NON-NATIVE CHEMISTRY OF METALLOENZYMES

Thesis by

Nathaniel Wood Goldberg

In Partial Fulfillment of the Requirements for the

Degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2022

(Defended February 16, 2022)

Nathaniel Wood Goldberg
ORCID: 0000-0003-0201-8856

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Prof. Frances Arnold, for her leadership and advice during my Ph.D. Proteins and enzymes weren't really even on my radar until I arrived at Caltech, so I am glad she was willing to take a chance on a naive first-year chemistry student. I am grateful that she has given me the freedom to pursue ambitious and challenging problems, while always making sure that I am asking the right questions. I would also like to thank the other members of my committee, Profs. Jonas Peters, Sarah Reisman, and Bil Clemons, for their support and guidance during my time here.

The Arnold lab has been my scientific home for the last five and a half years, and it would not have been such a supportive and welcoming environment without the contributions of all group members, past and present. I want to thank our fearless lab manager, Sabine Brinkmann-Chen for her expert stewardship of the lab, and for giving me a creative outlet by letting me (attempt to) tinker with and repair most of our lab equipment. Xiongyi Huang, Kai Chen, and Jenny Kan were my first mentors in the lab, and they got me off on the right foot while introducing me to the world of biocatalysis. Anders Knight is the person I have worked most closely with during my graduate work, and I am grateful to him both for teaching me everything I know about biology and also for having my back when I needed it most. Kelly Zhang, my scientific role model, is one of the most creative people I have ever met. Ella Watkins-Dulaney and I joined the Arnold lab at the same time and came up through the ranks together, and I am grateful to have been on this journey with someone as talented and thoughtful as she is. For friendship, support, and advice more generally, I would like to thank (non-exhaustively and in no particular order) Rusty Lewis, Zach Wu, Kari Hernandez, Bruce Wittmann, Patrick Almhjell, Kadina Johnston, David Romney, Tina Boville, Dave Miller, and Nick Porter.

My chemistry career did not begin at Caltech, so I would like to thank my high school teachers, Kari Newman and William Personette, for introducing me to this world and setting me off on my path. I also want to thank my undergraduate professors, especially my research

advisor Prof. Tobias Ritter, for inspiring me to pursue graduate school and a career in research. Finally, I would not be where I am today without the love and support of my family: my parents, Ken Goldberg and Sarah Wood, who have always believed in me, and my brothers Sam and Jacob, who always inspire me to do my best.

ABSTRACT

Metalloenzymes are important catalysts in biochemistry, but the scope of their naturally occurring activities is dwarfed by the range of chemistry achieved by synthetic transition-metal catalysts. To date, efforts to expand the catalytic repertoire of metalloproteins beyond their native activities have focused almost exclusively on heme-binding proteins, which have been engineered to catalyze a wide variety of carbene- and nitrene-transfer chemistry. Heme-binding proteins represent only a limited subset of the vast diversity of metalloproteins that exists in Nature, and the non-native chemistry of the rest of the metalloproteome remains largely unexplored. This thesis details the discovery and engineering of non-native catalytic abilities of non-heme metalloproteins. Chapter 1 introduces metalloproteins as biocatalysts in synthetic chemistry, and various approaches to expand their catalytic activities. Chapter 2 describes efforts towards enzyme-catalyzed hydrosilylation, including the curation and development of a diverse library of non-heme metalloproteins. In Chapter 3, a non-heme iron-dependent dioxygenase (*Pseudomonas savastanoi* ethylene-forming enzyme, *Ps*EFE) is found to catalyze nitrene-transfer chemistry, and is engineered by directed evolution to improve this non-native activity. The nitrene transfer activity and selectivity of *Ps*EFE can be modulated by small-molecule metal-coordinating ligands. Chapter 4 describes the discovery and development of a *Ps*EFE-catalyzed olefin aminoarylation reaction, a previously unknown reaction of sulfonyl azides and olefins. This reaction is unprecedented in the existing chemical literature, and displays a number of unusual mechanistic features. Together, the work described here represents the expansion of non-native chemistry to a new class of metalloenzymes, enabling the discovery of previously unknown catalytic activities.

PUBLISHED CONTENT AND CONTRIBUTIONS

† denotes equal contribution

Goldberg, N. W.†; Knight, A. M.†; Zhang, R. K.; Arnold, F. H. Nitrene Transfer Catalyzed by a Non-Heme Iron Enzyme and Enhanced by Non-Native Small-Molecule Ligands. *J. Am. Chem. Soc.* **2019**, *141*, 19585–19588. DOI:10.1021/jacs.9b11608

N.W.G. discovered initial activity for aziridination and C–H insertion. N.W.G. and A.M.K. performed directed evolution. N.W.G. performed chemical synthesis. N.W.G. and A.M.K. wrote the manuscript with input from all co-authors.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

METALLOPROTEINS IN BIOCATALYSIS

Abstract

Metalloenzymes are integral to natural biochemistry, catalyzing some of the most important reactions of life. Metalloproteins have also found use in synthetic biocatalysis, due largely to their remarkable selectivities. However, the synthetic utility of natural metalloenzymes is restricted by the limited scope of biologically accessible chemistry, especially compared to the catalytic repertoire of synthetic transition metal catalysts. A number of approaches have been developed to expand the reaction scope of metalloenzymes beyond what exists in nature. Artificial metalloenzymes incorporate a synthetic transition metal complex into a protein scaffold, combining the activities achievable with synthetic catalysts with the selectivities enabled by proteins. A complementary approach takes advantage of the non-natural catalytic abilities of naturally occurring metallocofactors and metalloproteins, using reagents rarely found or absent in nature. Many natural and engineered heme-binding proteins have been developed in this fashion, catalyzing a range of nitrene- and carbene-transfer reactions. Heme-binding proteins represent only a small fraction of the diversity of metalloproteins found in nature, however, and the non-natural catalytic abilities of non-heme metalloproteins have scarcely been explored.

1.1 Metalloproteins in natural biochemistry

In the broadest sense, metalloproteins encompass all proteins that contain a metal, accounting for at least 30% of all proteins in nature[1]. These metals play a variety of roles, including structural stabilization, metal ion transport and regulation, and Lewis acid and redox catalysis. This chapter will focus primarily on transition metal-containing metalloproteins, and their catalytic activities.

Metal-binding sites in proteins

Proteins coordinate metals in a variety of metal-binding sites. The simplest coordination mode is direct protein-metal coordination, in which a metal ion is coordinated by the polypeptide chain. In principle, nearly any heteroatom-containing amino acid side chain can coordinate a metal, but among the most common coordination modes are coordination at nitrogen by histidine, at oxygen by aspartate or glutamate, or at sulfur by cysteine (Figure 1-1A)[2]. Coordination by other amino acid side chains is also found, including at sulfur by methionine, at oxygen by tyrosine, and at oxygen by asparagine or glutamine (Figure 1-1B). The peptide main chain can also coordinate a metal, by a carbonyl oxygen, amide nitrogen, or *N*-terminal amine (Figure 1-1C).

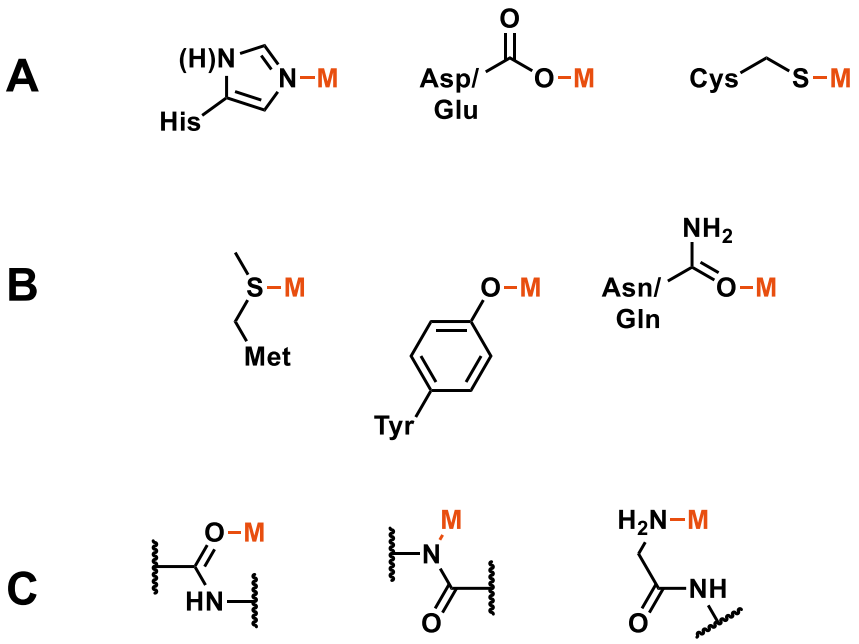


Figure 1-1. Examples of direct polypeptide metal coordination. (A) Common side-chain interactions, with histidine, aspartate/glutamate, and cysteine. (B) Less common side-chain interactions with other heteroatom-containing side chains. (C) Coordination by the peptide main chain.

Direct protein-metal coordination allows for a range of metal-binding sites, but is inherently limited by the fixed set of 20 canonical amino acids. To access a broader diversity of metal-binding sites, Nature has evolved metal complex cofactors, in which a protein-bound metal is coordinated by ligands other than the peptide chain. Some examples of natural metal complex cofactors are shown in Figure 1-2, including heme[3] (discussed in more detail in section 1.4), the iron/molybdenum cofactor in nitrogenase responsible for biological nitrogen fixation[4], the iron/nickel cofactor of NiFe hydrogenase[5], and the nickel cofactor of lactate racemase[6], featuring a biologically unusual carbon-metal bond.

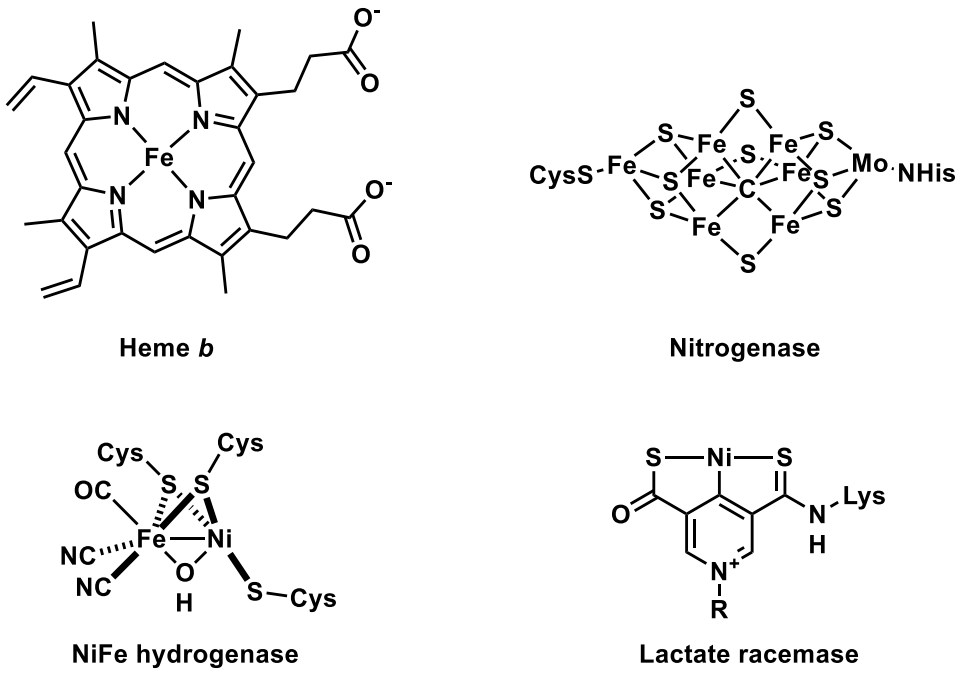


Figure 1-2. Examples of metal complex cofactors found in natural metalloenzymes.

Metalloenzymes in biochemistry

Primary metabolism

Metalloenzymes play a variety of roles in natural biochemistry, including in some of the most fundamental chemistry of life. A few examples are shown in Figure 1-3. Deoxyribonucleotides, the building blocks of DNA, are biosynthesized by ribonucleotide reductase (RNR). Several classes of RNR exist, but almost all are dependent on transition metals, including iron, manganese, and cobalt[7]. Biological nitrogen fixation, responsible for essentially all terrestrial nitrogen fixation until the 20th century, is catalyzed by nitrogenase, typically an iron and molybdenum-dependent enzyme[4]. Photosynthesis, responsible for nearly all biological carbon fixation, relies on a number of different enzymes, including several metalloenzymes, binding iron, manganese, and copper[8].

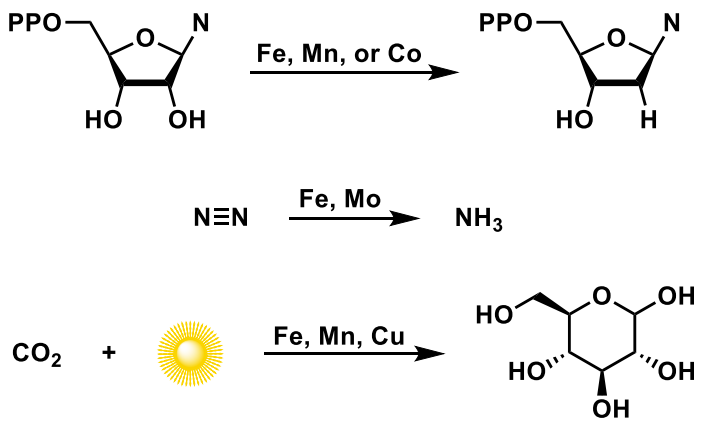


Figure 1-3. Examples of primary metabolism catalyzed by metalloenzymes: Synthesis of deoxyribonucleotides by ribonucleotide reductase; nitrogen fixation by nitrogenase; photosynthesis.

Secondary metabolism

Metalloenzymes, especially oxygenases[9,10], are also widely involved in secondary metabolism in Nature. The biosynthesis of many natural products, such as the bacterial polyketide antibiotic erythromycin[11] and the plant-derived terpenoid paclitaxel[12], include highly specific tailoring enzymes (Figure 1-4). In human biology, cytochromes P450 are responsible for approximately 75% of drug metabolism and degradation[13].

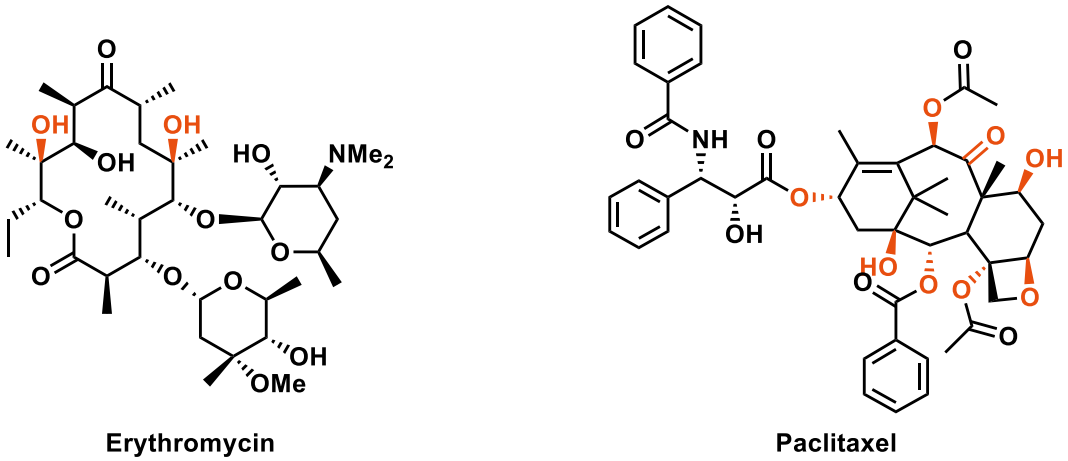


Figure 1-4. Examples of secondary metabolite natural products, with metalloenzyme-installed functionalities highlighted.

1.2 Native/near-native metalloprotein chemistry in biocatalysis

Some of the native activities of metalloenzymes, in particular those involved in secondary metabolite biosynthesis, are well-suited to address some of the challenges of human synthetic chemistry.

Industrial biocatalysis with metalloenzymes: nitrile hydratase

The largest-scale industrial biocatalytic process is catalyzed by a metalloenzyme. The hydration of acrylonitrile to acrylamide catalyzed by nitrile hydratase (Figure 1-5), an iron- or cobalt-dependent enzyme[14], is performed on a scale of hundreds of thousands of tons per year. Compared to previously employed chemical methods, the use of nitrile hydratase avoids over-hydrolysis to the carboxylate and suppresses polymerization. Nitrile hydratase is also used to catalyze the production of thousands of tons per year of nicotinamide, used in agricultural animal feed[15]. The nitrile hydratase active site features the catalytic metal bound by three cysteine residues, of which two are post-translationally modified to a sulfenate and sulfinate, and two backbone amide nitrogens. The sixth coordination site is where the nitrile substrate binds[16].

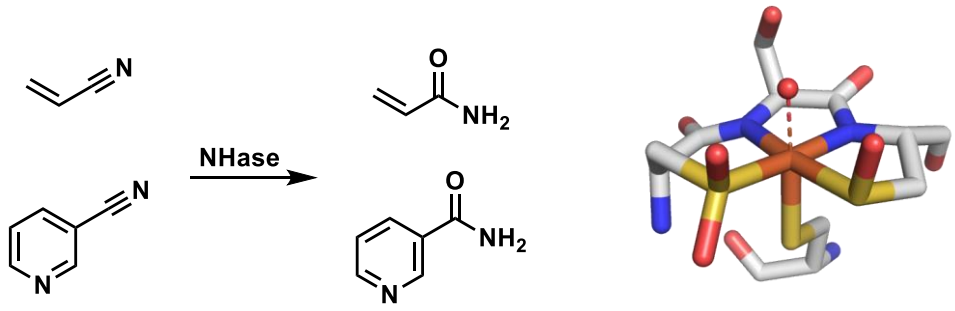


Figure 1-5. Production of acrylamide and nicotinamide catalyzed by nitrile hydratase, with a structural depiction of the active site with a water molecule bound at the substrate-binding site (PDB ID: 3A8O).

C–H functionalization

On the laboratory scale, synthetic chemists have been drawn to metalloprotein biocatalysts for their ability to perform C–H functionalization reactions with exquisite site- and stereoselectivity.

Hydroxylation with cytochrome P450

The cytochrome P450 family of enzymes are attractive catalysts for selective C–H hydroxylation. To this end, cytochrome P450 variants have been engineered for late-stage functionalization of complex molecules, such as steroids[17] or the antimalarial drug artemisinin[18], as well as for valorization of simple hydrocarbon feedstock chemicals[19] (Figure 1-6).

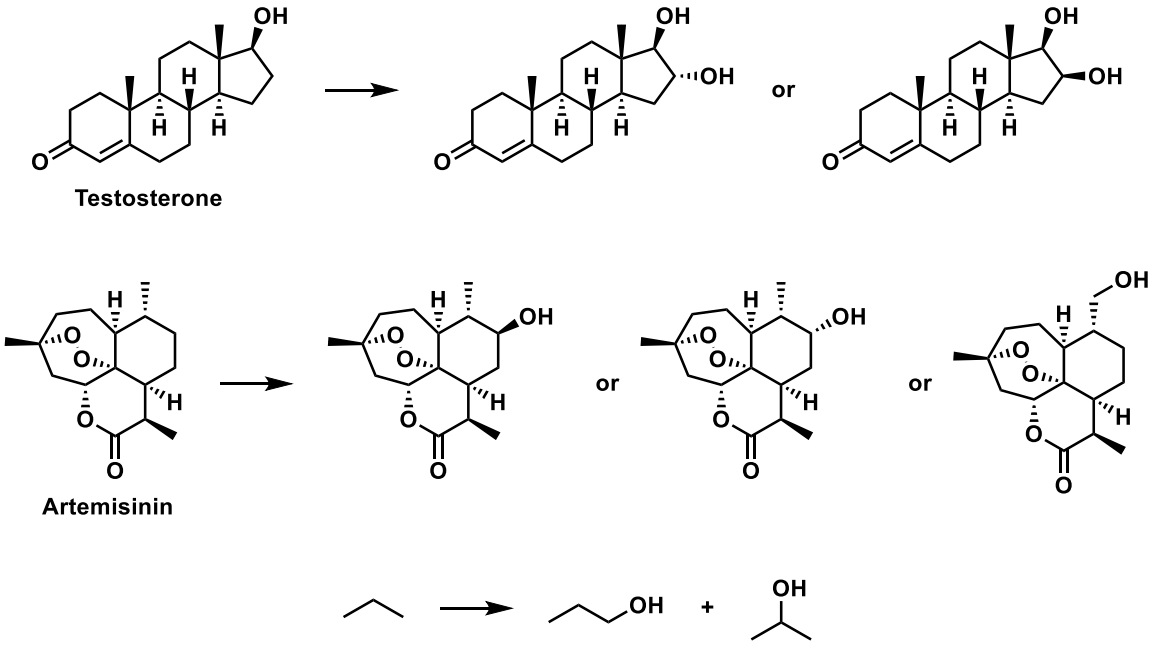


Figure 1-6. Examples of biocatalytic C–H hydroxylation reactions catalyzed by engineered variants of cytochrome $P450_{BM3}$.

Halogenations

Some non-heme iron dioxygenases display C–H halogenation activity, analogous to the C–H hydroxylation more typically observed. The Chang group has discovered a family of

halogenases capable of site-selective chlorination of various α-amino acids[20], and the Buller group has engineered variants of the enzyme WelO5 to perform regioselective late-stage chlorination of natural products, including the alkaloid martinelline[21] and the polyketide soraphen family[22] (Figure 1-7).

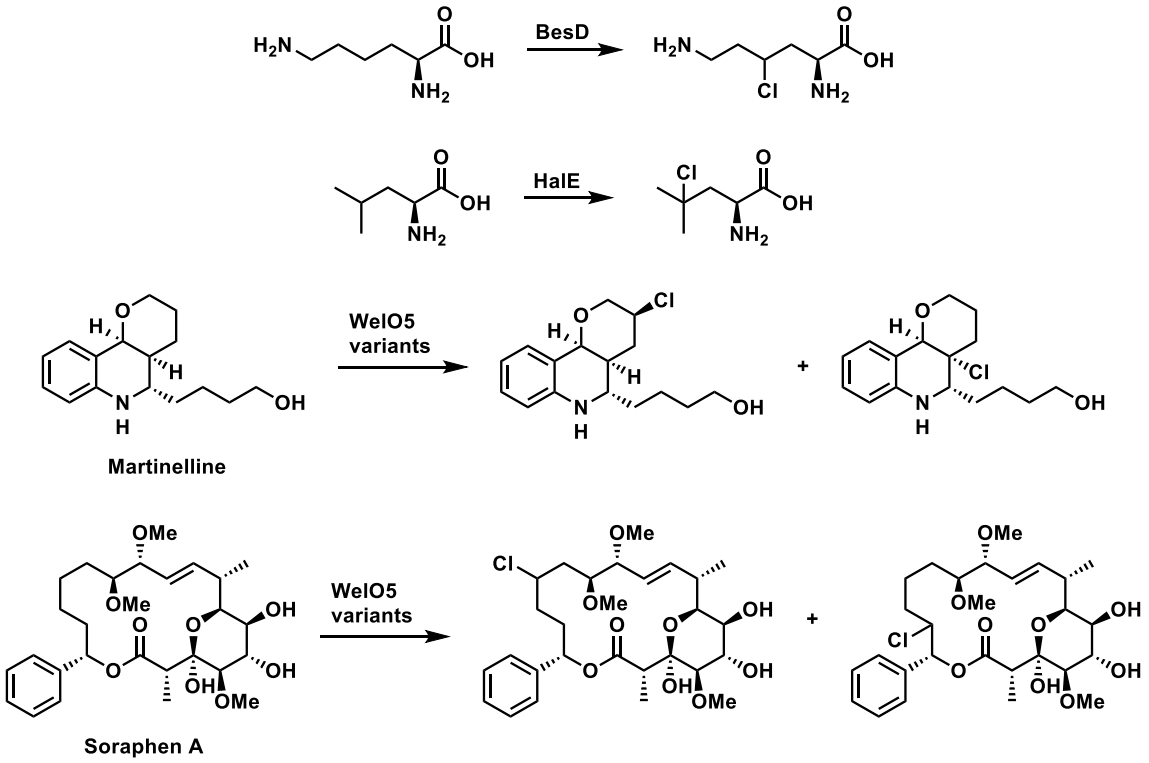


Figure 1-7. Examples of biocatalytic C–H halogenation.

Applications of metalloenzymes in total synthesis

With their remarkable selectivities, often superior to those of conventional small-molecule catalysts, metalloenzyme biocatalysts have been employed in the total synthesis of natural products. For example, the Stoltz group employed an engineered cytochrome P450 to effect a late-stage oxidation in their synthesis of nigelladine A[23], and the Renata group used a non-heme iron dioxygenase GriE to perform a stereoselective hydroxylation of L-leucine *en route* to manzacidin C[24] (Figure 1-8).

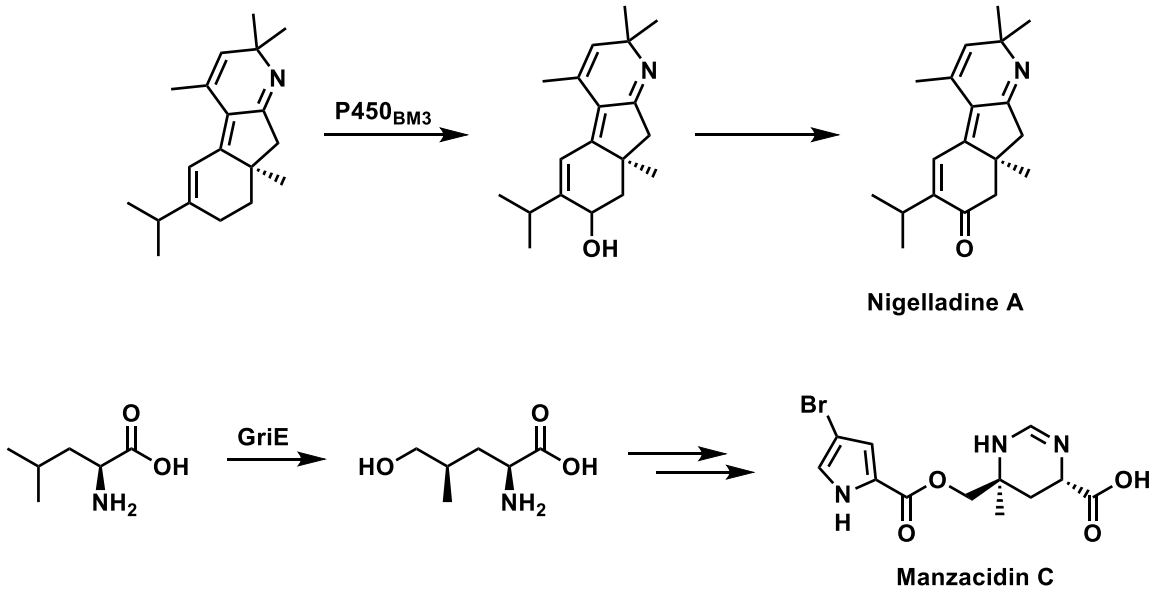


Figure 1-8. Examples of metalloenzyme-catalyzed C–H oxidations in natural product total syntheses.

1.3 Artificial metalloenzymes

While Nature's metalloenzymes can offer remarkably selective transformations, their catalytic scope is limited compared to that of synthetic transition metal complexes. In efforts to combine the activities of synthetic catalysts with the selectivities of enzymes, chemists have developed "artificial metalloenzymes," broadly defined as a synthetic metal complex inserted or attached to a protein scaffold[25,26]. Artificial metalloenzymes have been to expand the reactive scope of metalloenzymes beyond naturally occurring chemistry, though in many cases the reactivity is mostly or entirely due to the synthetic metallocofactor, with the protein serving simply as a large chiral ligand to impart stereoselectivity on the reaction. A number of techniques have been employed to create artificial metalloenzymes; some of the most widely used are discussed below.

Non-covalent binding

One method for formation of artificial metalloenzymes is conjugation of the metallocofactor by non-covalent binding to the protein, commonly using biotin/(strept)avidin binding. One

of the first reported artificial metalloenzymes, by Whitesides and coworkers in 1978, used this strategy to create an asymmetric rhodium-based hydrogenation catalyst[27]. Subsequent work in this area has led to the development of artificial metalloenzymes that catalyze a number of reactions uncommon or entirely unknown in nature, including iridium-catalyzed asymmetric transfer hydrogenation[28], ruthenium-catalyzed olefin metathesis[29], and asymmetric palladium-catalyzed Suzuki aryl-aryl cross-coupling[30] (Figure 1-9).

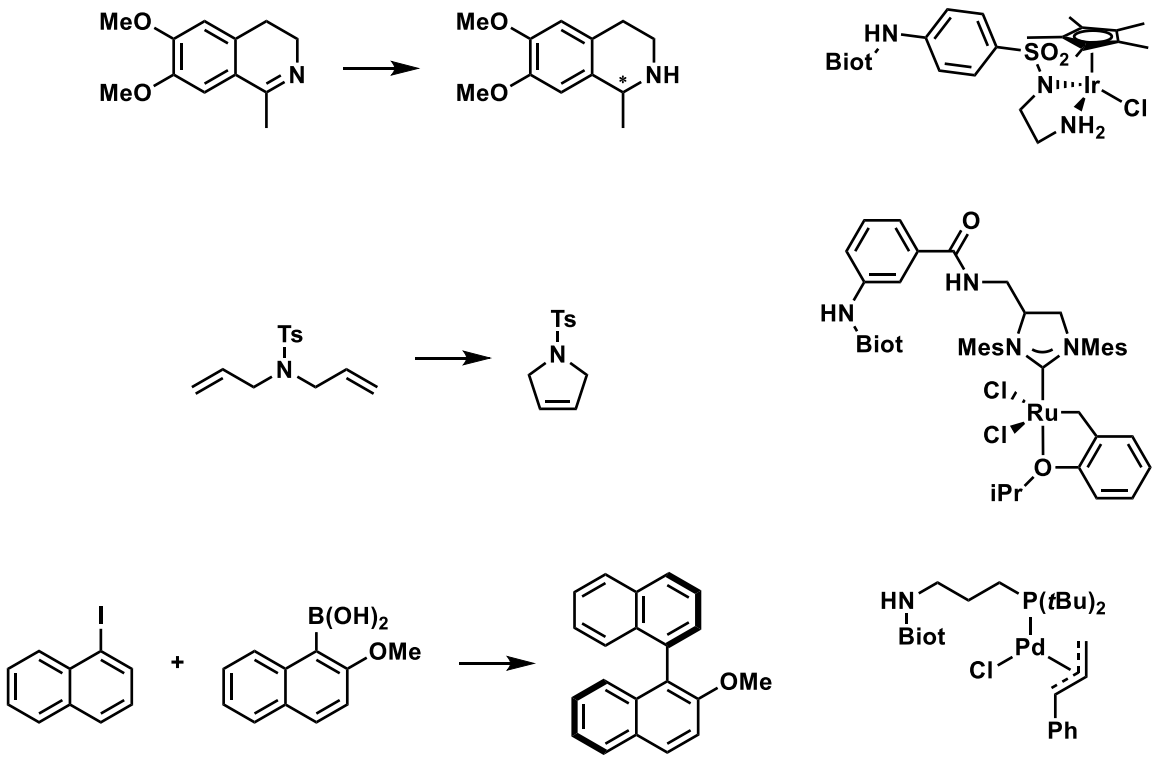


Figure 1-9. Examples of reactions catalyzed by (strept)avidin-based artificial metalloenzymes, and their corresponding synthetic metallocofactors. Biot = biotin.

Covalent conjugation

An alternative approach for the creation of artificial metalloenzymes is the use of covalent conjugation of the metallocofactor, using conventional bioconjugation techniques such as electrophilic addition to cysteines. This method enables the use of a much larger variety of protein scaffolds than non-covalent binding, as it can be used with proteins that do not have a known high-affinity small-molecule ligand. Examples of covalently conjugated artificial

metalloenzymes include a metathesis enzyme derived from an α-bromoacetamide-tethered ruthenium complex[31], a hydroformylation enzyme derived from a maleimide-tethered rhodium complex[32] (Figure 1-10), and an enantioselective Diels-Alderase derived from an α-bromoacetamide-tethered copper complex[33] (Figure 1-10).

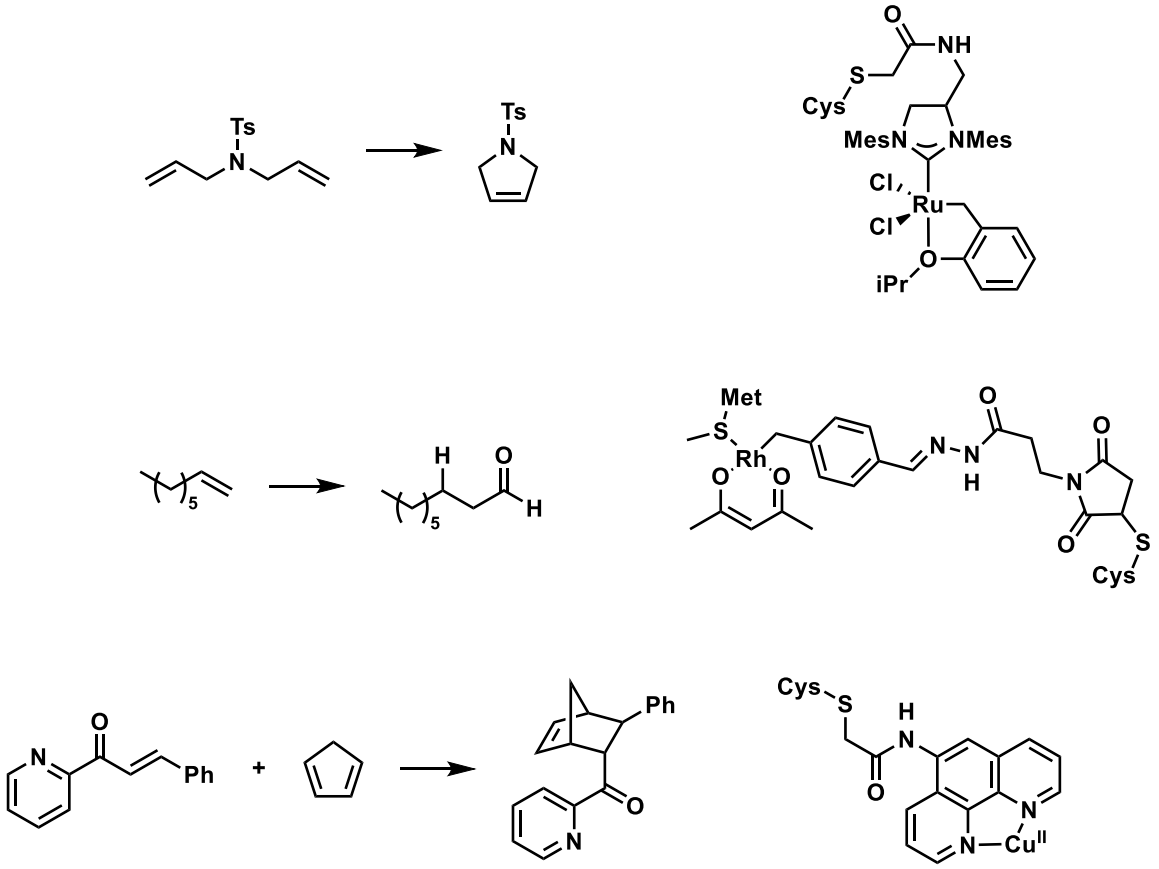


Figure 1-10. Examples of reactions catalyzed by covalently conjugated artificial metalloenzymes, and their corresponding metallocofactors.

Non-canonical amino acid-tethered artificial metalloenzymes

Site-specific non-canonical amino acid (ncAA) incorporation has been used to generate artificial metalloenzymes, either through direct coordination of the metal by the ncAA side chain, or by bioorthogonal conjugation (commonly azide/alkyne [3+2] cycloaddition)[34]. Examples include an asymmetric hydratase catalyzed by a bipyridine-containing ncAA-

copper complex[35] and olefin cyclopropanation catalyzed by an azide-tethered rhodium complex bound to an alkyne-containing ncAA[36] (Figure 1-11).

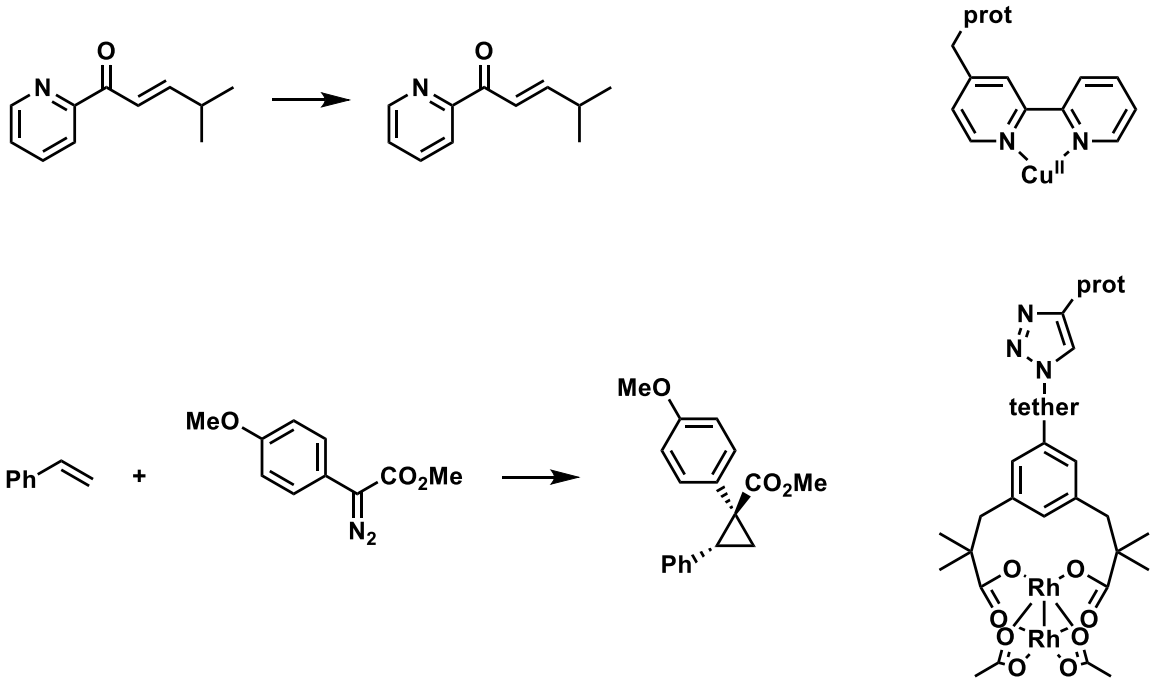


Figure 1-11. Examples of reactions catalyzed by ncAA-containing artificial metalloenzymes, and their metal-binding sites. Prot = protein.

1.4 Non-native chemistry with heme-binding proteins

As an alternative to the use of artificial metalloenzymes, researchers have recently turned to non-natural catalytic activities of naturally occurring metallocofactors, primarily using heme-binding proteins. This approach enables enzymatic catalysis of abiological reactions, but without requiring the synthesis and conjugation of an abiological metal complex. Heme-binding proteins play a variety of roles in biology, but perhaps their best-known catalytic roles are as oxygenases, as exemplified by the cytochrome P450 family. Cytochromes P450 bind dioxygen to form a high-valent iron oxo (Compound I), which can then react with a variety of substrates, such as addition to an olefin to form an epoxide, or insertion into a C–H bond in a hydroxylation reaction (Figure 1-12)[9]. Transition metal-porphyrin complexes (including iron-porphyrin complexes) are also well-known catalysts in the synthetic

literature, including as oxygenation catalysts[37], but also as carbene-[38] and nitrene-transfer[39] catalysts, among other roles. Over the last decade or so, researchers have found great success in merging the activities known with synthetic metal-porphyrin complexes with Nature's metal-porphyrin complexes, heme-binding proteins[40]. Notably, and in contrast to many artificial metalloenzymes, heme-binding proteins engineered for these non-natural activities are often substantially more active than unbound heme, and can catalyze reactions for which heme alone has no activity at all.

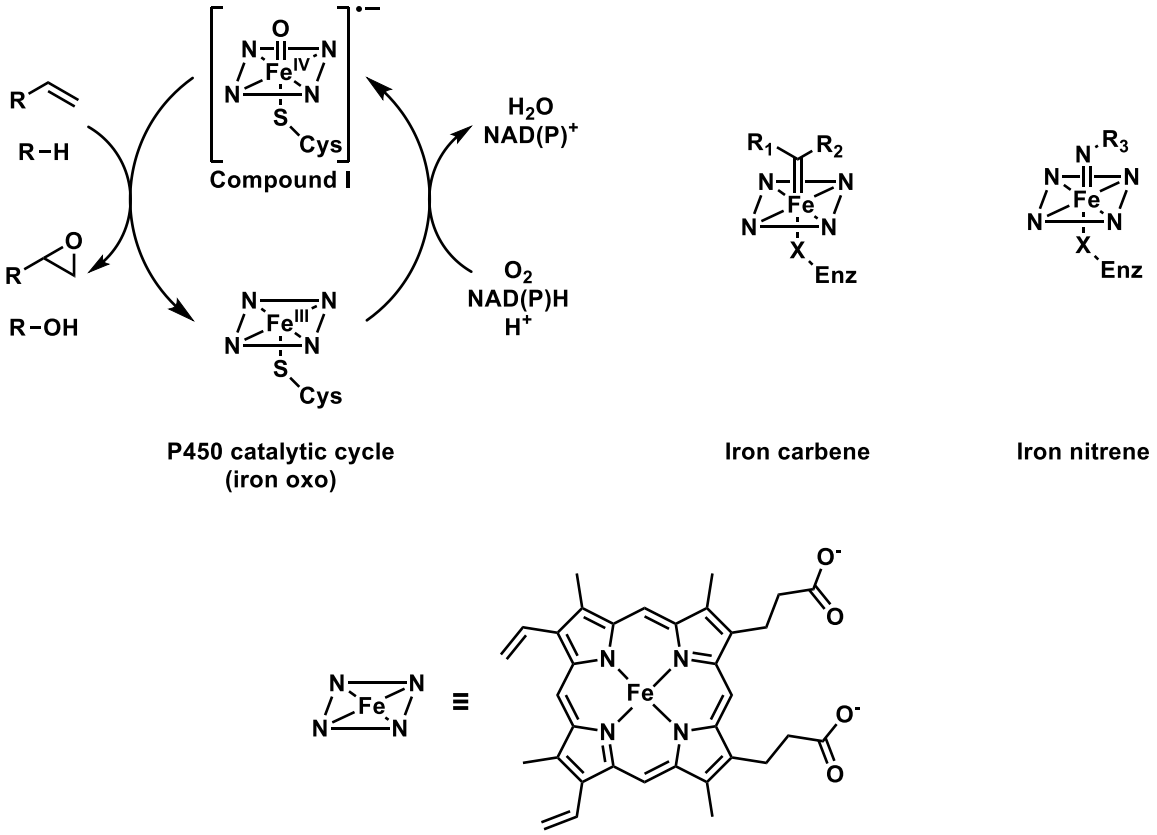


Figure 1-12. Comparison of native cytochrome P450 chemistry (left), proceeding through an iron oxo intermediate, with non-native iron carbene (center) and iron nitrene (right) intermediates. For visual clarity, the heme structure has been abbreviated.

Carbene transfer

The carbon analogue of the iron oxo is an iron carbene. Enzymatic carbene transfer reactions are typically performed using diazo compounds as carbene precursors.

Cyclopropanation

Addition of a carbene unit to an olefin, analogous to epoxidation, results in a cyclopropane. The enzyme-catalyzed cyclopropanation reaction of styrene derivatives and ethyl diazoacetate (EDA) was first reported by the Arnold group in 2013 using cytochrome P450 variants[41], followed by the Fasan group in 2015 with variants of myoglobin[42], a heme-binding protein with no native catalytic function (Figure 1-13A). Further work in this area has included the cyclopropanation of unactivated and electron-deficient olefins, with a panel of stereodivergent enzymes able to selectively access all possible stereoisomers of the cyclopropane product (Figure 1-13B)[43], as well as intramolecular cyclopropanation of diazoacetate esters and diazoacetamides to form bicyclic lactones and lactams, respectively (Figure 1-13C)[44,45].

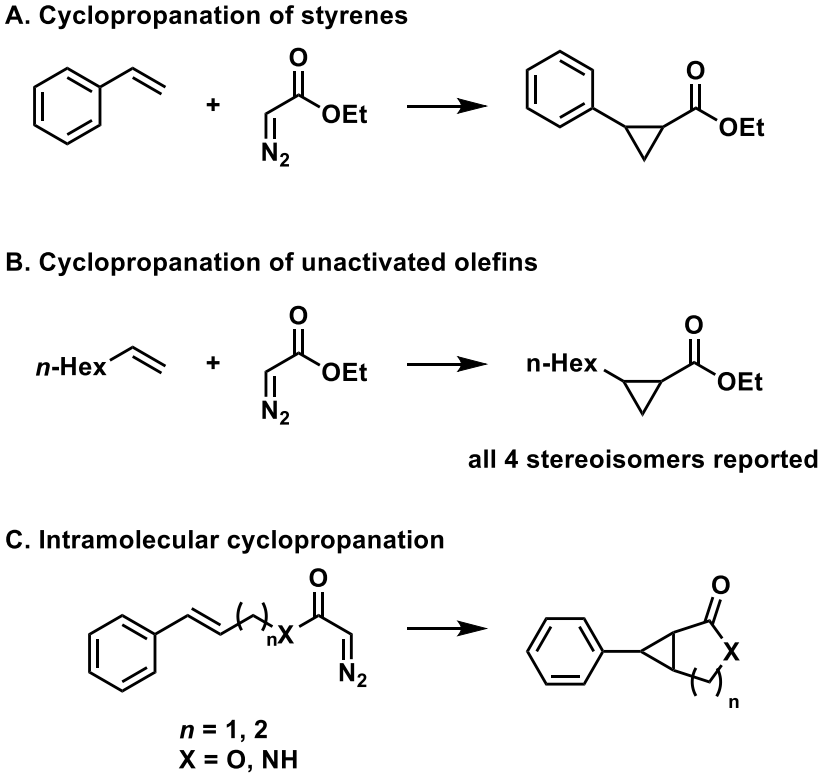


Figure 1-13. Examples of reported cyclopropanations catalyzed by heme-binding proteins.

In addition to olefins, alkynes have also been reported as substrates for enzymatic carbene transfer from EDA catalyzed by cytochrome P450 variants. Addition of a single carbene to an alkyne results in a cyclopropene[46], while addition of a second carbene unit yields a bicyclobutane (Figure 1-14)[47]. These products are highly strained, considerably more so than

an ordinary cyclopropane, making their synthesis by conventional small-molecule catalysts, particularly in a stereoselective fashion, quite difficult.

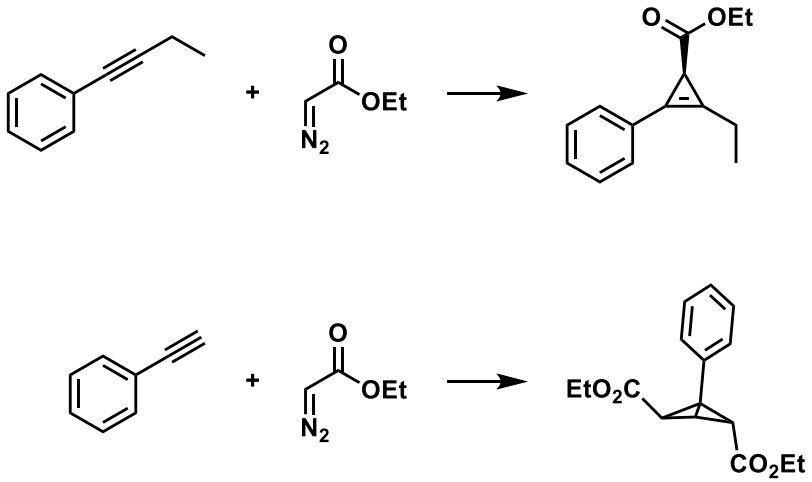


Figure 1-14. Carbene transfer to alkynes: cyclopropenation and bicyclobutanation.

C–H alkylation

An enzyme-bound carbene can insert into a C–H bond, analogous to the native hydroxylation activity of cytochromes P450. Site-selectivity and stereoselectivity are often challenging for small-molecule C–H functionalization catalysts, as most molecules contain many C–H bonds, but enzymes are well-suited to address these challenge. Indeed, several enzyme-catalyzed C–H alkylation reactions by carbene transfer have been reported, including with EDA[48], perfluoroalkyl-substituted carbenes[49], and lactone-derived carbenes[50] (Figure 1-15).

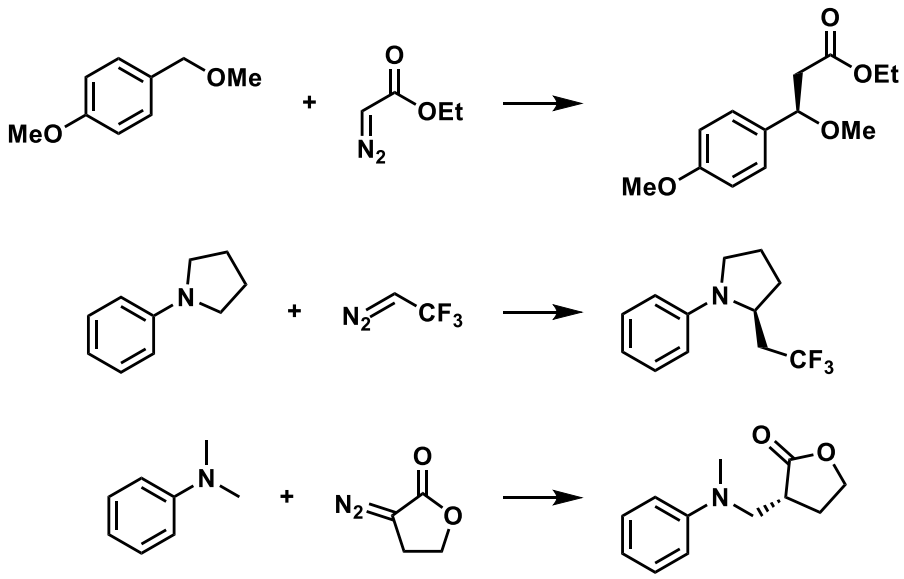


Figure 1-15. Examples of reported carbene C–H insertion reactions catalyzed by heme-binding proteins.

Other X–H insertions

Enzymatic carbene insertions are not limited to C–H bonds; variants of cytochrome P450, cytochrome *c*, and myoglobin have been reported to catalyze carbene insertions into other X–H bonds as well, including Si–H bonds[51], B–H bonds[52], N–H bonds[53], and S–H bonds[54] (Figure 1-16).

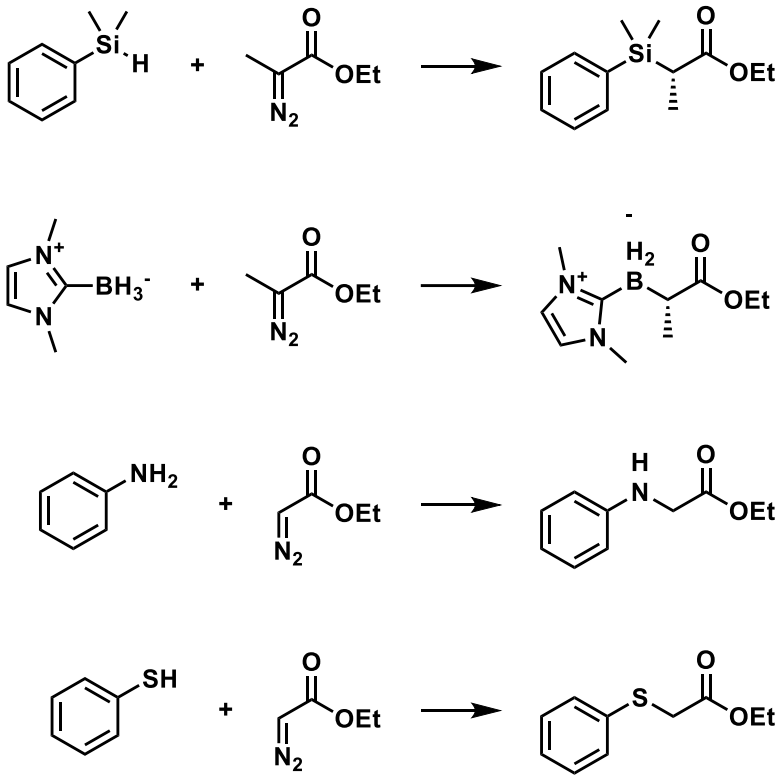


Figure 1-16. Examples of Si–H, B–H, N–H, and S–H carbene insertion reactions catalyzed by heme-binding proteins.

Nitrene transfer

The nitrogen analogue of a carbene is a nitrene. Commonly used nitrene precursors in biocatalysis include organoazides and hydroxylamine derivatives. A cytochrome P450 was recently discovered that natively catalyzes an aziridination reaction by nitrene transfer in a secondary metabolite biosynthetic pathway[55], so nitrene transfer is not strictly speaking "new-to-nature." Nevertheless, nitrene transfer is still quite rare in natural biochemistry, and the vast majority of heme-binding proteins have no native nitrene-transfer role.

Aziridination

Nitrene transfer to a carbon-carbon double bond results in an aziridine product, in the nitrene equivalent of cyclopropanation. Cytochrome P450 variants have been reported to catalyze the aziridination of styrene derivatives with *p*-toluenesulfonyl azide[56] (Figure 1-17A). When

an unprotected nitrene precursor is used, catalyzed by cytochrome *c* variants, the resulting aziridine product undergoes stereospecific hydrolysis to yield an aminohydroxylated product[57] (Figure 1-17B).

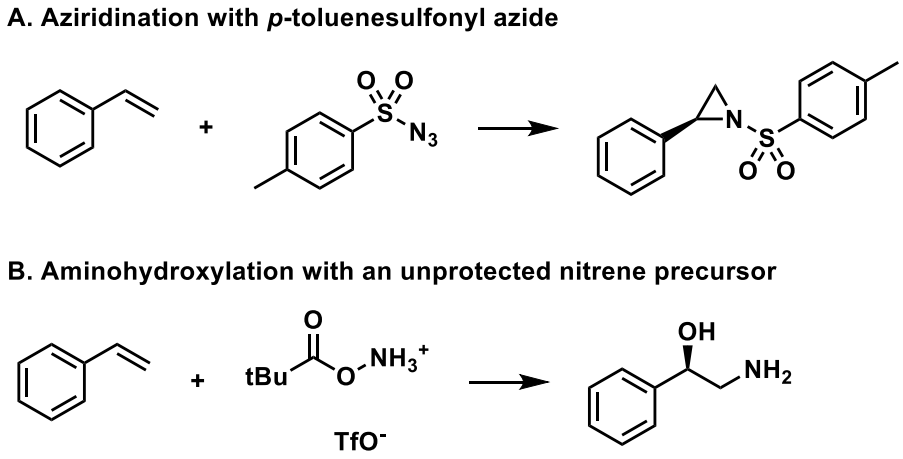


Figure 1-17. Aziridination reactions catalyzed by heme-binding proteins.

C–H amination

The first reports of enzymatic C–H amination were of intramolecular cyclizations of sulfonyl azides catalyzed by cytochrome P450[58] and myoglobin[59] variants. Subsequent work expanded the scope of P450-catalyzed C–H amination to intermolecular synthesis of sulfonamides[60], carboxamides[61], and unprotected primary amines[62] (Figure 1-18).

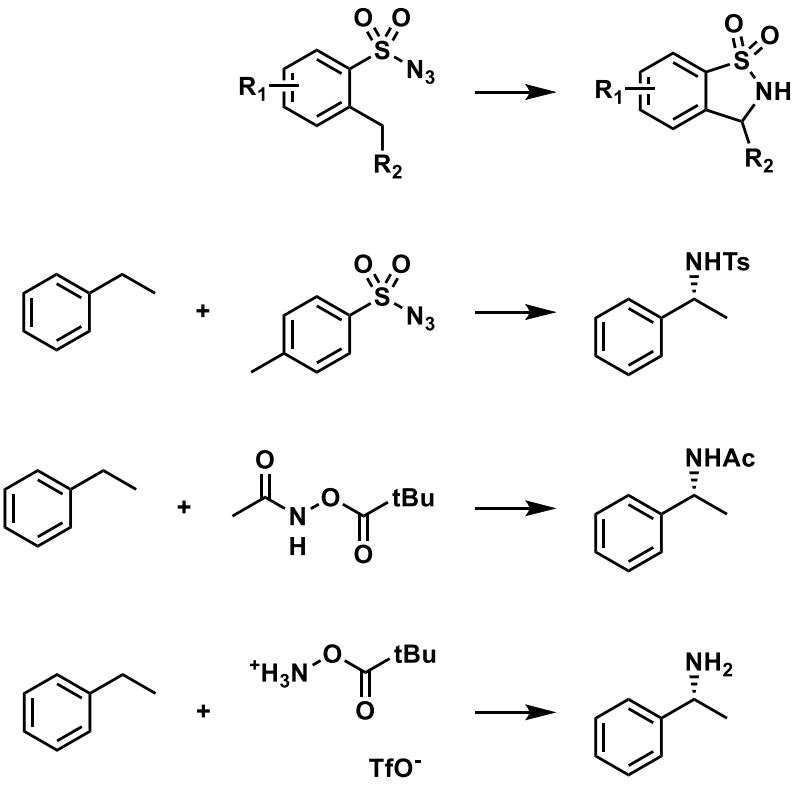


Figure 1-18. Reported examples of nitrene C–H insertion catalyzed by heme-binding proteins.

Non-native chemistry with other metalloenzymes

There have been a few reports of non-native catalysis with native metalloproteins other than heme-binding proteins, though this area remains much less developed than catalysis with heme-binding proteins. The Lewis group recently reported that engineered variants of a cobalamin-dependent enzyme could catalyze the β-alkylation of styrene derivatives with diazo compounds, showing orthogonal reactivity to heme-binding enzymes, which typically catalyze cyclopropanation[63]. The Hartwig group has reported that wild-type carbonic anhydrase, a zinc-dependent enzyme, can catalyze the asymmetric reduction of ketones by silanes[64] (Figure 1-19).

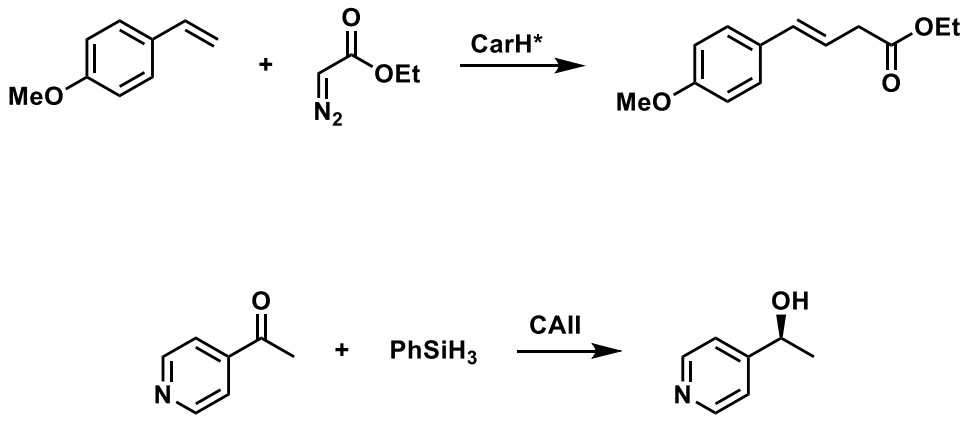


Figure 1-19. Examples of non-native chemistry catalyzed by non-heme metalloenzymes.

1.5 Conclusions

Metalloenzymes play many important roles in natural biochemistry, including some of the most fundamental and most impressive chemistry of life. Human chemists have been drawn to metalloenzymes for synthetic biocatalysis, but have been limited by the available catalytic repertoire of natural metalloenzymes, which is much narrower than that of synthetic transition-metal catalysts. Artificial metalloenzymes, formed by conjugation of a synthetic, abiological metal complex to a protein, have expanded the range of chemistry achievable with biocatalysis, but these conjugates can be laborious to prepare, and often fail to take full advantage of the remarkable catalytic abilities of enzymes. More recently, researchers have begun to explore non-native activities of natural metalloenzymes, primarily heme-binding proteins. Heme proteins, however, represent only a small fraction of the rich biodiversity of natural metalloenzymes, and the non-native chemistry of non-heme metalloproteins remains underexplored.

1.6 References for Chapter 1

[1] Degtyarenko, K. Metalloproteins. In *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics*; Jorde, L., Little, P., Dunn, M., Subramaniam, S., Eds.; John Wiley & Sons, Ltd.: New York, 2005.

[2] Holm, R. H.; Kennepohl, P.; Solomon, E. I. Structural and Functional Aspects of Metal Sites in Biology. *Chem. Rev.* **1996**, *96*, 2239–2314.

[3] Poulos, T. L. Heme Enzyme Structure and Function. *Chem. Rev.* **2014**, *117*, 3919–3962.

[4] Einsle, O.; Rees, D. C. Structural Enzymology of Nitrogenase Enzymes. *Chem. Rev.* **2020**, *120*, 4969–5004.

[5] Volbeda, A.; Garcin, E.; Piras, C.; de Lacy, A. L.; Fernandez, V. M.; Hatchikian, E. C.; Frey, M.; Fontecilla-Camps, J. C. Structure of the [NiFe] Hydrogenase Active Site: Evidence for Biologically Uncommon Fe Ligands. *J. Am. Chem. Soc*. **1996**, *118*, 12989–12996.

[6] Desguin, B.; Zhang, T.; Soumillion, P.; Hols, P.; Hu, J.; Hausinger, R. P. A tethered niacin-derived pincer complex with a nickel-carbon bond in lactate racemase. *Science* **2015**, *349*, 66–69.

[7] Greene, B. L.; Kang, G.; Cui, C.; Bennati, M.; Nocera, D. G.; Drennan, C. L.; Stubbe, J. Ribonucleotide Reductases: Structure, Chemistry, and Metabolism Suggest New Therapeutic Targets. *Annu. Rev. Biochem.* **2020**, *89*, 45–75.

[8] Yruela, I. Transition metals in plant photosynthesis. *Metallomics* **2013**, *5*, 1090–1109.

[9] Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I. Structure and Chemistry of Cytochrome P450. *Chem. Rev.* **2005**, *105*, 2253–2278.

[10] Islam, M. D.; Leissing, T. M.; Chowdhury, R.; Hopkinson, R. J.; Schofield, C. J. 2-Oxoglutarate-Dependent Oxygenases. *Annu. Rev. Biochem*. **2018**, *87*, 585–620.

[11] Staunton, J.; Wilkinson, B. Biosynthesis of Erythromycin and Rapamycin. *Chem. Rev.* **1997**, *97*, 2611–2630.

[12] Croteau, R.; Ketchum, R. E. B.; Long, R. M.; Kaspera, R.; Wildung, W. R. Taxol biosynthesis and molecular genetics. *Phytochem. Rev.* **2006**, *5*, 75–97.

[13] Guengerich, F. P. Cytochrome P450 and Chemical Toxicology. *Chem. Res. Toxicol.* **2008**, *21*, 70–83.

[14] Yamada, H.; Kobayashi, M. Nitrile Hydratase and Its Application to Industrial Production of Acrylamide. *Biosci., Biotechnol., Biochem.* **1996**, *60*, 1391–1400.

[15] Nagasawa, T.; Mathew, C. D.; Mauger, J.; Yamada, H. Nitrile Hydratase-Catalyzed Production of Nicotinamide from 3-Cyanopyridine in *Rhodococcus rhodochrous* J1. *Appl. Environ. Microbiol.* **1988**, *54*, 1766–1769.

[16] Yamanaka, Y.; Hashimoto, K.; Ohtaki, A.; Noguchi, K.; Yohda, M.; Odaka, M. Kinetic and structural studies on roles of the serine ligand and a strictly conserved tyrosine residue in nitrile hydratase. *J. Biol. Inorg. Chem.* **2010**, *15*, 655–665.

[17] Acevedo-Rocha, C. G.; Gamble, C. G.; Lonsdale, R.; Li, A.; Nett, N.; Hoebenreich, S.; Lignau, J. B.; Wirtz, C.; Fares, C.; Hinrichs, H.; Deege, A.; Mulholland, A. J.; Nov, Y.; Leys, D.; McLean, K. J.; Munro, A. W.; Reetz, M. T. P450-Catalyzed Regio- and Diastereoselective Steroid Hydroxylation: Efficient Directed Evolution Enabled by Mutability Landscaping. *ACS Catal.* **2018**, *8*, 3395–3410.

[18] Zhang, K.; Shafer, B. M.; Demars, M. D. II.; Stern, H. A.; Fasan, R. Controlled Oxidation of Remote sp^3 C–H Bonds in Artemisinin via P450 Catalysts with Fine-Tuned Regio- and Stereoselectivity. *J. Am. Chem. Soc.* **2012**, *134*, 18695–18704.

[19] Fasan, R.; Chen, M. M.; Crook, N. C.; Arnold, F. H. Engineered Alkane-Hydroxylating Cytochrome $P450_{BM3}$ Exhibiting Nativelike Catalytic Properties. *Angew. Chem. Int. Ed.* **2007**, *46*, 8414 –8418.

[20] Neugebauer, M. E.; Sumida, K. H.; Pelton, J. G.; McMurry, J. L.; Marchand, J. A.; Chang, M. C. Y. A family of radical halogenases for the engineering of amino-acid-based products. *Nat. Chem. Biol.* **2019**, *15*, 1009–1016.

[21] Hayashi, T.; Ligibel, M.; Sager, E.; Voss, M.; Hunziker, J.; Schroer, K.; Snajdrova, R.; Buller, R. Evolved Aliphatic Halogenases Enable Regiocomplementary C−H Functionalization of a Pharmaceutically Relevant Compound. *Angew. Chem. Int. Ed.* **2019**, *58*, 18535–18539.

[22] Büchler, J.; Malca, S. H.; Patsch, D.; Voss, M.; Turner, N. J.; Bornscheuer, U. T.; Allemann, O.; Le Chapelain, C.; Lumbroso, A.; Loiseleur, O.; Buller, R. Algorithm-aided engineering of aliphatic halogenase WelO5* for the asymmetric late-stage functionalization of soraphens. *Nat. Commun.* **2022**, *13*, 371.

[23] Loskot, S. A.; Romney, D. K.; Arnold, F. H.; Stoltz, B. M. Enantioselective Total Synthesis of Nigelladine A *via* Late-Stage C–H Oxidation Enabled by an Engineered P450 Enzyme. *J. Am. Chem. Soc.* **2017**, *139*, 10196–10199.

[24] Zwick, C. R. III.; Renata, H. Remote C–H Hydroxylation by an α-Ketoglutarate-Dependent Dioxygenase Enables Efficient Chemoenzymatic Synthesis of Manzacidin C and Proline Analogs. *J. Am. Chem. Soc.* **2018**, *140*, 1165–1169.

[25] Schwizer, F.; Okamoto, Y.; Heinisch, T.; Gu, Y.; Pellizzoni, M. M.; Lebrun, V.; Reuter, R.; Köhler, V.; Lewis, J. C.; Ward, T. R. Artificial Metalloenzymes: Reaction Scope and Optimization Strategies. *Chem. Rev.* **2018**, *118*, 142–231.

[26] Davis, H. J.; Ward, T. R. Artificial Metalloenzymes: Challenges and Opportunities. *ACS Cent. Sci.* **2019**, *5*, 1120–1136.

[27] Wilson, M. E.; Whitesides, G. M. Conversion of a protein to a homogeneous asymmetric hydrogenation catalyst by site-specific modification with a diphosphinerhodium(I) moiety. *J. Am. Chem. Soc.* **1978**, *100*, 306–307.

[28] Dürrenberger, M.; Heinisch, T.; Wilson, Y. M.; Rossel, T.; Nogueira, E.; Knörr, L.; Mutschler, A.; Kersten, K.; Zimbron, M. J.; Pierron, J.; Schirmer, T.; Ward, T. R. Artificial

Transfer Hydrogenases for the Enantioselective Reduction of Cyclic Imines. *Angew. Chem. Int. Ed.* **2011**, *50*, 3026–3029.

[29] Lo, C.; Ringenberg, M. R.; Gnandt, D.; Wilson, Y.; Ward, T. R. Artificial metalloenzymes for olefin metathesis based in the biotin-(strept)avidin technology. *Chem. Commun.* **2011**, *47*, 12065– 12067.

[30] Chatterjee, A.; Mallin, H.; Klehr, J.; Vallapurackal, J.; Finke, A. D.; Vera, L.; Marsh, M.; Ward, T. R. An enantioselective artificial Suzukiase based on the biotin–streptavidin technology. *Chem. Sci.* **2016**, *7*, 673–677.

[31] Mayer, C.; Gillingham, D. G.; Ward, T. R.; Hilvert, D. H. An artificial metalloenzyme for olefin metathesis. *Chem. Commun.* **2011**, *47*, 12068–12070.

[32] Jarvis, A. G.; Obrecht, L.; Deuss, P. J.; Laan, W.; Gibson, E. K.; Wells, P. P.; Kamer, P. C. J. Enzyme Activity by Design: An Artificial Rhodium Hydroformylase for Linear Aldehydes. *Angew. Chem. Int. Ed.* **2017**, *56*, 13596–13600.

[33] Bos, J.; Fusetti, F.; Driessen, A. J. M.; Roelfes, G. Enantioselective Artificial Metalloenzymes by Creation of a Novel Active Site at the Protein Dimer Interface. *Angew. Chem. Int. Ed.* **2012**, *51*, 7472–7475.

[34] Lewis, J. C. Metallopeptide Catalysts and Artificial Metalloenzymes Containing Unnatural Amino Acids. *Curr. Opin. Chem. Biol*. **2015**, *25*, 27–35.

[35] Drienovská, I.; Alonso-Cotchico, L.; Vidossich, P.; Lledós, A.; Maréchal, J.-D.; Roelfes, G. Design of an enantioselective artificial metallo-hydratase enzyme containing an unnatural metal-binding amino acid. *Chem. Sci.* **2017**, *8*, 7228–7235.

[36] Srivastava, P.; Yang, H.; Ellis-Guardiola, K.; Lewis, J. C. Engineering a dirhodium artificial metalloenzyme for selective olefin cyclopropanation. *Nat. Commun.* **2015**, *6*, 7789.

[37] Baglia, R. A.; Zaragoza, J. P. T.; Goldberg, D. P. Biomimetic Reactivity of Oxygen-derived Manganese and Iron Porphyrinoid Complexes. *Chem. Rev.* **2017**, *117*, 13320–13352.

[38] Zhu, S.-F.; Zhou, Q.-L. Iron-catalyzed transformations of diazo compounds. *Natl. Sci. Rev.* **2014**, *1*, 580–603.

[39] Fantauzzi, S.; Caselli, A.; Gallo, E. Nitrene transfer reactions mediated by metallo-porphyrin complexes. *Dalton Trans.* **2009**, 5434–5443.

[40] Liu, Z.; Arnold. F. H. New-to-nature chemistry from old protein machinery: carbene and nitrene transferases. *Curr. Opin. Biotechnol.* **2021**, *69*, 43–51.

[41] Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. Olefin Cyclopropanation *via* Carbene Transfer Catalyzed by Cytochrome P450 Variants. *Science* **2013**, *339*, 307–310.

[42] Bordeaux, M.; Tyagi, V.; Fasan, R. Highly Diastereoselective and Enantioselective Olefin Cyclopropanation Using Engineered Myoglobin-Based Catalysts. *Angew. Chem. Int. Ed.* **2015**, *54*, 1744–1748.

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

[44] Chandgude, A. L.; Ren, X.; Fasan, R. Stereodivergent Intramolecular Cyclopropanation Enabled by Engineered Carbene Transferases. *J. Am. Chem. Soc.* **2019**, *141*, 9145–9150.

[45] Ren, X.; Chandgude, A. L.; Fasan, R. Highly Stereoselective Synthesis of Fused Cyclopropane-γ-Lactams via Biocatalytic Iron-Catalyzed Intramolecular Cyclopropanation. *ACS Catal.* **2020**, *10*, 2308–2313.

[46] Chen, K.; Arnold, F. H. Engineering Cytochrome P450s for Enantioselective Cyclopropenation of Internal Alkynes. *J. Am. Chem. Soc.* **2020**, *142*, 6891–6895.

[47] Chen, K.; Huang, X.; Kan, S. B. J.; Zhang, R. K.; Arnold, F. H. Enzymatic construction of highly strained carbocycles. *Science* **2018**, *360*, 71–75.

[48] Zhang, R. K.; Chen, K.; Huang, X.; Wohlschlager, L.; Renata, H.; Arnold, F. H. Enzymatic assembly of carbon–carbon bonds via iron-catalysed sp^3 C–H functionalization. *Nature* **2019**, *565*, 67–72.

[49] Zhang, J.; Huang, X.; Zhang, R. K.; Arnold. F. H. Enantiodivergent α-Amino C-H Fluoroalkylation Catalyzed by Engineered Cytochrome P450s. *J. Am. Chem. Soc.* **2019**, *141*, 9798–9802.

[50] Zhou, A. Z.; Chen, K.; Arnold, F. H. Enzymatic Lactone-Carbene C–H Insertion to Build Contiguous Chiral Centers. *ACS Catal.* **2020**, *10*, 5393–5398.

[51] 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* **2016**, *354*, 1048 – 1051.

[52] Kan, S. B. J.; Huang, X.; Gumulya, Y.; Chen, K.; Arnold, F. H. Genetically programed chiral organoborane synthesis. *Nature* **2017**, *552*, 132–136.

[53] Wang, Z. J.; Peck, N. E.; Renata, H.; Arnold, F. H. Cytochrome P450-catalyzed insertion of carbenoids into N–H bonds. *Chem. Sci.* **2014**, *5*, 598–601.

[54] Tyagi, V.; Bonn, R. B.; Fasan, R. Intermolecular carbene S–H insertion catalysed by engineered myoglobin-based catalysts. *Chem. Sci.* **2015**, *6*, 2488–2494.

[55] Tsutsumi, H.; Katsuyama, Y.; Izumikawa, M.; Takagi, M.; Fujie, M.; Satoh, N.; Shin-ya, K.; Ohnishi, Y. Unprecedented Cyclization Catalyzed by a Cytochrome P450 in Benzastatin Biosynthesis. *J. Am. Chem. Soc.* **2018**, *140*, 6631–6639.

[56] Farwell, C. C.; Zhang, R. K.; McIntosh, J. A.; Hyster, T. K.; Arnold, F. H. Enantioselective Enzyme-Catalyzed Aziridination Enabled by ActiveSite Evolution of a Cytochrome P450. *ACS Cent. Sci.* **2015**, *1*, 89–93.

[57] Cho, I.; Prier, C. K.; Jia, Z.-J.; Zhang, R. K.; Görbe, T.; Arnold, F. H. Enantioselective Aminohydroxylation of Styrenyl Olefins Catalyzed by an Engineered Hemoprotein *Angew. Chem. Int. Ed.* **2019**, *58*, 3138–3142.

[58] McIntosh, J. A.; Coelho, P. S.; Farwell, C. C.; Wang, Z. J.; Lewis, J. C.; Brown, T. R.; Arnold, F. H. Enantioselective Intramolecular C–H Amination Catalyzed by Engineered Cytochrome P450 Enzymes *In Vitro* and *In Vivo*. *Angew. Chem. Int. Ed.* **2013**, *52*, 9309–9312.

[59] Bordeaux, M.; Singh, R.; Fasan, R. Intramolecular C(sp^3) –H amination of arylsulfonyl azides with engineered and artificial myoglobin-based catalysts. *Bioorg. Med. Chem.* **2014**, *22*, 5697–5704.

[60] Prier, C. K.; Zhang, R. K.; Buller, A. R.; Brinkmann-Chen, S.; Arnold, F. H. Enantioselective, Intermolecular Benzylic C–H Amination Catalysed by an Engineered Iron-Haem Enzyme. *Nat. Chem.* **2017**, *9*, 629–634.

[61] Athavale, S. V.; Gao, S.; Liu, Z.; Mallojjala, S. C.; Hirschi, J. S.; Arnold. F. H. Biocatalytic, Intermolecular C−H Bond Functionalization for the Synthesis of Enantioenriched Amides. *Angew. Chem. Int. Ed.* **2021**, *60*, 24864–24869.

[62] Jia, Z.-J.; Gao, S.; Arnold, F. H. Enzymatic Primary Amination of Benzylic and Allylic C(sp^3)–H Bonds. *J. Am. Chem. Soc.* **2020**, *142*, 10279–10283.

[63] Yang, X.; Gerroll, B. H. R.; Jiang, Y.; Kumar, A.; Zubi, Y. S.; Baker, L. A.; Lewis, J. C. Controlling Non-Native Cobalamin Reactivity and Catalysis in the Transcription Factor CarH. *ACS Catal.* **2022**, *12*, 935–942.

[64] Ji, P.; Park, J.; Gu, Y.; Clark, D. S.; Hartwig, J. F. Abiotic reduction of ketones with silanes catalysed by carbonic anhydrase through an enzymatic zinc hydride. *Nat. Chem.* **2021**, *13*, 312–318.

Chapter 2

EFFORTS TOWARDS ENZYME-CATALYZED HYDROSILYLATION

The work described in this chapter was performed in collaboration with Anders M. Knight, S. B. Jennifer Kan, and Sabine Brinkmann-Chen. N.W.G. and A.M.K. performed molecular biology, protein expression and purification, and reaction screening. N.W.G. performed synthetic chemistry. A.M.K. developed the systematic structural database search and optimized analytical methods. A.M.K., N.W.G, S.B.J.K., and S.B.-C. performed further narrowing of the protein search set.

Abstract

Hydrosilylation, the formal addition of a silane Si–H bond across a double bond, is a widely used reaction for the synthesis of functionalized organosilicon compounds. A wide variety of small-molecule transition metal complexes have been reported to catalyze hydrosilylation, but industrial hydrosilylation still relies almost exclusively on platinum-based catalysts, which are financially and environmentally costly. We sought to discover and engineer a protein biocatalyst for hydrosilylation. Towards this goal, we collected a set of metalloproteins drawn from natural diversity and tested them for a range of hydrosilylation reactions, but we were unable to find detectable protein-catalyzed hydrosilylation activity.

2.1 Hydrosilylation

Hydrosilylation is the formal addition of a silicon-hydrogen bond across a π-system, typically a carbon-carbon double/triple bond, or a carbon-oxygen double bond. Hydrosilylation of olefins is a widely used reaction in the silicones industry, where it is used to synthesize functionalized monomers and crosslink polymer chains for the curing of end products[1,2]. Several varieties of hydrosilylation have been reported, but the most common and the most widely used involve transition-metal catalysts. A variety of different catalysts have been developed for hydrosilylation[3], including complexes of noble metals such as platinum[4,5], rhodium[6], and iridium[7], as well as base metals[8] including iron[9], nickel[10], and copper[11] (Figure 2-1). Despite the diversity of reported catalysts for laboratory-scale hydrosilylation, however, most are air- or water-sensitive and are not sufficiently robust or active for large-scale use. As a result, industrial hydrosilylation relies almost exclusively on platinum catalysis, consuming several tons of platinum per year which generally cannot be recovered or reused[12]. Development of a cheap and sustainable hydrosilylation catalyst that is sufficiently active and stable for industrial use remains an outstanding challenge.

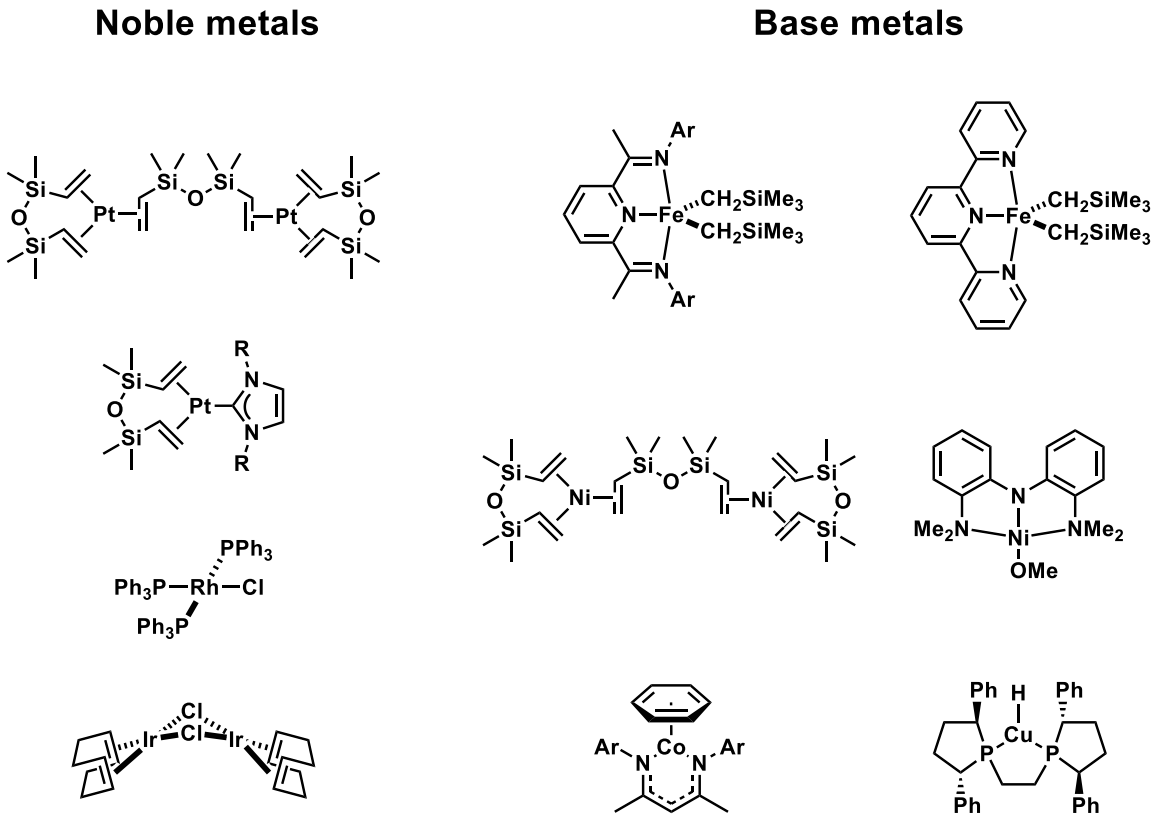


Figure 2-1. Examples of reported transition-metal catalysts for olefin hydrosilylation.

Classical hydrosilylation mechanisms

While a number of mechanisms are known for olefin hydrosilylation, among the most well-known for classical hydrosilylation catalysts are the Chalk-Harrod and modified Chalk-Harrod mechanisms[13,14]. In the Chalk-Harrod mechanism, the catalytic metal center undergoes oxidative addition into the silane Si–H bond. Following olefin association, migratory insertion of the hydride into the olefin and C–Si reductive elimination, yields the hydrosilylated product. The modified Chalk-Harrod mechanism is similar, except the silyl group performs migratory insertion followed by C–H reductive elimination (Figure 2-2). Other reported mechanisms involve a metal-silylene complex[15] or σ-bond metathesis of a metal-hydride complex[11].

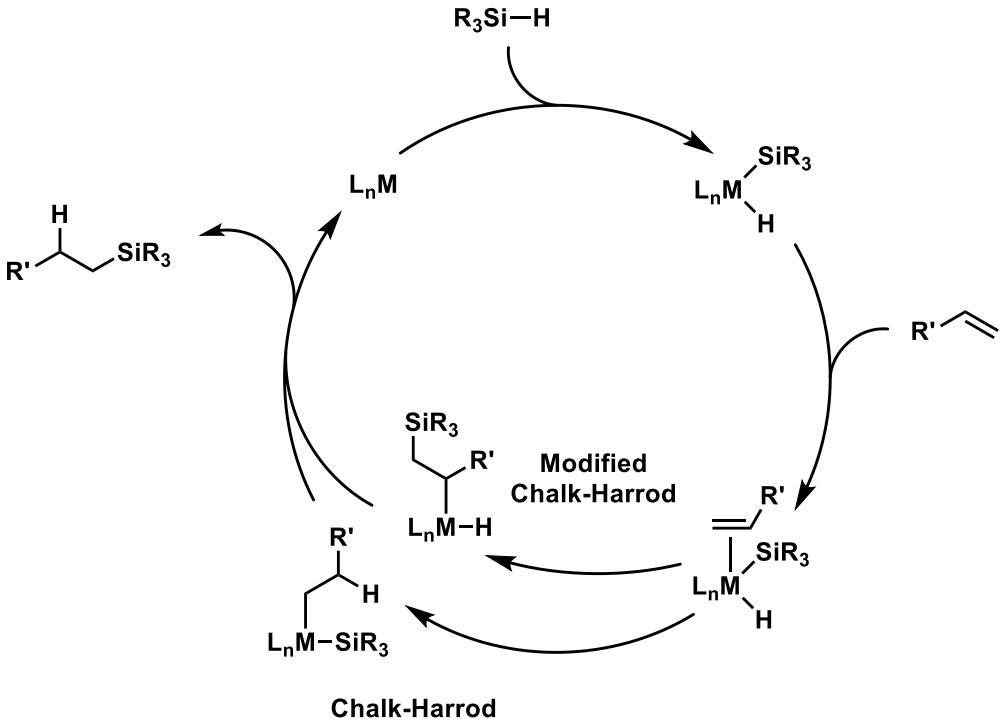


Figure 2-2. Chalk-Harrod and modified Chalk-Harrod mechanisms for olefin hydrosilylation.

Potential biocatalytic hydrosilylation

Biocatalysis offers a potential solution to the challenge of sustainable hydrosilylation. Biocatalysts are often cheaper and greener than traditional chemical catalysts[16], and an enzymatic hydrosilylation catalyst could potentially overcome some of the limitations of existing methods, providing stereo- and regioselective base-metal catalysts that function under mild, ambient conditions. As many hydrosilylation catalysts are transition metal-based, including the most industrially relevant catalysts, we reasoned that metalloproteins would be the most promising candidates to test. Since classical hydrosilylation mechanisms require multiple available coordination sites on the metal, heme-binding or other metalloporphyrin-binding proteins were deemed unlikely to be successful. A large fraction (30–50%) of all proteins bind a metal in some way. These metals are largely drawn from the first-row transition metals, and they display a variety of biological roles, including structural stabilization, electrochemistry, and Lewis-acid or redox catalysis[17]. We proposed to explore the natural diversity of metalloproteins in the search for hydrosilylation activity.

2.2 Initial set of metalloproteins

We first assembled a collection of metalloproteins drawn from natural diversity, listed in Table 2-1.

Table 2-1. Initial protein set, grouped by native metal.

Protein	**Species**	**UniProt ID**	**Native metal**
Amicyanin	*Paracoccus denitrificans*	P22364	Cu
Auracyanin D	*Chloroflexus aurantiacus*	A9WFS1	Cu
Azurin	*Pseudomonas aeruginosa*	P00282	Cu
Azurin	*Pseudomonas putida*	P34097	Cu
Azurin iso-2	*Methylomonas sp.*	P12335	Cu
Laccase	*Thermus thermophilus*	Q72HW2	Cu
Nitrosocyanin	*Nitrosomonas europaea*	Q820S6	Cu
Plantacyanin	*Cucumus sativus*	P00303	Cu
Plastocyanin	*Chlamodymonas reinhardtii*	P18068	Cu
Plastocyanin	*Dryopteris crassirhizoma*	Q7SIB8	Cu
Plastocyanin	*Phormodium laminosum*	Q51883	Cu
Pseudoazurin	*Methylobacterium extorquens*	P04171	Cu
Pseudoazurin	*Rhizobium meliloti*	Q92M26	Cu
Rusticyanin	*Alcaligenes faecalis*	P0C918	Cu
Cupin-like protein	*Thermotoga maritima*	Q9X1H0	Mn
Superoxide dismutase	*Pyrobaculum aerophilum*	O93724	Mn
Superoxide dismutase	*Thermus thermophilus*	P61503	Mn
Acireductone dioxygenase	*Bacillus anthracis*	Q81MI9	Ni
Quercetinase	*Streptomyces sp.*	A2VA43	Ni
Taurine dioxygenase	*Escherichia coli*	P37610	Fe
Leucine hydroxylase	*Gluconobacter oxydans*	Q5FQD2	Fe

Leucine hydroxylase	*Streptomyces muensis*	A0A0E3URV8	Fe
Arginine hydroxylase	*Streptomyces vinaceus*	Q6WZB0	Fe

Cupredoxins

We included a number of cupredoxins and other related copper-binding proteins, as synthetic copper complexes are known catalysts for both carbonyl[18] and olefin hydrosilylation[11,19]. Cupredoxins are natively electron-transport proteins, and have been extensively studied for their biophysical properties[20]. In the typical cupredoxin metal site, the copper is coordinated equatorially by two histidines and a cysteine, with an axially coordinating methionine (Figure 2-3). The azurins have an additional axial interaction with a glycine main-chain carbonyl oxygen.

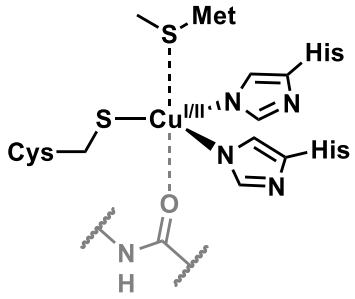


Figure 2-3. Generalized cupredoxin metal-binding site.

Ligand-mutated cupredoxins

The copper center in cupredoxins is coordinatively saturated in the resting state, so at least one coordinating residue would have to dissociate for catalysis to occur. To facilitate this, we chose two well-expressing cupredoxins (*P. aeruginosa* azurin and *P. denitrificans* amicyanin) and performed site-saturation mutagenesis on the axial methionine residue and the solvent-facing coordinating histidine residue, screening visually (by blue color) for the ability to bind copper selectively in cell lysate. We found that the *Pae* azurin H117D and M121G variants and *Pde* amicyanin M98Q were able to bind copper, so they were included for reaction screening.

α-Ketoglutarate-dependent non-heme iron dioxygenases

Our set included 4 α-ketoglutarate-dependent non-heme iron dioxygenases. These enzymes bind iron in a two histidine/one carboxylate binding site, and natively perform oxidation chemistry, commonly C–H hydroxylation reactions (Figure 2-4). This family of enzymes is discussed in more detail in Chapter 3, section 3.1.

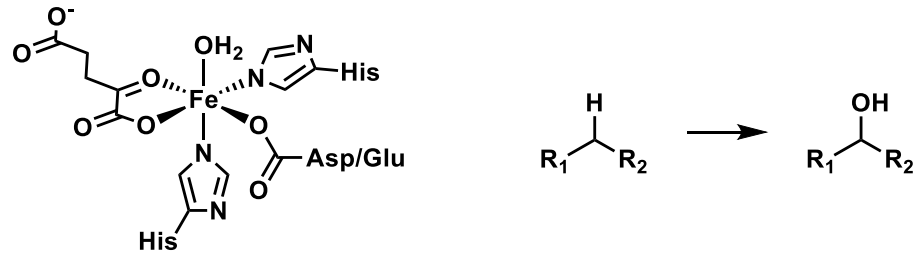


Figure 2-4. α-Ketoglutarate-dependent iron dioxygenase metal binding site and canonical native catalytic function.

Superoxide dismutases

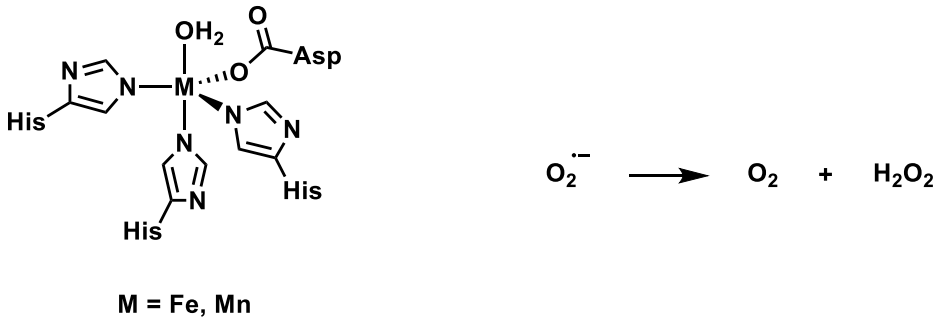


Figure 2-5. Superoxide dismutase metal-binding site and native catalytic function.

Our set included two superoxide dismutases from hyperthermophilic source organisms. These enzymes are natively iron- or manganese-dependent, and catalyze the disproportionation of superoxide to dioxygen and hydrogen peroxide (Figure 2-5). Notably, these proteins are produced as apoprotein when expressed in mesophilic heterologous hosts such as *E. coli*, and only bind a metal upon heat treatment[21]. This property enabled us to easily prepare these proteins bound with manganese, iron, cobalt, nickel, and copper.

Furthermore, as these proteins are quite thermostable, we tested them for hydrosilylation at 60 °C as well as at room temperature.

Other proteins

Our collection also included a few cupins and related proteins[22,23]. The proteins in our collection natively bind nickel or manganese, and have various native functions, mostly oxidation-related. Their metal-binding site structures are shown in Figure 2-6.

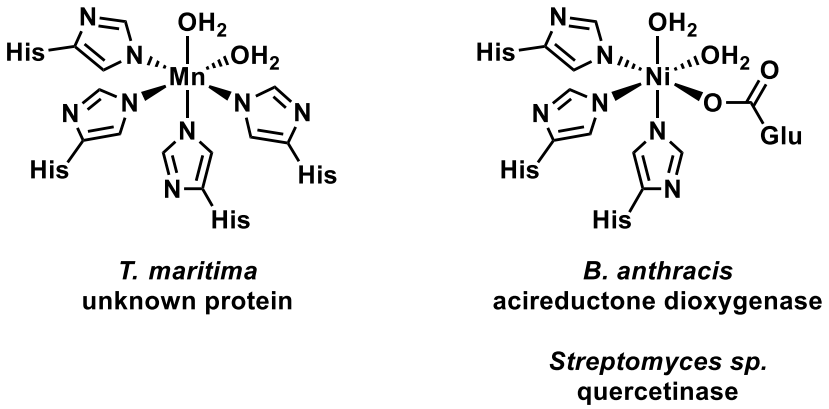


Figure 2-6. Metal binding site structures of cupin-like proteins in our collection.

2.3 Hydrosilylation reaction screening with initial protein set

We tested the proteins against a panel of olefin and carbonyl hydrosilylation reactions. We chose a set of silanes known to be stable in water, and olefins and carbonyl compounds of varied electronic and steric characters (Figure 2-7). All pairwise combinations of one silane and one olefin or carbonyl were tested. We did not find detectable activity for any of these reactions with any of the proteins tested.

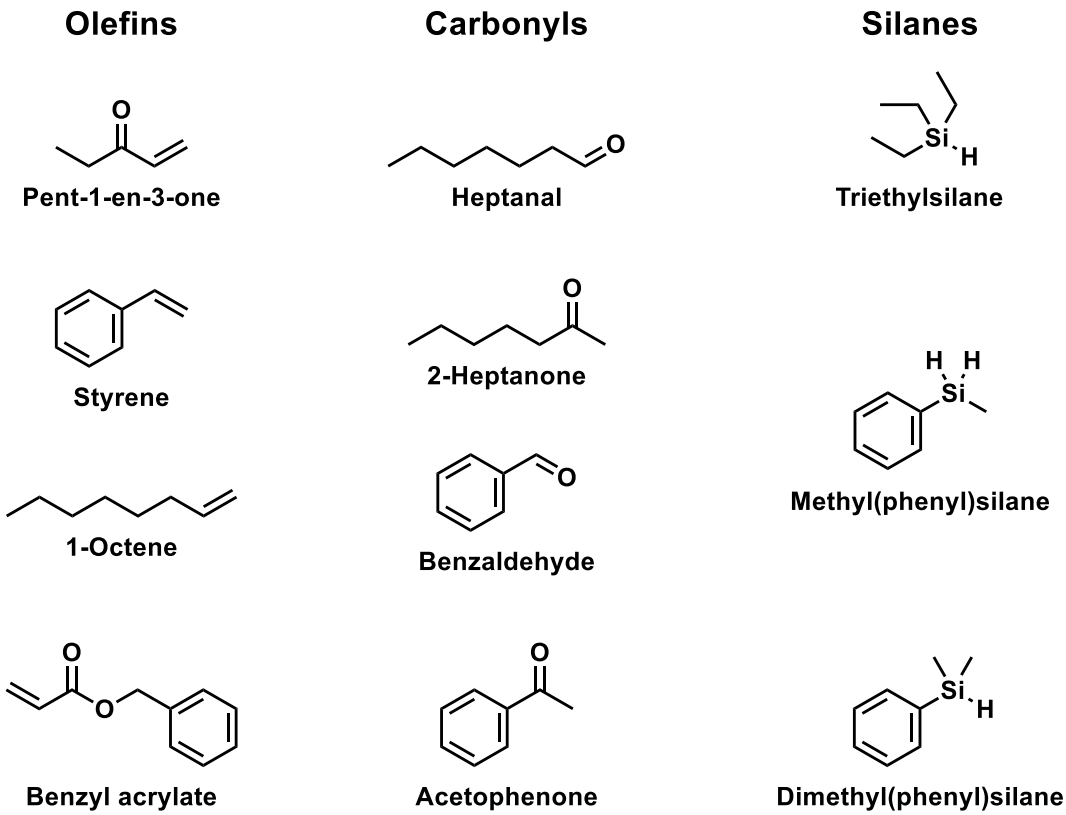


Figure 2-7. Substrates tested for hydrosilylation activity. Reaction conditions: MOPS buffer (20 mM pH 7.0) with 5% acetonitrile as co-solvent, 23 °C (also 60 °C for superoxide dismutases), anaerobic, 18–24 h. 10 mM each substrate, ~50 μM protein, 1 mM metal salt. Reactions performed both with and without added sodium dithionite (reductant).

2.4 Expanded metalloprotein set

As we did not find detectable hydrosilylation activity with our initial protein set, we decided to expand our protein search space. Our initial protein set sampled only a small subset of the diversity of metal-binding sites found in naturally occurring proteins. To sample a broader range of metalloproteins, we decided to perform a systematic search of protein structural databases, specifically the Protein Data Bank (PDB)[24] and MetalPDB[25], a database of annotated information on the coordinating residues and coordination geometry of metal-containing structures in the PDB.

Structural database search

To generate a list of candidate structures from the PDB, we developed a set of search criteria:

1. The structure must only consist of protein macromolecules. This criterion eliminates structures containing DNA or RNA, or other non-protein macromolecules.
2. The protein must contain a first-row transition-metal ion (Mn, Fe, Co, Ni, or Cu). This search criterion eliminates proteins containing metal complex cofactors such as heme, cobalamin, or others. While there are also proteins that bind second- and third-row transition metals, these proteins often have complex metal-binding sites that increase the difficulty of expression and manipulation.
3. The protein must be 600 amino acids or shorter. Smaller proteins are generally preferable as the genes are cheaper to synthesize, and they are often better behaved and more highly expressing than large proteins.
4. The protein must have been previously expressed in *E. coli*. We wanted to maximize the likelihood that proteins would express and fold properly when expressed recombinantly in a prokaryotic host.
5. The protein list will give representative structures at 95% sequence identity. Some proteins have been extensively studied and may have multiple deposited structures; this criterion eliminates duplicate and near-duplicate proteins.

In addition to the PDB search described above, we searched the MetalPDB for protein structures containing first-row transition metal ions (Cu, Fe, Ni, Mn, and Co). These searches resulted in 2071 structures from the PDB and 1086 structures from the MetalPDB, which combined to 2572 unique candidate protein structures. As 2572 proteins are far too many to screen experimentally, we refined this set by visual inspection of every structure, looking for a well-defined metal binding site, with an accessible cavity that could potentially accommodate the binding of organic substrates. Based on these criteria, we selected 179 of the most promising candidate proteins. We narrowed this set further by analyzing the reported methods for expression and purification, selecting proteins unlikely to present difficulties or complications, resulting in a final set of 62 proteins. A plurality of these proteins are reported to bind iron (Figure 2-8), reflecting the abundance of iron in natural

biochemical pathways. The remainder are approximately equally distributed among the other metals in our search set, although we intentionally chose fewer copper-binding proteins as copper-binding proteins were well-represented in our initial protein set. A full list is given in Table 2-2, and selected active site structures are shown in Figure 2-9, highlighting unusual metal-binding sites or those dissimilar to the proteins in our initial set.

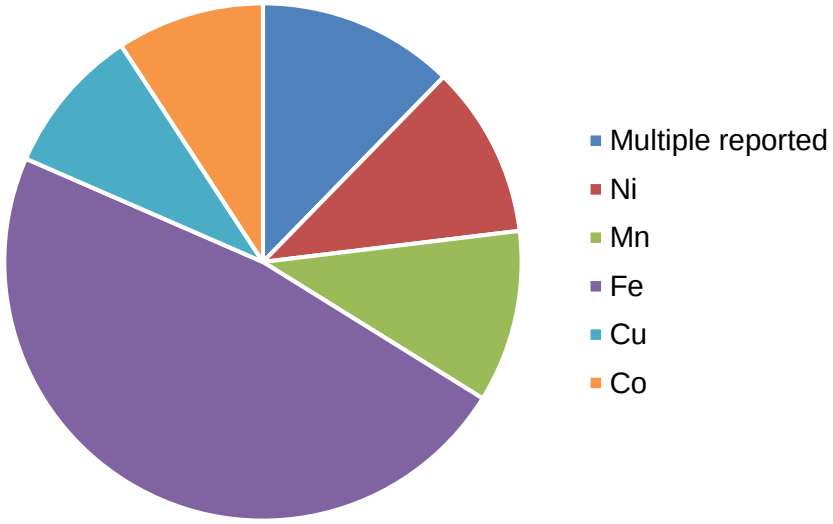


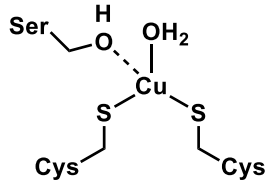
Figure 2-8. Reported bound metal of the expanded protein set.

Table 2-2. Expanded set of metalloproteins.

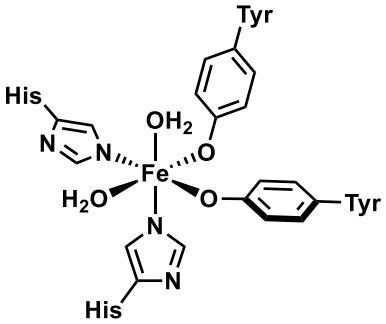
#	UniProt ID	Protein name	Source organism	Metal
1	Q9REI7	2,4'-Dihydroxyacetophenone dioxygenase	*Alcaligenes sp.*	Fe
2	P17109	Succinyl isochorismate synthase	*Escherichia coli*	Mn
3	P77072	Organomercurial lyase MerB D99S	*Escherichia coli*	Cu
4	P0A9S1	Lactaldehyde:1,2-propanediol oxidoreductase	*Escherichia coli*	Fe
5	P0AC81	Glyoxalase	*Escherichia coli*	Ni/Co
6	A0A067YX61	α-Ketoglutarate-dependent halogenase	*Hapalosiphon welwitschii*	Fe
7	Q68RJ8	Glyoxalase I	*Leishmania major*	Ni
8	Q99JT9	Acireductone dioxygenase	*Mus musculus*	Ni/Fe
9	G7CFI3	Ergothioneine-biosynthetic sulfoxide synthase	*Mycobacterium thermoresistibile*	Fe
10	A9A2G4	Purple Cupredoxin	*Nitrosopumilus maritimus*	Cu
11	Q7MZL9	plu4264 Protein, unknown function	*Photorhabdus luminescens*	Ni
12	Q70AC7	Transcarboxylase	*Propionibacterium freudenreichii*	Co
13	O50580	D-Tagatose 3-epimerase	*Pseudomonas cichorii*	Mn
14	O58810	Superoxide reductase	*Pyrococcus horikoshii*	Fe

15	Q6N272	Functionally unknown protein RPA4178	*Rhodopseudomonas palustris*	Ni
16	Q9X034	Amidohydrolase	*Thermotoga maritima*	Ni
17	Q9X113	Oxalate decarboxylase	*Thermotoga maritima*	Mn
18	Q5AR53	Fe(II)/α-ketoglutarate-dependent dioxygenase	*Aspergillus nidulans*	Fe
19	B1L4V6	Hydrolase, DNA repair enzyme	*Korarchaeum cryptofilum*	Mn/Co/Cu
20	V6TJK7	Superoxide reductase	*Giardia intestinalis*	Fe
21	C7R4I0	Chitinase	*Jonesia denitrificans*	Cu
22	Q74MF3	Superoxide reductase	*Nanoarchaeum equitans*	Fe
23	P13280	Glycogenin	*Oryctolagus cuniculus*	Mn
24	F8LWI3	Cambialistic superoxide dismutase	*Streptococcus thermophilus*	Fe/Mn
25	P80857	Superoxide dismutase	*Sulfolobus solfataricus*	Fe
26	Q9WYP7	L-Ketose-3-epimerase	*Thermotoga maritima*	Mn/Ni
27	A0A0M3KL01	Phytanoyl-CoA dioxygenase	*Micromonospora carbonacea*	Fe
28	A5VWI3	Persulfide dioxygenase	*Pseudomonas putida*	Fe
29	A6VKV4	Phosphoenolpyruvate carboxykinase	*Actinobacillus succinogenes*	Mn
30	B3PJ79	Lytic polysaccharide monooxygenase	*Cellvibrio japonicus*	Cu
31	C6FI44	Hydroxyquinol 1,2-dioxygenase	*Pseudomonas putida*	Fe
32	C6RPG2	Catechol 1,2 dioxygenase	*Acinetobacter radioresistens*	Fe
33	D0VX22	α-Ketoglutarate-dependent dioxygenase	*Streptomyces sp.*	Fe
34	D6EWM4	Lytic polysaccharide monooxygenase	*Streptomyces lividans*	Cu
35	F0QXN6	Phosphotriesterase-like lactonase	*Vulcanisaeta moutnovskia*	Co
36	O53512	3-Deoxy-D-arabino-heptulosonate 7-phosphate synthase	*Mycobacterium tuberculosis*	Mn
37	O58691	Dipeptidase	*Pyrococcus horikoshii*	Co
38	O87198	Homocitrate synthase	*Thermus thermophilus*	Co/Cu
39	P0A434	Phosphotriesterase	*Brevundimonas diminuta*	Mn
40	P05050	2-oxoglutarate iron(II) dependent dioxygenase	*Escherichia coli*	Fe
41	P29082	Sulfur oxygenase/reductase	*Acidianus ambivalens*	Fe
42	P30967	Phenylalanine hydroxylase	*Chromobacterium violaceum*	Fe
43	P32021	Ethylene-forming enzyme	*Pseudomonas savastanoi*	Fe
44	P42106	Quercetin 2,3-dioxygenase	*Bacillus subtilis*	Fe
45	P53608	Arginase	*[Bacillus] caldovelox*	Mn
46	P70080	Trytophan hydroxylase	*Gallus gallus*	Fe
47	P74334	Apocarotenoid cleavage oxygenase	*Synechocystis sp. PCC 6803*	Fe
48	Q1D4C9	Metallo-β-lactamase	*Myxococcus xanthus*	Fe
49	Q6REQ5	Catechol dioxygenase	*Rhodococcus sp. DK17*	Fe
50	Q7WYF5	2,3-Dioxygenase LapB	*Pseudomonas alkylphenolica*	Fe
51	Q9RHA2	Fructose-1,6-bisphosphate aldolase	*Thermus aquaticus*	Co
52	Q9WZS1	N-Acetylglucosamine-6-phosphate deacetylase	*Thermotoga maritima*	Fe
53	Q9WZS7	NADH-dependent butanol dehydrogenase	*Thermotoga maritima*	Fe
54	Q9X0P5	Metallo-β -lactamase H8A	*Thermotoga maritima*	Ni

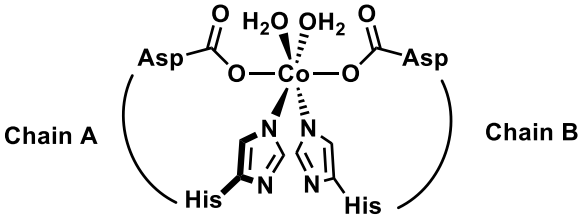
55	Q9X1T8	Transcription regulator	*Thermotoga maritima*	Ni
56	Q672W7	Peptide deformylase	*Helicobacter pylori*	Co
57	Q838S1	Polysaccharide monooxygenase	*Enterococcus faecalis*	Cu
58	Q53586	Dioxygenase	*Streptomyces avermitilis*	Fe
59	Q56185	Hydroxypropylphosphonic acid epoxidase	*Streptomyces wedmorensis*	Co
60	Q96323	Anthocyanidin synthase	*Arabidopsis thaliana*	Fe
61	Q07XY2	Putative hydrolase	*Shewanella frigidimarina*	Ni
62	Q9UXT7	tRNA N6-adenosine threonylcarbamoyltransferase	*Pyrococcus abyssi*	Fe



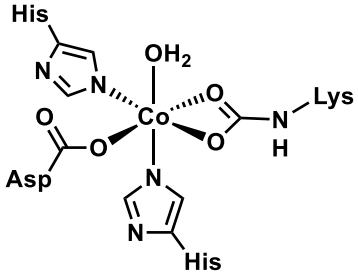
***E. coli* MerB D99S**
PDB: 5C0U



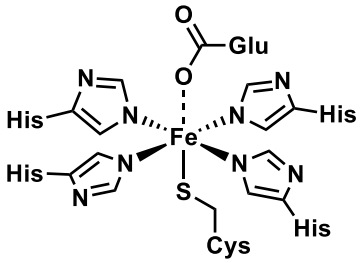
***Streptomyces sp*. SirexAA-E dioxygenase**
PDB: 4ILV



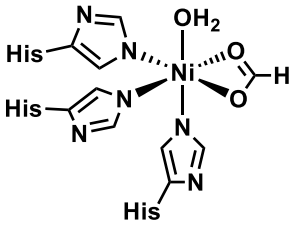
***E. coli* glyoxalase I**
PDB: 1FA6



***P. freudenreichii* transcarboxylase**
PDB: 1RQB



***P. horikoshii* superoxide reductase**
PDB: 2HVB

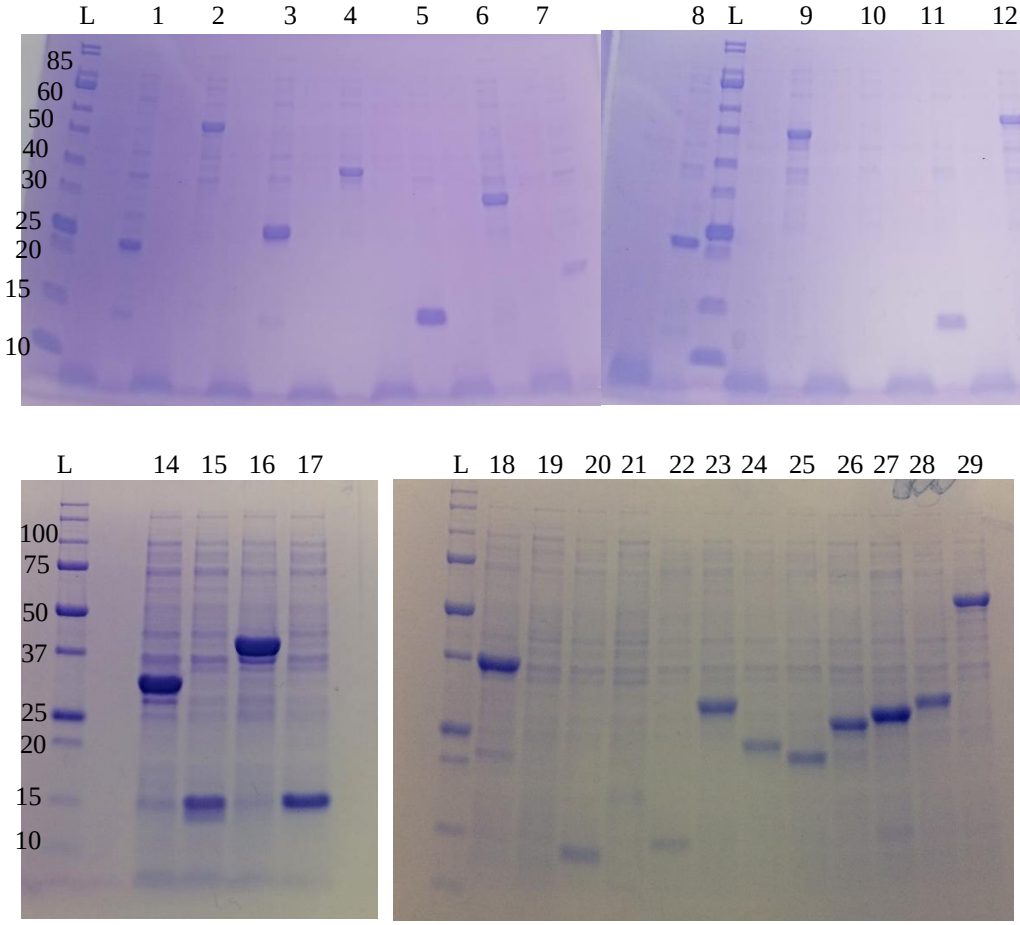


***R. palustris* unknown protein**
PDB: 3LAG

Figure 2-9. Selected metal-binding site structures of proteins in the expanded set.

Expression and purification of expanded protein set

Genes encoding the proteins listed in Table 2-2 were purchased as assembled plasmids from Twist Bioscience. Before embarking on large-scale expression and purification of the entire set of 62 proteins, we performed small-scale tests to qualitatively determine expression level. Fortunately, 55 of the 62 proteins had detectable expression, with 35–40 highly overexpressed (Figure 2-10).



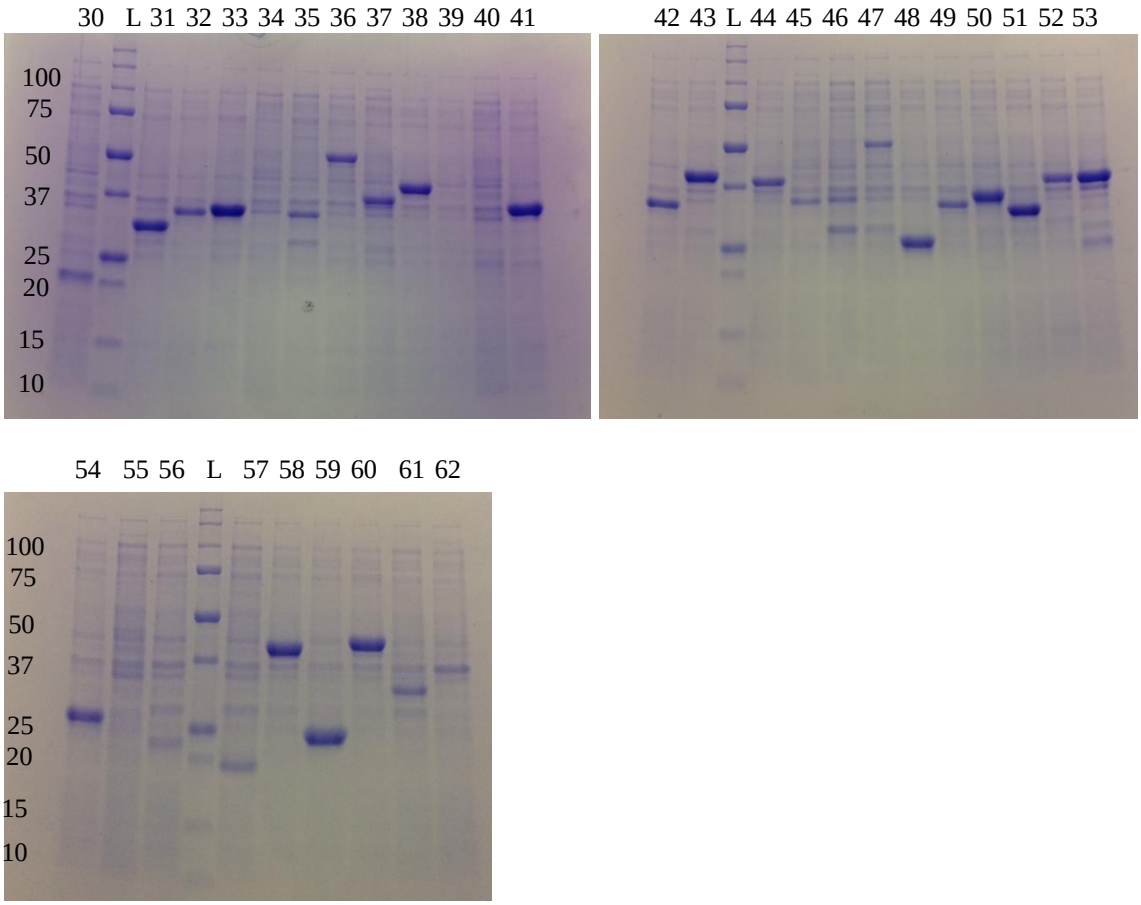


Figure 2-10. SDS-PAGE gels of *E. coli* cells recombinantly expressing proteins from the expanded metalloprotein set. L = ladder of molecular weight standards (labeled in kilodaltons, kD). Proteins are numbered as shown in Table 2-2.

From the total set of 62 proteins, we successfully expressed and purified 35 proteins. The remainder either did not express well on large scale or presented challenges in purification.

Analysis of metal binding by ICP-MS

As there are a number of metals present inside *E. coli* cells during protein expression, we wanted to determine the metal content of our purified proteins. While some proteins (e.g. the cupredoxins) display a characteristic spectroscopic signal upon metal binding, most of the proteins in our collection do not. To measure metalation, we used inductively coupled plasma-mass spectrometry (ICP-MS). This method measures all metals in solution in a given sample, not just protein-bound metals, so we used the final flow-through from the buffer exchange for each protein to determine the background concentration of non-protein-bound metals in solution. Of the proteins analyzed by ICP-MS, all showed enrichment of the desired

metal except for some α-ketoglutarate-dependent non-heme iron dioxygenases. Some proteins showed enrichment of multiple metals, either due to non-specific metal binding on the protein surface, metal binding in the His_6 purification tag, or competitive binding of multiple metals in the active site. To maximize properly metalated protein in our reactions, we further supplemented the reaction mixtures with excess exogenous metal salts.

2.5 Reaction screening with expanded protein set

Given the larger number of proteins to be tested compared to the initial set (Section 2.3), we chose a somewhat reduced set of substrates. We screened all pairwise combinations of a silane and silane acceptor substrate shown in Figure 2-11, as well as a substrate for intramolecular olefin hydrosilylation.

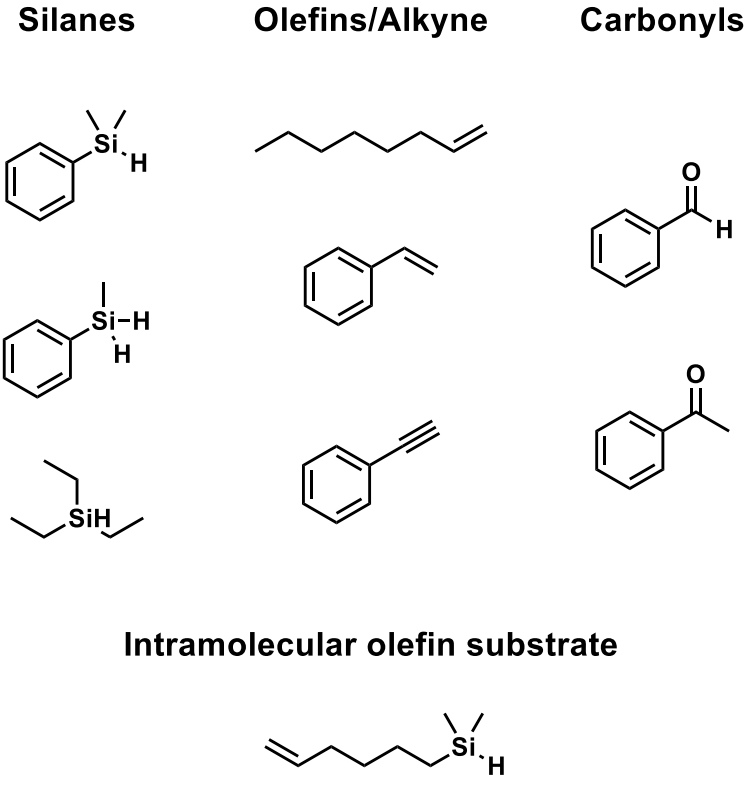


Figure 2-11. Panel of substrates tested for hydrosilylation. Reaction conditions: MOPS buffer (20 mM pH 7.0) with 5% acetonitrile as co-solvent, 23 °C, anaerobic, 18–24 h. 10 mM each substrate, ~50 μM protein, 1 mM metal salt. Reactions performed both with and without added sodium dithionite (reductant).

To maximize the likelihood of detecting successful hydrosilylation, we synthesized authentic standards of the expected products (experimental details are in section 2.8) to determine their GC-MS retention times and characteristic ionization patterns (Figure 2-12). These data were used to prepare product-specific GC-MS methods, using single ion monitoring (SIM) for the product ions, increasing the effective sensitivity of the instrument for the products of interest.

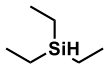
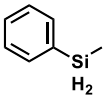
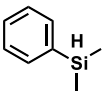
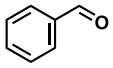
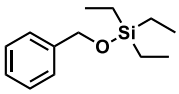
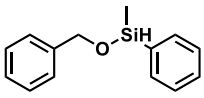
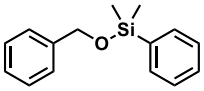
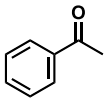
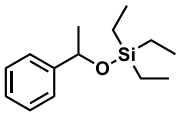
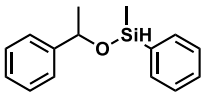
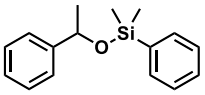
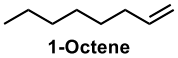
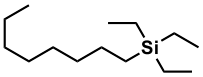
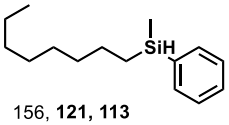
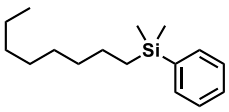
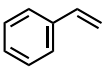
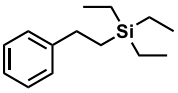
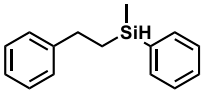
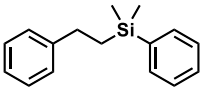
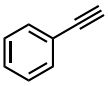
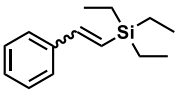
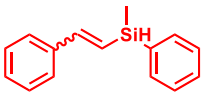
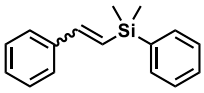
	 **Triethylsilane** Too volatile	 **Methyl(phenyl)silane** **122, 121**, 107, 105 2.53 min	 **Dimethyl(phenyl)silane** 136, 135, **121,** 105 2.93 min
 **Benzaldehyde** **106, 105, 77** 3.04 min	 **193**, 163, 135, **91** 5.22 min	 197, 183, 149, **137, 91** 5.73 min	 227, 197, **164**, 148, **91** 5.88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

Figure 2-12. Substrate/product matrix for the finalized reaction set, showing retention times and characteristic mass spectrometry ions. Samples analyzed on an Agilent GC-MS (7820 GC oven, 5977B single-quadrupole MS) with a 30 m × 0.25 mm ID × 0.25 μm film thickness DB-5MS column and the following program: hold at 80 °C for 1.7 min, ramp at 45 °C/min to 325 °C, hold at 325 °C for 2 min.

To ensure that hydrosilylation was possible under typical biocatalytic conditions (aqueous, ambient temperature), and that our substrates and products were stable and detectable under these conditions, we performed aqueous control reactions using Karstedt's catalyst[5] and Speier's catalyst[4], both well-known platinum hydrosilylation catalysts, for our olefin/alkyne hydrosilylation test reactions (3 silanes × 3 olefins/alkyne × 2 catalysts; 18 reactions in total). In all cases, we observed high consumption of the silane starting material (>50%). Reactions with the tertiary silanes (triethylsilane and dimethyl(phenyl)silane) were quite clean, with corresponding >50% yield of the expected anti-Markovnikov hydrosilylation products. Reactions with methyl(phenyl)silane showed more complex product mixtures, including hydrosilylation products as well as probable silanols. Silanols are not observed in the absence of a catalyst, suggesting that they are not simply due to background hydrolysis or degradation of the silane. As a negative control, we also performed test reactions with the metal salts used to supplement our protein-based reactions ($MnCl_2$, $Fe(NH_4)_2(SO_4)_2$, $CoCl_2$, $NiCl_2$, $CuSO_4$), to determine if there was any background reactivity from free metal ions. In no cases, did we observe any detectable hydrosilylation products from the negative controls.

We screened the 35 proteins successfully purified from the expanded protein set, as well as several proteins from the initial set, against this reaction array, but we did not find detectable protein-catalyzed hydrosilylation activity for any reaction with any of the proteins tested.

2.6 Copper/*E. coli* interaction with silanes

While we never observed detectable enzyme-catalyzed hydrosilylation, we did make some interesting observations of cellular interactions with silanes in the presence of copper.

Conjugate reduction

While searching for hydrosilylation activity with copper-binding proteins, we observed trace yields of conjugate reduction of α,β-unsaturated carbonyl compounds (Figure 2-13). Further investigation revealed that this activity was not due to any recombinantly expressed protein, and was mediated by the *E. coli* cellular background in the presence of copper salts. The

reduction is silane-specific; when triethylsilane is used instead no conjugate reduction is observed.

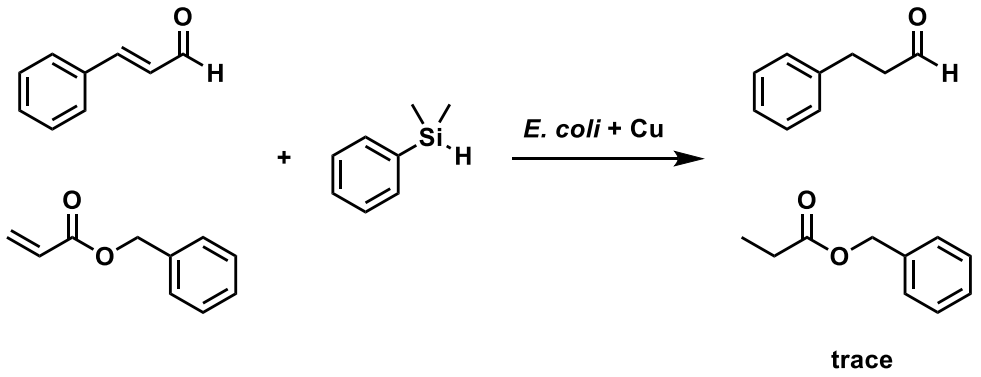


Figure 2-13. Conjugate reduction of cinnamaldehyde and benzyl acrylate by dimethyl(phenyl)silane, mediated by copper-supplemented *E. coli* lysate.

Red species

We also observed that copper-supplemented *E. coli* cell lysate, when treated with sodium dithionite and dimethyl(phenyl)silane, develops an intense red color (Figure 2-14). Similar to the observed conjugate reduction described above, other aryldimethylsilanes produce the same red color, but triethylsilane has no observed effect. The red species is unstable to oxygen, and decays over the course of a few hours when exposed to air. It does not pass through a 3 kD-cutoff centrifugal filter, suggesting that it is protein-associated, or otherwise macromolecular. We were unable to identify the cellular component that is responsible for this red color, and it is unclear how or if the red species is related to the observed conjugate reduction, but there does appear to be some kind of copper/protein-mediated silane activation by *E. coli*.

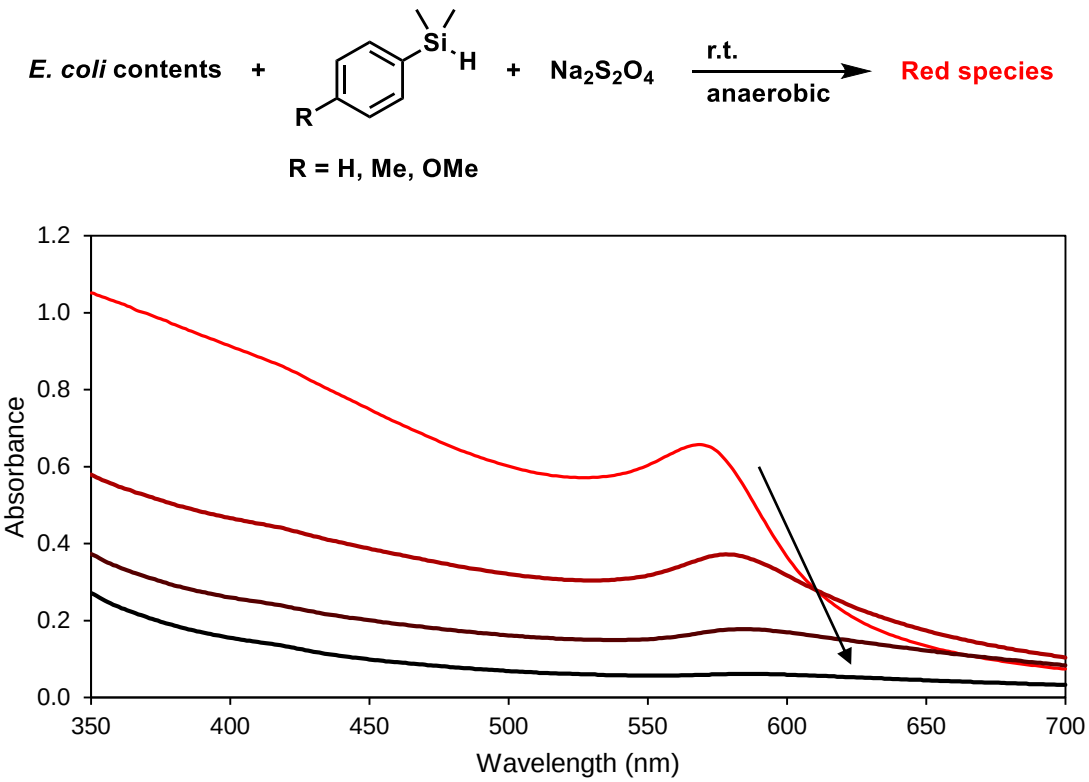


Figure 2-14. UV-vis spectrum of copper-supplemented *E. coli* lysate treated with sodium dithionite and aryldimethylsilanes.

2.7 Conclusions

We collected and assembled a diverse collection of metalloproteins drawn from the diversity evolved in Nature. These proteins include a range of metal-binding sites and natively bind a number of different first-row transition metals. We tested these proteins for a range of different olefin and carbonyl hydrosilylation reactions, but were unable to find any protein-catalyzed hydrosilylation activity.

2.8 Supplementary information for Chapter 2

Cloning

Genes encoding proteins of interest were codon-optimized for *Escherichia coli* and purchased as synthetic DNA. These genes were inserted into the pET22b(+) vector (a standard commercial protein expression vector) via Gibson assembly[26]. Genes encoding the expanded protein set (Table 2-2) were directly purchased as pre-assembled plasmids. As the intracellular concentration of free copper is negligible, genes encoding copper-binding proteins were cloned with a signal peptide (pelB leader sequence) that will localize these proteins to the periplasm, where most native *E. coli* copper-binding proteins are found. The assembled plasmids were used to transform *E. coli* strain BL21(DE3) cells by electroporation. Site-directed mutagenesis was performed by polymerase chain reaction (PCR) with primers containing the desired mutation (Table 2-3). Site-saturation mutagenesis was performed with the 22-codon method[27].

Table 2-3. Primers for mutagenesis. Mutation sites are denoted here as *NNN* for simplicity; in practice, they are a 12:9:1 mixture of NDT:VHG:TGG for site saturation.

Primer name	Sequence
Pae azurin H117X forward	GTTCTTCTGCACATTCCCTGGA***NNN***TCCGCTCTGATGAAGG GAACG
Pae azurin H117X reverse	TCCAGGGAATGTGCAGAAGAACATATACTGTTCCCCCTC
Pae azurin M121G forward	CATTCCCTGGACACTCCGCTCTG***NNN***AAGGGAACGCTGAC ACTTAAA
Pae azurin M121G reverse	CAGAGCGGAGTGTCCAGGGAATGTGCAGAAGAACATATAC TG
Pde amicyanin H95X forward	CGACTATCACTGCACGCCG***NNN***CCTTTCATGCGCGGTAAA GTAGTA
Pde amicyanin H95X reverse	CGGCGTGCAGTGATAGTCGTAAGTACCGGCTTCCGTG
Pde amicyanin M98X forward	ACTGCACGCCGCACCCTTTC***NNN***CGCGGTAAAGTAGTAGT CGAA
Pde amicyanin M98X reverse	GAAAGGGTGCGGCGTGCAGTGATAGTCGTAAGTACC

General protein expression protocol

The following general protein expression protocol was used for most proteins: *E. coli* BL21(DE3) cells carrying a plasmid encoding the protein of interest were grown overnight in Luria-Bertani supplemented with ampicillin (LB_{amp}) medium at 37 °C and 220 rpm. The preculture was used to inoculate an expression culture in Terrific Broth supplemented with ampicillin (TB_{amp}) medium (1% v/v preculture). The expression culture was grown at 37 °C and 220 rpm for 2–3 hours to an optical density at 600 nm of 0.7–0.9 and then cooled on ice. Isopropyl β-D-glucopyranoside (IPTG, 50–500 μM final concentration) was added to induce protein expression, along with the appropriate metal salt for non-iron proteins (e.g. $NiCl_2$, 1 mM final concentration), and the proteins were expressed at 22 °C and 170 rpm overnight. Following expression, the cultures were centrifuged at 4 °C and 5,000 *g* for 5 min. The cell pellets were then resuspended in MOPS buffer (20 mM, pH 7.0, approx. 5 mL g^{-1} wet cells) containing the appropriate metal salt (e.g. $NiCl_2$, 5 mM final concentration). These suspensions were used directly for whole-cell reactions. For reactions in lysate, the cell suspensions (approx. 5 mL g^{-1} wet cells) were sonicated and the resulting lysate was clarified by centrifugation (4 °C, 20,000 *g*, 10 min) to remove precipitated cell debris.

Protein purification

For protein purification, protein constructs with a C-terminal His_6 tag for Ni^{2+}-affinity chromatography were expressed as described above in 1 L cultures, harvested, and resuspended in binding buffer (20 mM Tris·HCl pH 7.5, 20 mM imidazole, 100 mM NaCl). The resuspended cells were lysed via sonication as described above. The clarified lysate was loaded onto a Ni^{2+}-NTA column using an Äkta FPLC system. The column-bound protein was washed with binding buffer (20 mM Tris·HCl pH 7.5, 20 mM imidazole, 100 mM NaCl) and then eluted by a linear gradient with elution buffer (20 mM Tris·HCl pH 7.5, 500 mM imidazole, 100 mM NaCl). Fractions containing the desired protein were pooled, concentrated, and subjected to three exchanges of the storage buffer (20 mM MOPS, pH 7.0) by centrifugal filtration (4 °C, 20 min, 4,000 *g*). Concentrated proteins were divided into 100 μL aliquots and flash-frozen on dry ice for storage at –80 °C.

Superoxide dismutase metalation

Superoxide dismutases from hyperthermophilic organisms (*Thermus thermophilus* and *Pyrobaculum aerophilum*) were metalated by heat treatment, as they express as apoproteins in mesophilic hosts such as *E. coli*[21]. Purified proteins were diluted to approximately 100 μM in MOPS buffer (20 mM, pH 7.0) and a metal salt ($MnCl_2$, $(NH_4)_2Fe(SO_4)_2$, $CoCl_2$, $NiCl_2$, or $CuSO_4$) was added to a final concentration of 10 mM. Manipulations involving iron(II) were performed anaerobically to minimize precipitation of iron(III). The resulting solution was heated in a water bath to 70 °C (*T. thermophilus*) or 75 °C (*P. aerophilum*) for 45 minutes, and then cooled to room temperature. Unbound metal ions were removed by buffer exchange with MOPS (20 mM, pH 7.0) by iterative centrifugal concentration as described above.

Protein inductively coupled plasma – mass spectrometry (ICP-MS)

Metalation of proteins of interest was determined via inductively coupled plasma – mass spectrometry (ICP-MS). Purified protein samples are diluted 100-fold to ca. 5–10 μM in 2% nitric acid and MQ-H_2O. The negative control for each sample is the final "flow-through" buffer from that protein's centrifugal concentration and should have approximately the same background metal concentration as the protein sample. These samples were run on an Agilent 8800 triple quadrupole ICP-MS. The counts per second (CPS) for each ion of interest are compared against calibration curves prepared from an ICP-MS standard transition-metal mixture (diluted from 100 μg mL^{-1} in 2% nitric acid).

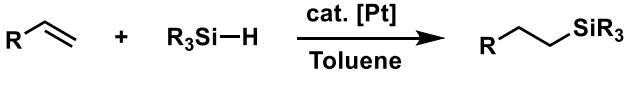
Hydrosilylation reactions

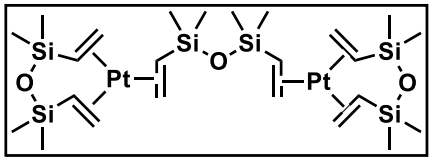
Reactions were performed with whole cells, cell lysates, or purified proteins. All reactions were set up in an anaerobic chamber (<10 ppm O_2) and sealed until workup. Reactions were performed in MOPS buffer (20 mM, pH 7.0), with 5% (v/v) acetonitrile as co-solvent to facilitate dissolution of organic compounds. After the reaction, the organic compounds were extracted into an organic solvent (typically 1:1 cyclohexane/ethyl acetate) and analyzed by GC-MS.

Synthesis of authentic hydrosilylation product standards

Authentic standards of the hydrosilylation products were synthesized so they could be qualitatively identified by GC-MS. The authentic standards were analyzed directly from reaction mixtures and were not isolated or purified.

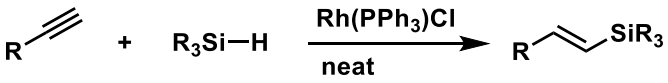
Olefin hydrosilylation products





Under air, olefin (1.00 mmol, 1.00 equiv.) and silane (3.00 mmol, 3.00 equiv.) were dissolved in toluene (1 mL) in an 8 mL glass vial. Karstedt's catalyst (2.2% w/v Pt in xylenes, 104 µL, 0.01 mmol, 0.01 equiv.) was added, and the reaction mixture was heated to 60 °C for 4 hours. The reaction was cooled to room temperature, then partitioned between water and ethyl acetate (10 mL each). The organic layer was washed with water (2×10 mL) and brine (10 mL), then dried over magnesium sulfate. This solution was used directly for qualitative GC-MS analysis.

Alkyne hydrosilylation products

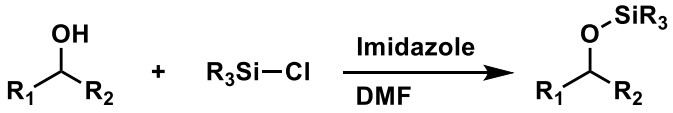


Caution: This reaction should only be performed with tertiary silanes. The use of a secondary silane resulted in a violent exotherm.

Under argon, tris(triphenylphosphine)rhodium(I) chloride (9.25 mg, 0.01 mmol, 0.01 equiv.) was suspended in alkyne (1.00 mmol, 1.00 equiv.) in a 2 mL glass vial. Silane (1.00 mmol, 1.00 equiv.) was then added, and the reaction mixture was stirred at room temperature for 16 hours. The reaction mixture was partitioned between water and ethyl acetate (10 mL each). The organic layer was washed with water (2×10 mL) and brine (10 mL), then dried over magnesium sulfate. This solution was used directly for qualitative GC-MS analysis.

Carbonyl hydrosilylation products

Carbonyl hydrosilylation products were synthesized by silylation of the corresponding alcohol, rather than by direct hydrosilylation.



Alcohol (1.00 mmol, 1.00 equiv.) and imidazole (204 mg, 3.00 mmol, 3.00 equiv.) were dissolved in *N*,*N*-dimethylformamide (2 mL) and cooled to 0 °C. Chlorosilane (1.50 mmol, 1.50 equiv.) was added dropwise. The reaction was warmed to room temperature and stirred for 30 minutes. The organic layer was washed with water (2×10 mL) and brine (10 mL), then dried over magnesium sulfate. This solution was used directly for qualitative GC-MS analysis.

2.9 References for Chapter 2

[1] Lewis, L.N.; Stein, J.; Gao, Y.; Colborn, R.E.; Hutchins, G. Platinum Catalysts Used in the Silicones Industry. *Platinum Metals Rev.*, **1997**, *41*, 66-75.

[2] Lukin, R. Y.; Kuchkaev, A. M.; Sukhov, A. V.; Bekmukhamedov, G. E.; Yakhvarov, D. G. Platinum-Catalyzed Hydrosilylation in Polymer Chemistry. *Polymers* **2020**, *12*, 2174.

[3] Nakajima, Y.; Shimada, S. Hydrosilylation reaction of olefins: recent advances and perspectives. *RSC Adv*. **2015**, *5*, 20603–20616.

[4] Speier, J. L.; Webster, J. A.; Barnes, G. H. The Addition of Silicon Hydrides to Olefinic Double Bonds. Part II. The Use of Group VIII Metal Catalysts. *J. Am. Chem. Soc.* **1957**, *79*, 974–979.

[5] Karstedt, B. D. Platinum complexes of unsaturated siloxanes and platinum containing organopolysiloxanes. US 3775452A, November 27, 1973.

[6] Onopchenko, A.; Sabourin, E. T.; Beach, D. L. Rhodium(I)-catalyzed hydrosilylation of styrene. *J. Org. Chem.* **1983**, *48*, 5101–5105.

[7] Apple, D. C.; Brady, K. A.; Chance, J. M.; Heard, N. E.; Nile, T. A. Iridium Complexes as Hydrosilylation Catalysts. *J. Mol. Catal.* **1985**, *29*, 55–64.

[8] Du, X.; Huang, Z. Advances in Base-Metal-Catalyzed Alkene Hydrosilylation. *ACS Catal.* **2017**, *7*, 1227–1243.

[9] Tondreau, A. M.; Hojilla Atienza C. C.; Weller, K. J.; Nye, S. A.; Lewis, K. M.; Delis, J. G.; Chirik, P. J. Iron Catalysts for Selective Anti-Markovnikov Alkene Hydrosilylation Using Tertiary Silanes. *Science* **2012**, *335*, 567–570.

[10] Maciejewski, H.; Marciniec, B.; Kownacki, I. Catalysis of hydrosilylation: Part XXXIV. High catalytic efficiency of the nickel equivalent of Karstedt catalyst [{Ni(η-CH_2=CHSiMe_2)$_2$O}$_2${μ-(η-CH_2=CHSiMe_2)$_2$O}]. *J. Organomet. Chem.* **2000**, *597*, 175–181.

[11] Gribble, Jr., M. W.; Pirnot, M. T.; Bandar, J. S.; Liu, R. Y.; Buchwald, S. L. Asymmetric Copper Hydride-Catalyzed Markovnikov Hydrosilylation of Vinylarenes and Vinyl Heterocycles. *J. Am. Chem. Soc.* **2017**, *139*, 2192–2195.

[12] Troegel, D.; Stohrer, J. Recent advances and actual challenges in late transition metal catalyzed hydrosilylation of olefins from an industrial point of view. *Coord. Chem. Rev*. **2011**, *255*, 1440–1459.

[13] Chalk, A. J.; Harrod, J. F. Homogeneous Catalysis. II. The Mechanism of the Hydrosilation of Olefins Catalyzed by Group VIII Metal Complexes. *J. Am. Chem. Soc.* **1965**, *87*, 16–21.

[14] Marciniec, B., Ed. Hydrosilylation: A Comprehensive Review on Recent Advances. Springer: Berlin, 2009.

[15] Glaser, P.B.; Tilley, T.D. Catalytic Hydrosilylation of Alkenes by a Ruthenium Silylene Complex. Evidence for a New Hydrosilylation Mechanism. *J. Am. Chem. Soc.*, **2003**, *125*, 13640–13641.

[16] Bornscheuer, U.T.; Huisman, G.W.; Kazlauskas, R.J.; Lutz, S.; Moore, J.C.; Robins, K. Engineering the third wave of biocatalysis. *Nature* **2012**, *485*, 185–194.

[17] Degtyarenko, K. Metalloproteins. In *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics*; Jorde, L., Little, P., Dunn, M., Subramaniam, S., Eds.; John Wiley & Sons, Ltd.: New York, 2005.

[18] Díez-González, S.; Nolan, S. P. Copper, Silver, and Gold Complexes in Hydrosilylation Reactions. *Acc. Chem. Res.* **2008**, *41*, 349–358.

[19] Wang, Z.-L.; Zhang, F.-L.; Xu, J.-L.; Shan, C.-C.; Zhao, M.; Xu, Y.-H. Copper-Catalyzed *Anti*-Markovnikov Hydrosilylation of Terminal Alkynes. *Org. Lett.* **2020**, *22*, 7735–7742.

[20] Solomon, E. I.; Hadt, R. G. Recent advances in understanding blue copper proteins. *Coord. Chem. Rev.* **2011**, *255*, 744–789.

[21] Whittaker, J.W. The irony of manganese superoxide dismutase. *Biochem. Soc. Trans.* **2003**, *31*, 1318–1321.

[22] Jaroszewski, L.; Schwarzenbacher, R.; von Delft, F.; McMullan, D.; Brinen, L. S.; Canaves, J. M.; Dai, X.; Deacon, A. M.; DiDonato, M.; Elsliger, M. A.; Eshagi, S.; Floyd, R.; Godzik, A.; Grittini, C.; Grzechnik, S. K.; Hampton, E.; Levin, I.; Karlak, C.; Klock, H. E.; Koesema, E.; Kovarik, J. S.; Kreusch, A.; Kuhn, P.; Lesley, S. A.; McPhillips, T. M.; Miller, M. D.; Morse, A.; Moy, K.; Ouyang, J.; Page, R.; Quijano, K.; Reyes, R.; Rezezadeh, F.; Robb, A.; Sims, E.; Spraggon, G.; Stevens, R. C.; van den Bedem, H.; Velasquez, J.; Vincent, J.; Wang, X.; West, B.; Wolf, G.; Xu, Q.; Hodgson, K. O.; Wooley, J.; Wilson, I. A. Crystal structure of a novel manganese-containing cupin (TM1459) from Thermotoga maritima at 1.65 A resolution. *Proteins* **2004**, *56*, 611–614.

[23] Jeoung, J. H.; Nianios, D.; Fetzner, S.; Dobbek, H. Quercetin 2,4-Dioxygenase Activates Dioxygen in a Side-On O_2-Ni Complex. *Angew. Chem. Int. Ed.* **2016**, *55*, 3281–3284.

[24] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

[25] Andreini, C.; Cavallaro, G.; Lorenzini, S.; Rosato, A. MetalPDB: a database of metal sites in biological macromolecular structures. *Nucleic Acids Res.* **2013**, *41*, D312–D319.

[26] Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **2009**, *6*, 343–345.

[27] Kille, S.; Acevedo-Rocha, C. G.; Parra, L. P.; Zhang, Z.-G.; Opperman, D. J.; Reetz, M. T.; Acevedo, J. P. Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synth. Biol.*, **2013**, *2*, 83–92.

Chapter 3

NON-HEME IRON ENZYME-CATALYZED NITRENE TRANSFER

Content in this chapter is adapted from published work:

† denotes equal contribution

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

N.W.G. and R.K.Z. performed initial aziridination reaction discovery. N.W.G. and A.M.K. performed further reaction characterization, performed directed evolution for aziridination, and developed analytical methods. N.W.G. discovered and characterized C–H insertion activity and performed synthetic chemistry.

Abstract

α-Ketoglutarate-dependent dioxygenases are a family of non-heme iron-dependent dioxygenases that in nature catalyze similar chemistry to cytochromes P450 and other heme-dependent dioxygenases. We found that *Pseudomonas savastanoi* ethylene-forming enzyme (*Ps*EFE), an α-ketoglutarate dependent dioxygenase, can catalyze olefin aziridination and C–H sulfamidation by nitrene transfer, through a non-native catalytic mechanism. We have improved these non-natural activities of *Ps*EFE by directed evolution. The non-heme iron center allows for the coordination of small-molecule ligands to the active-site iron, which can exert significant effects on the activity and selectivity of *Ps*EFE for nitrene transfer chemistry.

3.1 α-Ketoglutarate-dependent iron dioxygenases

α-Ketoglutarate (αKG)-dependent iron dioxygenases are a family of non-heme iron enzymes that catalyze a variety of oxidative transformations in nature, using molecular oxygen and α-ketoglutarate as co-substrates[1]. The catalytic mechanism of α-KG-dependent dioxygenases is somewhat similar to that of the heme-dependent cytochrome P450 family, in which a high-valent iron oxo complex is the key oxidizing catalytic intermediate. Unlike cytochromes P450, however, which perform O–O bond scission to generate the iron oxo by reduction with NAD(P)H, αKG-dependent dioxygenases use α-ketoglutarate as a co-substrate, which is oxidatively decarboxylated to succinate to generate the iron oxo (Figure 3-1).

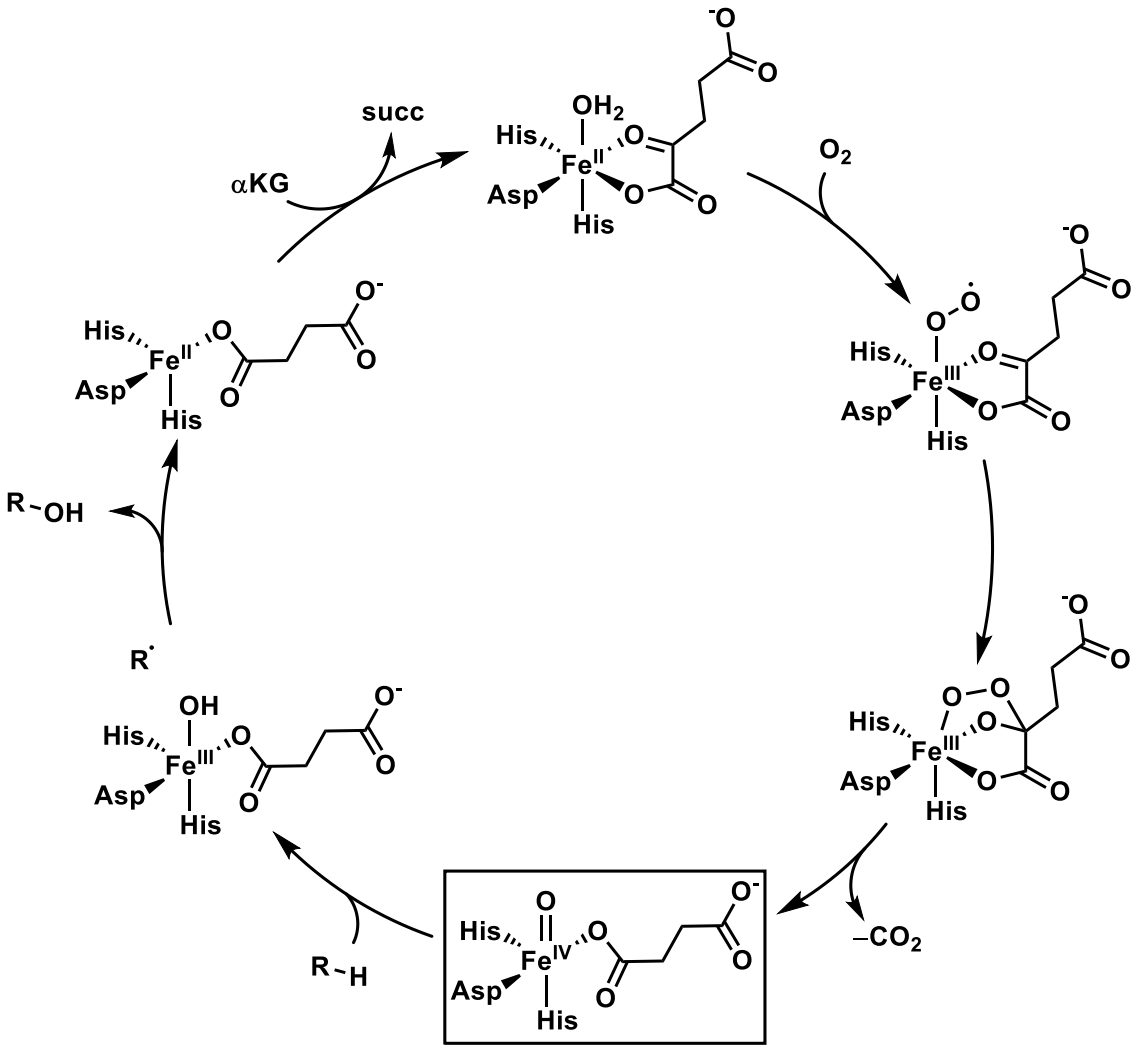


Figure 3-1. Canonical mechanism of α-KG-dependent iron dioxygenases, shown here for C–H hydroxylation.

α-KG-dependent dioxygenases are well-known for C–H hydroxylation[2]. In hydroxylation reactions, the iron-oxo intermediate abstracts a hydrogen atom from the substrate, and the resulting carbon radical recombines with the iron hydroxyl to form the hydroxylated product (Figure 3-1). The most common post-translational modification in human proteins is the hydroxylation of proline residues in collagen catalyzed by an α-KG-dependent hydroxylase[3], and other members of this family also catalyze a range of other protein modifications. Other hydroxylases of this family are also involved in a number of secondary metabolite biosynthetic pathways[1,4] (Figure 3-2).

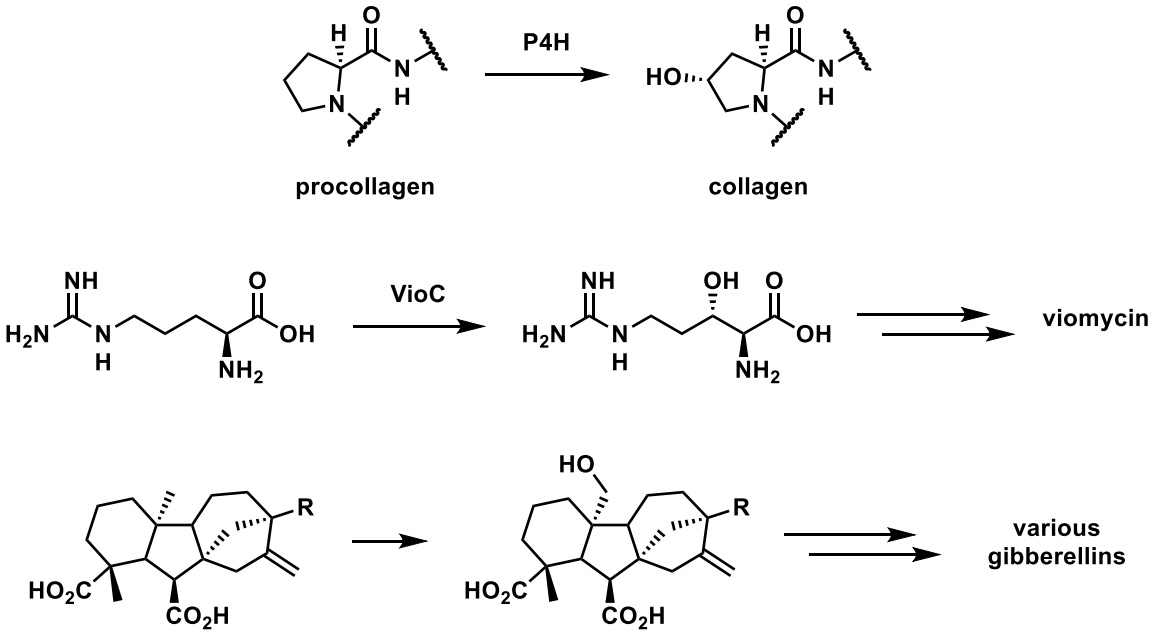


Figure 3-2. Examples of hydroxylations catalyzed by α-KG-dependent dioxygenases.

In addition to hydroxylation, α-KG-dependent dioxygenases catalyze a wide variety of other oxidative transformations in nature (Figure 3-3). Some catalyze halogenation[5], where radical rebound occurs with an iron-bound halide instead of the iron hydroxyl. Others catalyze various ring-forming or ring-expanding reactions, as exemplified by the biosyntheses of clavaminate[6] and cephalosporin antibiotics[7].

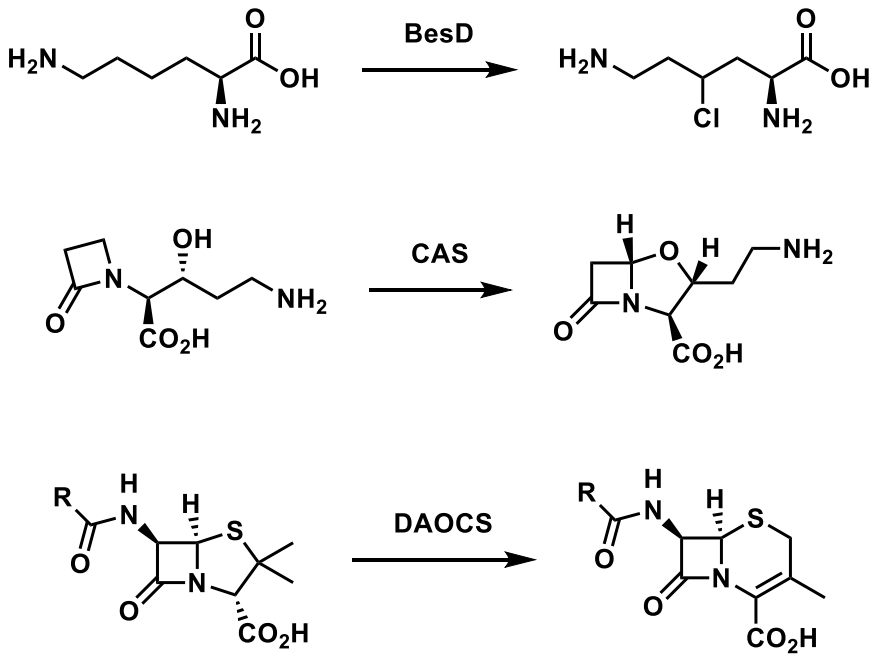


Figure 3-3. Examples of non-hydroxylase oxidations catalyzed by αKG-dependent dioxygenases.

In addition to their roles in natural biochemistry, α-KG-dependent dioxygenases have also found use in synthetic biocatalysis. Many of these enzymes natively catalyze the site- and stereoselective hydroxylation of amino acids, which form useful intermediates for the total synthesis of many bioactive natural products. The Renata group has shown that GriE, natively an L-leucine hydroxylase, can accept a range of analogs of the native substrate, allowing the chemoenzymatic total synthesis of manzacidin C[8] (Figure 3-4A). The same group has also employed various L-lysine hydroxylases, with complementary selectivity, in the total syntheses of tambromycin[9] and cepafungin I[10] (Figure 3-4B, C).

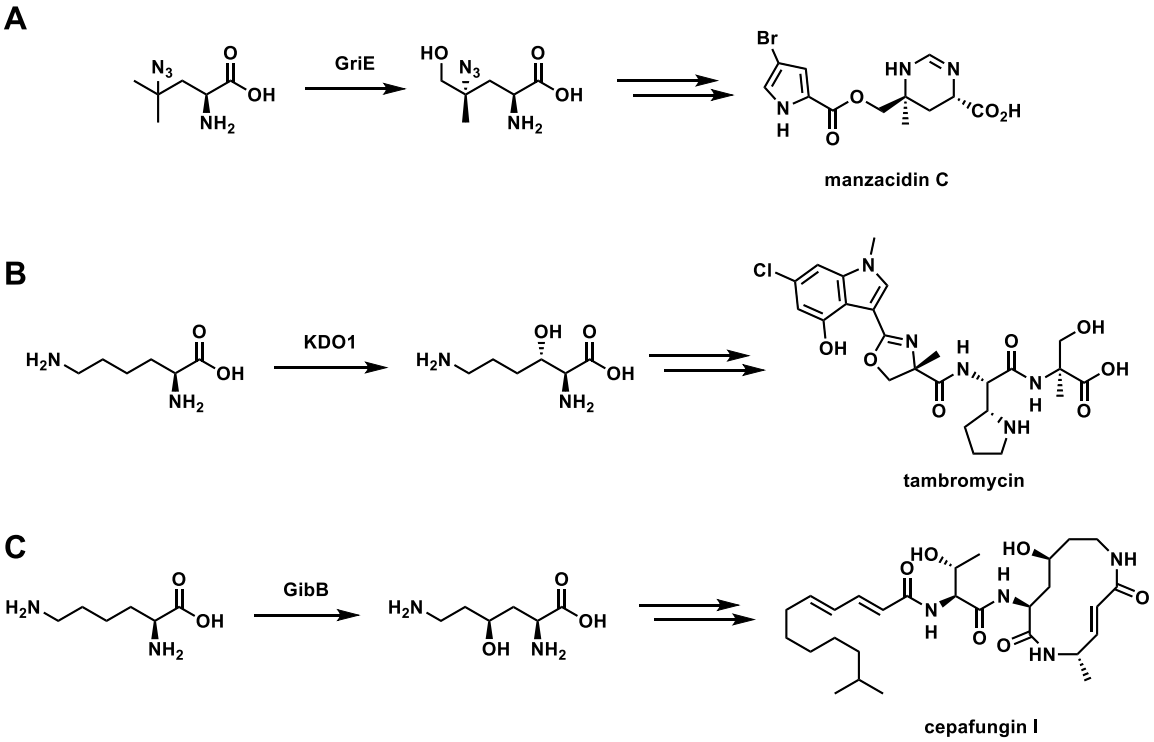


Figure 3-4. Examples of α-KG dioxygenase-catalyzed hydroxylations in the chemoenzymatic total synthesis of natural products.

The α-KG-dependent halogenases have also attracted attention for their potential uses in synthetic biocatalysis. The Bollinger/Krebs group has shown that SyrB2, natively a chlorinase, can catalyze nitration and azidation as well, when nitrate and azide are present instead of chloride[11] (Figure 3-5).

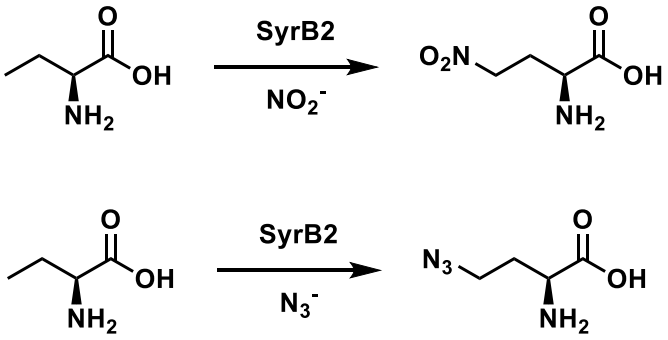


Figure 3-5. Biosynthetic applications of α-KG-dependent halogenases.

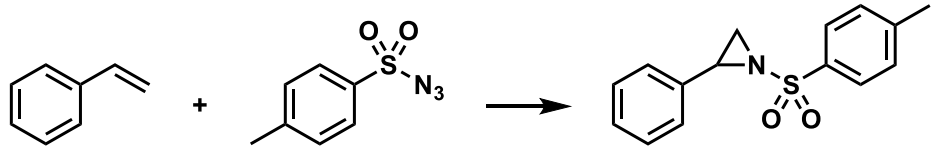
All activities of α-KG-dependent dioxygenases previously demonstrated, either natural or engineered, proceed by essentially the native mechanism, through the key iron-oxo

intermediate. Given the similarity to heme-dependent oxygenases and the breadth of non-native chemistry demonstrated with them, we asked if α-KG-dependent dioxygenases could also be persuaded to catalyze abiological reactions through non-native mechanistic pathways.

3.2 Initial discovery of nitrene-transfer activity

We screened a set of seven wild-type α-KG-dependent dioxygenases that we had previously expressed and purified for hydrosilylation reaction screening (details are in chapter 2) against the intermolecular aziridination of styrene and *p*-toluenesulfonyl azide. Bovine serum albumin (BSA) was included as a negative control to determine non-specific background activity. Activity was determined by LC-MS, measuring at *m/z* = +274 ($[M+H]^+$ for the aziridine product). From this set, one enzyme, *Pseudomonas savastanoi* ethylene-forming enzyme (*Ps*EFE), displayed aziridination activity significantly higher than background (Table 3-1).

Table 3-1. Initial aziridination activity search.



Enzyme	UniProt ID	Relative activity
***P. savastanoi* ethylene-forming enzyme**	**P32021**	**12.0**
Streptomyces sp. 2-aminobutyric acid chlorinase	D0VX22	0.93
A. thaliana anthocyanidin synthase	Q96323	0.54
G. oxydans leucine dioxygenase	Q5FQD2	1.11
E. coli taurine dioxygenase	P37610	0.61
S. vinaceus arginine hydroxylase	Q6WZB0	0.57
S. muensis leucine hydroxylase	A0A0E3URV8	0.61
Bovine serum albumin (negative control)	P02769	1.00

Reaction conditions: 50 μM enzyme, 1 mM $Fe(NH_4)_2(SO_4)_2$, 1 mM αKG, 1 mM L-ascorbic acid, and 10 mM styrene and *p*-toluenesulfonyl azide. Activities are normalized to the negative control, bovine serum albumin.

*Ps*EFE is a rather atypical member of the αKG-dependent iron dioxygenase family. In nature it catalyzes two reactions (Figure 3-6)[12]. The first, typical of enzymes in this family, is the hydroxylation of L-arginine in the 5-position. The resulting hydroxylated product spontaneously eliminates urea and cyclizes to form pyrroline-5-carboxylate (P5C). The second reaction catalyzed by *Ps*EFE, the formation of ethylene, is more unusual. After generation of the iron-oxo intermediate, the bound succinate is oxidatively decarboxylated to ethylene. Structurally, *Ps*EFE is somewhat distinct from other members of its family. It displays a hybrid fold, with elements of both a type I and a type II α-KG dependent enzyme[13]. It binds α-KG in an unusual conformation in an unusually hydrophobic pocket, which is likely responsible for the atypical catalytic activity[14].

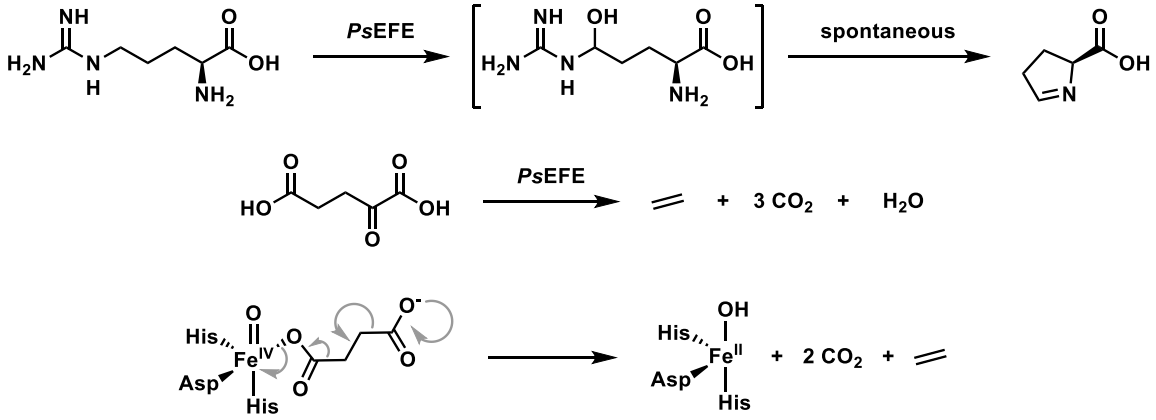


Figure 3-6. Native activities of *Ps*EFE.

3.3 Reaction characterization and controls

As the active site of *Ps*EFE is quite different from those of the heme-binding proteins previously reported to catalyze similar nitrene transfer chemistry, and as nitrene transfer is quite different from the native activities of *Ps*EFE, we sought to characterize the necessary components of the nitrene transfer reaction.

We prepared the H189A D191A double mutant, in which two of the three metal-coordinating residues have been mutated to alanine, which cannot coordinate a metal (Figure 3-7). This variant is not expected to bind iron and indeed does not catalyze aziridination. Furthermore,

purified wild-type apoenzyme is inactive, and aziridination activity is restored by addition of an Fe(II) salt (Table 3-2). Together, these observations demonstrate that the observed nitrene transfer activity is catalyzed by an iron ion bound in the enzyme active site.

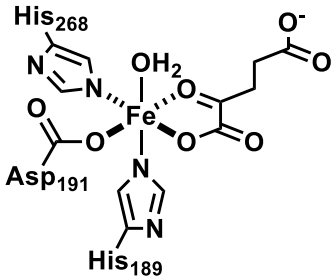


Figure 3-7. Metal-binding site of wild-type *Ps*EFE, depicted with water and αKG bound.

While α-ketoglutarate is required as a stoichiometric (co-)substrate for the native dioxygenase activities of *Ps*EFE, it is not expected to be consumed in the course of nitrene transfer. Indeed, while addition of αKG does improve activity twofold, other carboxylate additives provide much higher activity enhancements. In particular, *N*-oxalylglycine (NOG) is a structural analog of αKG and a competitive inhibitor of the native activities of *Ps*EFE and other enzymes of its family[15], but it gives 8-fold higher activity for nitrene transfer than αKG and 16-fold higher activity than no additive. Acetate, which is ubiquitous in the cell and trivially inexpensive, performs nearly as well as NOG (Table 3-2).

Table 3-2. Aziridination reaction controls with wild-type *Ps*EFE.

Deviation from standard conditions	Aziridine yield (%)
None	0.56%
No iron	0.01%
No ascorbate	0.50%
No acetate	0.04%
αKG instead of acetate	0.08%
Succinate instead of acetate	0.11%
N-oxalylglycine instead of acetate	0.64%
H189A D191A mutant	<0.01%

Standard conditions: 20 μM purified apoenzyme, 1 mM $Fe(NH_4)_2(SO_4)_2$, 1 mM acetate, 1 mM L-ascorbic acid, and 10 mM styrene and *p*-toluenesulfonyl azide.

We then measured the aziridination activity of *Ps*EFE in whole *E. coli* cells and in clarified lysate. Protein purification is laborious and time-consuming, and is prohibitive for any kind of medium- to high-throughput screening workflow, so activity in the unpurified cellular milieu is highly desirable. Fortunately, *Ps*EFE is active for nitrene transfer in both whole cells and lysate (Figure 3-8). The activity is dependent on the cell density, as expected, and is approximately two-fold higher in lysate than in whole cells. The activity drops at lower iron supplementation, particularly in whole cells, indicating that the natural iron concentration in the cells is insufficient to fully metallate the expressed enzyme. Cells expressing a tryptophan synthase variant (TrpB), a non-metal-dependent enzyme, were included in this experiment as a negative control to determine the background cellular activity.

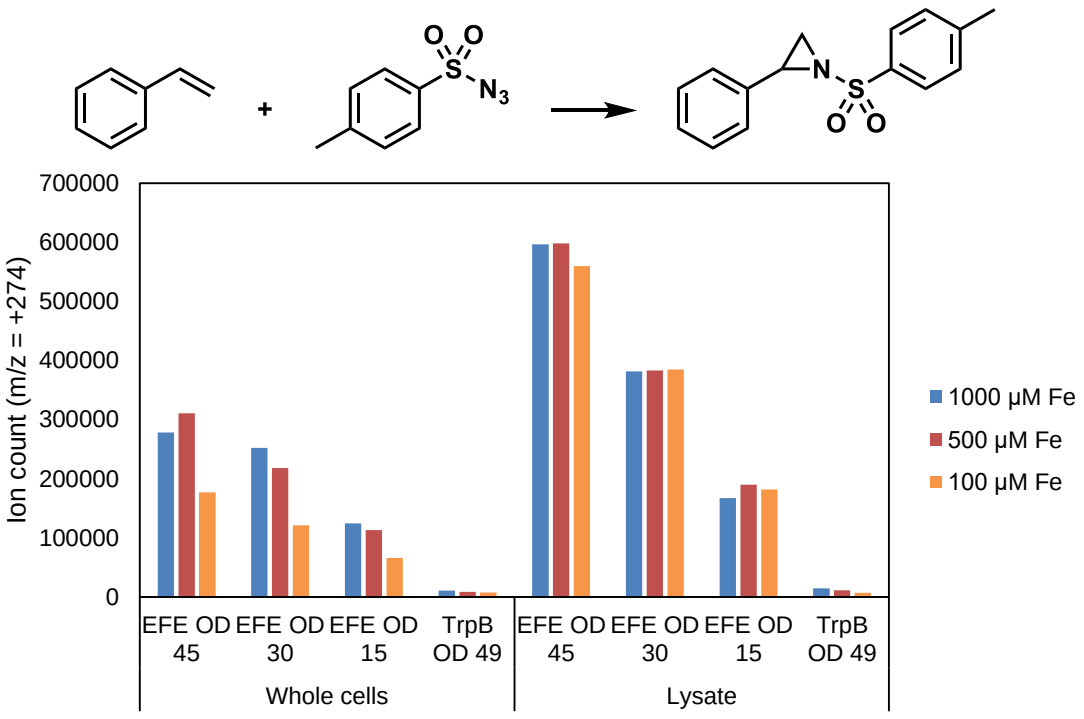


Figure 3-8. Aziridination activity in whole *E. coli* cells expressing *Ps*EFE and the corresponding lysate. Reaction conditions: 100–1000 μM $Fe(NH_4)_2(SO_4)_2$, 1 mM L-ascorbic acid, and 10 mM styrene and *p*-toluenesulfonyl azide.

3.4 Directed evolution for improved aziridination activity

We sought to improve activity for the aziridination of styrene and *p*-toluenesulfonyl azide by directed evolution of *Ps*EFE. We chose a strategy of site-saturation mutagenesis, in which residues in or near the active site (as determined by published structural data) are randomized to all 20 canonical amino acids. While mutations throughout the entire protein can be beneficial, active-site residues are more likely to have a significant effect on activity, so the incidence of beneficial mutations in focused site-saturation libraries is generally higher than in libraries generated by random mutagenesis. Our activity assay relied on LC-MS screening, which permits analysis of hundreds of samples at a time, less than the thousands or tens of thousands generally required for successful random mutagenesis screening.

In the first round of mutagenesis, we created 24 single-site saturation libraries (targeting amino-acid positions 84, 86, 87, 91, 171, 173, 175, 186, 189, 191, 192, 198, 206, 228, 268, 270, 277, 279, 281, 283, 314, 316, 317, and 318). Of these, we successfully cloned, expressed, and screened 15 (positions 84, 86, 87, 91, 171, 173, 186, 189, 191, 192, 277, 283, 314, 316, and 317) (Figure 3-9). 80 variants from each library were screened, along with 8 parent clones and 8 sterile controls, giving approximately four-fold oversampling. In this round, reactions were screened anaerobically with supplemented α-ketoglutarate. Potentially improved variants were regrown in shake flasks and their aziridination activity was measured in lysate.

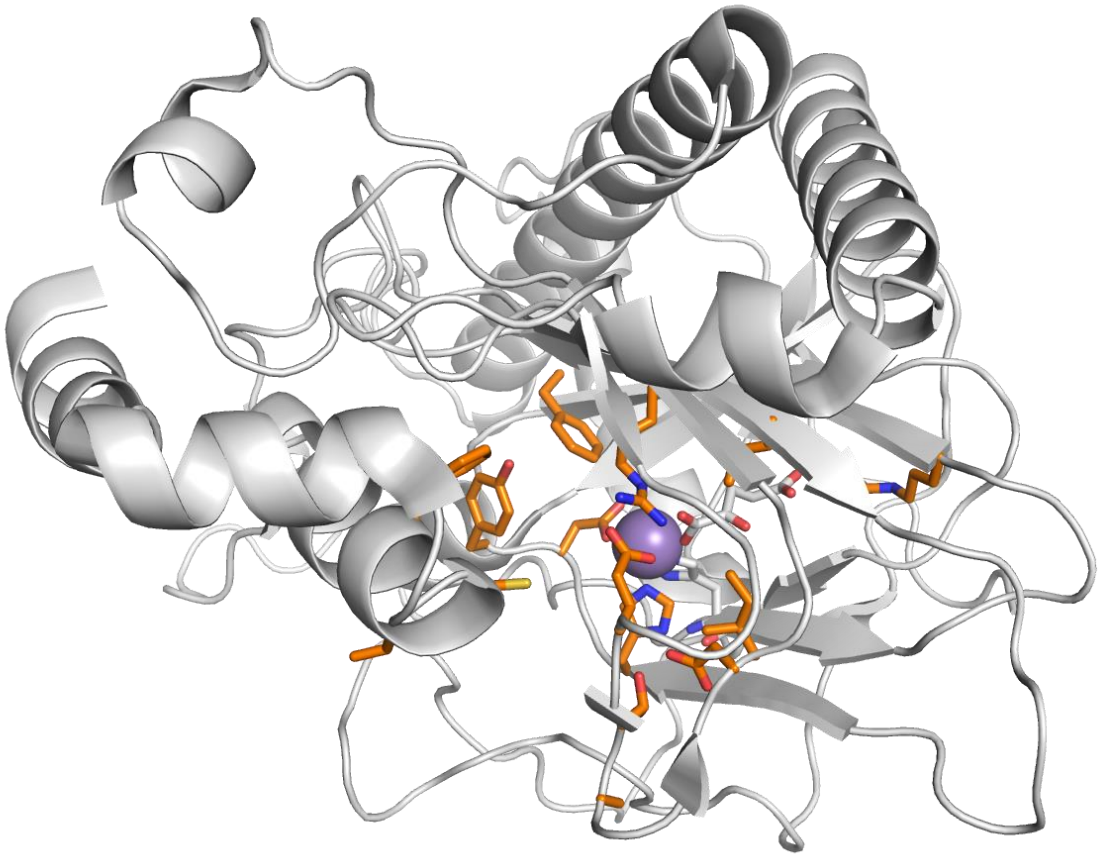


Figure 3-9. Structural representation of *Ps*EFE complexed with manganese (purple sphere) and αKG (sticks) (PDB ID: 5VKB). The side chains of the 15 sites successfully targeted in the first round are highlighted in orange.

In this round, several beneficial mutations were discovered, but of these, C317M gave by far the largest increase in activity (Figure 3-10). We chose this variant as the parent for the next round of mutagenesis.

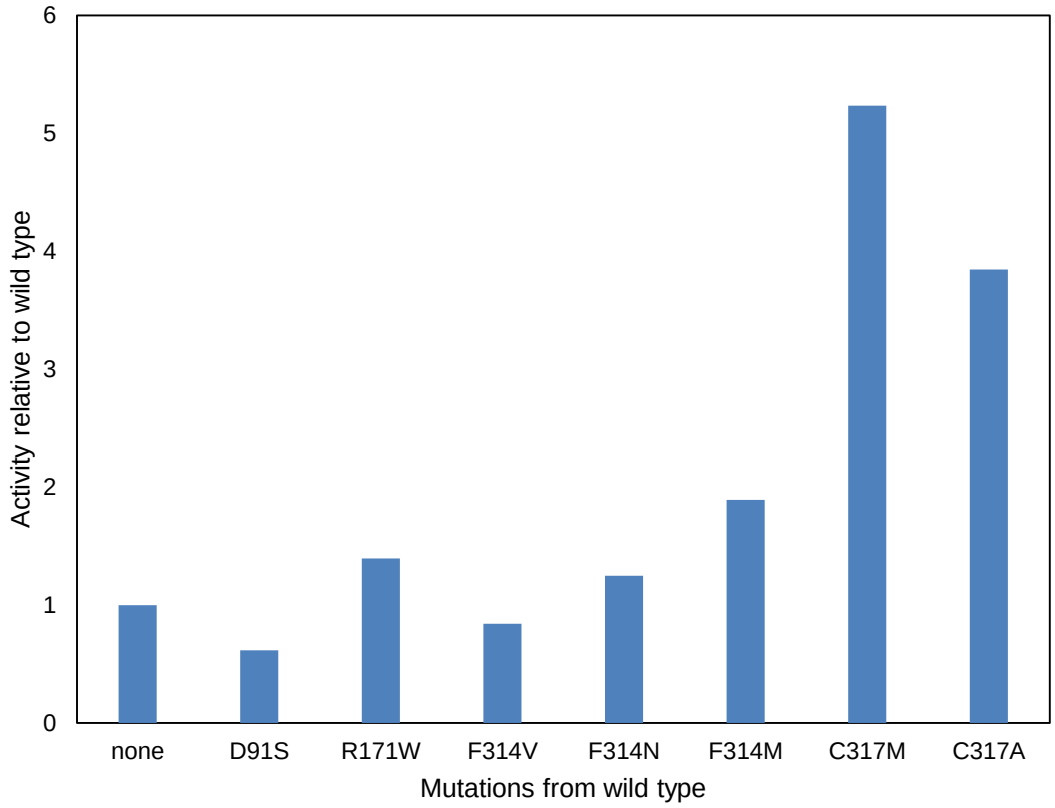


Figure 3-10. Aziridination activity of variants found in the first round of directed evolution. Reactions performed in clarified *E. coli* lysate with 5% ethanol co-solvent, 1 mM $Fe(NH_4)_2(SO_4)_2$, 1 mM acetate, 1 mM L-ascorbic acid, and 10 mM styrene and *p*-toluenesulfonyl azide.

The second round of directed evolution was performed with *Ps*EFE C317M as the parent. In this round, we targeted 10 residues for site-saturation mutagenesis (positions 84, 86, 91, 171, 186, 192, 277, 283, 314, 316). These are primarily sites that showed some modest improvement in the first round, as sites that were intolerant of mutations in the first round were deemed unlikely to yield improvements in the second round. Screening reactions for this round were performed aerobically with supplemented acetate, both for practical simplicity and to select for air-tolerant variants. We found beneficial mutations at sites 171, 277, and 314, as well as an unprogrammed mutation at site 97 found alongside a mutation at site 314 (by coincidence, sites 97 and 314 are in close structural contact) (Figure 3-11). These variants display a range of enantioselectivities, with variants giving up to modest selectivity for either enantiomer of the aziridine product (Table 3-3).

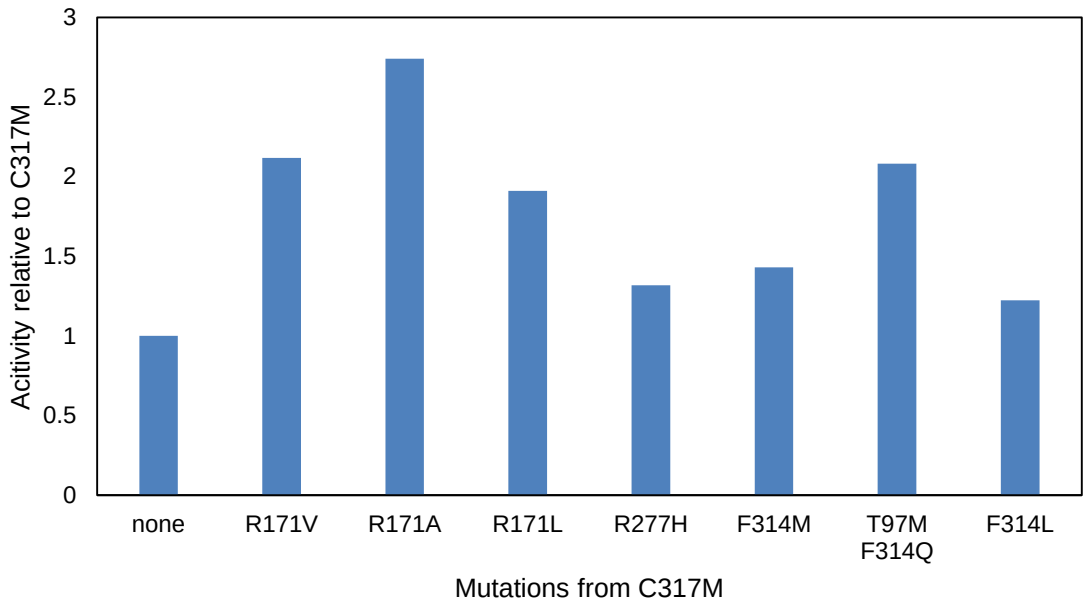


Figure 3-11. Aziridination activity of variants found in the second round of evolution. Reactions performed in clarified *E. coli* lysate with 5% ethanol co-solvent, 1 mM $Fe(NH_4)_2(SO_4)_2$, 1 mM acetate, 1 mM L-ascorbic acid, and 10 mM styrene and *p*-toluenesulfonyl azide.

Table 3-3. Enantioselectivity of variants from the second round of evolution.

Mutations from C317M	Enantiomeric excess (%)
none	-12.5
R171V	8.6
R171A	38.1
R171L	-49.4
R277H	-32.0
F314M	-6.3
T97M F314Q	-54.2
F314L	-52.1

With several beneficial mutations found at several sites and none clearly superior to the rest, we chose to recombine the mutations from round 2. The resulting recombination library has 64 possible variants, including the parent amino acid at each position. We screened 320 clones (4 96-well plates), giving five-fold oversampling, and found a number of variants with much higher activity than parent (Figure 3-12). Of these, the most promising was the T97M R171L R277H F314M C317M variant (MLHMM). While it did not have the highest total

activity in the validation reactions, it expressed at a lower level than some of the higher-performing variants, so its relative activity was among the highest. Furthermore, it displayed the highest enantioselectivity of any variant observed to date (Table 3-4).

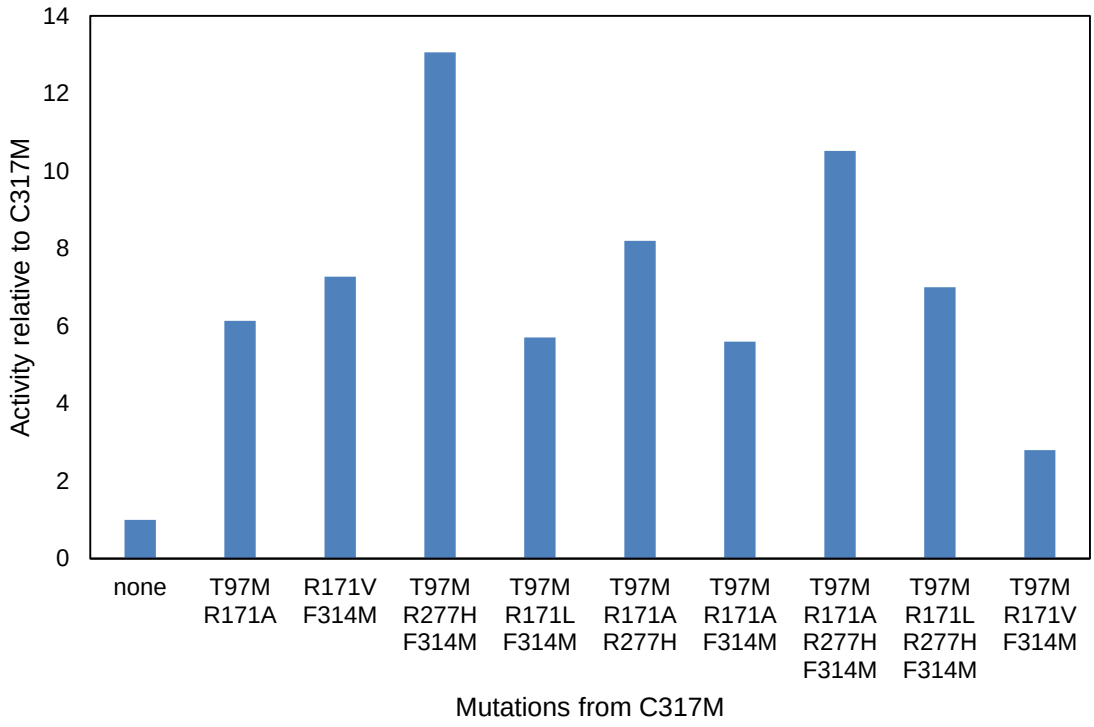


Figure 3-12. Aziridination activity of variants found in round 2 recombination. Reactions performed in clarified *E. coli* lysate with 5% ethanol co-solvent, 1 mM $Fe(NH_4)_2(SO_4)_2$, 1 mM acetate, 1 mM L-ascorbic acid, and 10 mM styrene and *p*-toluenesulfonyl azide.

Table 3-4. Enantioselectivity of variants from round 2 recombination.

Mutations from C317M	Enantiomeric excess (%)
none	-5.5
T97M R171A	26.9
R171V F314M	-31.6
T97M R277H F314M	-63.0
T97M R171L F314M	-69.8
T97M R171A R277H	8.1
T97M R171A F314M	16.6
T97M R171A R277H F314M	-1.9
T97M R171L R277H F314M	-85.4
T97M R171V F314M	-21.5

We performed a third round of mutagenesis with MLHMM as parent, targeting sites 84, 95, 173, 175, 186, 192, 270, and 283 for saturation mutagenesis. From this round, we found a variant with an additional I186L mutation (MLLHMM) that displayed comparable activity to MLHMM in lysate but with modestly increased enantioselectivity (90.4% ee).

After three rounds of evolution, we characterized the variants in the lineage with reactions using purified protein, allowing quantitative determination of protein concentration and total turnover number (TTN). The MLHMM variant, resulting from recombination of mutations found in round 2, had the highest activity, with 122.8 TTN (Table 3-5). The MLLHMM variant from round 3 had lower activity, but maintained its higher enantioselectivity. The observed activity in lysate was presumably a result of higher expression compared to MLHMM.

Table 3-5. Summary of evolutionary lineage with purified protein reactions.

Variant	Yield (%)	TTN	ee (*R*) (%)
wt	1.6	8.1	50.2
C317M	3.7	18.5	55.3
MLHMM	24.6	122.8	87.5
MLLHMM	9.8	48.9	90.4

Reaction conditions: 20 μM purified apoenzyme, 1 mM $Fe(NH_4)_2(SO_4)_2$, 1 mM acetate, 1 mM L-ascorbic acid, and 10 mM styrene and *p*-toluenesulfonyl azide.

The most active variant, *Ps*EFE MLHMM, has five mutations from wild type. Four of these mutations (T97M, R171L, F314M, and C317M) are in the region of the protein that in the wild type binds the native substrate L-arginine. The substrates for the aziridination reaction, styrene and *p*-toluenesulfonyl azide are much more hydrophobic than L-arginine, and the beneficial mutations found are generally to more hydrophobic amino acids. The fifth mutation, R277H, is in the region that binds the (co-)substrate α-ketoglutarate in the wild-type enzyme. The effects of this mutation are discussed in more detail in section 3.6, but it likely disrupts the native binding mode of α-ketoglutarate.

3.5 Nitrene C–H insertion with engineered *Ps*EFE variants

With a set of diverse variants of *Ps*EFE generated in the course of evolution for aziridination, we sought to explore other types of nitrene transfer reactions. Of particular interest are C–H insertion reactions: C–H bonds are ubiquitous in organic compounds but are often unreactive; when they do react their ubiquity poses a major selectivity challenge. Selective functionalization of C–H bonds remains an attractive target for synthetic chemistry.

The most diverse set of *Ps*EFE variants we had was the recombination library from round 2 of the aziridination evolution. We screened this library against the intramolecular nitrene C–H insertion reaction of 2-ethylbenzenesulfonyl azide and the intermolecular reaction of *p*-toluenesulfonyl azide with either 4-ethylanisole or 4-isopropylanisole (Figure 3-13).

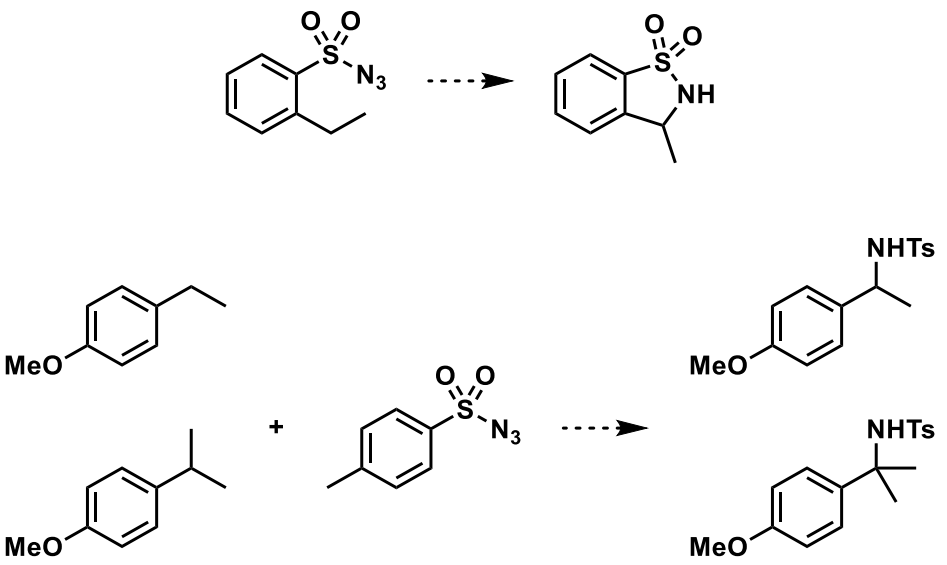


Figure 3-13. Intra- and intermolecular nitrene C–H insertion reactions tested.

While no product was detected for the intermolecular C–H insertion with any *Ps*EFE variant tested, a number of variants were able to catalyze the intramolecular reaction with high activity (Figure 3-14).

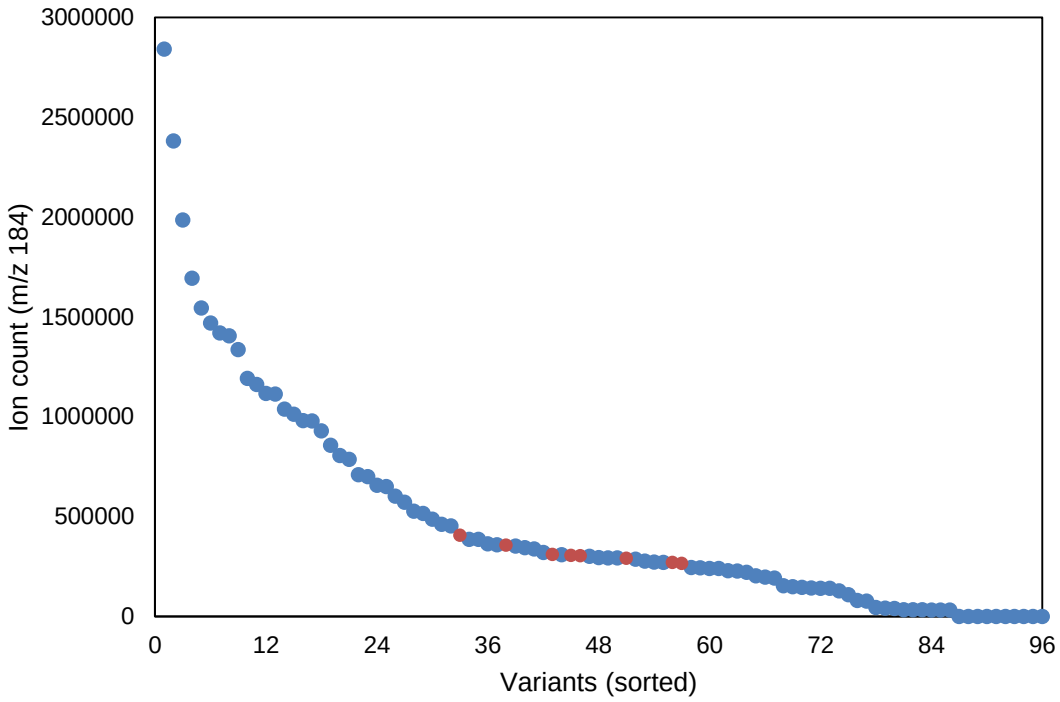


Figure 3-14. Ranked plot of recombination variants for intramolecular C–H insertion of 2-ethylbenzenesulfonyl azide. Clones of the library parent (C317M) are depicted in orange.

Validation of the top variants revealed *Ps*EFE R171V F314M C317M (VMM) and R171V R277H F314M C317M (VHMM) as the best-performing variants for this reaction.

3.6 Metal-coordinating ligand effects

One of the most intriguing features of *Ps*EFE and other non-heme metalloenzymes with multiple available coordination sites, in contrast to heme-dependent enzymes, is the ability to easily modify the coordination environment of the metal with simple reaction additives. In *Ps*EFE, protein amino acid side chains occupy three coordination sites of the iron. The putative iron-nitrene intermediate is expected to occupy only one coordination site, leaving potentially two sites available for other ligands. For the native dioxygenase activity, α-ketoglutarate or succinate occupies these sites, but we have shown that these ligands are not required for nitrene transfer activity.

One of the beneficial mutations found in the course of directed evolution for aziridination was R277H. In the wild-type enzyme, the side-chain guanidine of arginine-277 coordinates to the distal carboxylate of αKG by an ionic interaction (Figure 3-15). Mutation of this residue to histidine likely disrupts this interaction and weakens binding of αKG and other 1,5-dicarboxylate analogues.

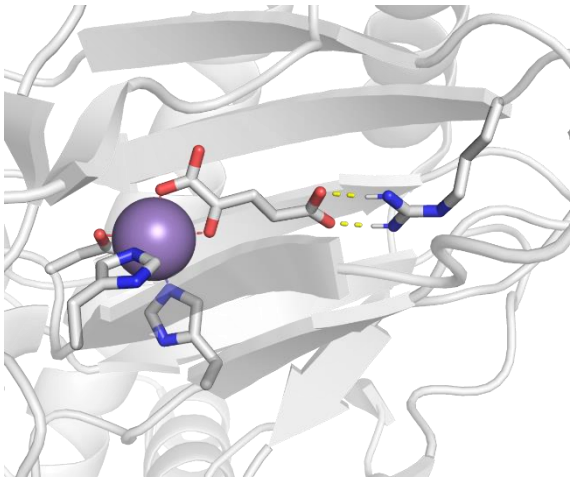
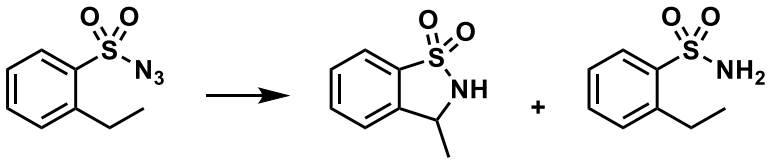


Figure 3-15. Structural representation of wild-type *Ps*EFE complexed with manganese and αKG, showing the interaction between R277 (right) and αKG (PDB ID: 6CBA).

The two most active variants for C–H insertion differ only by this R277H mutation, allowing a direct comparison. *Ps*EFE VHMM, with the R277H mutation, shows no meaningful difference when either αKG or NOG is added, but a nearly ten-fold activity increase with acetate. The VMM variant, in contrast, shows a five-fold enhancement with αKG, a 12-fold enhancement with acetate, and a nearly 30-fold enhancement with NOG, attaining 730 TTN with 100:1 chemoselectivity for insertion over reduction. VMM also displays higher enantioselectivity with αKG and NOG bound than with acetate, likely due to more rigid binding of the ligand in the active site.

Table 3-6. Ligand effects on C–H insertion activity.

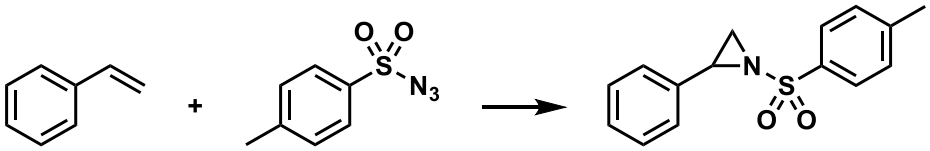


*Ps*EFE variant	Additive	Yield (%)	TTN	ee (%)	Insertion/reduction
VHMM	None	5.4	27	n.d.[1]	3.4
VHMM	αKG	6.2	31	n.d.[1]	3.8
VHMM	Acetate	48	240	7.3	32
VHMM	NOG	6.6	33	n.d.[1]	4.1
VMM	None	5.0	25	n.d.[1]	0.9
VMM	αKG	26	130	61	9.0
VMM	Acetate	62	310	9.4	24
VMM	NOG	90	450	48	105
VMM[2]	NOG	73	730	47	100

Reaction conditions: 20 μM purified apoenzyme, 1 mM $Fe(NH_4)_2(SO_4)_2$, 1 mM additive, 1 mM L-ascorbic acid, and 10 mM 2-ethylbenzenesulfonyl azide (maximum 500 TTN). TTNs are shown for insertion only. [1]Not determined due to low conversion. [2]10 μM enzyme concentration (max. 1000 TTN).

The most active aziridination variant (MLHMM) contains the R277H mutation and shows no change when αKG or NOG is added, consistent with the observed C–H insertion activity of VHMM (Table 3-7). As a comparison, we prepared the corresponding 277R variant, *Ps*EFE T97M R171L F314M C317M (MLMM). While MLMM is somewhat less active for aziridination (30% lower activity with acetate bound) overall than MLHMM, it shows the same general ligand/activity trend as VMM does for C–H insertion, albeit less dramatically. Notably, with NOG as additive, MLMM gives 98% enantiomeric excess.

Table 3-7. Ligand effects on aziridination activity.



Variant	Additive	Yield	TTN	ee (%)
MLHMM	None	2.2	11	n.d.
MLHMM	αKG	2.6	13	n.d.
MLHMM	Acetate	24	120	88
MLHMM	NOG	2.4	12	n.d.
MLMM	None	2.8	14	n.d.
MLMM	αKG	6.8	34	76
MLMM	Acetate	17	84	80
MLMM	NOG	20	98	98

Reaction conditions: 20 μM apoenzyme, 1 mM $Fe(NH_4)_2(SO_4)_2$, 1 mM additive, 1 mM L-ascorbic acid, and 10 mM styrene and *p*-toluenesulfonyl azide (maximum 500 TTN).

In whole *E. coli* cells and in unpurified lysate, there are a number of carboxylate-containing compounds that could conceivably coordinate to the catalytic iron, including acetate and αKG, both of which are abundant primary metabolites. Indeed, both *Ps*EFE VMM and VHMM display high activity for C–H insertion in both whole cells and lysate with no external additive. *N*-Oxalylglycine has a modest beneficial effect for *Ps*EFE VMM in lysate, both for yield and chemoselectivity, but otherwise external additives have little effect (Figure 3-16). The observed chemoselectivities in whole cells and lysate are significantly lower than in reactions with purified protein, presumably due to increased reduction by the cellular background.

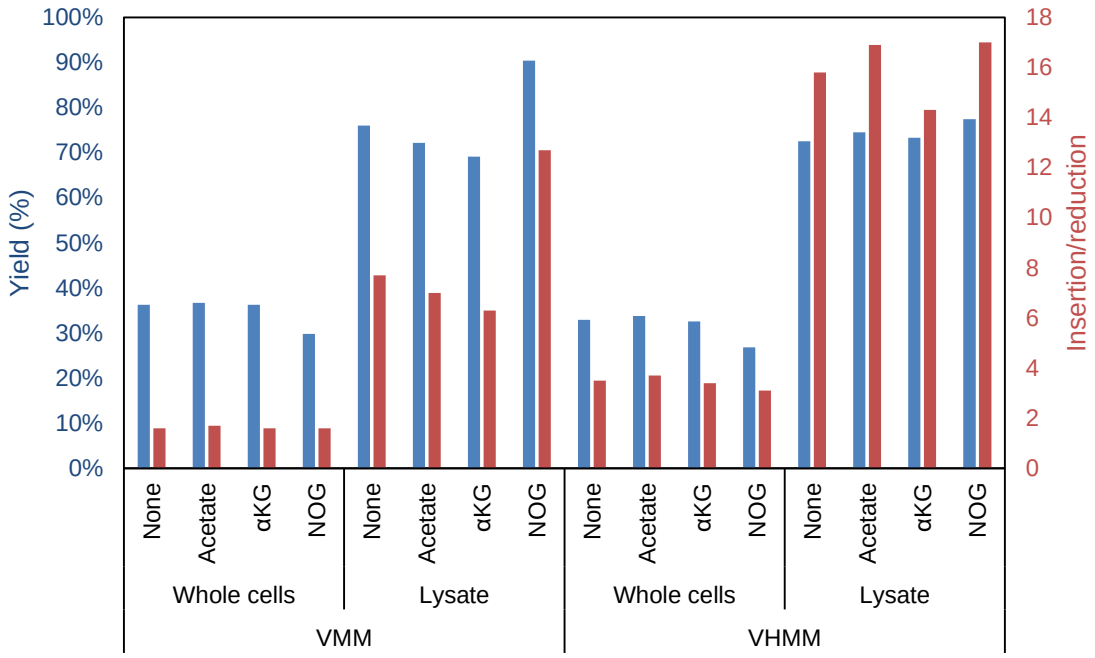


Figure 3-16. Ligand effects on C–H insertion activity in whole cells and lysate. Reactions performed with cell suspensions at OD_{600} = 30 and with the corresponding clarified lysate, with 1 mM $Fe(NH_4)_2(SO_4)_2$, 1 mM additive, 1 mM L-ascorbic acid, and 10 mM 2-ethylbenzenesulfonyl azide.

We also tested C–H insertion with ligands beyond simple carboxylates and α-KG mimics, including *N*- and *S*-donors. Of the ligands tested, only glycine displayed a modest activity enhancement, with both VMM and VHMM (Table 3-8).

Table 3-8. C–H insertion with expanded ligand set.

Variant	Ligand	TTN	Insertion/reduction
VMM	Imidazole	109	5.3
VMM	Pyridine	103	5.3
VMM	Picolinic acid	7.7	2.4
VMM	1-3-Dimethylimidazolium-2-carboxylate	105	5.2
VMM	Glycine	294	15.5
VMM	*N*,*N*-dimethylglycine	101	5.2
VMM	Thioglycolic acid	106	1.7
VMM	Thioacetate	17	0.2
VHMM	Imidazole	54	7.0
VHMM	Pyridine	52	7.3
VHMM	Picolinic acid	22	4.5

VHMM	1-3-Dimethylimidazolium-2-carboxylate	59	7.8
VHMM	Glycine	108	12.8
VHMM	*N*,*N*-dimethylglycine	60	7.9
VHMM	Thioglycolic acid	70	1.4
VHMM	Thioacetate	18	0.3

Reaction conditions: 20 μM purified apoenzyme, 1 mM $Fe(NH_4)_2(SO_4)_2$, 1 mM additive, 1 mM L-ascorbic acid, and 10 mM 2-ethylbenzenesulfonyl azide (maximum 500 TTN).

3.7 Conclusions

We discovered promiscuous nitrene-transfer activity of *Ps*EFE, an α-KG-dependent dioxygenase, which can catalyze the aziridination reaction of styrene and *p*-toluenesulfonyl azide. This is the first demonstration of nitrene transfer catalyzed by a non-heme iron enzyme, either natural or engineered. Through directed evolution, with two rounds of mutagenesis and one round of recombination, we increased aziridination activity 15-fold, up to 120 TTN with 88% enantiomeric excess. Variants from the aziridination evolution were also found to perform intramolecular nitrene C–H insertion with high activity. As nitrene transfer does not proceed through the native catalytic mechanism, the native cofactor αKG is no longer strictly required. Indeed, while binding αKG does provide a modest activity enhancement for nitrene transfer, we found that the addition of other ligands in place of αKG, in particular acetate and *N*-oxalylglycine, resulted in a much larger enhancement of activity and selectivity. This kind of small-molecule-controlled reactivity is unique to non-heme metalloenzymes, as the iron center in heme is coordinatively saturated.

3.8 Supplementary information for Chapter 3

Synthetic chemistry

All manipulations were performed using oven-dried glassware (130 °C for a minimum of 12 hours) and standard Schlenk techniques under an atmosphere of argon, unless otherwise stated.

Solvents

ACS- and HPLC-grade solvents were purchased from Fisher Chemical. Anhydrous tetrahydrofuran was obtained by filtration through a drying column and a deoxygenation column on a Pure Process Technologies solvent system. High-purity water for PCR and HPLC was distilled after filtration through a deionizing column and organic removal column. Deuterated solvents were purchased from Cambridge Isotope Laboratories.

Chromatographic materials

Thin layer chromatography (TLC) was performed using EMD TLC plates pre-coated with 250 μm thickness silica gel 60 F_{254} and visualized by fluorescence quenching under UV light and staining with potassium permanganate or cerium ammonium molybdate. Preparative flash chromatography was performed using a Biotage Isolera automated chromatography instrument using columns hand-packed with silica gel (230–400 mesh, Silicyle Inc.).

Starting materials

All compounds were used as received from commercial suppliers, unless otherwise stated.

Analytical instrumentation

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

NMR spectra were recorded on a Varian Unity/Inova 500 spectrometer operating at 500 MHz and 125 MHz for ^{1}H and ^{13}C, respectively, or a Bruker Avance 400 spectrometer

operating at 400 MHz and 100 MHz for ^{1}H and ^{13}C, respectively. NMR data were analyzed in MestReNova (MestreLab Research). Chemical shifts are reported in ppm with the solvent resonance as the internal standard. For ^{1}H NMR: $CDCl_3$, δ 7.26. For ^{13}C NMR: $CDCl_3$, δ 77.16. Data are reported as follows: s = singlet, d = doublet, t = triplet, q = quartet, spt = septet, m = multiplet, br = broad; coupling constants in Hz; integration.

Biology and biocatalytic reactions

Materials

Oligonucleotides were purchased from IDT DNA. PCRs were run with Phusion® High-Fidelity PCR Kit (New England Biolabs). Gibson assembly mix[16] is prepared with isothermal master mix in-house and enzymes T5 exonuclease, Phusion® DNA polymerase, and Taq DNA ligase purchased from New England Biolabs. Oligonucleotide, DNA, and amino-acid sequences used in this work are given in the supplemental Excel document.

Cloning

Plasmids encoding *Pseudomonas savastanoi* ethylene-forming enzyme (Uniprot ID P32021), *Streptomyces sp*. 2-aminobutyric acid chlorinase (UniProt ID D0VX22), and *Arabidopsis thaliana* anthocyanidin synthase (UniProt ID Q96323), with the coding sequences codon-optimized for *Escherichia coli* were purchased from Twist Biosciences. Plasmids encoding *Gluconobacter oxydans* leucine dioxygenase (UniProt ID Q5FQD2), *Streptomyces vinaceus* arginine hydroxylase (UniProt ID Q6WZB0), and *Streptomyces muensis* leucine hydroxylase (UniProt ID A0A0E3URV8) were obtained from the laboratory of Prof. Hans Renata (Scripps Research Institute). The plasmid encoding *Escherichia coli* taurine dioxygenase (UniProt ID P37610) was obtained from the laboratory of Prof. Harry Gray (Caltech). All genes were encoded with a C-terminal His_6-tag for purification and inserted between the NdeI and XhoI cut sites in the pET-22b(+) vector (Novagen).

Plasmids were used to transform *E. cloni* BL21(DE3) cells (Lucigen) by electroporation. SOC medium (0.75 mL) was added and the cells were incubated at 37 °C for 45 minutes before being plated on Luria-Bertani medium (Research Products International) supplemented with ampicillin (100 μg mL^{-1}, LB-amp) agar plates.

Plasmids were isolated from stationary-phase cultures by miniprep (Qiagen) and Sanger sequencing was performed by Laragen, Inc. (Culver City, CA) using T7 promoter and T7 terminator primers.

Protein expression and purification

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

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

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

100 μL), flash-frozen on powdered dry ice, and stored at −80 °C. Protein concentration was determined by Bradford assay (Bio-Rad Quick Start Bradford).

Site-saturation mutagenesis and library screening

Site-saturation mutagenesis was performed using the 22-codon method[17]. Oligonucleotides including the three 22-codon trick codons (NDT, VHG, TGG) and oligonucleotides within the ampicillin resistance cassette were used to amplify the plasmid in two pieces, with an overlap for Gibson assembly in the gene encoding the protein of interest (and where the mutation is introduced) and an overlap for Gibson assembly in the gene encoding β-lactamase (which confers ampicillin resistance). This two-piece assembly can prevent mis-assembled constructs from conferring antibiotic resistance. The oligonucleotide sequences which generated the variants reported in in this work are listed in Table 3-9. PCR products were loaded on 1% agarose gels with loading dye containing SYBR Gold nucleic acid gel stain (Thermo Fisher) and visualized on a blue transilluminator. The DNA bands at the expected size were excised and the DNA was extracted with a Zymoclean Gel DNA recovery kit. The two linear PCR products for a given site-saturation mutagenesis library were assembled *via* isothermal Gibson assembly (50 °C, 1 hour). *E. cloni* BL21(DE3) cells (Lucigen) were transformed by electroporation with the Gibson assembly product without further purification. SOC medium (0.75 mL) was added to the electroporated cells and the cells were incubated at 37 °C for 45 minutes before being plated on LB-amp agar plates. The LB-amp agar plates with the plated cells were incubated at 37 °C for 12–18 hours and stored at 4 °C until the libraries were picked. Single colonies from the agar plates were picked with sterile toothpicks and used to inoculate starter cultures (0.5 mL LB-amp) in 96 deep-well plates. The starter culture plates were grown at 37 °C, 250 rpm, and 80% humidity in a Multitron Infors shaker overnight (14–16 hours). The starter cultures (50 μL) were used to inoculate expression cultures (1 mL TB-amp) in 96 deep-well plates. In parallel, glycerol stock plates were prepared for long-term storage by mixing starter culture (50 μL) with sterile glycerol (50% v/v, 50 μL) and frozen at −80 °C. The expression cultures were grown at 37 °C, 250 rpm, and 80% humidity for three hours, then cooled on ice for 20 minutes. Protein expression was induced by addition of IPTG (0.5 mM). Cultures were incubated at 22 °C

and 220 rpm overnight (18–20 hours). Cells were pelleted (5000×*g*, 5 minutes) and the cell pellets were frozen at −20 °C for at least 24 hours prior to use.

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

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

Wells which showed an apparent enhancement in aziridine product formation were streaked out (from the glycerol stock plates prepared in parallel) onto LB-amp agar plates and

incubated at 37 °C. Starter cultures (5 mL LB-amp) were inoculated from single colonies and incubated overnight at 37 °C. Plasmids were isolated from these cultures and sequenced as described above (Cloning). Unique variants were then regrown in Erlenmeyer flasks as described above (Protein expression and purification) and assayed in clarified cell lysate for nitrene-transfer activity.

Table 3-9. Oligonucleotides used for mutagenesis. Mutated codons are denoted here as NNN for simplicity; in practice they are a 12:9:1 ratio of NDT:VHG:TGG for site saturation or the appropriate single codon for site-directed mutagenesis. ampR forward and reverse oligonucleotides are used together with the mutagenesis oligonucleotides to amplify the plasmid in two pieces. ampR forward is used with reverse mutagenesis primers and ampR reverse is used with forward mutagenesis primers.

Mutations relative to wild type	Direction	Sequence
T97X	Forward	CCGACTTCCCCGAAATTTTC*NNN*GTCTGCAAAGATCTTTC
T97X	Reverse	GAAAATTTCGGGGAAGTCGGGCTTTCCAGCAGTCACCTC
R171X	Forward	GATGGATGGCACCACATG*NNN*GTGTTGCGTTTTCCGCC
R171X	Reverse	CATGTGGTGCCATCCATCGCGGGTCAAATCTG
R277X	Forward	GGTGAAACTTAATACACGTGAG*NNN*TTTGCTTGCGCGTAC TTCCATGAGCCG
R277X	Reverse	CACGTGTATTAAGTTTCACCTTATGCGGAGTGCTAAGTAA CTGTCCCCCG
F314X C317M	Forward	CACTATGGGGAACATTTCACGAACATG*NNN*ATGCGTATG TATCCTGACCG
F314X	Reverse	CATGTTCGTGAAATGTTCCCCATAGTGAATGCGCTCATTG GCC
C317X	Forward	TCACGAACATGTTCATGCGT*NNN*TATCCTGACCGCATTAC CACACAGC
C317X	Reverse	CATGAACATGTTCGTGAAATGTTCCCCATAGTGAATGCGC TC
ampR	Forward	CCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGC TAACCGCTTTTTTGC
ampR	Reverse	CGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACT CATGGTTATGGCAG

Analytical-scale biocatalytic aziridination reactions

Biocatalytic reactions were set up in 2-mL screw-cap vials (Agilent). Purified apoprotein (350 μL, 22.9 μM in 20 mM MOPS pH 7.0, final concentration 20 μM) was added to the vial. Solutions of ferrous ammonium sulfate and L-ascorbic acid were prepared immediately

prior to use. Reactions to be set up anaerobically were brought into a Coy vinyl anaerobic chamber (nitrogen atmosphere, 0–10 ppm oxygen). To each reaction was added in order ferrous ammonium sulfate (40 mM in water, 10 μL, 1 mM final concentration), sodium acetate or other additive (40 mM in water, 10 μL, 1 mM final concentration), and L-ascorbic acid (40 mM in water, 10 μL, 1 mM final concentration). Each reaction was then charged with styrene (400 mM in ethanol, 10 μL, 10 mM final concentration) immediately followed by *p*-toluenesulfonyl azide (400 mM in ethanol, 10 μL, 10 mM final concentration, 500 max. TTN). The reactions were sealed and shaken at room temperature for three hours unless otherwise noted. To quench the reactions, acetonitrile (350 μL) was added to each vial, followed by internal standard propiophenone (0.1% v/v in acetonitrile, 50 μL). The sample was transferred to a 1.7-mL Eppendorf tube, vortexed, and then centrifuged (20817×*g*, 5 minutes). 250 μL of the supernate was transferred to HPLC vial inserts for reverse-phase HPLC analysis. The remaining supernate was partially concentrated *in vacuo* to remove acetonitrile and ethanol. Cyclohexane (500 μL) was added to the resulting aqueous suspension. The mixture was thoroughly shaken and then centrifuged (20817×*g*, 5 minutes). 250 μL of the organic layer was transferred to HPLC vial inserts for normal-phase chiral HPLC analysis.

Analytical-scale biocatalytic C–H insertion reactions

Biocatalytic reactions were set up in 2-mL screw-cap vials (Agilent). Purified apoprotein (360 μL, 22.2 μM in 20 mM MOPS pH 7.0, final concentration 20 μM) was added to the vial. Solutions of ferrous ammonium sulfate and L-ascorbic acid were prepared immediately prior to use. Reactions to be set up anaerobically were brought into a Coy vinyl anaerobic chamber (nitrogen atmosphere, 0–10 ppm oxygen). To each reaction was added in order ferrous ammonium sulfate (40 mM in water, 10 μL, 1 mM final concentration), sodium acetate or other additive (40 mM in water, 10 μL, 1 mM final concentration), L-ascorbic acid (40 mM in water, 10 μL, 1 mM final concentration). Each reaction was then charged with 2-ethylbenzenesulfonyl azide (400 mM in ethanol, 10 μL, 10 mM final concentration, 500 max. TTN). The reactions were sealed and shaken at room temperature for six hours unless otherwise noted. To quench the reactions, acetonitrile (350 μL) was added to each vial, followed by internal standard propiophenone (0.5% v/v in acetonitrile, 50 μL). The sample was transferred to a 1.7-mL Eppendorf tube, vortexed, and then centrifuged (20817×*g*,

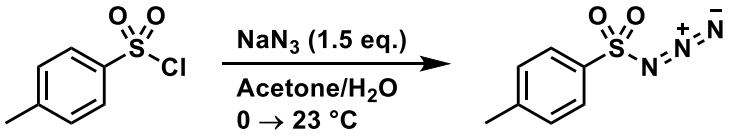
5 minutes). 250 μL of the supernate was transferred to HPLC vial inserts for reverse-phase HPLC analysis. The remaining supernate was partially concentrated *in vacuo* to remove acetonitrile and ethanol. Hexanes (250 μL, HPLC grade) and ethyl acetate (250 μL, HPLC grade) were added. The resulting mixture was thoroughly shaken and then centrifuged (20817×*g*, 5 minutes). 250 μL of the organic layer was transferred to HPLC vial inserts for normal-phase chiral HPLC analysis.

Synthesis of sulfonyl azide substrates

Safety statement

Organic azides are potentially explosive compounds. We have not observed any problems in our handling of the compounds described, but care should be taken, especially on large scales.

p-Toluenesulfonyl azide



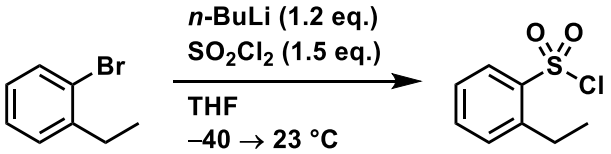
Under air, *p*-toluenesulfonyl chloride (19.1 g, 100 mmol, 1.00 equiv.) was dissolved in acetone (200 mL) in a 500 mL round-bottomed flask with a magnetic stir bar and cooled to 0 °C. A solution of sodium azide (9.75 g, 150 mmol, 1.50 equiv.) in water (60 mL) was added dropwise over one hour with stirring. Once addition was complete, the reaction was allowed to warm to room temperature and stirred for 16 hours. The acetone was removed *in vacuo*, and the resulting mixture was extracted with diethyl ether (2×100 mL). The combined organic layers were washed with water (2×100 mL), saturated aqueous sodium bicarbonate (100 mL), and brine (100 mL), then dried over magnesium sulfate and concentrated *in vacuo* to afford the title compound as a colorless oil that solidified upon storage at −20 °C (19.1 g, 97%).

NMR Spectroscopy:

^{1}H NMR (500 MHz, $CDCl_3$, 23 ºC, δ): 7.80 (d, *J* = 8.5 Hz, 2 H), 7.38 (d, *J* = 8.5 Hz, 2 H), 2.45 (s, 3 H)

^{13}C NMR (125 MHz, $CDCl_3$, 23 ºC, δ): 146.3, 135.3, 130.3, 127.4, 21.7

2-Ethylbenzenesulfonyl chloride[18]

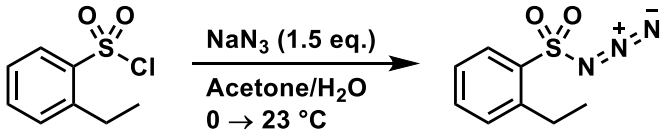


Under argon, 1-bromo-2-ethylbenzene (2.07 mL, 2.78 g, 15.0 mmol, 1.00 equiv.) was dissolved in anhydrous tetrahydrofuran (30 mL) in a 100 mL round-bottomed flask with magnetic stirring and cooled to −40 °C. *n*-Butyllithium (2.5 M solution in hexanes, 7.20 mL, 18.0 mmol, 1.20 equiv.) was added dropwise by syringe over two minutes. The reaction was stirred at −40 °C for thirty minutes, then sulfuryl chloride (1.82 mL, 3.04 g, 22.5 mmol, 1.5 equiv.) was added dropwise by syringe over two minutes. The reaction was allowed to warm to room temperature and stirred for 16 hours. The reaction was cooled to 0 °C, then carefully quenched by the addition of ice-cold water (50 mL). The resulting mixture was extracted with diethyl ether (2×50 mL). The combined organic layers were dried over sodium sulfate and concentrated *in vacuo*. The residue was purified by flash column chromatography on silica (50 g), eluting with a gradient of 0 to 20% diethyl ether/hexanes, to afford the title compound as a slightly yellow oil (1.10 g, 36%).

NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, 23 ºC, δ): 8.07 (d, *J* = 8.1 Hz, 1 H), 7.66 (t, *J* = 7.5 Hz, 1 H), 7.49 (d, *J* = 7.6 Hz, 1 H), 7.41 (t, *J* = 7.9 Hz, 1 H), 3.20 (q, *J* = 7.5 Hz, 2 H), 1.37 (t, *J* = 7.5 Hz, 3 H)

2-Ethylbenzenesulfonyl azide[18]



Under air, 2-ethylbenzenesulfonyl chloride **S1** (1.00 g, 4.89 mmol, 1.00 equiv.) was dissolved in acetone (8 mL) in a 20 mL scintillation vial with magnetic stirring and cooled

to 0 °C. A solution of sodium azide (476 mg, 7.33 mmol, 1.50 equiv.) in water (2.5 mL) was added dropwise over two minutes with stirring. Once addition was complete, the reaction was allowed to warm to room temperature and stirred for six hours. The acetone was removed *in vacuo*, and the resulting mixture was extracted with diethyl ether (2×15 mL). The combined organic layers were washed with water, saturated aqueous sodium bicarbonate, and brine (15 mL each), then dried over magnesium sulfate and concentrated *in vacuo* to afford the title compound as a yellow oil (1.01 g, 98%).

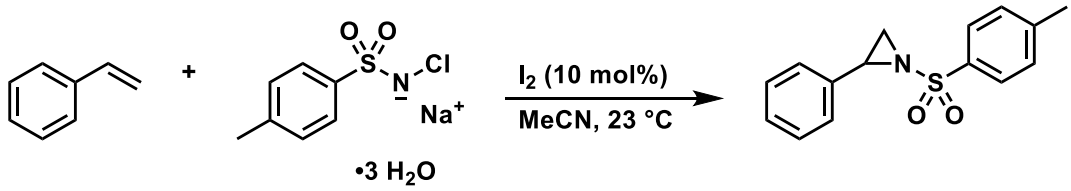
NMR Spectroscopy:

^{1}H NMR (500 MHz, $CDCl_3$, 23 ºC, δ): 8.01 (dd, *J* = 8.0, 1.3 Hz, 1 H), 7.62 (td, *J* = 7.6, 1.3 Hz, 1 H), 7.46 (dd, *J* = 7.7, 0.7 Hz, 1 H), 7.37 (td, *J* = 8.0, 1.2 Hz, 1 H), 3.03 (q, *J* = 7.5 Hz, 2 H), 1.31 (t, *J* = 7.5 Hz, 3 H)

^{13}C NMR (125 MHz, $CDCl_3$, 23 ºC, δ): 144.5, 136.3, 134.9, 131.4, 129.4, 126.4, 26.1, 15.2

Synthesis of authentic product standards

2-Phenyl-1-(p-toluenesulfonyl)aziridine



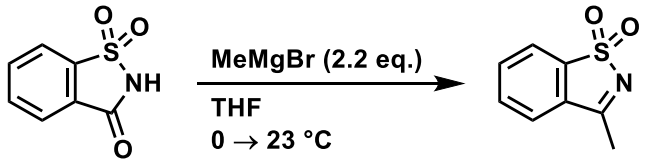
Under argon, chloramine-T trihydrate (4.23 g, 15.0 mmol, 1.00 equiv.) and iodine (381 mg, 1.50 mmol, 0.100 equiv.) were dissolved in acetonitrile (100 mL). Styrene (3.45 mL, 3.13 g, 30.0 mmol, 2.00 equiv.) was added dropwise, and the reaction was stirred at room temperature for 18 hours. The reaction mixture was partitioned between water (50 mL) and dichloromethane (100 mL), and the layers were separated. The aqueous layer was extracted with dichloromethane (2×100 mL). The combined organic layers were concentrated *in vacuo* and the residue was purified by flash column chromatography on silica (100 g) eluting with hexanes/ethyl acetate (6:1 v/v) containing 1% triethylamine to afford the title compound as a colorless solid (3.55 g, 87%).

NMR Spectroscopy:

^{1}H NMR (500 MHz, $CDCl_3$, 23 ºC, δ): 7.87 (d, *J* = 8.3 Hz, 2 H), 7.33 (d, *J* = 8.3 Hz, 2 H), 7.30–7.25 (m, 3 H), 7.24–7.20 (m, 2 H), 3.78 (dd, *J* = 7.2, 4.5 Hz, 1 H), 2.99 (d, *J* = 7.2 Hz, 1 H), 2.44 (s, 3 H), 2.40 (d, *J* = 4.4 Hz, 1 H)

^{13}C NMR (125 MHz, $CDCl_3$, 23 ºC, δ): 144.8, 135.1, 135.0, 129.9, 128.7, 128.4, 128.1, 126.7, 41.2, 36.1, 21.8

3-Methylbenzo[d]isothiazole 1,1-dioxide[19]

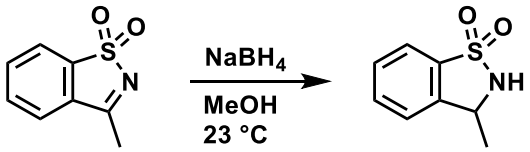


Under argon, saccharin (1.00 g, 5.46 mmol, 1.00 equiv.) was dissolved in anhydrous tetrahydrofuran (25 mL) in a 100 mL round-bottomed flask with magnetic stirring and cooled to 0 °C. Methylmagnesium bromide (3 M solution in diethyl ether, 4.00 mL, 12.0 mmol, 2.20 equiv.) was added dropwise by syringe over 10 minutes. After addition, the reaction was stirred at 0 °C for five minutes, then slowly warmed to room temperature and stirred at room temperature for 16 hours. The reaction was then cooled to 0 °C and carefully poured into ice-cold hydrochloric acid (1 M, 30 mL). The resulting mixture was extracted with dichloromethane (2×50 mL). The combined organic layers were dried over sodium sulfate and concentrated *in vacuo* to afford the title compound as a colorless solid (crude yield 1.04 g, 105%) which was used in the next step without further purification.

NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, 23 ºC, δ): 7.94–7.90 (m, 1 H), 7.77–7.73 (m, 2 H), 7.71–7.67 (m, 1 H), 2.67 (s, 3 H)

3-Methyl-2,3-dihydrobenzo[d]isothiazole 1,1-dioxide



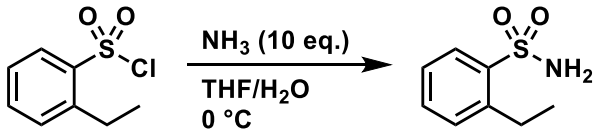
Under air, 3-methyl-[*d*]isothiazole 1,1-dioxide **S2** (500 mg, 2.76 mmol, 1.00 equiv.) was dissolved in methanol (20 mL) in a 50 mL round-bottomed flask with magnetic stirring. Sodium borohydride (522 mg, 13.8 mmol, 5.00 equiv.) was slowly added over two minutes. The reaction mixture bubbled vigorously and became warm to the touch. The reaction was stirred at room temperature for thirty minutes to ensure complete reaction. The reaction was cooled to 0 °C, then poured carefully into cold hydrochloric acid (2.5 M, 40 mL). The methanol was removed *in vacuo* and the resulting mixture was extracted with dichloromethane (3×25 mL). The combined organic layers were dried over sodium sulfate and concentrated *in vacuo*. The residue was purified by flash column chromatography on silica (25 g), eluting with a gradient of 10 to 60% ethyl acetate/hexanes to afford the title compound as a colorless solid (378 mg, 75%).

NMR Spectroscopy:

^{1}H NMR (500 MHz, $CDCl_3$, 23 ºC, δ): 7.74 (d, *J* = 8.0 Hz, 1 H), 7.61 (td, *J* = 7.6, 1.0 Hz, 1 H), 7.50 (t, *J* = 7.6 Hz, 1 H), 7.38 (d, *J* = 7.7 Hz, 1 H), 5.15 (br d, *J* = 4.8 Hz, 1 H), 4.78 (qd, *J* = 6.7, 4.8 Hz, 1 H), 1.59 (d, *J* = 6.7 Hz, 3 H)

^{13}C NMR (125 MHz, $CDCl_3$, 23 ºC, δ): 141.8, 135.4, 133.3, 129.2, 124.0, 121.2, 53.5, 21.5

2-Ethylbenzenesulfonamide



Under air, 2-ethylbenzenesulfonyl chloride **S1** (50.0 mg, 244 μmol, 1.00 equiv.) was dissolved in tetrahydrofuran (1 mL) in a 4 mL vial with magnetic stirring and cooled to 0 °C.

Ammonia (28% w/v in water, 149 μL, 2.44 mmol, 10.0 equiv.) was added dropwise over one minute. After stirring for five minutes, the reaction mixture was partitioned between water and ethyl acetate (10 mL each). The layers were separated and the aqueous layer was extracted with ethyl acetate (2×10 mL). The combined organic layers were washed with brine (10 mL), dried over sodium sulfate, and concentrated *in vacuo* to afford the title compound as a colorless solid (41.8 mg, 92%).

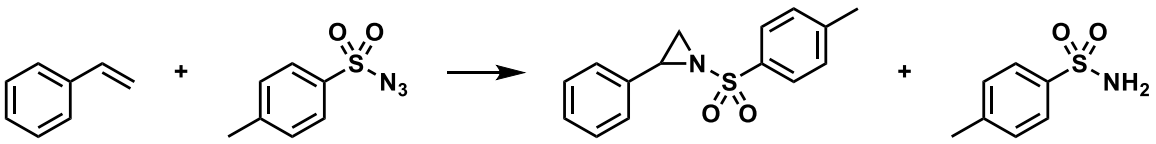
NMR Spectroscopy:

^{1}H NMR (500 MHz, $CDCl_3$, 23 ºC, δ): 7.99 (d, *J* = 8.0 Hz, 1 H), 7.51 (t, *J* = 7.5 Hz, 1 H), 7.39 (d, *J* = 7.6 Hz, 1 H), 7.29 (t, *J* = 7.7 Hz, 1 H), 5.02 (br s, 2 H), 3.07 (q, *J* = 7.5 Hz, 2 H), 1.33 (t, *J* = 7.5 Hz, 3 H)

^{13}C NMR (125 MHz, $CDCl_3$, 23 ºC, δ):143.0, 139.6, 133.1, 130.7, 128.3, 126.2, 26.1, 15.3

HPLC analytical methods and calibration curves

Aziridination reaction



Samples for HPLC calibration curves were prepared as simulated reaction samples. To MOPS buffer (20 mM pH 7.0, 380 μL) was added a solution of the appropriate reaction product in acetonitrile (0–100 μM, 20 μL, final concentration 0–5 mM). To this sample was added the internal standard propiophenone (0.1% v/v in acetonitrile, 50 μL) and acetonitrile (350 μL). The product concentration in the curves below corresponds to the concentration in the reaction mixture; the final analytical sample is two-fold diluted.

Analysis was performed on an Agilent 1200 series HPLC with water/acetonitrile mobile phase (1 mL min^{-1} flow), with an Agilent Poroshell 120 EC-C18 column (4 μm packing, 2.1×50 mm) fitted with a Poroshell 120 guard column (1.7 μm packing, 2.1×5 mm), injecting

5 μL. Detection was at 230 nm (16 nm bandwidth). The gradient program and retention times are given in Table 3-10 and, Table 3-11, respectively.

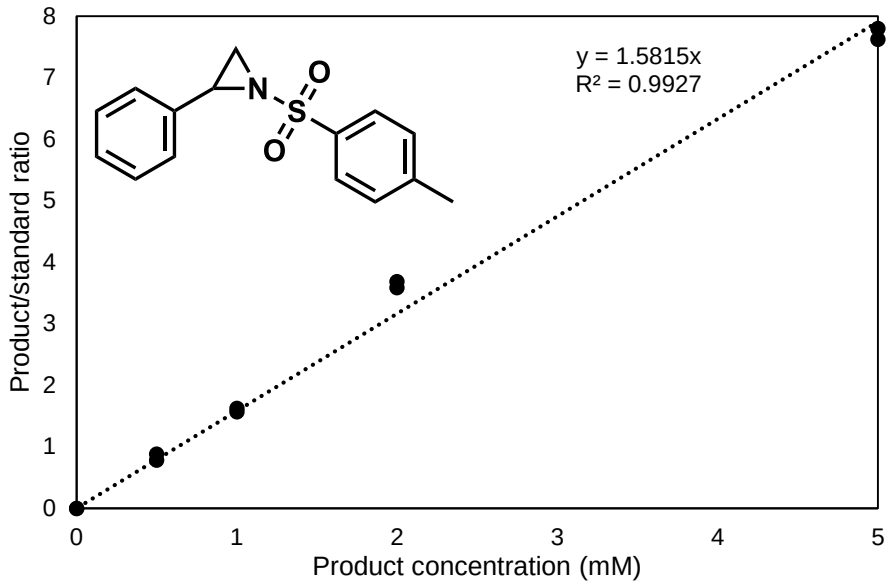
Table 3-10. HLPC gradient program for aziridination analysis.

Time (minutes)	% Acetonitrile
0.00	20
0.50	20
1.00	40
5.00	65
5.50	95
6.00	95
6.01	20
7.00	20

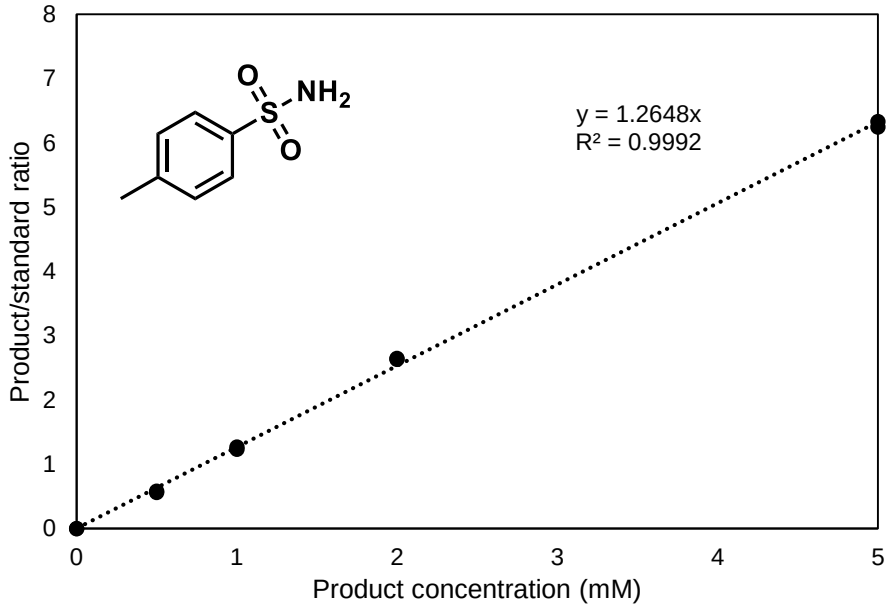
Table 3-11. HPLC retention times for aziridination analysis.

Compound	Retention time (minutes)
p-Toluenesulfonamide	0.58
Propiophenone	2.24
p-Toluenesulfonyl azide	2.81
Styrene	3.01
2-Phenyl-1-(*p*-toluenesulfonyl)aziridine	3.45

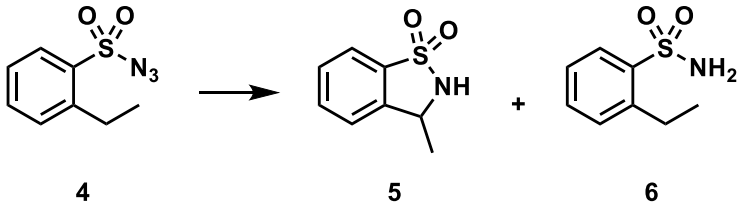
2-Phenyl-1-(p-toluenesulfonyl)aziridine calibration curve



p-Toluenesulfonamide calibration curve



C–H insertion reaction



4 5 6

Calibration curve samples were prepared as described above for the aziridination reaction, except the internal standard used was propiophenone (0.5% v/v in acetonitrile, 50 μL) and the final product concentrations ranged from 1–10 mM. The product concentration shown corresponds to the concentration in the reaction mixture; the final analytical sample is twofold diluted.

Analysis was performed on an Agilent 1260 Infinity II HPLC instrument with water/acetonitrile mobile phase (1 mL min^{-1} flow), with an Agilent Poroshell 120 EC-C18 column (4 μm packing, 2.1×50 mm) fitted with a Poroshell 120 guard column (1.7 μm packing, 2.1×5 mm), injecting 5 μL. Detection was at 220 nm (4 nm bandwidth). The gradient program and retention times are given in Table 3-12 and Table 3-13, respectively.

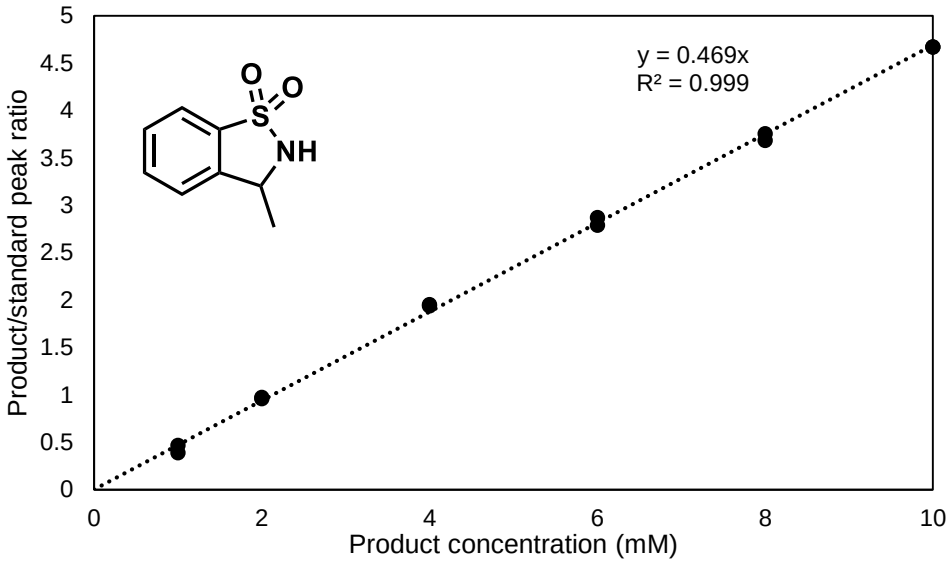
Table 3-12. HPLC gradient program for C–H insertion analysis.

Time (minutes)	% Acetonitrile
0.00	12
1.00	12
3.50	95
4.00	95
4.01	12
5.00	12

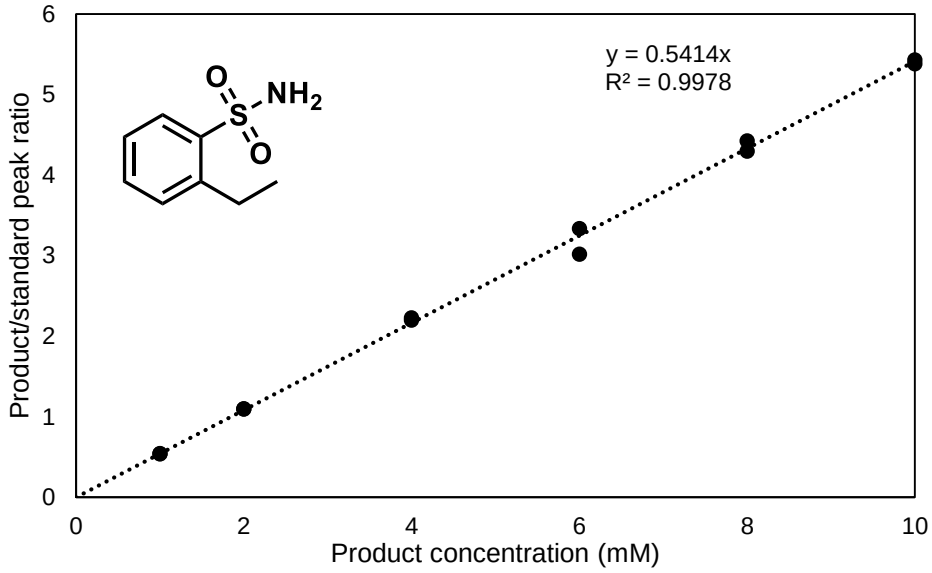
Table 3-13. HPLC retention times for C–H insertion analysis.

Compound	Retention time (minutes)
3-Methyl-2,3-dihydrobenzo[*d*]isothiazole 1,1-dioxide	1.05
2-Ethylbenzenesulfonamide	2.38
Propiophenone	3.00
2-Ethylbenzenesulfonyl azide	3.48

3-Methyl-2,3-dihydrobenzo[d]isothiazole 1,1-dioxide calibration curve



2-Ethylbenzenesulfonamide calibration curve

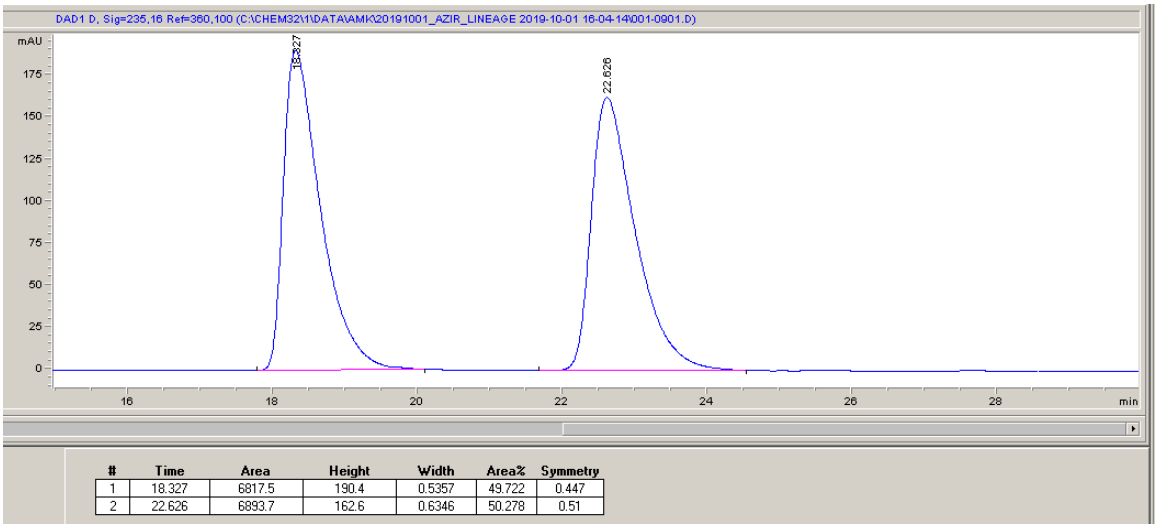


Chiral analysis

Chiral analysis was performed by HPLC with a chiral stationary phase, using a Hewlett Packard Series 1100 HPLC instrument with hexanes/2-propanol mobile phase (1 mL min^{-1} flow).

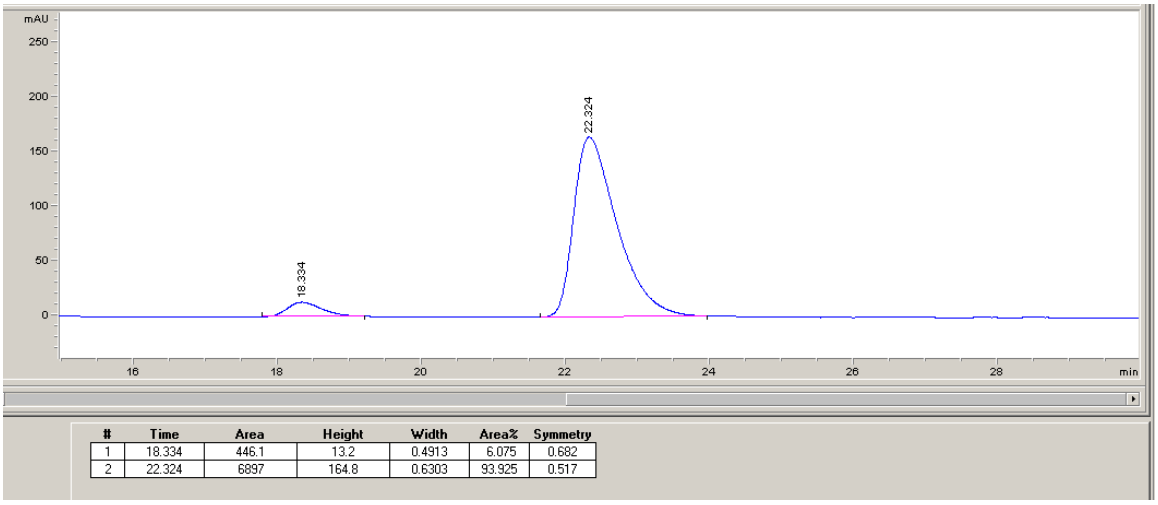
2-Phenyl-1-(p-toluenesulfonyl)aziridine

Analysis was performed with a Daicel Chiralcel OJ-H column, (5 μm packing, 4.6×250 mm), with an isocratic 30% 2-propanol/70% hexanes mobile phase. The peak areas were analyzed at 235 nm (16 nm bandwidth). Absolute configuration was assigned by reference to the literature[20]. The enantiomers elute in the order (*S*), (*R*).



#	Time	Area	Height	Width	Area%	Symmetry
1	18.327	6817.5	190.4	0.5357	49.722	0.447
2	22.626	6893.7	162.6	0.6346	50.278	0.51

Figure 3-17. Chiral HPLC trace of *rac*-2-phenyl-1-(*p*-toluenesulfonyl)aziridine.

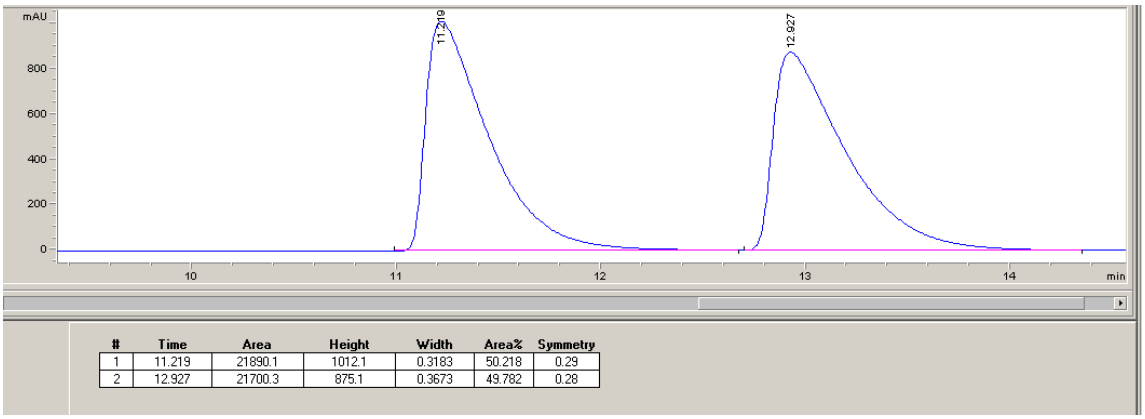


#	Time	Area	Height	Width	Area%	Symmetry
1	18.334	446.1	13.2	0.4913	6.075	0.682
2	22.324	6897	164.8	0.6303	93.925	0.517

Figure 3-18. Chiral HPLC trace of *Ps*EFE MLHMM-catalyzed 2-phenyl-1-(*p*-toluene-sulfonyl)aziridine.

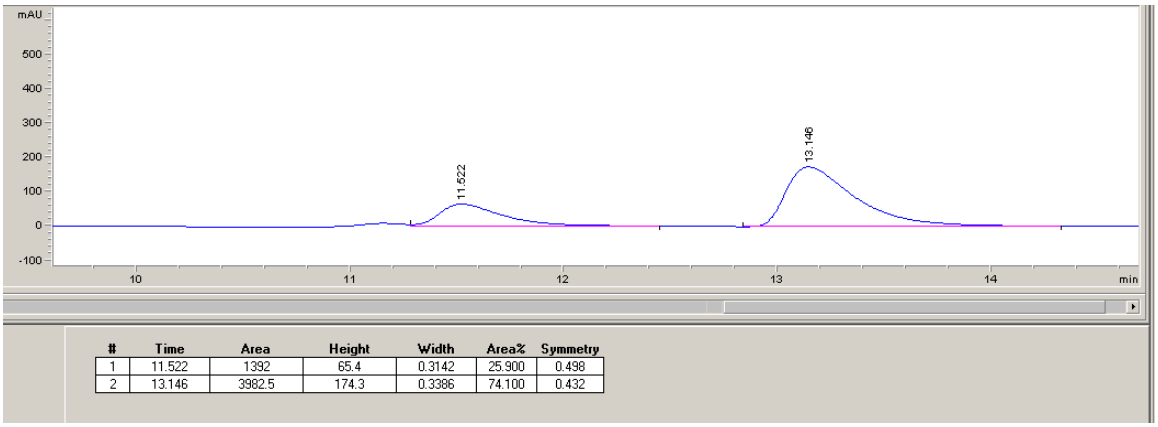
3-Methyl-2,3-dihydrobenzo[d]isothiazole 1,1-dioxide

Analysis was performed with a Daicel Chiralpak IB column (5 μm packing, 4.6×250 mm), with an isocratic 25% 2-propanol/75% hexanes mobile phase. The peak areas were analyzed at 220 nm (16 nm bandwidth).



#	Time	Area	Height	Width	Area%	Symmetry
1	11.219	21890.1	1012.1	0.3183	50.218	0.29
2	12.927	21700.3	875.1	0.3673	49.782	0.28

Figure 3-19. Chiral HPLC trace of *rac*-3-methyl-2,3-dihydrobenzo[*d*]isothiazole 1,1-dioxide.



#	Time	Area	Height	Width	Area%	Symmetry
1	11.522	1392	65.4	0.3142	25.900	0.498
2	13.146	3982.5	174.3	0.3386	74.100	0.432

Figure 3-20. Chiral HPLC trace of *Ps*EFE VMM-catalyzed 3-methyl-2,3-dihydrobenzo[*d*]isothiazole 1,1-dioxide.

3.9 References for Chapter 3

[1] Islam M. D.; Leissing, T. M.; Chowdhury, R.; Hopkinson, R. J.; Schofield, C. J. 2-Oxoglutarate-Dependent Oxygenases. *Annu. Rev. Biochem*. **2018**, *87*, 585–620.

[2] Hausinger, R. P. Fe(II)/α-Ketoglutarate-Dependent Hydroxylases and Related Enzymes. *Crit. Rev. Biochem. Mol. Biol*. **2004**, *39*, 21–68.

[3] Gorres, K. L.; Raines, R. T. Prolyl 4-hydroxylase. *Crit. Rev. Biochem. Mol. Biol.* **2010**, *45*, 106–124.

[4] Yin, X.; Zabriskie, T. M. VioC is a Non-Heme Iron, α-Ketoglutarate-Dependent Oxygenase that Catalyzes the Formation of 3*S*-Hydroxy-L-Arginine during Viomycin Biosynthesis. *ChemBioChem* **2004**, *5*, 1274–1277.

[5] Marchand, J. A.; Neugebauer, M. E.; Ing, M. C.; Lin, C.-I.; Pelton, J. G.; Chang, M. C. Y. Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature* **2019**, *567*, 420–424.

[6] Salowe, S. P.; Krol, W. J.; Iwata-Reuyl D.; Townsend, C. A. Elucidation of the Order of Oxidations and Identification of an Intermediate in the Multistep Clavaminate Synthase Reaction. *Biochemistry* **1991**, *30*, 2281–2292.

[7] Dotzlaf, J. E.; Yeh, W.-K. Purification and Properties of Deacetoxycephalosporin C Synthase from Recombinant *Escherichia coli* and Its Comparison with the Native Enzyme Purified from *Streptomyces clavuligerus*. *J. Biol. Chem.* **1989**, *264*, 10219–10227.

[8] Zwick, C. R. III.; Renata, H. Remote C−H Hydroxylation by an α-Ketoglutarate-Dependent Dioxygenase Enables Efficient Chemoenzymatic Synthesis of Manzacidin C and Proline Analogs. *J. Am. Chem. Soc.* **2018**, *140*, 1165–1169.

[9] Zhang, X.; King-Smith, E.; Renata, H. Total Synthesis of Tambromycin by Combining Chemocatalytic and Biocatalytic C−H Functionalization. *Angew. Chem. Int. Ed.* **2018**, *57*, 5037–5041.

[10] Amatuni, A.; Shuster, A.; Adibekian, A.; Renata, H. Concise Chemoenzymatic Total Synthesis and Identification of Cellular Targets of Cepafungin I. *Cell Chem. Biol.* **2020**, *27*, 1318–1326.

[11] Matthews, M. L.; Chang, W.-c.; Layne, A. P.; Miles, L. A.; Krebs, C.; Bollinger, J. M. Jr. Direct nitration and azidation of aliphatic carbons by an iron-dependent halogenase. *Nat. Chem. Biol.* **2014**, *10*, 209–215.

[12] Fukuda, H.; Ogawa, T.; Tazaki, M.; Nagahama, K.; Fujii, T.; Tanase, S.; Morino, Y. Molecular cloning in *Escherichia coli*, expression, and nucleotide sequence of the gene for the ethylene-forming enzyme of *Pseudomonas syringae* pv. *phaseolicola* PK2. *Biochem. Biophys. Res. Commun.* **1992**, *188*, 483–489.

[13] Zhang Z., Smart T. J.; Choi, H.; Hardy, F.; Lohans, C. T.; Abboud, M. I.; Richardson, M. S. W.; Paton, R. S.; McDonough, M. A.; Schofield, C. J. Structural and stereoelectronic insights into oxygenase-catalyzed formation of ethylene from 2-oxoglutarate. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 4467–4672.

[14] Martinez, S.; Fellner, M.; Ferr, C. Q.; Ritchie, A.; Hu, J.; Hausinger, R. P. Structures and Mechanisms of the Non-Heme Fe(II)- and 2-Oxoglutarate-Dependent Ethylene-Forming Enzyme: Substrate Binding Creates a Twist. *J. Am. Chem. Soc.* **2017**, *139*, 11980–11988.

[15] Hausinger, R. P. Fe(II)/α-Ketoglutarate-Dependent Hydroxylases and Related Enzymes. *Crit. Rev. Biochem. Mol. Biol*. **2004**, *39*, 21–68.

[16] Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A. III; Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **2009**, *6*, 343–345.

[17] Kille, S.; Acevedo-Rocha, C. G.; Parra, L. P.; Zhang, Z.-G..; Opperman, D. J.; Reetz, M. T.; Acevedo, J. P. Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synth. Biol*. **2013**, *2*, 83–92.

[18] Ichinose, M.; Suematsu, H.; Yasutomi, Y.; Nishioka, Y.; Uchida, T.; Katsuki, T. Enantioselective Intramolecular Benzylic C–H Bond Amination: Efficient Synthesis of Optically Active Benzosultams. *Angew. Chem. Int. Ed.* **2011**, *50*, 9884–9887.

[19] Li, B.; Chen, J.; Zhang, Z.; Gridnev, I. Y., Zhang. W. Nickel-Catalyzed Asymmetric Hydrogenation of *N*-Sulfonyl Imines. *Angew. Chem. Int. Ed.* **2019**, *58*, 7329–7334.

[20] Takeda Y.; Ikeda, Y.; Kuroda, A.; Tanaka, S.; Minakata, S. Pd/NHC-Catalyzed Enantiospecific and Regioselective Suzuki–Miyaura Arylation of 2-Arylaziridines: Synthesis of Enantioenriched 2-Arylphenethylamine Derivatives. *J. Am. Chem. Soc*. **2014**, *136*, 8544–8547.

Chapter 4

OLEFIN AMINOARYLATION CATALYZED BY A NON-HEME IRON ENZYME

N.W.G. performed initial reaction discovery and all directed evolution experiments. N.W.G. performed mechanistic studies and reaction controls. N.W.G. performed all synthetic chemistry and analytical method development and data processing. R.K.Z. postulated the Smiles rearrangement mechanistic hypothesis. A.M.K. designed the “tile” region saturation libraries.

Abstract

Organic azides are well known to react as nitrene donors with small-molecule and enzymatic transition metal catalysts, including both heme-dependent and non-heme-dependent iron enzymes. I have discovered a previously unknown reactive mode of sulfonyl azides catalyzed by variants of *Pseudomonas savastanoi* ethylene-forming enzyme (*Ps*EFE), a non-heme iron enzyme, in which they react with electron-rich styrene derivatives to give a 1,2-aminoarylated product. In this reaction, the olefin acts as both the aminoarylation substrate and aryl donor, with loss of one vinyl group, and the azide acts as the amine donor. This reaction is unlike any previously reported transformation of sulfonyl azides, and it displays a number of unusual mechanistic features. The aminoarylation activity of *Ps*EFE was improved 75-fold by directed evolution, though the final activity was still rather low. Further engineering efforts did not yield meaningfully improved variants.

4.1 Introduction: Nitrene-transfer reactions of sulfonyl azides

Nitrene transfer

Sulfonyl azides are primarily known to react with transition metal catalysts, including metalloenzymes, as nitrene donors. In a typical nitrene transfer reaction, the azide binds to the catalytic metal and eliminates dinitrogen to form a metal nitrenoid, which then transfers the nitrene moiety to a second substrate, typically either an olefin to give an aziridine product[1], or a C–H bond-containing substrate in a C–H amination reaction[2] (Figure 4-1).

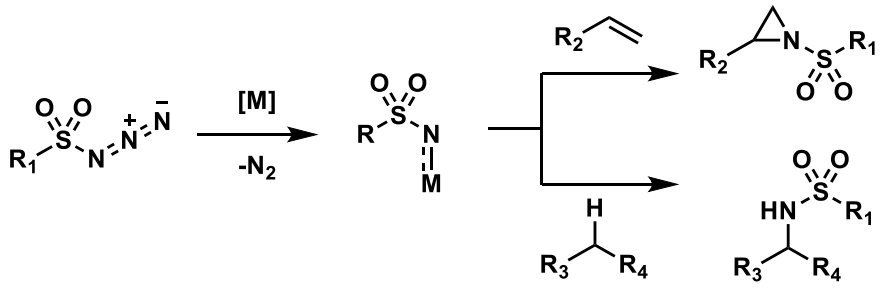


Figure 4-1. General reaction scheme of transition-metal catalyzed nitrene transfer from sulfonyl azides.

Nitrene transfer from sulfonyl azides has been reported with a variety of small-molecule transition metal catalysts, including complexes of precious metals such as iridium[3], rhodium[4], or ruthenium[5], as well as base metals such as copper[6] and cobalt[7] (Figure 4-2).

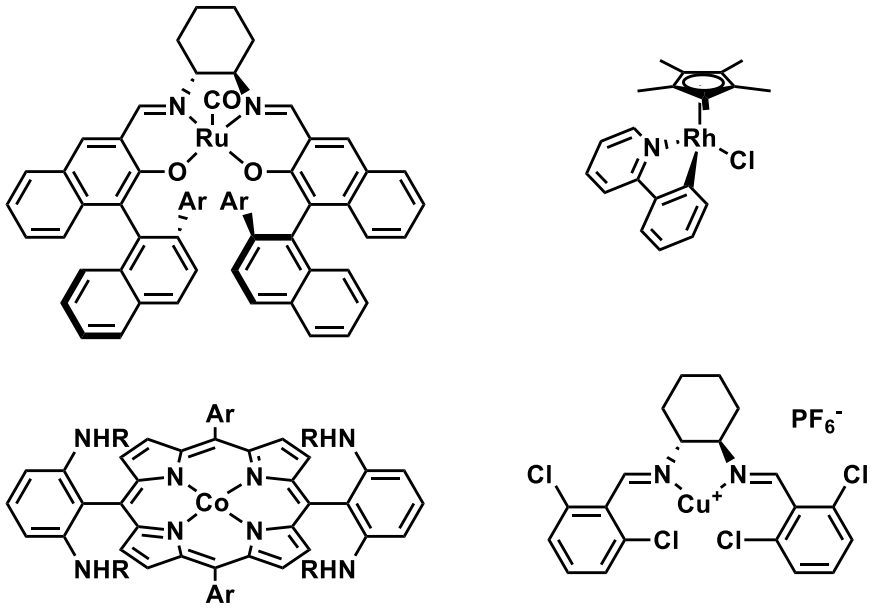


Figure 4-2. Examples of small-molecule catalysts for nitrene transfer from sulfonyl azides.

Metalloenzymes, including heme-binding enzymes[8,9,10] and non-heme iron enzymes (Chapter 3)[11,12], have also been reported to catalyze nitrene transfer reactions (both aziridination and C–H amination) with sulfonyl azides (Figure 4-3).

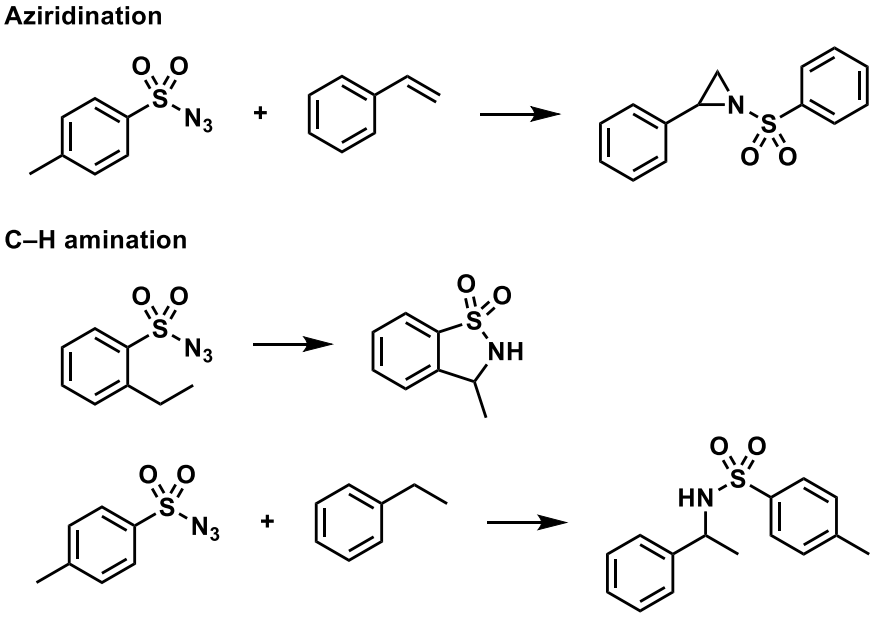


Figure 4-3. Enzyme-catalyzed reactions of sulfonyl azides.

4.2 Initial mechanistic hypothesis and activity discovery

Investigation of enzymatic olefin aminoarylation was inspired by a related reaction reported by Stephenson and coworkers using photoredox catalysis[13]. In the proposed enzyme-catalyzed aminoarylation mechanism (Figure 4-4), a sulfonyl azide reacts with an enzyme-bound iron center to form an iron nitrenoid. The nitrenoid can then react with a styrenyl olefin (here shown as 4-methoxystyrene) in a concerted manner to give an aziridine product, as has previously been reported, or it can add stepwise to give a benzylic radical. This radical can then cyclize onto the arene substituent of the sulfonyl group, and through a Smiles rearrangement, yield the aminoarylated product with loss of SO_2 and a two-electron reduction. While the Smiles rearrangement hypothesis was ruled out by later mechanistic studies (Section 4.4), it guided my initial activity discovery and early directed evolution (in particular the choice of 4-methoxybenzenesulfonyl azide as substrate).

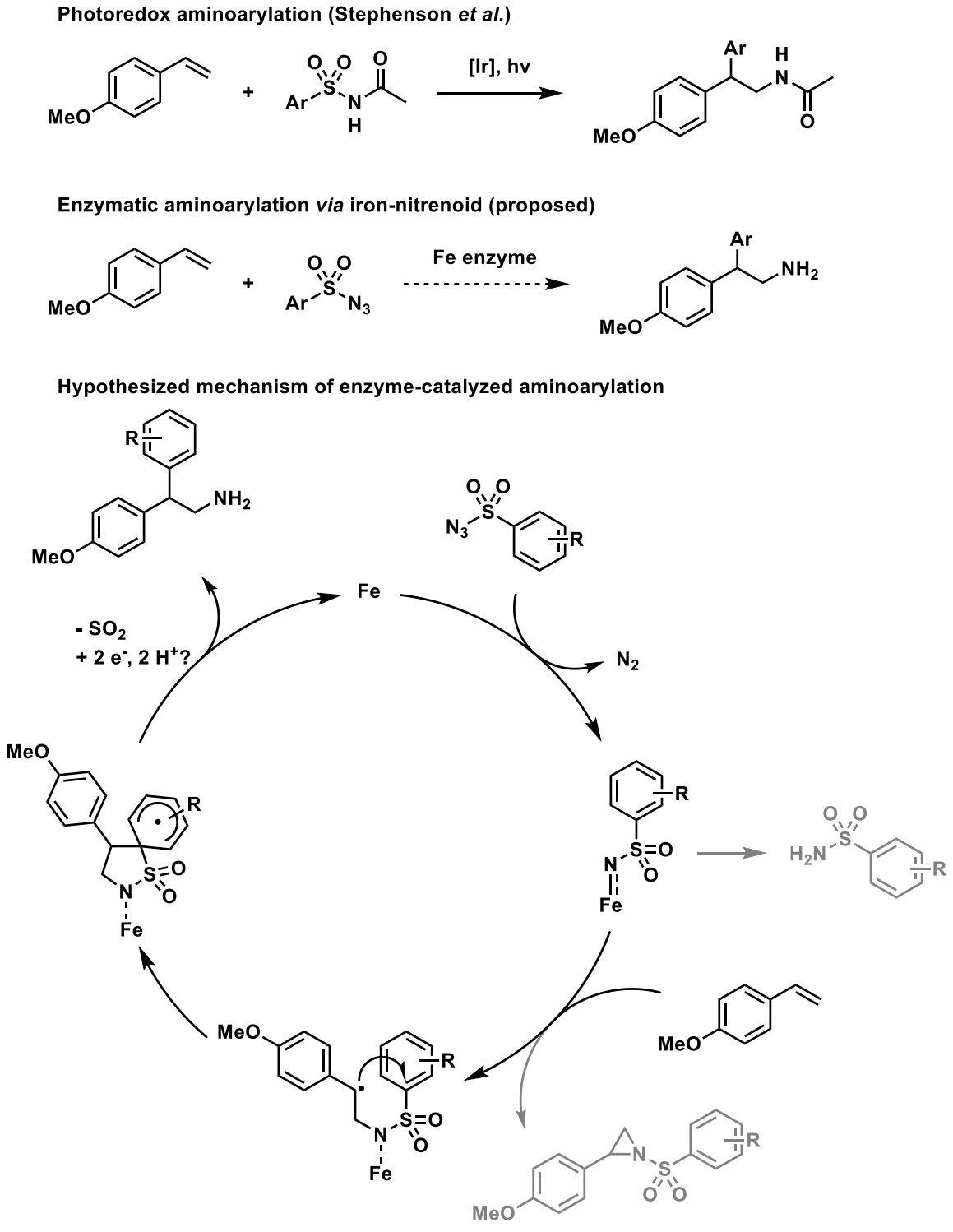


Figure 4-4. Initial mechanistic hypothesis for olefin aminoarylation: stepwise nitrene transfer/Smiles rearrangement pathway.

The initial activity screen was conducted with a recombination library of *Ps*EFE variants previously developed for the aziridination of styrene and *p*-toluenesulfonyl azide (Chapter 3). The parent for this library was *Ps*EFE C317M, and the recombined mutations were T97M, R171A, R171V, R171L, R277H, F314M, F314Q, and F314L, including the parent residue at each position. The reaction screen was conducted in 96-well plates, as described in section 4.6 (Site-saturation mutagenesis and library screening), using 4-methoxystyrene and a set of arylsulfonyl azides as substrates. Naphthalene-1-sulfonyl azide (a model substrate used by Stephenson and coworkers) was tested, as well as 4-nitro- and 4-methoxybenzenesulfonyl azide, to include electron-withdrawing and electron-donating substituents (Figure 4-5).

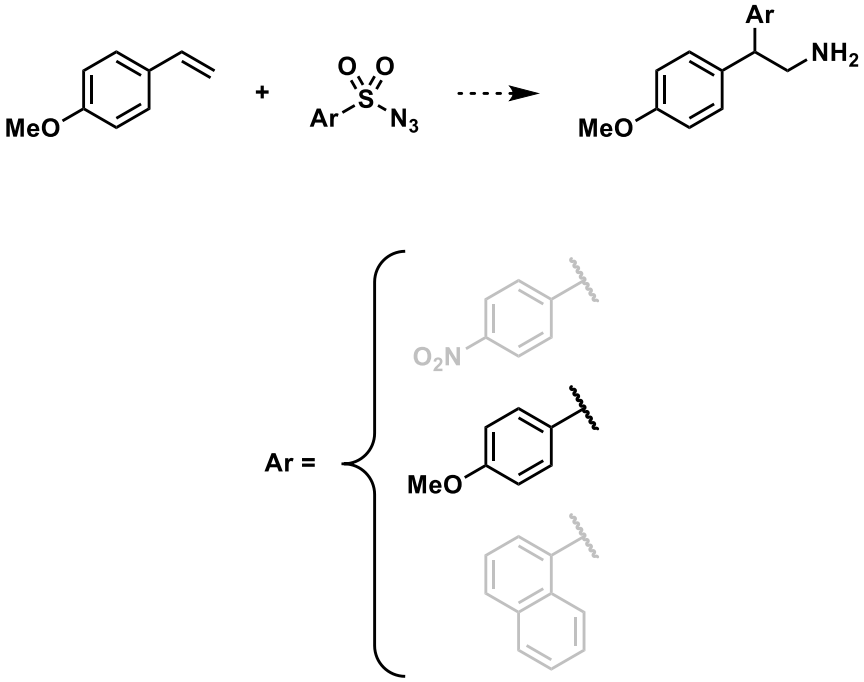


Figure 4-5. Substrates tested for aminoarylation in the initial activity screen.

Of the substrates evaluated, the putative aminoarylation product was observed only with 4-methoxybenzenesulfonyl azide. Later experiments would reveal that both arenes in the aminoarylation product in fact derive from the styrene substrate, which is why the expected product from the Smiles rearrangement pathway is only observed when the sulfonyl azide bears the same arene substituent. A number of variants in the initial screen formed this product detectably, of which the most active were re-expressed in Erlenmeyer flasks and evaluated in lysate. Several related variants, all containing the R171A mutation, showed

promising activity (Figure 4-6). *Ps*EFE R171A F314M C317M (AMM) was chosen as the parent for directed evolution due to its sequence similarity to previously studied *Ps*EFE variants.

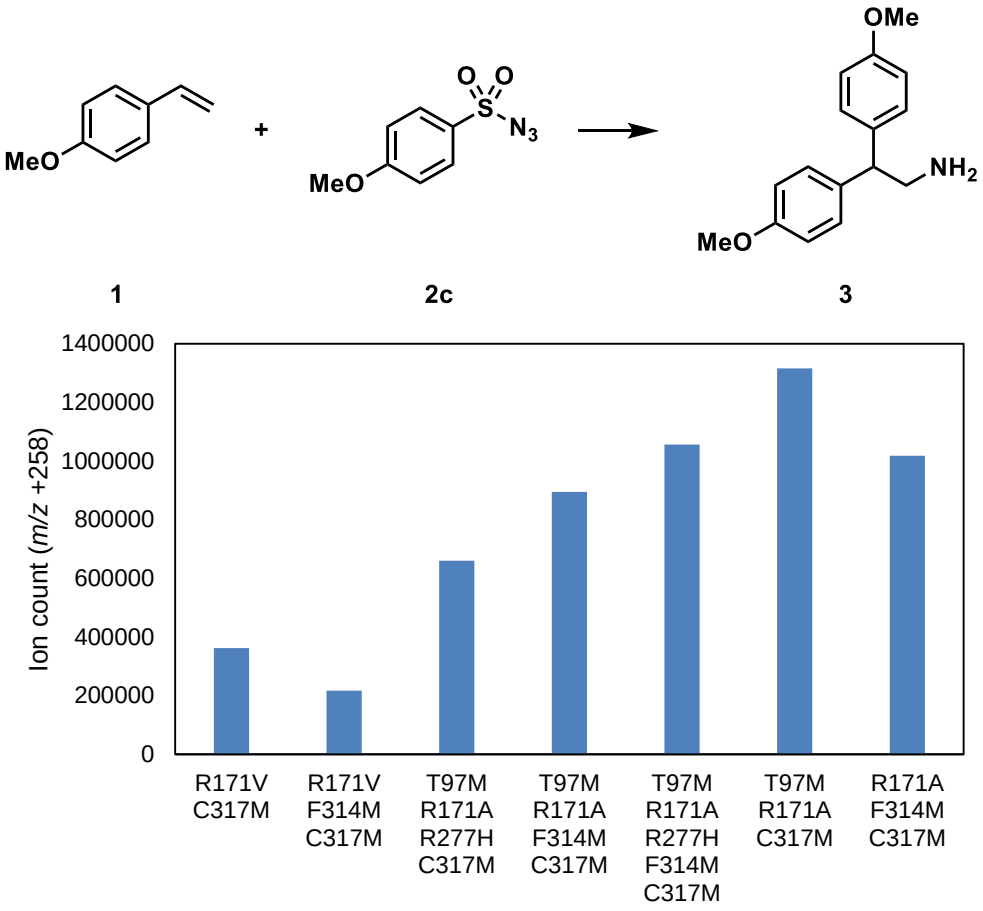


Figure 4-6. Evaluation of aziridination recombination variants for aminoarylation activity. Activity was determined by mass spectrometry (*m*/*z* = +258).

4.3 Directed evolution for improved aminoarylation activity

I subjected *Ps*EFE AMM to directed evolution by site-saturation mutagenesis, to improve the aminoarylation activity. Screening reactions and lysate validation reactions for directed evolution were performed with 4-methoxybenzenesulfonyl azide as substrate, thus the

apparent improvements are less pronounced than with 4-trifluoromethylbenzenesulfonyl azide, which was used as the substrate in later studies.

Site-saturation mutagenesis round 1

Sites targeted for site-saturation in the first round were E84, T86, D91, T97, A171, I186, Y192, R277, F283, M314, R316, and M317 (Figure 4-7). These sites were all previously targeted in directed evolution for aziridination (Chapter 3).

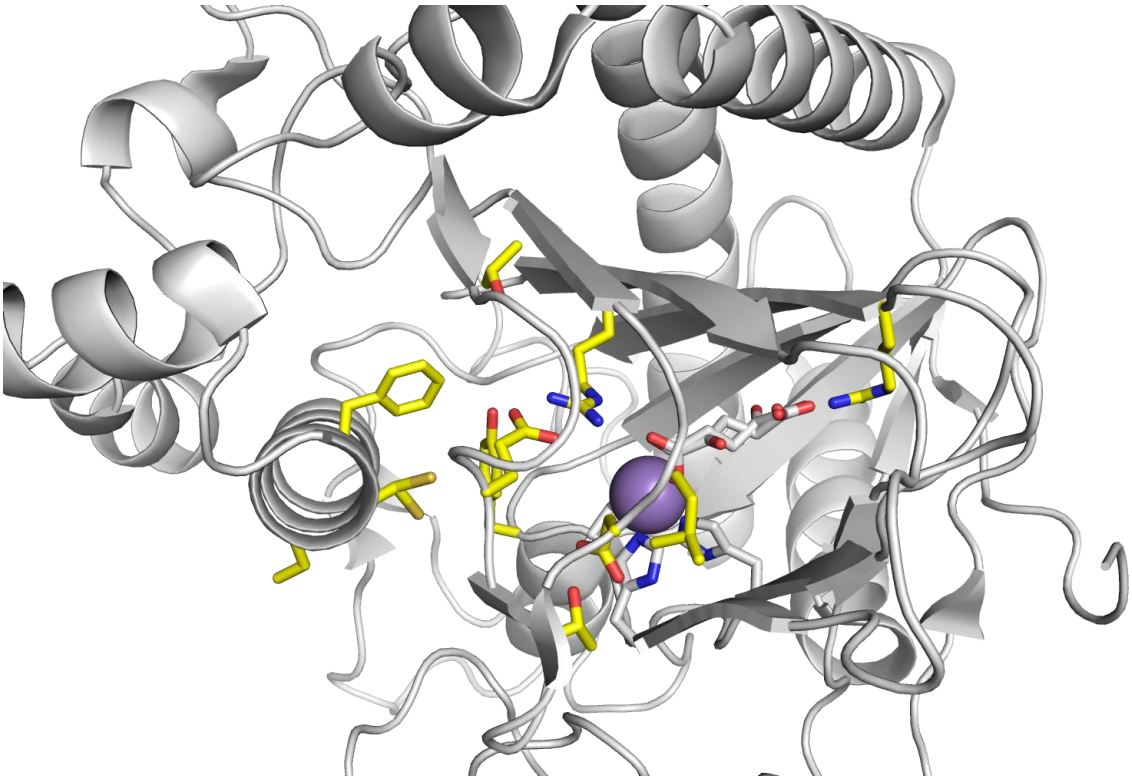


Figure 4-7. Structural representation of *Ps*EFE complexed with α-ketoglutarate (white sticks) and manganese (purple sphere) (PDB ID: 5V2X). Metal-coordinating residues (H189, D191, H268) are depicted with white sticks and the side chains of residues targeted for site-saturation mutagenesis in round 1 are depicted with yellow sticks.

Apparent beneficial mutations found in the site-saturation libraries were then recombined by StEP PCR. The recombined mutations were A171Y, A171F, I186F, M314F, M314L, M317I, and M317V, including the parent residue at each position. The most improved variant from this library was *Ps*EFE R171Y F314M R316E C317M (YMEM), adding the mutations A171Y and R316E (Figure 4-8).

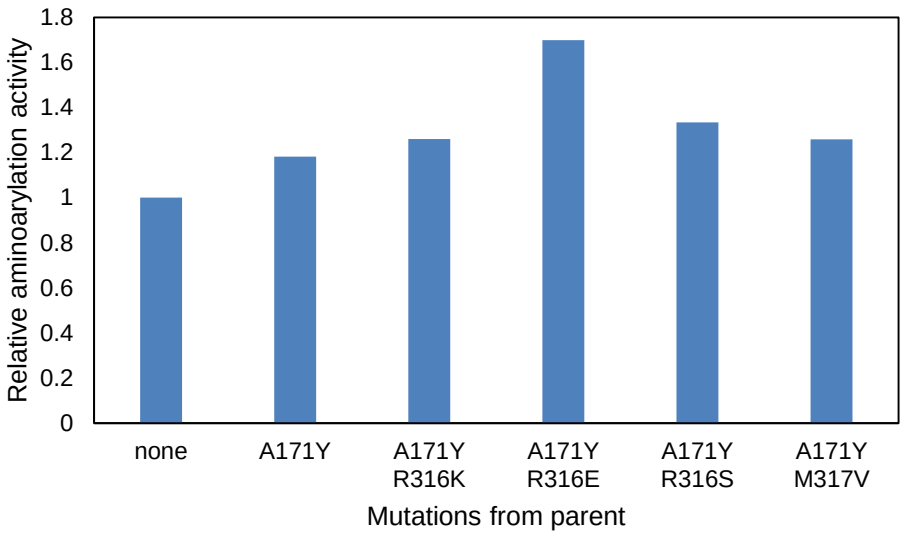


Figure 4-8. Evaluation of candidate variants from recombination of site-saturation round 1 mutations as cell lysate, with 4-methoxybenzenesulfonyl azide as nitrogen donor. Yields determined by mass spectrometry (*m*/*z* = +241).

Site-saturation mutagenesis round 2

The most active variant for aminoarylation from the first round of mutagenesis/ recombination, *Ps*EFE YMEM, was then subjected to a second round of site-saturation mutagenesis, targeting the same positions but omitting positions 171 and 316. The most improved variant from these libraries was *Ps*EFE T97V R171Y F314M R316E C317M (VYMEM), adding the T97V mutation (Figure 4-9).

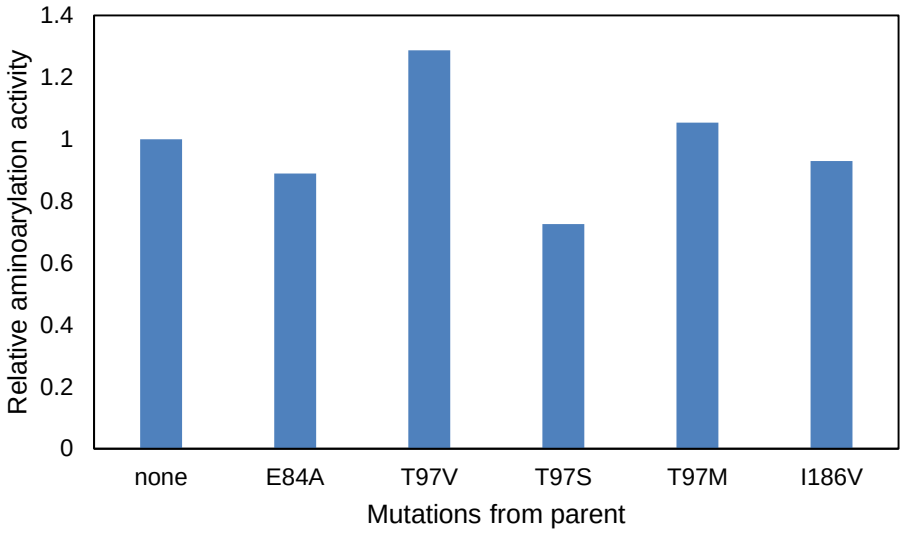


Figure 4-9. Evaluation of candidate variants from site-saturation round 2 as cell lysate, with 4-methoxybenzenesulfonyl azide as nitrogen donor. Yields determined by mass spectrometry (*m*/*z* = +241).

Further evolutionary attempts

Beneficial mutations can occur globally throughout a protein, and are not limited to residues immediately in or near the active site. Since site-saturation rounds 1 and 2 targeted most of the readily apparent residues in the first shell of the active site, I decided to turn to less focused mutagenesis strategies, in order to find mutations at sites that are not obvious from visual inspection of *Ps*EFE structures. Despite extensive screening efforts, however, no significantly improved variant was found. This suggests that *Ps*EFE VYMEM is at or near a local fitness maximum in protein sequence space, and further evolutionary improvements would require exploration beyond closely related variants.

I first constructed and screened region saturation libraries of *Ps*EFE VYMEM, as described by Knight[14]. These libraries are essentially pooled site-saturation libraries of every position within a "tile" of sequence space. Like established random mutagenesis techniques such as error-prone PCR[15] or SeSAM[16], tile libraries provide access to mutations at every position within the tile, but they contain every possible amino acid substitution, including those inaccessible by single-nucleotide mutations. I created two such libraries, targeting positions

180–228 and 293–342. Together, these two tiles cover much of the active site, the flexible "lid-loop" region, and other flexible linkers (Figure 4-10).

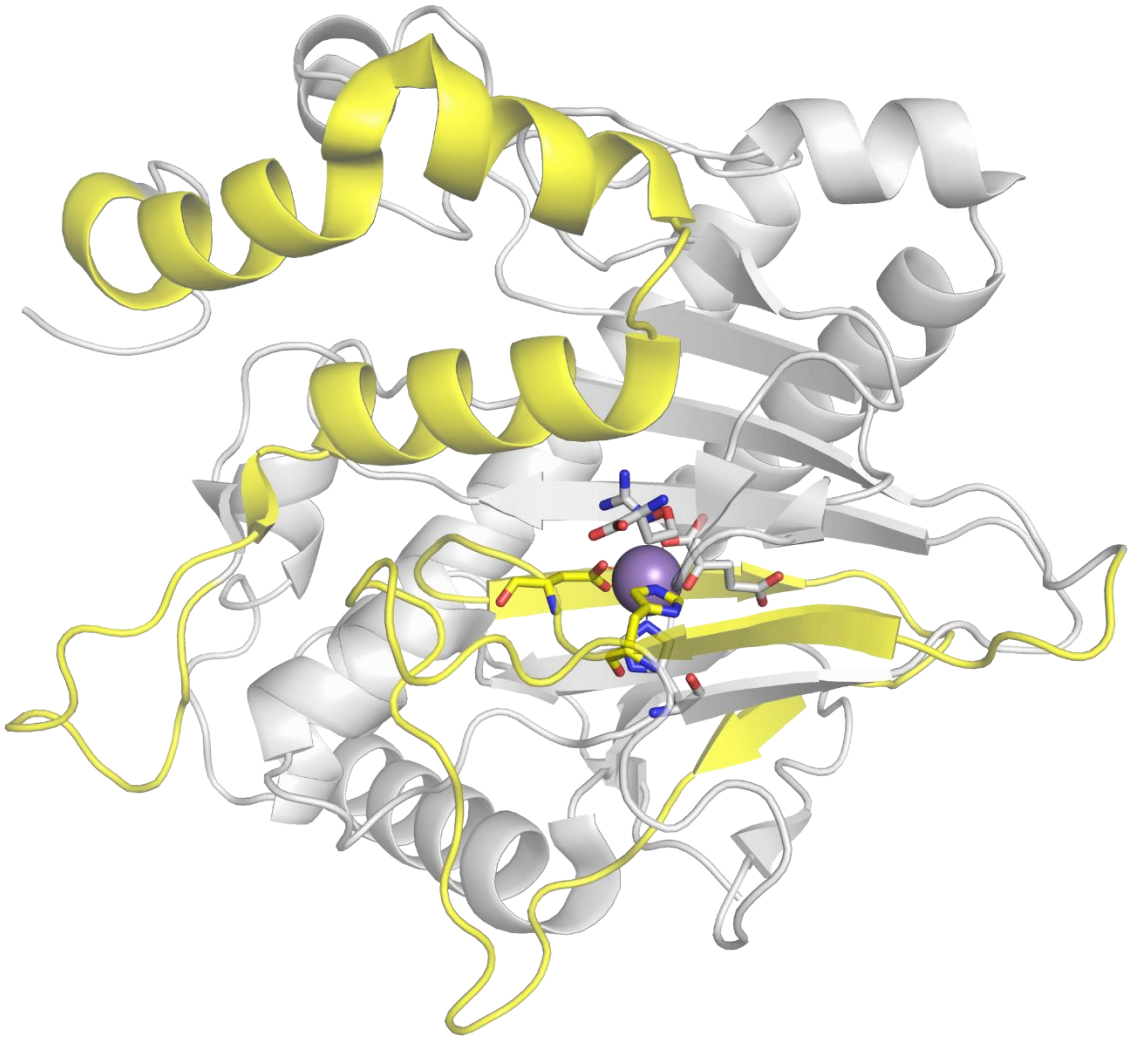


Figure 4-10. Structural representation of *Ps*EFE (PDB ID: 6CBA). Regions targeted by tile saturation libraries are depicted in yellow.

Screening of these libraries revealed apparent beneficial mutations at position 181 (T181S), 182 (L182Y), 184 (R184K), 190 (T190Q, T190D), 298 (P298R), and 316 (E316R reversion). Recombination of these mutations yielded a variant, *Ps*EFE T97V R171Y T181S L182Y T190D P298R F314M C317M (VYSYDRMM), which displayed ~40% higher aminoarylation activity than *Ps*EFE VYMEM in cell lysate, although slightly lower activity with purified protein.

Nevertheless, I sought to engineer *Ps*EFE VYSYDRMM further using random mutagenesis by error-prone PCR. Screening of ~1000 clones from random mutagenesis libraries gave a few apparent improved variants, but upon validation these variants displayed essentially parent-like activity.

Summary of evolutionary lineage

Although the final evolved variant, *Ps*EFE VYMEM still possessed low aminoarylation activity, this represents a 75-fold improvement from the parent, *Ps*EFE AMM, with a corresponding increase in the chemoselectivity, when 4-trifluoromethylbenzenesulfonyl azide is used as substrate (Table 4-1).

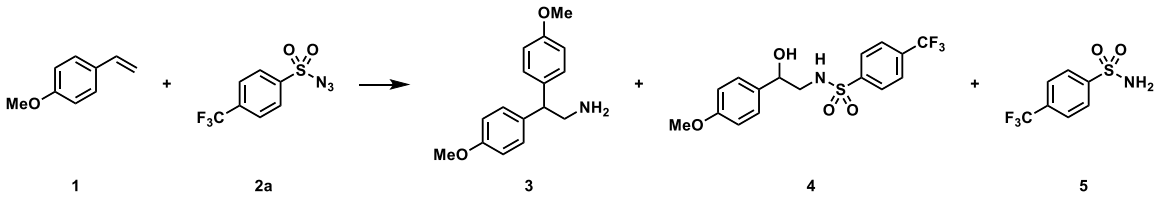


Table 4-1. Evaluation of evolutionary lineage variants as purified protein. Reactions were performed as described in Section 4.6 (Analytical-scale biocatalytic reactions) with sodium acetate as additive, 15 mM 4-methoxystyrene, 10 mM 4-trifluoromethylbenzenesulfonyl azide, and 50 μM purified apoenzyme. Yields were determined by HPLC-UV at 230 nm, except the yield of **3** with variant AMM, which was determined by mass spectrometry (*m*/*z* = +241), and are reported as the average of triplicate reactions.

Variant	**3** (μM)	**4** (mM)	**5** (mM)
R171A F314M C317M	2	0.825	0.971
R171Y F314M R316E C317M	46	0.900	1.421
T97V R171Y F314M R316E C317M	121	0.808	1.674

4.4 Mechanistic investigations and reaction controls

Isotopic labeling experiments

Isotope labeling experiments rule out the Smiles rearrangement hypothesis, as both arenes in the product derive from the 4-methoxystyrene substrate. The amine nitrogen derives from N_3 of the sulfonyl azide, in contrast to conventional nitrene transfer chemistry (including the

Smiles hypothesis) which transfers N_1. A label at the vinylic β-position is incorporated once into the product, suggesting that the 2-carbon aliphatic fragment derives from the vinyl group of one styrene molecule; the other is lost. (Figure 4-11).

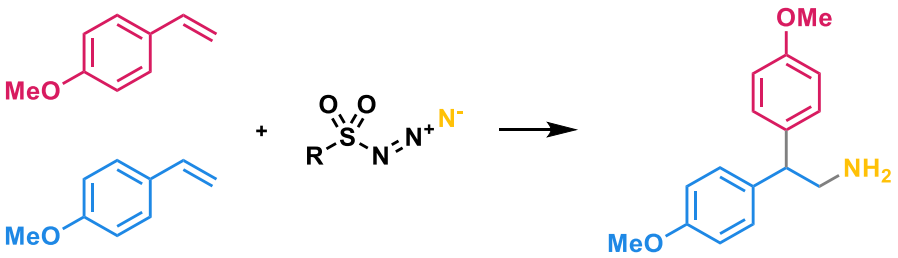
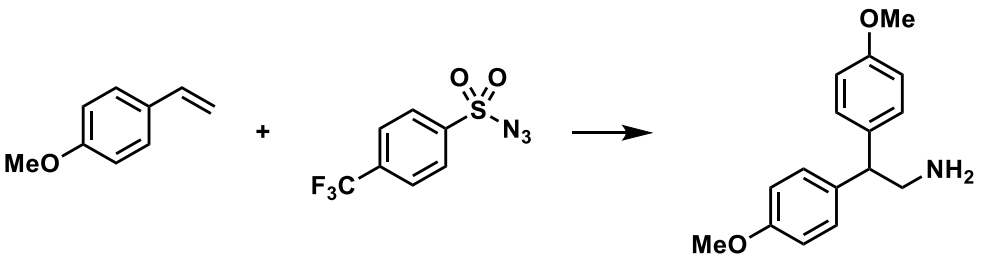
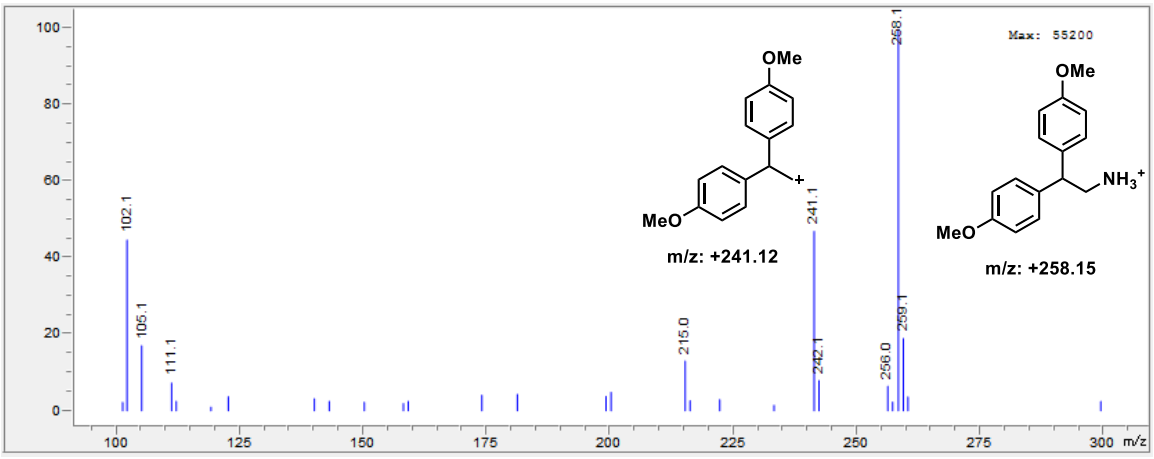


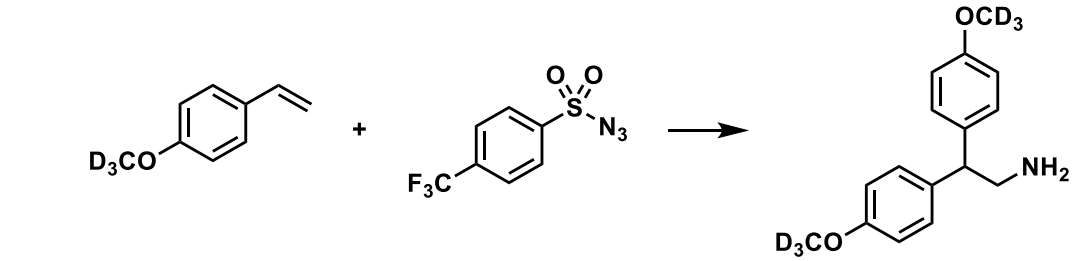
Figure 4-11. Atom mapping determined by isotope labeling.

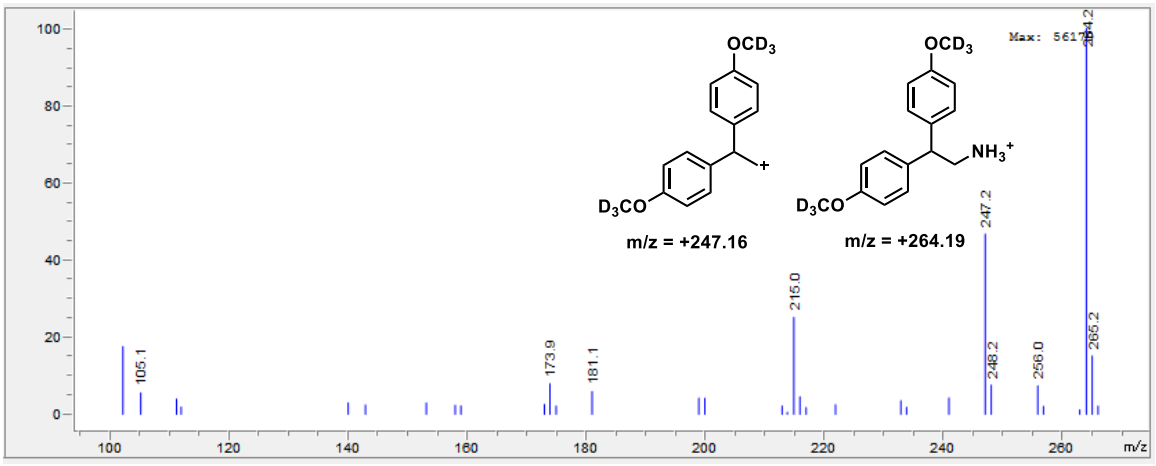
Unlabeled reaction



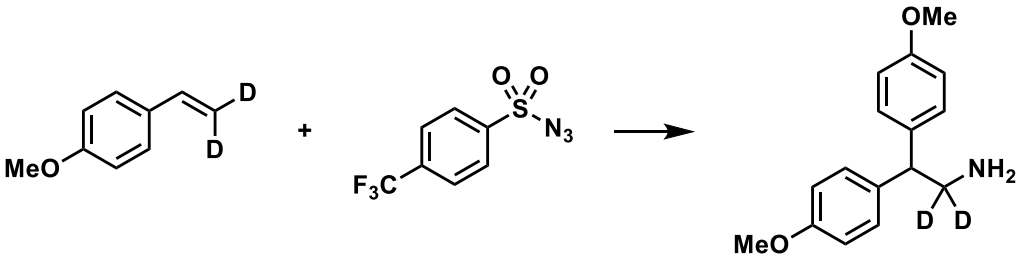


4-(Methoxy-d_3)styrene label

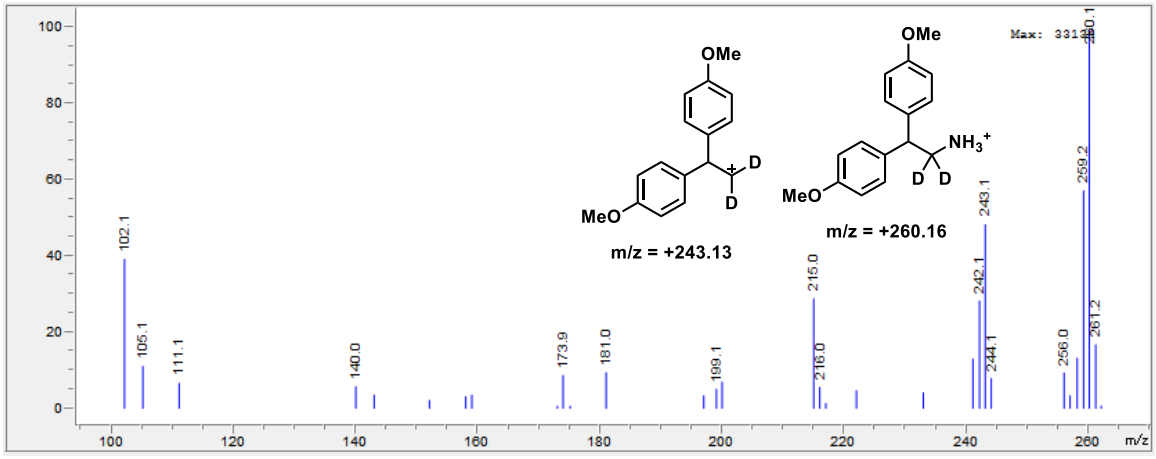




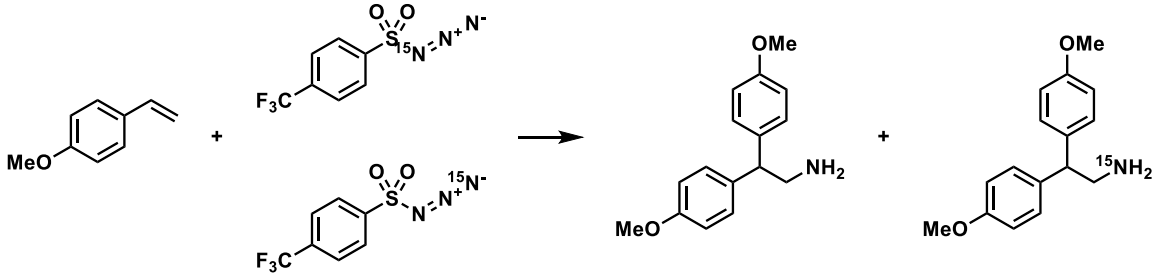
4-Methoxystyrene-β,β-d_2



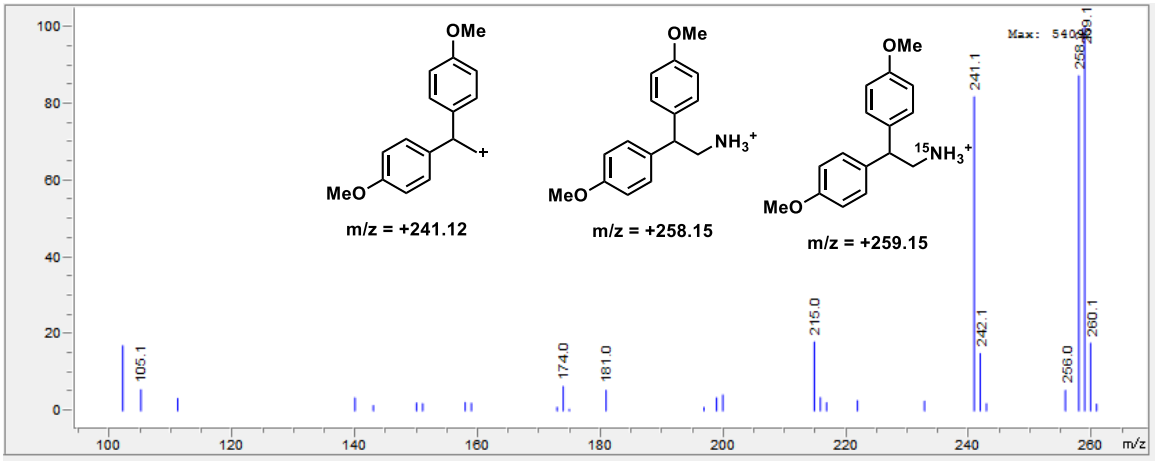
74% labeling at each position



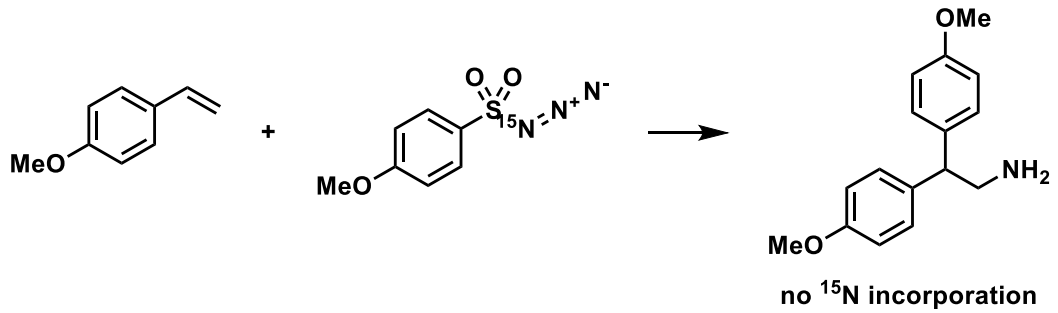
4-Trifluoromethylbenzenesulfonyl azide-(1,3)-^{15}N



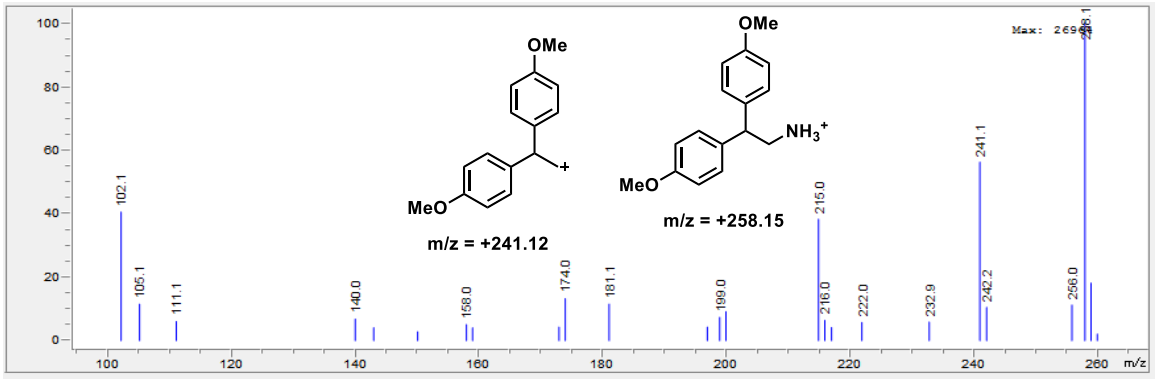
50% ^{15}N incorporation



4-Methoxybenzenesulfonyl azide-1-^{15}N



no ^{15}N incorporation



Substrate scope

I performed a limited substrate scope study to gain insights into the requirements of the reaction (Figure 4-12). A range of arylsulfonyl azides are tolerated in the reaction, with aminoarylation increasing and aziridination decreasing with increasingly electron-withdrawing substituents. The olefin substrate, however, is less tolerant of substitution. Styrene derivatives without a 4-methoxy substituent did not undergo detectable aminoarylation, yielding only aziridine products. Using mixtures of 4-methoxystyrene and other styrene derivatives did not give the corresponding mixed aminoarylation product, only the bis(4-methoxyphenyl) product, suggesting that the 4-methoxy substituent is important for both coupling partners in the reaction. A methyl group at the vinylic β-position is tolerated, albeit with substantially reduced yield, but not at the α-position.

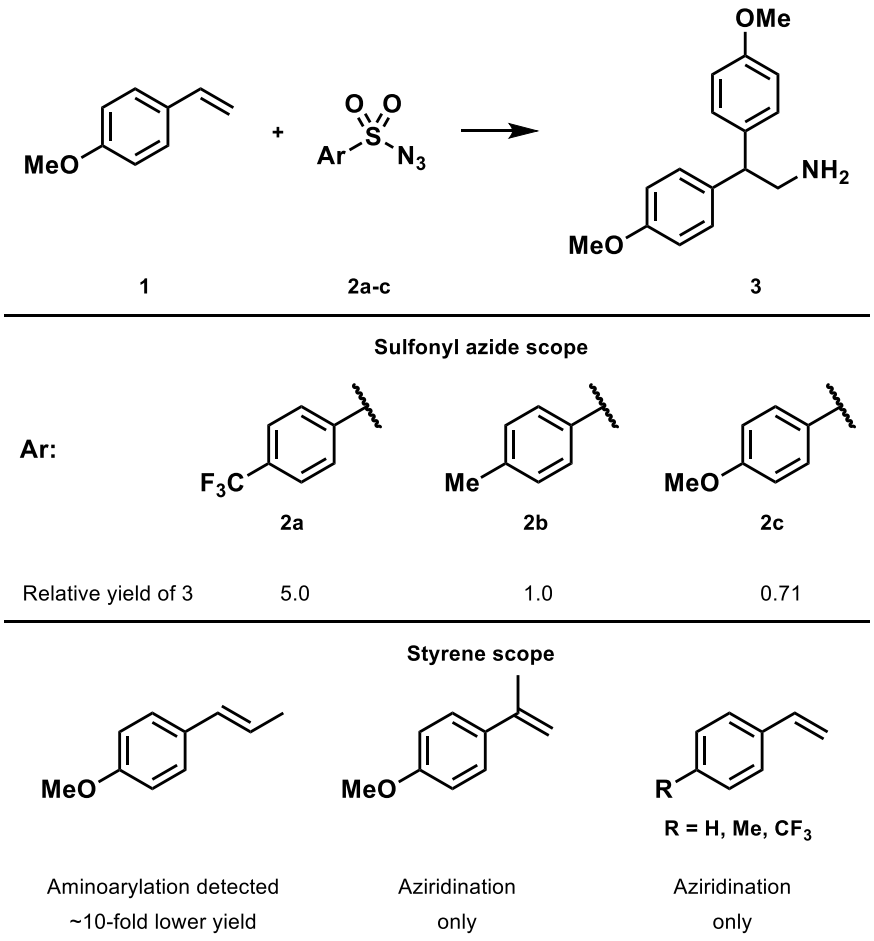


Figure 4-12. Aminoarylation substrate scope.

Reaction controls and additive studies

I performed a set of reaction control experiments (Table 4-2). As expected, in the absence of either enzyme or added iron(II), no detectable aminoarylation product **3** or aziridination product **4** is observed, with only a small amount of reduction of the azide to the sulfonamide **5**. Under aerobic conditions, aziridination still proceeds, albeit with somewhat reduced yield, similar to that seen previously with other *Ps*EFE variants (Chapter 3), but there is no detectable aminoarylation. I also tested a number of carboxylate-containing reaction additives as potential metal-coordinating ligands. Of the additives tested, acetate was clearly the best for promoting aminoarylation, giving nearly two-fold higher yield than the next best, *N*-oxalylglycine. The other additives tested gave little to no aminoarylation yield.

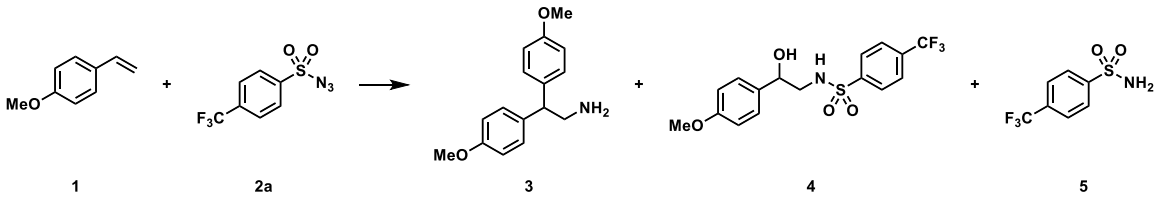


Table 4-2. Reaction condition controls. Under the standard conditions, reactions were performed as described in Section 4.6 (Analytical-scale biocatalytic reactions) with sodium acetate as additive, 15 mM **1**, 10 mM 4-**2a**, and 50 μM *Ps*EFE VYMEM. Yields were determined by HPLC-UV at 230 nm, and are reported as the average of triplicate reactions.

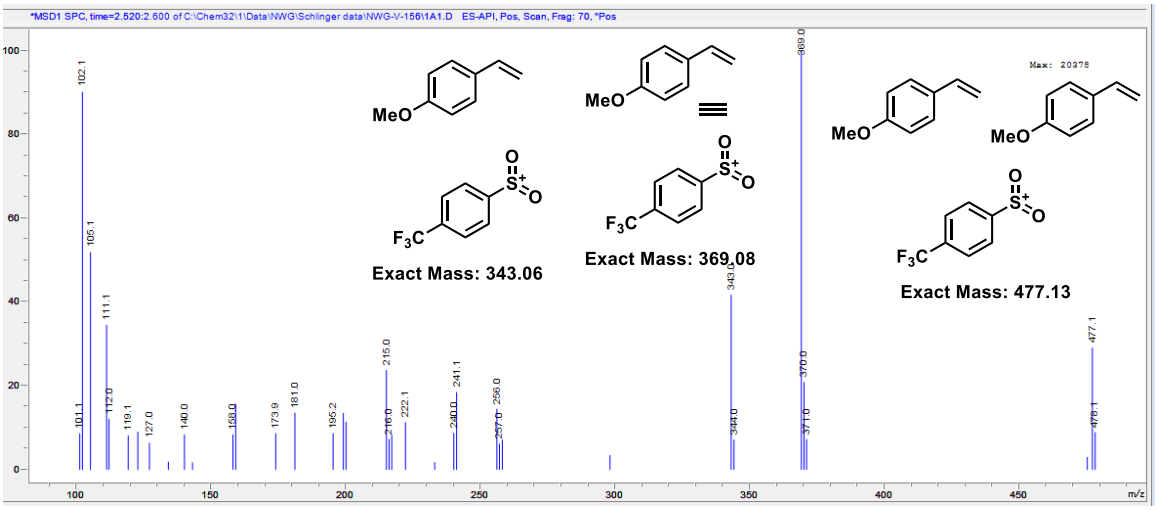
Deviation from standard conditions	**3** (μM)	**4** (mM)	**5** (mM)
None	121	0.808	1.67
No enzyme	n.d.	n.d.	0.233
No iron	n.d.	n.d.	0.356
No acetate	<0.1	0.193	0.569
Aerobic reaction	n.d.	0.382	0.840
N-Oxalylglycine (disodium salt) instead of sodium acetate	62.7	0.667	3.89
α-Ketoglutarate (disodium salt) instead of sodium acetate	<0.1	0.371	0.944
Glycolic acid instead of sodium acetate	14	1.80	1.80
Glycine instead of sodium acetate	1	0.244	0.453
Disodium succinate instead of sodium acetate	1	0.157	0.474
Glutaric acid instead of sodium acetate	2	0.345	0.703

Unknown side products

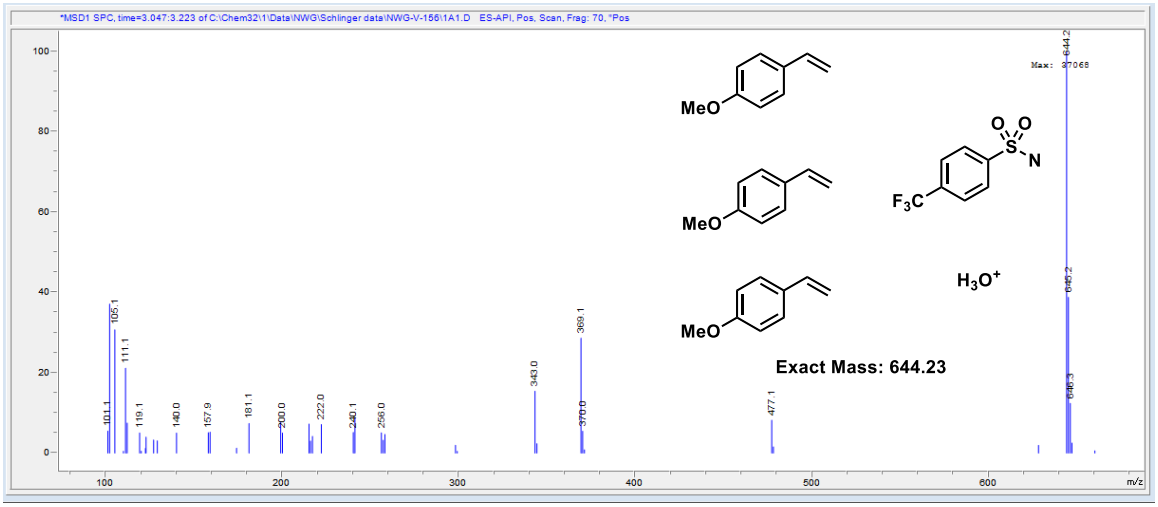
A number of related side products can be observed by mass spectrometry. All of the observed side products incorporate one sulfonyl fragment from the azide substrate, and several (up to 4) 4-methoxystyrene fragments, although the exact structures and connectivities are unknown. It is quite unlikely that 5 substrate molecules can fit into the active site of the enzyme, suggesting that there may be a reactive intermediate that is able to escape into free solution, and initiate some kind of oligomerization of 4-methoxystyrene. Notably, a large

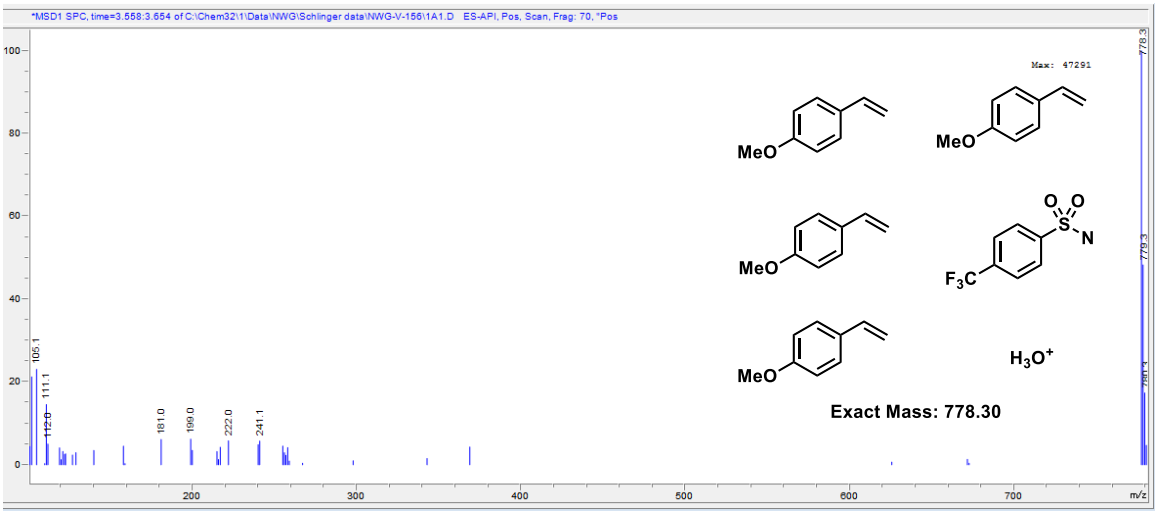
fraction (often >50%) of the 4-methoxystyrene mass balance cannot be accounted for by known products or remaining starting material. The mass spectra of some of these products are shown below, along with proposed mass fragments. Some products appear in pairs, differing by one 4-methoxystyrene moiety. The proposed 4-methoxystyrene multiplicities are corroborated by isotope labeling experiments.

Side product #1

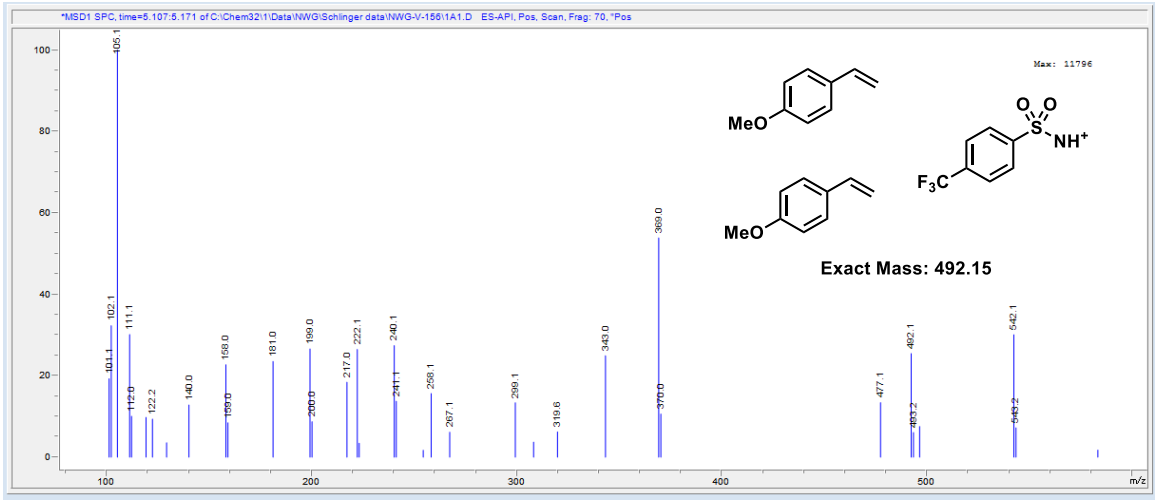


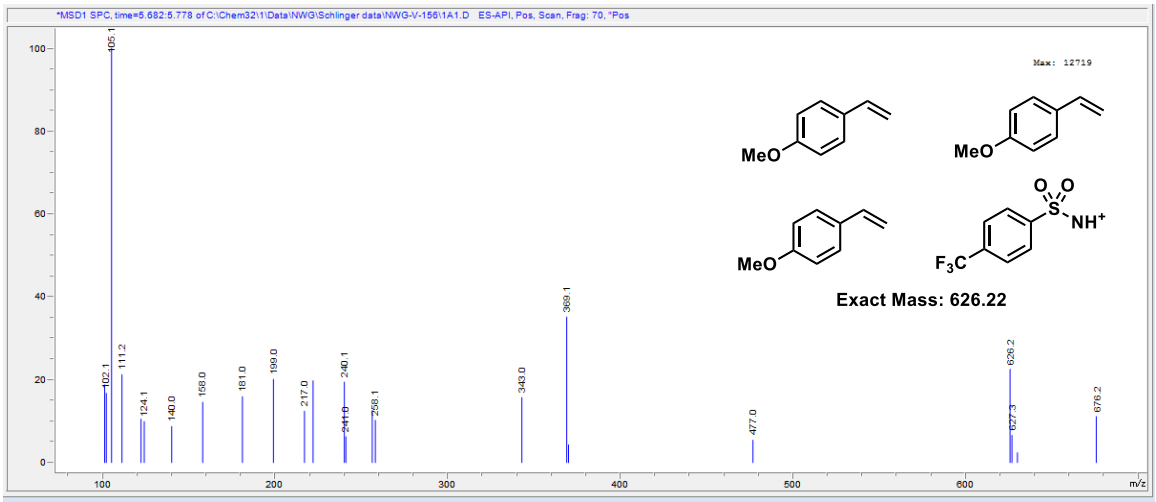
Side products #2a and 2b





Side products #3a and 3b





Possible aminoarylation mechanistic pathway

While I do not have sufficient data to reasonably propose a complete aminoarylation mechanism, I can offer some speculative possibilities (Figure 4-13). The incorporation of N_3 into the aminoarylation product instead of N_1, as observed with conventional nitrene transfer, suggests that the mechanistic branch point between these two pathways could be the regioselectivity of the binding of the sulfonyl azide to iron. Binding at N_1, with subsequent elimination of dinitrogen, leads to nitrene transfer products, while binding at N_3 leads to aminoarylation. The N_3-sulfonyl azide-iron complex could then break down to give a sulfonyl radical, which could escape the enzyme active site and initiate oligomerization of the styrene substrate, and some kind of iron-nitrogen species, which could react with 4-methoxystyrene to form the aminoarylation product.

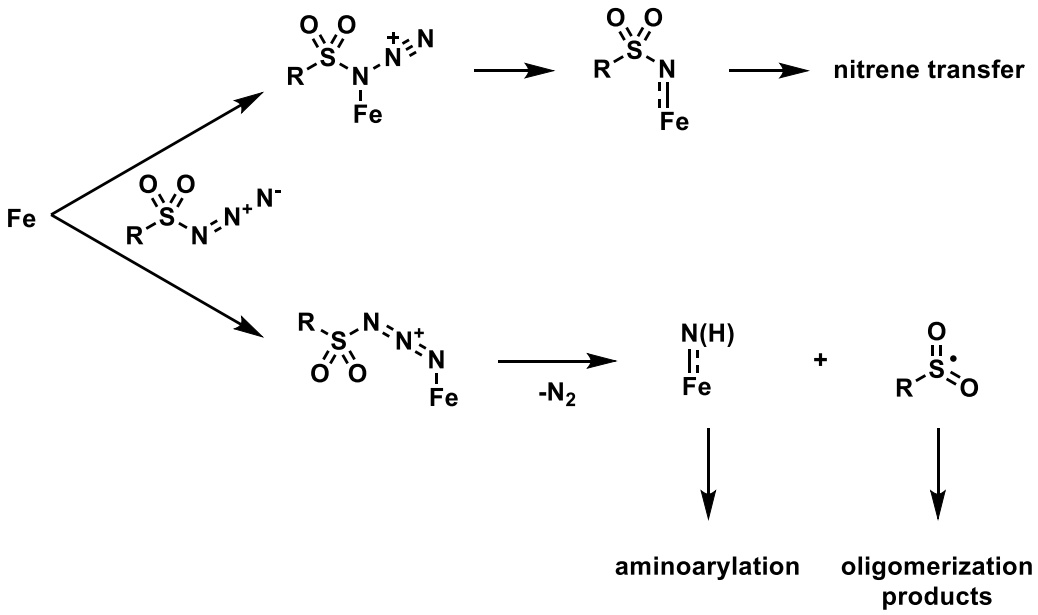


Figure 4-13. Mechanistic speculation.

4.5 Conclusions

I have discovered a novel olefin aminoarylation reaction of electron-rich olefins and sulfonyl azides, catalyzed by a non-heme iron enzyme. I was able to improve the activity significantly by directed evolution, although the final activity is still relatively low. The specific details of the mechanism remain unclear, but the data I do have indicate that this transformation is quite unusual, bearing little resemblance to any known reactions of these substrates.

4.6 Supplementary information for Chapter 4

Synthetic chemistry

Air-sensitive manipulations were performed using flame-dried glassware and standard Schlenk techniques under an atmosphere of argon.

Solvents

ACS- and HPLC-grade solvents were purchased from Fisher Chemical. Anhydrous tetrahydrofuran was obtained by filtration through a drying column and a deoxygenation column on a Pure Process Technologies solvent system. High-purity water for PCR and HPLC was distilled after filtration through a deionizing column and organic removal column. Deuterated solvents were purchased from Cambridge Isotope Laboratories.

Chromatographic materials

Thin layer chromatography (TLC) was performed using EMD TLC plates pre-coated with 250 μm thickness silica gel 60 F_{254} and visualized by fluorescence quenching under UV light and staining with potassium permanganate or cerium ammonium molybdate. Preparative flash chromatography was performed using a Biotage Isolera automated chromatography instrument using columns hand-packed with silica gel (230–400 mesh, Silicyle Inc.).

Starting materials

All compounds were used as received from commercial suppliers, unless otherwise stated.

Analytical instrumentation

HPLC-MS analysis for initial activity determination on an Agilent 1260 Infinity HPLC with an Agilent 6120 quadrupole mass spectrometer. Quantitative HPLC-UV analysis was performed on an Agilent 1260 Series Infinity II HPLC connected to an Agilent iQ MSD, using an Agilent Poroshell 120 EC-C18 column (4 μm packing, 2.1×50 mm) fitted with a Poroshell 120 guard column (1.7 μm packing, 2.1×5 mm). Library screening was performed on an Agilent 1260 Series Infinity II HPLC connected to an Agilent iQ MSD, using an Agilent Poroshell SB-C18 column (2.7 μm packing, 4.6×5 mm). Isotope labeling experiments were performed on an Agilent 1290 UPLC-MS equipped with a C18 silica column (1.8 μm packing, 2.1×50 mm).

NMR spectra were recorded on a Varian Unity/Inova 500 spectrometer operating at 500 MHz and 125 MHz for ^{1}H and ^{13}C, respectively, a Bruker Avance 400 spectrometer operating at 400 MHz and 100 MHz for ^{1}H and ^{13}C, respectively, or a Varian Mercury 300 spectrometer operating at 272 MHz for ^{19}F. NMR data were analyzed in MestReNova

(MestreLab Research). Chemical shifts are reported in ppm with the solvent resonance as the internal standard. For ^{1}H NMR: $CDCl_3$, δ 7.26, CD_3CN, δ 1.94, D_2O, δ 4.79. For ^{13}C NMR: $CDCl_3$, δ 77.16, CD_3CN, δ 1.32. Data are reported as follows: s = singlet, d = doublet, t = triplet, q = quartet, spt = septet, m = multiplet, br = broad; coupling constants in Hz; integration.

Biology and biocatalytic reactions

Materials

Oligonucleotides were purchased from IDT DNA. PCRs were run with Phusion® High-Fidelity PCR Kit (New England Biolabs). Gibson assembly mix[17] is prepared with isothermal master mix in-house and enzymes T5 exonuclease, Phusion® DNA polymerase, and *Taq* DNA ligase purchased from New England Biolabs. Oligonucleotide, DNA, and amino-acid sequences used in this work are given in the supplemental Excel document.

Cloning

The plasmid encoding wild-type *Pseudomonas savastanoi* ethylene-forming enzyme (Uniprot ID P32021), was purchased from Twist Biosciences, encoded with a C-terminal His_6-tag for purification and inserted between the NdeI and XhoI cut sites in the pET-22b(+) vector (Novagen).

Plasmids were used to transform *E. cloni* BL21(DE3) cells (Lucigen) by electroporation. SOC medium (0.75 mL) was added and the cells were incubated at 37 °C for 45 minutes before being plated on Luria-Bertani medium (Research Products International) supplemented with ampicillin (100 μg mL^{-1}, LB-amp) agar plates.

Plasmids were isolated from stationary-phase cultures by miniprep (Qiagen) and Sanger sequencing was performed by Laragen, Inc. (Culver City, CA) using T7 promoter and T7 terminator primers.

Protein expression and purification

Starter cultures of LB-amp were inoculated from a single *E. coli* colony on an agar plate harboring a plasmid encoding the protein of interest and grown overnight to stationary phase at 37 °C. Expression cultures of Terrific Broth (Research Products International)

supplemented with ampicillin (100 mg L^{-1}, TB-amp) were inoculated from the starter cultures (1% v/v) and shaken at 37 °C and 160 rpm in a Multitron Infors incubator. When the expression cultures reached OD_{600} ~ 0.8 (typically 2–3 hours), they were cooled to room temperature. Protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.1 mM). Cultures were incubated at 22 °C and 160 rpm overnight (16–24 hours). Cells were pelleted by centrifugation (5000×*g*, 10 minutes).

For reactions with whole cells, cell pellets were resuspended in MOPS buffer (20 mM pH 7.0) to OD_{600} 40. For reactions with cell lysate, the whole cell suspensions were lysed by sonication (QSonica Q500 sonicator, 25% amplitude, 33% duty cycle, 3 minutes). The lysate was clarified by centrifugation (20,817×*g*, 10 minutes).

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

Site-saturation mutagenesis and library screening

Site-saturation mutagenesis was performed using the 22-codon method[18]. Oligonucleotides including the three 22-codon trick codons (NDT, VHG, TGG) and oligonucleotides within the ampicillin resistance cassette were used to amplify the plasmid in two pieces, with an

overlap for Gibson assembly in the gene encoding the protein of interest (and where the mutation is introduced) and an overlap for Gibson assembly in the gene encoding β-lactamase (which confers ampicillin resistance). This two-piece assembly can prevent mis-assembled constructs from conferring antibiotic resistance. The oligonucleotide sequences which generated the variants reported in in this work are listed in Table 4-3. PCR products were loaded on 1% agarose gels with loading dye containing GelGreen nucleic acid gel stain (Biotium) and visualized on a blue transilluminator. The DNA bands at the expected size were excised and the DNA was extracted with a Zymoclean Gel DNA recovery kit. The two linear PCR products for a given site-saturation mutagenesis library were assembled *via* isothermal Gibson assembly (50 °C, 1 hour).

For recombination, a plasmid template was isolated from pooled stationary-phase cultures of the variants to be recombined by miniprep (Qiagen). This template was used for staggered extension process (StEP) PCR[19] with *Taq* DNA polymerase (New England Biolabs) to amplify the insert encoding *Ps*EFE, using the insert amplification primers in Table 4-4. This insert was then assembled with the amplified vector (amplified with the vector amplification primers in Table 4-4) *via* isothermal Gibson assembly (50 °C, 1 hour).

E. cloni BL21(DE3) cells (Lucigen) were transformed by electroporation with the Gibson assembly product without further purification. SOC medium (0.75 mL) was added to the electroporated cells and the cells were incubated at 37 °C for 45 minutes before being plated on LB-amp agar plates. The LB-amp agar plates with the plated cells were incubated at 37 °C for 12–18 hours and stored at 4 °C until the libraries were picked. Single colonies from the agar plates were picked with sterile toothpicks and used to inoculate starter cultures (0.5 mL LB-amp) in polypropylene 96 deep-well plates. The starter culture plates were grown at 37 °C, 220 rpm, and 80% humidity in a Multitron Infors shaker overnight (14–16 hours). Expression cultures (950 μL TB-amp) were inoculated with overnight culture (50 μL), and grown at 37 °C, 220 rpm, and 80% humidity for three hours, then cooled on ice for 20 minutes. Protein expression was induced by addition of IPTG (0.1 mM). Cultures were incubated at 22 °C and 220 rpm overnight (18–20 hours). Cells were pelleted (5000×*g*, 5 minutes) and the cell pellets were frozen at −20 °C for at least 24 hours prior to use.

Cells were resuspended in MOPS buffer (20 mM pH 7.0, 360 μL) containing 1 mM sodium acetate and brought into a Coy vinyl anaerobic chamber (nitrogen atmosphere, 0–10 ppm oxygen, 2–3% hydrogen). Ferrous ammonium sulfate (40 mM in water, 10 μL, 1 mM final concentration, prepared immediately before use), L-ascorbic acid (40 mM in water, 10 μL, 1 mM final concentration, prepared immediately before use), 4-methoxystyrene (400 mM in ethanol, 10 μL, 10 mM final concentration), and 4-methoxybenzenesulfonyl azide (400 mM in acetonitrile, 10 μL, 10 mM final concentration) were added to each well. The plates were sealed with foil covers and shaken at room temperature for two hours. To quench the reactions, acetonitrile (400 μL) was added and the reaction plate was shaken for an additional 30 minutes. Insoluble material was pelleted by centrifugation (6000×*g*, 10 minutes) and 200 μL of the supernate was filtered through a 0.2 μm PTFE 96-well filter plate into a 96-well microplate (3000×*g*, 2 minutes). The microplate was sealed with a pierceable cover and analyzed via HPLC-MS (Analytical instrumentation), detecting m/z = +241 ($[M–NH_2]^+$ for the aminoarylation product).

Wells which showed an apparent enhancement in aminoarylation product formation were streaked out from the overnight cultures onto LB-amp agar plates and incubated at 37 °C. Starter cultures (5 mL LB-amp) were inoculated from single colonies and incubated overnight at 37 °C. Plasmids were isolated from these cultures and sequenced as described above (Cloning). Unique variants were then regrown in Erlenmeyer flasks as described above (Protein expression and purification) and assayed in clarified cell lysate for aminoarylation activity.

Table 4-3. Oligonucleotides used for mutagenesis. Mutated codons are denoted here as *NNN* for simplicity; in practice they are a 12:9:1 ratio of NDT:VHG:TGG for site saturation or the appropriate single codon for directed mutagenesis.

Mutations relative to wild type	Direction	Sequence
T97X	Forward	CCGACTTCCCCGAAATTTTC*NNN*GTCTGCAAAGATCTTTC
T97X	Reverse	GAAAATTTCGGGGAAGTCGGGCTTTCCAGCAGTCACCTC
R171X	Forward	GATGGATGGCACCACATG*NNN*GTGTTGCGTTTTCCGCC
R171X	Reverse	CATGTGGTGCCATCCATCGCGGGTCAAATCTG
F314X C317M	Forward	CACTATGGGGAACATTTCACGAACATG*NNN*ATGCGTATGTATCCTGACCG
F314X	Reverse	CATGTTCGTGAAATGTTCCCCATAGTGAATGCGCTCATTGGCC
F314M C317M R316X	Forward	ATTTCACGAACATGATGATG*NNN*ATGTATCCTGACCGCATTACCACAC
F314M R316X	Reverse	CATCATCATGTTCGTGAAATGTTCCCCATAGTGAATGCGCTC
C317X	Forward	TCACGAACATGTTCATGCGT*NNN*TATCCTGACCGCATTACCACACAGC
C317X	Reverse	CATGAACATGTTCGTGAAATGTTCCCCATAGTGAATGCGCTC

Table 4-4. Miscellaneous oligonucleotides.

Description	Direction	Sequence
ampR	Forward	CCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGC
ampR	Reverse	CGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAG
Insert amplification	Forward	GAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATG
Insert amplification	Reverse	GCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAG
Vector amplification	Forward	CTCGAGCACCACCACCACCACCACTGAGATCCGGC
Vector amplification	Reverse	CATATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTC

Analytical-scale biocatalytic reactions

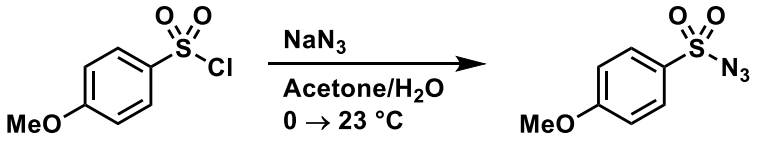
Biocatalytic reactions were set up in 2-mL screw-cap vials (Agilent). Purified apoprotein (360 μL, 55.6 μM in 20 mM MOPS pH 7.0, final concentration 50 μM) was added to the vial. Solutions of ferrous ammonium sulfate were prepared immediately prior to use. Reactions were brought into a Coy vinyl anaerobic chamber (nitrogen atmosphere, 0–10 ppm oxygen, 2–3% hydrogen). To each reaction was added in order ferrous ammonium sulfate (40 mM in water, 10 μL, 1 mM final concentration), and sodium acetate or other additive (40 mM in water, 10 μL, 1 mM final concentration). Each reaction was then charged with 4-methoxystyrene (600 mM in ethanol, 10 μL, 15 mM final concentration) immediately followed by 4-trifluoromethylbenzenesulfonyl azide or other sulfonyl azide (400 mM in acetonitrile, 10 μL, 10 mM final concentration). The reactions were sealed and shaken at room temperature for two hours unless otherwise noted. To quench the reactions, acetonitrile (350 μL) was added to each vial, followed by internal standard 4-hydroxybenzoic acid (0.5% v/v in acetonitrile, 50 μL). The sample was transferred to a 1.7-mL Eppendorf tube, vortexed, and then centrifuged (20817×*g*, 5 minutes). 250 μL of the supernate was transferred to HPLC vial inserts for reverse-phase HPLC analysis.

Synthesis of substrates

Safety statement

Organic azides are potentially explosive compounds. I have not observed any problems in my handling of the compounds described, but appropriate precautions should be taken, especially on large scales.

4-Methoxybenzenesulfonyl azide (2c)



2c

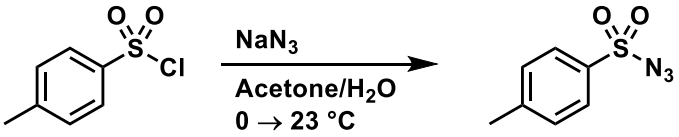
Under air, 4-methoxybenzenesulfonyl chloride (10.0 g, 48.4 mmol, 1.00 equiv.) was dissolved in acetone (100 mL) in a 250 mL round-bottomed flask with a magnetic stir bar and cooled to 0 °C. A solution of sodium azide (4.72 g, 72.6 mmol, 1.50 equiv.) in water (30 mL) was added dropwise over 10 minutes with stirring. Once addition was complete, the reaction was allowed to warm to room temperature and stirred for 2 hours. The acetone was removed *in vacuo*, and the resulting mixture was extracted with diethyl ether (3×30 mL). The combined organic layers were washed with water (30 mL), saturated aqueous sodium bicarbonate (30 mL), and brine (30 mL), then dried over magnesium sulfate and concentrated *in vacuo* to afford the title compound as a colorless solid (9.99 g, 97%).

NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, 23 ºC, δ): 7.90 (d, *J* = 9.1 Hz, 2 H), 7.06 (d, *J* = 9.1 Hz, 2 H), 3.91 (s, 3 H)

^{13}C NMR (100 MHz, $CDCl_3$, 23 ºC, δ): 164.8, 130.1, 129.9, 115.0, 56.0

p-Toluenesulfonyl azide (2b)



2b

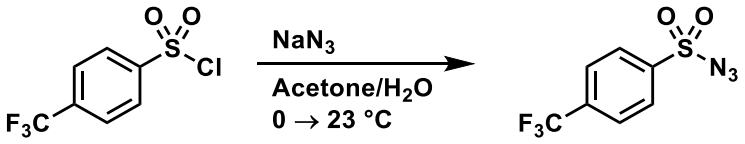
Under air, *p*-toluenesulfonyl chloride (19.1 g, 100 mmol, 1.00 equiv.) was dissolved in acetone (200 mL) in a 500 mL round-bottomed flask with a magnetic stir bar and cooled to 0 °C. A solution of sodium azide (9.75 g, 150 mmol, 1.50 equiv.) in water (60 mL) was added dropwise over one hour with stirring. Once addition was complete, the reaction was allowed to warm to room temperature and stirred for 16 hours. The acetone was removed *in vacuo*, and the resulting mixture was extracted with diethyl ether (2×100 mL). The combined organic layers were washed with water (2×100 mL), saturated aqueous sodium bicarbonate (100 mL), and brine (100 mL), then dried over magnesium sulfate and concentrated *in vacuo* to afford the title compound as a colorless oil that solidified upon storage at −20 °C (19.1 g, 97%).

NMR Spectroscopy:

^{1}H NMR (500 MHz, $CDCl_3$, 23 ºC, δ): 7.80 (d, *J* = 8.5 Hz, 2 H), 7.38 (d, *J* = 8.5 Hz, 2 H), 2.45 (s, 3 H)

^{13}C NMR (125 MHz, $CDCl_3$, 23 ºC, δ): 146.3, 135.3, 130.3, 127.4, 21.7

4-Trifluoromethylbenzenesulfonyl azide (2a)



2a

Under air, 4-trifluoromethylbenzenesulfonyl chloride (12.23 g, 50 mmol, 1.00 equiv.) was dissolved in acetone (75 mL) in a 250 mL round-bottomed flask with a magnetic stir bar and cooled to 0 °C. A solution of sodium azide (4.88 g, 75 mmol, 1.50 equiv.) in water (20 mL) was added dropwise over one hour with stirring. Once addition was complete, the reaction

was allowed to warm to room temperature and stirred for 1.5 hours. The acetone was removed *in vacuo*, and the resulting mixture was extracted with diethyl ether (2×30 mL). The combined organic layers were washed with saturated aqueous sodium bicarbonate (50 mL), and brine (50 mL), then dried over magnesium sulfate and concentrated *in vacuo* to afford the title compound as a colorless solid (12.0 g, 96%).

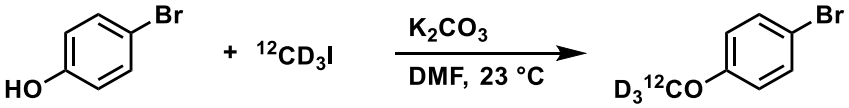
NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, 23 ºC, δ): 8.13–8.08 (m, 2 H), 7.92–7.87 (m, 2 H)

^{13}C NMR (100 MHz, $CDCl_3$, 23 ºC, δ): 142.0 (q, J_{CF} = 1.3 Hz), 136.5 (q, J_{CF} = 33.8 Hz), 128.2, 127.1 (q, J_{CF} = 3.7 Hz), 122.9 (q, J_{CF} = 272 Hz)

^{19}F NMR (282 MHz, $CDCl_3$, 23 °C, δ): –63.4 (s)

1-Bromo-4-(methoxy-^{12}C-d_3)benzene

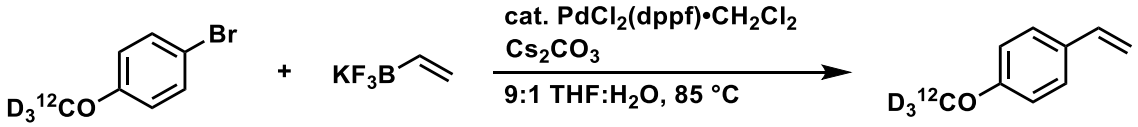


Under air, 4-bromophenol (1.73 g, 10.0 mmol, 1.00 equiv.) and potassium carbonate (2.76 g, 20.0 mmol, 2.00 equiv.) were dissolved/suspended in *N*,*N*-dimethylformamide (30 mL) in a 50 mL round-bottomed flask with magnetic stirring. ^{12}*C*-Iodomethane-d_3 (747 µL, 12.0 mmol, 1.20 equiv.) was added, and the reaction was stirred at room temperature for 16 hours. The reaction mixture was partitioned between water and diethyl ether (50 mL each). The aqueous layer was extracted with additional ether (2×25 mL). The combined ether layers were washed with 5% (w/v) aqueous lithium chloride (2×50 mL) and brine (50 mL), and concentrated to a pale yellow oil (1.87 g, 98%, >99% labeling by NMR).

NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, 23 ºC, δ): 7.41–7.35 (m, 2 H), 6.81–6.75 (m, 2H)

4-(Methoxy-^{12}C-d_3)styrene



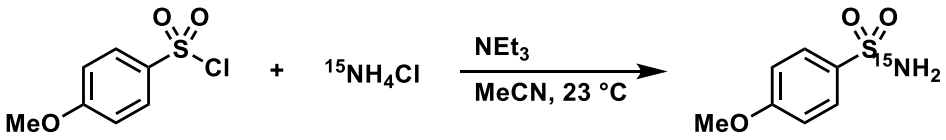
Following a reported procedure[20], 1-bromo-4-(methoxy-^{12}C-d_3)benzene (570 mg, 3.00 mmol, 1.00 equiv.), potassium vinyltrifluoroborate (402 mg, 3.00 mmol, 1.00 equiv.), cesium carbonate (2.93 g, 9.00 mmol, 3.00 equiv.) and [1,1'-bis(diphenylphosphino)ferrocene]dichloropalladium(II) dichloromethane complex (49.0 mg, 60.0 μmol, 0.02 equiv.) were dissolved in 9:1 (v/v) tetrahydrofuran:water (30 mL) in a 100 mL Schlenk tube. The reaction mixture was sparged with argon, then sealed and heated to 85 °C for 22 hours. The reaction was cooled to room temperature and filtered through cotton to remove precipitated palladium black. The filtrate was partitioned between water and diethyl ether (50 mL each), and the aqueous layer was extracted with additional ether (25 mL). The combined organic layers were washed with brine (50 mL), dried over magnesium sulfate, and concentrated *in vacuo*. The residue was purified by automated flash column chromatography on silica gel (25 g), eluting with a 0 to 5% ethyl acetate/hexanes gradient, to afford the title compound as a colorless oil (99 mg, 24%).

NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, 23 ºC, δ): 7.38–7.33 (m, 2 H), 6.90–6.84 (m, 2 H), 6.67 (dd, *J* = 17.6, 10.9 Hz), 5.62 (dd, *J* = 17.6, 1.0 Hz), 5.13 (dd, *J* = 10.9, 1.0 Hz)

^{13}C NMR (100 MHz, $CDCl_3$, 23 ºC, δ): 159.5, 136.4, 130.5, 127.5, 114.0, 111.7

4-Methoxybenzenesulfonamide-^{15}N



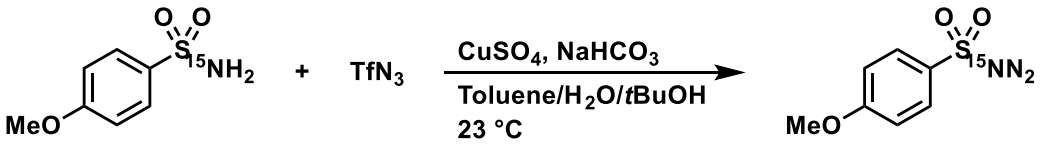
Under air, 4-methoxybenzenesulfonyl chloride (2.07 g, 10.0 mmol, 1.00 equiv.) and ^{15}N-ammonium chloride (817 mg, 15.0 mmol, 1.50 equiv.) were suspended in acetonitrile (50 mL) in a 100 mL round-bottomed flask with magnetic stirring. Triethylamine (4.19 mL, 30.0 mmol, 3.00 equiv.) was added dropwise over 3 minutes, and the reaction was stirred for 30 minutes. The reaction mixture was partitioned between water and ethyl acetate (100 mL each) and the aqueous layer was extracted with additional ethyl acetate (50 mL). The combined organic layers were washed with saturated aqueous sodium bicarbonate (100 mL) and brine (100 mL), then dried over sodium sulfate and concentrated *in vacuo*. The residue was purified by automated flash column chromatography on silica gel (50 g), eluting with a gradient 0 to 60% ethyl acetate/hexanes, to afford the title compound as a colorless crystalline solid (795 mg, 42%).

NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, 23 ºC, δ): 7.89–7.84 (m, 2 H), 7.01–6.95 (m, 2 H), 4.83 (d, J = 80.2 Hz, 2 H), 3.87 (s, 3 H)

^{13}C NMR (100 MHz, $CDCl_3$, 23 ºC, δ): 163.1, 133.8 (d, J_{CN} = 4.5 Hz), 128.8, 114.4, 55.8

4-Methoxybenzenesulfonyl azide-1-^{15}N



Following a reported procedure[21], sodium azide (692 mg, 10.6 mmol, 1.33 equiv.) was dissolved in water (2.7 mL) in an 8 mL vial. Toluene (1.8 mL) was added and the mixture was cooled to 0 °C in an ice/water bath. Triflic anhydride (1.35 mL, 8.00 mmol, 1.00 equiv.) was added dropwise over 5 minutes. The reaction was stirred at 0 °C for 2 hours, then saturated aqueous sodium bicarbonate (1 mL) was added slowly. When bubbling had ceased, the organic layer was removed, and the aqueous layer was extracted with toluene (2×1.8 mL). The combined toluene layers were used for the next step.

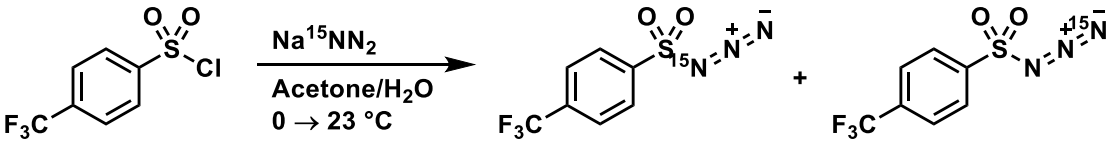
4-Methoxybenzenesulfonamide-^{15}N (500 mg, 2.66 mmol, 1.00 equiv.), sodium bicarbonate, 893 mg, 10.63 mmol, 4.00 equiv.), and copper(II) sulfate (17.0 mg, 106 μmol, 0.04 equiv.) were suspended in water (3.2 mL). The triflyl azide toluene solution was added, followed by tert-butanol (22 mL). The reaction mixture was stirred for 48 hours, then the volatiles were removed *in vacuo* with xylenes in the trap to prevent accumulation of concentrated triflyl azide. The remaining mixture was partitioned between water and ethyl acetate (25 mL each). The aqueous layer was extracted with additional ethyl acetate (2×10 mL). The combined organic layers were washed with brine (2×25 mL), dried over sodium sulfate, and concentrated *in vacuo*. The residue was purified by manual flash column chromatography on silica gel, with a gradient 0 to 60% ethyl acetate/hexanes, to afford the title compound as a colorless solid (546 mg, 96% from the sulfonamide).

NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, 23 ºC, δ): 7.92–7.87 (m, 2 H), 7.08–7.03 (m, 2 H), 3.91 (s, 3 H)

^{13}C NMR (100 MHz, $CDCl_3$, 23 ºC, δ): 164.6, 129.9, 129.8 (d, J_{CN} = 4.0 Hz), 114.9, 55.9

4-Trifluoromethylbenzenesulfonyl azide-(1,3)-^{15}N



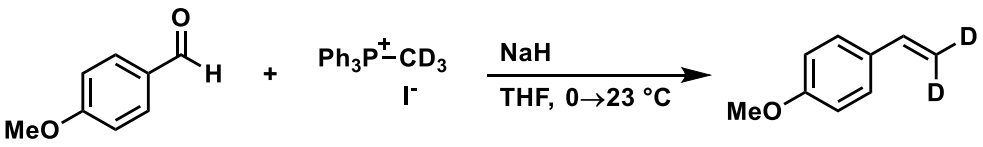
Under air, 4-trifluoromethylbenzenesulfonyl chloride (367 mg, 1.50 mmol, 1.00 equiv.) was dissolved in acetone (3 mL) in an 8 mL vial and cooled to 0 °C. Sodium azide-1-^{15}N (119 mg, 1.80 mmol, 1.20 equiv.), as a solution in water (1 mL), was added dropwise to the sulfonyl chloride solution over 3 minutes. The reaction was warmed to room temperature and stirred for 30 minutes, then partitioned between water and diethyl ether (10 mL each). The aqueous layer was extracted with additional ether (10 mL). The combined organic layers were washed with saturated aqueous sodium bicarbonate (20 mL) and brine (20 mL), then dried over magnesium sulfate and concentrated *in vacuo* to afford the title compound as a colorless solid as a mixture of isotopomers (362 mg, 96%).

NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, 23 ºC, δ): 8.13–8.08 (m, 2 H), 7.92–7.87 (m, 2 H)

^{13}C NMR (100 MHz, $CDCl_3$, 23 ºC, δ): 142.0, 136.5 (q, J_{CF} = 33.8 Hz), 128.2, 127.1 (q, J_{CF} = 3.7 Hz), 122.9 (q, J = 274 Hz)

4-Methoxystyrene-β,β-d_2



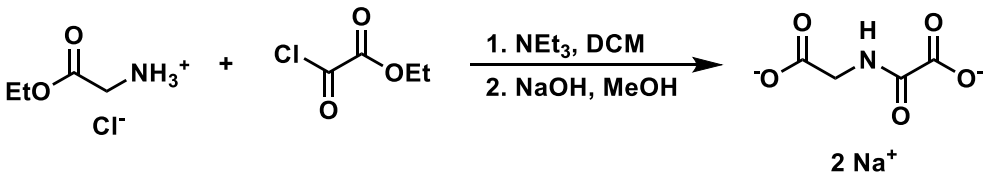
Under argon, (methyl-d_3)triphenylphosphonium iodide (2.04 g, 5.00 mmol, 1.00 equiv.) was suspended in anhydrous tetrahydrofuran (20 mL) in a 50 mL round-bottomed flask with magnetic stirring, and cooled to 0 °C. Under positive argon flow, sodium hydride (60% dispersion in mineral oil, 300 mg, 7.50 mmol, 1.50 equiv.) was added. The reaction was allowed to warm to room temperature and was stirred for 30 minutes, then cooled back to 0 °C. *p*-Anisaldehyde (608 μL, 5.00 mmol, 1.00 equiv.) was added dropwise, then the reaction was allowed to warm to room temperature and stirred for 16 hours. The reaction was cooled to 0 °C, and saturated aqueous ammonium chloride (10 mL) was added dropwise to quench. The resulting mixture was partitioned between water and diethyl ether (25 mL each) and the aqueous layer was extracted with additional ether (25 mL). The combined ether layers were washed with brine (30 mL), dried over magnesium sulfate, and concentrated *in vacuo*. The residue was purified by automated flash column chromatography on silica gel (50 g), eluting with a gradient 0 to 20% ether/hexanes, to afford the title compound as a colorless oil (52 mg, 8%, 74% labeled at each position by NMR).

NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, 23 ºC, δ): 7.39–7.33 (m, 2 H), 6.90–6.85 (m, 2 H), 6.72–6.62 (m, 1 H) 5.65–5.58 (m, 0.26 H), 5.16–5.10 (m, 0.26 H), 3.82 (s, 3 H)

^{13}C NMR (100 MHz, $CDCl_3$, 23 ºC, δ): 159.5, 136.3, 136.2, 130.5, 127.5, 114.0, 55.4

N-Oxalylglycine disodium salt



Under air, glycine ethyl ester hydrochloride (1.00 g, 7.16 mmol, 1.00 equiv.) and triethylamine (2.50 mL, 17.9 mmol, 2.50 equiv.) were dissolved in dichloromethane (50 mL) in a 100 mL round-bottomed flask with magnetic stirring. Ethyl oxalyl chloride (875 µL, 7.88 mmol, 1.10 equiv.) was added dropwise over 3 minutes, and the reaction was stirred at room temperature for 15 hours. The reaction mixture was then poured into saturated aqueous sodium bicarbonate (50 mL) and vigorously mixed. The organic layer was washed with saturated aqueous sodium bicarbonate (50 mL), dried over sodium sulfate, and concentrated to a brown oil (1.29 g).

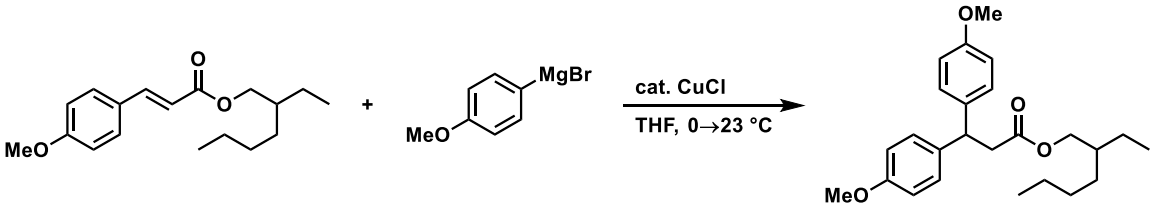
The brown oil was dissolved in methanol (50 mL) and cooled to 0 °C in an ice/water bath. 1 M aqueous sodium hydroxide (13.0 mL, 13.0 mmol, 2.05 equiv.) was added dropwise over 3 minutes. Stirring continued at 0 °C for 5 minutes, then at room temperature for 1 hour, at which point there was substantial precipitated solid. The reaction mixture was cooled to 0 °C to maximize precipitation, and the solid was removed by vacuum filtration. The filter cake was washed with cold methanol (5 mL) and diethyl ether (5 mL), to afford the title compound as a colorless solid (1.02 g, 74% from glycine ethyl ester).

^{1}H NMR (400 MHz, D_2O, 23 ºC, δ): 3.82 (s, 2 H)

^{13}C NMR (100 MHz, D_2O, 23 ºC, δ): 176.4, 166.0, 164.8, 43.2

Synthesis of authentic product standards

2-Ethylhexyl 3,3-bis(4-methoxyphenyl)propionate

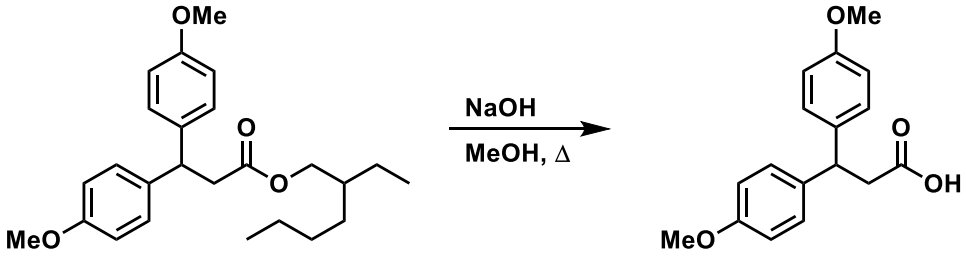


Under argon, a 100 mL round-bottomed flask with a magnetic stir bar was charged with copper(I) chloride (170 mg, 1.72 mmol, 0.10 equiv.) and 4-methoxyphenylmagnesium bromide (0.5 M solution in tetrahydrofuran, 44.8 mL, 22.4 mmol, 1.30 equiv.), then cooled to 0 °C in an ice-water bath. 2-Ethylhexyl 4-methoxycinnamate (4.95 mL, 17.2 mmol, 1.00 equiv.) was added dropwise over 15 minutes. When addition was complete, the reaction was warmed to room temperature, and stirring was continued for 1 hour. The reaction was cooled to 0 °C, then saturated aqueous ammonium chloride (25 mL) was added slowly to quench. The resulting mixture was partitioned between water and diethyl ether (100 mL each). The aqueous layer was extracted with additional diethyl ether (2×50 mL). The combined organic layers were washed with water (2×100 mL) and brine (100 mL), dried over magnesium sulfate, and concentrated *in vacuo*. The residue was purified by automated flash column chromatography on silica gel (100 g), eluting with a gradient 0 to 30% ethyl acetate/hexanes, to yield the title compound as a colorless oil (2.72 g, 40%).

NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, 23 ºC, δ): 7.17–7.11 (m, 4 H), 6.84–6.78 (m, 4 H), 4.45 (t, *J* = 8.1 Hz, 1 H), 3.89 (dd, *J* = 5.7, 1.9 Hz, 2 H), 3.76 (s, 6 H), 2.99 (d, *J* = 8.1 Hz, 2 H), 1.28–1.15 (m, 9 H), 0.88 (t, *J* = 7.0 Hz, 3H), 0.80 (t, *J* = 7.4 Hz, 3H)

3,3-bis(4-Methoxyphenyl)propionic acid

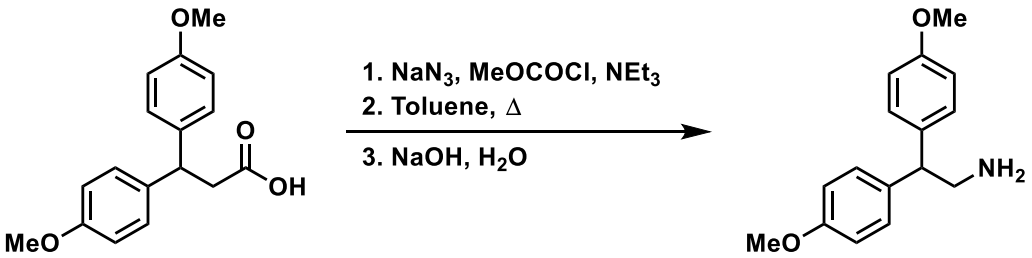


Under air, 2-ethylhexyl 3,3-bis(4-methoxyphenyl)propionate (2.72 g, 6.82 mmol, 1.00 equiv.) was dissolved in methanol (50 mL) in a 250 mL round-bottomed flask with magnetic stirring. 10% (w/v) aqueous sodium hydroxide (27.3 mL, 68.2 mmol, 10.0 equiv.) was added and the reaction was heated to reflux for 14 hours. The reaction was cooled to 0 °C, and 2.5 M hydrochloric acid (40 mL) was added. The resulting mixture was extracted with ether (3×50 mL), and the combined ether layers were washed with water (2×100 mL) and brine (100 mL), dried over magnesium sulfate, and concentrated *in vacuo*. The residue was purified by automated flash column chromatography on silica gel (100 g), eluting with a gradient 0 to 50% ethyl acetate/hexanes, to afford the title compound as a colorless solid (1.34 g, 69%).

NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, 23 ºC, δ): 7.15–7.10 (m, 4 H), 6.84–6.79 (m, 4 H), 4.43 (t, *J* = 7.9 Hz, 1H), 3.76 (s, 6 H), 3.02 (d, *J* = 8.0 Hz, 2H)

2,2-bis(4-Methoxyphenyl)ethylamine (3)



3

Under air, 3,3-bis(4-methoxyphenyl)propionic acid (1.34 g, 4.68 mmol, 1.00 equiv.) was dissolved in acetone (25 mL) in a 100 mL round-bottomed flask with magnetic stirring, and cooled to 0 °C in an ice/water bath. Triethylamine (750 µL, 5.38 mmol, 1.15 equiv.) and methyl chloroformate (471 µL, 6.08 mmol, 1.30 equiv.) were added, and the reaction was stirred for 20 minutes. Sodium azide (456 mg, 7.02 mmol, 1.50 equiv.), as a solution in water (2 mL) was added dropwise, and stirring was continued for 20 minutes. The reaction mixture was poured into ice-cold water (50 mL), and the resulting mixture was extracted with diethyl ether (3×50 mL). The combined ether layers were washed with brine (100 mL), dried over magnesium sulfate, and concentrated *in vacuo* behind a blast shield to a colorless oil.

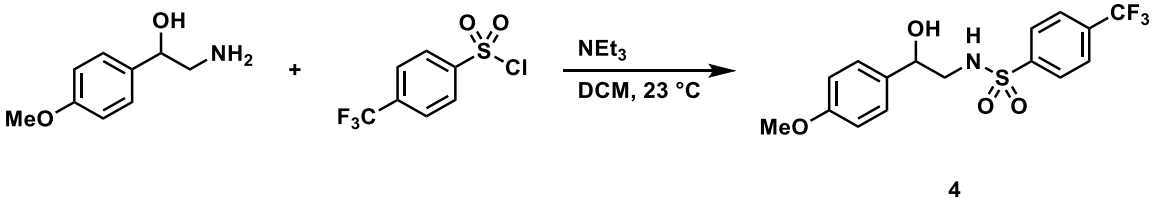
The crude acyl azide was dissolved in toluene (30 mL) in a 100 mL round-bottomed flask, and heated until gentle bubbling was observed (60 °C). Visible bubbling ceased after 1 hour, and heating was continued for an additional hour to ensure complete reaction. The reaction mixture was cooled to room temperature and concentrated *in vacuo* to a slightly brown oil. 10% (w/v) aqueous sodium hydroxide (25 mL) was added and the mixture was heated to 85 °C for 14 hours. The reaction mixture was cooled to room temperature and acidified with concentrated hydrochloric acid to pH <1, water was added to a total volume ~50 mL, and the mixture was heated to 95 °C. Undissolved solids were removed by gravity filtration of the hot solution. The filtrate was cooled to room temperature, then washed with diethyl ether (2×50 mL). The acidic aqueous solution was then basified with 8 M aqueous sodium hydroxide to pH >13 to precipitate the amine. The resulting suspension was extracted with diethyl ether (3×50 mL). The combined ether layers were washed with brine (50 mL), dried over sodium sulfate, and concentrated to a colorless crystalline solid (510 mg, 42% from the carboxylic acid).

NMR Spectroscopy:

^{1}H NMR (400 MHz, CD_3CN, 23 ºC, δ): 7.19–7.14 (m, 2 H), 6.87–6.82 (m, 2 H), 3.83 (t, *J* = 7.6 Hz, 1 H). 3.74 (s, 6 H), 3.14 (d, *J* = 7.7 Hz, 2 H)

^{13}C NMR (100 MHz, CD_3CN, 23 ºC, δ):159.1, 137.1, 129.8, 114.8, 55.8, 54.5, 48.0

2-(4-Trifluoromethylbenzenesulfonamido)-1-(4-methoxyphenyl)ethanol (4)



4

Under air, 2-amino-1-(4-methoxyphenyl)ethanol (33.0 mg, 197 μmol, 1.00 equiv.) and triethylamine (55.1 μL, 395 μmol, 2.00 equiv.) were dissolved in dichloromethane (1.5 mL) in a 4 mL vial with magnetic stirring. 4-Trifluoromethylbenzenesulfonylchloride (53.1 mg, 217 μmol, 1.10 equiv.) was dissolved in dichloromethane (1 mL), then added dropwise to the amino alcohol solution. After 10 minutes, the reaction mixture was diluted with dichloromethane (10 mL) and washed with water (10 mL). The organic layer was dried over sodium sulfate and concentrated *in vacuo*. The residue was purified by automated flash column chromatography on silica gel (10 g), eluting with a 0 to 50% ethyl acetate/hexanes gradient, to afford the title compound as a colorless solid (46.2 mg, 61%).

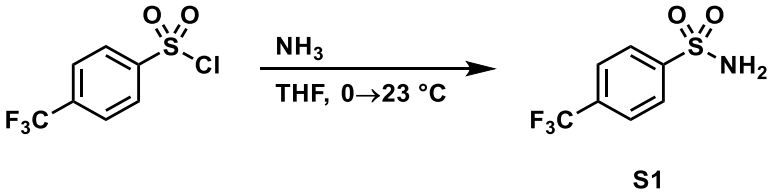
NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, 23 ºC, δ): 7.97–7.92 (m, 2 H), 7.78–7.73 (m, 2 H), 7.21–7.16 (m, 2 H), 6.87–6.82 (m, 2 H), 5.26 (dd, *J* = 8.0, 4.4 Hz, 1H), 4.77 (dd, *J* = 8.5, 3.8 Hz, 1H), 3.78 (s, 3 H), 3.26 (ddd, *J* = 13.0, 7.7, 3.7 Hz, 1H), 3.06 (ddd, *J* = 12.7, 8.5, 3.9 Hz, 1H), 2.47 (br. s, 1 H)

^{13}C NMR (100 MHz, $CDCl_3$, 23 ºC, δ): 159.8, 143.6 (q, J_{CF} = 1.4 Hz), 134.5 (q, J_{CF} = 33.1 Hz), 132.7, 127.7, 127.3, 126.4 (q, J_{CF} = 3.7 Hz), 123.3 (q, J_{CF} = 273 Hz), 114.3, 72.6, 55.4, 50.2

^{19}F NMR (272 MHz, $CDCl_3$, 23 ºC, δ): –63.1 (s)

4-Trifluoromethylbenzenesulfonamide (5)



Under air, 4-trifluoromethylbenzenesulfonyl chloride (489 mg, 2.00 mmol, 1.00 equiv.) was dissolved in tetrahydrofuran (10 mL) in a 20 mL scintillation vial with magnetic stirring, and cooled to 0 °C. 30% (w/v) aqueous ammonia (1.14 mL, 20.0 mmol, 10.0 equiv.) was added dropwise, then the reaction mixture was warmed to room temperature and stirring continued for 5 minutes. The reaction mixture was partitioned between water and ethyl acetate (20 mL each). The aqueous layer was extracted with additional ethyl acetate (10 mL). The combined organic layers were washed with saturated aqueous sodium bicarbonate (25 mL) and brine (25 mL), then dried over sodium sulfate and concentrated *in vacuo* to afford the title compound as a colorless solid (449 mg, 99%).

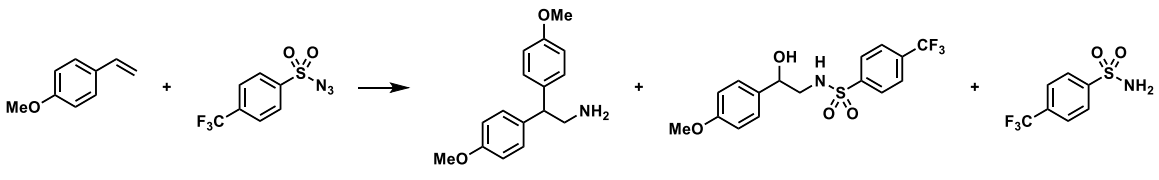
NMR Spectroscopy:

^{1}H NMR (400 MHz, CD_3CN, 23 ºC, δ): 8.07–8.02 (m, 2 H), 7.90–7.85 (m, 2 H), 5.84 (br. s, 2 H)

^{13}C NMR (100 MHz, CD_3CN, 23 ºC, δ): 147.9, 134.1 (q, J_{CF} = 32.5 Hz), 127.9, 127.3 (q, J_{CF} = 3.9 Hz), 124.6 (q, J_{CF} = 272 Hz)

^{19}F NMR (272 MHz, CD_3CN, 23 °C, δ): –63.6 (s)

HPLC analytical method and calibration curves



Samples for HPLC calibration curves were prepared as simulated reaction samples. To MOPS buffer (20 mM pH 7.0, 380 μL) was added a solution of the appropriate reaction product in acetonitrile (0–200 μM, 20 μL, final concentration 0–10 mM). To this sample was added the internal standard 4-hydroxybenzoic acid (0.5% v/v in acetonitrile, 50 μL) and acetonitrile (350 μL). The product concentration in the curves below corresponds to the concentration in the reaction mixture; the final analytical sample is two-fold diluted.

Analysis was performed on an Agilent 1260 Series Infinity II HPLC connected to an iQ MSD with water/acetonitrile mobile phase containing 0.1% acetic acid (1 mL min^{-1} flow), with an Agilent Poroshell 120 EC-C18 column (4 μm packing, 2.1×50 mm) fitted with a Poroshell 120 guard column (1.7 μm packing, 2.1×5 mm), injecting 1 μL. Detection was at 230 nm (4 nm bandwidth). The gradient program and retention times are given in Table 4-5 and Table 4-6, respectively, and an example chromatogram is shown in Figure 4-14.

Table 4-5. HPLC gradient program.

Time (minutes)	% Acetonitrile
0	5
0.5	5
4	95
4.5	95
4.51	5
5	5

Table 4-6. HPLC retention times.

Compound	Retention time (minutes)
4-Hydroxybenzoic acid	0.98
2,2-bis(4-Methoxyphenyl)ethylamine	2.45
4-Trifluoromethylbenzenesulfonamide	2.69
2-(4-Trifluoromethylbenzenesulfonamido)-1-(4-methoxyphenyl)ethanol	3.34
4-Methoxystyrene	3.44
4-Trifluoromethylbenzenesulfonyl azide	3.55

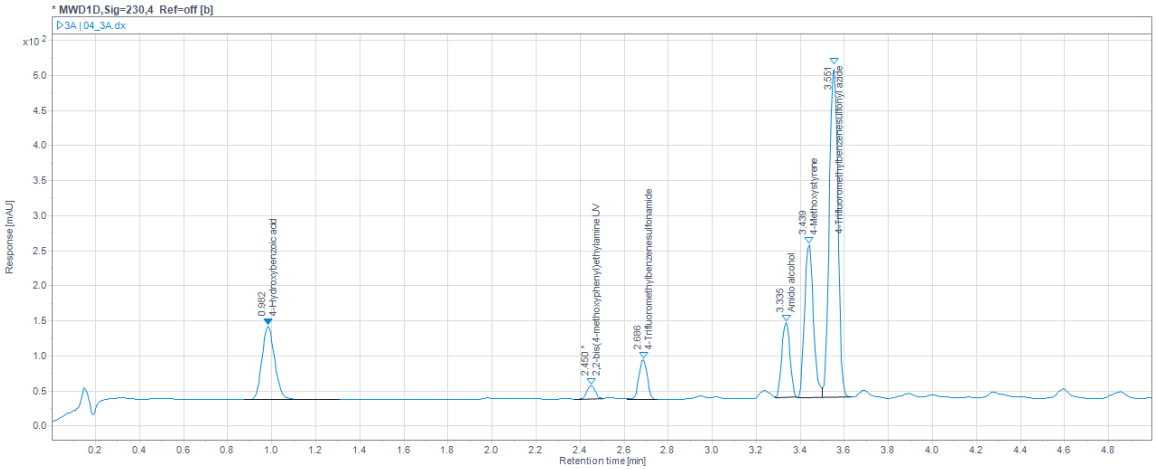
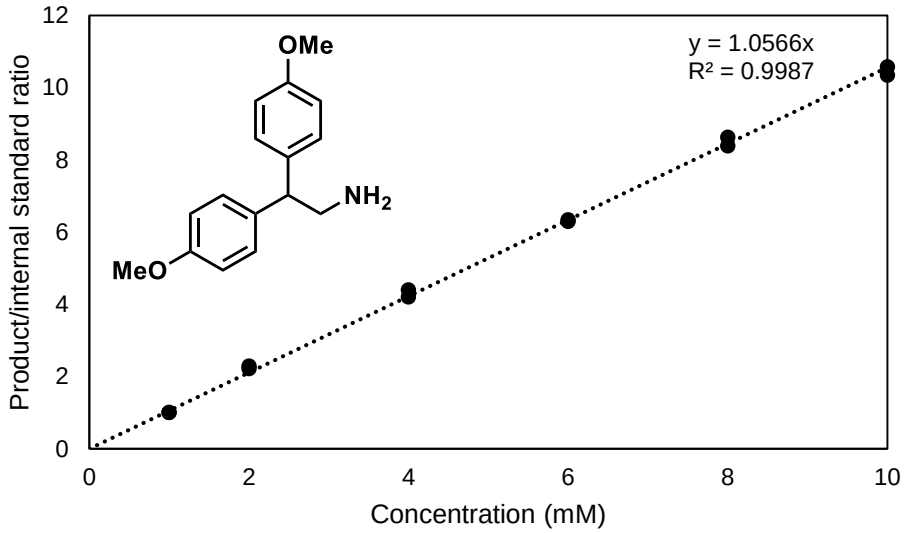
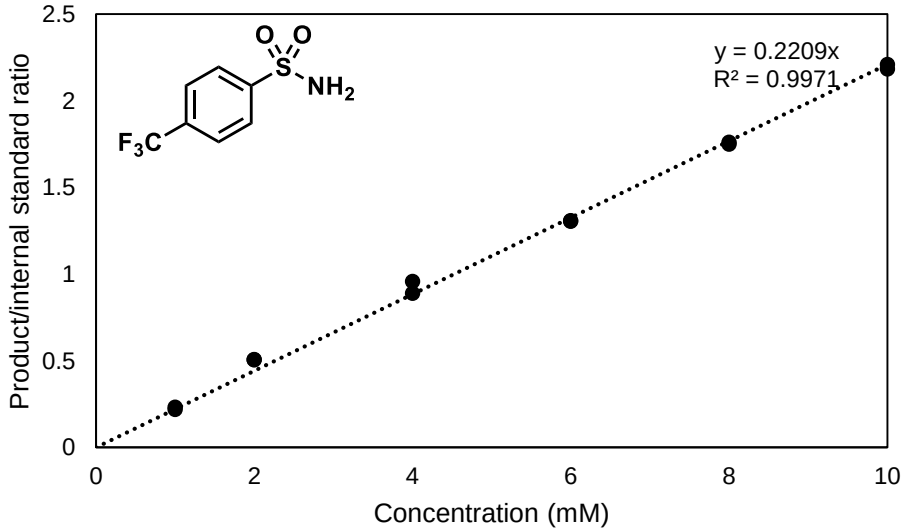


Figure 4-14. Representative HPLC-UV chromatogram of an enzymatic reaction sample.

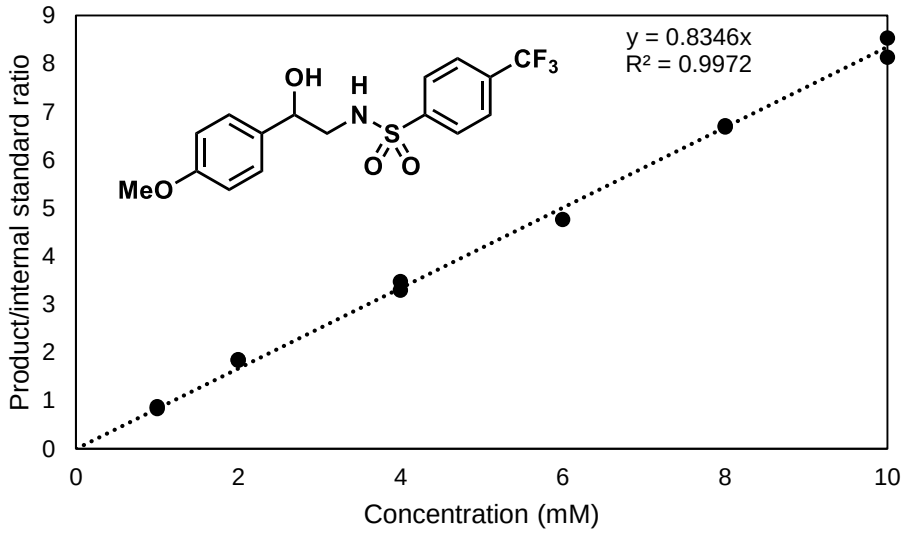
2,2-bis(4-Methoxyphenyl)ethylamine calibration curve



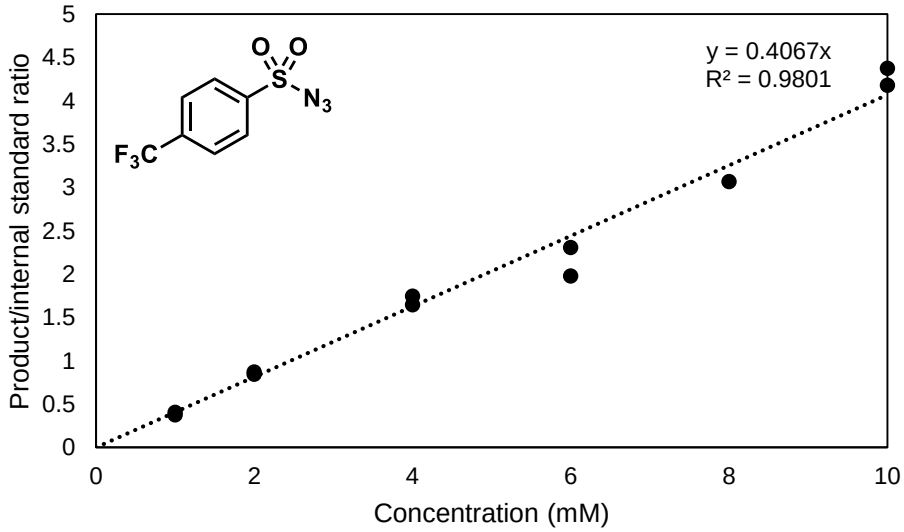
4-Trifluoromethylbenzenesulfonamide calibration curve



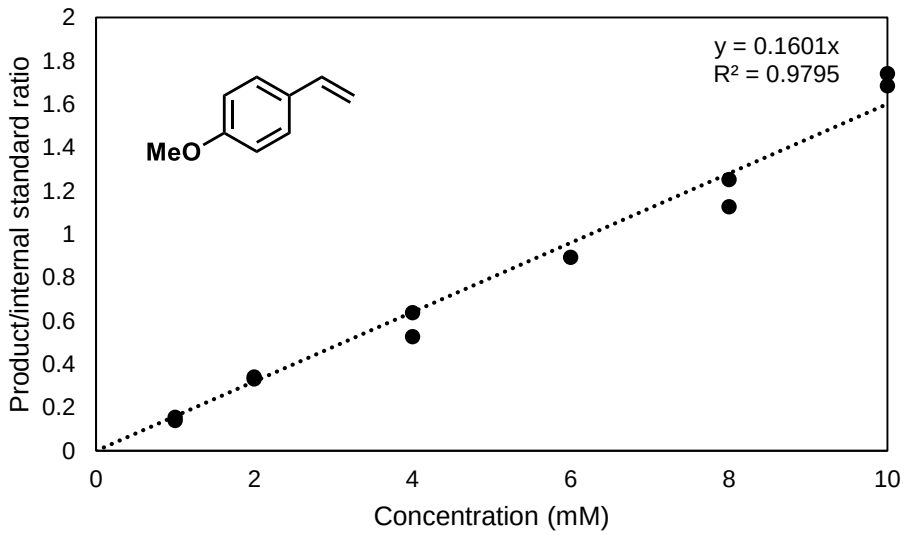
2-(4-Trifluoromethylbenzenesulfonamido)-1-(4-methoxyphenyl)ethanol calibration curve



4-Trifluoromethylbenzenesulfonyl azide calibration curve



4-Methoxystyrene calibration curve



4.7 References for Chapter 4

[1] Degennaro, L.; Trinchera, P.; Luisi, R. Recent Advances in the Stereoselective Synthesis of Aziridines. *Chem. Rev.* **2014**, *114*, 7881–7929.

[2] Ju, M.; Schomaker, J. M. Nitrene transfer catalysts for enantioselective C–N bond formation. *Nat. Rev. Chem.* **2021**, *5*, 580–594.

[3] Zhang, T.; Hu, X.; Dong, X.; Li, G.; Lu, H. Iridium-Catalyzed Unreactive C(sp^3)–H Amination with 2,2,2-Trichloroethoxycarbonyl Azide. *Org. Lett.* **2018**, *20*, 6260–6264.

[4] Park, S. H.; Kwak, J.; Shin, K.; Ryu, J.; Park, Y.; Chang, S. Mechanistic Studies of the Rhodium-Catalyzed Direct C–H Amination Reaction Using Azides as the Nitrogen Source. *J. Am. Chem. Soc.* **2014**, *136*, 2492–2502.

[5] Omura, K.; Murakami, M.; Uchida, T.; Irie, R.; Katsuki, T. Enantioselective Aziridination and Amination Using *p*-Toluenesulfonyl Azide in the Presence of Ru(salen)(CO) Complex. *Chem. Lett*. **2003**, *32*, 354–355.

[6] Li, Z.; Quan, R. W.; Jacobsen, E. N. Mechanism of the (Diimine)copper-Catalyzed Asymmetric Aziridination of Alkenes. Nitrene Transfer via Ligand-Accelerated Catalysis. *J. Am. Chem. Soc*. **1995**, *117*, 5889–5890.

[7] Subbarayan, V.; Ruppel, J. V.; Zhu, S.; Perman, J. A.; Zhang, X. P. Highly asymmetric cobalt-catalyzed aziridination of alkenes with trichloroethoxysulfonyl azide ($TcesN_3$). *Chem. Commun*. **2009**, 4266–4268.

[8] McIntosh, J. A.; Coelho, P. S.; Farwell, C. C.; Wang, Z. J.; Lewis, J. C.; Brown, T. R.; Arnold. F. H. Enantioselective Intramolecular C =–H Amination Catalyzed by Engineered Cytochrome P450 Enzymes *In Vitro* and *In Vivo*. *Angew. Chem. Int. Ed.* **2013**, *52*, 9309–9312.

[9] Farwell, C. C.; Zhang, R. K.; McIntosh, J. A.; Hyster, T. K.; Arnold, F. H. Enantioselective Enzyme-Catalyzed Aziridination Enabled by Active-Site Evolution of a Cytochrome P450. *ACS Cent. Sci.* **2015**, *1*, 89–93.

[10] Prier, C. K.; Zhang, R. K.; Buller, A. R.; Brinkmann-Chen, S.; Arnold, F. H. Enantioselective, Intermolecular Benzylic C–H Amination Catalysed by an Engineered Iron-Haem Enzyme. *Nature Chem.* **2017**, *9*, 629–634.

[11] Goldberg, N. W.; Knight, A. M.; Zhang, R. K. Nitrene Transfer Catalyzed by a Non-Heme Iron Enzyme and Enhanced by Non-Native Small-Molecule Ligands. *J. Am. Chem. Soc.* **2019**, *141*, 19585–19588.

[12] Vila, M. A.; Steck, V.; Giordano, S. R.; Carrera, I.; Fasan, R. C–H Amination *via* Nitrene Transfer Catalyzed by Mononuclear Non-Heme Iron-Dependent Enzymes. *ChemBioChem* **2020**, *21*, 1981–1987.

[13] Monos, T. M.; McAtee, R. C.; Stephenson, C. R. J. Arylsulfonylacetamides as bifunctional reagents for alkene aminoarylation. *Science* **2018**, *361*, 1369–1373.

[14] Knight, A. M. Expanding the Scope of Metalloprotein Families and Substrate Classes in New-to-Nature Reactions. Ph.D. Dissertation, California Institute of Technology, 2020.

[15] Vanhercke, T.; Ampe, C.; Tirry, L.; Denolf, P. Reducing mutational bias in random protein libraries. *Anal. Biochem.* **2005**, *339*, 9–14.

[16] Wong. T. S.; Tee, K. L.; Hauer, B.; Schwaneberg, U. Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. *Nucleic Acids Res.* **2004**, *32*, e26.

[17] Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A. III; Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **2009**, *6*, 343–345.

[18] Kille, S.; Acevedo-Rocha, C. G.; Parra, L. P.; Zhang, Z.-G..; Opperman, D. J.; Reetz, M. T.; Acevedo, J. P. Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synth. Biol*. **2013**, *2*, 83–92.

[19] Zhao, H.; Giver, L.; Shao, Z.; Affholter, J. A.; Arnold, F. H. Molecular Evolution by Staggered Extension Process (StEP) *in vitro* Recombination. *Nat. Biotechnol.* **1998**, *16*, 258–261.

[20] Molander, G. A.; Brown, A. R. Suzuki−Miyaura Cross-Coupling Reactions of Potassium Vinyltrifluoroborate with Aryl and Heteroaryl Electrophiles. *J. Org. Chem.* **2006**, *71*, 9681–9686.

[21] Raushel, J.; Pitram, S. M.; Folkin, V. V. Efficient Synthesis of Sulfonyl Azides from Sulfonamides. *Org. Lett.* **2008**, *10*, 3385–3388.