

Evolution and characterization of carbene
transferases for cyclopropanation and
carbon–silicon bond formation

Thesis by
Russell DeRieux Lewis

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

2019
Defended May 3, 2019

© 2019

Russell DeRieux Lewis
ORCID: 0000-0002-5776-7347

ACKNOWLEDGEMENTS

I would first like to thank my advisor Frances Arnold, and my committee members Mikhail Shapiro, Dennis Dougherty, and Brian Stoltz for their support and advice during my time at Caltech.

In the Arnold lab, I've had the chance to learn from some amazing mentors—Hans Renata, Sabine Brinkmann-Chen, Andrew Buller, and Jenny Kan. I've also had the pleasure of collaborating with a number of fantastic people both inside and outside the Arnold lab, including Marc Garcia-Borràs, Allison Tang, and Matt Chalkley. I would not have been able to make it without all of your help and guidance.

I'd also like to give particular thanks to those people who helped make graduate school fun—my boyfriend Josh, my roommate Nikkie, and my fellow graduate students Kari Hernandez, Zach Wu, Anders Knight, Sam Ho, and Kelly Zhang. Finally, I'd like to thank my family for supporting me throughout all the steps of my education.

ABSTRACT

Heme proteins have recently been demonstrated to catalyze cyclopropanation reactions via a putative carbene transfer mechanism. Carbene transfer reactions are not known to occur in natural biological systems, but are highly useful synthetic reactions. There is growing interest in developing new “carbene transferases” that bring new chemical reactions into the realm of biology, and growing interest in engineering these enzymes for use in organic synthesis. Additionally, the mechanistic details of iron porphyrin-catalyzed carbene transfer reactions are largely unknown, especially with regards to how the enzyme environment influences the outcome of a carbene transfer reaction. This thesis details both the engineering of carbene transferases with novel catalytic capabilities and investigations into how these enzymes catalyze carbene transfer reactions. Chapter 1 introduces heme protein-catalyzed carbene transfer reactions and describes the directed evolution of new enzymes that allow access to a range of useful cyclopropane products. Chapter 2 describes the evolution of an enzyme that performs carbene transfer to silicon–hydrogen bonds, resulting in a highly efficient and selective carbon–silicon bond-forming enzyme, the first of its kind. Chapter 3 focuses on the characterization of a key reactive intermediate, the iron-porphyrin carbene, in the active site of the evolved carbon–silicon bond-forming enzyme. This study provides an explanation of the remarkable enantioselectivity of the enzyme and provides a foundation from which to investigate the enzyme reaction mechanism. The mechanism of carbon–silicon bond formation is elucidated in Chapter 4, and the is then used to explain how the enzyme achieves chemoselectivity, which in turn guides the evolution of enzyme variants with altered chemoselectivity. Finally, two off-cycle catalytic pathways that cause inactivation of the carbene transferase are characterized, and methods to prevent and/or circumvent inactivation are investigated (Chapter 5). Overall, the work presented here expands the repertoire of enzyme-catalyzed reactions and facilitates the continuing development of new carbene transferases by developing our mechanistic understanding of this novel class of enzymes.

PUBLISHED CONTENT AND CONTRIBUTIONS

(† indicates equal contributions by authors)

1. Renata, H.[†], **Lewis, R.D.**[†], Sweredoski, M.J., Moradian, A., Hess, S., Wang, Z.J., Arnold, F.H. Identification of Mechanism-Based Inactivation in P450-Catalyzed Cyclopropanation Facilitates Engineering of Improved Enzymes. *J. Am. Chem. Soc.* **138**, 12527-12533 (2016). doi:10.1021/jacs.6b06823

R.D.L participated in the design and execution of research, including the mass spectrometry, circular dichroism, and protein engineering. R.D.L. participated in the writing of the manuscript.

2. Hernandez, K.E., Renata, H., **Lewis, R.D.**, Kan, S.B.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.* 7810–7813 (2016). doi:10.1021/acscatal.6b02550

R.D.L participated in the design and execution of the protein engineering associated with this project. R.D.L. participated in the writing of the manuscript.

3. Kan, S.B.J., **Lewis, R.D.**, Chen, K., Arnold, F.H. Directed evolution of cytochrome *c* for carbon–silicon bond formation: Bringing silicon to life. *Science* **354**, 1048–1051 (2016). doi: 10.1126/science.aah6219

R.D.L participated in the conception, design, and execution of the research, focusing mainly on protein engineering and cytochrome *c* biochemistry. R.D.L. participated in the writing of the manuscript.

4. 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.* **4**, 372–377 (2018). doi: 10.1021/acscentsci.7b00548

R.D.L helped conceive the project, and participated in the design and execution of experiments including protein engineering and chemical synthesis. R.D.L. participated in the writing of the manuscript.

5. **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.* **115**, 7308–7313 (2018). doi: 10.1073/pnas.1807027115

R.D.L participated in the conception, design, and execution of the research. R.D.L solved and analyzed the crystal structures, and participated in the analysis of

spectroscopic and computational data. R.D.L. prepared the manuscript.

PUBLISHED CONTENT NOT INCLUDED IN THESIS

6. 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). doi:10.1073/pnas.1901979116

R.D.L participated in the conception of the project, and helped design the molecular cloning strategy.

TABLE OF CONTENTS

Acknowledgments	iv
Abstract.....	v
Published Content and Contributions	vi
Table of Contents	viii
List of Tables and Schemes.....	xii
List of Figures.....	xiii
Abbreviations.....	xvi
Chapter I: Beyond cytochrome P450: introduction to heme protein	
catalyzed carbene transfer reactions	1
1.1 Abstract	1
1.2 Enzymes and directed evolution	2
1.3 New reactions and catalytic promiscuity.....	2
1.4 Cyclopropanation via carbene transfer	3
1.5 Results and discussion.....	5
1.5.1 Cyclopropanation to produce a key intermediate to ticagrelor.....	5
1.5.2 Cyclopropanation of unactivated alkenes.....	7
1.6 Conclusions	10
1.7 References	11
Appendix A: Supporting Information for Chapter 1	14
A.1 Materials and methods	14
A.2 Cloning and library generation	15
A.3 Hemochrome assay	16
A.4 Growth and induction in 96-well plates.....	16
A.5 Growth and induction in flasks	17
A.6 Reactions screening in 96-well plates.....	17
A.6.1 Ticagrelor	17
A.6.2 Unactivated alkenes	18
A.7 Small scale biocatalytic reactions	18
A.7.1 Ticagrelor	18
A.7.2 Unactivated alkenes	19
A.8 Proteins investigated and homology modeling.....	20
A.9 Synthesis and characterization of authentic products.....	20
A.9.1 Ticagrelor	20
A.9.2 Unactivated alkenes	21
A.10 Calibration curves	24
A.11 Nucleotide and amino acid sequences	26

A.12	References	26
Chapter II: Directed evolution of cytochrome c for carbon–silicon bond formation.....		
	formation.....	28
2.1	Abstract	28
2.2	Introduction	29
2.3	Results and discussion.....	29
2.3.1	Heme catalyzed carbon–silicon bond formation.....	29
2.3.2	Evolution of <i>Rhodothermus marinus</i> cytochrome <i>c</i>	30
2.3.3	Substrate scope.....	33
2.3.4	Chemoselectivity changes over the course of evolution	33
2.3.5	Hypothesized substrate binding modes	35
2.4	Conclusions	36
2.5	References	37
Appendix B: Supporting Information for Chapter 2.....		
B.1	Materials and methods.....	40
B.2	Plasmid construction	41
B.3	Cytochrome <i>c</i> expression and purification.....	41
B.4	P450 and globin expression and purification.....	42
B.5	Hemochrome assay.....	43
B.6	Ferrous assay	44
B.7	Library construction	45
B.8	Enzyme library screening.....	45
B.9	Protein lysate preparation.....	46
B.10	Small-scale biocatalytic reactions	47
B.11	Supporting tables	48
B.12	Supporting figures	52
B.13	References	54
Chapter III: Catalytic iron-carbene intermediate revealed in a cytochrome c carbene transferase		
	carbene transferase	57
3.1	Abstract	57
3.2	Introduction	58
3.3	Results and discussion.....	59
3.3.1	Structural characterization of engineered carbene transferase, <i>Rma</i> TDE.....	59
3.3.2	Crystallographic observation of an iron porphyrin carbene in <i>Rma</i> TDE.....	61
3.3.3	Iron porphyrin carbene electronic structure and spectroscopy	64
3.3.4	Computational models of the iron porphyrin carbene	65
3.4	Summary and conclusions	66
3.5	References	66

Appendix C: Supporting Information for Chapter 3	70
C.1 Materials and methods.....	70
C.2 Plasmid construction	71
C.3 Cytochrome c expression and purification.....	72
C.4 ⁵⁷ Fe-labelled cytochrome c expression and purification.....	73
C.5 Hemochrome assay.....	75
C.6 Synthesis and characterization of authentic organosilicon product..	75
C.7 Biocatalytic reactions	76
C.8 UV-Visible spectroscopy	77
C.9 X-ray crystallography	77
C.10 Mössbauer spectroscopy	79
C.10.1 General description of Mössbauer spectroscopy.....	79
C.10.2 Sample preparation for Mössbauer spectroscopy.....	79
C.11 Electron paramagnetic resonance (EPR) spectroscopy	80
C.11.1 General description of EPR spectroscopy	80
C.11.2 Sample preparation for EPR spectroscopy	80
C.12 Supporting tables and figures.....	81
C.13 References	95
Chapter IV: Origins of chemoselectivity in cytochrome c-catalyzed carbene insertions into silicon–hydrogen bonds	98
4.1 Abstract	98
4.2 Introduction	99
4.3 Results and discussion.....	100
4.3.1 Quantum mechanical characterization of the reaction mechanism	100
4.3.2 Rate-determining step of the enzymatic reaction.....	103
4.3.3 Silane binding to <i>Rma</i> cyt c.....	105
4.3.4 Rationalization of chemoselectivity	107
4.3.5 Engineering inverted chemoselectivity in <i>Rma</i> TDE.....	109
4.4 Conclusions	112
4.5 References	113
Appendix D: Supporting Information for Chapter 4	116
D.1 Materials and methods	116
D.2 Plasmid construction	117
D.3 Cytochrome c expression and purification	118
D.4 Hemochrome assay	119
D.5 Library construction	120
D.6 Enzyme library screening.....	120
D.7 Protein lysate preparation.....	121
D.8 Biocatalytic reactions	121
D.8.1 Kinetic time course reactions.....	122
D.9 Synthesis and characterization of organosilicon compounds.....	123

D.10	Supporting figures and tables.....	125
D.11	References	132
Chapter V: Mechanism based inactivation in heme-dependent carbene transferases.....		
		133
5.1	Abstract	133
5.2	Introduction	134
5.3	Results	135
5.3.1	EDA-induced inactivation of P450 BM3-Hstar.....	135
5.3.2	Heme cofactor modification	136
5.3.3	Modification of protein side chains and structural changes.....	139
5.3.4	Identification of modified residues.....	140
5.3.5	Site-directed mutagenesis of modified residues improves catalyst performance.....	143
5.3.6	Characterization of BM3-HStar H92N H100N	144
5.3.7	Replacement of the modified heme cofactor.....	145
5.4	Discussion	148
5.4	Conclusion.....	151
5.5	References	152
Appendix E: Supporting Information for Chapter 5.....		
		157
E.1	Materials and methods.....	157
E.2	Generation of enzyme variants.....	158
E.3	Hemochrome binding assay	158
E.4	Determination of ϵ_{\max} for ferrous histidine-ligated P450 enzymes.....	158
E.5	Ferrous concentration assay.....	159
E.6	P450 expression and purification	160
E.7	Small-scale whole-cell bioconversions	160
E.8	Small-scale cyclopropanation reaction with purified enzymes	161
E.9	Analysis of diazo-induced inactivation	162
E.10	Liquid chromatography-mass spectrometry (LCMS) analysis of cofactor modification	163
E.11	Isolation of modified porphyrin cofactor	163
E.12	Analysis of enzyme modification by top-down proteomics	164
E.13	Alkylation analysis by bottom-up proteomics	164
E.14	Circular dichroism analysis	165
E.15	Whole-cell reaction of styrene with EDA catalyzed by BM3-HStar variants.....	166
E.16	Whole-cell reaction of aniline with EDA catalyzed by BM3-HStar variants.....	166
E.17	Immobilization of <i>Bsu</i> TrHb LAA	167
E.18	Protein-bead reactions	167
E.19	Protein-bead heme removal and replacement	168
E.20	Supporting tables and figures	169
E.21	References.....	175

LIST OF TABLES AND SCHEMES

<i>Table Number</i>	<i>Page</i>
A-1	Heme binding proteins for unactivated alkene cyclopropanation.....20
B-1	Catalytic systems for enantioselective carbene insertion into silicon–hydrogen bonds.....48
B-2	Preliminary experiments with heme and heme proteins.....50
B-3	Carbon–silicon bond formation catalyzed by <i>Rma</i> cyt <i>c</i> variants51
C-1	Data collection and refinement statistics for <i>Rma</i> TDE81
C-2	Bond lengths of interest.....82
C-3	Out-of-plane distortions for various IPC structures.....93
E-1	Amino acids substitutions relative to WT P450-BM3169
E-2	TTN and selectivity for whole-cell BM3-HStar variants169
E-3	TTN and selectivity for purified BM3-Hstar variants169
E-4	Representative deconvolution for purified, intact protein172
E-5	Representative deconvolution of EDA-treated intact protein.....172
<i>Scheme Number</i>	<i>Page</i>
4-1	General reaction catalyzed by <i>Rma</i> TDE100
4-2	Schematic representation of the IPC insertion into the Si–H bond.....104
D-1	How silane binding might affect the enzyme rate-determining step....131
5-1	Cyclopropanation catalyzed by BM3-Hstar136
5-2	Proposed method for replacing modified heme cofactor146

LIST OF FIGURES

<i>Number</i>	<i>Page</i>
1-1 Cytochrome P450-catalyzed cyclopropanation and epoxidation.....	4
1-2 Ticagrelor and key synthetic intermediates.....	5
1-3 Site selection in <i>Bacillus subtilis</i> TrHb	7
1-4 Test reactions for unactivated alkene cyclopropanation	9
1-5 Substrate scope of <i>ApePgb</i> AGW and <i>RmaNOD</i> Q52V.....	10
A-1 Achiral GC calibration curve of 6	25
A-2 Achiral GC calibration curve of 12	25
A-3 Achiral GC calibration curve of 13	25
2-1 Heme protein-catalyzed carbon–silicon bond formation	32
2-2 Scope of <i>Rma</i> TDE-catalyzed carbon–silicon bond formation.....	34
2-3 Chemoselectivity of evolved <i>Rma</i> <i>cyt c</i>	35
2-4 Proposed model for carbene orientation.....	36
B-1 SDS-PAGE of purified <i>Rma</i> <i>cyt c</i>	52
B-2 Circular dichroism spectra of purified <i>Rma</i> TDE.....	52
B-3 Hammett analysis of <i>Rma</i> TDE-catalyzed carbon–silicon bond formation	53
3-1 The scope of laboratory-evolved biocatalytic carbene transfer reactions.....	59
3-2 Carbene transfer silylation reaction and structures of resting and carbene-bound <i>Rma</i> TDE.....	63
C-1 Active-site front loop deviation	82
C-2 UV-vis spectra.....	84
C-3 EPR spectroscopic analysis.....	86
C-4 Mössbauer spectroscopy studies	87
C-5 GC standard curve for ethyl 2-(dimethylphenylsilyl)propanoate.....	94
C-6 ¹ H NMR for ethyl 2-(dimethylphenylsilyl)propanoate	94

C-7	Separation of ethyl 2-(dimethylphenylsilyl)propanoate	95
4-1	Computed reaction pathway for carbene Si–H insertion reaction between PhMe ₂ SiH and Me-EDA	102
4-2	Intrinsic reaction coordinate for the lowest energy Si–H carbene insertion transition state.....	102
4-3	Silane binding and MD simulations with <i>Rma</i> TDE and WT <i>Rma</i> cyt <i>c</i>	106
4-4	Comparison of amination and silylation transition states	108
4-5	Evolution of an amination-selective variant of <i>Rma</i> cyt <i>c</i>	111
D-1	Kinetic time course data for <i>Rma</i> cyt <i>c</i> variants.....	125
D-2	¹ H NMR for phenyl(dimethyl)silane- <i>d</i>	129
D-3	GC standard curve for ethyl 2-(dimethylphenylsilyl)propanoate	130
5-1	On-cycle and off-cycle catalytic pathways for carbene transfer	135
5-2	LC-MS monitoring of heme content over time	138
5-3	Analysis of modified porphyrin isolated after the cyclopropanation.	138
5-4	CD spectra of BM3-Hstar before and after EDA treatment.....	140
5-5	Representative LC-MS/MS analysis of alkylated peptides.....	141
5-6	Location of alkylated residues relative to the heme cofactor.....	142
5-7	Site-directed mutagenesis of BM3 Hstar to improve cyclopropanation yield and selectivity	145
5-8	Use and re-use of agarose bead immobilized <i>Bsu</i> TrHb LAA.....	148
E-1	Pre-incubation experiments with EDA and BM3-Hstar	170
E-2	Comparison of isolated porphyrin with PPIX standard	171
E-3	Whole protein MS of BM3-Hstar before and after EDA treatment ..	171
E-4	Polar contact between K69 and heme cofactor	172
E-5	TTN comparison for styrene cyclopropanation and N–H insertion between BM3-HStar and BM3-HStar H92N H100N.....	173
E-6	Whole cell time course using HStar variants	173
E-7	Circular dichroism of HStar H92N H100N before and after reaction	174

E-8	GC calibration curve for aniline/EDA product.....	174
E-9	Control reactions with immobilized <i>Bsu</i> TrHb LAA.....	175

ABBREVIATIONS

Å	angstrom(s)
AcOH	acetic acid
ALA	5-aminolevulinic acid
amp	ampicillin
<i>ApePgb</i>	protoglobin from <i>Aeropyrum pernix</i>
<i>Bsu</i>	<i>Bacillus subtilis</i>
BM3-Hstar	a histidine-ligated P450 variant engineered for cyclopropanation
cyt <i>c</i>	cytochrome <i>c</i>
CD	circular dichroism
chlor	chloramphenicol
CSS	closed-shell singlet
DCM	dichloromethane
DFT	density functional theory
d.r.	diastereomeric ratio
<i>ee</i>	enantiomeric excess
EDA	ethyl diazoacetate
EtOH	ethanol
FID	flame ionization detection
FPLC	fast protein liquid chromatography
GC	gas chromatography
HB	hyperbroth medium
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
IPC	iron porphyrin carbene
LB	Luria-Bertani medium
LC-MS	liquid chromatography with mass spectrometry
M9-N	M9 minimal media with no nitrogen source added
MD	molecular dynamics
Me-EDA	ethyl 2-diazopropanoate (methyl-EDA)
MeCN	acetonitrile
MeOH	methanol
MOPS	3-(N-morpholino)propanesulfonic acid
NaPi	sodium phosphate
OD ₆₀₀	optical density at 600 nm
OSS	open-shell singlet
P450 BM3	Cytochrome P450 from <i>Bacillus megaterium</i> (CYP102A1)
PDB	Protein Data Bank
QM/MM	quantum mechanics / molecular mechanics
<i>Rma</i>	<i>Rhodothermus marinus</i>
<i>RmaNOD</i>	nitric oxide dioxygenase from <i>Rhodothermus marinus</i>

<i>Rma</i> TDE	<i>Rma</i> cytochrome c V75T M100D M103E
<i>Rma</i> TDFPI	<i>Rma</i> cytochrome c V75T M100D (N80F M99P E103I)
RT	room temperature
SFC	supercritical fluid chromatography
tR	retention time
TrHb	truncated hemoglobin
TS	transition state
TTN	total turnover number
WT	wild type

*Chapter 1*BEYOND CYTOCHROME P450: INTRODUCTION TO HEME
PROTEIN CATALYZED CARBENE TRANSFER REACTIONS

Material from this chapter appears in “Hernandez KE, Renata H, Lewis RD, Kan SBJ, Zhang C, Forte J, Rozzell D, McIntosh JA, Arnold FH (2016) **Highly stereoselective biocatalytic synthesis of key cyclopropane intermediate to ticagrelor.** *ACS Catalysis*, 6:7810-7813. doi: 10.1021/acscatal.6b02550” and “Knight AM, Kan SBJ, Lewis RD, Brandenburg OF, Chen K, Arnold FH (2018) **Diverse engineered heme proteins enable stereodivergent cyclopropanation of unactivated alkenes.** *ACS Cent Sci*, 4:372-377. doi: 10.1021/acscentsci.7b00548” Reprinted with permission from the American Chemical Society.

1.1 Abstract

Heme proteins have been discovered to catalyze cyclopropanation reactions via a putative carbene transfer mechanism, which is not known to occur in nature. Although initially discovered through studies of cytochrome P450 enzymes, trace cyclopropanation activity is found in many heme proteins, including those not known to have any natural catalytic function. Through directed evolution, these heme proteins can be engineered into highly efficient and selective cyclopropanation enzymes, and the natural diversity of heme proteins provides a functionally diverse panel of starting points from which we can rapidly engineer biocatalysts to produce new and useful cyclopropane products. This work highlights how directed evolution, reaction mechanism, and the natural diversity of proteins can be leveraged to develop biocatalysts for synthetically important reactions, which provides new opportunities to replace traditional chemical syntheses of cyclopropanes with efficient, clean, and highly selective enzyme-catalyzed processes.

1.2 Enzymes and directed evolution

Over the billions of years that life has been evolving on earth, nature has engineered enzymes to catalyze a multitude of biochemical reactions. Through the natural process of evolution, enzymes have become finely tuned for their specific functions and are highly efficient for their native reactions. Because of their remarkable efficiency and selectivity, enzymes are attractive candidates for use in industrial chemical synthesis^{1,2}. However, the enzymes found in nature are not optimized for use by human chemists—natural enzymes may be poor catalysts under industrial conditions or may have low activity with non-native substrates².

In order to create enzymes optimized for industrial purposes, researchers turn to directed evolution, a laboratory process designed to mimic evolution in the natural world^{3,4}. In a directed evolution experiment, researchers start with an enzyme that has some measureable desired property (or properties) that they wish to optimize. Next, a library of genes is created, each encoding an enzyme containing a few mutations (≥ 1) relative to the “parent” enzyme. The variants are then evaluated for the desired property (“fitness”). When researchers identify a mutation that improves enzyme fitness, this variant can serve as a new parent enzyme, and the directed evolution cycle can be repeated. Iteration of the directed evolution cycle leads to an accumulation of beneficial mutations in the enzyme, ultimately optimizing the enzyme for the desired property.

1.3 New reactions and catalytic promiscuity

Directed evolution is the method of choice for creating enzymes of high value to chemists: enzymes with improved stability, activity with non-native substrates, and/or altered selectivity¹. More recently, directed evolution has been applied to solve a new problem—creating enzymes that catalyze reactions not known to occur in biology^{5,6}. Many useful reactions known in synthetic chemistry have no biological counterpart, and developing enzymes to catalyze these reactions would provide new green routes for chemical synthesis^{5,6}.

A good strategy for identifying and evolving an enzyme to catalyze a synthetic reaction is to take a mechanism-guided approach⁵. Enzymes often display “catalytic promiscuity”, meaning that in addition to catalyzing their native reaction they also display low-level activity for mechanistically similar reactions⁷⁻⁹. Thus, by looking for similarity between the mechanism of a desired synthetic reaction and enzyme mechanisms, we can identify candidate enzymes to screen for the new, desired catalytic activity⁵. Once an enzyme with the desired catalytic activity is identified, directed evolution can be used to re-optimize the enzyme for its new function.

1.4 Cyclopropanation via carbene transfer

In a seminal example, Coelho et al. investigated the ability of cytochrome P450 proteins to catalyze a non-natural cyclopropanation reaction between ethyl diazoacetate (EDA, **1**) and styrene¹⁰ (**2**, **Figure 1-1a**). Mechanistic studies of small molecule iron porphyrin cyclopropanation catalysts had shown that the reaction proceeds by formation of a reactive iron porphyrin carbene (IPC) intermediate, and subsequent transfer of the carbene to an alkene, yielding a three-membered ring (cyclopropane) product¹¹ (**Figure 1-1c**). Mechanistic studies have also demonstrated that cytochrome P450 catalyze epoxidation reactions by a similar mechanism¹². Cytochrome P450 proteins use an iron porphyrin (heme) cofactor, and catalyze epoxidation by first forming a reactive iron-oxo (oxene equivalent) intermediate (**Figure 1-1b**). The oxene equivalent is then transferred to an alkene, yielding a three-membered ring (epoxide) product.

By providing proteins with a carbene precursor (**1**) and an alkene (**2**), Coelho et. al found that cytochrome P450 proteins displayed some promiscuous activity for the cyclopropanation reaction. After discovering the initial reactivity, the researchers were able to evolve cytochrome P450 variants with improved cyclopropanation activity and evolved variants that allowed access to different stereoisomers of the cyclopropane product. Interestingly, Coelho et. al also found measureable activity for styrene cyclopropanation in other heme binding proteins that were not typically considered “enzymes”, including heme-dependent electron- and oxygen-transport proteins (cytochrome *c* and globin proteins). These

proteins are small, well-studied proteins that have been shown to be robust to mutation and can be easily expressed to high titers in *E. coli*—making them promising scaffolds for biocatalysts^{13,14}.

Following the initial report of heme protein catalyzed cyclopropanation, researchers from the lab of Rudi Fasan (University of Rochester) began to engineer sperm whale myoglobin for cyclopropanation activity. Fasan and coworkers targeted mutations to the distal oxygen binding pocket of the protein, and after introducing only two mutations, they produced a variant which catalyzed cyclopropanation reactions with efficiency and selectivity rivalling the best P450-derived cyclopropanation catalysts¹³. Their work showed that myoglobin was an easily engineered scaffold for cyclopropanation reactions, and suggested that other members of the globin superfamily might also be easily engineered into cyclopropanation enzymes.

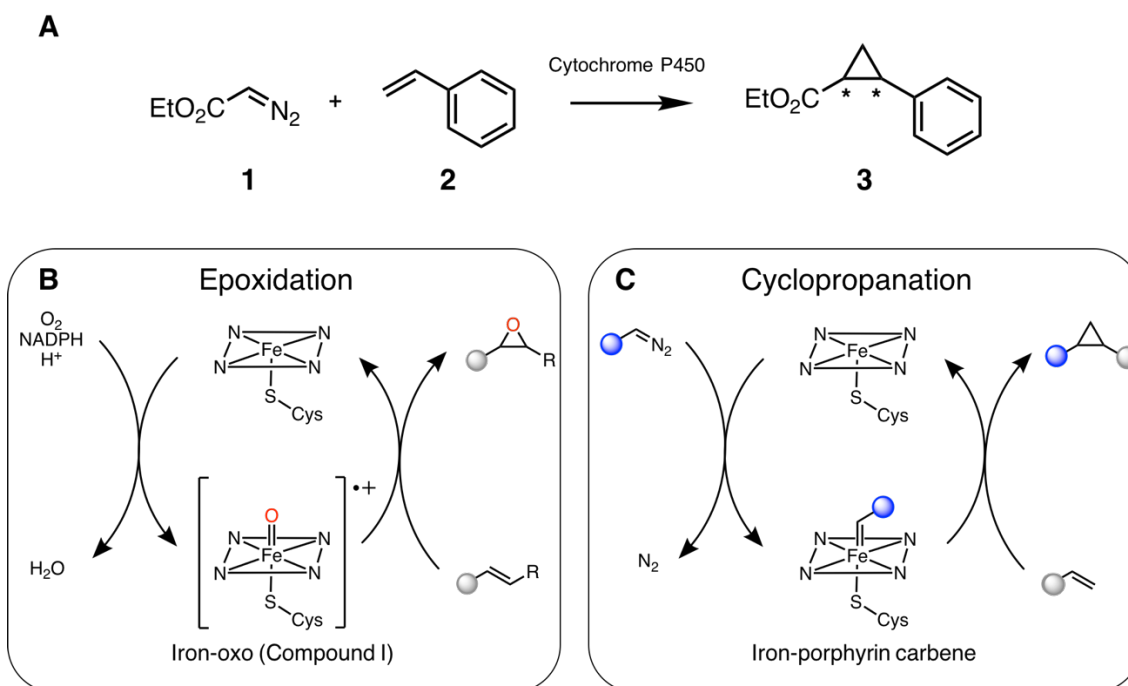


Figure 1-1. Cytochrome P450 catalyzed cyclopropanation and epoxidation. **A.** The reaction between ethyl diazoacetate and styrene produces a cyclopropane with two chiral centers. **B.** Epoxidation is a natural reaction catalyzed by cytochrome P450 proteins, which proceeds through the key catalytic intermediate Compound I, an iron-oxo species. **C.** Proposed mechanism for P450-catalyzed cyclopropanation. A diazo compound serves as a carbene precursor, and reacts with the heme cofactor to form an iron-porphyrin carbene intermediate. *Figure adapted from Coelho et al. This is reference 11.*

1.5 Results and discussion

1.5.1 Cyclopropanation to produce a key intermediate to ticagrelor

Cyclopropyl motifs are present in many bioactive and synthetic compounds and serve as versatile intermediates in organic synthesis^{15,16}. Thus, after the initial reports of heme protein catalyzed cyclopropanation reactions, we sought to apply enzymatic cyclopropanation to a pharmaceutically relevant synthetic problem. Ticagrelor (**3**) is an FDA-approved, multibillion-dollar small-molecule therapeutic for which most of the patented syntheses use a key chiral cyclopropane intermediate, (1*R*,2*S*)-2-(3,4-difluorophenyl)cyclopropan-1-amine¹⁷ (**4**). This key chiral intermediate can be accessed by a single-step derivatization of ethyl-(1*R*, 2*R*)-2-(3,4-difluorophenyl)-cyclopropane-1-carboxylate (**6**), which closely resembles products made already made by P450- and globin-derived cyclopropanation enzymes.

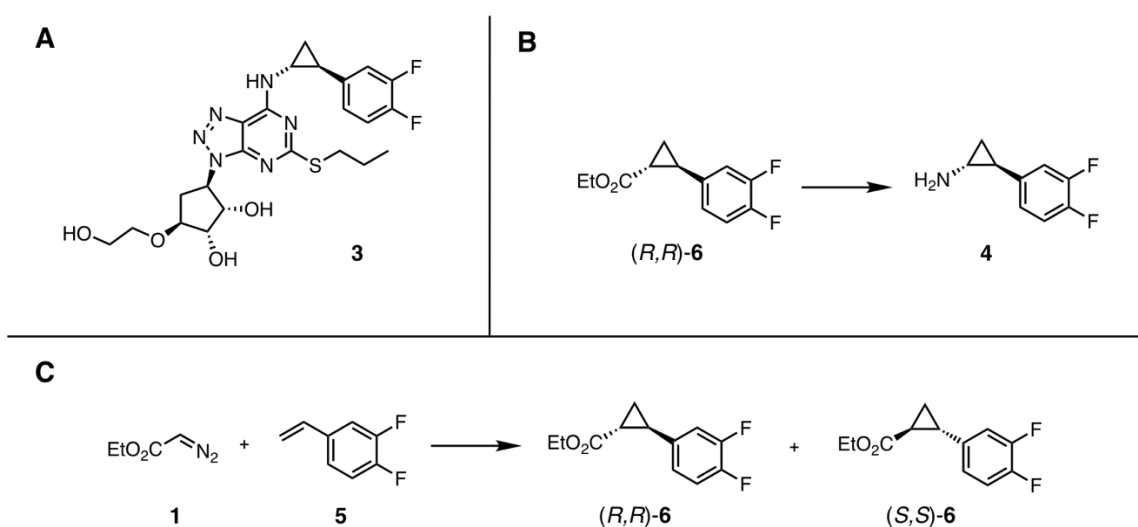


Figure 1-2. Ticagrelor and key synthetic intermediates. **A.** Structure of ticagrelor (**3**) **B.** The cyclopropyl ester precursor **6** can be converted to the cyclopropylamine precursor (**4**) by a Curtius rearrangement or Hofmann rearrangement of the corresponding amide **C.** Cyclopropanation reaction between ethyl diazoacetate (**1**) and 3,4-difluorostyrene (**5**) makes the cyclopropyl ester (**6**).

In order to evolve a cyclopropanation enzyme that could be used in the synthesis of ticagrelor, we began screening heme proteins for their ability to selectively produce the desired *R,R* isomer. We screened proteins previously engineered to catalyze cyclopropanation reactions, as well as wild-type globin and cytochrome *c* proteins. We identified a wild-type globin protein that produced the desired *trans* isomer with >99:1 diastereomeric ratio (d.r.) and 34% enantiomeric excess (*ee*) for the desired (*R,R*) enantiomer. The protein, a type II truncated hemoglobin (TrHb) from *Bacillus subtilis*, is a small (132 residues, 15 kDa) monomeric protein with high oxygen affinity and a published crystal structure¹⁸ (PDB ID: 1UX8). By aligning¹⁹ the crystal structure of TrHb with the structure of sperm whale myoglobin²⁰, we chose residues to mutate in TrHb that were analogous to those targeted by the Fasan lab in sperm whale myoglobin (T45 and Q49 in TrHb, **Figure 1-3**). The two residues were mutated simultaneously to one of three residues (leucine, phenylalanine, and alanine) and the mutant proteins were screened for improved cyclopropane yield and enantioselectivity. The TrHb T45A Q49A double mutant stood out as highly selective, yielding the desired product in 95% *ee* with an analytical yield of 57%. To further improve enantioselectivity, site-saturation mutagenesis was performed at position A45, and the putative distal heme ligand Y25. While mutations at A45 failed to improve enantioselectivity, screening of the Y25 library led to the discovery of variant TrHb Y25L T45A Q49A, which maintained the yield of the parent protein while affording >99% d.r. and 97% *ee* at 40 mM **1** and 20 mM **5**.

Like Fasan's earlier work with sperm whale myoglobin, our work provided an example of a wild-type, non-enzyme heme protein that could be easily evolved into a highly efficient and selective cyclopropanation catalyst. Together, our work and Fasan's work demonstrated that cyclopropanation enzymes with *divergent stereospecificity* can be easily evolved from parent proteins found in nature's repertoire of non-enzyme heme proteins. More generally, the results suggest that the natural diversity of heme proteins can be leveraged to discover functionally diverse cyclopropanation catalysts.

A

<i>B. subtilis</i> TrHb	---N-----APYEAI GEELLSQLVDTFYERVA-SHPLLKPIF-PS-----
<i>P. macrocephalus</i> Mb	VLSEGEWQLVLHVWAK-VEADVAGHGQDILIRLFKSHPETLEKFD RFKHLKTEAEMKASE
<i>B. subtilis</i> TrHb	DLTE T ARK Q KQFLTQYLGGPPLYTEEHGHP--MLRARHL-PPFITNERADAWLSCMKDAM
<i>P. macrocephalus</i> Mb	DLKK H GVT V LTA LGAILKKKGHHE----AELKPLAQSHATKHKIPIKYLEFISEAIIHVL
<i>B. subtilis</i> TrHb	DHV-G--LEGEIREFLFGRLELTARHM---V----N
<i>P. macrocephalus</i> Mb	HSRHPGDFGADAQGAMNKALELFRKDI AAKYKELGY

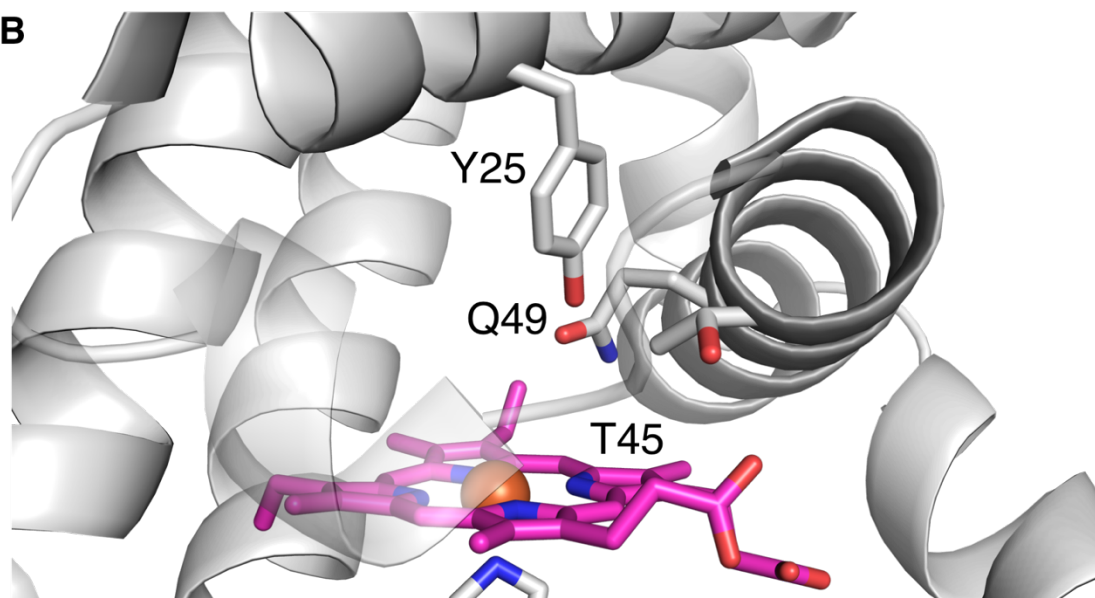
B

Figure 1-3. Site selection in *Bacillus subtilis* TrHb. **A.** Structure-guided sequence alignment between sperm whale myoglobin (*Physeter macrocephalus* Mb) and *Bacillus subtilis* TrHb. Residues highlighted in red were targeted for mutagenesis by the Fasan lab¹³, and the analogous residues in *Bacillus subtilis* TrHb, T45 and Q49, were targeted for mutagenesis in this work. **B.** Positions of the Y25, T45, and Q49 residues near the heme iron in the *B. subtilis* wild-type protein (PDB ID: 1UX8).

1.5.2 Cyclopropanation of unactivated alkenes

Although capable of high turnover and selectivity, the first cyclopropanation enzymes were limited to activity on aryl alkenes^{10,13,21-23}. 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 stereocontrol²⁴. We therefore set out to discover heme proteins that could cyclopropanate unactivated alkenes, and test our hypothesis that we could leverage the natural diversity of heme proteins to find functionally diverse cyclopropanation catalysts.

We collected a panel of eleven heme proteins from thermophilic and hyperthermophilic bacteria and archaea to test for unactivated alkene cyclopropanation (Appendix A.6, **Table A-1**). Proteins from thermophiles were selected because thermostable proteins are better able to withstand the destabilizing effects of mutations, and are therefore more “evolvable”²⁵. They are also often easier to work with and better tolerate polar organic solvents used to solubilize substrates. We screened this panel of proteins for cyclopropanation activity on 1-octene (**7**) with ethyl diazoacetate (**1**), and found that *Aeropyrum pernix* protoglobin (*ApePgb* WT, UniProt ID: Q9YFF4) and *Rhodothermus marinus* nitric oxide dioxygenase (*RmaNOD* WT, UniProt ID: D0MGT2) had low but measurable activity for this reaction, catalyzing the reaction with 18 and 27 total turnovers (**Figure 1-4a**). Notably, *ApePgb* WT and *RmaNOD* WT displayed complementary diastereoselectivity, preferentially producing *cis* (*R,S*)-**8** and *trans* (*S,S*)-**8**, respectively.

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 carbene formation and substrate binding are predicted to take place (Appendix A.6, **Table A-1**). Individual site-saturation libraries were screened for increased activity and diastereoselectivity using 4-phenyl-1-butene (**9**) and EDA (**1**) as substrates (**9** was chosen as a substrate for library screening because the UV-visible phenyl group enables screening by HPLC-UV). Variants with enhanced diastereoselectivity in the production of **10** were regrown in larger scale, and their activities were tested for cyclopropanation of 1-octene (**7**) and 4-phenyl-1-butene (**9**) and with EDA. Enzyme variants with the greatest overall selectivity enhancements were used as parents in the next rounds of site-saturation mutagenesis and screening. A single mutation (Q52V) gave *RmaNOD* near-

perfect stereoselectivity for producing *trans* (*S,S*)-**7**. Three mutations (W59A Y60G F145W, or “AGW”) gave *ApePgb* the ability to make *cis* (*R,S*)-**7** with 89:11 d.r. and 99% *ee*.

In addition to activity on the two unactivated alkene substrates **7** and **9**, *RmaNOD* Q52V and *ApePgb* AGW were found to have even higher activity for cyclopropanation of an acrylate compound, an electron-deficient alkene (**Figure 1-5a**). *ApePgb* AGW was also found to have activity on a variety of other alkene substrates including sterically hindered alkenes and small electron-deficient alkenes (**Figure 1-5b**). Though activity and stereoselectivity differed on different substrates, the broad range of alkenes accepted by *ApePgb* AGW suggests that this variant could serve as a good parent protein from which to evolve cyclopropanation catalysts that accept specific alkene substrates to make a specific desired product, as has been shown in many other directed evolution studies^{3,4}.

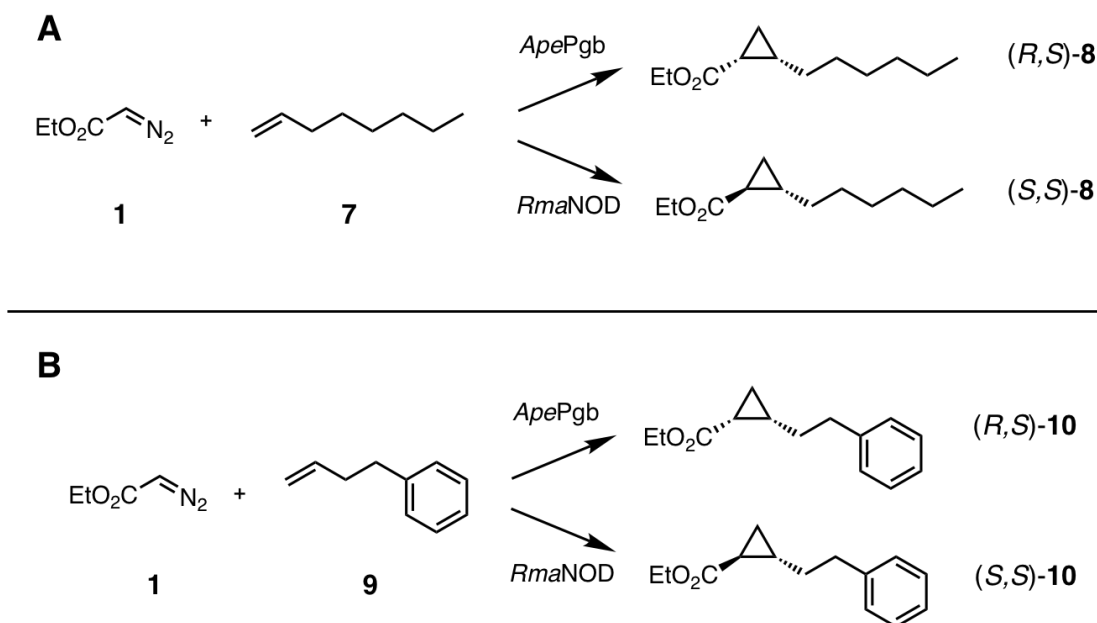


Figure 1-4. Test reactions for unactivated alkene cyclopropanation. **A.** Cyclopropanation of 1-octene (**7**) with ethyl diazoacetate (**1**). **B.** Cyclopropanation of 4-phenyl-1-butene (**9**) with ethyl diazoacetate (**1**).

A		<i>ApePgb</i> AGW (<i>R,S</i>)	<i>RmaNOD</i> Q52V (<i>S,S</i>)
8		490 TTN 89:11 dr 99% ee	100 TTN <1:99 dr >99% ee
10		2400 TTN 84:16 dr 95% ee	210 TTN 3:97 dr 98% ee
11		5400 TTN 71:29 dr 98% ee	3700 TTN <1:99 dr >99% ee

B			
12		1100 TTN 92:8 dr 99% ee	
13		550 TTN 76:24 dr 96% ee	
14		610 TTN 96% ee	
15		610 TTN 89:11 dr 99% ee	

Figure 1-5. Substrate scope of *ApePgb* AGW and *RmaNOD* Q52V. Reactions were performed under anaerobic conditions using whole *E. coli* cells at $OD_{600} = 5.0$ in M9-N buffer with 25 mM glucose, 5% ethanol as a cosolvent, 10 mM alkene, and 20 mM **1**. **A.** *ApePgb* AGW and *RmaNOD* Q52V are able to produce cyclopropane products using the unactivated alkene screening substrates, as well as using an electron deficient alkene to produce product **10**. **B.** Activity and selectivity for a range cyclopropane products produced by *ApePgb* AGW. Reaction conditions for product **13** used cells at $OD_{600} = 10$ and for product **15** $OD_{600} = 20$.

1.6 Conclusions

While biocatalysts often possess very high selectivity, this selectivity can be synthetically limiting. A single enzyme may make only a single isomer, and/or may only accept a certain range of substrates, but a synthetic chemist may want to access a wide range of substrates and a variety of product isomers. Therefore, for broad synthetic utility, an array of

biocatalysts may be needed to access a broad range of chemical transformations. For natural reactions, the natural diversity of enzymes has been leveraged to meet this challenge: genome mining has been used to discover lipases, ketoreductases, and transaminases with different stereoselectivities and substrate preferences^{26–28}. However, when engineering enzymes for a new-to-nature reaction, this genome mining approach may be impractical. In the case of cyclopropanation reactions, there is no established family of “carbene transferases” to use as a template for genome mining. Here, we have addressed this problem by demonstrating that heme-binding globin proteins are a rich source of functionally diverse cyclopropanation catalysts. With just a few mutations, they can catalyze cyclopropanation reactions with high efficiency and selectivity, despite the fact that they may not have any known natural catalytic activity. Thus, by surveying the natural diversity of heme binding proteins, we can find good starting points for evolution and rapidly engineer novel cyclopropanation biocatalysts. This approach has allowed us to develop catalysts for pharmaceutical precursors as well as cyclopropanes derived from electronically challenging substrates. As these cyclopropanation enzymes function via a carbene transfer mechanism, they may also display promiscuous activity for new carbene transfer reactions and serve as starting points from which to evolve novel reactivity in biological systems.

1.7 References

1. Hughes, G. & Lewis, J. C. Introduction: biocatalysis in industry. *Chem. Rev.* **118**, 1–3 (2018).
2. Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485**, 185–194 (2012).
3. Turner, N. J. Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* **5**, 567–573 (2009).
4. Cobb, R. E., Chao, R. & Zhao, H. Directed evolution: past, present and future. *AIChE J.* **59**, 1432–1440 (2013).
5. 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.* **54**, 3351–3367 (2015).
6. Prier, C. K. & Arnold, F. H. Chemomimetic biocatalysis: exploiting the synthetic potential of cofactor-dependent enzymes to create new catalysts. *J. Am. Chem. Soc.* **137**, 13992–14006 (2015).
 7. Brien, P. J. O. & Herschlag, D. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **6**, R91–R105 (1999).
 8. Copley, S. D. Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Curr. Opin. Chem. Biol.* **7**, 265–272 (2003).
 9. Khersonsky, O., Roodveldt, C. & Tawfik, D. S. Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.* **10**, 498–508 (2006).
 10. Coelho, P. S., Brustad, E. M., Kannan, A. & Arnold, F. H. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* **339**, 307–310 (2013).
 11. Li, Y., Huang, J., Zhou, Z., Che, C. & You, X. Remarkably stable iron porphyrins bearing nonheteroatom-stabilized carbene or (alkoxycarbonyl)carbenes: isolation, X-ray crystal structures, and carbon atom transfer reactions with hydrocarbons. *J. Am. Chem. Soc.* **124**, 13185–13193 (2002).
 12. Guengerich, F. P. Mechanisms of cytochrome P450 substrate oxidation. *J. Biochem. Mol. Toxicol.* **21**, 163–168 (2007).
 13. Bordeaux, M., Tyagi, V. & Fasan, R. Highly diastereoselective and enantioselective olefin cyclopropanation using engineered myoglobin-based catalysts. *Angew. Chem. Int. Ed.* **54**, 1744–1748 (2015).
 14. Vazquez-Duhalt, R. Cytochrome c as a biocatalyst. *J. Mol. Cat. - B Enzymatic* **7**, 241–249 (1999).
 15. Chen, D. Y., Pouwer, H. & Richard, J. Recent advances in the total synthesis of cyclopropane-containing natural products. *Chem. Soc. Rev.* **41**, 4631–4642 (2012).
 16. Reissig, H.-U. & Zimmer, R. Donor–acceptor-substituted cyclopropane derivatives and their application in organic synthesis. *Chem. Rev.* **103**, 1151–1196 (2003).
 17. Hugentobler, K. G., Sharif, H., Rasparini, M., Heath, R. S. & Turner, N. J. Biocatalytic approaches to a key building block for the anti-thrombotic agent ticagrelor. *Org. Biomol. Chem.* **14**, 8064–8067 (2016).
 18. Giangiacomo, L., Ilari, A., Boffi, A., Morea, V. & Chiancone, E. The truncated

- oxygen-avid hemoglobin from *Bacillus subtilis*: X-ray structure and ligand binding properties. *J. Biol. Chem.* **280**, 9192–9202 (2005).
19. Pei, J., Kim, B. H. & Grishin, N. V. PROMALS3D: A tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, 2295–2300 (2008).
 20. Vojtěchovský, J., Chu, K., Berendzen, J., Sweet, R. M. & Schlichting, I. Crystal structures of myoglobin-ligand complexes at near-atomic resolution. *Biophys. J.* **77**, 2153–2174 (1999).
 21. Coelho, P. S. *et al.* A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo. *Nat. Chem. Biol.* **9**, 485–7 (2013).
 22. Wang, Z. J. *et al.* Improved cyclopropanation activity of histidine-ligated cytochrome P450 enables the enantioselective formal synthesis of levomilnacipran. *Angew. Chem. Int. Ed.* **53**, 6810–6813 (2014).
 23. Hernandez, K. E. *et al.* Highly stereoselective biocatalytic synthesis of key cyclopropane intermediate to ticagrelor. *ACS Catal.* 7810–7813 (2016). doi:10.1021/acscatal.6b02550
 24. Coombs, J. R. & Morken, J. P. Catalytic enantioselective functionalization of unactivated terminal alkenes. *Angew. Chem. Int. Ed.* **55**, 2636–2649 (2016).
 25. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proc. Nat. Acad. Sci. USA* **103**, 5869–5874 (2006).
 26. Kaluzna, I. A., Matsuda, T., Sewell, A. K. & Stewart, J. D. Systematic investigation of *Saccharomyces cerevisiae* enzymes catalyzing carbonyl reductions. *J. Am. Chem. Soc.* **126**, 12827–12832 (2007).
 27. 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.* **6**, 807–813 (2010).
 28. Ferrer, M., Martínez-Abarca, F. & Golyshin, P. N. Mining genomes and ‘metagenomes’ for novel catalysts. *Curr. Opin. Biotechnol.* **16**, 588–593 (2005).

Appendix A

SUPPLEMENTARY INFORMATION FOR CHAPTER 1

Material from this chapter appears in “Hernandez KE, Renata H, Lewis RD, Kan SBJ, Zhang C, Forte J, Rozzell D, McIntosh JA, Arnold FH (2016) **Highly stereoselective biocatalytic synthesis of key cyclopropane intermediate to ticagrelor.** *ACS Catalysis*, 6:7810-7813. doi: 10.1021/acscatal.6b02550” and “Knight AM, Kan SBJ, Lewis RD, Brandenberg OF, Chen K, Arnold FH (2018) **Diverse engineered heme proteins enable stereodivergent cyclopropanation of unactivated alkenes.** *ACS Cent Sci*, 4:372-377. doi: 10.1021/acscentsci.7b00548” Reprinted with permission from the American Chemical Society.

A.1 Materials and methods

All chemicals and reagents were purchased from chemical suppliers (Sigma-Aldrich, Fischer Scientific, Combi-Blocks, Alfa Aesar, TCI) and used without purification. Silica gel chromatography purifications were carried out using AMD Silica Gel 60, 230-400 mesh. ^1H and ^{13}C NMR spectra were recorded on a Bruker Prodigy 400 MHz instrument and are internally referenced to the residual solvent peak (chloroform). Data for ^1H NMR are reported in the conventional form: chemical shift (δ ppm), integration, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet), coupling constant (Hz). Data for ^{13}C are reported in terms of chemical shift (δ ppm). High-resolution mass spectra were obtained with a JEOL JMS-600H High Resolution Mass Spectrometer at the California Institute of Technology Mass Spectral Facility. Sonication was performed using a Qsonica Q500 sonicator. Chemical reactions were monitored using thin layer chromatography (Merck 60 silica gel plates) and a UV-lamp for visualization. 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. Optical rotation data were collected on a JASCO P-2000 Polarimeter.

A.2 Cloning and library generation

Plasmid pET22b was used as a cloning vector, cloning was performed using Gibson assembly¹. Cells were grown using Luria-Bertani medium or HyperBroth (AthenaES) with 100 µg/mL ampicillin (LB_{amp} or HB_{amp}). Electrocompetent *Escherichia coli* cells were prepared using E. cloni[®] Express BL-21 (DE3) cells following the protocol of Sambrook *et al.*² T5 exonuclease, Phusion polymerase, and *Taq* ligase were purchased from New England Biolabs (NEB, Ipswich, MA). Electroporation was carried out in 2 mm electroporation cuvettes using the Gene Pulser XCell Electroporation System (Bio-Rad). Codon-optimized gBlocks of the various heme binding proteins were ordered from Integrated DNA Technologies (IDT). As the putative distal ligand could interfere with the initially low substrate binding affinities, some proteins were ordered with the distal axial ligands mutated to smaller, nonpolar residues, found to be beneficial for vinylarene cyclopropanation in myoglobin³. Some gBlocks (**Table A-1**) were also ordered with a mutation in a putative entrance tunnel residue previously found to enhance styrene cyclopropanation activity in myoglobin³.

Site-saturation mutagenesis was performed using the 22-codon method⁴. 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 together via Gibson assembly. The Gibson assembly product was transformed into electrocompetent *E. cloni* EXPRESS BL21(DE3) cells (Lucigen, Middleton, WI). Aliquots of SOC medium (750 µL) were added and the cells were incubated at 37°C for 45 minutes before being plated on LB-ampicillin (100 µg mL⁻¹) agar plates.

A.3 Hemochrome assay

An aliquot of cells for protein concentration determination was sonicated for 2 minutes, 1 second on, 1 second off at 25% amplitude. Cells expressing ApePgb or RmaNOD variants had 0.1 eq. Bugbuster 10X protein extraction reagent (EMD Millipore) added prior to sonication. The sonicated lysate was clarified via centrifugation at $4500\times g$ and $4^{\circ}C$ for 10 minutes.

At room temperature, 1.75 mL pyridine was combined with 0.75 mL of 1M NaOH. The solution was vortexed for 30 seconds to ensure complete mixing then centrifuged for 1 minute at 5000 rpm to remove excess base from the pyridine solution. 0.75 mL of protein lysate in M9-N minimal medium was combined with 0.25 mL of the pyridine solution along with about 2.0 mg of sodium dithionite, the cuvette's contents were mixed with a pipet and sealed with parafilm. The content's absorbance was immediately read. The concentration of *B. subtilis* TrHb was determined using an extinction coefficient of $196 \text{ mM}^{-1} \text{ cm}^{-1}$ at 420 nm. Baseline absorbance was determined through extrapolation of two points on either side of the hemochrome peak at 390 and 450 nm.

For ApePgb and Rma NOD, 500 μL clarified lysate was sterile filtered and added to a cuvette. 500 μL of solution I (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 $23.98 \text{ mM}^{-1} \text{ cm}^{-1}$ for $(557 \text{ nm}_{\text{reduced}} - 540 \text{ nm}_{\text{oxidized}})^5$.

A.4 Growth and induction in 96 well plates

Single colonies from the LB-ampicillin agar plates were picked using sterile toothpicks and grown in 300 μL LB-ampicillin in 2 mL 96 deep-well plates at $37^{\circ}C$, 250 rpm, 80% humidity overnight (12-18 hours). Multi-channel pipettes were used to transfer 30 μL 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 μL) and 50% (v/v) sterile

glycerol (100 μ 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 isopropyl β -D-1-thiogalactopyranoside (IPTG) and supplemented with 1 mM 5-aminolevulinic acid (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 .

A.5 Growth and induction in flasks

Overnight cultures were prepared in LB_{amp} medium (100 $\mu\text{g}/\text{mL}$ ampicillin) from glycerol stocks of protein variants. The overnight cultures were incubated at 37°C for 12 hours. Flasks containing HB_{amp} medium (100 $\mu\text{g}/\text{mL}$ ampicillin) were inoculated with 1% volume of the overnight culture. Flasks were incubated at 37°C until they reached $\text{OD}_{600}=0.6-0.8$. Flasks were cooled on ice for 20 minutes, then induced with 0.5 mM final concentration IPTG and 1 mM final concentration ALA. Flasks were incubated at 25°C for 18-24 hours. Cells were harvested via centrifugation at 4000 rpm, 10 minutes, at 4°C .

A.6 Reaction screening in 96 well-plates

A.6.1 Ticagrelor

After harvesting cells, pellets were resuspended in 350 μL M9-N minimal media. In the anaerobic chamber 50 μL of 3,4 difluorostyrene (0.16 M in ethanol) and EDA (0.32 M in ethanol) were added to the whole cells in each well. The plate was sealed with a gas-impermeable membrane and allowed to shake in the anaerobic chamber at room temperature at 400 rpm for 1 hour. The plate was then removed from the anaerobic chamber, and 1 mL of cyclohexane was added to each well using a multichannel pipette and mixed up and down three times. The plates were spun for 5 minutes at 4000 rpm and 20°C to separate the organic and aqueous layers. 600 μL of the organic layer was transferred into a shallow 96-well plate and screened for enantioselectivity on the SFC. Analytical chiral supercritical fluid chromatography was performed with a JACSO 2000 series instrument using 2% *i*-PrOH and 98% supercritical CO_2 as the mobile phase (2.5 mL/min), with visualization at 210 and 254 nm. The column used was Chiralpak AD-H (4.6 mm x 250 mm).

A.6.2 Unactivated Alkenes

After harvesting cells, pellets were resuspended in 400 μ L M9-N minimal media. In an anaerobic chamber, 50 μ L reactant mixture in ethanol (final concentrations in a 450 μ L reaction: 20 mM EDA **1**, 20 mM 4-phenyl-1-butene **9**) were added to the reaction plate. The reaction plate was covered with a gas-impermeable membrane and allowed to shake in the anaerobic chamber at room temperature at 500 rpm for 3 hours. To quench the reaction and extract the substrates, 400 μ L of a mixture of acetonitrile (49 mL) and 3 M HCl (1 mL) was 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 supernatant was filtered through a 0.2 μ 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 **10** formation via HPLC using a Kromasil 100-5-C18 column, 4.6x50 mm with a 71% acetonitrile isocratic method (3 minutes). In later screening with higher enzymatic activity, the separation of *cis*- and *trans*-isomers of **10** was improved with the use of an Eclipse XDB-C18 column, 5 μ m particle size, 4.6x150 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.

A.7 Small scale biocatalytic reactions

A.7.1 Ticagrelor

Small-scale (400 μ L) reactions took place in 2 mL glass crimp vials. Whole cells expressing the enzyme were resuspended in M9-N minimal medium to the OD₆₀₀ specified in the reaction conditions. The headspace of the vials was purged during gas cycling of an anaerobic chamber. The open vials were allowed to shake in the anaerobic chamber for one minute before reagents were added. 10 μ L of 3,4-difluorostyrene (800 mM in ethanol) was

added to the reactions followed by 10 μL of ethyl diazoacetate (1.6 M in ethanol). Vials were crimp-sealed with silicone septa and removed from the anaerobic chamber and allowed to shake at 400 rpm at room temperature for 1 hour. After reaction completion, reactions to be screened with gas chromatography (GC) had 0.9 mL cyclohexane and 20 μL acetophenone (40 mM in cyclohexane) added to them. The vials were vortexed for 30 seconds two times and then centrifuged at 13000 rpm for 5 minutes at 20 $^{\circ}\text{C}$. The supernatant was transferred to a clean 2 mL glass vial for analysis by GC. Reactions to be screened via supercritical fluid chromatography (SFC) had 0.9 ml of cyclohexane added to them then were vortexed twice for 30 seconds and centrifuged at 13000 rpm for 5 minutes at 20 $^{\circ}\text{C}$. The supernatant was moved to a clean 2 mL glass vial for analysis on the SFC. Enantioselectivity was screened using SFC, while yield and diastereoselectivity were determined by GC. All reactions were performed in triplicate (unless otherwise noted).

A.7.2 Unactivated alkenes

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) was added to the vials and they were brought into a Coy anaerobic chamber (\sim 0-10 ppm O_2). 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. 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 \times g for 5 minutes at RT and the organic layer was used for chromatographic analysis.

A.8 Proteins investigated and homology modeling

ApePgb was modeled with the *Methanosarcina acetivorans* protoglobin (PDB ID: 3ZJL) and RmaNOD was modeled with *Alcaligenes eutrophus* flavohemoglobin (PDB ID: 1CQX) using SWISS-MODEL⁶

Table A-1. 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

A.9 Synthesis and characterization of authentic products

A.9.1 Ticagrelor

Racemic ticagrelor cyclopropane *rac*-(**6**) was prepared according to the following procedure: To hemin (1.63 g, 2.5 mmol) in water (20 mL) under Ar was added 3,4-difluorostyrene (1.72 g, 10 mmol). Ethyl diazoacetate (2.42 mL, 20 mmol) in DCM (3 mL)

was added dropwise to the vigorously stirred mixture, and the reaction was left to stir for 12 h. The reaction was extracted with hexane (20 mL x 3), and the organic extracts were washed with brine (50 mL), dried (MgSO₄) and concentrated *in vacuo*. Purification by silica flash chromatography (2 to 20% EtOAc in hexane) separated the desired *trans*-cyclopropane from its *cis*-isomer and gave *rac*-(**5**) as a colorless oil (1.43 g, 6.3 mmol, 63%). Absolute configuration was determined by comparing HPLC retention times to those reported in literature⁷.

¹H NMR (400 MHz, CDCl₃) δ 7.05 (1H, td, *J* = 10.3, 8.5 Hz, H₁₁), 6.94–6.74 (2H, m, H₈ and H₁₂), 4.17 (2H, q, *J* = 7.2 Hz, H₂), 2.47 (1H, ddd, *J* = 9.4, 6.4, 4.2 Hz, H₆), 1.84 (1H, ddd, *J* = 8.5, 5.3, 4.2 Hz, H₄), 1.59 (1H, ddd, *J* = 9.2, 5.3, 4.8 Hz, H_{5a}), 1.28 (3H, t, *J* = 7.1 Hz, H₁), 1.24 (1H, ddd, *J* = 8.4, 6.4, 4.8 Hz, H_{5b}).

¹³C NMR (100 MHz, CDCl₃) δ 173.1, 151.0 (dd, ¹*J*_{C-F} = 118.7 Hz, ²*J*_{C-F} = 12.8 Hz), 148.6 (dd, ¹*J*_{C-F} = 117.5 Hz, ²*J*_{C-F} = 12.7 Hz), 137.3 (dd, ³*J*_{C-F} = 5.9 Hz, ⁴*J*_{C-F} = 3.7 Hz), 122.5 (dd, ³*J*_{C-F} = 6.2 Hz, ⁴*J*_{C-F} = 3.4 Hz), 117.3 (dd, ²*J*_{C-F} = 17.2 Hz, ³*J*_{C-F} = 0.2 Hz), 115.3 (d, ²*J*_{C-F} = 17.6 Hz), 61.0, 25.3 (d, ⁴*J*_{C-F} = 1.7 Hz), 24.3, 17.1, 14.4.

A.9.2 Unactivated alkenes

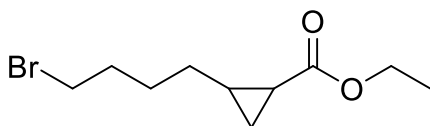
General Procedure A: Rhodium acetate dimer (10 μmol, 4.4 mg) and a stir bar were added to a 5 mL dram vial and it was sealed with a septum. The sealed vial was purged with three cycles of vacuum and argon. Neat olefin (8 mmol) was added to the vial. EDA (2 mmol) was added in a 2-hour slow addition on 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 μmol, 17 mg) and a stir bar was 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.

Absolute configurations of **8** were confirmed by comparison to literature chiral GC⁸. Cyclosil-B column, 90°C isothermal, absolute configuration of products elute in order: *cis* (*1S*, *2R*)-**8**, (*1R*, *2S*)-**8**, *trans*: (*1R*, *2R*)-**8**, (*1S*, *2S*)-**8**. 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 in the section below (**A.8**).

Ethyl 2-(4-bromobutyl)cyclopropane-1-carboxylate (**12**)



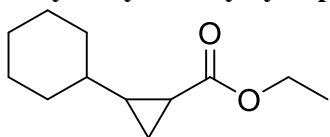
12 was synthesized with procedure A, d.r. not determined.

Cis-**12** ¹H 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). ¹³C 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 $[M - \text{CH}_3\text{CH}_2\text{O}]^+ \text{C}_8\text{H}_{12}^{79}\text{BrO}$, calculated 203.0072, found 203.0026.

12 fraction containing 71:29 *trans/cis* diastereomeric mixture, reporting shifts characterized as *trans*-**12**. ^1H 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). ^{13}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.

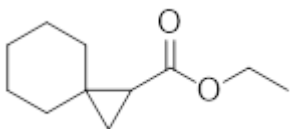
Ethyl 2-cyclohexylcyclopropane-1-carboxylate (13)



13 was synthesized with general procedure A; d.r. not determined.

Cis-**13**: ^1H NMR (300 MHz, Chloroform-*d*) δ 4.18 – 4.04 (m, 1H), 1.85 – 1.51 (m, 6H), 1.29 – 0.92 (m, 2H). ^{13}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+): $[\text{M}]^+ \text{C}_{12}\text{H}_{20}\text{O}_2$, calculated 196.1463, found 196.1464.

Ethyl spiro[2.5]octane-1-carboxylate (14)



14 was synthesized with general procedure A.

^1H NMR (400 MHz, Chloroform-*d*) δ 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). ^{13}C NMR (101 MHz, Chloroform-*d*) δ 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+): $[\text{M}]^+ \text{C}_{11}\text{H}_{18}\text{O}_2$, calculated 182.1307, found 182.1282

A.10 Calibration curves

The calibration curve for **6** was generated by performing extraction on mock reactions (spiked with authentic standard) with final concentrations of product ranging from 0 – 25 mM. Mock reactions were then extracted in 0.9 mL of cyclohexane, and 20 μ L of acetophenone (40 mM in cyclohexane). Samples were vortexed for 30 seconds two times then spun at 14,000 rpm and 20 °C for 5 minutes. The cyclohexane supernatant was transferred to 2 mL glass vials and quantified using the GC. The average of duplicate extractions was used for the GC standard curve.

Calibration curves for the unactivated alkene-derived proteins were prepared with analogous conditions to the analytical scale reactions. Cyclopropane product standards were diluted in ethanol and added (20 μ L) to M9-N buffer (380 μ 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 μ L, 3 M stock) and internal standard (16 μ L of 40 mM acetophenone in cyclohexane). Cyclohexane (700 μ 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.

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 total turnover number (TTN) per enzyme.

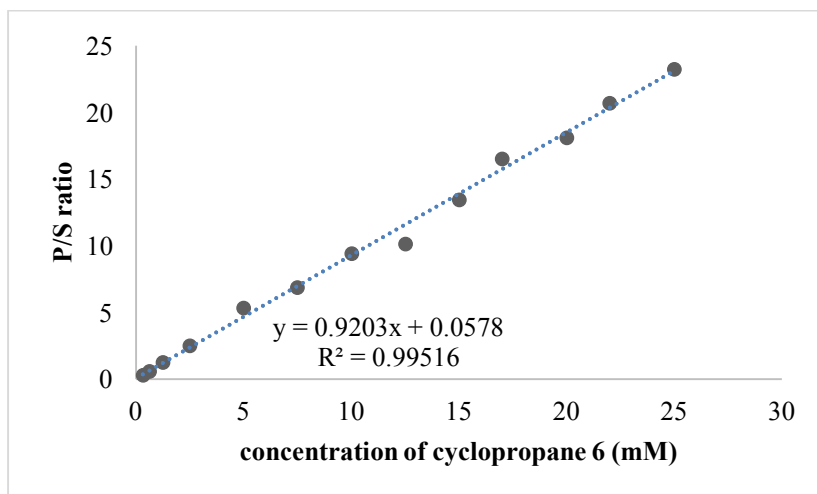


Figure A-1. Achiral GC calibration curve of **6**.

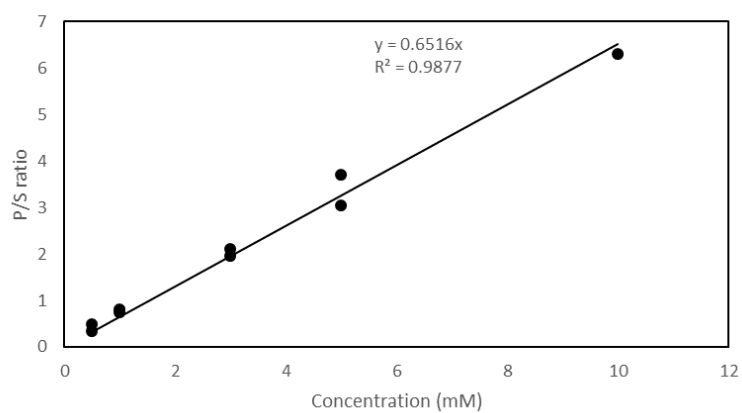


Figure A-2. Achiral GC calibration curve of **12**.

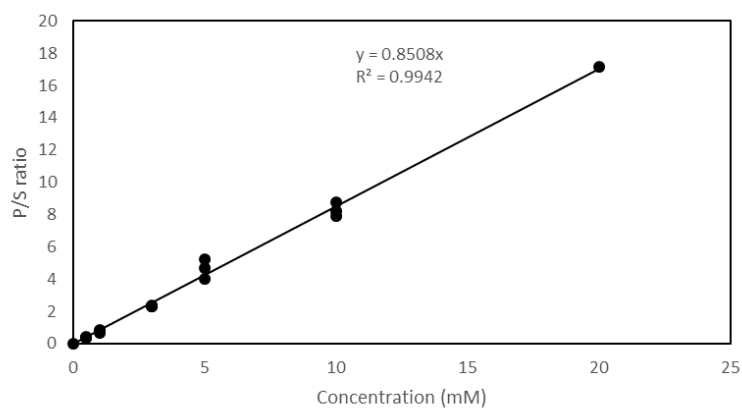


Figure A-3. Achiral GC calibration curve of **13**.

A.11 Nucleotide and amino acid sequences

B. subtilis truncated globin Y25L T45A Q49A amino acid sequence

MGQSFNAPYEAI GEELLSQLVDTFLE RVASHPLLKPIF PSDLTEAARKAKQFLTQYLGGPPLYTEEHGHPMLR
ARHLPFPITNERADAWLSCMKDAMDHVGLGEGEREFLFGRLELTARHMVNQTEAEDRSSLEHHHHHH*

B. subtilis truncated globin Y25L T45A Q49A nucleotide sequence

ATGGGTCAAAGTTTCAATGCACCATACGAAGCCATCGGGGAAGAGCTGCTCAGTCAGCTCGTTGATACGTT
CCTGGAGCGTGTGGCGTCTCATCCGCTGCTCAAACCGATTTTTCCGTCTGATCTGACGGAAGCCGCACGTA
AAGCGAAGCAGTTTTCTGACGCAATATCTCGGTGGTCCGCCGTTATACACTGAAGAACATGGTCACCCAATG
CTTCGCGCGCGGCACCTCCCATTCCCCGATACCAATGAGCGTGCAGGACGCTTGGCTCTCCTGTATGAAAGA
CGCAATGGATCACGTGGGACTGGAAGGTGAAATTCGTGAATTCCTGTTTGGTTCGTTTAGAACTGACGGCGC
GCCACATGGTCAATCAGACCGAAGCAGAAGACCGGAGCAGCCTCGAGCACCACCACCACCACCCTGA

ApePgb AGW amino acid sequence

MTPSDIPGYDYGRVEKSPI TDLEFDLLKKTVM LGEKDV MYLKKACDVLKDQVDEILDLAGGWVASNEHLIYYF
SNPDTGEP IKEYLERVRARFGAWILD TTRDYNREWL DYQYEVGLRHRSKKGVT DGVRTVPHIPLRYLIAWI
YPITATIKPFLAKKGGSPEDIEGMYNAWFKSVVLQVAIWSHPYTKENDWLEHHHHHH*

ApePgb AGW nucleotide sequence

ATGACTCCCTCGGACATCCCGGATATGATTATGGGCGTGTGCGAGAAGTCACCCATCACGGACCTTGAGTTTG
ACCTTCTGAAGAAGACTGTCA TGTTAGGTGAAAAGGACGTAATGTACTTGAAAAGGCGTGTGACGTTCTGAA
AGATCAAGTTGATGAGATCCTTGACTTGGCGGGTGGTTGGGTAGCATCAAATGAGCATTGATTTATTACTTC
TCCAATCCGGATACAGGAGAGCCTATTAAGGAATACCTGGAACGTGTACGCGCTCGCTTTGGAGCCTGGATTC
TGGACACTACCTGCCGCGACTATAACCGTGAATGGTTAGACTACCAGTACGAAGTTGGGCTTCGTCATCACCG
TTCAAAGAAAGGGGTCACAGACGGAGTACGCACCGTGCCCATATCCCACTTCGTTATCTTATCGCATGGATC
TATCCTATCACCGCCACTATCAAGCATT TTTGGCTAAGAAAGGTGGCTCTCCGGAAGACATCGAAGGGATGT
ACAACGCTTGGTTCAAGTCTGAGTTTTACAAGTTGCCATCTGGTACACCCCTTATACTAAGGAGAATGACTG
GCTCGAGCACCACCACCACCACCCTGA

RmaNOD Q52V amino acid sequence

MAPTLSEQTRQLVRASVPALQKHSVAISATMYRLLFERYPETRSLFELPERVIHKLASALLAYARSIDNPSAL
QAAIRRMVLSHARAGVQAVHYPLVWECLRD AIKEVLGPDATETLLQAWKEAYDFLAHLLSTKEAQVYAVLAEL
EHHHHHH*

RmaNOD Q52V nucleotide sequence

ATGGCGCCGACCCTGTCGGAACAGACCCGTCAGTTGGTACGTGCGTCTGTGCCTGCACTGCAGAAACACTCAG
TCGCTATTAGCGCCACGATGTATCGGCTGCTTTTCGAACGGTATCCCGAAACCGGAGCTTGTGTTGAACCTCC
TGAGAGAGTTATACACAAGCTTGCGTCGGCCCTGTTGGCCTACGCCCGTAGTATCGACAACCCATCGGCGTTA
CAGGCGGCCATCCGCCGCATGGTGCTTTCCACGCACGCGCAGGAGTGCAGGCCGTCATTATCCGCTGGTTT
GGGAATGTTTGAGAGACGCTATAAAAAGAGTCTTGGGCCCCGATGCCACCGAGACCCTTCTGCAGGCGTGGAA
GGAAGCCTATGATTTTTTAGCTCATTTACTGTCTACCAAGGAAGCGCAAGTCTACGCTGTGTTAGCTGAAC
GAGCACCACCACCACCACCCTGA

A.12 References

1. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
2. Sambrook, J. & W Russell, D. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harb. Lab. Press. Cold Spring Harb. NY 999 (2001). doi:10.1016/0092-8674(90)90210-6

3. Bordeaux, M., Tyagi, V. & Fasan, R. Highly diastereoselective and enantioselective olefin cyclopropanation using engineered myoglobin-based catalysts. *Angew. Chem. Int. Ed.* **54**, 1744–1748 (2015).
4. Kille, S. *et al.* Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.* **2**, 83–92 (2013).
5. Berry, E. A. & Trumppower, B. L. Simultaneous determination of hemes a, b, and c from pyridine hemochrome spectra. *Anal. Biochem.* **161**, 1–15 (1987).
6. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* **31**, 3381–3385 (2003).
7. Hugentobler, K. G., Sharif, H., Rasparini, M., Heath, R. S. & Turner, N. J. Biocatalytic approaches to a key building block for the anti-thrombotic agent ticagrelor. *Org. Biomol. Chem.* **14**, 8064–8067 (2016).
8. 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* **534**, 534–537 (2016).

*Chapter 2*DIRECTED EVOLUTION OF CYTOCHROME C FOR CARBON–
SILICON BOND FORMATION

Material from this chapter appears in “Kan SBJ, Lewis RD, Chen K, Arnold FH. (2016) **Directed evolution of cytochrome *c* for silicon–carbon bond formation: Bringing silicon to life.** *Science*, 354:1048-1051. doi: 10.1126/science.aah6219” Reprinted with permission from the American Association for the Advancement of Science.

2.1 Abstract

Enzymes that catalyze carbon–silicon bond formation are unknown in nature, despite the natural abundance of both elements. Such enzymes would expand the catalytic repertoire of biology, enabling living systems to access chemical space previously only open to synthetic chemistry. We have discovered that heme proteins catalyze the formation of organosilicon compounds under physiological conditions via carbene insertion into silicon–hydrogen bonds. In particular, we found that the cytochrome *c* from *Rhodothermus marinus* is capable of catalyzing the reaction with high enantioselectivity. Using directed evolution, we enhanced the catalytic function of cytochrome *c* from *Rhodothermus marinus* to achieve more than 15-fold higher turnover than state-of-the-art synthetic catalysts, and find that the evolved protein can produce a broad range of enantiopure organosilicon products, and displays high chemoselectivity for functionalization of the silicon–hydrogen bond. This carbon–silicon bond-forming biocatalyst offers an environmentally friendly, efficient and selective route to producing organosilicon molecules.

2.2 Introduction

Silicon is ubiquitous in the earth's biosphere, but biochemistry is almost completely devoid of reactions that involve silicon^{1,2}. Some organisms use silicate as an integral structural component in their biology, but even in these organisms, enzymes that can form bonds between carbon and silicon atoms are unknown^{1,2}. Despite the absence of organosilicon compounds in the biological world, synthetic chemists have found that organosilicon compounds have distinctive and desirable properties that have led to their broad utility in synthetic chemistry and material science^{3,4}. Organosilicon compounds also have potential uses in medicinal chemistry; as a biocompatible carbon isostere, silicon can also be used to optimize and repurpose the pharmaceutical properties of bioactive molecules^{5,6}.

Despite utility of organosilicon compounds and the natural abundance of silicon, there is a dearth of sustainable methods for synthesizing organosilicon compounds⁷⁻⁹. In particular, methods that can enantioselectively introduce silicon motifs to organic molecules via carbon-silicon bond formation rely on multistep synthetic campaigns to prepare and optimize chiral reagents or catalysts, and precious metals are also sometimes needed to achieve the desired activity¹⁰⁻¹⁶. Carbene transfer to silicon-hydrogen bonds is one such carbon-silicon bond-forming reaction that can be rendered enantioselective using chiral transition metal complexes based on rhodium^{12,13}, iridium¹⁴, and copper^{15,16}. These catalysts can provide optically pure products, but not without limitations: they have limited turnovers (<100), require halogenated solvents, and sometimes require low temperatures for optimum selectivity (Appendix B, **Table B-1**).

2.3 Results and discussion

2.3.1 Heme catalyzed carbon-silicon bond formation

Based on previous work demonstrating that heme proteins can function as carbene transfer biocatalysts for cyclopropanation reactions¹⁷⁻²¹, we hypothesized that heme proteins might also catalyze carbene insertion into silicon-hydrogen bonds. Because iron is not known to catalyze this transformation (using iron, only stoichiometric carbene transfer to Si-

H bonds has been reported²²), we first examined whether free heme could function as a catalyst in aqueous media. Initial experiments showed that the reaction between phenyldimethylsilane and ethyl 2-diazopropanoate (Me-EDA) in neutral buffer (M9-N minimal medium, pH 7.4) at room temperature gave racemic organosilicon product **3** at very low levels, a total turnover number (TTN) of 4 (**Figure 2-1A**). No product formation was observed in the absence of heme, and the organosilicon product was stable under the reaction conditions.

We next investigated whether heme proteins could catalyze the same carbon–silicon bond-forming reaction. We screened a panel of heme proteins, including both wild-type proteins and those previously evolved to catalyze carbene transfer reactions. We observed that many heme proteins could catalyze our test reaction with more turnovers than the control catalysts, (hemin and hemin with bovine serum albumin (BSA)) but almost all proteins tested had negligible enantioinduction (Table S4). However, we observed that the wild-type cytochrome *c* from *Rhodothermus marinus* (*Rma cyt c*) catalyzed the reaction with 97% enantiomeric excess (*ee*), indicating that the reaction took place in an environment where the protein exerted excellent stereocontrol. *Rma cyt c* is highly thermostable protein, and in its native environment it serves as an electron-transfer protein in the respiratory electron transport chain²³. Like other bacterial cytochrome *c* proteins, it is not known to have any catalytic function in living systems²⁴, but when we tested other cytochrome *c* proteins (both bacterial and eukaryotic) for the carbon–silicon bond forming reaction, all had substantially lower enantioselectivity. We thus chose *Rma cyt c* as the platform for evolving a carbon–silicon bond-forming enzyme.

2.3.2 Evolution of *Rhodothermus marinus* cytochrome *c*

The crystal structure of wild-type *Rma cyt c* [Protein Data Bank (PDB) ID: 3CP5]²³ reveals that the heme prosthetic group resides in a hydrophobic pocket, with the iron axially coordinated to a proximal His (H49) and a distal Met (M100), the latter of which is located on a loop (**Figure 2-1B, 2-1C**). The distal Met, common in cytochrome *c* proteins, is coordinately labile^{25,26}. We hypothesized that M100 must be displaced in order to allow for iron-carbene formation, and that mutation of this amino acid could facilitate formation of

this adventitious “active site” and yield an improved carbon–silicon bond-forming biocatalyst. Therefore, a variant library made by site-saturation mutagenesis of M100 was cloned and recombinantly expressed in *E. coli*. After protein expression, the bacterial cells were heat treated (75°C for 10 min) before screening in the presence of phenyldimethylsilane (10 mM), Me-EDA (10 mM), and sodium dithionite ($\text{Na}_2\text{S}_2\text{O}_4$, 10 mM) as a reducing agent, at room temperature under anaerobic conditions. The M100D mutation stood out as highly activating: this first-generation mutant provided chiral organosilicon **3** as a single enantiomer in 550 TTN, a 12-fold improvement over the wild-type protein (**Figure 2-1D**).

Further analysis of the wild-type *Rma* cyt *c* crystal structure revealed that amino acid residues V75 and M103 reside close (within 7Å) to the heme iron center in wild-type *Rma* cyt *c*. Sequential site-saturation mutagenesis at these positions in the M100D mutant led to the discovery of triple-mutant V75T M100D M103E, which catalyzed carbon–silicon bond formation in >1500 turnovers and >99% ee. This level of activity is more than 15 times the total turnovers reported for the best synthetic catalysts for this class of reaction (Appendix B, **Table B-1**). As standalone mutations, both V75T and M103E are activating for wild-type *Rma* cyt *c*, and the beneficial effects increase with each combination (Appendix B, **Table B-3**). Comparison of the initial reaction rates established that each round of evolution enhanced the rate: relative to the wild-type protein, the evolved triple mutant catalyzes the reaction more than seven times faster, with turnover frequency (TOF) of 46 min^{-1} (**Figure 2-1E**).

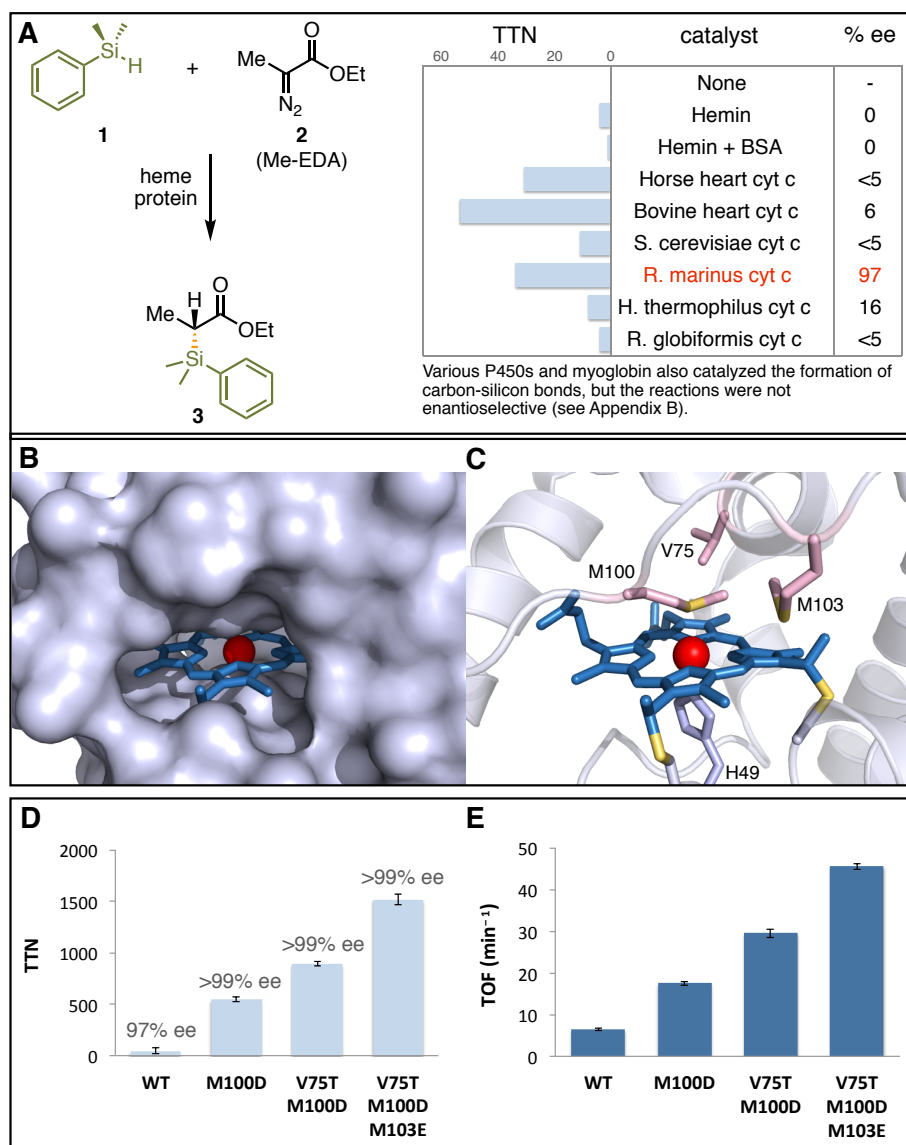


Figure 2-1. Heme protein-catalyzed carbon–silicon bond formation. **A.** Carbon–silicon bond formation catalyzed by heme and purified heme proteins. **B.** Surface representation of the heme-binding pocket of wild-type *Rma* cyt *c* (PDB ID: 3CP5). **C.** “Active site” structure of wild-type *Rma* cyt *c* showing a covalently bound heme cofactor ligated by axial ligands H49 and M100. Amino acid residues M100, V75, and M103 residing close to the heme iron were subjected to site- saturation mutagenesis. **D.** Directed evolution of *Rma* cyt *c* for carbon–silicon bond formation [reaction shown in (A)]. Experiments were performed using lysates of *E. coli* expressing *Rma* cyt *c* variant (OD₆₀₀ = 15; heat-treated at 75°C for 10 min), 10 mM silane, 10 mM diazo ester, 10 mM Na₂S₂O₄, 5 vol % MeCN, M9-N buffer (pH 7.4) at room temperature under anaerobic conditions for 1.5 hours. Reactions were done in triplicate. **E.** Carbon–silicon bond forming rates over four generations of *Rma* cyt *c*.

2.3.3 Substrate scope

Assaying the new enzyme against a panel of silicon and diazo reagents, we found that the mutations were broadly activating for enantioselective carbon–silicon bond formation. The reaction substrate scope was surveyed with the use of heat-treated lysates of *E. coli* expressing *Rma cyt c* V75T M100D M103E (*Rma* TDE), with 10 mM of both silane and diazo ester to determine TTN. Whereas many natural enzymes excel at catalyzing reactions on only their native substrates and little else (especially primary metabolic enzymes), *Rma* TDE catalyzed the formation of 20 silicon-containing products, most of which were obtained cleanly as single enantiomers, demonstrating the broad substrate scope of this reaction with just a single variant of the enzyme (**Figure 2-2**).

2.3.4 Chemoselectivity changes over the course of evolution

The evolved *Rma cyt c* also exhibits high specificity for carbon–silicon bond formation. Even in the presence of carbene-reactive functional groups (e.g. alkenes, alkynes, O–H bonds, N–H bonds), the enzyme selectively reacts with the Si–H functionality. In contrast, when the same reactants were subjected to rhodium catalysis [1 mol % $\text{Rh}_2(\text{OAc})_4$], O–H and N–H insertions were the predominant reaction pathways, and copper catalysis [10 mol% $\text{Cu}(\text{OTf})_2$] gave complex mixtures of products.

We next asked whether all *Rma cyt c* variants would catalyze carbon–silicon bond formation selectively over insertion of the carbene into an N–H bond in the same substrate. We reexamined the evolutionary lineage and tested all four generations of *Rma cyt c* (wild-type, M100D, V75T M100D, and V75T M100D M103E) with Me-EDA and 4-(dimethylsilyl)aniline (**23**), a reagent that could serve as both nitrogen and silicon donor, to probe the proteins' bond-forming preferences. The wild-type cytochrome *c* exhibited a slight preference for forming amination product **24** over organosilicon product **22**. Even though silane **23** was not used for screening, and the *Rma cyt c*, therefore, did not undergo direct selection for chemoselectivity, each round of evolution effected a distinct shift from amination to carbon–silicon bond-forming activity (**Figure 2-3**). This evolutionary path that focused solely on increasing desired product formation culminated in a catalyst that

channeled most of the reactants (97%) through carbon–silicon bond formation (>30-fold improved with respect to the wild type), presumably by improving the orientation and binding of the silicon donor.

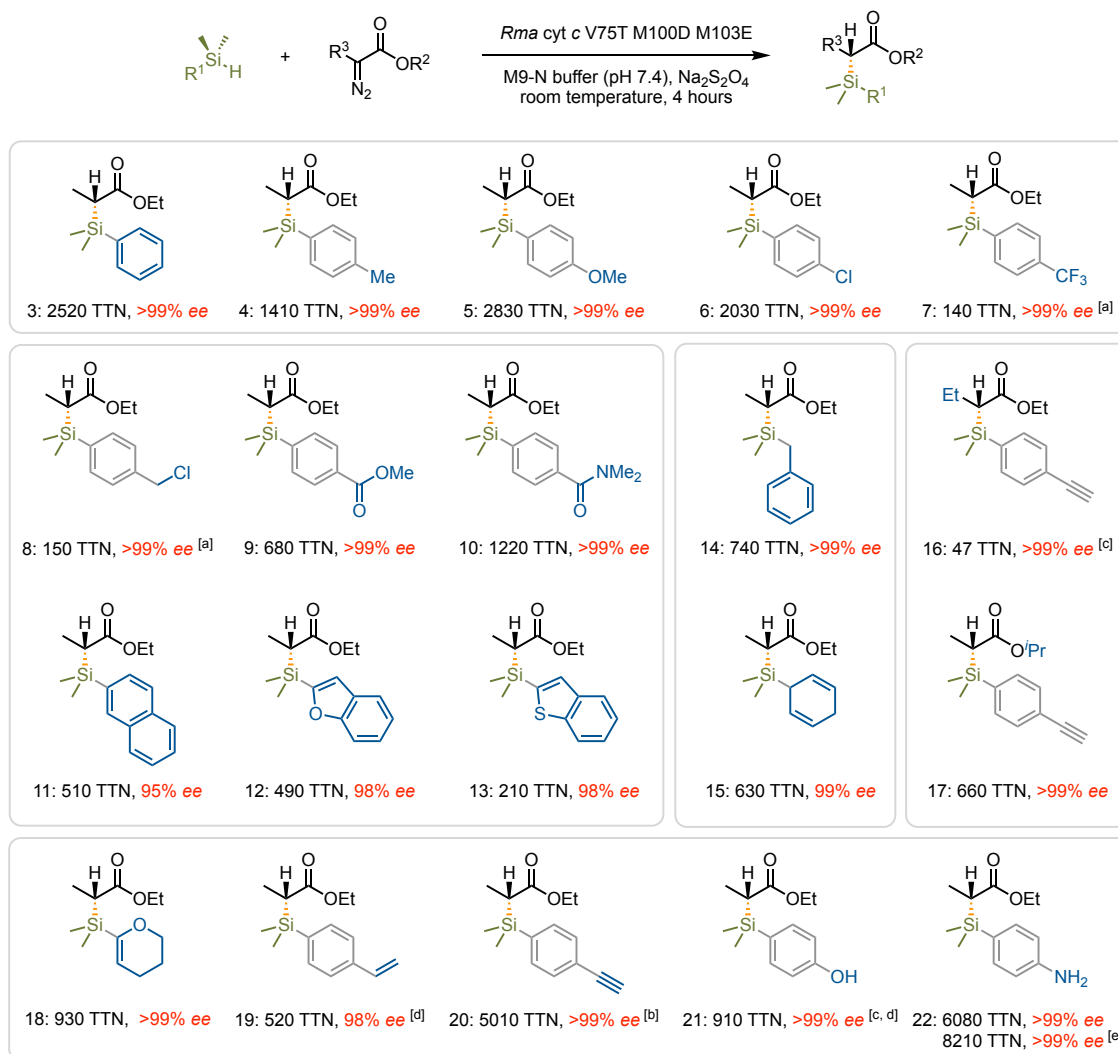


Figure 2-2. Scope of *Rma cyt c* V75T M100D M103E-catalyzed carbon–silicon bond formation. Standard reaction conditions: lysate of *E. coli* expressing *Rma cyt c* V75T M100D M103E ($OD_{600} = 1.5$; heat-treated at 75°C for 10 min), 20 mM silane, 10 mM diazo ester, 10 mM $Na_2S_2O_4$, 5% v/v MeCN, M9-N buffer (pH 7.4) at room temperature under anaerobic conditions. Reactions performed in triplicate. [a] $OD_{600} = 5$ lysate. [b] $OD_{600} = 0.5$ lysate. [c] $OD_{600} = 15$ lysate. [d] 10 mM silane. [e] $OD_{600} = 0.15$ lysate.

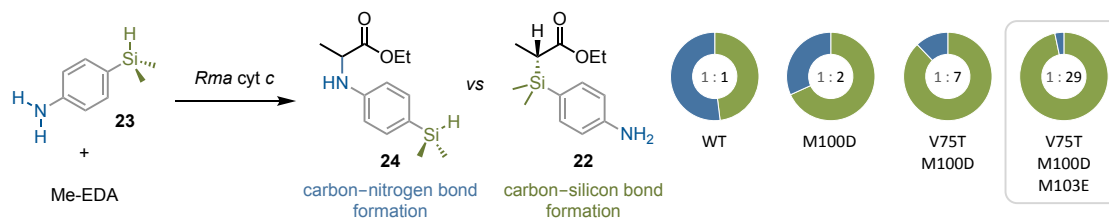


Figure 2-3. Chemoselectivity of evolved *Rma cyt c*. Chemoselectivity for carbene Si–H insertion over N–H insertion increased markedly during directed evolution of *Rma cyt c*. Standard reaction conditions as described in Figure 2-2. Reactions were performed in duplicate using heat-treated lysates of *E. coli* expressing *Rma cyt c* with protein concentration normalized across variants. Product distribution was quantified after 2 hours of reaction time (before complete conversion, no double insertion product was observed under these conditions).

2.3.5 Hypothesized substrate binding modes

From our work on the substrate scope of *Rma cyt c* V75T M100D M103E (*Rma* TDE), we concluded that *Rma* TDE has an overall high tolerance for different silane and carbene precursor substrates, except for substitution of the α -methyl group on Me-EDA (**Figure 2-4A**). This suggested that the methyl group may be buried in the protein, and larger groups cause steric clashes with the protein that impede catalysis. Based on this information and the published crystal structure of the wild-type protein, we hypothesized that the iron porphyrin carbene forms in the enzyme such that the methyl group on the carbene occupies the same space as the methyl group of residue M100 in the wild-type structure (**Figure 2-4B, 2-4C**). Then, if the silane substrate approaches from the more solvent-exposed face of the carbene (the M103 face) the transition state would be expected to be pro-*R*, which matches the observed absolute configuration of the organosilicon products produced by *Rma* TDE.

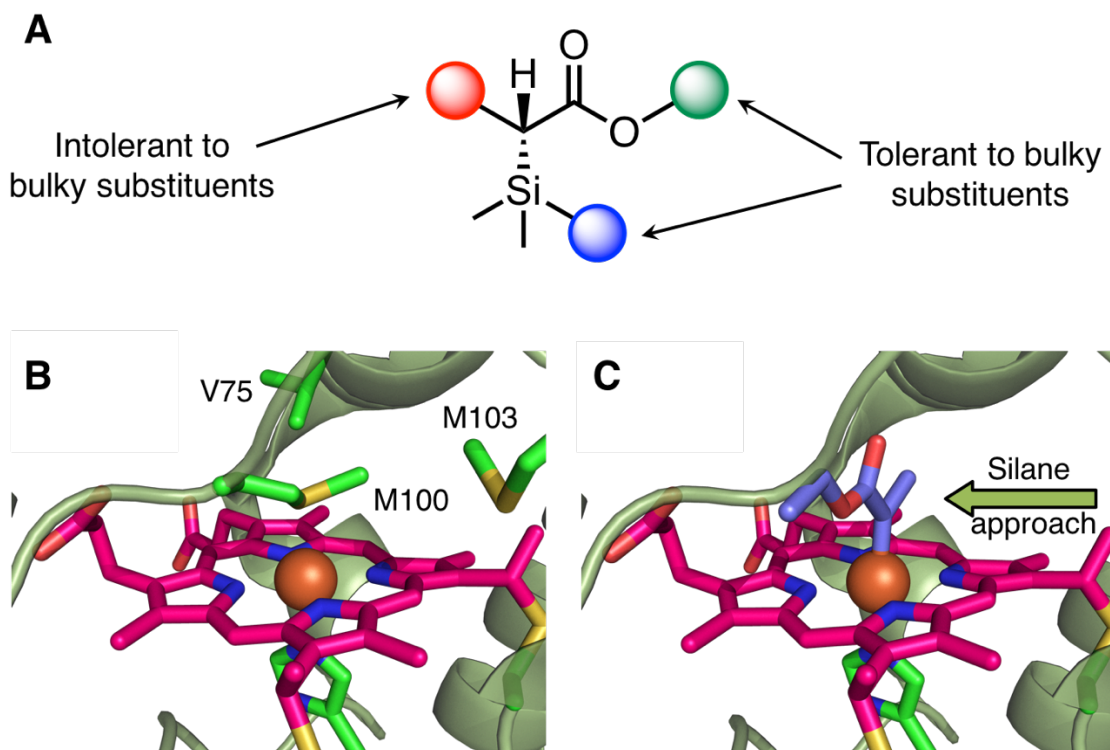


Figure 2-4. Proposed model for carbene orientation. **A.** Groups on the silane and a group on the ester are found to be highly tolerant to substitution, which may suggest that these groups are not in the active site and solvent-exposed during catalysis. **B.** Structure of wild-type *Rma* cyt *c* showing residues V75, M100, and M103 (PDB ID: 3CP5). **C.** Proposed binding mode for the iron-carbene, where the carbene forms in the place of the distal methionine. The silane is proposed to approach from the more solvent-exposed side in the wild-type protein, which would result in the observed *R* stereochemistry of the organosilicon products. The V75T, M100D, and M103E mutations may promote reactivity by improving solvent and substrate access to the iron center.

2.4 Conclusions

This work developing a highly efficient and selective iron-dependent enzyme that catalyzes carbon–silicon bond formation affirms the notion that nature’s protein repertoire is highly evolvable and poised for adaptation, including proteins that have no known catalytic activity. With only a few mutations, existing heme proteins can be repurposed to efficiently perform carbene transfer reactions. The substrate scope of the evolved protein and available crystal structure of the wild-type protein provide clues as to how the active site of the protein may be structured, and how the reaction occurs in the active site. As *Rma* cytochrome *c* does

not naturally bind substrates or have biologically relevant enzymatic function, further investigations into the effects of the mutations introduced to the protein may provide insight into how new families of enzymes evolve in nature, and how evolution can access to areas chemical space that have not yet been explored by biological systems.

2.5 References

1. Frampton, M. B. & Zelisko, P. M. Organosilicon biotechnology. *Silicon* **1**, 147–163 (2009).
2. Frampton, M. B. & Zelisko, P. M. Biocatalysis in Silicon Chemistry. *Chem. Asian J.* **12**, 1153–1167 (2017).
3. Rappoport, Z. & Apeloig, Y. *The Chemistry of Organic Silicon Compounds. The Chemistry of Organic Silicon Compounds* (2003). doi:10.1002/0470857250
4. Ponomarenko, S. A. & Kirchmeyer, S. in *Silicon Polymers* (ed. Muzafarov, A. M.) 33–110 (Springer Berlin Heidelberg, 2011). doi:10.1007/12_2009_48
5. Showell, G. A., Mills, J. S. & Showell, G. A. Silicon isosteres in drug discovery, *DrugDiscToday* 2003. **8**, 551–556 (2003).
6. Franz, A. K. & Wilson, S. O. Organosilicon molecules with medicinal applications. *J. Med. Chem.* **56**, 388–405 (2013).
7. Anastas, P. T. & Warner, J. C. *Green Chemistry: Theory and Practice. Oxford Univ. Press. New York* 30 (1998).
8. Tondreau, A. M. *et al.* Iron catalysts for selective anti-Markovnikov alkene hydrosilylation using tertiary silanes. *Science* **335**, 567–570 (2012).
9. Toutov, A. A. *et al.* Silylation of C-H bonds in aromatic heterocycles by an Earth-abundant metal catalyst. *Nature* **518**, 80–84 (2015).
10. Marciniak, B., Maciejewski, H. & Pietraszuk, P. P. *Hydrosilylation: A Comprehensive Review on Recent Advances. Springer* **1**, (2013).
11. Lee, T. & Hartwig, J. F. Rhodium-Catalyzed Enantioselective Silylation of Cyclopropyl C–H Bonds. *Angew. Chem. Int. Ed.* **55**, 8723–8727 (2016).
12. Sambasivan, R. & Ball, Z. T. Metallopeptides for asymmetric dirhodium catalysis. *J. Am. Chem. Soc.* **132**, 9289–9291 (2010).

13. Chen, D., Zhu, D. X. & Xu, M. H. Rhodium(I)-Catalyzed Highly Enantioselective Insertion of Carbenoid into Si-H: Efficient Access to Functional Chiral Silanes. *J. Am. Chem. Soc.* **138**, 1498–1501 (2016).
14. Yasutomi, Y., Suematsu, H. & Katsuki, T. Iridium(III)-catalyzed enantioselective Si-H bond insertion and formation of an enantioenriched silicon center. *J. Am. Chem. Soc.* **132**, 4510–4511 (2010).
15. Zhang, Y. Z., Zhu, S. F., Wang, L. X. & Zhou, Q. L. Copper-catalyzed highly enantioselective carbenoid insertion into Si-H bonds. *Angew. Chem. Int. Ed.* **47**, 8496–8498 (2008).
16. Hyde, S. *et al.* Copper-Catalyzed Insertion into Heteroatom-Hydrogen Bonds with Trifluorodiazalkanes. *Angew. Chem. Int. Ed.* **55**, 3785–3789 (2016).
17. Coelho, P. S., Brustad, E. M., Kannan, A. & Arnold, F. H. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* **339**, 307–310 (2013).
18. Bordeaux, M., Tyagi, V. & Fasan, R. Highly diastereoselective and enantioselective olefin cyclopropanation using engineered myoglobin-based catalysts. *Angew. Chem. Int. Ed.* **54**, 1744–1748 (2015).
19. Coelho, P. S. *et al.* A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo. *Nat. Chem. Biol.* **9**, 485–7 (2013).
20. Wang, Z. J. *et al.* Improved cyclopropanation activity of histidine-ligated cytochrome P450 enables the enantioselective formal synthesis of levomilnacipran. *Angew. Chem. Int. Ed.* **53**, 6810–6813 (2014).
21. Hernandez, K. E. *et al.* Highly stereoselective biocatalytic synthesis of key cyclopropane intermediate to ticagrelor. *ACS Catal.* 7810–7813 (2016). doi:10.1021/acscatal.6b02550
22. Scharrer, E. & Brookhart, M. Insertion reactions of electrophilic iron carbene complexes with organosilanes: a synthetic and mechanistic study. *J. Organomet. Chem.* **497**, 61–71 (1995).
23. Stelter, M. *et al.* A novel type of monoheme cytochrome c: Biochemical and structural characterization at 1.23 Å resolution of *Rhodothermus marinus* cytochrome c. *Biochemistry* **47**, 11953–11963 (2008).
24. Kleingardner, J. G. & Bren, K. L. Biological Significance and Applications of Heme c Proteins and Peptides. *Acc. Chem. Res.* **48**, 1845–1852 (2015).
25. Levin, B. D., Walsh, K. A., Sullivan, K. K., Bren, K. L. & Elliott, S. J. Methionine ligand lability of homologous monoheme cytochromes c. *Inorg. Chem.* **54**, 38–46

(2015).

26. Zaidi, S., Hassan, M. I., Islam, A. & Ahmad, F. The role of key residues in structure, function, and stability of cytochrome-c. *Cellular and Molecular Life Sciences* **71**, 229–255 (2014).

Appendix B

SUPPORTING INFORMATION FOR CHAPTER 2

Material from this chapter appears in “Kan SBJ, Lewis RD, Chen K, Arnold FH. (2016) **Directed evolution of cytochrome *c* for silicon–carbon bond formation: Bringing silicon to life.** *Science*, 354:1048-1051. doi: 10.1126/science.aah6219” Reprinted with permission from the American Association for the Advancement of Science.

B.1 Materials and methods

Unless otherwise noted, all chemicals and reagents for chemical reactions were obtained from commercial suppliers (Acros, Arch Bioscience, Fisher Scientific, Sigma-Aldrich, TCI America) and used without further purification. The following proteins were all purchased from Sigma-Aldrich: bovine serum albumin (BSA), cytochrome *c* (from bovine, equine heart and *S. cerevisiae*), peroxidase II (from horseradish), and catalase (from *C. glutamicum*). Silica gel chromatography purifications were carried out using AMD Silica Gel 60, 230-400 mesh. ^1H and ^{13}C NMR spectra were recorded on a Bruker Prodigy 400 MHz instrument and are internally referenced to the residual solvent peak (chloroform). ^{29}Si NMR spectra were recorded on the same instrument and referenced to tetramethoxysilane ($\delta -78.9$ ppm). Data for ^1H NMR are reported in the conventional form: chemical shift (δ ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, hept = heptet, m = multiplet, br = broad, app = appears as), coupling constant (Hz), integration. Data for ^{13}C and ^{29}Si are reported in terms of chemical shift (δ ppm). High-resolution mass spectra were obtained with a JEOL JMS-600H High Resolution Mass Spectrometer at the California Institute of Technology Mass Spectral Facility. Sonication was performed using a Qsonica Q500 sonicator. Chemical reactions were monitored using thin layer chromatography (Merck 60 silica gel plates) and a UV-lamp for visualization. Gas chromatography (GC) analyses were carried out using a Shimadzu GC-17A gas chromatograph, a FID detector, and J&W HP-5 (30 m x 0.32 mm, 0.25 μm film; 90 $^\circ\text{C}$ hold 1 min, 90 to 110 $^\circ\text{C}$ at 15 $^\circ\text{C}/\text{min}$, 110 to 280 $^\circ\text{C}$ at 60 $^\circ\text{C}/\text{min}$, 280 $^\circ\text{C}$ hold 1 min, 6.2 min total). Analytical chiral supercritical fluid chromatography (SFC)

was performed with a JACSO 2000 series instrument using *i*-PrOH and supercritical CO₂ as the mobile phase, with visualization at 210 nm. The following chiral columns were used: Daicel Chiralpak IC, Chiralpak AD-H, or Chiralcel OD-H (4.6 mm x 25 cm).

Plasmid pET22 was used as a cloning vector, and cloning was performed using Gibson assembly¹. The cytochrome *c* maturation plasmid pEC86² was used as part of a two-plasmid system to express prokaryotic cytochrome *c* proteins. Cells were grown using Luria-Bertani medium or HyperBroth (AthenaES) with 100 µg/mL ampicillin and 20 µg/mL chloramphenicol (LB_{amp/chlor} or HB_{amp/chlor}). Cells without the pEC86 plasmid were grown with 100 µg/mL ampicillin (LB_{amp} or HB_{amp}). Primer sequences are available upon request. Electrocompetent *Escherichia coli* cells were prepared following the protocol of Sambrook *et al.*³. T5 exonuclease, Phusion polymerase, and *Taq* ligase were purchased from New England Biolabs (NEB, Ipswich, MA). M9-N minimal medium (abbreviated as M9-N buffer; pH 7.4) was used as a buffering system for whole cells, lysates, and purified proteins, unless otherwise specified. M9-N buffer was used without a carbon source; it contains 47.7 mM Na₂HPO₄, 22.0 mM KH₂PO₄, 8.6 mM NaCl, 2.0 mM MgSO₄, and 0.1 mM CaCl₂.

B.2 Plasmid construction

All variants described in this paper were cloned and expressed using the pET22(b)+ vector (Novagen). The gene encoding *Rma* cyt *c* (UNIPROT ID B3FQS5)⁴ was obtained as a single gBlock (IDT), codon-optimized for *E. coli*, and cloned using Gibson assembly into pET22(b)+ (Novagen) between restriction sites *Nde*I and *Xho*I in frame with an *N*-terminal pelB leader sequence (to ensure periplasmic localization and proper maturation; MKYLLPTAAAGLLLLAAQPAMA) and a *C*-terminal 6xHis-tag. This plasmid was co-transformed with the cytochrome *c* maturation plasmid pEC86² into *E. coli*[®] EXPRESS BL21(DE3) cells (Lucigen).

B.3 Cytochrome *c* expression and purification

Purified cytochrome *c* proteins were prepared as follows. One liter HB_{amp/chlor} in a 4 L flask was inoculated with an overnight culture (20 mL, LB_{amp/chlor}) of recombinant *E. coli*[®]

EXPRESS BL21(DE3) cells containing a pET22(b)+ plasmid encoding the cytochrome *c* variant, and the pEC86 plasmid. The culture was shaken at 37 °C and 200 rpm (no humidity control) until the OD₆₀₀ was 0.7 (approximately 3 hours). The culture was placed on ice for 30 minutes, and isopropyl β-D-1-thiogalactopyranoside (IPTG) and 5-aminolevulinic acid (ALA) were added to final concentrations of 20 μM and 200 μM respectively.

The incubator temperature was reduced to 20 °C, and the culture was allowed to shake for 20 hours at 200 rpm. Cells were harvested by centrifugation (4 °C, 15 min, 4,000xg), and the cell pellet was stored at -20 °C until further use (at least 24 hours). The cell pellet was resuspended in buffer containing 100 mM NaCl, 20 mM imidazole, and 20 mM Tris-HCl buffer (pH 7.5 at 25 °C) and cells were lysed by sonication (2 minutes, 2 seconds on, 2 seconds off, 40% duty cycle; Qsonica Q500 sonicator). Cell lysate was placed in a 75 °C heat bath for 10 minutes, and cell debris was removed by centrifugation for 20 min (5000xg, 4 °C). Supernatant was sterile filtered through a 0.45 μm cellulose acetate filter and purified using a 1 mL Ni-NTA column (HisTrap HP, GE Healthcare, Piscataway, NJ) using an AKTA purifier FPLC system (GE healthcare). The cytochrome *c* protein was eluted from the column by running a gradient from 20 to 500 mM imidazole over 10 column volumes.

The purity of the collected cytochrome *c* fractions was analyzed using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Pure fractions were pooled and concentrated using a 3 kDa molecular weight cut-off centrifugal filter and dialyzed overnight into 0.05 M phosphate buffer (pH = 7.5) using 3 kDa molecular weight cut-off dialysis tubing. The dialyzed protein was concentrated again, flash-frozen on dry ice, and stored at -20 °C.

The concentration of cytochrome *c* was determined in triplicate using the ferrous assay described in section **B.5**.

B.4 P450 and globin expression and purification

Purified P450s and globins were prepared differently from the cytochrome *c* proteins, and described as follows. One liter HB_{amp} in a 4 L flask was inoculated with an overnight culture (20 mL, LB_{amp}) of recombinant *E. cloni*[®] EXPRESS BL21(DE3) cells containing a

pET22(b)+ plasmid encoding the P450 or globin variant. The culture was shaken at 37 °C and 200 rpm (no humidity control) until the OD₆₀₀ was 0.7 (approximately 3 hours). The culture was placed on ice for 30 minutes, and IPTG and 5-ALA were added to final concentrations of 0.5 mM and 1 mM, respectively. The incubator temperature was reduced to 20 °C, and the culture was allowed to shake for 20 hours at 200 rpm. Cells were harvested by centrifugation (4 °C, 15 min, 4,000xg), and the cell pellet was stored at -20 °C until further use (at least 24 hours). The cell pellet was resuspended in buffer containing 100 mM NaCl, 20 mM imidazole, and 20 mM Tris-HCl buffer (pH 7.5 at 25 °C). Hemin (30 mg/mL, 0.1 M NaOH; Frontier Scientific) was added to the resuspended cells such that 1 mg of hemin was added for every 1 gram of cell pellet. Cells were lysed by sonication (2 minutes, 1 seconds on, 2 seconds off, 40% duty cycle; Qsonica Q500 sonicator). Cell debris was removed by centrifugation for 20 min (27,000xg, 4 °C). Supernatant was sterile filtered through a 0.45 µm cellulose acetate filter, and purified using a 1 mL Ni-NTA column (HisTrap HP, GE Healthcare, Piscataway, NJ) using an AKTA purifier FPLC system (GE healthcare). The P450 and globin proteins were eluted from the column by running a gradient from 20 to 500 mM imidazole over 10 column volumes.

The purity of the collected protein fractions was analyzed using SDS-PAGE. Pure fractions were pooled and concentrated using a 10 kDa molecular weight cut-off centrifugal filter and buffer-exchanged with 0.1 M phosphate buffer (pH = 8.0). The purified protein was flash-frozen on dry ice and stored at -20 °C. P450 and globin concentrations were determined in triplicate using the hemochrome assay described in section **B.5**

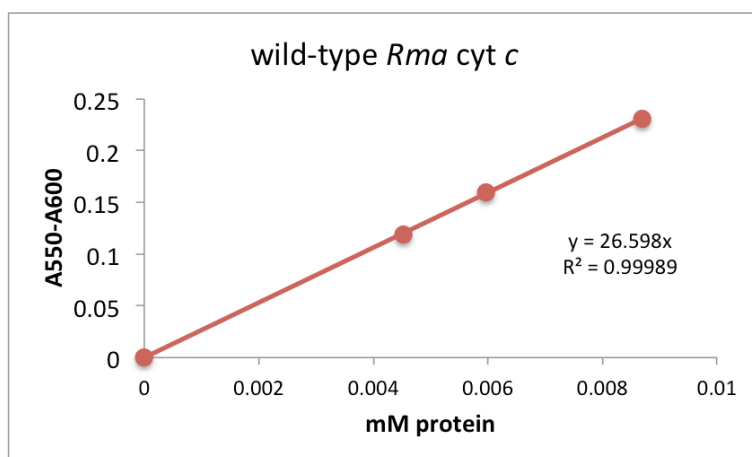
B.5 Hemochrome assay

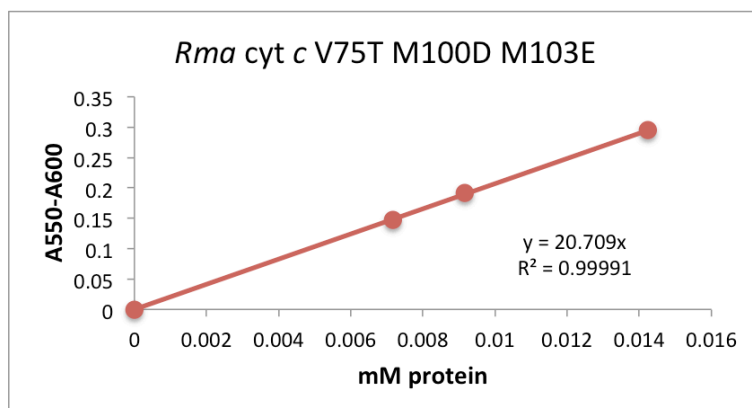
A solution of sodium dithionite (10 mg/mL) was prepared in M9-N buffer. Separately, a solution of 1 M NaOH (0.4 mL) was mixed with pyridine (1 mL), followed by centrifugation (10,000xg, 30 seconds) to separate the excess aqueous layer gave a pyridine-NaOH solution. To a cuvette containing 700 µL protein solution (purified protein or heat-treated lysate) in M9-N buffer, 50 µL of dithionite solution and 250 µL pyridine-NaOH solution were added. The cuvette was sealed with Parafilm, and the UV-Vis spectrum was recorded immediately.

Cytochrome *c* concentration was determined using $\epsilon_{550-535} = 22.1 \text{ mM}^{-1}\text{cm}^{-1}$. Protein concentrations determined by the hemochrome assay were in agreement with that determined by the bicinchoninic acid (BCA) assay (Thermo Fisher) using bovine serum albumin (BSA) for standard curve preparation.

B.6 Ferrous assay

To a cuvette containing 700 μL protein solution in M9-N buffer was added 50 μL of dithionite solution (10 mg/mL in M9-N buffer). The cuvette was sealed with Parafilm, and the UV-Vis spectrum was recorded immediately. The absorbance value for the peak at 550 nm was recorded, and background absorbance at 600 nm was subtracted. Using the protein concentration as determined by the hemochrome assay, ferrous $\epsilon_{550-600}$ was determined to be $27 \text{ mM}^{-1}\text{cm}^{-1}$ for wild-type *Rma cyt c*, and $21 \text{ mM}^{-1}\text{cm}^{-1}$ for *Rma V75T M100D M103E* (see calibration curves shown on the following page). Concentrations of *Rma M100D* and *V75T M100D M103E* were determined using the extinction coefficient calculated for *V75T M100D M103E*.





B.7 Library construction

Cytochrome *c* site-saturation mutagenesis libraries were generated using a modified version of the 22-codon site-saturation method⁶. For each site-saturation library, oligonucleotides were ordered such that the coding strand contained the degenerate codon NDT, VHG or TGG. The reverse complements of these primers were also ordered. The three forward primers were mixed together in a 12:9:1 ratio, (NDT:VHG:TGG) and the three reverse primers were mixed similarly. Two PCRs were performed, pairing the mixture of forward primers with a pET22(b)+ internal reverse primer, and the mixture of reverse primers with a pET22b internal forward primer. The two PCR products were gel purified, ligated together using Gibson assembly, and transformed into *E. coli*[®] EXPRESS BL21(DE3) cells.

B.8 Enzyme library screening

Single colonies were picked with toothpicks off of LB_{amp/chlor} agar plates, and grown in deep-well (2 mL) 96-well plates containing LB_{amp/chlor} (400 μ L) at 37 °C, 250 rpm shaking, and 80% relative humidity overnight. After 16 hours, 30 μ L aliquots of these overnight cultures were transferred to deep-well 96-well plates containing HB_{amp/chlor} (1 mL) using a 12-channel EDP3-Plus 5-50 μ L pipette (Rainin). Glycerol stocks of the libraries were prepared by mixing cells in LB_{amp/chlor} (100 μ L) with 50% v/v glycerol (100 μ L). Glycerol stocks were stored at -78 °C in 96-well microplates. Growth plates were allowed to shake

for 3 hours at 37 °C, 250 rpm shaking, and 80% relative humidity. The plates were then placed on ice for 30 min. Cultures were induced by adding 10 μ L of a solution, prepared in sterile deionized water, containing 2 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) and 20 mM ALA. The incubator temperature was reduced to 20 °C, and the induced cultures were allowed to shake for 20 hours (250 rpm, no humidity control). Cells were pelleted (4,000xg, 5 min, 4 °C) and resuspended in 500 μ L M9-N buffer. For cell lysis, plates were placed in a 75 °C water bath for 10 min, followed by centrifugation (4,000xg, 5 min, 4 °C) to remove cell debris. The resulting heat-treated lysates (340 μ L) were then transferred to deep-well plates for biocatalytic reactions. In an anaerobic chamber, to deep-well plates of heat-treated lysates were added Na₂S₂O₄ (40 μ L per well, 100 mM in dH₂O), PhMe₂SiH (10 μ L per well, 400 mM in MeCN) and Me-EDA (10 μ L per well, 400 mM in MeCN). The plates were sealed with aluminum sealing tape, removed from the anaerobic chamber, and shaken at 400 rpm for 1.5 h. After quenching with cyclohexane (1 mL), internal standard was added (20 μ L of 20 mM methyl 2-phenylacetate in cyclohexane) and the reaction mixtures were pipetted up and down to thoroughly mix the organic and aqueous layers. The plates were centrifuged (4,000xg, 5 min) and the organic layer (400 μ L) was transferred to shallow-well 96-well plates for SFC analysis. Hits from library screening were confirmed by small-scale biocatalytic reactions, which were analyzed by GC and SFC for accurate determination of turnovers and enantioselectivities.

B.9 Protein lysate preparation

Protein lysates for biocatalytic reactions were prepared as follow: *E. coli* cells expressing *Rma* cyt *c* variant were pelleted (4,000xg, 5 min, 4 °C), resuspended in M9-N buffer and adjusted to the appropriate OD₆₀₀. The whole-cell solution was heat-treated (75 °C for 10 min) then centrifuged (14,000xg, 10 min, 4 °C) to remove cell debris. The supernatant was sterile filtered through a 0.45 μ m cellulose acetate filter into a 6 mL crimp vial, crimp sealed, and the head space of the crimp vial was degassed by bubbling argon through for at least 10 min. The concentration of cytochrome *c* protein lysate was determined using the ferrous assay described in section (E). Using this protocol, the protein concentrations we typically

observed for $OD_{600} = 15$ lysates are in the 8-15 μM range for wild-type *Rma* cyt *c* and 2-10 μM for other *Rma* cyt *c* variants.

B.10 Small-scale biocatalytic reactions

In an anaerobic chamber, protein lysate (340 μL) in a 2 mL crimp vial was added 40 μL $\text{Na}_2\text{S}_2\text{O}_4$ (100 mM in dH_2O), 10 μL PhMe_2SiH (400 or 800 mM in MeCN) and 10 μL Me-EDA (400 mM in MeCN). The vial was crimp sealed, removed from the anaerobic chamber, and shaken at 400 rpm at room temperature for the stated reaction time. At the end of the reaction, the crimp vial was opened and the reaction was quenched with cyclohexane (1 mL). Internal standard was added (20 μL of 20 mM 2-phenylethanol in cyclohexane) and the reaction mixture was transferred to a microcentrifuge tube, vortexed (10 seconds, 3 times), then centrifuged (14,000 \times g, 5 min) to completely separate the organic and aqueous layers (the vortex-centrifugation step was repeated if complete phase separation was not achieved). The organic layer (750 μL) was removed for GC and SFC analysis. All biocatalytic reactions were performed in triplicate unless otherwise stated. The total turnover numbers (TTNs) reported are calculated with respect to the protein catalyst and represent the total number of turnovers that is possible to obtain from the catalyst under the stated reaction conditions.

B11. Supporting tables

Table B-1. Summary of known catalytic systems for enantioselective carbene insertion into silicon–hydrogen bonds

Chiral catalytic system	Ref	Reaction condition	Substrate scope	TTN	% ee
Copper					
A	7	CH₂Cl₂, rt, 18 h	4	17 to 25	17 to 98
B	8	C ₆ H ₆ , 0 °C, 13.5 h	1	8 to 10	29 to 78
C	9	CH₂Cl₂, –60 to 0 °C, 2-12 h	24	3 to 19	12 to 99
B	10	CH ₂ Cl ₂ , –40 to 0 °C, 48-72 h	8	5 to 9	49 to 88
B	11	CH ₂ Cl ₂ , –40 to –10 °C	9	5 to 8	40 to 84
Iridium					
D	12	CH₂Cl₂, –78 or –30 °C, 24 h	15	24 to 50	94 to 99
E	13	CH ₂ Cl ₂ , –78 °C, 24 h	7	75 to 94	72 to 91
Rhodium					
F	14	CH₂Cl₂, rt to 40 °C, 6-12 h	34	9 to 30	77 to 99
G	15	CH ₂ Cl ₂ , –78 °C or rt, 3-12 h	5	8 to 45	20 to 63
H	8	C ₅ H ₁₂ , –78 to –75 °C, 24 h	1	24 to 70	48 to 97
I	16	CF₃CH₂OH, –35 °C	10	51 to 97	20 to 99
H	17	CH ₂ Cl ₂ , –78 °C then rt, 23 h	6	22 to 54	77 to 94
various Rh(II)- carboxylate	18	CH ₂ Cl ₂ , rt or –78 °C, 0.5 to 24 h to overnight	6	<1 to 23	6 to 76
J	19	CH ₂ Cl ₂ , –78 °C to rt, 23 h	2	35 to 40	38 to 58
H	20	CH ₂ Cl ₂ , –78 °C, 1 to 1.5 h	4	34 to 43	35 to 72
H	21	C ₅ H ₁₂ , –78 °C, 24 h	5	13 to 19	75 to 95
various Rh(II)- carboxylate	22	CH ₂ Cl ₂ , rt or reflux	1	8 to 17	6 to 47

Catalytic systems that could yield enantiopure products are highlighted in blue. rt = room temperature. In cases where the reaction times were not documented in the original literature, this information is not shown in the table above.

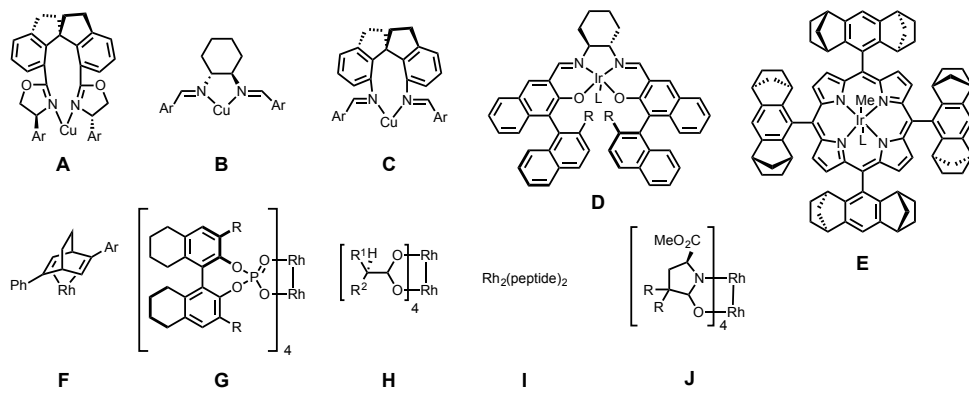
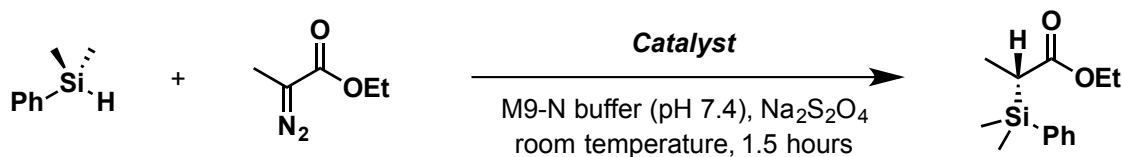


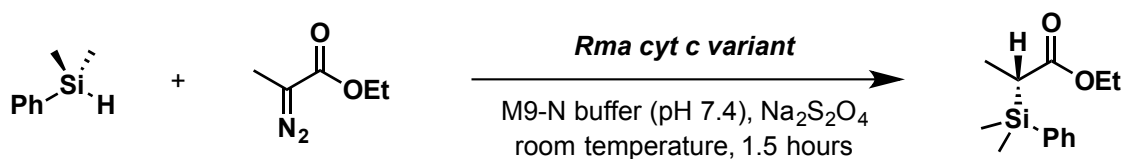
Table B-2. Preliminary experiments with heme and purified heme proteins

Catalyst	TTN	% ee
Controls		
None	0	-
Hemin	4 ± 1	0
Hemin + BSA	1 ± 1	0
P450s		
BM3 P450 T268A ²³	44 ± 15	0
BM3 P450 T268A C400H ²⁴	45 ± 3	<5
BM3 P450 CIS I263F C400S T438S ²⁵	24 ± 9	<5
BM3 P450 F87A T268A C400S ²⁶	40 ± 15	<5
BM3 P450 Hstar H92N H100N ²⁷	46 ± 7	0
Myoglobins		
Sperm whale Mb	12 ± 1	0
Sperm whale Mb H64V V68A ²⁸	17 ± 1	0
Cytochromes c		
Horse heart cyt <i>c</i>	31 ± 10	<5
Bovine heart cyt <i>c</i>	54 ± 2	6
<i>S. cerevisiae</i> cyt <i>c</i>	11 ± 1	<5
<i>R. marinus</i> cyt <i>c</i>	34 ± 10	97
<i>H. thermophilus</i> cyt <i>c</i>	8 ± 2	16
<i>R. globiformis</i> cyt <i>c</i>	4 ± 1	<5
Others		
Horse radish peroxidase	0	-
<i>C. glutamicum</i> catalase	0	-

Heme proteins that are available commercially or in our laboratory inventory were screened to identify the most enantioselective protein variant as starting point for directed evolution. Experiments with heme proteins were performed using 10 μM purified heme protein, 10 mM silane, 10 mM diazo ester, 10 mM Na₂S₂O₄, 5 vol% MeCN, M9-N buffer

at room temperature under anaerobic conditions for 1.5 h. Experiments with hemin were performed using 100 μ M hemin. Experiments with hemin and BSA were performed using 100 μ M hemin in the presence of BSA (0.75 mg/mL). Reactions were performed in triplicate. TTNs reported are the average of three experiments. Within instrument detection limit, variability in % *ee* was not observed. Unreacted starting materials were observed at the end of all reactions and no attempt was made to optimize these reactions.

Table B-3. Carbon–silicon bond formation catalyzed by *Rma cyt c* variants



<i>Rma cyt c</i>	TTN
WT	44 ± 27
M100D	549 ± 24
V75T	150 ± 48
M103E	70 ± 21
V75T M100D	892 ± 20
V75T M100E	154 ± 37
M100D M103E	520 ± 88
V75T M100D M103E	1518 ± 51

Experiments were performed using lysates of *E. coli* expressing *Rma cyt c* variant (OD_{600} = 15; heat-treated at 75 °C for 10 min), 10 mM silane, 10 mM diazo ester, 10 mM $Na_2S_2O_4$, 5 vol% MeCN, M9-N buffer at room temperature under anaerobic conditions for 1.5 h. Reactions were performed in triplicate. TTNs reported are the average of three experiments.

B.12 Supporting Figures

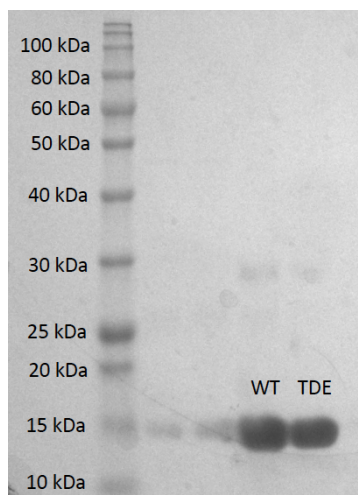


Figure B-1. Representative SDS-PAGE gel of purified wild-type *Rma cyt c* and its V75T M100D M103E variant (TDE) in comparison to a standard protein ladder. The second and third lanes from the left are the same samples loaded at lower protein concentration.

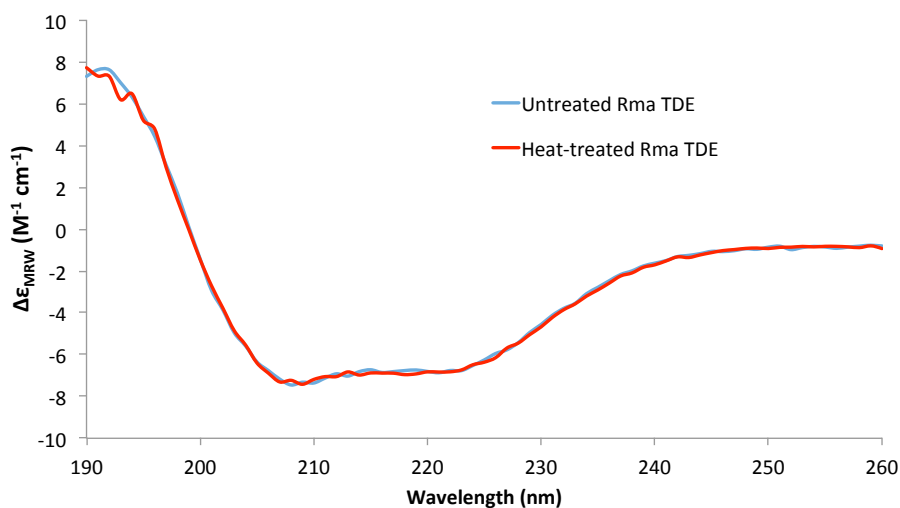


Figure B-2. Circular dichroism (CD) spectra of purified *Rma cyt c* V75T M100D M103E (Rma TDE). *Rma cyt c* V75T M100D M103E was purified without performing the heat treatment step. Two identical samples were prepared in M9-N buffer, and one was heat treated at 75 °C for 10 minutes. After cooling to room temperature, the samples were

analyzed by CD. The CD spectra of heat-treated and untreated *Rma cyt c* V75T M100D M103E are identical, suggesting that heat treatment at 75 °C for 10 minutes does not cause irreversible denaturation of the protein. The $\Delta\epsilon_{MRW}$ values shown here are similar to previously published values for wild-type *Rma cyt c*⁴, which suggests mutations V75T, M100D, and M103E are not highly disruptive to the protein secondary structure.

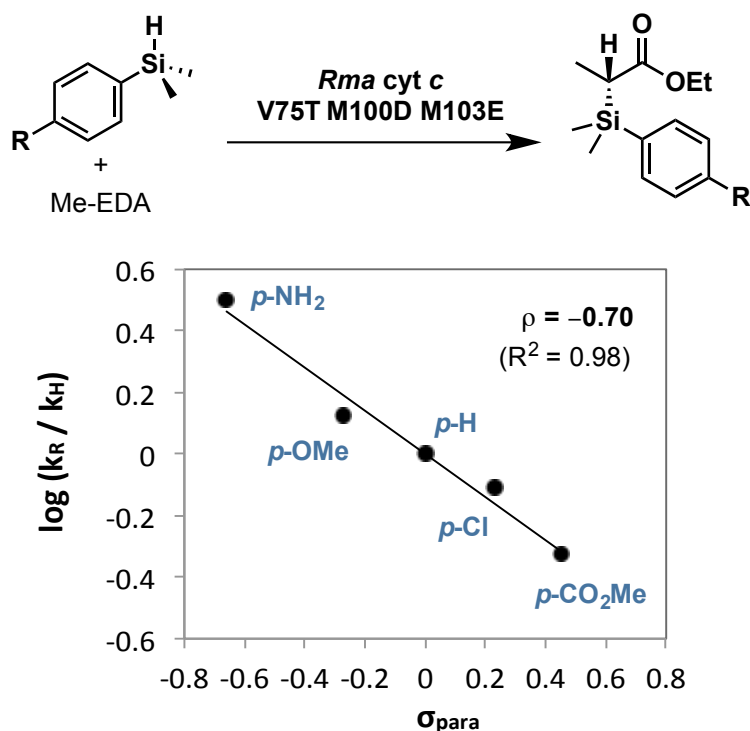


Figure B-3. Hammett analysis of *Rma cyt c* V75T M100D M103E catalyzed carbon–silicon bond formation. The Hammett plot suggests a small build-up of positive charge on the silane in the reaction transition state. This observation is similar to that reported for carbene insertion into Si–H bond catalyzed by copper ($\rho = -0.54$)¹⁰ and rhodium ($\rho = -0.31$)³⁰

B.13 References

1. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
2. Arslan, E., Schulz, H., Zufferey, R., Kunzler, P. & Thony-Meyer, L. Overproduction of the Bradyrhizobium japonicum c-type cytochrome subunits of the cbb(3) oxidase in Escherichia coli. *Biochem. Biophys. Res. Commun.* **251**, 744–747 (1998).
3. Sambrook, J. & W Russell, D. Molecular Cloning: A Laboratory Manual. *Cold Spring Harb. Lab. Press. Cold Spring Harb. NY* 999 (2001). doi:10.1016/0092-8674(90)90210-6
4. Stelter, M. *et al.* A novel type of monoheme cytochrome c: Biochemical and structural characterization at 1.23 Å resolution of Rhodothermus marinus cytochrome c. *Biochemistry* **47**, 11953–11963 (2008).
5. Berry, E. A. & Trumpower, B. L. Simultaneous determination of hemes a, b, and c from pyridine hemochrome spectra. *Anal. Biochem.* **161**, 1–15 (1987).
6. Kille, S. *et al.* Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.* **2**, 83–92 (2013).
7. Hyde, S. *et al.* Copper-Catalyzed Insertion into Heteroatom-Hydrogen Bonds with Trifluorodiazalkanes. *Angew. Chem. Int. Ed.* **55**, 3785–3789 (2016).
8. Wu, J. & Panek, J. S. Total Synthesis of (–)-Virginiamycin M2: Application of Crotylsilanes Accessed by Enantioselective Rh(II) or Cu(I) Promoted Carbenoid Si–H Insertion. *J. Org. Chem.* **76**, 9900–9918 (2011).
9. Zhang, Y. Z., Zhu, S. F., Wang, L. X. & Zhou, Q. L. Copper-catalyzed highly enantioselective carbenoid insertion into Si–H bonds. *Angew. Chem. Int. Ed.* **47**, 8496–8498 (2008).
10. Dakin, L. A., Ong, P. C., Panek, J. S., Staples, R. J. & Stavropoulos, P. Speciation and Mechanistic Studies of Chiral Copper(I) Schiff Base Precursors Mediating Asymmetric Carbenoid Insertion Reactions of Diazoacetates into the Si–H Bond of Silanes. *Organometallics* **19**, 2896–2908 (2000).
11. Dakin, L. A., Schaus, S. E., Jacobsen, E. N. & Panek, J. S. Carbenoid Insertions into the Silicon-Hydrogen Bond Catalyzed by Chiral Copper(I) Schiff Base Complexes. *Tetrahedron Lett.* **39**, 8947–8950 (1998).
12. Yasutomi, Y., Suematsu, H. & Katsuki, T. Iridium(III)-catalyzed enantioselective Si–H bond insertion and formation of an enantioenriched silicon center. *J. Am. Chem.*

- Soc.* **132**, 4510–4511 (2010).
13. Wang, J. C. *et al.* Highly enantioselective intermolecular carbene insertion to C-H and Si-H bonds catalyzed by a chiral iridium(III) complex of a D₄-symmetric Halterman porphyrin ligand. *Chem. Commun.* **48**, 4299–4301 (2012).
 14. Chen, D., Zhu, D. X. & Xu, M. H. Rhodium(I)-Catalyzed Highly Enantioselective Insertion of Carbenoid into Si-H: Efficient Access to Functional Chiral Silanes. *J. Am. Chem. Soc.* **138**, 1498–1501 (2016).
 15. Hrdina, R., Guénée, L., Moraleda, D. & Lacour, J. Synthesis, Structural Analysis, and Catalytic Properties of Tetrakis(binaphthyl or octahydrobinaphthyl phosphate) Dirhodium(II,II) Complexes. *Organometallics* **32**, 473–479 (2013).
 16. Sambasivan, R. & Ball, Z. T. Metallopeptides for asymmetric dirhodium catalysis. *J. Am. Chem. Soc.* **132**, 9289–9291 (2010).
 17. Ge, M. & Corey, E. J. A method for the catalytic enantioselective synthesis of 6-silylated 2-cyclohexenones. *Tetrahedron Lett.* **47**, 2319–2321 (2006).
 18. Buck, R. T. *et al.* Asymmetric rhodium carbene insertion into the Si-H bond: Identification of new dirhodium(II) carboxylate catalysts using parallel synthesis techniques. *Tetrahedron Asymmetry* **14**, 791–816 (2003).
 19. Doyle, M.P., Hu, W., Phillips, I.M. Moody, C.J., Pepper, A.G., Slawin, A.M.Z. Reactivity enhancement for chiral dirhodium (II) tetrakis(carboxamidates). *Adv. Synth. Catal.* **343**, 112–117 (2001).
 20. Kitagaki, S., Kinoshita, M., Takeba, M., Anada, M. & Hashimoto, S. Enantioselective Si-H insertion of methyl phenyldiazoacetate catalyzed by dirhodium(II) carboxylates incorporating N-phthaloyl-(S)-amino acids as chiral bridging ligands. *Tetrahedron Asymmetry* **11**, 3855–3859 (2000).
 21. Davies, H. M. L., Hansen, T., Rutberg, J. & Bruzinski, P. R. Rhodium(II) (S)-N-(arylsulfonyl)prolinate catalyzed asymmetric insertions of vinyl- and phenylcarbenoids into the Si-H bond. *Tetrahedron Lett.* **38**, 1741–1744 (1997).
 22. Buck, R. T. *et al.* Asymmetric rhodium carbenoid insertion into the Si-H bond. *Tetrahedron Lett.* **37**, 7631–7634 (1996).
 23. Coelho, P. S., Brustad, E. M., Kannan, A. & Arnold, F. H. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* **339**, 307–310 (2013).
 24. Wang, Z. J. *et al.* Improved cyclopropanation activity of histidine-ligated cytochrome P450 enables the enantioselective formal synthesis of levomilnacipran. *Angew. Chem. Int. Ed.* **53**, 6810–6813 (2014).

25. 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.* **136**, 15505–15508 (2014).
26. Coelho, P. S. *et al.* A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo. *Nat. Chem. Biol.* **9**, 485–7 (2013).
27. Renata, H. *et al.* Identification of Mechanism-Based Inactivation in P450-Catalyzed Cyclopropanation Facilitates Engineering of Improved Enzymes. *J. Am. Chem. Soc.* **138**, 12527–12533 (2016).
28. Bordeaux, M., Tyagi, V. & Fasan, R. Highly diastereoselective and enantioselective olefin cyclopropanation using engineered myoglobin-based catalysts. *Angew. Chem. Int. Ed.* **54**, 1744–1748 (2015).
29. Landais, Y. & Planchenault, D. Preparation of optically active α -Silylcarbonyl compounds using asymmetric alkylation of α -Silylacetic esters and asymmetric metal-carbene insertion into the Si-H bond. *Tetrahedron* **53**, 2855–2870 (1997).
30. Landais, Y., Parra-Rapado, L., Planchenault, D. & Weber, V. Mechanism of metal-carbenoid insertion into the Si-H bond. *Tetrahedron Lett.* **38**, 229–232 (1997).

*Chapter 3***CATALYTIC IRON-CARBENE INTERMEDIATE REVEALED IN A
CYTOCHROME C CARBENE TRANSFERASE**

Material from this chapter appears in “Lewis RD[†], Garcia-Borràs M[†], Chalkley MJ, Buller AR, Houk KN, Kan SBJ, Arnold FH. (2018) **Catalytic Iron-Carbene Intermediate Revealed in a Cytochrome *c* Carbene Transferase.** *PNAS*, 115:7308-7313. doi: 10.1073/pnas.1807027115” Reprinted with permission from the Proceedings of the National Academy of Sciences. [†]Denotes equal author contribution.

3.1 Abstract

Reactive iron porphyrin carbenes (IPCs) are presumed to be the key catalytic intermediate common to an array of abiological, but synthetically useful carbene transfer reactions catalyzed by wild-type and engineered heme proteins. Despite growing interest in developing new and highly-active carbene transfer enzymes, the IPC intermediate has never been observed in a protein. Using crystallographic, spectroscopic, and computational methods, we have captured and studied a catalytic IPC intermediate in the active site of an enzyme derived from thermostable *Rhodothermus marinus* (*Rma*) cytochrome *c*. High-resolution crystal structures and computational methods reveal how directed evolution created an active site for carbene transfer in an electron transfer protein, and how the laboratory-evolved enzyme achieves perfect carbene transfer stereoselectivity by holding the catalytic IPC in a single orientation. We also discovered that the IPC in *Rma* cytochrome *c* has a singlet ground electronic state, and that the protein environment uses geometrical constraints and non-covalent interactions to influence different IPC electronic states. Knowledge gained by studying this versatile intermediate provides a foundation for studying the mechanisms of carbene transfer reactions and will facilitate the engineering of new carbene transfer enzymes.

3.2 Introduction

Carbenes are formally neutral, divalent carbon species with two non-bonded electrons. They are typically reactive due to their incomplete octet electron configuration¹. Highly stabilized (“persistent”) carbenes exist in biological systems, most notably as the deprotonated thiazolium group of thiamine in pyruvate oxidases². In contrast, reactive metal-carbenes have not been found in nature, despite their versatility and utility in chemical synthesis^{3–5}. Recent protein engineering efforts have revealed that heme proteins can catalyze a myriad of formal carbene transfer reactions including alkene cyclopropanation^{6–9}, alkyne cyclopropanation and bicyclobutanation¹⁰, carbonyl olefination¹¹, as well as carbon–boron¹², –nitrogen¹³, –silicon¹⁴, and –sulfur¹⁵ bond formation (**Figure 3-1**). The proteins enable the heme center to make products that iron porphyrin alone does not^{10,12}, and their genetically-encoded structures direct selectivities that can be switched by evolution^{6,8–10,12,14}, making it possible to create whole new biocatalytic pathways to important molecules.

By analogy to known transition-metal catalyzed reactions that use diazo reagents as carbene precursors^{3–5}, the emerging enzymatic carbene transfer activities of heme proteins are all thought to involve a reactive iron porphyrin carbene (IPC) intermediate; transfer of the carbene to a second substrate and product release regenerate the protein catalyst^{6–15}. Our current understanding of enzyme-catalyzed carbene transfer reactions comes primarily from experimental characterizations of model IPC compounds^{16–21} and their computational studies^{22–27}. Determining the structures and properties of reactive intermediates is central to deciphering how enzymes enhance reaction rates and dictate selectivities²⁸, and recent mechanistic studies of cyclopropanation enzymes^{24,26,27} have also underscored the importance of the IPC intermediate and the need to characterize it within a protein scaffold.

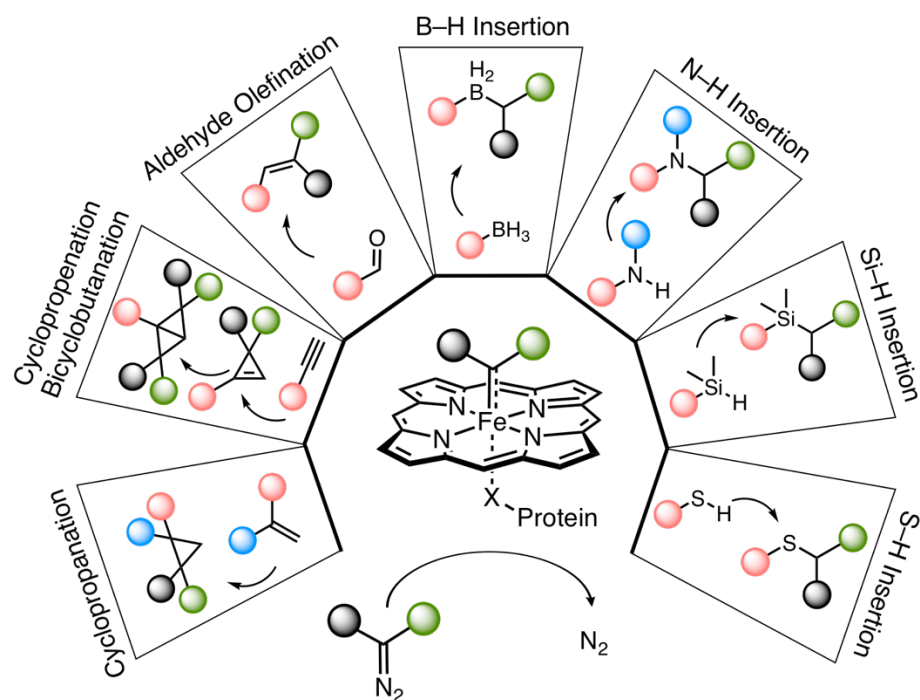


Figure 3-1. The scope of laboratory-evolved biocatalytic carbene transfer reactions. Engineered heme proteins catalyze a broad range of abiological reactions with diazo reagents. These reactions are assumed to proceed through a catalytic IPC intermediate. X indicates porphyrin proximal ligand, usually serine or histidine. All published examples of heme-dependent carbene transfer enzymes accept diazo reagents having one or more electron-withdrawing groups (e.g., COOR, CF₃) as carbene precursors. The active resting state of the catalyst is Fe(II)-porphyrin.

3.3 Results and discussion

3.3.1 Structural characterization of engineered carbene transferase, *Rma* TDE

Rma cyt *c* is an electron transfer protein (Protein Data Bank (PDB) accession number 3CP5²⁹) previously engineered to catalyze the formation of carbon–silicon¹⁴ and carbon–boron¹² bonds by inserting a reactive iron-carbene into the corresponding Si–H and B–H bonds of a silane or borane substrate. We set out to capture and characterize the IPC intermediate in *Rma* cyt *c* by turning to “*Rma* TDE”, a variant previously engineered to catalyze silylation reactions with high efficiency¹⁴ This protein is capable of catalyzing thousands of rounds of carbene transfer using ethyl 2-diazopropanoate (Me-EDA, **Figure 3-2a**) as a carbene precursor, presumably through the intermediacy of IPC 1 (**Figure 3-2a**). To

visualize the active site for carbene transfer, we obtained an X-ray crystal structure of *Rma* TDE at 1.47 Å (PDB 6CUK, Appendix C, **Table C-1**).

The original construct of *Rma* TDE had a C-terminal His-Tag to facilitate easy purification, but efforts to crystallize this variant were not successful. In contrast, the crystals of the wild-type *Rma* cyt *c* had been obtained using an untagged protein²⁹, which suggested that the C-terminal His-Tag might be interfering with our ability to grow crystals of *Rma* TDE. Close examination of the crystal packing in the structure of wild-type *Rma* cyt *c* revealed that the C-terminus resides in close proximity to residues 23-27 of an adjacent symmetry mate, suggesting that additional residues on the C-terminus might disrupt this crystal contact. In contrast, the first seven N-terminal residues of the mature *Rma* cyt *c* were unresolved in the crystal structure, suggesting that these residues were not involved in crystal contacts, and thus not necessary to obtain crystals of the protein. Therefore, we generated two new constructs for *Rma* TDE: a tag-less construct and an N-terminally tagged construct, where the first seven residues of the mature protein were replaced by six histidine residues. By performing fine screening around the various conditions reported to generate crystalline material for the wild-type protein³⁰, we found that small crystals of the tag-less *Rma* TDE could be grown using a well solution containing 1.8-1.9 M (NH₄)₂SO₄ and 2% v/v isopropanol. Through additive screening, we found that addition of NaSCN greatly improved the size of crystals, but single crystals were difficult to obtain. Ultimately, changing the protein from the tag-less *Rma* TDE to the N-terminal tagged *Rma* TDE and using 1.43-1.48 M (NH₄)₂SO₄, 300-400 mM NaSCN, and 1.5% isopropanol yielded large, well-diffracting single crystals.

The structure obtained of *Rma* TDE reveals a pocket on the distal face of the heme created by the three mutations (**Figure 3-2b**). In *Rma* TDE, the native distal methionine has been mutated to an aspartate (M100D), the α -carbon of which is displaced by 4.8 Å relative to its position in the wild-type structure (**Figure C-1**). This displacement produces a cavity in the space previously occupied by the M100 sidechain and allows a water molecule (or hydroxide ion) to coordinate to the heme iron in the enzyme resting state. Overall, directed evolution created an entirely new active site in *Rma* TDE where substrates bind, the carbene forms, and catalysis takes place.

3.3.2 Crystallographic observation of an iron porphyrin carbene in *Rma* TDE

To observe an IPC in the protein scaffold, crystals of *Rma* TDE were soaked with Me-EDA at room temperature for 30 minutes under air, reduced with sodium dithionite, and flash-frozen in liquid nitrogen prior to diffraction. Under these conditions, we determined an X-ray crystal structure at 1.29 Å that shows a new species bound to the iron porphyrin (PDB 6CUN, Appendix C, **Table C-1**). We observed contiguous electron density from the iron porphyrin to an active-site ligand, consistent with an enzymatic IPC intermediate, rather than iron-bound Me-EDA (**Figure 3-2c**). The observed electron density is in agreement with a single carbene species in a single conformation (Appendix C, section **C.9**), suggesting this structure represents the dominant form of the IPC intermediate in solution during catalysis. The high occupancy of the carbene indicates the rate of carbene formation greatly exceeds the rate of carbene decay in the absence of a second substrate, making the direct characterization of an enzymatic IPC intermediate possible for the first time.

The crystal structure of carbene-bound *Rma* TDE displays some similarities to previously studied small-molecule IPC complexes. The Fe–C bond length of 1.9 Å is consistent with the previously determined structure of the IPC 1-methylimidazole-ligated meso-tetrakis(pentafluorophenyl)porphyrin (TPFPP) diphenylcarbene, where the reported Fe–C bond length is 1.83 Å²⁰. The Fe–N (imidazole) bond length of 2.1 Å in the protein is also consistent with that in the small molecule structure (2.17 Å)²⁰. Out-of-plane distortions, particularly ruffling distortions, are observed in both the imidazole-ligated TPFPP diphenylcarbene structure and the carbene-bound *Rma* TDE structure. Whereas the mean out-of-plane deviation for TPFPP diphenylcarbene was 1.03 Å (Appendix C, **Table C-2**), it is smaller for carbene-bound *Rma* TDE (0.6 Å; the deviation for *Rma* TDE alone is 0.7 Å). The difference in distortion between both *Rma* TDE structures is small, and both values are within the range reported for c-type cytochrome proteins (0.3-1.2 Å)³¹. Ruffling is known to be the main distortion in these proteins and is thought to be induced by covalent attachment of the heme to the protein³².

In contrast to what we observed in *Rma* TDE in the absence of Me-EDA, amino acid residues D100, T101, and D102 are unresolved in the carbene-bound *Rma* TDE structure, indicating that this loop region is flexible and the carbene is solvent-accessible (**Figure 3-**

2c). The crystal structure also reveals close, non-polar contacts between the iron-carbene and five amino acid side chains (T75, M76, P79, I83 and M89 lie 3.5-4.3 Å from the carbene, **Figure 3-2c**), suggesting that the new active site is strongly hydrophobic, despite its solvent accessibility.

We had previously predicted an orientation of the carbene in the context of the *Rma* cytochrome *c* protein (**Figure 2-4**). When the observed orientation of the carbene is compared with the orientation proposed in Chapter 2, (**Figure 2-4, Figure 3-2D**) it is clear that the positioning of the α -methyl substituent on the carbene is close to correct, differing by only a 20° rotation around the Fe–C bond. The predicted carbene orientation was based on the hypothesis that the α -methyl substituent of the carbene occupied the space left when the native distal heme ligand, M100 is displaced or mutated. When the carbene-bound *Rma* TDE structure is compared to the wild-type crystal structure, it is clear that the two methyl groups do indeed occupy the same space.

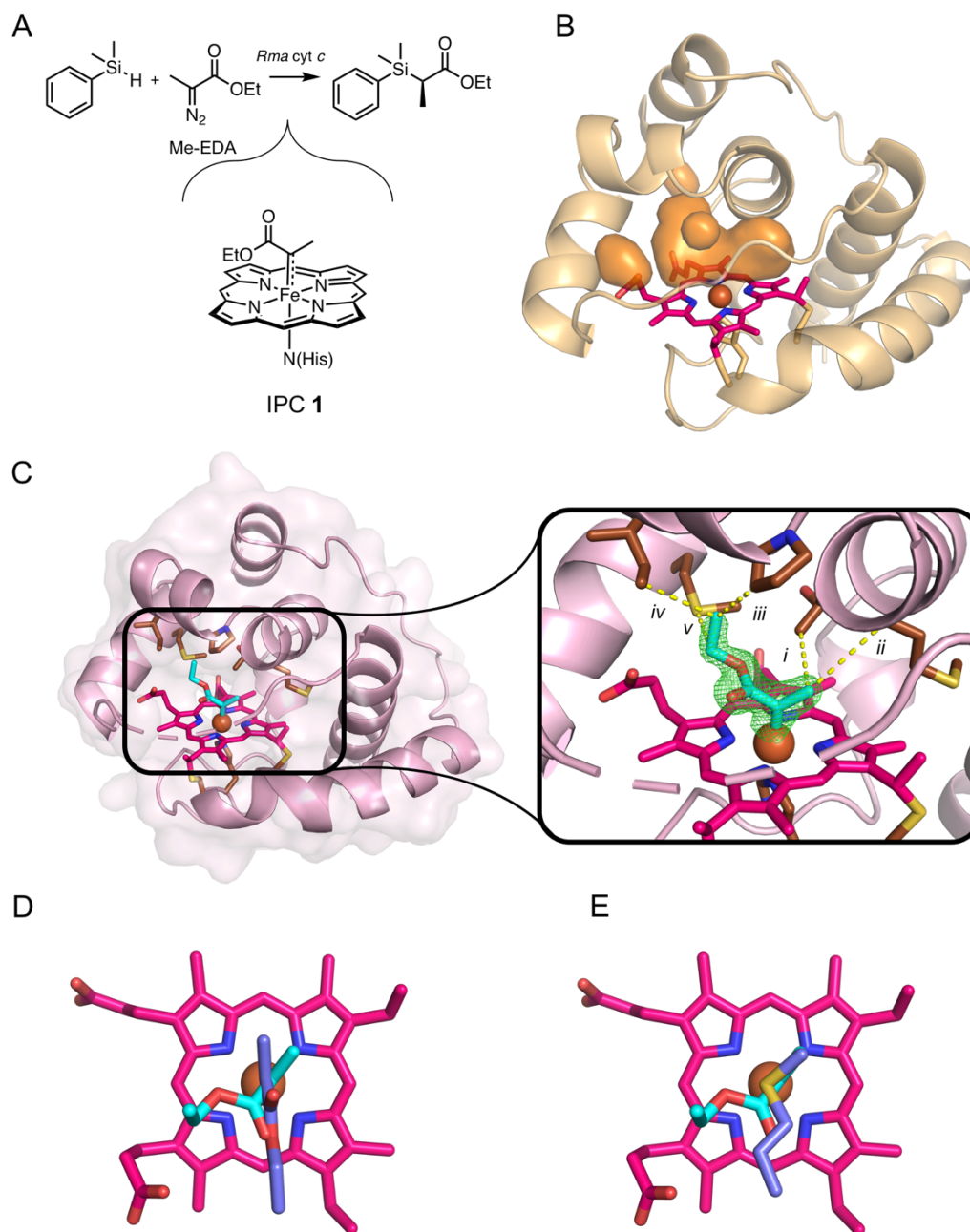


Figure 3-2. Carbene transfer silylation reaction and structures of resting and carbene-bound *Rma* TDE. **A.** The C–Si bond-forming reaction catalyzed by *Rma* cytochrome c proceeds via intermediate IPC 1. **B.** Mutations V75T, M100D, and M103E (*Rma* TDE) in *Rma* cytochrome c increased its carbene transfer activity (14). The crystal structure of *Rma* TDE with no Me-EDA bound (1.47 Å; PDB ID code 6CUK) reveals a cavity distal to the heme (orange surfaces) that is not present in wild-type *Rma* cyt c. **C.** The crystal structure of carbene-bound *Rma* TDE (1.29 Å; PDB ID code 6CUN), with the carbene species in cyan. Residues D100, T101, and D102 are unresolved. (Inset) Omit map in green ($F_o - F_c$) of the carbene electron density contoured at 3σ . Interactions between the carbene and amino acid

residues are shown with yellow lines. (i) T75, 3.7 Å. (ii) M76, 4.3 Å. (iii) P79, 3.5 Å. (iv) I83, 3.6 Å. (v) M89, 3.5 Å. **D.** Comparison of the predicted carbene orientation (purple, proposed in Chapter 2) and the crystallographically observed carbene (cyan). **E.** Comparison of the positioning of residue M100 in wild-type *Rma* cyt *c* (purple, PDB ID: 3CP5) and the observed carbene (cyan). The methyl groups of the carbene and residue M100 occupy the same space, as previously hypothesized.

3.3.3 Iron porphyrin carbene electronic structure and spectroscopy

When first proposed as a reactive enzyme intermediate, the IPC in cytochrome P450 was presented as analogous to oxo-ferryl intermediates found in the natural P450 catalytic cycle⁶. This assignment was supported by the Fe(IV)-like Mössbauer parameters obtained for small molecule IPC complexes^{20,21,33}. However, Zhang and co-workers have argued that IPCs are better described as Fe(II) closed-shell singlets, lower in energy than the triplet states^{24,26,27}. Most recently, Shaik and co-workers studied a model IPC complex, and concluded that the lowest energy state for their IPC was an open-shell singlet, best described as an $S = 1/2$ Fe(III) that is antiferromagnetically coupled to a carbene radical²³. Although the IPC open-shell singlet is similar to the Fe(III)-superoxide electronic structure in heme oxygen binding³⁴, no experimental evidence has been published to support the relevance of an open-shell singlet configuration in IPCs. Computational studies from both the Zhang and Shaik groups suggest that the energy differences between electronic states are small^{22-24,27}, and thus the open-shell singlet, closed-shell singlet, and triplet states should all be considered plausible ground states for the IPC intermediate in the environment of an enzyme active site.

Encouraged by our success in capturing the IPC in the enzyme crystals, we sought to use spectroscopy to characterize the electronic structure of the IPC in *Rma* TDE. Under conditions similar to those where *Rma* TDE is catalytically active, we prepared solution samples of both *Rma* TDE and Me-EDA-treated *Rma* TDE and analyzed them with UV-Vis (Appendix C, **Figure C-2**), electron paramagnetic resonance (EPR), and Mössbauer spectroscopy. Neither sample gave EPR signals in either parallel (10 mW) or perpendicular (2 mW) mode at 4 K (Appendix C, **Figure C-3**). While the absence of EPR signals under these conditions does not prove that the samples are diamagnetic, these results are consistent with our assignment of the resting enzyme and carbene-bound enzyme as low-spin species

(*vide infra*). Mössbauer spectra of *Rma* TDE at 80 K were best fit as a single species with an isomer shift (δ_{Fe}) of $0.45 \text{ mm}\cdot\text{s}^{-1}$ and a quadrupole splitting ($|\Delta E_{\text{Q}}|$) of $1.18 \text{ mm}\cdot\text{s}^{-1}$ (Appendix C, **Figure C-4**), consistent with an $S = 0$ low spin Fe(II) resting-state species. This observation is in contrast to other Fe(II) hemes in proteins lacking an endogenous distal ligand to iron, which have been shown to favor high spin ($S = 2$) states^{35,36}. Mössbauer spectra of Me-EDA-treated *Rma* TDE at 80 K indicated the presence of a major (88% abundance) and a minor species (12% abundance), the latter of which was assigned as ferrous *Rma* TDE. The major species was assigned as carbene-bound *Rma* TDE based on the Mössbauer parameters obtained ($\delta_{\text{Fe}} = 0.18 \text{ mm}\cdot\text{s}^{-1}$, $|\Delta E_{\text{Q}}| = 1.83 \text{ mm}\cdot\text{s}^{-1}$) and the lack of observed EPR signal (Supplementary Fig. 3, 4). The symmetric Mössbauer lineshape of Me-EDA-treated *Rma* TDE and its lack of EPR signal suggested this is an $S = 0$ species, or an $S = 1$ species with a large zero-field splitting. Thus, this reactive intermediate can be made in high yields and studied in both freeze-quenched solutions and crystals to determine its electronic and geometric structure.

3.3.4 Computational models of the iron porphyrin carbene

Density functional theory (DFT) calculations were used to model the three potential electronic structures of IPC **1**. The DFT calculations shows that the triplet state is disfavored, and the triplet structure is not consistent with the observed IPC structure. DFT calculations were unable to distinguish between the two singlet states; both singlet states are close in energy, and both have structures consistent with the observed IPC structure.

Molecular dynamics (MD) simulations were also used to characterize the structure of IPC **1** in *Rma* TDE. MD simulations showed that the IPC favors a structure like that found in the crystal structure, but also revealed a persistent hydrogen bonding interaction between a water molecule and the IPC in a region where the electron density of the X-ray crystal structure is poor (and the water molecule is not observed). Integrating the newly identified water molecule into quantum mechanical calculations on the IPC reveals that the hydrogen bonding interactions causes stabilization of the singlet states relative to the triplet.

3.4 Summary and conclusions

In conclusion, characterization of a catalytically active IPC intermediate inside the active site of an enzyme has provided a foundation from which to study enzyme-catalyzed carbene transfer reactions and explain the reactivity and selectivity of *Rma* TDE-catalyzed carbene transfer. Structural characterization of the *Rma* TDE enzyme showed that the three mutations introduced during directed evolution led to dramatic changes in a loop distal to the heme and generated a protein structure pre-organized for substrate binding and IPC formation. The experimental IPC structure and computational modelling showed that *Rma* TDE stabilizes a single conformation of the IPC that gives the enzyme complete stereoselectivity control during carbene transfer. The conformational restraints imposed by the enzyme also favor the formation of a singlet iron-carbene. This convergent approach to characterizing the emerging family of carbene transferases enables informed engineering that will enhance the generality and utility of these powerful catalysts.

3.5 References

1. Bourissou, D., Guerret, O., Gabbai, F. P. & Bertrand, G. Stable Carbenes. *Chem. Rev.* **100**, 39–91 (2000).
2. Meyer, D., Neumann, P., Ficner, R. & Tittmann, K. Observation of a stable carbene at the active site of a thiamin enzyme. *Nat. Chem. Biol.* **9**, 488–90 (2013).
3. Kornecki, K. P. *et al.* Direct Spectroscopic Characterization of a Transitory Dirhodium Donor-Acceptor Carbene Complex. *Science* **342**, 351–354 (2013).
4. Doyle, M. P. Catalytic Methods for Metal Carbene Transformations. *Chem. Rev.* **86**, 919–939 (1986).
5. Ford, A. *et al.* Modern Organic Synthesis with α -Diazocarbonyl Compounds. *Chem. Rev.* **115**, 9981–10080 (2015).
6. Coelho, P. S., Brustad, E. M., Kannan, A. & Arnold, F. H. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* **339**, 307–310 (2013).
7. Bordeaux, M., Tyagi, V. & Fasan, R. Highly diastereoselective and enantioselective olefin cyclopropanation using engineered myoglobin-based catalysts. *Angew. Chem.*

- Int. Ed.* **54**, 1744–1748 (2015).
8. Gober, J. G. *et al.* Mutating a Highly Conserved Residue in Diverse Cytochrome P450s Facilitates Diastereoselective Olefin Cyclopropanation. *ChemBioChem* **17**, 394–397 (2016).
 9. Knight, A. M. *et al.* Diverse Engineered Heme Proteins Enable Stereodivergent Cyclopropanation of Unactivated Alkenes. *ACS Cent. Sci.* **4**, 372–377 (2018).
 10. Chen, K., Huang, X. H., Kan, S. B. J., Zhang, R. K. & Arnold, F. H. Enzymatic Construction of Highly Strained Carbocycles. *Science* **360**, 71–75 (2018).
 11. Weissenborn, M. J. *et al.* Enzyme-Catalyzed Carbonyl Olefination by the E. coli Protein YfeX in the Absence of Phosphines. *ChemCatChem* **8**, 1636–1640 (2016).
 12. Kan, S. B. J., Huang, X., Gumulya, Y., Chen, K. & Arnold, F. H. Genetically programmed chiral organoborane synthesis. *Nature* **552**, 132 (2017).
 13. Wang, Z. J., Peck, N. E., Renata, H. & Arnold, F. H. Cytochrome P450-catalyzed insertion of carbenoids into N–H bonds. *Chem. Sci.* **5**, 598–601 (2014).
 14. Kan, S. B. J., Lewis, R. D., Chen, K. & Arnold, F. H. Directed evolution of cytochrome c for carbon–silicon bond formation: Bringing silicon to life. *Science* **354**, 1048–1051 (2016).
 15. Tyagi, V., Bonn, R. B. & Fasan, R. Intermolecular carbene S–H insertion catalysed by engineered myoglobin-based catalysts. *Chem. Sci.* **6**, 2488–2494 (2015).
 16. Mansuy, D. *et al.* Dichlorocarbene Complexes of Iron(II)-Porphyrins- Crystal and Molecular Structure of Fe(TPP)(CCl₂)(H₂O). *Angew. Chem. Int. Ed.* **17**, 781–782 (1978).
 17. Chevrier, B., Weiss, R., Lange, M., Chottard, J. C. & Mansuy, D. An Iron(III)-Porphyrin Complex with a Vinylidene Group Inserted into an Iron-Nitrogen Bond: Relevance to the Structure of the Active Oxygen Complex of Catalase. *J. Am. Chem. Soc.* **103**, 2899–2901 (1981).
 18. Olmstead, M. M., Cheng, R. J. & Balch, A. L. X-ray Crystallographic Characterization of an Iron Porphyrin with a Vinylidene Carbene Inserted into an Iron-Nitrogen Bond. *Inorg. Chem.* **21**, 4143–4148 (1982).
 19. Artaud, I., Gregoire, N., Leduc, P. & Mansuy, D. Formation and Fate of Iron-Carbene Complexes in Reactions between a Diazoalkane and Iron-Porphyrins: Relevance to the Mechanism of Formation of N-Substituted Hemes in Cytochrome P-450 Dependent Oxidation of Sydnonones. *J. Am. Chem. Soc.* **112**, 6899–6905 (1990).
 20. Li, Y., Huang, J., Zhou, Z., Che, C. & You, X. Remarkably stable iron porphyrins

- bearing nonheteroatom-stabilized carbene or (alkoxycarbonyl)carbenes: isolation, X-ray crystal structures, and carbon atom transfer reactions with hydrocarbons. *J. Am. Chem. Soc.* **124**, 13185–13193 (2002).
21. Liu, Y. *et al.* Electronic Configuration and Ligand Nature of Five-Coordinate Iron Porphyrin Carbene Complexes: An Experimental Study. *J. Am. Chem. Soc.* **139**, 5023–5026 (2017).
 22. Khade, R. L. *et al.* Iron Porphyrin Carbenes as Catalytic Intermediates: Structures, Mössbauer and NMR Spectroscopic Properties, and Bonding. *Angew. Chem. Int. Ed.* **53**, 7574–7578 (2014).
 23. 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.* **138**, 9597-9610 (2016). doi:10.1021/jacs.6b04636
 24. Wei, Y., Tinoco, A., Steck, V., Fasan, R. & Zhang, Y. Cyclopropanations via Heme Carbenes: Basic Mechanism and Effects of Carbene Substituent, Protein Axial Ligand, and Porphyrin Substitution. *J. Am. Chem. Soc.* jacs.7b09171 (2017). doi:10.1021/jacs.7b09171
 25. Tatsumi, K. & Hoffmann, R. Metalloporphyrins with Unusual Geometries. 2. Slipped and Skewed Bimetallic Structures, Carbene and Oxo Complexes, and Insertions into Metal-Porphyrin Bonds. *Inorg. Chem.* **20**, 3771–3784 (1981).
 26. Khade, R. L. & Zhang, Y. Catalytic and Biocatalytic Iron Porphyrin Carbene Formation: Effects of Binding Mode, Carbene Substituent, Porphyrin Substituent, and Protein Axial Ligand. *J. Am. Chem. Soc.* **137**, 7560–7563 (2015).
 27. Khade, R. L. & Zhang, Y. C-H Insertions by Iron Porphyrin Carbene: Basic Mechanism and Origin of Substrate Selectivity. *Chem. Eur. J.* 17654–17658 (2017). doi:10.1002/chem.201704631
 28. Rittle, J. & Green, M. T. Cytochrome P450 Compound I: Capture, Characterization, and C-H Bond Activation Kinetics. *Science* **330**, 933–937 (2010).
 29. Stelter, M. *et al.* A novel type of monoheme cytochrome c: Biochemical and structural characterization at 1.23 Å resolution of *Rhodothermus marinus* cytochrome c. *Biochemistry* **47**, 11953–11963 (2008).
 30. Stelter, M., Melo, A. M., Saraiva, L. M., Teixeira, M. & Archer, M. Crystallization and X-ray analysis of *Rhodothermus marinus* cytochrome c at 1.23 Å resolution. *Protein Pept Lett* **14**, 1038–1040 (2007).
 31. Graves, A. B., Graves, M. T. & Liptak, M. D. Measurement of Heme Ruffling Changes in MhuD Using UV-vis Spectroscopy. *J. Phys. Chem. B* **120**, 3844–3853 (2016).

32. Kleingardner, J. G. & Bren, K. L. Biological Significance and Applications of Heme c Proteins and Peptides. *Acc. Chem. Res.* **48**, 1845–1852 (2015).
33. English, D. R., Hendrickson, D. N. & Suslick, K. S. Mössbauer Spectra of Oxidized Iron Porphyrins. *Inorg. Chem.* **22**, 367–368 (1983).
34. Wilson, S. A. *et al.* X-ray absorption spectroscopic investigation of the electronic structure differences in solution and crystalline oxyhemoglobin. *Proc. Natl. Acad. Sci.* **110**, 16333–16338 (2013).
35. Ran, Y. *et al.* Spectroscopic identification of heme axial ligands in HtsA that are involved in heme acquisition by *Streptococcus pyogenes*. *Biochemistry* **49**, 2834–2842 (2010).
36. Engler, N., Prusakov, V., Ostermann, A. & Parak, F. G. A water network within a protein: temperature-dependent water ligation in H64V-metmyoglobin and relaxation to deoxymyoglobin. *Eur. Biophys. J. with Biophys. Lett.* **31**, 595–607 (2003).

Appendix C

SUPPORTING INFORMATION FOR CHAPTER 3

Material from this chapter appears in “Lewis RD[†], Garcia-Borràs M[†], Chalkley MJ, Buller AR, Houk KN, Kan SBJ, Arnold FH. (2018) **Catalytic Iron-Carbene Intermediate Revealed in a Cytochrome *c* Carbene Transferase.** *PNAS*, 115:7308-7313. doi: 10.1073/pnas.1807027115” Reprinted with permission from the Proceedings of the National Academy of Sciences. [†]Denotes equal author contribution.

C.1 Materials and methods

Unless otherwise noted, all chemicals and reagents for chemical reactions were obtained from commercial suppliers (Acros, Arch Bioscience, Fisher Scientific, Sigma-Aldrich, TCI America, VWR) and used without further purification. Multitron shakers (Infors, Annapolis Junction, MD) were used for cell growth. UV-Vis spectroscopy was performed using a Shimadzu UV-1800 spectrophotometer (Shimadzu, Carlsbad, CA). Sonication was performed using a Qsonica Q500 sonicator. Silica gel chromatography purifications were carried out using AMD Silica Gel 60, 230-400 mesh. ¹H NMR spectra were recorded on a Bruker Prodigy 400 MHz instrument and are internally referenced to the residual solvent peak (chloroform). Data for ¹H NMR are reported in the conventional form: chemical shift (δ ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, hept = heptet, m = multiplet, br = broad, app = appears as), coupling constant (Hz), integration. Gas chromatography (GC) analyses were carried out using a Shimadzu GC-17A gas chromatograph, a flame ionization detector (FID), and J&W HP-5 (30 m x 0.32 mm, 0.25 μ m film; 90 °C hold 1 min, 90 to 110 °C at 15 °C/min, 110 to 280 °C at 60 °C/min, 280 °C hold 1 min, 6.2 min total). Analytical chiral high pressure liquid chromatography (HPLC) was performed with a HP-Agilent 1100 series instrument using *i*-PrOH and hexanes as the

mobile phase, with visualization at 210 nm. A Chiralcel OD-H column was used for chiral separations (4.6 mm x 25 cm).

Plasmid pET22 was used as a cloning vector, and cloning was performed using Gibson assembly¹. The cytochrome *c* maturation plasmid pEC86² was used as part of a two-plasmid system to express prokaryotic cytochrome *c* proteins. Cells were grown using Lysogeny Broth medium or HyperBroth (AthenaES, Baltimore, MD) with 100 µg/mL ampicillin and 20 µg/mL chloramphenicol (LB_{amp/chlor} or HB_{amp/chlor}). Primer sequences are available upon request. Electrocompetent *Escherichia coli* cells were prepared following the protocol of Sambrook *et al.*³. T5 exonuclease, Phusion polymerase, and *Taq* ligase were purchased from New England Biolabs (NEB, Ipswich, MA). Sodium phosphate (NaPi, 20 mM, pH 7.5) was used as a buffering system for purified proteins, unless otherwise specified. M9-N minimal medium (abbreviated as M9-N buffer; pH 7.4) was used without a carbon source; it contains 47.7 mM Na₂HPO₄, 22.0 mM KH₂PO₄, 8.6 mM NaCl, 2.0 mM MgSO₄, and 0.1 mM CaCl₂.

C.2 Plasmid construction

All variants described in this paper were cloned and expressed using the pET22(b)+ vector (MilliporeSigma, St. Louis, MO). The gene encoding *Rhodothermus marinus* cyt *c* V75T M100D M103E (*Rma* TDE)⁴ was obtained as a single gBlock (Integrated DNA Technologies, Coralville, IA), codon-optimized for *E. coli*, and cloned using Gibson assembly¹ into pET22(b)+ between restriction sites *Nde*I and *Xho*I in frame. The gene encoding *Rma* TDE contained an *N*-terminal pelB leader sequence and 6xHisTag (MKYLLPTAAAGLLLLAAQPAMAHHHHHH) and had the first seven amino acids of mature, wild-type *Rma* cyt *c* removed (TESGTAA). This plasmid was co-transformed with the cytochrome *c* maturation plasmid pEC86² into *E. coli*[®] EXPRESS BL21(DE3) cells (Lucigen, Middleton, WI).

DNA coding sequence of *Rma* cyt *c* V75T M100D M103E with an *N*-terminal pelB leader sequence and 6xHisTag:

```
ATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCAGCCG
GCGATGGCCATCATCATCACCACCAAGACCCGGAAGCACTGGCAGCGGAAATTGG
TCCGGTCAAACAGGTGAGCCTGGGTGAACAGATTGATGCGGCCCTGGCGCAACAGGGAG
```

AACAGCTCTTCAACACGTATTGTACTGCGTGCCACCGTCTGGATGAGCGTTTTATCGGA
 CCGGCCCTGCGCGATGTTACCAAACGTCGTGGGCCGGTTTACATCATGAACACGATGCT
 GAACCCGAATGGGATGATCCAGCGTCATCCGGTGATGAAACAGCTCGTGCAGGAATATG
 GGACCATGGATACCGATGAGGCCCTGAGTGAAGAACAAGCGCGCAATTCTGGAGTAT
 CTGCGCCAGGTTGCGGAAAACCAGTAATGA

Amino acid sequence of *Rma* cyt *c* V75T M100D M103E with an *N*-terminal pelB leader sequence and 6xHisTag:

MKYLLPTAAAGLLLLLAAQPAMAHHHHHHQQDPEALAAEIGPVKQVSLGEQIDAALAQQ
 GEQLFNTYCTACHRLDERFIGPALRDVTKRRGPVYIMNTMLNPNNGMIQRHPVMKQLVQE
 YGTMDDTDEALSEEQARAILEYLRQVAENQ

C.3 Cytochrome *c* expression and purification

Purified cytochrome *c* proteins were prepared as follows. In a 4 L flask, one liter HyperBroth (AthenaES) containing 100 µg/mL ampicillin and 20 µg/mL chloramphenicol (HB_{amp/chlor}) was inoculated with an overnight culture (20 mL, Lysogeny Broth, with 100 µg/mL ampicillin, 20 µg/mL chloramphenicol, LB_{amp/chlor}) of recombinant *E. cloni*[®] EXPRESS BL21(DE3) cells containing a pET22(b)+ plasmid encoding the cytochrome *c* variant, and the pEC86 plasmid. The culture was shaken at 37°C and 200 rpm (no humidity control) until the OD₆₀₀ was 0.7 (approximately 3 hours). The culture was placed on ice for 30 minutes, and isopropyl β-D-1-thiogalactopyranoside (IPTG) and 5-aminolevulinic acid (ALA) were added to final concentrations of 20 µM and 200 µM, respectively.

The incubator temperature was reduced to 25°C, and the culture was allowed to shake for 20 hours at 200 rpm. Cells were harvested by centrifugation (4 °C, 15 min, 4,000xg), and the cell pellet was stored at -20°C until further use (at least 24 hours). The cell pellet was resuspended in buffer containing 100 mM NaCl, 20 mM imidazole, and 20 mM Tris-HCl buffer (pH 7.5 at 25 °C) and cells were lysed by sonication (4 minutes, 1 second on, 1 second off, 30% duty cycle; Qsonica Q500 sonicator). Cell lysate was placed in a 75 °C heat bath for 10 minutes, and cell debris was removed by centrifugation for 20 min (30,000xg, 4 °C). Supernatant was sterile filtered through a 0.45 µm cellulose acetate filter and purified using a 1 mL Ni-NTA column (HisTrap HP, GE Healthcare, Piscataway, NJ) using an AKTA purifier FPLC system (GE Healthcare). The cytochrome *c* protein was

eluted from the column by running a gradient from 20 to 500 mM imidazole over 10 column volumes.

The purity of the collected cytochrome *c* fractions was analyzed using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Pure fractions were pooled and concentrated using a 3 kDa molecular weight cut-off centrifugal filter. Protein for crystallography was dialyzed overnight into 20 mM Tris-HCl buffer (pH 7.5 at 25 °C) using 3.5 kDa molecular weight cut-off dialysis tubing. Protein for non-crystallography use was dialyzed overnight into 20 mM sodium phosphate buffer (pH 7.5) using 3.5 kDa molecular weight cut-off dialysis tubing. The dialyzed protein was further concentrated using a 3 kDa molecular weight cut-off centrifugal filter, flash-frozen on dry ice, and stored at -20°C .

The concentration of cytochrome *c* was determined in triplicate using the hemochrome assay described in section C.5.

C.4 ^{57}Fe -labelled cytochrome *c* expression and purification

In order to prepare protein samples for Mössbauer spectroscopy, ^{57}Fe -labelled *Rma* cyt *c* was expressed and purified, using a protocol adapted from Liptak *et al*⁵.

For 1 L ^{57}Fe media:

950 mL	basic salts buffer
10 mL	trace metals mix
10 mL	vitamin cocktail
300 μL	^{57}Fe / H_2SO_4 solution
10 mL	1 M MgCl_2
5 g	L/D-glucose
1 g	NH_4Cl
16.8 mg	5-aminolevulinic acid (ALA)
30 mg	thiamine hydrochloride
1 mL	20 mg/mL chloramphenicol in ethanol
1 mL	100 mg/mL ampicillin in water

For 950 mL basic salts buffer:

950 mL	water
13.0 g	KH_2PO_4
13.1 g	$\text{K}_2\text{HPO}_4 \cdot (3\text{H}_2\text{O})$
17.1 g	$\text{Na}_2\text{HPO}_4 \cdot (7\text{H}_2\text{O})$

3.6 g Na₂SO₄

Adjusted pH to 7.4 with 5M NaOH, autoclaved, and stored at room temperature.

For 100 mL Trace metals mix:

90 mL water

600 mg calcium chloride, dihydrate

120 mg manganese chloride, tetrahydrate

80 mg cobalt chloride, hexahydrate

70 mg zinc sulfate, heptahydrate

30 mg copper chloride, dihydrate

2 mg boric acid

500 mg EDTA

Adjusted pH to 7.4 with 5M NaOH, and added water until final volume was 100 mL. Solution was deep purple. The solution was sterile filtered through a 0.2 μm PES membrane and stored at -20°C.

For 100 mL vitamin cocktail:

100 mL water

10 mg biotin

10 mg choline chloride

10 mg folic acid

10 mg nicotinamide

10 mg calcium D-pantothenate

10 mg pyridoxal phosphate

1 mg riboflavin

The solution was sterile filtered through a 0.2 μm PES membrane, and stored at -20°C.

For 600 μL ⁵⁷Fe / H₂SO₄ solution:

25.0 mg ⁵⁷Fe (95% enrichment, Isotrex USA, San Francisco)

600 μL 2 M H₂SO₄

H₂SO₄ was added to the iron powder under anaerobic conditions, and allowed to shake overnight, until the iron was dissolved. The solution was pale blue, and was stored at room temperature under anaerobic conditions.

Purified ⁵⁷Fe cytochrome *c* proteins were prepared as follows. In a 125 mL flask, 30 mL of HB_{amp/chor} was inoculated with an overnight culture (1 mL, LB_{amp/chor}) of recombinant *E. cloni*[®] EXPRESS BL21(DE3) cells containing a pET22(b)+ plasmid encoding the cytochrome *c* variant, and the pEC86 plasmid. The culture was shaken at 37°C and 200 rpm (no humidity control) until the OD₆₀₀ was 0.7 (approximately 3 hours). In a 4 L flask, one liter of ⁵⁷Fe medium was inoculated with 20 mL of the late-log phase HB_{amp/chor} culture. The culture was again shaken at 37°C and 200 rpm (no humidity control) until the OD₆₀₀ was 0.3 (approximately 4 more hours). The culture was placed on

ice for 30 minutes, and isopropyl β -D-1-thiogalactopyranoside (IPTG) and 5-aminolevulinic acid (ALA) were added to final concentrations of 20 μ M and 200 μ M, respectively.

After addition of IPTG and ALA, the culture was returned to the shaker, allowed to shake for 24 hours at 37 °C, 200 rpm. Cells were harvested and lysed using the same procedure as described in C.2. 57 Fe-labelled protein was then purified as described in C.3.

C.5 Hemochrome assay

A solution of sodium dithionite (10 mg/mL) was prepared in M9-N buffer. Separately, a solution of 1 M NaOH (0.4 mL) was mixed with pyridine (1 mL), followed by centrifugation (10,000xg, 30 seconds) to separate the excess aqueous layer gave a pyridine-NaOH solution. To a cuvette containing 10 μ L purified protein solution and 690 μ L M9-N buffer, 50 μ L of dithionite solution and 250 μ L pyridine-NaOH solution were added. The cuvette was sealed with Parafilm, and the UV-Vis spectrum was recorded immediately. Cytochrome *c* concentration was determined using $\epsilon_{550-535} = 22.1 \text{ mM}^{-1} \text{ cm}^{-1}$ ^{4,6}.

C.6 Synthesis and characterization of authentic organosilicon product

Ethyl 2-(dimethylphenylsilyl)propanoate was prepared in a similar manner to that published in Kan et al. 2016⁴. In a 25 mL round bottom flask, 0.281 g of dimethylphenylsilane (2.0 mmol, 2.0 equiv.) were added, followed by, Rh₂(OAc)₄ (1.1 mg, 0.25 mol%) and 5 mL of dichloromethane. The mixture was cooled to 0°C, after which 0.128 g of ethyl-2-diazopropanoate (1.0 mmol, 1.0 equiv.) was added dropwise to the solution. The reaction was allowed to slowly warm up to room temperature over 4 hours and stirred at room temperature for another 1 hour. Evaporation of the organic solvent and purification by silica column chromatography using EtOAc and hexane as eluents afforded the organosilicon product ethyl 2-(dimethylphenylsilyl)propanoate with 54% yield. ¹H NMR was used to confirm product identity and purity.

C.7 Biocatalytic reactions

In an anaerobic chamber, 10 μL of 80 μM purified protein was added to 350 μL of 20 mM sodium phosphate buffer (NaPi, pH 7.5) in a 2 mL crimp vial. Next, the protein was reduced by adding 20 μL $\text{Na}_2\text{S}_2\text{O}_4$ (200 mM in NaPi, pH 7.5), and the reaction was initiated with the addition of reactants, 10 μL PhMe₂SiH (400 mM in acetonitrile) and 10 μL Me-EDA (400 mM in MeCN). The vial was crimp sealed, removed from the anaerobic chamber, and shaken at 400 rpm at room temperature for 2 hours. At the end of the reaction, the crimp vial was opened and the reaction was quenched with cyclohexane (1 mL). Internal standard was added (20 μL of 1,2,3-trimethoxybenzene in toluene) and the reaction mixture was transferred to a microcentrifuge tube, vortexed (1 minute), then centrifuged (14,000xg, 5 min) to completely separate the organic and aqueous layers. The organic layer (750 μL) was removed for GC and chiral HPLC analysis. All biocatalytic reactions were performed in triplicate. The total turnover numbers (TTNs) reported are calculated with respect to the protein catalyst and represent the total number of turnovers obtained using the catalyst under the stated reaction conditions.

Concentration of organosilicon product was calculated based on the ratio of areas between the product peak and internal standard peak, as measured by GC-FID. Stock solutions of chemically synthesized ethyl 2-(dimethylphenylsilyl)propanoate were prepared at various concentrations (20 to 200 mM in MeCN). To a microcentrifuge tube were added 360 μL 20 mM sodium phosphate buffer, 20 μL $\text{Na}_2\text{S}_2\text{O}_4$ (200 mM in NaPi, pH 7.5), 20 μL organosilicon product, 20 μL internal standard (20 mM 1,2,3-trimethoxybenzene in toluene) and 1 mL cyclohexane. The mixture was vortexed (1 minute) then centrifuged (14,000xg, 5 min) to completely separate the organic and aqueous layers. The organic layer (750 μL) was removed for GC analysis. Each concentration was prepared in duplicate, and both data points are shown for each concentration (**Figure C-5**). The standard curves plot product concentration in mM (x-axis) against the ratio of product area to internal standard area on the GC (y-axis).

Enantiomeric excess for biocatalytically synthesized products was determined using chiral HPLC, with a Chiralcel OD-H column as the stationary phase and 1% isopropanol, 99% hexanes as the mobile phase, at 40°C for 30 minutes.

C.8 UV-Visible spectroscopy

UV-Visible spectra were collected at room temperature between 600 and 380 nm on a UV-1800 Shimadzu spectrophotometer in quartz cuvettes sealed with Parafilm (aerobic conditions) or a screw cap with a pierceable septum (anaerobic conditions). Sodium phosphate (NaPi, 20 mM, pH 7.5) was used as a buffer system. Spectra of reduced *Rma* TDE was taken in buffer supplemented with 50 mM sodium dithionite to ensure complete reduction of the ferrous iron, and various co-solvents were added to demonstrate that the spectral characteristics are similar in 1% isopropanol, 5% acetonitrile, and 15% dimethylsulfoxide. Spectra of reduced, Me-EDA-treated *Rma* TDE was taken in buffer containing 50 mM sodium dithionite, 10 mM ethyl 2-diazo propanoate, and 1% isopropanol. Co-solvents were added to demonstrate that the addition of 5% acetonitrile or 15% dimethylsulfoxide did not substantially impact the spectrum. For spectra taken under anaerobic conditions, the headspace of the cuvettes was de-gassed with argon for 5 minutes prior to the addition of sodium dithionite and Me-EDA.

C.9 X-ray crystallography

Fine screening for *Rma* TDE crystallization was performed around crystallography conditions previously reported to yield crystals of wild-type *Rma* cyt *c*⁷. Crystals were grown against a 1-mL reservoir with mother liquor comprised of 2.0 μ L of 14.0 mg/mL *Rma* TDE (in 20 mM Tris HCl, pH 7.5) and 1.0 μ L of well solution. Very small crystals were observed after two weeks in 1.8-1.9 M $(\text{NH}_4)_2\text{SO}_4$, 2% isopropanol. Further optimization failed to yield larger crystals, but showed that the protein crystallized within the range of 1.4-1.9 M $(\text{NH}_4)_2\text{SO}_4$, 1-3% isopropanol. Additive screening was performed using an additive screening kit (Hampton Research, Aliso Viejo, CA) and with mother liquor comprised of 2.0 μ L of 14.0 mg/mL *Rma* TDE, 1.0 μ L of well solution (1.7 M $(\text{NH}_4)_2\text{SO}_4$, 1.5% isopropanol), and 0.2 μ L of additive solution. Additive screening showed that addition of NaSCN yielded larger and more numerous crystals. Further fine screening around this condition revealed that the best crystals were obtained using 1.43-1.48 M $(\text{NH}_4)_2\text{SO}_4$, 300-400 mM NaSCN, and 1.5% isopropanol (2.0 μ L of 14.0 mg/mL *Rma* TDE, 1.0 μ L well solution). Under these

conditions, crystals formed over 3-5 days, and reached maximum size after 7-10 days. Crystals were cryo-protected by addition of ethylene glycol to the well solution to a final concentration of 20%. Crystals were flash frozen and stored in liquid N₂ until diffraction.

For crystals soaked with the carbene precursor Me-EDA, the mother liquor was supplemented with 1 M Me-EDA in isopropanol to a final concentration of 40 mM Me-EDA. Crystals were soaked at room temperature for 30 minutes before moving the samples to 4°C. At 4°C, the crystals were soaked for approximately 1 minute in a solution of mother liquor supplemented with 40 mM Me-EDA and 50 mM sodium dithionite before flash-freezing in liquid nitrogen. Prolonged soaking in either of the Me-EDA supplemented solutions caused the crystals to dissolve.

Diffraction data were collected remotely at the Stanford Synchrotron Radiation Laboratories on beamline 12-2, with a wavelength of 0.97946 Å, and at a temperature of 100 K. Crystals routinely diffracted at or below 1.7 Å, and the data were integrated and scaled using XDS⁸ and AIMLESS⁹. A resolution cutoff of $I/\sigma > 2.5$ was applied along all axes of diffraction. These data contributed to model quality as judged by R_{free} in the final bin < 0.27 . The structure was solved using molecular replacement with PHASER¹⁰, as implemented in CCP4¹¹. For *Rma* TDE, the search model comprised a single monomer of wild-type *Rma* cyt *c* (PDB ID: 3CP5)¹² subjected to 10 cycles of geometric idealization in Refmac5¹³ and removal of all ligands. For carbene-bound *Rma* TDE, the search model comprised a single monomer of *Rma* TDE (PDB ID: 6CUK) subjected to 10 cycles of geometric idealization in Refmac5 and removal of all ligands. Model building was performed in Coot¹⁴ beginning with data processed at 2.2 Å, followed by subsequent inclusion of increasingly higher resolution shells of data with relaxed geometric constraints. Refinement was performed using Refmac5. The MolProbity server¹⁵ was used to identify rotamer flips and clashes. After the protein, ligand, and solvent atoms were built, TLS operators were added to refinement¹⁶, which resulted in substantial improvements in R_{free} for the models. Crystallographic and refinement statistics are reported in Table S1. The carbene ligand was modelled in a single conformation at 100% occupancy, which resulted in the atoms comprising the Fe–C bond to have B-factors of 14.74 and 21.36, respectively. Occupancy was not refined by B-factor matching, as the electron density suggested only a single conformation, and B-factors on both atoms were both sufficiently low and similar to indicate high occupancy in the crystal. The carbene-

bound *Rma* TDE structure is deposited with PDB ID: 6CUN, and the *Rma* TDE structure without the carbene ligand is deposited with PDB ID: 6CUK. Analysis of the protein structures with MolProbity showed that the carbene-bound *Rma* TDE structure had a MolProbity score of 1.0, and the *Rma* TDE structure had a MolProbity score of 0.9, putting both structures in the highest percentile rank for MolProbity scores. The carbene-bound *Rma* TDE structure contains 98.15% Ramachandran-favored rotamers, with no Ramachandran outliers, and the *Rma* TDE structure contains 99.12% Ramachandran-favored rotamers and no Ramachandran outliers.

C.10 Mössbauer spectroscopy

C.10.1 General description of Mössbauer spectroscopy

Mössbauer spectra were recorded on a spectrometer from SEE Co. (Edina, MN) operating in the constant acceleration mode in a transmission geometry. The sample was kept in an SVT-400 cryostat from Janis (Wilmington, MA). The quoted isomer shifts are relative to the centroid of the spectrum of a metallic foil of α -Fe at room temperature (RT). Solution samples were transferred to a sample cup and chilled to 77 K inside of the glovebox, and unless noted otherwise, quickly removed from the glovebox and immersed in liquid N₂ until mounted in the cryostat. Data analysis was performed using version 4 of the program WMOSS (www.wmoss.org) and quadrupole doublets were fit to Lorentzian lineshapes except where noted. See discussion below for detailed notes on the fitting procedure.

C.10.2 Sample preparation for Mössbauer spectroscopy

Sample preparation conditions for Mössbauer samples were designed to be similar to those conditions used to assay the biocatalytic activity of *Rma* TDE, as shown in Supplementary Figure 4F. In an anaerobic chamber, an aliquot of ⁵⁷Fe-labelled purified *Rma* TDE (4.00 mM) was thawed. For the reduced *Rma* TDE sample, 341 μ L of protein (4.00 mM in 20 mM NaPi, pH 7.5) was added to a Mössbauer cup containing 3.7 mg sodium dithionite. Acetonitrile (18 μ L) was added to the Mössbauer cup, and the sample was pipetted up and down to ensure the sample was homogenous (approximately 10 seconds). The sample

was then flash-frozen on liquid nitrogen. Final concentrations: 3.8 mM protein, 19 mM NaPi (pH 7.5), 50 mM sodium dithionite, 5% acetonitrile.

For the reduced Me-EDA-treated *Rma* TDE sample, 341 μ L of protein (4.0 mM in 20 mM NaPi, pH 7.5) was added to a Mössbauer cup containing 3.7 mg sodium dithionite. 18 μ L of 1.0 M Me-EDA in acetonitrile was added to the Mössbauer cup, and the sample was pipetted up and down to ensure the sample was homogenous (approximately 10 seconds). The sample was then flash-frozen on liquid nitrogen. Final concentrations: 3.8 mM protein, 19mM NaPi (pH 7.5), 50 mM sodium dithionite, 50 mM Me-EDA, 5% acetonitrile.

C.11 Electron paramagnetic resonance (EPR) spectroscopy

C.11.1 General description of EPR spectroscopy




EPR spectra were recorded on a Bruker EMX X-band CW-EPR Spectrometer (Bruker, Billerica, MA) at 4 K. using Bruker Win-EPR software (ver. 3.0). Data were collected with in parallel mode with a power of 10 mW, and in perpendicular mode with a power of 2 mW.

C.11.2 Sample preparation for EPR spectroscopy

In an anaerobic chamber, an aliquot of purified *Rma* TDE was thawed and diluted to 1.00 mM with 20 mM NaPi, pH 7.5 with 10% v/v DMSO. Solid sodium dithionite was added to the purified *Rma* TDE. For the reduced *Rma* TDE sample, 250 μ L of the reduced protein (1.00 mM in 20 mM NaPi, pH 7.5) was added to an EPR tube containing 13 μ L DMSO, vortexed briefly, and flash-frozen on liquid nitrogen. Final concentrations: 950 μ M protein, 30 mM sodium dithionite, 19 mM NaPi, 15% v/v DMSO. For the reduced, Me-EDA treated *Rma* TDE sample, 250 μ L of reduced protein was added to an EPR tube containing 13 μ L 1M Me-EDA in DMSO, vortexed briefly, and flash-frozen on liquid nitrogen. Final concentrations: 950 μ M protein, 50 mM Me-EDA, 30 mM sodium dithionite, 19 mM NaPi, 15% v/v DMSO. Empty cavity measurements were performed without putting an EPR tube into the machine.

C.12 Supporting tables and figures

Table C-1. Data collection and refinement statistics of *Rma* TDE

	Carbene-bound <i>Rma</i> TDE ^a (6CUN)	<i>Rma</i> TDE ^a (6CUK)
Data collection		
Space group	P3 ₂ 21	P3 ₂ 21
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	60.3, 60.3, 77.2	60.0, 60.0, 77.3
  	90.0, 90.0, 120.0	90.0, 90.0, 120.0
(°)		
Resolution (Å)	38.62-1.29 (1.32-1.29) ^b	38.65-1.47 (1.50-1.47) ^b
<i>R</i> _{merge} , <i>R</i> _{meas} , <i>R</i> _{pim}	0.070, 0.072, 0.016 (1.2, 1.2, 0.30)	0.067, 0.073, 0.028 (0.425, 0.469, 0.270)
<i>I</i> / <i>σ</i> (<i>I</i>)	21.4 (2.9)	15.2 (3.7)
<i>CC</i> _{1/2}	1.0 (0.95)	1.0 (0.95)
Completeness (%)	100 (100)	99.8 (98.9)
Multiplicity	19.4 (16.6)	6.6 (5.6)
Refinement		
Resolution (Å)	1.29	1.47
No. reflections	39554	27920
<i>R</i> _{work} / <i>R</i> _{free}	0.182 / 0.202	0.183 / 0.209
No. atoms	997	1076
Protein	847	949
Heme	43	43
Carbene	7	
Water	100	83
<i>B</i> factors	15.67	23.62
Protein	12.95	23.09

Heme	16.36	16.00
Carbene	26.17	
Water	37.66	33.57
R.M.S. deviations		
Bond lengths (Å)	0.002	0.017
Bond angles (°)	1.814	1.719

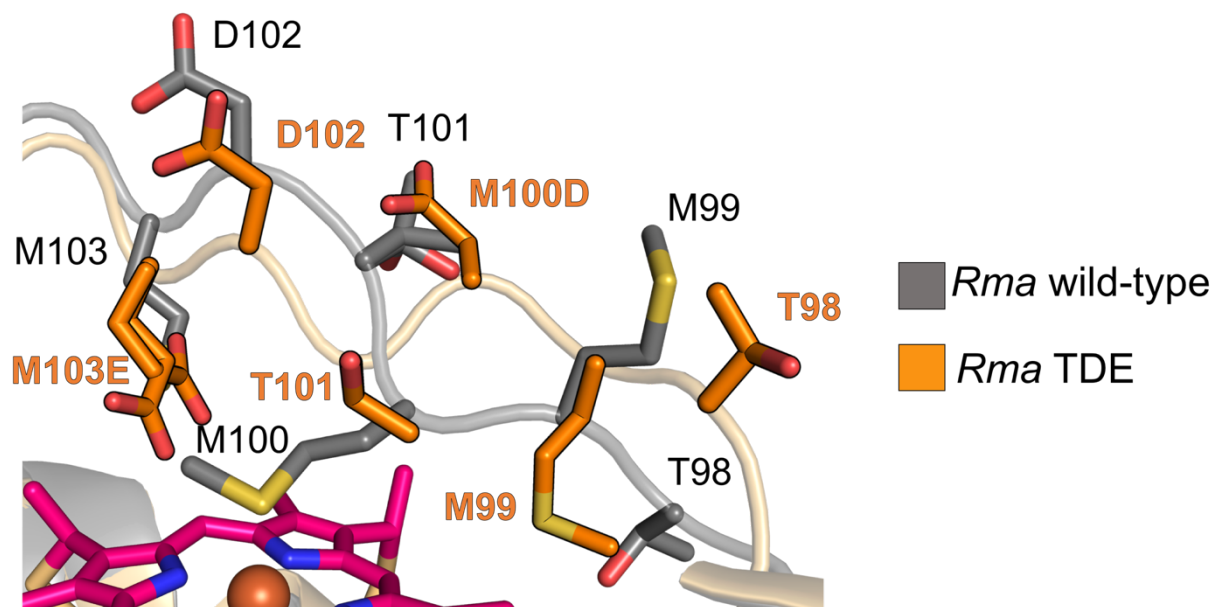
^a Data were collected on a single crystal

^b Values in parentheses are for highest-resolution shell.

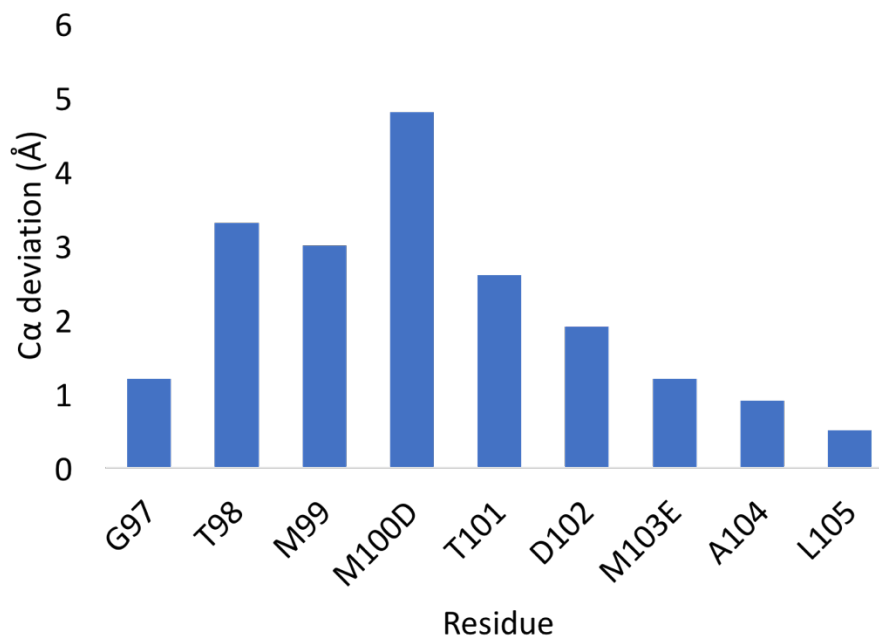
Table C-2. Bond lengths of interest

Crystal Structure	Fe–N(Imz) (Å)	Fe–O (Å)	Fe–C (Å)
6CUK	2.0	2.1	N/A
6CUN	2.1	N/A	1.9
Fe(TPFPP)(CPh ₂)(MeIm)	2.17	N/A	1.83

Figure C-1. Active-site front loop deviation. The active site front loop in *Rma* TDE is substantially deviated from its positioning in wild-type *Rma* cyt *c*.

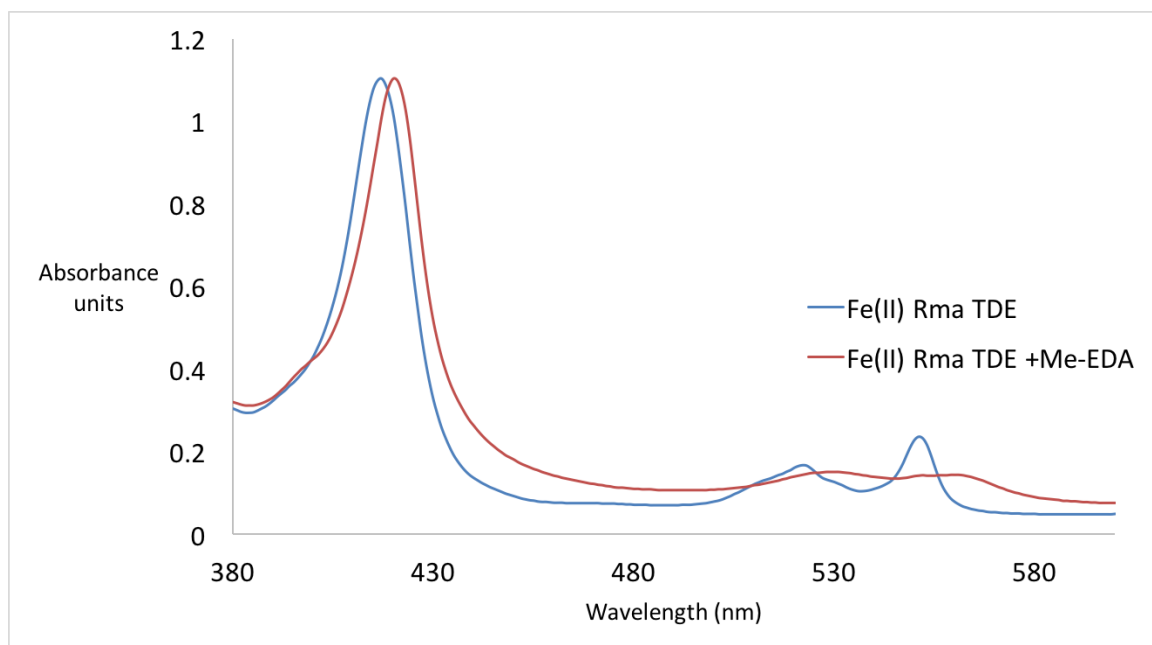


A. Alignment of *Rma* wild-type (PDB: 3CP5) and *Rma* TDE (PDB: 6CUK). The two structures superimposed on each other show that residues 98, 99, 100, and 101 have shifted the projection of their sidechains by approximately 180 degrees.

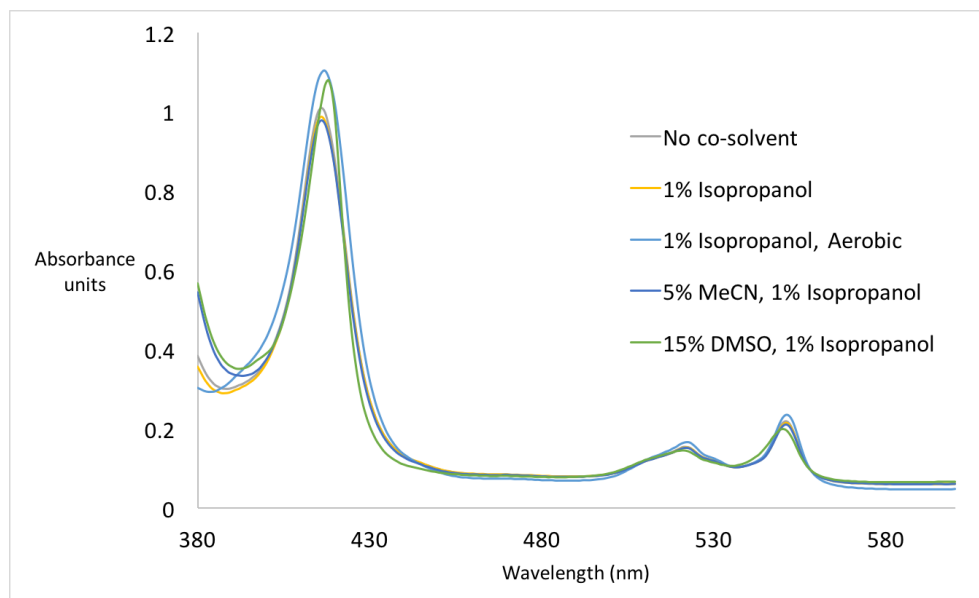


B. Measurement of alpha-carbon deviations between *Rma* wild-type (PDB: 3CP5) and *Rma* TDE (PDB: 6CUK). Distances were measured between the corresponding alpha carbons in wild-type *Rma* cyt *c*, and *Rma* TDE. The deviation measurements show that T98, M99, M100D, and T101 are substantially different in the two structures, with the largest deviation observed between M100 and D100 (4.8 Å).

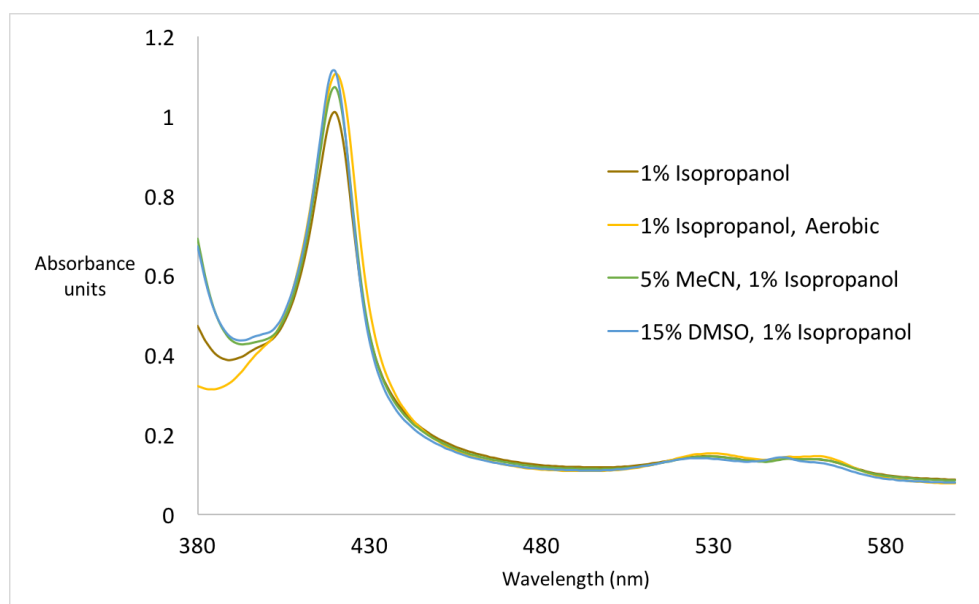
Figure C-2. UV-Vis spectra. UV-Vis spectra of reduced *Rma* TDE taken with and without the presence of the carbene precursor Me-EDA show that the spectrum changes upon addition of the carbene precursor.



A. Aerobic spectra of reduced *Rma* TDE, and reduced *Rma* TDE treated with Me-EDA. The UV-Vis spectra of 7 μ M reduced *Rma* TDE and 7 μ M reduced *Rma* TDE, treated with Me-EDA were taken under the ambient atmosphere. The reduced spectrum shows a Soret band at 419 nm, a Q-band peak at 551 nm, and a Q-band minimum at 534 nm. The Me-EDA treated sample shows a Soret band at 416 nm, and two Q-band peaks at 529 and 561 nm. Previously characterized small- molecule IPC complexes have Soret bands at 403-408 nm and two Q-bands at 522-525 nm and 550-556 nm¹⁷.

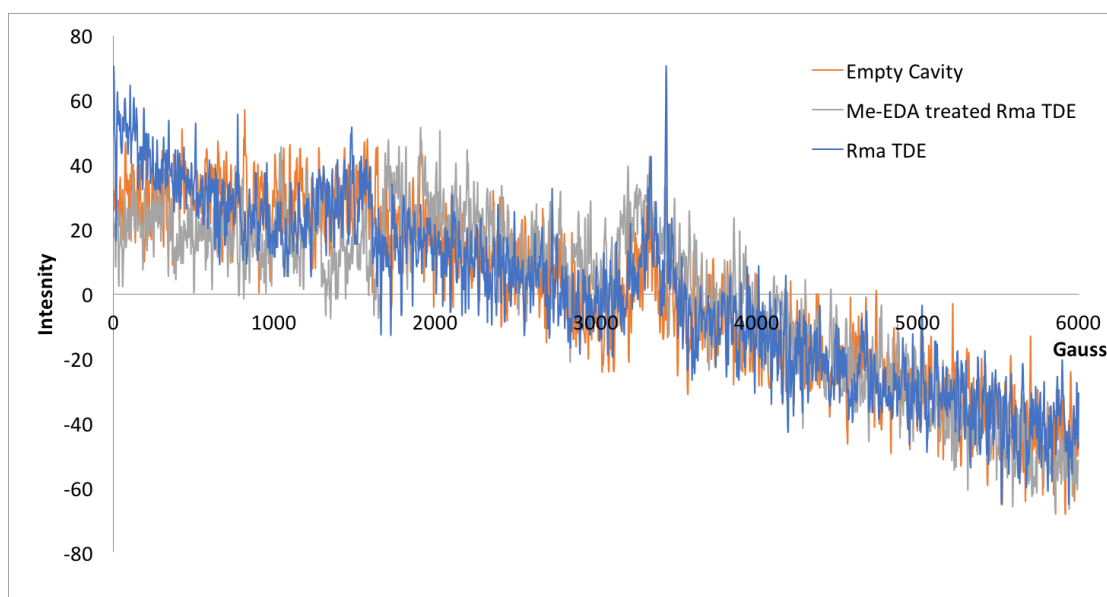


B. Spectra of reduced *Rma* TDE under various conditions. The UV-Vis spectra of 7 μM reduced *Rma* TDE were taken in the presence of various solvents, and unless specified, under an argon atmosphere. The reduced spectra show no substantial differences with the presence of isopropanol, acetonitrile (MeCN), or dimethylsulfoxide (DMSO). The spectrum also appears to undergo no substantial changes between an aerobic atmosphere and an argon atmosphere.

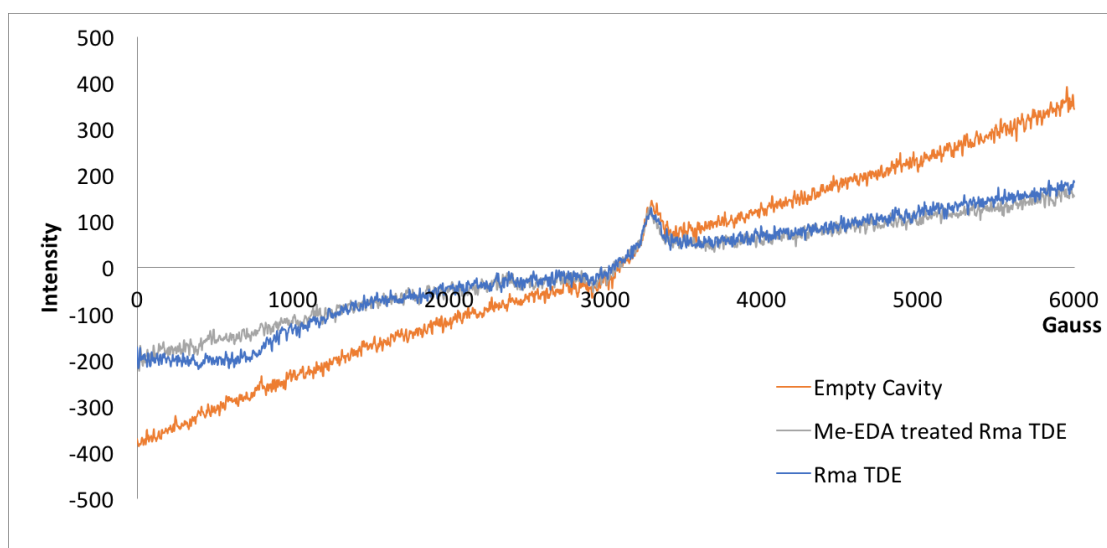


C. Spectra of reduced *Rma* TDE treated with Me-EDA under various conditions. The UV-Vis spectra of 7 μM reduced *Rma* TDE treated with Me-EDA were taken in the presence of various solvents, and unless specified, under an argon atmosphere. The spectra show no substantial differences with the presence of acetonitrile (MeCN) or dimethylsulfoxide (DMSO). The spectrum also appears to undergo no substantial changes between an aerobic atmosphere and an argon atmosphere.

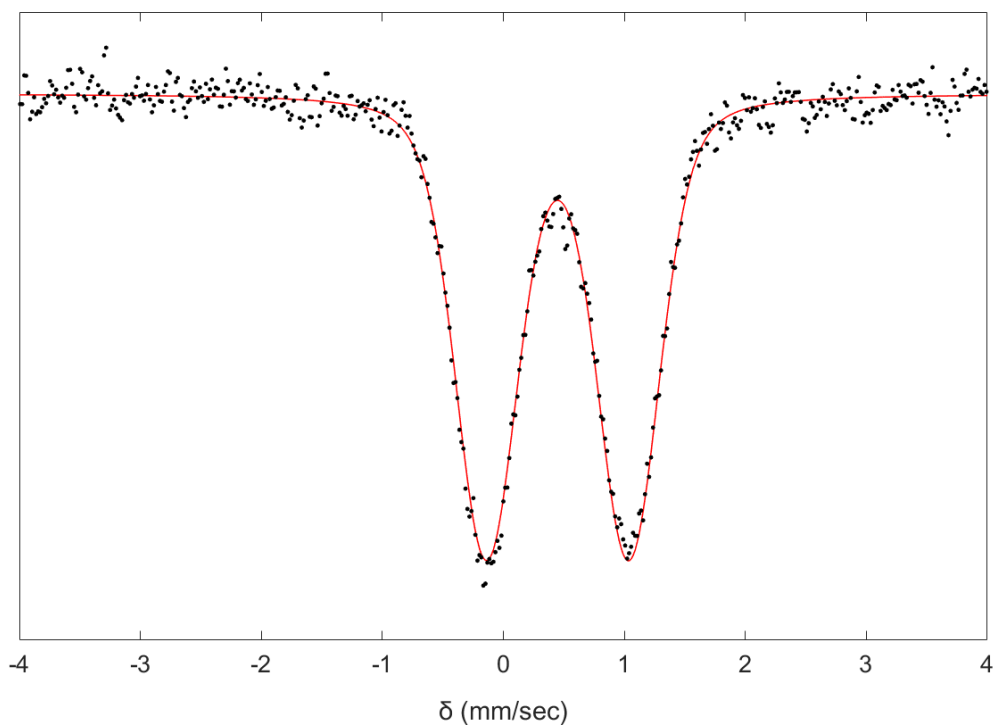
Figure C-3. EPR spectroscopic analysis.



A. EPR spectra in perpendicular mode. The spectra for all three samples look highly similar. No signals are discernible from these spectra, indicating that the protein complexes present have integer spin states.



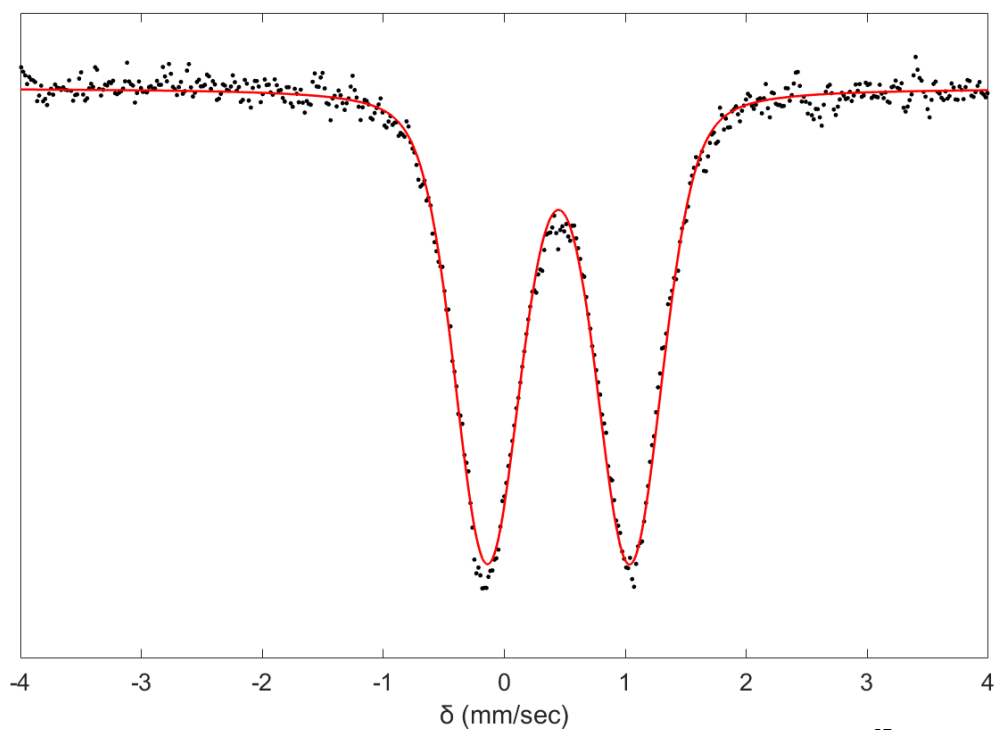
B. EPR spectra in parallel mode. The spectra for all three samples look highly similar. A signal is observed (at ~3300 Gauss) in all three spectra, including the empty cavity spectrum, which suggests this signal does not come from the protein. The observed signal is likely due to Cu(II) salts formed in the instrument as a result of oxidized conductors. Another small signal is observed (at ~800 Gauss) in the *Rma* TDE spectrum. This signal is typical of high spin Fe(II), but due to its low intensity, the signal likely comes from a small amount of Fe(II) contamination in the spectrum, rather than from *Rma* TDE. The lack of substantial protein-derived signal suggests that both protein complexes may be in low-spin states.

Figure C-4. Mössbauer spectroscopy studies.

A. Zero-field Mössbauer spectrum at 80 K of ^{57}Fe -protein (3.8 mM) in frozen NaPi (19 mM, pH 7.5) and MeCN (5% v/v) in the presence of sodium dithionite (50 mM).

Fitting details for Figure C-4A. This Mössbauer was modelled as one species with the parameters seen below. In this case, the data was best fit using a Voigtian lineshape. The Voigtian lineshape is a Gaussian distribution of the typical Lorentzian lineshape observed in Mössbauer spectra. Although less typically observed for solution samples, small differences in the protein (e.g. folding), in the local active-site environment (e.g. solvent, H-bonding network), or small domains in the sample caused by rapid freezing could all result in such a lineshape. Alternative fits such as asymmetric Lorentzian lineshapes or adding in a second component lead to worse fits. Application of a small magnetic field (50 mT, Figure C-4B) rules out incipient magnetic behavior as the cause of this behavior. The isomer shift and quadrupole splitting of this species are in agreement with its assignment as a $S = 0$, heme species¹⁸.

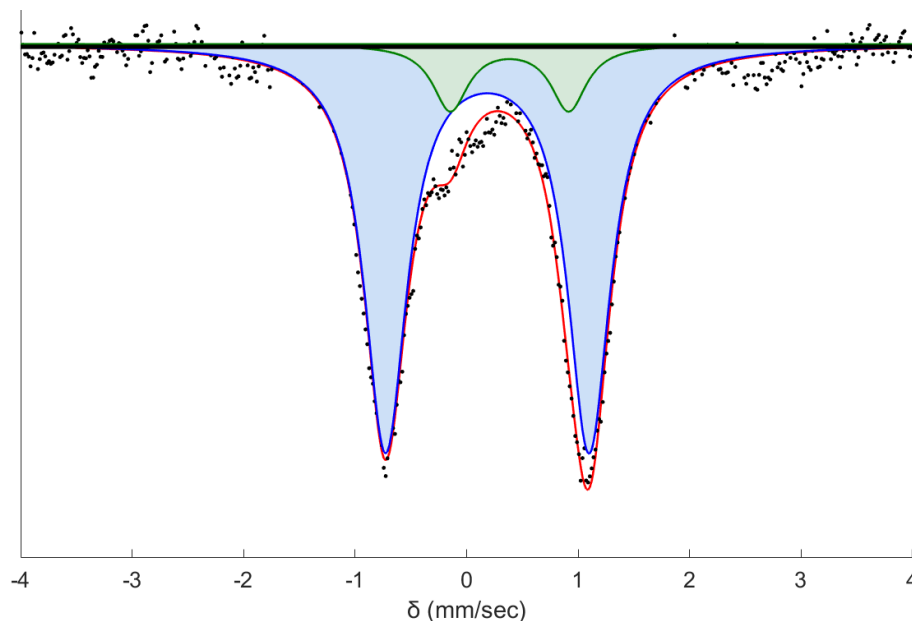
Component	δ (mm s ⁻¹)	ΔE_Q (mm s ⁻¹)	Linewidths, Γ_L/Γ_R (mm s ⁻¹)
Red line	0.45 ± 0.02	1.18 ± 0.02	-0.60/-0.60



B. Parallel applied 50-mT magnetic field Mössbauer spectrum at 80 K of ^{57}Fe -protein (3.8 mM) in NaPi (19 mM, pH 7.5) and MeCN (5% v/v) in the presence of sodium dithionite (50 mM).

Fitting details for Figure C-4B. As expected given the proposed diamagnetic nature of this species, the application of a small magnetic field does not result in any changes to the signal observed in the zero-field spectrum.

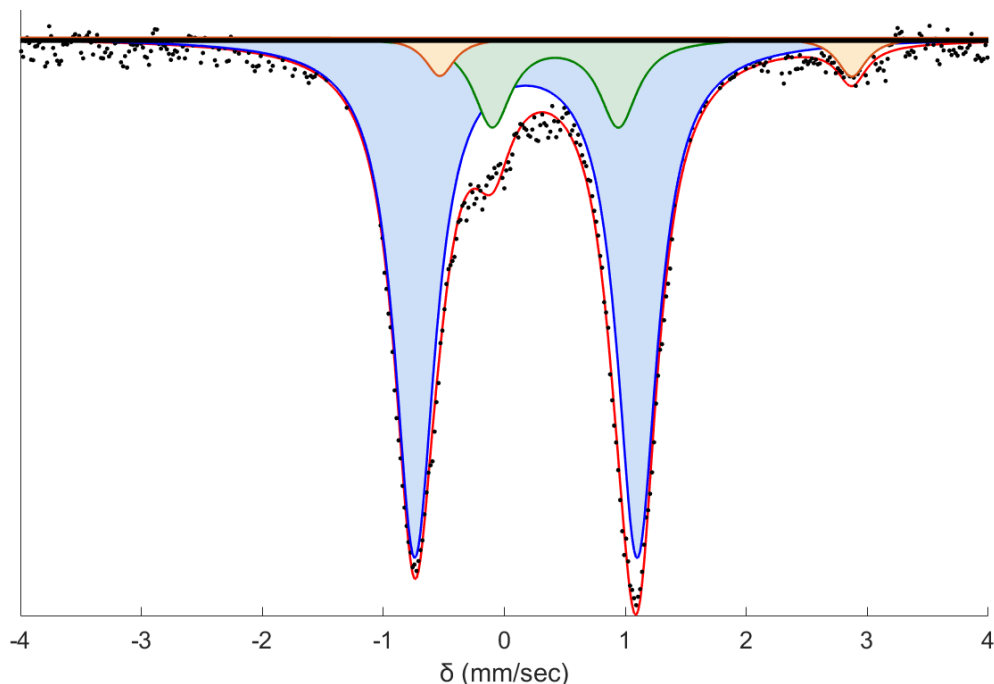
Component	δ (mm s $^{-1}$)	ΔE_Q (mm s $^{-1}$)	Linewidths, Γ_L/Γ_R (mm s $^{-1}$)
Red line	0.45 ± 0.02	1.18 ± 0.02	$-0.63/-0.63$



C. Zero-field Mössbauer spectrum at 80 K ^{57}Fe -protein (3.8 mM) in frozen NaPi (19 mM, pH 7.5) and MeCN (5% v/v) in the presence of sodium dithionite (50 mM) and Me-EDA (50 mM).

Fitting details for Figure C-4C. This Mössbauer was modelled as two species. The minor component is clearly defined by the absorbance feature at approximately -0.1 mm/sec. This fixes one half of that quadrupole doublet. The other absorbance is not clearly defined with respect to the major feature so the exact isomer shift and quadrupole splitting of this component is not well-defined. This is reflected in the slightly different parameters that are found for this feature in the spectrum taken with a 50 mT magnetic field applied. However, the fit featured here is reasonable for an Fe(II), $S = 0$ heme species which could be starting material, a reaction intermediate, or a carbene decomposition product. Forcing this feature to have the parameters of the $S = 0$ Fe(II) starting material ($\delta_{\text{Fe}} = 0.45$ and $|\Delta E_{\text{Q}}| = 1.18$) shown in Table C-4B did not substantially change the quality of the curve-fitting. To avoid over-interpretation this fit has not been enforced. The major absorbance is in agreement with Mössbauer parameters previously reported for formally Fe(IV), $S = 0$ heme species¹⁸.

Component	δ (mm s ⁻¹)	ΔE_{Q} (mm s ⁻¹)	Linewidths, $\Gamma_{\text{L}}/\Gamma_{\text{R}}$ (mm s ⁻¹)	Relative area
A (blue)	0.18 ± 0.02	1.83 ± 0.04	0.46/0.46	0.88
B (green)	0.38 ± 0.02	1.06 ± 0.02	0.38/0.38	0.12



D. Parallel applied 50-mT magnetic field Mössbauer spectrum at 80 K of ^{57}Fe -protein (3.8 mM) in frozen NaPi (19 mM, pH 7.5) and MeCN (5% v/v) in the presence of sodium dithionite (50 mM) and Me-EDA (50 mM).

Fitting details for Figure C-4D. This Mössbauer was modelled as three species. Again, the major component (modelled in blue) is consistent with a $S = 0$, formally Fe(IV) heme species. Again, there is a second species as seen in Figure S4.C (modelled in green) that shows parameters consistent with an $S = 0$, Fe(II) impurity. The third species here (modelled in orange) has isomer shift and quadrupole splitting characteristic of high-spin ($S = 2$) Fe(II) that is likely not bound to the heme. This species may be poorly resolved due to broadening in the absence of the applied magnetic splitting explaining why it is not seen in Figure C-4C.

Component	δ (mm s^{-1})	ΔE_Q (mm s^{-1})	Linewidths, Γ_L/Γ_R (mm s^{-1})	Relative area
A (blue)	0.18 ± 0.02	1.84 ± 0.04	0.41/0.41	0.82
B (green)	0.42 ± 0.02	1.04 ± 0.02	0.38/0.38	0.13
C (orange)	1.17 ± 0.02	3.40 ± 0.07	0.29/0.29	0.04

E. Comparison of Calculated and Experimental Mössbauer parameters

Entry Number	System	Method	Electronic state ^c	δ_{Fe} (mm/s)	ΔE_{Q} (mm/s)	Reference
1	<i>Rma</i> TDE (CMeCOOEt)	Experimental	CSS	0.18	(-) ^a 1.83	(This work)
2	Fe(Por)(CPh ₂)(Melm)	Calculated	CSS	0.19	-1.76	Khade <i>et al.</i> ¹⁹
3	Fe(TPFPP)(CPh ₂) ^b	Experimental	CSS	0.03	(-) ^a 2.34	Li <i>et al.</i> ¹⁷
4	Fe(Por)(CPh ₂)	Calculated	CSS	0.10	-2.37	Khade <i>et al.</i> ¹⁹

Im = imidazole

^a Sign not determined experimentally

^b TPFPP = 5,10,15,20-Tetrakis(pentafluorophenyl)porphyrin

The major species in Me-EDA treated *Rma* TDE was present at 88% abundance, a similarly high occupancy to that of the carbene observed in the crystal structure. This major species was best fit with $\delta_{\text{Fe}} = 0.18 \text{ mm}\cdot\text{s}^{-1}$ and $|\Delta E_{\text{Q}}| = 1.83 \text{ mm}\cdot\text{s}^{-1}$ (**Figure C-4E**, Entry 1). Although the isomer shift is higher than those typically observed for Fe(IV)–oxo species, the negative shift relative to the starting material is indicative of a short Fe–C bond. Additionally, the quadrupole splitting lies between the values typically observed for Fe(IV)–oxo species ($1\text{--}1.5 \text{ mm}\cdot\text{s}^{-1}$) and the values typically observed for Fe(IV)–hydroxo species ($2\text{--}2.1 \text{ mm}\cdot\text{s}^{-1}$)^{20–22}. Unlike the oxo species, both the hydroxo and carbene species are axially dissymmetric (d_{xz} and d_{yz} are no longer equivalent) and therefore are expected to display larger quadrupole splittings. Thus, based on its high occupancy and distinctive Mössbauer parameters, we conclude that the major species observed in the Me-EDA-treated sample is carbene-bound *Rma* TDE.

Previously, Khade and Zhang created a model of the IPC Fe(Por)(CPh₂) in a closed shell singlet state and simulated its Mössbauer parameters (**Figure C-4E**, Entry 3). Their simulations yielded numbers similar to those obtained experimentally for the complex Fe(TPFPP)(CPh₂) (**Figure C-4E**, Entry 4). Khade and Zhang also predicted parameters for the *N*-methylimidazole-ligated version of this complex, Fe(Por)(CPh₂)(Melm) (**Figure C-4E**, Entry 2). Although further experimental characterization such as high-field Mössbauer or SQUID magnetometry would have to be conducted to unambiguously assign the spin state, the data thus far suggest that the enzyme IPC, like small molecule IPC complexes, exists in a singlet state.

F. Comparison of sample preparation conditions between Mössbauer samples and biocatalytic reaction samples

Component	Final concentration in Me-EDA Mössbauer sample	Final concentration in biocatalytic reaction sample
<i>Rma</i> TDE	3800 μ M	2.0 μ M
NaPi, pH 7.5	19 mM	19 mM
Sodium dithionite	50 mM	10 mM
Me-EDA	50 mM	10 mM
Dimethylphenylsilane	0 mM	10 mM
Acetonitrile	5% (v/v)	5% (v/v)

To demonstrate the catalytic relevance of the captured carbene species, Mössbauer samples were prepared under conditions similar to those used to assay the activity and enantioselectivity of *Rma* TDE. The same buffer system and co-solvent were used for both the biocatalytic samples and the Mössbauer samples. Increased concentrations of purified *Rma* TDE protein were necessary in the Mössbauer samples to achieve sufficient signal. Similarly, the concentration of sodium dithionite was increased in the Mössbauer samples to ensure complete reduction of the protein, and the concentration of the carbene-precursor Me-EDA was also increased to ensure that the protein was subjected to multi-turnover conditions, with excess carbene precursor.

Table C-3. Out-of-plane heme distortion parameters for various IPC structures. Normal-Coordinate Structural Decomposition (NCS) quantitative analysis of the out-of-plane heme deformations using vibrational normal coordinates and complete-basis, as proposed by Jentzen and Shelnut²³, and implemented in the online script by Liptak and co-workers²⁴. The out-of-plane normal deformations considered are: mean deviation out-of-plane (Doo_p), saddling (B2u), ruffling (B1u), doming (A2u), waving (Eg(x), Eg(y)), and propelling (A1u). Values are given in Å. The values used for discussion in the manuscript are highlighted in bold.

Species	Doo _p	Saddling B2u	Ruffling B1u	Doming A2u	Waving (x) Eg(x)	Waving (y) Eg(y)	Propelling A1u
<i>Rma</i> TDE ^a	0.709	0.162	0.620	0.115	0.037	0.277	0.016
Carbene-bound <i>Rma</i> TDE ^b	0.603	0.243	0.491	0.164	0.099	0.162	0.026
Fe(TPFPP)(CPh ₂)(Melm _z) ^b	1.028	0.345	0.744	0.078	0.046	0.615	0.019

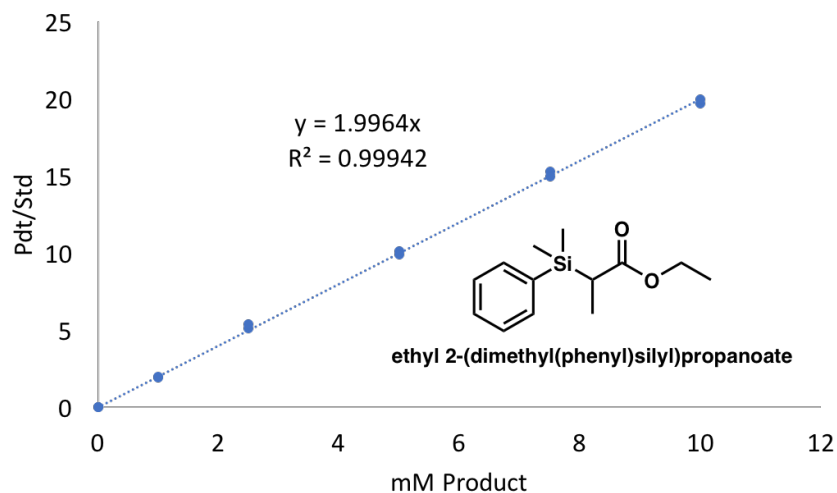
^a Measured from 6CUK PDB structure.

^b Measured from 6CUN PDB structure.

^b Measured from 1208784.cif crystal structure reported in Ref 17.

Out-of-plane distortion analysis on *Rma* TDE with and without the carbene ligand showed that for both structures, the primary out-of-plane distortion is ruffling, although the structure without the carbene ligand (6CUK, 0.620 Å) displays more ruffling than the structure with the carbene ligand (6CUN, 0.491 Å). Among the truncated model DFT-optimized structures, the triplet state is the least consistent with the out-of-plane deviations observed in carbene-bound *Rma* TDE, as the saddling deviation of the triplet is much lower than is crystallographically observed (0.010 Å vs 0.243 Å), and the ruffling deviation is much higher than is crystallographically observed (0.784 Å vs 0.491 Å). When considering the optimized structures by QM/MM, which take into account conformational restraints imposed by the protein on porphyrin ruffling, the out-of-plane distortion parameters for all three electronic structures are similar to that observed in the crystal structure due to the rigidity imposed by the protein to the heme.

Figure C-5. GC Standard curve for ethyl 2-(dimethylphenylsilyl)propanoate product.



Total turnover for reactions with 2 μ M *Rma* TDE was determined to be 828 +/- 15 TTN.

Pdt / std = ratio of product to internal standard (see section C.7)

Figure C-6. ^1H NMR for Ethyl 2-(dimethylphenylsilyl)propanoate product. ^1H NMR (400 MHz, CDCl_3) δ 7.55-7.46 (m, 2H), 7.43-7.32 (m, 3H), 4.02 (q, $J = 7.1$ Hz, 2H), 2.25 (q, $J = 7.1$ Hz, 1H), 1.17-1.11 (m, 6H), 0.37 (d, $J = 0.6$ Hz, 6H).

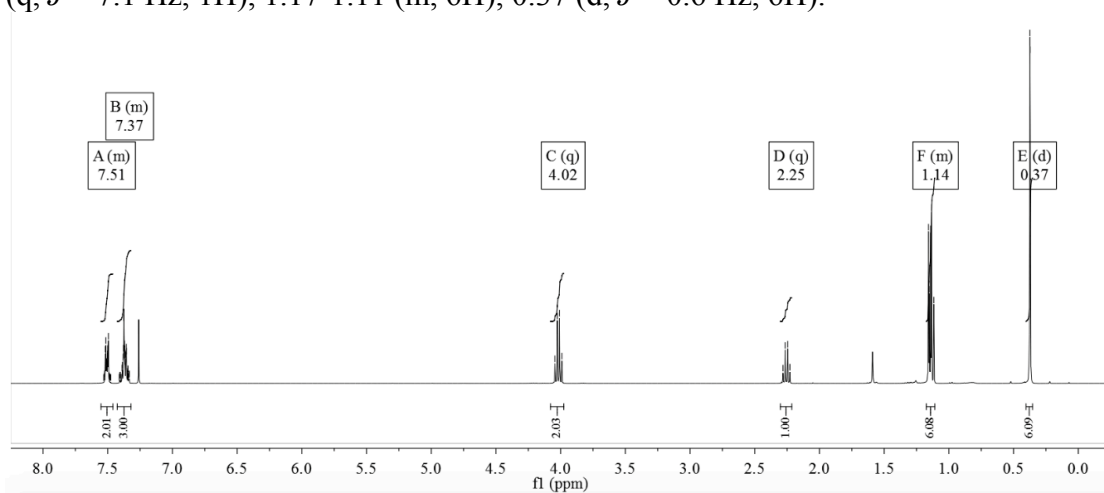
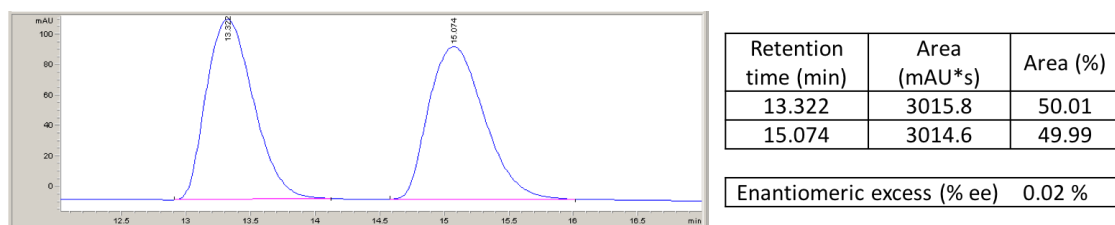
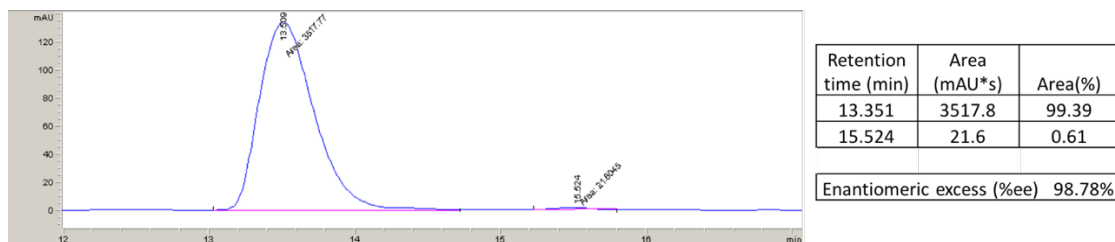


Figure C-7. Separation of ethyl 2-(dimethylphenylsilyl)propanoate.



A. Separation of racemic ethyl 2-(dimethylphenylsilyl)propanoate. A representative chiral HPLC trace of chemically-synthesized ethyl 2-(dimethylphenylsilyl)propanoate shows that the product is racemic, as expected.



B. Separation of enzymatically-produced ethyl 2-(dimethylphenylsilyl)propanoate. A representative chiral HPLC trace of ethyl 2-(dimethylphenylsilyl)propanoate produced from a reaction catalyzed by *Rma* TDE shows that the enzyme produces product with 99% *ee*. The product was assigned as having the *R* absolute configuration based on comparison of retention times to the literature.^{4,25}

C.13 References

- Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343 (2009).
- Arslan, E., Schulz, H., Zufferey, R., Kunzler, P. & Thony-Meyer, L. Overproduction of the Bradyrhizobium japonicum c-type cytochrome subunits of the cbb(3) oxidase in Escherichia coli. *Biochem. Biophys. Res. Commun.* **251**, 744–747 (1998).
- Sambrook, J. & W Russell, D. Molecular Cloning: A Laboratory Manual. *Cold Spring Harb. Lab. Press.* Cold Spring Harb. NY 999 (2001). doi:10.1016/0092-8674(90)90210-6
- Kan, S. B. J., Lewis, R. D., Chen, K. & Arnold, F. H. Directed evolution of cytochrome c for carbon–silicon bond formation: Bringing silicon to life. *Science* **354**, 1048–1051 (2016).
- Liptak, M. D., Wen, X. & Bren, K. L. NMR and DFT investigation of heme ruffling: Functional implications for cytochrome c. *J. Am. Chem. Soc.* **132**, 9753–9763 (2010).

6. Berry, E. A. & Trumpower, B. L. Simultaneous determination of hemes a, b, and c from pyridine hemochrome spectra. *Anal. Biochem.* **161**, 1–15 (1987).
7. Stelter, M., Melo, A. M., Saraiva, L. M., Teixeira, M. & Archer, M. Crystallization and X-ray analysis of *Rhodothermus marinus* cytochrome c at 1.23 Å resolution. *Protein Pept Lett* **14**, 1038–1040 (2007).
8. Kabsch, W. I XDS. *Acta Crystallogr. Sect. D* **66**, 125–132 (2010).
9. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
10. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
11. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* **67**, 235–242 (2011).
12. Stelter, M. et al. A novel type of monoheme cytochrome c: Biochemical and structural characterization at 1.23 Å resolution of *Rhodothermus marinus* cytochrome c. *Biochemistry* **47**, 11953–11963 (2008).
13. Winn, M. D., Murshudov, G. N. & Papiz, M. Z. Macromolecular TLS Refinement in REFMAC at Moderate Resolutions. *Methods Enzymol.* **374**, 300–321 (2003).
14. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
15. Chen, V. B. et al. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21 (2010).
16. Painter, J. & Merritt, E. A. TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* **39**, 109–111 (2006).
17. Li, Y., Huang, J., Zhou, Z., Che, C. & You, X. Remarkably stable iron porphyrins bearing nonheteroatom-stabilized carbene or (alkoxycarbonyl)carbenes: isolation, X-ray crystal structures, and carbon atom transfer reactions with hydrocarbons. *J. Am. Chem. Soc.* **124**, 13185–13193 (2002).
18. English, D. R., Hendrickson, D. N. & Suslick, K. S. Mössbauer Spectra of Oxidized Iron Porphyrins. *Inorg. Chem.* **22**, 367–368 (1983).
19. Khade, R. L. et al. Iron Porphyrin Carbenes as Catalytic Intermediates: Structures, Mössbauer and NMR Spectroscopic Properties, and Bonding. *Angew. Chem. Int. Ed.* **53**, 7574–7578 (2014).

20. Stone, K. L., Hoffart, L. M., Behan, R. K., Krebs, C. & Green, M. T. Evidence for two ferryl species in chloroperoxidase compound II. *J. Am. Chem. Soc.* **128**, 6147–6153 (2006).
21. Behan, R. K., Hoffart, L. M., Stone, K. L., Krebs, C. & Green, M. T. Evidence for basic ferryls in cytochromes P450. *J. Am. Chem. Soc.* **128**, 11471–11474 (2006).
22. Yosca, T. H. et al. Iron(IV)hydroxide pKa and the Role of Thiolate Ligation in C–H Bond Activation by Cytochrome P450. *Science* **342**, 825–829 (2013).
23. Jentzen, W., Song, X.-Z. & Shelnut, J. A. Structural Characterization of Synthetic and Protein-Bound Porphyrins in Terms of the Lowest-Frequency Normal Coordinates of the Macrocycle. *J. Phys. Chem. B* **101**, 1684–1699 (2002).
24. Graves, A. B., Graves, M. T. & Liptak, M. D. Measurement of Heme Ruffling Changes in MhuD Using UV-vis Spectroscopy. *J. Phys. Chem. B* **120**, 3844–3853 (2016).
25. 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.* **138**, 9597–9610 (2016). doi:10.1021/jacs.6b04636
26. Chen, D., Zhu, D. X. & Xu, M. H. Rhodium(I)-Catalyzed Highly Enantioselective Insertion of Carbenoid into Si-H: Efficient Access to Functional Chiral Silanes. *J. Am. Chem. Soc.* **138**, 1498–1501 (2016).

*Chapter 4*ORIGINS OF CHEMOSELECTIVITY IN CYTOCHROME C-CATALYZED
CARBENE INSERTIONS INTO SILICON–HYDROGEN BONDS

Material from this chapter is adapted from a manuscript in preparation which will be submitted with the following author list: Garcia-Borràs M[†], Lewis RD[†], Tang A, Jimenez-Osés G, Kan SBJ, Arnold FH, Houk KN.

[†] Denotes equal author contribution.

4.1 Abstract

Heme proteins have recently been engineered to catalyze the formation of carbon–silicon bonds via carbene insertion into Si–H bonds. Although this reactivity was unprecedented in the biological world, engineered cytochrome *c* proteins are now the most efficient and selective catalysts available for this transformation. Using computational and experimental tools, we create a detailed model of how the reaction occurs in the active site of the protein. Our model reveals that enzyme activity and chemoselectivity are modulated by the conformational ensemble of the cytochrome *c* protein front loop. Using mechanistic understanding to guide directed evolution, we further showed that the cytochrome *c* carbene transferase can switch chemoselectivity from silylation to amination. This work demonstrates that information on protein structure, dynamics, and enzyme mechanism can provide explanations of how non-natural chemical transformations become possible in the biological world, can guide evolution in the laboratory and can propel the discovery of new enzyme functions.

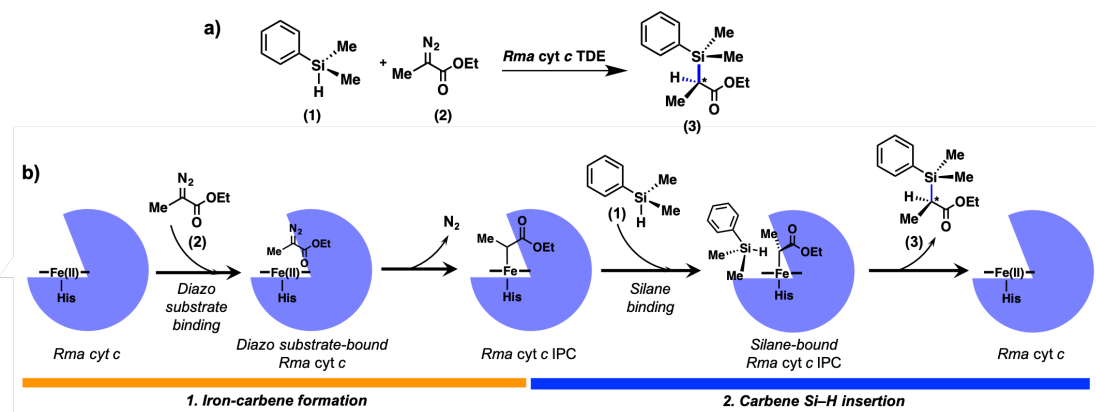
4.2 Introduction

The discovery of novel carbene transfer enzymes has largely been driven by repurposing existing heme proteins, and evolving these proteins for improved activity for a specific, novel reaction¹⁻¹². Fundamentally, the process of repurposing and evolving enzymes for novel reactivity is an alteration of enzyme chemoselectivity, which makes engineering carbene transfer enzyme chemoselectivity key to the development of new carbene transfer enzymes. Engineering enzymes for chemoselectivity is also important for developing useful biocatalysts that can selectively functionalize compounds without the need for protection and deprotection steps^{8,13,14}. Despite the importance of engineering altered chemoselectivity in carbene transfer enzymes, how these enzymes achieve chemoselectivity remains poorly understood⁸. Metal carbene intermediates are innately reactive towards a broad array of functional groups, and yet carbene transfer enzymes are able to achieve high degrees of chemoselectivity^{1,7,8}.

Previously, we discovered that a protein engineered to forge carbon–silicon bonds (*Rma* TDE, see Chapter 2) is a remarkably chemoselective carbene transfer enzyme¹. *Rma* TDE is capable of differentiating between the functionalization of Si–H and N–H bonds with 97% selectivity for silylation, even though selection pressure for chemoselectivity was not applied during the directed evolution of this enzyme¹. This preference for Si–H functionalization is the opposite of what a small molecule catalyst would typically prefer^{1,15}, and the opposite of what is observed in myoglobin-based (Mb) carbene transfer competition experiments⁸. Interestingly, wild-type *Rma* cyt *c* functionalizes N–H and Si–H bonds with approximately equal preference, suggesting that the mutations in *Rma* TDE (V75T M100D M103E) somehow encode chemoselectivity for silicon functionalization¹.

In order to understand how the protein achieves chemoselectivity in this new-to-nature reaction, we must first elucidate precisely how the reaction occurs in the protein. Heme protein-catalyzed carbene insertion reactions are proposed to occur through the formation of a reactive iron porphyrin carbene (IPC) intermediate, followed by transfer of the carbene equivalent to a second substrate, such as an olefin, Si–H bond, or an amine N–H bond¹⁶⁻¹⁹ (**Scheme 4-1**). Previous studies of iron porphyrin-catalyzed carbene transfer reactions have suggested that three distinct IPC electronic states could be involved in these reactions: the

closed-shell singlet (CSS), open-shell singlet (OSS), and triplet (T) states^{17,18,20–22}. The protein structure has been shown to influence the relative stabilities of various electronic states and therefore has to be considered when determining the overall reaction mechanism¹⁹. Herein, using experimental and computational methods, we investigate the mechanism of iron porphyrin-catalyzed carbene Si–H insertion and determine how the reaction mechanism is affected by the protein scaffold and mutations introduced by directed evolution. Using our understanding of how the reaction occurs in the enzyme, we elucidate how the enzyme controls the selectivity of carbene transfer to competing Si–H and N–H bonds, and further use our findings to guide enzyme engineering to switch silylation-amination chemoselectivity.



Scheme 4-1. a) Stereoselective carbon–silicon bond formation catalyzed by *Rma* cytochrome *c* V75T M100D M103E variant (*Rma* TDE). *Rma* TDE was specifically evolved to catalyze carbene insertions into Si–H bonds using silane 1 and ethyl 2-diazopropanoate (Me-EDA, 2) as substrates. b) Schematic description of the proposed mechanism for *Rma* cytochrome *c*-catalyzed carbon–silicon bond formation via 1) iron-carbene formation; followed by 2) carbene Si–H insertion.

4.3 Results and Discussion

4.3.1 Quantum mechanical characterization of the reaction mechanism

To understand how iron porphyrin catalyzes carbene insertion into Si–H bonds, we first applied Density Functional Theory (DFT) calculations to a simple truncated model (imidazole-ligated iron porphyrin). Along the reaction pathway, three possible electronic states for the iron center were considered: closed-shell singlet (CSS), open-shell singlet (OSS), and triplet (T) (see Appendix D, **D.2**). For iron-carbene formation from the ethyl 2-

diazopropanoate substrate (Me-EDA, **2**), DFT calculations suggest that the closed-shell singlet electronic state **TS1-CSS** is lowest in energy ($\Delta G^\ddagger = 27.5 \text{ kcal}\cdot\text{mol}^{-1}$) (**Figure 4-1**). Hybrid quantum mechanics/molecular mechanics (QM/MM) calculations (simulating the iron-carbene-forming transition state in *Rma* TDE) again show that the closed-shell transition state **TDE-TS1-CSS** has the lowest energy barrier of the three electronic states.

We next investigated the mechanism of the C–Si bond formation step, and used the truncated model to characterize the electronic and geometric features of the transition state (**TS2**) using dimethyl(phenyl)silane **1** as substrate. The closed-shell singlet transition state for Si–H insertion (**TS2-CSS**) was found to be lowest in energy, with $\Delta G^\ddagger = 13.0 \text{ kcal}\cdot\text{mol}^{-1}$. This is $12.5 \text{ kcal}\cdot\text{mol}^{-1}$ lower in Gibbs energy than the optimized triplet transition state (**Figure 4-1, TS2-T**). Optimization of the open-shell singlet transition state **TS2-OSS** was not successful using this truncated model (see Appendix D, **D.2**). The intrinsic reaction coordinate (IRC) for **TS2-CSS** shows that Si–H insertion is concerted but highly asynchronous, consisting of partial hydride-transfer from the silane to the carbene carbon, followed by C–Si bond formation (**Figure 4-2**). In contrast, **TS2-T** favors a radical stepwise pathway. A partial positive charge is accumulated on the silicon atom in **TS2-CSS**, consistent with previous Hammett analysis of the *Rma* TDE-catalyzed reaction: when the relative reaction rates of various *para*-substituted aryldimethylsilanes were plotted against the Hammett σ values for the corresponding *para*-substituents, the calculated ρ value for *Rma* TDE was -0.70 (indicating a partial positive charge build-up on the Si atom in the rate-determining transition state)¹. This ρ value is larger than those reported for Rh and Cu-based catalysts, for which ρ is -0.31 and -0.54 , respectively^{23,24}.

QM/MM calculations were also used to model the C–Si bond formation step in the context of *Rma* TDE. The concerted closed-shell singlet **TDE-TS2-CSS** has a $\Delta E^\ddagger = 20.0 \text{ kcal}\cdot\text{mol}^{-1}$, which is similar in energy to the open-shell singlet **TDE-TS2-OSS** ($\Delta\Delta E^\ddagger = 0.4 \text{ kcal}\cdot\text{mol}^{-1}$), and lower in energy than the triplet **TDE-TS2-T** ($\Delta\Delta E^\ddagger = 10.1 \text{ kcal}\cdot\text{mol}^{-1}$) transition states.

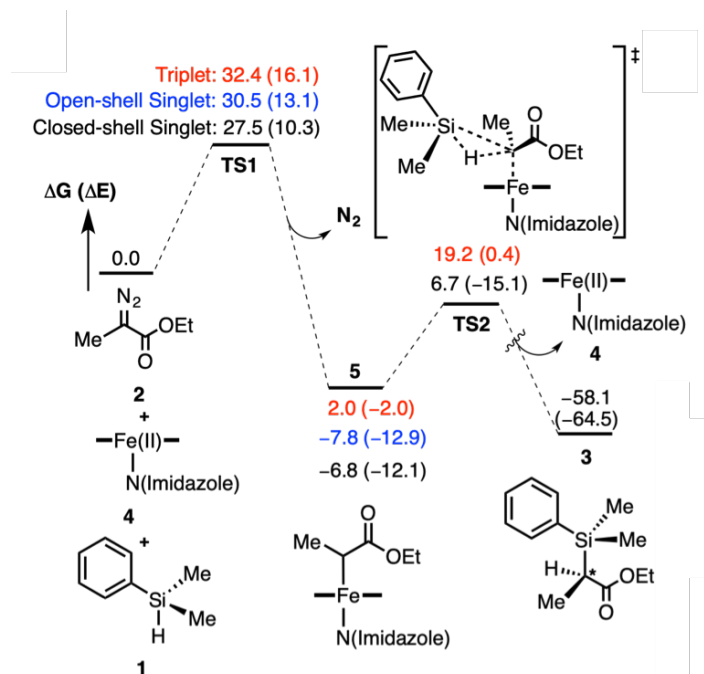


Figure 4-1. Computed reaction pathway for carbene Si-H insertion reaction between PhMe_2SiH **1** and Me-EDA **2** catalyzed by a model imidazole-ligated iron porphyrin (truncated QM-model). Results with three iron electronic states are given. Electronic and Gibbs free energies are given in $\text{kcal}\cdot\text{mol}^{-1}$.

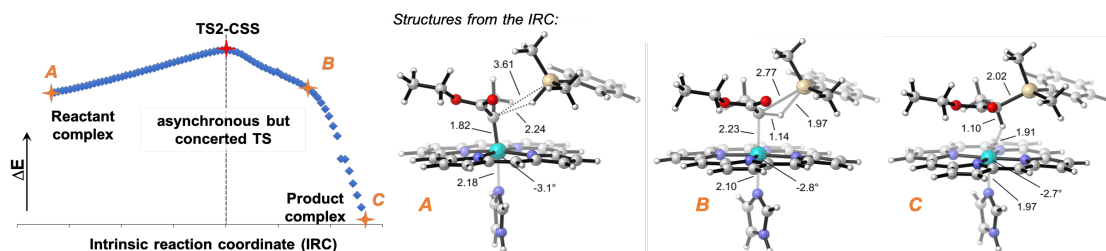


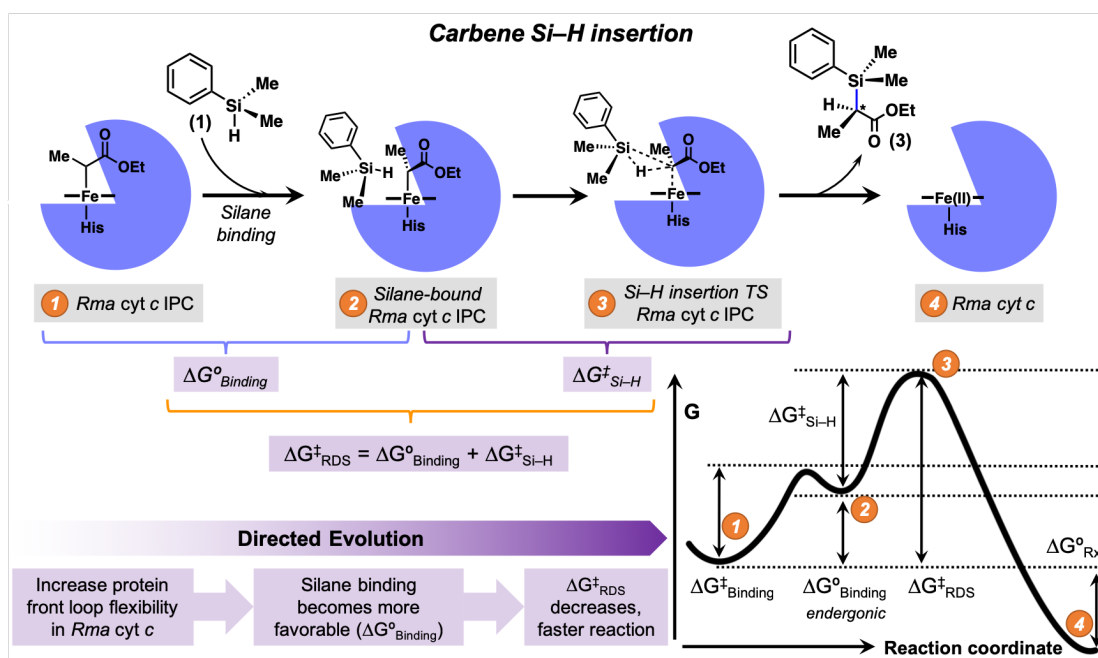
Figure 4-2. Intrinsic reaction coordinate (IRC) calculation for the lowest energy Si-H carbene insertion transition state (TS2-CSS) calculated using truncated models. Key distances and angles are shown in Å and degrees.

4.3.2 Rate-determining step of the enzymatic reaction

QM calculations on the model system and QM/MM calculations on *Rma* TDE both predict that the energy barrier for C–Si bond formation is lower than that for carbene formation, suggesting formation of the carbene is intrinsically rate-determining in the overall reaction. However, this contrasts with our earlier Hammett analysis, where we observed that substitutions to the silane can increase the initial rate of the reaction, which suggests that the Si–H insertion step is rate-determining. To further interrogate the reaction mechanism and rate-determining step (RDS) experimentally, the initial-rate primary kinetic isotope effect (KIE) of the *Rma* TDE-catalyzed reaction was determined using dimethyl(phenyl)silane (PhMe₂Si-H) and dimethyl(phenyl)silane-*d* (PhMe₂Si-D). The KIE value obtained for *Rma* TDE was 2.1 ± 0.2 (see Appendix D, **Figure D.1**), demonstrating Si–H bond cleavage is rate-determining in an experimental system, while carbene formation is not. The experimental initial-rate KIE for *Rma* TDE is comparable to that computed for **TDE-TS2-CSS** (1.74). The predicted KIE values for **TDE-TS2-OSS** and **TDE-TS2-T** are 2.80 and 2.95, respectively (**Table S1**). In addition to the initial-rate KIE value, the end-point KIE value was also measured for *Rma* TDE, and determined to be 1.34 ± 0.08 . This is similar to the end-point KIE values reported for Rh-, Ir-, and Cu-catalyzed reactions (1.29 and 1.5 for Rh, 1.6 for Ir, 1.08 for Cu), where the C–Si bond formation step is believed to be both rate-limiting and concerted^{23–25}. Overall, our experimental results support a concerted carbene Si–H insertion mechanism, and suggest the C–Si bond formation step is rate-limiting in *Rma* TDE. A concerted reaction mechanism is in agreement with our QM/MM calculations, which suggest that the concerted and nonradical **TDE-TS2-CSS** is the lowest energy transition state. Additionally, a concerted mechanism also supports our previous hypothesis on how the enzyme controls stereoselectivity: the stereochemical outcome of C–Si bond formation is solely determined by how the silane substrate approaches the IPC intermediate in the enzyme active site¹⁹.

Experimental evidence demonstrates that carbene Si–H insertion is rate-determining, although our quantum mechanical results show that the intrinsic energy barrier for C–Si bond formation is lower than that of carbene formation. This apparent discrepancy may arise because our quantum mechanical calculations use a substrate-protein complex in

which the silane substrate is already optimally positioned to approach the carbene-bound protein (the Michaelis-Menten complex). If the rate of dissociation of the Michaelis-Menten complex is much greater than the rate of C–Si bond formation, the C–Si bond formation step will be rate-limiting²⁶. Fast dissociation of the Michaelis-Menten complex suggests that its formation is thermodynamically unfavorable, and that the enzymatic rate-determining step ($\Delta G^{\ddagger}_{\text{RDS}}$) depends on both the quantum mechanical energy barrier ($\Delta G^{\ddagger}_{\text{Si-H}}$) and the binding energy ($\Delta G^{\circ}_{\text{Binding}}$) associated with the energy required for the enzyme to bind silane substrate in a catalytically competent pose (**Scheme 4-2** and **Scheme D-1**). In the next section, we will address how silane binding occurs and how it may affect the rate of enzymatic silylation.



Scheme 4-2. Schematic representation of the reaction profile of *Rma cyt c* IPC insertion into Si–H bond, the rate-determining step (RDS) of enzymatic C–Si bond formation. The RDS energy barrier consists of two components: *i*) the silane substrate binding affinity ($\Delta G^{\circ}_{\text{Binding}}$), which is >0 when binding is endergonic; and *ii*) the Si–H insertion energy barrier ($\Delta G^{\ddagger}_{\text{Si-H}}$). When silane binding becomes more favorable (*less endergonic*), $\Delta G^{\ddagger}_{\text{RDS}}$ decreases, leading to an overall faster reaction. Silane binding is more favorable for *Rma* TDE IPC than for wild-type *Rma* IPC because of the more flexible protein front loop in the evolved variant (see *Section 3* for details).

4.3.3 Silane binding to *Rma* cyt *c*

Previously, we found that one effect of the mutations in *Rma* TDE is to modulate the conformation of the protein front loop (residues 98–103) in such a way that facilitates access to the active site of carbene-bound *Rma* TDE by the silane substrate (**Figure 4-3a**)¹⁹. We hypothesized that the changes in the protein front loop conformation might also modulate the energetics of silane binding, and thus the rate-determining step of the reaction (**Scheme 2**). Using umbrella sampling MD (US-MD) simulations (see Appendix D, **D.4**), we compared the energetic cost of silane binding in carbene-bound wild-type *Rma* cyt *c* (WT IPC) and carbene-bound *Rma* TDE (TDE IPC) (**Figure 4-3b**). In both cases, the protein front loop must change conformation to accommodate the incoming silane substrate, and in both cases the silane substrate is only loosely bound by the protein: it is highly solvent exposed and not completely surrounded by the protein residues (**Figure 4-3c**, left structure). As a consequence, the silane binding process is found to be thermodynamically unfavorable, with positive binding energies in both cases ($\Delta G^{\circ}_{\text{Binding}} > 0$) (**Figure 4-3b** and **Scheme 4-2**). Because *Rma* cyt *c* does not naturally bind any substrates, it is perhaps unsurprising that both wild-type and *Rma* TDE have a poor ability to bind silane **1** after the IPC is formed in the cyt *c* active site. However, the silane binding energy in *Rma* TDE is ca. 2 kcal·mol⁻¹ lower (more favorable) than in the wild-type system. (**Figure 4-3b**). This improvement in the ability of *Rma* TDE to bind the silane may arise from increased front loop flexibility in IPC-bound *Rma* TDE IPC relative to IPC-bound *Rma* wild-type, or could also arise from the increased solvent exposure of D100 and E103, (**Figure 4-3d**), which might cause the protein front loop to adopt a conformation that better favors silane binding.

Based on these results, the C–Si bond formation step is dependent on the concentration of a thermodynamically unfavorable Michaelis-Menten complex, which likely is what causes the C–Si bond formation step to be rate-determining in the experimental system. Because of the improved silane binding in *Rma* TDE with respect to *Rma* wild-type, the concentration of the Michaelis-Menten complex would be higher in the *Rma* TDE system, resulting in higher turnover frequency.

From kinetics experiments, we have determined that *Rma* TDE has a 7-fold increase in initial rate relative to the wild-type protein (**Figure D-1**) In the transition state theory (TST) framework, a 7-fold increase in rate represents an energy difference of ca. $1.2 \text{ kcal}\cdot\text{mol}^{-1}$ between the RDS barriers for wild-type *Rma* cyt *c* and *Rma* TDE. This experimentally estimated energy difference ($\Delta\Delta G_{\text{RDS}}^\ddagger$) of $1.2 \text{ kcal}\cdot\text{mol}^{-1}$ agrees with the computationally estimated energy difference in silane binding ($\Delta\Delta G_{\text{Binding}}^\circ$) between wild-type *Rma* cyt *c* and *Rma* TDE (ca. $2 \text{ kcal}\cdot\text{mol}^{-1}$ more favorable in *Rma* TDE).

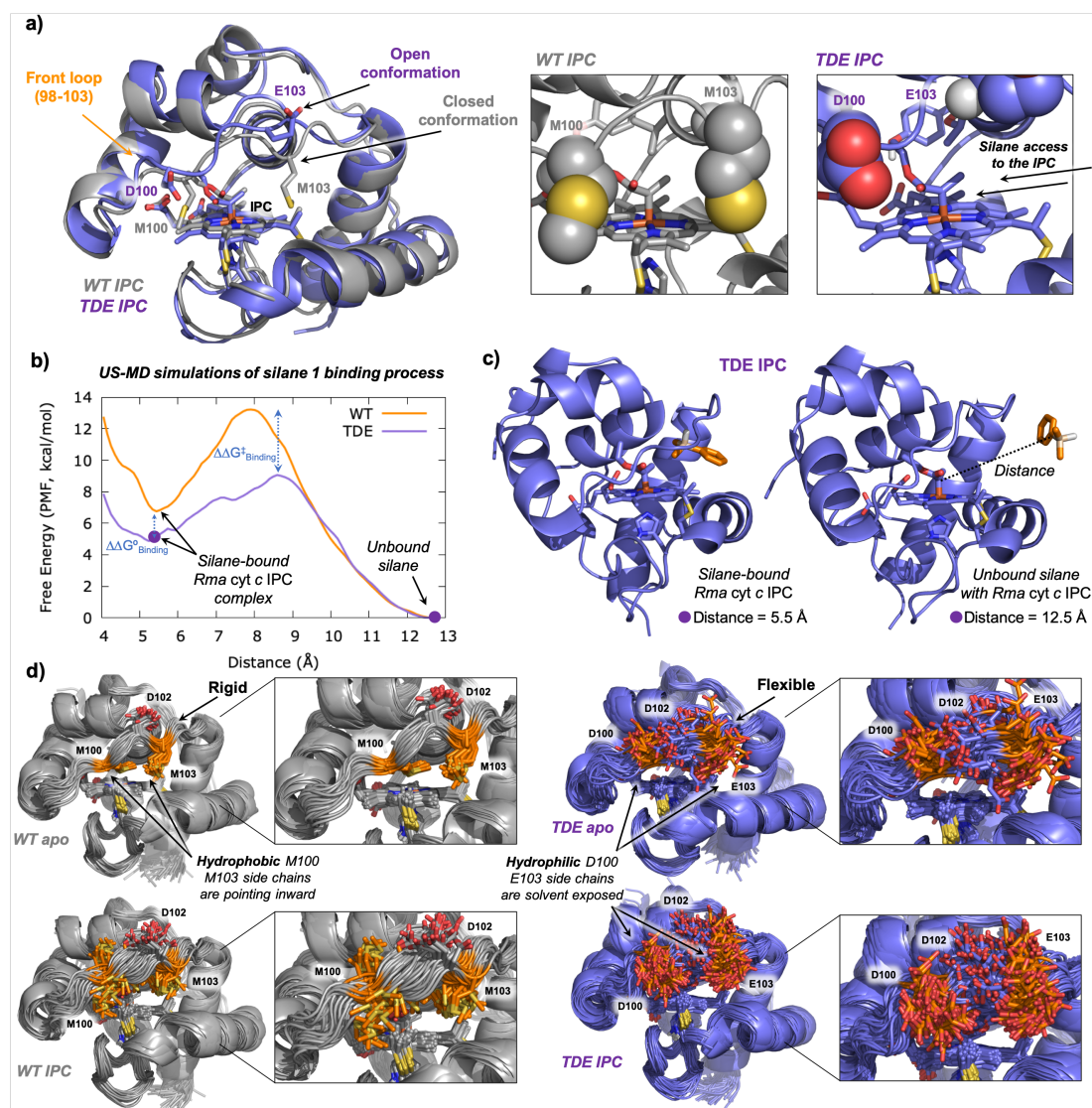


Figure 4-3. a) Representative snapshots obtained from MD simulations of IPC bound in wild-type *Rma* cyt *c* (gray), and the *Rma* TDE variant (purple). Front loop M100/D100 and M103/E103 residues are represented in space filling format. Hydrophobic methionines

(M100 and M103) close the substrate channel in wild-type *Rma* cyt *c*, but mutations M100D and M103E introduce hydrophilic residues that are exposed to the solvent in *Rma* TDE along the MD simulations, allowing an easier access of the silane substrate to interact with the IPC. V75T creates new key H-bond with Y71 that induce a more hydrophobic environment for the IPC intermediate. b) Potential of mean force (PMF) measured along the reaction coordinate defined as the distance between the IPC central C atom and the center of mass of the silane substrate 1 for wild-type *Rma* cyt *c* and *Rma* TDE IPC variants. c) Representative snapshots of the PMF reaction coordinate obtained from the *Rma* TDE PMF simulations. d) Overlay of 50 snapshots (every 10 ns) obtained from 500 ns MD simulations for wild-type *Rma* cyt *c* and *Rma* TDE variants in its apo and IPC-bound states. Residues mutated in the front loop (original M100 and M103 in WT, and D100 and E103 in TDE variant) are shown in orange. Insets highlight the flexibility acquired by the front loop due to new introduced mutations (going from “WT apo” to “TDE apo”) and the formation of the IPC (going from “WT apo” and “TDE apo” to “WT IPC” and “TDE IPC”, respectively).

4.3.4 Rationalization of chemoselectivity

With our model of the Si–H functionalization mechanism in satisfactory agreement with our experimental results, we compared the mechanisms of imidazole-ligated iron porphyrin-catalyzed carbene transfer to N–H and Si–H bonds (**Figures 4-3a–b**). N–H functionalization can occur through a concerted N–H insertion mechanism involving hydrogen atom transfer (HAT) (**Figure 4-3c**), or through nucleophilic attack by the nitrogen lone pair followed by subsequent ylide rearrangement^{5,20} (**Figure 4-3b**). Using DFT calculations to model the reaction between imidazole-ligated IPC and 4-(dimethylsilyl)aniline **6**, the preferred amination pathway was determined to be *N*-nucleophilic attack followed by ylide rearrangement, which agrees with previous mechanistic work on this reaction²⁰. We found that silylation and amination have geometrically distinct transition states, as the required substrate approach to the carbene carbon are different in these reactions, and that in the absence of the protein scaffold, amination has a lower energy barrier than Si–H carbene insertion ($\Delta\Delta G^\ddagger = 3.5 \text{ kcal}\cdot\text{mol}^{-1}$) (**Figures 4-3a–c**).

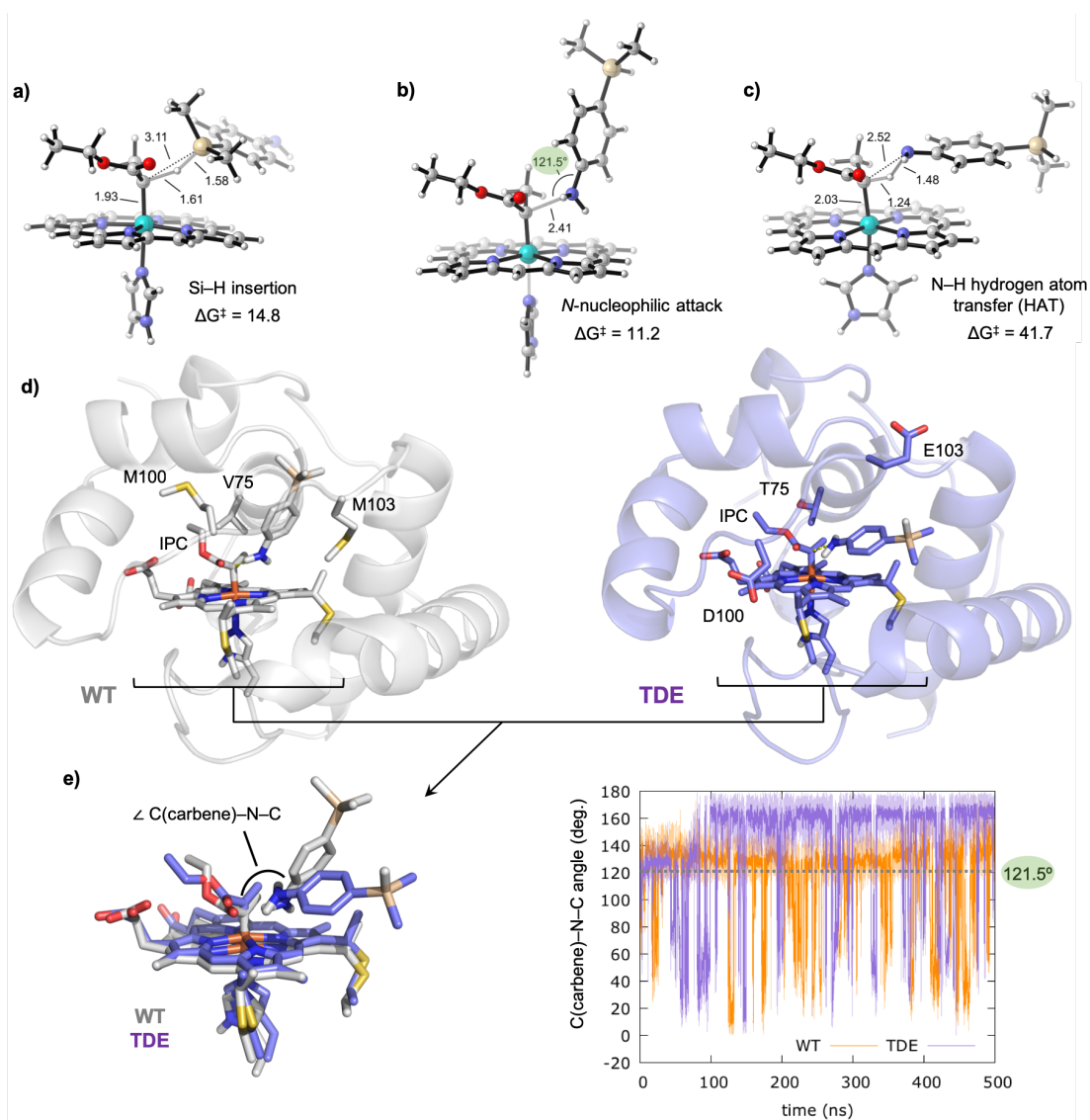


Figure 4-4. DFT optimized transition states using 4-(dimethylsilyl)aniline 6 as substrate: a) Si-H carbene insertion; b) *N*-nucleophilic attack to the electrophilic IPC carbon atom; and c) direct N-H carbene insertion. Gibbs free energies are estimated from lowest in energy OSS IPC 5, and are given in kcal·mol⁻¹. Key distances and angles are given in Å and degrees, respectively. Different orientations of the substrate towards the carbene are required. Analysis of the substrate binding pose and near attack conformations through constrained MD simulations (N(silane) – C(carbene) distance constrained along 500 ns MD trajectories). Representative snapshot of constrained MD simulations (constrained N – C distance) in the d) wild-type and e) TDE IPC *Rma* enzyme active sites. (f) C(carbene)-N(silane)-C(silane) attack angles explored along the constrained MD simulations. The ideal C(carbene)-N(silane)-C(silane) attack angle value for an effective *N*-nucleophilic attack is 121.5°, determined from DFT optimized TS.

We hypothesized that the enzyme might achieve chemoselectivity by distinguishing between the geometrically distinct transition states. Using carbene-bound wild-type *Rma* cyt *c* and *Rma* TDE as starting structures, we docked 4-(dimethylsilyl)aniline **6** into both enzyme active sites, with conformation and orientation of **6** resembling how this substrate attacks the carbene in the DFT-optimized *N*-nucleophilic attack transition state (**Figure 4-4b**). The docked structures then served as starting points for MD simulations, in which the N(substrate **6**) – C(carbene) distance was kept constrained during the entire simulation. These constrained MD simulations showed that in wild-type *Rma* cyt *c*, **6** prefers a bound near attack conformation that resembles the *N*-nucleophilic attack TS, in which the measured \angle C(carbene)-N(substrate **6**)-C(substrate **6**) angle is 121.5° (**Figures 4-4d** and **4-4f**). In contrast, in *Rma* TDE, **6** prefers a binding pose that resembles the near attack conformation for the energetically disfavored DFT optimized N–H hydrogen abstraction TS ($\Delta G^\ddagger = 41.7 \text{ kcal}\cdot\text{mol}^{-1}$, **Figure 4-4c**) with an enlarged \angle C(carbene)-N(substrate **6**)-C(substrate **6**) angle (**Figures 4-4e–f**), rather than the corresponding *N*-nucleophilic attack TS. Taken together, DFT calculations and MD simulations suggest that the front loop conformation ensemble in *Rma* TDE prevents amination by disfavoring the *N*-nucleophilic attack transition state while at the same time, facilitating the silylation reaction.

4.3.5 Engineering inverted chemoselectivity in *Rma* TDE

As a mechanistic understanding of *Rma* TDE helped us explain chemoselectivity, we next sought to demonstrate that such understanding can be leveraged to engineer chemoselectivity, allowing these enzymes to be easily tailored for new catalytic applications. For this demonstration, we sought to create an enzyme that favors the amination product of 4-(dimethylsilyl)aniline (**7**) over the competing silylation product (**8**), the opposite to what *Rma* TDE prefers. We hypothesized that by mutating residues that help control the conformational dynamics of the front loop, we could identify a protein that prefers the amination transition state. Toward this end, three amino acid residues in *Rma* TDE were chosen for site-saturation mutagenesis: E103, M99, and N80. While front loop residue E103 has been experimentally shown to affect the chemoselectivity of the enzyme¹, residue M99, which also resides on the front loop, has not, but it does affect front loop

conformation as observed in our MD simulations. N80 is an amino acid residue found on the D helix of *Rma* cyt *c*. Through MD simulations, N80 was found to persistently interact with the front loop through hydrogen bonding interactions with the main chains of M99, M100D, T101, and D102, as well as the D102 side chain. We therefore predict N80 could modulate the conformation and dynamics of the front loop, and thus the reaction chemoselectivity.

Using *Rma* TDE as our parent enzyme, we generated site-saturation libraries at each of the three sites, and screened for chemoselectivity, as measured by products ratios of carbene transfer to an amine substrate (4-isopropylaniline) and a silane substrate (dimethyl(phenyl)silane, **1**) in a single competition reaction. Variants displaying improved selectivity for the amination reaction were then evaluated for their chemoselectivity on 4(dimethylsilyl)-aniline (**6**). The most amination-selective variants from each library were determined to be E103I, M99P and N80F, which produced the amination product as 27–34% of the total product, as compared to 3% for the parent *Rma* TDE (**Figure 4-5a**). When we combined these three mutations in a single protein (*Rma* TD N80F M99P E103I, abbreviated as *Rma* TDFPI), the resulting catalyst produces the amination product as 90% of the total product, suggesting that the front loop now occupies a new and distinct conformation that selectively favors amination over silylation.

A computational model for carbene-bound *Rma* TDFPI was then generated and used to perform MD simulations to understand the effects of the new mutations. Our simulations revealed a significant change in front loop conformation in *Rma* TDFPI compared to *Rma* TDE (**Figure 4-5b**). Notably, the M99P and E103I mutations work synergistically to disfavor the silylation transition state by changing the conformational preference of the front loop (M99P) and blocking the preferred binding pose required for the silane to approach the IPC (E103I) (**Figure 4-5c**). The new active site found in *Rma* TDFPI has a substrate access tunnel between the D helix and the protein front loop that guides substrate **6** to approach the carbene-bound heme from the top, a trajectory that favors a near attack, amination conformation via an *N*-nucleophilic attack transition state (**Figure 4-5c**). This trajectory is also favored by the N80F mutation: because of steric repulsion between the phenylalanine side chain and the protein front loop, the front loop is now distanced from the D helix where N80F resides. Finally, by docking substrate **6** into carbene-bound *Rma*

TDFPI with an amination-like conformation and constraining the N(substrate **6**) – C(carbene) distance (2.5 Å), constrained MD simulations were used to analyze the catalytically competent binding poses of substrate **6** in *Rma* TDFPI. These simulations showed that the binding pose of substrate **6** is modulated by hydrophobic interactions between E103I and the aromatic ring of **6**, which stabilizes the substrate in a near attack conformation that leads to the *N*-nucleophilic attack TS (**Figure 4-5c**).

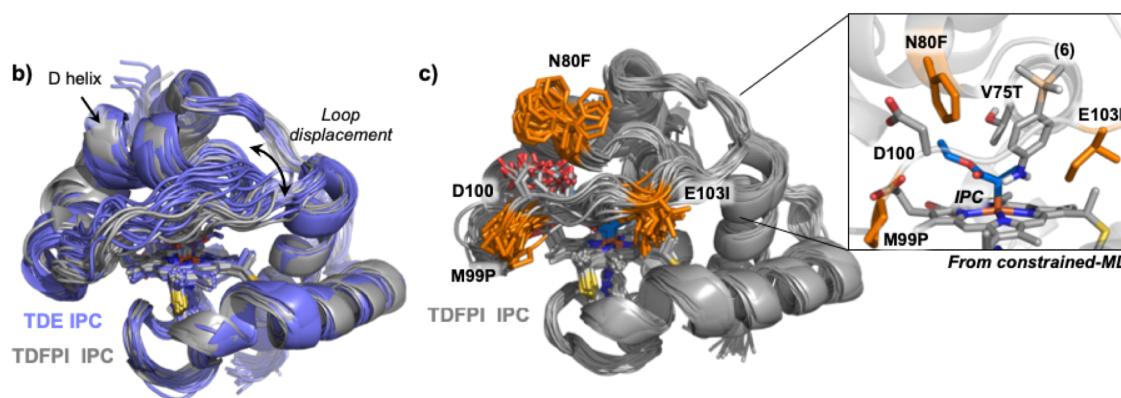
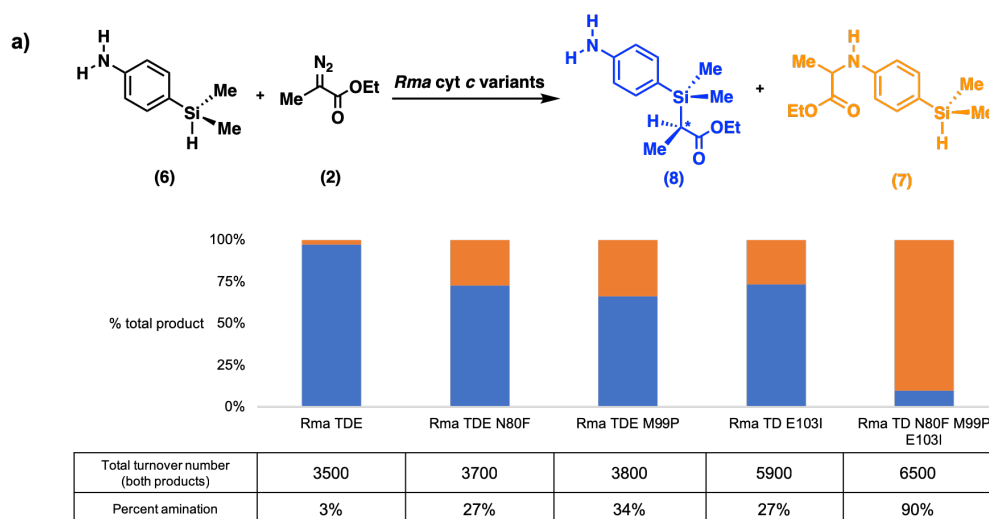


Figure 4-5. a) Site saturation mutagenesis of *Rma* TDE for chemoselective amination over Si–H carbene insertion using 4(dimethylsilyl)-aniline **6** and diazo compound **2**. Reactions were performed with heat-treated cell lysates with 1 μM *Rma* cyt *c* protein, 10 mM **6**, 10 mM **2**, and 10 mM sodium dithionite in M9–N buffer. Chemoselectivity was measured as the percent of amination product over total product formed, and was determined by gas chromatography. Total turnover number (TTN) is defined as the

combined concentration of **7** and **8**, over protein concentration. See SI for full description of the experimental procedures. **b**) Overlay of different snapshots obtained from 500 ns MD trajectories of *Rma* TDE (purple) and *Rma* TDFPI (gray) IPC-bound complexes. The most important structural changes correspond to the preferred “*top-open*” conformation adopted by the front loop (residues 98-103) in *Rma* TDFPI as compared to the parent *Rma* TDE variant. **c**) Overlay of snapshots obtained from 500 ns MD trajectory for *Rma* TDFPI IPC-bound complex. The mutations N80F, M99P, and E103I are highlighted in orange. Inset: Snapshot from constrained-MD simulation of *Rma* TDFPI IPC-bound complex with substrate **6**, showing how substrate **6** prefers to adopt a binding pose that favors amination via *N*-nucleophilic attack.

4.4 Conclusions

The catalytic versatility and selectivity observed in engineered carbene transfer enzymes has led to growing interest in developing these systems for chemical catalysis. Understanding the catalytic mechanism of these engineered enzymes and how the protein scaffold influence the reaction outcomes offers fundamental insights into how non-natural enzymes function and how we might design new ones. Here, we have characterized, through computations and experiments, how *Rma* *cyt c* variants catalyze the formation of carbon–silicon bonds. We determined that the enzyme catalyzes carbene insertion into Si–H bonds through a concerted, nonradical mechanism, in which closed-shell singlet transition states are lowest in energy for both carbene formation and Si–H insertion. Organosilane binding to the enzyme requires unfavorable displacement of the protein front loop, rendering C–Si bond formation the rate-determining step of the overall reaction. We reasoned that directed evolution transformed *Rma* TDE into a faster enzyme by making silane binding more favorable in *Rma* TDE than in the wild-type protein, which reduces the rate-determining energy barrier of carbene Si–H insertion. Importantly, the mutations in *Rma* TDE also confer chemoselectivity to the enzyme by stabilizing a substrate binding pose that favors one reaction (silylation) over another (amination), through modulation of the protein front loop conformation and dynamics. Finally, taking advantage of the mechanistic details uncovered in this study, we predictably switch the enzyme

chemoselectivity from silylation to amination, demonstrating how insights into protein conformational dynamics and reaction mechanism can guide the evolution of new enzymatic reactions, offering a framework for future mechanism-guided engineering of useful and novel carbene transfer enzymes.

4.5 References

1. Kan, S. B. J., Lewis, R. D., Chen, K. & Arnold, F. H. Directed evolution of cytochrome c for carbon–silicon bond formation: Bringing silicon to life. *Science* **354**, 1048–1051 (2016).
2. Kan, S. B. J., Huang, X., Gumulya, Y., Chen, K. & Arnold, F. H. Genetically programmed chiral organoborane synthesis. *Nature* **552**, 132–136 (2017).
3. Tyagi, V., Bonn, R. B. & Fasan, R. Intermolecular carbene S–H insertion catalysed by engineered myoglobin-based catalysts. *Chem. Sci.* **6**, 2488–2494 (2015).
4. Sreenilayam, G. & Fasan, R. Myoglobin-catalyzed intermolecular carbene N–H insertion with arylamine substrates. *Chem. Commun.* **51**, 1532–1534 (2015).
5. Wang, Z. J., Peck, N. E., Renata, H. & Arnold, F. H. Cytochrome P450-catalyzed insertion of carbenoids into N–H bonds. *Chem. Sci.* **5**, 598–601 (2014).
6. Chen, K., Huang, X., Jennifer Kan, S. B., Zhang, R. K. & Arnold, F. H. Enzymatic construction of highly strained carbocycles. *Science*. **360**, 71–75 (2018).
7. Zhang, R. K. *et al.* Enzymatic assembly of carbon–carbon bonds via iron-catalysed sp³ C–H functionalization. *Nature* **565**, 67–72 (2019).
8. Moore, E. J., Steck, V., Bajaj, P. & Fasan, R. Chemoselective Cyclopropanation over Carbene Y–H Insertion Catalyzed by an Engineered Carbene Transferase. *J. Org. Chem.* **83**, 7480–7490 (2018).
9. 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* **534**, 534–537 (2016).
10. 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.* **139**, 1750–1753 (2017).
11. Huang, X. *et al.* A Biocatalytic Platform for Synthesis of Chiral α -Trifluoromethylated Organoborons. *ACS Cent. Sci.* **5**, 270–276 (2019).

12. Jeschek, M., Panke, S. & Ward, T. R. Artificial Metalloenzymes on the Verge of New-to-Nature Metabolism. *Trends in Biotechnology* **36**, 60–72 (2018).
13. Padwa, A. *et al.* Control of Chemoselectivity in Catalytic Carbenoid Reactions. Dirhodium(II) Ligand Effects on Relative Reactivities. *J. Am. Chem. Soc.* **114**, 1874–1876 (1992).
14. Padwa, A. *et al.* Ligand Effects on Dirhodium(II) Carbene Reactivities. Highly Effective Switching between Competitive Carbenoid Transformations. *J. Am. Chem. Soc.* **115**, 8669–8680 (1993).
15. Keipour, H. & Ollevier, T. Iron-Catalyzed Carbene Insertion Reactions of α -Diazoesters into Si-H Bonds. *Org. Lett.* **19**, 5736–5739 (2017).
16. Hayashi, T. *et al.* Capture and characterization of a reactive haem–carbenoid complex in an artificial metalloenzyme. *Nat. Catal.* **1**, 578–584 (2018).
17. Khade, R. L. *et al.* Iron Porphyrin Carbenes as Catalytic Intermediates: Structures, Mössbauer and NMR Spectroscopic Properties, and Bonding. *Angew. Chem. Int. Ed.* **53**, 7574–7578 (2014).
18. Khade, R. L. & Zhang, Y. Catalytic and Biocatalytic Iron Porphyrin Carbene Formation: Effects of Binding Mode, Carbene Substituent, Porphyrin Substituent, and Protein Axial Ligand. *J. Am. Chem. Soc.* **137**, 7560–7563 (2015).
19. Lewis, R. D. *et al.* Catalytic iron-carbene intermediate revealed in a cytochrome c carbene transferase. *Proc. Natl. Acad. Sci.* **115**, 7308–7313 (2018).
20. 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.* **138**, 9597–9610 (2016). doi:10.1021/jacs.6b04636
21. Khade, R. L. & Zhang, Y. C-H Insertions by Iron Porphyrin Carbene: Basic Mechanism and Origin of Substrate Selectivity. *Chem. Eur. J.* 17654–17658 (2017). doi:10.1002/chem.201704631
22. Wei, Y., Tinoco, A., Steck, V., Fasan, R. & Zhang, Y. Cyclopropanations via Heme Carbenes: Basic Mechanism and Effects of Carbene Substituent, Protein Axial Ligand, and Porphyrin Substitution. *J. Am. Chem. Soc.* **140**, 1649–1662 (2018). doi:10.1021/jacs.7b09171
23. Dakin, L. A., Ong, P. C., Panek, J. S., Staples, R. J. & Stavropoulos, P. Speciation and Mechanistic Studies of Chiral Copper(I) Schiff Base Precursors Mediating Asymmetric Carbenoid Insertion Reactions of Diazoacetates into the Si–H Bond of Silanes. *Organometallics* **19**, 2896–2908 (2000).
24. Chen, D., Zhu, D. X. & Xu, M. H. Rhodium(I)-Catalyzed Highly Enantioselective

- Insertion of Carbenoid into Si-H: Efficient Access to Functional Chiral Silanes. *J. Am. Chem. Soc.* **138**, 1498–1501 (2016).
25. Yasutomi, Y., Suematsu, H. & Katsuki, T. Iridium(III)-catalyzed enantioselective Si-H bond insertion and formation of an enantioenriched silicon center. *J. Am. Chem. Soc.* **132**, 4510–4511 (2010).
26. Murdoch, J. R. What is the rate-limiting step of a multistep reaction? *J. Chem. Educ.* **58**, 32 (1981).

Appendix D

SUPPLEMENTARY INFORMATION FOR CHAPTER 4

D.1 Materials and methods

Unless otherwise noted, all chemicals and reagents for chemical reactions were obtained from commercial suppliers (Acros, Arch Bioscience, Fisher Scientific, Sigma-Aldrich, TCI America, VWR) and used without further purification. Multitron shakers (Infors, Annapolis Junction, MD) were used for cell growth. UV-Vis spectroscopy was performed using a Shimadzu UV-1800 spectrophotometer (Shimadzu, Carlsbad, CA). Sonication was performed using a Qsonica Q500 sonicator. Silica gel chromatography purifications were carried out using AMD Silica Gel 60, 230-400 mesh. ^1H NMR spectra were recorded on a Bruker Prodigy 400 MHz instrument and are internally referenced to the residual solvent peak (chloroform). Data for ^1H NMR are reported in the conventional form: chemical shift (δ ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, hept = heptet, m = multiplet, br = broad, app = appears as), coupling constant (Hz), integration. Gas chromatography (GC) analyses were carried out using an Agilent 7820A gas chromatograph, a flame ionization detector (FID), and J&W HP-5 (30 m x 0.32 mm, 0.25 μm film; 90 $^\circ\text{C}$ hold 1 min, 90 to 110 $^\circ\text{C}$ at 15 $^\circ\text{C}/\text{min}$, 110 to 280 $^\circ\text{C}$ at 60 $^\circ\text{C}/\text{min}$, 280 $^\circ\text{C}$ hold 1 min, 6.2 min total).

Plasmid pET22 was used as a cloning vector, and cloning was performed using Gibson assembly.¹ The cytochrome *c* maturation plasmid pEC86² was used as part of a two-plasmid system to express prokaryotic cytochrome *c* proteins. Cells were grown using Lysogeny Broth medium or HyperBroth (AthenaES, Baltimore, MD) with 100 $\mu\text{g}/\text{mL}$ ampicillin and 20 $\mu\text{g}/\text{mL}$ chloramphenicol ($\text{LB}_{\text{amp/chor}}$ or $\text{HB}_{\text{amp/chor}}$). Primer sequences are available upon request. Electrocompetent *Escherichia coli* cells were prepared following the protocol of Sambrook *et al.*³ T5 exonuclease, Phusion polymerase, and *Taq* ligase were purchased from New England Biolabs (NEB, Ipswich, MA). M9-N minimal medium (abbreviated as M9-N

buffer; pH 7.4) was used as a buffering system unless otherwise specified. M9-N minimal medium (abbreviated as M9-N buffer; pH 7.4) was used without a carbon source; it contains 47.7 mM Na₂HPO₄, 22.0 mM KH₂PO₄, 8.6 mM NaCl, 2.0 mM MgSO₄, and 0.1 mM CaCl₂.

D.2 Plasmid construction

All variants described in this paper were cloned and expressed using the pET22(b)+ vector (MilliporeSigma, St. Louis, MO). The gene encoding *Rhodothermus marinus* *cyt c* V75T M100D M103E (*Rma* TDE) was obtained as a single gBlock (Integrated DNA Technologies, Coralville, IA),⁴ codon-optimized for *E. coli*, and cloned using Gibson assembly¹ into pET22(b)+ between restriction sites *Nde*I and *Xho*I in frame. The gene encoding *Rma* TDE contained an *N*-terminal pelB leader sequence and 6xHisTag (MKYLLPTAAAGLLLLAAQPAMAHHHHHH) and had the first seven amino acids of mature, wild-type *Rma* *cyt c* removed (TESGTAA). This plasmid was co-transformed with the cytochrome *c* maturation plasmid pEC86² into *E. coli*[®] EXPRESS BL21(DE3) cells (Lucigen, Middleton, WI).

DNA coding sequence of *Rma* *cyt c* V75T M100D M103E with an *N*-terminal pelB leader sequence and 6xHisTag:

```
ATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCAGCCGGC
GATGGCCCATCATCATCACCACCAAGACCCGGAAGCACTGGCAGCGGAAATTGGTC
CGGTCAAACAGGTGAGCCTGGGTGAACAGATTGATGCGGCCCTGGCGCAACAGGGAGAA
CAGCTCTTCAACACGTATTGTA CTGCGTGCCACCGTCTGGATGAGCGTTTTATCGGACC
GGCCCTGCGCGATGTTACCAAACGTCGTGGGCCGGTTTACATCATGAACACGATGCTGA
ACCCGAATGGGATGATCCAGCGTCATCCGGTGATGAAACAGCTCGTGCAGGAATATGGG
ACCATGGATACCGATGAGGCCCTGAGTGAAGAACAAGCGCGCGCAATTCTGGAGTATCT
GCGCCAGGTTGCGGAAAACCAGTAATGA
```

Amino acid sequence of *Rma* *cyt c* V75T M100D M103E with an *N*-terminal pelB leader sequence and 6xHisTag:

```
MKYLLPTAAAGLLLLAAQPAMAHHHHHHQPPEALAAEIGPVKQVSLGEQIDAALAQQGE
QLFNTYCTACHRLDERFIGPALRDVTKRRGPVYIMNTMLNPNNGMIQRHPVMKQLVQVEYG
TMDTDEALSEEQARAILEYLRQVAENQ*
```

DNA coding sequence of *Rma cyt c* V75T N80F M99P M100D M103I with an *N*-terminal pelB leader sequence and 6xHisTag:

```
ATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCAGCCGGC
GATGGCCATCATCATCATCACCACCAAGACCCGGAAGCACTGGCAGCGGAAATTGGTC
CGGTCAAACAGGTGAGCCTGGGTGAACAGATTGATGCGGCCCTGGCGCAACAGGGAGAA
CAGCTCTTCAACACGTATTGTACTGCGTGCCACCGTCTGGATGAGCGTTTTATCGGACC
GGCCCTGCGCGATGTTACCAAACGTCGTGGGCCGGTTTACATCATGAACACGATGCTGA
ACCCGTTTGGGATGATCCAGCGTCATCCGGTGATGAAACAGCTCGTGCAGGAATATGGG
ACCATGGATCCGGATATTGCCCTGAGTGAAGAACAAGCGCGCGCAATTCTGGAGTATCT
GCGCCAGGTTGCGGAAAACCAGTAATGA
```

Amino acid sequence of *Rma cyt c* V75T N80F M99P M100D M103I with an *N*-terminal pelB leader sequence and 6xHisTag:

```
MKYLLPTAAAGLLLLLAAQPAMAHHHHHHQQDPEALAAEIGPVKQVSLGEQIDAALAQQGE
QLFNTYCTACHRLDERFIGPALRDVTKRRGPVYIMNTMLNPFMIQRHPVMKQLVQVEYG
TMDPDIALSEEQARAILEYLRQVAENQ*
```

D.3 Cytochrome *c* expression and purification

Purified cytochrome *c* proteins were prepared as follows. In a 4 L flask, one liter HyperBroth (AthenaES) containing 100 µg/mL ampicillin and 20 µg/mL chloramphenicol (HB_{amp/chlor}) was inoculated with an overnight culture (20 mL, Lysogeny Broth, with 100 µg/mL ampicillin, 20 µg/mL chloramphenicol, LB_{amp/chlor}) of recombinant *E. cloni*[®] EXPRESS BL21(DE3) cells containing a pET22(b)+ plasmid encoding the cytochrome *c* variant, and the pEC86 plasmid. The culture was shaken at 37 °C and 200 rpm (no humidity control) until the OD₆₀₀ was 0.7 (approximately 3 hours). The culture was placed on ice for 30 minutes, and isopropyl β-D-1-thiogalactopyranoside (IPTG) and 5-aminolevulinic acid (ALA) were added to final concentrations of 20 µM and 200 µM, respectively.

The incubator temperature was reduced to 25 °C, and the culture was allowed to shake for 20 hours at 200 rpm. Cells were harvested by centrifugation (4 °C, 15 min, 4,000xg), and the cell pellet was stored at -20 °C until further use (at least 24 hours). The cell pellet was resuspended in buffer containing 100 mM NaCl, 20 mM imidazole, and 20 mM sodium phosphate buffer (pH 7.5) and cells were lysed by sonication (4 minutes, 1 second

on, 1 second off, 30% duty cycle; Qsonica Q500 sonicator). Cell lysate was placed in a 75 °C heat bath for 10 minutes, and afterwards cell debris was removed by centrifugation for 20 min (30,000xg, 4 °C). Supernatant was sterile-filtered through a 0.45 µm cellulose acetate filter and purified using a 1 mL Ni-NTA column (HisTrap HP, GE Healthcare, Piscataway, NJ) using an AKTA purifier FPLC system (GE Healthcare). The cytochrome *c* protein was eluted from the column by running a gradient from 20 to 500 mM imidazole over 10 column volumes.

The purity of the collected cytochrome *c* fractions was analyzed using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Pure fractions were pooled and concentrated using a 3 kDa molecular weight cut-off centrifugal filter. Protein was dialyzed overnight into 20 mM sodium phosphate buffer (pH 7.5) using 3.5 kDa molecular weight cut-off dialysis tubing. The dialyzed protein was further concentrated using a 3 kDa molecular weight cut-off centrifugal filter, flash-frozen on dry ice, and stored at -80 °C. The concentration of cytochrome *c* was determined in triplicate using the hemochrome assay described in section **D.7**.

D.4 Hemochrome assay

A solution of sodium dithionite (100 mM) was prepared in M9-N buffer. Separately, a solution of 1 M NaOH (0.4 mL) was mixed with pyridine (1 mL), followed by centrifugation (10,000xg, 30 seconds) to separate the excess aqueous layer gave a pyridine-NaOH solution. To a cuvette containing 10 µL purified protein solution and 640 µL M9-N buffer, 100 µL of dithionite solution and 250 µL pyridine-NaOH solution were added. The cuvette was sealed with Parafilm, and the UV-Vis spectrum was recorded immediately at room temperature between 600 and 380 nm on a UV-1800 Shimadzu spectrophotometer. Cytochrome *c* concentration was determined using $\epsilon_{550-535} = 22.1 \text{ mM}^{-1} \text{ cm}^{-1}$.^{4,5}

D.5 Library construction

Cytochrome *c* site-saturation mutagenesis libraries were generated using a modified version of the 22-codon site-saturation method.⁶ For each site-saturation library, oligonucleotides were ordered such that the coding strand contained the degenerate codon NDT, VHG or TGG. The reverse complements of these primers were also ordered. The three forward primers were mixed together in a 12:9:1 ratio, (NDT:VHG:TGG) and the three reverse primers were mixed similarly. Two PCRs were performed, pairing the mixture of forward primers with a pET22(b)+ internal reverse primer, and the mixture of reverse primers with a pET22b internal forward primer. The two PCR products were gel purified, ligated together using Gibson assembly,¹ and transformed into *E. coli*[®] EXPRESS BL21(DE3) cells. Primer sequences are available upon request.

D.6 Enzyme library screening

Single colonies were picked with toothpicks off of LB_{amp/chor} agar plates, and grown in deep-well (2 mL) 96-well plates containing LB_{amp/chor} (400 μ L) at 37 °C, 250 rpm shaking, and 80% relative humidity overnight. After 16 hours, 30 μ L aliquots of these overnight cultures were transferred to deep-well 96-well plates containing HB_{amp/chor} (1 mL) using a 12-channel EDP3-Plus 5-50 μ L pipette (Rainin). Glycerol stocks of the libraries were prepared by mixing cells in LB_{amp/chor} (100 μ L) with 50% v/v glycerol (100 μ L). Glycerol stocks were stored at -78 °C in 96-well microplates. Growth plates were allowed to shake for 3 hours at 37 °C, 250 rpm shaking, and 80% relative humidity. The plates were then placed on ice for 30 min. Cultures were induced by adding 10 μ L of a solution, prepared in sterile deionized water, containing 2 mM IPTG and 20 mM ALA. The incubator temperature was reduced to 24 °C, and the induced cultures were allowed to shake for 20 hours (250 rpm, no humidity control). Cells were pelleted (4,000xg, 5 min, 4 °C) and resuspended in 450 μ L M9-N buffer. For cell lysis, plates were placed in a 75 °C water bath for 10 min, followed by centrifugation (4,000xg, 5 min, 4 °C) to remove cell debris. The resulting heat-treated lysates (320 μ L) were then transferred to deep-well plates for

biocatalytic reactions. In an anaerobic chamber, to deep-well plates of heat-treated lysates were added 40 μL $\text{Na}_2\text{S}_2\text{O}_4$ (100 mM in dH_2O), and 40 μL reactant solution (100 mM phenyl(dimethyl)silane, 100 mM 4-isopropylaniline, 150 mM Me-EDA in MeCN). The plates were sealed with aluminum sealing tape, removed from the anaerobic chamber, and shaken at 400 rpm for 1.5 h. Reactions were quenched with acetonitrile (400 μL), and allowed to sit at room temperature for 30 minutes to allow proteins to precipitate from solution. The plates were centrifuged (5,000 \times g, 10 min) and the organic layer (400 μL) was transferred to shallow-well 96-well plates for reverse phase HPLC analysis. Protein variants that displayed improved selectivity for the 4-isopropylaniline substrate were then regrown in shake flasks, and evaluated for their activity and selectivity with 4-dimethylsilylaniline in small-scale biocatalytic reactions, which were analyzed by GC for accurate determination of turnovers and chemoselectivity.

D.7 Protein lysate preparation

Protein lysates for biocatalytic reactions were prepared as follow: *E. coli* cells expressing *Rma* cyt *c* variant were pelleted (4,000 \times g, 5 min, 4 $^\circ\text{C}$), resuspended in M9-N buffer and adjusted to the appropriate optical density at 600 nm (OD_{600}). The whole-cell solution was sonicated, (1 minute, 1 second on, 1 second off, 30% duty cycle) heat-treated (75 $^\circ\text{C}$ for 10 min) and then centrifuged (14,000 \times g, 10 min, 4 $^\circ\text{C}$) to remove cell debris. The supernatant was sterile filtered through a 0.45 μm cellulose acetate filter, concentration of cytochrome *c* protein in the lysate was determined using the hemochrome assay described in section D.7. Using this protocol, the protein concentrations we typically observed for $\text{OD}_{600} = 30$ lysates were in the 3-15 μM range.

D.8 Biocatalytic reactions

In an anaerobic chamber, heat-treated cell lysate containing an *Rma* cyt *c* variant was diluted to concentration of 1.2 μM with M9-N buffer, and 340 μL of this solution was transferred to a 2 mL crimp vial. Next, the protein was reduced by adding 40 μL $\text{Na}_2\text{S}_2\text{O}_4$ (100 mM in M9-N buffer), and the reaction was initiated with the addition of reactants, 10

μL the appropriate silane solution (400 mM in MeCN) and 10 μL Me-EDA (400 mM in MeCN). The vial was crimp sealed, removed from the anaerobic chamber, and shaken at 400 rpm at room temperature for 2 hours. At the end of the reaction, the crimp vial was opened and the reaction was quenched with cyclohexane (1 mL). Internal standard was added (20 μL of 20 mM 1,2,3-trimethoxybenzene in toluene or 20 μL of 20 mM 2-phenylethanol in toluene) and the reaction mixture was transferred to a microcentrifuge tube, vortexed (1 minute), then centrifuged (14,000xg, 5 min) to completely separate the organic and aqueous layers. The organic layer (800 μL) was removed for GC and chiral HPLC analysis. All biocatalytic reactions were performed in triplicate. The total turnover numbers (TTNs) reported are calculated with respect to the protein catalyst and represent the total number of turnovers obtained using the catalyst under the stated reaction conditions.

Concentration of organosilicon product was calculated based on the ratio of areas between the product peak and internal standard peak, as measured by GC-FID.

Stock solutions of chemically synthesized authentic standard organosilicon products were prepared at various concentrations (20 to 200 mM in MeCN). To a microcentrifuge tube were added 340 μL M9-N buffer, 40 μL $\text{Na}_2\text{S}_2\text{O}_4$ (100 mM in M9-N buffer), 20 μL organosilicon product, 20 μL internal standard (20 mM 1,2,3-trimethoxybenzene in toluene or 20 μL of 20 mM 2-phenylethanol in toluene) and 1 mL cyclohexane. The mixture was vortexed (1 minute) then centrifuged (14,000xg, 5 min) to completely separate the organic and aqueous layers. The organic layer (800 μL) was removed for GC analysis. Each concentration was prepared in duplicate, and both data points are shown for each concentration (**Figure D-3**). The standard curves plot product concentration in mM (x-axis) against the ratio of product area to internal standard area on the GC (y-axis).

D.8.1 Kinetic time course reactions

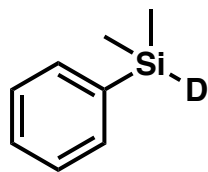
In an anaerobic chamber, purified protein was added to 3400 μL of M9-N buffer in a 6 mL headspace vial with a magnetic stir bar. Next, the protein was reduced by adding 400 μL $\text{Na}_2\text{S}_2\text{O}_4$ (100 mM in M9-N) and the reaction was initiated with the addition of

reactants, 100 μL PhMe_2SiH (400 mM in MeCN) and 100 μL Me-EDA (400 mM in MeCN). At two minute intervals (2, 4, 6, 8, and 10 min) 400 μL of the reaction solution was removed from the 6 mL vial, and was transferred to a microcentrifuge tube containing 1000 μL cyclohexane and an internal standard (20 μL of 20 mM 1,2,3 trimethoxybenzene in toluene). The microcentrifuge tube was vortexed for 30 seconds, and then allowed to sit until all samples from all time points had been produced. At the end of the kinetic time course reaction, the microcentrifuge tubes were centrifuged (14,000 \times g, 1 min) to completely separate the organic and aqueous layers. The organic layer (800 μL) was removed for GC analysis. All biocatalytic reactions were performed in triplicate or quadruplicate. The initial rates reported are calculated with respect to the protein catalyst and represent the number of turnovers per minute per enzyme, obtained using the enzyme under the stated reaction conditions.

D.9 Synthesis and characterization of organosilicon compounds

Phenyl(dimethyl)-silane-*d* (Compound D-1)

Phenyl(dimethyl)-silane-*d* was prepared by adapting a procedure for the hydrogenation of chlorosilanes by NaBH_4 .⁷ In a 100 mL-round bottom flask, 2.5 g sodium borodeuteride (NaBD_4 , 60 mmol, 2.0 equiv., Cambridge Isotope Labs) were added. The flask was purged with argon, and 30 mL anhydrous acetonitrile was added. Under anaerobic conditions, 5.0 g phenyl(dimethyl)silyl chloride (30 mmol, 1 equiv.) was added dropwise over 30 minutes at room temperature, with continuous stirring. The round bottom flask became very hot over the addition of phenyl(dimethyl)silyl chloride, and so slower addition than performed here, and/or addition at 0 $^\circ\text{C}$ is recommended for a safer procedure. After completing the addition of the phenyl(dimethyl)silyl chloride, the reaction was allowed to stir at room temperature for an additional 30 min. The reaction was then cooled to 0 $^\circ\text{C}$, after which the flask was opened to aerobic atmosphere and quenched with dropwise addition of saturated aqueous ammonium chloride (NH_4Cl , approximately 10 mL). Purification by silica column chromatography using 100% pentane as an eluent afforded phenyl(dimethyl)-silane-*d* with 61% yield.



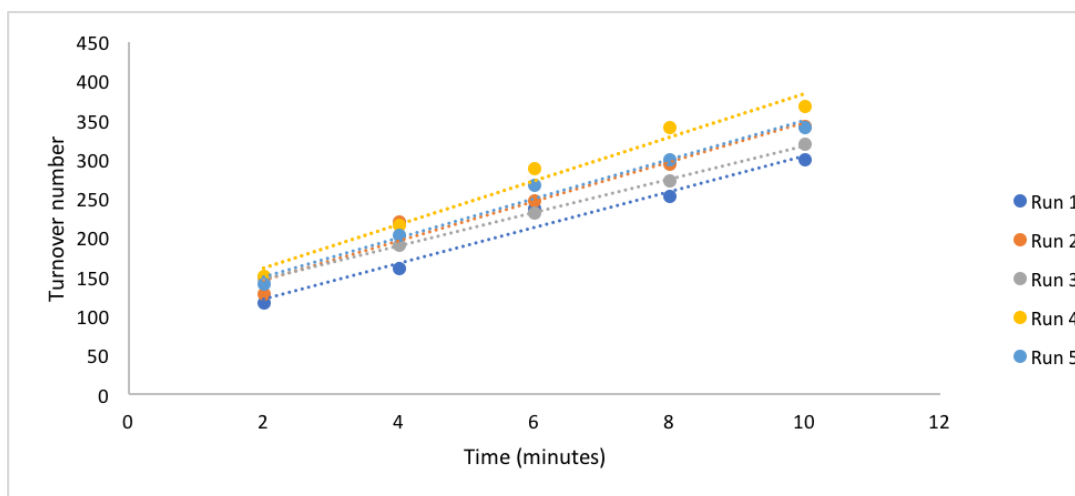
Compound D-1. Phenyl(dimethyl)silane-*d*

^1H NMR (400 MHz, CDCl_3) δ 7.48-7.62 (m, 2H), 7.30-7.44 (m, 3H), 0.35 (app q, $J = 0.6$ Hz, 6H).

All other compounds were synthesized as described in ref. (4).

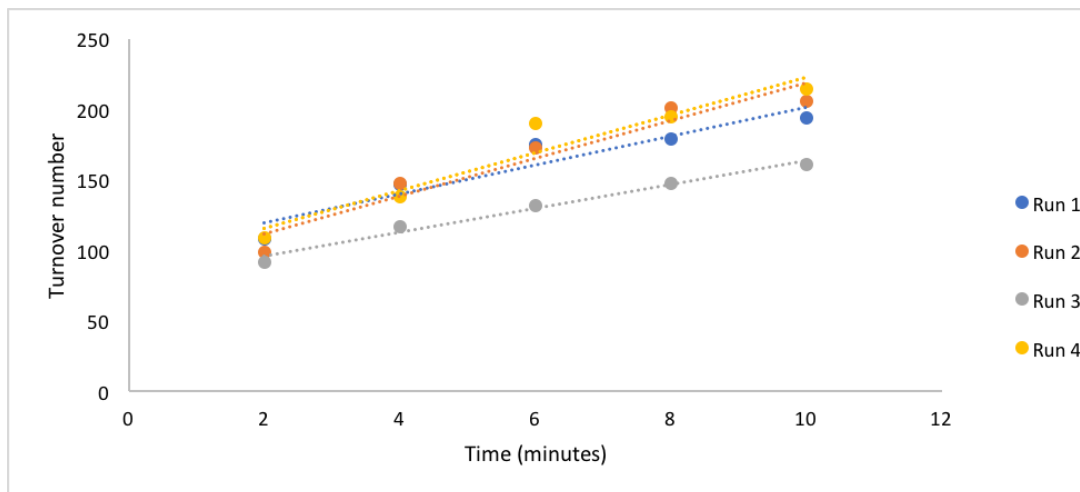
D.10 Supporting figures and tables

Figure D-1. Kinetic time course data for reaction between ethyl 2-diazopropanoate and phenyl(dimethyl)-silane (PhMe₂Si-H or PhMe₂Si-D, as specified), catalyzed by *Rma* cytochrome *c* variants.



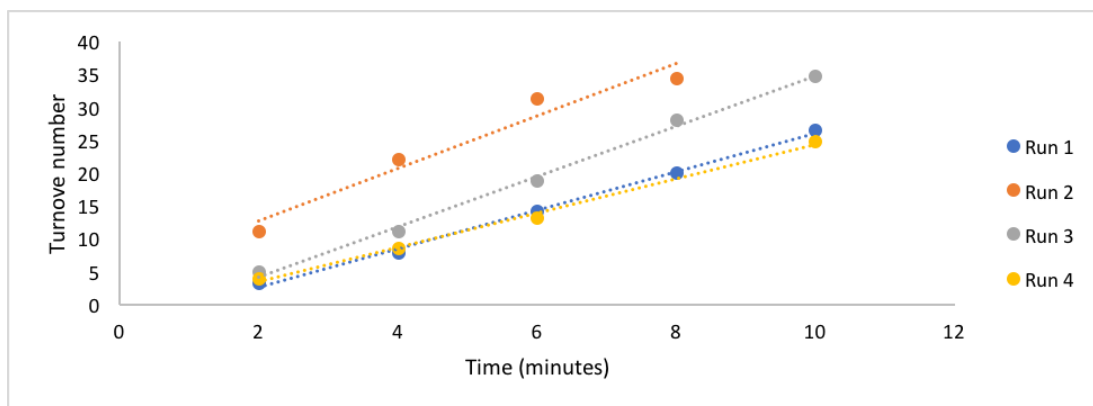
A. Kinetic time course data for reaction using *Rma* TDE and PhMe₂Si-H. Five independent runs are shown, and the corresponding turnover frequencies are summarized in the table below.

<i>Rma</i> TDE (PhMe ₂ Si-H)	
Run 1	25.0 /min
Run 2	21.2 /min
Run 3	27.9 /min
Run 4	25.0 /min
Run 5	22.9 /min
Average	24.4 /min



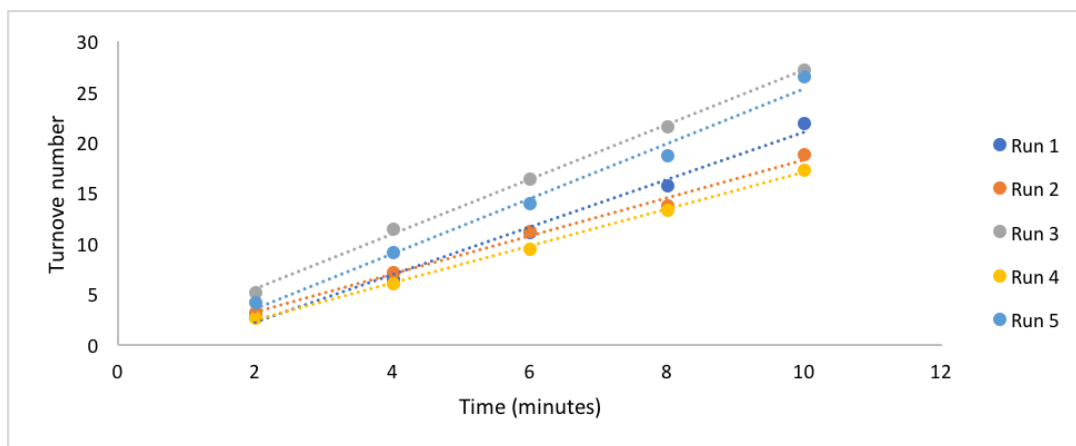
B. Kinetic time course data for reaction using *Rma* TDE and PhMe₂Si–D. Four independent runs are shown, and the corresponding turnover frequencies are summarized in the table below.

<i>Rma</i> TDE (PhMe ₂ Si–D)	
Run 1	10.2 /min
Run 2	13.4 /min
Run 3	8.4 /min
Run 4	13.0 /min
Average	11.3 /min



C. Kinetic time course data for reaction using wild-type *Rma* cyt *c* and PhMe₂Si-H. Four independent runs are shown, and the corresponding turnover frequencies are summarized in the table below.

Wild-type <i>Rma</i> cyt <i>c</i> (PhMe ₂ Si-H)	
Run 1	4.0 /min
Run 2	3.8 /min
Run 3	2.6 /min
Run 4	2.9 /min
Average	3.3 /min



D. Kinetic time course data for reaction using wild-type *Rma* cyt *c* and PhMe₂Si-H. Five independent runs are shown, and the corresponding turnover frequencies are summarized in the table below.

Wild-type <i>Rma</i> cyt <i>c</i> (PhMe ₂ Si-D)	
Run 1	2.3 /min
Run 2	1.9 /min
Run 3	2.7 /min
Run 4	1.8 /min
Run 5	2.7 /min
Average	2.3 /min

Figure D-2. ^1H NMR for phenyl(dimethyl)silane and phenyl(dimethyl)silane-*d*.
 ^1H NMR (400 MHz, CDCl_3)

Top spectrum: δ 7.48-7.62 (m, 2H), 7.30-7.44 (m, 3H), 0.35 (app q, $J = 0.6$ Hz, 6H).
Bottom spectrum: δ 7.48-7.62 (m, 2H), 7.30-7.44 (m, 3H), 4.36-4.50 (m, 1H), 0.28-0.42 (m, 6H).
The absence of peak C (4.36-4.50 ppm) in the top spectrum indicates that the Si-H bond has been replaced by a Si-D bond.

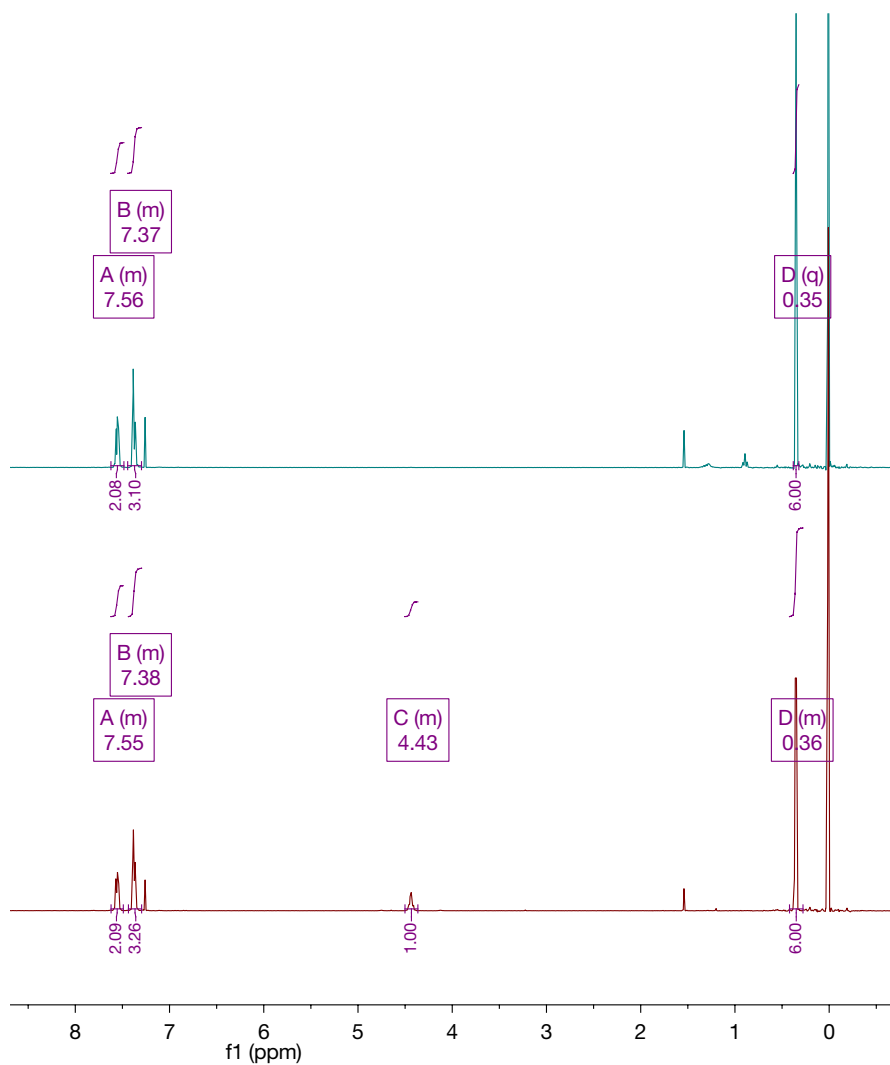
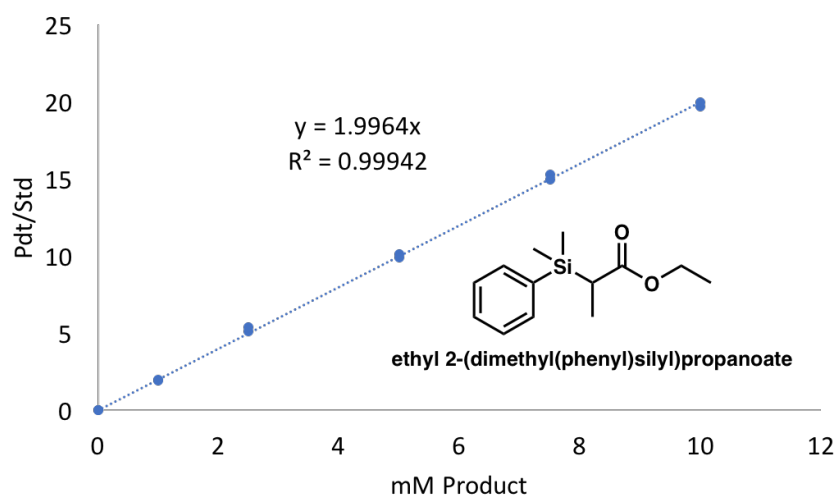


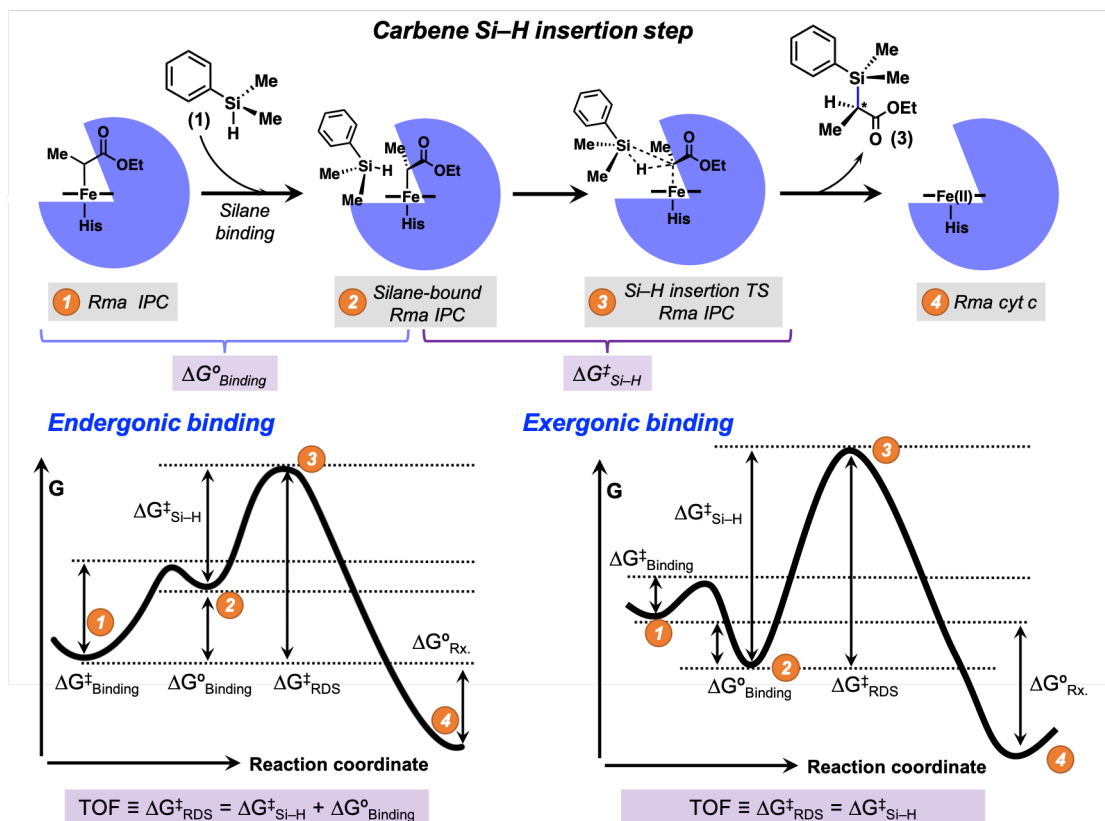
Figure D-3. GC Standard curve for ethyl 2-(dimethylphenylsilyl)propanoate product.

The analysis of product formation in enzymatic reactions was performed based on GC standard curves. The standard curve use to quantify the kinetic time course reactions is shown below; all other standard curves were prepared and used as described in reference (4).

Stock solutions of chemically synthesized organosilicon products at various concentrations (20 to 200 mM in MeCN) were prepared. To a microcentrifuge tube were added 340 μL M9-N buffer, 40 μL $\text{Na}_2\text{S}_2\text{O}_4$ (100 mM in dH_2O), 20 μL organosilicon product, 20 μL internal standard (20 mM of 1,2,3 trimethoxybenzene in toluene) and 1 mL cyclohexane. The mixture was vortexed (10 seconds, 3 times) then centrifuged (14,000 \times g, 5 min) to completely separate the organic and aqueous layers. The organic layer (750 μL) was removed for GC analysis. All data points represent the average of duplicate runs. The standard curves plot product concentration in mM (x-axis) against the ratio of product area to internal standard area on the GC (y-axis).



Scheme D-1. Schematic representation of the Si–H insertion enzymatic rate-determining step (RDS) profile and how silane binding might affect to the enzyme-catalyzed RDS barrier. The enzyme-catalyzed RDS barrier consists of two components: *i*) the silane substrate binding affinity ($\Delta G^{\circ}_{\text{Binding}}$); and *ii*) the Si–H insertion energy barrier ($\Delta G^{\ddagger}_{\text{Si-H}}$), assuming a previously formed silane-enzyme bound complex. In the case of an endergonic silane binding, the rate-determining step observed depends on both the endergonic silane-enzyme complex formation ($\Delta G^{\circ}_{\text{Binding}} > 0$) and the carbene Si–H insertion energy barrier ($\Delta G^{\ddagger}_{\text{Si-H}}$). On the other hand, an exergonic silane binding ($\Delta G^{\circ}_{\text{Binding}} < 0$) would result into an RDS step that only depends on the carbene Si–H insertion energy barrier ($\Delta G^{\ddagger}_{\text{Si-H}}$).



D.11 References

1. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343 (2009).
2. Arslan, E., Schulz, H., Zufferey, R., Kunzler, P. & Thony-Meyer, L. Overproduction of the Bradyrhizobium japonicum c-type cytochrome subunits of the cbb(3) oxidase in Escherichia coli. *Biochem. Biophys. Res. Commun.* **251**, 744–747 (1998).
3. Sambrook, J. & W Russell, D. Molecular Cloning: A Laboratory Manual. Cold Spring Harb. Lab. Press. *Cold Spring Harb. NY* **999** (2001). doi:10.1016/0092-8674(90)90210-6
4. Kan, S. B. J.; Lewis, R. D.; Chen, K.; Arnold, F. H., Directed evolution of cytochrome c for carbon–silicon bond formation: Bringing silicon to life. *Science* **354**, 1048–1051 (2016).
5. Berry, E. A.; Trumpower, B. L., Simultaneous determination of hemes a, b, and c from pyridine hemochrome spectra. *Anal. Biochem.* **161**, 1–15 (1987).
6. 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.* **2**, 83–92 (2013).
7. Ito, M.; Itazaki, M.; Abe, T.; Nakazawa, H., Hydrogenation of Chlorosilanes by NaBH₄. *Chem. Lett.* **45**, 1434–1436 (2016).

*Chapter 5*MECHANISM BASED INACTIVATION IN HEME-DEPENDENT
CARBENE TRANSFERASES

Material from this chapter appears in “Renata H[†], Lewis RD[†], Sweredoski MJ, Moradian A, Hess S, Wang ZJ, Arnold FH. (2016) **Identification of mechanism-based inactivation in P450-catalyzed cyclopropanation facilitates engineering of improved enzymes.** *JACS*, 138:12527-12533. doi:10.1021/jacs.6b06823”. Reproduced with permission from the American Chemical Society. [†]Denotes equal author contribution.

5.1 Abstract

Following the recent discovery that heme proteins can catalyze carbene transfer reactions with high efficiency and selectivity, interest in engineering these proteins for use in chemical synthesis has burgeoned. Many of these “carbene transferases” are derived from cytochrome P450 enzymes, but natural P450 enzymes are subject to mechanism-based inhibition by carbene precursors. Here, we present evidence for two inactivation pathways that are operative in engineered carbene transferases. Using a combination of UV-vis, mass spectrometry, and proteomic analyses, we show that the heme cofactor and several nucleophilic side chains undergo covalent modification by the carbene precursor ethyl diazoacetate (EDA). Substitution of two of the affected residues with less-nucleophilic amino acids led to a more than twofold improvement in cyclopropanation performance (total turnover number, TTN). Substitution of the damaged heme cofactor after the reaction results in modest improvements in TTN, indicating that both pathways are involved in enzyme inactivation. Elucidating the inactivation pathways of heme protein-based carbene transfer catalysts should aid in the optimization of this new biocatalytic function by providing information to guide engineering efforts.

5.2 Introduction

Over the past few years, we¹⁻⁶ and others⁷⁻¹¹ have reported that engineered P450 enzymes and other heme proteins^{4,5,7,11-13} can catalyze non-natural carbene transfer reactions, e.g., olefin cyclopropanation (**Figure 5-1A**), with high efficiency and selectivity. The proposed catalytic cycle initiates by reduction of the Fe³⁺ heme enzyme to its Fe²⁺ counterpart, which then reacts with a diazo molecule to form an iron carbene intermediate¹⁴. Transfer of the carbene to a second substrate forms product and regenerates the active Fe²⁺ enzyme. Although a myriad of carbene transfer biocatalysts have been engineered from cytochrome P450s^{1,2,6,15}, carbenes are implicated in the mechanism of suicide inactivation observed with natural P450s in the presence of halothane¹⁶⁻¹⁸ and CCl₄¹⁹⁻²¹. Similarly, rat liver P450s are known to be inactivated by the decomposition of 3- [[2-(2,4,6-trimethylphenyl)thio]ethyl]-4-methylsydnone, which is believed to occur via formation an iron carbene intermediate that reacts with the heme cofactor to form N-vinylprotoporphyrin IX²².

Because understanding inactivation pathways can facilitate development of more stable and efficient enzymes²³⁻²⁵, we investigated whether P450-derived carbene transferases are similarly inactivated by undesired reactions of the iron carbene intermediate. Our goal is to understand the mechanism of inactivation and delay or circumvent it.

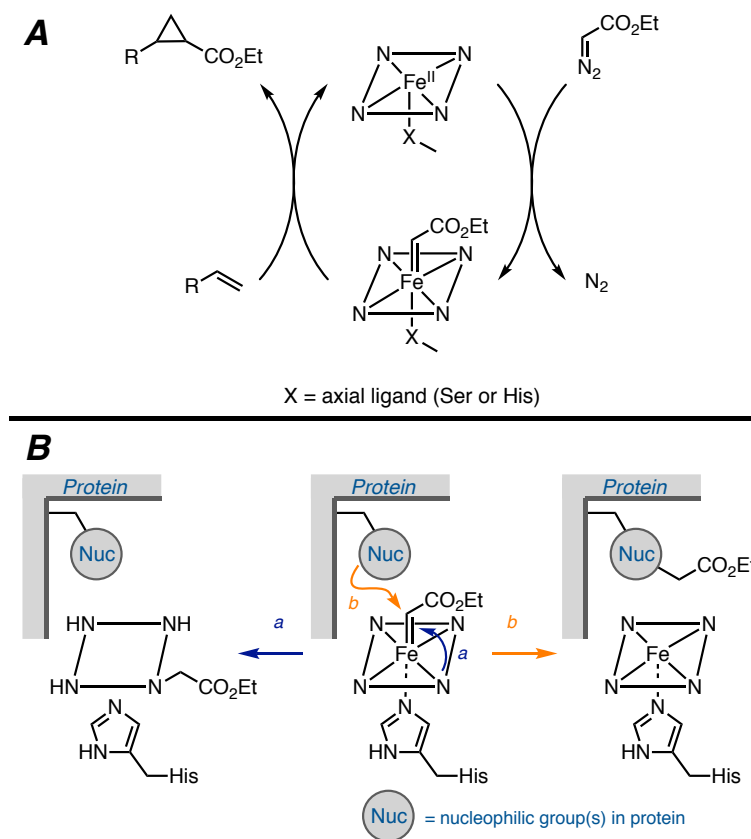


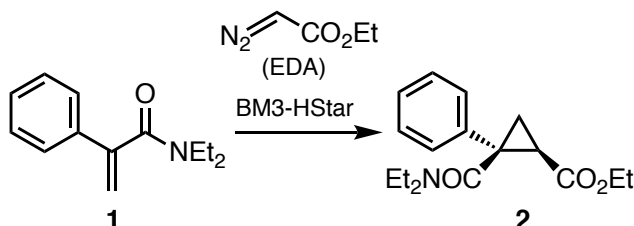
Figure 5-1. On-cycle and off-cycle catalytic pathways for carbene transfer **A.** Proposed catalytic cycle for heme protein-catalyzed carbene transfer to an olefin starts from the reduced ferrous state, which undergoes reaction with a diazo compound to form the putative iron carbene. The carbene reacts with an olefin, forming the cyclopropane product and regenerating the ferrous heme protein. **B.** Proposed mechanism-based inactivation via (a) porphyrin and (b) side chain modification involving the same iron carbene intermediate that mediates productive olefin cyclopropanation.

5.3 Results

5.3.1 EDA-induced inactivation of P450 BM3-HStar

We previously discovered that substitution of the proximal cysteine ligand with histidine in cytochrome P450-BM3 led to a highly active cyclopropanation catalyst³. This finding enabled the engineering of a histidine-ligated P450 (BM3-HStar: P450-BM3 V78M L181V T268A C400H L437W) that catalyzes the cyclopropanation of phenylacrylamide **1** in a highly stereoselective fashion (**Scheme 1**) en route to the synthesis of levomilnacipran, a

United States Food and Drug Administration (FDA)-approved antidepressant³. As BM3-HStar produced a useful pharmaceutical precursor, we hoped to improve the utility of the enzyme by abrogating any possible enzyme inactivation.



Scheme 5-1. Cyclopropanation reaction catalyzed by BM3-HStar.

We chose to investigate inactivation of the isolated BM3-HStar enzyme in order to avoid possible complications from cellular components. From control experiments, we noted that purified enzyme that was preincubated with ethyl diazoacetate (EDA) in the absence of $\text{Na}_2\text{S}_2\text{O}_4$ retained virtually all of its catalytic activity, whereas preincubation with EDA and reductant led to rapid inactivation (Appendix E, **Figure E-1**). The fact that inactivation requires reductant suggests that carbene formation is the first step in the inactivation pathway, as the heme-catalyzed carbene transfer reactions typically require an Fe^{2+} heme^{1-3,26}.

5.3.2 Heme cofactor modification

Previous reports of mechanism-based P450 inactivation indicate that carbene precursors can lead to covalent modification of the heme prosthetic group or protein side chain(s)^{16-22,27-33} (**Figure 5-2B**). In particular, Ortiz de Montellano and co-workers observed porphyrin modification in the reaction of P450 and model Fe-porphyrin complexes with 3-[[2-(2,4,6-trimethylphenyl)thio]ethyl]-4-methylsydnone²² or diazoacetophenone³² and proposed an intramolecular carbene insertion reaction to account for their observations. To investigate whether an analogous pathway is operative in our system, we analyzed the porphyrins from the reaction of EDA and **1** using purified BM3-HStar and quenched after 15 s and 1 min.

After extraction from the reaction mixture, the crude organics were analyzed by liquid chromatography–mass spectrometry (LC-MS) using detection at 330 nm to minimize background signal. In agreement with the observed dependence of enzyme inactivation on presence of the reductant, extracts from reaction mixture that did not contain $\text{Na}_2\text{S}_2\text{O}_4$ yielded a single species with a mass spectrum and retention time that correspond to unmodified Fe-protoporphyrin IX (Fe-PPIX) cofactor, as determined by LC-MS analysis of authentic standard (**Figure 5-2**). However, the crude extracts obtained after 15 s of reaction of **1** and EDA using reduced BM3-HStar showed diminished signal for Fe-PPIX and the presence of additional species. Strikingly, after 1 min, only modified Fe-PPIX remained in the reaction mixture. Similar analyses performed with other heme protein-based carbene transfer catalysts, myoglobin (Mb) mutant H64V V68A⁷ and a Ser-ligated P450-BM3², also revealed loss of Fe-PPIX.

Thin-layer chromatography (TLC) analysis of the reaction mixture indicated the presence of a new red-fluorescent band, which was subsequently purified by preparative TLC and analyzed by UV–vis spectroscopy and mass spectrometry. Absorption spectra obtained for the isolated porphyrin species showed that it is distinct from PPIX, both as the free porphyrin (**Figure E-2**) and as the Zn complex (**Figure 5-3**). The absorption spectrum of the modified porphyrin after complexation with Zn^{2+} yielded a Soret peak at 430 nm with a slight shoulder, which is a characteristic feature of N-alkylprotoporphyrin derivatives²². Furthermore, the mass spectrum of the isolated porphyrin exhibited two major peaks with m/z of 735.3374 and 821.3753, corresponding to loss of Fe^{3+} and addition of two and three ethyl acetylidene fragments ($\text{C}_4\text{H}_6\text{O}_2$, $m/z = 86.1$ Da) to the heme cofactor, respectively. Together with previous literature on P450 cofactor modification, our evidence suggests that the heme cofactor is subject to N-alkylation during the carbene transfer reaction^{22,32,34}.

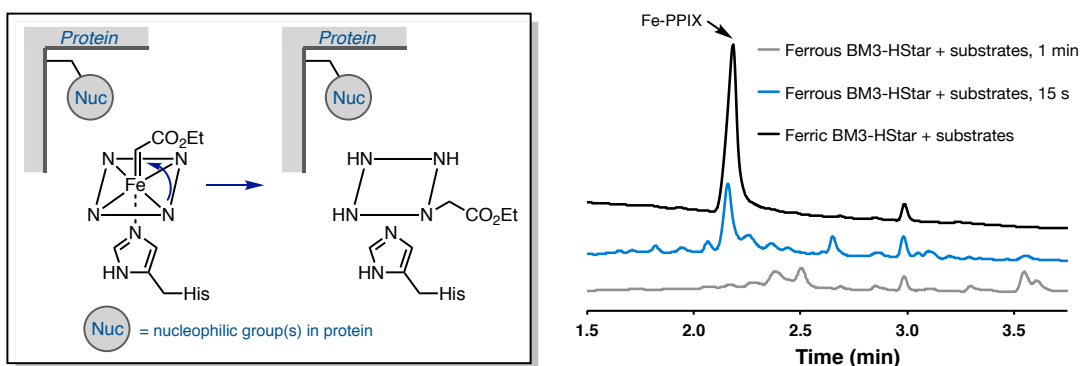


Figure 5-2. LC-MS monitoring of heme content of the reaction mixture (MWD set at 330 nm) at different reaction times: Amount of unmodified heme decreases during the course of the reaction, and new species appear.

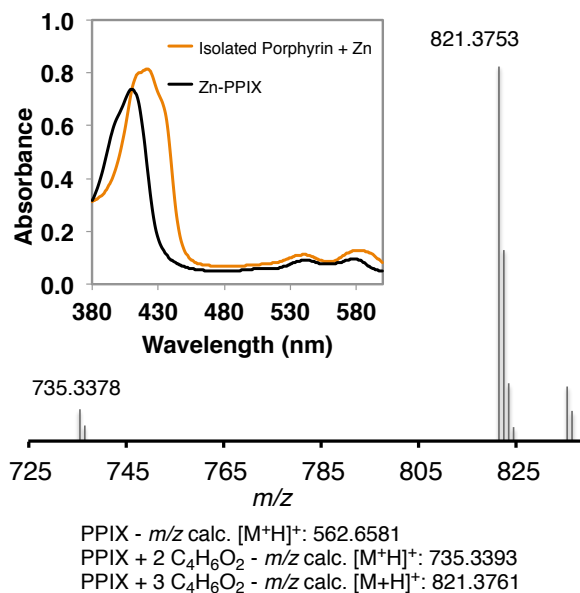


Figure 5-3. High-resolution mass spectrometry (HRMS) analysis of modified porphyrin isolated after the cyclopropanation reaction, showing major peaks with m/z of 735.3378 and 821.3753, which correspond to multiple additions of $C_4H_6O_2$ fragment. **Inset:** Electronic absorption spectrum of porphyrin isolated after the cyclopropanation reaction as the Zn-complexed form indicates modification; spectrum of Zn-PPIX standard is provided for comparison.

5.3.3 Modification of protein side chains and structural changes

Given the precedent of protein side chain alkylation during oxidation of acetylenic compounds²⁹⁻³¹, we investigated whether a similar phenomenon occurs during the enzyme-catalyzed cyclopropanation reaction. To this end, cyclopropanation was performed with purified BM3-HStar, quenched after 15 s, and the crude reaction mixture was submitted to mass spectrometry (MS) analysis. Extreme heterogeneity of the enzyme could already be observed (**Figure E-3**). Deconvolution of the mass spectrum afforded multiple species with molecular weights that correspond to integer number additions of the carbene fragment (86 Da). Under the denaturing conditions used for MS analysis, the heme cofactor is extruded from the protein. Thus, any observed mass difference in the spectra can be attributed solely to modification of the polypeptide chain.

Circular dichroism (CD) spectroscopy of EDA-treated enzyme also showed changes relative to unreacted enzyme in the far-UV region (**Figure 5-4**). We noted reduced molar ellipticity at 222 nm, which suggests decreased α -helical content and increased random coil content of the protein. Taking into account both the MS and CD spectral observations, we propose a model wherein EDA modifies the cofactor and multiple protein side chains, leading to disruption of secondary structures and partial unfolding. Comparison of these data and LC-MS observation of cofactor modification suggest that the side chain and porphyrin alkylations happen on similar time scales, although we cannot determine the exact order of events.

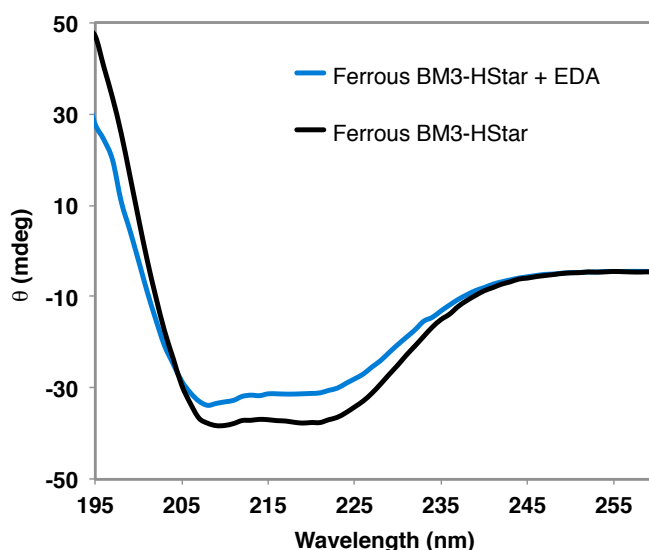


Figure 5-4. Circular dichroism (CD) spectra of ferrous BM3-HStar before (black) and after reaction with EDA (blue). Decrease in molar ellipticity at 222 nm suggests a change in the secondary structure after the reaction, specifically a decrease in α -helical content and an increase in random coil.

5.3.4 Identification of modified residues

To identify the residues that are modified during the initial stages of the reaction, protein from the crude cyclopropanation reaction mixture after 5 s was analyzed by trypsin digestion followed by tandem LC-MS/MS analysis. From this analysis (**Figure 5-5**), we observed modifications (underlined residues indicate the site of modification) on peptide fragments EACDESRFDK (residues 60–69), NLSQALKFMR (residues 70–79), DFAGDGLFTSWTHEK (residues 80–94), and AHNILLPSFSQQAMK (residues 99–113). Modifications were addition of a $C_4H_6O_2$ fragment (86.1 Da), corresponding to the ethyl acetylidene moiety, onto residues C62, S72, K76, R79, H92, and H100. We also observed modification of residue K69 on fragment FDKNLSQALK (residues 67–76) in samples taken at later time points. Control experiments with CO-treated enzyme (inactive for cyclopropanation) revealed much-diminished signals for modified peptides during the proteomic analysis, again suggesting that the iron carbene generated from the ferrous heme is responsible for the modifications. That the affected residues are all nucleophilic in nature

is consistent with the proposed electrophilic character of the carbene species^{14,35–37}.

Modifications presumably occur by way of O-, N-, and S- alkylation by the carbene fragment.

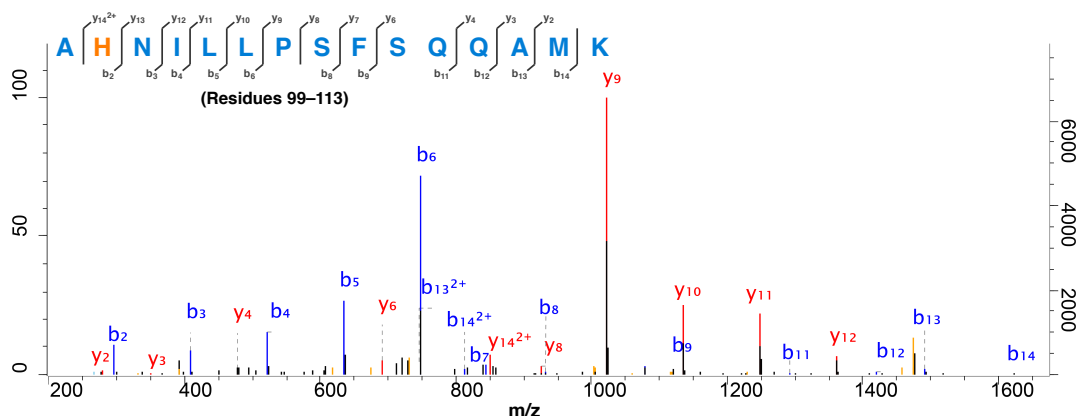


Figure 5-5. Representative LC-MS/MS analysis of peptide fragments from crude cyclopropanation mixture. Modified peptides from inactivated BM3- HStar were analyzed using mass spectrometry after incubation with $\text{Na}_2\text{S}_2\text{O}_4$, EDA and 1 for 5 s. The modified residue is highlighted in orange.

In the absence of a crystal structure for His-ligated P450-BM3, the crystal structure of wild-type (WT) BM3 (PDB ID: 2IJ2) was used to map the locations of these residues relative to the heme cofactor (**Figure 5-6**). To our surprise, other than K69, which makes a polar contact with the propionate moiety of the heme cofactor, all the alkylated residues are more than 10 Å away from the iron center. One possible explanation is that the Hstar active site is devoid of nucleophilic residues, and that all likely alkylation targets are distant. It is worth noting, however, that residues S72, K76, and R79 are situated in the B'-helix and H92 is in the B'/C loop. The B'-helix has been proposed as an important recognition site that controls substrate specificity in P450s³⁸. It has also been shown crystallographically that a single mutation (F81L) in this region can affect its collapse toward the active site³⁹. Moreover, our recent report on the X-ray crystal structure of His-ligated CYP119 shows notable rearrangements in the protein to compensate for non-native ligation, including unresolved

portions in the B'- and F-helices and increased disorder in the C-helix⁴⁰. If similar rearrangements occur upon mutation of the axial residue from Cys to His in P450-BM3, it is possible that the region where S72, K76, R79, and H92 reside undergoes large-scale movements that bring them closer to the active site. Furthermore, the P450s possess remarkable structural plasticity^{41,42}, and iron carbene formation could also induce structural and dynamic changes³⁷ that bring this region in closer contact.

The remaining three alkylated residues also deserve mention, as we surmise that they would play vital roles in maintaining the folded structure of P450-BM3. As previously noted, K69 is involved in a polar contact with the heme prosthetic group (**Figure E-4**). H100 is implicated in a hydrogen bonding interaction with the FMN domain of the fused reductase partner and is conserved as a basic amino acid residue in P450s⁴³. Mutation of C62 in P450-BM3 has been reported to lower thermostability⁴⁴. Thus, modifications of these residues could result in disruption of key interactions and loss of structural integrity.

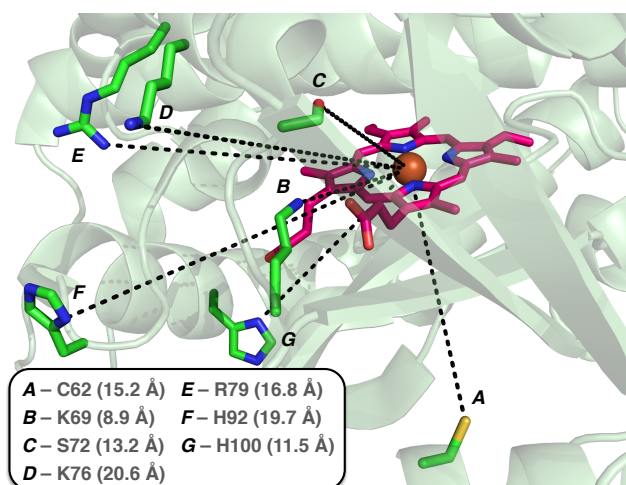


Figure 5-6. Location of alkylated residues relative to the heme cofactor as mapped onto WT BM3 crystal structure (PDB ID: 2IJ2). Measured distances to the Fe center (in parentheses) are determined based on the wild-type P450-BM3 crystal structure.

5.3.5 Site-directed mutagenesis of modified residues improves catalyst performance

The observations outlined above indicate that these modifications have deleterious effects on the secondary structure of the enzyme and the efficiency of the cyclopropanation reaction. We hypothesized that even though BM3-HStar will eventually be inactivated through modification of the heme prosthetic group, mutations of the alkylated residues to attenuate their nucleophilicity might better preserve structure and function during the reaction and, ultimately, improve carbene transfer efficiency to the substrate (measurable in terms of total turnover number to product). We therefore created seven single variants of BM3-HStar in which the alkylated residues were replaced with their non-nucleophilic isosteres. This design entails mutation of His to Asn (mutations H92N and H100N), Arg to Gln (mutation R79Q), Cys and Ser to Ala (mutations C62A and S72A), and Lys to Arg (mutations K69R and K76R). The seven variants were first tested for cyclopropanation activity in cell lysate. Three of the variants, C62A, K69R, and K76R, were poorly expressed and not studied further. Two of the variants, H92N and H100N, gave improved total turnover numbers (1.2- and 1.4-fold improvement, respectively) for the formation of **2**, while the remaining two failed to yield any improvements. Following this observation, we characterized the performance of the purified H92N and H100N variants and were pleased to observe similar enhancements in cyclopropanation activity. While purified BM3-HStar showed only modest conversion and poor selectivity, variants H92N and H100N in purified form gave much improved reaction yields and selectivities for cyclopropanation of **1** (**Figure 5-7A**). Moreover, these two mutations could be combined in a double variant H92N H100N that, as purified enzyme, afforded up to 560 TTN (0.1% enzyme loading, 1.8- fold improvement over BM3-HStar) and 98:2 d.r. and 87% *ee*; selectivity numbers that are very close to those previously observed in reactions with whole cells expressing BM3-HStar (**Scheme 5-1**).

Next, we evaluated the performance of the new variants in whole-cell reactions, as the enzymes have hitherto provided superior performance when used as whole-cell catalysts. Gratifyingly, the improvement was even more significant (**Figure 5-7B**). In particular,

variant H92N H100N gave more than 20,000 TTN in the production of **2** (2.7-fold improvement over BM3-HStar), while maintaining the excellent diastereo- and enantioselectivity of the parent catalyst. Variants H92N, H100N, and H92N H100N express at similar levels to parent enzyme BM3-HStar (0.20–0.25 μM at $\text{OD}_{600} = 10$, see the Appendix E, section E.5 for determination of protein concentration in whole-cell suspension). Thus, the observed increase in TTN translates to similar improvements in product per dry cell weight.

Finally, the improvement in activity imparted by mutations H92N and H100N could be generalized to other carbene transfer reactions, including styrene cyclopropanation and carbene insertion into the N–H bond of aniline (**Figure E-5**). Under whole-cell reaction conditions, HStar H92N H100N yielded 55% and 30% increases in styrene cyclopropanation and aniline N–H insertion activities over the parent BM3-HStar.

5.3.6 Characterization of BM3-HStar H92N H100N

Comparing the kinetics of formation of **2** with BM3-HStar and HStar H92N H100N under whole-cell reaction conditions at 10 mM substrate concentrations and ca. 0.2 μM enzyme concentration shows that the engineered variant supports faster product formation, but both rates decrease after ~30 min (**Figure E-6**). We performed similar inactivation analyses on BM3-HStar H92N H100N to find out whether the inactivation mechanism had changed. We noted changes in the CD spectral features upon treatment of ferrous BM3-HStar H92N H100N with EDA, similar to BM3-HStar (**Figure E-7**). In addition, bottom-up proteomic analyses indicated that residues K69, S72, K76, S89, K97, S304, and K306 of BM3-HStar H92N H100N were alkylated during the initial stages of this reaction, but no modification occurred at N92 or N100. Of these, only K69, S72, and K76 were observed to undergo modification in the parent enzyme BM3-HStar. Lastly, LC-MS analysis of purified BM3-HStar H92N H100N after cyclopropanation showed that the porphyrin cofactor underwent identical alkylative modifications as in BM3-HStar.

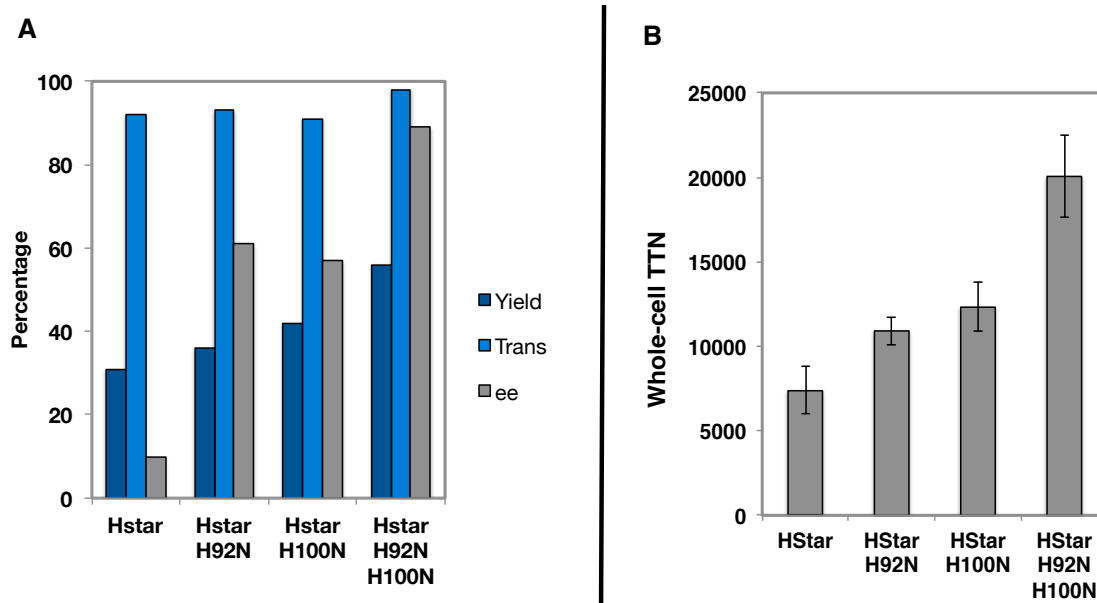
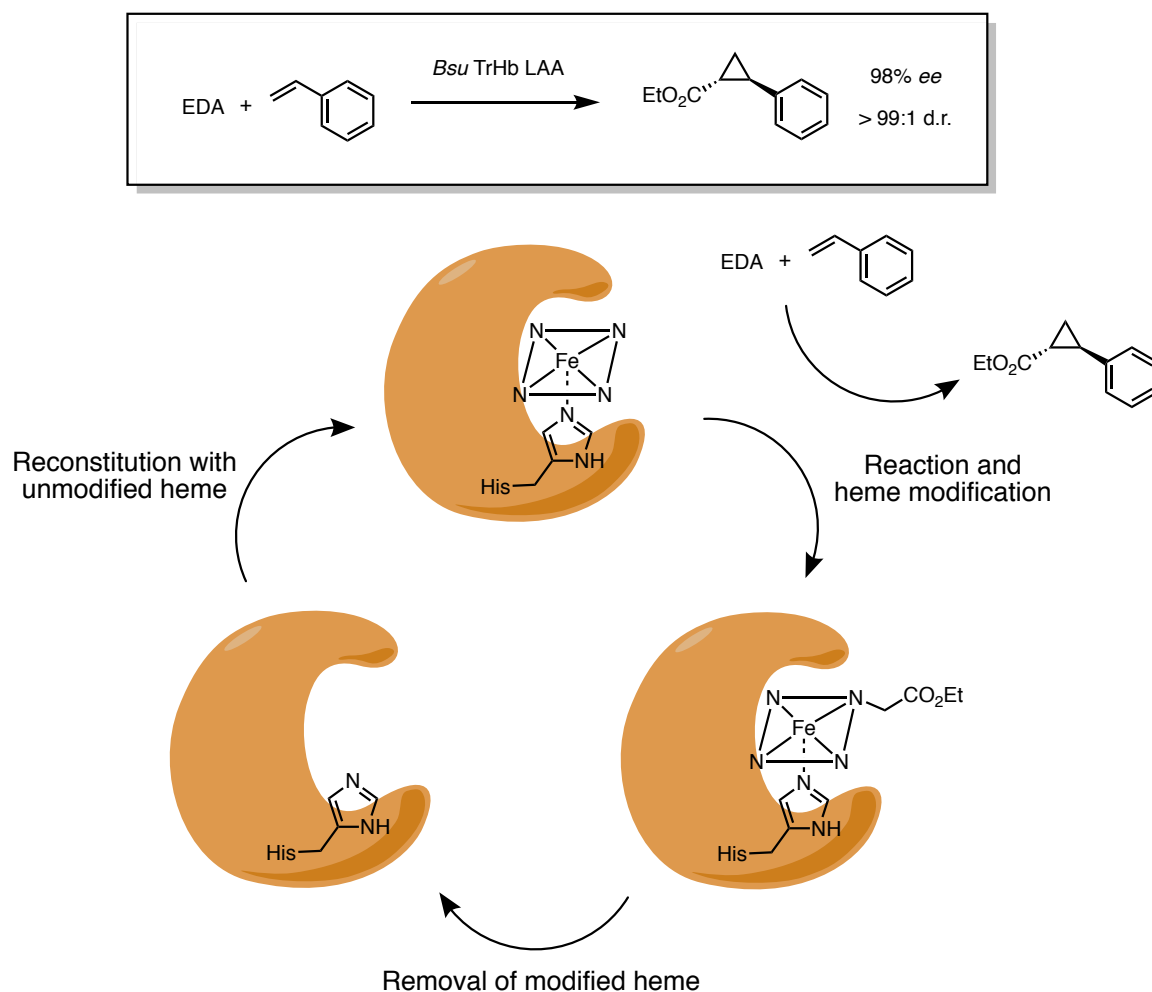


Figure 5-7. Site-directed mutagenesis of BM3-HStar to improve yield and selectivity of cyclopropanation of acrylamide **1** with EDA. **A.** Reactions were performed with 10 μ M enzyme, 10 mM **1**, 20 mM EDA, and 10 mM $\text{Na}_2\text{S}_2\text{O}_4$ in 100 mM KPi buffer (pH = 8.0). Yields and diastereoselectivity were determined by GC calibrated for **2**. Enantioselectivity was determined by chiral SFC (see Appendix E, E.7 and E.8). **B.** Total turnover numbers (TTNs) for cyclopropanation of **1** with whole cells expressing BM3-HStar variants. Reactions were performed with whole cells expressing the enzyme (resuspended in M9-N media to OD600 = 10), 10 mM **1**, and 20 mM EDA. Enzyme concentration in whole-cell suspension was determined by ferrous expression assay (Appendix E); product concentration at the end of the reaction was determined by GC calibrated for **2**.

5.3.7 Replacement of the modified heme cofactor

Our engineering efforts targeting alkylated side chains in BM3-HStar could not prevent inactivation entirely, but yielded some improvement to the turnover and selectivity of the enzymes. As our strategy did not investigate the effects side chain alkylation in the absence heme cofactor modification, it is difficult to conclude if both factors lead to inactivation, or if one proceeds the other. Therefore, we next sought to develop a method that could help address the problem of heme modification. We hypothesized that, if after a reaction, we could remove the modified heme cofactor and replace it with unmodified heme, it could regenerate

the enzyme and allow it to be reused (**Scheme 5-2**). In order to develop a system to replace the heme cofactor, we chose to focus on a protein other than BM3-HStar, as we hypothesized that the poor folding associated with histidine-ligated P450 enzymes might lead to problems when attempting to remove and replace the heme cofactor⁴⁰. Instead, we chose to focus on *Bacillus subtilis* truncated globin Y25L T45A Q49A (*Bsu* TrHb LAA), a carbene transfer heme enzyme engineered to catalyze a cyclopropanation reaction, producing an intermediate of the anti-stroke agent ticagrelor⁴.



Scheme 5-2. Proposed method for replacing modified heme to circumvent enzyme inactivation.

Initially, we sought to be able to recover our enzymes from the reaction solution, so that they could be used multiple times and their activity loss over time could be monitored. We

expressed and purified *Bsu* TrHb LAA, and covalently immobilized the protein on a solid support (See Appendix E, section **E.20**). Using the reaction between styrene and EDA as a model reaction (**Scheme 5-2**), we ran successive reactions with the beads, recovering the beads after each reaction. Over the multiple uses, we observed a substantial decrease in activity.

We then sought to develop a method to denature the protein on the beads, remove the heme cofactor, and then refold the protein with unmodified heme cofactor (**Scheme 5-2**). After screening a panel of denaturants and buffer conditions, we found that heme could be extracted off the beads in a neutral buffer solution containing 50% v/v acetone (Appendix E, section **E.21**).

Then we looked for conditions that enabled denatured enzyme to properly refold and reincorporate heme, as measured by the proteins ability to catalyze the model reaction with high selectivity and high activity. We emphasized high enantioselectivity to ensure that the protein and heme cofactor were together and folded properly, since the native *Bsu* TrHb LAA displays very high selectivity for the cyclopropanation of styrene, and poorly folded protein displays low selectivity. Ultimately, we discovered that the protein could be refolded in M9-N buffer, supplemented with heme, imidazole, and guanadinium chloride (Appendix E, section **E.21**).

Control experiments demonstrated that the protein-beads are essentially inactive after heme removal, and that the reconstituted protein-beads have activity and selectivity comparable to the native protein-beads (Appendix E, **Figure E-9**). These control experiments indicated that our heme removal and heme reconstitution steps are highly efficient, and allow us to fully replace any modified heme cofactor. When the protein-beads were subjected to a cycle of heme removal and heme reconstitution before each reaction (treated protein-beads), we observed that the treated protein-beads displayed activity and selectivity comparable to untreated protein-beads for the first two reactions. For the third and fourth reactions, we observed that the treated protein-beads had 2-fold and 3-fold higher activity than the untreated protein-beads (respectively), although the treated protein-beads still displayed a marked decrease in activity, suggestive of enzyme inactivation.

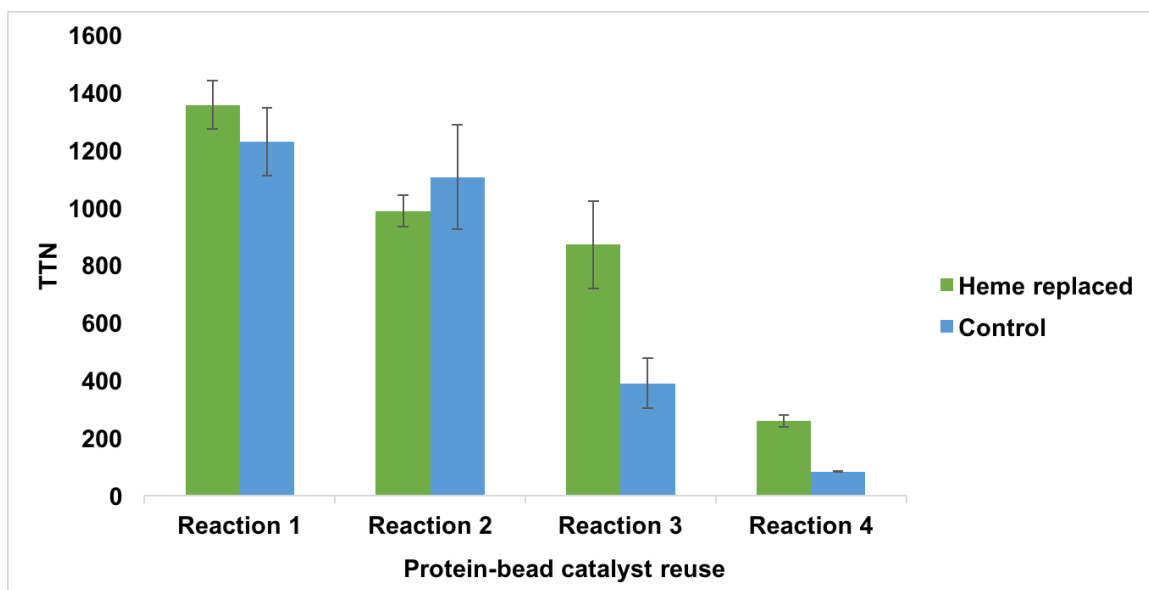


Figure 5-8. Use and re-use of the biocatalyst *Bsu* TrHb LAA immobilized on agarose beads. Replacement of the heme cofactor before each use yields some improvement over simple re-use of the beads, but the enzyme is still subject to inactivation under both conditions. TTN calculated relative to total amount of protein immobilized to beads. All reactions generated product with 98%-99% *ee* and >99:1 d.r.

5.4 Discussion

Various reports of irreversible inhibition of P450s by carbene precursors led us to investigate whether such pathways compete with productive catalysis in engineered “carbene transferases” derived from heme-containing proteins. This new enzyme activity involves a reactive intermediate that can damage the cofactor and protein if the carbene is not efficiently transferred to the substrate. Mutations to the protein sequence could assist productive catalysis, either by promoting productive transfer through improved binding and orientation of the substrate or by disfavoring non-productive transfer to the protein itself. Evolution finds both types of solutions by molding the protein active site and dynamics to promote productive catalysis, even when highly reactive species are generated during the catalytic cycle. Native P450 reactions that go through Compound I provide an excellent example of how an enzyme can generate a highly reactive species and direct it (most of the time) to a substrate. In contrast to P450-catalyzed monooxygenations, however, the newly-developed

carbene transfer activities of heme proteins have not yet been highly tuned by evolution.

We are starting to investigate opportunities for enhancing this non-natural activity by directed evolution¹ and rational engineering based on understanding some of the failure modes.

We found that the heme cofactor and several nucleophilic residues are modified by transfer of the reactive carbene to the enzyme rather than the olefin substrate during cyclopropanation of phenylacrylamide with EDA. Up to three additions of ethyl acetylidene fragment on the porphyrin were observed, and spectroscopic analysis of the modified porphyrin species suggests the formation of *N*-alkylporphyrin products. Presumably, this modification also interferes with metal binding, leading to the expulsion of Fe³⁺ as seen in the mass spectrum. These results suggest erosion of the active heme species, eventually leading to complete loss of the cofactor and catalytic function. Similar cofactor loss during reactions with other heme protein-based catalysts (myoglobin and a Ser-ligated P450) suggests that this off-cycle reaction is an inherent feature of porphyrin-based protein catalysts of carbene transfer reactions.

We also obtained evidence that carbene formation can result in the modification of protein side chains and disruption of BM3-HStar secondary structure. The observation of multiply modified species in the proteomic analyses implies that BM3-HStar is still capable of reacting with EDA even after initial modification. However, the iron carbene species might possess increasingly diminished reactivity and/or selectivity for productive cyclopropane formation due to the cumulative effects of protein structural modifications, accounting for the observed erosion in activity selectivity of BM3-HStar during pre-incubation with EDA.

We reasoned that reducing the nucleophilicity of some of the modified residues could reduce protein damage and possibly lead to better secondary structure preservation during the reaction and less inactivation. We identified two beneficial mutations in BM3-HStar, H92N and H100N, that improve carbene transfer activity and selectivity. That BM3-HStar double mutant H92N H100N performed better in other types of carbene transfer reactions—styrene cyclopropanation and N–H insertion—indicates that the improvements are general rather than specific to substrate **1**. A possible explanation for the effects of the mutations is

that the reaction ensemble contains more active form of the enzyme. BM3-HStar H92N H100N, however, still undergoes irreversible inactivation: even though the mutations avert potentially deleterious modification at residues 92 and 100, the reactive iron carbene species nonetheless inflicts alkylative damage on the porphyrin and other nucleophilic residues, including several that did not appear to be alkylated in BM3-HStar.

Our results indicate that EDA-induced inactivation is operative in reactions with purified enzyme and whole-cell catalyst. Yet, the relative rates of inactivation are strikingly different, with purified enzyme suffering much more rapid inactivation as well as inferior initial activity and selectivity. The origins of this difference still unclear, but several possible explanations can be put forth. One is that cellular components contribute to higher enzyme stability through the protective effects of, for example, molecular chaperones^{45,46}, macromolecular crowding⁴⁷⁻⁴⁹, or even cytoplasmic molecules acting as sacrificial carbene acceptors. Another possibility is that of diffusion across the cell membrane acting causes a “slow addition” of the carbene precursor to the enzyme, which is known to decrease side reactions in synthetic systems.

Finally, we were able to develop a system to replace the heme after each reaction, which allowed us to separate the cofactor modification pathway from the peptide modification pathway. Even in the system that allows for complete replacement of the heme cofactor, we still observe catalyst inactivation, although to a lesser degree than without cofactor replacement. It remains to be seen whether further engineering efforts, possibly through a combination of directed evolution, targeted mutagenesis, and cofactor replacement strategies, can generate a carbene transferase that is highly resistant to inactivation—performing hundreds of thousands of turnovers, and only inactivating in a manner that is easily reversible.

5.5 Conclusions

Exploiting the ability of proteins to evolve and adapt to new challenges, our laboratory has developed variants of heme proteins as platforms for non-natural reactions with xenobiotic diazo compounds. However, introduction of xenobiotic reagents to natural enzymes can lead to undesired repercussions such as inactivation^{23–25}. We have provided clear evidence for mechanism-based inactivation of a His-ligated P450-BM3 variant by ethyl diazoacetate that results in the modification of the heme prosthetic group and several protein side chains, which degrades enzyme performance and causes irreversible activity loss. Site-directed mutagenesis of some of the affected residues to amino acids that are less easily modified created superior variants with improved cyclopropanation activity as both purified enzymes and whole-cell catalysts. An alternate strategy, replacement of the modified cofactor, also yielded improvements in catalyst performance, but was also unable to completely avoid catalyst inactivation.

Classical directed evolution could also be applied to address the problem of catalyst inactivation—simply screening enough variants until an enzyme with the desired turnover and selectivity is discovered. However, screening can be slow for some reactions, especially when high enantioselectivity is desired, and so other engineering methods can be introduced to complement directed evolution.

Tracing the catalyst's fate during the reaction enables the identification of inactivation “hotspots”, and allows for the discovery of beneficial mutations at more distant residues. Similarly, understanding that damage to the cofactor is a reversible source of inactivation may allow for the development of industrial processes that help improve catalyst utility. Echoing a recent report on mitigation of substrate inhibition in a DERA enzyme²⁵, our results demonstrate the merits of mechanistic analyses in improving the overall activity of enzymes for biotechnological applications, especially ones that are employed in non-native environments. As interest grows in developing new and useful carbene transferases, we expect that a holistic mechanistic understanding of these enzymes will serve as an invaluable guide for their engineering.

5.6 References

1. Coelho, P. S., Brustad, E. M., Kannan, A. & Arnold, F. H. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* **339**, 307–310 (2013).
2. Coelho, P. S. *et al.* A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo. *Nat. Chem. Biol.* **9**, 485–7 (2013).
3. Wang, Z. J. *et al.* Improved cyclopropanation activity of histidine-ligated cytochrome P450 enables the enantioselective formal synthesis of levomilnacipran. *Angew. Chem. Int. Ed.* **53**, 6810–6813 (2014).
4. Hernandez, K. E. *et al.* Highly stereoselective biocatalytic synthesis of key cyclopropane intermediate to ticagrelor. *ACS Catal.* 7810–7813 (2016). doi:10.1021/acscatal.6b02550
5. Knight, A. M. *et al.* Diverse Engineered Heme Proteins Enable Stereodivergent Cyclopropanation of Unactivated Alkenes. *ACS Cent. Sci.* **4**, 372–377 (2018).
6. Chen, K., Huang, X., Jennifer Kan, S. B., Zhang, R. K. & Arnold, F. H. Enzymatic construction of highly strained carbocycles. *Science* **360**, 71–75 (2018).
7. Bordeaux, M., Tyagi, V. & Fasan, R. Highly diastereoselective and enantioselective olefin cyclopropanation using engineered myoglobin-based catalysts. *Angew. Chem. Int. Ed.* **54**, 1744–1748 (2015).
8. Sreenilayam, G. & Fasan, R. Myoglobin-catalyzed intermolecular carbene N-H insertion with arylamine substrates. *Chem. Commun.* **51**, 1532–1534 (2015).
9. Tyagi, V., Sreenilayam, G., Bajaj, P., Tinoco, A. & Fasan, R. Biocatalytic Synthesis of Allylic and Allenyl Sulfides through a Myoglobin-Catalyzed Doyle–Kirmse Reaction. *Angew. Chem. Int. Ed.* **55**, 13562–13566 (2016).
10. Gober, J. G. *et al.* Mutating a Highly Conserved Residue in Diverse Cytochrome P450s Facilitates Diastereoselective Olefin Cyclopropanation. *ChemBioChem* **17**, 394–397 (2016).
11. Weissenborn, M. J. *et al.* Enzyme-Catalyzed Carbonyl Olefination by the E. coli Protein YfeX in the Absence of Phosphines. *ChemCatChem* **8**, 1636–1640 (2016).
12. Kan, S. B. J., Lewis, R. D., Chen, K. & Arnold, F. H. Directed evolution of cytochrome c for carbon–silicon bond formation: Bringing silicon to life. *Science* **354**, 1048–1051 (2016).

13. Kan, S. B. J., Huang, X., Gumulya, Y., Chen, K. & Arnold, F. H. Genetically programmed chiral organoborane synthesis. *Nature* **552**, 132–136 (2017).
14. Khade, R. L. *et al.* Iron Porphyrin Carbenes as Catalytic Intermediates: Structures, Mössbauer and NMR Spectroscopic Properties, and Bonding. *Angew. Chem. Int. Ed.* **53**, 7574–7578 (2014).
15. Zhang, R. K. *et al.* Enzymatic assembly of carbon–carbon bonds via iron-catalysed sp³ C–H functionalization. *Nature* **565**, 67–72 (2019).
16. Mansuy, D., Nastainczyk, W. & Ullrich, V. The mechanism of halothane binding to microsomal cytochrome P450. *Naunyn. Schmiedebergs. Arch. Pharmacol.* **285**, 315–324 (1974).
17. Nastainczyk, W., Ullrich, V. & Sies, H. Effect of oxygen concentration on the reaction of halothane with cytochrome P450 in liver microsomes and isolated perfused rat liver. *Biochem. Pharmacol.* **27**, 387–392 (1978).
18. Krieter, P. A. & Van Dyke, R. A. Cytochrome P-450 and halothane metabolism. Decrease in rat liver microsomal P-450 in vitro. *Chem. Biol. Interact.* **44**, 219–235 (1983).
19. Wolf, C. R., Mansuy, D., Nastainczyk, W., Deutschmann, G. & Ullrich, V. The Reduction of Polyhalogenated Cytochrome Methanes by Liver Microsomal. *Mol. Pharmacol.* **13**, 698–705 (1977).
20. Manno, M., Reed, C., King, L. J. & De Matteis, F. Enzymatic and/or Non-Enzymatic ‘Suicidal’ Activation of Carbon Tetrachloride by Heme and Cytochrome P-450. *Arch. Toxicol., Suppl.* **12**, 315–317 (1988).
21. Manno, M., De Matteis, F. & King, L. J. The Mechanism of the Suicidal, Reductive Inactivation of Microsomal Cytochrome P-450 by Carbon Tetrachloride. *J. Biochem. Pharmacol.* **37**, 1981–1990 (1988).
22. Ortiz de Montellano, P. R. & Grab, L. A. Inactivation of Cytochrome P-450 during Catalytic Oxidation of a 3-[(Arylthio)ethyl]sydnone: N-Vinyl Heme Formation via Insertion into the Fe-N Bond. *J. Am. Chem. Soc.* **108**, 5584–5589 (1986).
23. Estell, D. A., Graycar, T. P. & Wells, J. A. Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. *J. Biol. Chem.* **260**, 6518–6521 (1985).
24. Kim, S. J., Joo, J. C., Song, B. K., Yoo, Y. J. & Kim, Y. H. Engineering a horseradish peroxidase C stable to radical attacks by mutating multiple radical coupling sites. *Biotechnol. Bioeng.* **112**, 668–676 (2015).

25. Dick, M. *et al.* Mechanism-based inhibition of an aldolase at high concentrations of its natural substrate acetaldehyde: Structural insights and protective strategies. *Chem. Sci.* **7**, 4492–4502 (2016).
26. Wolf, J. R., Kodadek, T., Hamaker, C. G., Djukic, J. P. & Woo, L. K. Shape and Stereoselective Cyclopropanation of Alkenes Catalyzed by Iron Porphyrins. *J. Am. Chem. Soc.* **117**, 9194–9199 (1995).
27. Manno, M., Cazzaro, S. & Rezzadore, M. The mechanism of the suicidal reductive inactivation of microsomal cytochrome P-450 by halothane. *Arch. Toxicol.* **65**, 191–198 (1991).
28. Groves, J. T., Avaria-Neisser, G. E., Fish, K. M., Imachi, M. & Kuczkowski, R. L. Hydrogen-Deuterium Exchange during Propylene Epoxidation by Cytochrome P-450. *J. Am. Chem. Soc.* **108**, 3837–3838 (1986).
29. Kent U.M., Jushchhyshyn M.I. & Hollenberg P.F. Mechanism-Based Inactivators as Probes of Cytochrome P450 Structure and Function. *Curr. Drug Metab.* **2**, 29 (2001).
30. Hollenberg, P. F., Kent, U. M. & Bumpus, N. N. Mechanism-based inactivation of human cytochromes P450s: Experimental characterization, reactive intermediates, and clinical implications. *Chem. Res. Toxicol.* **21**, 189–205 (2008).
31. Hirao, H., Cheong, Z. H. & Wang, X. Pivotal role of water in terminating enzymatic function: A density functional theory study of the mechanism-based inactivation of cytochromes P450. *J. Phys. Chem. B* **116**, 7787–7794 (2012).
32. Komives, E. A., Olmstead, M. M., Ortiz de Montellano, P. R. & Tew, D. Models for Cytochrome P450 Prosthetic Heme Alkylation. Reaction of Diazoacetophenone with (Tetraphenylporphyrinato)iron(II) Chloride. *Inorg. Chem.* **27**, 3112–3117 (1988).
33. Ortiz de Montellano, P. R., Kunze, K. L., Beilan, H. S. & Wheeler, C. Destruction of Cytochrome P-450 by Vinyl Fluoride, Fluroxene, and Acetylene. Evidence for a Radical Intermediate in Olefin Oxidation. *Biochemistry* **21**, 1331–1339 (1982).
34. Setsune, J., Iida, T. & Kitao, T. The Reaction of Cobalt(III) Porphyrins with α -Diazo- β -Dicarbonyl Compounds. *Chem. Lett.* **18**, 885–888 (2006).
35. Khade, R. L. & Zhang, Y. Catalytic and Biocatalytic Iron Porphyrin Carbene Formation: Effects of Binding Mode, Carbene Substituent, Porphyrin Substituent, and Protein Axial Ligand. *J. Am. Chem. Soc.* **137**, 7560–7563 (2015).
36. 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. **138**, 9597-9610 (2016). doi:10.1021/jacs.6b04636

37. Lewis, R. D. *et al.* Catalytic iron-carbene intermediate revealed in a cytochrome c carbene transferase. *Proc. Natl. Acad. Sci.* **115**, 7308–7313 (2018).
38. Li, H. & Poulos, T. L. The structure of the cytochrome P450 BM3 haem domain complexed with the fatty acid substrate, palmitoleic acid. *Nat. Struct. Biol.* **4**, 140–146 (1997).
39. Brustad, E. M. *et al.* Structure-guided directed evolution of highly selective P450-based magnetic resonance imaging sensors for dopamine and serotonin. *J. Mol. Biol.* **422**, 245–262 (2012).
40. McIntosh, J. A., Heel, T., Buller, A. R., Chio, L. & Arnold, F. H. Structural Adaptability Facilitates Histidine Heme Ligation in a Cytochrome P450. *J. Am. Chem. Soc.* **137**, 13861–13865 (2015).
41. Whitehouse, C. J. C., Bell, S. G. & Wong, L. L. P450 BM3 (CYP102A1): Connecting the dots. *Chem. Soc. Rev.* **41**, 1218–1260 (2012).
42. Jung, S. T., Lauchli, R. & Arnold, F. H. Cytochrome P450: Taming a wild type enzyme. *Current Opinion in Biotechnology* **22**, 809–817 (2011).
43. Sevrioukova, I. F., Hazzard, J. T., Tollin, G. & Poulos, T. L. The FMN to heme electron transfer in cytochrome P450BM-3. Effect of chemical modification of cysteines engineered at the FMN-heme domain interaction site. *J. Biol. Chem.* **274**, 36097–36106 (1999).
44. Tran, N. H. *et al.* A series of hybrid P450 BM3 enzymes with different catalytic activity in the light-initiated hydroxylation of lauric acid. *J. Inorg. Biochem.* **115**, 50–56 (2012).
45. Ganea, E. & Harding, J. J. Molecular Chaperones Protect Against Glycation-Induced Inactivation of Glucose-6-Phosphate Dehydrogenase. *Eur. J. Biochem.* **231**, 181–185 (1995).
46. Hook, D. W. A. & Harding, J. J. Alpha-crystallin acting as a molecular chaperone protects catalase against steroid-induced inactivation. *FEBS Lett.* **382**, 281–284 (1996).
47. Zimmerman, S. B. & Pfeiffer, B. H. Macromolecular crowding allows blunt-end ligation by DNA ligases from rat liver or *Escherichia coli*. *Proc. Natl. Acad. Sci.* **80**, 5852–5856 (1983).

48. Guo, Z. F., Jiang, M., Zheng, S. & Guo, Z. Structural change of the enterobactin synthetase in crowded solution and its relation to crowding-enhanced product specificity in nonribosomal enterobactin biosynthesis. *Bioorganic Med. Chem. Lett.* **20**, 3855–3858 (2010).
49. Moran-Zorzano, M. T. *et al.* Escherichia coli AspP activity is enhanced by macromolecular crowding and by both glucose-1,6-bisphosphate and nucleotide-sugars. *FEBS Lett.* **581**, 1035–1040 (2007).

SUPPORTING INFORMATION FOR CHAPTER 5

Material from this chapter appears in “Renata H[†], Lewis RD[†], Sweredoski MJ, Moradian A, Hess S, Wang ZJ, Arnold FH. (2016) **Identification of mechanism-based inactivation in P450-catalyzed cyclopropanation facilitates engineering of improved enzymes.** *JACS*, 138:12527-12533. doi:10.1021/jacs.6b06823”. Reproduced with permission from the American Chemical Society. [†]Denotes equal author contribution.

E.1 Materials and methods

Unless otherwise noted, all chemicals and reagents for chemical reactions were obtained from commercial suppliers (Sigma-Aldrich, Acros) and used without further purification. Sonication was performed using a Qsonica Q500 sonicator. For UV-vis analyses, spectra were collected between 650 and 390 nm on a UV1800 Shimadzu spectrophotometer. Circular dichroism (CD) analyses were performed using Aviv Biomedical Model 410 CD spectrophotometer. Gas chromatography (GC) analyses were carried out using a Shimadzu GC-17A gas chromatograph with FID detector and J&W HP-5 column (30 m x 0.32 mm, 0.25 μ m film) and 2-phenylethanol as an internal standard. Analytical SFC was performed with a Mettler SFC supercritical CO₂ analytical chromatography system with isopropanol and liquid CO₂ as the mobile phase. Enantiomers were separated using a Chiralpak AS column (4.6 mm x 25 cm) obtained from Daicel Chemical Industries, Ltd. High resolution mass spectrometry (HRMS) was performed with Agilent 6200 Series TOF with an Agilent G1978A multimode source in electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), or mixed (MM) ionization mode. Enzymes (DpnI, Phusion polymerase) were purchased from New England Biolabs (NEB, Ipswich, MA).

E.2 Generation of enzyme variants

Plasmid pET22b(+) was used as a cloning and expression vector for all enzymes used in this study. Site-directed mutagenesis was accomplished by modified QuikChangeTM protocol using primers bearing the desired mutations (IDT, San Diego, CA). The PCR products were gel purified, digested with DpnI, repaired using Gibson assembly¹, and used directly to transform *E. coli* strain BL21(DE3). Variants were stored as glycerol stocks at $-80\text{ }^{\circ}\text{C}$.

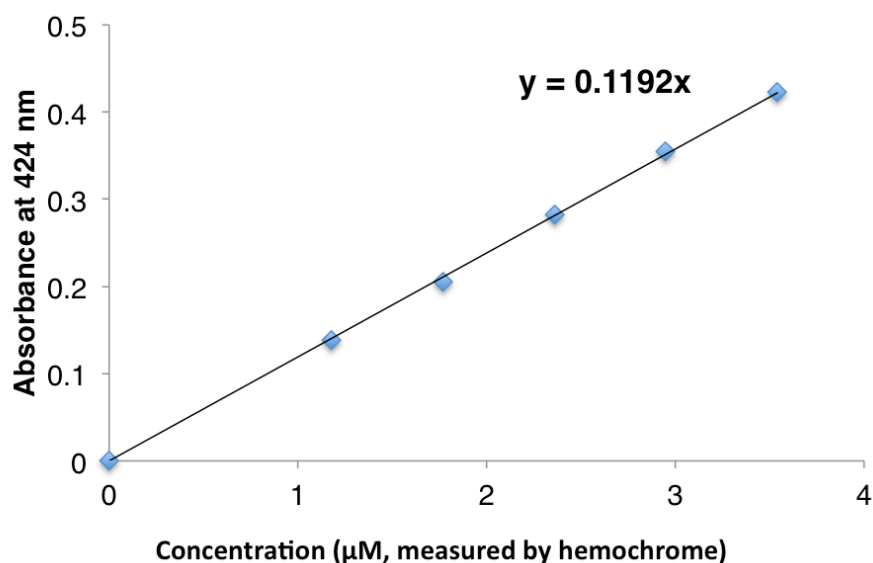
E.3 Hemochrome binding assay

A solution of pyridine was made by combining 1.75 mL pyridine with 0.75 mL 1 M NaOH. The solution mixed at room temperature was centrifuged for 30 s at 5000 rcf to remove excess aqueous base. To a cuvette containing 0.75 mL of protein solution in phosphate buffer (0.1 M, pH 8.0), 0.25 mL of the pyridine solution was added followed by a few grains (less than 2.0 mg) of sodium dithionite. The cuvette was sealed with parafilm and a UV-vis spectrum was recorded immediately. P450 concentration was determined from the absorbance of the hemochrome complex using extinction coefficient $\epsilon_{418} = 196\text{ mM}^{-1}\text{ cm}^{-1}$. Absorbance was assigned as the difference between the peak max at 418 nm and the baseline at 420 nm as determined by extrapolating from two points on either side of the hemochrome peak.

E.4 Determination of ϵ_{max} for ferrous histidine-ligated P450 enzymes

BM3-HStar holoprotein with known concentration (determined by hemochrome binding assay) was used for extinction coefficient determination. An appropriate amount of the protein was transferred to a semi-micro cuvette fitted with a screw septum cap (Starna Cells, product number 9-SOG-10-GL14-S). The cuvette was made anaerobic by flushing with argon. In 6 mL crimp-sealed vials, a phosphate buffer solution (0.1 M, pH = 8.0) and a 12.5 mM solution of sodium dithionite in phosphate buffer (0.1 M, pH = 8.0) were degassed by bubbling with argon. To the semi-micro cuvette, 700 μL of the degassed phosphate buffer

and 250 μL of the degassed $\text{Na}_2\text{S}_2\text{O}_4$ solution were added, and then the UV-vis spectrum was recorded. Absorbance was assigned as the difference between the peak max at 424 nm and the baseline as determined by extrapolating from two points on either side of the ferrous peak (390 nm and 460 nm). The measurement was repeated for five different enzyme concentrations and a calibration curve was constructed to determine the ferrous ϵ_{max} .



From the above calibration curve, we determine $\epsilon_{\text{max}} = 119 \text{ mM}^{-1}\text{cm}^{-1}$

E.5 Ferrous concentration assay

Ferrous concentration assay was used to determine the concentration of His-ligated P450-BM3 variant in crude lysate. To an anaerobic cuvette containing 800 μL of P450 lysate of unknown concentration was added 200 μL of degassed $\text{Na}_2\text{S}_2\text{O}_4$ solution (2 mg/mL), and the UV-vis spectrum was recorded. Absorbance was assigned as the difference between the peak max at 424 nm and the baseline as determined by extrapolating from two points on either side of the ferrous peak (390 nm and 460 nm). P450 concentration was determined from the absorbance of the ferrous species using an extinction coefficient of $\epsilon = 119 \text{ mM}^{-1}\text{cm}^{-1}$ for histidine-ligated P450-BM3.

E.6 P450 expression and purification

One liter Hyperbroth_{amp} (AthenaES©) was inoculated with an overnight culture (25 mL, LB_{amp}) of recombinant *E. coli* BL21(DE3) cells harboring a pET22b(+) plasmid encoding the P450 variant. The cultures were shaken at 200 rpm at 37 °C for roughly 3 h or until an optical density of 0.7–1.0 was reached. Cultures were cooled on ice (20 min) and then induced by adding IPTG and aminolevulinic acid to a final concentration of 0.5 mM and 1.0 mM, respectively. The cultures were allowed to continue for another 20 hours at 20 °C and shaking at 200 rpm. Cells were harvested by centrifugation (4 °C, 15 min, 3,000xg), and the cell pellet was stored at –20 °C or below for at least 2 h.

Purification was performed with an AKTA purifier FPLC system (GE Healthcare). For the purification of 6XHis tagged P450s, the thawed cell pellet was resuspended in Ni-NTA buffer A (25 mM Tris.HCl, 200 mM NaCl, 25 mM imidazole, pH 8.0, 4 mL/g of cell wet weight) and lysed by sonication (2x1 min, 30% duty cycle). The lysate was centrifuged at 27,000xg for 20 min at 4 °C to remove cell debris. The collected supernatant was subjected to a Ni-NTA chromatography step using a Ni Sepharose column (HisTrap-HP, GE healthcare, Piscataway, NJ). The P450 was eluted from the Ni Sepharose column using 25 mM Tris.HCl, 200 mM NaCl, 300 mM imidazole, pH 8.0. Ni-purified protein was buffer exchanged into 0.1 M phosphate buffer (pH = 8.0) using a 10 kDa MW cut-off centrifugal filter. Protein concentrations were determined by hemochrome binding assay described above. For storage, proteins were portioned into 300 µL aliquots and stored at –80 °C.

E.7 Small-scale whole-cell bioconversions

E. coli BL21(DE3) cells with plasmid coding for a P450 mutant were grown from glycerol stock overnight (37 °C, 250 rpm) in 5 ml LB_{amp}. A portion of the pre-culture (1.25 mL) was used to inoculate 50 mL of Hyperbroth medium (1 L Hyperbroth prepared from powder from AthenaES©, 0.1 mg mL⁻¹ ampicillin) in a 250 mL Erlenmeyer flask and this culture was incubated at 37 °C, 200 rpm for approximately 3 h. At OD₆₀₀ = 0.7–1.0, the cultures were cooled on ice (20 min) before inducing with IPTG (0.5 mM final concentration)

and δ -aminolevulinic acid (1.0 mM final concentration). Shaking was continued at 20 °C, 200 rpm. Cultures were harvested after 20 h and resuspended in nitrogen-free M9-N medium (1 L: 31 g Na_2HPO_4 , 15 g KH_2PO_4 , 2.5 g NaCl , 0.24 g MgSO_4 , 0.01 g CaCl_2) until the indicated OD_{600} was reached ($\text{OD}_{600} = 30$ unless otherwise noted). Aliquots of the cell suspension were used for determination of the P450 expression level (2–3 mL) after lysis.

Anaerobic reaction conditions: *E. coli* cells of the appropriate OD_{600} were transferred to a crimped 6 mL vial and degassed with argon for 5-10 min. In parallel, a solution of glucose in M9-N (250 mM) was degassed with argon for the same period of time. All solutions were uncapped and transferred to an anaerobic chamber. Glucose solution (50 μL , 250 mM) was added to 2 mL crimp vials, followed by cell suspension (425 μL), the olefin substrate (12.5 μL of a 400 mM solution in EtOH), and EDA (12.5 μL of a 800 mM solution in EtOH). The vials were sealed, removed from the anaerobic chamber and shaken on a tabletop shake plate at room temperature for 1 h at 350 rpm. The reactions were quenched by addition of 20 μL of 3 M HCl, followed by 20 μL of the internal standard (20 mM 2-phenylethanol solution in cyclohexane) and 1 mL cyclohexane. The mixture was transferred to a 1.7 mL Eppendorf tube, vortexed and then centrifuged (10,000x rcf, 30 s). The organic layer was removed and analyzed by GC to determine yield and chiral SFC to determine enantioselectivity.

E.8 Small-scale cyclopropanation reactions with purified enzymes

Small-scale (400 μL) reactions were carried out in 2 mL glass crimp vials (Agilent Technologies, San Diego, CA). P450 solution (60 μL , 67 μM) was added to an unsealed crimp vial. In parallel, a 12.5 mM solution of sodium dithionite in phosphate buffer (0.1 M, pH = 8.0) was degassed by bubbling with argon in a 6 mL crimp-sealed vial. All solutions were uncapped and transferred to an anaerobic chamber. The buffer/dithionite solution (320 μL) was added to each reaction vial containing P450 solution, followed by the olefin substrate (10 μL , 400 mM in EtOH) and EDA (10 μL , 800 mM in EtOH). The reaction vials were capped, removed from the anaerobic chamber, and then placed in a tray on a plate shaker and left to shake at 350 rpm for 2 h at room temperature. The final concentrations of the reagents were typically: 10 olefin, 20 mM EDA, 10 mM $\text{Na}_2\text{S}_2\text{O}_4$, and 10 μM purified

enzyme. The reaction was quenched by the addition of 3 M HCl (20 μ L), followed by 20 μ L of the internal standard (20 mM 2-phenylethanol solution in cyclohexane) and 1 mL cyclohexane. The mixture was transferred to a 1.7 mL Eppendorf tube and vortexed and centrifuged (10,000x rcf, 30 s). The organic layer was then analyzed by GC to determine yield and chiral SFC to determine enantioselectivity. GC analysis of product **2** was performed using J&W HP-5 column (30 m x 0.32 mm, 0.25 μ M film) with the method 90 °C hold 2 min, 90-110 at 6 °C/min, 110-190 at 40 °C/min, 190-280 at 20 °C/min, 280 °C hold 1 min, 12.8 min total): internal standard (3.55 min), trans-**2** (9.55 min), and cis-**2** (9.69 min). Yields of cyclopropanation products were determined using calibration curves previously described in reference 1. Analytical SFC of product **2** was performed on a Chiralpak AS column, eluting with 2% iPrOH at 2.5 mL/min: λ = 210 nm, t_R (min): major = 8.50, minor = 9.48.

E.9 Analysis of diazo-induced inactivation

Small-scale reaction mixtures with whole cells expressing the appropriate enzyme variants or purified enzymes were prepared according to the general procedures outlined above. To the reaction vial was added 10 μ L of EDA solution (800 mM in EtOH). After the specified pre-incubation time, 10 μ L of olefin solution (400 mM in EtOH) was added. For determination of olefin-induced inactivation, the order of addition was reversed. The reactions were shaken on a tabletop shake plate at room temperature for 1 h at 350 rpm. The reactions were quenched by addition of 20 μ L of 3 M HCl, followed by 20 μ L of the internal standard (20 mM 2-phenylethanol solution in cyclohexane) and 1 mL cyclohexane. The mixture was transferred to a 1.7 mL Eppendorf tube, vortexed and then centrifuged (10,000x rcf, 30 s). The organic layer was removed and analyzed by GC to determine yield and by chiral SFC to determine enantioselectivity.

E.10 Liquid chromatography-mass spectrometry (LCMS) analysis of cofactor modification

To a 6 mL crimp-sealed vial was added BM3-HStar (final concentration = 10 μ M) and the vial was made anaerobic by degassing the headspace with argon for 10 min. In a separate 6 mL crimp-sealed vial, a 12.5 mM solution of Na₂S₂O₄ in phosphate buffer (0.1 M, pH = 8.0) was degassed by bubbling with argon for 10 min. The buffer/dithionite solution (1.8 mL) was then added to the reaction vial via syringe, followed by 50 μ L of 200 mM EDA solution. For investigation of reductant effects, degassed phosphate buffer solution (1.8 mL) was added in place of buffer/dithionite solution. The reaction was quenched at the appropriate time points by adding 100 μ L of 3 M HCl, followed by 2.5 mL of 9:1 dichloromethane:MeOH. The mixture was transferred to a 15 mL Falcon tube, vortexed and centrifuged (5000 rpm, 5 min). The organics were collected, and the extraction procedure was repeated one more time. The organics were combined, concentrated *in vacuo*, redissolved in MeCN, and submitted for LCMS analysis. For reaction progress monitoring, the analyses were performed at 15 s and 1 min time points.

E.11 Isolation of modified porphyrin cofactor

Cofactor was extracted from the reaction mixture of isolated BM3-HStar and EDA using the procedure outlined above. Following extraction, the crude organics were concentrated *in vacuo* and then purified by preparative TLC (10:90:0.1 MeOH:DCM:AcOH), isolating the red fluorescent band ($R_f = 0.3$). The isolated porphyrin was dissolved in MeOH and then submitted to HRMS analysis. UV-vis analysis was performed using a solution of isolated porphyrin in DCM. For UV-vis analysis of Zn²⁺-complexed porphyrin, isolated porphyrin was dissolved in 0.5% ZnCl₂ solution in 5:95 MeOH/DCM and submitted for analysis.

E.12 Analysis of enzyme modification by top-down proteomics

Small-scale reactions were performed following the general procedure for catalysis using purified enzyme. The protein samples were analyzed on an LC-MSD SL 1100 series (Agilent) liquid chromatography mass spectrometry system. The samples were subjected to reverse phase chromatography on a 2.1 x 150 mm Zorbax 300SB-C3 column (Agilent) using a gradient consistent of 0.2% formic acid (solvent A) and 0.2% formic acid in acetonitrile (89.8%) and methanol (10%) (solvent B). Eluted intact proteins were ionized using electrospray ionization and measured in the single quadrupole mass spectrometer and quantified.

E.13 Alkylation analysis by bottom-up proteomics

Small-scale reactions were performed following the general procedure for catalysis using purified enzyme. Appropriate amounts of the reaction aliquot were lyophilized to afford 10–20 μg of proteins. Dried pellets were dissolved in 40 μL of 8 M urea/100 mM Tris-HCl (pH 8.5) mixture. Proteins were reduced by addition of 0.25 μL of 500 mM TCEP (tris(2-carboxyethyl)-phosphine hydrochloride) and alkylated by addition of 0.9 μL of 500 mM iodoacetamide. For sample digestion, 1 μL of 0.1 gL^{-1} of lysyl endopeptidase (from *A. lyticus*) in 100 mM Tris-HCl (pH 8.5) was added, and the samples were incubated at room temperature for 4 h in the dark. The samples were diluted with 120 μL of 100 mM Tris-HCl (pH 8.5) and 1.6 μL of 100 mM CaCl_2 , followed by addition of 1 μL of 0.5 gL^{-1} of trypsin. After overnight incubation at room temperature, the digests were diluted to 5% formic acid concentration, desalted and lyophilized, and then resuspended in 0.2% formic acid solution for analysis. In-solution digested samples were subjected to LC-MS/MS analysis on a nanoflow LC system, EASY-nLC II, (Thermo Fisher Scientific) coupled to a LTQ Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a Nanospray Flex ion source (Thermo Fisher Scientific).

For the EASY-nLC II system, solvent A consisted of 97.8% H_2O , 2% MeCN, and 0.2% formic acid and solvent B consisted of 19.8% H_2O , 80% MeCN, and 0.2% formic acid.

Samples were directly loaded onto a 16-cm analytical HPLC column (75 mm ID) packed in-house with ReproSil-Pur C18AQ 3 μ m resin (120Å pore size, Dr. Maisch, Ammerbuch, Germany). The column was heated to 45° C. The peptides were separated with a 90 min gradient at a flow rate of 350 nL/min. The gradient was as follows: 2–30% Solvent B (90 min), 30–100% B (1 min), and 100% B (9 min). Eluted peptides were then ionized using a standard coated silica tip (New Objective, Woburn, MA) as an electrospray emitter and introduced into the mass spectrometer. The LTQ Orbitrap was operated in a data-dependent mode, automatically alternating between a full-scan (m/z 300-1700) in the Orbitrap and subsequent MS/MS scans of the 15 most abundant peaks in the linear ion trap (Top15 method). Data acquisition was controlled by Xcalibur 2.0.7 and Tune 2.4 software (Thermo Fisher Scientific).

Raw files were searched using MaxQuant (v 1.5.1.2) using default parameters except where noted. UniProt *E. coli* entries along with expressed protein sequence and a contaminant database were considered in the search. A decoy database was generated and searched in parallel to estimate the false discovery rate. Trypsin was specified as the digestion enzyme with up to two missed cleavages allowed. Precursor mass tolerance was 4.5 ppm after recalibration and fragment ion mass tolerance was 0.5 Da. Variable modifications included carbamidomethylation of cysteine (+57.0215 Da), oxidation of methionine (+15.9949 Da), protein N-terminal acetylation (+42.0106 Da), and the custom modification (+86.0368 Da) on lysine, cysteine, histidine, serine, and arginine. PSM, Protein, and Site FDRs were all less than 1% as estimated by the target-decoy approach.

E.14 Circular dichroism analysis

Small-scale reactions were performed following the general procedure for catalysis using the heme domain of the appropriate enzyme variants after isolation and purification by HisTrap-HP column (final concentrations: 10 μ M protein, 2 mM substrates, 2 mM $\text{Na}_2\text{S}_2\text{O}_4$). CD samples were obtained by mixing 20 μ L aliquot of the reaction with 1.0 mL of 10 mM phosphate buffer (pH 8), and 2.0 mL of H_2O . CD spectra were recorded in the far UV region,

190–260 nm, using a quartz cell with 1 cm optical path length. Measurements were taken every 1 nm at a scan rate of 1 nm/sec.

E.15 Whole-cell reaction of styrene with EDA catalyzed by BM3-HStar variants

Following the procedure for small-scale whole-cell reactions under anaerobic conditions, *E. coli* expressing BM3-HStar, and BM3-HStar H92N H100N were grown and resuspended to $OD_{600} = 10$. To determine the P450 concentration of each sample, 2 mL of each cell stock was removed and analyzed by ferrous concentration assays following the general procedure outlined above. Small-scale reactions were performed following the general procedure with 10 mM final concentrations each of styrene and EDA. Reactions were shaken for 2 h at room temperature. GC analysis of product **2** was performed using J&W HP-5 column (30 m x 0.32 mm, 0.25 μ M film) with the method 90 °C hold 2 min, 90-190 at 6 °C/min, 190-230 °C at 40 °C/min, 230 °C hold 2 min (internal standard: 3.19 min, *Z* product 8.97 min, *E* product 9.99 min). Yields of cyclopropanation products were determined using calibration curves previously described in Wang et al. (2014)².

E.16 Whole-cell reaction of aniline with EDA catalyzed by BM3-HStar variants

Following the procedure for small-scale whole-cell reactions under anaerobic conditions, *E. coli* expressing BM3-HStar, and BM3-HStar H92N H100N were grown and resuspended to $OD_{600} = 10$. To determine the P450 concentration of each sample, 2 mL of each cell stock was removed and analyzed by ferrous concentration assays following the general procedure outlined above. Small-scale reactions were performed following the general procedure with 20 mM and 10 mM final concentrations of aniline and EDA, respectively. Reactions were shaken for 2 h at room temperature. GC analysis of product **2** was performed using J&W HP-5 column (30 m x 0.32 mm, 0.25 μ M film) with the method 90 °C hold 2 min, 90-190 at 6 °C/min, 190-230 °C at 40 °C/min, 230 °C hold 2 min (internal standard: 3.19 min, product 9.84 min). Yields of N–H insertion product were determined using calibration curves described below.

E.17 Immobilization of *Bsu* TrHb LAA

1 mL of purified *Bsu* TrHb LAA³ (expressed and purified as described in Appendix A, section A.5, purified as described in Appendix E, section E.6) concentration of 5 mg/mL, 310 mM) was added to 78.5 mg of N-hydroxysuccinimide-linked agarose resin beads (ThermoFisher) in a 1.7 mL Eppendorf tube. The tube was rotated end-over-end for 90 minutes at room temperature. The solution was then transferred to a spin column, centrifuged at 1000xg for 1 min, and the flow-through was collected. The beads were then washed with 1 mL of 20 mM sodium phosphate buffer pH 8, centrifuged at 1000xg for 1 min, and the flow-through was again collected. The two flow-through solutions were combined and concentrated to 300 uL. Finally, 1 mL of 1 M Tris-HCl pH 8 was added to the beads, and then the column was rotated end-over-end for 1 hr at room temperature. Then column was then centrifuged at 1 g for 1 min, and the flow-through was discarded. The beads were transferred to a new 1.7 mL Eppendorf tube, resuspended to a volume of 1.1 mL with sterile M9-N buffer, and stored at 4°C.

To determine the coupling efficiency of the beads, a bicinchoninic acid assay (using bovine serine albumin as a standard) was used to measure the concentration of protein in the flow-through as well as the starting protein concentration. The coupling efficiency was calculated by subtracting the amount of flow-through protein from that of the starting protein to find the amount of protein immobilized, then dividing that number by the amount of starting protein. The coupling efficiency was 83.7%, with a protein bead concentration of 178 uM.

E.18 Protein-bead reactions

In triplicate, 25.6 uL of beads (178 uM protein) was added to a 2 mL crimp vial with 399.4 uL of M9-N to a total volume of 425 uL. Under anaerobic conditions, 50 uL of 100 mM sodium dithionite in M9-N, and then 25 uL of 1 M styrene in ethanol, was added to each vial. After shaking each vial briefly, 6.25 uL of 8 M ethyl diazoacetate was added. The final reaction conditions were 9.11 uM protein, 50 mM styrene, 100 mM ethyl diazoacetate, 10 mM dithionite, and 5% ethanol in M9-N. The vials were left to shake at

room temperature at 60 rpm for 90 min, after which 20 uL of 40 mM ethyl phenylacetate and 1 mL of cyclohexane were added. The solution was transferred to a 1.7 mL Eppendorf tube and centrifuged at 1000 g for 1 min. The organic layer was extracted and analyzed via GC.

For bead recovery, residual cyclohexane was first removed from the tubes. The reaction vials were also washed with ~500 uL of M9-N to collect any remaining beads and their solutions were transferred to the tubes. The tubes were centrifuged at 1 g for 1 min, and then the supernatant was removed. The beads were resuspended with 418.8 uL of M9-N, then 139.6 uL of the solution were transferred to each of three new crimp vials. This was repeated three times, so each vial had a volume of 418.8 uL and the beads were evenly distributed. Reactions were then conducted in the conditions described in the previous paragraph. This was repeated until the beads had lost most of their activity, which ended up being four uses.

E.19 Protein-bead heme removal and replacement

A denaturation buffer was prepared with 500 uL acetone, 500 uL of a 20 mM MOPS buffer, pH 7.5, with 400 mM imidazole, and 1 uL of 2 M dithiothreitol. 1 mL of this solution was added to ~100 uL of beads into a 1.7 mL Eppendorf tube. The tube was then shaken at room temperature for 40 min, after which it was centrifuged at 1000xg for 1 min, the supernatant removed, and the denaturation buffer replaced. This denaturation was performed three times, after which the solution was centrifuged at 1000 g for 1 min and the supernatant was removed.

A renaturation buffer was prepared with 25 uL of 1 M imidazole in M9-N, 500 uL of 4 M guanidinium chloride, 1 uL of 2 M dithiothreitol, 424 uL of M9-N, and 50 uL of 1 mM hemin in dimethylsulfoxide. This buffer was added to ~100 uL of beads and the tube was shaken at room temperature for 1 hr. The solution was then centrifuged at 1000 g for 1 min, and the supernatant was removed. Protein-beads were rinsed once with 500 uL of M9-N before use.

E.20 Supporting tables and figures

Table E-1. Amino acid sequences of mutants, relative to wild type P450-BM3 (WT). All mutations listed below are in the heme domain. There are no mutations present in the reductase domain.

Enzyme	Amino acid substitution with respect to WT
BM3-HStar	V78M, L181V, T268A, C400H, L437W
BM3-HStar H92N	V78M, H92N, L181V, T268A, C400H, L437W
BM3-HStar H100N	V78M, H100N, L181V, T268A, C400H, L437W
BM3-HStar H92N H100N	V78M, H92N, H100N, L181V, T268A, C400H, L437W

Table E-2. Total turnover numbers, diastereoselectivities, and enantioselectivities of whole-cell bioconversions of **1** and EDA. Experiments were performed at room temperature using whole cells expressing the BM3-HStar variant, resuspended to $OD_{600} = 10$, with 10 mM **1**, and 20 mM EDA. Results are the average of experiments performed with three independent batches of *E. coli* cultures, each used to perform three independent chemical reactions (total of 9 reactions for each enzyme variant). *ee* = enantiomeric excess.

<i>Variant</i>	<i>TTN</i>	<i>%trans</i>	<i>%ee</i>
BM3-Hstar	7400	91	61
Hstar H92N	10900	91	82
Hstar H100N	12300	95	69
Hstar H92N H100N	20100	96	83

Table E-3. Total turnover numbers, diastereoselectivities, and enantioselectivities of reaction of **1** and EDA with purified enzymes. Results are the average of three independent chemical reactions, performed with 10 μ M enzyme concentration, 10 mM of **1**, 20 mM of EDA, and 10 mM of $Na_2S_2O_4$ at room temperature

<i>Variant</i>	<i>TTN</i>	<i>%trans</i>	<i>%ee</i>
BM3-Hstar	310	92	10
Hstar H92N	360	93	61
Hstar H100N	420	91	57
Hstar H92N H100N	560	98	89

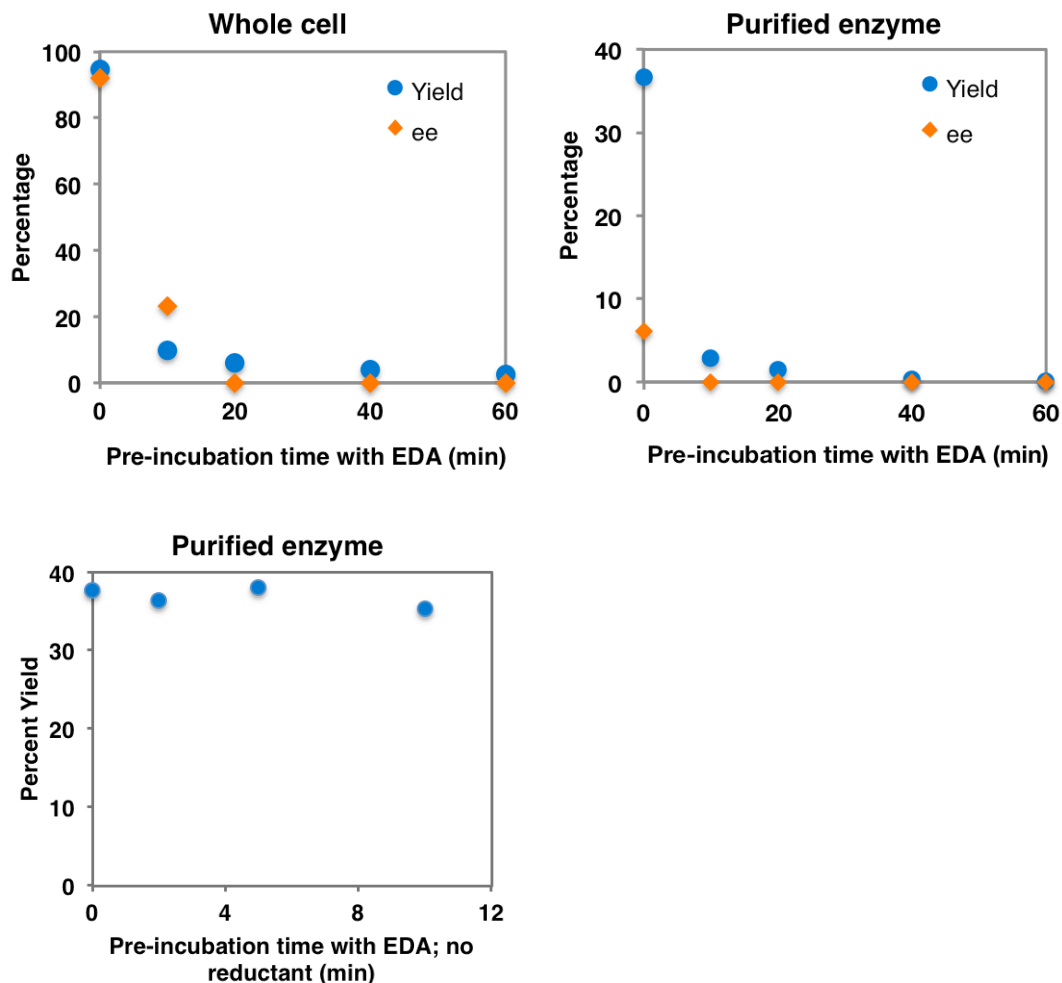


Figure E-1. Pre-incubation experiments with EDA (top left, reaction with whole cells expressing the enzyme; top right, reaction with purified enzyme) showed rapid loss in activity and enantioselectivity, suggesting EDA-induced inactivation. For whole-cell experiments, reactions were performed with whole *E. coli* cell suspension ($OD_{600} = 60$), 10 mM of **1**, and 20 mM of EDA. For experiments with purified enzyme, reactions were performed with purified BM3-HStar (10 μ M), 10 mM of **1**, and 20 mM of EDA. Pre-incubation experiments with EDA in the absence of $Na_2S_2O_4$ (bottom graph) showed that the enzyme maintained its activity in the absence of reductant; reaction conditions: purified BM3-HStar (10 μ M), 10 mM of **1**, and 20 mM of EDA, with $Na_2S_2O_4$ solution added concurrently with **1**.

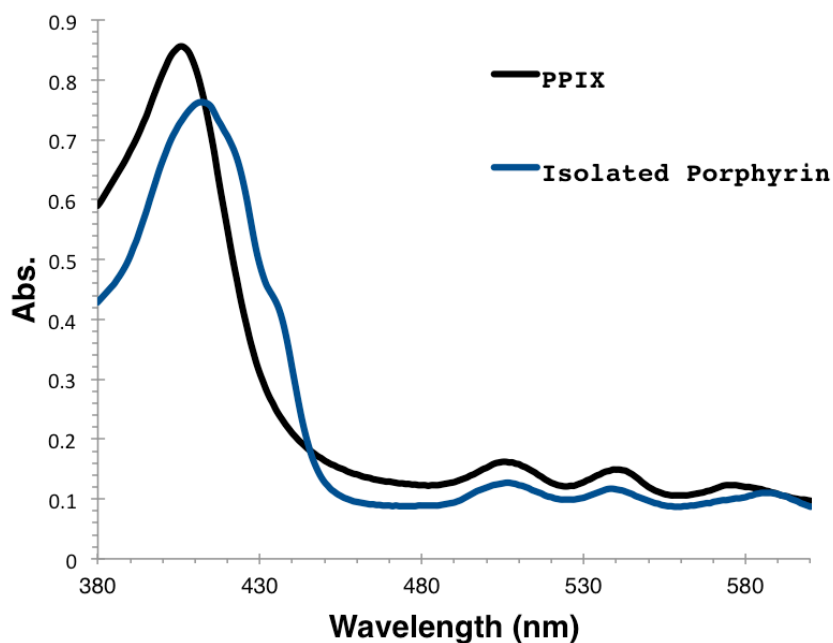


Figure E-2. Comparison of absorption spectra of isolated porphyrin species from BM3-HStar after cyclopropanation reaction with that of authentic protoporphyrin IX (PPIX) standard.

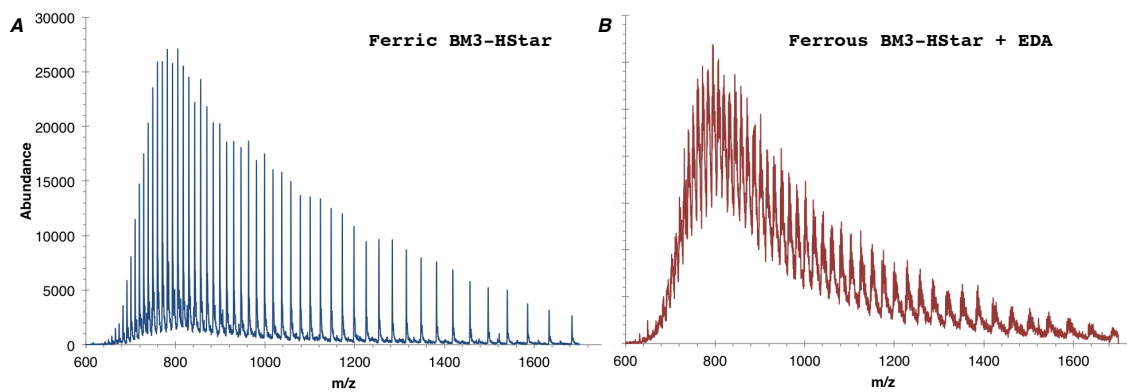


Figure E-3. Whole protein mass spectrometry analysis of BM3-HStar heme domain at the end of reaction with **1** and EDA. Mass spectrum exhibits extreme heterogeneity, which can be deconvoluted to reveal the presence of multiply-modified protein species.

Table E-4. Representative deconvolution of purified, intact protein. Error on the LC-MSD SL 1100 series is approximately 100 ppm, or ± 5.4 Da in a 54000 Da protein.

<i>Expected mass (Da)</i>	<i>Observed mass (Da)</i>	<i>EDA equivalents</i>
53866.49	53871.29	0

Table E-5. Representative deconvolution of EDA-treated, intact protein. Error on the LC-MSD SL 1100 series is approximately 100 ppm, or ± 5.4 Da in a 54000 Da protein.

<i>Expected mass (Da)</i>	<i>Observed mass (Da)</i>	<i>EDA equivalents</i>
53866.49	53870.49	0
53952.59	53952.75	1
54038.69	54039.62	2
54124.79	54125.00	3
54210.89	54212.2	4
54296.99	54300.46	5
54383.09	not detected	6
54469.19	54477.2	7

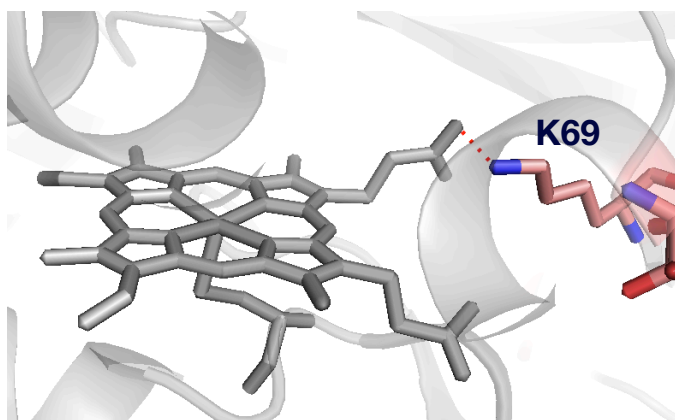


Figure E-4. Polar contact between K69 and the propionate side-chain of the heme in P450-BM3.

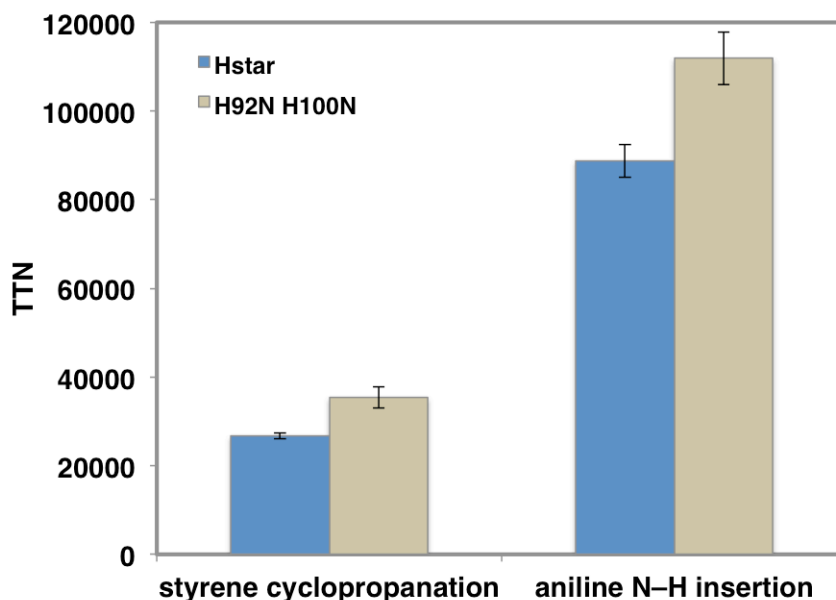


Figure E-5. Comparison of TTNs for styrene cyclopropanation and carbene insertion into the N-H bond of aniline between HStar and HStar H92N H100N. Reaction conditions for cyclopropanation: whole *E. coli* cells expressing the enzyme variant (suspension in M9-N, $OD_{600} = 10$), 10 mM of styrene, 10 mM of EDA, room temperature, 2 h. Reaction conditions for N-H insertion: whole *E. coli* cells expressing the enzyme variant (suspension in M9-N, $OD_{600} = 10$), 20 mM of aniline, 10 mM of EDA, room temperature, 2 h.

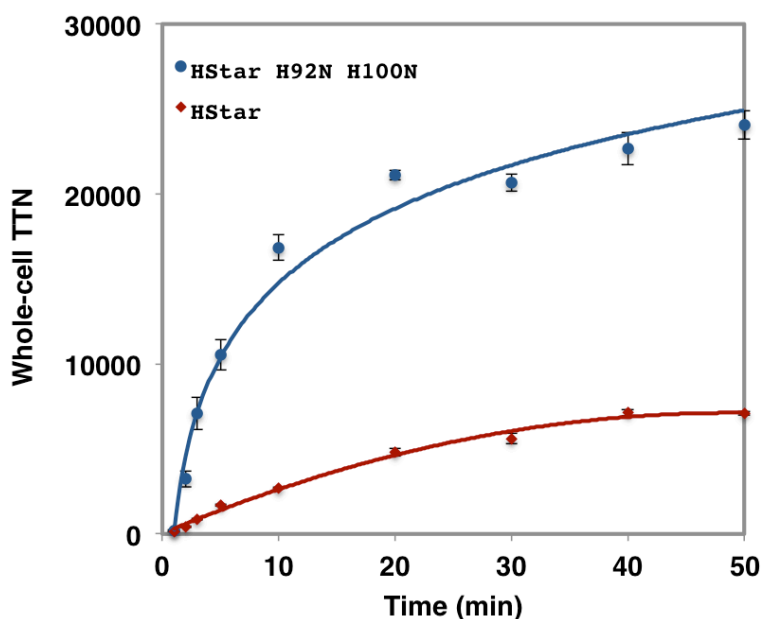


Figure E-6. Time-course of cyclopropanation of **1** with EDA using whole *E. coli* cells expressing BM3-HStar variants. Reaction conditions: whole *E. coli* cells expressing enzyme variants ($OD_{600} = 10$, $0.2 \mu\text{M}$ enzyme), 10 mM of **1**, 20 mM of EDA, room temperature.

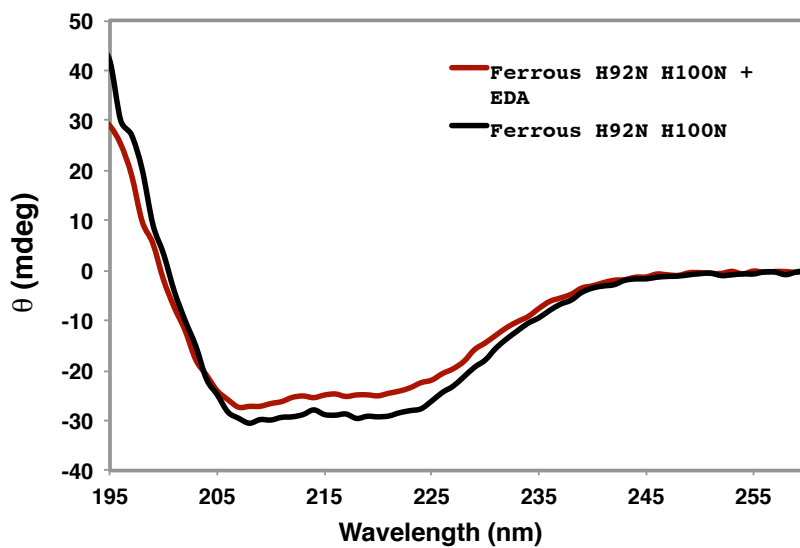


Figure E-7. Circular dichroism (CD) spectra of ferrous BM3-HStar H92N H100N before (black) and after reaction with EDA (red).

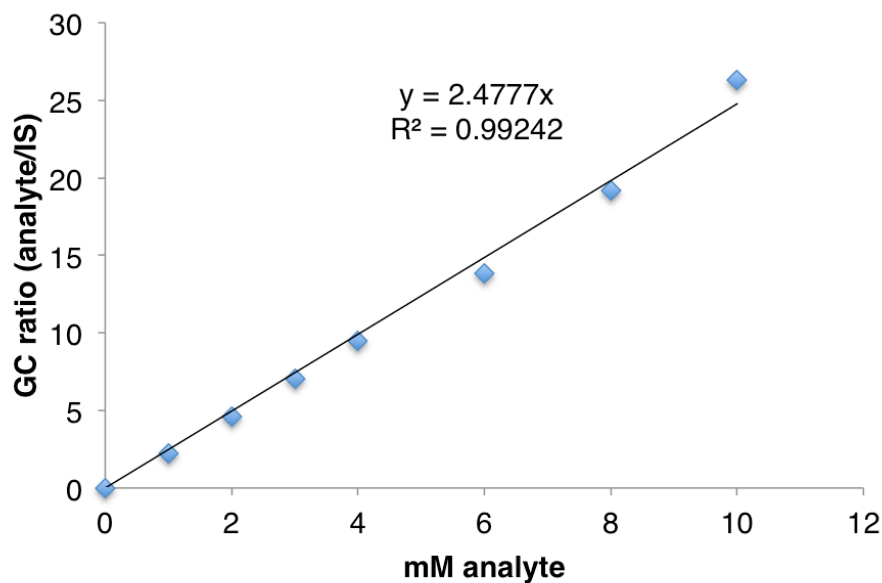


Figure E-8. GC calibration curve for the reaction product between aniline and EDA.

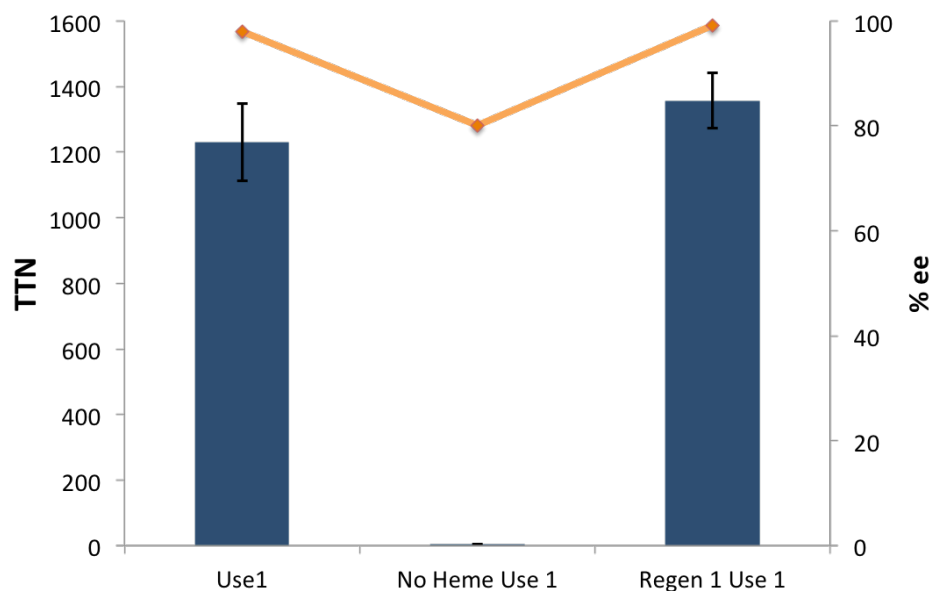


Figure E-9. Control reactions with immobilized *Bsu* TrHb LAA show that almost all activity is lost in the denaturation process, and that all activity is regained in regeneration.

E.21 References

1. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
2. Wang, Z. J. *et al.* Improved cyclopropanation activity of histidine-ligated cytochrome P450 enables the enantioselective formal synthesis of levomilnacipran. *Angew. Chem. Int. Ed.* **53**, 6810–6813 (2014).
3. Hernandez, K. E. *et al.* Highly stereoselective biocatalytic synthesis of key cyclopropane intermediate to ticagrelor. *ACS Catal.* 7810–7813 (2016). doi:10.1021/acscatal.6b02550
4. 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.* **136**, 15505–15508 (2014).