–**3a**, *trans*: (1*R*, 2*R*)–**3a**, (1*S*, 2*S*)–**3a**. The absolute configurations of other compounds in this study were not determined, but one could infer them by analogy, assuming the facial selectivity of the diazo reagents and olefins from which these products were made remains the same for each protein variant. The inferred absolute configurations provided here should be used with caution, understanding that substrate effects could have inverted the absolute stereochemistry. The chiral separation conditions for all cyclopropane products are detailed below and can be utilized to compare the enzymatic products to an absolute configuration authentic standard in future studies. The optical rotations of some isolated products were collected for additional characterization.

### Ethyl 2-hexylcyclopropane-1-carboxylate (**3a**)

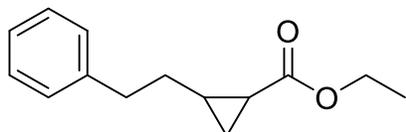


**3a** was synthesized with procedure B from 1-octene **2a** and **1** to give **3a** as a 58:42 mixture of *trans/cis*-isomers. Both *trans*- and *cis*-**3a** are known compounds.<sup>15,16</sup>

*Cis*-**3a**: <sup>1</sup>H NMR (300 MHz, Chloroform-*d*) δ 4.13 (q, *J* = 7.1 Hz, 2H), 1.66 (ddd, *J* = 8.9, 7.8, 5.5 Hz, 1H), 1.56 – 1.43 (m, 2H), 1.38 – 1.22 (m, 12H), 1.03 – 0.90 (m, 2H), 0.90 – 0.82 (m, 3H).

*Trans*-**3a**: <sup>1</sup>H NMR (300 MHz, Chloroform-*d*) δ 4.11 (q, *J* = 7.1 Hz, 2H), 1.36 – 1.22 (m, 15H), 1.14 (dt, *J* = 8.7, 4.5 Hz, 1H), 0.90 – 0.84 (m, 3H), 0.68 (ddd, *J* = 7.9, 6.1, 4.0 Hz, 1H). HR-MS (FAB<sup>+</sup>): fragment ion, loss of ethoxy group [M – CH<sub>3</sub>CH<sub>2</sub>O]<sup>+</sup> C<sub>10</sub>H<sub>17</sub>O calculated 153.1279, found 153.127.

### Ethyl 2-phenethylcyclopropane-1-carboxylate (**3b**)



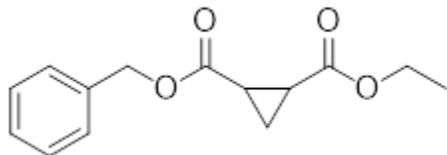
**3b** was synthesized with procedure B from 4-phenyl-1-butene **2b** and **1** to give **3b** as a 57:43 mixture of *trans/cis*-isomers. *Cis*-**3b** is a known compound.<sup>14</sup>

*Cis*-**3b**: <sup>1</sup>H NMR (300 MHz, Chloroform-*d*) δ 7.37 – 7.22 (m, 5H), 4.20 (q, *J* = 7.1 Hz, 2H), 2.70 (t, *J* = 7.3 Hz, 2H), 1.93 (tq, *J* = 14.1, 6.7 Hz, 2H), 1.80 – 1.66 (m, 1H), 1.34 (td, *J* = 7.1, 0.7 Hz, 4H), 1.14 – 0.96 (m, 2H). <sup>13</sup>C NMR (75 MHz, Chloroform-*d*) δ 173.2, 142.1, 128.6, 128.4, 125.9, 60.5, 36.0, 29.0, 21.5, 18.4, 14.5, 13.6.

*Trans*-**3b** <sup>1</sup>H NMR (400 MHz, Chloroform-*d*) δ 7.32 – 7.23 (m, 2H), 7.24 – 7.13 (m, 3H), 4.11 (q, *J* = 7.1 Hz, 2H), 2.72 (t, *J* = 7.7 Hz, 2H), 1.71 – 1.54 (m, 2H), 1.45 – 1.31 (m, 2H), 1.26 (t, *J* = 7.1 Hz, 3H), 1.16 (ddd, *J* = 8.8, 4.8, 4.1 Hz, 1H), 0.69 (ddd, *J* = 8.2, 6.3, 4.1 Hz, 1H). <sup>13</sup>C NMR (101 MHz, Chloroform-*d*) δ 174.5, 141.8, 128.5, 128.5, 126.0, 60.5, 35.6,

35.2, 22.6, 20.4, 15.6, 14.4. HR-MS (FAB+):  $[M+H]^+$   $C_{14}H_{19}O_2$ , calculated 219.1385, found 219.1382.

### 1-Benzyl 2-ethyl cyclopropane-1,2-dicarboxylate (**3c**)



General procedure A only produced trace amounts of the cyclopropane product **3c**. As each final protein variant produced **3c** with high activity, preparative-scale reactions were carried out for each protein. The purified products were confirmed via NMR and used for chiral separation method development and analytical-scale calibration curves.

*Cis-3c* (prepared from P411-UA V87C)  $^1H$  NMR (400 MHz, Chloroform-*d*)  $\delta$  7.42 – 7.28 (m, 5H), 5.15 (d,  $J = 12.3$  Hz, 1H), 5.10 (d,  $J = 12.3$  Hz, 1H), 4.08 (qd,  $J = 7.1, 1.0$  Hz, 2H), 2.16 – 1.99 (m, 2H), 1.71 (td,  $J = 6.7, 5.0$  Hz, 1H), 1.28 – 1.23 (m, 1H), 1.20 (t,  $J = 7.2$  Hz, 3H).  $^{13}C$  NMR (101 MHz, Chloroform-*d*)  $\delta$  169.9, 169.9, 135.9, 128.7, 128.5, 128.4, 67.0, 61.2, 21.9, 21.7, 14.3, 11.9. HR-MS:  $[M+H]^+$   $C_{14}H_{17}O_4$ , calculated 249.1127, found 249.1125. *Cis-3c* optical rotation from *ApePgb* AGW product:  $[\alpha]_D^{22} = -17.0^\circ$  ( $c$  0.1, EtOAc). *Cis-3c* optical rotation from P411<sub>BM3</sub>-UA V87C product:  $[\alpha]_D^{22} = +9.8^\circ$  ( $c$  0.1, EtOAc).

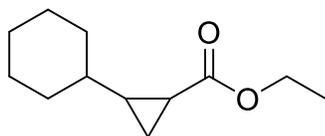
*Trans-3c* (prepared from P411-UA V87F)  $^1H$  NMR (400 MHz, Chloroform-*d*)  $\delta$  7.43 – 7.29 (m, 5H), 5.13 (s, 2H), 4.14 (q,  $J = 7.1$  Hz, 2H), 2.27 – 2.15 (m, 2H), 1.46 (ddd,  $J = 8.3, 6.2, 1.8$  Hz, 2H), 1.26 (t,  $J = 7.1$  Hz, 3H).  $^{13}C$  NMR (101 MHz, Chloroform-*d*)  $\delta$  171.8, 171.8, 135.6, 128.8, 128.5, 128.5, 67.0, 61.3, 22.7, 22.5, 15.7, 14.3. HR-MS :  $[M+H]^+$   $C_{14}H_{17}O_4$ , calculated 249.1127, found 249.1123. *Trans-3c* optical rotation from P411<sub>BM3</sub>-UA V87F product:  $[\alpha]_D^{22} = -124.1^\circ$  ( $c$  0.1, EtOAc).

**Ethyl 2-(4-bromobutyl)cyclopropane-1-carboxylate (3d)**

Both *cis*- and *trans*-**3d** are known compounds.<sup>37</sup> **3d** was synthesized with Procedure A from 6-bromo-1-hexene **2d** and EDA **1** to give **3d**. The crude *cis:trans* ratio of this reaction was not determined.

*Cis*-**3d** <sup>1</sup>H NMR (400 MHz, Chloroform-*d*) δ 4.14 (q, *J* = 7.1 Hz, 2H), 3.40 (t, *J* = 6.8 Hz, 2H), 1.93 – 1.80 (m, 2H), 1.74 – 1.38 (m, 4H), 1.56 (s, 1H), 1.33 – 1.15 (m, 4H), 1.02 (td, *J* = 8.0, 4.5 Hz, 1H), 0.92 (ddd, *J* = 7.2, 5.4, 4.5 Hz, 1H). <sup>13</sup>C NMR (101 MHz, Chloroform-*d*) δ 173.1, 60.5, 34.0, 32.6, 28.4, 26.3, 21.7, 18.3, 14.5, 13.5. HR-MS (FAB+): Fragment ion, loss of ethoxy group [M - CH<sub>3</sub>CH<sub>2</sub>O]<sup>+</sup> C<sub>8</sub>H<sub>12</sub><sup>79</sup>BrO, calculated 203.0072, found 203.0026.

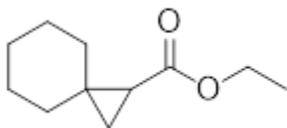
**3d** fraction containing 71:29 *trans/cis* diastereomeric mixture, reporting shifts characterized as *trans*-**3d**. <sup>1</sup>H NMR (400 MHz, Chloroform-*d*) δ 4.21 – 4.04 (m, 2H), 3.40 (td, *J* = 6.8, 3.6 Hz, 2H), 1.94 – 1.79 (m, 2H), 1.72 – 1.46 (m, 3H), 1.39 – 1.31 (m, 2H), 1.31 – 1.19 (m, 4H), 1.22 – 1.12 (m, 1H), 0.74 – 0.64 (m, 1H). <sup>13</sup>C NMR (101 MHz, Chloroform-*d*) δ 174.5, 60.5, 33.8, 32.5, 32.3, 27.8, 22.6, 20.3, 15.6, 14.4.

**Ethyl 2-cyclohexylcyclopropane-1-carboxylate (3e)**

Synthesized with general procedure A from vinylcyclohexane **2e** and EDA **1** to give **3e**. The crude *trans:cis* ratio of this reaction was not determined. *Cis*-**3e** is a known compound.<sup>14</sup>

*Cis*-**3e**: <sup>1</sup>H NMR (300 MHz, Chloroform-*d*) δ 4.18 – 4.04 (m, 1H), 1.85 – 1.51 (m, 6H), 1.29 – 0.92 (m, 2H). <sup>13</sup>C NMR (75 MHz, Chloroform-*d*) δ 173.2, 60.2, 35.8, 33.3, 33.2, 28.6, 26.4, 26.1, 25.9, 17.9, 14.4, 12.5. HR-MS (FAB+): [M]<sup>+</sup> C<sub>12</sub>H<sub>20</sub>O<sub>2</sub>, calculated 196.1463, found 196.1464.

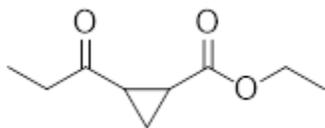
### Ethyl spiro[2.5]octane-1-carboxylate (**3f**)



Synthesized with general procedure A from methylenecyclohexane **2f** and EDA **1**. **3f** is a known compound.<sup>38</sup>

<sup>1</sup>H NMR (400 MHz, Chloroform-*d*)  $\delta$  4.11 (qd,  $J = 7.1, 1.0$  Hz, 2H), 1.62 – 1.27 (m, 11H), 1.24 (t,  $J = 7.1$  Hz, 3H), 1.06 (dd,  $J = 5.4, 4.3$  Hz, 1H), 0.79 (dd,  $J = 7.8, 4.3$  Hz, 1H). <sup>13</sup>C NMR (101 MHz, Chloroform-*d*)  $\delta$  173.1, 60.2, 37.4, 30.7, 28.8, 26.2 (2 overlapping carbons), 25.7, 25.7, 20.6, 14.4. HR-MS (FAB<sup>+</sup>): [M]<sup>+</sup> C<sub>11</sub>H<sub>18</sub>O<sub>2</sub>, calculated 182.1307, found 182.1282.

### Ethyl 2-propionylcyclopropane-1-carboxylate (**3g**)



Running general procedure B with 1-penten-3-one **2h** produced no detectable cyclopropane product **3g**; the primary products were diethyl fumarate, diethyl maleate, and an apparent 1,3-dipolar [2+3] cycloaddition product (ethyl 3-propionyl-4,5-dihydro-1*H*-pyrazole-5-carboxylate). *Trans*-**3g** was synthesized on a 5-mmol scale with a racemic strategy adapted from the chiral synthesis previously reported.<sup>39</sup> Co(II) tetraphenylporphyrin (5 mol%, 168 mg) and 1,1-dimethylaminopyridine (DMAP, 305 mg, 2.5 mmol) were added to a 40-mL scintillation vial with stir bar. The vial was sealed with a septum and placed under an Argon atmosphere. Dichloromethane (18 mL) was added, followed by 1-penten-3-one **2g** (510  $\mu$ L, 5 mmol). EDA **1** (725  $\mu$ L, 6 mmol) was added via syringe pump over 2 hours. The crude reaction mixture was concentrated *in vacuo* and separated via flash chromatography with a SNAP Ultra 25g column using a gradient of 0-20% diethyl ether in pentane. Fractions containing *trans*-**3g** were concentrated *in vacuo* to give the desired product.

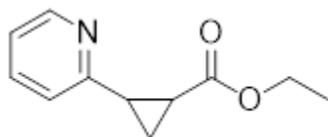
*Trans*-**3g** <sup>1</sup>H NMR (400 MHz, Chloroform-*d*)  $\delta$  4.14 (q,  $J = 7.2$  Hz, 2H), 2.63 (q,  $J = 7.3$  Hz, 2H), 2.49 – 2.39 (m, 1H), 2.21 – 2.11 (m, 1H), 1.45 – 1.35 (m, 2H), 1.27 (t,  $J = 7.1$  Hz, 3H),

1.08 (t,  $J = 7.3$  Hz, 3H).  $^{13}\text{C}$  NMR (101 MHz, Chloroform- $d$ )  $\delta$  208.2, 172.3, 61.2, 37.3, 28.9, 24.1, 17.2, 14.3, 7.8. HR-MS (ESI+):  $[\text{M}+\text{Na}]^+$   $\text{C}_9\text{H}_{14}\text{O}_3\text{Na}$ , calculated 193.0841 found 193.0819.

*Cis-3g* was synthesized based on a modified literature procedure of the methyl ketone analog.<sup>40</sup> 3-Oxabicyclo[3.1.0]hexane-2,4-dione (10 mmol) was added to anhydrous diethyl ether under Argon. Ethylmagnesium bromide (9 mL of 1 M in diethyl ether, diluted from 3 M solution in diethyl ether, 9 mmol) was added dropwise over 20 min at  $-78$  °C. The reaction mixture was allowed to warm to room temperature over 2 hours and stirred for 18 hours. The reaction was then quenched with aqueous ammonium chloride and extracted by diethyl ether (4 x 40 mL), dried over sodium sulfate and concentrated *in vacuo*. The crude product was used for the esterification step without further purification. Half of the crude product was dissolved in ethanol (20 mL), thionyl chloride (1 mL) was added, and the reaction proceeded at room temperature for 12 hours. The reaction was then quenched with aqueous sodium bicarbonate and diluted with water and extracted with diethyl ether. The crude mixture was purified via flash chromatography using a gradient of 0–50% diethyl ether in pentane. Fractions containing *cis-3g* were concentrated *in vacuo* to give the desired product.

*Cis-3g*  $^1\text{H}$  NMR (400 MHz, Chloroform- $d$ )  $\delta$  4.20 – 4.04 (m, 2H), 2.69 – 2.47 (m, 2H), 2.22 (ddd,  $J = 9.3, 8.2, 6.7$  Hz, 1H), 2.08 (ddd,  $J = 9.2, 8.4, 6.6$  Hz, 1H), 1.70 (tdd,  $J = 6.7, 4.7, 0.6$  Hz, 1H), 1.23 (t,  $J = 7.2$  Hz, 3H), 1.22 – 1.15 (m, 1H), 1.07 (td,  $J = 7.3, 0.5$  Hz, 3H).  $^{13}\text{C}$  NMR (101 MHz, Chloroform- $d$ )  $\delta$  206.6, 170.1, 61.1, 36.9, 27.9, 23.4, 14.3, 12.2, 7.9.

### Ethyl 2-(pyridin-2-yl)cyclopropane-1-carboxylate (**3h**)



Both *trans*- and *cis-3h* are known compounds.<sup>41</sup> *Cis*- and *trans-3h* were produced by refluxing 2-vinylpyridine **2h** (10 mmol) with EDA **1** (10 mmol) in toluene (20 mL) overnight. The reaction yielded a 65:35 *trans/cis* mixture of diastereomers. The crude mixture was concentrated *in vacuo* and the product was purified using flash chromatography

with a SNAP Ultra 10g Biotage column, 0–100% hexane:EtOAc gradient. Fractions containing pure *trans*-**3h** and *cis*-**3h** were separable and concentrated *in vacuo* to give the desired products (391.7 mg *trans*-**3h**, 290.7 mg *cis*-**3h**, 36% yield).

*Cis*-**3h**:  $^1\text{H}$  NMR (400 MHz, Chloroform-*d*)  $\delta$  8.49 (ddd,  $J = 4.9, 1.9, 0.9$  Hz, 1H), 7.58 (td,  $J = 7.7, 1.8$  Hz, 1H), 7.32 – 7.22 (m, 1H), 7.10 (ddd,  $J = 7.4, 4.8, 1.1$  Hz, 1H), 3.90 (qd,  $J = 7.2, 1.5$  Hz, 2H), 2.72 (td,  $J = 9.0, 7.6$  Hz, 1H), 2.22 – 2.11 (m, 1H), 1.87 – 1.77 (m, 1H), 1.47 – 1.36 (m, 1H), 1.02 (td,  $J = 7.1, 0.9$  Hz, 3H).  $^{13}\text{C}$  NMR (101 MHz, Chloroform-*d*)  $\delta$  171.1, 156.7, 148.9, 135.9, 123.6, 121.6, 60.3, 27.2, 21.7, 14.0, 11.7.

*Trans*-**3h**  $^1\text{H}$  NMR (400 MHz, Chloroform-*d*)  $\delta$  8.43 (ddd,  $J = 4.9, 1.9, 1.0$  Hz, 1H), 7.54 (td,  $J = 7.7, 1.8$  Hz, 1H), 7.21 (dt,  $J = 7.8, 1.1$  Hz, 1H), 7.07 (ddd,  $J = 7.5, 4.9, 1.2$  Hz, 1H), 4.15 (q,  $J = 7.2$  Hz, 2H), 2.56 (ddd,  $J = 8.9, 6.1, 3.9$  Hz, 1H), 2.23 (ddd,  $J = 8.4, 5.5, 3.9$  Hz, 1H), 1.59 (dddd,  $J = 13.8, 9.0, 5.8, 3.8$  Hz, 2H), 1.26 (t,  $J = 7.1$  Hz, 3H).  $^{13}\text{C}$  NMR (101 MHz, Chloroform-*d*)  $\delta$  173.4, 158.9, 149.4, 136.0, 122.5, 121.3, 60.7, 27.2, 24.4, 17.3, 14.3. HR-MS (FAB+):  $[\text{M}]^+$   $\text{C}_{11}\text{H}_{13}\text{NO}_2$ , calculated 191.0946, found 191.092.

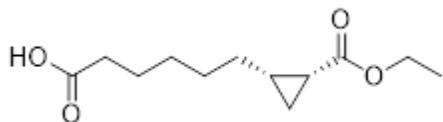
### **Cis Ethyl 2-(6-hydroxyhexyl)cyclopropane-1-carboxylate (*cis*-**3i**)**



Cyclopropanation of the free alcohol **2i** was performed via preparative-scale enzymatic reaction (*ApePgb* AGW) to give *cis*-**3i**.

*cis* -**3i**  $^1\text{H}$  NMR (400 MHz, Chloroform-*d*)  $\delta$  4.12 (q,  $J = 7.1$  Hz, 2H), 3.62 (t,  $J = 6.6$  Hz, 2H), 1.78 – 1.61 (broad m, 1H), 1.66 (ddd,  $J = 8.9, 7.8, 5.5$  Hz, 1H), 1.60 – 1.43 (m, 4H), 1.41 – 1.29 (m, 6H), 1.26 (t,  $J = 7.1$  Hz, 3H), 1.25 – 1.16 (m, 1H), 1.05 – 0.95 (m, 1H), 0.91 (ddd,  $J = 7.2, 5.5, 4.4$  Hz, 1H).  $^{13}\text{C}$  NMR (101 MHz, Chloroform-*d*)  $\delta$  173.3, 63.1, 60.4, 32.8, 29.7, 29.2, 27.0, 25.8, 22.1, 18.3, 14.5, 13.5. HR-MS (FAB+):  $[\text{M}+\text{H}]^+$   $\text{C}_{12}\text{H}_{23}\text{O}_3$ , calculated 215.1647, found 215.1648.

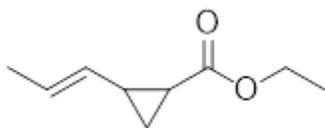
**Cis 6-(2-(ethoxycarbonyl)cyclopropyl)hexanoic acid (*cis*-**3j**)**



Cyclopropanation of the free carboxylic acid **2j** was performed via preparative-scale enzymatic reaction (*ApePgb* AGW) to give *cis*-**3j**.

*Cis*-**3j**  $^1\text{H}$  NMR (400 MHz, Chloroform-*d*)  $\delta$  12.00 – 9.50 (broad, 1H), 4.13 (q,  $J = 7.1$  Hz, 2H), 2.38 – 2.29 (m, 2H), 1.71 – 1.58 (m, 3H), 1.58 – 1.40 (m, 2H), 1.40 – 1.30 (m, 4H), 1.29 – 1.16 (m, 1H), 1.26 (t, 3H), 0.99 (td,  $J = 8.1, 4.4$  Hz, 1H), 0.91 (ddd,  $J = 7.2, 5.4, 4.5$  Hz, 1H). *Trans*-**3j** cyclopropane isomers (minor product in *cis*-**3j**  $^1\text{H}$  NMR): 1.14 (tdd,  $J = 7.1, 4.8, 2.5$  Hz), 0.68 (ddd,  $J = 7.8, 5.9, 4.0$  Hz).  $^{13}\text{C}$  NMR (101 MHz, Chloroform-*d*)  $\delta$  179.4, 173.3, 60.4, 34.0, 29.4, 28.8, 26.9, 24.8, 21.9, 18.3, 14.5, 13.5. HR-MS (FAB+):  $[\text{M}+\text{H}]^+ \text{C}_{12}\text{H}_{21}\text{O}_4$ , calculated 229.1440, found 229.1448.

**Ethyl (*E*)-2-(prop-1-en-1-yl)cyclopropane-1-carboxylate (**3k**)**



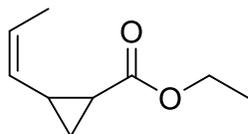
Both *trans*- and *cis*-**3k** are known compounds.<sup>42</sup> Terminal alkene cyclopropanation of (*E*)-penta-1,3-diene **2k** and EDA **1** accomplished via preparative-scale enzymatic reaction gave *cis*-**3k** (*ApePgb* AGW) and *trans*-**3k** (*RmaNOD* Q52V).

*Cis*-**3k**  $^1\text{H}$  NMR (400 MHz, Chloroform-*d*)  $\delta$  5.66 (dq,  $J = 15.4, 6.5$  Hz, 1H), 5.41 (ddq,  $J = 15.3, 8.6, 1.7$  Hz, 1H), 4.22 – 4.03 (m, 2H), 1.89 – 1.78 (m, 2H), 1.67 (dd,  $J = 6.5, 1.7$  Hz, 3H), 1.26 (t, 3H), 1.23 – 1.10 (m, 2H).

*Trans*-**3k**  $^1\text{H}$  NMR (400 MHz, Chloroform-*d*)  $\delta$  5.60 (dq,  $J = 15.1, 6.5, 0.8$  Hz, 1H), 5.03 (ddq,  $J = 15.2, 8.3, 1.7$  Hz, 1H), 4.12 (q,  $J = 7.1$  Hz, 2H), 1.95 (dddd,  $J = 9.4, 8.5, 6.3, 4.0$  Hz, 1H), 1.65 (dd,  $J = 6.5, 1.7$  Hz, 3H), 1.57 – 1.53 (m, 1H), 1.31 (ddd,  $J = 9.2, 5.1, 4.3$  Hz, 1H), 1.26 (t,  $J = 7.1$  Hz, 3H), 0.91 (ddd,  $J = 8.3, 6.3, 4.3$  Hz, 1H).  $^{13}\text{C}$  NMR (101 MHz, Chloroform-*d*)  $\delta$  173.9, 130.9, 126.1, 60.7, 25.0, 21.8, 17.9, 15.6, 14.4. HR-MS (EI+):  $[\text{M}]^+$

$C_9H_{14}O_2$ , calculated 154.0994, found 154.0974. *Trans-3k* optical rotation from *RmaNOD* Q52V product:  $[\alpha]_D^{23} = +164^\circ$  (*c* 0.1, EtOAc).

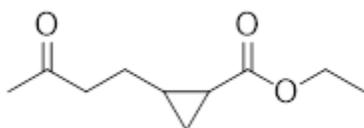
### Ethyl (Z)-2-(prop-1-en-1-yl)cyclopropane-1-carboxylate (**3l**)



Both *trans*- and *cis*-**3l** are known compounds.<sup>42</sup> General procedure A was used to produce a mixture of 8 possible cyclopropanes (internal and terminal) for chiral GC standard from (*Z*)-penta-1,3-diene **2l** and EDA **1**. Stereoselective, terminal alkene cyclopropanation accomplished via preparative-scale enzymatic reaction (*ApePgb* AGW).

*Cis-3l*  $^1H$  NMR (400 MHz, Chloroform-*d*)  $\delta$  5.56 (dq,  $J = 10.9, 6.9, 1.1$  Hz, 1H), 5.34 (ddq,  $J = 11.0, 9.3, 1.8$  Hz, 1H), 4.21 – 4.04 (m, 2H), 2.16 – 2.03 (m, 1H), 1.92 (ddd,  $J = 8.8, 7.8, 6.0$  Hz, 1H), 1.72 (dd,  $J = 6.8, 1.8$  Hz, 3H), 1.25 (t,  $J = 7.1$  Hz, 3H), 1.23 – 1.17 (m, 2H).  $^{13}C$  NMR (101 MHz, Chloroform-*d*)  $\delta$  172.3, 127.2, 126.3, 60.5, 21.0, 19.2, 14.5, 14.5, 13.3. *Cis-3l* optical rotation from *ApePgb* AGW product:  $[\alpha]_D^{23} = -260^\circ$  (*c* 0.1, EtOAc).

### Ethyl 2-(3-oxobutyl)cyclopropane-1-carboxylate (**3m**)



*Cis-3m* is a known compound.<sup>16</sup> General procedure B was used to produce a mixture of 55:45 *trans/cis* **3m**, and a preparative-scale enzymatic reaction (*ApePgb* AGW) produced *cis-3m*.

*Cis-3m*  $^1H$  NMR (400 MHz, Chloroform-*d*)  $\delta$  4.13 (q,  $J = 7.1$  Hz, 2H), 2.46 (td,  $J = 7.3, 2.3$  Hz, 2H), 2.13 (s, 3H), 1.94 – 1.62 (m, 3H), 1.26 (t,  $J = 7.1$  Hz, 4H), 1.06 – 0.86 (m, 2H).  $^{13}C$  NMR (101 MHz, Chloroform-*d*)  $\delta$  208.9, 173.2, 60.7, 43.7, 30.2, 21.8, 21.3, 18.6, 14.7, 13.8. HR-MS (FAB+):  $[M+H]^+$   $C_{10}H_{17}O_3$ , calculated 185.1178, found 185.1171. *Cis-3m* optical rotation from *ApePgb* AGW product:  $[\alpha]_D^{23} = -52^\circ$  (*c* 0.1, EtOAc).

### Compound chiral separation conditions

Table 2-14. Chromatographic separation conditions. Column dimensions: Cyclosil-B 30 m length x 0.32 mm ID x 0.25  $\mu\text{m}$  film thickness; Chiraldex G-TA 30 m length x 0.25 mm ID x 0.12  $\mu\text{m}$  film thickness; DB-WAXETR (15/30 m) length x 0.32 mm ID x 0.25  $\mu\text{m}$  film thickness; Chiralpak IC 4.6 mm x 250 mm x 5  $\mu\text{m}$ ; Chiralpak IA 4.6 mm x 250 mm x 5  $\mu\text{m}$ .

Compound	Instrument / column	Method	Retention times
<b>3a</b>	GC-FID, Cyclosil-B	90 °C isothermal	<i>Cis</i> 67.9 min, 69.5 min; <i>trans</i> 81.7 min, 83.5 min
<b>3b</b>	GC-FID, Cyclosil-B	80 °C hold 2 min, 30 °C $\text{min}^{-1}$ to 140 °C hold 40 min	<i>cis</i> 30.2 min, 30.9 min; <i>trans</i> 35.7 min (unresolved)
	HPLC, Chiralpak IC	Hexane: 0.8% isopropanol isocratic	<i>cis</i> 7.1 min, 7.3 min (unresolved); <i>trans</i> 8.9 min, 10.3 min
<b>3c</b>	GC-FID, Cyclosil-B	80 °C hold 2 min, 30 °C $\text{min}^{-1}$ to 170 °C hold 30 min	<i>Trans</i> 24.5 min (unresolved); <i>cis</i> 26.4 min, 26.8 min
	HPLC, Chiralpak IC	Hexane: 16% isopropanol isocratic	<i>Cis</i> 4.6 min (unresolved); <i>trans</i> 6.4 min, 7.2 min
<b>3d</b>	GC-FID, Chiraldex G-TA	100 °C, 2 °C $\text{min}^{-1}$ to 170 °C	<i>Cis</i> 25.1 min, 25.3 min; <i>trans</i> 27.1 min, 27.2 min
<b>3e</b>	GC-FID, Cyclosil-B	80 °C hold 2 min, 30 °C $\text{min}^{-1}$ to 100 °C hold 80 min	<i>Cis</i> 44.7 min, 46.8 min; <i>trans</i> 76.6 min, 78.2 min
<b>3f</b>	GC-FID, Chiraldex G-TA	80 °C isothermal	50.05 min, 51.0 min
<b>3g</b>	HPLC, Chiralpak IC	Hexane: 9% isopropanol isocratic	<i>Trans</i> 8.7 min, 9.7 min (second peak coelution with EDA dimer)
	GC-FID, Cyclosil-B	80 °C hold 2 min, 30 °C $\text{min}^{-1}$ to 125 °C hold 12 min	<i>Cis</i> 12.5 min, 12.7 min
<b>3h</b>	GC-FID, Cyclosil-B	80 °C hold 2 min, 20 °C $\text{min}^{-1}$ to 120 °C, 6 °C $\text{min}^{-1}$ to 180 °C	<i>Trans</i> 13.7 min, 13.8 min
<b>3i</b>	GC-FID, DB-WAXETR (30 m)	80 °C hold 2 min, 10 °C $\text{min}^{-1}$ to 170 °C	2i 8.4 min; EDA dimer (byproducts) 8.9 min, 10.1 min; <i>cis</i> -3i 14.5 min, <i>trans</i> -3i 14.6 min
<b>3ia</b>	HPLC, Chiralpak IA	Hexane: 0.5% isopropanol isocratic	<i>Cis</i> 8.7 min, 9.0 min
<b>3j</b>	GC-FID, DB-WAXETR (15 m)	90 °C hold 2 min, 10 °C $\text{min}^{-1}$ to 250 °C	EDA dimer (byproducts) 5.0 min, 6.2 min, 2j 9.7 min, <i>cis</i> -3j 16.6 min, <i>trans</i> -3j 16.7 min
<b>3ja</b>	HPLC, Chiralpak IC	Hexane: 1% isopropanol isocratic	<i>Cis</i> 24.9 min, 26.2 min

Compound	Instrument / column	Method	Retention times
<b>3k</b>	GC-FID, Cyclosil-B	80 °C hold 2 min, 5 °C min <sup>-1</sup> to 140 °C	Terminal cyclopropane products: <i>cis</i> 9.5 min, 9.6 min; <i>trans</i> 9.9 min, 10.1 min
<b>3l</b>	GC-FID, Cyclosil-B	80 °C hold 2 min, 10 °C min <sup>-1</sup> to 85 °C hold 20 min	Terminal cyclopropane products: <i>cis</i> 15.8 min, 16.3 min; <i>trans</i> 17.6 min, 20.4 min
<b>3m</b>	GC-FID, ChiralDEX G-TA	90 °C isothermal	<i>Cis</i> 112.7 min, 114.6 min

### Calibration curves for analytical-scale TTN determination

Calibration curves were prepared with analogous conditions to the analytical scale reactions. Cyclopropane product standards were diluted in ethanol and added (20  $\mu\text{L}$ ) to M9-N buffer (380  $\mu\text{L}$ ), with final concentrations of product ranging from 0–20 mM (with two to three technical replicates). The samples were worked up the same as analytical-scale reactions, by the addition of HCl (16  $\mu\text{L}$ , 3 M stock) and internal standard (16  $\mu\text{L}$  of 40 mM acetophenone in cyclohexane). Cyclohexane (700  $\mu\text{L}$ ) was added, and the samples were transferred into 1.7-mL Eppendorf tubes for extraction. The extraction was carried out with a Retsch MM 301 mixing mill (1 minute, 30 Hz/1800 rpm). Samples were centrifuged at 20000 $\times$ g for 5 minutes at RT and the organic layer was used for chromatographic analysis. Cyclopropane products for methylenecyclohexane **3g** and 2-vinylpyridine **3h** were set up in the same way, with the exception of running ten single replicates over the 0–20 mM product range.

The ratio of the product area and internal standard (P/S ratio) was determined for each sample, and a linear regression was performed for the P/S ratio versus the known concentration in each reaction, with the y–intercept set to 0. The resulting slopes were used to determine the concentration of product in the analytical-scale reactions, which is in turn used to calculate the TTN per enzyme.

Table 2-15. Information on engineered proteins in the chapter. All variants for biocatalytic reactions were in pET22b(+) vectors lacking the pelB signal peptide sequence and expressed in the E. coli strain EXPRESS BL21(DE3). DNA and amino-acid sequence information are deposited on CaltechDATA (DOI: 10.22002/D1.1437).

Protein	Original organism	Mutation(s) with respect to WT
<i>ApePgb</i> WT	<i>Aeropyrum pernix</i>	
<i>ApePgb</i> Y60G	<i>Aeropyrum pernix</i>	Y60G
<i>ApePgb</i> W59A Y60G	<i>Aeropyrum pernix</i>	W59A Y60G
<i>ApePgb</i> W59A Y60G F145W ("AGW")	<i>Aeropyrum pernix</i>	W59A Y60G F145W
<i>RmaNOD</i> WT	<i>Rhodothermus marinus</i>	
<i>RmaNOD</i> Q52V	<i>Rhodothermus marinus</i>	Q52V
P411 <sub>BM3</sub> -CIS	<i>Bacillus megaterium</i>	V78A, F87V, P142S, T175I, A184V, S226R, H236Q, E252G, T268A, A290V, L353V, I366V, C400S, E442K
P411 <sub>BM3</sub> -UA	<i>Bacillus megaterium</i>	P411-CIS, L75Y, L181I, L437F, T438Q
P411 <sub>BM3</sub> -UA-V87C	<i>Bacillus megaterium</i>	P411-CIS, L75Y, V87C, L181I, L437F, T438Q
P411 <sub>BM3</sub> -UA-V87F	<i>Bacillus megaterium</i>	P411-CIS, L75Y, V87F, L181I, L437F, T438Q

## 2.7 References for Chapter 2

- (1) Finefield, J. M.; Sherman, D. H.; Kreitman, M.; Williams, R. M. Enantiomeric Natural Products: Occurrence and Biogenesis. *Angew. Chem., Int. Ed.* **2012**, *51*, 4802–4836. <https://doi.org/10.1002/anie.201107204>.
- (2) Mishra, B. B.; Tiwari, V. K. Natural Products: An Evolving Role in Future Drug Discovery. *Eur. J. Med. Chem.* **2011**, *46*, 4769–4807. <https://doi.org/10.1016/j.ejmech.2011.07.057>.
- (3) Shi, S.-L.; Wong, Z. L.; Buchwald, S. L. Copper-Catalysed Enantioselective Stereodivergent Synthesis of Amino Alcohols. *Nature* **2016**, *532*, 353–356. <https://doi.org/10.1038/nature17191>.
- (4) H. Brooks, W.; C. Guida, W.; G. Daniel, K. The Significance of Chirality in Drug Design and Development. *Curr. Top. Med. Chem.* **2011**, *11*, 760–770. <https://doi.org/10.2174/156802611795165098>.
- (5) Krautwald, S.; Carreira, E. M. Stereodivergence in Asymmetric Catalysis. *J. Am. Chem. Soc.* **2017**, *139*, 5627–5639. <https://doi.org/10.1021/jacs.6b13340>.
- (6) Miller, C. A. Advances in Enzyme Discovery Technology: Capturing Diversity. *Inform* **2000**, *11*, 489–496.
- (7) Kaluzna, I. A.; Matsuda, T.; Sewell, A. K.; Stewart, J. D. Systematic Investigation of

- Saccharomyces cerevisiae Enzymes Catalyzing Carbonyl Reductions. *J. Am. Chem. Soc.* **2004**, *126*, 12827–12832. <https://doi.org/10.1021/ja0469479>.
- (8) Höhne, M.; Schätzle, S.; Jochens, H.; Robins, K.; Bornscheuer, U. T. Rational Assignment of Key Motifs for Function Guides in Silico Enzyme Identification. *Nat. Chem. Biol.* **2010**, *6*, 807–813. <https://doi.org/10.1038/nchembio.447>.
- (9) Ferrer, M.; Martinez-Abarca, F.; Golyshin, P. N.; Mining Genomes and ‘Metagenomes’ for Novel Catalysts. *Curr. Opin. Biotechnol.* **2005**, *16*, 588–593. <https://doi.org/10.1016/j.copbio.2005.09.001>.
- (10) Wessjohann, L. A.; Brandt, W.; Thiemann, T. Biosynthesis and Metabolism of Cyclopropane Rings in Natural Compounds. *Chem. Rev.* **2003**, *103*, 1625–1648. <https://doi.org/10.1021/cr0100188>.
- (11) Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. Olefin Cyclopropanation via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes. *Science* **2013**, *339*, 307–310. <https://doi.org/10.1126/science.1231434>.
- (12) Bordeaux, M.; Tyagi, V.; Fasan, R. Highly Diastereoselective and Enantioselective Olefin Cyclopropanation Using Engineered Myoglobin-Based Catalysts. *Angew. Chem., Int. Ed.* **2015**, *54*, 1744–1748. <https://doi.org/10.1002/anie.201409928>.
- (13) Gober, J. G.; Rydeen, A. E.; Gibson-O’Grady, E. J.; Leuthaeuser, J. B.; Fetrow, J. S.; Brustad, E. M. Corrigendum: Mutating a Highly Conserved Residue in Diverse Cytochrome P450s Facilitates Diastereoselective Olefin Cyclopropanation. *ChemBioChem* **2016**, *17*, 2099–2099. <https://doi.org/10.1002/cbic.201600528>.
- (14) Brandenburg, O. F.; Fasan, R.; Arnold, F. H. Exploiting and Engineering Hemoproteins for Abiological Carbene and Nitrene Transfer Reactions. *Curr. Opin. Biotechnol.* **2017**, *47*, 102–111. <https://doi.org/10.1016/j.copbio.2017.06.005>.
- (15) Wang, Z. J.; Renata, H.; Peck, N. E.; Farwell, C. C.; Coelho, P. S.; Arnold, F. H. Improved Cyclopropanation Activity of Histidine-Ligated Cytochrome P450 Enables the Enantioselective Formal Synthesis of Levomilnacipran. *Angew. Chem., Int. Ed.* **2014**, *53*, 6810–6813. <https://doi.org/10.1002/anie.201402809>.
- (16) Bajaj, P.; Sreenilayam, G.; Tyagi, V.; Fasan, R. Gram-Scale Synthesis of Chiral Cyclopropane-Containing Drugs and Drug Precursors with Engineered Myoglobin Catalysts Featuring Complementary Stereoselectivity. *Angew. Chem., Int. Ed.* **2016**, *55*, 16110–16114. <https://doi.org/10.1002/anie.201608680>.
- (17) Hernandez, K. E.; Renata, H.; Lewis, R. D.; Kan, S. B. J. J.; Zhang, C.; Forte, J.; Rozzell, D.; McIntosh, J. A.; Arnold, F. H. Highly Stereoselective Biocatalytic Synthesis of Key Cyclopropane Intermediate to Ticagrelor. *ACS Catal.* **2016**, *6*, 7810–7813. <https://doi.org/10.1021/acscatal.6b02550>.
- (18) Kim, T.; Kassim, A. M.; Botejue, A.; Zhang, C.; Forte, J.; Rozzell, D.; Huffman, M. A.; Devine, P. N.; McIntosh, J. A. Hemoprotein-Catalyzed Cyclopropanation En Route to the Chiral Cyclopropanol Fragment of Grazoprevir. *ChemBioChem* **2019**, *20*, 1129–1132. <https://doi.org/10.1002/cbic.201800652>.
- (19) Coombs, J. R.; Morken, J. P. Catalytic Enantioselective Functionalization of Unactivated Terminal Alkenes. *Angew. Chem., Int. Ed.* **2016**, *55*, 2636–2649. <https://doi.org/10.1002/anie.201507151>.
- (20) Tinoco, A.; Steck, V.; Tyagi, V.; Fasan, R. Highly Diastereo- and Enantioselective Synthesis of Trifluoromethyl-Substituted Cyclopropanes via Myoglobin-Catalyzed

- Transfer of Trifluoromethylcarbene. *J. Am. Chem. Soc.* **2017**, *139*, 5293–5296. <https://doi.org/10.1021/jacs.7b00768>.
- (21) Suematsu, H.; Kanchiku, S.; Uchida, T.; Katsuki, T. Construction of Aryliridium–Salen Complexes: Enantio- and Cis -Selective Cyclopropanation of Conjugated and Nonconjugated Olefins. *J. Am. Chem. Soc.* **2008**, *130*, 10327–10337. <https://doi.org/10.1021/ja802561t>.
- (22) Key, H. M.; Dydio, P.; Clark, D. S.; Hartwig, J. F. Abiological Catalysis by Artificial Haem Proteins Containing Noble Metals in Place of Iron. *Nature* **2016**, *534*, 534–537. <https://doi.org/10.1038/nature17968>.
- (23) Key, H. M.; Dydio, P.; Liu, Z.; Rha, J. Y. E.; Nazarenko, A.; Seyedkazemi, V.; Clark, D. S.; Hartwig, J. F. Beyond Iron: Iridium-Containing P450 Enzymes for Selective Cyclopropanations of Structurally Diverse Alkenes. *ACS Cent. Sci.* **2017**, *3*, 302–308. <https://doi.org/10.1021/acscentsci.6b00391>.
- (24) Xu, X.; Wang, Y.; Cui, X.; Wojtas, L.; Zhang, X. P. Metalloradical Activation of  $\alpha$ -Formyldiazoacetates for the Catalytic Asymmetric Radical Cyclopropanation of Alkenes. *Chem. Sci.* **2017**, *8*, 4347–4351. <https://doi.org/10.1039/C7SC00658F>.
- (25) Minuth, T.; Boysen, M. Carbohydrate-Derived Bis(Oxazoline) Ligand in the Total Synthesis of Grenadamide. *Synthesis* **2010**, *2010*, 2799–2803. <https://doi.org/10.1055/s-0030-1258143>.
- (26) Wolf, J. R.; Hamaker, C. G.; Djukic, J.-P.; Kodadek, T.; Woo, L. K. Shape and Stereoselective Cyclopropanation of Alkenes Catalyzed by Iron Porphyrins. *J. Am. Chem. Soc.* **1995**, *117*, 9194–9199. <https://doi.org/10.1021/ja00141a011>.
- (27) Natoli, S. N.; Hartwig, J. F. Noble–Metal Substitution in Hemoproteins: An Emerging Strategy for Abiological Catalysis. *Acc. Chem. Res.* **2019**, *52*, 326–335. <https://doi.org/10.1021/acs.accounts.8b00586>.
- (28) Sreenilayam, G.; Moore, E. J.; Steck, V.; Fasan, R. Metal Substitution Modulates the Reactivity and Extends the Reaction Scope of Myoglobin Carbene Transfer Catalysts. *Adv. Synth. Catal.* **2017**, *359*, 2076–2089. <https://doi.org/10.1002/adsc.201700202>.
- (29) Nuss, P.; Eckelman, M. J. Life Cycle Assessment of Metals: A Scientific Synthesis. *PLoS One* **2014**, *9*, e101298. <https://doi.org/10.1371/journal.pone.0101298>.
- (30) Prier, C. K.; Zhang, R. K.; Buller, A. R.; Brinkmann-Chen, S.; Arnold, F. H. Enantioselective, Intermolecular Benzylic C–H Amination Catalysed by an Engineered Iron-Haem Enzyme. *Nat. Chem.* **2017**, *9*, 629–634. <https://doi.org/10.1038/nchem.2783>.
- (31) Kan, S. B. J. J.; Lewis, R. D.; Chen, K.; Arnold, F. H. Directed Evolution of Cytochrome c for Carbon–Silicon Bond Formation: Bringing Silicon to Life. *Science* **2016**, *354*, 1048–1051. <https://doi.org/10.1126/science.aah6219>.
- (32) Kan, S. B. J.; Huang, X.; Gumulya, Y.; Chen, K.; Arnold, F. H. Genetically Programmed Chiral Organoborane Synthesis. *Nature* **2017**, *552*, 132–136. <https://doi.org/10.1038/nature24996>.
- (33) Zhang, R. K.; Chen, K.; Huang, X.; Wohlschlager, L.; Renata, H.; Arnold, F. H. Enzymatic Assembly of Carbon–Carbon Bonds via Iron-Catalysed Sp<sup>3</sup> C–H Functionalization. *Nature* **2019**, *565*, 67–72. <https://doi.org/10.1038/s41586-018-0808-5>.

- (34) Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H. Protein Stability Promotes Evolvability. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 5869–5874. <https://doi.org/10.1073/pnas.0510098103>.
- (35) Coelho, P. S.; Wang, Z. J.; Ener, M. E.; Baril, S. A.; Kannan, A.; Arnold, F. H.; Brustad, E. M. A Serine-Substituted P450 Catalyzes Highly Efficient Carbene Transfer to Olefins in Vivo. *Nat. Chem. Biol.* **2013**, *9*, 485–487. <https://doi.org/10.1038/nchembio.1278>.
- (36) Brandenberg, O. F.; Prier, C. K.; Chen, K.; Knight, A. M.; Wu, Z.; Arnold, F. H. Stereoselective Enzymatic Synthesis of Heteroatom-Substituted Cyclopropanes. *ACS Catal.* **2018**, *8*, 2629–2634. <https://doi.org/10.1021/acscatal.7b04423>.
- (37) Vottero, E.; Rea, V.; Lastdrager, J.; Honing, M.; Vermeulen, N. P. E.; Commandeur, J. N. M. Role of Residue 87 in Substrate Selectivity and Regioselectivity of Drug-Metabolizing Cytochrome P450 CYP102A1 M11. *JBIC J. Biol. Inorg. Chem.* **2011**, *16*, 899–912. <https://doi.org/10.1007/s00775-011-0789-4>.
- (38) Turner, N. J. Directed Evolution Drives the next Generation of Biocatalysts. *Nat. Chem. Biol.* **2009**, *5*, 567–573. <https://doi.org/10.1038/nchembio.203>.
- (39) Cobb, R. E.; Chao, R.; Zhao, H. Directed Evolution: Past, Present, and Future. *AIChE J.* **2013**, *59*, 1432–1440. <https://doi.org/10.1002/aic.13995>.
- (40) Zhang, J.; Huang, X.; Zhang, R. K.; Arnold, F. H. Enantiodivergent  $\alpha$ -Amino C–H Fluoroalkylation Catalyzed by Engineered Cytochrome P450s. *J. Am. Chem. Soc.* **2019**, *141*, 9798–9802. <https://doi.org/10.1021/jacs.9b04344>.
- (41) Huang, X.; Garcia-Borràs, M.; Miao, K.; Kan, S. B. J.; Zutshi, A.; Houk, K. N.; Arnold, F. H. A Biocatalytic Platform for Synthesis of Chiral  $\alpha$ -Trifluoromethylated Organoborons. *ACS Cent. Sci.* **2019**, *5*, 270–276. <https://doi.org/10.1021/acscentsci.8b00679>.
- (42) Zhou, A. Z.; Chen, K.; Arnold, F. H. Enzymatic Lactone-Carbene C–H Insertion to Build Contiguous Chiral Centers. *ACS Catal.* **2020**, *10*, 5393–5398. <https://doi.org/10.1021/acscatal.0c01349>.
- (43) Chen, K.; Huang, X.; Zhang, S.-Q.; Zhou, A.; Kan, S. B. J.; Hong, X.; Arnold, F. Engineered Cytochrome C-Catalyzed Lactone-Carbene B–H Insertion. *Synlett* **2019**, *30*, 378–382. <https://doi.org/10.1055/s-0037-1611662>.
- (44) Chen, Y.; Ruppel, J. V.; Zhang, X. P. Cobalt-Catalyzed Asymmetric Cyclopropanation of Electron-Deficient Olefins. *J. Am. Chem. Soc.* **2007**, *129*, 12074–12075. <https://doi.org/10.1021/ja074613o>.
- (45) Cerutti-Delasalle, C.; Mehiri, M.; Cagliero, C.; Rubiolo, P.; Bicchi, C.; Meierhenrich, U. J.; Baldovini, N. The (+)-Cis- and (+)-Trans-Olibanic Acids: Key Odorants of Frankincense. *Angew. Chem., Int. Ed.* **2016**, *55*, 13719–13723. <https://doi.org/10.1002/anie.201605242>.
- (46) Jin, C.; Decker, A. M.; Huang, X.-P.; Gilmour, B. P.; Blough, B. E.; Roth, B. L.; Hu, Y.; Gill, J. B.; Zhang, X. P. Synthesis, Pharmacological Characterization, and Structure–Activity Relationship Studies of Small Molecular Agonists for the Orphan GPR88 Receptor. *ACS Chem. Neurosci.* **2014**, *5*, 576–587. <https://doi.org/10.1021/cn500082p>.
- (47) Noels, A. F.; Demonceau, A.; Petiniot, N.; Hubert, A. J.; Teyssié, P. Transition-Metal-Catalyzed Reaction of Diazocompounds, Efficient Synthesis of

- Functionalized Ethers by Carbene Insertion into the Hydroxylic Bond of Alcohols. *Tetrahedron* **1982**, *38*, 2733–2739. [https://doi.org/10.1016/0040-4020\(82\)80031-8](https://doi.org/10.1016/0040-4020(82)80031-8).
- (48) Nardini, M.; Pesce, A.; Thijs, L.; Saito, J. A.; Dewilde, S.; Alam, M.; Ascenzi, P.; Coletta, M.; Ciaccio, C.; Moens, L.; Bolognesi, M. Archaeal Protoglobin Structure Indicates New Ligand Diffusion Paths and Modulation of Haem-reactivity. *EMBO Rep.* **2008**, *9*, 157–163. <https://doi.org/10.1038/sj.embor.7401153>.
- (49) Jancarik, J.; Kim, S. H. Sparse Matrix Sampling: A Screening Method for Crystallization of Proteins. *J. Appl. Crystallogr.* **1991**, *24*, 409–411. <https://doi.org/10.1107/S0021889891004430>.
- (50) Carson, M.; Johnson, D. H.; McDonald, H.; Brouillette, C.; DeLucas, L. J. His-Tag Impact on Structure. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2007**, *63*, 295–301. <https://doi.org/10.1107/S0907444906052024>.
- (51) Malakhov, M. P.; Mattern, M. R.; Malakhova, O. A.; Drinker, M.; Weeks, S. D.; Butt, T. R. SUMO Fusions and SUMO-Specific Protease for Efficient Expression and Purification of Proteins. *J. Struct. Funct. Genomics* **2004**, *5*, 75–86. <https://doi.org/10.1023/B:JSFG.0000029237.70316.52>.
- (52) Teh, A.-H.; Saito, J. A.; Baharuddin, A.; Tuckerman, J. R.; Newhouse, J. S.; Kanbe, M.; Newhouse, E. I.; Rahim, R. A.; Favier, F.; Didierjean, C.; Sousa, E. H. S.; Stott, M. B.; Dunfield, P. F.; Gonzalez, G.; Gilles-Gonzalez, M.-A.; Najimudin, N.; Alam, M. Hell's Gate Globin I: An Acid and Thermostable Bacterial Hemoglobin Resembling Mammalian Neuroglobin. *FEBS Lett.* **2011**, *585*, 3250–3258. <https://doi.org/10.1016/j.febslet.2011.09.002>.
- (53) Danley, D. E. Crystallization to Obtain Protein–Ligand Complexes for Structure-Aided Drug Design. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2006**, *62*, 569–575. <https://doi.org/10.1107/S0907444906012601>.
- (54) Ehrmann, F. R.; Stojko, J.; Metz, A.; Debaene, F.; Barandun, L. J.; Heine, A.; Diederich, F.; Cianférani, S.; Reuter, K.; Klebe, G. Soaking Suggests “Alternative Facts”: Only Co-Crystallization Discloses Major Ligand-Induced Interface Rearrangements of a Homodimeric tRNA-Binding Protein Indicating a Novel Mode-of-Inhibition. *PLoS One* **2017**, *12*, e0175723. <https://doi.org/10.1371/journal.pone.0175723>.
- (55) Müller, I. Guidelines for the Successful Generation of Protein–Ligand Complex Crystals. *Acta Crystallogr. Sect. D Struct. Biol.* **2017**, *73*, 79–92. <https://doi.org/10.1107/S2059798316020271>.
- (56) Sharon, D. A.; Mallick, D.; Wang, B.; Shaik, S. Computation Sheds Insight into Iron Porphyrin Carbenes' Electronic Structure, Formation, and N–H Insertion Reactivity. *J. Am. Chem. Soc.* **2016**, *138*, 9597–9610. <https://doi.org/10.1021/jacs.6b04636>.
- (57) Khade, R. L.; Zhang, Y. C–H Insertions by Iron Porphyrin Carbene: Basic Mechanism and Origin of Substrate Selectivity. *Chem. Eur. J.* **2017**, *23*, 17654–17658. <https://doi.org/10.1002/chem.201704631>.
- (58) Hayashi, T.; Tinzl, M.; Mori, T.; Kregel, U.; Proppe, J.; Soetbeer, J.; Klose, D.; Jeschke, G.; Reiher, M.; Hilvert, D. Capture and Characterization of a Reactive Haem–Carbenoid Complex in an Artificial Metalloenzyme. *Nat. Catal.* **2018**, *1*, 578–584. <https://doi.org/10.1038/s41929-018-0105-6>.
- (59) Lewis, R. D.; Garcia-Borràs, M.; Chalkley, M. J.; Buller, A. R.; Houk, K. N.; Kan,

- S. B. J.; Arnold, F. H. Catalytic Iron-Carbene Intermediate Revealed in a Cytochrome c Carbene Transferase. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 7308–7313. <https://doi.org/10.1073/pnas.1807027115>.
- (60) Tao, Z.; Goodisman, J.; Souid, A.-K. Oxygen Measurement via Phosphorescence: Reaction of Sodium Dithionite with Dissolved Oxygen. *J. Phys. Chem. A* **2008**, *112*, 1511–1518. <https://doi.org/10.1021/jp710176z>.
- (61) Wu, Z.; Kan, S. B. J.; Lewis, R. D.; Wittmann, B. J.; Arnold, F. H. Machine Learning-Assisted Directed Protein Evolution with Combinatorial Libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 8852–8858. <https://doi.org/10.1073/pnas.1901979116>.
- (62) Freitas, T. A. K.; Hou, S.; Dioum, E. M.; Saito, J. A.; Newhouse, J.; Gonzalez, G.; Gilles-Gonzalez, M.-A.; Alam, M. Ancestral Hemoglobins in Archaea. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 6675–6680. <https://doi.org/10.1073/pnas.0308657101>.
- (63) Pesce, A.; Bolognesi, M.; Nardini, M. Protoglobin. In *Advances in Microbial Physiology*; 2013; Vol. 63, pp 79–96. <https://doi.org/10.1016/B978-0-12-407693-8.00003-0>.
- (64) Shaw Stewart, P. D.; Kolek, S. A.; Briggs, R. A.; Chayen, N. E.; Baldock, P. F. M. Random Microseeding: A Theoretical and Practical Exploration of Seed Stability and Seeding Techniques for Successful Protein Crystallization. *Cryst. Growth Des.* **2011**, *11*, 3432–3441. <https://doi.org/10.1021/cg2001442>.
- (65) Till, M.; Robson, A.; Byrne, M. J.; Nair, A. V.; Kolek, S. A.; Shaw Stewart, P. D.; Race, P. R. Improving the Success Rate of Protein Crystallization by Random Microseed Matrix Screening. *J. Vis. Exp.* **2013**, No. 78, 1–7. <https://doi.org/10.3791/50548>.
- (66) Yamada, K. D.; Kunishima, N.; Matsuura, Y.; Nakai, K.; Naitow, H.; Fukasawa, Y.; Tomii, K. Designing Better Diffracting Crystals of Biotin Carboxyl Carrier Protein from *Pyrococcus horikoshii* by a Mutation Based on the Crystal-Packing Propensity of Amino Acids. *Acta Crystallogr. Sect. D Struct. Biol.* **2017**, *73*, 757–766. <https://doi.org/10.1107/S2059798317010932>.
- (67) Derewenda, Z. S. Application of Protein Engineering to Enhance Crystallizability and Improve Crystal Properties. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 604–615. <https://doi.org/10.1107/S090744491000644X>.
- (68) Sevrioukova, I. F.; Li, H.; Zhang, H.; Peterson, J. A.; Poulos, T. L. Structure of a Cytochrome P450-Redox Partner Electron-Transfer Complex. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 1863–1868. <https://doi.org/10.1073/pnas.96.5.1863>.
- (69) Joyce, M. G.; Ekanem, I. S.; Roitel, O.; Dunford, A. J.; Neeli, R.; Girvan, H. M.; Baker, G. J.; Curtis, R. A.; Munro, A. W.; Leys, D. The Crystal Structure of the FAD/NADPH-Binding Domain of Flavocytochrome P450 BM3. *FEBS J.* **2012**, *279*, 1694–1706. <https://doi.org/10.1111/j.1742-4658.2012.08544.x>.
- (70) Zhang, H.; Yokom, A. L.; Cheng, S.; Su, M.; Hollenberg, P. F.; Southworth, D. R.; Osawa, Y. The Full-Length Cytochrome P450 Enzyme CYP102A1 Dimerizes at Its Reductase Domains and Has Flexible Heme Domains for Efficient Catalysis. *J. Biol. Chem.* **2018**, *293*, 7727–7736. <https://doi.org/10.1074/jbc.RA117.000600>.
- (71) Su, M.; Chakraborty, S.; Osawa, Y.; Zhang, H. Cryo-EM Reveals the Architecture of the Dimeric Cytochrome P450 CYP102A1 Enzyme and Conformational Changes

- Required for Redox Partner Recognition. *J. Biol. Chem.* **2020**, *295*, 1637–1645. <https://doi.org/10.1074/jbc.RA119.011305>.
- (72) Grygorenko, O. O.; Artamonov, O. S.; Komarov, I. V.; Mykhailiuk, P. K. Trifluoromethyl-Substituted Cyclopropanes. *Tetrahedron* **2011**, *67*, 803–823. <https://doi.org/10.1016/j.tet.2010.11.068>.
- (73) Khanal, A.; Yu McLoughlin, S.; Kershner, J. P.; Copley, S. D. Differential Effects of a Mutation on the Normal and Promiscuous Activities of Orthologs: Implications for Natural and Directed Evolution. *Mol. Biol. Evol.* **2015**, *32*, 100–108. <https://doi.org/10.1093/molbev/msu271>.
- (74) Murciano-Calles, J.; Romney, D. K.; Brinkmann-Chen, S.; Buller, A. R.; Arnold, F. H. A Panel of TrpB Biocatalysts Derived from Tryptophan Synthase through the Transfer of Mutations That Mimic Allosteric Activation. *Angew. Chem., Int. Ed.* **2016**, *55*, 11577–11581. <https://doi.org/10.1002/anie.201606242>.
- (75) Li, Y.; Drummond, D. A.; Sawayama, A. M.; Snow, C. D.; Bloom, J. D.; Arnold, F. H. A Diverse Family of Thermostable Cytochrome P450s Created by Recombination of Stabilizing Fragments. *Nat. Biotechnol.* **2007**, *25*, 1051–1056. <https://doi.org/10.1038/nbt1333>.
- (76) Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Cassarino, T. G.; Bertoni, M.; Bordoli, L.; Schwede, T. SWISS-MODEL: Modelling Protein Tertiary and Quaternary Structure Using Evolutionary Information. *Nucleic Acids Res.* **2014**, *42*, W252–W258. <https://doi.org/10.1093/nar/gku340>.
- (77) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat. Methods* **2009**, *6*, 343–345. <https://doi.org/10.1038/nmeth.1318>.
- (78) Kille, S.; Acevedo-Rocha, C. G.; Parra, L. P.; Zhang, Z.-G.; Opperman, D. J.; Reetz, M. T.; Acevedo, J. P. Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synth. Biol.* **2013**, *2*, 83–92. <https://doi.org/10.1021/sb300037w>.
- (79) Barr, I.; Guo, F. Pyridine Hemochromagen Assay for Determining the Concentration of Heme in Purified Protein Solutions. *Bio. Protoc.* **2015**, *5*. <https://doi.org/10.21769/BioProtoc.1594>.
- (80) Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 125–132. <https://doi.org/10.1107/S0907444909047337>.
- (81) Terwilliger, T. C.; Adams, P. D.; Read, R. J.; McCoy, A. J.; Moriarty, N. W.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Zwart, P. H.; Hung, L.-W. Decision-Making in Structure Solution Using Bayesian Estimates of Map Quality: The PHENIX AutoSol Wizard. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2009**, *65*, 582–601. <https://doi.org/10.1107/S0907444909012098>.
- (82) Adams, P. D.; Afonine, P. V.; Bunkóczy, 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. PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 213–221. <https://doi.org/10.1107/S0907444909052925>.

- (83) Evans, P. R.; Murshudov, G. N. How Good Are My Data and What Is the Resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69*, 1204–1214. <https://doi.org/10.1107/S0907444913000061>.
- (84) Bunkóczi, G.; Echols, N.; McCoy, A. J.; Oeffner, R. D.; Adams, P. D.; Read, R. J. Phaser.MRage : Automated Molecular Replacement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69*, 2276–2286. <https://doi.org/10.1107/S0907444913022750>.
- (85) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 486–501. <https://doi.org/10.1107/S0907444910007493>.

DIVERSITY-ORIENTED ENZYMATIC SYNTHESIS OF  
CYCLOPROPANE BUILDING BLOCKS

Content in this chapter is adapted from submitted work:

† denotes equal contribution

Wittmann, B. J. †; **Knight, A. M.** †; Hofstra, J. L.; Reisman, S. E.; Kan, S. B. J.; Arnold, F. H. Diversity-oriented enzymatic synthesis of cyclopropane building blocks. *ACS Catal.* **2020**, *10*, 7112–7116. DOI: 10.1021/acscatal.0c01888.

A.M.K. and S.B.J.K. conceived the project. A.M.K. prepared compilation plates for the initial activity screen and determined enzymes with initial activity. A.M.K. designed, constructed, and screened the libraries for the cis-selective lineage. A.M.K. ran multi-gram scale enzymatic reactions. A.M.K. developed the column-free purification procedure via trifluoroboration methods. A.M.K. and B.J.W wrote the manuscript with input from all authors.

## Abstract

While biocatalysis is increasingly incorporated into drug development pipelines, most biocatalysts in pharmaceutical production were engineered after drug development was completed. Incorporating biocatalysts into early stages of drug discovery requires the ability to rapidly build molecular diversity. By engineering proteins to produce chiral motifs with derivatizable functional handles, biocatalysts can be used to help generate diverse building blocks for drug discovery. Here we show the engineering of the heme protein *Rhodothermus marinus* nitric oxide dioxygenase (*RmaNOD*) for the diastereodivergent formation of a pinacolboronate-substituted cyclopropane, which can be derivatized using cross-coupling reactions to generate diverse stereopure cyclopropane building blocks. These cyclopropylboronates can be prepared at gram scale using lyophilized biocatalysts and purified in a column-free approach through distillation or trifluoroboration.

### 3.1 Leveraging biocatalysis in the drug development pipeline

Since the beginning of the second wave of biocatalysis,<sup>1</sup> enzymes have been engineered for and applied to the synthesis of myriad molecules of industrial interest, from low-value, high-volume fuels<sup>2,3</sup> and specialty chemicals<sup>4</sup> to value-added fine chemicals such as flavors and fragrances.<sup>5</sup> One application where the advantages of biocatalysis have truly shone is the synthesis of pharmaceutical compounds. The chemo-, regio-, and stereoselectivity required in the synthesis of complex pharmaceuticals can be met by a single enzyme engineered for this specific function. The high value of pharmaceutical compounds offsets the substantial cost of protein engineering and biocatalyst production. It is therefore unsurprising that intensive academic and industrial research and development have gone into the discovery and engineering of enzymes to synthesize pharmaceutical compounds.<sup>6-8</sup>

The early stages of a drug discovery and development pipeline require the ability to rapidly generate tens to hundreds of grams of a candidate drug for further studies.<sup>9</sup> For biocatalysts to outperform the competing chemical processes, the speed at which they can carry a project from initial activity determination to industrial production is crucial.<sup>8-10</sup> The field of protein engineering has made great strides toward faster and more efficient biocatalyst development, which has begun to bring biocatalysis into the drug development pipeline, albeit after the drug candidate has been identified and is in clinical trials. Integrating biocatalysis even earlier in the drug development pipeline requires examining and overcoming current limitations of biocatalysts, namely their narrow substrate scope and the time required to engineer them for the desired activity. To determine where biocatalysts can fit earlier in the drug discovery pipeline, we must look at the workflow of the drug discovery pipeline and examine what is needed at each stage (Figure 3-1). The drug discovery development pipeline starts with identifying a target for treatment, typically a key protein for which an inhibitor could be developed. Following this, a screen to assay fitness (e.g. protein inhibition) is developed, and large libraries of small molecules are tested. Molecules identified at this stage (lead fragments) are used as a core motif for further derivatization and optimization. When a molecule has been optimized and has met the features required for its function, it will move into clinical trials for approval by the FDA and other regulatory agencies.



Figure 3-1. Drug development pipeline. Target identification: determination of which biomolecule of interest will be targeted for the therapeutic effect. Compound screening and hit validation: screen small-molecule libraries and determine which have the desired therapeutic effect. Lead identification and optimization: lead fragments are derivatized and decorated to improve their therapeutic effect. Clinical trials: drug candidates are tested for their efficacy and safety to be approved by regulatory bodies (e.g. FDA).

The lead identification and optimization stage of drug discovery, in which lead fragments are diversified to improve their efficacy, is a highly attractive stage for the incorporation of biocatalysts. At this stage, a lead fragment might already require several synthetic steps to produce and generating decorated analogs requires either selective functionalization of the lead fragment or modifying the synthetic route to incorporate diverse functional groups. Enzymes have been shown to excel at selective late-stage functionalization in natural product and pharmaceutical syntheses, installing functional groups with high selectivity.<sup>9,11</sup> Developing a set of enzymes to selectively produce dozens or hundreds of decorated analogs of the lead fragment will take a long time – at a stage of drug development in which speed is a crucial factor. Instead, employing an enzyme to produce a single analog of the complex chiral lead fragment with high selectivity, which is readily derivatized by commonly used methods in diversity-oriented synthesis, would then allow for the use of standard medicinal chemistry diversification libraries and pipelines which have been developed and optimized over decades (Figure 3-2).<sup>12</sup>

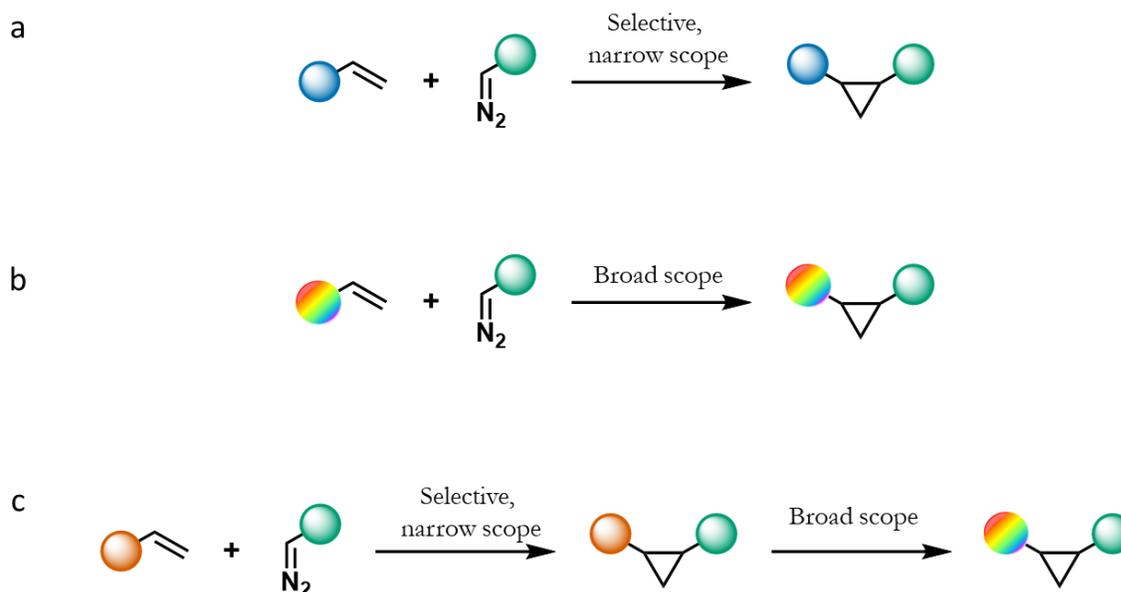


Figure 3-2. Expanding cyclopropane diversity through selective installation of derivatizable functional groups. (a) Enzymes can catalyze reactions with high specificity and selectivity, but often suffer from a narrow substrate scope. (b) Combinatorial chemistry approaches can be used to install diverse functional groups, but can lack the stereoselectivity control needed to generate the target compound. (c) Enzymes with high activity and selectivity for a single transformation can be coupled with combinatorial chemistry approaches to generate a diverse array of substituted cyclopropanes.

Combining traditional synthesis and enzymatic synthesis has been a cornerstone of the biocatalysis field for decades; these chemoenzymatic approaches use the specificity and selectivity of enzymes together with well-established synthetic steps, where the optimal catalyst for each step is used. In many cases, a key synthetic step which sets one or more stereocenters is performed enzymatically, preceded or followed by chemical synthetic steps.<sup>13-15</sup> Coupling enzymes with combinatorial chemistry has precedent as well; researchers have shown the use of chemoenzymatic approaches to diversify small molecules, with decarboxylases generating substrates for olefin metathesis,<sup>16</sup> halogenases generating substrates for cross-coupling reactions,<sup>17,18</sup> and carbene transferases installing nitrile moieties for conversion to a variety of functional groups.<sup>19</sup>

Substituted cyclopropanes are prevalent in pharmaceutical and agrochemical compounds.<sup>20-22</sup> The ability to rapidly produce derivatives of a stereopure cyclopropane could be used in lead fragment optimization to assist future drug discovery and development efforts. Boronate ester moieties are ubiquitous in medicinal chemistry due to their robust activity in the presence of a wide range of functional groups and their efficacy in convergent synthesis of complex molecules.<sup>12</sup> Therefore, I envisioned that an enzymatically produced pinacolboronate (Bpin)-substituted chiral cyclopropane could be used as a substrate for Suzuki-Miyaura cross-coupling reactions to generate a diverse array of substituted cyclopropanes. In light of our previous studies that showed that heme proteins can be engineered to generate the cyclopropane-containing pharmaceuticals,<sup>23-26</sup> I chose to develop a chemoenzymatic strategy to produce a cyclopropane motif with a functional handle, which could then be derivatized to form substituted chiral cyclopropanes. To realize this approach, I set out to engineer heme proteins to catalyze the stereoselective cyclopropanation via carbene transfer of ethyl diazoacetate (EDA) and vinylboronic acid pinacol ester, generating 2-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-cyclopropanecarboxylic acid ethyl ester (Figure 3-3).



Figure 3-3. Proposed enzymatic reaction of vinylboronic acid pinacol ester (**1**) and ethyl diazoacetate (**2**) to form the cyclopropylboronate ester (**3**).

### 3.2 Initial activity determination and enhancement via directed evolution

The first step in a directed evolution campaign is the identification of parent-protein candidates with promiscuous activity for the reaction of interest. The initial-activity screen for the reaction of **1** and **2** was carried out using a panel of heme-binding proteins which had been engineered for a variety of heme-catalyzed, new-to-nature reactions.<sup>27</sup> This collection of enzymes includes cytochromes P450<sub>BM3</sub>, serine-ligated cytochromes P411<sub>BM3</sub>,

cytochromes *c*, and globins (Supplementary Information, Table 3-1). The activity was determined by GC-MS, with mass fragmentation and retention time of the cyclopropylboronate compared to the authentic product standard. The best initial activities for the formation of *cis*- and *trans*- cyclopropylboronate were shown by *Aeropyrum pernix* protoglobin W59A Y60G F145W (*ApePgb* AGW) and *Rhodothermus marinus* nitric oxide dioxygenase Q52A (*RmaNOD* Q52A), respectively. Both variants were developed during the engineering for the cyclopropanation of linear, aliphatic alkenes.<sup>28</sup> These proteins were also tested for their ability to use alternative commercially available boronate esters such as MIDA ester and dibutyl ester, but only the pinacol ester showed detectable cyclopropanation activity. The initial stereoselectivity was good, with *ApePgb* AGW catalyzing the formation of the cyclopropylboronate ester with 96:4 diastereomeric ratio (dr) and 96% enantiomeric excess (ee, determined by chiral GC-FID).

Single site-saturation libraries were generated for *ApePgb* AGW at amino-acid residues 63, 86, and 90. Hits were identified in screening, but subsequent validation of protein variants identified as hits in screening failed to verify improved variants. In parallel, screening *RmaNOD* Q52A for enhanced *trans*-cyclopropylboronate activity yielded variant *RmaNOD* Y32T Q52A, with 420 TTN and an inversion of diastereoselectivity from 17:83 *cis:trans* to 90:10 *cis:trans* (Figure 3-4). *RmaNOD* Y32T Q52A produced the same major *cis*-enantiomer as *ApePgb* AGW with 99% ee, but with higher activity. We therefore used the *RmaNOD* Y32T Q52A variant for further evolution.

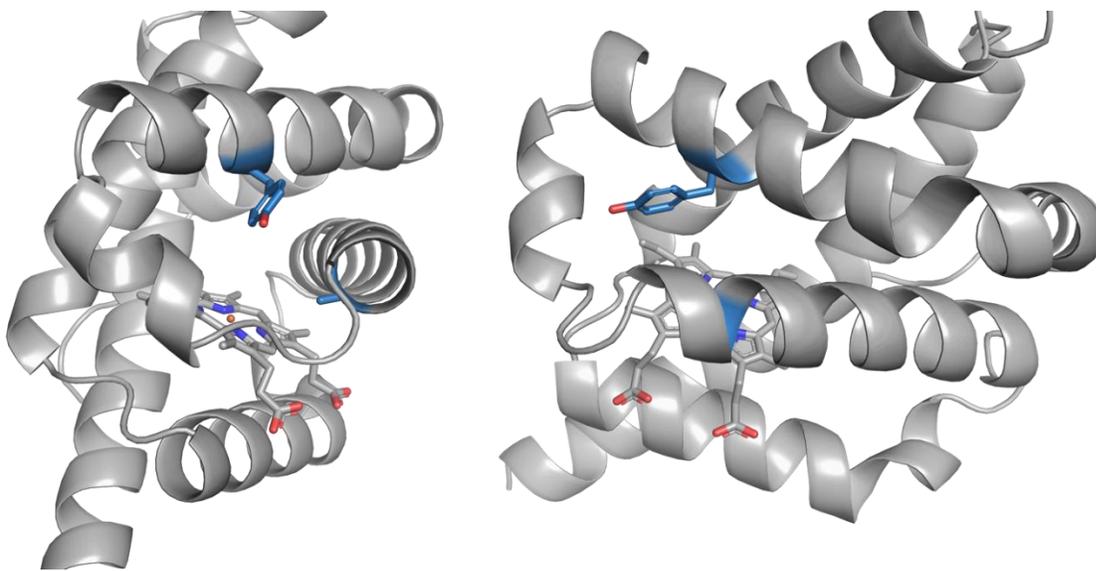


Figure 3-4. Location of residues 32 and 52 in the *RmaNOD* scaffold, based on the *RmaNOD* Q52V crystal structure (PDB ID: 6WK3). Residues Y32 and A52 are shown in blue. The heme-bound acetate is omitted for clarity.

As we had thus far only targeted a small subset of the active-site residues, we expanded the search to include several more active-site residues. In order to screen a larger sequence space without increasing the overall screening time with a medium-throughput GC-based screening method, we opted to reduce redundancy by screening 44 clones per library rather than the usual 88. Screening only 44 clones for a 22-member library still has an 86% library coverage (compared to 98% coverage for 88 clones per library) while allowing for higher sequence-space coverage (Supplementary Information, Figure 3-11).

Using *RmaNOD* Y32T Q52A as parent, in a second round of site-saturation mutagenesis, we targeted residues M31, F36, Y39, F46, L48, P49, I53, L56, R79, R86, V89, and L101, screening 44 colonies per library (Figure 3-5). This round of engineering yielded beneficial mutations at positions 39, 48, 79, and 86. Residues 39 and 48 are in the distal heme pocket, but R79 and R86 coordinate with the heme carboxylate; the underlying cause for the activity enhancement at these positions is unclear. It is noteworthy that the mutations at R79 enhanced activity at the cost of diastereoselectivity, while mutation of residues 39 and 48 enhanced diastereoselectivity. We therefore recombined the mutations at these positions; the

combination of mutations Y32T, Y39H, L48R, Q52A, and R79W was found under screening conditions to both improve activity and stereoselectivity. This variant, *RmaNOD* THRAW, displays 1300 total turnovers (TTN) while maintaining high dr (94:6 *cis:trans*) and greater than 99% ee (Figure 3-6a).

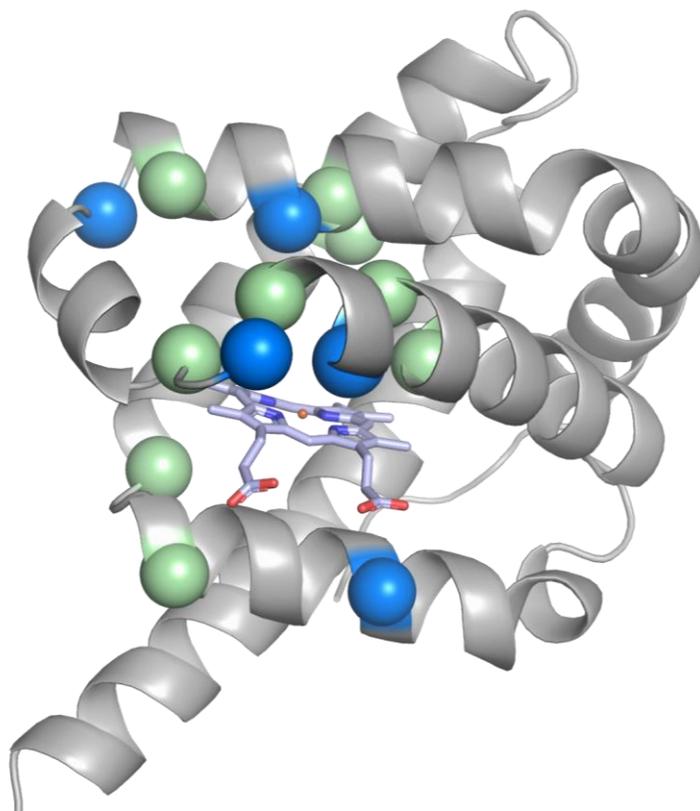


Figure 3-5. *RmaNOD* scaffold (PDB ID: 6WK3) displaying residues targeted in the *cis*-selective lineage. Residues targeted for mutagenesis are shown as  $\alpha$ -carbon spheres. Positions at which the final variant contained a mutation relative to *RmaNOD* WT are shown in blue; other targeted positions are shown in green.

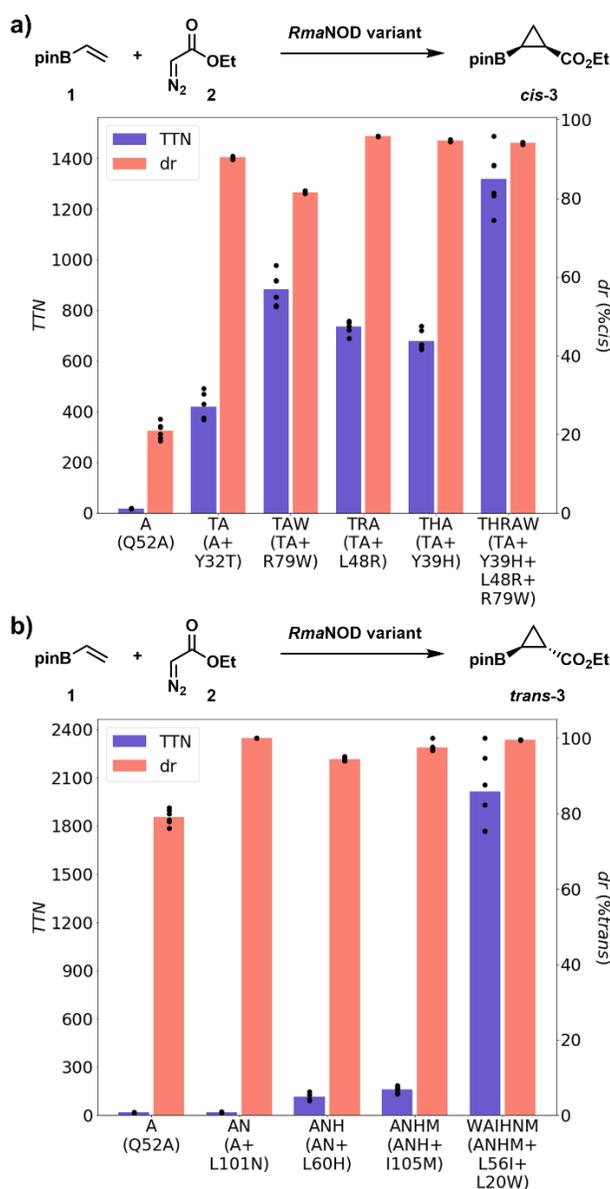


Figure 3-6. Activity and diastereoselectivity of the *cis*- and *trans*- specific lineages for the formation of **3**. Variant names are listed as parent protein plus new amino-acid substitution(s). Data points are from biological duplicates of technical triplicates.

In parallel to the engineering which generated the *cis*-selective variant *RmaNOD* THRAW, we targeted the *RmaNOD* scaffold to engineer a *trans*-selective protein variant. The *trans*-selective lineage evolution was led by my collaborator, Bruce Wittmann. Site-saturation mutagenesis libraries were generated at positions 31, 36, 46, 49, 53, 56, 79, 89, and 101 using

*RmaNOD* Q52A as the parent (Figure 3-7). We screened 44 colonies per library and identified mutation L101N, which improved diastereoselectivity for the *trans*-product without significantly affecting enzyme activity (Figure 3-6b). In the second generation, site-saturation mutagenesis libraries were generated at positions 35, 42, 53, 56, 60, 79, 96, 97, and 125, using *RmaNOD* Q52A L101N as the parent. Here, we identified beneficial mutations at positions 56 and 60, with L60H as the most beneficial one (defined as enhancing activity the most while retaining diastereoselectivity).

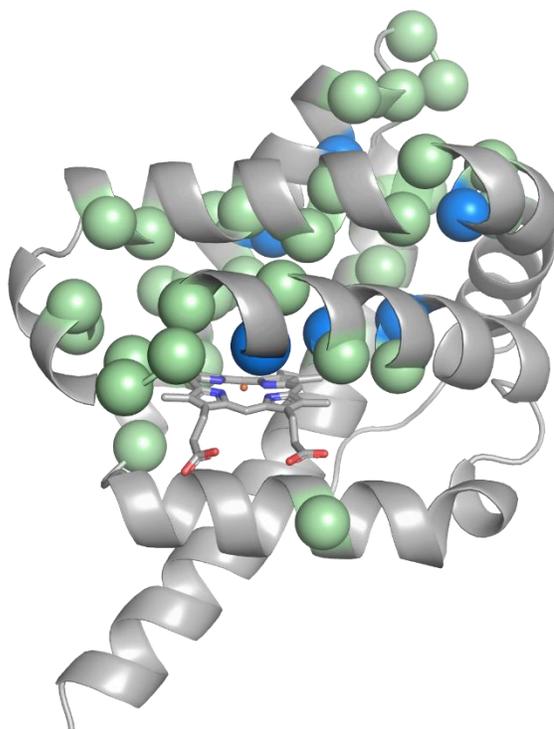


Figure 3-7. *RmaNOD* scaffold (PDB ID: 6WK3) displaying residues targeted in the *trans*-selective lineage. Residues targeted for mutagenesis are shown as  $\alpha$ -carbon spheres. Positions at which the final variant contained a mutation relative to *RmaNOD* WT are shown in blue; other targeted positions are shown in green.

Because the activity of *RmaNOD* Q52A L60H L101N was still low, we continued with two further rounds of site-saturation mutagenesis. Using *RmaNOD* Q52A L60H L101N as the parent, in the first round, we targeted positions 27, 28, 43, 47, 48, 55, 59, 94, 105, 111, and 121, identifying beneficial mutations at positions 55 and 105, with mutation I105M being

most beneficial. In the final round of site-saturation mutagenesis, we increased the coverage of the first- and second-shell residues by targeting positions 16, 19, 20, 23, 24, 27, 28, 31, 36, 46, 49, 56, 79, 98, 100, 108, 109, 112, 117, and 118, now using *RmaNOD* Q52A L60H L101N I105M as the parent. From these libraries, we discovered the beneficial mutations L20W, M31F, L56A, and L56I. In a recombination library of these four mutations, we identified *RmaNOD* L20W Q52A L56I L60H L101N I105M as the combination with the greatest enhancement in activity with 2300 TTN, while maintaining greater than 99:1 *trans:cis* dr and 99% ee (Figure 3-6b).

### 3.3 Derivatization of cyclopropylboronates via cross-coupling reactions

We successfully engineered heme proteins to catalyze the formation of one *cis* enantiomer and one *trans* enantiomer of the cyclopropylboronate, but we still lacked enzymes for the other two stereoisomers. We then looked for a chemoenzymatic solution to achieve stereodivergence. We hypothesized that a combination of stereoretentive and stereoinvertive cross-coupling reactions could be used to access each of the four stereoisomers of the derivatized cyclopropyl ester (Figure 3-8). This would require the stereocenter adjacent to the ester to have inverted to form the opposite diastereomer. In contrast, inversion at the stereocenter adjacent to the Bpin moiety would not lead to all four stereoisomers through stereoretentive and stereoinvertive cross-coupling reactions. Using chiral GC, we determined that *RmaNOD* Y32T Q52A and *ApePgb* AGW formed the same stereoisomer of the cyclopropylboronate product. As *ApePgb* AGW and *RmaNOD* Q52V were known to catalyze the cyclopropanation of 1-octene and EDA to form stereoisomers differing at the ester stereocenter, we hypothesized that the *cis*- and *trans*- *RmaNOD* variants also had inversion at the ester stereocenter, and could therefore be used to access all substituted cyclopropane stereoisomers.

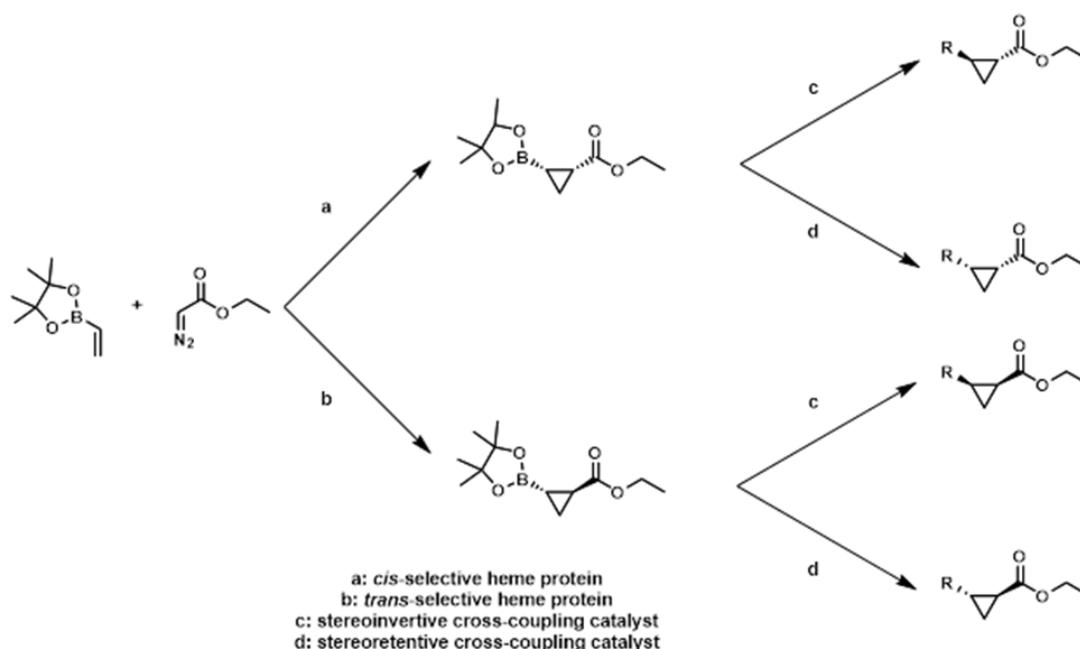


Figure 3-8. Proposed access to all four stereoisomers of derivatized cyclopropane esters. Accessing all four stereoisomers requires the ability to either enzymatically generate all cyclopropylboronate stereoisomers or generate cyclopropylboronate esters with an inverted stereocenter at the ester and access to stereoretentive and stereoinvertive cross-coupling reactions.

Biscoe and coworkers reported reaction conditions for the enantiodivergent cross-coupling of diverse trifluoroborates.<sup>29</sup> As the interconversion between pinacolborane and trifluoroborate functional groups is a well-established single-step synthesis,<sup>30</sup> we hypothesized that applying these cross-coupling conditions to trifluoroborated derivatives of our cyclopropylboronates would enable access to all four stereoisomers without requiring the discovery of two more enzymes with complementary stereoselectivity. To this end, we began a collaboration with Julie Hofstra, a graduate student studying cross-coupling reactions in the Reisman lab at Caltech. To test the trifluoroborate-based cross-coupling reactions, she prepared the trifluoroborate-substituted cyclopropane from the *RmaNOD* THRAW enzymatic product (Figure 3-9a). Unlike the cyclopropylboronate, which was an oil, the trifluoroborated product was solid; we hypothesized that this product could be crystallized to confirm the absolute configuration of our product. Indeed, the trifluoroborate-substituted cyclopropane from *RmaNOD* THRAW crystallized and its structure was solved

via X-ray crystallography; the absolute configuration determination of (*1R,2S*) is in agreement with the configuration proposed by analogy to the cyclopropanation of 1-octene and EDA<sup>28</sup> and lends further evidence toward the hypothesis that the stereocenter inverted between the *cis*- and *trans*-*Rma*NOD biocatalysts is adjacent to the ester. We attempted to crystallize the *trans*-trifluoroborated cyclopropane product to confirm this, but crystallization attempts were unsuccessful.

With trifluoroborate-substituted cyclopropanes in hand, the stereodivergent cross-coupling reaction conditions were tested. While the stereoretentive cross-coupling reactions proceeded well, the cross-coupling reaction conditions reported by Biscoe and coworkers to be stereoinvertive resulted in either stereoretentive cross-coupling or no reaction. Discussions with the authors revealed that the steric hindrance of small carbocycles prevented stereoinvertive cross-coupling; as such only the stereoretentive cross-coupling reactions would demonstrate activity. Despite the challenges encountered in performing the stereoinvertive cross-coupling reactions, both trifluoroborate- and Bpin-substituted cyclopropanes can be used in stereoretentive cross-coupling reactions to generate substituted cyclopropanes (Figure 3-9b). A potential future application of enzymatic diversity-oriented synthesis would be the installation of trifluoroborate moieties on molecules other than sterically hindered carbocycles, thus allowing use of the enantiodivergent cross-coupling reactions.<sup>29</sup>

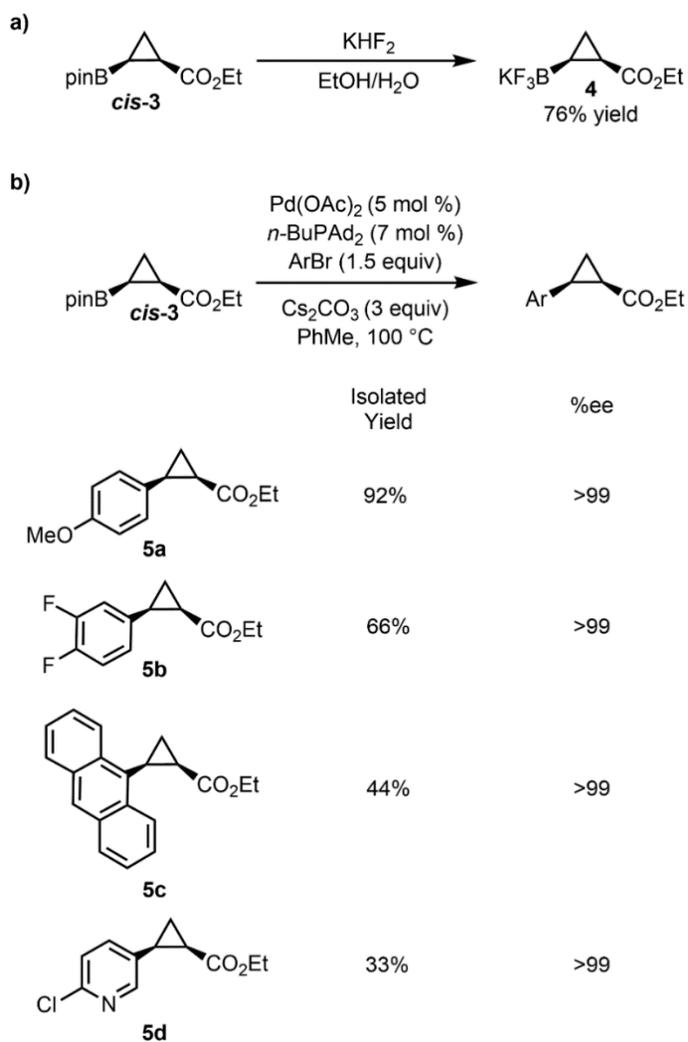


Figure 3-9. Derivatizations of **3**. (a) **3** can be reacted with potassium bifluoride to form the trifluoroborated salt **4**. (b) Cyclopropylboronate **3** was converted to multiple arylboronates through Pd-catalyzed cross-coupling reactions while retaining high enantioselectivity. The yields reported are for the cross-coupling step starting from isolated (*1R,2S*)-**3**.

### **3.4 Reaction engineering for preparative-scale production of cyclopropylboronates**

The efficient production and purification of large quantities of a compound are crucial for a diversity-oriented synthesis approach with the goal to generate derivatives of a lead fragment. I therefore investigated potential improvements to both the reaction engineering and downstream processing of the enzymatic product.

#### **3.4.1 Buffer optimization**

During directed evolution for cyclopropylboronate production, we noticed a nonlinearity in our calibration curves, with decreased rate of response at low product concentrations. In preparative-scale reactions, we observed the hydrolysis of the pinacolboronate to the free boronic acid, and thus attributed the nonlinearity of our calibration curves to this loss in product formation. As M9-N buffer has a high concentration of magnesium (which could promote hydrolysis), I hypothesized that buffer optimization could remove the undesired hydrolysis. Using MOPS buffer in place of M9-N did slightly decrease the nonlinearity (Figure 3-10), but the hydrolyzed cyclopropylboronic acid was still observed in preparative-scale reactions. Instead of furthering buffer optimization efforts, we focused on approaches to reprotect the cyclopropylboronic acid. A mixture of cyclopropylboronate ester and cyclopropylboronic acid can be efficiently reprotected in organic solvent with the addition of excess (2 equivalents) pinacol in the presence of magnesium sulfate.

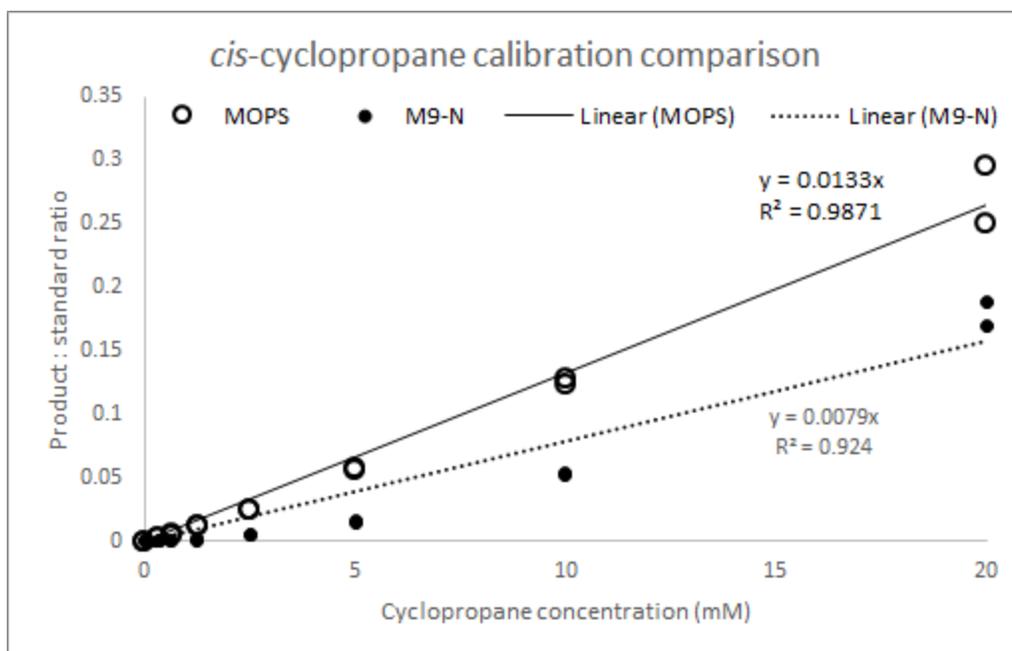


Figure 3-10. Nonlinear response behavior of the cyclopropylboronate compounds.

### 3.4.2 Lyophilized biocatalyst formulation

To enable access to these engineered biocatalysts for chemists without cell culture experience, I tested the formulation of the enzymes as lyophilized whole cells. Proteins have long been lyophilized to improve their shelf life and storage,<sup>31</sup> and many commercially available enzymes sold today are formulated as lyophilized powders. Rather than preparing biocatalysts as fresh cells each time a reaction is required, the biocatalyst is produced in a large-scale fermentation, which eliminates batch-to-batch variability in specific expression and activity levels and allows for robust process optimization. Lyophilizing biocatalysts as whole *Escherichia coli* cells rather than cell lysate or purified protein decreases cost by removing additional lysis and purification steps, removes the requirement to exogenously add expensive redox cofactors, and in some cases provides enhanced stability to the biocatalyst relative to purified lyophilisate.<sup>32,33</sup> I therefore prepared large-scale cultures (1 L) of *RmaNOD THRAW* and *RmaNOD WAIHNM* and lyophilized the whole-cell pellets. The resultant lyophilisate was ground into a fine powder with a mortar and pestle to facilitate weighing and adding catalyst to reaction vessels. The lyophilized cell powder could be dissolved into aqueous buffer and remained active for cyclopropylboronate formation.

Lyophilized biocatalysts allow for reactions being performed in the presence of minimal amounts of water. While water is viewed as a green solvent, starting the reaction in it simply adds a second waste stream if the downstream processing requires extracting the compounds into organic solvent. Biocatalysts have been used extensively in organic solvent systems. Lipase reactions are a great example, as their reaction equilibrium is dependent on the ratio of water and substrate. Performing reactions directly in organic solvent is also beneficial in cases where the substrate is poorly soluble or unstable in water. Rother and others have shown multiple examples of the efficacy of biocatalytic reactions run in organic solvent and 1-5% water using lyophilized whole cells, either directly in solution<sup>32</sup> or within “teabag” membranes for easy catalyst removal.<sup>34</sup> Analytical-scale experiments with lyophilized *RmaNOD* THRAW and *RmaNOD* WAIHNM cells in methyl *tert*-butyl ether with 1-5% aqueous buffer added showed that the biocatalysts were still active under micro-aqueous conditions, though the drastically different reaction conditions prevent product extraction and comparison to the product standard calibration curves generated from aqueous samples. Further work on micro-aqueous preparative-scale reaction optimization could enable an alternative biocatalytic synthesis and product isolation workflow; this could be particularly promising for cyclopropylboronate formation, as performing the reaction under low water content has the potential to decrease pinacolboronate hydrolysis.

### 3.4.3 EDA addition optimization

In scale-up reactions, whether with fresh or lyophilized whole cells, a pad of cell debris would form at the surface where the EDA was being added dropwise. Switching to a long needle that inserted into the solution instead of dropping EDA on top of it prevented the formation of this cell debris pad. I noted the distinct formation of gas bubbles at the point the needle tip inserted into solution, which indicated that the EDA was rapidly consumed at the point of addition, evolving nitrogen at that point to form a bubble. I hypothesized that the local concentration of EDA at the needle tip (and thus the concentration of the EDA solution added via slow addition) was a crucial parameter. Previously, the concentration of EDA in the syringe was 800 mM, the same as we used in our screening assay. After testing the solubility limit of EDA in water:ethanol mixtures, I diluted the EDA solution being added

via slow addition to 163 mM (increasing the water content of the EDA solution) while maintaining the same final reaction volume. The resultant *Rma*NOD THRAW preparative-scale reaction had decreased EDA dimer formation and was used in the trifluoroboration product isolation described below to yield the trifluoroborated enzymatic product in 52% isolated yield.

#### **3.4.4. Efficient product isolation**

Downstream processing is a crucial consideration for any product synthesized in large scale; an efficient process to separate and purify the product from the catalyst and byproducts needs to be developed. At laboratory scale, this is typically performed with some combination of acid-base extractions, solvent extractions, and silica column chromatography.

In the case of cyclopropylboronate synthesis, the final reaction mixture contains whole-cell biocatalysts, cyclopropylboronate product, the EDA dimer byproducts diethyl maleate and diethyl fumarate, and unreacted starting material. Due to the product's partial hydrolysis to the cyclopropylboronic acid, silica column chromatography resulted in poor isolated yields. We examined alternative methods to efficiently isolate the cyclopropylboronate product. One option was distillation; the vinylboronate starting material, EDA dimers, and cyclopropylboronate product have different boiling points. Using rotary evaporation to remove starting material and a Vigreux distillation column to remove EDA dimers, 3.7 g of cyclopropylboronate were purified from a crude reaction extract in a column free approach (36% isolated yield, 41% isolated yield based on recovered starting material, 95:5 dr, 99% ee). Further iterative optimization of the reaction and downstream processing will improve the yields, while decreased starting material and byproducts in the final reaction mixture will make downstream processing more efficient.

A second option explored was the trifluoroboration of the cyclopropylboronate to the trifluoroborate-substituted cyclopropane. Previously, the trifluoroboration was carried out on purified cyclopropylboronate, but by performing the trifluoroboration on a crude reaction mixture, both the cyclopropylboronate and the free boronic acid can be converted to the trifluoroborate salt. This enables an alternative purification procedure for our cyclopropane

product. First, organic solvent is used to extract the cyclopropane, starting material, and EDA dimer byproducts from the aqueous reaction. Concentrating this mixture *in vacuo* removes the organic solvent and the volatile starting material. This remaining crude mixture, consisting of primarily cyclopropylboronate, cyclopropylboronic acid, and EDA dimer byproducts, is used as the starting material for a well-established trifluoroboration reaction.<sup>30</sup> The reaction can then be lyophilized to remove the more volatile organic compounds, and the cyclopropane product can be separated from the fluoride salt with an acetone extraction. This technique minimizes loss of cyclopropane product from pinacol deprotection and provides a column-free and scalable approach to the trifluoroborate salt. Using this approach, lyophilized whole cells harboring *RmaNOD* THRAW were used in a 1-mmol scale reaction to produce the cyclopropylboronate; the cyclopropylboronate was purified as the trifluoroborate salt with a 52% isolated yield across both steps. The trifluoroboration and distillation approaches can both be performed at large scale with further optimization.

### 3.4.5 Assessment of carbene precursor scope

The ability to generate diverse cyclopropanes from the cyclopropylboronate is limited in part by the carbene precursor used; we were therefore interested in testing diazo compounds in addition to EDA. *RmaNOD* THRAW and *RmaNOD* WAIHNM were examined for their ability to form and transfer five additional diazo esters and ketones to vinyl Bpin to produce alternate cyclopropylboronates. In each case, however, there was no formation of the cyclopropylboronate detected. To expand the cyclopropylboronate repertoire, it appears that a substrate-walk engineering approach is needed, where one first screens for carbene-transfer activity on styrene and other model substrates. Once the enzyme has been evolved to accept alternative diazo compounds as carbene precursors, these variants can be assayed for their ability to accept vinylboronate substrates. A similar approach was required for organoborane synthesis, with a separate engineered lineage for each diazo compound.<sup>35</sup>

## 3.5 Conclusions

Chemoenzymatic approaches can couple the specificity of biocatalysts with the broad substrate scope of small-molecule catalysts to rapidly generate chiral small molecule

libraries. Variants of *RmaNOD* were engineered for diastereodivergent cyclopropanation, stereoselectively forming **3**. We demonstrated that **3** could be further derivatized with standard Suzuki-Miyaura cross-coupling reactions. The reaction conditions and downstream processing protocol were modified to increase the isolated yield and improve the ease of isolation of **3**. Further condition optimization could improve the yield and/or productivity of the process, generating **3** at large scale.

### 3.6 Supplementary information for Chapter 3

#### Materials and instrumentation

Solvents and reagents were ordered from Sigma Aldrich, TCI, Aurum, or Alfa Aesar and used without further purification. GC-FID data were collected on Shimadzu GC-17A and Agilent 7820A GC systems. GC-MS data were collected on a Shimadzu GCMS-QP2010 SE and Agilent 7820A with 5977B MSD.  $^1\text{H}$  NMR,  $^{11}\text{B}$  NMR, and  $^{13}\text{C}$  NMR spectra were recorded on a Bruker Prodigy 400 MHz instrument;  $^1\text{H}$  NMR spectra were recorded at 400 MHz,  $^{11}\text{B}$  NMR spectra were recorded at 128 MHz, and  $^{13}\text{C}$  NMR spectra were recorded at 101 MHz. For  $^1\text{H}$  NMR and  $^{13}\text{C}$  NMR, chemical shifts were normalized to the deuterated solvent's proto impurity ( $\text{CHCl}_3$ :  $^1\text{H}$  NMR 7.26 ppm,  $^{13}\text{C}$  NMR 77.16 ppm; DMSO- $d_5$ :  $^1\text{H}$  NMR 2.50 ppm,  $^{13}\text{C}$  NMR 39.52 ppm).  $^{19}\text{F}$  NMR were recorded on a Varian 300 MHz instrument;  $^{19}\text{F}$  NMR spectra were recorded at 282 MHz.

#### Site-saturation library construction

Site-saturation mutagenesis was performed using the 22-codon trick.<sup>36</sup> Two methods were used for primer design and cloning: kinase-ligase-*DpnI* (KLD) (New England Biolabs) and Gibson assembly. In KLD cloning, primers were designed for blunt-end cloning with the KLD kit, where the degenerate codons were placed at the blunt end. The PCR product was then ligated via the KLD enzyme mix for 1 hour and used to transform electrocompetent cells. In the second approach, primers were designed for Gibson assembly, with the degenerate codons placed within either the forward or reverse primer. Two PCRs were performed for each library, the first containing a mixture of degenerate primers (12:9:1 NDT:VHG:TGG) and the second containing a reverse primer with a complementary

overhang to the degenerate primer. The second primer for each PCR is located in the ampicillin cassette to select against off-target PCR products. The two PCR products were gel-purified with the Zymoclean Gel DNA Recovery Kit (Zymo Research Corp, Irvine, CA) and ligated via Gibson assembly, followed by transformation.

### **Transformation of genes**

Electrocompetent *Escherichia coli* (*E. coli*, E. cloni EXPRESS BL21(DE3) cells, Lucigen, Middleton, WI) were transformed with Gibson assembly products using a Gene Pulser Xcell (Bio-Rad, Hercules, CA). SOC medium (750  $\mu\text{L}$ ) was added, and the cells were plated on LB-ampicillin (100  $\mu\text{g mL}^{-1}$ ) (LB-amp) agar plates. Overnight cultures (5 mL LB-amp in culture tubes) were grown at 37 °C and 230 rpm for 12–18 hours. Overnight cultures were used to inoculate flask cultures, prepare glycerol stocks, and isolate plasmids. Plasmids were isolated with Qiagen Miniprep kits, and the genes were sequence verified (T7 promoter / terminator sequencing primers, Laragen, Inc.).

### **96-Well plate library expression**

Single colonies from the LB-amp agar plates were picked using sterile toothpicks and grown in 400  $\mu\text{L}$  LB-amp in 2-mL 96-well deep-well plates at 37 °C, 250 rpm, 80% humidity overnight (12–18 hours). Libraries were screened at a depth of 44 colonies per library (Figure 3-11). Multi-channel pipettes were used to transfer 50  $\mu\text{L}$  of starter culture into deep-well plates containing 1 mL Hyperbroth (AthenaES) with 100  $\mu\text{g mL}^{-1}$  ampicillin (HB-amp) per well. The deep-well expression culture plate was incubated at 37 °C, 250 rpm, 80% humidity for 2.5 hours. The plate was then chilled on ice for 30 minutes. Expression of the protein of interest was induced with isopropyl- $\beta$ -D-thiogalactoside (IPTG), and cellular heme production was increased with 5-aminolevulinic acid (ALA); an induction mixture containing IPTG and ALA in HB-amp (50  $\mu\text{L}$ ) was added to each well for a final concentration of 0.5 mM IPTG and 1 mM ALA, 1.1 mL culture volume. The plate was incubated at 22 °C and 250 rpm overnight.

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B						Parents	sterile					
C												
D			Library 1							Library 2		
E												
F												
G												
H												

Figure 3-11. Representative 96-well plate layout for screening two libraries per plate. Keeping four positive controls (parent protein) and four negative controls (sterile wells) allowed for 44 colonies per library to be screened.

### 96-Well plate reactions and screening

96-Well deep-well plates containing *E. coli* expressing a protein variant were centrifuged at  $4000 \times g$  for 10 minutes at 4 °C. The supernatant was discarded, and pellets were resuspended in nitrogen-free M9 minimal media (M9-N, 380  $\mu$ L) and brought into a Coy vinyl anaerobic chamber (0–20 ppm O<sub>2</sub>). To each well were added 4,4,5,5-tetramethyl-2-vinyl-1,3,2-dioxaborolane (**1**) (10  $\mu$ L, 400 or 800 mM in ethanol) and EDA (**2**) (10  $\mu$ L, 400 or 800 mM in ethanol) and sealed with an adhesive foil cover. The reactions were shaken in the Coy chamber at room temperature for 16 hours. The plate was then removed from the Coy chamber, and the reactions were worked up: to each reaction, internal standard (20 mM 4-methyl anisole in 1:1 hexane:ethyl acetate, 20  $\mu$ L) and 1:1 hexane:ethyl acetate (480  $\mu$ L) were added. A silicone plate sealing mat (AWSM1003S, ArcticWhite) was used to cover the plate, and the plate was vortexed. The plate was then centrifuged ( $4000 \times g$ , 10 minutes) and the organic layer was transferred to GC vial inserts in GC vials. The samples were assayed for product formation via GC.

Wells with improved activity relative to the parent protein were streaked out onto LB-amp agar plates. A single colony from the plate was picked and grown in 5 mL LB-amp overnight (230 rpm, 37 °C). These overnight cultures were used in flask protein expression and small-scale biocatalytic reactions to verify enhanced activity and/or selectivity relative to the parent sequence.

### **Medium- and large-scale protein expression**

Cultures of 50 mL HB-amp (50 mL in 250-mL unbaffled flasks or 1 L in 2.8-L unbaffled flasks) were inoculated 1% (v/v) with saturated overnight cultures and shaken in an Inova 42 shaker at 230 rpm, 37 °C or an NFors shaker at 160 rpm, 37 °C. At an approximate OD<sub>600</sub> of 1.0 – 1.2, cultures were chilled on ice for 30 minutes. Protein expression was induced with 1 M IPTG to a final concentration of 0.5 mM, and heme production was enhanced with supplementation of 1 M ALA to a final concentration of 1 mM. The cultures were shaken in an Innova 42 shaker at 230 rpm (160 rpm for Infors) at 22 °C overnight (18–24 hours). Cells were transferred to 50-mL Falcon tubes or 500-mL centrifuge buckets and pelleted via centrifugation at 4000 × g for 10 minutes at 4 °C.

### **Lineage validation reactions**

Small-scale reactions were set up in 2-mL GC crimp vials. *E. coli* expressing the appropriate heme protein catalyst (380 µL, adjusted to the appropriate optical density with 20 mM MOPS buffer, pH=7) was added to the vials; the vials were then brought into a Coy anaerobic chamber (0 – 20 ppm O<sub>2</sub>). To each vial was added **1** (10 µL in ethanol, final concentration 20 mM) followed by **2** (10 µL in ethanol final concentration 40 mM). Directly following addition of **2**, the reaction vial was crimped and shaken at 500 rpm at room temperature. Reactions were worked up for GC analysis. For GC analysis, 20 µL 100% ethanol and an internal standard (20 µL 4-ethyl anisole, 20 mM in 1:1 hexane:ethyl acetate) were added to the vial. A 1:1 mixture of hexane:ethyl acetate (480 µL) was added, and the reaction was transferred into 1.7-mL Eppendorf tubes for extraction. The liquid-liquid extraction was carried out with a Retsch MM 301 mixing mill (1 minute, 30 Hz / 1800 rpm). Samples were centrifuged at 20817 × g for 5 minutes at RT, and the organic layer was used for chromatographic analysis.

### **Sonicated lysate preparation and hemochrome assay**

An aliquot of cells adjusted to a density of OD<sub>600</sub>= 30 for protein concentration determination was sonicated (QSonica Q500 Sonicator, 1/8 in tip) for 4 minutes, 1 second on, 1 second off at 30% amplitude. Following sonication, a 0.1 volume equivalent Bugbuster 10X protein

extraction reagent (EMD Millipore) was added; the lysate-Bugbuster mixture was incubated for 15 minutes at room temperature to improve cell lysis. The sonicated lysate was clarified via centrifugation at  $20817 \times g$  and  $4\text{ }^{\circ}\text{C}$  for 10 minutes. The concentration of heme-loaded protein was determined with the pyridine hemochromagen (hemochrome) assay.<sup>37</sup> Briefly, sonicated and clarified lysate ( $500\text{ }\mu\text{L}$ ) was added to a cuvette. Solution I ( $500\text{ }\mu\text{L}$  of  $0.2\text{ M}$  NaOH,  $40\%$  (v/v) pyridine,  $500\text{ }\mu\text{M}$  potassium ferricyanide) were added, and the spectrum of this oxidized sample was taken from  $350 - 600\text{ nm}$ . Sodium dithionite ( $10\text{ }\mu\text{L}$  of  $0.5\text{ M}$  solution in  $0.5\text{ M}$  NaOH) was added, and the reduced spectrum was taken from  $350 - 600\text{ nm}$ . The pyridine hemochromagen concentration was determined using its Q bands, with the literature extinction coefficient of  $23.98\text{ mM}^{-1}\text{ cm}^{-1}$  for ( $557\text{ nm}_{\text{reduced}} - 540\text{ nm}_{\text{oxidized}}$ ).

### **Preparation of lyophilized whole-cell biocatalysts**

Cultures of  $1\text{ L}$  HB-amp in unbaffled  $2.8\text{-L}$  Erlenmeyer flasks were inoculated  $1\%$  (v/v) with saturated overnight cultures and shaken in an Infors shaker at  $160\text{ rpm}$ ,  $37\text{ }^{\circ}\text{C}$ . At an approximate  $\text{OD}_{600}$  of  $1.0$ , cultures were chilled on ice for 30 minutes. Protein expression was induced with  $1\text{ M}$  IPTG to a final concentration of  $0.5\text{ mM}$ , and heme production was enhanced with supplementation of  $1\text{ M}$  ALA to a final concentration of  $1\text{ mM}$ . The cultures were shaken in an Infors shaker at  $230\text{ rpm}$  and  $22\text{ }^{\circ}\text{C}$  overnight ( $18\text{--}24\text{ hours}$ ). Cells were transferred to  $500\text{-mL}$  centrifuge buckets and pelleted via centrifugation at  $4000 \times g$  for 10 minutes at  $4\text{ }^{\circ}\text{C}$ . The pellet was transferred into a  $50\text{-mL}$  Falcon tube and flash-frozen on powdered dry ice. The lid of the  $50\text{-mL}$  Falcon tube was replaced by a Kimwipe held on with a rubber band and put in a lyophilization bottle. Pellets were lyophilized for at least 36 hours on a Labconco 4.5 plus freezezone benchtop lyophilizer (Labconco, Kansas City, MO) to ensure complete lyophilization. Following lyophilization, the cells were transferred into a sealable plastic bag and ground into a fine powder using a mortar and pestle. The lyophilized cell powder was stored at  $-20^{\circ}\text{C}$  until use.

### **Preparative-scale biocatalytic reactions with lyophilized whole cells**

To a  $250\text{-mL}$  round-bottom flask, lyophilized whole cells harboring *RmaNOD* THRAW ( $500\text{ mg}$ ,  $10\text{ mg mL}^{-1}$  final loading) were added, and the catalyst was reconstituted in aqueous

buffer (KPi 50 mM pH 7.5, 37.6 mL) by stirring for 10 minutes. The round-bottom flask was sealed with a septum and cycled with Argon and vacuum three times. To the reconstituted catalysts **2** (174.8  $\mu$ L neat substrate, 1 mmol, final concentration 20 mM) was added. Using a syringe pump, EDA solution (5 mL EtOH, 7 mL dH<sub>2</sub>O, 241.8  $\mu$ L EDA, solution concentration 163 mM, final concentration 40 mM) was added over a period of 2 hours, using a long Luer lock needle such that the needle sat below the solution surface during addition. The reaction was stirred at room temperature overnight (16-hour total reaction time). Upon reaction completion, the reaction mixture was stored at -20 °C until workup.

#### **Procedure for column-free purification of trifluoroborated product (4)**

The trifluoroboration procedure was adapted from Bagutski et al.<sup>30</sup> Following completion of the reaction, the product was extracted three times with 1:1 hexane:ethyl acetate (1 volume equivalent per extraction). Deprotected boronic acid was reprotected by stirring the organic layer with pinacol (236 mg, 2 mmol, 2.0 equivalents) in the presence of magnesium sulfate (1 g). The solution was filtered to remove magnesium sulfate and concentrated *in vacuo* to remove solvent and remaining unreacted starting material (**1**). The crude reaction mixture was then resuspended in 1:1 methanol:water (8 mL total). Potassium bifluoride solution (1.67 mL of 3 M solution in water, 5.0 equivalents) was added dropwise, and the reaction was left to proceed overnight. The solution was concentrated *in vacuo* to remove methanol, flash-frozen, and lyophilized to remove water. Acetone (20 mL) was added to the 40-mL scintillation vial containing the lyophilized crude product. The product was resuspended in the sealed scintillation vial via sonication bath. The solution was vacuum filtered, and the vial was washed three times with acetone (10 mL). The acetone was removed *in vacuo*, and the product was dried on high vacuum.

#### **Determination of initial cyclopropanation activity**

A panel of heme proteins previously engineered for carbene or nitrene transfer reactions<sup>27</sup> was tested for their activity in formation of **3**; proteins included in these tests were expressed in 96-well plates (*96-Well plate library expression*). The cell pellets were resuspended with 400  $\mu$ L 1x-M9-N, and the 96-well deep-well plates were brought into a Coy anaerobic

chamber. A substrate solution (50  $\mu$ L of 180 mM **1** and **2**, 20 mM final concentration of each substrate) was added to each well. The plates were sealed with an adhesive foil cover and shaken at 500 rpm and room temperature for 6 hours. The foil covers were removed, 500  $\mu$ L cyclohexane were added to each well, and the plates were sealed with silicone sealing mats and vortexed for 1 minute. The organic layer was transferred to GC vial inserts in GC vials, and formation of **3** was analyzed via GC-MS.

The highest initial activity for *cis*-selective cyclopropanation was exhibited by *ApePgb* W59A Y60G F145W and its intermediate variants. The highest initial activity for *trans*-selective cyclopropanation was produced by *RmaNOD* Q52A, *RmaNOD* Q52V, and Mb H64V V68A; *ApePgb* W59A Y60G F145W and *RmaNOD* Q52V were variants engineered for unactivated alkene cyclopropanation.<sup>28</sup> The myoglobin was engineered by the Fasan lab for styrenyl alkene cyclopropanation reactions.<sup>38</sup>

Table 3-1. Protein scaffolds tested for formation of **3** in the initial-activity screening.

UniProt ID of wild type	Organism	Annotation	Number of variants tested	Formation of <b>3</b>
B3FQS5	<i>Rhodothermus marinus</i>	Cytochrome <i>c</i>	9	Trace detected
P15452	<i>Hydrogenobacter thermophilus</i>	Cytochrome <i>c</i> -552	2	n.d.
P00080	<i>Rhodopila globiformis</i>	Cytochrome <i>c</i> 2	2	n.d.
D0MGT2	<i>Rhodothermus marinus</i>	Nitric oxide dioxygenase	10	Detected
P02185	<i>Physeter catodon</i>	Myoglobin	1	Detected
B3DUZ7	<i>Methylophilum infernorum</i>	Hell's gate globin	2	Trace detected
O66586	<i>Aquifex aeolicus</i>	Thermoglobin	1	Trace detected
G7VHJ7	<i>Pyrobaculum ferrireducens</i>	Protoglobin	1	Trace detected
Q0PB48	<i>Campylobacter jejuni</i>	Truncated hemoglobin	1	n.d.
Q9YFF4	<i>Aeropyrum pernix</i>	Protoglobin	10	Detected
O31607	<i>Bacillus subtilis</i>	Truncated hemoglobin	1	n.d.
P14779	<i>Bacillus megaterium</i>	Cytochrome P450 <sub>BM3</sub>	50	n.d.

Table 3-2. Activity of top variants from initial screening. Activity is given as integrated area for the target ions 182 m/z and 195 m/z. \*Variant was found during protein engineering for the referenced project. n.d.: not detected in given ion channel.

Variant	Reference	<i>cis</i> (182 m/z)	<i>trans</i> (182 m/z)	<i>cis</i> (195 m/z)	<i>trans</i> (195 m/z)
<i>ApePgb</i> WT	28	122	n.d.	n.d.	n.d.
<i>ApePgb</i> Y60G	28	1865	n.d.	n.d.	n.d.
<i>ApePgb</i> Y60G F145N	28*	32478	896	2620	1314
<i>ApePgb</i> W59A Y60G	28	5375	n.d.	351	n.d.
<i>ApePgb</i> W59A Y60G F145W	28	30043	n.d.	2513	204
Mb H64V V68A	38	423	1095	n.d.	1437
<i>RmaNOD</i> WT	28	n.d.	576	n.d.	251
<i>RmaNOD</i> Q52V	28	193	1039	n.d.	1074
<i>RmaNOD</i> Q52A	28*	n.d.	1417	n.d.	1767

### Calibration curves for analytical-scale concentration determinations

To generate a calibration curve for analytical-scale reactions, solutions at varying concentrations of the cyclopropane authentic standard were prepared in *E. coli* cell resuspensions (20 mM MOPS, pH = 7.01, 4.8% ethanol, OD<sub>600</sub> = 28 for *cis*-**3**, OD<sub>600</sub> = 30 for *trans*-**3**). Two 420- $\mu$ L aliquots of each solution at each concentration were worked up following the protocols in *small-scale, whole-cell biocatalytic reaction preparation and work-up*, skipping the addition of 20  $\mu$ L 100% EtOH, and the worked-up sample was analyzed by GC-MS. The y-axis of each calibration curve gives the product:internal standard ratio, calculated as total ion count (TIC) of the GC-MS peak corresponding to the target product divided by the TIC of the GC-MS peak corresponding to the internal standard. The x-axis of each calibration curve is the concentration corresponding to the product:internal standard ratio. The resulting calibration curves reproducibly display non-linearity, in which low concentrations of cyclopropane show a diminished response. While this indicates a product loss event, the standard curve reproducibility shows that it can be utilized to reliably calculate analytical-scale yields after fitting with an exponential function.



**Ethyl (*trans*)-2-((*S*)-phenylsulfinyl)cyclopropane-1-carboxylate**

The cyclopropane *trans*-S1 (1.97 g) was added to a round-bottom flask, dissolved in 90 mL of CH<sub>2</sub>Cl<sub>2</sub>, and cooled to 0 °C. The *m*-CPBA (99%, 1.63 g) was added in three portions over 30 minutes and then continued to stir for 2 hours at 0 °C. Aqueous KOH (3 M, 35 mL) was added, and the layers were separated. The aqueous layer was extracted with CH<sub>2</sub>Cl<sub>2</sub> (5 × 35 mL). The combined organic layers were dried with MgSO<sub>4</sub>, filtered, and concentrated. The crude material was purified by column chromatography (silica, 1:2 hexane:Et<sub>2</sub>O) to provide 1.39 g of *trans*-S2 (66% yield).

**Ethyl (*cis*)-2-((*S*)-phenylsulfinyl)cyclopropane-1-carboxylate**

The cyclopropane *cis*-S1 (1.78 g) was added to a round-bottom flask, dissolved in 80 mL of CH<sub>2</sub>Cl<sub>2</sub> and cooled to 0 °C. The *m*-CPBA (99%, 1.44 g) was added in three portions over 30 minutes and then continued to stir for 2 hours at 0 °C. Aqueous KOH (3 M, 30 mL) was added, and the layers were separated. The aqueous layer was extracted with CH<sub>2</sub>Cl<sub>2</sub> (5 × 30 mL). The combined organic layers were dried with MgSO<sub>4</sub>, filtered, and concentrated. The crude material was purified by column chromatography (silica, 1:2 hexane: diethyl ether) to provide 1.20 g of *cis*-S2 (63% yield).

**Ethyl (*cis*)-2-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)cyclopropane-1-carboxylate**

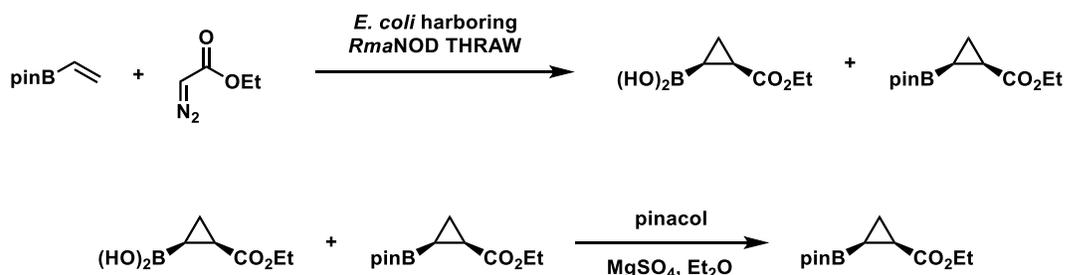
The cyclopropane *trans*-S2 (1.29 g) was added to a round-bottom flask and dissolved in 54 mL of THF. The reaction was cooled to -78 °C, and then the *i*PrMgCl (2 M, 4.4 mL) was added dropwise. The reaction continued to stir for 10 minutes before the addition of 2-isopropoxy-4,4,5,5-tetramethyl-1,3,2-dioxaborolane (2.2 mL). The reaction was warmed to 0 °C and continued to stir for 6 hours before being warmed to room temperature. Methanol (15 mL) was added to quench the reaction and stirred for 5 minutes. Then, 150 mL of diethyl ether were added, the mixture was filtered and concentrated. The crude mixture was purified by column chromatography (florisil, 10% to 50% diethyl ether : hexane) to provide 480 mg of *trans*-3 (37% yield). <sup>1</sup>H NMR (400 MHz, chloroform-*d*) δ 4.19 – 3.98 (m, 2H), 1.83 – 1.73 (m, 1H), 1.26 – 1.21 (m, 12H), 1.12 – 1.04 (m, 1H), 1.04 – 0.96 (m, 1H), 0.43 – 0.31

(m, 1H).  $^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  174.5, 83.6, 60.6, 25.0, 25.0, 17.8, 14.4, 11.3.  $^{11}\text{B}$  NMR (128 MHz,  $\text{CDCl}_3$ )  $\delta$  32.1.

**Ethyl (*cis*)-2-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)cyclopropane-1-carboxylate**

The cyclopropane *cis*-**S2** (1.10 g) was added to a round-bottom flask and dissolved in 46 mL of THF. The reaction was cooled to  $-78\text{ }^\circ\text{C}$ , and then the *i*PrMgCl (2 M, 3.5 mL) was added dropwise. The reaction continued to stir for 10 minutes before the addition of 2-isopropoxy-4,4,5,5-tetramethyl-1,3,2-dioxaborolane (1.9 mL). The reaction was warmed to  $0\text{ }^\circ\text{C}$  and continued to stir for 6 hours before being warmed to room temperature. Methanol (15 mL) was added to quench the reaction and stirred for 5 minutes. Then, 150 mL of diethyl ether were added, the mixture was filtered and concentrated. The crude mixture was purified by column chromatography (florisil, 10% to 50% diethyl ether: hexane) to provide 440 mg of *cis*-**3** (39% yield).  $^1\text{H}$  NMR (400 MHz, chloroform-*d*)  $\delta$  4.18 – 4.04 (m, 2H), 1.79 – 1.69 (m, 1H), 1.21 (d,  $J = 1.7\text{ Hz}$ , 12H), 1.02 – 0.92 (m, 1H), 0.62 – 0.51 (m, 1H).  $^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  174.3, 83.5, 60.5, 24.8, 24.7, 18.6, 14.3, 13.0.  $^{11}\text{B}$  NMR (128 MHz,  $\text{CDCl}_3$ )  $\delta$  32.58.

**Ethyl (1*R*,2*S*)-2-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)cyclopropane-1-carboxylate (enzymatic)**

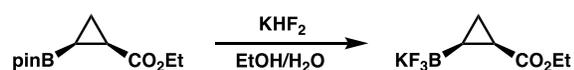


Whole *E. coli* cells were grown fresh as described above (*Medium- and large-scale protein expression*); two 1-L cultures in 2.8-L non-baffled flasks were grown for *RmaNOD THRAW*. The cells were resuspended in 1X M9-N buffer ( $\text{OD}_{600} = 30$ , 664 mL) and transferred into a 1-L round-bottom flask with a stir bar. The round-bottom flask was sealed

with a septum and attached to a Schlenk line; three cycles of Argon purging and vacuum evacuation were used to remove oxygen from the reaction flask. The mixture remained under argon flow throughout the entire reaction. Substrate (**1**) (neat stock, 7.5 mL, 60 mM final concentration) was added via syringe. EDA (**2**) (10.41 mL diluted in 35.9 mL ethanol cosolvent, 120 mM final concentration) was added over 2 hours via syringe pump slow addition. The reaction continued to stir overnight (16 hours) at room temperature. The reactions were then extracted three times with 1:1 hexane:ethyl acetate. The organic fractions were concentrated *in vacuo* and used in the subsequent protection step without purification. To the crude mixture of **3** and the deprotected cyclopropylboronic acid (**6**) were added pinacol (1.18 g), anhydrous magnesium sulfate (15 g), and anhydrous diethyl ether (100 mL). The reaction flask was sealed with a rubber septum and stirred for 2 hours at room temperature. After 2 hours, an aliquot was analyzed by  $^1\text{H}$  NMR spectroscopy to ensure complete consumption of **6**. The reaction was filtered over a plug of anhydrous  $\text{MgSO}_4$ , concentrated, and purified by distillation.

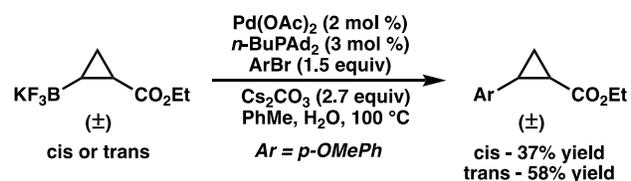
The crude mixture was purified via fractional distillation using a short path distillation setup attached to a small Vigreux column under reduced pressure. The residual alkene starting material was isolated at 10 torr and 40 °C to obtain 820 mg of the reagent as a colorless oil. The Vigreux column was then removed and a mixture of the desired product **3**, diethyl fumarate, and diethyl maleate were co-distilled at 5 torr and 70–100 °C. The obtained mixture was transferred to a round-bottom flask and vigorously stirred while heated under vacuum (40 °C at 1 torr for 24 hours, then 45 °C at 1 torr for 24 hours, then 50 °C at 1 torr for 3.5 hours) to remove the undesired diethyl fumarate and diethyl maleate side products. Product **3** was isolated as a colorless oil (3.7 g, 95% purity, 36% yield, 41% yield based on recovered starting material).

**Ethyl (1R,2S)-2-(trifluoro-1-*i*-boranyl)cyclopropane-1-carboxylate, potassium salt**  
 ((1R, 2S)-4))



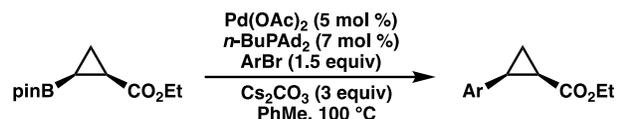
To a dram vial equipped with a magnetic stir bar were added **3** (50 mg, 0.208 mmol, 1 equiv) and 1 mL of ethanol. A solution of  $\text{KHF}_2$  (78 mg, 1 mmol, 5 equiv) in 0.22 mL of water was added to the vial, then sealed with a screw cap, and stirred overnight at room temperature. Upon completion, the solvent was removed *in vacuo* and the  $\text{BF}_3\text{K}$  salt was extracted with acetone ( $3 \times 4$  mL). The combined organic layers were filtered through a plug of cotton and concentrated to provide 35 mg of the desired cyclopropyl  $\text{BF}_3\text{K}$  salt (76% yield). A small sample (~1 mg) was recrystallized from slow evaporation of acetone to provide crystals suitable for X-ray diffraction.  $^1\text{H}$  NMR (400 MHz,  $\text{DMSO-}d_6$ )  $\delta$  3.92 (qd,  $J = 7.1, 1.2$  Hz, 2H), 1.32 – 1.22 (m, 1H), 1.12 (t,  $J = 7.1$  Hz, 3H), 0.75 (ddd,  $J = 8.8, 4.9, 2.5$  Hz, 1H), 0.56 (dddt,  $J = 10.8, 8.2, 2.4, 1.1$  Hz, 1H), -0.13 (dddt,  $J = 14.7, 10.3, 8.7, 4.3$  Hz, 1H).  $^{13}\text{C}$  NMR (101 MHz,  $\text{DMSO}$ )  $\delta$  174.9, 59.1, 40.6, 40.4, 40.2, 40.0, 39.8, 39.6, 39.3, 17.3, 17.3, 17.3, 17.3, 14.7, 10.3, 10.3, 10.2, 10.2.  $^{11}\text{B}$  NMR (128 MHz,  $\text{DMSO-}d_6$ )  $\delta$  3.43, 3.00.  $^{19}\text{F}$  NMR (282 MHz,  $\text{DMSO-}d_6$ )  $\delta$  -135.1, -135.2.

#### Suzuki-Miyaura cross-coupling with **4**



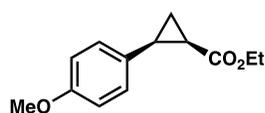
In a glovebox under an inert atmosphere, **4** (4.4 mg, 0.02 mmol, 1 equiv) was dissolved in 110  $\mu\text{L}$  of a mixture of 10:1 toluene:degassed water in a half dram vial equipped with a magnetic stir bar. The 4-bromoanisole (5.6 mg, 1.5 equiv),  $\text{Pd(OAc)}_2$  (0.09 mg, 0.02 equiv),  $n\text{-BuPAD}_2$  (0.22 mg, 0.03 equiv), and  $\text{Cs}_2\text{CO}_3$  (17.6 mg, 2.7 equivalents) were added sequentially. The vial was sealed with a Teflon cap, wrapped with electrical tape, and stirred at 100  $^\circ\text{C}$  for 20 hours. The reaction was cooled to room temperature, loaded onto a packed silica column, and analyzed by  $^1\text{H}$  NMR against an internal standard to determine the product yield.

### General procedure: Suzuki-Miyaura cross-coupling with *cis*-**3**



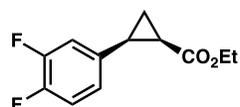
In a glovebox under an inert atmosphere, **3** (48 mg, 0.2 mmol, 1 equiv), aryl bromide (1.5 equiv), Pd(OAc)<sub>2</sub> (2.2 mg, 0.05 equiv), *n*-BuPAd<sub>2</sub> (5 mg, 0.07 equiv), Cs<sub>2</sub>CO<sub>3</sub> (3.0 equiv), and toluene (2 mL) were sequentially added to a 2-dram vial equipped with a magnetic stir bar. The vial was sealed with a Teflon cap, wrapped with electrical tape, and stirred at 100 °C for 36 hours. The reaction was cooled to room temperature, loaded onto a packed silica column, and purified by column chromatography to afford the desired product.

### Ethyl (1*R*,2*S*)-2-(4-methoxyphenyl)cyclopropane-1-carboxylate (**5a**)



Prepared from (1*R*,2*S*)-**3** (48 mg, 1.0 equiv, 0.2 mmol) and 1-bromo-4-methoxybenzene (56 mg, 1.5 equiv, 0.3 mmol). The crude residue was purified by column chromatography (silica, 10% Et<sub>2</sub>O/hexanes) to yield 40.4 mg of **5a** (92% yield). The product was analyzed by chiral SFC and no minor enantiomer peak was detected, confirming the product is formed in >99% ee. This is a known compound, and the spectroscopic data are in agreement with the literature.<sup>38</sup> Chiral SFC: (OJ-H, 2.5 mL/min, 5% IPA in CO<sub>2</sub>, λ = 254 nm): *t*<sub>R</sub> (major) = 3.9 min, *t*<sub>R</sub> (minor) = 5.5 min. <sup>1</sup>H NMR (400 MHz, chloroform-*d*) δ 7.21 – 7.15 (m, 2H), 6.80 (dt, *J* = 8.9, 2.5 Hz, 2H), 3.89 (q, *J* = 7.1 Hz, 2H), 3.77 (s, 3H), 2.52 (q, *J* = 8.7, 8.0 Hz, 1H), 2.03 (ddd, *J* = 9.2, 7.8, 5.6 Hz, 1H), 1.70 – 1.61 (m, 1H), 1.30 (ddd, *J* = 8.7, 7.8, 5.0 Hz, 1H), 1.02 (t, *J* = 7.1 Hz, 3H). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 171.2, 158.4, 130.4, 128.7, 113.4, 60.3, 55.3, 25.0, 21.8, 14.2, 11.4. HRMS (EI<sup>+</sup>, *m/z*): calculated for C<sub>13</sub>H<sub>16</sub>O<sub>3</sub> [M]<sup>+</sup>: 220.1100, found 220.1095.

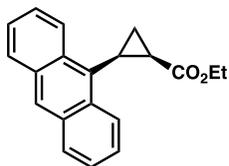
### Ethyl (1*R*,2*S*)-2-(3,4-difluorophenyl)cyclopropane-1-carboxylate (**5b**)



Prepared from (1*R*,2*S*)-**3** (48 mg, 1.0 equiv, 0.2 mmol) and 4-bromo-1,2-difluorobenzene (58 mg, 1.5 equiv, 0.3 mmol). The crude residue was purified by column chromatography (silica, 10% Et<sub>2</sub>O/hexanes) to yield 29.7 mg of **5b** (66% yield). The product was analyzed by chiral SFC and no minor enantiomer peak was

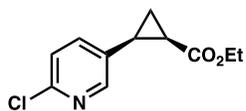
detected, confirming the product is formed in >99% ee. Chiral SFC: (OJ-H, 2.5 mL/min, 0% IPA in CO<sub>2</sub>,  $\lambda$  = 210 nm):  $t_R$  (major) = 2.9 min,  $t_R$  (minor) = 3.4 min. <sup>1</sup>H NMR (400 MHz, chloroform-*d*)  $\delta$  7.13 – 6.93 (m, 3H), 3.92 (q,  $J$  = 7.1 Hz, 2H), 2.56 – 2.45 (m, 1H), 2.07 (ddd,  $J$  = 9.3, 7.9, 5.7 Hz, 1H), 1.63 (ddd,  $J$  = 7.4, 5.6, 5.2 Hz, 1H), 1.34 (ddd,  $J$  = 8.7, 7.9, 5.3 Hz, 1H), 1.05 (t,  $J$  = 7.1 Hz, 3H). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>)  $\delta$  170.8, 151.2, 151.1, 150.7, 150.5, 148.7, 148.6, 148.2, 148.1, 133.9, 133.8, 133.8, 133.8, 133.8, 125.5, 125.4, 125.4, 125.4, 118.4, 118.3, 116.8, 116.6, 60.6, 24.7, 24.7, 21.9, 14.2, 11.7. <sup>19</sup>F NMR (282 MHz, CDCl<sub>3</sub>)  $\delta$  -138.7 (m), -140.7 (m). HRMS (EI+,  $m/z$ ): calculated for C<sub>12</sub>H<sub>12</sub>F<sub>2</sub>O<sub>2</sub> [M]<sup>+</sup>: 226.0805, measured 226.0805.

### Ethyl (1*R*,2*S*)-2-(anthracen-9-yl)cyclopropane-1-carboxylate (**5c**)



Prepared from (1*R*,2*S*)-**3** (48 mg, 1.0 equiv, 0.2 mmol) and 9-bromoanthracene (77 mg, 1.5 equiv, 0.3 mmol). The crude residue was purified by column chromatography (silica, 5-15% Et<sub>2</sub>O/hexanes) to yield 25.5 mg of **5c** (44% yield). The product was analyzed by chiral SFC and no minor enantiomer peak was detected, confirming the product is formed in >99% ee. Chiral SFC: (OJ-H, 2.5 mL/min, 15% IPA in CO<sub>2</sub>,  $\lambda$  = 254 nm):  $t_R$  (major) = 7.3 min,  $t_R$  (minor) = 7.9 min. <sup>1</sup>H NMR (400 MHz, Chloroform-*d*)  $\delta$  8.63 (d,  $J$  = 8.9 Hz, 1H), 8.50 (d,  $J$  = 8.7 Hz, 1H), 8.36 (s, 1H), 7.98 (t,  $J$  = 8.6 Hz, 2H), 7.51 – 7.42 (m, 4H), 3.55 – 3.34 (m, 2H), 3.09 (q,  $J$  = 8.3, 7.9 Hz, 1H), 2.57 (td,  $J$  = 8.3, 5.6 Hz, 1H), 2.01 (ddd,  $J$  = 8.8, 8.1, 4.6 Hz, 1H), 1.82 (ddd,  $J$  = 7.8, 5.6, 4.7 Hz, 1H), 0.54 (t,  $J$  = 7.1 Hz, 3H). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>): 171.8, 131.9, 131.8, 131.2, 129.7, 129.4, 129.0, 127.0, 125.6, 125.5, 125.3, 125.1, 125.0, 124.8, 60.1, 22.5, 20.9, 16.7, 13.6.

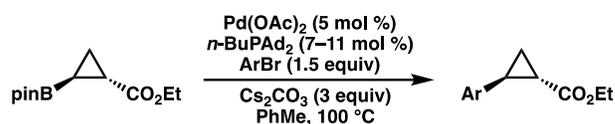
### Ethyl (1*R*,2*S*)-2-(6-chloropyridin-3-yl)cyclopropane-1-carboxylate (**5d**)



Prepared from (1*R*,2*S*)-**3** (48 mg, 1.0 equiv, 0.2 mmol) and 5-bromo-2-chloropyridine (58 mg, 1.5 equiv, 0.3 mmol). The crude residue was purified by column chromatography (silica, 5-40% Et<sub>2</sub>O/hexanes) to yield 15 mg of **5d** (33% yield) as a colorless oil. The product was analyzed by chiral SFC, confirming the product is formed in >99% ee. Chiral SFC: (OJ-H, 2.5 mL/min, 10% IPA in

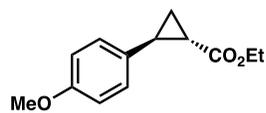
CO<sub>2</sub>,  $\lambda = 254$  nm):  $t_R$  (major) = 2.3 min,  $t_R$  (minor) = 2.6 min. <sup>1</sup>H NMR (400 MHz, chloroform-*d*)  $\delta$  8.30 (dt,  $J = 2.5, 0.7$  Hz, 1H), 7.53 (ddd,  $J = 8.2, 2.5, 0.7$  Hz, 1H), 7.25 – 7.19 (m, 1H), 3.94 (qt,  $J = 7.1, 3.6$  Hz, 2H), 2.48 (q, 1H), 2.14 (ddd,  $J = 9.0, 7.9, 5.7$  Hz, 1H), 1.67 (dt,  $J = 7.5, 5.5$  Hz, 1H), 1.42 (ddd,  $J = 8.7, 8.0, 5.2$  Hz, 1H), 1.06 (t,  $J = 7.1$  Hz, 3H). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>)  $\delta$  170.6, 150.9, 149.8, 139.5, 131.5, 123.5, 60.8, 22.2, 21.8, 14.3, 11.4. HRMS (FAB,  $m/z$ ): calculated for C<sub>11</sub>H<sub>12</sub>ClNO<sub>2</sub> [M]<sup>+</sup>: 225.0557, measured 225.0706.

### General procedure: Suzuki-Miyaura cross-coupling of *trans*-4



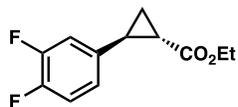
In a glovebox under an inert atmosphere, *trans*-3 (4.8 mg, 0.02 mmol, 1 equiv), aryl bromide (1.5 equiv), Pd(OAc)<sub>2</sub> (0.22 mg, 0.005 equiv), *n*-BuPAD<sub>2</sub> (0.5–0.8 mg, 0.07–0.11 equiv), Cs<sub>2</sub>CO<sub>3</sub> (3.0 equiv), and toluene (0.2 mL) were sequentially added to a 1/2-dram vial equipped with a magnetic stir bar. The vial was sealed with a Teflon cap, wrapped with electrical tape, and stirred at 100 °C for 24 hours. The reaction was cooled to room temperature, flushed through a pipet over a short pad of silica gel, and concentrated *in vacuo*. The crude reaction was dissolved in deuterated chloroform and analyzed by NMR versus an added standard (2,3,5,6-tetrachloronitrobenzene) to obtain an NMR yield of the product.

### Ethyl *trans*-2-(4-methoxyphenyl)cyclopropane-1-carboxylate



Prepared from *trans*-3 (4.8 mg, 1.0 equiv, 0.02 mmol) and 1-bromo-4-methoxybenzene (5.6 mg, 1.5 equiv, 0.03 mmol). The crude residue was analyzed by NMR spectroscopy against an internal standard (2.0 mg of 2,3,5,6-tetrachloronitrobenzene, singlet at 7.73 ppm) to afford the reaction yield (56%). <sup>1</sup>H NMR (500 MHz, Chloroform-*d*)  $\delta$  7.06 – 7.01 (m, 2H), 6.84 – 6.80 (m, 2H), 4.17 (q,  $J = 7.1$  Hz, 2H), 3.79 (s, 5H), 2.49 (ddd,  $J = 9.2, 6.5, 4.2$  Hz, 1H), 1.83 (ddd,  $J = 8.4, 5.2, 4.1$  Hz, 1H), 1.56 (ddd,  $J = 9.4, 5.2, 4.4$  Hz, 1H), 1.29 (t,  $J = 7.1$  Hz, 3H), 1.28 – 1.24 (m, 1H).

### Ethyl *trans*-2-(3,4-difluorophenyl)cyclopropane-1-carboxylate



Prepared from *trans*-**3** (4.8 mg, 1.0 equiv, 0.02 mmol) and 4-bromo-1,2-difluorobenzene (5.8 mg, 1.5 equiv, 0.03 mmol). The crude residue was analyzed by NMR spectroscopy against an internal standard (3.5 mg of 2,3,5,6-tetrachloronitrobenzene, singlet at 7.73 ppm) to afford the reaction yield (61%). <sup>1</sup>H NMR (500 MHz, Chloroform-*d*)  $\delta$  7.11 (dt,  $J = 10.2, 8.3$  Hz, 1H), 6.94 (ddd,  $J = 11.4, 7.5, 2.3$  Hz, 1H), 6.89 (ddt,  $J = 8.2, 3.9, 1.8$  Hz, 1H), 4.23 (qd,  $J = 7.2, 0.8$  Hz, 2H), 2.53 (ddd,  $J = 9.3, 6.4, 4.1$  Hz, 1H), 1.91 (ddd,  $J = 8.5, 5.3, 4.2$  Hz, 1H), 1.70 – 1.62 (m, 1H), 1.37 – 1.33 (m, 3H), 1.32 – 1.29 (m, 1H).

### Compound separation conditions

Table 3-3. Achiral separation conditions for compound **3**.

Compound	Instrument; column	Method program	Retention times
<b>3</b>	GC-FID; HP-5	90 °C isothermal 1 minute 15 °C/min to 110 °C 60 °C/min to 280 °C isothermal 1 minute	Internal standard: 2.9 minutes <i>Cis</i> - <b>3</b> : 4.6 min <i>Trans</i> - <b>3</b> : 4.7 min
<b>3</b>	GC-MS; HP-5ms UI	90 °C isothermal 1 minute 75 °C/min to 325 °C isothermal 1.5 minutes	Internal standard: 2.3 min <i>Cis</i> - <b>3</b> : 3.55 min <i>Trans</i> - <b>3</b> : 3.76 min

Table 3-4. Chiral separation conditions. Due to the worse separation conditions for *trans*-**3** enantiomers, a separate method was prepared when only *cis*-**3** enantioselectivity was to be measured.

Compound	Instrument; column	Method program	Retention times
<i>Cis</i> - <b>3</b>	GC-MS; CP-Chirasil Dex CB (25 m, 0.25 mm ID, 0.25 $\mu$ m film thickness)	80 °C isothermal 2 minutes 10 °C/min to 130 °C, isothermal 15 minutes	<i>Cis</i> minor: 14.3 min, major: 14.8 min ( <i>trans</i> unresolved): 19.0 min
<i>Trans</i> - <b>3</b>	GC-MS; CP-Chirasil Dex CB (25 m, 0.25 mm ID, 0.25 $\mu$ m film thickness)	90 °C isothermal 180 minutes	<i>Cis</i> major (1R, 2S): 87.4 min, minor: 98.9 min <i>Trans</i> major: 172.8 min, minor: 177.8 min

### 3.7 References for Chapter 3

- (1) Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. Engineering the Third Wave of Biocatalysis. *Nature*. 2012, pp 185–194. <https://doi.org/10.1038/nature11117>.
- (2) Clomburg, J. M.; Gonzalez, R. Biofuel Production in *Escherichia Coli*: The Role of Metabolic Engineering and Synthetic Biology. *Appl. Microbiol. Biotechnol.* **2010**, *86*, 419–434. <https://doi.org/10.1007/s00253-010-2446-1>.
- (3) Yeoman, C. J.; Han, Y.; Dodd, D.; Schroeder, C. M.; Mackie, R. I.; Cann, I. K. O. Thermostable Enzymes as Biocatalysts in the Biofuel Industry. *Adv. Appl. Microbiol.* **2010**, *70*, 1–55. [https://doi.org/10.1016/S0065-2164\(10\)70001-0](https://doi.org/10.1016/S0065-2164(10)70001-0).
- (4) Prasad, S.; Chand, T. Nitrile Hydratases ( NHases ): At the Interface of Academia and Industry. *Biotechnol. Adv.* **2010**, *28*, 725–741. <https://doi.org/10.1016/j.biotechadv.2010.05.020>.
- (5) Forti, L.; Mauro, S. Di; Cramarossa, M. R.; Filippucci, S.; Turchetti, B.; Buzzini, P. Non-Conventional Yeasts Whole Cells as Efficient Biocatalysts for the Production of Flavors and Fragrances. *Molecules* **2015**, 10377–10398. <https://doi.org/10.3390/molecules200610377>.
- (6) Sheldon, R. A.; Brady, D.; Bode, M. L. The Hitchhiker’s Guide to Biocatalysis: Recent Advances in the Use of Enzymes in Organic Synthesis. *Chem. Sci.* **2020**, *11*, 2587–2605. <https://doi.org/10.1039/C9SC05746C>.
- (7) Hughes, D. L. Biocatalysis in Drug Development - Highlights of the Recent Patent Literature. *Org. Process Res. Dev.* **2018**, *22*, 1063–1080. <https://doi.org/10.1021/acs.oprd.8b00232>.
- (8) Fryszkowska, A.; Devine, P. N. Biocatalysis in Drug Discovery and Development. *Curr. Opin. Chem. Biol.* **2020**, *55*, 151–160. <https://doi.org/10.1016/j.cbpa.2020.01.012>.
- (9) Devine, P. N.; Howard, R. M.; Kumar, R.; Thompson, M. P.; Truppo, M. D.; Turner, N. J. Extending the Application of Biocatalysis to Meet the Challenges of Drug Development. *Nat. Rev. Chem.* **2018**, *2*, 409–421. <https://doi.org/10.1038/s41570-018-0055-1>.
- (10) Truppo, M. D. Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Med. Chem. Lett.* **2017**, *8*, 476–480. <https://doi.org/10.1021/acsmedchemlett.7b00114>.
- (11) Li, F.; Zhang, X.; Renata, H. Enzymatic C–H Functionalizations for Natural Product Synthesis. *Curr. Opin. Chem. Biol.* **2019**, *49*, 25–32. <https://doi.org/10.1016/j.cbpa.2018.09.004>.
- (12) Brown, D. G.; Boström, J. Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? *J. Med. Chem.* **2016**, *59*, 4443–4458. <https://doi.org/10.1021/acs.jmedchem.5b01409>.
- (13) Ma, S. K.; Gruber, J.; Davis, C.; Newman, L.; Gray, D.; Wang, A.; Grate, J.; Huisman, G. W.; Sheldon, R. A. A Green-by-Design Biocatalytic Process for Atorvastatin Intermediate. *Green Chem.* **2010**, *12*, 81–86. <https://doi.org/10.1039/B919115C>.
- (14) Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.;

- Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J. Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science* **2010**, *329*, 305–309. <https://doi.org/10.1126/science.1188934>.
- (15) Nawrat, C. C.; Whittaker, A. M.; Huffman, M. A.; McLaughlin, M.; Cohen, R. D.; Andreani, T.; Ding, B.; Li, H.; Weisel, M.; Tschaen, D. M. Nine-Step Stereoselective Synthesis of Islatravir from Deoxyribose. *Org. Lett.* **2020**, *22*, 2167–2172. <https://doi.org/10.1021/acs.orglett.0c00239>.
- (16) Gómez Baraibar, Á.; Reichert, D.; Mügge, C.; Seger, S.; Gröger, H.; Kourist, R. A One-Pot Cascade Reaction Combining an Encapsulated Decarboxylase with a Metathesis Catalyst for the Synthesis of Bio-Based Antioxidants. *Angew. Chem., Int. Ed.* **2016**, *55*, 14823–14827. <https://doi.org/10.1002/anie.201607777>.
- (17) Latham, J.; Henry, J. M.; Sharif, H. H.; Menon, B. R. K.; Shepherd, S. A.; Greaney, M. F.; Micklefield, J. Integrated Catalysis Opens New Arylation Pathways via Regiodivergent Enzymatic C-H Activation. *Nat. Commun.* **2016**, *7*, 1–8. <https://doi.org/10.1038/ncomms11873>.
- (18) Frese, M.; Schnepel, C.; Minges, H.; Voß, H.; Feiner, R.; Sewald, N. Modular Combination of Enzymatic Halogenation of Tryptophan with Suzuki-Miyaura Cross-Coupling Reactions. *ChemCatChem* **2016**, *8*, 1799–1803. <https://doi.org/10.1002/cctc.201600317>.
- (19) Chandgude, A. L.; Fasan, R. Highly Diastereo- and Enantioselective Synthesis of Nitrile-Substituted Cyclopropanes by Myoglobin-Mediated Carbene Transfer Catalysis. *Angew. Chem., Int. Ed.* **2018**, *57*, 15852–15856. <https://doi.org/10.1002/anie.201810059>.
- (20) Talele, T. T. The “Cyclopropyl Fragment” Is a Versatile Player That Frequently Appears in Preclinical/Clinical Drug Molecules. *J. Med. Chem.* **2016**, *59*, 8712–8756. <https://doi.org/10.1021/acs.jmedchem.6b00472>.
- (21) Taylor, R. D.; Maccoss, M.; Lawson, A. D. G. Rings in Drugs. *J. Med. Chem.* **2014**, *57*, 5845–5859. <https://doi.org/10.1021/jm4017625>.
- (22) Katsuda, Y. Development of and Future Prospects for Pyrethroid Chemistry. *Pestic. Sci.* **1999**, *55*, 775–782. <https://doi.org/10.1002/ps.2780550803>.
- (23) Wang, Z. J.; Renata, H.; Peck, N. E.; Farwell, C. C.; Coelho, P. S.; Arnold, F. H. Improved Cyclopropanation Activity of Histidine-Ligated Cytochrome P450 Enables the Enantioselective Formal Synthesis of Levomilnacipran. *Angew. Chem., Int. Ed.* **2014**, *53*, 6810–6813. <https://doi.org/10.1002/anie.201402809>.
- (24) Bajaj, P.; Sreenilayam, G.; Tyagi, V.; Fasan, R. Gram-Scale Synthesis of Chiral Cyclopropane-Containing Drugs and Drug Precursors with Engineered Myoglobin Catalysts Featuring Complementary Stereoselectivity. *Angew. Chem., Int. Ed.* **2016**, *55*, 16110–16114. <https://doi.org/10.1002/anie.201608680>.
- (25) Hernandez, K. E.; Renata, H.; Lewis, R. D.; Kan, S. B. J. J.; Zhang, C.; Forte, J.; Rozzell, D.; McIntosh, J. A.; Arnold, F. H. Highly Stereoselective Biocatalytic Synthesis of Key Cyclopropane Intermediate to Ticagrelor. *ACS Catal.* **2016**, *6*, 7810–7813. <https://doi.org/10.1021/acscatal.6b02550>.
- (26) Kim, T.; Kassim, A. M.; Botejue, A.; Zhang, C.; Forte, J.; Rozzell, D.; Huffman, M. A.; Devine, P. N.; McIntosh, J. A. Hemoprotein-Catalyzed Cyclopropanation En

- Route to the Chiral Cyclopropanol Fragment of Grazoprevir. *ChemBioChem* **2019**, *20*, 1129–1132. <https://doi.org/10.1002/cbic.201800652>.
- (27) Brandenburg, O. F.; Fasan, R.; Arnold, F. H. Exploiting and Engineering Hemoproteins for Abiological Carbene and Nitrene Transfer Reactions. *Curr. Opin. Biotechnol.* **2017**, *47*, 102–111. <https://doi.org/10.1016/j.copbio.2017.06.005>.
- (28) Knight, A. M.; Kan, S. B. J.; Lewis, R. D.; Brandenburg, O. F.; Chen, K.; Arnold, F. H. Diverse Engineered Heme Proteins Enable Stereodivergent Cyclopropanation of Unactivated Alkenes. *ACS Cent. Sci.* **2018**, *4*, <https://doi.org/10.1021/acscentsci.7b00548>.
- (29) Zhao, S.; Gensch, T.; Murray, B.; Niemeyer, Z. L.; Sigman, M. S.; Biscoe, M. R. Enantiodivergent Pd-Catalyzed C–C Bond Formation Enabled through Ligand Parameterization. *Science* **2018**, *362*, 670–674. <https://doi.org/10.1126/science.aat2299>.
- (30) Bagutski, V.; Ros, A.; Aggarwal, V. K. Improved Method for the Conversion of Pinacolboronic Esters into Trifluoroborate Salts: Facile Synthesis of Chiral Secondary and Tertiary Trifluoroborates. *Tetrahedron* **2009**, *65*, 9956–9960. <https://doi.org/10.1016/j.tet.2009.10.002>.
- (31) Maa, Y.-F.; Prestrelski, S. Biopharmaceutical Powders Particle Formation and Formulation Considerations. *Curr. Pharm. Biotechnol.* **2000**, *1*, 283–302. <https://doi.org/10.2174/1389201003378898>.
- (32) Jakoblinert, A.; Rother, D. A Two-Step Biocatalytic Cascade in Micro-Aqueous Medium: Using Whole Cells to Obtain High Concentrations of a Vicinal Diol. *Green Chem.* **2014**, *16*, 3472–3482. <https://doi.org/10.1039/c4gc00010b>.
- (33) Wachtmeister, J.; Rother, D. Recent Advances in Whole Cell Biocatalysis Techniques Bridging from Investigative to Industrial Scale. *Curr. Opin. Biotechnol.* **2016**, *42*, 169–177. <https://doi.org/10.1016/j.copbio.2016.05.005>.
- (34) Wachtmeister, J.; Jakoblinert, A.; Kulig, J.; Offermann, H.; Rother, D. Whole-Cell Teabag Catalysis for the Modularisation of Synthetic Enzyme Cascades in Micro-Aqueous Systems. *ChemCatChem* **2014**, *6*, 1051–1058. <https://doi.org/10.1002/cctc.201300880>.
- (35) Kan, S. B. J.; Huang, X.; Gumulya, Y.; Chen, K.; Arnold, F. H. Genetically Programmed Chiral Organoborane Synthesis. *Nature* **2017**, *552*, 132. <https://doi.org/10.1038/nature24996>.
- (36) Kille, S.; Acevedo-Rocha, C. G.; Parra, L. P.; Zhang, Z. G.; Opperman, D. J.; Reetz, M. T.; Acevedo, J. P. Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synth. Biol.* **2013**, *2*, 83–92. <https://doi.org/10.1021/sb300037w>.
- (37) Barr, I.; Guo, F. Pyridine Hemochromagen Assay for Determining the Concentration of Heme in Purified Protein Solutions. *Bio. Protoc.* **2015**, *5*, <https://doi.org/10.21769/bioprotoc.1594>.
- (38) Bordeaux, M.; Tyagi, V.; Fasan, R. Highly Diastereoselective and Enantioselective Olefin Cyclopropanation Using Engineered Myoglobin-Based Catalysts. *Angew. Chem., Int. Ed.* **2015**, *54*, 1744–1748. <https://doi.org/10.1002/anie.201409928>.
- (39) Chawner, S. J.; Cases-Thomas, M. J.; Bull, J. A. Divergent Synthesis of Cyclopropane-Containing Lead-Like Compounds, Fragments and Building Blocks

through a Cobalt Catalyzed Cyclopropanation of Phenyl Vinyl Sulfide. *Eur. J. Org. Chem.* **2017**. <https://doi.org/10.1002/ejoc.201701030>.

## EXPANDING NEW-TO-NATURE REACTIONS BEYOND HEME PROTEINS

Content in this chapter is adapted from published work:

† denotes equal contribution

Goldberg, N. W.;<sup>†</sup> **Knight, A. M.**;<sup>†</sup> Zhang, R. K.; Arnold, F. H. Nitrene Transfer Catalyzed by a Non-Heme Iron Enzyme and Enhanced by Non-Native Small-Molecule Ligands. *J. Am. Chem. Soc.* **2019**, *141*, 19585-19588.

A.M.K. prepared a list of potential metalloproteins based on structural information, assembled biochemical information available for the top candidates, and designed codon-optimized genes for order. A.M.K. and N.W.G. transformed all plasmids into an expression strain of *Escherichia coli* (*E. coli*). A.M.K. performed whole-cell expression tests of the library. A.M.K. and N.W.G. expressed and purified non-heme metalloproteins and assayed their hydrosilylation activity. N.W.G. and R.K.Z. found initial nitrene-transfer activity with *Pseudomonas savastanoi* ethylene-forming enzyme (*PsEFE*). A.M.K. and N.W.G. ran controls to confirm the activity to be enzymatic. A.M.K. carried out alternative ligand experiments. A.M.K. designed site-saturation mutagenesis libraries. A.M.K. and N.W.G. constructed and A.M.K. screened and analyzed these libraries, and N.W.G. performed shake-flask validation of potentially improved variants. A.M.K. characterized evolved variants via thermostability assays, and time course experiments. A.M.K. generated crystallography constructs of C–H insertion variants and purified the constructs. A.M.K. and N.J.P. crystallized the variants. A.M.K. collected data and refined crystal structures.

## Abstract

Proteins that use metal ion cofactors catalytically have been underutilized to explore new-to-nature enzymatic reactions, despite the biocatalytic potential of these metalloproteins. To facilitate the use of this protein diversity, I have curated a collection of metalloproteins which bind redox-active metal ion cofactors, have solvent-accessible active sites, and have been shown to express heterologously in *E. coli*. One of these proteins, *Pseudomonas savastanoi* ethylene-forming enzyme (*PsEFE*), a non-heme iron enzyme, can catalyze olefin aziridination and nitrene C–H insertion reactions, and was improved by directed evolution. As this nitrene-transfer reaction does not require  $\alpha$ -ketoglutarate ( $\alpha$ KG) as a co-substrate, non-native ligands can be substituted in place of  $\alpha$ KG to modulate the enzyme's activity and selectivity.

## 4.1 Expanding new-to-nature metalloprotein chemistry beyond heme proteins

While enzymes from myriad functional classes have been engineered to enhance their stability and substrate scope in their native reaction classes, metalloproteins engineered for new-to-nature reactions have been almost exclusively dependent on metal complex cofactors, either native or nonnative. Carbene and nitrene transfer has been demonstrated with heme proteins binding their native cofactor, analogous iron cofactors such as chlorin,<sup>1</sup> or metal-substituted cofactors such as iridium,<sup>2</sup> rhodium,<sup>3</sup> and ruthenium porphyrins.<sup>4</sup> Beyond heme proteins, research groups have inserted non-native metal complexes into proteins using bioconjugations methods like biotin/streptavidin complexes. The Ward and Lewis groups have pioneered the use of bioconjugated rhodium complexes in protein scaffolds for new-to-nature reactions such as carbene transfer, hydrogenation, and olefin metathesis.<sup>5,6</sup> In each of these cases, the majority of the metal's primary coordination sphere is pre-formed as part of the cofactor. However, many metalloproteins in Nature consist of metal ions coordinated directly by the protein via side-chain and main-chain interactions. The nomenclature distinguishing metal cofactors which are single ions or part of coordination complexes is imprecise. For clarity, I will refer to metallocofactors with the metal tightly bound within a non-proteinaceous coordination complex as metal complex cofactors, and metallocofactors with the metal ion bound directly by the protein (via side-chain or main-chain interactions) as metal ion cofactors.

The ability to perform new-to-nature biocatalytic reactions with proteins that bind metal ion cofactors rather than metal complex cofactors would expand the degree to which the metal's primary coordination sphere can be modulated. Whereas a heme protein's coordination sphere can only be modified by mutating the single axial coordinating ligand or synthesizing and incorporating a non-native heme cofactor, the primary coordination sphere of a metal-ion-bound metalloprotein can be modified through multiple coordinating residue mutations and loosely associated co-substrates (e.g.  $\alpha$ -ketoglutarate ( $\alpha$ KG), Figure 4-1). Heme also lacks a *cis*-coordination site, having only two open coordination sites distal to each other. As certain reactions require two adjacent coordination sites, metalloporphyrins like heme are

advantageous for preventing these side reactions and improving reaction selectivity. If the reaction of interest requires two adjacent coordination sites, however, heme or non-native metalloporphyrins will not be capable of performing the transformation.

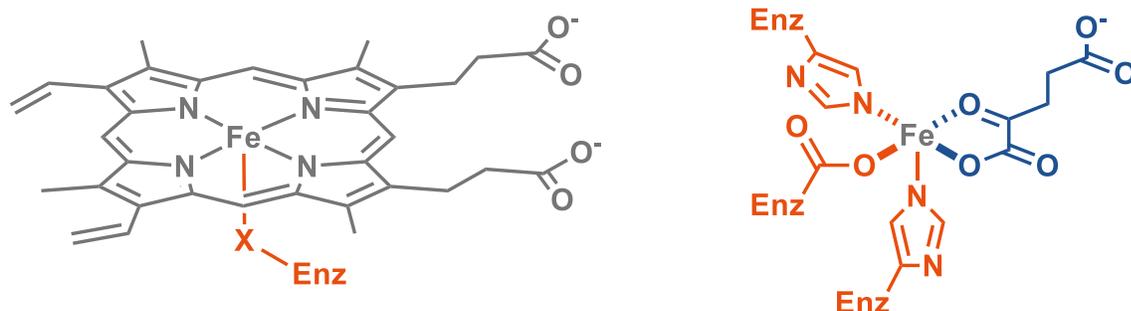


Figure 4-1. Potential primary coordination sphere plasticity in heme proteins and non-heme metalloproteins. In metal complex cofactor metalloproteins such as heme proteins, four metal coordination sites are occupied by the (natural or non-natural) porphyrin ring, requiring porphyrin synthesis and substitution to alter most of the primary coordination sphere. In metal ion cofactor metalloproteins such as  $\alpha$ KG-dependent enzymes, multiple side-chain interactions can be modified with mutagenesis, and the metal and ligand can both be substituted through modified reaction conditions.

One strategy which has been used successfully is modulating enzyme activity and selectivity using non-iron metalloporphyrins in heme proteins.<sup>2</sup> Like hemoproteins, metalloproteins featuring a metal ion cofactor can be metallated with a non-native metal ion (xenometallation) to modulate their activity. Under physiological conditions, metalloproteins are either metallated through cellular metal delivery systems or are metallated based on local metal concentrations and the protein's specificity.<sup>7-9</sup> Generally, mismetallated proteins are simply inactive, but in some cases enzymes have been shown to have activity for their native reaction with multiple metals (cambialism). Ancestral nitrogenases have been predicted to be cambialistic,<sup>10</sup> and several superoxide dismutases have been shown to be cambialistic, performing their native reactions with either iron or manganese.<sup>11</sup> In biocatalytic reactions, where parameters such as metal ion concentrations can be easily adjusted, an apoenzyme of interest can be xenometallated by adding an excess of a different metal for which the enzyme has at least a weak binding affinity. Xenometallation has been studied in copper proteins for

decades, with native and engineered copper proteins shown to bind cobalt,<sup>12</sup> iron,<sup>13</sup> nickel,<sup>14</sup> and even noble metals.<sup>15</sup>

The Arnold laboratory and others have developed many interesting and useful new-to-nature biocatalytic reactions,<sup>16,17</sup> primarily through screening variants of a small set of heme protein scaffolds. The world of metalloprotein diversity is vast, with an estimated 30% of all proteins binding a metal; the potential for new-to-nature biotransformations using non-heme metalloproteins is likewise vast. We therefore set out to develop a collection of non-heme metalloproteins which could be tested for promiscuous activity for new-to-nature reactions of interest.

## **4.2 Curating a collection of non-heme metalloproteins for discovering new-to-nature activities**

When selecting a starting protein for a protein engineering campaign, one generally chooses a protein with the desired catalytic activity on a substrate similar to the target substrate. Databases such as UniProt<sup>18</sup> and BRENDA<sup>19</sup> have sequence and functional information on (most if not) all known enzyme classes. Selecting proteins with reported function on compounds similar to the substrate of interest and testing those proteins for initial activity on this compound, while not guaranteed to work, is a tried-and-true method to find initial enzymatic activity which one can improve with directed evolution.

For a desired new-to-nature activity, however, one cannot simply look in a functional database to find an enzyme with the catalytic activity on a similar substrate. When first reported, carbene and nitrene transfer using a heme protein would not appear in a functional database like BRENDA or UniProt; the only result matching the search term “carbene transfer” in UniProt is *Rma* cytochrome *c* (UniProt ID B3FQS5), which was engineered in the Arnold laboratory for this new-to-nature activity.<sup>20</sup> Inspiration to search for these novel biocatalytic activities came from the chemical catalysis literature. Chemists have been using metalloporphyrins for C–H activation reactions for decades,<sup>21</sup> having in turn been inspired by the remarkable breadth of oxene-transfer reactions performed by heme proteins.<sup>22</sup> Coelho

and coworkers therefore took cytochrome P450<sub>BM3</sub>, the workhorse heme protein in the Arnold lab, and assayed its styrene cyclopropanation activity, a model reaction for carbene transfer. The activity and selectivity of the protein are modulated by the heme's axial ligand and the surrounding protein environment, but the coordination geometry and the primary electronic contribution is fixed by the porphyrin ring of the heme cofactor. Indeed, heme itself will catalyze styrene cyclopropanation and many more reported transformations, albeit at low levels.

The difficulty of selecting a suitable protein scaffold increases when the primary coordination sphere of the metal is composed entirely of protein interactions and labile interactions with water or small molecules. A reaction of interest might have a known catalyst in the chemistry literature, but designing a catalyst's primary coordination sphere into a protein active site is computationally expensive and infeasible in most cases currently, though researchers have made some progress in the field of *de novo* enzyme design.<sup>23</sup> Rather than attempting to computationally design a new catalytic active site into a protein structure, we instead decided to rely on the prevalence of enzyme catalytic promiscuity<sup>24</sup> and looked to Nature's diversity of metal-binding proteins for candidate protein scaffolds for new-to-nature reactions. We initially purchased, subcloned, and successfully expressed the genes encoding 14 copper-binding proteins to use in testing non-heme metalloprotein reactions. These were selected because copper complexes are well known in the synthetic literature and the mononuclear cupredoxins have been extensively studied by bioinorganic chemists. We also requested and received plasmids encoding non-heme metalloproteins of interest from other research groups; these were primarily non-heme iron and copper-binding proteins from Professors Hans Renata (Scripps Research) and Harry Gray (California Institute of Technology). We quickly realized that the collection should be expanded to cover more protein classes and sequence diversity and set out to take a more systematic approach in the curation of this collection.

#### 4.2.1 Criteria selection and bioinformatic analysis of the Protein Data Bank

To ensure we were looking at as large of a pool of potential proteins as possible, I expanded our search for any protein binding redox-active metal ions of interest. When protein engineers look for a starting point, they can usually turn to databases where protein activities are annotated. As we are looking for non-natural activities, we cannot rely on functional annotations, but rather must hypothesize based on active-site structures and geometries. We therefore opted to narrow the search by looking at proteins with a reported crystal structure, as this was the easiest way to see if the (putative) active site might accommodate our substrates of interest. I used the Protein Data Bank (PDB), a collection of over 140,000 crystal structures,<sup>25</sup> and the MetalPDB, a database with annotated coordination geometries of metals within proteins in the PDB.<sup>26</sup> To narrow the structures in the PDB and MetalPDB databases down to potentially interesting proteins, I used several search criteria. First, structures were required to contain a redox-active metal ion; cobalt, copper, iron, nickel, and manganese were included. Zinc, while bound broadly by metalloproteins, was omitted as it is redox inactive and only used in enzymes as a Lewis acid.<sup>27</sup> I composed a list of all PDB ligand IDs for these metal ion cofactors while omitting metal complex cofactors which contain these metals, ensuring that only metal ion cofactors were included as results. MetalPDB metal ion cofactor search queries were similar and included specifying coordination geometries. The proteins selected were also kept to fewer than 580 amino acids in length; while there is a slight trend of larger proteins exhibiting lower stability,<sup>28</sup> this cutoff was a practical choice due to the price and turn-around time increasing above sequences longer than 1800 nucleotides for the chosen gene synthesis service. The search was also limited to proteins reported to have been expressed in *E. coli* for crystallization; this would ensure that the protein of interest had been successfully expressed in our preferred expression host. Finally, sequence identity for representative structures was set to 95%, preventing engineered protein variants or commonly studied proteins from being overrepresented in the results.

Combining the results from the MetalPDB and PDB searches yielded a collection of 2572 unique candidate protein structures. To narrow down these 2572 potential proteins, I decided

the best approach was visual inspection of the proteins rather than relying on any further annotated information. I also wanted a diverse array of protein structures and as such did not want to further narrow down the candidate proteins solely via structural criteria beyond the initial criteria listed above. To make the visual inspection of large numbers of protein structures more efficient, I generated structure viewing files (PyMOL sessions) with settings to make it easy and clear to see the geometry of the active site, its solvent exposure and accessibility, and other factors which could be beneficial or detrimental. Candidate proteins were chosen based on the solvent accessibility of the metal site and interesting features of the active-site region. This analysis resulted in a list of 179 candidate proteins.

I then reviewed the literature on the proteins on this short list to determine if there are any which should be ruled out based on previous reported results (toxicity, poor expression, unstable, etc.). For many proteins submitted, this information was not available. A lack of information on poor protein stability or poor expression did not preclude a given protein, but literature documentation of instability or toxicity would remove a given candidate. Many proteins of interest were from structural genomics projects for which there are expression and purification methods, but no peer-reviewed research associated. Proteins from thermophilic organisms were preferred as they are generally thermostable, and thermostability has been shown to be a key trait when choosing a starting point for evolution.<sup>29</sup> 65 sequences were codon-optimized for *E. coli* and sent for gene synthesis, of which 62 were successfully synthesized and cloned into pET expression vectors. Overall, the proteins selected represent substantial diversity in the metalloprotein sequence space; the highest pairwise sequence identity was only 49%, with a mean sequence identity of 10%.

#### **4.2.2 Construct preparation and initial expression tests**

To efficiently turn a list of candidate metalloproteins into expressed proteins ready for testing, we decided to order the collection of genes already inserted into an *E. coli* expression vector. As Twist Biosciences did not have an in-house protein expression vector which used our standard T7lac promoter expression system, a custom pET expression vector was onboarded with the company. All genes encoding proteins of interest were cloned by Twist

Biosciences into this expression vector. For proteins which were predicted to natively require a signal peptide to form the holoprotein *in vivo* (based on UniProt annotations and the SignalP server),<sup>30</sup> the gene was synthesized with the nucleotides to encode an N-terminal pelB signal peptide. This is particularly important for genes such as lytic polysaccharide monooxygenase (Uniprot ID D6EWM4), for which the amino group and side chain of the N-terminal His residue is part of the primary coordination sphere.<sup>31</sup>

As Twist Biosciences did not have an in-house strain of *E. coli* with the  $\lambda$ DE3 lysogen (necessary for expression with the T7lac promoter in our vector), *E. coli* BL21(DE3) electrocompetent cells were transformed with the plasmids from Twist Biosciences. Each protein had been previously reported to express heterologously in *E. coli*, but it was important to confirm that the proteins would express in our plasmid construct under standard expression conditions. *E. coli* harboring each plasmid were grown at small (5 mL) scale and expression of the protein of interest was assayed with protein gel electrophoresis (Supplementary Information, Figure 4-29). Under standard expression conditions, 40 of the 62 proteins were strongly overexpressed, and only seven had no detectable expression. Of these seven, four are pelB-tagged proteins, which in our experience have enhanced expression at reduced IPTG induction concentrations.

The names, bound metals, and UniProt IDs for 86 non-heme metalloproteins (including the initial cupredoxins ordered and cloned in house, the clonal genes acquired from other research groups, and the clonal genes ordered from Twist Biosciences) are in the Supplementary Information (Table 4-4 and Table 4-5). The DNA and amino-acid sequences are deposited as a supplemental worksheet through CaltechDATA (DOI: 10.22002/D1.1437).

This collection of non-heme metalloproteins is available to the biocatalysis community.

### **4.3 Toward enzymatic hydrosilylation with non-heme metalloproteins**

One class of catalytic mechanisms in which we were interested are reactions which go through oxidative addition and reductive elimination pathways. Despite the ubiquity of these

elementary steps in the synthetic catalytic repertoire, enzymatic mechanisms using these steps are limited to complex systems such as nitrogenases.<sup>32</sup> Key examples include the Suzuki-Miyaura cross-coupling reactions discussed in Chapter 3, hydroamination, hydroboration, and hydrosilylation. These reactions generally require *cis*-coordination sites inaccessible to heme-protein catalyzed reactions. While the Arnold laboratory reported the first enzyme known to form a carbon–silicon bond by using an engineered heme protein, many of the organosilicon compounds ubiquitous to the modern world are generated through industrial-scale hydrosilylation reactions. To this end, we set out with the intention of engineering metalloproteins as catalysts for hydrosilylation, in collaboration with Dow, Inc.

In many initial discovery experiments, a large set of proteins, either variants developed from previous directed evolution campaigns or newly cloned constructs, is used to screen for one specific desired reaction. Our goal was to find promiscuous activity for any hydrosilylation reaction in a large and diverse enzyme library (curated as described in Section 4.2); once activity was found for a hydrosilylation reaction, we reasoned that substrate walking could expand to diverse hydrosilylation reactions utilizing the same catalytic mechanism. We assembled a set of model substrates, including activated, unactivated, and electron-deficient alkenes (Figure 4-2). The silanes used were limited by the fact that alkoxy silanes and silanes with fewer substituents will hydrolyze under aqueous conditions.

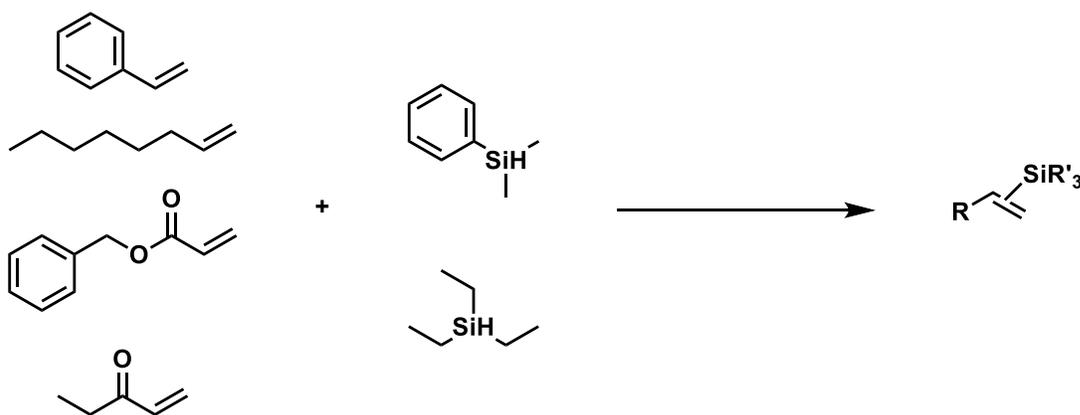


Figure 4-2. Initial hydrosilylation reaction tests. Each pairwise combination of alkene and silane was tested individually. Reaction conditions: anaerobic, RT, 10 mM silane, 10 mM alkene, whole *E. coli* cells, cell lysate, and cell lysate with 5 mM sodium dithionite.

The first group of enzymes we expressed and used in whole *E. coli* cells were copper enzymes (Table 4-4). As phosphate buffers rapidly precipitated copper from solution, alternative buffers were tested to minimize metal salt precipitation. MOPS buffer displayed less copper precipitation and was used for non-heme metalloprotein reactions for the remainder of the project. We determined by UV-visible spectroscopy that these proteins were metallated *in vivo* when expressed in the periplasm (Figure 4-3). *E. coli* cells expressing these copper proteins would turn vivid shades of blue following addition of copper sulfate. These proteins were tested against the initial set of four alkenes and two silanes as whole cells, lysate, and lysate with sodium dithionite added.

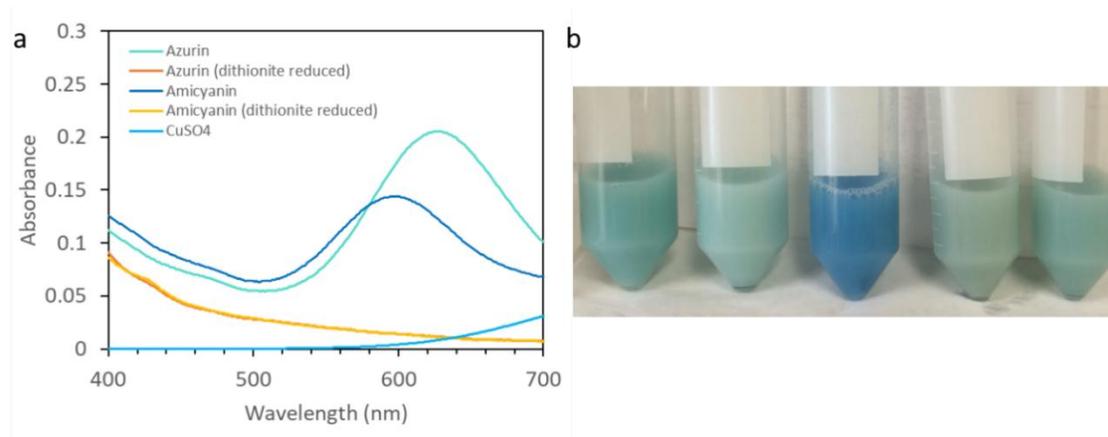


Figure 4-3. Copper-binding protein metallation. (a) Example UV-visible spectra of the copper binding proteins azurin and amicyanin. Copper sulfate and dithionite-reduced holoprotein samples do not display the characteristic absorbance peaks from 600 – 620 nm. (b) Copper protein metallation in whole *E. coli* cells. From left to right: *Pae* azurin, *Afe* rusticyanin, *Pde* amicyanin, *Ppu* azurin, *Rme* pseudoazurin.

Preliminary hydrosilylation reaction screening using whole cells and crude cell lysates containing the copper proteins did not yield detectable hydrosilylation activity. As there is literature precedent that the coordinating sidechains in azurins can be mutated without complete loss of copper binding,<sup>33,34</sup> Nat Goldberg also explored mutants which would have an additional vacant coordination site (i.e. bound to water rather than an amino-acid side chain). These variants still had soluble expression and bound copper, but were also not found to perform the tested hydrosilylation reactions.

We therefore decided to pursue further screening using purified proteins. The use of purified proteins offers several advantages over whole cells or lysates: it allows for a known quantity of catalyst in each reaction and normalizing for variable expression. Purified protein reactions can also be run at higher catalyst loading, such that even single-digit enzyme turnover or stoichiometric activity would be detectable in sensitive GC-MS screening. The removal of the cellular milieu from the reactions also prevents any potential cellular background reactions from taking place and removes small-molecule cellular contaminants from the analysis, leading to cleaner GC-MS spectra.

Since protein purification is both time and resource intensive, we wanted to maximize protein production and minimize the quantity of each protein of interest required. We first pursued a strategy to maximize the reaction space explored by pooling several substrates in each well (Figure 4-4). This allowed us to screen several potential hydrosilylation reactions using high catalyst loading, without requiring prohibitively large quantities of each protein of interest. We chose a set of olefins and carbonyls of varied electronic and geometric character, as well as a set of three different silanes that are stable under the reaction conditions. We tested reactions at three different pH values (pH 5, 7, and 9 by using MES, MOPS, and CHES buffers), as this can alter the protonation state of the coordinating residues, which affects how easily they can dissociate from the metal center and vacate substrate binding site(s) for hydrosilylation. We also examined the absence or presence of a reductant, sodium dithionite, as the oxidation state of the metal center is key for catalytic activity.

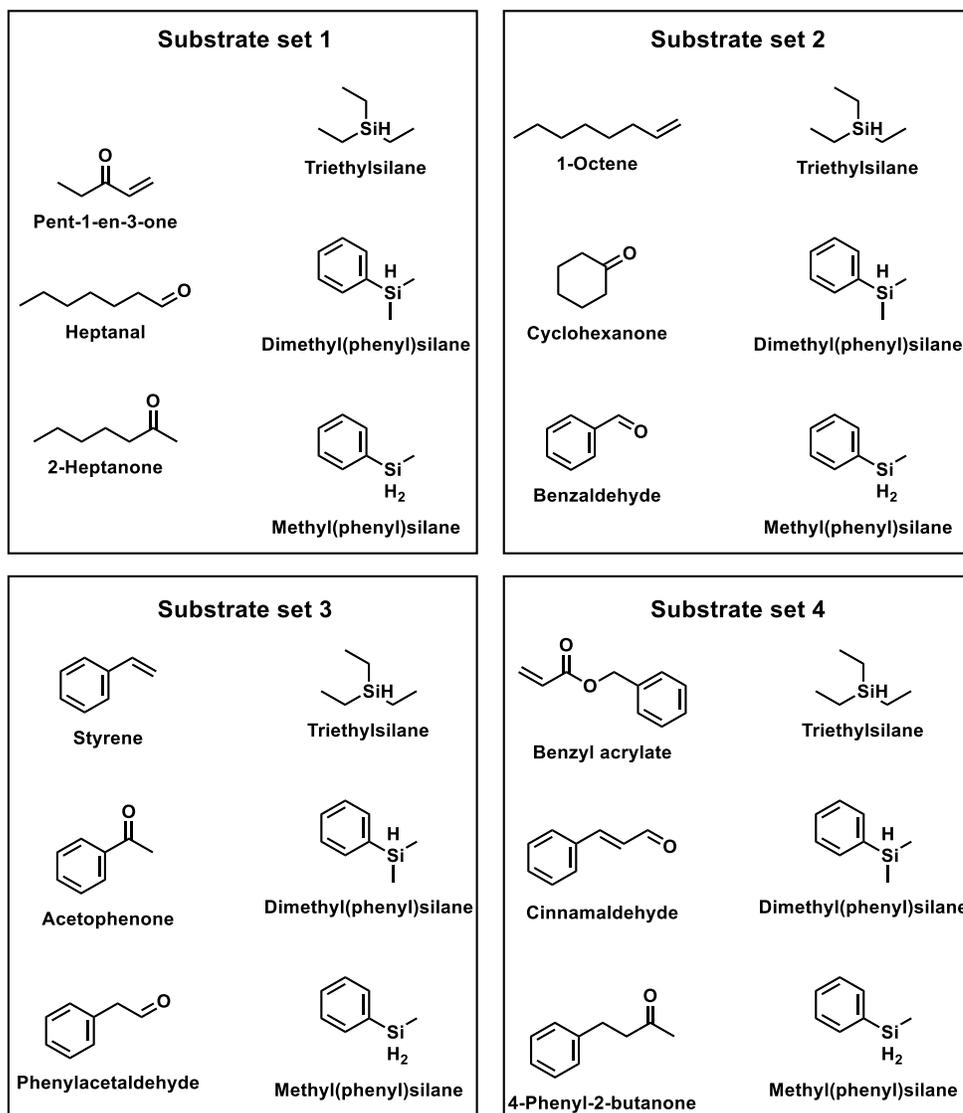


Figure 4-4. Pooled substrate sets. Each substrate set was added as a mixture for a single biocatalytic reaction. GC-MS analytics allowed for mass-fragmentation-specific analysis; any potential activity could then be followed up on with reactions containing a single substrate-silane pair.

These pooled substrate reactions were used in screening wild-type and ligand-mutated copper proteins. In each case that a potential hydrosilylation product peak was detected, follow-up controls showed that the peak was an artifact or non-enzymatic background (false positive). We re-evaluated the pooled reaction approach, deciding to focus on a single

substrate-silane pair in each reaction to reduce the sample complexity and potential for cross reactivity (Figure 4-5).

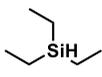
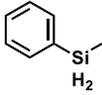
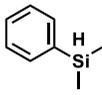
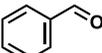
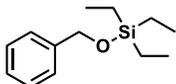
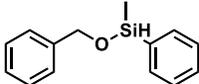
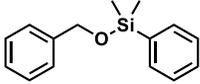
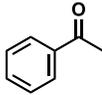
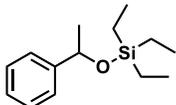
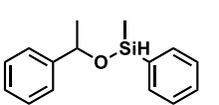
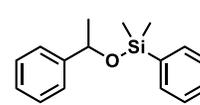
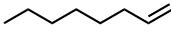
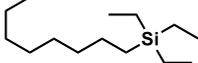
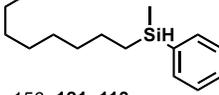
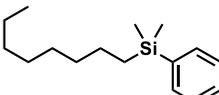
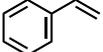
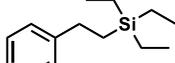
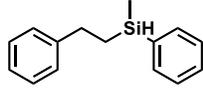
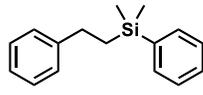
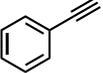
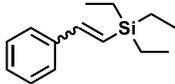
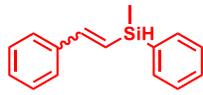
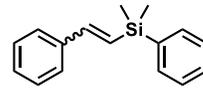
	 <p><b>Triethylsilane</b> Too volatile</p>	 <p><b>Methyl(phenyl)silane</b> 122, 121, 107, 105 2.53 min</p>	 <p><b>Dimethyl(phenyl)silane</b> 136, 135, 121, 105 2.93 min</p>
 <p><b>Benzaldehyde</b> 106, 105, 77 3.04 min</p>	 <p>193, 163, 135, 91 5.22 min</p>	 <p>197, 183, 149, 137, 91 5.73 min</p>	 <p>227, 197, 164, 148, 91 5.88 min</p>
 <p><b>Acetophenone</b> 120, 105, 77 3.62 min</p>	 <p>207, 103, 75 5.14 min</p>	 <p>227, 197, 167, 149, 137 5.65, 5.69 min</p>	 <p>241, 178, 167, 137 5.79 min</p>
 <p><b>1-Octene</b> 112, 83, 70, 55 2.01 min</p>	 <p>199, 171, 115, 87 5.18 min</p>	 <p>156, 121, 113 5.70 min</p>	 <p>233, 170, 137, 135, 121 5.83 min</p>
 <p><b>Styrene</b> 104, 78, 51 2.61 min</p>	 <p>191, 163, 135, 87 5.48 min</p>	 <p>226, 121, 105 5.76 min</p>	 <p>225, 162, 147, 135, 121 6.19 min</p>
 <p><b>Phenylacetylene</b> 102, 76 2.51 min</p>	 <p>189, 161, 131, 105 5.34, 5.58 min</p>	 <p><b>Synthesis failed</b></p>	 <p>238, 223, 145, 135, 121, 105 6.00, 6.34 min</p>

Figure 4-5. Final hydrosilylation substrate matrix. The characteristic mass fragmentation pattern and retention times are given for the GC-MS method on an Agilent GC-MS (7820 GC oven, 5977B single-quadrupole MS) with a 30m × 0.25 mm ID × 0.25 μm film thickness DB-5ms column and the following program: hold at 80 °C for 1.7 min, ramp at 45 °C/min to 325 °C, hold at 325 °C for 2 min.

At this time, the larger metalloprotein collection ordered through Twist Biosciences (Table 4-5) arrived. While the copper-binding proteins have characteristic UV-visible spectra upon metallation, many of the other proteins in this new collection do not. We therefore investigated mass spectrometric techniques to determine protein metallation. Protein LC-MS and protein TOF were both tested using azurin holoprotein confirmed by UV-visible spectroscopy to be metallated. These native protein approaches invariably resulted in apoprotein detection; the protein's instability to the analytical conditions appeared to result in demetallation. As an alternative metal detection route, we found that inductively coupled plasma-mass spectrometry (ICP-MS) is a reliable way of determining enrichment of the metal of interest in our protein samples. It should be noted that this will detect all metals in solution and not only those specifically bound to the protein. To determine background metal concentrations, we used a sample of the final flow through from each protein's purification as negative control. The analysis from our first round of ICP-MS data collection showed that the deionized water used for the protein purification and storage buffers had high background in some metals, most notably iron. We have since then adapted our protein purification protocol to use distilled deionized water, which lowered this background substantially (an average of 19-fold lower  $^{56}\text{Fe}$  background). There was consistently high Ni content in the protein samples, which we attributed to the C-terminal polyhistidine tag and the Ni affinity chromatography used to purify the proteins. Using protein ICP-MS, we have confirmed the enrichment of the expected metal ligands in 13 of 17 proteins analyzed. The proteins which did not have enrichment were exclusively non-heme iron proteins which also use  $\alpha\text{KG}$  as a coordinating ligand to the iron. In the absence of  $\alpha\text{KG}$ , it is possible that the binding affinity for the metal is not strong enough to keep the metal bound throughout the purification process. Further reactions with  $\alpha\text{KG}$  binding proteins should therefore be tested with  $\alpha\text{KG}$  or an analogous small-molecule ligand added.

To minimize chances of false positives due to cross contamination from other in-house reaction screening, we moved to using a GC-MS DB-5ms column dedicated to these reactions. The GC-MS methods were also optimized for minimal inlet split ratios, maximizing the signal-to-noise ratio while maintaining peak shapes. Using authentic

standards that Nat synthesized, we were able to confirm the retention time and mass fragmentation for each compound. I prepared separate GC-MS methods for each substrate pair, using single ion monitoring (SIM) mode to monitor the characteristic mass fragments expected in the hydrosilylation products (Figure 4-5). The reactions with this  $3 \times 5$  substrate matrix were set up at only one pH (MOPS pH 7) in the presence and absence of sodium dithionite, resulting in 30 reactions per protein. Two proteins could therefore be set up in a single 96-well deep-well plate while leaving the outer wells empty, avoiding potential differences in assay conditions sometimes observed (edge effects).

One key control was ensuring that hydrosilylation was possible under conditions typical for biocatalytic reactions (aqueous, room temperature, anaerobic) and that our substrates and potential products are not degraded under these conditions and can be detected by our methods. To this end, we performed hydrosilylation reactions of each of our model silanes and olefins/alkyne in aqueous buffer with Karstedt's catalyst and Speier's catalyst, both known platinum catalysts for hydrosilylation, for a total of 18 reactions. In all 18 reactions, we observed high consumption of the silane starting material (>50%). Reactions with triethylsilane and dimethyl(phenyl)silane with both olefins and alkyne were generally very clean, with at least 50% yield of the expected anti-Markovnikov hydrosilylation products. The corresponding silanol byproducts were not observed in significant quantities. Reactions with methyl(phenyl)silane also showed hydrosilylation products, but with more uncharacterized side products, including probable silanols. Silanol formation from methyl(phenyl)silane is not observed in the absence of a catalyst, suggesting that the catalysts are activating this silane rather than a background degradation.

Control reactions were key to ensuring any hydrosilylation product detected was enzymatic. In most cases, a false positive was ruled out by comparing to negative control reactions or testing whether the peak area was proportional to the enzyme loading. In one case, the false positive came from an unexpected component of the screening workflow. The hydrosilylation of dimethyl(phenyl)silane and styrene was observed under screening conditions, but we failed to reproduce this activity when running validation reactions. A key

difference was that screening reactions were prepared in 96-well deep-well plates and validation was run in 2-mL glass vials. Control reactions showed that otherwise identical reactions set up in the plates showed activity when the glass vial reactions did not. Through a process of elimination, we determined that the activity was coming from the plate seal applied during the solvent extraction post reaction; the plate seal, a PTFE-coated silicone plate-sealing mat material (ArcticWhite LLC, part number AWSM-1003S), was used such that the solvent extraction could be performed through vortexing. The silicone mat was presumably manufactured via industrial hydrosilylation and contained trace catalyst. The silicone mat was confirmed to catalyze the reaction by adding a piece of this plate sealing mat to a glass vial containing aqueous buffer, styrene, and dimethyl(phenyl)silane — only reactions containing the silicone sealing mat showed the hydrosilylation product (Figure 4-6). This unexpected background reaction was removed by performing solvent extractions through pipette mixing rather than vortexing. This shows both the ubiquity of hydrosilylation products in the world and the importance of control reactions.

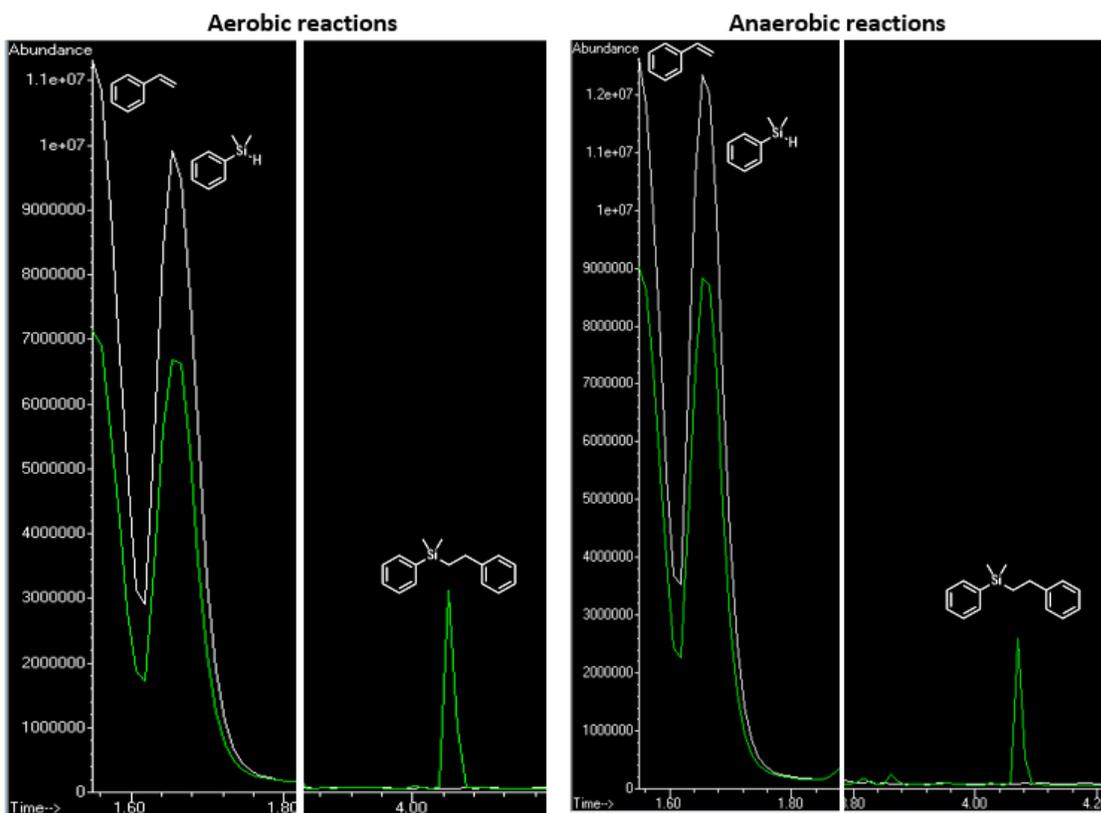


Figure 4-6. Background reactivity caused by a silicone sealing mat. When the substrates dimethyl(phenyl)silane and styrene were added to aqueous buffer in the presence of PTFE-coated silicone plate-sealing mat material (ArcticWhite LLC, part number AWSM-1003S), a peak with the expected retention time and mass fragmentation pattern was present (green). The peak was not observed when an otherwise identical reaction vial was prepared in the absence of the plate-sealing mat material (white).

With the  $3 \times 5$  silane / substrate matrix (Figure 4-5) and updated GC-MS analytics, we screened an additional 36 purified proteins from the metalloprotein collection. In this screening, there were no reactions which displayed hydrosilylation product formation. Future work on enzymatic hydrosilylation could continue screening native metalloprotein diversity. One particular area of metalloprotein diversity that can be explored are proteins which include a metal-hydride intermediate as part of their native catalytic mechanism, such as nitrogenases<sup>32</sup> and hydrogenases.<sup>35</sup> The complexity of the cofactor synthesis pathways and holoprotein maturation for these proteins (as well as extreme oxygen sensitivity) makes them challenging to work with and poor candidates for biotechnological applications. Many

research groups are focused on improving the expression and stability of these proteins in model organisms, however, and these more robust expression platforms could be tested as potential starting points. As more non-heme metalloproteins are engineered for other new-to-nature reactions on model substrates like those tested in this section, those variants can be screened for their activity on hydrosilylation and related reactions. Once initial activity is found, it has been shown countless times that directed evolution can improve that activity – all that is needed is a starting point and a good screen.<sup>36</sup>

#### **4.4 Nitrene-transfer activity with $\alpha$ KG enzymes**

In addition to the suite of reactions tested as part of the Dow collaboration, we were interested in testing non-heme metalloproteins for their ability to perform carbene- and nitrene-transfer reactions similar to the transformations for which heme proteins have been previously engineered.<sup>16</sup> There are a few rationales for the decision to pursue reactions already catalyzed by engineered heme proteins: these new-to-nature reactions are known to be possible to perform enzymatically under physiological conditions, the starting materials are inexpensive and/or already available, and the analytical instrumentation and methods are established.

The  $\alpha$ KG-dependent enzymes in the curated collection natively catalyze oxene-transfer reactions via an iron-oxo intermediate to hydroxylate their substrates, consuming molecular oxygen and converting one equivalent of  $\alpha$ KG to succinate in the process (Figure 4-7). As Coelho and coworkers hypothesized that the catalytic oxene-transfer machinery of P450<sub>BM3</sub> could also form iron-carbenoid and iron-nitrenoid species,<sup>37</sup> we hypothesized that the analogous machinery in non-heme iron proteins could perform carbene- and/or nitrene-transfer reactions.

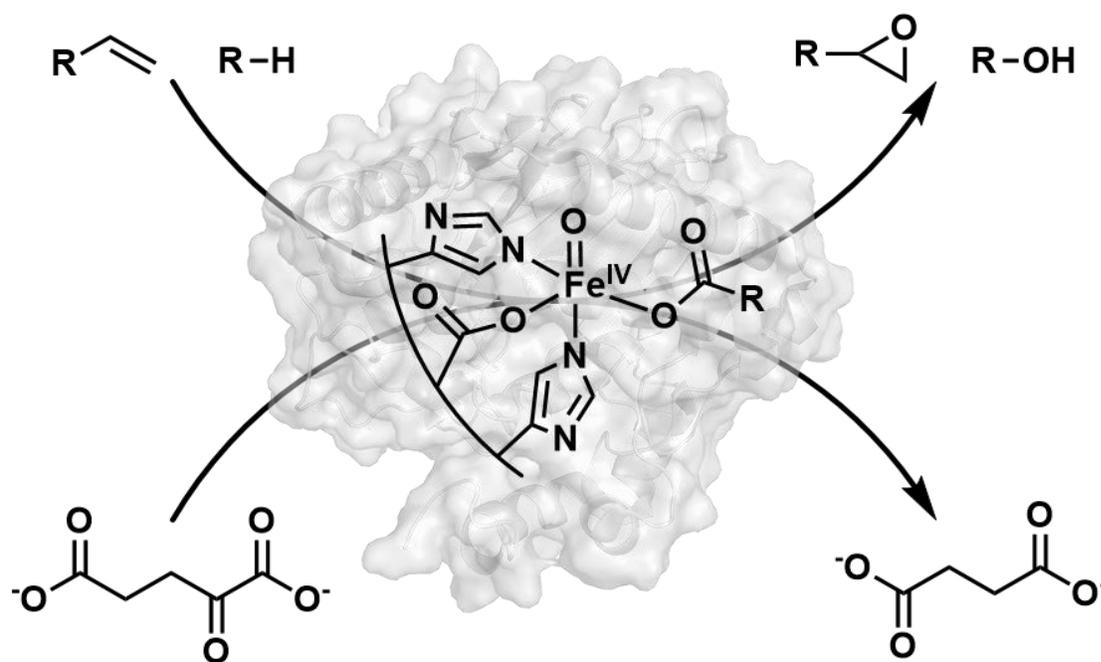


Figure 4-7. Typical oxene-transfer reactions carried out by non-heme metalloenzymes.

While testing our collection of  $\alpha$ KG-dependent enzymes for nitrene-transfer activity, we found that one protein, wild-type ethylene-forming enzyme from *Pseudomonas savastanoi* (*PsEFE* WT), showed greater than background activity for the aziridination of styrene and tosyl azide (Figure 4-8,

Table 4-1).

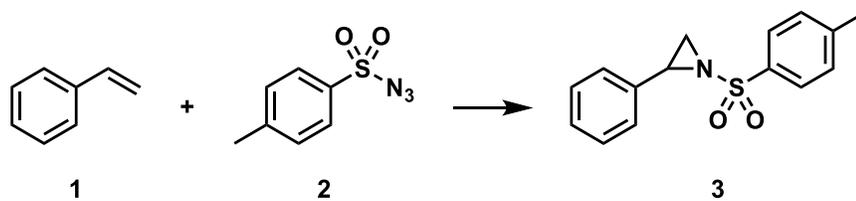


Figure 4-8. Nitrene-transfer reaction catalyzed by non-heme iron proteins.

Table 4-1. Activities of  $\alpha$ -KG-dependent iron enzymes toward aziridination to form **3**.

Enzyme	Relative activity
<b><i>P. savastanoi</i> ethylene-forming enzyme</b>	<b>12.0</b>
<i>Streptomyces</i> sp. 2-aminobutyric acid chlorinase	0.93
<i>A. thaliana</i> anthocyanidin synthase	0.54
<i>G. oxydans</i> leucine dioxygenase	1.11
<i>E. coli</i> taurine dioxygenase	0.61
<i>S. vinaceus</i> arginine hydroxylase	0.57
<i>S. muensis</i> leucine hydroxylase	0.61
Bovine serum albumin (negative control)	1.00

*PsEFE* WT has been studied both for the potential biotechnological application of efficient ethylene production<sup>38</sup> as well as the fact that it has two native activities.<sup>39</sup> In its first, more classical  $\alpha$ KG-dependent enzyme activity, an equivalent of  $\alpha$ KG is used to oxidize L-Arg to form pyrroline-5-carboxylic acid, an intermediate in proline synthesis (Figure 4-9a). Its second native reaction is the conversion of  $\alpha$ KG into ethylene (Figure 4-9b).

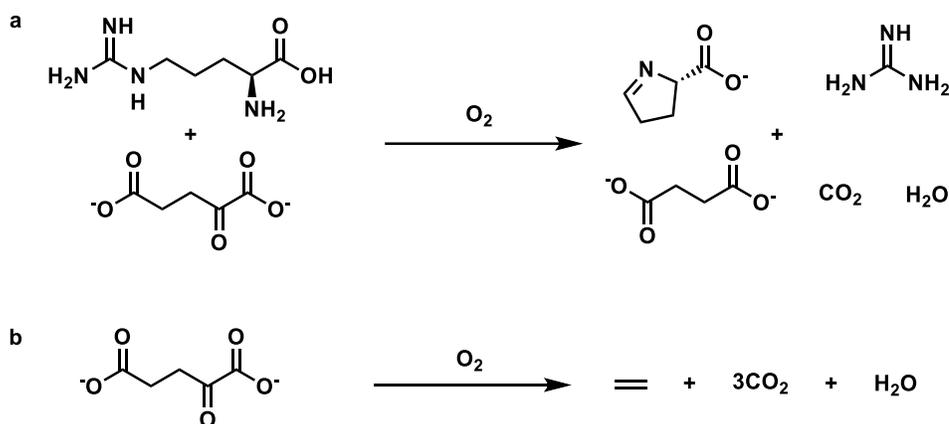


Figure 4-9. Reactions catalyzed natively by *Pseudomonas savastanoi* ethylene-forming enzyme (*PsEFE*). (a) An equivalent of  $\alpha$ KG is used to oxidize L-Arg to form pyrroline-5-carboxylic acid, an intermediate in proline synthesis, with guanidine, succinate, carbon dioxide, and water as byproducts. (b)  $\alpha$ KG is converted into ethylene with water and three equivalents of carbon dioxide as byproducts.

When testing new reactions, it is critical to ensure that product formation is enzyme catalyzed and not a background reaction. Several additional controls were carried out to ensure that the activity was due to *PsEFE* WT and not due to background activity. These controls include running the reaction in the absence of key reagents such as iron, ascorbate (reductant), and  $\alpha$ KG (co-substrate). In addition, we generated a protein variant which lacked two of the coordinating amino-acid residues (*PsEFE* H189A D191A). This double mutant protein still solubly expressed and purified, but was expected to have a greatly disrupted metal-binding site. These control reactions showed that omission of iron or the use of the double mutant abolishes activity, indicating that the reaction is enzymatic and occurs in the iron-binding site (Table 4-2). The activity is also proportional to the catalyst loading, lending further evidence that it is enzymatic. Another test is whether the product is enantioenriched; while the enzyme might form racemic product, it is unlikely that a background reaction would form an enantioenriched product. Chiral HPLC analysis showed that *PsEFE* WT was (*R*)-selective, with 25% enantiomeric excess (ee); the P411<sub>BM3</sub> variant previously engineered for this nitrene-transfer reaction was (*S*)-selective.

Table 4-2. Control and ligand-substitution reactions for *PsEFE* WT aziridination to form **3**.

Deviation from standard conditions	Aziridine yield (%)
H189A D191A mutant (disrupted iron binding)	<0.01%
No iron	0.01%
No $\alpha$ KG or analog	0.04%
None (standard conditions)	0.08%
Succinate instead of $\alpha$ KG	0.11%
Acetate instead of $\alpha$ KG	0.56%
<i>N</i> -oxalylglycine instead of $\alpha$ KG	0.64%
No ascorbate, acetate instead of $\alpha$ KG	0.50%

Given the ability to catalyze the aziridination of tosyl azide and styrene, we investigated the scope of nitrene-transfer reactions that *PsEFE* WT could catalyze. The tests focused on the aziridination of unactivated alkenes, which had not been reported with the P411<sub>BM3</sub> aziridination platform.<sup>40</sup> We also tested 4-ethyl anisole, a model substrate for benzylic C–H amination.<sup>41</sup> No products were detected for the intermolecular C–H amination of 4-ethyl

anisole or the aziridination of 1-octene, but the aziridination of allylbenzene and 4-phenyl-1-butene were detected, with 1.5 to 2-fold higher signal than the background and negative control reactions (Figure 4-10).

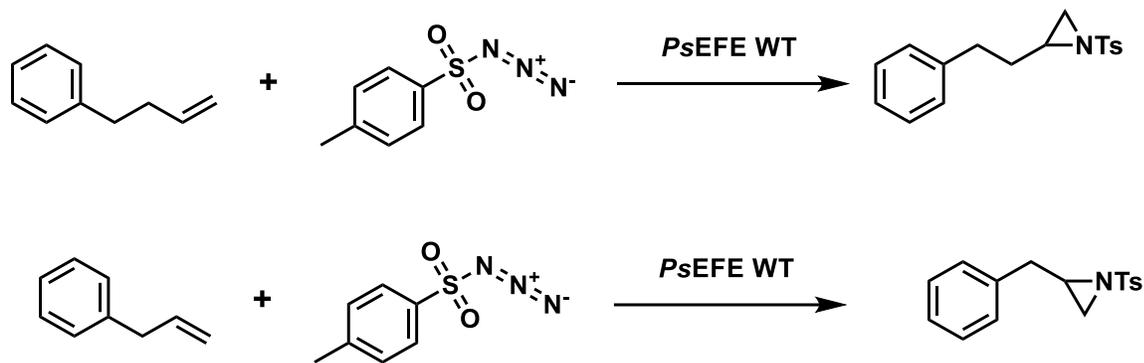


Figure 4-10. Aziridination of unactivated alkenes. Reaction conditions: 50  $\mu$ M *PsEFE* WT purified protein, 1 mM ferrous ammonium sulfate, 1 mM  $\alpha$ KG, 10 mM alkene, 10 mM tosyl azide, 5% ethanol cosolvent, 2 hours, room temperature.

In addition to testing for further nitrene-transfer reactions, an array of carbene-transfer reactions were tested with *PsEFE* WT. The formation of the metal-carbenoid intermediate is a crucial prerequisite for the carbene-transfer reaction to occur, and its formation is heavily dependent on the catalyst and the specific diazo compound used as a carbene precursor. Four diazo compounds with varied steric and electronic properties were tested (Figure 4-11). Xenometallation of *PsEFE* was also attempted, adding first-row transition metals (Mn, Ni, Co, Cu) in place of Fe in reactions. To limit the combinatorial expansion of test reactions, styrene was used as the alkene in all carbene-transfer reaction attempts. GC-MS traces of organic extracts of these reactions were analyzed for mass spectra corresponding to both carbene transfer to styrene and for dimerization of the diazo molecules; no carbene transfer activity above the negative control background was observed, and no clear evidence of carbene formation (e.g. nitrogen evolution) was noted in the reactions.

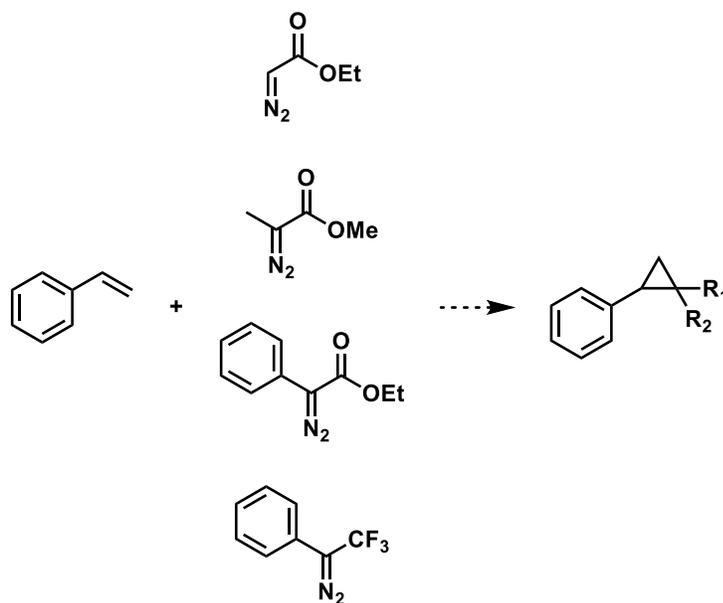


Figure 4-11. Attempted carbene transfer reactions. Reaction conditions: 50  $\mu\text{M}$  purified *PsEFE* WT, 100  $\mu\text{M}$  metal salt, 1 mM  $\alpha$ -KG, 1 mM sodium dithionite, 10 mM styrene, 10 mM diazo reagent, 5% ethanol cosolvent. Reactions were analyzed via GC-MS.

Following the initial evaluation of *PsEFE* WT's new-to-nature reaction scope, we decided to focus on aziridination of styrene with tosyl azide as our model reaction in a directed evolution campaign to develop *PsEFE* as a new-to-nature biocatalytic platform. As has been observed countless times, mutations which are found to be beneficial when evolving for one function can also provide the starting activity for other reactions of interest;<sup>42</sup> engineering *PsEFE* for this aziridination model reaction would likely enable access to additional new-to-nature reactions.

#### 4.4.1 Analogs of $\alpha$ -ketoglutarate modulate *PsEFE*'s aziridination activity

Ethylene-forming enzyme requires  $\alpha$ -ketoglutarate for both of its native functions, but the nitrene-transfer reaction should not require an equivalent of the co-substrate to go through the catalytic cycle. I therefore hypothesized that  $\alpha$ -ketoglutarate would not be strictly required and could potentially be replaced with other compounds to modulate the activity of *PsEFE*. This change would be analogous to the ability of cytochrome P411<sub>BM3</sub> to catalyze carbene- and nitrene-transfer reactions in the absence of the FAD domain (or, equivalently,



should prevent the competing native oxene transfer from occurring, even under aerobic conditions. The ability to substitute the ligand by simply changing the reaction conditions and without requiring the synthesis of a non-native metal complex cofactor enables the facile modulation of the primary coordination sphere.

#### **4.5 Evolution of enzymes with enhanced nitrene-transfer activity**

The initial activity found for the *PsEFE*-catalyzed aziridination of styrene provided a starting point for evolving a non-heme metalloprotein into a platform for non-native activity; the next step was enhancing this activity through directed evolution.

When designing our protein engineering strategy, I considered our expected screening throughput (maximum number of samples screened per day) and what factors might affect *PsEFE*'s aziridination activity. As the primary screen would be via LC-MS, the screen would be approximately 3–4 minutes per sample (approximately 400 clones per day at maximum). With this medium-throughput screening capability, I opted for site-saturation mutagenesis at positions which might reasonably modulate enzymatic activity. As the nitrene-transfer reaction putatively proceeds within the arginine substrate binding pocket, I primarily focused on mutations within that region of the protein. This scaffold had not previously been engineered for nitrene-transfer activity and the effects of mutations throughout the active site were unknown. I therefore ordered primers for 24 site-saturation libraries targeting residues in the first shell of the arginine and  $\alpha$ -ketoglutarate binding sites (amino acid positions 84, 86, 87, 91, 171, 173, 175, 186, 189, 191, 192, 198, 206, 228, 268, 270, 277, 279, 281, 283, 314, 316, 317, 318, highlighted in Figure 4-13). Out of these 24 libraries, 15 were successfully cloned, expressed, and screened on the first pass (84, 86, 87, 91, 171, 173, 186, 189, 191, 192, 277, 283, 314, 316, 317, highlighted in Figure 4-14). Libraries which were not generated on the first pass were revisited in future rounds of mutagenesis.

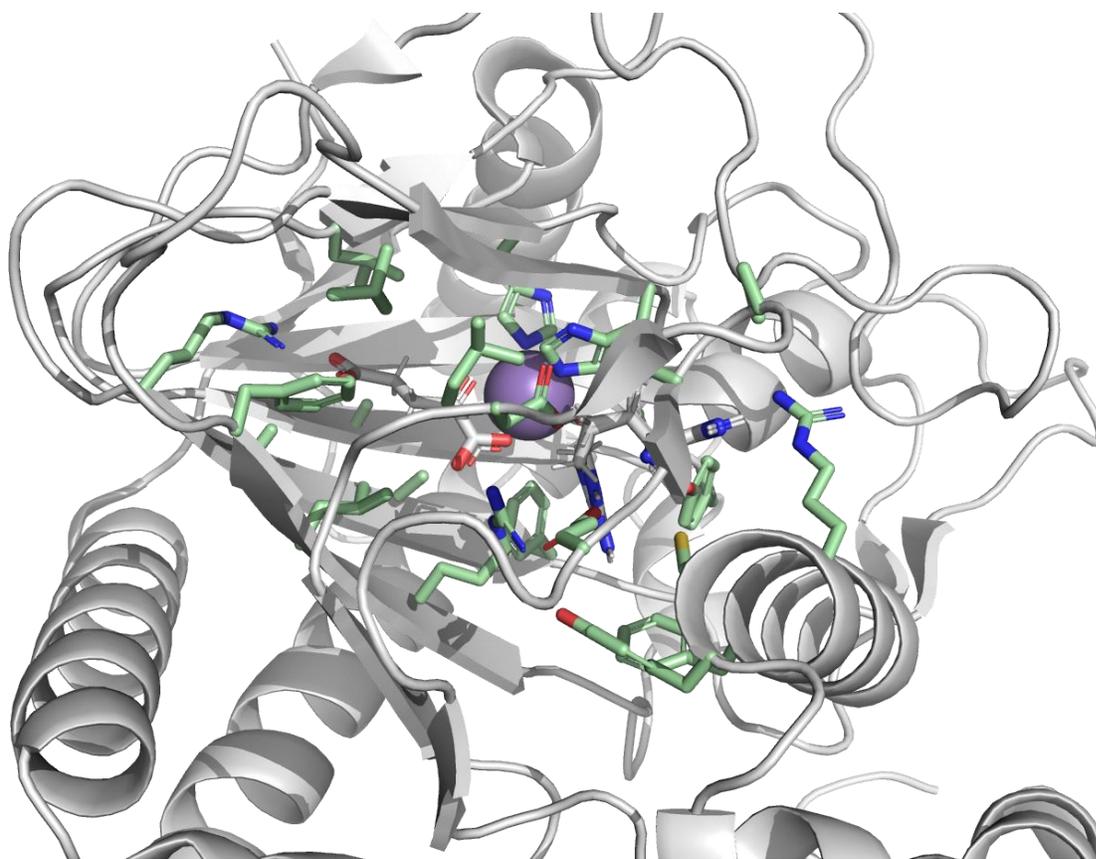


Figure 4-13. Amino-acid residues targeted for site-saturation mutagenesis in the first round of *PsEFE* engineering. Structural representation of *PsEFE* complexed with  $\alpha$ -ketoglutarate (represented in sticks) and Mn (the metal with which the protein was crystallized, purple sphere) (PDB ID: 5VKB). 24 positions are displayed for which primers for site-saturation mutagenesis were designed and ordered; side chains are represented as sticks and shown in pale green.

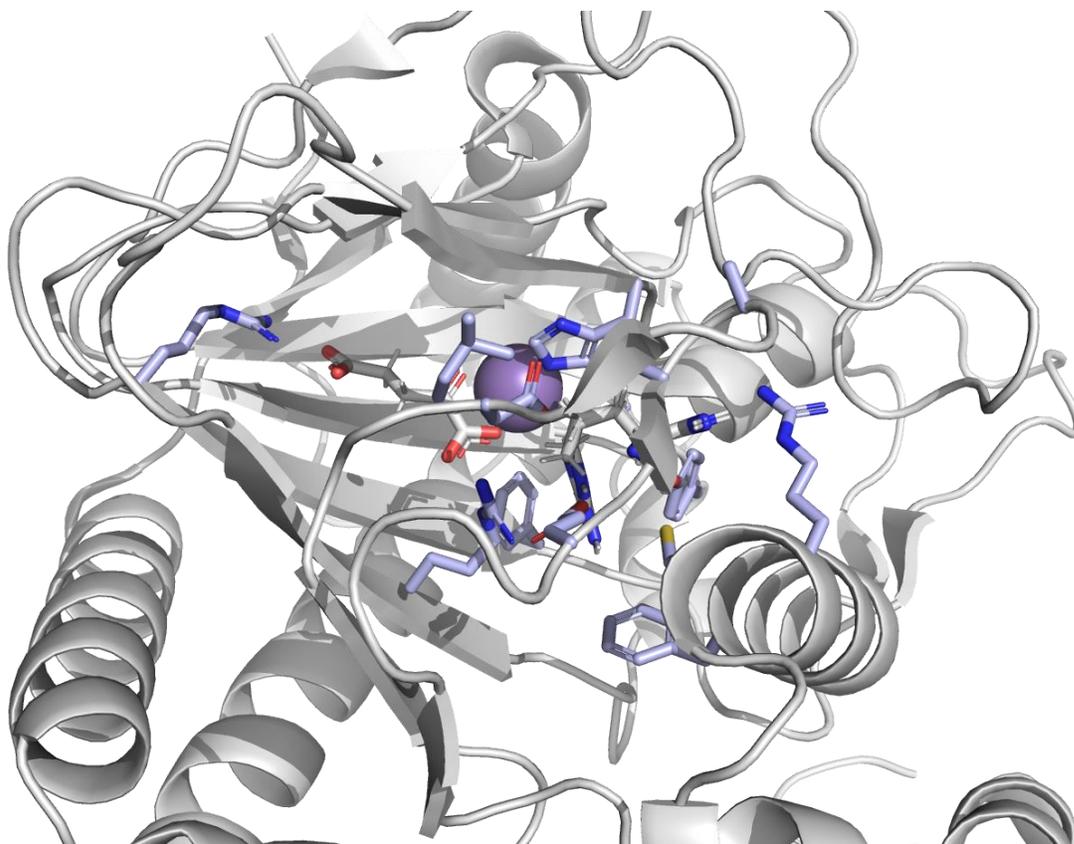


Figure 4-14. Amino-acid residues targeted for site-saturation mutagenesis and screened via LC-MS in the first round of *PsEFE* engineering. Structural representation of *PsEFE* complexed with  $\alpha$ -ketoglutarate (represented in sticks) and Mn (the metal with which the protein was crystallized, purple sphere) (PDB ID: 5VKB). 15 positions are displayed for which site-saturation mutagenesis libraries were cloned, expressed, and screened for aziridination activity; side chains are represented as sticks and shown in light blue.

Site-saturation mutagenesis libraries were screened at a rate of 80 clones per library, as well as eight clones of the parent and eight sterile controls. The screening was performed on the LC-MS using the mass selective detector (MSD) signal at  $m/z = +274$  ( $[M+H]^+$ ) to quantify aziridine product formation. The high number of parent and sterile controls and screening depth is due to the high variability observed when testing the *PsEFE* WT activity; eight parent and eight sterile controls spread across the plate enabled us to monitor the variability within each plate.

In parallel to screening the libraries from round 1, I had determined that purified wild-type *PsEFE* has enhanced activity when adding acetate in place of  $\alpha$ -ketoglutarate and that the reaction also works under aerobic conditions. Rather than modifying the screening conditions partway through a single round of directed evolution, the remaining libraries were screened under anaerobic conditions with  $\alpha$ -ketoglutarate as the supplemented ligand. Variants with enhanced aziridination activity relative to the parent protein (*PsEFE* WT) were then sequenced and expressed at 40-mL scale. These validation reactions were performed with clarified cell lysate, testing the variants under anaerobic and aerobic conditions and supplementing with either acetate or  $\alpha$ KG. While there were variants with modest enhancement from other site-saturation libraries, *PsEFE* C317M using acetate as the ligand was by far the variant with the highest enhancement in aziridination activity, including under acetate-supplemented conditions (Figure 4-15). We therefore fixed *PsEFE* C317M as the parent for the second round of site-saturation mutagenesis.

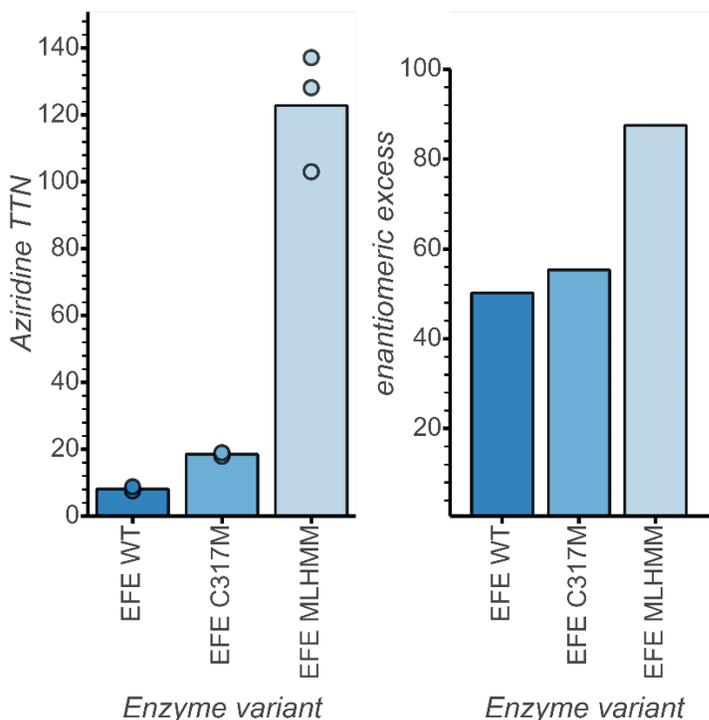


Figure 4-15. Aziridination activity and selectivity of *PsEFE* wild type and engineered variants.

In the second round of evolution, we targeted 10 positions for site-saturation mutagenesis (84, 86, 91, 171, 186, 192, 277, 283, 314, 316, highlighted in Figure 4-16). The amino-acid positions chosen for mutagenesis were ones which either failed in cloning or expression during round 1 or showed some variants with modest improvements in activity during round 1 screening. Most of these positions could be targeted using the primers purchased in the first round of mutagenesis. For positions 314 and 316, new primers were purchased to include the C317M mutation (otherwise the mutation would be overwritten in PCR amplification). These libraries were screened under aerobic conditions and supplemented with acetate in place of  $\alpha$ -ketoglutarate. Of these libraries, the largest enhancements in aziridination activity in the screening data were found at the libraries targeting positions 171, 277, and 314. In sequencing the variants with beneficial mutations, we also found a non-programmed single nucleotide substitution (C290T) which introduced the (serendipitously beneficial) amino-acid substitution T97M.

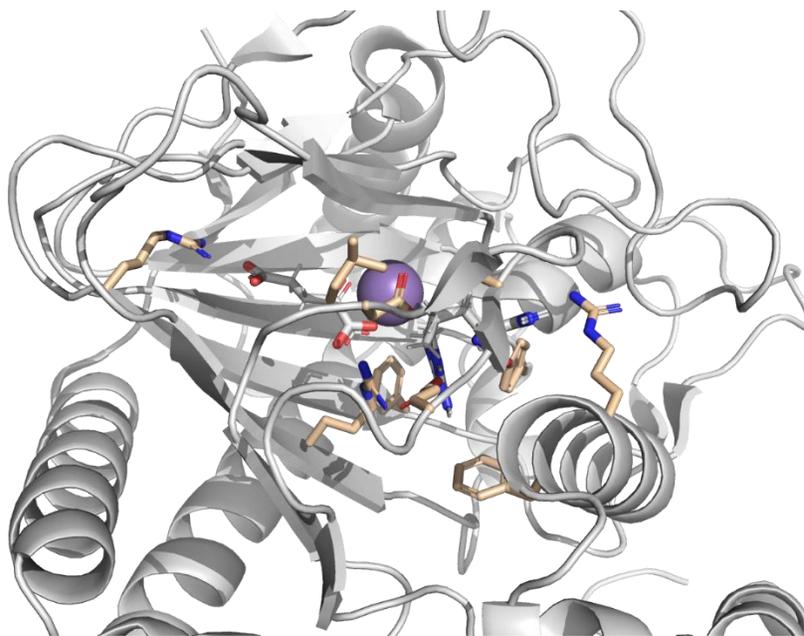


Figure 4-16. Amino-acid residues targeted in the second round of mutagenesis. Ten residues for which site-saturation mutagenesis libraries were cloned, expressed, and screened for aziridination activity; side chains are represented as sticks and shown in tan.

As the second round of evolution had multiple modest improvements (1.5–2-fold activity enhancements) rather than one variant which was clearly the best, we prepared a recombination library of the mutations found in the best-performing variants in the round 2 validation reactions. In preparing a recombination library, we introduced the mutations via PCR similar to the site-saturation mutagenesis cloning strategies. The mutations included in the recombination library were T97M, R171A, R171V, R171L, R277H, F314M, F314Q, F314L; at each position, the parent residue was also included in the library, resulting in 64 possible variants.

As the four amino-acid residues were spread out across the gene, we were able to generate three PCR fragments which included the programmed mutations at positions 97, 171, and 314 (Figure 4-17). To introduce the mutation at position 277, we used an equimolar mixture of plasmids *PsEFE* C317M and *PsEFE* R277H C317M as the template for the PCR. These three PCR products were gel extracted and assembled via Gibson assembly.

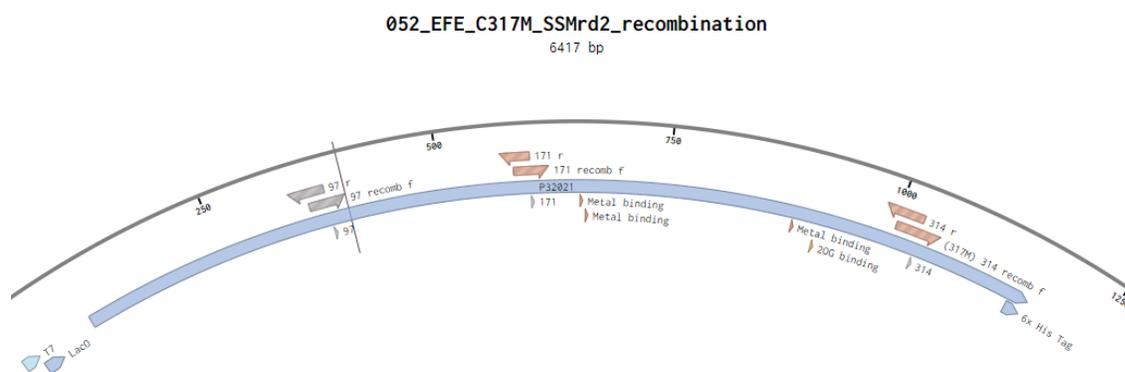


Figure 4-17. *PsEFE* recombination library primer design. The promoter (T7lac), gene (P32021, *PsEFE*), and primer binding sites are annotated. The plasmid map was generated via Benchling software (Benchling, San Francisco, CA).

To get high coverage and redundancy of this recombination library, four 96-well plates of this recombination library were screened (320 clones). This resulted in *PsEFE* T97M R171L R277H F314M C317M (*PsEFE* MLHMM), a protein variant which catalyzes the formation

of the aziridine product with 120 TTN and 88% ee. The TTN is modest compared to values reported for cytochromes P411<sub>BM3</sub>, but the high expression of *PsEFE* WT and variants enables the facile preparation of large quantities of enzyme (Supplementary Information, Figure 4-32).

A third round of site-saturation mutagenesis was carried out targeting eight additional positions (84, 95, 173, 175, 186, 192, 270, and 283), which resulted in the variant *PsEFE* MLLHMM with comparable activity in lysate and 90% ee, a slight enhancement over *PsEFE* MLHMM. During the validation of the variant's activity as purified protein, however, the enhanced enantioselectivity was maintained, but the TTN decreased two-fold relative to *PsEFE* MLHMM. The increased enantioselectivity was insufficient to justify the loss in activity. Further evolution could be carried out on *PsEFE* MLHMM if there was an application for this specific aziridine production. However, as the goal here was the development of a biocatalytic platform using non-heme metalloproteins, *PsEFE* MLHMM was chosen as the final aziridination variant in this evolution campaign.

The second round of site-saturation mutagenesis and subsequent recombination were both performed with acetate supplemented in place of  $\alpha$ KG. One mutation identified, R277H, is at the amino-acid residue which coordinates the distal carboxylate of  $\alpha$ KG, and therefore potentially disrupted the native ligand binding (Figure 4-18). Indeed, aziridination reactions prepared with *PsEFE* MLHMM and  $\alpha$ KG or NOG had the same activity as reactions with no ligand added, while acetate enhanced the activity by an order of magnitude (Figure 4-19). This was a strong indication that the *PsEFE* MLHMM aziridination catalyst no longer productively binds the native ligand,  $\alpha$ KG. The *PsEFE* variants catalyzing nitrene transfer are no longer  $\alpha$ KG-dependent enzymes, but *PsEFE* MLHMM, along with other variants with mutations at R277, appear to no longer be  $\alpha$ KG-binding enzymes.

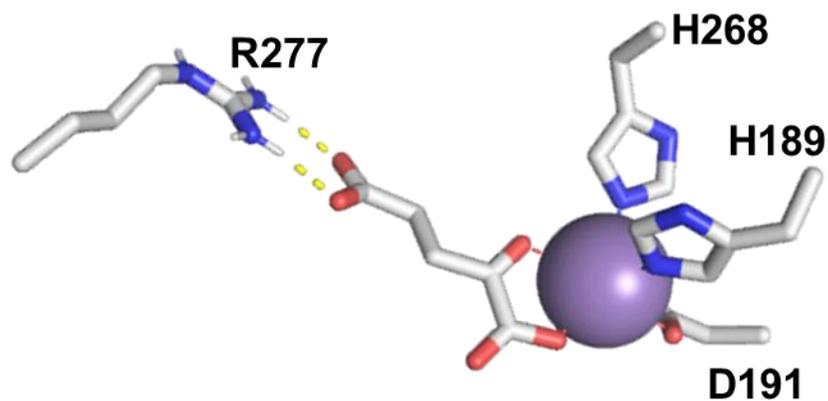


Figure 4-18. Position of R277 relative to  $\alpha$ KG. R277 interacts with the distal carboxylate of  $\alpha$ KG and  $\alpha$ KG analogs in reported crystal structures.

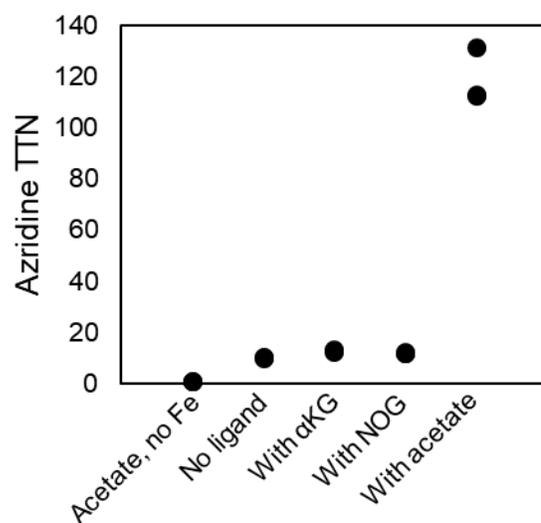


Figure 4-19. Aziridination activity with *PsEFE* MLHMM under different reaction conditions.

#### 4.5.1 Exploring engineered PsEFE sequence for other nitrene-transfer reactions

The recombination plates generated using beneficial mutations in the second round of site-saturation mutagenesis libraries represented a large amount of active-site residue sequence diversity which had been shown to be beneficial for nitrene-transfer reactions. As the active site has been remodeled since the initial reactions testing alternative nitrene-transfer reactions, we hypothesized that this recombination library might contain protein variants which could perform alternative nitrene-transfer reactions. In this screen, activity on the intramolecular C–H insertion of 2-ethylbenzenesulfonyl azide was found (Figure 4-20), with the best variant, *PsEFE* R171V F314M C317M (*PsEFE* VMM) catalyzing the reaction with 310 TTN, but only 9.4% ee with the screening ligand (acetate). Its activity and selectivity were further increased by using NOG as a ligand, with up to 730 TTN and 47% ee. Not only does the activity and stereoselectivity vary drastically through ligand substitution, but also the chemoselectivity: *PsEFE* without a ligand slightly favors formation of the sulfonamide over the sultam, while the addition of NOG increases the ratio of sultam:sulfonamide to up to 105:1, with only trace reduction product detectable (Table 4-3).

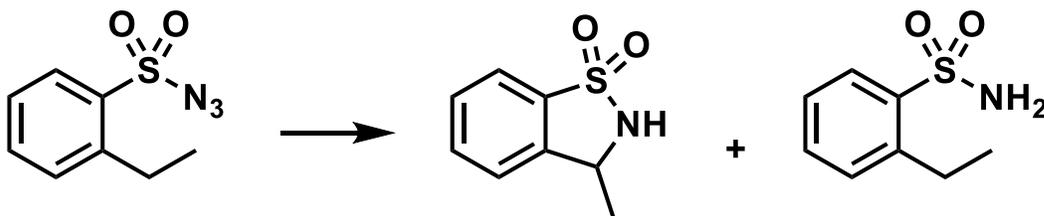


Figure 4-20. Intramolecular C–H insertion reaction. 2-ethylbenzenesulfonyl azide can be converted to the sultam via C–H insertion or reduced to the 2-ethylbenzenesulfonamide byproduct.

Table 4-3. Intramolecular C–H insertion activity with ligand substitutions in *PsEFE* VMM. Reaction conditions: 20  $\mu$ M purified *PsEFE* VMM, 10 mM 2-ethylbenzenesulfonyl azide, anaerobic, room temperature, 6 hours. <sup>1</sup>10  $\mu$ M enzyme concentration.

<i>PsEFE</i> variant	Ligand	TTN	ee (%)	Insertion / reduction
VMM	None	25	n.d.	0.9
VMM	$\alpha$ KG	130	61	9.0
VMM	Acetate	310	9.4	24
VMM	NOG	450	48	105
VMM <sup>1</sup>	NOG	730	47	100

A second variant in the recombination plate which was identified to have C–H insertion activity was *PsEFE* VHMM, with the only difference being the R277H mutation. As the aziridination variant *PsEFE* MLHMM contained this R77H mutation and had no activity enhancement from  $\alpha$ KG nor NOG, I hypothesized that *PsEFE* VHMM would only be activated by acetate. Indeed, while *PsEFE* VMM exhibits enhanced activity with  $\alpha$ KG and NOG, only acetate enhances *PsEFE* VHMM's C–H insertion activity (Figure 4-21). The best variant for this reaction is still *PsEFE* VMM with NOG as a ligand, but the ability to modulate ligand binding through mutations at R277, including preventing the native ligand from productively binding, could be useful for further *PsEFE* protein engineering.

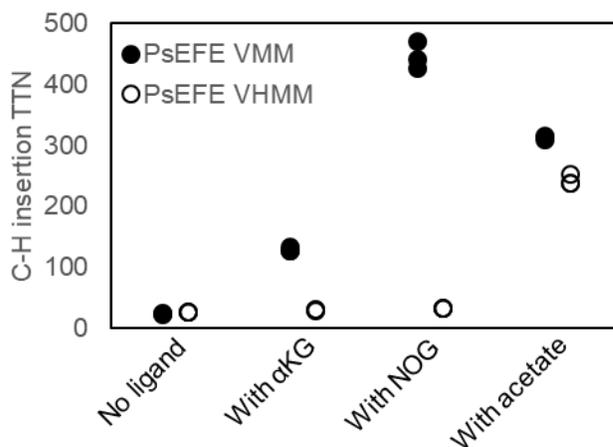


Figure 4-21. Intramolecular C–H insertion with variants *PsEFE* VMM and *PsEFE* VHMM. *PsEFE* VMM is enhanced by  $\alpha$ KG, NOG, and acetate, while *PsEFE* VHMM is only enhanced by acetate. Reaction conditions: 20  $\mu$ M purified *PsEFE* variant, 10 mM 2-ethylbenzenesulfonyl azide, anaerobic, room temperature, 6 hours.

#### 4.5.2 Analysis of evolved variants' stability and lifetime

A time course of the aziridination reaction was run with *PsEFE* WT and *PsEFE* MLLHMM to better understand the reaction kinetics and enzyme stability. Reactions containing purified protein were set up in triplicate both anaerobically and aerobically. Time points were taken at 15 minutes, 30 minutes, 1 hour, 2 hours, 3 hours, 4 hours, and 8 hours, at which point those reactions were quenched by addition of acetonitrile (350  $\mu\text{L}$ ) and the internal standard propiophenone (1  $\mu\text{L mL}^{-1}$  in acetonitrile, 50  $\mu\text{L}$ ).

As seen in Figure 4-22, the reaction with wild-type *PsEFE* is almost done by the 15-minute mark, while the yield with MLLHMM continues to increase for approximately 2 hours. The reaction appears to proceed for a longer time anaerobically than aerobically for *PsEFE* MLLHMM, but the low activity for the aerobic reaction makes such comparisons challenging.

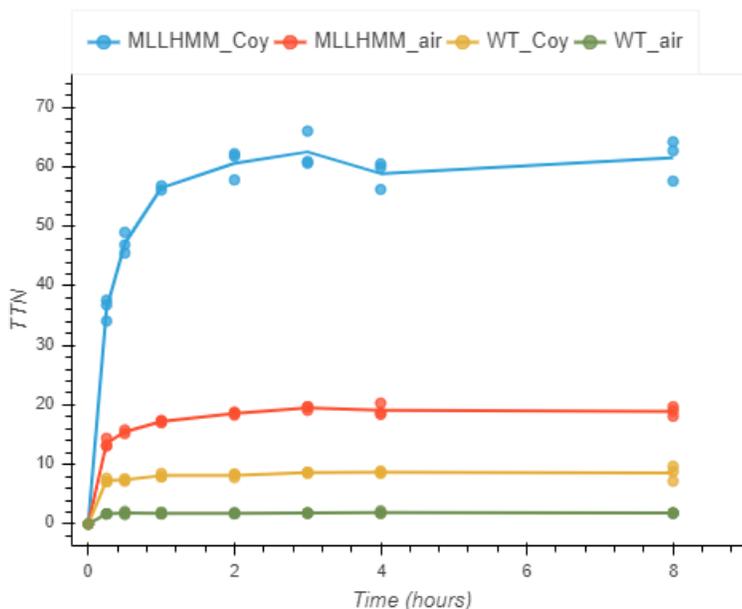


Figure 4-22. Aziridination time course. Time courses denoted “air” were set up aerobically; time courses denoted “Coy” were set up anaerobically.

This rapid enzyme inactivation could be indicative of enzyme instability. Indeed, in purified protein reactions, what is ostensibly protein precipitation is observed to increase over time. The thermostability of *PsEFE* WT has been measured with ITC;<sup>45</sup> not surprisingly, the protein is reported to have increased stability in the presence of iron and  $\alpha$ -ketoglutarate. I used a thermal shift assay<sup>46</sup> using SYPRO orange (Thermo Fisher Scientific) to measure the effect iron, acetate, and  $\alpha$ KG have on the stability of each protein variant. It is clear from the data that, even though beneficial mutations were only chosen based on activity and stereoselectivity, the protein's stability improved from wild type to the final variants. There is also a significant enhancement in thermostability for early variants upon addition of  $\alpha$ KG, which is not observed in the later evolved variants. This indicates that the mutations introduced, despite being within the active site, enhanced *PsEFE*'s thermostability, and that screening for higher TTN also likely exerted pressure for mutations to stabilize the protein.

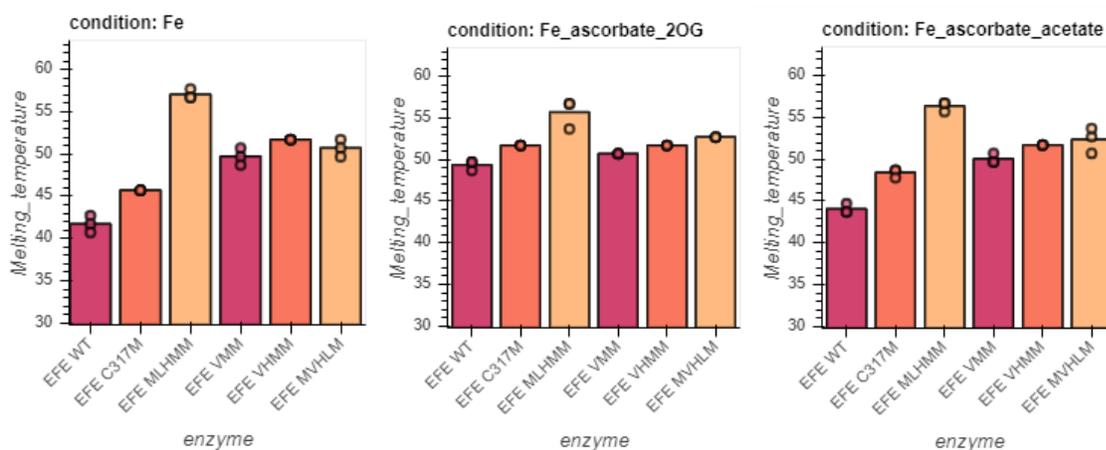


Figure 4-23. Thermostability of wild-type and evolved *PsEFE* variants for aziridination and intramolecular C-H insertion. The thermostability of the proteins was assayed anaerobically in the presence of Fe, in the presence of Fe, ascorbate, and  $\alpha$ KG (noted as 2OG for 2-oxoglutarate), and in the presence of Fe, ascorbate, and acetate.

#### 4.6 Structural characterization of *PsEFE* variants

To understand the effect of the mutations and non-native ligands in the *PsEFE* protein variants, I turned to protein crystallography. The structure of *PsEFE* WT with L-Arg,  $\alpha$ KG,

and analogs of both molecules have been reported by the Schofield and Hausinger groups using manganese in place of iron to allow for aerobic crystallography.<sup>47,48</sup> In these structures, *PsEFE* is seen to adopt an open conformation with a highly solvent-accessible active site in the absence of its substrate, L-Arg. The lid loop closes the active site with multiple polar contacts to L-Arg, reducing active-site volume and orienting the substrate. The substrates used for aziridination and C–H insertion lack these polar contacts, and as such, it is unclear whether the lid-loop region would close. The solvent-exposed active site could explain why *PsEFE* was found to have initial activity.

I cloned N-terminally polyhistidine-SUMO-tagged constructs of *PsEFE* VMM and *PsEFE* VHMM to investigate the effects these mutations have on *PsEFE*'s structure, using the same cloning strategy used in the generation of heme protein crystallization constructs in Chapter 2; crystallization constructs with N-terminally polyhistidine-SUMO tags were also used by Schofield and coworkers to purify *PsEFE* WT.<sup>48</sup> I expressed and purified both proteins as tagless constructs (Supplementary Information, Figure 4-30). Sparse-matrix screening of the purified protein supplemented with manganese resulted in initial crystallization conditions in multiple PEG 3350 and PEG 4000 conditions, similar to those reported for the wild-type protein.<sup>48</sup> Drop-volume ratio screening resulted in large clusters of crystals in 1-2 days in several ratios of protein and well solution (PEG 4000, 0.1 M Tris pH 8.0) (Figure 4-24). While *PsEFE* VHMM did not initially crystallize under these conditions, streak-seeding from *PsEFE* VMM crystals using a cat whisker (a gift from Crick Boville) resulted in *PsEFE* VHMM crystals. Short (15 minute) soaking experiments as well as overnight soaking experiments were carried out, adding  $\alpha$ KG, *N*-oxalylglycine, and acetate to *PsEFE* VMM crystals and acetate to *PsEFE* VHMM crystals.



Figure 4-24. *PsEFE* VMM crystals. Using polarized film, the crystal's birefringence is clearly seen, indicating it is proteinogenic (right).

Crystals were cryoprotected with ethylene glycol and shipped to SSRL 12-2 for X-ray diffraction studies. The *PsEFE* crystals consistently diffracted to 1.5–2.0 Å resolution, and data sets were collected for each soaking experiment. Both *PsEFE* VMM and *PsEFE* VHMM crystallized in the open conformation, with a highly solvent-exposed active site (Figure 4-25a). There are no apparent tertiary structure changes near the mutations relative to a *PsEFE* WT structure (PDB ID 5V2Y, 0.44 Å RMSD over 2039 atoms, Figure 4-25b). Looking at the side chains, however, one can begin to rationalize potential reasons why R171V, F314M, and C317M are beneficial mutations. Both R171V and F314M increase the active-site volume; it is reasonable that the intermolecular reaction of two aromatic substrates would require more active-site space (Figure 4-25c). The C317M mutation extends the thioether methyl group 1.4 Å closer (relative to the C317 thiol in PDB 5V2Y) to the guanidino group of L-Arg bound in *PsEFE* WT (Figure 4-25d); the presence of this thioether could disfavor L-Arg binding. Disfavoring the native substrate binding might be key in this case, as directed evolution screening was carried out in freeze-thawed whole cells, *PsEFE* WT exhibits a  $K_M$  of 37 μM for L-Arg, and *E. coli* intracellular L-Arg concentration is 570 μM.<sup>49</sup>

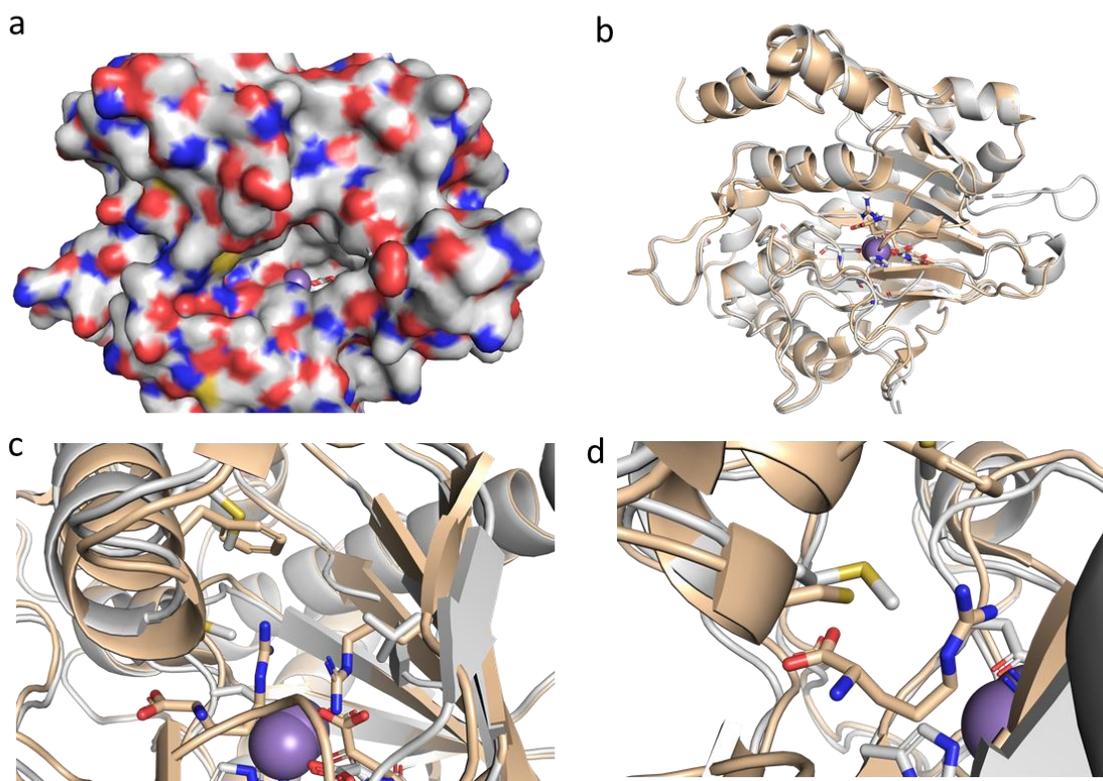


Figure 4-25. Structural comparison of Mn- and NOG-bound *PsEFE* VMM and Mn- and  $\alpha$ KG-bound *PsEFE* WT (PDB ID 5V2Y). (a) Surface representation of *PsEFE* VMM, showing the solvent-exposed active site; the manganese ion is shown in purple. (b) Cartoon representation of aligned *PsEFE* WT (tan) and *PsEFE* VMM (gray). (c) Mutations at positions R171 and F314 to smaller amino-acid side chains increase the active site's volume. (d) The C317M mutation in *PsEFE* VMM extends a thioether moiety toward the guanidino group of L-Arg bound in *PsEFE* WT.

Using ethylene glycol as cryoprotectant with no ligand soak or short ligand soaking experiments resulted in bidentate metal binding by ethylene glycol. This led to testing additional hydroxy acids as reaction additives, including glycolic acid, which further enhanced activity relative to acetate in *PsEFE* VHMM (Supplementary Information, Figure 4-31). In the data set collected on a *PsEFE* crystal soaked overnight with *N*-oxalylglycine, however, there is density in agreement with the NOG ligand binding in the active site. Interestingly, the electron density supports two conformations of NOG (Figure 4-26). A similar set of multiple conformations is also observed in a structure of  $\alpha$ KG-bound *PsEFE*

WT,<sup>47</sup> but while the second  $\alpha$ KG conformation would not be biochemically productive, it is possible that NOG bound in either configuration would modulate nitrene-transfer activity.

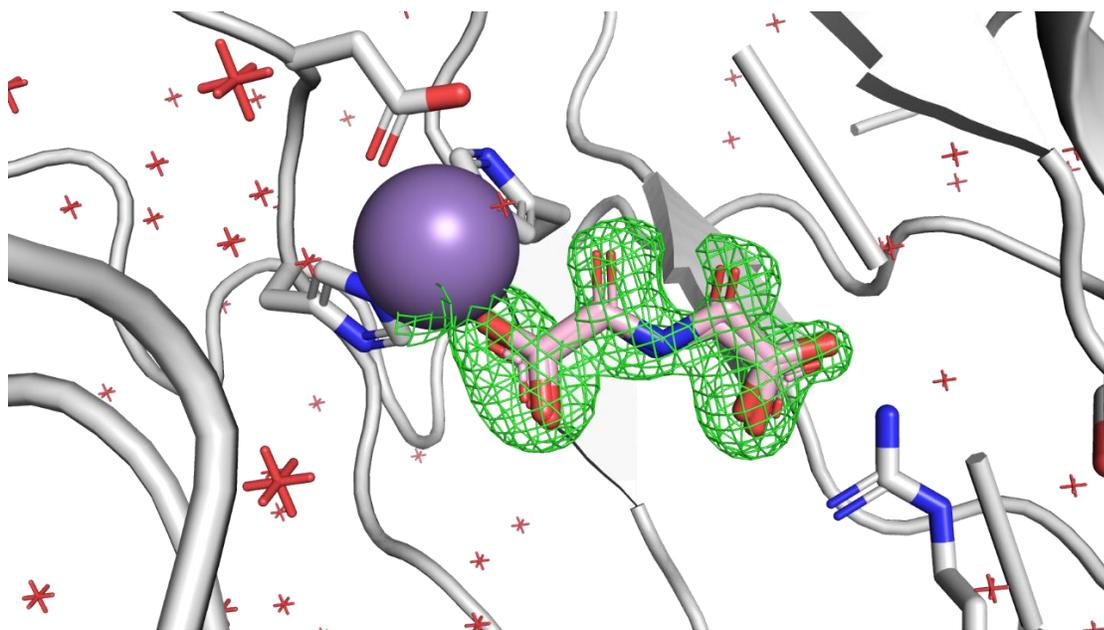


Figure 4-26. Electron density of NOG-bound *PsEFE* VMM matches NOG in two conformations. Electron density map of the ligand was generated via polder.

Given the ability to obtain crystal structures of the Mn-NOG-bound *PsEFE* VMM, I was interested in obtaining a structure with molecules relevant to nitrene-transfer reactions. The open active site and multiple potential coordination sites on the active-site metal increase potential binding modes for each substrate, and insight on these binding modes would be beneficial for future site-saturation mutagenesis library design. Due to the large conformational changes associated with L-Arg binding, I was concerned that performing soaking experiments might result in a structure which is not biochemically relevant and opted for co-crystallization of *PsEFE* VMM with reaction substrates or products instead. I chose the intramolecular C–H insertion substrate and sultam product as the C–H insertion reaction had higher activity than the aziridination reaction, and the intramolecular reaction only required binding a single molecule in the active site for either starting material or product. I

set up co-crystallization experiments, in which *PsEFE* VMM, NOG, manganese chloride, and either the sulfonyl azide or sultam were incubated with the protein, followed by centrifugal filtration to remove any precipitated protein. The purified protein solution was used to set up drop-volume ratio crystallization refinement screens with streak-seeding, and crystals grew with the same clustered needle morphology over the same time frame. These crystals were cryoprotected with PEG 400 in place of ethylene glycol in the attempt to prevent ethylene glycol binding in the active site. While ethylene glycol was no longer found in the active site, there was also no electron density supporting the binding of either the sulfonyl azide or sultam product. The lid loop also remained open in the structures collected from crystals in the co-crystallization attempts. It is possible that the crystallization conditions prevent the lid-loop region from closing, though similar conditions are used by Schofield to crystallize *PsEFE* WT. It is also possible that the substrate or product binding do not induce the protein to adopt the closed conformation, or that the co-crystallization conditions were not conducive to substrate or product binding.

To investigate whether the lid loop still closes, nitrene-transfer reactions using *PsEFE* variants could be performed in the presence of different amounts of L-Arg. This would provide insight into whether L-Arg binding has been impacted in the evolved variants. If L-Arg inhibits *PsEFE* nitrene-transfer activity, co-crystallization of *PsEFE* VMM with L-Arg would show whether the lid closing mechanism has been disrupted and whether L-Arg's binding mode has been affected by the mutants as rationalized above.

#### **4.7 Adaptation of *PsEFE* for high-throughput screening**

*PsEFE* is a potential platform for biocatalytic new-to-nature reactions due to its high level of expression, demonstrated ability to perform multiple nitrene-transfer biotransformations, and the ability to modulate the activity of *PsEFE* variants through ligand substitution. Two of the most important features of useful biocatalysts are their expression and stability; poor expression increases the cost of producing the catalyst, while poor stability makes both engineering and routine use of the biocatalyst more challenging. *PsEFE* already expresses well, but we have seen that its stability could be further improved (Figure 4-23). It is also

clear that expanding the diversity of protein variants with activity for new-to-nature reactions has led to the discovery of starting points for new reactions for which the wild-type protein exhibited no activity.<sup>42</sup>

Looking to new *PsEFE* library generation, I noted that many active-site residues have already been targeted for mutagenesis during the evolution of the aziridination enzyme variant *PsEFE* MLHMM, but only 7% of the sites within *PsEFE* have been targeted. Expanding the sequence diversity targeted in future rounds of evolution will improve the probability of finding beneficial mutations. Using the tile-based mutagenesis strategy laid out in Chapter 5, I designed and ordered a mutagenesis library consisting of a total of 127 amino-acid sites targeted over three tiles in *PsEFE* VMM (Figure 4-27). The tiles include almost all active-site residues (excluding residues whose side chains comprise the metal's primary coordination sphere), the flexible lid-loop region, and the C-terminus, which is highly flexible based on crystallographic B-factors and could contribute to the protein's instability. The goal of this library is to expand the sequence breadth targeted in the next rounds of *PsEFE* evolution on new-to-nature reactions and the identification of residues which stabilize or generally activate the *PsEFE* scaffold. The tiles were also chosen such that three additional tiles covering complementary sequence space could be ordered and used to generate libraries targeting most of the remaining sequence space.

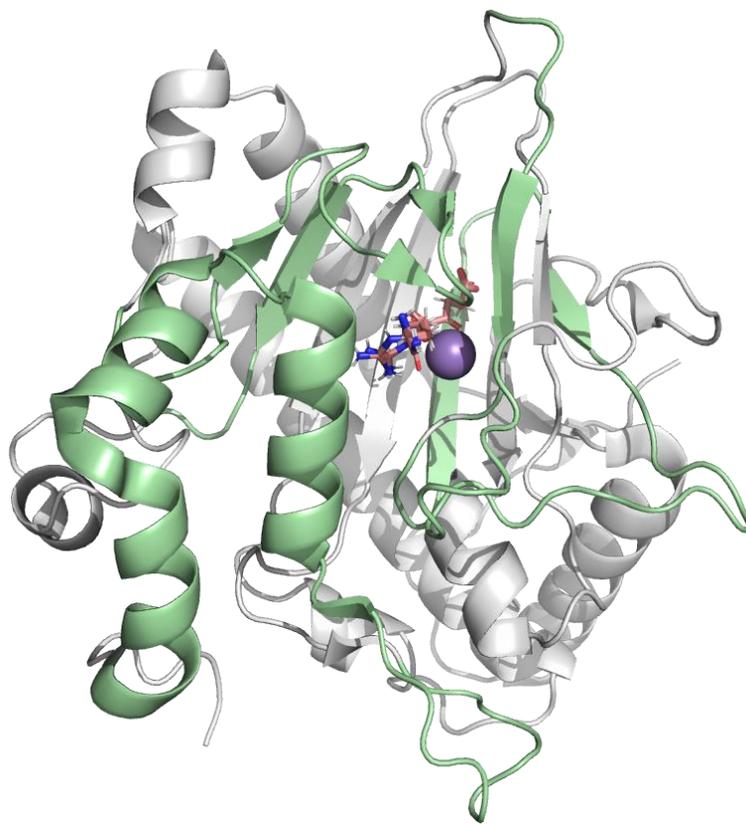


Figure 4-27. *PsEFE* VMM 127-site-saturation library. Green regions show amino-acid residues targeted for site-saturation mutagenesis in the Twist library for *PsEFE* VMM.

With any large library design, it is important to have a screening method to effectively search the library's sequence space for the desired fitness enhancements. As protein stability is a potential issue for future *PsEFE* engineering, developing a rapid screen for protein expression and stability would enable the enrichment of stable *PsEFE* variants, which could then be screened for new-to-nature activity. Unlike heme proteins, which can be quantified in lysate via carbon monoxide binding assays,<sup>50</sup> spectrophotometric assays developed for  $\alpha$ KG-dependent enzymes typically rely on succinate formation. As my initial goal was to assay protein expression and stability, it was only necessary to determine whether the full-length protein is present and soluble. To this end, I employed a split-fluorescent protein system, in which the protein of interest is tagged with a short (16 amino acid) fragment of a fluorescent protein. When the tagged protein of interest and the complementary fluorescent

protein fragment are co-expressed, the two fluorescent protein fragments self-assemble, and the protein's fluorescence is enhanced. This approach was developed and used to assay soluble protein expression<sup>51</sup> and has more recently been used to normalize enzyme concentrations in biocatalytic reactions to enable separately analyzing protein expression and specific activity.<sup>52</sup>

mNeonGreen2(1-10) (mNG2(1-10), complemented by mNG2(11)) was chosen for the fluorescent protein system as it was engineered for decreased background fluorescence in the absence of the complementary 11<sup>th</sup>  $\beta$ -strand, and was readily available through Addgene.<sup>53</sup> The construct from Addgene expressed mNG2(1-10) under IPTG induction control; to allow for orthogonal induction control, I subcloned the mNG2(1-10) encoding gene into the pBAD33 vector under arabinose induction control. Separately, pET22b constructs containing IPTG-inducible *PsEFE* VMM with and without C-terminal polyhistidine tags were C-terminally tagged with mNG2(11). The orthogonal induction allows for the independent expression optimization of both proteins. In cases when the fluorescence output is not needed, the cellular metabolic load can be reduced by only expressing the protein of interest. The arabinose induction is tightly regulated and further repressed under high glucose concentrations, leading to minimal expression of mNG2(1-10) without addition of arabinose.

In cases when the stability of the fluorescent protein is greater than that of a protein to which it is fused, the stability of the protein of interest can be assayed by a heat challenge followed by centrifugation. The denaturation and aggregation of the protein of interest pulls the soluble fluorescent protein into the pellet.<sup>54</sup> This approach was feasible with our fluorescent system as mNG2(1-10) linked to mNG2(11) is stable at 60 °C for over 1 hour. Lysate from cells co-expressing *PsEFE* VMM-His<sub>6</sub>-mNG2(11) showed fluorescence, but following incubation in a water bath at 55 °C for one hour and centrifugation, the supernate fluorescence was depleted (Figure 4-28). The fluorescent protein was still folded and stable, as the pellet was brightly fluorescent.

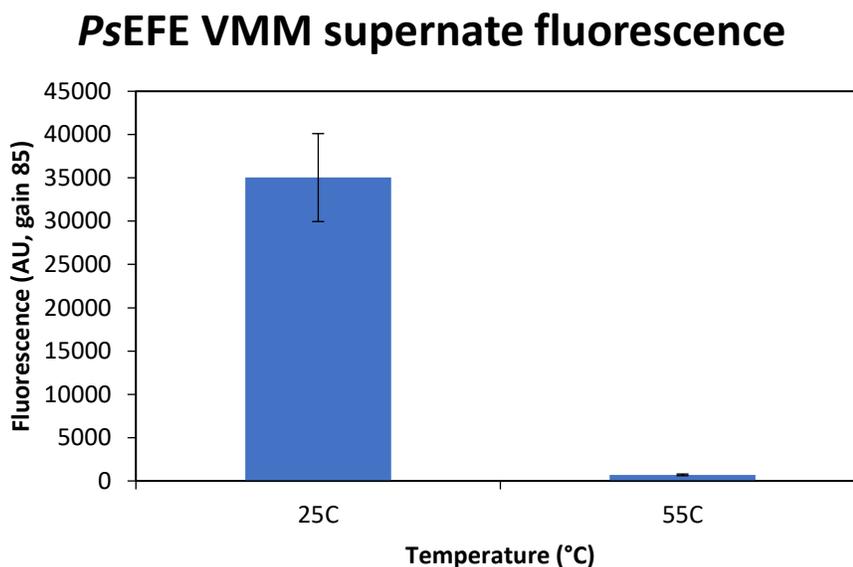


Figure 4-28. Split-fluorescent protein labeling enables measurement of *PsEFE* folding and aggregation. Fluorescence intensity is proportional to the protein of interest in the solution; as *PsEFE* VMM is denatured and aggregates below 55 °C, the associated split-fluorescent protein pellets out with the aggregated *PsEFE* VMM to reduce supernate fluorescence.

This fluorescent stability assay is amenable to laboratory automation. With automated colony picking and 96-tip head liquid handlers, each liquid transfer can be performed robotically, increasing throughput potential. With the addition of single-tip liquid handling, variants with improved stability by this screen can be robotically re-arrayed into 96-well plates and assayed for an activity of interest.

## 4.8 Conclusions

A collection of non-heme metalloproteins was developed with the goal of expanding the scope of new-to-nature reactions. These non-heme metalloproteins were tested for initial activity in the hydrosilylation of alkenes, alkynes, and carbonyls, but no enzymatic activity was found during screening. The non-heme iron proteins were also tested for carbene- and nitrene-transfer activity. *PsEFE*, an  $\alpha$ KG-dependent enzyme, was found to have aziridination activity. This activity was enhanced by site-saturation mutagenesis of active-site residues, and mutations found for aziridination also enhanced *PsEFE*'s activity for intramolecular C–

H insertion. The aziridination and C–H insertion activity and selectivity could be enhanced by adding non-native ligands NOG or acetate in place of  $\alpha$ KG. To aid in further evolution of *PsEFE* variants, I prepared a mutagenesis library targeting 127 residues in the active site and flexible C-terminus and tested the use of a split-fluorescent protein thermostability assay. I anticipate that this *PsEFE* comprehensive active-site saturation library and the mNG2-split-fluorescent protein reporter for protein expression and stability will be helpful for additional engineering of new, stabilized *PsEFE* variants for new-to-nature reactions.

## 4.9 Supplementary information for Chapter 4.

Table 4-4. Collection of non-heme metalloproteins acquired as synthetic gene fragments from IDT DNA or as clonal constructs from other research groups. The DNA and amino-acid sequences for these proteins are deposited in CaltechDATA (DOI: 10.22002/D1.1437).

UniProt ID	Protein name	Source organism	Metal
P22364	Amicyanin	<i>Paracoccus denitrificans</i>	Cu
A9WFS1	Auracyanin D	<i>Chloroflexus aurantiacus</i>	Cu
P00282	Azurin	<i>Pseudomonas aeruginosa</i>	Cu
P34097	Azurin	<i>Pseudomonas putida</i>	Cu
P12335	Azurin iso-2	<i>Methylomonas sp.</i>	Cu
Q72HW2	Laccase	<i>Thermus thermophiles</i>	Cu
Q820S6	Nitrosocyanin	<i>Nitrosomonas europaea</i>	Cu
P00303	Plantacyanin	<i>Cucumis sativus</i>	Cu
P18068	Plastocyanin	<i>Chlamodymonas reinhardtii</i>	Cu
Q7SIB8	Plastocyanin	<i>Dryopteris crassirhizoma</i>	Cu
Q51883	Plastocyanin	<i>Phormodium laminosum</i>	Cu
P04171	Pseudoazurin	<i>Methylbacterium extorquens</i>	Cu
Q92M26	Pseudoazurin	<i>Rhizobium meliloti</i>	Cu
P0C918	Rusticyanin	<i>Alcaligenes faecalis</i>	Cu
Q9X1H0	Cupin-like protein	<i>Thermotoga maritima</i>	Mn
O93724	Superoxide dismutase	<i>Pyrobaculum aerophilum</i>	Mn
P61503	Superoxide dismutase	<i>Thermus thermophilus</i>	Mn
Q81MI9	Acireductone dioxygenase	<i>Bacillus anthracis</i>	Ni
A2VA43	Quercetinase	<i>Streptomyces sp.</i>	Ni
C7QJ42	Lysine 3-hydroxylase	<i>Catenulispora acidiphila</i>	Fe
P37610	Taurine dioxygenase	<i>Escherichia coli</i>	Fe
Q6WZB0	Arginine hydroxylase	<i>Streptomyces vinaceus</i>	Fe
A0A0E3URV8	Leucine 5-hydroxylase	<i>Streptomyces muensis</i>	Fe
Q5FQD2	Leucine 5-hydroxylase	<i>Gluconobacter oxydans</i>	Fe

Table 4-5. Collection of non-heme metalloproteins acquired as clonal constructs from Twist Biosciences. The process through which these proteins were selected is detailed in section 4.2. The DNA and amino-acid sequences for these proteins are deposited in CaltechDATA (DOI: 10.22002/D1.1437).

#	UniProt ID	Protein name	Source organism	Metal
1	Q9REI7	2,4'-Dihydroxyacetophenone dioxygenase	<i>Alcaligenes sp.</i>	Fe
2	P17109	Succinyl isochorismate synthase	<i>Escherichia coli</i>	Mn
3	P77072	Organomercurial lyase MerB D99S	<i>Escherichia coli</i>	Cu
4	P0A9S1	Lactaldehyde:1,2-propanediol oxidoreductase	<i>Escherichia coli</i>	Fe
5	P0AC81	Glyoxalase	<i>Escherichia coli</i>	Ni/Co
6	A0A067YX61	$\alpha$ -Ketoglutarate-dependent halogenase	<i>Hapalosiphon welwitschii</i>	Fe
7	Q68RJ8	Glyoxalase I	<i>Leishmania major</i>	Ni
8	Q99JT9	Acireductone dioxygenase	<i>Mus musculus</i>	Ni/Fe
9	G7CFI3	Ergothioneine-biosynthetic sulfoxide synthase	<i>Mycobacterium thermoresistibile</i>	Fe
10	A9A2G4	Purple Cupredoxin	<i>Nitrosopumilus maritimus</i>	Cu
11	Q7MZL9	plu4264 Protein, unknown function	<i>Photorhabdus luminescens</i>	Ni
12	Q70AC7	Transcarboxylase	<i>Propionibacterium freudenreichii</i>	Co
13	O50580	D-Tagatose 3-epimerase	<i>Pseudomonas cichorii</i>	Mn
14	O58810	Superoxide reductase	<i>Pyrococcus horikoshii</i>	Fe
15	Q6N272	Functionally unknown protein RPA4178	<i>Rhodopseudomonas palustris</i>	Ni
16	Q9X034	Amidohydrolase	<i>Thermotoga maritima</i>	Ni
17	Q9X113	Oxalate decarboxylase	<i>Thermotoga maritima</i>	Mn
18	Q5AR53	Fe(II)/ $\alpha$ -ketoglutarate-dependent dioxygenase	<i>Aspergillus nidulans</i>	Fe
19	B1L4V6	Hydrolase, DNA repair enzyme	<i>Korarchaeum cryptofilum</i>	Mn/Co/Cu
20	V6TJK7	Superoxide reductase	<i>Giardia intestinalis</i>	Fe
21	C7R4I0	Chitinase	<i>Jonesia denitrificans</i>	Cu
22	Q74MF3	Superoxide reductase	<i>Nanoarchaeum equitans</i>	Fe
23	P13280	Glycogenin	<i>Oryctolagus cuniculus</i>	Mn

#	UniProt ID	Protein name	Source organism	Metal
24	F8LWI3	Cambialistic superoxide dismutase	<i>Streptococcus thermophilus</i>	Fe/Mn
25	P80857	Superoxide dismutase	<i>Sulfolobus solfataricus</i>	Fe
26	Q9WYP7	L-Ketose-3-epimerase	<i>Thermotoga maritima</i>	Mn/Ni
27	A0A0M3KL01	Phytanoyl-CoA dioxygenase	<i>Micromonospora carbonacea</i>	Fe
28	A5VWI3	Persulfide dioxygenase	<i>Pseudomonas putida</i>	Fe
29	A6VKV4	Phosphoenolpyruvate carboxykinase	<i>Actinobacillus succinogenes</i>	Mn
30	B3PJ79	Lytic polysaccharide monooxygenase	<i>Cellvibrio japonicus</i>	Cu
31	C6FI44	Hydroxyquinol 1,2-dioxygenase	<i>Pseudomonas putida</i>	Fe
32	C6RPG2	Catechol 1,2 dioxygenase	<i>Acinetobacter radioresistens</i>	Fe
33	D0VX22	$\alpha$ -Ketoglutarate-dependent dioxygenase	<i>Streptomyces sp.</i>	Fe
34	D6EWM4	Lytic polysaccharide monooxygenase	<i>Streptomyces lividans</i>	Cu
35	F0QXN6	Phosphotriesterase-like lactonase	<i>Vulcanisaeta moutnovskia</i>	Co
36	O53512	3-Deoxy-D-arabino-heptulosonate 7-phosphate synthase	<i>Mycobacterium tuberculosis</i>	Mn
37	O58691	Dipeptidase	<i>Pyrococcus horikoshii</i>	Co
38	O87198	Homocitrate synthase	<i>Thermus thermophilus</i>	Co/Cu
39	P0A434	Phosphotriesterase	<i>Brevundimonas diminuta</i>	Mn
40	P05050	2-oxoglutarate iron(II) dependent dioxygenase	<i>Escherichia coli</i>	Fe
41	P29082	Sulfur oxygenase/reductase	<i>Acidianus ambivalens</i>	Fe
42	P30967	Phenylalanine hydroxylase	<i>Chromobacterium violaceum</i>	Fe
43	P32021	Ethylene-forming enzyme	<i>Pseudomonas savastanoi</i>	Fe
44	P42106	Quercetin 2,3-dioxygenase	<i>Bacillus subtilis</i>	Fe
45	P53608	Arginase	<i>[Bacillus] caldovelox</i>	Mn
46	P70080	Tryptophan hydroxylase	<i>Gallus gallus</i>	Fe
47	P74334	Apocarotenoid cleavage oxygenase	<i>Synechocystis sp. PCC 6803</i>	Fe
48	Q1D4C9	Metallo- $\beta$ -lactamase	<i>Myxococcus xanthus</i>	Fe

#	UniProt ID	Protein name	Source organism	Metal
49	Q6REQ5	Catechol dioxygenase	<i>Rhodococcus sp. DK17</i>	Fe
50	Q7WYF5	2,3-Dioxygenase LapB	<i>Pseudomonas alkylphenolica</i>	Fe
51	Q9RHA2	Fructose-1,6-bisphosphate aldolase	<i>Thermus aquaticus</i>	Co
52	Q9WZS1	N-Acetylglucosamine-6-phosphate deacetylase	<i>Thermotoga maritima</i>	Fe
53	Q9WZS7	NADH-dependent butanol dehydrogenase	<i>Thermotoga maritima</i>	Fe
54	Q9X0P5	Metallo- $\beta$ -lactamase H8A	<i>Thermotoga maritima</i>	Ni
55	Q9X1T8	Transcription regulator	<i>Thermotoga maritima</i>	Ni
56	Q672W7	Peptide deformylase	<i>Helicobacter pylori</i>	Co
57	Q838S1	Polysaccharide monooxygenase	<i>Enterococcus faecalis</i>	Cu
58	Q53586	Dioxygenase	<i>Streptomyces avermitilis</i>	Fe
59	Q56185	Hydroxypropylphosphonic acid epoxidase	<i>Streptomyces wedmorensis</i>	Co
60	Q96323	Anthocyanidin synthase	<i>Arabidopsis thaliana</i>	Fe
61	Q07XY2	Putative hydrolase	<i>Shewanella frigidimarina</i>	Ni
62	Q9UXT7	tRNA N6-adenosine threonylcarbamoyltransferase	<i>Pyrococcus abyssi</i>	Fe

The MetalPDB queries included the following information:

molecule type: protein. site type: mononuclear. representatives only: yes. Metal: any. Geometry: selected from the following list: square pyramid, trigonal bipyramid with a vacancy (axial / equatorial), trigonal bipyramid, square plane with a vacancy, square plane, square pyramid with a vacancy, square antiprism, square antiprism with a vacancy, tetrahedron, tetrahedron with a vacancy, trigonal plane, linear, octahedron, trigonal prism with a vacancy, trigonal prism, irregular, trigonal plane with a vacancy.

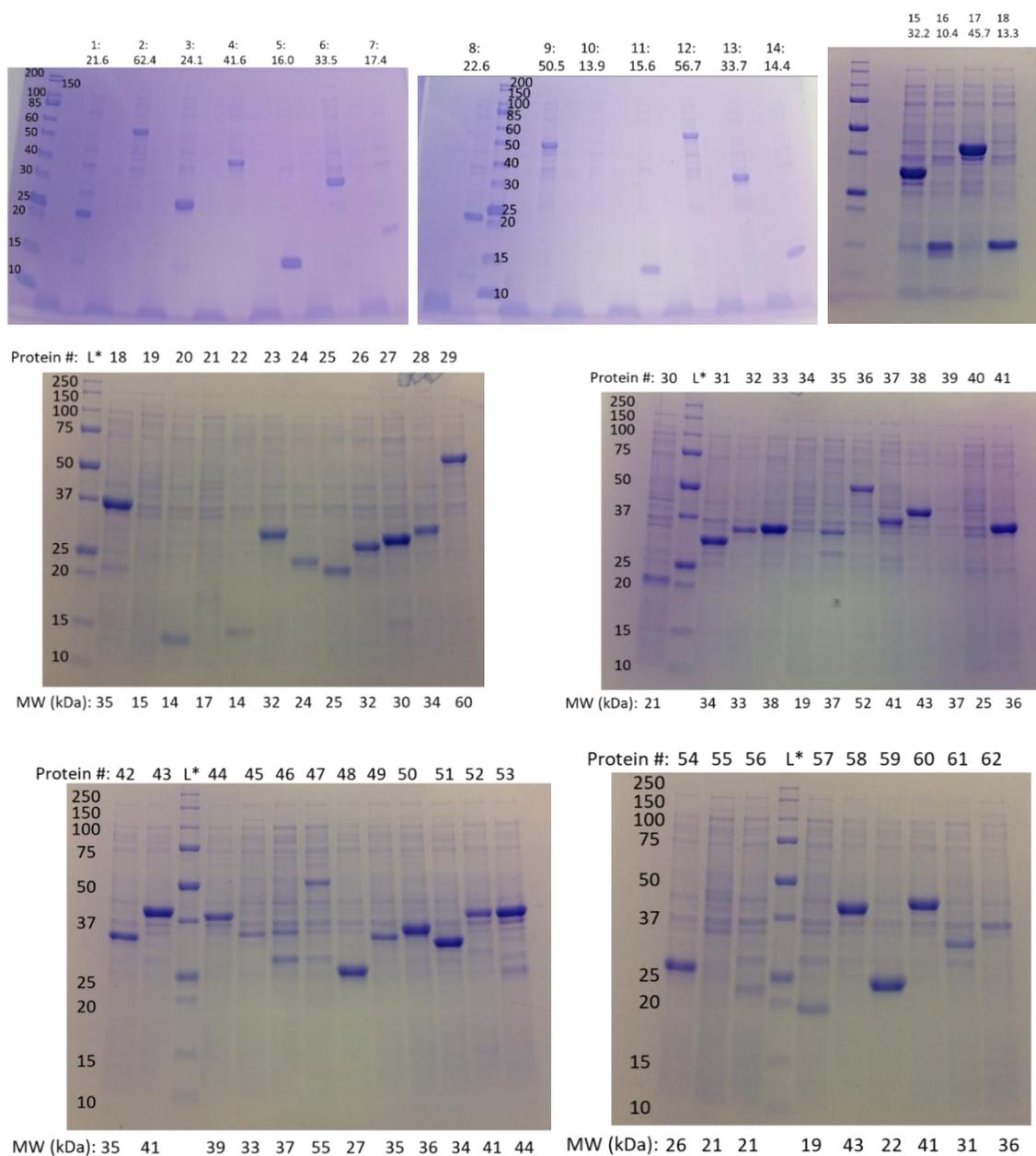


Figure 4-29. Expression tests of the non-heme metalloprotein collection. Protein numbers refer to numbering in Table 4-5. The molecular weight for each protein is given in kDa. L\* denotes the lane with the molecular weight ladder on each gel.

### **Analytical instrumentation**

HPLC-MS analysis for initial activity determination was performed on an Agilent 1290 UPLC-MS equipped with a C18 silica column (1.8  $\mu\text{m}$  packing, 2.1 $\times$ 50 mm). HPLC-MS analysis of site-saturation mutagenesis libraries was performed on an Agilent 1260 Infinity HPLC with an Agilent 6120 quadrupole mass spectrometer. Reverse-phase HPLC-UV analysis was performed with an Agilent 1200 series HPLC or an Agilent 1260 Series Infinity II HPLC using an Agilent Poroshell 120 EC-C18 column (4  $\mu\text{m}$  packing, 2.1 $\times$ 50 mm) fitted with a Poroshell 120 guard column (1.7  $\mu\text{m}$  packing, 2.1 $\times$ 5 mm). Normal-phase HPLC-UV analysis for chiral separations was performed with a Hewlett Packard Series 1100 HPLC instrument using a Daicel Chiralcel OJ-H column, (5  $\mu\text{m}$  packing, 4.6 $\times$ 250 mm) or a Daicel Chiralpak IB column (5  $\mu\text{m}$  packing, 4.6 $\times$ 250 mm).

NMR spectra were recorded on a Varian Unity/Inova 500 spectrometer operating at 500 MHz and 125 MHz for  $^1\text{H}$  and  $^{13}\text{C}$ , respectively, or a Bruker Avance 400 spectrometer operating at 400 MHz and 100 MHz for  $^1\text{H}$  and  $^{13}\text{C}$ , respectively. NMR data were analyzed in MestReNova (Mestrelab Research). Chemical shifts are reported in ppm with the solvent resonance as the internal standard. For  $^1\text{H}$  NMR:  $\text{CDCl}_3$ ,  $\delta$  7.26. For  $^{13}\text{C}$  NMR:  $\text{CDCl}_3$ ,  $\delta$  77.16. Data are reported as follows: s = singlet, d = doublet, t = triplet, q = quartet, spt = septet, m = multiplet, br = broad; coupling constants in Hz; integration.

### **Materials**

Oligonucleotides were purchased from IDT DNA. PCRs were run with Phusion® High-Fidelity PCR Kit (New England Biolabs). Gibson assembly mix<sup>55</sup> is prepared with isothermal master mix in-house and enzymes T5 exonuclease, Phusion® DNA polymerase, and Taq DNA ligase purchased from New England Biolabs.

### **Cloning**

Plasmids encoding *Pseudomonas savastanoi* ethylene-forming enzyme (UniProt ID P32021), *Streptomyces sp.* 2-aminobutyric acid chlorinase (UniProt ID D0VX22), and *Arabidopsis thaliana* anthocyanidin synthase (UniProt ID Q96323), with the coding sequences codon-optimized for *Escherichia coli* were purchased from Twist Biosciences.

Plasmids encoding *Gluconobacter oxydans* leucine dioxygenase (UniProt ID Q5FQD2), *Streptomyces vinaceus* arginine hydroxylase (UniProt ID Q6WZB0), and *Streptomyces muensis* leucine hydroxylase (UniProt ID A0A0E3URV8) were obtained from the laboratory of Prof. Hans Renata (Scripps Research Institute). The plasmid encoding *E. coli* taurine dioxygenase (UniProt ID P37610) was obtained from the laboratory of Prof. Harry Gray (Caltech). All genes were encoded with a C-terminal His<sub>6</sub>-tag for purification and inserted between the NdeI and XhoI cut sites in the pET-22b(+) vector (Novagen).

Plasmids were used to transform *E. coli* BL21(DE3) cells (Lucigen) by electroporation. SOC medium (0.75 mL) was added and the cells were incubated at 37 °C for 45 minutes before being plated on Luria-Bertani medium (Research Products International) supplemented with ampicillin (100 µg mL<sup>-1</sup>, LB-amp) agar plates. Plasmids were isolated from stationary-phase cultures by miniprep (Qiagen) and Sanger sequencing was performed by Laragen, Inc. (Culver City, CA) using T7 promoter and T7 terminator primers.

### **Protein expression and purification**

Starter cultures of LB-amp were inoculated from a single *E. coli* colony on an agar plate harboring a plasmid encoding the protein of interest and grown overnight to stationary phase at 37 °C. Expression cultures of Terrific Broth (Research Products International) supplemented with ampicillin (100 mg L<sup>-1</sup>, TB-amp) were inoculated from the starter cultures (1% v/v) and shaken at 37 °C and 160 rpm in a Multitron Infors incubator. When the expression cultures reached OD<sub>600</sub> ~ 0.8 (typically 2–3 hours), they were cooled on ice for 20 minutes. Protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.5 mM). Cultures were incubated at 22 °C and 110 rpm overnight (16–24 hours). Cells were pelleted by centrifugation (5000×g, 10 minutes).

For reactions with whole cells, cell pellets were resuspended in MOPS buffer (20 mM pH 7.0) to OD<sub>600</sub> 30. For reactions with cell lysate, the whole cell suspensions were lysed by sonication (QSonica Q500 sonicator, 25% amplitude, 33% duty cycle, 3 minutes). The lysate was clarified by centrifugation (20,817×g, 10 minutes).

For purification, cell pellets were frozen at  $-20\text{ }^{\circ}\text{C}$  for at least 24 hours. Cells were resuspended in binding buffer (20 mM Tris-HCl, 100 mM sodium chloride, 20 mM imidazole, pH 7.0,  $\sim 5\text{ mL/g}$  wet cells) and lysed by sonication (QSonica Q500 sonicator, 25% amplitude, 33% duty cycle, 4 minutes). The lysate was clarified by centrifugation (20,817  $g$ , 10 minutes) followed by filtration (0.45  $\mu\text{m}$  syringe filter). The protein was purified using an Äkta Purifier with a HisTrap HP column (GE Healthcare), eluting with a gradient of 20–500 mM imidazole. Fractions containing the protein of interest were pooled and dialyzed at  $4\text{ }^{\circ}\text{C}$  against MOPS buffer (20 mM pH 7.0) containing 1 mM EDTA ( $>100:1$  v/v) (Spectrum Laboratories Spectra/Por 12–14 kD membrane) for four hours, then against MOPS buffer (20 mM pH 7.0) overnight (12–16 hours). The dialyzed protein was concentrated by centrifugal filtration (Amicon Ultra-15 10 kD MWCO) to a final concentration of 40–100  $\text{mg mL}^{-1}$ . The concentrated protein was divided into aliquots (50–100  $\mu\text{L}$ ), flash-frozen on powdered dry ice, and stored at  $-80\text{ }^{\circ}\text{C}$ . Protein concentration was determined by Bradford assay (Bio-Rad Quick Start Bradford).

### **Site-saturation mutagenesis and library screening**

Site-saturation mutagenesis was performed using the 22-codon method.<sup>56</sup> Oligonucleotides including the three 22-codon trick codons (NDT, VHG, TGG) and oligonucleotides within the ampicillin resistance cassette were used to amplify the plasmid in two pieces, with an overlap for Gibson assembly in the gene encoding the protein of interest (and where the mutation is introduced) and an overlap for Gibson assembly in the gene encoding  $\beta$ -lactamase (which confers ampicillin resistance). This two-piece assembly can prevent mis-assembled constructs from conferring antibiotic resistance. PCR products were loaded on 1% agarose gels with loading dye containing SYBR Gold nucleic acid gel stain (Thermo Fisher) and visualized on a blue transilluminator. The DNA bands at the expected size were excised and the DNA was extracted with a Zymoclean Gel DNA recovery kit. The two linear PCR products for a given site-saturation mutagenesis library were assembled *via* isothermal Gibson assembly ( $50\text{ }^{\circ}\text{C}$ , 1 hour). *E. coli* BL21(DE3) cells (Lucigen) were transformed by electroporation with the Gibson assembly product without further purification. SOC medium (0.75 mL) was added to the electroporated cells and the cells were incubated at  $37\text{ }^{\circ}\text{C}$  for 45

minutes before being plated on LB-amp agar plates. The LB-amp agar plates with the plated cells were incubated at 37 °C for 12–18 hours and stored at 4 °C until the libraries were picked. Single colonies from the agar plates were picked with sterile toothpicks and used to inoculate starter cultures (0.5 mL LB-amp) in 96 deep-well plates. The starter culture plates were grown at 37 °C, 250 rpm, and 80% humidity in a Multitron Infors shaker overnight (14–16 hours). The starter cultures (50 µL) were used to inoculate expression cultures (1 mL TB-amp) in 96 deep-well plates. In parallel, glycerol stock plates were prepared for long-term storage by mixing starter culture (50 µL) with sterile glycerol (50% v/v, 50 µL) and frozen at –80 °C. The expression cultures were grown at 37 °C, 250 rpm, and 80% humidity for three hours, then cooled on ice for 20 minutes. Protein expression was induced by addition of IPTG (0.5 mM). Cultures were incubated at 22 °C and 220 rpm overnight (18–20 hours). Cells were pelleted (5000×g, 5 minutes) and the cell pellets were frozen at –20 °C for at least 24 hours prior to use.

In site-saturation mutagenesis round 1, cells were resuspended in MOPS buffer (20 mM pH 7.0) and brought into the Coy anaerobic chamber. Ferrous ammonium sulfate (40 mM in water, 10 µL, 1 mM final concentration, prepared immediately before use), disodium  $\alpha$ -ketoglutarate (40 mM in water, 10 µL, 1 mM final concentration), L-ascorbic acid (40 mM in water, 10 µL, 1 mM final concentration, prepared immediately before use), styrene (400 mM in ethanol, 10 µL, 10 mM final concentration), and *p*-toluenesulfonyl azide (400 mM in ethanol, 10 µL, 10 mM final concentration) were added to each well. The plates were sealed with foil covers and shaken at room temperature for two hours. To quench the reactions, acetonitrile (400 µL) was added and the reaction plate was shaken for an additional 30 minutes. Insoluble material was pelleted by centrifugation (6000×g, 10 minutes) and 200 µL of the supernate was filtered through a 0.2 µm PTFE 96-well filter plate into a 96-well microplate (3000×g, 2 minutes). The microplate was sealed with a pierceable cover and analyzed via HPLC-MS (Analytical instrumentation).

After site-saturation mutagenesis round 1, cells were resuspended in MOPS buffer (20 mM pH 7.0) containing 1 mM sodium acetate. Under air, ferrous ammonium sulfate (40 mM in

water, 10  $\mu\text{L}$ , 1 mM final concentration, prepared immediately before use), L-ascorbic acid (40 mM in water, 10  $\mu\text{L}$ , 1 mM final concentration, prepared immediately before use), styrene (400 mM in ethanol, 10  $\mu\text{L}$ , 10 mM final concentration), and *p*-toluenesulfonyl azide (400 mM in ethanol, 10  $\mu\text{L}$ , 10 mM final concentration) were added to each well. The plates were sealed with foil covers and shaken at room temperature for two hours. To quench the reactions, acetonitrile (400  $\mu\text{L}$ ) was added and the reaction plate was shaken for an additional 30 minutes. Insoluble material was pelleted by centrifugation (6000 $\times g$ , 10 minutes) and 200  $\mu\text{L}$  of the supernate was filtered through a 0.2  $\mu\text{m}$  PTFE 96-well filter plate into a 96-well microplate (3000 $\times g$ , 2 minutes). The microplate was sealed with a pierceable cover and analyzed via HPLC-MS (Analytical instrumentation).

Wells which showed an apparent enhancement in aziridine product formation were streaked out (from the glycerol stock plates prepared in parallel) onto LB-amp agar plates and incubated at 37  $^{\circ}\text{C}$ . Starter cultures (5 mL LB-amp) were inoculated from single colonies and incubated overnight at 37  $^{\circ}\text{C}$ . Plasmids were isolated from these cultures and sequenced as described above (Cloning). Unique variants were then regrown in Erlenmeyer flasks as described above (Protein expression and purification) and assayed in clarified cell lysate for nitrene-transfer activity.

### **Analytical-scale biocatalytic aziridination reactions**

Biocatalytic reactions were set up in 2-mL screw-cap vials (Agilent). Purified apoprotein (350  $\mu\text{L}$ , 22.9  $\mu\text{M}$  in 20 mM MOPS pH 7.0, final concentration 20  $\mu\text{M}$ ) was added to the vial. Solutions of ferrous ammonium sulfate and L-ascorbic acid were prepared immediately prior to use. Reactions to be set up anaerobically were brought into a vinyl anaerobic chamber (Coy Laboratory Products, nitrogen atmosphere, 0–10 ppm oxygen). To each reaction was added in order ferrous ammonium sulfate (40 mM in water, 10  $\mu\text{L}$ , 1 mM final concentration), sodium acetate or other additive (40 mM in water, 10  $\mu\text{L}$ , 1 mM final concentration), and L-ascorbic acid (40 mM in water, 10  $\mu\text{L}$ , 1 mM final concentration). Each reaction was then charged with styrene (400 mM in ethanol, 10  $\mu\text{L}$ , 10 mM final concentration) immediately followed by *p*-toluenesulfonyl azide (400 mM in ethanol, 10  $\mu\text{L}$ ,

10 mM final concentration, 500 max. TTN). The reactions were sealed and shaken at room temperature for three hours unless otherwise noted. To quench the reactions, acetonitrile (350  $\mu\text{L}$ ) was added to each vial, followed by internal standard propiophenone (0.1% v/v in acetonitrile, 50  $\mu\text{L}$ ). The sample was transferred to a 1.7-mL Eppendorf tube, vortexed, and then centrifuged (20817 $\times g$ , 5 minutes). 250  $\mu\text{L}$  of the supernate was transferred to HPLC vial inserts for reverse-phase HPLC analysis. The remaining supernate was partially concentrated *in vacuo* to remove acetonitrile and ethanol. Cyclohexane (500  $\mu\text{L}$ ) was added to the resulting aqueous suspension. The mixture was thoroughly shaken and then centrifuged (20817 $\times g$ , 5 minutes). 250  $\mu\text{L}$  of the organic layer was transferred to HPLC vial inserts for normal-phase chiral HPLC analysis.

### **Analytical-scale biocatalytic C–H insertion reactions**

Biocatalytic reactions were set up in 2-mL screw-cap vials (Agilent). Purified apoprotein (360  $\mu\text{L}$ , 22.2  $\mu\text{M}$  in 20 mM MOPS pH 7.0, final concentration 20  $\mu\text{M}$ ) was added to the vial. Solutions of ferrous ammonium sulfate and L-ascorbic acid were prepared immediately prior to use. Reactions to be set up anaerobically were brought into a Coy vinyl anaerobic chamber (nitrogen atmosphere, 0–10 ppm oxygen). To each reaction was added in order ferrous ammonium sulfate (40 mM in water, 10  $\mu\text{L}$ , 1 mM final concentration), sodium acetate or other additive (40 mM in water, 10  $\mu\text{L}$ , 1 mM final concentration), L-ascorbic acid (40 mM in water, 10  $\mu\text{L}$ , 1 mM final concentration). Each reaction was then charged with 2-ethylbenzenesulfonyl azide (400 mM in ethanol, 10  $\mu\text{L}$ , 10 mM final concentration, 500 max. TTN). The reactions were sealed and shaken at room temperature for six hours unless otherwise noted. To quench the reactions, acetonitrile (350  $\mu\text{L}$ ) was added to each vial, followed by internal standard propiophenone (0.5% v/v in acetonitrile, 50  $\mu\text{L}$ ). The sample was transferred to a 1.7-mL Eppendorf tube, vortexed, and then centrifuged (20817 $\times g$ , 5 minutes). 250  $\mu\text{L}$  of the supernate was transferred to HPLC vial inserts for reverse-phase HPLC analysis. The remaining supernate was partially concentrated *in vacuo* to remove acetonitrile and ethanol. Hexanes (250  $\mu\text{L}$ , HPLC grade) and ethyl acetate (250  $\mu\text{L}$ , HPLC grade) were added. The resulting mixture was thoroughly shaken and then centrifuged

(20817×g, 5 minutes). 250 µL of the organic layer was transferred to HPLC vial inserts for normal-phase chiral HPLC analysis.

### **Thermostability measurement via thermal shift assay**

Thermal shift assay samples were prepared in triplicate anaerobically, adding all components within a vinyl anaerobic chamber (Coy Laboratory Products). To a PCR tube with purified *PsEFE* wild type or a *PsEFE* variant (stripped and dialyzed, 10–15 µM final concentration) was added (to a final concentration of 1.25 mM each) either:

- ferrous ammonium sulfate
- ferrous ammonium sulfate, L-ascorbic acid, and  $\alpha$ -ketoglutarate
- ferrous ammonium sulfate, L-ascorbic acid, and sodium acetate

Following these additions, to each tube was added 5 µL SYPRO orange (25-fold diluted in water). The PCR tubes were sealed, brought out of the anaerobic chamber, and analyzed on a Stratagene Mx3005P qPCR machine (Agilent Technologies, Inc.). The temperature program ran from 25 °C to 99 °C, holding for 30 seconds per degree before measuring fluorescence on the SYPRO channel and increasing temperature. The melting temperature for a given temperature was taken as the maximum of the numerical first derivative.

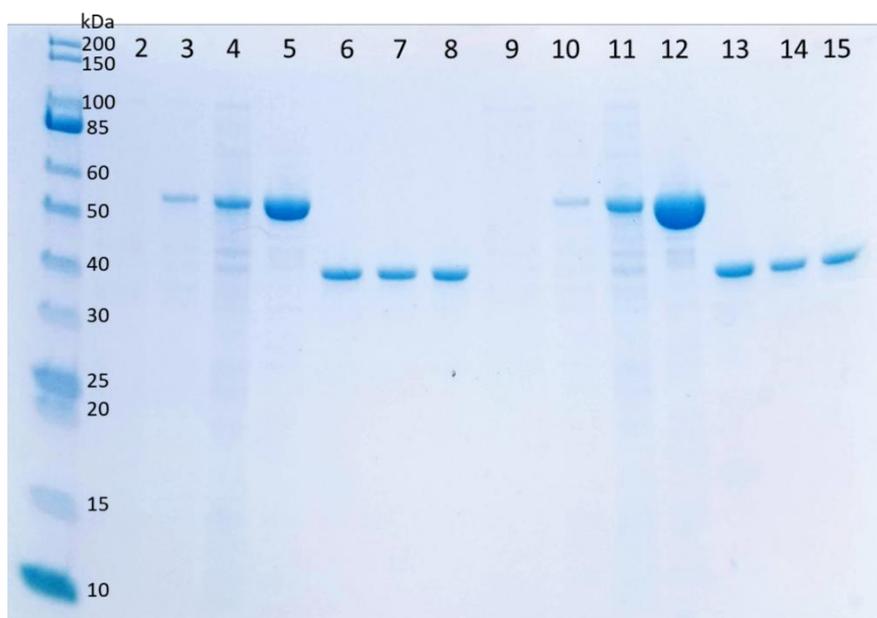


Figure 4-30. SDS-PAGE showing purification of the cleavable *N*-His<sub>6</sub>-SUMO-*PsEFE* variant constructs. Lane 1 is the molecular weight ladder, with molecular weights labeled in kDa. Lanes 2-8 are *PsEFE* VMM at induction, harvest, clarified lysate, Ulp1 digested, resin flow-through, and dialysis. Lanes 10-15 are *PsEFE* VHMM at induction, harvest, clarified lysate, Ulp1 digested, resin flow-through, and dialysis.

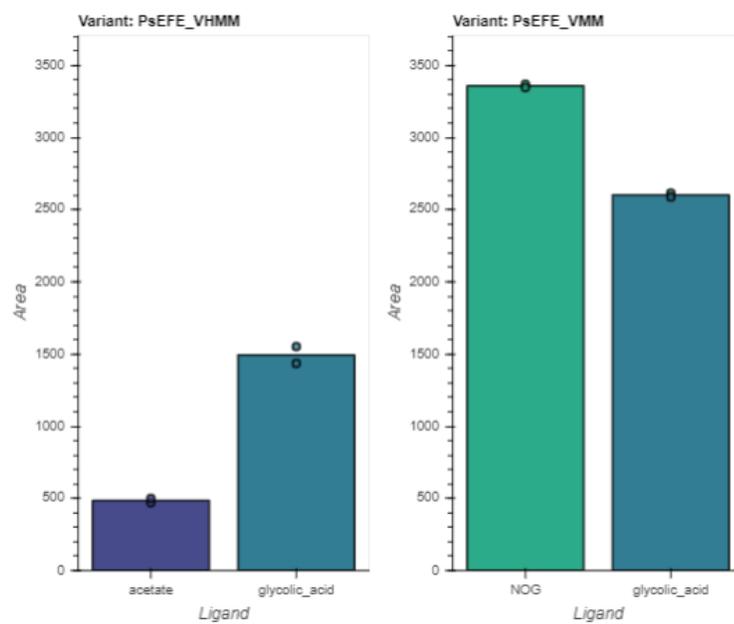


Figure 4-31. C-H insertion activity for *PsEFE* VHMM and *PsEFE* VMM with glycolic acid as an alternative ligand.

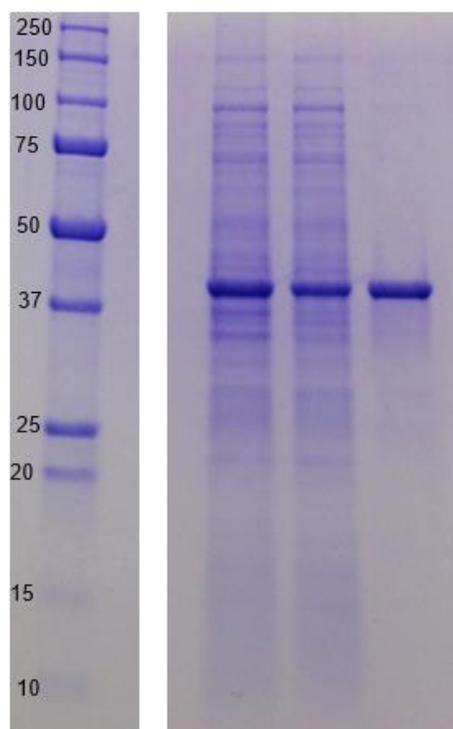


Figure 4-32 SDS-PAGE of PsEFE aziridination variant T97M R171L R277H F314M C317M (PsEFE MLHMM). The protein is shown from left to right as whole-cell sample, clarified lysate sample, and purified protein sample. Whole cells and lysates were diluted 25-fold; purified protein was diluted 50-fold (each dilution is prior to addition of 2X Laemmli loading dye). The ladder and sample were run on the same gel; unrelated protein samples were cropped out for image clarity. The SDS-PAGE image brightness was increased in Microsoft Word for image clarity and is not being used for quantitation.

#### 4.10 References for Chapter 4

- (1) Sreenilayam, G.; Moore, E. J.; Steck, V.; Fasan, R. Stereoselective Olefin Cyclopropanation under Aerobic Conditions with an Artificial Enzyme Incorporating an Iron-Chlorin E6 Cofactor. *ACS Catal.* **2017**, *7*, 7629–7633. <https://doi.org/10.1021/acscatal.7b02583>.
- (2) Natoli, S. N.; Hartwig, J. F. Noble–Metal Substitution in Hemoproteins: An Emerging Strategy for Abiological Catalysis. *Acc. Chem. Res.* **2019**, *52*, 326–335. <https://doi.org/10.1021/acs.accounts.8b00586>.
- (3) Sreenilayam, G.; Moore, E. J.; Steck, V.; Fasan, R. Metal Substitution Modulates the Reactivity and Extends the Reaction Scope of Myoglobin Carbene Transfer Catalysts. *Adv. Synth. Catal.* **2017**, *359*, 2076–2089. <https://doi.org/10.1002/adsc.201700202>.
- (4) Wolf, M. W.; Vargas, D. A.; Lehnert, N. Engineering of RuMb: Toward a Green Catalyst for Carbene Insertion Reactions. *Inorg. Chem.* **2017**, *56*, 5623–5635.

- <https://doi.org/10.1021/acs.inorgchem.6b03148>.
- (5) Liang, A. D.; Serrano-Plana, J.; Peterson, R. L.; Ward, T. R. Artificial Metalloenzymes Based on the Biotin–Streptavidin Technology: Enzymatic Cascades and Directed Evolution. *Acc. Chem. Res.* **2019**, *52*, 585–595. <https://doi.org/10.1021/acs.accounts.8b00618>.
  - (6) Lewis, J. C. Beyond the Second Coordination Sphere: Engineering Dirhodium Artificial Metalloenzymes To Enable Protein Control of Transition Metal Catalysis. *Acc. Chem. Res.* **2019**, *52*, 576–584. <https://doi.org/10.1021/acs.accounts.8b00625>.
  - (7) Cotruvo, Jr, J. A.; Stubbe, J. Metallation and Mismetallation of Iron and Manganese Proteins in Vitro and in Vivo: The Class I Ribonucleotide Reductases as a Case Study. *Metallomics* **2012**, *4*, 1020. <https://doi.org/10.1039/c2mt20142a>.
  - (8) Barber-Zucker, S.; Shaanan, B.; Zarivach, R. Transition Metal Binding Selectivity in Proteins and Its Correlation with the Phylogenomic Classification of the Cation Diffusion Facilitator Protein Family. *Sci. Rep.* **2017**, *7*, 16381. <https://doi.org/10.1038/s41598-017-16777-5>.
  - (9) Foster, A. W.; Osman, D.; Robinson, N. J. Metal Preferences and Metallation. *J. Biol. Chem.* **2014**, *289*, 28095–28103. <https://doi.org/10.1074/jbc.R114.588145>.
  - (10) Boyd, E. S.; Peters, J. W. New Insights into the Evolutionary History of Biological Nitrogen Fixation. *Front. Microbiol.* **2013**, *4*, 1–12. <https://doi.org/10.3389/fmicb.2013.00201>.
  - (11) Garcia, Y. M.; Barwinska-Sendra, A.; Tarrant, E.; Skaar, E. P.; Waldron, K. J.; Kehl-Fie, T. E. A Superoxide Dismutase Capable of Functioning with Iron or Manganese Promotes the Resistance of *Staphylococcus Aureus* to Calprotectin and Nutritional Immunity. *PLoS Pathog.* **2017**, *13*, e1006125. <https://doi.org/10.1371/journal.ppat.1006125>.
  - (12) Salgado, J.; Jimenez, H. R.; Donaire, A.; Moratal, J. M. 1 H-NMR Study of a Cobalt-Substituted Blue Copper Protein: *Pseudomonas Aeruginosa* Co(II)-Azurin. *Eur. J. Biochem.* **1995**, *231*, 358–369. <https://doi.org/10.1111/j.1432-1033.1995.tb20708.x>.
  - (13) McLaughlin, M. P.; Retegan, M.; Bill, E.; Payne, T. M.; Shafaat, H. S.; Peña, S.; Sudhamsu, J.; Ensign, A. A.; Crane, B. R.; Neese, F.; Holland, P. L. Azurin as a Protein Scaffold for a Low-Coordinate Nonheme Iron Site with a Small-Molecule Binding Pocket. *J. Am. Chem. Soc.* **2012**, *134*, 19746–19757. <https://doi.org/10.1021/ja308346b>.
  - (14) Moratal, J. M.; Romero, A.; Salgado, J.; Perales-Alarcon, A.; Jimenez, H. R. The Crystal Structure of Nickel(II)-Azurin. *Eur. J. Biochem.* **1995**, *228*, 653–657. <https://doi.org/10.1111/j.1432-1033.1995.0653m.x>.
  - (15) McLaughlin, M. P.; Darrah, T. H.; Holland, P. L. Palladium(II) and Platinum(II) Bind Strongly to an Engineered Blue Copper Protein. *Inorg. Chem.* **2011**, *50*, 11294–11296. <https://doi.org/10.1021/ic2017648>.
  - (16) Brandenburg, O. F.; Fasan, R.; Arnold, F. H. Exploiting and Engineering Hemoproteins for Abiological Carbene and Nitrene Transfer Reactions. *Curr. Opin. Biotechnol.* **2017**, *47*, 102–111. <https://doi.org/10.1016/j.copbio.2017.06.005>.
  - (17) Chen, K.; Arnold, F. H. Engineering New Catalytic Activities in Enzymes. *Nat. Catal.* **2020**, *3*, 203–213. <https://doi.org/10.1038/s41929-019-0385-5>.

- (18) Bateman, A.; Martin, M. J.; O'Donovan, C.; Magrane, M.; Alpi, E.; Antunes, R.; Bely, B.; Bingley, M.; Bonilla, C.; Britto, R.; Bursteinas, B.; Bye-AJee, H.; Cowley, A.; Da Silva, A.; De Giorgi, M.; Dogan, T.; Fazzini, F.; Castro, L. G.; Figueira, L.; Garmiri, P.; Georghiou, G.; Gonzalez, D.; Hatton-Ellis, E.; Li, W.; Liu, W.; Lopez, R.; Luo, J.; Lussi, Y.; MacDougall, A.; Nightingale, A.; Palka, B.; Pichler, K.; Poggioli, D.; Pundir, S.; Pureza, L.; Qi, G.; Rosanoff, S.; Saidi, R.; Sawford, T.; Shypitsyna, A.; Speretta, E.; Turner, E.; Tyagi, N.; Volynkin, V.; Wardell, T.; Warner, K.; Watkins, X.; Zaru, R.; Zellner, H.; Xenarios, I.; Bougueleret, L.; Bridge, A.; Poux, S.; Redaschi, N.; Aimo, L.; ArgoudPuy, G.; Auchincloss, A.; Axelsen, K.; Bansal, P.; Baratin, D.; Blatter, M. C.; Boeckmann, B.; Bolleman, J.; Boutet, E.; Breuza, L.; Casal-Casas, C.; De Castro, E.; Coudert, E.; CuChe, B.; Doche, M.; Dornevil, D.; Duvaud, S.; Estreicher, A.; Famiglietti, L.; Feuermann, M.; Gasteiger, E.; Gehant, S.; Gerritsen, V.; Gos, A.; Gruaz-Gumowski, N.; Hinz, U.; Hulo, C.; Jungo, F.; Keller, G.; Lara, V.; Lemercier, P.; Lieberherr, D.; Lombardot, T.; Martin, X.; Masson, P.; Morgat, A.; Neto, T.; Nospikel, N.; Paesano, S.; Pedruzzi, I.; Pilbout, S.; Pozzato, M.; Pruess, M.; Rivoire, C.; Roechert, B.; Schneider, M.; Sigrist, C.; Sonesson, K.; Staehli, S.; Stutz, A.; Sundaram, S.; Tognolli, M.; Verbregue, L.; Veuthey, A. L.; Wu, C. H.; Arighi, C. N.; Arminski, L.; Chen, C.; Chen, Y.; Garavelli, J. S.; Huang, H.; Laiho, K.; McGarvey, P.; Natale, D. A.; Ross, K.; Vinayaka, C. R.; Wang, Q.; Wang, Y.; Yeh, L. S.; Zhang, J. UniProt: The Universal Protein Knowledgebase. *Nucleic Acids Res.* **2017**, *45*, D158–D169. <https://doi.org/10.1093/nar/gkw1099>.
- (19) Placzek, S.; Schomburg, I.; Chang, A.; Jeske, L.; Ulbrich, M.; Tillack, J.; Schomburg, D. BRENDA in 2017: New Perspectives and New Tools in BRENDA. *Nucleic Acids Res.* **2017**, *45*, D380–D388. <https://doi.org/10.1093/nar/gkw952>.
- (20) Kan, S. B. J. J.; Lewis, R. D.; Chen, K.; Arnold, F. H. Directed Evolution of Cytochrome c for Carbon–Silicon Bond Formation: Bringing Silicon to Life. *Science* **2016**, *354*, 1048–1051. <https://doi.org/10.1126/science.aah6219>.
- (21) Lu, H.; Zhang, X. P. Catalytic C–H Functionalization by Metalloporphyrins: Recent Developments and Future Directions. *Chem. Soc. Rev.* **2011**, *40*, 1899–1909. <https://doi.org/10.1039/c0cs00070a>.
- (22) Groves, J. T.; Nemo, T. E.; Myers, R. S. Hydroxylation and Epoxidation Catalyzed by Iron-Porphine Complexes. Oxygen Transfer from Iodosylbenzene. *J. Am. Chem. Soc.* **1979**, *101*, 1032–1033. <https://doi.org/10.1021/ja00498a040>.
- (23) Huang, P.-S.; Boyken, S. E.; Baker, D. The Coming of Age of de Novo Protein Design. *Nature* **2016**, *537*, 320–327. <https://doi.org/10.1038/nature19946>.
- (24) Khersonsky, O.; Tawfik, D. S. Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. *Annu. Rev. Biochem.* **2010**, *79*, 471–505. <https://doi.org/10.1146/annurev-biochem-030409-143718>.
- (25) Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. <https://doi.org/10.1093/nar/28.1.235>.
- (26) Putignano, V.; Rosato, A.; Banci, L.; Andreini, C. MetalPDB in 2018: A Database of Metal Sites in Biological Macromolecular Structures. *Nucleic Acids Res.* **2018**, *46*, D459–D464. <https://doi.org/10.1093/nar/gkx989>.
- (27) McCall, K. A.; Huang, C.; Fierke, C. A. Function and Mechanism of Zinc

- Metalloenzymes. *J. Nutr.* **2000**, *130*, 1437S-1446S.  
<https://doi.org/10.1093/jn/130.5.1437S>.
- (28) Jarzab, A.; Kurzawa, N.; Hopf, T.; Moerch, M.; Zecha, J.; Leijten, N.; Bian, Y.; Musiol, E.; Maschberger, M.; Stoehr, G.; Becher, I.; Daly, C.; Samaras, P.; Mergner, J.; Spanier, B.; Angelov, A.; Werner, T.; Bantscheff, M.; Wilhelm, M.; Klingenspor, M.; Lemeer, S.; Liebl, W.; Hahne, H.; Savitski, M. M.; Kuster, B. Meltome Atlas—Thermal Proteome Stability across the Tree of Life. *Nat. Methods* **2020**, *17*, 495–503. <https://doi.org/10.1038/s41592-020-0801-4>.
- (29) Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H. Protein Stability Promotes Evolvability. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 5869–5874. <https://doi.org/10.1073/pnas.0510098103>.
- (30) Petersen, T. N.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions. *Nat. Methods* **2011**, *8*, 785–786. <https://doi.org/10.1038/nmeth.1701>.
- (31) Chaplin, A. K.; Wilson, M. T.; Hough, M. A.; Svistunenko, D. A.; Hemsworth, G. R.; Walton, P. H.; Vijgenboom, E.; Worrall, J. A. R. Heterogeneity in the Histidine-Brace Copper Coordination Sphere in Auxiliary Activity Family 10 (AA10) Lytic Polysaccharide Monooxygenases. *J. Biol. Chem.* **2016**, *291*, 12838–12850. <https://doi.org/10.1074/jbc.M116.722447>.
- (32) Harris, D. F.; Lukoyanov, D. A.; Kallas, H.; Trncik, C.; Yang, Z. Y.; Compton, P.; Kelleher, N.; Einsle, O.; Dean, D. R.; Hoffman, B. M.; Seefeldt, L. C. Mo-, V-, and Fe-Nitrogenases Use a Universal Eight-Electron Reductive-Elimination Mechanism to Achieve N<sub>2</sub> Reduction. *Biochemistry* **2019**, *58*, 3293–3301. <https://doi.org/10.1021/acs.biochem.9b00468>.
- (33) Sieracki, N. A.; Tian, S.; Hadt, R. G.; Zhang, J. L.; Woertink, J. S.; Nilges, M. J.; Sun, F.; Solomon, E. I.; Lu, Y. Copper-Sulfenate Complex from Oxidation of a Cavity Mutant of *Pseudomonas aeruginosa* Azurin. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 924–929. <https://doi.org/10.1073/pnas.1316483111>.
- (34) Alagaratnam, S.; Meeuwenoord, N. J.; Navarro, J. A.; Hervás, M.; De La Rosa, M. A.; Hoffmann, M.; Einsle, O.; Ubbink, M.; Canters, G. W. Probing the Reactivity of Different Forms of Azurin by Flavin Photoreduction. *FEBS J.* **2011**, *278*, 1506–1521. <https://doi.org/10.1111/j.1742-4658.2011.08067.x>.
- (35) Reijerse, E. J.; Pham, C. C.; Pelmeshnikov, V.; Gilbert-Wilson, R.; Adamska-Venkatesh, A.; Siebel, J. F.; Gee, L. B.; Yoda, Y.; Tamasaku, K.; Lubitz, W.; Rauchfuss, T. B.; Cramer, S. P. Direct Observation of an Iron-Bound Terminal Hydride in [FeFe]-Hydrogenase by Nuclear Resonance Vibrational Spectroscopy. *J. Am. Chem. Soc.* **2017**, *139*, 4306–4309. <https://doi.org/10.1021/jacs.7b00686>.
- (36) Arnold, F. H. Innovation by Evolution: Bringing New Chemistry to Life (Nobel Lecture). *Angew. Chem., Int. Ed.* **2019**, *58*, 14420–14426. <https://doi.org/10.1002/anie.201907729>.
- (37) Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. Olefin Cyclopropanation via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes. *Science* **2013**, *339*, 307–310. <https://doi.org/10.1126/science.1231434>.
- (38) Eckert, C.; Xu, W.; Xiong, W.; Lynch, S.; Ungerer, J.; Tao, L.; Gill, R.; Maness, P.-C.; Yu, J. Ethylene-Forming Enzyme and Bioethylene Production. *Biotechnol.*

- Biofuels* **2014**, *7*, 33. <https://doi.org/10.1186/1754-6834-7-33>.
- (39) Martinez, S.; Hausinger, R. P. Biochemical and Spectroscopic Characterization of the Non-Heme Fe(II)- and 2-Oxoglutarate-Dependent Ethylene-Forming Enzyme from *Pseudomonas Syringae* Pv. Phaseolicola PK2. *Biochemistry* **2016**, *55*, 5989–5999. <https://doi.org/10.1021/acs.biochem.6b00890>.
- (40) Farwell, C. C.; Zhang, R. K.; McIntosh, J. A.; Hyster, T. K.; Arnold, F. H. Enantioselective Enzyme-Catalyzed Aziridination Enabled by Active-Site Evolution of a Cytochrome P450. *ACS Cent. Sci.* **2015**, *1*, 89–93. <https://doi.org/10.1021/acscentsci.5b00056>.
- (41) Prier, C. K.; Zhang, R. K.; Buller, A. R.; Brinkmann-Chen, S.; Arnold, F. H. Enantioselective, Intermolecular Benzylic C–H Amination Catalysed by an Engineered Iron-Haem Enzyme. *Nat. Chem.* **2017**, *9*, 629–634. <https://doi.org/10.1038/nchem.2783>.
- (42) Renata, H.; Wang, Z. J.; Arnold, F. H. Expanding the Enzyme Universe: Accessing Non-Natural Reactions by Mechanism-Guided Directed Evolution. *Angew. Chem., Int. Ed.* **2015**, *54*, 3351–3367. <https://doi.org/10.1002/anie.201409470>.
- (43) Brandenburg, O. F.; Chen, K.; Arnold, F. H. Directed Evolution of a Cytochrome P450 Carbene Transferase for Selective Functionalization of Cyclic Compounds. *J. Am. Chem. Soc.* **2019**, *141*, 8989–8995. <https://doi.org/10.1021/jacs.9b02931>.
- (44) Zhang, R. K.; Chen, K.; Huang, X.; Wohlschlager, L.; Renata, H.; Arnold, F. H. Enzymatic Assembly of Carbon–Carbon Bonds via Iron-Catalysed Sp<sup>3</sup> C–H Functionalization. *Nature* **2019**, *565*, 67–72. <https://doi.org/10.1038/s41586-018-0808-5>.
- (45) Li, M.; Martinez, S.; Hausinger, R. P.; Emerson, J. P. Thermodynamics of Iron(II) and Substrate Binding to the Ethylene-Forming Enzyme. *Biochemistry* **2018**, *57*, 5696–5705. <https://doi.org/10.1021/acs.biochem.8b00730>.
- (46) Ericsson, U. B.; Hallberg, B. M.; DeTitta, G. T.; Dekker, N.; Nordlund, P. Thermofluor-Based High-Throughput Stability Optimization of Proteins for Structural Studies. *Anal. Biochem.* **2006**, *357*, 289–298. <https://doi.org/10.1016/j.ab.2006.07.027>.
- (47) Martinez, S.; Fellner, M.; Herr, C. Q.; Ritchie, A.; Hu, J.; Hausinger, R. P. Structures and Mechanisms of the Non-Heme Fe(II)- and 2-Oxoglutarate-Dependent Ethylene-Forming Enzyme: Substrate Binding Creates a Twist. *J. Am. Chem. Soc.* **2017**, *139*, 11980–11988. <https://doi.org/10.1021/jacs.7b06186>.
- (48) Zhang, Z.; Smart, T. J.; Choi, H.; Hardy, F.; Lohans, C. T.; Abboud, M. I.; Richardson, M. S. W.; Paton, R. S.; McDonough, M. A.; Schofield, C. J. Structural and Stereoelectronic Insights into Oxygenase-Catalyzed Formation of Ethylene from 2-Oxoglutarate. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 4667–4672. <https://doi.org/10.1073/pnas.1617760114>.
- (49) Bennett, B. D.; Kimball, E. H.; Gao, M.; Osterhout, R.; Van Dien, S. J.; Rabinowitz, J. D. Absolute Metabolite Concentrations and Implied Enzyme Active Site Occupancy in *Escherichia Coli*. *Nat. Chem. Biol.* **2009**, *5*, 593–599. <https://doi.org/10.1038/nchembio.186>.
- (50) Otey, C. R. High-Throughput Carbon Monoxide Binding Assay for Cytochromes P450. In *Directed Enzyme Evolution*; Humana Press: New Jersey, 2003; Vol. 230,

- pp 137–140. <https://doi.org/10.1385/1-59259-396-8:137>.
- (51) Cabantous, S.; Waldo, G. S. In Vivo and in Vitro Protein Solubility Assays Using Split GFP. *Nat. Methods* **2006**, *3*, 845–854. <https://doi.org/10.1038/nmeth932>.
- (52) Santos-Aberturas, J.; Dörr, M.; Waldo, G. S.; Bornscheuer, U. T. In-Depth High-Throughput Screening of Protein Engineering Libraries by Split-GFP Direct Crude Cell Extract Data Normalization. *Chem. Biol.* **2015**, *22*, 1406–1414. <https://doi.org/10.1016/j.chembiol.2015.08.014>.
- (53) Feng, S.; Sekine, S.; Pessino, V.; Li, H.; Leonetti, M. D.; Huang, B. Improved Split Fluorescent Proteins for Endogenous Protein Labeling. *Nat. Commun.* **2017**, *8*, 370. <https://doi.org/10.1038/s41467-017-00494-8>.
- (54) Moreau, M. J. J.; Morin, I.; Schaeffer, P. M. Quantitative Determination of Protein Stability and Ligand Binding Using a Green Fluorescent Protein Reporter System. *Mol. BioSyst.* **2010**, *6*, 1285. <https://doi.org/10.1039/c002001j>.
- (55) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat. Methods* **2009**, *6*, 343–345. <https://doi.org/10.1038/nmeth.1318>.
- (56) Kille, S.; Acevedo-Rocha, C. G.; Parra, L. P.; Zhang, Z.-G.; Opperman, D. J.; Reetz, M. T.; Acevedo, J. P. Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synth. Biol.* **2013**, *2*, 83–92. <https://doi.org/10.1021/sb300037w>.

## INCREASING ROBUSTNESS OF CYTOCHROME P411<sub>BM3</sub> TO FACILITATE ENGINEERING FOR NEW-TO-NATURE REACTIONS

A.M.K conceived of the project. A.M.K. designed comprehensive site-saturation mutagenesis libraries for truncated cytochrome P411<sub>BM3</sub> variant C10. A.M.K. generated the split-fluorescent P411<sub>BM3</sub>-C10-mNG2(11) construct and evaluated its utility for stability and expression assays.

## Abstract

The self-sufficient *Bacillus megaterium* cytochrome P450 (P450<sub>BM3</sub>) has been engineered to perform selective hydroxylation reactions on myriad substrates. Recently, P450<sub>BM3</sub> and its serine-ligated variant P411<sub>BM3</sub> have been established as a platform for new-to-nature carbene- and nitrene-transfer reactions. Even though the enzyme platform has undergone many rounds of mutagenesis, only a small fraction of the possible sequence space has been touched, leaving the vast majority unexplored. In this Chapter, I describe the design criteria used in the generation of a tile-based mutagenesis approach to generate amino-acid substitutions at 504 residues in P411<sub>BM3</sub>, propose a combination of expression and stability screening methods to evaluate these variants, and describe a method for recombining beneficial mutations within tiles, between tiles, and into other P411<sub>BM3</sub> variants.

## 5.1 Effect of protein stability in directed evolution

The main mantra in the field of directed evolution is, “You get what you screen for.”<sup>1</sup> The converse is also true: if one does not screen or select for a current feature of a protein, it is possible to lose that feature (you might not keep what you do not screen for). As new-function mutations have a high probability of being destabilizing,<sup>2</sup> starting with a poorly stable protein as the parent and/or not monitoring the potentially destabilizing effects of introduced mutations can rapidly result in a protein for which any mutations are deleterious. A prime example of this was reported on the evolutionary lineage to convert cytochrome P450<sub>BM3</sub> from a fatty acid hydroxylase to a propane monooxygenase. Researchers found that in order to continue evolution toward the desired activity, they had to halt these efforts and, instead, perform a round of evolution focused on stabilizing the variant.<sup>3</sup> In contrast, thermostable proteins – naturally or engineered – tend to be highly evolvable;<sup>4</sup> empirically, proteins stable to far above operating conditions are desirable starting points for directed evolution, because they can tolerate the introduction of more deleterious mutations than a less stable starting point.<sup>5</sup> While the desired feature is a protein’s long-term stability to its reaction conditions, thermostability is positively correlated to a protein’s robustness when exposed to harsh reaction conditions such as, for example, high organic solvent concentrations for a long period of time,<sup>6,7</sup> Even though this correlation does not equate causation, protein engineers often choose a naturally thermostable parent enzyme or, if no thermostable parent enzyme has been found, thermostabilize a mesostable one before any other engineering efforts are made. Current methods for evolving protein stability involve a combination of standard mutagenesis library construction and screening, data-driven modeling of protein stability,<sup>8</sup> and rational design to predict stabilizing mutations.<sup>9,10</sup>

## 5.2 Cytochrome P411<sub>BM3</sub> as a platform for evolving new-to-nature biocatalysts

Cytochrome P450<sub>BM3</sub> was the first protein scaffold demonstrated to perform carbene transfer, as described in Coelho et. al's seminal report of enzymatic styrene cyclopropanation.<sup>11</sup> Shortly thereafter, it was shown that the mutation of the axial cysteine to serine (C400S = P411), known since 2002 to abolish native monooxygenase activity in mammalian

cytochromes P450,<sup>12</sup> enhanced carbene transfer activity and enabled the use of *in vivo* reductants.<sup>13</sup> Globins, cytochromes *c*, and other members of the cytochrome P450 family also have been engineered for a variety of carbene- and nitrene-transfer reactions,<sup>14,15</sup> but the majority of these new-to-nature transformations has been reported with variants of P411<sub>BM3</sub>. One P411<sub>BM3</sub> variant of note as part of this biocatalytic platform is P411<sub>BM3</sub>-C10, which was discovered as part of the C–H alkylation lineage.<sup>16</sup> Due to its broad reaction and substrate scope, P411<sub>BM3</sub>-C10 has been used as a starting point for several successful engineering campaigns in the Arnold laboratory, including the cyclopropanation of internal alkynes, synthesis of substituted lactones, and further unpublished work.

Despite being the target for several directed evolution projects aiming to expand the repertoire of carbene- and nitrene-transfer biocatalysis, most sequence diversification in the P411<sub>BM3</sub> evolutionary family so far has been focused on a handful of amino-acid residues in the distal heme pocket. This makes sense because these positions are known to have a major effect on the enzyme's stereo- and regioselectivity<sup>17</sup>, likely through modulating substrate binding. At the same time, mutations at these positions have also been shown to negatively affect the stability of P450<sub>BM3</sub> and P411<sub>BM3</sub>.<sup>11</sup> Unlike cytochromes *c* and protoglobins used in other carbene- and nitrene-transfer projects, P411<sub>BM3</sub> and its variants are not particularly thermostable. Efforts to determine thermostabilizing mutations in P411<sub>BM3</sub> could offset the destabilization of new gain-of-function mutations,<sup>2,11</sup> further improving the utility of this biocatalytic platform. Additionally, improved robustness will make this versatile biocatalytic platform more useful and attractive for industrial applications.

### **5.3 High-throughput techniques to assay protein fitness**

Efficient sampling of a large fraction of the sequence space typically generated with mutagenesis libraries requires a screen which can rapidly assess variants for fitness. The highest throughput techniques are selection-based approaches; if protein fitness can be tied to gene expression in some way, these techniques can be applied to enrich the library with beneficial mutations in a rapid and facile fashion. Continuous evolution systems such as

phage-assisted continuous evolution (PACE)<sup>18</sup> and error-prone orthogonal replication (OrthoRep)<sup>19</sup> are prominent examples.

If the fitness can be linked to fluorescence, (e.g. by using a fluorogenic substrate or producing a fluorescent protein), ultra-high throughput sorting techniques such as fluorescence-activated cell sorting<sup>20</sup> and fluorescence-activated droplet sorting<sup>21</sup> can be used to screen 10<sup>3</sup> samples per second. Fitness linked to absorbance has also been shown to work in microfluidic droplet-sorting methods.<sup>22</sup> Spectrophotometric-based measurements rely on the change in absorbance intensity at a particular wavelength. Even minor (5–10 nm) shifts in absorbance spectra or changes in extinction coefficients can be sufficient to develop a rapid and effective assay. Such spectrophotometric methods have been used in the study of transaminases<sup>23</sup> and tryptophan synthase,<sup>24</sup> and indirect assays based on cofactor depletion can be applied to many enzyme classes.<sup>25</sup>

So far, most screening methods used for protein engineering campaigns with P411<sub>BM3</sub> as a scaffold have been based on chromatographic separations such as liquid chromatography (LC) or gas chromatography (GC), which typically are considered low- to medium-throughput. These chromatographic methods enable a sensitive and quantitative measurement of the molecule of interest, rather than using an indirect readout or surrogate molecules – ensuring that “you get what you screen for.” While highly optimized analytical methods and rapid techniques such as multiple injections in a single experimental run (MISER) exist<sup>26</sup>, the majority of LC and GC screens takes on the order of minutes per sample, compared to chromogenic or fluorescent assays which can be screened in a matter of seconds, or even faster, per sample. This lengthy screening time has led to the almost exclusive use of site-saturation mutagenesis techniques and the generation of very small libraries to generate protein variants of P411<sub>BM3</sub>. As a result, amino-acid positions distant from the heme-binding pocket in P411<sub>BM3</sub> have been largely neglected, and the few reported mutations introduced into the reductase domain of P411<sub>BM3</sub> have been the result of random mutagenesis via error-prone PCR.<sup>27</sup> In fact, only two efforts to engineer P411<sub>BM3</sub> variants have utilized random mutagenesis, an approach that generally requires the screening of

substantially more samples than a targeted one. In the first, variants were screened via HPLC for C–H alkylation activity; this resulted in the identification of additional positions which were targeted with site-saturation mutagenesis in following rounds of evolution.<sup>16</sup> In the second effort, a high-throughput assay was developed based on a UV-visible absorbance shift between 1-methylindole and the C<sub>3</sub>-alkylated product, enabling screening thousands of variants in a single experimental day.<sup>27</sup>

In order to search the so far untouched sequence space of the P411<sub>BM3</sub> for stabilizing mutations that do not abolish the activities of the enzymes, I envisioned applying the mutagenesis and screening approach described below to the P411<sub>BM3</sub> scaffold.

## **5.4 Developing a screen to discover stabilizing mutations in P411<sub>BM3</sub>**

### **variants**

Cytochromes P450 and P411 are named for the wavelength at which they display a characteristic spectral peak in difference spectroscopy measurements in the presence and absence of carbon monoxide (CO-binding assay). Researchers have taken advantage of this behavior to quantify the concentration of properly folded cytochrome P450 in solution<sup>28</sup> and to screen for thermostabilizing mutations.<sup>8</sup> This absorbance-based screen can be performed rapidly in lysate on the order of one sample per second in 96-well plate format with a plate reader, enabling the assessment of the proper folding of thousands of P411<sub>BM3</sub> variants per day.

The ability to only pick colonies known to contain the fully translated protein of interest increases the chances of finding hits because misassembled cloning constructs variants will be excluded from the screen. Toward this goal, I prepared a colony-based pre-screen in form of a direct fluorescence report for protein expression. Colonies which can express full-length proteins can then be picked, cultured, heat challenged, and assayed with CO-binding assay. For the colony-based prescreen, I chose a split-fluorescent protein system, in which the protein of interest is tagged with a short (16 amino acid) fragment of a fluorescent protein. When the tagged protein of interest and the complementary fluorescent protein fragment are

co-expressed, the two fluorescent protein fragments self-assemble, and the protein's fluorescence is enhanced. This approach was developed and used to assay soluble protein expression<sup>29</sup> and has more recently been used to normalize enzyme concentrations in biocatalytic reactions to enable separately analyzing protein expression and specific activity.<sup>30</sup> Fluorescence is only reported in colonies which have a full-length expressed P411<sub>BM3</sub>-C10 as the split-fluorescent protein fragment is on the C-terminus of the protein – any incorrectly assembled or truncated proteins will not fluoresce. As this is a split-fluorescent protein under orthogonal induction conditions rather than a protein fusion, culture growth can be performed with only the target protein expressed, decreasing metabolic load. Sequential induction of the tagged protein of interest followed by the complementary split fluorescent protein fragment can also be used to ensure the protein of interest is soluble upon expression.<sup>29</sup>

As discussed in Chapter 4, mNeonGreen2(1-10) (mNG2(1-10)), complemented by mNG2(11) was chosen as the split-fluorescent protein of choice because it was engineered for decreased background fluorescence in the absence of the complementary 11<sup>th</sup>  $\beta$ -strand.<sup>31</sup> The construct, as available from Addgene, expressed mNG2(1-10) under IPTG induction control; to allow for orthogonal induction control, the gene encoding mNG2(1-10) was subcloned into the pBAD33 vector under arabinose induction control. By C-terminally tagging P411<sub>BM3</sub>-C10 with mNeonGreen2(11) (mNG2(11))<sup>31</sup> and co-expressing mNeonGreen2(1-10) (mNG2(1-10)), expression can be detected via fluorescence on LB-agar plates (Figure 5-1).

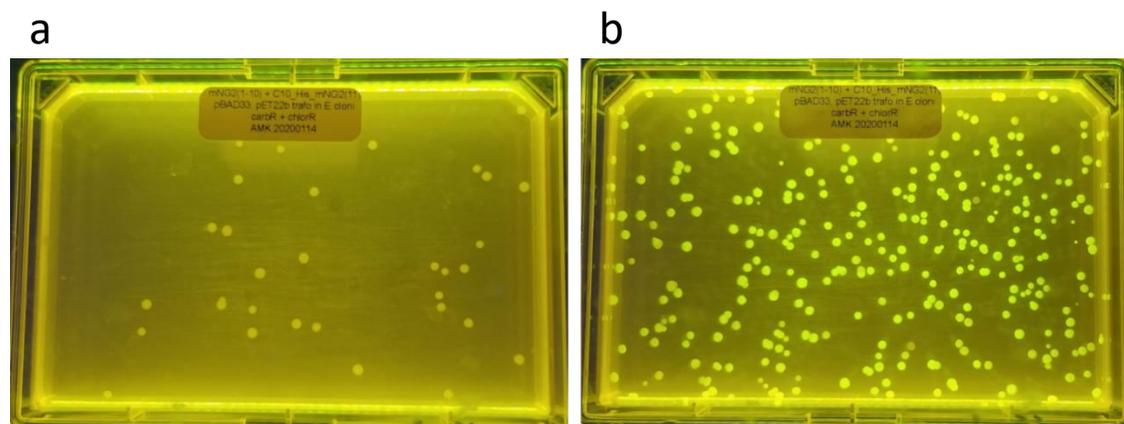


Figure 5-1. Colony-based detection of P411<sub>BM3</sub> expression via split-fluorescent protein tagging. (a) *E. coli* colonies transformed with plasmids encoding mNG2(1-10) and P411<sub>BM3</sub>-C10-mNG2(11), grown on LB-carb-chlor-agar plates. (b) *E. coli* colonies transformed with plasmids encoding mNG2(1-10) and P411<sub>BM3</sub>-C10-mNG2(11), grown on LB-carb-chlor-agar plates, followed by the addition of arabinose and IPTG to induce expression of mNG2(1-10) and P411<sub>BM3</sub>-C10-mNG2(11), respectively.

If the stability of the fluorescent protein is greater than that of a protein to which it is fused, the stability of the protein of interest can be assayed by a heat challenge followed by centrifugation. The denaturation and aggregation of the protein of interest pulls the soluble fluorescent protein into the pellet; the loss of supernate fluorescence is proportional to the fraction of the protein of interest which aggregated.<sup>32</sup> This approach was feasible with our fluorescent system as mNG2(1-10) linked to mNG2(11) is stable at 60 °C for over 1 hour. In contrast, full-length P411<sub>BM3</sub>-C10 has a  $T_M$  of 50 °C (determined by CO binding assay)<sup>28</sup>, and the  $T_M$  of the heme domain of P411<sub>BM3</sub>-CIS had been determined to be 60.6 °C.<sup>13</sup> While P411<sub>BM3</sub>-C10's broad reaction scope has led to its identification as a potential starting point in new projects, its stability — and thus robustness to destabilizing mutations — has played an important part in its further evolution toward new activities.

When testing for loss of fluorescence for P411<sub>BM3</sub>-C10-mNG2 following a heat challenge (at 55 °C based on P411<sub>BM3</sub>-C10's measured  $T_M$ ), however, I was surprised to find that the fluorescence increased by a factor of two (Figure 5-2); fluorescence increased similarly following heat treatments at 75 °C and 95 °C. Further absorbance and fluorescence spectral

scans led to the conclusion that this increased fluorescence is due to the release of flavin mononucleotide (FMN) from P411<sub>BM3</sub> and oxidation in solution.<sup>33</sup> Indeed, the fluorescence is lost upon addition of sodium dithionite (reducing FMN), while addition of sodium dithionite to lysate containing mNG2-tagged *PsEFE* VMM does not attenuate fluorescence. As the absorbance and fluorescence spectra for mNG2 and FMN are completely overlapping, incomplete FMN reduction would confound the mNG2 fluorescence measurements, assaying thermostability with mNG2 is challenging for P411<sub>BM3</sub> constructs still containing the FMN domain. The mNG2 system can still be used as a colony-based pre-screen to a CO-binding assay to determine expression and stability. In addition, in the absence of the fully assembled and fluorescent mNG2, the increased fluorescence from oxidized FMN can be used together with the CO-binding assay to monitor P411<sub>BM3</sub> stability.

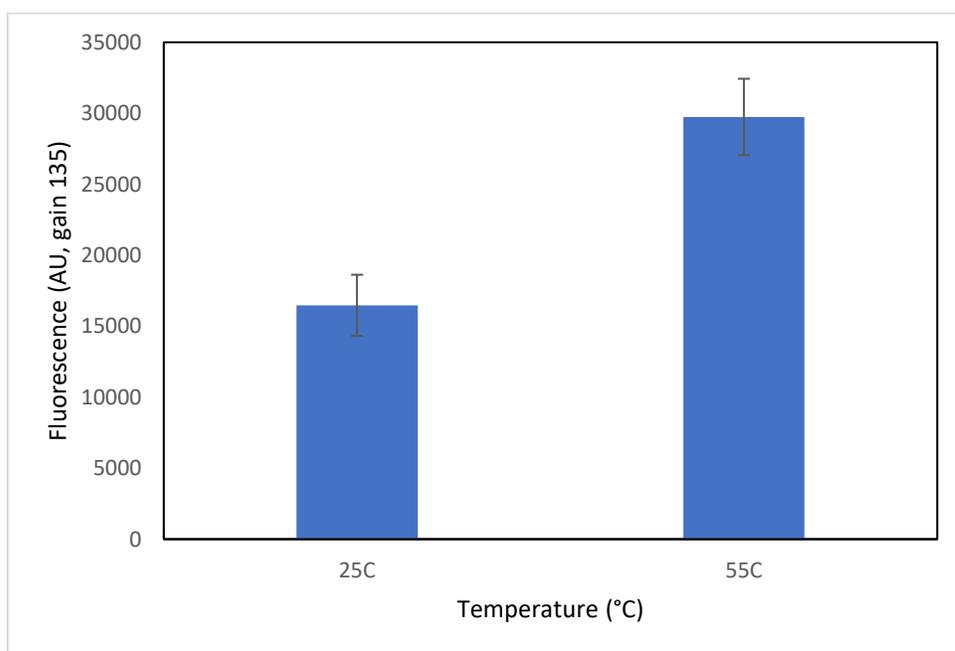


Figure 5-2. Supernate fluorescence of P450<sub>BM3</sub>-C10 tagged with mNG2(11) and co-expressed with the complementary mNG2(1-10) fragment. The increased fluorescence can be attributed to FMN release and oxidation in solution, as it disappears with the addition of sodium dithionite.

To further accelerate directed evolution experiments, the CO-binding assay is fully amenable to robotic automation steps. An automated colony picker and 96-head liquid handlers can perform all transfer operations, from inoculating cultures with single *E. coli* colonies on a Petri dish to preparing clarified cell lysate for CO-binding analysis. With a total analysis time of approximately 5 minutes per 96-well plate and a plate reader with a plate stacker to measure tens of plates in series, the throughput bottleneck becomes the space available to grow 96-deep-well plates of libraries. This combination of automation and high-throughput assay reduces screening effort and enables screening of tens of thousands of variants per week.

## **5.5 Design of tile-based comprehensive site-saturation mutagenesis libraries**

Despite the focus on a subset of active-site residues in P411<sub>BM3</sub> engineering, protein fitness has been shown to be globally encoded.<sup>34</sup> Random mutagenesis and other library designs exploring global sequence space often uncover beneficial amino-acid substitutions at positions unlikely to be prioritized by structure-guided mutagenesis strategies. My goal was to design a library which could be inexpensively constructed, provide sequence coverage of the majority of P411<sub>BM3</sub> amino-acid positions, and have the potential to be used for various other P411 variants and more than a single mutagenesis campaign.

There are myriad methods developed to perform targeted or random mutagenesis. Techniques like random mutagenesis via error-prone PCR will cover the entirety of a protein sequence, but single nucleotide substitution only enables access to a subset of the available amino-acid sequence space. Unlike mutagenesis methods such as error-prone PCR and sequence-saturation mutagenesis (SeSaM)<sup>35</sup>, which introduce single-nucleotide substitutions, libraries with single-codon substitutions enable access to all single steps in protein sequence space. The concept of single-codon substitutions throughout a full gene has been established in the protein engineering community since the early 2000s.<sup>36,37</sup> While the cost to synthesize full genes continues to drop, the price of a commercially purchased whole-gene site-saturation library is still cost-prohibitive (several thousand dollars at minimum),

particularly for academic work. As the cost for DNA synthesis is typically priced per base pair, synthesizing shorter segments of DNA is a straightforward way to reduce the financial strain of library preparation. This cost is further reduced when purchasing these oligonucleotides as a single, pooled sample. The use of such commercial oligo pools to generate sets of site-saturation mutagenesis libraries has been demonstrated with nicking mutagenesis<sup>38</sup> and orthogonal replication (OrthoRep)<sup>19</sup> systems.

Inspired by these commercial oligo pool mutagenesis methods and SCHEMA library design, in which homologous proteins are divided into blocks which are recombined,<sup>39</sup> I envisioned splitting the P411<sub>BM3</sub> primary sequence into several distinct tiles encoding comprehensive single amino-acid substitutions. Using this tile-based mutagenesis approach, 12 mutagenesis tiles were designed spanning the length of P411<sub>BM3</sub>-C10, targeting 504 of the 665 residues (Figure 5-3). The tile set was designed such that each tile was up to 200 nucleotides in length; oligos longer than 200 nucleotides increased the price and turnaround time for the oligo pool and paired-end reads on next-generation sequencing platforms will not give full sequencing coverage if the sequenced region is longer than 200 nucleotides. Conversely, shorter tiles (<120 bp) have reduced yields from PCR amplification and gel extraction protocols. The primer-binding sites between tiles were designed to have secondary structure predictions conducive to efficient PCR amplification, which were then validated experimentally (Supplementary Information, Figure 5-6). The tile positions were also chosen to maximize the number of commonly targeted residues in the primer binding sites between tiles, such that primers complementary to the parent variant can be used to amplify a tile. The oligo pool was designed such that each position targeted for saturation has 19 codon substitutions, using the most abundant codon in *E. coli* to encode each amino acid. The use of 19 single-codon substitutions in place of degenerate codons (22-codon trick or NNK degeneracy) reduces the library size and necessary screening effort without decreasing the protein sequence space covered. The entire library of 12 P411<sub>BM3</sub>-C10 tiles was ordered as a single pool and was generously provided by Twist Biosciences.

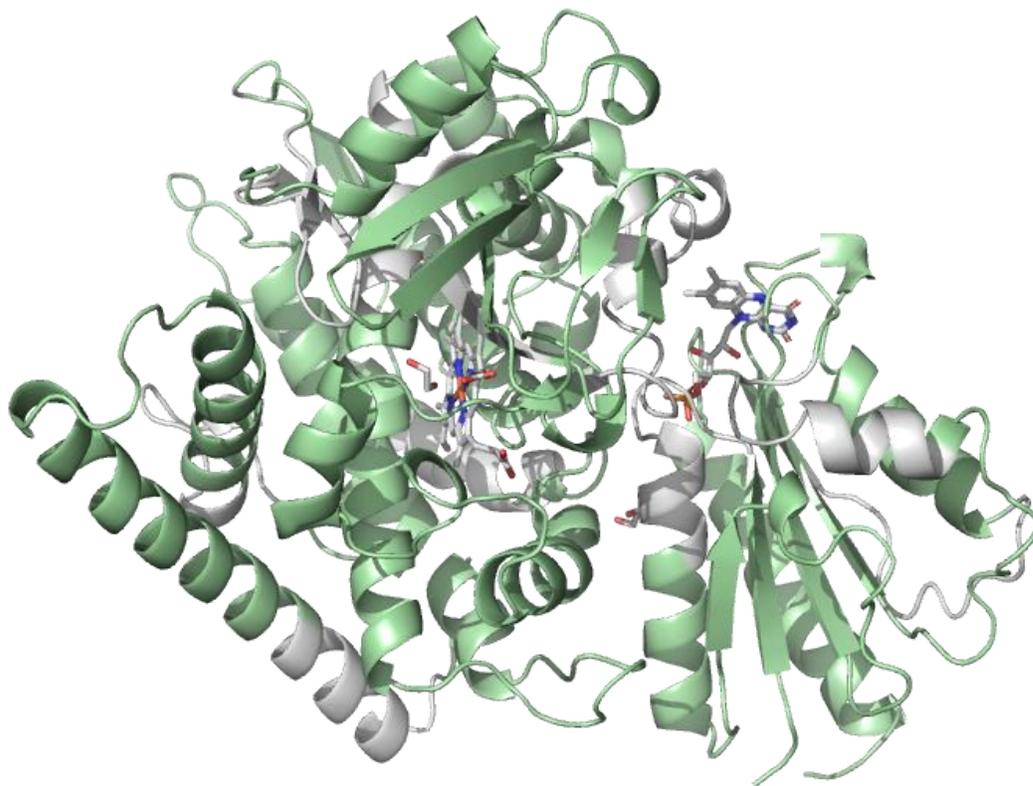


Figure 5-3. Overview of sites targeted in cytochrome P450<sub>BM3</sub> heme and FMN domains (PDB ID: 1BVY). Positions which are mutated as part of a tile are shown in green. The heme domain has 9 tiles and the FMN domain has 3. Together these 12 tiles constitute 504 sites at which all amino-acid substitutions are encoded.

The individual tiles, each containing 39–45 positions at which mutations are programmed, can be amplified out of the synthetic oligo pool via PCR with tile-specific amplification primers (Figure 5-4). Following PCR amplification, the tiles can be assembled with the complementary DNA fragment containing the remainder of the protein and vector via a standard Gibson assembly molecular biology pipeline.<sup>40</sup> As Gibson assembly is amenable to multiple fragment assemblies, multiple degenerate tiles can be combined within the same construct to generate large combinatorial libraries.

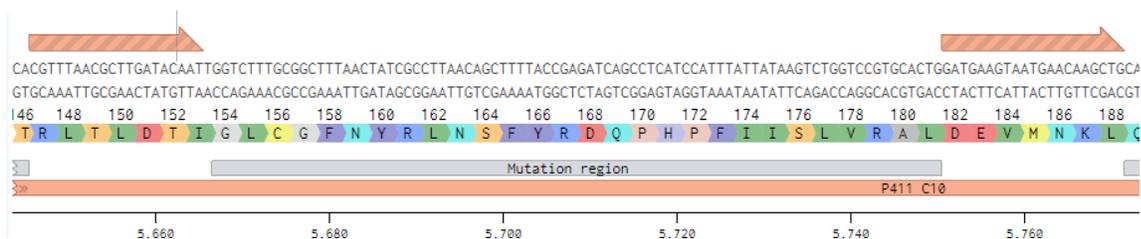


Figure 5-4. Example tile in P411<sub>BM3</sub>-C10. A mutation region of 40-45 codons is flanked by two overlapping regions without programmed mutations. The tile can be amplified via PCR with complementary primers and assembled to generate a plasmid with full-length protein.

Library coverage during screening is always a concern in a directed evolution campaign. With 19 single-codon substitutions per position targeted, the single-tile libraries designed contain 741–855 programmed single amino-acid substitutions. Screening eight 96-well plates (704 colonies after factoring in eight control wells per plate) per tile mutagenesis library (39–45 positions targeted per tile) would result in 50–60% library coverage. Although single site-saturation mutagenesis libraries have been screened at 86% to 98% library coverage, the amount of additional screening required to reach these library coverages is not worth the fractional increase in coverage. In random mutagenesis approaches, one screens only a tiny fraction of the complete library in a round of directed evolution, and machine learning approaches to directed evolution have been carried out collecting data with library coverage of 5% or less.<sup>41</sup> A combinatorial mutagenesis library built from two tiles targeting 39 residues each would contain 549,081 variants (two single-substitution libraries ( $39 \times 19 = 741$ ), combined pairwise ( $741^2 = 549,081$ )), which could still be screened if ultra-high-throughput screening methods are available or low library coverage is acceptable. Combining larger tile libraries or more than two tiles simultaneously would further increase the library size.

When ordering oligonucleotides and larger libraries, the ability to reuse the primers for future mutagenesis on the same or related directed evolution projects becomes a concern. I designed the tile overlap regions to coincide with the amino-acid positions targeted most often for site-saturation mutagenesis libraries in P411<sub>BM3</sub> variants in our lab. This way, the tiles can be

used in mutagenesis projects to complement the generation of site-saturation mutagenesis libraries at commonly targeted positions. This allows for great flexibility: one can either first target the positions most often mutated in P411s, followed by use of the tiles (or a subset of the tile mutations found to be beneficial), or vice versa. The plasmid containing a tile mutagenesis library can be stored in the -20 °C freezer and can be amplified and assembled for reuse with a new P411<sub>BM3</sub> variant.

## 5.6 Identification and recombination of stabilizing mutations

Due to the costs associated with Sanger sequencing, only a handful of the best variants are sequenced in each round of traditional directed evolution experiments. The increased use of machine learning methods for protein engineering applications has led to the push for collecting sequence information for all variants screened, so that machine learning models can be built on data that show which mutations are beneficial and which are detrimental to the target function.<sup>42</sup> Collecting sequence information for all variants tested is also useful to determine the true library coverage when screening a large, defined library at low redundancy. A multiplexed approach to make sequencing the full libraries economically viable has been developed recently in the lab, which would enable sequencing of all tile-based mutagenesis library plates. I have contributed code to extend its use to tile-based mutagenesis libraries: the initial code requires user specified mutation positions, which are not known within a tile mutagenesis library. The code I wrote removes the need to specify the mutation position; codon frequencies at each position are compared to the reference sequence codon, and codons which do not match the reference are reported, making the rapid analysis of mutations feasible.

When beneficial mutations are found, they can be recombined in libraries to find combinations of mutations which are better than individual mutations. There are several options based on library size and complexity. For small libraries, recombination can be performed with standard PCR amplification with degenerate primers, which can be challenging to design with larger libraries. In larger libraries, early recombination was performed with random techniques such as DNA shuffling<sup>43</sup> and staggered extension

processes;<sup>44</sup> as with other library synthesis strategies, the decreased cost of DNA synthesis makes it feasible to synthesize combinatorial libraries which would be too complex to generate with PCR-based approaches. For tile-based mutagenesis libraries, it is possible to generate combinatorial libraries within a single tile, between multiple tiles, or a combination of the two. To generate a recombination library of multiple mutations within the same tile, it is most time efficient to order an additional oligo pool library consisting of each combination of mutations desired within the tile (Figure 5-5a). This tile can then be assembled and screened as a standard tile mutagenesis library. To generate a recombination library with individual mutations from different tiles, one can purchase the synthetic library as above. Alternatively, one can grow a culture for each variant to be combined from the same tile and isolate the plasmid. This mixture of plasmids can be used as a PCR template to prepare that tile library. Multi-fragment Gibson assembly can be used to assemble multiple tile libraries together, whether the libraries are each composed of single mutations (Figure 5-5b) or are combinatorial libraries (Figure 5-5c).

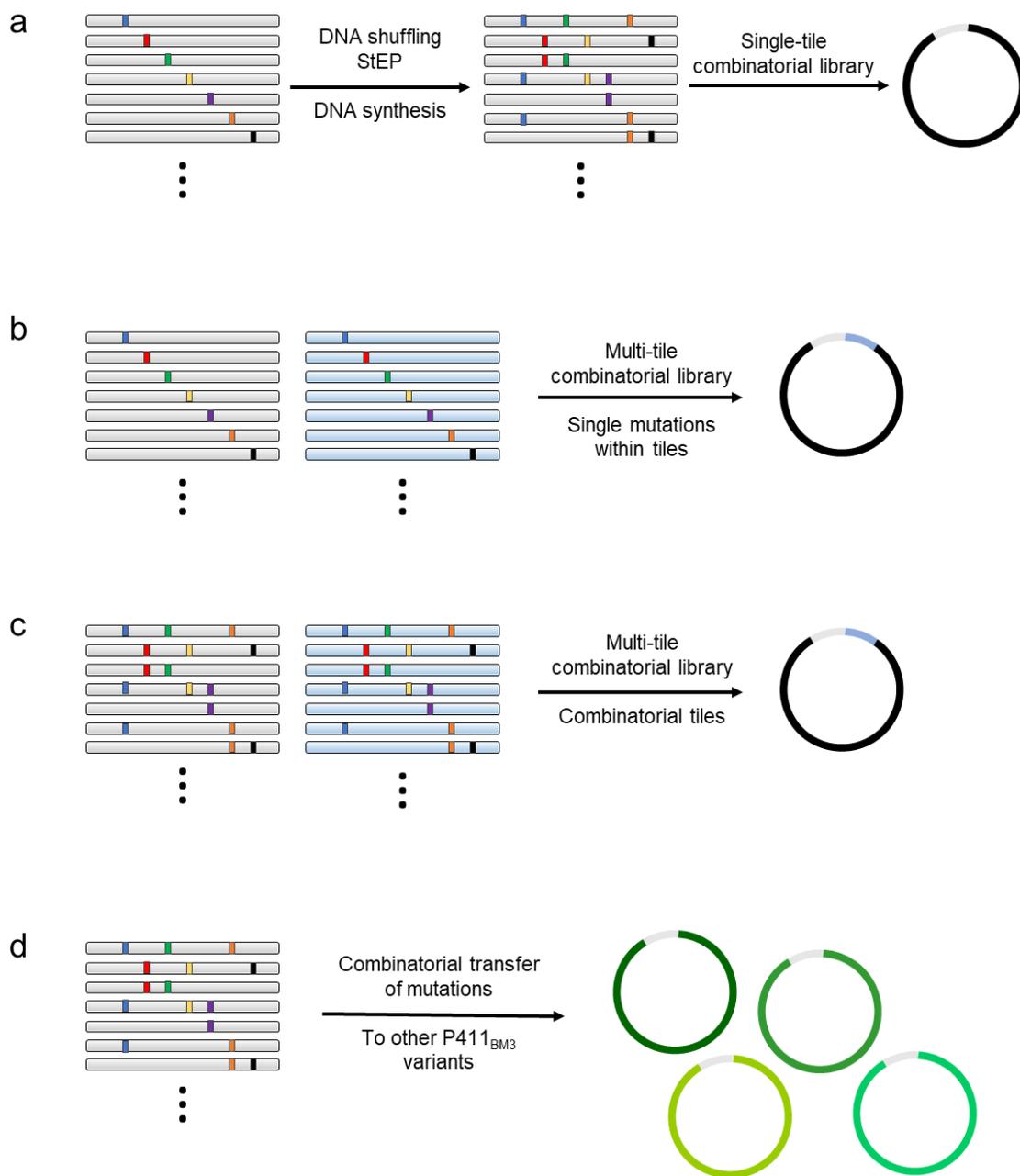


Figure 5-5. Potential recombination methods to generate combinatorial diversity from beneficial mutations found in tile libraries. (a) Recombination within a single region of DNA (e.g. gene or tile) can be performed with several molecular biology methods or by direct DNA synthesis, then assembled into the plasmid. (b) Single mutations in multiple tiles can be assembled to generate combinatorial libraries. (c) Combinatorial tile libraries can also be assembled to generate more complex, multi-tile combinatorial libraries. (d) Libraries (whether single mutations or combinatorial) can be transferred into other P411<sub>BM3</sub> variants if the variant does not have any key mutations within the span of the tile.

Mutations found to enhance protein expression and/or stability can also be transferred into other P411<sub>BM3</sub> variants through the same process; this transfer of mutations could be applied to stabilize variants already engineered for other functions. If there are no other mutations in the P411<sub>BM3</sub> variant in the tile's region, the tile (with any mutations to be transferred) can be amplified and assembled with the P411<sub>BM3</sub> variant (Figure 5-5d). These mutations could also be transferred to P450<sub>BM3</sub> variants using the same approach. As a new mutation in the context of a different P411<sub>BM3</sub> scaffold will not necessarily provide the same fitness enhancement, it is vital to test this new variant – ideally transferring a combination of mutations as a library and screening for enhanced fitness.

Sequencing all variants screened in these libraries will generate a large data set mapping sequence to fitness globally on P411<sub>BM3</sub>-C10. These data will be extremely valuable for machine-learning approaches to directed evolution to both predict beneficial combinations of mutations which were observed in the library and predict variants which were not observed, but could display enhanced fitness. As the anticipated library coverage is 50% and models can be trained on 5% or less of the data,<sup>41</sup> it is possible to compare machine learning predictions to a large experimental test data set. This ground truth dataset could therefore be used to compare the accuracy of different models on global mutational fitness predictions.

## 5.7 Conclusions

Protein stability is a crucial parameter for the evolution of enzymes for synthetic applications. The biocatalytic platform based on P411<sub>BM3</sub> variants could be further enhanced by focusing on stabilizing the P411<sub>BM3</sub> scaffold. In this chapter, I have shown a tile-based mutagenesis library design which enables the exploration of largely untapped sequence space in P411<sub>BM3</sub> and demonstrated fluorescent and spectrophotometric assays which can be used to efficiently screen the library for mutations which enhance stability and expression. I lay out methods which could be used to recombine these mutations and transfer beneficial mutations to other P411<sub>BM3</sub> variants and the potential utility of the sequence-fitness data set to benchmark different machine-learning models for protein fitness prediction. Together, these approaches

and methods have the potential to improve the robustness of the P411<sub>BM3</sub> platform and ensure that it can continue to be used to develop more remarkable new-to-nature biotransformations.

## 5.8 Supplementary information for Chapter 5

### Materials

Oligonucleotide primers were obtained from IDT DNA. Synthetic oligonucleotide pool libraries were obtained from Twist Biosciences. PCRs were run with Phusion® High-Fidelity PCR Kit (New England Biolabs). Gibson assembly mix<sup>40</sup> is prepared with isothermal master mix in-house, and enzymes T5 exonuclease, Phusion® DNA polymerase, and Taq DNA ligase purchased from New England Biolabs.

### Cloning

Plasmids were used to transform *E. coli* BL21(DE3) cells (Lucigen) by electroporation. SOC medium (0.75 mL) was added, and the cells were incubated at 37 °C for 45 minutes before being plated on Luria-Bertani medium (Research Products International) agar plates supplemented with carbenicillin (100 µg mL<sup>-1</sup>, LB-carb-agar). To prepare *E. coli* harboring two plasmids, *E. coli* BL21(DE3) cells were co-transformed with both plasmids; future library transformations should be performed using competent *E. coli* BL21(DE3) cells already harboring the second plasmid to improve efficiency. Plasmids were isolated from stationary-phase cultures by miniprep (Qiagen) and Sanger sequencing was performed by Laragen, Inc. (Culver City, CA) using T7 promoter and T7 terminator primers.

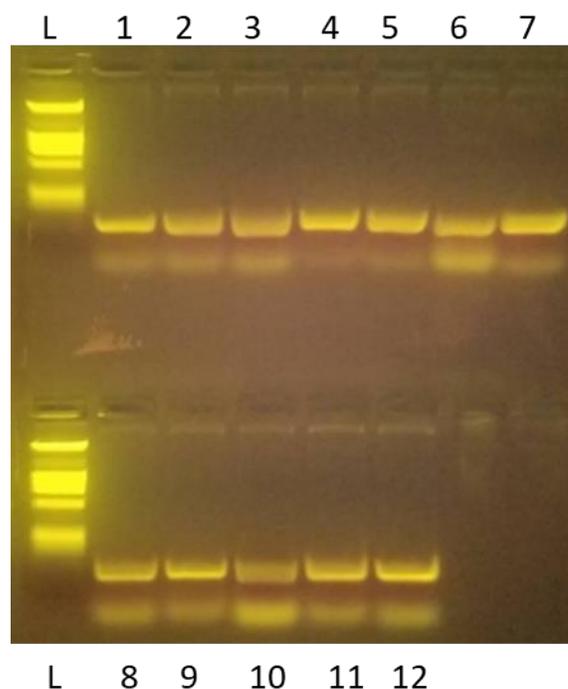


Figure 5-6. Agarose gel of PCR-amplified tile regions of P411<sub>BM3</sub>. L: DNA ladder (1 kbp ladder, NEB), 1-12: PR products from P411<sub>BM3</sub>-C10 tile regions. The larger molecular-weight band is the tile fragment; the lower band is primer dimer. Gel excision of the DNA fragment and Gibson assembly with linear DNA containing the expression vector and the remainder of the gene encoding P411<sub>BM3</sub>-C10 can be used to transform *E. coli*.

### Protein expression

Starter cultures of LB-amp were inoculated from a single *E. coli* colony on an agar plate harboring a plasmid encoding the protein of interest with a sterile toothpick and grown overnight to stationary phase at 37 °C. Expression cultures of Terrific Broth (Research Products International) supplemented with ampicillin (100 mg L<sup>-1</sup>, TB-amp) were inoculated from the starter cultures (1% v/v) and shaken at 37 °C and 160 rpm in a Multitron Infors incubator. When the expression cultures reached OD<sub>600</sub> ~ 0.8 (2–3 hours), they were cooled on ice for 20 minutes. Protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.5 mM). Cultures were incubated at 22 °C and 110 rpm overnight (16–24 hours). Cells were pelleted by centrifugation (5000×g, 10 minutes).

### **Agar plate induction of protein expression**

*E. coli* BL21(DE3) cells were co-transformed with pET22b\_P411<sub>BM3</sub>C10\_mNG2(11) and pBAD33\_mNG2(1-10) plasmids and grown on LB agar supplemented with carbenicillin (100  $\mu\text{g mL}^{-1}$ ) and chloramphenicol (25  $\mu\text{g mL}^{-1}$ ) (LB-carb-chlor-agar) at 37 °C for 12–16 hours. Following colony formation, the plates were stored at 4 °C until induction. Using an Echo acoustic liquid handler (Labcyte), induction media (18%(w/w) L-arabinose, 100 mM IPTG) was patterned across the single-well LB-agar plate in 1536 applications of 50-nL droplets. The plates were then incubated at room temperature for 24 hours. Colony fluorescence was observed using a blue-light transilluminator.

### **Cell lysis**

Cells grown in 96-well plates were frozen at -20 °C for at least 24 hours and rapidly thawed in room-temperature water. Lysis buffer (400  $\mu\text{L}$  of 1 $\times$ M9-N buffer supplemented with 1 mg  $\text{mL}^{-1}$  hen egg white lysozyme and 0.1 mg  $\text{mL}^{-1}$  Dnase I) was added, and the plate was vortexed until all cell pellets were visually resuspended. The plate was incubated at 37 °C for 1 hour, then centrifuged (5000 $\times g$ , 10 minutes) to clarify the lysate.

### **Thermostability assay**

Purified protein or clarified cell lysate were incubated in a water bath set to the specified temperature for 30 or 60 minutes. Single samples were incubated in 1.7-mL plastic tubes. 96-Well plate format samples were incubated using semi-skirted PCR plates to facilitate even heat transfer across the wells. Following heat incubation, the samples were centrifuged (10 minutes; single samples at 20817 $\times g$ , 96-well plates at 5000 $\times g$ ). The supernate was used for CO-binding assays.

### **Carbon monoxide binding assay**

The carbon monoxide binding assay was performed as described previously.<sup>11</sup> Briefly, sodium dithionite (40  $\mu\text{L}$  of 150 mM solution, 30 mM final concentration) was added to clarified cell lysate containing P411<sub>BM3</sub> variants (160  $\mu\text{L}$ ). Absorbance measurements at 411 nm and 490 nm were collected using a Spark multimode microplate reader (Tecan), and the

microtiter plates were placed in a vacuum chamber. The chamber was sealed, evacuated to approximately -15 inHg, filled with carbon monoxide gas to atmospheric pressure, and incubated for 30 min. The plates were then removed, and the absorbance at 411 nm and 490 nm was again recorded using the plate reader. The difference spectra were then used to determine the P411<sub>BM3</sub> lysate concentration in each well.<sup>28</sup>

Table 5-1. P411<sub>BM3</sub>-C10 tile locations. Indices are based on nucleotide sequence and are 0-indexed (i.e., the first nucleotide of the start-codon is A0). Tile start and end indices include the overlap regions in which primer binding sites were designed. Mutation indices show the region in which codon substitutions are introduced

Tile name	Tile start index	Tile end index	Mutation start index	Mutation end index	Tile length	Mutation region length	Number of codons mutated
P411C10_tile_01	1	199	33	167	199	135	45
P411C10_tile_02	168	363	201	332	196	132	44
P411C10_tile_03	334	517	366	482	184	117	39
P411C10_tile_04	484	681	519	647	198	129	43
P411C10_tile_05	648	845	684	806	198	123	41
P411C10_tile_06	808	998	846	968	191	123	41
P411C10_tile_07	970	1167	1002	1139	198	138	46
P411C10_tile_08	1142	1338	1170	1292	197	123	41
P411C10_tile_09	1295	1489	1341	1457	195	117	39
P411C10_tile_10	1459	1652	1491	1616	194	126	42
P411C10_tile_11	1619	1802	1653	1769	184	117	39
P411C10_tile_12	1770	1955	1803	1934	186	132	44

Table 5-2. Oligonucleotide sequences to amplify the 12 tile-based mutagenesis libraries for P411<sub>BM3</sub>-C10.

Tile name	Primer direction	Amplification primer sequence
P411C10_tile_01	Forward	TGACAATTAAGAAATGCCTCAGCCAAAAACG
P411C10_tile_02	Forward	CGTCTAATTAAGAAGCATGCGATGAATCACG
P411C10_tile_03	Forward	CAATGAAAGGCTATCATGCGAGTATGGTGC
P411C10_tile_04	Forward	GCCTTAACAGCTTTTACCGAGATCAGCC
P411C10_tile_05	Forward	GTAGATAAAATTATTGCAGATCGCAAAGCAAGGG
P411C10_tile_06	Forward	CAAGTGGTCTTTTATCATTGCGCTGTATTTCTTAGTG
P411C10_tile_07	Forward	GCTTATGGCCAACGGTTCCTTATTTTCCC
P411C10_tile_08	Forward	AAATCCAAGTGCATTCCGCAGCATG
P411C10_tile_09	Forward	CGATATTAAGAAGACTGCAGACGTTAAAACCTAAAGGC
P411C10_tile_10	Forward	ACGGTTCAAATATGGGTACCGCTGAAGGA
P411C10_tile_11	Forward	TCCGCCTGATAACGCAAAGCAATTTGTC
P411C10_tile_12	Forward	GCCGCTAAAGGGGCAGAAAACATCG
P411C10_tile_01	Reverse	CGTGATTCATCGCATGCTTCTTTAATTAGACG
P411C10_tile_02	Reverse	CGACCATACTCGCATGATAGCCTTTCATTG
P411C10_tile_03	Reverse	GGATGAGGCTGATCTCGGTAAAAGCTGTAAAG
P411C10_tile_04	Reverse	CCCTTGCTTTGCGATCTGCAATAATTTTATCTAC
P411C10_tile_05	Reverse	CACTAAGAAATACAGCGCAAATGATAAAAAGACCACTT G
P411C10_tile_06	Reverse	GGAAAATAAGGAACCGTTGGCCATAAGCG
P411C10_tile_07	Reverse	CATGCTGCGGAATCGCACTTGGATTT
P411C10_tile_08	Reverse	CCACAAAGCCTTTAGGTTTAAACGTCTGCAG
P411C10_tile_09	Reverse	GTTCCTTCAGCGGTACCCATATTTGAACC
P411C10_tile_10	Reverse	CCAGTCGACAAATTGCTTTGCGTTATCAGG
P411C10_tile_11	Reverse	GCGGTCAGCGATGTTTTCTGCC
P411C10_tile_12	Reverse	CGCGGCGCTGTCGACAAATTG

## 5.9 References for Chapter 5

- (1) Arnold, F. H. Innovation by Evolution: Bringing New Chemistry to Life (Nobel Lecture). *Angew. Chem., Int. Ed.* **2019**, *58*, 14420–14426. <https://doi.org/10.1002/anie.201907729>.
- (2) Tokuriki, N.; Tawfik, D. S. Stability Effects of Mutations and Protein Evolvability. *Curr. Opin. Struct. Biol.* **2009**, *19*, 596–604. <https://doi.org/10.1016/j.sbi.2009.08.003>.
- (3) Fasan, R.; Meharena, Y. T.; Snow, C. D.; Poulos, T. L.; Arnold, F. H. Evolutionary History of a Specialized P450 Propane Monooxygenase. *J. Mol. Biol.* **2008**, *383*, 1069–1080. <https://doi.org/10.1016/j.jmb.2008.06.060>.
- (4) Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H. Protein Stability Promotes Evolvability. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 5869–5874.

- <https://doi.org/10.1073/pnas.0510098103>.
- (5) Besenmatter, W.; Kast, P.; Hilvert, D. Relative Tolerance of Mesostable and Thermostable Protein Homologs to Extensive Mutation. *Proteins: Struct., Funct., Bioinf.* **2006**, *66*, 500–506. <https://doi.org/10.1002/prot.21227>.
  - (6) Cowan, D. A. Thermophilic Proteins: Stability and Function in Aqueous and Organic Solvents. *Comp. Biochem. Physiol. Part A Physiol.* **1997**, *118*, 429–438. [https://doi.org/10.1016/S0300-9629\(97\)00004-2](https://doi.org/10.1016/S0300-9629(97)00004-2).
  - (7) Polizzi, K. M.; Bommaris, A. S.; Broering, J. M.; Chaparro-Riggers, J. F. Stability of Biocatalysts. *Curr. Opin. Chem. Biol.* **2007**, *11*, 220–225. <https://doi.org/10.1016/j.cbpa.2007.01.685>.
  - (8) Romero, P. A.; Krause, A.; Arnold, F. H. Navigating the Protein Fitness Landscape with Gaussian Processes. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E193–E201. <https://doi.org/10.1073/pnas.1215251110>.
  - (9) Goldenzweig, A.; Goldsmith, M.; Hill, S. E.; Gertman, O.; Laurino, P.; Ashani, Y.; Dym, O.; Unger, T.; Albeck, S.; Prilusky, J.; Lieberman, R. L.; Aharoni, A.; Silman, I.; Sussman, J. L.; Tawfik, D. S.; Fleishman, S. J. Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Mol. Cell* **2016**, *63*, 337–346. <https://doi.org/10.1016/j.molcel.2016.06.012>.
  - (10) Bommaris, A. S.; Paye, M. F. Stabilizing Biocatalysts. *Chem. Soc. Rev.* **2013**, *42*, 6534. <https://doi.org/10.1039/c3cs60137d>.
  - (11) Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. Olefin Cyclopropanation via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes. *Science* **2013**, *339*, 307–310. <https://doi.org/10.1126/science.1231434>.
  - (12) Vatsis, K. P.; Peng, H.-M.; Coon, M. J. Replacement of Active-Site Cysteine-436 by Serine Converts Cytochrome P450 2B4 into an NADPH Oxidase with Negligible Monooxygenase Activity. *J. Inorg. Biochem.* **2002**, *91*, 542–553. [https://doi.org/10.1016/S0162-0134\(02\)00438-5](https://doi.org/10.1016/S0162-0134(02)00438-5).
  - (13) Coelho, P. S.; Wang, Z. J.; Ener, M. E.; Baril, S. A.; Kannan, A.; Arnold, F. H.; Brustad, E. M. A Serine-Substituted P450 Catalyzes Highly Efficient Carbene Transfer to Olefins in Vivo. *Nat. Chem. Biol.* **2013**, *9*, 485–487. <https://doi.org/10.1038/nchembio.1278>.
  - (14) Brandenburg, O. F.; Fasan, R.; Arnold, F. H. Exploiting and Engineering Hemoproteins for Abiological Carbene and Nitrene Transfer Reactions. *Curr. Opin. Biotechnol.* **2017**, *47*, 102–111. <https://doi.org/10.1016/j.copbio.2017.06.005>.
  - (15) Chen, K.; Arnold, F. H. Engineering New Catalytic Activities in Enzymes. *Nat. Catal.* **2020**, *3*, 203–213. <https://doi.org/10.1038/s41929-019-0385-5>.
  - (16) Zhang, R. K.; Chen, K.; Huang, X.; Wohlschlag, L.; Renata, H.; Arnold, F. H. Enzymatic Assembly of Carbon–Carbon Bonds via Iron-Catalysed Sp<sup>3</sup> C–H Functionalization. *Nature* **2019**, *565*, 67–72. <https://doi.org/10.1038/s41586-018-0808-5>.
  - (17) Vottero, E.; Rea, V.; Lastdrager, J.; Honing, M.; Vermeulen, N. P. E.; Commandeur, J. N. M. Role of Residue 87 in Substrate Selectivity and Regioselectivity of Drug-Metabolizing Cytochrome P450 CYP102A1 M11. *J. Biol. Inorg. Chem.* **2011**, *16*, 899–912. <https://doi.org/10.1007/s00775-011-0789-4>.
  - (18) Esvelt, K. M.; Carlson, J. C.; Liu, D. R. A System for the Continuous Directed

- Evolution of Biomolecules. *Nature* **2011**, *472*, 499–503.  
<https://doi.org/10.1038/nature09929>.
- (19) Ravikumar, A.; Arzumanyan, G. A.; Obadi, M. K. A.; Javanpour, A. A.; Liu, C. C. Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell* **2018**, *175*, 1946–1957.e13.  
<https://doi.org/10.1016/j.cell.2018.10.021>.
- (20) Cormack, B. P.; Valdivia, R. H.; Falkow, S. FACS-Optimized Mutants of the Green Fluorescent Protein (GFP). *Gene* **1996**, *173*, 33–38. [https://doi.org/10.1016/0378-1119\(95\)00685-0](https://doi.org/10.1016/0378-1119(95)00685-0).
- (21) Baret, J.-C.; Miller, O. J.; Taly, V.; Ryckelynck, M.; El-Harrak, A.; Frenz, L.; Rick, C.; Samuels, M. L.; Hutchison, J. B.; Agresti, J. J.; Link, D. R.; Weitz, D. A.; Griffiths, A. D. Fluorescence-Activated Droplet Sorting (FADS): Efficient Microfluidic Cell Sorting Based on Enzymatic Activity. *Lab Chip* **2009**, *9*, 1850–1858. <https://doi.org/10.1039/b902504a>.
- (22) Gielen, F.; Hours, R.; Emond, S.; Fischlechner, M.; Schell, U.; Hollfelder, F. Ultrahigh-Throughput-Directed Enzyme Evolution by Absorbance-Activated Droplet Sorting (AADS). *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E7383–E7389. <https://doi.org/10.1073/pnas.1606927113>.
- (23) Schätzle, S.; Höhne, M.; Redestad, E.; Robins, K.; Bornscheuer, U. T. Rapid and Sensitive Kinetic Assay for Characterization of  $\omega$ -Transaminases. *Anal. Chem.* **2009**, *81*, 8244–8248. <https://doi.org/10.1021/ac901640q>.
- (24) Buller, A. R.; Brinkmann-Chen, S.; Romney, D. K.; Herger, M.; Murciano-Calles, J.; Arnold, F. H. Directed Evolution of the Tryptophan Synthase  $\beta$ -Subunit for Stand-Alone Function Recapitulates Allosteric Activation. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 14599–14604. <https://doi.org/10.1073/pnas.1516401112>.
- (25) Klingenberg, M. Nicotinamide-Adenine Dinucleotides (NAD, NADP, NADH, NADPH). In *Methods of Enzymatic Analysis*; Elsevier, 1974; pp 2045–2072. <https://doi.org/10.1016/B978-0-12-091304-6.50060-4>.
- (26) Welch, C. J.; Gong, X.; Schafer, W.; Pratt, E. C.; Brkovic, T.; Pirzada, Z.; Cuff, J. F.; Kosjek, B. MISER Chromatography (Multiple Injections in a Single Experimental Run): The Chromatogram Is the Graph. *Tetrahedron: Asymmetry* **2010**, *21*, 1674–1681. <https://doi.org/10.1016/j.tetasy.2010.05.029>.
- (27) Brandenburg, O. F.; Chen, K.; Arnold, F. H. Directed Evolution of a Cytochrome P450 Carbene Transferase for Selective Functionalization of Cyclic Compounds. *J. Am. Chem. Soc.* **2019**, *141*, 8989–8995. <https://doi.org/10.1021/jacs.9b02931>.
- (28) Otey, C. R. High-Throughput Carbon Monoxide Binding Assay for Cytochromes P450. In *Directed Enzyme Evolution*; Humana Press: New Jersey, 2003; Vol. 230, pp 137–140. <https://doi.org/10.1385/1-59259-396-8:137>.
- (29) Cabantous, S.; Waldo, G. S. In Vivo and in Vitro Protein Solubility Assays Using Split GFP. *Nat. Methods* **2006**, *3*, 845–854. <https://doi.org/10.1038/nmeth932>.
- (30) Santos-Aberturas, J.; Dörr, M.; Waldo, G. S.; Bornscheuer, U. T. In-Depth High-Throughput Screening of Protein Engineering Libraries by Split-GFP Direct Crude Cell Extract Data Normalization. *Chem. Biol.* **2015**, *22*, 1406–1414. <https://doi.org/10.1016/j.chembiol.2015.08.014>.
- (31) Feng, S.; Sekine, S.; Pessino, V.; Li, H.; Leonetti, M. D.; Huang, B. Improved Split

- Fluorescent Proteins for Endogenous Protein Labeling. *Nat. Commun.* **2017**, *8*, 370. <https://doi.org/10.1038/s41467-017-00494-8>.
- (32) Moreau, M. J. J.; Morin, I.; Schaeffer, P. M. Quantitative Determination of Protein Stability and Ligand Binding Using a Green Fluorescent Protein Reporter System. *Mol. BioSyst.* **2010**, *6*, 1285. <https://doi.org/10.1039/c002001j>.
- (33) Macheroux, P. UV-Visible Spectroscopy as a Tool to Study Flavoproteins. In *Flavoprotein Protocols*; Humana Press: New Jersey, 1999; Vol. 131, pp 1–8. <https://doi.org/10.1385/1-59259-266-X:1>.
- (34) Wrenbeck, E. E.; Azouz, L. R.; Whitehead, T. A. Single-Mutation Fitness Landscapes for an Enzyme on Multiple Substrates Reveal Specificity Is Globally Encoded. *Nat. Commun.* **2017**, *8*, 15695. <https://doi.org/10.1038/ncomms15695>.
- (35) Wong, T. S. Sequence Saturation Mutagenesis (SeSaM): A Novel Method for Directed Evolution. *Nucleic Acids Res.* **2004**, *32*, 26e – 26. <https://doi.org/10.1093/nar/gnh028>.
- (36) Kretz, K. A.; Richardson, T. H.; Gray, K. A.; Robertson, D. E.; Tan, X.; Short, J. M. Gene Site Saturation Mutagenesis: A Comprehensive Mutagenesis Approach. In *Methods in Enzymology*; 2004; pp 3–11. [https://doi.org/10.1016/S0076-6879\(04\)88001-7](https://doi.org/10.1016/S0076-6879(04)88001-7).
- (37) Gray, K. A.; Richardson, T. H.; Kretz, K.; Short, J. M.; Bartnek, F.; Knowles, R.; Kan, L.; Swanson, P. E.; Robertson, D. E. Rapid Evolution of Reversible Denaturation and Elevated Melting Temperature in a Microbial Haloalkane Dehalogenase. *Adv. Synth. Catal.* **2001**, *343*, 607–617. [https://doi.org/10.1002/1615-4169\(200108\)343:6/7<607::AID-ADSC607>3.3.CO;2-D](https://doi.org/10.1002/1615-4169(200108)343:6/7<607::AID-ADSC607>3.3.CO;2-D).
- (38) Medina-Cucurella, A. V.; Steiner, P. J.; Faber, M. S.; Beltrán, J.; Borelli, A. N.; Kirby, M. B.; Cutler, S. R.; Whitehead, T. A. User-Defined Single Pot Mutagenesis Using Unamplified Oligo Pools. *Protein Eng. Des. Sel.* **2019**, *32*, 41–45. <https://doi.org/10.1093/protein/gzz013>.
- (39) Silberg, J. J.; Endelman, J. B.; Arnold, F. H. SCHEMA-Guided Protein Recombination. In *Methods in Enzymology*; 2004; Vol. 388, pp 35–42. [https://doi.org/10.1016/S0076-6879\(04\)88004-2](https://doi.org/10.1016/S0076-6879(04)88004-2).
- (40) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat. Methods* **2009**, *6*, 343–345. <https://doi.org/10.1038/nmeth.1318>.
- (41) Wu, Z.; Kan, S. B. J.; Lewis, R. D.; Wittmann, B. J.; Arnold, F. H. Machine Learning-Assisted Directed Protein Evolution with Combinatorial Libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 8852–8858. <https://doi.org/10.1073/pnas.1901979116>.
- (42) Yang, K. K.; Wu, Z.; Arnold, F. H. Machine-Learning-Guided Directed Evolution for Protein Engineering. *Nat. Methods* **2019**, *16*, 687–694. <https://doi.org/10.1038/s41592-019-0496-6>.
- (43) Stemmer, W. P. C. Rapid Evolution of a Protein in Vitro by DNA Shuffling. *Nature* **1994**, *370*, 389–391. <https://doi.org/10.1038/370389a0>.
- (44) Zhao, H.; Giver, L.; Shao, Z.; Affholter, J. A.; Arnold, F. H. Molecular Evolution by Staggered Extension Process (StEP) in Vitro Recombination. *Nat. Biotechnol.* **1998**, *16*, 258–261. <https://doi.org/10.1038/nbt0398-258>.

## APPENDIX A: SEQUENCE INFORMATION

The sequences of expression vectors (DNA sequence) and proteins (DNA and amino-acid sequence) used in this study are deposited in csv format on CaltechDATA. The oligonucleotide primer sequences used in this work are also deposited to CaltechDATA. These data can be found at DOI: 10.22002/D1.1437.

Table A-1. Proteins (and expression vectors) for which sequences are given in the worksheet deposited to CaltechDATA (DOI: 10.22002/D1.1437). The UniProt ID for the wild-type protein is listed for each entry.

Sequence_name	WT_UniProt
pET22_vector_sequence	n_a
pSUMO_empty_vector	n_a
pBAD33_empty_vector	n_a
P411 <sub>BM3</sub> -CIS	P14779
P411 <sub>BM3</sub> -C10	P14779
P411 <sub>BM3</sub> -CIS	P14779
P411 <sub>BM3</sub> -UA	P14779
P411 <sub>BM3</sub> -UA-V87C	P14779
P411 <sub>BM3</sub> -UA-V87F	P14779
SUMO-tagged P411 <sub>BM3</sub> -UA-V87C heme domain only	P14779
SUMO-tagged P411 <sub>BM3</sub> -UA-V87F heme domain only	P14779
P411 <sub>BM3</sub> _C10_His_mNG2(11)	P14779
P411 <sub>BM3</sub> _C10_mNG2(11)	P14779
His_mNG2(1-10)	A0A1S4NYF2
<i>RmaNOD</i> Y32G	D0MGT2
<i>RmaNOD</i> wild type (WT)	D0MGT2
<i>RmaNOD</i> Q52V	D0MGT2
SUMO-tagged <i>RmaNOD</i> Q52V	D0MGT2
<i>RmaNOD</i> Q52A	D0MGT2
<i>RmaNOD</i> Y32T Q52A	D0MGT2
<i>RmaNOD</i> Y32T Y39H L48R Q52A R79W (THRAW)	D0MGT2
<i>RmaNOD</i> Q52A L101N	D0MGT2
<i>RmaNOD</i> Q52A 60H L101N	D0MGT2
<i>RmaNOD</i> Q52A 60H L101N I105M	D0MGT2
<i>RmaNOD</i> L20W Q52A L56I 60H L101N I105M (WAIHNM)	D0MGT2
<i>ApePgb</i> WT	Q9YFF4
<i>ApePgb</i> Y60G	Q9YFF4
<i>ApePgb</i> W59A Y60G	Q9YFF4
<i>ApePgb</i> W59A Y60G F145W (AGW)	Q9YFF4
<i>ApePgb</i> Y60G F145N	Q9YFF4
SUMO-tagged <i>ApePgb</i> AGW	Q9YFF4

Sequence_name	WT_UniProt
<i>PsEFE</i> WT	P32021
<i>PsEFE</i> C317M	P32021
<i>PsEFE</i> H189A D191A	P32021
<i>PsEFE</i> R171V F314M C317M	P32021
<i>PsEFE</i> R171V R277H F314M C317M	P32021
SUMO-tagged <i>PsEFE</i> R171V F314M C317M	P32021
SUMO-tagged <i>PsEFE</i> R171V R277H F314M C317M	P32021
<i>PsEFE</i> T97M R171L R277H F314M C317M	P32021
<i>PsEFE</i> T97M R171L I186L R277H F314M C317M	P32021
<i>PsEFE</i> T97M R171V R277H R314L C317M	P32021
<i>Streptomyces vinaceus</i> Arginine hydroxylase	Q6WZB0
<i>Streptomyces muensis</i> Leucine hydroxylase	A0A0E3URV8
<i>Gluconobacter oxydans</i> Leucine hydroxylase	Q5FQD2
<i>Escherichia coli</i> Taurine dioxygenase	P37610
<i>Arabidopsis thaliana</i> Anthocyanidin synthase	Q96323
<i>Streptomyces sp.</i> 2-aminobutyric acid chlorinase	D0VX22
<i>Paracoccus denitrificans</i> Amicyanin	P22364
<i>Chloroflexus aurantiacus</i> Auracyanin D	A9WFS1
<i>Pseudomonas aeruginosa</i> Azurin	P00282
<i>Pseudomonas putida</i> Azurin	P34097
<i>Methylomonas sp.</i> Azurin iso-2	P12335
<i>Thermus thermophilus</i> Laccase	Q72HW2
<i>Nitrosomonas europaea</i> Nitrosocyanin	Q820S6
<i>Cucumis sativus</i> Plantacyanin	P00303
<i>Chlamydomonas reinhardtii</i> Plastocyanin	P18068
<i>Dryopteris crassirhizoma</i> Plastocyanin	Q7SIB8
<i>Phormodium laminosum</i> Plastocyanin	Q51883
<i>Methylbacterium extorquens</i> Pseudoazurin	P04171
<i>Rhizobium meliloti</i> Pseudoazurin	Q92M26
<i>Alcaligenes faecalis</i> Rusticyanin	P0C918
<i>Thermotoga maritima</i> Cupin-like protein	Q9X1H0
<i>Pyrobaculum aerophilum</i> Superoxide dismutase	O93724
<i>Thermus thermophilus</i> Superoxide dismutase	P61503
<i>Bacillus anthracis</i> Acireductone dioxygenase	Q81MI9
<i>Streptomyces sp.</i> Quercetinase	A2VA43
<i>Alcaligenes sp.</i> dihydroxyacetophenone dioxygenase	Q9REI7
<i>Escherichia coli</i> 2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexadiene-1-carboxylate synthase	P17109
<i>Escherichia coli</i> Organomercurial Lyase MerB	P77072
<i>Escherichia coli</i> Lactaldehyde:propanediol oxidoreductase	P0A9S1
<i>Escherichia coli</i> Glyoxalase	P0AC81
<i>Hapalosiphon welwitschii</i> 2-OG Halogenase	A0A067YX61
<i>Leishmania major</i> Glyoxalase I	Q68RJ8
<i>Mus musculus</i> Acireductone dioxygenase	Q99JT9

Sequence_name	WT_UniProt
<i>Mycobacterium thermoresistibile</i> Ergothioneine-biosynthetic sulfoxide synthase	G7CFI3
<i>Nitrosopumilus maritimus</i> Purple Cupredoxin	A9A2G4
<i>Photorhabdus laumondii</i> cupin	Q7MZL9
<i>Propionibacterium freudenreichii</i> transcarboxylase	Q70AC7
<i>Pseudomonas cichorii</i> D-tagatose 3-epimerase	O50580
<i>Pyrococcus horikoshii</i> superoxide reductase	O58810
<i>Rhodopseudomonas palustris</i> cupin	Q6N272
<i>Thermotoga maritima</i> Amidohydrolase	Q9X034
<i>Thermotoga maritima</i> Oxalate decarboxylase	Q9X113
<i>Aspergillus nidulans</i> Fe(II)/(alpha)ketoglutarate-dependent dioxygenase AsqJ	Q5AR53
<i>Candidatus Korarchaeum cryptofilum</i> hydrolase DNA repair enzyme	B1L4V6
<i>Giardia intestinalis</i> Superoxide reductase	V6TJK7
<i>Jonesia denitrificans</i> Chitinase	C7R4I0
<i>Nanoarchaeum equitans</i> Superoxide reductase	Q74MF3
<i>Oryctolagus cuniculus</i> Glycogenin	P13280
<i>Streptococcus thermophilus</i> cambialistic superoxide dismutase	F8LWI3
<i>Sulfolobus solfataricus</i> superoxide dismutase	P80857
<i>Thermotoga maritima</i> L-ketose-3-epimerase	Q9WYP7
<i>Micromonospora carbonacea</i> EvdO2	A0A0M3KL01
<i>Pseudomonas putida</i> Persulfide dioxygenase	A5VWI3
<i>Actinobacillus succinogenes</i> Phosphoenolpyruvate Carboxykinase	A6VKV4
<i>Cellvibrio japonicus</i> Lytic Polysaccharide Monooxygenase	B3PJ79
<i>Pseudomonas putida</i> Hydroxyquinol dioxygenase	C6FI44
<i>Acinetobacter radioresistens</i> catechol dioxygenase	C6RPG2
<i>Streptomyces</i> sp. Alpha-Ketoglutarate dependent enzyme	D0VX22
<i>Streptomyces lividans</i> lytic polysaccharide monooxygenase	D6EWM4
<i>Vulcanisaeta moutnovskia</i> Lactonase	F0QXN6
<i>Mycobacterium tuberculosis</i> 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase	O53512
<i>Pyrococcus horikoshii</i> Dipeptidase	O58691
<i>Thermus thermophilus</i> homocitrate synthase	O87198
<i>Brevundimonas diminuta</i> Phosphotriesterase	P0A434
<i>Escherichia coli</i> alpha-ketoglutarate dependent dioxygenase	P05050
<i>Acidianus ambivalens</i> sulfur oxygenase	P29082
<i>Chromobacterium violaceum</i> Phenylalanine hydroxylase	P30967
<i>Pseudomonas savastanoi</i> Ethylene forming enzyme	P32021
<i>Bacillus subtilis</i> Quercetin dioxygenase	P42106
<i>Bacillus caldovelox</i> arginase	P53608
<i>Gallus gallus</i> Tryptophan hydroxylase	P70080
<i>Synechocystis</i> sp. PCC 6803 apocarotenoid cleavage oxygenase	P74334
<i>Myxococcus xanthus</i> Persulfide Dioxygenase	Q1D4C9
<i>Rhodococcus</i> sp. DK17 Catechol dioxygenase	Q6REQ5
<i>Pseudomonas alkylphenolica</i> dioxygenase LapB	Q7WYF5
<i>Thermus aquaticus</i> fructose bisphosphate aldolase	Q9RHA2
<i>Thermotoga maritima</i> N-acetylglucosamine-6-phosphate deacetylase	Q9WZS1

Sequence_name	WT_UniProt
<i>Thermotoga maritima</i> NADH-dependent butanol dehydrogenase	Q9WZS7
<i>Thermotoga maritima</i> metallo-beta-lactamase	Q9X0P5
<i>Thermotoga maritima</i> Transcription regulator	Q9X1T8
<i>Helicobacter pylori</i> Peptide deformylase	Q672W7
<i>Enterococcus faecalis</i> Polysaccharide monooxygenase	Q838S1
<i>Streptomyces avermitilis</i> dioxygenase	Q53586
<i>Streptomyces wedmorensis</i> Hydroxypropylphosphonic acid epoxidase	Q56185
<i>Arabidopsis thaliana</i> Anthocyanidin synthase	Q96323
<i>Shewanella frigidimarina</i> Putative hydrolase	Q07XY2
<i>Pyrococcus abyssi</i> tRNA N6-adenosine threonylcarbamoyltransferase	Q9UXT7